Work in progress - please don't cite without the authors' permission.

Mapping from EQ-5D-Y-5L to EQ-5D-Y-3L: An exploratory study of methods

Mina Bahrampour*[1], Tianxin Pan*[2], Juan Manuel Ramos-Goni[3,4], Brendan Mulhern[1], Nancy Devlin[2], Aureliano Finch[5], Mark Oppe[5], Yan Feng[6]**

1: Centre for Health Economics Research and Evaluation, University of Technology Sydney, Australia
2: Melbourne School of Population of Global Health, University of Melbourne, Australia
3: Maths In Health B.V, Kimmen, The Netherlands
4: Decision Analysis and Support Unit, SGH, Warsaw School of Economics, Poland
5: EuroQol Research Foundation, Rotterdam, The Netherlands
6: Queen Mary University of London, United Kingdom

*Joint first authors
** Corresponding author: yan.feng@qmul.ac.uk

Abstract

Background: The EQ-5D-Y-3L (Y-3L) and EQ-5D-Y-5L (Y-5L) are both available to measure generic health-related quality of life for children and adolescents. There is a gap of a mapping algorithm between the two youth-version of the EQ-5D instruments. To fill in this research gap, the EuroQol Research Foundation recently approved a project titled "Developing a function to map EQ-5D-Y-5L to EQ-5D-Y-3L". This paper reports our exploration on the statistical methods for developing a mapping algorithm between the Y-3L and Y-5L using the Australian Paediatric Multi-Instrument Comparison study data.

Methods: We explored three statistical methods, including ordered logit regression, multinomial logic regression and a non-parametric method. Regression models were estimated with and without the inclusion of a latent factor (which recognised the within-respondent correlation) and three covariates (i.e. self- /proxy-reported instrument version, age, and gender). The primary criterion for model selectin was the prediction accuracy.

Results: A total of 8,920 observations were included. To recognise the within-respondent correlation significantly improved the prediction accuracy, while the inclusion of three covariates also slightly improved the predictive ability of the models. The non-parametric model performed better than models without a latent factor, but less well than models with a latent factor.

Conclusion: The best model to map the Y-5L to Y-3L, using the Australian P-MIC data, is the ordered logit model with a latent factor and including the self- /proxy- reported variable as a covariate. Our methodological exploration is still ongoing. The preliminary results highlighted the importance of collecting primary data from children who are ill or very ill for the Y-5L to Y-3L study.

1. Introduction

The EQ-5D-Y (called Y-3L from now on) is a generic preference accompanied measure of health, designed for administration in children and adolescents aged 8 to 15 (Wille et al., 2010). It comprises of two components, a descriptive system and a value set. The descriptive system measures health using 5 dimensions, namely mobility, looking after myself, usual activities, pain/discomfort and being sad, worried, or unhappy. Each dimension is associated with three levels of severity, "no problems/ not", "some problems/ a bit" and "a lot of problems /very". The value set assigns a score that reflect adults' strength of preferences for the combination of dimensions and levels, taking a 10 years old perspective. Numerous value sets have been published for the Y-3L, including recent examples such as those for Indonesia, Hungary, and Netherlands (Fitriana et al., 2022; Rencz et al., 2022; Roudijk et al., 2022). An overview of existing value sets for the Y-3L is available in Devlin et al., (2022).

Evidence on the Y-3L shows it is generally valid and responsive (Rowen et al., 2021). Yet, as it was adapted from the adult EQ-5D-3L it may have similar issues reported by the adult instrument (Finch et al., 2017, Longworth et al., 2014). In adults, the development of a five-level version i.e., EQ-5D-5L (Herdman et al., 2011) considerably improved measurement properties (Feng et al., 2021; Buchholz et al., 2018).

The EQ-5D-Y-5L (called Y-5L from now on) was developed as an extension of the Y-3L (Kreimeier et al, 2019). The Y-5L descriptive system covers the same dimensions of the Y-3L, but with five severity levels i.e., "no/no problems", "a little bit/ a little bit of problems", "some /some problems/quite", "a lot/ a lot of problems/ really", "cannot/extreme(ly)". The generation of a youth instrument with increased granularity served two purposes. On the one hand, this was intended to ensure the transfer of the learnings from the adult to the youth measures (Kreimeier et al, 2019). On the other hand, availability of a five-level version of the youth instrument was meant to promote comparability of data collected in adults and children, as the EQ-5D-5L is more frequently used than the EQ-5D-3L.

To facilitate the uptake of Y-5L on launch, or shortly after, a scoring approach that enables the estimation of Y-5L utility values is in need. One way to fill in this gap is to develop a mapping algorithm between the Y-3L and Y-5L (Longworth and Rowen, 2013). The mapping algorithm would enable the large suite of recently developed Y-3L value sets being meaningfully used while allowing time for the development of a protocol for the direct valuation of the Y-5L.

The Executive Committee of the EuroQol Research Foundation recently approved a study (called "Y-5L to Y-3L" study from now on). It aims to (1) examine different mapping approaches between the Y-3L and the Y-5L and (2) develop a mapping function that can be used internationally for deriving health utilities from the Y-5L (reference 1650-RA). This paper is part of the funded project. It focused on aim (1) of the project. To be specific, this paper aims to explore different statistical methods for developing the mapping algorithm between the Y-3L and Y-5L using the Australian Paediatric Multi-Instrument Comparison study data.

2. Methods

2.1 Data

Our study used data from the Australian Paediatric Multi-Instrument Comparison (P-MIC) study. The P-MIC study is part of a wider research programme titled "Quality of Life in Kids: Key Evidence for Decision Makers in Australia" or QUOKKA. A detailed description of the P-MIC data collection is reported in Jones et al (2023). The P-MIC study includes three key samples of children, i.e. (1) a sample recruited via general or specialised hospital service, (2) a sample of general population recruited via online panel, and (3) a sample of children with 11 health conditions recruited primarily via online panels. Children aged between 5 and 18 years old from all three samples were asked to complete the Y-3L and Y-5L in the same appointment. The order of the two instruments presented to the study participants was randomised. Children aged between 7 and 18 years were invited to self-report their own health. The proxy version of the Y-3L and Y-5L was applied if a child was not able to report their health themselves, either due to age (younger than 7 years) or health problems. The EQ Visual Analogue Scale (EQ-VAS) score was completed once, alongside the first EQ-5D youth instrument presented to participants. All participants received a follow-up survey either at 2 days (children who were recruited through general hospital service in sample 1) or 4-weeks (the rest of the participants) after completing the initial survey. The P-MIC study also collected data using other HRQoL instruments, however, those data were not the focus of our study.

2.2 Statistical Methods

For developing a mapping algorithm between the Y-5L and Y-3L, in this study, we explored three statistical methods. It includes two regression-based methods and a non-parametric method. We selected the best model primarily based on the prediction accuracy, while other criteria were also considered such as theoretical arguments and practical concerns. Also, by applying the best model, we compared the observed 3L utilities and the predicted 5L utilities across sub-samples defined by participants' gender, age, EQ-VAS sub-sets, EQ-5D-3L utility sub-sets, and instrument version (proxy- versus self- report).

Regression-based methods

Ordered logit regression (van Hout and Shaw, 2021) and multinomial logit regression (Gray et al., 2006) were widely applied in mapping studies to EQ-5D. Both methods are suitable for modelling categorial variables. As our dependent variable, the Y-3L is an ordinal variable, the ordered logit regression would be our "natural" choice for modelling. However, the method relies on parallel regression assumption (Long, 1997, p140). It means, for each dimension of the Y-3L, the ordered logit regression generates two binary response models for the three ordered EQ-5D level response. The assumption asks the coefficients for the independent variables in those binary response models to be the same. This is a strong assumption which is not required by the multinomial logit regression method. As an alternative approach, we also fitted in our data using the multinomial logit regression. It should be noted that this approach does not recognise the ordered nature of the Y-3L data. For both methods, our aim was to predict the probabilities of responses in Y-3L from the Y-5L. We used the Stata command *mlogit* and *ologit* for the multinomial and ordered logit regressions respectively.

For both methods, we started with a 20-parameter model where independent variables took the 20 Y-5L dummies (levels 2 to 5 for each of the five Y-5L dimensions). We also explored three alternative specifications for each method: (1) included instrument version (proxy- versus self- report) as a covariate, (2) recognised the within-respondent correlation through a latent factor model i.e., allowing the correlation between responses from the five Y-3L dimensions by the same respondent (Stata command *gsem*), (3) allowed the within-respondent correlation as well as including instrument version (proxy- versus self- report) as a covariate. Effectively, we ran the following eight model specifications:

- M1: 20-parameter ordered logit model
- M2: 20-parameter ordered logit model + self- /proxy- reported
- M3: 20-parameter ordered logit model + latent factor
- M4: 20-parameter ordered logit model + self- /proxy- reported + latent factor
- M5: 20-parameter multinomial logit model
- M6: 20-parameter multinomial logit model + self- /proxy- reported
- M7: 20-parameter multinomial logit model + latent factor
- M8: 20-parameter multinomial logit model + self- /proxy- reported + latent factor

In the mapping literature around the EQ-5D instruments, it is inconclusive on what covariate(s) to include and how they should be included. Based on the published EQ-5D mapping algorithms for adult population, age and gender were widely used as candidates for covariates if a study considered "main effects" and covariates in model specifications (Abdin et la., 2018, Peak et al., 2018, Siani et al., 2016). There is little empirical evidence on whether and how to manage covariates in the EQ-5D mapping studies for child population. In addition to address the instrument version variable as presented in M2, M4, M6 and M8, we explored one more specification by including age and gender as the covariates:

- The best performed specification between M1, M3, M5, M7 + age and gender

To predict the Y-5L utility, we applied the expected value method (Le and Doctor, 2011). Effectively, we multiplied the predicted probability of being in each of the response levels of the Y-3L by the corresponding value of the interim Australian Y-3L value set (Pan et al., 2024).

Non-parametric method

The non-parametric method was developed in the EQ-5D-5L to EQ-5D-3L mapping study (van Hout et al., 2012). We applied the same approach in this paper. For each health state described by the Y-5L (n = 3,125), we calculated its probability of reporting each of the 243 Y-3L states. A 3125x243 transition probability matrix was achieved by cross tabulating responses between the Y-3L and Y-5L. The predicted utility for each Y-5L state was the sum of the 243 weighted Y-3L utilities. We calculated the utilities for the 243 Y-3L states by applying the interim Australian Y-3L value set. For a given Y-5L state, the weight for each Y-3L state was determined by the transition probability.

Model performance

To assess the prediction accuracy for each model, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSA) were calculated. Essentially, we calculated the difference between the predicted Y-5L utilities (with our mapping algorithm applied) and the observed utilities based on the parallel Y-3L states (with the interim Australian Y-3L value set applied). We also reported model fit, measured by the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), for each model. The primary criterion to select the best model was the prediction accuracy, while other criteria were also considered such as theoretical arguments and practical concerns.

After selecting the best model, we calculated the observed Y-3L utilities using the interim Australian Y-3L value set and the predicted Y-5L utilities using our mapping algorithm for the full sample. Observed and predicted utilities were compared across sub-samples defined by instrument version (proxy- versus self- report), gender (female, male, and other), age groups (5-7 years, 8-11 years, and 12-18 years), EQ-VAS sub-sets (0-25, 25-50, 50-75, 75-100), and observed Y-3L utility sub-sets (0-0.25, 0.25-0.50, 0.50-0.75, 0.75-1).

Data analyses were performed in Stata18 (Stata Corp LP, College Station, TX, USA) and Microsoft Excel.

3. Results

3.1 Descriptives Analysis

Tabel 1 reports the summary statistics of the data used in this paper. The total sample includes 8,920 observations with complete information on the Y-3L and Y-5L profiles data. The average age of the sample was 10.82 (± 3.91) years old with 52.3% male. The sample reported average Y-3L utility and EQ-VAS as 0.865 and 76.63 respectively. There were 187 unique Y-3L states and 794 unique Y-5L states been used. It represents 77% (= 187/243) and 25% (= 794/3125) of the total available Y-3L and Y-5L states respectively. Just under 2/3 of the sample were self-reported (62.53%). The largest data source was the sample with 11 health conditions (61.72%), followed by the sample from the general population (23.43%), and the sample recruited via hospital service (14.85%). More than 70% of the sample came from the baseline. The summary statistics for the baseline data alone are presented in the third column.

Table 1: Summary statistics

Variable names	N (%) or mean (SD)	
	Full sample (n = 8,920)	Wave 1 (n = 6,336)
Age (mean)	10.82 (3.91)	10.84 (3.92)
Gender		
Male	4665 (52.30%)	3307 (52.19%)
Female	4128 (46.28%)	2935 (46.32%)
Transgender female	24 (0.27%)	18 (0.28%)
Transgender male	40 (0.45%)	31 (0.49%)
Not described (please specify)	42 (0.47%)	27 (0.43%)
Prefer not to answer	21 (0.24%)	18 (0.28%)
EQ-5D-3L utilities (mean)	0.865 (0.151)	0.863 (0.151)
EQ-VAS (mean)	76.63 (19.26)	76.60 (19.31)
N of unique 3L states reported	187	175
N of unique 5L states reported	794	664
Self-reported version	5578 (62.53%)	4099 (64.69%)
Proxy version	3342 (37.47%)	2237 (35.31%)
Sample 1 (collected from hospitals)	1325 (14.85%)	848 (13.38%)
Sample 2 (general population)	2090 (23.43%)	1561 (24.64%)
Sample 3 (with health conditions)	5505 (61.72%)	3927 (61.98%)
Wave one (baseline)	6336 (71.03%)	
Wave two (follow-up)	2584 (28.97%)	

Cross-tabulation of participants' responses to the Y-3L and Y-5L is shown in **Table 2**. Most majority of the participants reported no problem with the "mobility" dimension on the Y-3L and no or little bit of problems on the Y-5L (88%), followed by the "looking after self" dimension (78%), "usual activity" dimension (72%), "pain/discomfort" dimension (64%), and "feeling worries, sad and unhappy" dimension (49%). A very small proportion of the respondents reported a lot of problem on the Y-3L and a lot of problems or cannot/extreme on the Y-5L, from the largest of 3.46% on the "feeling worries, sad and unhappy" dimension to the smallest of 1.43% on the "mobility" dimension.

Different methods have been employed in existing literature to define logically consistent responses (van Hout et al., 2012; Janssen et al., 2008). Further information regarding the inconsistencies in P-MIC data can be found in Bahrampour et al (2024).

Table 2: Cross-tabulation of EQ-5D-Y-3L and EQ-5D-Y-5L responses by dimensions and levels

	EQ-5D-Y-5L				
EQ-5D-Y-3L					
Mobility	No problems	A little bit of problems	Some problems	A lot of problems of	Cannot
No problems	7,541	308	71	10	3
Some problems	157	377	223	50	1
A lot of problems	14	7	30	72	56
Looking after self	No problems	A little bit of problems	Some problems	A lot of problems of	Cannot
No problems	6,582	395	59	18	7
Some problems	183	775	432	85	7
A lot of problems	9	13	51	151	153
Usual Activities	No problems	A little bit of problems	Some problems	A lot of problems of	Cannot
No problems	5,776	614	133	25	12
Some problems	406	833	491	140	31
A lot of problems	17	51	111	178	102
Pain/Discomfort	No	A little bit	Some	A lot of	Extreme
No	4,999	716	97	13	7
Some	432	1597	677	129	5
A lot of	14	17	59	118	40
Feeling worries, sad and unhappy	Not	A little bit	Quite	Really	Extremely
Not	3,349	1014	75	21	6
A bit	566	2511	614	188	40
Very	12	72	143	165	144

3.2 Model performance

Table 3 presents the model performance for each of the nine models (M1-M9) by five statistics, including the mean predicted Y-5L utilities, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Table 3: Model performance between nine models (M1-M9)

Models	Mean utility	MAE	RMSE	AIC	BIC
Ordered logit models (OL)					
1: OL	0.86391	0.06201	0.08853	33811.58	34592.14
2: OL + self- /proxy- reported	0.86392	0.06199	0.08848	33774.59	34590.63
3: OL + latent factor	0.86785	0.03997	0.05778	32803.95	33619.99
4: OL + self- /proxy- reported + latent factor	**0.86787**	**0.03980**	**0.05754**	**32755.15**	**33606.68**
Multinomial logit model (MNL)					
5: MNL	0.86463	0.06188	0.08822	33479.41	34969.58
6: MNL + self- /proxy- reported	0.86463	0.06181	0.08809	33436.82	34997.95
7: MNL + latent factor	0.86748	0.04072	0.05861	32392.21	33953.34
8: MNL + self- /proxy- reported + latent factor	0.86749	0.04056	0.05839	32340.25	33972.34
Non-parametric model					
9: Non-parametric model	0.86463	0.05423	0.07806		

Note: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC). M4 is the best model which IS highlighted in bold text.

As reported in **Table 1**, the mean observed Y-3L utilities of the full sample was 0.8646 (±0.151). Between the nine models, the mean predicted Y-5L utilities based on the non-parametric method was the closest to the mean observed Y-3L utilities.

We compared performance between the nine models by looking at the MAE and RMSE statistics. Lower MAE and RMSE values suggest better prediction accuracy. Between the four ordered logit models (M1-M4), both MAE and RMSE were slightly improved when we included the self- /proxy- reported variable as a covariate, as shown in M2 (in comparison to M1) and M4 (in comparison to M3). Accounting for the within-respondent correlation by adding a latent factor as in M3 and M4, the MAE and RMSE significantly decreased compared to M1 and M2. A similar pattern was observed in the four multinominal logistic models (M5-M8).

The two ordered logit models without latent factor reported slightly higher MAE and RMSE values in comparison to their corresponding multinominal logit models (i.e. M1 versus M5, M2 versus M6). When we took latent factor into account, the two ordered logit models reported slight lower MAE and RMSE values than their corresponding multinominal logit models (i.e. M3 versus M7, M4 versus M8). The MAE and RMSE values reported from the non-parametric model were lower than those of the ordered logit and multinomial logit models without latent factor (i.e. M1, M2, M5, and M6), but higher than specifications with latent factor considered (i.e. M3, M4, M7 and M8). Between the nine models, the 20-parameter ordered logit model with the inclusion of self- /proxy-reported variable as a covariate and allowing for a latent factor (M4) reported the lowest MAE and RMSE, demonstrating superior predictive ability. The full results of the best performed model M4 are presented in **Table 4**.

Furthermore, we explored a specification by including age and gender as covariates in the best performed model between M1, M3, M5 and M7. As reported in **Table 3**, M3 reported the best prediction accuracy between the four models. Including age and gender as covariates in M3, we observed slight improvements in the MAE and RMSE compared to our best model M4.

Figure 1 presents the plots of the observed Y-3L states utilities against the predicted utilities based on the parallel Y-5L states (one figure for each of the nine models from M1-M9). Closer alignment to the red dash line suggests better prediction accuracy. **Figure 1** suggested that for both the ordered logit and multinomial logit models, the two models that recognised the within-respondent correlation reported better prediction accuracy compared to the other two models that did not consider this correlation.

Table 5 reported the comparison between observed Y-3L utilities and predicted Y-5L utilities across sub-samples defined by instrument version, gender, age groups, EQ-VAS sub-sets, and observed Y-3L utility sub-sets. While looking at the mean difference across sub-samples defined by instrument version, gender, and age groups, the predicted mean utility was consistently higher than the observed mean utility. The magnitude of the mean difference was small (0.000 to 0.011) and stable (between sub-samples). Comparing predicted and observed utilities across EQ-VAS and Y-3L utility sub-sets showed consistent overestimate of poor health. Furthermore, the level of mean difference was far greater at the poor health end. To define sub-samples by EQ-VAS range, we observed the largest mean difference for sub-set between 0 and 25 (= -0.042), followed by sub-set between 25 and 50 (= -0.019), sub-set between 50 and 75 (= -0.008), and sub-set between 75 and 100 (= 0.003). Similar pattern was observed between the Y-3L utility sub-sets. As far as we know, there is no evidence on the minimal clinically important difference (MCID) for the Y-3L. We are therefore unable to comment on the mean differences in comparing to the MCID. The MAE values were reported between 0.034 and 0.119, while the RMSE values were in a range from 0.047 to 0.133.

Table 4: 20-parameter ordered logit model with self-reported/proxy as a covariate and allowed for a latent factor

	Mobility		Self-Care		Usual Activity		Pain/Discomfort		Anxiety/Depression	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
MO2	3.850***	(0.169)	0.444***	(0.140)	0.664***	(0.129)	0.607***	(0.125)	-0.157	(0.115)
MO3	4.713***	(0.232)	0.315	(0.192)	0.853***	(0.186)	0.747***	(0.187)	-0.622***	(0.174)
MO4	7.139***	(0.365)	-0.006	(0.284)	0.726**	(0.295)	0.971***	(0.286)	-0.599**	(0.267)
MO5	9.943***	(0.708)	0.393	(0.560)	1.660***	(0.530)	2.085***	(0.461)	-0.299	(0.422)
SC2	0.638***	(0.157)	4.377***	(0.133)	1.042***	(0.111)	0.007	(0.108)	-0.025	(0.092)
SC3	0.689***	(0.196)	5.632***	(0.189)	1.176***	(0.152)	-0.128	(0.158)	0.133	(0.137)
SC4	0.642**	(0.257)	8.184***	(0.269)	1.804***	(0.217)	0.116	(0.222)	-0.029	(0.198)
SC5	2.117***	(0.328)	10.561***	(0.429)	2.796***	(0.309)	0.060	(0.299)	0.004	(0.267)
UA2	1.016***	(0.160)	0.669***	(0.119)	3.048***	(0.132)	0.704***	(0.101)	0.732***	(0.088)
UA3	0.990***	(0.200)	0.778***	(0.158)	3.991***	(0.180)	0.502***	(0.146)	0.829***	(0.128)
UA4	1.147***	(0.258)	0.954***	(0.218)	5.615***	(0.255)	0.681***	(0.205)	1.268***	(0.188)
UA5	0.656*	(0.344)	0.194	(0.314)	6.334***	(0.343)	0.894***	(0.290)	1.216***	(0.266)
PD2	1.041***	(0.153)	0.064	(0.110)	0.468***	(0.098)	3.606***	(0.109)	0.218***	(0.072)
PD3	1.351***	(0.187)	-0.054	(0.152)	0.379***	(0.136)	4.763***	(0.163)	0.429***	(0.113)
PD4	1.539***	(0.256)	-0.068	(0.237)	0.599***	(0.217)	7.043***	(0.256)	0.131	(0.193)
PD5	1.766***	(0.488)	0.166	(0.460)	0.343	(0.451)	8.192***	(0.477)	0.256	(0.402)
AD2	-0.197	(0.144)	0.117	(0.108)	0.669***	(0.101)	0.293***	(0.083)	2.986***	(0.086)
AD3	-0.380**	(0.193)	0.137	(0.156)	0.953***	(0.142)	0.222*	(0.130)	4.609***	(0.144)
AD4	-0.585**	(0.249)	0.030	(0.204)	1.303***	(0.187)	0.258	(0.176)	6.022***	(0.190)
AD5	0.280	(0.296)	0.316	(0.267)	1.775***	(0.245)	0.347	(0.234)	7.764***	(0.264)
SELF/PROXY	-0.184	(0.129)	0.353***	(0.093)	0.235***	(0.087)	-0.400***	(0.079)	-0.079	(0.064)
ID	1.000	(0.000)	0.729***	(0.077)	1.108***	(0.116)	0.859***	(0.084)	0.755***	(0.078)
VAR(ID)	1.862***	(0.282)								

Note: Standard errors (SE) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

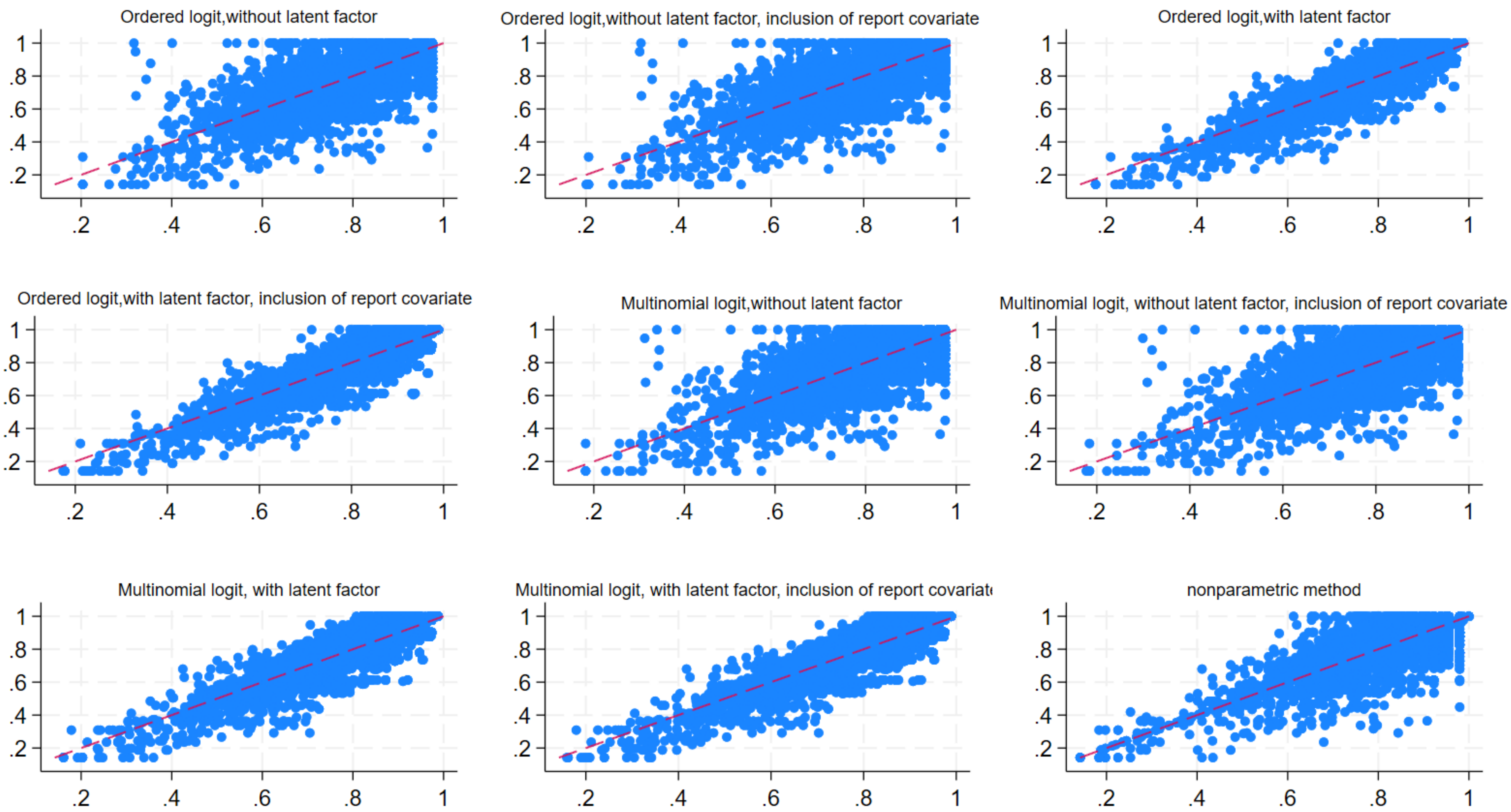

Figure 1: Predicted Y-5L states utilities (base on M1-M9) against the observed utilities based on the paralleled Y-3L states. Y axis is the observed Y-3L states utilities. X axis is the predicted Y-5L states utilities.

Table 5: Compare the observed and predicted utilities for sub-samples using the best performed model M4

	N	Observed mean (SD)	Predicted mean (SD)	Mean difference (Observed-Predicted)	MAE	RMSE
Proxy	3342	0.87 (0.15)	0.873 (0.132)	-0.003	0.038	0.055
Self-reported	5578	0.861 (0.152)	0.865 (0.129)	-0.003	0.041	0.059
Male	4665	0.873 (0.147)	0.875 (0.127)	-0.002	0.038	0.055
Female	4128	0.858 (0.153)	0.863 (0.131)	-0.005	0.041	0.059
Other gender types	127	0.758 (0.216)	0.769 (0.186)	-0.011	0.056	0.078
5-7 years old	2242	0.887 (0.135)	0.887 (0.118)	0.000	0.037	0.052
8-11 years old	2859	0.873 (0.14)	0.875 (0.12)	-0.002	0.039	0.057
12-18 years old	3819	0.846 (0.166)	0.851 (0.142)	-0.006	0.042	0.060
VAS [0-25)	206	0.559 (0.212)	0.601 (0.183)	-0.042	0.068	0.091
VAS [25-50)	1163	0.728 (0.173)	0.747 (0.149)	-0.019	0.053	0.074
VAS [50-75)	1887	0.814 (0.139)	0.822 (0.123)	-0.008	0.046	0.065
VAS [75-100]	5664	0.921 (0.103)	0.918 (0.085)	0.003	0.034	0.049
Y-3L [0-0.25)	42	0.194 (0.042)	0.313 (0.08)	-0.119	0.119	0.133
Y-3L [0.25-0.5)	229	0.408 (0.065)	0.5 (0.097)	-0.092	0.100	0.129
Y-3L [0.5-0.75)	1383	0.659 (0.071)	0.695 (0.086)	-0.036	0.060	0.081
Y-3L [0.75-1]	7266	0.922 (0.081)	0.916 (0.072)	0.006	0.034	0.047

Note: Standard Deviation (SD), Mean Absolute Error (MAE), Root Mean Square Error (RMSE)

4. Discussion

Summary of findings

The primary aim of this paper was to explore different statistical methods for developing a mapping algorithm between the Y-5L and Y-3L using the P-MIC data. The findings will be used to inform the development of a mapping function which can be used internationally for deriving health utilities from the Y-5L data. We applied three statistical methods in this study including the ordered logit regression, multinomial logit regression, and a non-parametric method. We also explored the impacts of including three covariates (i.e. self- /proxy- reported, age, and gender) and a latent factor (which recognised the within-respondent correlation) in the modelling.

The primary criterion for model selectin was the prediction accuracy. Our results suggest that the predictive ability was similar between the four ordered logit models and their corresponding multinomial logit models. To recognise the within-respondent correlation by including a latent factor significantly improved the prediction accuracy of the ordered logit and multinomial logit models. To include (1) self- /proxy- reported version variable or (2) age and gender as covariates slightly improved models' prediction accuracy. The non-parametric model performed better than models without a latent factor, and less well than models with a latent factor. The best model is the ordered logit model with a latent factor and including the self- /proxy- reported version variable as a covariate. The model showed overestimate of poor health.

Next steps

This study is working in progress. In the next step, we will complete the following tasks:

- The P-MIC dataset includes two waves of data. In the current study, we pooled the baseline and follow-up data together without recognising the repeated nature of the data. In the next step, we will explore the repeated responses by applying for a respondent random effect, for example, while analysing each dimension of the Y-3L.
- In the current study, we calculated the prediction accuracy for each model using the *full sample*. In the next step, we will perform cross-validation by randomly splitting our full sample into estimation sample and validation sample. We will fit our models to the estimation sample, and then apply the estimated coefficients that derived from the estimation sample to the validation sample. To account for variability, we will perform multiple rounds of cross validation. Average values for model fit and prediction accuracy from the repeated validation process will be calculated model by model. The primary criterion for model selection will be the prediction accuracy base on the *validation sample*.
- The copula model with mixture marginal was developed in recently years in a mapping study (Hernandez-Alava and Putney, 2017). The study aimed to map EQ-5D-3L to EQ-5D-5L. We will explore this method using the P-MIC data.

Three discussion points

The mapping algorithms to link EQ-5D-5L and EQ-5D-3L responses were published in van Hout et al., (2012) and van Hout and Shaw (2021). In the 3L to 5L study, authors compared two statistical methods, i.e. the non-parametric method and ordinal logistic regression. The recommended approach was based on an *ordinal logistic regression*. In the 5L to 3L study, authors selected the best model from four statistical methods including the linear regression, non-parametric statistics, ordered logistic regression, and item-response theory. The *non-*

parametric model was preferred. In our study, we compared the performance of three statistical methods including the non-parametric method, ordered logit model, and multinomial logit model. Our best model so far was based on the *ordered logit model*. Furthermore, van Hout and Shaw (2021) reported the effect of including *a latent factor* in the ordered logit model. They found the inclusion of a latent factor lowered the AIC and slightly improved predictive accuracy, while our finding suggested that the ordered logit and multinomial logit models with a latent factor significantly improved the prediction accuracy and the AIC. Finally, van Hout and Shaw (2021) explored the impacts of including age and gender as covariates in the ordered logit model, and suggested that the inclusion lowered the AIC but without improving predictions. Our best model does not include age and gender as covariates but for a different reason (see the next discussion point).

In this study, we explored the impacts of including three covariates in the mapping function, including age and gender of children and the instrument version (self- or proxy- reported version). Our results suggest that including those variables as covariates could slightly improve the model prediction accuracy. However, in our best model (M4), we only included the self-/proxy- reported version variable as a binary covariate. We argue that to include age and gender as covariates might potentially limit the use of our mapping algorithm, as it cannot be applied to Y-5L data alone for calculating utility values without age and gender. However, there is a ground to consider instrument version as a covariate. As the underling relationship between the Y-3L and Y-5L might be systematically different between these two types of responses. Proxies (usually be parents/carers) are used when it is not feasible to have children directly self-report their health. This could be due to language and communication limitations (e.g. a child with learning or behavioural disorders), difficulties with understanding abstract concepts (e.g. a child with neurocognitive conditions), ethical considerations (e.g. a child being too ill to self-report health), and the age of the child (under 7 years old). Studies showed discrepancies in self- versus proxy- reported health problems (Bahrampour et al., 2024; Khanna et al., 2022). In the next step of our study, we could explore different formats of the self- /proxy- reported version variable in our model, such as interacting with Y-5L variables. To keep the simplicity of our specification, the model with a binary instrument version variable (proxy- versus self- report) included is preferred.

None of the eight regression-based analyses in this paper reported convergence problem. However, convergence failure was an issue emerged in the ordered logit and multinomial logit models when we excluded logically inconsistent observations from the full sample (approaches as defined by either van Hout et al. (2012) or Janssen et al (2008)). This leads to two discussion points. First, in this paper, all analyses were based on the full sample without excluding logically inconsistent responses. It is arguable on whether we should exclude logically inconsistent observations as they could be treated as random errors. Furthermore, the two approaches used to define logically inconsistent responses for the adult instruments may not be directly applicable to the youth instruments (van Hout et al., 2012; Janssen et al., 2008). This is because level 4 of Y-5L is descriptively the same as the worst level (level 3) on Y-3L for four dimensions. An implication is that children or proxies might consider level 4 in the Y-5L as the worst possible level for those dimensions. Second, the P-MIC dataset reported small proportion of participants with severe problems. As shown in **Table 2**, between 1.43% and 3.46% of the participants reported a lot of problem on the Y-3L and a lot of problems or cannot/extreme on the Y-5L across the five dimensions. This data feature raised concern of the potential modelling problems for the Y-5L to Y-3L study, including issues like non-convergence, unstable estimates. To fill in the potential data "gaps", the EuroQol Research Foundation recently approved a primary data collection project. The project aims to collect parallel Y-5L and Y-3L responses from children who are unwell or very unwell. Our study team received permission to access to the primary data for our Y-5L to Y-3L study.

5. Conclusion

The best model to map the Y-5L to Y-3L, using the Australian P-MIC data, is the ordered logit model with a latent factor and including the self- /proxy- reported variable as a covariate. This methodological work is ongoing. Our preliminary results highlighted the importance of collecting primary data from children who are ill or very ill for the Y-5L to Y-3L study.

Acknowledgements

The “Developing a function to map EQ-5D-Y-5L to EQ-5D-Y-3L” study is funded by the EuroQol Research Foundation (reference 1650-RA). The “P-MIC extension” study was funded by the EuroQol Research Foundation (reference 361-RA). The views expressed are those of the authors and not necessarily those of the funder.

Reference

Abdin, E., Chong, S. A., Seow, E., Verma, S., Tan, K. B., Subramaniam, M. (2018). Mapping the Positive and Negative Syndrome Scale scores to EQ-5D-5L and SF-6D utility scores in patients with schizophrenia. Qual Life Res, 28 (1), 177-186.

Bahrampour, M., et al., A Comparison of the Psychometric Properties of the EQ-5D-Y-3L and EQ-5D-Y-5L Using Paediatric Multi-Instrument Comparison (P-MIC) Study Data. PharmacoEconomics, 2024. 42(1): p. 95-111.

Buchholz I, Janssen MF, Kohlmann T, Feng YS. A Systematic Review of Studies Comparing the Measurement Properties of the Three-Level and Five-Level Versions of the EQ-5D. Pharmacoeconomics. 2018 Jun;36(6):645-661. doi: 10.1007/s40273-018-0642-5.

Devlin, N., et al., Valuing EQ-5D-Y: the current state of play. Health and Quality of Life Outcomes, 2022. 20(1): p. 105.

Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res. 2021 Mar;30(3):647-673. doi: 10.1007/s11136-020-02688-y.

Finch AP, Brazier JE, Mukuria C, Bjorner JB. An Exploratory Study on Using Principal-Component Analysis and Confirmatory Factor Analysis to Identify Bolt-On Dimensions: The EQ-5D Case Study. Value Health. 2017 Dec;20(10):1362-1375. doi: 10.1016/j.jval.2017.06.002.

Fitriana TS, Roudijk B, Purba FD, Busschbach JJV, Stolk E. Estimating an EQ-5D-Y-3L Value Set for Indonesia by Mapping the DCE onto TTO Values. Pharmacoeconomics. 2022 Dec;40(Suppl 2):157-167. doi: 10.1007/s40273-022-01210-1.

Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. Med Decis Making. 2006;26(1):18-29. doi: 10.1177/0272989X05284108. PMID: 16495197.

Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011 Dec;20(10):1727-36. doi: 10.1007/s11136-011-9903-x.

Hernández-Alava, M. and S. Pudney, 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, 2017. 55: p. 139-152.

Janssen, M.F., et al., Comparing the standard EQ-5D three-level system with a five-level version. Value in Health, 2008. 11(2): p. 275-284.

Jones, R., et al., Australian Paediatric Multi-Instrument Comparison (P-MIC) Study: Technical Methods Paper 2023: The University of Melbourne, Australia.

Khanna, D., et al., Are We Agreed? Self- Versus Proxy-Reporting of Paediatric Health-Related Quality of Life (HRQoL) Using Generic Preference-Based Measures: A Systematic Review and Meta-Analysis. Pharmacoeconomics, 2022. 40(11): p. 1043-1067.

Kreimeier, S., et al., EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. Qual Life Res, 2019. 28(7): p. 1951-1961.

Le, Q. A., & Doctor, J. N. (2011). Probabilistic mapping of descriptive health status responses onto health state utilities using bayesian networks: An empirical analysis converting SF-12 into EQ-5D utility index in a national US sample. Medical Care, 49(5), 451–460.

Long, J. S. (1997). Regression models for categorical and limited dependent variables. Thousand Oaks, CA: Sage Publications, Inc.

Longworth, L. and D. Rowen, Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. Value in Health, 2013. 16(1): p. 202-210.

Longworth L, Yang Y, Young T, Mulhern B, Hernández Alava M, Mukuria C, Rowen D, Tosh J, Tsuchiya A, Evans P, Devianee Keetharuth A, Brazier J. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. Health Technol Assess. 2014 Feb;18(9):1-224. doi: 10.3310/hta18090.

Pan, Tianxin., et al., An Australian Value Set for the EQ-5D-Y-3L. 2024. Under peer review.

Peak, J., Goranitis, I., Day, E., Copello, A., Freemantle, N., Frew, E. (2018). Predicting health-related quality of life (EQ-5D-5 L) and capability wellbeing (ICECAP-A) in the context of opiate dependence using routine clinical outcome measures: CORE-OM, LDQ and TOP. Health Qual Life Outcomes, 16:106.

Rencz F, Ruzsa G, Bató A, Yang Z, Finch AP, Brodszky V. Value Set for the EQ-5D-Y-3L in Hungary. Pharmacoeconomics. 2022 Dec;40(Suppl 2):205-215. doi: 10.1007/s40273-022-01190-2.

Roudijk B, Sajjad A, Essers B, Lipman S, Stalmeier P, Finch AP. A Value Set for the EQ-5D-Y-3L in the Netherlands. Pharmacoeconomics. 2022 Dec;40(Suppl 2):193-203. doi: 10.1007/s40273-022-01192-0.

Rowen D, Keetharuth AD, Poku E, Wong R, Pennington B, Wailoo A. A Review of the Psychometric Performance of Selected Child and Adolescent Preference-Based Measures Used to Produce Utilities for Child and Adolescent Health. Value Health. 2021 Mar;24(3):443-460. doi: 10.1016/j.jval.2020.09.012.

Siani, C., de Peretti, C., Millier, A., Boyer, L., & Toumi, M. (2016). Predictive models to estimate utility from clinical questionnaires in schizophrenia: findings from EuroSC. Qual Life Res, 25(4), 925-934.

van Hout, B.A. and J.W. Shaw, Mapping EQ-5D-3L to EQ-5D-5L. Value in Health, 2021. 24(9): p. 1285-1293.

van Hout, B., et al., Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health, 2012. 15(5): p. 708-15.

Wille, N., et al., Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual Life Res, 2010. 19(6): p. 875-86.