Comparing Crosswalks for Mapping from EQ-5D-3L to EQ-5D-5L in Two Patient Cohorts

PCR170, Abstract #126262

Shaw J[1] , Bennett B[2] , Khan I[3] , van Hout BA[4] , Bryning S[5*] , Fox A[5] , Hirst A[5]

[1] Bristol Myers Squibb, Lawrenceville, NJ, USA, [2] Bristol Myers Squibb, Uxbridge, UK, [3] University of Warwick, UK, [4] University of Sheffield, Sheffield, UK, [5]Adelphi Values PROVE, Bollington, UK

*University of Liverpool, Liverpool, UK

Objectives

- The EQ-5D-5L and EQ-5D-3L are questionnaires used to measure health-related quality of life (HRQoL). Several HTA bodies have begun to demand the use of EQ-5D-5L utilities in submissions.[1] Due to this requirement and in situations where only EQ-5D-3L data may be collected, a crosswalk that accurately predicts EQ-5D-5L utility values from observed EQ-5D-3L values is needed.
- Two crosswalks that map EQ-5D-3L responses to EQ-5D-5L utility values have recently been published. One was developed by the NICE Decision Support Unit (DSU) (Hernandez-Alava & Pudney) and another more recently developed by members of the EuroQoL Group (van Hout & Shaw).[2,3]
- The objective of this study was to assess the performance of the two crosswalks for predicting EQ-5D-5L utilities and associated quality-adjusted life years (QALYs) from EQ-5D-3L data using two independent datasets. Predicted utilities and QALYs derived from EQ-5D-3L data were compared against observed utilities and QALYs derived from EQ-5D-5L data.

Methods

Dataset description

The research employed two datasets:

- A prospective non-interventional cohort study of patients (N=100) in the UK with non-small cell lung cancer (NSCLC) over a minimum period of 12 months. Both versions of the EQ-5D were administered monthly with the EQ-5D-5L completed at least one week after the EQ-5D-3L to avoid the potential for repetition bias or anchoring effects.
- A multicenter randomized controlled trial of dental patients (N=168) in the UK, assessing the effectiveness of pulpotomy compared to orthograde root canal treatment on the tooth pulp response. The study had a follow-up duration of 24-months. Both versions of the EQ-5D were measured at baseline, Days 1-7, 6 and 12 months post randomization, with the EQ-5D-3L completed first followed by the EQ-5D-5L. The EQ-5D was completed in a hospital setting unless completed at home during Days 2-7.

Overview of crosswalks

- Both crosswalks differ in their statistical approaches and the datasets in which they were developed. The EuroQoL crosswalk is based on an ordinal logistic regression accounting for unobserved heterogeneity using a latent variable and used a collection of datasets across eight disease populations.[3] The DSU crosswalk incorporates a copula-based mixture model including age and gender as predictors and was originally developed using a dataset of patients with rheumatoid arthritis.[2]
- Although the EuroQoL crosswalk has the advantage that several different value sets can be applied, the English value set was used to ensure comparability as the DSU crosswalk can only be used to map to the UK value set.

Methods

EuroQoL ordinal logistic regression

- The EuroQoL crosswalk in the base-case analysis corresponds to an ordinal logistic regression with the 5L dimension scores predicted by the 3L scores on the same dimension (EQ(2) in Table 1). The crosswalk was implemented using publicly available code provided by EuroQoL in their reverse crosswalk mapping tool.[4] In addition, eight alternative crosswalks were implemented, with the features defined in Table 1.

Table 1. Defined features of the alternative crosswalks

	Predicted on same and complementary dimensions	Predicted on age	Predicted on sex	Predicted on age^2	Predicted using a latent variable
EQ(1)	Neither	No	No	No	No
EQ(2)	Same only	No	No	No	No
EQ(3)	Yes	No	No	No	No
EQ(4)	Yes	Yes	Yes	No	No
EQ(5)	Yes	Yes	Yes	Yes	No
EQ(6)	Same only	Yes	No	No	Yes
EQ(7)	Yes	Yes	No	No	Yes
EQ(8)	Yes	Yes	Yes	No	Yes
EQ(9)	Yes	Yes	Yes	Yes	Yes

- The crosswalk recommended by van Hout & Shaw for use in applications is an ordinal logistic regression with dummy variables for complementary dimension, excluding age and sex, and a latent factor accounting for unobserved heterogeneity (EQ(7)).[3] However, as mentioned, EQ(2) was chosen as the base-case to ensure comparability with the DSU crosswalk, as it is the only EuroQoL crosswalk that is publicly available.

DSU copula approach

- Hernandez-Alava & Pudney developed a flexible bidirectional crosswalk that can be used to predict either 3L responses from 5L responses or 5L responses from 3L responses and did not impose necessarily strong assumptions.[2]
- For this research, the crosswalk was implemented using R code publicly available from the NICE DSU website with age and gender included as covariates.[5] Note that the latest version of code represents a refined version of the original analysis, undertaken by Alava et al. who re-estimated the model on a different dataset with a larger sample size.[6]
- As the DSU crosswalk required data on age and gender, those patients with missing data were excluded meaning the full dataset will have more observations than the analysis dataset.
- For comparability, the EuroQoL and DSU crosswalks were applied on the same dataset.

Methods

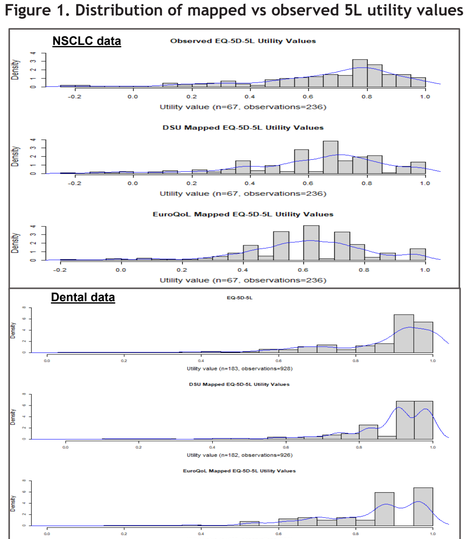
Assessment of predictive performance

- Predictive performance of each mapping crosswalk was assessed for both utility values and QALYs, measured using the root mean square error (RMSE) and mean absolute error (MAE).

Results

- Figure 1 compares the distribution of mapped vs observed 5L utility values. In the NSCLC dataset, the distribution of utility values using observed 5L responses exhibits a peak at 0.8 and is characterized by a smooth continuous distribution across the utility value range. In contrast, discontinuities and gaps are present in the distribution of mapped utility values, likely reflecting the high frequency of some 3L health states and absence of others. In the dental clinical trial dataset, the distribution of observed EQ-5D-5L utility values peaks at 0.9, also with a smooth continuous distribution across the utility value range. However, there are fewer discontinuities in the distribution of mapped utility values compared to the NSCLC dataset, likely reflecting a smaller number of observed health states and subsequently a narrower utility value range.

Figure 1. Distribution of mapped vs observed 5L utility values



Results

- Using either MAE or RMSE as the measure of predictive performance, the DSU crosswalk provided nominally more accurate predicted utility values in the NSCLC dataset, though the difference is marginal when compared to the most accurate EuroQoL crosswalk (Table 2).
- For the NSCLC dataset the DSU had a MAE and RMSE of 0.088 and 0.116 respectively, compared to 0.091 and 0.122 for EQ(8).
- There is evidence to suggest that including age and gender as predictors slightly improved predictive performance: four of the five crosswalks that provided the lowest MAE and RMSE did so.
- For the dental clinical trial dataset, several EuroQoL crosswalks provided the most accurate predicted utility values with comparable MAE and RMSE values, though differences compared to the DSU crosswalk were marginal.

Table 2. Predictive performance mapping crosswalk for utility values

Data	Measure	Mapping Crosswalk									
		DSU	EQ (1)	EQ (2)	EQ (3)	EQ (4)	EQ (5)	EQ (6)	EQ (7)	EQ (8)	EQ (9)
NSCLC	MAE	***0.088***	0.097	0.107	0.098	*0.097*	0.099	0.096	0.092	*0.091*	*0.092*
	RMSE	***0.116***	0.128	0.129	0.127	*0.123*	0.129	0.123	0.122	*0.122*	*0.123*
Dental	MAE	*0.057*	**0.044**	0.058	0.047	*0.049*	0.047	0.052	**0.044**	***0.044***	***0.044***
	RMSE	*0.091*	**0.074**	0.078	0.075	***0.074***	**0.074**	0.075	0.075	*0.075*	*0.075*

Bold font indicates the most accurate mapping crosswalk (lowest MAE/RMSE).
Italic font indicates crosswalks that included age and gender as predictors.

- When assessing the predictive performance of mapped QALY estimates (Table 3), results were largely consistent with those observed for utility values. For both datasets the DSU crosswalk produced a lower MAE, though differences compared to the EuroQoL crosswalks were marginal, particularly in the dental clinical trial.
- The non-parametric EuroQoL crosswalk produced the lowest RMSE in the NSCLC dataset (Table 3). In the dental clinical trial dataset, the MAE and RMSE appeared to be slightly more consistent across crosswalks and the magnitude of error was significantly larger than in the NSCLC dataset. This does not imply worse performance but reflects a difference of scale given the much higher utility values, and therefore QALY estimates, in the dental clinical trial population.

Table 3. Predictive performance mapping crosswalk for QALY values

Data	Measure	Mapping Crosswalk									
		DSU	EQ (1)	EQ (2)	EQ (3)	EQ (4)	EQ (5)	EQ (6)	EQ (7)	EQ (8)	EQ (9)
NSCLC	MAE	***0.043***	0.046	0.051	0.052	*0.050*	0.052	0.046	0.048	*0.047*	*0.048*
	RMSE	*0.086*	**0.085**	0.091	0.092	*0.089*	0.092	0.085	0.088	*0.088*	*0.089*
Dental	MAE	***0.127***	0.129	0.132	0.130	*0.130*	0.130	0.131	0.130	*0.129*	*0.129*
	RMSE	*0.393*	0.393	0.395	0.393	*0.391*	0.393	**0.388**	0.396	*0.396*	*0.396*

Bold font indicates the most accurate mapping crosswalk (lowest MAE/RMSE).
Italic font indicates crosswalks that included age and gender as predictors.

Conclusions

- This is the first study (to our knowledge) comparing the performance of the two crosswalks for predicting EQ-5D-5L utility values (using the UK value set) and associated QALYs.
- Applying the two crosswalks in two studies with different designs and patient populations reduced the risk that the results are influenced by the features of any one study and provided valuable insights into the factors that determine the predictive performance of mapping crosswalks.
- For the NSCLC dataset, the DSU crosswalk provided marginally more accurate predicted utility and QALY values, relative to the EuroQoL crosswalks. When varying the observed EQ-5D responses, age and gender values in the NSCLC dataset via simulations, the results were less conclusive suggesting that the comparative performance of the crosswalks may change if applied in similar datasets. For the dental clinical trial dataset, the EuroQoL crosswalks generally provided more accurate predicted utility and QALY values though the difference in prediction error was small.
- Both the NSCLC and dental clinical trial dataset were characterized by a relatively narrow coverage of 3L health states, which is likely to be a feature common to most datasets that researchers are attempting to implement mapping in.
- Algorithms that included age and gender as predictors tended to produce lower errors in the NSCLC dataset but this trend was less evident in the dental clinical trial dataset.

References

1. Kennedy-Martin M, Slaap B, Herdman M, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. Eur J Health Econ. 2020;21(8):1245-1257.
2. 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;55:139-152.
3. van Hout BA, Shaw JW. Mapping EQ-5D-3L to EQ-5D-5L. Value Health. 2021;24(9):1285-1293.
4. EuroQol. Cross-walk. https://euroqol.org/support/analysis-tools/cross-walk/. Updated 31/01/2022. Accessed 8th July, 2022.
5. NICE DSU. Mapping EQ-5D-5L to 3L. NICE Decision Support Unit. https://nicedsu.sites.sheffield.ac.uk/methods-development/mapping-eq-5d-5l-to-3l. Accessed 8th July, 2022.
6. Hernández-Alava M, Pudney S, Wailoo AJ. 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.

Acknowledgments

- We thank Prof Francesco Mannocci King's College London Dental School for allowing use of the data, Protocol IRAS : 237565. Data is based on an NIHR grant under the RfpB (Request for Patient Benefit Stream)

Declaration of interests

Adelphi Values PROVE were contracted by Bristol Myers Squibb to conduct and report this research. JS and BB are employees of Bristol Myers Squibb. AH and AF are employees of Adelphi Values PROVE. IK is an employee of University of Warwick. BvH is an employee of University Sheffield. SB is an employee at University of Liverpool.

Email: James.Shaw@bms.com