



This is a repository copy of *EQ5DMAP: a command for mapping between EQ-5D-3L and EQ-5D-5L*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/124872/>

Version: Published Version

Article:

Hernandez, M. orcid.org/0000-0003-4474-5883 and Pudney, S. (2018) EQ5DMAP: a command for mapping between EQ-5D-3L and EQ-5D-5L. *The Stata journal*, 18 (2). pp. 395-415. ISSN 1536-867X

Notwithstanding the Stata Journal copyright notice, StataCorp and the Stata Journal hereby agree that an electronic copy (PDF) of the following article can be made available by open access: Hernandez-Alava, M., and S. Pudney. 2018 "eq5dmap: A command for mapping between EQ-5D-3L and EQ-5D-5L". *Stata Journal* 18: 395-415. The Stata Journal website: <https://www.stata-journal.com>

Reuse

See Attached

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

THE STATA JOURNAL

Editors

H. JOSEPH NEWTON
Department of Statistics
Texas A&M University
College Station, Texas
editors@stata-journal.com

NICHOLAS J. COX
Department of Geography
Durham University
Durham, UK
editors@stata-journal.com

Associate Editors

CHRISTOPHER F. BAUM, Boston College
NATHANIEL BECK, New York University
RINO BELLOCCO, Karolinska Institutet, Sweden, and
University of Milano-Bicocca, Italy
MAARTEN L. BUIS, University of Konstanz, Germany
A. COLIN CAMERON, University of California–Davis
MARIO A. CLEVES, University of Arkansas for
Medical Sciences
MICHAEL CROWTHER, University of Leicester, UK
WILLIAM D. DUPONT, Vanderbilt University
PHILIP ENDER, University of California–Los Angeles
JAMES HARDIN, University of South Carolina
BEN JANN, University of Bern, Switzerland
STEPHEN JENKINS, London School of Economics and
Political Science
ULRICH KOHLER, University of Potsdam, Germany

FRAUKE KREUTER, Univ. of Maryland–College Park
PETER A. LACHENBRUCH, Oregon State University
STANLEY LEMESHOW, Ohio State University
J. SCOTT LONG, Indiana University
ROGER NEWSON, Imperial College, London
AUSTIN NICHOLS, Abt Associates, Washington, DC
MARCELLO PAGANO, Harvard School of Public Health
SOPHIA RABE-HESKETH, Univ. of California–Berkeley
J. PATRICK ROYSTON, MRC CTU at UCL, London, UK
MARK E. SCHAFFER, Heriot-Watt Univ., Edinburgh
PHILIPPE VAN KERM, LISER, Luxembourg
VINCENTO VERARDI, Université Libre de Bruxelles,
Belgium
IAN WHITE, MRC CTU at UCL, London, UK
RICHARD A. WILLIAMS, University of Notre Dame
JEFFREY WOOLDRIDGE, Michigan State University

Stata Press Editorial Manager

LISA GILMORE

Stata Press Copy Editors

ADAM CRAWLEY, DAVID CULWELL, and DEIRDRE SKAGGS

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*), *Scopus*, and *Social Sciences Citation Index*.

For more information on the *Stata Journal*, including information for authors, see the webpage

<http://www.stata-journal.com>

Subscriptions are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-782-8272, fax 979-696-4601, or online at

<http://www.stata.com/bookstore/sj.html>

Subscription rates listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
Printed & electronic		Printed & electronic	
1-year subscription	\$124	1-year subscription	\$154
2-year subscription	\$224	2-year subscription	\$284
3-year subscription	\$310	3-year subscription	\$400
1-year student subscription	\$ 89	1-year student subscription	\$119
1-year institutional subscription	\$375	1-year institutional subscription	\$405
2-year institutional subscription	\$679	2-year institutional subscription	\$739
3-year institutional subscription	\$935	3-year institutional subscription	\$1,025
Electronic only		Electronic only	
1-year subscription	\$ 89	1-year subscription	\$ 89
2-year subscription	\$162	2-year subscription	\$162
3-year subscription	\$229	3-year subscription	\$229
1-year student subscription	\$ 62	1-year student subscription	\$ 62

Back issues of the *Stata Journal* may be ordered online at

<http://www.stata.com/bookstore/sjj.html>

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

<http://www.stata-journal.com/archives.html>

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to sj@stata.com.



Copyright © 2018 by StataCorp LLC

Copyright Statement: The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LLC. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, **STATA**, Stata Press, Mata, **MATA**, and NetCourse are registered trademarks of StataCorp LLC.

eq5dmap: A command for mapping between EQ-5D-3L and EQ-5D-5L

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. In this article, we describe a new command, `eq5dmap`, for conditional prediction of the utility values of EQ-5D-5L (EQ-5D-3L) from observed or specified values of EQ-5D-3L (EQ-5D-5L) conditional on age and gender. Predictions can be made either from the five-item health descriptions 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 that have been fit to alternative reference datasets (the National Data Bank for Rheumatic Diseases and a EuroQol Group coordinated data-collection study). 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 that includes an investigation of the sensitivity of the mapping outcomes to the choice of reference dataset.

Keywords: `st0528`, `eq5dmap`, EQ-5D, EQ-5D-3L, EQ-5D-5L, 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 programs designed to improve health. The QALY allows healthcare 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 in its appraisals of health interventions in England (NICE 2014). Preference-based measures such as the EQ-5D-3L underpin the calculation of QALYs.

EQ-5D describes health states in terms of five dimensions: mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. The original EQ-5D, now called 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 or discomfort dimensions, and extreme problems in the anxiety or depression dimension. Valuation studies in different countries assigned an index or utility score to each of the

health states described in this way. Dolan (1997) published the first UK value set using general public preferences. Other countries have developed their own value sets, 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 or ceiling distortions of the EQ-5D-3L. The number of dimensions has remained unchanged, but the new EQ-5D-5L 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.¹ The number of discrete health states described by the new version is 3,125. Utility value sets for EQ-5D-5L have been released for England (Devlin et al. 2018),² Japan (Shiroiwa et al. 2016), Canada (Xie et al. 2016), Uruguay (Augustovski et al. 2016), Netherlands (Versteegh et al. 2016), Korea (Kim et al. 2016), China (Luo et al. 2017), and Indonesia (Purba et al. 2017), and similar work is underway in other countries.

Many studies now include EQ-5D-5L instead of the original EQ-5D-3L. We have shown in previous articles (Hernández-Alava and Pudney 2017; Hernández-Alava et al. 2018) that in the UK, EQ-5D-3L and EQ-5D-5L lead to different utility scores for the same underlying level of health; this has profound implications for economic evaluations. It is, therefore, inappropriate to mix the evidence collected using both instruments without adjusting for these differences. Because 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 (2017) developed a flexible model that 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 refit using two different datasets and results reported elsewhere (Hernández-Alava et al. 2018). In this article, we describe a new command, `eq5dmap`, for conditional prediction of the EQ-5D-5L (EQ-5D-3L) utility from observed EQ-5D-3L (EQ-5D-5L) responses or specified utility values and age and gender. The command predictions are based on the models in Hernández-Alava et al. (2018). This command is the only available method of translating evidence from EQ-5D-3L to EQ-5D-5L and vice versa using the individual health states, an individual utility value, or an approximate utility score. 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. Section 5 presents an illustrative example of the use of the command. Section 6 concludes.

1. See the EuroQol website <https://euroqol.org/eq-5d-instruments/> for examples of the question wording used in EQ-5D-3L and EQ-5D-5L.

2. There are some differences between the value set reported in the *Health Economics* article and the previous version reported in the Office of Health Economics research paper (Devlin et al. 2016).

2 The mapping method

The `eq5dmap` command allows mapping both from the older to the newer format (3L → 5L) and the reverse (5L → 3L). We explain the mapping methodology for the case of 3L → 5L, but the procedure is essentially the same for 5L → 3L.

Let $Y_{3d} \in \{1, 2, 3\}$ and $Y_{5d} \in \{1, 2, 3, 4, 5\}$ represent outcomes for the d th domain ($d = 1, \dots, 5$) of EQ-5D-3L and EQ-5D-5L, respectively. Define the vectors $\mathbf{Y}_3 = (Y_{31}, \dots, Y_{35})$ and $\mathbf{Y}_5 = (Y_{51}, \dots, Y_{55})$, and write the corresponding utility scoring scales $v_3(\cdot)$ and $v_5(\cdot)$. Our aim is to calculate the expectation of $v_5(\mathbf{Y}_5)$ conditional on the values of a vector of covariates \mathbf{X} and also on available information about \mathbf{Y}_3 . Depending on the form of that information, two types of mapping can be done.

2.1 A specified value for \mathbf{Y}_3

Let $S_5 = \{1, \dots, 5\}^5$ be the set of possible values that can be taken by the vector \mathbf{Y}_5 . If we know the conditioning value of \mathbf{Y}_3 , the expectation of v_5 can be computed as

$$E(v_5 | \mathbf{Y}_3 = y_3, \mathbf{X}) = \sum_{y_5 \in S_5} v_5(y_5) p(y_5 | y_3, \mathbf{X})$$

where $p(y_5 | y_3, \mathbf{X})$ is the form of conditional probability implied by the specified underlying statistical model for the joint distribution $\mathbf{Y}_3 | \mathbf{X}$ and $\mathbf{Y}_5 | \mathbf{X}$.

2.2 A specified (approximate) value for $v_3(\mathbf{Y}_3)$

In some cases, the user may know only the value of $v_3(\mathbf{Y}_3)$, rather than \mathbf{Y}_3 itself. Because the mapping $\mathbf{Y}_3 \rightarrow v_3$ 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 EQ-5D-3L utility score. We handle both problems by distance-weighted averaging within a neighborhood of the specified value, \hat{v}_3 .

Let C be a user-specified bandwidth. Define a set of vectors

$$\mathcal{S}(v) = \{\mathbf{Y}_3 : |v_3(\mathbf{Y}_3) - v| \leq C\}$$

and a weight function of Epanechnikov form:

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

The estimate of the expected value of v_5 is

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

where the summation over y_5 covers all 3,125 possible outcome vectors for the EQ-5D-5L descriptive system.

3 The underlying statistical model

The predictive distribution $\Pr(\mathbf{Y}_5|\mathbf{Y}_3, \mathbf{X})$ is derived from a model of the joint distribution of $\mathbf{Y}_3|\mathbf{X}$ and $\mathbf{Y}_5|\mathbf{X}$ developed by Hernández-Alava and Pudney (2017). That model is a system of 10 latent regressions, arranged in 5 groups, following the natural pairing of the dimensions in the 2 versions of EQ-5D, with domain d containing the equations for Y_{3d} and Y_{5d} ,

$$\left. \begin{aligned} Y_{3d}^* &= \mathbf{X}\boldsymbol{\beta}_{3d} + U_{3d} \\ Y_{5d}^* &= \mathbf{X}\boldsymbol{\beta}_{5d} + U_{5d} \end{aligned} \right\} \quad d = 1, \dots, 5$$

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

$$Y_{kd} = q \quad \text{iff} \quad \Gamma_{kqd} \leq Y_{kd}^* < \Gamma_{k(q+1)d}, \quad q = 1, \dots, Q_k, \quad \text{and} \quad k = 3, 5$$

where $Q_k = 3$ or 5 is the number of categories of Y_{kd} and the Γ_{kqd} are threshold parameters, with $\Gamma_{k1d} = -\infty$ and $\Gamma_{k(Q_k+1)d} = +\infty$.

To allow for background correlation between the five dimensions of EQ-5D, we decompose the residual U_{kid} 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 ε_{kid} correlated within but not between dimensions,

$$U_{kid} = \psi_{kd}V_i + \varepsilon_{kid} \quad k = 3, 5 \quad \text{and} \quad d = 1, \dots, 5$$

where the ψ_{kd} are a set of 10 parameters.

The model can be fit under various alternative assumptions about the joint distribution of the residuals ε_{3id} and ε_{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 ε_{3id} and ε_{5id} have bivariate normal distributions and $V_i \sim N(0, 1)$; and the copula specification, where the distribution of each pair ε_{3id} and ε_{5id} is specified in copula form with normal mixture marginals and V_i as a normal mixture (see Hernández-Alava and Pudney [2017] for details).

These two variants of the mapping model have been fit using two alternative datasets. To our knowledge, these are the only datasets large enough to undertake this type

of study. We present a concise summary of both datasets below. A more detailed discussion of the similarities and differences between the two datasets can be found in Pennington et al. (Forthcoming). Users of `eq5dmap` should consider carefully which reference dataset is better suited for their study.

The first dataset comes from the National Data Bank for Rheumatic Diseases (NDB), which is a register of patients with rheumatic disease, mainly referred by U.S. and Canadian rheumatologists (Wolfe and Michaud 2011). In 2011, there was a switch from the three-level to the five-level version of EQ-5D, and both versions were collected in parallel during the January 2011 wave, which we used to fit the reference model. 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 six 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 many 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 with 81% for NDB, in line with what is expected in a rheumatoid arthritis 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 0.703³ for EQ-5D-3L and EQ-5D-5L, respectively, versus 0.681 and 0.766⁴ in the NDB dataset. Fitted reference models for these datasets are described in Hernández-Alava and Pudney (2017) and Hernández-Alava et al. (2018).

4 The eq5dmap command

4.1 Syntax

```
eq5dmap outputvarname [if] [in] [weight], covariates(varlist)
      {items(varlist) | score(varname)} [model(modelname)
      direction(mappingdirection) values3(3Lvaluesetname)
      values5(5Lvaluesetname) bwidth(#)]
```

4.2 Description

`eq5dmap` is a community-contributed command that allows outcomes measured using EQ-5D-3L to be converted into (expected) utility values measured using the newer EQ-5D-5L or vice versa.

3. 0.712 using the valuation in Devlin et al. (2016).

4. 0.779 using the valuation in Devlin et al. (2016).

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

4.3 Output

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

4.4 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 in the interval [16, 100] and gender (coded as female = 0, male = 1). `covariates()` is required.

If the predictor is the five-dimensional EQ-5D health description:

`items(varlist)` specifies the variables that contain observed values for the five EQ-5D domain items to map from. They must be specified as a *varlist* containing five variables ordered as mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. For 3L → 5L mapping, the variables should all be coded on a scale 1, 2, 3, where 1 = no problems, . . . , 3 = extreme problems; for 5L → 3L mapping, the coding must be 1 = no problems, . . . , 5 = extreme problems. Either `items()` or `score()` is required, but not both.

If the predictor is a utility score rather than the health description:

`score(varname)` specifies a variable that contains the value of the utility score. Either `items()` or `score()` is required, but not both.

`model(modelname)` specifies the model to be used for the mapping. The available options for *modelname* are `NDBgauss`, `NDBcopula`, `EQGcopula`, or `EQGgauss`. The default is `model(EQGcopula)`.

`direction(mappingdirection)` specifies the direction of mapping: `direction(3L->5L)` specifies mapping from EQ-5D-3L to the newer EQ-5D-5L, while `direction(5L->3L)` specifies the reverse. The default is `direction(3->5)`.

`values3(3Lvaluesetname)` specifies one of the alternative EQ-5D-3L value sets offered. Currently, the only one offered is `UK`, which specifies the value set described by Dolan (1997). The default is `values3(UK)`. This option is used only when EQ-5D-3L is the target outcome or when mapping from an EQ-5D-3L utility score.

`values5(5Lvaluesetname)` specifies one of the alternative EQ-5D-5L value sets offered.

Currently, the only one offered is UK, which specifies the value set described by Devlin et al. (2018). The default is `values5(UK)`. This option is used only when EQ-5D-5L is the target outcome or when mapping from an EQ-5D-5L utility score.

`width(#)` is the bandwidth that controls the matching of the specified utility score to neighboring values. The default is `width(0)`, which enforces exact matching to a point on the chosen value set; if there is a multiplicity of points that match exactly, then their average is returned as the value for `outputvarname`. If there is no exact match within the neighborhood defined by the bandwidth, then `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 `eq5dmap` for $Y_3 \rightarrow v_5$ mapping and (exact) $v_3 \rightarrow v_5$ mapping. Section 5.3 gives recommendations on the choice of bandwidth for approximate $v_3 \rightarrow v_5$ mapping, and section 5.4 considers the potential sensitivity of results to the choice of reference dataset and model specification.

5.1 Mapping from EQ-5D-3L items to a EQ-5D-5L utility score

`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 $N = 243 \times 2 \times 85 = 41310$ records). The dataset includes a set of five ordinal variables Y3_1 to Y3_5 corresponding to the five EQ-5D-3L dimensions.⁵ The data are summarized below.

```
. // Load and examine the input dataset
. use eq5dmap_data
. summarize
```

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

5. The dataset also contains the corresponding utility value, `u3`, calculated using the UK value set; `age`; a dummy variable for gender; and two weight variables, `fwEQG` and `fwNDB`. We have also included an additional variable, `u3hat`, that contains an approximate EQ-5D-3L utility value such as those typically obtained from published articles or a previous mapping. These additional variables are used in later examples.

Our example uses the UK value sets reported in Devlin et al. (2018) and Dolan (1997) for the EQ-5D-5L and EQ-5D-3L value sets, respectively. We use the NDB reference dataset and the copula specification and predict the five-level utility score from the vector of EQ-5D-3L descriptive items, that is, a $\mathbf{Y}_3 \rightarrow v_5$ mapping. Note that it is not necessary to include the `utility5` option in the command, because the Devlin et al. (2018) UK score is the default (and currently only) choice.

```
. eq5dmap v5_y3, covariates(age male) model(NDBcopula)
> items(Y3_1 Y3_2 Y3_3 Y3_4 Y3_5)
No direction specified: default is 3->5
No 5L value set specified: default is UK
Summary of inputs to eq5dmap:
The 5-level value set 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
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v5_y3	41,310	.4755077	.262427	-.2247076	.9600279

The striking feature of the predicted utility scores is that they do not cover the full range $[-0.285, 1.000]$ of the Devlin et al. (2018) value set. Instead, they vary between -0.225 and 0.960 . This loss of dispersion is an inevitable feature of any minimum mean-squared error prediction based on the conditional expectation because the purely random component of $v_5(\mathbf{Y}_5)$ 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, because the conditional mean of QALYs is not affected by the loss of dispersion.⁶ However, loss of dispersion does become a problem when confidence intervals are to be computed or the distribution, rather than the mean or aggregate, of net benefit is required. See Hernández-Alava and Pudney (2017) for details of the full predictive distribution that can be used to handle such cases.

5.2 Mapping from an exact EQ-5D-3L score to an EQ-5D-5L score

The second type of mapping generates the predicted five-level utility score from a specified utility value (a $v_3 \rightarrow v_5$ mapping). In this example, we choose the precise utility score `u3` corresponding to the actual health state description in the dataset and a bandwidth of 0.001 , which gives exact mapping of utility scores.

6. Provided calculation of QALYs involves only linear operations, as is typically the case.

```
. eq5dmap v5_u3, covariates(age male) model(NDBcopula) score(u3) bwidth(0.001)
No direction specified: default is 3->5
No 5L value set specified: default is UK
No 3L value set specified: default is UK
Summary of inputs to eq5dmap:
The 5-level value set 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 value set is: UK
The 3-level score is contained in input variable: u3
The bandwidth is: .001
Unweighted mean of predicted 5L score within selected sample
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v5_u3	41,310	.4755077	.2607798	-.2247076	.9600279

The means of the two mapped variables, $v5_y3$ and $v5_u3$, are the same up to the seventh decimal place, but not their standard deviations (SDs). 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 $\mathbf{Y}_3 = (3, 3, 3, 3, 3)$ 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 $\mathbf{Y}_3 = (1, 2, 2, 1, 1)$ and $\mathbf{Y}_3 = (2, 1, 1, 1, 2)$ 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. Averaging across these equal-valued states reduces the SD slightly.

5.3 Mapping from an approximate utility score: choice of bandwidth

One can use `eq5dmap` to carry out a $\hat{v}_3 \rightarrow v_5$ mapping in situations where \hat{v}_3 is not a valid point in the EQ-5D-3L value set, but rather an inexact average or predicted utility score. Mapping is done using distance-weighted averaging of scale points within a neighborhood of \hat{v}_3 . The bandwidth parameter simultaneously defines the neighborhood over which averaging is to be done and the rate at which the weight declines with increasing distance.

There are potential pitfalls in this type of mapping, and it is important to consider two issues: What is the local character of the distribution of utilities that (implicitly) underlies the specified value \hat{v}_3 ? And does the mapping function $\hat{v}_3 \rightarrow v_5$ have the desirable property of monotonicity?

Given the coarse structure of the EQ-5D-3L and the nature of the UK value set reported in Dolan (1997), the EQ-5D-3L utility distributions in most trial datasets are very irregular, with large intervals of zero or nonzero probability and multiple modes. Consequently, the local averaging procedure may work well in some regions and not in others, and results can be very sensitive to the choice of bandwidth; indeed, with a small bandwidth, if \hat{v}_3 lies in one of the zero-probability intervals, there may be no scale

points near enough to average. Because we do not know the shape of the distribution implicitly underlying \widehat{v}_3 , we cannot implement an adaptive bandwidth procedure as is commonly used in nonparametric density estimation.

To illustrate this, we now repeat the mapping using the variable `u3hat`, which contains approximate utility values that do not exactly match any values found in the “official” UK EQ-5D-3L value set.⁷ We illustrate the effect of making alternative choices for the bandwidth using the mixed copula model and the NDB reference dataset. The following code is used:

```
. // Predicted UK utility scores:
. // loop over different bandwidths: 0.01, 0.03, 0.05, 0.1, and 0.2
. foreach c of numlist 1 3 5 10 20 {
2.   local bw = `c'/100
3.   display "Bandwidth = `bw', Specification = copula, Data = NDB"
4.   display "Predicted UK scores..."
5.   eq5dmap v5_u3_`c', covariates(age male)
> model(NDBcopula) score(u3hat) bwidth(`bw')
6. // Correlations with mapped actual score:
. correlate v5_u3_`c' v5_u3
7. }
```

The output from the first two passes of the loop is reproduced below; the full results are summarized in the first column of table 1, which also gives results for other model choices.

7. 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 EQ-5D-3L descriptive items. This model has been used here for illustration only; it has been shown not to be adequate for modeling EQ-5D-3L data (Hernández Alava, Wailoo, and Ara 2012).

Bandwidth = .01, Specification = copula, Data = NDB
 Predicted UK scores...
 No direction specified: default is 3->5
 No 5L value set specified: default is UK
 No 3L value set specified: default is UK
 Summary of inputs to eq5dmap:
 The 5-level value set 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 value set is: UK
 The 3-level score is contained in input variable: u3hat
 The bandwidth is: .01
 Unweighted mean of predicted 5L score within selected sample

Variable	Obs	Mean	Std. Dev.	Min	Max
v5_u3_1	38,590	.4973598	.2307445	-.2247076	.8946413

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

(obs=38,590)

	v5_u3_1	v5_u3
v5_u3_1	1.0000	
v5_u3	0.8458	1.0000

Bandwidth = .03, Specification = copula, Data = NDB
 Predicted UK scores...
 No direction specified: default is 3->5
 No 5L value set specified: default is UK
 No 3L value set specified: default is UK
 Summary of inputs to eq5dmap:
 The 5-level value set 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 value set is: UK
 The 3-level score is contained in input variable: u3hat
 The bandwidth is: .03
 Unweighted mean of predicted 5L score within selected sample

Variable	Obs	Mean	Std. Dev.	Min	Max
v5_u3_3	41,310	.4819085	.2487869	-.2247076	.8379605

(obs=41,310)

	v5_u3_3	v5_u3
v5_u3_3	1.0000	
v5_u3	0.8923	1.0000

(output omitted)

Table 1. Summary statistics for NDB and EQG mapped values for EQ-5D-5L

Statistic	Mapping model based on ...			
	Copula		Gauss	
	NDB	EQG	NDB	EQG
Mapping actual UK utility score				
Mean predicted EQ-5D-5L score	0.476	0.408	0.461	0.400
Minimum predicted EQ-5D-5L score	-0.225	-0.243	-0.218	-0.239
SD predicted EQ-5D-5L score	0.261	0.280	0.257	0.265
Mapping predicted UK utility score: bandwidth = 0.01, $n = 38590$				
Mean predicted EQ-5D-5L score	0.497	0.432	0.482	0.422
Minimum predicted EQ-5D-5L score	-0.225	-0.243	-0.218	-0.239
SD predicted EQ-5D-5L score	0.231	0.252	0.228	0.239
Correlation with mapped actual score	0.846	0.872	0.858	0.895
Mapping predicted UK utility score: bandwidth = 0.03, $n = 41310$				
Mean predicted EQ-5D-5L score	0.482	0.417	0.468	0.407
Minimum predicted EQ-5D-5L score	-0.225	-0.243	-0.218	-0.239
SD predicted EQ-5D-5L score	0.249	0.268	0.246	0.257
Correlation with mapped actual score	0.892	0.912	0.902	0.928
Mapping predicted UK utility score: bandwidth = 0.05, $n = 41310$				
Mean predicted EQ-5D-5L score	0.482	0.416	0.467	0.407
Minimum predicted EQ-5D-5L score	-0.225	-0.243	-0.218	-0.239
SD predicted EQ-5D-5L score	0.247	0.266	0.245	0.255
Correlation with mapped actual score	0.899	0.917	0.908	0.933
Mapping predicted UK utility score: bandwidth = 0.10, $n = 41310$				
Mean predicted EQ-5D-5L score	0.482	0.416	0.467	0.407
Minimum predicted EQ-5D-5L score	-0.189	-0.223	-0.187	-0.211
SD predicted EQ-5D-5L score	0.243	0.261	0.240	0.251
Correlation with mapped actual score	0.904	0.923	0.913	0.939

Using the smallest bandwidth, 0.01, `eq5dmap` gives a warning. For 7% of the observations, matches could not be found within the bandwidth, 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 substantial.⁸ Even this small number of missing values is enough to distort the mean because missingness is not symmetric across the upper and lower tails of the utility distribution. For that reason, it is usually unwise to use a very small bandwidth when utility values are approximate. In this case, a slightly larger bandwidth of 0.03 resolved the existence issue, and in our artificial example, larger bandwidths generated a result more highly correlated with the result of an exact mapping of actual scores: compare the rows of table 1 for bandwidths

8. The largest gap between two EQ-5D-3L values is found at the top of the distribution, between the full health value of 1 and its preceding value of 0.883. At the bottom of the distribution, the gaps are smaller but more numerous.

0.03, 0.05, and 0.1. The mean of the predicted scores is fairly stable for different bandwidths, and, as expected, the SD decreases as the bandwidth increases. Note that, for any bandwidth, the mean of the mapped approximate utility is systematically above the mean of the mapping from the actual utility score. This happens to be due to the arbitrary method we used to generate hypothetical utility values; it is not inherent in the mapping approach.⁹ Specific guidance on the bandwidth choice when mapping a single overall mean utility from EQ-5D-3L to EQ-5D-5L is given in Pennington et al. (Forthcoming). Based on the two available reference datasets, it is shown that small bandwidths, no larger than 0.1 but large enough to cover the full health value of 1, work well when the overall mean utility $E(v_3)$ in the trial is in the interval $(1, 0.7]$ (this guarantees that a full health response in EQ-5D-3L is included in the distance-weighted average in an area where there is a large gap in the EQ-5D-3L distribution). If the mean utility is in the interval $(0.7, 0.6]$, a larger bandwidth of 0.2 is preferred; for mean utilities below 0.6, an even larger bandwidth of about 0.4 is preferable. Figure 1 compares the kernel densities of the mapped values using exact matching of the EQ-5D-3L utility values and three alternative bandwidth choices: 0.03, 0.10, and 0.20. The figure illustrates how using too large a bandwidth when mapping from utility scores that are higher up in the distribution artificially pushes the expected values down because of the large gap between full health and the adjacent utility value and the relative high frequency of utility values at full health in this dataset.

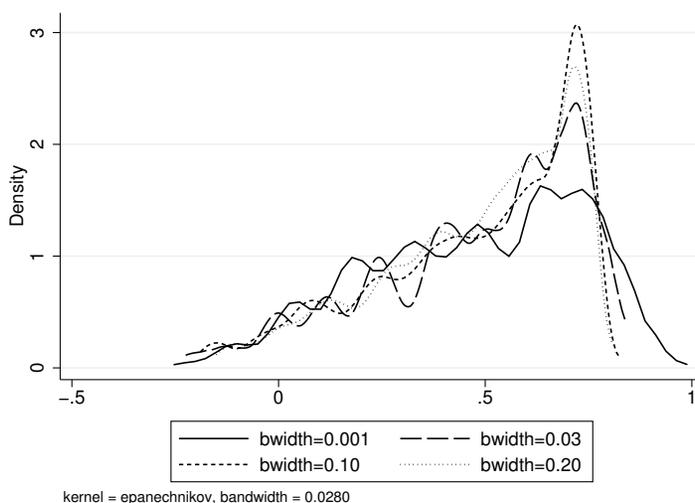


Figure 1. Kernel densities of mapped values for EQ-5D-5L using the copula mixture model and the NDB reference dataset; exact matching of individual EQ-5D-3L utility values versus approximate utility values for alternative bandwidths

9. The hypothetical utility values were generated using a tobit model that has been shown to produce biased predictions of EQ-5D-3L (Hernández Alava, Wailoo, and Ara 2012). Mapping from EQ-5D-3L values that are biased will generally result in poor EQ-5D-5L predictions.

The second important issue that should be considered in choosing a bandwidth for mapping from an approximate utility score \hat{v}_3 is monotonicity. If a trial establishes that procedure 1 gives a better expected EQ-5D-3L outcome than procedure 2, there will be serious difficulty in implementing cost-effectiveness analysis if the mapped EQ-5D-5L measure reverses that ranking, purely because of the choice of bandwidth.¹⁰ Perverse outcomes of this type can be avoided if the mapping function $\hat{v}_5(\hat{v}_3)$ generated by `eq5dmap` is monotonically increasing. The relationship between monotonicity and bandwidth choice can be illustrated by applying `eq5dmap` over a grid of values $\hat{v}_3 = -0.594, \dots, 1.0$ for alternative bandwidths; figure 2 plots the 3L \rightarrow 5L mapping function for alternative bandwidths, using the NDB copula-mixture model applied to a hypothetical 40-year-old male. In this case, the function is monotonic only for bandwidths of 0.207 or larger; if we switch to the same model based on the EQG reference dataset, monotonicity applies for bandwidths above 0.190.

The same procedure can be used to examine monotonicity of 5L \rightarrow 3L mapping. We would expect to be able to use a smaller bandwidth in that case because the UK EQ-5D-5L value set in Devlin et al. (2018) is much finer than the EQ-5D-3L value set in Dolan (1997). That proves to be the case, as shown in figure 2. For the same hypothetical individual, monotonicity is found for bandwidths above 0.041 (NDB) or 0.060 (EQG).

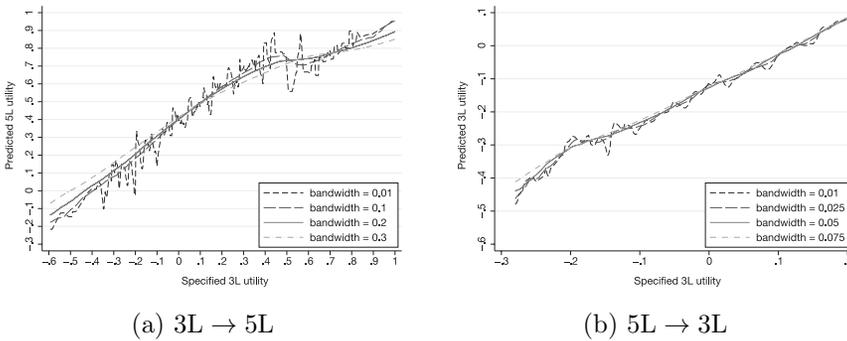


Figure 2. The effect of bandwidth choice on monotonicity of mapping functions (copula-mixture model; NDB reference data; 40-year-old male)

Although users of `eq5dmap` will want to investigate the sensitivity of results to alternative choices of bandwidth, we would strongly advise caution with mapping from an approximate utility score—it should be avoided in favor of exact individual-level mapping if at all possible; and, where unavoidable, it should not be done using bandwidths substantially less than 0.2 for 3L \rightarrow 5L mapping or 0.05 for 5L \rightarrow 3L mapping if monotonicity is required. As discussed above, exceptions to this rule of thumb should be made if the trial subjects are believed to be concentrated at either extreme of the health distribution. We recommend that sensitivity analysis always be carried out and reported.

10. Nonmonotonicity is also a possibility in exact $\mathbf{Y}_3 \rightarrow v_5$ mapping, and the possibility is inherent in the difference between the health description and utility tariff of 3L and 5L.

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 robustness. We compare results based on the NDB dataset, which covers North American patients with rheumatoid arthritis, with results based on EQG, which is an assemblage of ad hoc samples collected in several European countries. Do these very different samples give a similar picture of the relationship between responses to the three-level and five-level versions of EQ-5D? We also investigate the effects of model choice by comparing mapping results from the copula-mixture and Gaussian specifications.

We again examine the two types of mapping implemented in `eq5dmap`, beginning with $Y_3 \rightarrow v_5$ 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.

```
. foreach a of numlist 40 70 {
2.  forvalues m=0/1 {
3.  // Loop over model specification and dataset
.    foreach spec in gauss copula {
4.      foreach dat in NDB EQG {
5.        display "Age = `a`, Gender = `m`, Specification = `spec`, Data = `dat`"
6.        // 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`)
>          items(Y3_1 Y3_2 Y3_3 Y3_4 Y3_5)
7.      }
8.    }
9.  // Compute correlations between results for different reference datasets
.    correlate v5_gauss_NDB_`a`_`m` v5_gauss_EQG_`a`_`m` if age==`a` & male==`m`
10.   correlate v5_copula_NDB_`a`_`m` v5_copula_EQG_`a`_`m` if age==`a` & male==`m`
11.   }
12. }
```

(output omitted)

Table 2 summarizes the results in terms of the mean score, minimum score, and SD 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 the EQG dataset and the mixed copula rather than the 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 are slightly lower for older individuals.

Table 2. Summary statistics for NDB and EQG mapped values for EQ-5D-5L

Statistic	Mapping model based on ...			
	Copula		Gauss	
	NDB	EQG	NDB	EQG
Age 40, female				
Mean predicted EQ-5D-5L score	0.459	0.413	0.450	0.407
Minimum predicted EQ-5D-5L score	-0.225	-0.241	-0.218	-0.238
SD predicted EQ-5D-5L score	0.267	0.284	0.261	0.269
Correlation	0.971		0.980	
Age 40, male				
Mean predicted EQ-5D-5L score	0.475	0.423	0.466	0.413
Minimum predicted EQ-5D-5L score	-0.217	-0.239	-0.208	-0.235
SD predicted EQ-5D-5L score	0.264	0.285	0.259	0.269
Correlation	0.971		0.979	
Age 70, female				
Mean predicted EQ-5D-5L score	0.473	0.390	0.455	0.390
Minimum predicted EQ-5D-5L score	-0.220	-0.243	-0.215	-0.237
SD predicted EQ-5D-5L score	0.264	0.283	0.260	0.269
Correlation	0.958		0.970	
Age 70, male				
Mean predicted EQ-5D-5L score	0.489	0.402	0.471	0.396
Minimum predicted EQ-5D-5L score	-0.208	-0.240	-0.203	-0.234
SD predicted EQ-5D-5L score	0.260	0.284	0.257	0.269
Correlation	0.958		0.968	

Figure 3 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 copula-mixture model is illustrated), there are noticeable differences between the results for the two reference datasets. For a given reference dataset (the NDB is illustrated), the difference between the copula-mixture and Gaussian models is less pronounced, although there remain differences at the higher end of the utility values.

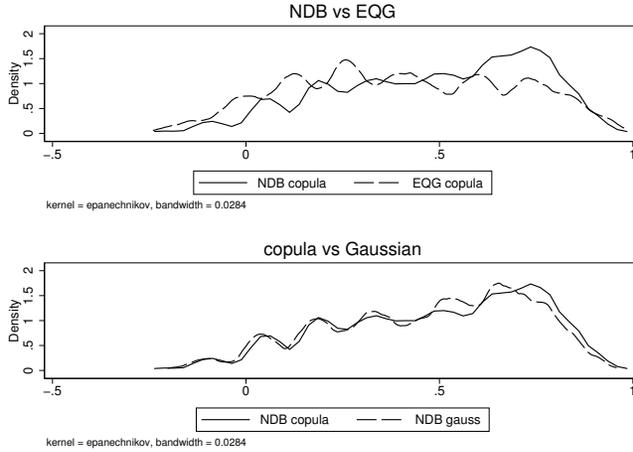


Figure 3. Comparison of kernel densities of mapped values for EQ-5D-5L for a 70-year-old male

Next, we repeat the exercise for a $v_3 \rightarrow v_5$ mapping using the entire dataset. First, we use `u3`, the variable containing the actual UK utility scores corresponding to the health states described by the variables `Y3_1` to `Y3_5`. The following code produces results that are summarized in the first section of table 1.

```
. // Actual UK 3-level utility scores:
. // loop over model specification and dataset
. foreach spec in gauss copula {
2.   foreach dat in NDB EQG {
3.     display "Specification = `spec`, Data = `dat`"
4.     display "Actual UK scores..."
5.     eq5dmap v5_`spec`_`dat`_actual,
>       covariates(age male) model(`dat` `spec`)
>       score(u3) bwidth(0.001)
6.   }
7. }
(output omitted)
```

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 bandwidth, the lower the correlations with the values mapped from the actual utility scores and also the smaller the SDs of the mapped values.

5.5 Mapping using weights

Our dataset contains a row for every possible EQ-5D-3L health state by age and gender. The mean three-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 three-level descriptive system with associated utility scores around and below 0 (equivalent to death). Even disease-specific samples in very severely affected subpopulations usually have much larger average three-level utility scores: for example, the mean three-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 only 8% and 4%. This contrasts with the almost 35% health states with negative utility values in the whole 3L descriptive system.

In this section, we demonstrate the use of the weighting option to calculate the mean mapped five-level utility for the populations in the EQG and the NDB samples. The variables `fwEQG` and `fwNDB` in the dataset contain frequency weights describing the demographic composition of the EQG and NDB data, 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 five-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], covariates(age male) model(NDBcopula)
> items(Y3_1 Y3_2 Y3_3 Y3_4 Y3_5)
No direction specified: default is 3->5
No 5L value set specified: default is UK
Summary of inputs to eq5dmap:
The 5-level value set 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	.7088379	.2298323	-.2170747	.9587727

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

```
. eq5dmap v5_y3NDB [fw = fwNDB], covariates(age male) model(NDBcopula)
> items(Y3_1 Y3_2 Y3_3 Y3_4 Y3_5)
No direction specified: default is 3->5
No 5L value set specified: default is UK
Summary of inputs to eq5dmap:
The 5-level value set 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	.7640676	.1689144	-.2246118	.9584147

The mean difference between three-level and five-level utility scores calculated for the EQG and NDB sample compositions are fairly small—0.081 and 0.083, respectively. In comparison, the average three-level or five-level difference is 0.339 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 (2017) and Hernández-Alava et al. (2018) of minor differences between three-level and five-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 $v_3 \rightarrow v_5$ mapped values in section 5.4 using the mixed copula model.

```
. // calculate correlations - unweighted
. correlate v5_copula_NDB_actual v5_copula_EQG_actual
(obs=41,310)
```

	v5_cop..	v5_cop..
v5_copula_..	1.0000	
v5_copula_..	0.9549	1.0000

```
. // calculate correlations - weighted
. correlate v5_copula_NDB_actual v5_copula_EQG_actual [fw = fwNDB]
(obs=5,205)
```

	v5_cop..	v5_cop..
v5_copula_..	1.0000	
v5_copula_..	0.9967	1.0000

```
. correlate v5_copula_NDB_actual v5_copula_EQG_actual [fw = fwEQG]
(obs=3,539)
```

	v5_cop..	v5_cop..
v5_copula_..	1.0000	
v5_copula_..	0.9959	1.0000

The unweighted correlation between the mapped values using different reference datasets is high (0.9549) 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 EQ-5D-3L and EQ-5D-5L 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 command, `eq5dmap`, that calculates the prediction of EQ-5D-3L utility scores from observed or specified values of EQ-5D-5L (individual items or utility score), age and gender, and vice versa. The predictive distribution was derived from a joint model of the two versions of EQ-5D developed in Hernández-Alava and Pudney (2017) and applied to two different reference datasets (Hernández-Alava et al. 2018). We illustrated use of the command through several examples. The first two examples demonstrated the basic use of the command. Further examples illustrated the consequences of a) using different bandwidths when mapping using approximate utility scores and b) using different reference datasets. The final example demonstrated the use of weights.

7 Acknowledgments

This work was supported by the National Institute for Health and Care Excellence with 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.

We thank Kaleb Michaud (University of Nebraska Medical Center and National Data Bank for Rheumatic Diseases), Frederick Wolfe (National Data Bank for Rheumatic Diseases), and the EuroQol Group for providing the data used in the calculation of the weighting variables in section 5.

8 References

- Augustovski, F., L. Rey-Ares, V. Irazola, O. U. Garay, O. Gianneo, G. Fernández, M. Morales, L. Gibbons, and J. M. Ramos-Goñi. 2016. An EQ-5D-5L value set based on Uruguayan population preferences. *Quality of Life Research* 25: 323–333.
- Devlin, N., K. Shah, Y. Feng, B. Mulhern, and B. van Hout. 2016. Valuing health-related quality of life: An EQ-5D-5L value set for England. Office of Health Economics, London, UK.
- Devlin, N. J., K. K. Shah, Y. Feng, B. Mulhern, and B. van Hout. 2018. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Economics* 27: 7–22.
- Dolan, P. 1997. Modeling valuations for EuroQol health states. *Medical Care* 35: 1095–1108.
- Hernández-Alava, M., and S. Pudney. 2017. Econometric modelling of multiple self-reports of health states: The switch from EQ-5D-3L to EQ-5D-5L in evaluating drug therapies for rheumatoid arthritis. *Journal of Health Economics* 55: 139–152.
- Hernández-Alava, M., A. Wailoo, S. Grimm, S. Pudney, M. Gomes, Z. Sadique, D. Meads, J. O’Dwyer, G. Barton, and L. Irvine. 2018. EQ-5D-5L versus EQ-5D-3L: The impact on cost effectiveness in the United Kingdom. *Value in Health* 21: 49–56.
- Hernández Alava, M., A. J. Wailoo, and R. Ara. 2012. Tails from the peak district: Adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. *Value in Health* 15: 550–561.
- van Hout, B., M. F. Janssen, Y.-S. Feng, T. Kohlmann, J. Busschbach, D. Golicki, A. Lloyd, L. Scalone, P. Kind, and A. S. Pickard. 2012. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health* 15: 708–715.
- Janssen, M. F., A. S. Pickard, D. Golicki, C. Gudex, M. Niewada, L. Scalone, P. Swinburn, and J. Busschbach. 2013. Measurement properties of the EQ-5D-5L compared

- to the EQ-5D-3L across eight patient groups: A multi-country study. *Quality of Life Research* 22: 1717–1727.
- Kim, S.-H., J. Ahn, M. Ock, S. Shin, J. Park, N. Luo, and M.-W. Jo. 2016. The EQ-5D-5L valuation study in Korea. *Quality of Life Research* 25: 1845–1852.
- Luo, N., G. Liu, M. Li, H. Guan, X. Jin, and K. Rand-Hendriksen. 2017. Estimating an EQ-5D-5L value set for China. *Value in Health* 20: 662–669.
- NICE. 2014. Developing NICE guidelines: The manual. National Institute for Health and Care Excellence. <https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview#updating-this-manual>.
- Pennington, B., M. Hernández-Alava, S. Pudney, and A. J. Wailoo. Forthcoming. Comparing the EQ-5D-3L and 5L versions. What are the implications for model-based cost effectiveness estimates? Decision Support Unit report.
- Purba, F. D., J. A. M. Hunfeld, A. Iskandarsyah, T. S. Fitriana, S. S. Sadarjoen, J. M. Ramos-Goñi, J. Passchier, and J. J. V. Busschbach. 2017. The Indonesian EQ-5D-5L value set. *Pharmacoeconomics* 35: 1153–1165.
- Shiroiwa, T., S. Ikeda, S. Noto, A. Igarashi, T. Fukuda, S. Saito, and K. Shimozuma. 2016. Comparison of value set based on DCE and/or TTO data: Scoring for EQ-5D-5L health states in Japan. *Value in Health* 19: 648–654.
- Versteegh, M. M., K. M. Vermeulen, S. M. A. A. Evers, G. Ardine de Wit, R. Prenger, and E. A. Stolk. 2016. Dutch tariff for the five-level version of EQ-5D. *Value in Health* 19: 343–352.
- Wolfe, F., and K. Michaud. 2011. The National Data Bank for rheumatic diseases: A multi-registry rheumatic disease data bank. *Rheumatology* 50: 16–24.
- Xie, F., E. Pullenayegum, K. Gaebel, N. Bansback, S. Bryan, A. Ohinmaa, L. Poissant, and J. A. Johnson. 2016. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Medical Care* 54: 98–105.

About the authors

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

Steve Pudney is Professor of Health Econometrics in the Health Economics and Decision Science section in SchARR, University of Sheffield, UK.