Testing alternative regression models to predict utilities: The case of mapping the QLQ-C30 onto the EQ-5D-5L and the SF-6D

Authors: Admassu N. Lamu [1], Jan Abel Olsen[1,2]*

[1]Department of Community Medicine, University of Tromsø, N-9037 Tromsø, Norway
[2]Centre for Health Economics, Monash University, VIC 3800, Australia

*Corresponding author:
E-mail: admassu.n.lamu@uit.no
Tel.: +47 41284336
ORCID: 0000-0001-6638-421x

Acknowledgment: This study was funded by The Australian National Health and Medical Research Council (grant number 1006334), the University of Tromsø, and The Research Council of Norway (grant number 221452).

Abstract

Purpose: Compare alternative statistical techniques to find the best approach for converting QLQ-C30 scores onto EQ-5D-5L and SF-6D utilities, and estimate the mapping algorithms that best predict these health state utilities.

Methods: 772 cancer patients described their health along the cancer-specific instrument (QLQ-C30) and two generic preference-based instruments (EQ-5D-5L and SF-6D). Seven alternative regression models were applied: ordinary least squares (OLS), generalized linear model (GLM), extended estimating equations (EEE), fractional regression model (FRM), beta binomial (BB) regression, logistic quantile regression (LQR), and censored least absolute deviation (CLAD). Normalized mean absolute error (NMAE), normalized root mean square error (NRMSE), r-squared (r^2) and concordance correlation coefficient (CCC) were used as model performance criteria. Cross-validation was conducted by randomly splitting internal dataset into two equally sized groups to test the generalizability of each model.

Results: In predicting EQ-5D-5L utilities, the BB regression performed best. It gave better predictive accuracy in terms of all criteria in the full sample, as well as in the validation sample. In predicting SF-6D, the EEE performed best. It outperformed in all criteria: NRMSE = 0.1004, NMAE = 0.0798, CCC = 0.842 and r^2 =72.7% in the full sample, and NRMSE=0.1037, NMAE=0.0821, CCC = 0.8345 and r^2=71.4% in cross-validation.

Conclusions: When only QLQ-C30 data are available, mapping provides an alternative approach to obtain health state utility data for use in cost-effectiveness analyses. Among seven alternative regression models, the BB and the EEE gave the most accurate predictions for EQ-5D-5L and SF-6D, respectively.

Keywords: Mapping, Regression models, QLQ-C30, EQ-5D-5L, SF-6D, QALYs

1. Introduction

Quality-adjusted life-years (QALYs) are becoming the standard outcome measure for health economic evaluations [1; 2]. To obtain the quality adjustment weight in the QALY, generic preference-based measures are used. However, many clinical trials commonly use a disease-specific outcome measure. In such circumstances, analysts would need to translate, or "map", disease specific scores onto generic preference-based values in order to express gains along a commensurable metric like the QALY.

Cancer is the second leading cause of death behind cardiovascular diseases [3], despite a strong fight against it. This fight is evidenced by recent developments in personalized medicine and novel treatment approaches such as immunotherapy [4; 5], which involve increasing pressures on healthcare budgets. For instance, a quarter of the technology appraisals produced by National Institute for Health and Care Excellence (NICE) has focused on cancer interventions [6]. Thus, there is a growing need for cost-effectiveness analyses in cancer treatment, not only to compare across the many new cancer interventions, but also to compare with resources spent in other diseases areas.

The European Organization for Research and Treatment Quality of Life Questionnaire Core 30 (EORTC QLQ–C30, hereafter QLQ-C30) [7] is a disease-specific instrument widely used to measure health related quality of life (HRQoL) in cancer patients. However, to enable comparisons of health outcomes across disease areas, a generic preference-based instrument is needed. The EQ-5D is the most widely used generic preference-based instrument for calculating QALYs [8; 9]. A new five levels version of the EQ-5D has recently been developed [10], intended to reduce ceiling effects and improve reliability and sensitivity as compared to the earlier three-level version [11]. National guidelines on health technology appraisals (HTA) submitted to the NICE in the United Kingdom (UK) [12] have recommended the use of the EQ-5D. The SF-6D is the second most widely used generic preference-based instrument. Recently, there is a rapid increase in mapping between 'source' (e.g., disease-specific) instruments and 'target' (e.g., generic preference-based) instruments, where the majority of mapping functions available have applied the EQ-5D as their target measures [13; 14].

Previous mapping studies between QLQ-C30 and EQ-5D were based on the EQ-5D-3L [15-19], except for one based on the EQ-5D-5L interim *cross-walk* value set [20]. Thus, there is a need to develop mapping algorithm with the publication of the new directly elicited EQ-5D-5L utilities. Similarly, mapping studies between QLQ-C30 and the SF-6D are sparse [16; 21]. Accordingly, this paper makes two important contributions. First, it produces optimal mapping functions for the EQ-5D-5L based on the recently published value set for England [22], as well as the UK value set for SF-6D. Second, we

make important methodological contributions, by comparing six regression models previously used in mapping studies as well as one model new to mapping research; the extended estimation equations (EEE) in the generalized linear model (GLM). This approach has a desirable property in that it allows estimation of flexible mean and variance functions using the data at hand, leading to consistent and efficient estimation [23]. This study followed the recently developed checklist of minimum reporting requirement for mapping studies [24]: 'MApping Preference-based Measures reporting Standards (MAPS)'.

2. Methods

2.1 Data

Data were obtained from the Multi Instrument Comparison (MIC) study, which was an online survey administered by a global panel company, CINT Australia Pty Ltd in six countries: Australia, Canada, Germany, Norway, UK , and the US. Considering the difficulty of direct control in the online survey, several edit procedures (e.g. exclusion of respondents with inconsistent responses on duplicated questions, and removal of respondents whose recorded completion time below 20 minutes) were conducted to ensure the quality of the data. For further details on the description of data, see Richardson *et al.* [25]. Among the seven chronic disease groups included in this comprehensive international study, the current paper is based on respondents who had been diagnosed with cancer (N=772). They described their health along the QLQ C-30 as well as the EQ-5D-5L and the SF-6D. The MIC data is an ideal source for deriving mapping algorithms from disease-specific outcome measures onto generic preference-based measures. So far it has been applied to develop several mapping algorithms in different chronic diseases, including asthma [26] depression [27; 28], heart diseases [29] and diabetes [30].

2.2 Measures of variables

EQ-5D-5L

The EQ-5D-5L is a validated instrument covering five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, each with five severity levels (no problems, slight problems, some problems, moderate problems, and unable to/extreme problems) [10]. Thus, the EQ-5D-5L produces 3,125 (5^5) health states. The described health state values were elicited from a sample of the English general public [22]. The utility values ranges from -0.285 for the worst health state (the 'pits') to 1.000 for full health. It serves as *target* or *dependent* variable.

SF-6D

The SF-6D is derived from the short form 36 items (SF-36)[31]. It comprises six dimensions: physical functioning, role functioning, social functioning, bodily pain, mental health, and vitality, each with four to six levels that define 18,000 unique health states. The preference weights have been elicited in the United Kingdom and ranges from 0.301 to 1.000. SF-6D is also a *target* or *dependent* variable.

QLQ-C30

The standard version of the QLQ-C30 recommended by the developers (QLQ-C30 version 3.0) was used, which comprises 30 items [7]. Each item has four response levels (i.e. 'not at all', 'a little', 'quite a bit' and 'very much') except the two items assessing global quality of life and overall health that use a seven-point scale. The QLQ-C30 covers 15 subscales: one *global health status* scale (GH) (2 items); five *functioning* scales – physical (PF), role (RF), emotional (EF), cognitive (CF), and social (SF) (15 items), and; nine *symptom* scales – fatigue (FA), nausea & vomiting (NV), pain (PN), dyspnea (DY), insomnia (SL), loss of appetite (AP), constipation (CO), diarrhea (DI), and financial difficulties (FI) (13 items). The score for each subscale was calculated by summing responses for all items in each subscale, and linearly transformed onto a [0 – 1] scale, with 0 indicating the worst, and 1 the best possible health state. These 15 subscales of QLQ-C30 are used as *source* or *independent* variables.

2.3 Statistical analyses

Exploratory analyses

Respondents' characteristics were described using mean and standard deviation (SD) or count and percentage share in the sample. The degree of conceptual overlap between the source and the target variables was examined with Spearman's rank correlation (rho, ρ) and exploratory factor analyses (EFA). The EFA was employed using principal axis factoring, which has been recommended as the preferred method of factor extraction [32]. An eigenvalue of greater than 1 was used as selection criterion to extract underlying factors. To account for potential correlations among factors, rotation was performed using an oblique Promax method [33].

Econometric models

Both direct and indirect (response) mapping approaches were explored. However, only results from the direct mapping were reported here[1]. The fifteen subscales of the QLQ-C30 together with

[1] Response mapping produced the highest error (measured in terms of MAE and RMSE) among other models. In response mapping, exact prediction of health state requires correct prediction for each dimension of the target

respondents' age and gender were used as a *source* to predict health state utility values on the *target* variables (EQ-5D-5L and SF-6D). The final predictors were determined via stepwise backward elimination that included only significant variables ($p < 0.10$). Variables with counter intuitive signs were excluded.

Seven econometric models were used: ordinary least square (OLS) regression, generalized linear model (GLM), the extended estimation equations (EEE) in the GLM, binomial beta (BB) regression, fractional regression model (FRM), logistic quantile regression (LQR), and censored least absolute deviation (CLAD). OLS is the most commonly used regression model in mapping studies [13], which requires data to be normally distributed with constant variance. If these assumptions are violated, the OLS formulation may not be appropriate.

The GLM is a flexible generalization of ordinary linear regression that allows for the outcome variables to have a non-normal error distributions [34] by a *priori* specifying the mean and variance functions. However, incorrect specifications of the mean and the variance functions can produce bias and inefficiency in estimation. To overcome this problem of misspecification, Basu and Rathouz [35] developed the EEE, which is a flexible techniques to estimate the mean and variance functions from the data at hand. They recommended power variance structure for continuous outcome variables.[2] This variance family is indexed by two parameters (θ_1, and θ_2), and defined as:

$$f(\mu_i, \theta_1, \theta_2) = \theta_1 \mu_i^{\theta_2}, \qquad (1)$$

where μ_i is the mean function, $E(Y|X = x)$, with Y a non-negative outcome variable and X a vector of predictors.

A parametric family of link functions indexed by lambda (λ) was defined as:

$$g(\mu_i;\lambda) = \begin{cases} (\mu_i^{\lambda} - 1)/\lambda, & if \ \lambda \neq 0 \\ \log(\mu_i), & if \ \lambda = 0 \end{cases} \qquad (2)$$

Following Basu and Rathouz [35], the predicted mean value was given as:[3]

instrument, which rarely achieved in practice. Thus, it can be severely penalized when incorrect prediction is made.

[2] Power variance structure is preferred as it includes the variance of several standard distributions (such as Poisson, gamma or inverse Gaussian) used for modelling health outcomes.

[3] The EEE is estimated by *pglm*, a user defined Stata command, which shows better convergence properties when the outcome variable is scaled by dividing by its *mean* [36].

$$\hat{\mu}(x)=((X\hat{\beta})*\hat{\lambda}+1)^{1/\hat{\lambda}},\ for\ all\ \hat{\lambda},\ \hat{\lambda}\neq 0, \quad (3)$$

where β is a vector of parameters to be estimated.

In both GLM and EEE, EQ-5D-5L disutility (where disutility = 1 – EQ-5D-5L utility) was used as an outcome variable to have non-negative values.

The FRM is a semi-parametric approach that appropriately model bounded dependent variables defined on [0, 1] interval [37]. It does not make any distributional assumption about an underlying structure used to obtain the outcome variable, but requires the correct specification of the conditional mean outcome [37; 38]. The *complementary loglog* (*cloglog*) is the best alternative functional form in both EQ-5D-5L and SF-6D prediction, and is used as a link function. The FRM model has been detailed elsewhere [30]. The BB regression is a fully parametric counterpart, which is flexible and capable of modeling dependent variables restricted between 0 and 1 [39]. Since it is not defined at 0 or 1, estimation with standard BB regression can be problematic, particularly in the presence of piling up of data at 0 or 1. Thus, previous studies have suggested a *zero-one inflated* BB model that can estimate probability masses at both 0 and 1 [40; 41]. As there is no 0 responses in the present study, a one-inflated BB (hereafter *BB* regression) model was applied. The BB regression with particular application to EQ-5D mapping study was detailed in Khan and Morris [42].

Like the FRM and BB, the LQR is modelling bounded data but uses quantiles (e.g. median) instead of mean. The boundary values need not be 0 and 1 in LQR. For instance, if Y is bounded from below by a known constant Y_{min} and from above by Y_{max}, then a *logistic* transformation can be applied to the outcome variable, and hence we obtain [43]:

$h(Y) = \log\left(\frac{Y - Y_{\min}}{Y_{\max} - Y}\right)$, which is equivalent to:

$$Q_Y(p) = \frac{\exp(X\beta)Y_{\max} + Y_{\min}}{1+\exp(X\beta)} \quad (4)$$

where $Q_Y(p)$ is defined as the conditional p^{th} quantile of Y given a set of X independent variables and p is the proportion between 0 and 1. In the present study, p = 0.50 (median) has been applied.

The CLAD model is more appropriate for outcome variables censored at lower or upper endpoints [44]. It is a semi-parametric estimator that is robust to distributional assumptions and heteroscedasticity, because it uses median values rather than means among similar groups, since medians are likely to be less affected by censoring.

2.4 Performance of mapping algorithms

Individual predictions were evaluated by examining differences between observed and predicted utilities, measured by the normalized root mean square error (NRMSE) and the normalized mean absolute error (NMAE), where lower values indicate better fit. We normalized both MAE and RMSE to the range of the observed data. Such normalization produce non-dimensional scale, and facilitate the comparison between datasets or models with different scales [30].

Model performance was also examined by the square of the correlation coefficient (r^2) between the observed and the predicted values. Furthermore, Lin's concordance correlation coefficient (CCC) [45] was considered as measures of model performance. The CCC measures absolute agreement, where a value close to unity implies a good concordance between predicted and observed measures (i.e. good prediction). The graphical display of the observed vs predicted utilities were visually compared for a series of models. Finally, to evaluate the bias and precision of prediction the mean, SD, median and other quintiles of predicted utilities were compared against the observed utilities.

2.5 Validation methods

Due to lack of external data, cross-validation was performed by randomly splitting the existing data into two equally sized groups (estimation vs validation) to evaluate the model fit in out-of-sample data. The model was fitted on the estimation sample, and the resulting parameters from the fitted model were then used to predict utilities on the validation sample. This procedure has been repeated by reversing the validation and estimation sample. The average of NRMSE, NMAE, CCC and r^2 from both iterations were reported for easy comparison of the models' predictive performance.

The preferred model was identified based on four criteria derived from both the full and the validation sample: the lowest combinations of NRMSE and NMAE values and the highest values of r^2 and CCC. The model that performed best in most of these criteria should be selected. Eventually, the best fitting model with good predictive accuracy was estimated as the preferred mapping algorithm using the full sample. All statistical analyses were conducted using Stata® version 15.1 (StataCorp LP, College Station, Texas, USA) except the EFA, which was carried out in SPSS version 24 (IBM Corp., Armonk, NY, USA).

3 Results

3.1 Descriptive statistics and conceptual overlap

Table 1 summarizes the description of variables. For EQ-5D-5L, the mean (SD) value was 0.781 (0.209). For SF-6D, it was 0.686 (0.133). The frequency distributions of both EQ-5D-5L and SF-6D were depicted in Figure 1. Among the QLQ-C30 subscales, the mean (SD) score ranges from 0.573 (0.242) for global health status subscale to 0.886 (0.215) for nausea and vomiting. The mean (SD) age was 58.2 (12.4) years, and there were 54% women in the sample.

Table 1

Sample characteristics (N=772)

Variable	Mean (SD)	Min	Max
Utility measures			
EQ-5D-5L	0.781 (0.209)	-0.281	1
SF-6D	0.686 (0.133)	0.301	1
QLQ C-30 subscales			
Global health status	0.573 (0.242)	0	1
Physical functioning	0.765 (0.234)	0	1
Role functioning	0.712 (0.306)	0	1
Emotional functioning	0.696 (0.264)	0	1
Cognitive functioning	0.776 (0.252)	0	1
Social functioning	0.663 (0.311)	0	1
Pain	0.655 (0.315)	0	1
Fatigue	0.605 (0.275)	0	1
Nausea & vomiting	0.886 (0.215)	0	1
Dyspnea	0.761 (0.293)	0	1
Insomnia	0.626 (0.330)	0	1
Loss of appetite	0.829 (0.283)	0	1
Constipation	0.848 (0.256)	0	1
Diarrhea	0.859 (0.251)	0	1
Financial difficulties	0.678 (0.348)	0	1
Socio-demographics			
Age	58.228 (12.358)	19	93
Female, n (%)	417 (54.0)		
Country, N (%)			
Australia	154 (20.0)		
Canada	138 (17.9)		
Germany	115 (14.9)		
Norway	80 (10.4)		
UK	137 (17.7)		
USA	148 (19.2)		

Spearman's rank correlation was presented in Table 2. Among EQ-5D-5L dimensions, usual activities dimension generally provide strong correlation with most QLQ-C30 subscales. The highest correlation was found between the pain dimension (for each target instrument) and cancer pain subscale (ρ is close to 0.80).

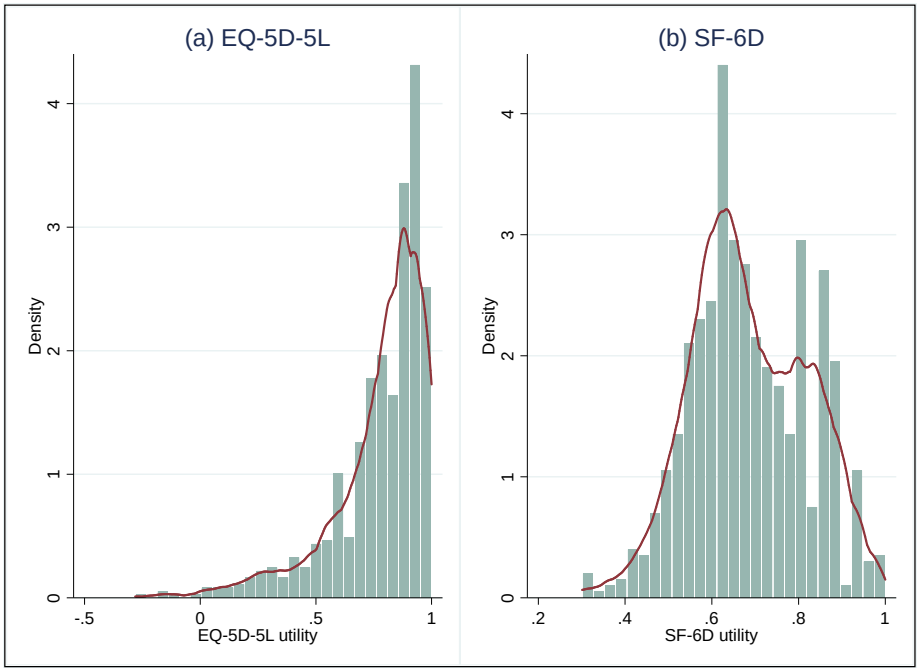

Figure 1
Frequency distribution of EQ-5D-5L and SF-6D utilities with kernel density overlaid

Table 2
Spearman's correlation coefficients between the QLQ-C30 the source and the target instruments

	EQ-5D dimensions					SF-6D dimensions					
QLQ-C30 domain scales	Mobility	Self-care	Usual activities	Pain/ Discomfort	Anxiety/ Depression	Physical function	Role function	Social function	Bodily pain	Mental function	Vitality
Global health status scale											
Global health status	0.528	0.332	0.612	0.536	0.453	0.553	0.539	0.594	0.585	0.505	0.615
Functioning scales											
Physical functioning	0.718	0.458	0.696	0.569	0.285	0.771	0.510	0.537	0.626	0.294	0.566
Role functioning	0.604	0.425	0.691	0.515	0.264	0.631	0.505	0.573	0.630	0.308	0.488
Emotional functioning	0.275	0.300	0.387	0.384	0.653	0.347	0.491	0.470	0.433	0.668	0.431
Cognitive functioning	0.293	0.312	0.403	0.315	0.375	0.379	0.414	0.404	0.367	0.368	0.348
Social functioning	0.489	0.351	0.576	0.466	0.360	0.562	0.485	0.622	0.554	0.374	0.471
Symptom scales											
Pain	0.596	0.399	0.600	0.785	0.332	0.564	0.469	0.491	0.803	0.360	0.454
Fatigue	0.508	0.354	0.597	0.535	0.417	0.593	0.562	0.597	0.607	0.427	0.615
Nausea & vomiting	0.229	0.260	0.330	0.314	0.296	0.304	0.318	0.397	0.333	0.302	0.289
Dyspnea	0.395	0.280	0.482	0.320	0.261	0.443	0.333	0.371	0.360	0.282	0.409
Insomnia	0.365	0.344	0.392	0.413	0.423	0.419	0.416	0.420	0.442	0.431	0.411
Loss of appetite	0.352	0.304	0.439	0.377	0.327	0.384	0.392	0.473	0.413	0.354	0.366
Constipation	0.246	0.253	0.257	0.292	0.221	0.277	0.268	0.230	0.326	0.231	0.194
Diarrhea	0.177	0.258	0.299	0.216	0.219	0.266	0.261	0.264	0.242	0.222	0.218
Financial difficulties	0.334	0.283	0.417	0.363	0.382	0.426	0.429	0.469	0.409	0.377	0.377

All coefficients are statistically significant at less than 1%.

The pattern matrices for EFA were reported in Appendix Table A1 and A2. The EFA analysis for the QLQ-C30 items and the EQ-5D-5L dimensions produced five underlying factors ('physical', 'emotional', 'symptom', 'pain', and 'self-care'), explaining 59% of the total variance. Two EQ-5D-5L dimensions (mobility, usual activities) were mainly loaded onto the same factor as the QLQ-C30 items that describe activities related to 'physical functioning'. One EQ-5D-5L dimension (self-care) was mainly loaded to the last factor, in which only one QLQ-C30 item (i.e., Do you need help with eating, dressing, washing yourself or using the toilet?) was mainly loaded onto. In general, the EQ-5D-5L dimensions mainly loaded onto all factors except the 'symptom' factor, which explains 55.6% of the total variance.

EFA results with SF-6D produced six factors; 'emotional ', 'physical ', 'symptom', 'energy', 'limitation' (in financial, family and social activities), and 'pain'. Surprisingly, the SF-6D 'role' and 'social' functioning dimensions were not mainly loaded onto anyone of the extracted factors. All other dimensions of SF-6D had conceptual overlap with the QLQ-C30 items that together extracted 'emotional', 'physical', 'pain' and 'energy' factors, which contribute about 54% to the total variance. Like the EQ-5D-5L, none of the SF-6D dimensions were mainly loaded to the 'symptom' factor.

3.2 Model performance

Table 3 summarizes the performance of models based on full sample and cross-validation. For EQ-5D-5L, the BB regression model performed best in terms of all criteria. The model produced r^2, NRMSE, NMAE and CCC of 68%, 0.0930, 0.0651 and 0.813, respectively in the full sample. Interestingly, this model revealed the best predictive accuracy in terms of all criteria in the cross-validation as well. Although the CLAD model performed best in terms of NMAE (0.0649) in the full-sample, it showed least predictive performance in all other criteria.

For the SF-6D, the EEE model consistently performed best both in the full sample and in cross-validation. For example, NRMSE (0.1004) and NMAE (0.0798) were minimal, and CCC (0.842) and r^2 (72.7%) were the highest, indicating low degree of predictive error with high level of accurate predictions in the full sample. The corresponding values in the validation sample were: NRMSE = 0.1037, NMAE = 0.0821, r^2 of 71.4%, and CCC = 0.8345.

Table 3

Model performance in the prediction of EQ-5D-5L and SF-6D utilities

	EQ-5D-5L						SF-6D					
Models	RMSE	MAE	NRMSE	NMAE	CCC	r^2	RMSE	MAE	NRMSE	NMAE	CCC	r^2
Panel A: Full-sample												
OLS	0.1218	0.0848	0.0951	0.0662	0.7960	0.6608	0.0732	0.0603	0.1047	0.0862	0.8250	0.7022
GLM	0.1207	0.0867	0.0942	0.0677	0.7950	0.6696	0.0709	0.0573	0.1014	0.0820	0.8390	0.7210
EEE	0.1199	**0.0834**	0.0936	0.0651	0.8000	0.6737	**0.0702**	**0.0558**	**0.1004**	**0.0798**	**0.8420**	**0.7269**
BB	**0.1191**	**0.0834**	**0.0930**	**0.0651**	**0.8130**	**0.6795**	0.0759	0.0633	0.1086	0.0906	0.8030	0.6857
FRM	0.1198	0.0835	0.0935	0.0652	0.8040	0.6724	0.0739	0.0607	0.1057	0.0869	0.8260	0.6974
LQR	0.1216	0.0835	0.0949	0.0652	0.8020	0.6662	0.0744	0.0603	0.1064	0.0863	0.8230	0.6931
CLAD	0.1267	0.0833	0.0989	0.0649	0.7170	0.6608	0.0754	0.0605	0.1079	0.0866	0.8260	0.6891
Panel B: Cross-validation												
OLS	0.1236	0.0858	0.0965	0.0670	0.7910	0.6494	0.0751	0.0610	0.1074	0.0873	0.8190	0.6967
GLM	0.1276	0.0894	0.0996	0.0698	0.7835	0.6437	0.0728	0.0584	0.1042	0.0835	0.8330	0.7069
EEE	0.1225	0.0858	0.0956	0.0670	0.7925	0.6563	**0.0725**	**0.0574**	**0.1037**	**0.0821**	**0.8345**	**0.7139**
BB	**0.1208**	**0.0835**	**0.0943**	**0.0652**	**0.8050**	**0.6698**	0.0764	0.0625	0.1093	0.0894	0.8010	0.6725
FRM	0.1217	0.0843	0.0950	0.0658	0.7995	0.6744	0.0747	0.0610	0.1069	0.0872	0.8200	0.6863
LQR	0.1241	0.0866	0.0969	0.0676	0.7980	0.6598	0.0757	0.0622	0.1083	0.0890	0.8140	0.6848
CLAD	0.1272	0.0856	0.0993	0.0668	0.7555	0.6402	0.0756	0.0612	0.1082	0.0876	0.8105	0.6856

Best results are in bold type

RMSE root mean squared error, *MAE* mean absolute error, *NRMSE* normalised RMSE, *NMAE* normalised MAE, r^2 square of correlation coefficient between predicted and observed utilities, *OLS* ordinary least square, *GLM* generalised linear model, *EEE* extended estimation equations, *BB* (one-inflated) beta binomial regression, *FRM* fractional regression model, *LQR* logistic quantile regression, *CLAD* censored least absolute deviation, *EQ-5D-5L* EuroQol five-dimensional five level questionnaire, *SF-6D* short-form six-dimensional questionnaire.

The predicted mean (SD) for the EQ-5D-5L ranges from 0.781 (0.170) for the OLS model to 0.801 (0.142) for the CLAD model, which is quite similar to the observed mean. Surprisingly, all models yield quite similar mean SF-6D prediction close to observed mean (0.686). However, over-prediction has been observed at severe health states (Figure 2a and 2b; and Appendix Table A3). For instance, the 1st, and 5th percentiles of the predicted EQ-5D-5L utility were 0.266 and 0.386 against 0.037 and 0.315 for the observed utility, respectively in the best fitting BB model (Appendix Table A3). The corresponding results for SF-6D were 0.467 and 0.497 for predicted against 0.395 and 0.482 for the observed values in the preferred EEE model. None of the seven models was able to predict the lowest utility score. Only OLS model predicted outside of the observed range (> 1) of the EQ-5D-5L.

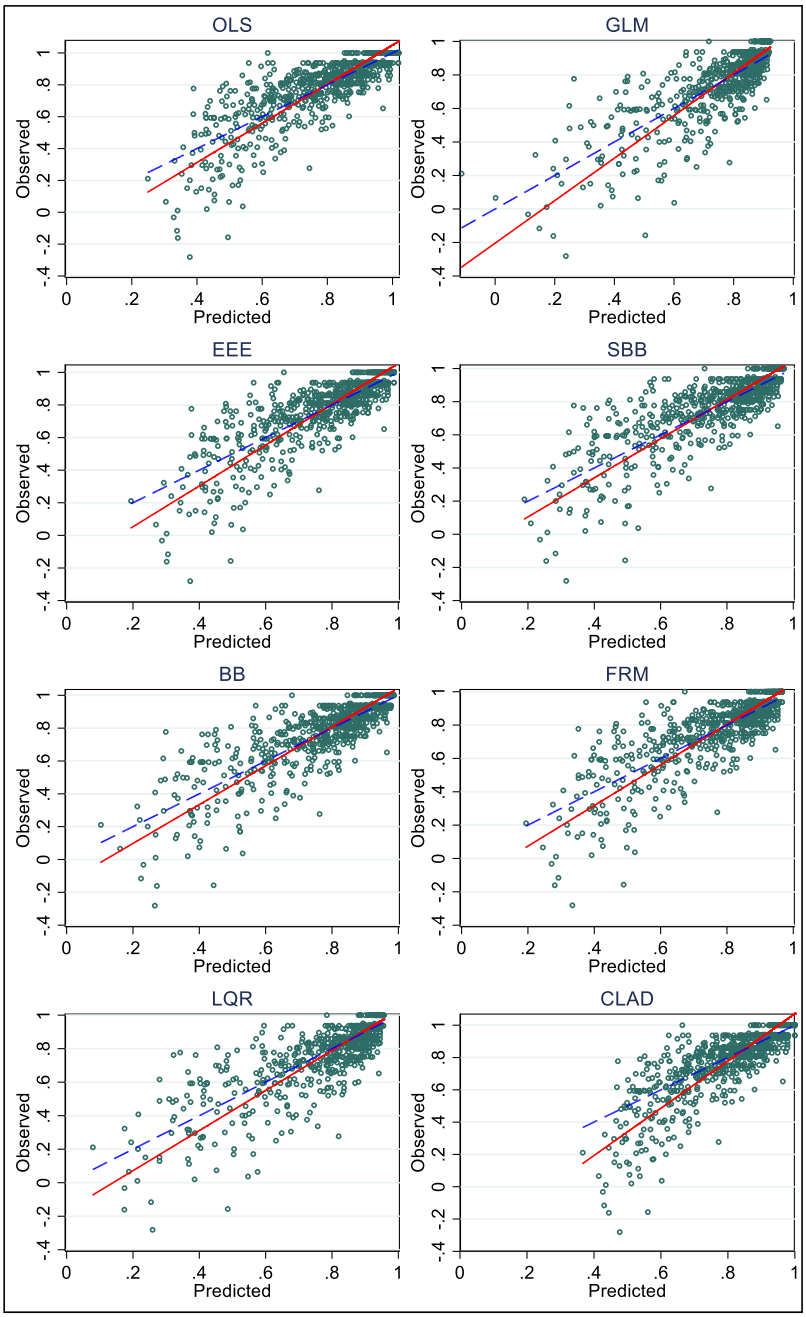

Figure 2a Scatter plot of observed versus predicted EQ-5D-5L

Solid (red) line depicts standard deviation (SD) line, which shows a measure of the centre of the data; broken (blue) line is a line along which observed utilities equal predicted utilities. Perfect prediction occurs when SD line and the line of equality overlaps. *OLS* ordinary least square, *GLM* generalized linear model, *EEE* extended estimation equations, *SBB* standard beta binomial regression, *BB* (one-inflated) beta binomial regression, *FRM* fractional regression model, *LQR* logistic quantile regression, *CLAD* censored least absolute deviation.

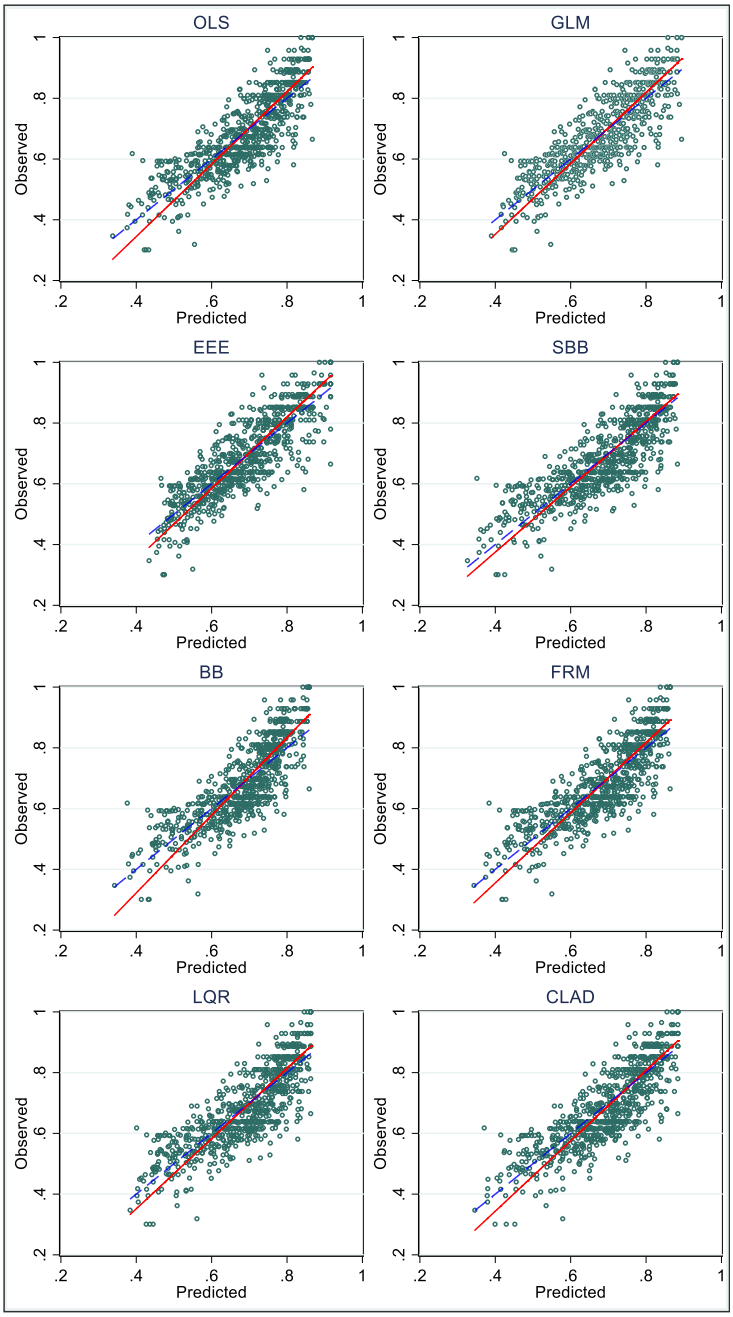

Figure 2b Scatter plot of observed versus predicted EQ-5D-5L

Solid (red) line depicts standard deviation (SD) line, which shows a measure of the centre of the data; broken (blue) line is a line along which observed utilities equal predicted utilities. Perfect prediction occurs when SD line and the line of equality overlaps. *OLS* ordinary least square, *GLM* generalized linear model, *EEE* extended estimation equations, *SBB* standard beta binomial regression, *BB* (one-inflated) beta binomial regression, *FRM* fractional regression model, *LQR* logistic quantile regression, *CLAD* censored least absolute deviation.

3.1 Regression results

The regression results are summarized in Table 4. The GH, PF, EF, and pain (PN) gave significant ($p < 0.01$) predictions of both EQ-5D-5L and SF-6D utilities. Further, six additional variables were significant predictors of SF-6D in the best model. The remaining QLQ-C30 sub-scales were either insignificant ($p > 0.10$) or yield logically inconsistent signs, and hence not reported. Neither age nor gender predicted EQ-5D-5L and SF-6D in the preferred models.

Table 4

Regression results for predicting EQ-5D-5L and SF-6D utilities from QLQ-C30 subscales

	EQ-5D-5L							SF-6D						
Variables	OLS	GLM	EEE	BB	FRM	QRM	CLAD	OLS	GLM	EEE	BB	FRM	QRM	CLAD
GH	0.0993*	-0.3965*	-0.5384*	0.6062*	0.3694*	0.8445*	0.0513‡	0.1160*	0.1664*	0.1396*	0.5187*	0.3586*	0.8929*	0.1531*
	(0.0312)	(0.1521)	(0.1095)	(0.1479)	(0.0784)	(0.1884)	(0.0293)	(0.0176)	(0.0259)	(0.0257)	(0.0769)	(0.0524)	(0.1534)	(0.0235)
PF	0.2763*	-0.7997*	-1.1290*	1.0456*	0.6574*	1.0271*	0.2058*	0.0925*	0.1610*	0.1951*	0.4389*	0.2593*	0.6522*	0.0893*
	(0.0354)	(0.1275)	(0.1186)	(0.1285)	(0.0820)	(0.1998)	(0.0365)	(0.0204)	(0.0311)	(0.0327)	(0.0884)	(0.0583)	(0.1667)	(0.0262)
RF				0.2856**			0.0453*	0.0318†	0.0476†	0.0646*		0.0797†		
				(0.1147)			(0.0289)	(0.0163)	(0.0252)	(0.0244)		(0.0482)		
EF	0.1444*	-0.4255*	-0.5998*	0.6723*	0.3874*	0.8293*	0.1322*	0.0898*	0.1483*	0.1664*	0.3727*	0.2592*	0.5890*	0.0969*
	(0.0244)	(0.1059)	(0.0874)	(0.1004)	(0.0553)	(0.1628)	(0.0282)	(0.0142)	(0.0216)	(0.0220)	(0.0612)	(0.0407)	(0.1438)	(0.0223)
SF								0.0242†	0.0387†		0.1047†	0.0729†		
								(0.0144)	(0.0227)		(0.0622)	(0.0423)		
PN	0.2487*	-1.0275*	-1.0294*	1.0199*	0.6458*	1.3563*	0.2149*	0.0722*	0.1093*	0.1118*	0.3478*	0.1984*	0.4670*	0.0826*
	(0.0209)	(0.0996))	(0.0775)	(0.0953)	(0.0503)	(0.1112)	(0.0273)	(0.0118)	(0.0181)	(0.0191)	(0.0500)	(0.0340)	(0.0959)	(0.0150)
SL			-0.1072†				0.0264†	0.0173†	0.0274†	0.0266†		0.0525†	0.1784†	
			(0.0631)				(0.0166)	(0.0102)	(0.0150)	(0.0152)		(0.0289)	(0.0939)	
FA								0.0507*	0.0813*	0.0765*	0.2720*	0.1702*	0.4050‡	0.0735‡
								(0.0175)	(0.0253)	(0.0257)	(0.0781)	(0.0499)	(0.1671)	(0.0267)
AP										0.0377†				
										(0.0224)				
FI						0.1830**		0.0242*	0.0371*	0.0461*	0.1123‡	0.0729*	0.2264*	0.0358*
						(0.0806)		(0.0090)	(0.0134)	(0.0130)	(0.0438)	(0.0258)	(0.0779)	(0.0137)
Female								-0.0112‡	-0.0137*		-0.0506**	-0.0322‡		
								(0.0054)	(0.0076)		(0.0256)	(0.0163)		
Constants														
Intercept	0.2492*	0.1081†	2.2707*	-0.7736*	-0.7683*	-0.9379*	-0.3280*	0.3597*	-0.9149*	-0.6233*	-0.5513*	-0.8011*	-2.0031*	0.0343*
	(0.0262)	(0.0655)	(0.0859)	(0.0926)	(0.0573)	(0.1578)	(0.0283)	(0.0143)	(0.0230)	(0.0239)	(0.0679)	(0.0436)	(0.1214)	(0.0155)
λ			0.7866*							-1.2830*				
			(0.0501)							(0.3004)				
Θ_1			0.2611*							0.0103*				
			(0.0164)							(0.0006)				
Θ_2			1.1484*							0.5896				
			(0.0923)							(0.3922)				

GH global health status, *PF* physical functioning, *RF* role functioning, *EF* emotional functioning, *SF* social functioning, *PN* pain, *SL* insomnia (trouble sleeping), *FA* fatigue, *AP* loss of appetite, *FI* financial difficulties, *EQ-5D-5L* EuroQol five dimensional five level questionnaire, *SF-6D* short form six dimensional questionnaire, *λ* Lambda, and *Θ* Theta.

* p < 0.01, ‡ p < 0.05, † p < 0.1. Standard errors in parentheses.

4 Discussion

We have estimated health state utilities for the most widely used cancer-specific (QLQ-C30) instrument by mapping it onto each of the two most widely used generic preference-based instruments (EQ-5D-5L and SF-6D). The findings revealed the BB and the EEE as the best performing models for predicting EQ-5D-5L and SF-6D utilities, respectively.

When conducting mapping studies some degree of conceptual overlap between the source and the target values is important. In the present study, strong correlations were observed between similar dimensions of the QLQ-C30 sub-scales and the two target instruments (EQ-5D-5L and SF-6D), confirming the existence of a substantial amount of conceptual overlap between source and target instruments. Results from EFA demonstrated similar findings, establishing the foundations for mapping. However, none of the EQ-5D-5L and SF-6D dimensions mainly loaded onto the 'symptom' factor, which comprises lack of appetite, feeling nauseated, vomiting, and diarrhea, and they turned out to be insignificant predictors in the regression models as well. Although the QLQ-C30 comprises several items that describe limitations in daily activities, leisure activities, financial problems, family and social activities, the SF-6D dimensions of 'role limitation' and 'social functioning' were not mainly loaded onto any constructs formed by these QLQ-C30 items. This implies that these items appear to be measuring other aspects of role and social functioning than those described in the SF-6D.

The inclusion of socio-demographics may improve the accuracy of mapping algorithms [13]. Previous studies found that age was a significant predictor when mapping QLQ-C30 onto the EQ-5D-3L [15; 46], but not when predicting SF-6D utilities [21]. Results for gender are mixed. Gender was significant in some studies mapping QLQ-C30 onto EQ-5D utilities [47; 48] and SF-6D utilities [21], but not in others [48]. In the present study, neither age nor gender were significant in the preferred models for predicting EQ-5D-5L and SF-6D utilities. Such variations across studies could be attributable to the type of cancer in relation to gender and the age of the populations included in the studies.

This study also assessed the empirical performance of alternative regression models. In the mapping between QLQ-C30 subscales and the EQ-5D-5L, the CLAD model performed well in terms of NMAE as compared to other models. However, CLAD gives poor prediction in terms of all other criteria (NRMSE, r^2 and CCC). The FRM and EEE models performed well after BB regression. As for the mapping between

QLQ-C30 subscales and SF-6D, the EEE consistently outperformed other models, followed by the GLM model with *log* link. The CLAD and BB models performed relatively poorly in the prediction of SF-6D. The novelty of the BB, FRM and LQR models is that they are more appropriate for data that are bounded, as is the case in the EQ-5D-5L and SF-6D. They also accounted for the non-linearity in the relationship between the source and the target instruments.

For predicting EQ-5D utilities, the finding that the BB model is the best corroborates with previous studies in the field [20; 42]. Studies indicated that the relationship between QLQ-C30 and EQ-5D may be better understood with a non-linear model [42; 49]. It is also clear that EQ-5D is characterised by two key properties: bounded nature of the data, and piling-up of observations at one (perfect health). Consequently, the effect of predictor variables cannot be constant throughout its entire range [50]. Thus, the novelty of the BB model is that it is non-linear and more appropriate for naturally bounded data as the case in EQ-5D.

For predicting SF-6D, the EEE model is the best. This model has superior statistical properties in terms of accuracy and efficiency, which makes it a powerful and flexible mapping algorithm in health economic evaluation [41; 42]. The EEE is a more flexible method because it not only identifies an appropriate link function from the data and suggests an underlying distribution for a specific application, but also serves as a robust estimator when no specific distribution for the outcome measure can be identified [35]. That is, the EEE method performs well in terms of bias and efficiency when the distribution of the outcome variable is not known, and when there is ambiguity about the appropriate link function.

The current study differs from previous studies in several important aspects [15-21; 47; 49]. First, by comparing seven distinct econometric models, we investigated the merits of alternative analytical approaches addressing the characteristics of the data, such as censoring, non-linearity, problems of normality and heterogeneity of variance. Second, previous studies have either applied the 3L version of the EQ-5D [15-17; 19; 42; 49] or the 5L based on an interim *cross-walk* tariff [20]. In the current study, we applied the new EQ-5D-5L value set [22], which is directly elicited by members of the general public. Our goodness of fit measures in the preferred models (r^2) were better than or comparable with other mapping studies. For instance, in similar studies aiming to develop mapping relationships between QLQ-C30 and EQ-5D, the explanatory power (r^2) ranges between 0.40 to 0.75 [16-18; 20; 42; 47; 49]. For the SF-6D , the r^2 was 0.63 in Kontodimopoulos [51] and 0.75 in Wong *et al.* [21].

Furthermore, the predictive ability of our preferred model, as measured by RMSE and MAE, were close to that reported in other mapping study involving QLQ-C30 and EQ-5D-5L interim cross-walk tariff [20]. For the SF-6D, RMSE is 0.080 and MAE is 0.065 in [21], and RMSE is about 0.077 and MAE is close to 0.060 in [51]. RMSE is 0.0702 and MAE is 0.0558 in the present study, indicating better predictive performance of the preferred (EEE) model. These discrepancies could be attributable to the nature and size of the sample. Most previous mapping studies have relied on samples with only one cancer type, while our sample includes different cancers. Other potential reasons for these discrepancies might be differences in methodological approach and predictor variables used as well as variation in the target instruments employed.

Our findings are consistent with the claim that the mapping algorithms tended to under-predict the true utilities for patients in better health and over-predict utilities for those in poorer health [51-53]. This was observed in all mapping algorithms considered and can be seen in Figure 2a and 2b. This effect was small and had little influence on the overall mean in SF-6D. The reasons why such non-linearities were more pronounced in EQ-5D than SF-6D prediction might be i) the ceiling effect at the upper limit of the scale for the EQ-5D, and ii) the substantial decrements in preference weights that occur at the severe EQ-5D health states [54].

This study has several strengths. First, it explored alternative regression models that provide consistent and robust estimates under misspecification of errors related to non-normality and heteroscedasticity. To our knowledge, this is the first study to explore the predictive performance of the EEE approach in the field of mapping studies. Second, we have normalized both RMSE and MAE for differences in scale to facilitate comparison between instruments or models with different scales, which is usually ignored in other mapping studies. Third, the use of several model performance criteria demonstrate the consistency of our results. Yet, the present study is not without limitations. In all online surveys, self-selection bias is one potential problem. Although our mapping algorithms were tested on the internal dataset and performed well, further validation is warranted using external samples. The data set on which the modelling is based include subjects from six Western countries. However, the English EQ-5D-5L and the UK SF-6D utilities have been applied. If cancer patients from each country would describe their problems differently on the source (QLQ-C30) and the target (EQ-5D-5L and SF-6D) instruments, the regression coefficients in the mapping functions should be interpreted with some caution.

In summary, the QLQ-C30 can be mapped onto the EQ-5D-5L and the SF-6D utilities with good predictive accuracies. The BB regression model was preferred for EQ-5D-5L, while the more flexible EEE method in the generalized linear model was preferred for SF-6D. Thus, in the absence of generic preference-based instruments, these mapping algorithms can predict health state utilities that are required in the calculation of QALY gains, thereby enabling comparisons of the relative cost-effectiveness of cancer interventions as compared to spending the resources in other disease areas.

Compliance with Ethical Standards

Conflicts of interest: The authors declare that they have no conflict of interest.

Ethical Approval: Ethical approval was granted by the Monash University Human Research Ethics Committee (Reference No. CF11/ 3192–2011001748). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

1. Brazier, J., Ratcliffe, J., Salomon, J. A., & Tsuchiya, A. (2017). *Measuring and Valuing Health Benefits for Economic Evaluation*: Oxford University Press.
2. Harris, R. A., Washington, A. E., Nease, R. F., & Kuppermann, M. (2004). Cost utility of prenatal diagnosis and the risk-based threshold. *The Lancet*, 363(9405), 276-282. DOI: https://doi.org/10.1016/S0140-6736(03)15385-8.
3. GBD 2015 Mortality and Causes of Death Collaborators. (2016). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*, 388(10053), 1459-1544. DOI: 10.1016/s0140-6736(16)31012-1.
4. Lowy, D. R., & Collins, F. S. (2016). Aiming High — Changing the Trajectory for Cancer. *New England Journal of Medicine*, 374(20), 1901-1904. DOI: 10.1056/NEJMp1600894.
5. Hanahan, D. (2014). Rethinking the war on cancer. *Lancet*, 383(9916), 558-563. DOI: 10.1016/s0140-6736(13)62226-6.

6. Garau, M., Shah, K. K., Mason, A. R., Wang, Q., Towse, A., & Drummond, M. F. (2011). Using QALYs in Cancer. *PharmacoEconomics*, 29(8), 673-685. DOI: 10.2165/11588250-000000000-00000.
7. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., & de Haes, J. C. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*, 85. DOI: 10.1093/jnci/85.5.365.
8. Richardson, J., McKie, J., & Bariola, E. (2014). Multi attribute utility instruments and their use. In A. J. Culyer (Ed.), *Encyclopedia of health economics* (pp. 341-357). San Diego: Elsevier Science.
9. Wisløff, T., Hagen, G., Hamidi, V., Movik, E., Klemp, M., & Olsen, J. A. (2014). Estimating QALY Gains in Applied Studies: A Review of Cost-Utility Analyses Published in 2010. *Pharmacoeconomics*, 32(4), 367-375. DOI: http://10.1007/s40273-014-0136-z.
10. Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G., & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*, 20(10), 1727-1736. DOI: 10.1007/s11136-011-9903-x.
11. Rabin, R., Oemar, M., Oppe, M., Janssen, B., & Herdman, M. (2011). EQ-5D-5L User Guide: Basic Information on how to use the EQ-5D-5L Instrument: EuroQoL Group.
12. NICE (National Institute for Health and Care Excellence). (2013). Guide to the methods of technology appraisal. London: National Health Service. Retrieved September 18, 2017, from http://www.nice.org.uk.
13. Brazier, J. E., Yang, Y., Tsuchiya, A., & Rowen, D. L. (2010). A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*, 11. DOI: 10.1007/s10198-009-0168-z.
14. Dakin, H., Abel, L., Burns, R., & Yang, Y. (2018). Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. *Health and Quality of Life Outcomes*, 16(1), 31. DOI: 10.1186/s12955-018-0857-3.
15. Kim, S. H., Jo, M.-W., Kim, H.-J., & Ahn, J.-H. (2012). Mapping EORTC QLQ-C30 onto EQ-5D for the assessment of cancer patients. *Health and Quality of Life Outcomes*, 10(1), 151. DOI: 10.1186/1477-7525-10-151.
16. Kontodimopoulos, N., Aletras, V. H., Paliouras, D., & Niakas, D. (2009). Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. *Value Health*, 12. DOI: 10.1111/j.1524-4733.2009.00569.x.
17. Marriott, E. R., van Hazel, G., Gibbs, P., & Hatswell, A. J. (2017). Mapping EORTC-QLQ-C30 to EQ-5D-3L in patients with colorectal cancer. *J Med Econ*, 20(2), 193-199. DOI: 10.1080/13696998.2016.1241788.
18. McKenzie, L., & van der Pol, M. (2009). Mapping the EORTC QLQ C-30 onto the EQ-5D Instrument: The Potential to Estimate QALYs without Generic Preference Data. *Value in Health*, 12(1), 167-171. DOI: http://dx.doi.org/10.1111/j.1524-4733.2008.00405.x.
19. Versteegh, M. M., Leunis, A., Luime, J. J., Boggild, M., Uyl-de Groot, C. A., & Stolk, E. A. (2012). Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. *Med Decis Making*, 32. DOI: 10.1177/0272989x11427761.
20. Khan, I., Morris, S., Pashayan, N., Matata, B., Bashir, Z., & Maguirre, J. (2016). Comparing the mapping between EQ-5D-5L, EQ-5D-3L and the EORTC-QLQ-C30 in non-small cell lung cancer patients. *Health and Quality of Life Outcomes*, 14, 60. DOI: 10.1186/s12955-016-0455-1.
21. Wong, C. K. H., Lam, C. L. K., Wan, Y. F., & Rowen, D. (2013). Predicting SF-6D from the European Organization for Treatment and Research of Cancer Quality of Life Questionnaire Scores in

Patients with Colorectal Cancer. *Value in Health*, 16(2), 373-384. DOI: http://doi.org/10.1016/j.jval.2012.12.004.
22. Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B., & van Hout, B. (2017). Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Econ*. DOI: 10.1002/hec.3564.
23. Basu, A., & Manca, A. (2012). Regression estimators for generic health-related quality of life and quality-adjusted life years. *Med Decis Making*, 32(1), 56-69. DOI: 10.1177/0272989x11416988.
24. Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., & Gray, A. (2015). The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. *Pharmacoeconomics*, 33(10), 993-1011. DOI: 10.1007/s40273-015-0312-9.
25. Richardson, J., Iezzi, A., & Maxwell, A. (2012). Cross-National Comparison of Twelve Quality of Life Instruments: MIC Paper 1 Background, Questions, Instruments. Research Paper 76. Retrieved November 23, 2017, from https://www.aqol.com.au/papers/researchpaper76.pdf.
26. Kaambwa, B., Chen, G., Ratcliffe, J., Iezzi, A., Maxwell, A., & Richardson, J. (2017). Mapping Between the Sydney Asthma Quality of Life Questionnaire (AQLQ-S) and Five Multi-Attribute Utility Instruments (MAUIs). *Pharmacoeconomics*, 35(1), 111-124. DOI: 10.1007/s40273-016-0446-4.
27. Mihalopoulos, C., Chen, G., Iezzi, A., Khan, M. A., & Richardson, J. (2014). Assessing outcomes for cost-utility analysis in depression: comparison of five multi-attribute utility instruments with two depression-specific outcome measures. *Br J Psychiatry*, 205(5), 390-397. DOI: 10.1192/bjp.bp.113.136036.
28. Gamst-Klaussen, T., Lamu, A. N., Chen, G., & Olsen, J. A. (2018). Assessment of outcome measures for cost-utility analysis in depression: mapping depression scales onto the EQ-5D-5L. *BJPsych Open*, 4(4), 160-166. DOI: 10.1192/bjo.2018.21.
29. Chen, G., McKie, J., Khan, M. A., & Richardson, J. R. (2014). Deriving health utilities from the MacNew Heart Disease Quality of Life Questionnaire. *European Journal of Cardiovascular Nursing*, 14(5), 405-415. DOI: 10.1177/1474515114536096.
30. Lamu, A., Chen, G., Gamst-Klaussen, T., & Olsen, J. (2018). Do country-specific preference weights matter in the choice of mapping algorithms? The case of mapping the Diabetes-39 onto eight country-specific EQ-5D-5L value sets. *Qual Life Res*. DOI: 10.1007/s11136-018-1840-5.
31. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, 21(2), 271-292. DOI: http://dx.doi.org/10.1016/S0167-6296(01)00130-8.
32. Russell, D. W. (2002). In Search of Underlying Dimensions: The Use (and Abuse) of Factor Analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, 28(12), 1629-1646. DOI: doi:10.1177/014616702237645.
33. Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. DOI: Doi 10.1037//1082-989x.4.3.272.
34. Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models*: SAGE Publications.
35. Basu, A., & Rathouz, P. J. (2005). Estimating marginal and incremental effects on health outcomes using flexible link and variance function models. *Biostatistics*, 6(1), 93-109. DOI: 10.1093/biostatistics/kxh020.
36. Basu, A. (2005). Extended generalized linear models: Simultaneous estimation of flexible link and variance functions. *The Statal Journa*, 5(4), 501–516.
37. Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *J Appl Econometrics*, 11(6), 619-632. DOI: 3.0.co;2-1.

38. Ramalho, E. A., Ramalho, J. J. S., & Murteira, J. M. R. (2011). Alternative Estimating and Testing Empirical Strategies for Fractional Regression Models. *Journal of Economic Surveys*, 25(1), 19-68. DOI: 10.1111/j.1467-6419.2009.00602.x.
39. Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta Regression or Modeling Rates and Proportions. *J Appl Statist*, 31(7), 799-815. DOI: 10.1080/0266476042000214501.
40. Swearingen, C. J., Castro, M. S. M., & Bursac, Z. (2012). Inflated Beta Regression: Zero, One, and Everything in Between. Paper presented at the SAS Global Forum.
41. Ospina, R., & Ferrari, S. L. P. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6), 1609-1623. DOI: https://doi.org/10.1016/j.csda.2011.10.005.
42. Khan, I., & Morris, S. (2014). A non-linear beta-binomial regression model for mapping EORTC QLQ- C30 to the EQ-5D-3L in lung cancer patients: a comparison with existing approaches. *Health and Quality of Life Outcomes*, 12, 163. DOI: 10.1186/s12955-014-0163-7.
43. Bottai, M., Cai, B., & McKeown, R. E. (2010). Logistic quantile regression for bounded outcomes. *Statistics in Medicine*, 29(2), 309-317. DOI: 10.1002/sim.3781.
44. Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25(3), 303-325. DOI: http://dx.doi.org/10.1016/0304-4076(84)90004-6.
45. Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1), 255-268.
46. Wu, E. Q., Mulani, P., Farrell, M. H., & Sleep, D. (2007). Mapping FACT-P and EORTC QLQ-C30 to Patient Health Status Measured by EQ-5D in Metastatic Hormone-Refractory Prostate Cancer Patients. *Value in Health*, 10(5), 408-414. DOI: http://dx.doi.org/10.1111/j.1524-4733.2007.00195.x.
47. Kim, E. J., Ko, S. K., & Kang, H. Y. (2012). Mapping the cancer-specific EORTC QLQ-C30 and EORTC QLQ-BR23 to the generic EQ-5D in metastatic breast cancer patients. *Qual Life Res*, 21. DOI: 10.1007/s11136-011-0037-y.
48. Sullivan, P. W., & Ghushchyan, V. (2006). Mapping the EQ-5D Index from the SF-12: US General Population Preferences in a Nationally Representative Sample. *Medical decision making : an international journal of the Society for Medical Decision Making*, 26(4), 401-409. DOI: 10.1177/0272989X06290496.
49. Crott, R., & Briggs, A. (2010). Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *Eur J Health Econ*, 11. DOI: 10.1007/s10198-010-0233-7.
50. Ramalho, J. J. S., & da Silva, J. V. (2009). A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms. *Quantitative Finance*, 9(5), 621-636. DOI: 10.1080/14697680802448777.
51. Kontodimopoulos, N. (2015). The potential for a generally applicable mapping model between QLQ-C30 and SF-6D in patients with different cancers: a comparison of regression-based methods. *Qual Life Res*, 24(6), 1535-1544. DOI: 10.1007/s11136-014-0857-7.
52. Barton, G. R., Sach, T. H., Jenkinson, C., Avery, A. J., Doherty, M., & Muir, K. R. (2008). Do estimates of cost-utility based on the EQ-5D differ from those based on the mapping of utility scores? *Health and Quality of Life Outcomes*, 6, 51-51. DOI: 10.1186/1477-7525-6-51.
53. Versteegh, M. M., Rowen, D., Brazier, J. E., & Stolk, E. A. (2010). Mapping onto EQ-5D for patients in poor health. *Health Qual Life Outcomes*, 8. DOI: 10.1186/1477-7525-8-141.
54. Olsen, J. A., Lamu, A. N., & Cairns, J. (2017). In search of a common currency: A comparison of seven EQ-5D-5L value sets. *Health Econ*. DOI: 10.1002/hec.3606.

Appendix Table A1 EFA for the QLQ-C30 items and the SF-6D dimensions

	Factors					
QLQ-C30 items and EQ-5D-5L dimensions	EF	PF	SM	EG	LM	PN
QLQ-C30 items						
1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or suitcase?		0.652				
2. Do you have any trouble taking a long walk?		0.722				
3. Do you have any trouble taking a short walk outside of the house?		0.770				
4. Do you need to stay in bed or a chair during the day?		0.405				
5. Do you need help with eating, dressing, washing yourself or using the toilet?		0.655		-0.433		
During the last week						
6. Were you limited in doing either your work or other daily activities?					0.487	
7. Were you limited in pursuing your hobbies or other leisure time activities?					0.575	
8. Were you short of breath?						
9. Have you had pain?						0.927
10. Did you need to rest?				0.410		
11. Have you had trouble sleeping?						
12. Have you felt weak?				0.485		
13. Have you lacked appetite?			0.517			
14. Have you felt nauseated?			0.829			
15. Have you vomited?			0.932			
16. Have you been constipated?						
17. Have you had diarrhea?			0.564			
18. Were you tired?				0.546		
19. Did pain interfere with your daily activities?						0.870
20. Have you had difficulty in concentrating on things, like reading a newspaper or watching television?						
21. Did you feel tense?	0.785					
22. Did you worry?	0.819					
23. Did you feel irritable?	0.696					
24. Did you feel depressed?	0.800					
25. Have you had difficulty remembering things?						
26. Has your physical condition or medical treatment interfered with your family life?					0.960	
27. Has your physical condition or medical treatment interfered with your social activities?					0.898	
28. Has your physical condition or medical treatment caused you financial difficulties?					0.457	
29. How would you rate your overall health?				-0.747		
30. How would you rate your overall quality of life?				-0.695		
SF-6D dimensions						
1. Physical functioning		0.770				
2. Role limitation	[0.341]			[0.351]		
3. Social functioning					[0.249]	
4. Bodily pain						0.795
5. Mental health	0.683					
6. Vitality				0.821		

Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization. Loadings lower than 0.40 was suppressed, except for the 'role' and 'social functioning' dimensions of SF-6D in which the largest loading was reported in a bracket. *EM* emotional functioning, *PF* physical functioning, *SM* symptom, *EG* energy, *LM* limitation, *PN* pain, *SF-6D* short form six dimensional questionnaire, *QLQ-C30* Quality of Life Questionnaire Core 30, *EFA* exploratory factor analysis.

Appendix Table A2 EFA for the QLQ-C30 items and the SF-6D dimensions

QLQ-C30 items and EQ-5D-5L dimensions	Factors					
	EF	PF	SM	EG	LM	PN
QLQ-C30 items						
1. Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or suitcase?		0.652				
2. Do you have any trouble taking a long walk?		0.722				
3. Do you have any trouble taking a short walk outside of the house?		0.770				
4. Do you need to stay in bed or a chair during the day?		0.405				
5. Do you need help with eating, dressing, washing yourself or using the toilet?		0.655		-0.433		
During the last week						
6. Were you limited in doing either your work or other daily activities?					0.487	
7. Were you limited in pursuing your hobbies or other leisure time activities?					0.575	
8. Were you short of breath?						
9. Have you had pain?						0.927
10. Did you need to rest?				0.410		
11. Have you had trouble sleeping?						
12. Have you felt weak?				0.485		
13. Have you lacked appetite?			0.517			
14. Have you felt nauseated?			0.829			
15. Have you vomited?			0.932			
16. Have you been constipated?						
17. Have you had diarrhea?			0.564			
18. Were you tired?				0.546		
19. Did pain interfere with your daily activities?						0.870
20. Have you had difficulty in concentrating on things, like reading a newspaper or watching television?						
21. Did you feel tense?	0.785					
22. Did you worry?	0.819					
23. Did you feel irritable?	0.696					
24. Did you feel depressed?	0.800					
25. Have you had difficulty remembering things?						
26. Has your physical condition or medical treatment interfered with your family life?					0.960	
27. Has your physical condition or medical treatment interfered with your social activities?					0.898	
28. Has your physical condition or medical treatment caused you financial difficulties?					0.457	
29. How would you rate your overall health?				-0.747		
30. How would you rate your overall quality of life?				-0.695		
SF-6D dimensions						
1. Physical functioning		0.770				
2. Role limitation	[0.341]			[0.351]		
3. Social functioning					[0.249]	
4. Bodily pain						0.795
5. Mental health	0.683					
6. Vitality				0.821		

Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization. Loadings lower than 0.40 was suppressed, except for the 'role' and 'social functioning' dimensions of SF-6D in which the largest loading was reported in a bracket. *EM* emotional functioning, *PF* physical functioning, *SM* symptom, *EG* energy, *LM* limitation, *PN* pain, *SF-6D* short form six dimensional questionnaire, *QLQ-C30* Quality of Life Questionnaire Core 30, *EFA* exploratory factor analysis.

Appendix Table A3 Distributions of observed vs. predicted EQ-5D-5L and SF-6D utilities at different severity levels

Variable	Mean	SD	iqr	p1	p5	p10	p25	p50	p75	p90	p95	p99
EQ-5D-5L												
Observed	0.781	0.209	0.230	0.037	0.315	0.512	0.708	0.838	0.937	1.000	1.000	1.000
Predicted												
OLS	0.781	0.170	0.241	0.353	0.448	0.518	0.674	0.823	0.915	0.977	0.999	1.018
GLM	0.776	0.164	0.161	0.195	0.415	0.543	0.726	0.839	0.887	0.908	0.915	0.920
EEE	0.781	0.166	0.228	0.316	0.438	0.519	0.683	0.830	0.911	0.957	0.971	0.987
BB	0.778	0.176	0.204	0.266	0.386	0.509	0.700	0.833	0.904	0.957	0.972	0.982
FRM	0.781	0.171	0.226	0.294	0.416	0.502	0.689	0.843	0.915	0.948	0.958	0.966
LQR	0.793	0.174	0.186	0.215	0.396	0.511	0.731	0.860	0.917	0.941	0.948	0.955
CLAD	0.801	0.142	0.206	0.445	0.525	0.575	0.709	0.835	0.915	0.964	0.981	1.000
SF-6D												
Observed	0.686	0.133	0.206	0.395	0.482	0.529	0.593	0.673	0.799	0.852	0.894	0.965
Predicted												
OLS	0.686	0.112	0.155	0.406	0.473	0.514	0.619	0.701	0.774	0.819	0.838	0.859
GLM	0.686	0.114	0.167	0.435	0.481	0.517	0.609	0.691	0.777	0.832	0.858	0.883
EEE	0.686	0.114	0.170	0.467	0.497	0.527	0.603	0.681	0.773	0.841	0.872	0.916
BB	0.684	0.104	0.141	0.406	0.478	0.528	0.624	0.703	0.765	0.796	0.822	0.854
FRM	0.686	0.115	0.164	0.402	0.463	0.507	0.616	0.704	0.780	0.822	0.838	0.857
LQR	0.689	0.115	0.169	0.419	0.463	0.509	0.614	0.707	0.783	0.827	0.840	0.863
CLAD	0.695	0.115	0.161	0.405	0.474	0.525	0.625	0.709	0.786	0.835	0.856	0.884

p1 1st percentile, *p5* 5th percentile, …, *p99* 99th percentile, *SD* standard deviation, *iqr* inter-quantile range, *EQ-5D-5L* EuroQol five-dimensional five level questionnaire, *SF-6D* short form six-dimensional questionnaire, *OLS* ordinary least square, *GLM* generalized linear model, *EEE* extended estimating equations, *BB* (one-inflated) beta binomial regression, *FRM* fractional regression model, *LQR* logistic quantile regression, *CLAD* censored least absolute deviations. In each model, EQ-5D-5L and SF-6D were target or dependent variables.