



This is a repository copy of *Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from a UK population study*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/194398/>

Version: Published Version

Article:

Hernández Alava, M., Pudney, S. and Wailoo, A. orcid.org/0000-0002-9324-1617 (2023) Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from a UK population study. *Pharmacoeconomics*, 41 (2). pp. 199-207. ISSN 1170-7690

<https://doi.org/10.1007/s40273-022-01218-7>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

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/>



Estimating the Relationship Between EQ-5D-5L and EQ-5D-3L: Results from a UK Population Study

Mónica Hernández Alava¹ · Steve Pudney¹ · Allan Wailoo¹

Accepted: 2 November 2022
© The Author(s) 2022

Abstract

Objectives The aim of this study was to estimate the relationship between EQ-5D-3L and EQ-5D-5L, in both directions, using a single model.

Methods An online survey containing both variants of EQ-5D, with randomised ordering, was administered to a large UK sample in 2020. A joint statistical model of the ten EQ-5D responses (five at 5L, five at 3L), using a multi-equation ordinal regression framework was estimated. The joint model ensures mappings in either direction are fully consistent with the information in the sample and satisfy Bayes' rule. Three extensions enhance model flexibility: a copula specification allows differing degrees of correlation between the 3L and 5L responses at the upper and lower extremes of health; a normal mixture residual distribution gives flexibility in the distributional form of responses; and a common factor captures correlations in responses across the five dimensions.

Results Almost 50,000 responses were received. Thirty-five percent of respondents reported an existing medical condition. Ninety percent of possible 3L and 43% of possible 5L health states were observed. The preferred model specification includes age, sex and the responses to the EQ-5D instrument. Close alignment to the observed data was observed both in within-sample and out-of-sample comparisons.

Conclusion The results from this study provide a means of translating evidence to or from EQ-5D-3L to or from 5L based on a large-scale UK population survey with randomised ordering. Mapping can be performed either using descriptive system responses, individual utility scores or summary statistics.

Key Points for Decision Makers

This paper presents a method for mapping between EQ-5D-3L and 5L, in either direction, and from responses to the descriptive system, individual utility scores or summary statistics.

The analysis is based on a large UK sample of respondents where the survey was designed to minimise potential biases.

Good model performance is demonstrated in and out of sample with pre-programmed software allowing easy implementation of the model results for non-specialist analysts.

1 Introduction

EQ-5D is one of the outcome measures most widely used for the calculation of quality-adjusted life years (QALYs) in economic evaluation. It comprises a descriptive system and a set of values ('utilities') for each health state that can be described. The descriptive system allows respondents to indicate their health state on five dimensions: mobility, ability to self-care, ability to undertake usual activities, pain and discomfort, and anxiety and depression. In the three-level version (3L), respondents indicate the degree of impairment on each dimension according to three levels (no problems, some problems, extreme problems). A newer, five-level version (5L) includes five levels of severity for each dimension (no problems, slight problems, moderate problems, severe problems, and extreme problems).

It has previously been demonstrated that there are substantial differences between the way in which 3L and 5L estimate QALY gains for health technologies and consequently their cost effectiveness. These differences occur

✉ Allan Wailoo
a.j.wailoo@sheffield.ac.uk

¹ School of Health and Related Research, University of Sheffield, Sheffield S1 4DA, UK

both in terms of the responses individuals give to the two descriptive systems and the valuation of health states. This is true across a wide range of disease areas and health technologies, in trial-based and model-based economic evaluations both in the UK [1, 2] and internationally [3]. Whilst the development of improved means of assessing health-related quality of life for economic evaluation is both welcome and inevitable, it raises challenges for consistent decision making. Since the two EQ-5D variants cannot be treated as if they were interchangeable, evidence gathered using one variant will need to be translated to the other. This is a requirement irrespective of whether 3L or 5L is seen as the preferred option and will remain a requirement potentially for decades, as the mass of 3L evidence diminishes in relevance and is gradually replaced by studies that measure health-related quality of life using the 5L instrument.

There are two requirements to generate the tools to allow analysts to easily translate evidence. First is a dataset sufficiently robust for the task in terms of sample size and other design features. Second is the application of appropriate analytical techniques that allow the mapping of 3L to 5L and vice versa in a way that remains consistent in both directions. For the approach to be of practical benefit, it needs to be capable of translating both individual patient responses to the descriptive system of either instrument to the value set of the other, and be able to use tariff scores or mean tariff scores (as would typically be reported in published evidence from clinical studies) to perform the mapping.

There are two existing approaches [4, 5] that map between 3L and 5L and both have some substantial limitations relative to the potential impact of their use. Van Hout et al. [4] provide a method to map from the responses provided to the 5L descriptive system to 3L tariff scores based on data from a EuroQoL Group (EQG) coordinated study carried out in six countries: Denmark, England, Italy, the Netherlands, Poland and Scotland. The questionnaire introduced the five-level version of EQ-5D first, followed by a few background questions (age, gender, education, etc.), then the three-level version of EQ-5D. The sample included eight broad patient groups (cardiovascular disease, respiratory disease, depression, diabetes mellitus, liver disease, personality disorders, arthritis, and stroke) and a student cohort. All countries used paper and pencil questionnaires, apart from England, which used an online version. There are four main limitations to the EQG data used in this study.

First, the sample of 3691 responses is small relative to the numbers of health states the mapping covers, leading to concerns about the degree of extrapolation involved. The 3L instrument describes $3^5 = 243$ logically possible health states, of which 51% are observed. The 5L instrument

describes $5^5 = 3125$ possible states, of which just 11% are observed.

Second, the use of an international sample, including responses from several non-English speaking countries, may be misleading for any single country. For use in the UK, the issue of ambiguity between ‘severe problems’ (level 4) and ‘extreme problems’ (level 5) of the 5L may raise particular concerns [6], though the impact of this potential ambiguity may be less in the context of mapping compared with valuation studies, since the descriptions are presented to respondents in the intended order of severity.

Third, valid mapping requires that responses to each EQ-5D variant are independent of the presence of the other in the same survey. The validity of this assumption is likely to depend on the degree of separation within the survey of the two instruments, and possibly also on the mode of administration. Responses to the 3L version encountered later in the EQG questionnaire may be contaminated in some way by recollection of responses to the earlier 5L version if the two instruments are nearly contiguous. The degree of separation between the 5L and 3L instruments in the EQG survey was limited: only a small number of questions were asked between the two EQ-5D variants, so there is a risk that 3L responses may be distorted (relative to what would be observed in a 3L-only survey).

Fourth, it has been established in a randomised experiment [7], undertaken as a pilot specifically to inform survey design for mapping, that the ordering of the two variants can have a material influence on the responses that are given. As with the third limitation, this relates to the fact that we are seeking to minimise the impact on responses to one variant of EQ-5D to the presence of the other variant in the same survey. In the EQG study, all respondents were presented with the 5L instrument before they encountered the 3L instrument.

The van Hout et al. analysis [4] provides a method for mapping from the 5L descriptive system to the 3L value set. An alternative approach developed by Hernández Alava and Pudney [5] goes beyond this to facilitate mapping in either direction within a single model (from 3L to 5L and vice versa) and using either responses to the descriptive system or tariff scores (whether these relate to a unique health state or are simply a mean value obtained from a sample of responses). This work originally applied the method to 3L and 5L data collected from a North American rheumatic disease registry and was then subsequently applied to the same EuroQoL dataset. Many of the data limitations described above are equally relevant.

In this paper, we present results from a UK study designed to provide data for the purposes of mapping between the two EQ-5D variants, and the UK/English value sets [8, 9]

that correct for the limitations of data that underpin existing approaches.

2 Methods

2.1 Data Collection

A new data collection exercise was designed and administered online in English for samples of the general population in the UK. We refer to this as the EEPURU (Policy Research Unit in Economic Methods of Evaluation in Health and Social Care Interventions) survey. Online administration was included in a pilot study in the Understanding Society ‘Innovation Panel Study 10’ (IP) which used a mixture of web interview (CAWI) and computer-assisted self-interviewing (CASI) in a sample of almost 3000 individuals [6]. Building on that independent pilot, data for the present study was collected by a specialist polling company (OnePoll) using existing UK panels during April 2020. Ethics approval for the study was granted by the University of Sheffield.

Members of the OnePoll panel are typically highly engaged. Participants are paid to complete surveys on a wide variety of topics and are removed from the panel if there is evidence that they are not providing considered responses (for example, if they continuously pick the same ordered option). There were no responses that we were prepared to deem logically inconsistent: we did not impose any requirement on how the 3L and 5L instruments could be answered. Therefore, no responses were ruled out as invalid in our survey.

The OnePoll panel is designed to be representative of the UK population in many measurable aspects; it should be noted that we were not seeking a representative sample of UK opinions. Rather, the concern was to obtain sufficient observations across the EQ-5D severity range. Oversampling of those in ill health was considered, but the preferred option was to use a sample from the overall panel that would be large enough to cover an appropriately wide range of health states.

There is no formal method for determining an appropriate sample size in this situation. The 3L instrument describes 243 logically possible health states, while the 5L instrument describes 3125 possible states, and therefore $15^5 = 759,375$ possible 3L–5L combinations. Although we might expect a significant proportion of those combinations to be rare in practice, it is clear that a large sample survey is required to adequately represent the range of combinations likely to be encountered in practice and to avoid excessive extrapolation.

As guidance, we used the General Practitioner Patient Survey (GPPS), a large-scale cross-sectional study supported by NHS England, which records the 5L version of EQ-5D. Data from 2012 to 2015 with a complete 5L

response ($n = 792,571$) shows that 2464 of the 3125 possible 5L health states were observed. We drew random subsamples (without replacement within each subsample) of varying sizes from the data, with 10 repetitions each. The mean number of distinct 5L health states by sample size is displayed in Fig. 1. For good coverage of 5L health states, a large sample is required. A sample size of 5000 would be expected to observe only 450 5L health states, 18% of those observed in the whole GPPS sample. A sample of 50,000 was targeted as this would be expected to yield observations relating to 1245 5L states, marginally above 50% of the states that appear in the GPPS sample.

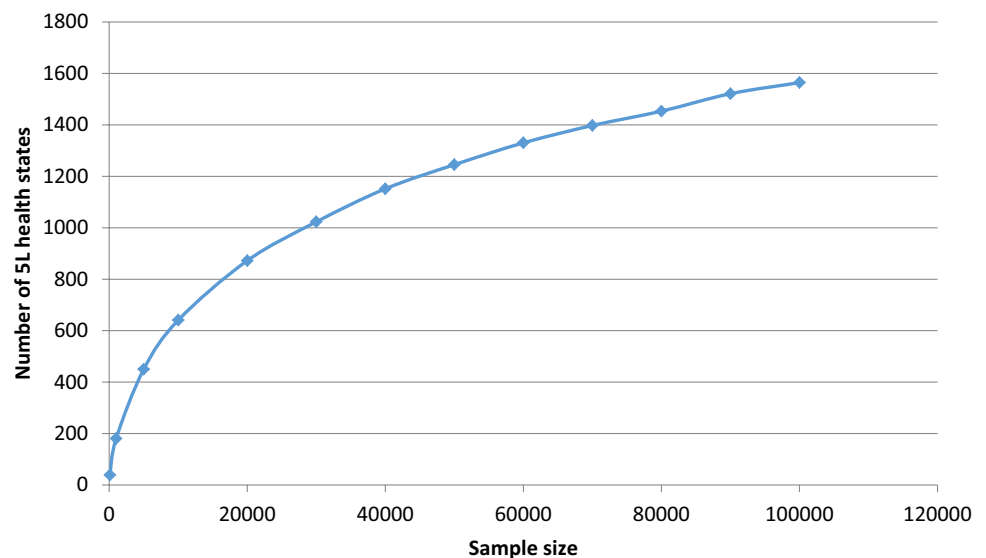
The survey design meant respondents would see variant 1 (3L or 5L) followed by the EuroQoL visual analogue scale (VAS), then a series of questions (age group, sex, family circumstances, educational achievement, existing medical conditions, use of medication, caring responsibilities, life and health satisfaction), followed by the second variant (5L or 3L) and a repeat of the VAS instrument. The inclusion of the VAS and its repetition was a requirement of the EuroQoL group.

2.2 Statistical Analysis

Full details of the modelling methods are reported in Hernández Alava and Pudney [5]. The approach is based on a joint statistical model of the ten EQ-5D responses (five at 5L, five at 3L), using a multi-equation ordinal regression framework. These ten equations are arranged in five groups, each including both 3L and 5L responses to a single dimension. Only by estimating a joint model as the basis for mapping in both directions we can ensure that the mappings (i.e. the conditional distributions for 3L|5L and 5L|3L) are fully consistent with the information in the sample and satisfy Bayes’ rule. In other words, using a joint model guarantees that the ratio of the conditional distributions is identical to the ratio of the marginal distributions in any population we apply them to. Estimating separate models will not in general satisfy Bayes’ rule. In addition, three special features were used to enhance the flexibility of the model.

First, within each of the five groups of equations, we use a copula specification to allow differing degrees of correlation between the 3L and 5L responses at the upper and lower extremes of health [10]. We allow these copulas to differ across the five different dimensions of EQ-5D. The copulas considered for selection allowed for different types of dependence and included the Gaussian, Clayton, Frank, Gumbel and Joe copulas. Second, we use normal mixture marginal distributions to give flexibility in the distributional form of responses. Mixture distributions are very flexible and can approximate many different distributional forms. Third, we include a latent factor to capture correlations in responses across the five dimensions of EQ-5D due to

Fig. 1 Coverage of 5L health states and sample size



common unobserved influences affecting the individual's responses (for example, a person's general tendency to "look on the bright side" or "take a dim view").

The entire sample is used for estimation, without editing or over-riding the data in any way.

We used two datasets for out-of-sample testing of model results: FORWARD Databank and the IP datasets. FORWARD Databank, the National Data Bank for Rheumatic Diseases (NDB), is a register of patients of US and Canadian rheumatologists [11]. During a switch from 3L to 5L, both were collected in the January 2011 wave, the 5L version early in the interview and the 3L version towards the end. The NDB dataset has 5295 complete cases. The IP [12] is a subset of the full "Understanding Society" household panel annual survey reserved for experimentation of various kinds. Wave 11, conducted in June to September 2018, included both 3L and 5L (in a randomised ordering) with complete responses obtained for 1705 of the sample. There were no other suitable, large-scale datasets available for analysis.

In the current absence of an EQ-5D-5L value set for England that is approved for use by the National Institute for Health and Care Excellence (NICE), our reporting focusses on mapping 5L responses to the 3L value set.

3 Results

3.1 Characteristics of the Response Sample

A sample of $n = 49,999$ responses was received and included in the analysis. Summary information is provided in Table 1. The sample is well balanced between males and females. All respondents were over 18 years. The largest of the seven

age categories was 35–44 years, but there was good spread across age groups between 25 and 74 years, and almost 2000 responses were obtained from people aged 75 years or over. On average, sample members are well educated; the most common level of educational attainment was undergraduate degree.

Despite being designed as a general-population sample survey, there is extensive coverage of people with impaired health—35% of respondents reported an existing diagnosed medical condition and 52% reported taking some type of medication. Good coverage of both 3L and 5L health states is achieved; 90% of possible 3L states and 43% of possible 5L health states were observed. Over half the 5L states observed in the GPPS dataset were reported here, consistent with the calculations used to guide sample size. Proportionally few responses are observed from individuals in the most severe categories of impairment measured by EQ-5D. However, the large overall sample size ensures there are adequate numbers of observations even in the most extreme categories. The lowest and highest numbers of respondents in the most severe category of impairment were seen for the self-care (427 and 250 for 3L and 5L, respectively) and anxiety/depression (2387 and 1148) domains.

3.2 Modelling Results

The best fitting specification included age and sex and a two-component mixture for the error terms. The Clayton copula was found to give the best fit for three of the five dimensions (mobility, self-care and usual activities). The Clayton copula allows only positive dependence, with strong left tail dependence and relatively weak right tail dependence. Therefore, in these three dimensions of EQ-5D, the dependence between 3L and 5L responses is weaker at poor

Table 1 Sample characteristics

Characteristics	EEPRU data	
	<i>n</i>	%
Sex		
Male	24,309	48.62
Education		
Secondary education (GCSE/O-Levels)	10,194	20.39
Post-secondary education (College, A-Level)	8678	17.36
Vocational qualification (Diploma, Cert)	8908	17.82
Undergraduate degree (BA, BSc, etc.)	13,011	26.02
Postgraduate degree (MA, MSc, etc.)	6564	13.13
Doctorate (PhD)	1607	3.21
None of the above	1037	2.07
Family status		
Single + no children	9366	18.73
Single + children	2004	4.01
In a relationship + no children	3600	7.2
In a relationship + children	2204	4.41
Cohabiting + no children	2203	4.41
Cohabiting + children	2217	4.43
Married + no children	5121	10.24
Married + children	18,704	37.41
Divorced + no children	749	1.5
Divorced + children	2357	4.71
Widowed + no children	366	0.73
Widowed + children	1108	2.22
Health		
Pre-existing diagnosed condition	17,688	35.38
Taking medication	26,098	52.2
Age (y)		
< 18		
18–24	4210	8.42
25–34	8729	17.46
35–44	9767	19.53
45–54	8640	17.28
55–64	9167	18.33
65–74	7519	15.04
75+	1967	3.93
EQ-5D-3L		
Coverage among all 243 health states	219	90.12
EQ-5D-5L		
Coverage among all 3125 health states	1341	42.91

EEPRU Policy Research Unit in Economic Methods of Evaluation in Health and Social Care Interventions

health states than at good ones. The best fitting copulas in the pain/discomfort and anxiety/depression dimensions were Frank and Gumbel, respectively. The Frank copula is symmetric and generates 3L–5L dependence that is weaker in the tails and stronger in the centre of the distribution of

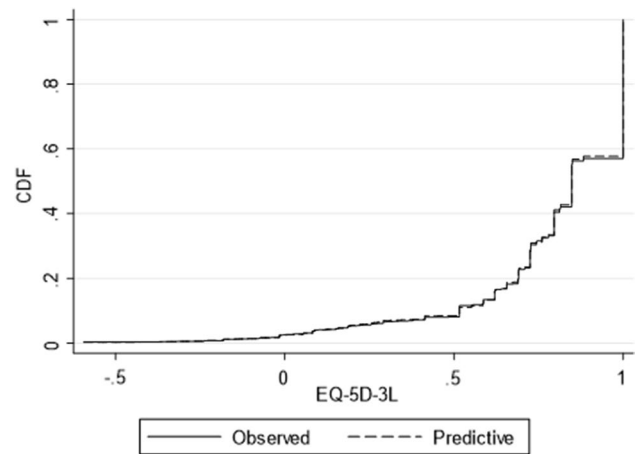


Fig. 2 Cumulative distribution function, within-sample

pain/discomfort responses. The Gumbel copula for anxiety/depression displays weak 3L–5L dependence in cases with good mental health but stronger dependence for those in a poorer state.

3.3 Within-Sample Fit

Figure 2 plots the empirical cumulative distribution function (cdf) of the EQ-5D-3L utility scores together with the cdf generated by the estimated model. There is no part of the distribution where any significant deviation between the data and the model is observed. Averaging across the sample, the mean error was 0.002 (95% confidence interval [CI] 0.001–0.003), mean absolute error 0.073 (0.072–0.074) and root mean squared error 0.130 (0.127–0.132) (See Table 3).

There is not a strictly monotonic correspondence between 3L and 5L responses in the data, as Table 2 shows (for example, the mean 3L utility score for respondents with 5L state 41111 is higher than the mean score for respondents with state 31111). The estimated model reflects this feature quite accurately. Table 2 uses example health states from the ‘mobility’ and ‘pain’ domains, taking the best state 11111 and middle state 33333 and varying the 5L mobility or pain response from 1 to 5. The 3L utility predicted by the model is compared with the sample average utility score for that state. Model predictions are based on the average for the sample. Some health states have very few observations, but the comparison demonstrates that the profiles for the important mobility and pain domains at different levels of general health are captured rather well by the model.

We also examined model fit at the level of the descriptive system by comparing the predicted proportions in each of the EQ-5D-3L categories for the five domains and with the observed proportions in the sample (see Fig. 3). Differences were very small. The largest difference was <0.006 (0.330

Table 2 Predicted and observed mean 3L utility scores for illustrative health states

5L state	3L	
	Mean observed	Mean predicted
Mobility domain		
11111	0.99	0.99
21111	0.88	0.92
31111	0.82	0.91
41111	0.95 ^a	0.95
51111	0.93 ^b	0.95
13333	0.45 ^b	0.48
23333	0.43	0.45
33333	0.49	0.43
43333	0.42 ^b	0.42
53333	0.33 ^a	0.37
Pain domain		
11111	0.99	0.99
11121	0.87	0.87
11131	0.80	0.78
11141	0.60 ^b	0.81
11151	0.82 ^a	0.83
33313	0.63 ^a	0.52
33323	0.45 ^b	0.50
33333	0.49	0.43
33343	0.25	0.21
33353	− 0.04	0.08

^a Based on 5 or fewer observations

^b Based on 25 or fewer observations

predicted vs 0.335 observed) which occurred in the “some problems” category for anxiety/depression.

3.4 Out-of-Sample Fit

Table 3 shows the overall summary measures of fit in the two out-of-sample options available (the FORWARD and IP datasets). Both MAE and RMSE are higher in the FORWARD dataset than the in-sample values but, surprisingly, lower in the IP dataset. Figure 4 shows a similar pattern of very close fit in the IP study data. The only observable divergence occurs in the range of EQ-5D-3L from approximately zero to 0.5. The model predicts a slightly higher proportion of the data to lie within this range than is observed. The plot for the FORWARD dataset also shows a small divergence higher up the EQ-5D scale.

Figure 3 shows the differences between mean observed and predicted probabilities of each 3L response within the five health domains. The largest difference (0.058) is in the probability of moderate problems with pain/discomfort using the FORWARD data. The model fits closer to the IP data than the FORWARD data.

Overall, however, the model exhibits very close alignment to these out-of-sample sources.

3.5 Comparison with van Hout et al. [4]

The FORWARD and IP datasets were used for comparisons of out-of-sample fit. Point estimates of summary measures of error over the whole sample favoured the van Hout et al. [4] approach, but all 95% confidence intervals overlapped. These same measures were examined for subgroups of the FORWARD data defined in terms of age and sex categories. A mixed picture resulted. The better performing model varied across these categories and also by the measure of error used.

We also calculated the differences between the predicted mean probability of being in each of the three 3L response categories and the corresponding sample proportions of those categories in the FORWARD and IP datasets for the van Hout et al. [4] approach (see Supplementary Information Figure 1 in the electronic supplementary material [ESM]). In FORWARD, prediction errors are relatively close to zero for our model in the domains of mobility, self-care and usual activities at all levels, and show closer alignment to the data than van Hout et al. However, the domains of pain and anxiety/depression show more mixed patterns. For pain, the model performs noticeably worse than van Hout et al. at levels 1 and 2, tending to under-predict “no pain” and over-predict “some pain”, but is better at extreme pain level 3.

Using the IP dataset, the picture that emerges is rather different. Mean errors are smaller for both mappings than they are in the FORWARD data, so there is not much to choose between them. It is striking that the results for pain and anxiety/depression are quite different from those in the FORWARD data, and mostly favour our mapping model.

We constructed cdfs using the van Hout et al. [4] approach (see Supplementary Information Fig. 2 in the ESM). Between EQ-5D values of 0.5 and 1.0, the van Hout et al. approach performs noticeably worse in both the datasets. Between 0.1 and 0.5 there is slight underestimation that is more apparent in the FORWARD dataset. The cdf for van Hout et al. deviates from the data more than the cdf for our model, but the difference is not large.

4 Discussion

We were motivated to design a survey specifically to provide a robust method for estimating the relationship between EQ-5D-3L and 5L in the UK. Currently, this is most likely to be of use in translating EQ-5D-5L results from clinical studies to the 3L value set but the requirement to map in the other direction, in a consistent manner, could be a requirement for many years as a means of linking decades of evidence



Fig. 3 Difference between observed and predicted probabilities for the 3L level within sample and in the FORWARD and Innovative Panel (IP) data

Table 3 Summaries of within-sample and out-of-sample fit

Mapping	Mean error	Mean absolute error	Root mean square error
Within sample	0.002 [0.001–0.003]	0.073 [0.072–0.074]	0.130 [0.127–0.132]
FORWARD dataset	0.010 [0.005–0.014]	0.100 [0.097–0.103]	0.150 [0.144–0.155]
IP dataset	0.007 [0.002–0.012]	0.064 [0.059–0.068]	0.109 [0.100–0.118]

95% bootstrap confidence intervals (500 replications)

from clinical studies using EQ-5D-3L to a 5L value set. The study sought to overcome data and design limitations of previous studies, including our own, and drew on pilot

studies to inform this new design. Sample size and coverage of health states are clearly improved but other changes such as the randomised ordering, separation of the two EQ-5D

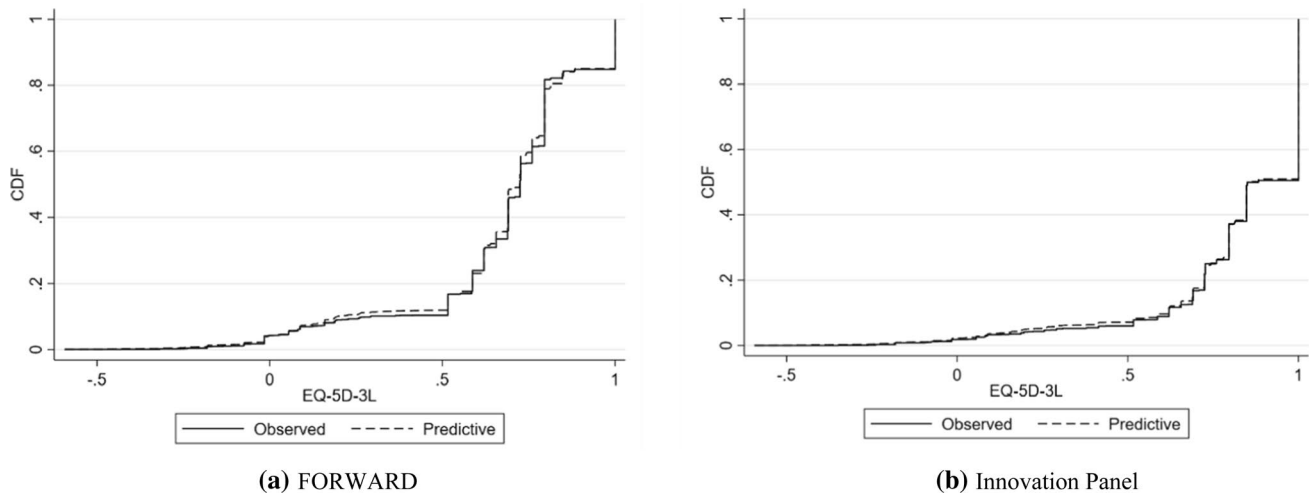


Fig. 4 Cumulative distribution functions out-of-sample using **a** FORWARD and **b** Innovation Panel validation datasets

variants, the use of a general population, UK sample single language variant are equally important features. These aspects of design may hold important lessons for mapping studies more generally.

We applied an econometric model to simultaneously model the 3L and 5L responses, providing a coherent approach for mapping in either direction. Mapping models that predict 3L and 5L separately do not provide consistency. The approach demonstrates good fit to observed data both within sample and in out-of-sample testing in two smaller datasets. Crucially, there is no evidence of any systematic deviation from observed data at any particular part of the distribution of health-related quality of life. Within this general picture, it is of note that there are differences between the datasets, lending further weight to the importance of the design of data collection studies in this area.

Comparisons with the van Hout et al. approach are made in two out-of-sample datasets. The FORWARD and IP datasets are certainly not representative of all patient populations of interest to cost-effectiveness analysts, and both surveys have some of the drawbacks that the Policy Research Unit in Economic Methods of Evaluation of Health and Social Care Interventions (EEPRU) data collection sought to eliminate. We find mixed evidence of the relative performance of the van Hout et al. approach and our approach in these datasets. The van Hout et al. approach performs slightly better in terms of whole-sample summary measures. However, differences are small in relation to the sampling variability in their estimates, with overlapping confidence intervals. A much more complex picture emerges when looking at fit across subgroups of the samples, the predicted probabilities by domain responses and the cdf. A more detailed set of comparisons, including with earlier versions of the Hernández et al. work [5], are reported in Hernández et al [13].

All these results are easily implemented for analysts, with pre-programmed functions and examples in Stata, R and Excel [14] allowing the prediction of EQ-5D-3L scores from EQ-5D-5L data and vice versa (currently based on the published 5L value set for England [9]). The mapping can be performed using responses to the descriptive system, a health utility score, or a summary statistic such as a mean utility score which may not align to any unique health state. Therefore, the results provide a mapping solution for the variety of situations that analysts face, whether they have access to patient level data (e.g. when conducting an economic evaluation alongside a clinical trial) or, when published, summary statistics are used as is typically the case for decision analytic models.

There are limitations. We observed a non-monotonic relationship in a few isolated examples, which may relate to ambiguity in the wording of the descriptive system. Alternative versions of the model that impose a monotonic relationship are feasible but, having taken care to generate a fit-for-purpose data collection exercise, to then over-ride these observations in favour of the relationship assumed by the designers of the EQ-5D instruments would provide misleading results.

We currently focus on the UK value set. Since the model is based on a form of response mapping, focussing on the observed responses to the descriptive systems, it is straightforward to substitute value sets from other countries. This would allow mapping in those countries that have both 3L and 5L value sets, or to translate reported results using one value set and variant of EQ-5D, to other value sets and the other variant. For researchers conducting analyses drawing on international evidence, this may be of particular value. Of course, the sample is drawn exclusively from the UK and was conducted in English. Replication studies using

international samples and different languages may be desirable for some decision makers.

5 Conclusion

As preference-based health-related quality-of-life instruments are refined, developed and enter into use in clinical studies, a challenge arises in maintaining consistency in decision making, that is, the need to link evidence gathered from different descriptive systems and value sets. Mapping is the general approach that allows this but in the case of EQ-5D-3L and 5L, unlike typical mappings that link from a set of clinical outcomes to the target preference-based measure, there is a need to be able to do so in both directions. The work presented here reports on a large scale, UK data collection exercise and the analysis of that data to provide such a mapping approach. The results are shown to perform well both in and out of sample and allow analysts to easily link the two variants of EQ-5D, in either direction, and from either the descriptive system responses, individual utility scores or summary statistics.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40273-022-01218-7>.

Acknowledgments The authors wish to thank all survey participants and acknowledge the work of OnePoll in conducting the survey. We are grateful for comments on previous drafts from Laura Gray, Alan Lamb and Rosie Lovett. For ongoing permission to use data collected from the FORWARD Databank we are extremely grateful to Kaleb Michaud. Any remaining errors are the responsibility of the authors. This paper has not been submitted elsewhere.

Declarations

Author contributions All authors made substantial contributions to obtaining funding, design of the study, analysis and interpretation of results. All authors were involved in drafting of the manuscript and approved the final version.

Data availability The data used to develop the model are available from the authors upon request subject to requirements for maintaining participant confidentiality.

Ethics This study was approved by the University of Sheffield Research Ethics Committee on 10/1/2020 Ref: 032503

Competing interests None.

Funding source This research is funded by the National Institute for Health Research (NIHR) Policy Research Programme, conducted through the Policy Research Unit in Economic Methods of Evaluation in Health and Social Care Interventions, PR-PRU-1217-20401. The views expressed are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the Department of Health and Social Care or its arm's length bodies, or other UK government departments.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. Pennington B, Hernández Alava M, Pudney S, Wailoo A. The impact of moving from EQ-5D-3L to -5L in NICE technology appraisals. *Pharmacoeconomics*. 2019;37(1):75–84.
2. Hernández Alava M, Wailoo A, Grimm S, Pudney S, Gomes M, Sadique Z, Meads D, O'Dwyer J, Barton G, Irvine L. EQ-5D-5L versus 3L: the impact on cost-effectiveness in the UK. *Value Health*. 2018;21(1):49–56.
3. Wailoo A, Hernández Alava M, Pudney S, Barton G, O'Dwyer J, Gomes M, Irvine L, Meads D, Sadique Z. An international comparison of EQ-5D-5L and 3L for use in cost effectiveness analysis. *Value Health*. 2021;24(4):568–74.
4. Van Hout B, Janssen MF, Feng Y, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard S. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15:708–15.
5. Hernández Alava M, Pudney S. 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. *J Health Econ*. 2017;55:139–52.
6. Craig BM, Pickard AS, Rand-Hendriksen K. Do health preferences contradict ordering of EQ-5D labels? *Qual Life Res*. 2015;24(7):1759–65.
7. Hernández Alava M, Pudney S. Mapping between EQ-5D-3L and EQ-5D-5L. A survey experiment on the validity of multi-instrument data. *Health Econ*. 2022;31:923–39.
8. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–108.
9. Devlin N, Shah K, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ 5D 5L value set for England. *Health Econ*. 2018;27(1):7–22.
10. Trivedi PK, Zimmer DM. Copula modeling: an introduction for practitioners. *Found Trends Economet*. 2005;1:1–111.
11. Wolfe F, Michaud K. The National Data Bank for rheumatic diseases: a multi-registry rheumatic disease data bank. *Rheumatology*. 2011;50:16–24.
12. University of Essex, Institute for Social and Economic Research. Understanding Society: Innovation Panel, Waves 1-13, 2008-2020. [data collection]. 11th Edition. UK Data Service. SN: 6849. 2021. <https://doi.org/10.5255/UKDA-SN-6849-14>.
13. Hernández Alava M, Pudney S, Wailoo A. Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from an English population study. Policy Research Unit in Economic Evaluation of Health and Care Interventions. Universities of Sheffield and York. 2020 Report 063.
14. Available at <https://www.sheffield.ac.uk/nice-dsu/methods-development/mapping-eq-5d-5l-3l>. Accessed 23 Nov 2022.