

WITHDRAWN: Mapping the FACT-G to EQ-5D-3L utility index in cancer: data from a cross-sectional study in Guangxi, China

ZhiKui He

guang xi yi ke da xue: Guangxi Medical University <https://orcid.org/0000-0002-5693-0261>

WenJie Liang

Guangxi Medical University

WenJia Xu

Guangxi Medical University

WenXiu Huang

Guangxi Medical University

XiaoMin Wang

Guangxi Medical University

KaiYong Huang

Guangxi Medical University

Li Yang

yangli8290@hotmail.com

Guangxi Medical University <https://orcid.org/0000-0001-6663-4629>

Research

Keywords: Mapping, Cancer, EQ-5D-3L, FACT-G, Health-related quality of life, Health utility

Posted Date: June 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-615349/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

EDITORIAL NOTE:

The full text of this preprint has been withdrawn by the authors while they make corrections to the work. Therefore, the authors do not wish this work to be cited as a reference. Questions should be directed to the corresponding author.

Abstract

Objective

This study aimed to develop a function for mapping the cancer-specific instrument (FACT-G) to a preference-based measure (EQ-5D-3L) utility index for HRQoL, in which the utility scores were generated using the Chinese value set.

Method

The data are based on a cross-sectional survey of 243 patients in China with different cancer types. Cancer patients who completed the EQ-5D-3L and the FACT-G questionnaire, and patient demographics and clinical characteristics were included in this study. Regression models were used to predict the EQ-5D-3L utility index values based on four subscale scores of the FACT-G using the ordinary least squares (OLS) model, generalized linear models (GLM), censored least absolute deviations (CLAD), Tobit model, and two-part model (TPM) regression approaches. The performance and predictive power of each model were also evaluated using r^2 and adj- r^2 , mean absolute error (MAE) and root mean squared error (RMSE). Linear equating is a mapping technique that avoids regression to the mean.

Results

The introduction of the square term and the interaction term improves the accuracy of the model. The social well-being subscale of the FACT-G was not associated with the EQ-5D-3L utility index, whereas, the physical, emotional, functional well-being, and FACT-G total scores were derived when mapping the FACT-G to the EQ-5D-3L utility index. For the FACT-G, the OLS model was the best at predicting mean EQ-5D-3L values among all the regression models and has 70% explanatory power for the observed EQ-5D-3L variation. Nevertheless, the OLS model overpredicted utilities for poorer health states and underestimated utilities for those with better health. The result of the OLS model, GLM, and CLAD models were similar to the MAE, MSE, and RMSE. The results showed that the mean prediction accuracy of the OLS models was higher than that of the other models, and the OLS model's linear equating predicted values were much closer to the observed values.

Conclusion

The algorithm based on Chinese population development for mapping the FACT-G into the EQ-5D-3L utility index can be realized. We also recommend that OLS models be used to assess the economic evaluation of patients' health-related quality of life when the population is in moderate to good health for further cost-utility analysis in China.

Introduction

Cancer is considered the leading cause of human death. With the rapid growth and aging of the population, the incidence and mortality rates of cancer are rapidly growing worldwide. According to the *GLOBOCAN 2018* estimates indicate, there would be 18.1 million new cases of cancer and 9.6 million deaths from cancer [1]. Cancer has also caused serious public health problems and represents a significant economic burden in China. However, with the continuous improvement of medicine and technology, the survival rate of cancer patients has improved markedly. Although early screening and treatment, as well as advanced medical technologies, can significantly improve the survival rate of cancer patients, these can place a huge socio-psychological and economic burden on patients. Therefore, it is very important to know the cancer patient's health-related quality of life (HRQoL), which has become increasingly popular in health economic evaluations and has recently gained significant attention in cancer studies [2; 3].

HRQoL refers to the state of physical, mental, and social well-being of an individual, and an accurate impact analysis of cancer-specificity on HRQoL is expected to contribute to the health economic evaluation [4; 5]. Utility is part of quality-adjusted life years (QALYs), as QALY is the product of morbidity (measured by health state utility instruments, e.g. EQ-5D), in which is scaled such that 1 indicates full health and, 0 indicates death. It also allows for less than 0 (negative), which represents a state of health worse than death [6; 7]. The three-level EuroQol-5-dimension questionnaire (EQ-5D-3L) is a standardized instrument

used to measure preference-based HRQoL and highly recommended in health-economic evaluations[8]. Liu et al. successfully developed Chinese population-specific EQ-5D-3L health states using the time trade-off (TTO) method in 2014 [9]. Thus, we used the value set based on the Chinese general population algorithm for this study.

However, preference-based instruments are not always available in clinical trials because many dimensions may not be relevant or sensitive to therapeutic effects. The Functional Assessment of Cancer Therapy-General (FACT-G) is one of the most widely used cancer-specific HRQoL instruments as well as a non-preference-based measure for cancer patients [10]. The FACT-G is used in clinical trials to assess the quality of life (QoL), with higher values representing a better QoL, and its reliability and validity have been proven [11]. Disease-specific instruments are mostly used to measure the HRQoL rather than generic preference-based measures as they can provide more specific details about the patients' assessment of a particular disease. However, it does not calculate the health utility scores and QALYs directly so that limit the development of health economics research [12]. One solution is to use the development of a mapping algorithm that maps scores from the HRQoL data collected by non-preference-based instruments to general preference-based instruments [13]. Growing literature studies have suggested that a mapping function from disease-specific instruments to generic preference-based measures using regression models in health economic evaluation is available [4; 13–15].

There are a few studies mapping from FACT-G to EQ-5D-3L that have been developed and evaluated using regression model analysis in health economic research. A study from Canada performed a mapping function to both the EQ-5D-3L and SF-6D health utility indices from the FACT-G [16]. Meanwhile, a study was conducted to evaluate the validity of both FACT-G and preference-based instruments (including the EQ-5D-3L, SF-6D, HUI-2, and HUI-3) in assessing cancer severity levels in Canadian patient data [17]. Another mapping from the FACT-G to the EQ-5D-3L health utility index in Singapore shows that a single equation can be applied to different versions of the FACT-G [18]. However, no studies are available to convert the FACT-G to EQ-5D-3L with mapping algorithms in Chinese population due to the inconsistency between utility value sets of different countries [19]. Therefore, it is necessary to develop a health utility value mapping from FACT-G to EQ-5D-3L for Chinese patients. Some studies have shown that mapping can improve the accuracy of models with socio-demographic and clinical factors among the instruments, thus affecting health utility in cost-utility analysis [10]. These studies also compare different regression methods with more accurate models [14].

In 2014, Fayers and Hays have pointed out that regression-based models typically under-predict high scores and over-predict low scores when mapping profile-based measures to preference-based measures, because of regression to the mean [20]. According to a review of mapping studies, predicted values from the mapping functions tend to have lower levels of variance than the original observed values due to regression to mean[21]. Thus, they have suggested the use of linking strategies such as simple linear equating, equipercetile equating, or item-response theory (IRT) methodology as alternatives. While regression-based models seek to predict the most likely true preference-based score using the profile-based score, linking seeks to find the preference-based score that is equivalent to the profile-based score by aligning the score distributions of the two scales [22]. In a few mapping studies, linear equating was used to predict utility using a regression-based method and then scale aligning between predicted and observed values to force them to have the same mean and variance [23].

In general, the objective of the present study was to develop a mapping algorithm to estimate EQ-5D-3L health utility values from the FACT-G for the Chinese population, select the best appropriate model to better estimate the patients' health status from a single assessment using the FACT-G, and make recommendations for future mapping studies.

Method And Materials

Study design and data collection

The Cancer Screening Program in Urban China, a major public health service project supported by the central government of China beginning in August 2012, was designed to screening programs for lung, breast, colorectal, liver, stomach, and esophageal cancers [24]. Meanwhile, a multicenter cross-sectional study was conducted in 12 provinces between September 2013 and December 2014, with appropriate screening interventions targeted at specific types of cancer [25]. The study

protocol was approved by the Institutional Review Board of the Cancer Hospital of the Chinese Academy of Medical Sciences (Approval No. 15-071/998). All participants gave their [written] informed consent.

This study involved 243 subjects according to the following criteria: 40-74 years old; be diagnosed with either lung, breast, stomach, esophagus, colorectal, or liver cancer; without any mental disorder and able to understand the survey procedure and complete the survey questionnaire; completed both the EQ-5D-3L and FACT-G scales and subscales. Data collection came from one center and multiple oncology hospitals in China. The questionnaire survey was conducted through a face-to-face interview between the investigator and the followed-up subjects.

For information regarding age, gender, marital status, level of education, family population, employment, family financial pressure, and significant life events, patients were required to complete a health and demographic questionnaire. In addition, clinical characteristics were retrieved from medical records, such as tumor site, treatment protocols, age at diagnosis, time point of the survey and clinical stage for cancer according to the 6th edition of the American Joint Committee on Cancer/International Union Against Cancer staging system [35].

HRQoL Instruments

EQ-5D-3L Scale

The EQ-5D-3L scale is a generic preference-based instrument that provides a simple and universal health measurement method for clinical and economic evaluation. The EQ-5D-3L descriptive system consists of five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/ depression), plus a vertical, 0-100-point visual analogue scale for rating the overall health status. The dimensions are characterized by three levels of health (i.e. no problems, some or moderate problems, and extreme problems) [26]. The EQ-5D-3L index scores were calculated using an algorithm based on societal preferences from the general population-based valuation. Due to the Chinese value set available in the crosswalk project, the most widely used utility algorithm was based on a TTO survey of 1147 Chinese respondents in China [9]. Using this value, the EQ-5D-3L utility index ranges from -0.149 to 1, where 1 indicates full health, 0 indicates a state equivalent to death, and a negative value implies that the respondent's health state is worse than death [27].

FACT-G Scale

The FACT-G produces four subscale scores consists of 27 items that reflect the patient's QoL: physical wellbeing (PWB) (7 items), social/family well-being (SFWB) (7 items), emotional well-being (EWB) (6 items), and functional well-being (FWB) (7 items) [28]. All items were rated on a 5-point Likert scale, with higher scores indicating better HRQoL. An overall score and four dimension scores are obtained by summing the responses to the individual items they comprise. The reliability and validity of the instrument have been well demonstrated in cancer trials and clinical settings.

Statistical analyses

The objectives of statistical analysis is to apply a direct mapping algorithm by regressing the EQ-5D-3L utility into the FACT-G domain scales (i.e. physical, emotional, functional, and social/family) to predicted the EQ-5D-3L health utility from different variables. Similarly, the squared terms and interaction terms of subscale scores were explored in our study. In previous studies [29; 30], the majority of the health utility scores had non-normal distributions with negative skew, as well as a ceiling effect, which violated the assumptions of the ordinary least squares method [OLS]. Despite the fact that its theoretical assumptions were not available, OLS outperformed other regression-based models [30]. In our sample, the distribution of EQ-5D-3L utility is negatively skewed, and there is a large ceiling effect in EQ-5D-3L (62.7% at ceiling). As a result, the different regression models tend to find the best fit for mapping the FACT-G to the EQ-5D-3L health utility.

Here, five functions mapping from the FACT-G to the EQ-5D-3L (i.e., the ordinary least squares [OLS] model, generalized linear model [GLM], Tobit model, the censored least absolute deviations [CLAD], and the two-part model [TPM]) were included in this study. Most of the previous studies suggested that, the OLS model was considered the most commonly used method in

mapping algorithms, but it may not be appropriate when preference-based scores are highly skewed [16]. The ceiling effect may also invalidate the normality assumption of OLS [31]. The GLM, on the other hand, relaxes the assumption of the OLS, allowing for a skewed distribution of utility values and accommodate the non-linear relationship with the predictor variables. We wonder if the GLM produced more accurate predictions than the OLS. The Tobit model is an alternative model that accounts for the ceiling effect, thus limiting predictions within a credible range. However, it is sensitive to normal distribution and heteroscedasticity. The CLAD model assumes that the median is more resistant than the mean to ceiling effects and is a possible solution to the heteroscedasticity problem as well, which minimizes the sum of absolute differences between observed and predicted values [31; 32]. The TPM is specifically designed to deal with limited dependent variables, which divide the data into two parts to predict responders in perfect health and those who are, not. The TPM with logistic regression is used to predict the probability of EQ-5D-3L utility at the ceiling in the first part, a truncated OLS to predict EQ-5D-3L index for those individuals whose EQ-5D-3L utility is below the ceiling in the second part and combined they obtain the overall utility value [32; 33]. The OLS model, GLM, Tobit model, CLAD model, and TPM were performed in five different models. We increased the squared terms and the interaction terms to improve the model accuracy for this study, as suggested in the literature [4]. Model 1 uses the FACT-G overall scores to regress the EQ-5D-3L utility indices, Model 2 uses all domain scores on the FACT-G. Model 3 includes only statistically significant domains. Statistically significant terms from Model 2 are equivalent to Model 3. Then Model 4 is Model 3 + their squared terms; Model 5 is Model 4 + interaction of linear terms. The covariate adjustment (i.e., age, gender) was omitted from in this study, because it may reduce the practical value of the developed mapping function in the regression analysis [30].

The purpose of prediction is usually to predict the most likely true preference-based score using the profile-based score that is known about the respondent. However, regression-based mapping model predictions result in biased estimates with lower levels of variance than the original observed values due to regression to mean [20]. Simple linear equating, which involves equating the mean and standard deviation of the two scales, can help to alleviate the typical problem of over-prediction of low scores and under-prediction of high scores [34]. Linear equating attempts to find a preference-based score that is equal to the score based on the profile by aligning the score distributions of the two scales [22]. As a result, we used linear equating force to predict the EQ-5D-3L preference-based scores of the regression models that had been linearly converted to have the same mean and standard deviation as the observed EQ-5D-3L scores. In other words, given observed EQ-5D-3L utility and predicted utility (Y_R), and predicted linear equating (Y_{LE}) estimated values are as follows: Where μ_R and σ_R were the mean and standard deviation of the observed EQ-5D-3L utility scores, and μ_{LE} and σ_{LE} were the mean and standard deviation of the predicted EQ-5D-3L utility scores obtained from the regression model. Then, using scatter plots and the Bland-Altman plot, we recoded predictions (equivalent) scores and observed scores to show the difference between them [23].

We calculated the goodness of fit of each model to assess how well the responses to the FACT-G predicted EQ-5D-3L utility. Examining the difference between predicted and observed values to the predictive performance of regression-based models is a better method for evaluating mapping functions [21]. Model goodness of fit was measured using mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), Akaike information criteria (AIC), and Bayes information criteria (BIC) to examine the differences between mean observed and predicted EQ-5D-3L utility, in which lower values indicate better model performance. The coefficient of determination, R^2 , and adjusted R^2 also estimated how well the model explained the values in OLS. However, it is not available for other regression models. Instead, we computed the square of the correlation coefficient (r) between the observed and predicted values of each model, with r^2 being equivalent to R^2 in OLS [19]. To penalize the complexity of the model, we defined the adjusted r^2 as follows: adjusted

$$r^2 = 1 - \frac{(n-1)}{(n-p-1)} (1 - r^2),$$

where n represents the sample size and p is the number of parameters in the model. Predictive ability was evaluated by a paired t -test to compare the differences in the distributions between the observed and mapped EQ-5D-3L utility scores. The different EQ-5D-3L utility scores from different models with demographic and clinical features were examined by non-parametric analysis. Moreover, we selected the lowest MAE/RMSE and the highest r^2 and adjusted r^2 as the best performing models. The EQ-5D-3L observed values and predicted utility values were compared in patients with different demographic and clinical characteristics by the Wilcoxon test and the Kruskal-Wallis H test in the different models. Finally,

scatter plots and the Bland-Altman plot were used to visualize each model's predictive performance between observed and predicted EQ-5D-3L scores. All statistical analyses were performed in STATA version 14.1 and all hypothetical tests were two-tailed, and p -value < 0.05 was considered statistically significant in this study.

Results

A total of 243 cancer patients were included in this analysis. Demographic and clinical features are summarized in Table 1. The average age of participants was 56.34 years (SD = 8.36 years) and the majority were female (71.6%). Among all age groups, the 55-59 years age group (19.3%) comprising the largest proportion of participants. Most of the cancer patients were married (80.7%) and had completed secondary education or less (72.9%). Most (55.1%) did not place financial stress on the family because of the illness, and the majority (94.7%) reported no significant life events. Of all the patients, 46 (18.9%) patients underwent surgical treatment, 122 (50.2%) patients received heteropathy, and 75 (30.9%) patients received other therapies. There were no statistically significant differences ($p > 0.05$) in the demographic and clinical characteristics of the patients except for family financial pressure and the significant life events, which were examined by the Wilcoxon test and the Kruskal-Wallis H test from the EQ-5D-3L, as appropriate.

Table 1 Patient characteristics at baseline (N = 243)

Characteristics of the sample	N (%)	<i>EQ-5D-3L median(Q25,Q75)</i>	<i>p</i>
Age: mean (SD)	56.34(8.36)	-	
Gender			
Male	57(28.4)	1(0.875,1)	0.799
Female	137(71.6)	1(0.875,1)	
Age, years			
<44	23(9.5)	1(0.875,1)	0.674
45–49	36(14.8)	1(0.875,1)	
50–54	46(18.9)	1(0.876,1)	
55–59	47(19.3)	1(0.875,1)	
60–64	40(16.5)	1(0.875,1)	
65–69	35(14.4)	1(0.875,1)	
>70	16(6.6)	0.938(0.872,1)	
Marital status			
Never married or cohabit	28(11.5)	0.875(0.869,1)	0.092
Married	196(80.7)	1(0.875,1)	
Widowed/divorced	19(7.8)	1(0.869,1)	
Level of Education			
Primary or below	21(8.6)	1(0.856,1)	0.327
Secondary	177(72.9)	1(0.875,1)	
Postsecondary	45(18.5)	1(0.887,1)	
Complication			
Yes	9(3.7)	1(0.875,1)	0.763
No	234(96.3)	1(0.875,1)	
Medical insurance			
Urban employees	110(45.3)	1(0.875,1)	0.772
Urban residents	89(36.3)	1(0.875,1)	
New rural cooperative scheme	35(14.4)	1(0.869,1)	
Other	9(3.7)	1(1,1)	
Occupation			
In employed	72(29.6)	1(0.875,1)	0.567
Self-employed entrepreneurs	93(38.3)	1(0.869,1)	
Unemployed/Retired	78(32.1)	1(0.875,1)	
Family financial pressure			
Almost no effect	134 55.1	1(0.875,1)	<0.001

Barely acceptable	46 18.9	1(1,1)	
under pressure	37 15.2	0.887(0.869,1)	
under great pressure	26 10.7	0.869(0.783,1)	
Significant life events			
Yes	13 5.3	1(0.875,1)	0.008
No	230 94.7	0.869(0.869,1)	
Tumor site			
Lung	28(11.5)	1(0.875,1)	
Breast	41(16.9)	1(0.869,1)	0.798
Colorectal	26(10.7)	1(0.875,1)	
Esophageal	25(10.3)	1(0.875,1)	
Stomach	42(17.3)	1(0.869,1)	
Hepatobiliary	81(33.3)	1(0.875,1)	
House income, CNY			
<40,000	66(27.2)	1(0.869,1)	0.592
40,000-79,999	112(46.1)	1(0.869,1)	
>80,000	65(26.7)	1(0.875,1)	
Treatment protocols			
Surgical treatment	46(18.9)	1(0.875,1)	0.546
heteropathy	122(50.2)	1(0.875,1)	
others	75(30.9)	1(0.875,1)	
EQ-5D-3L,mean(SD)	0.935(0.099)	1(0.875,1)	-

P Wilcoxon rank-sum tests for two categories or KruskalWallis test for more than two categories, SD standard deviation, EQ-5D-3L Three-level EuroQol-5-dimension questionnaire

The distributions of the EQ-5D-3L utility index and the FACT-G total and subscale scores are described in Table 2. The mean values of EQ-5D-3L and FACT-G utility scores were 0.935 and 82.7, respectively. The value of EQ-5D-3L ranged from 0.364 to 1.000, and the median was 1.000, with 62.7% of the subjects having the highest score. All FACT-G scores reached their ceiling levels but the FACT-G total score was negligible, with notable values for the PWB (27.2%), EWB (6.2%), SFWB (7.8%), FWB (5.8%), and FACT-G total (0.8%). Cronbach's α of all scales exceeded the threshold value ($\alpha \geq 0.7$), which was considered satisfactory. Likewise, all the scales exceeded the threshold for good reliability ($\alpha \geq 0.8$), except for the SFWB subscale ($\alpha = 0.683$). Both the EQ-5D-3L utility and FACT-G scores were negatively skewed.

Table 2 Descriptive Statistics for the FACT Scales and EQ-5D-3L Scales scores

Scores	N	Mean	SD	Median	Min	Max	Celling effect (%)	Cronbach's α
EQ-5D-3L	243	0.935	0.099	1.000	0.364	1.000	62.7	-
Physical	243	24.5	4.2	26	4	28	27.2	0.865
Emotional	243	19.7	5.9	20	1	28	6.2	0.886
Social	243	19.8	3.1	21	10	24	7.8	0.683
Functional	243	18.7	6.3	20	0	28	5.8	0.934
FACT-G	243	82.7	14.8	86	42	108	0.8	0.921

SD standard deviation, Min minimum, Max maximum, EQ-5D-3L Three-level EuroQol-5-dimension questionnaire, FACT-G Functional Assessment of Cancer Therapy-General, Cronbach's α values > 0.7 indicate adequate scale reliability

Spearman's rank correlation coefficients between the EQ-5D-3L and FACT-G (including the total and four subscales) are shown in Table 3. Most of the scores showed moderate and high correlations. The correlation coefficient between the EQ-5D-3L utility index and FACT-G total scores was 0.5382. The correlation coefficient between SFWB and other scores showed negligible correlations except for the FACT-G total, EWB, and FWB scores. All correlations are significant at the 0.05 level after Bonferroni correction.

Table 3 Convergent validity of EQ-5D-3L and FACT-G using Spearman's rank correlation

Dimension	EQ-5D-3L	FACT-G	PWB	SFWB	EWB	FWB
EQ-5D-3L	1.0000					
FACT-G	0.5382*	1.0000				
PWB	0.7307*	0.6305*	1.0000			
SFWB	0.1295	0.7480*	0.0698	1.0000		
EWB	0.5331*	0.7195*	0.5948*	0.3028*	1.0000	
FWB	0.3889*	0.8935*	0.3861*	0.6325*	0.5129*	1.0000

EQ-5D-3L Three-level EuroQol-5-dimension questionnaire, FACT-G Functional Assessment of Cancer Therapy-General, PWB Physical well-being, SFWB Social/family well-being, EWB Emotional well-being, FWB Functional well-being

All correlations are significant at 0.05 level after Bonferroni corrections:* $p < 0.05$.

Tables 4 and 5 present the results of the regression analyses using the Chinese value set. Model 2 performed better than Model 1 within each domain concerning the FACT-G scores. Although Model 3 was relatively consistent with Model 2, Model 3 included only statistically significant domains; hence, Model 3 was more concise. Although the model with squared terms and interaction terms performed better, effectively improving the accuracy of models and performance in Models 4 and 5, some of these terms were not significant and the coefficient estimates were small. The results of r^2 , adjusted r^2 , MAE, MSE, and MAPE of the OLS, CLAD, and GLM models were similar. The OLS model generally had the largest adjusted r^2 (0.2867 to 0.6085), and the TPM had the smallest adjusted r^2 (0.5129 to 5444). Among the five models, Model 5 had the largest adjusted r^2 using different regression methods compared to the others, which significantly improved the fitting degree of the model. The OLS model also had the smallest AIC and BIC values, ranging from (-649.1294 and, -614.1988) to (-511.148 and, -504.162), respectively. In the sample, the models with lowest MAE and RMSE were the CLAD (0.0423) and the OLS (0.0624) models, respectively. In terms of the accuracy of prediction in MAE, the CLAD model performed better than the other models. The best performing algorithms were identified for each scale according to the lowest MAE/RMSE scores, namely Models 5 for FACT-G in the CLAD model. In terms of all criteria, the OLS model performed the best among the regression-based models, while the

Tobit model performed the worst. For observed utilities ≥ 1 , the predictive accuracy of OLS, GLM, and CALD models was fairly similar. This result was also supported by scatter plots of observed and predicted EQ-5D-3L utility for best fitting models in Fig. 1. Broken line is the best fitting regression line between observed value and predicted value. The facts were also evident in Bland-Altman plots. Compared with other models, the linear equating model has similar goods-of-fit results (RMSE = 0.2481, MAE = 0.0516, MSE = 0.0047, MAPE=0.0590). In terms of raw Adj r^2 , MSE, MAE, and RMSE, OLS model regression performed better than other models, revealing superior predictive accuracy. However, after scale adjustment, the best mapping algorithm predicted tended to reduce the bias of estimated values.

Table 4 Regression of EQ-5D-3L utility index upon FACT-G scores

Variable	OLS				
	Model1	Model2	Model3	Model4	Model5
FACT-G Total	0.004*				
PWB		0.015*	0.015*	0.038*	0.014
SFWB		-0.001			
EWB		0.004*	0.004*	0.006*	0.005*
FWB		0.001*	0.001*	0.001*	0.013*
PWB^2				-0.001*	-0.001*
EWB^2				-0.0002*	-0.001*
FWB^2				0.0001*	0.0001
PWB*EWB					0.002*
PWB*FWB					-0.0004
EWB*FWB					-0.000
Constant	0.636*	0.477*	0.475*	0.232	0.410*
R-squared	0.2896	0.5545	0.5544	0.5767	0.6230
Adj R-squared	0.2867	0.5470	0.5489	0.5660	0.6085
Root MSE	0.0842	0.0671	0.0670	0.0657	0.0624
AIC	-511.148	-618.5225	-620.5041	-626.9642	-649.1294
BIC	-504.162	-601.0572	-606.5318	-602.5127	-614.1988
MAE	0.0614	0.0482	0.0482	0.0488	0.0448
MSE	0.0070	0.0044	0.0044	0.0042	0.0037
MAPE	0.0737	0.0570	0.0570	0.0569	0.0522

Table 4 Regression of EQ-5D-3L utility index upon FACT-G score (*Continued*)

Variable	GLM				
	Model1	Model2	Model3	Model4	Model5
FACT-G	0.004*				
PWB		0.017*	0.017*	0.049*	0.022*
SFWB		0.000			
EWB		0.001*	0.001*	-0.007*	0.003*
FWB		0.001*	0.001*	0.001*	0.013*
PWB^2				-0.001*	-0.001*
EWB^2				0.000	-0.001
FWB^2				0.000	0.000

PWB*EWB						0.002*
PWB*FWB						-0.001*
EWB*FWB						-0.000
Constant	0.632*	0.467*	0.470*	0.238	0.345*	
R-squared	0.2896	0.5506	0.5508	0.5703	0.6176	
Adj-squared	0.2867	0.5430	0.5451	0.5593	0.6028	
Root MSE	0.2726	0.2406	0.2401	0.2439	0.2336	
AIC	456.669	461.7348	459.735	465.5624	471.3904	
BIC	463.6551	479.2001	473.7073	490.0138	506.321	
MAE	0.0614	0.0482	0.0481	0.0499	0.0453	
MSE	0.0070	0.0045	0.0044	0.0043	0.0038	
MAPE	0.0736	0.0567	0.0566	0.0578	0.0523	

FACT-G Functional Assessment of Cancer Therapy-General, OLS Ordinary least squares, GLM Generalized linear models, PWB Physical well-being, SFWB Social/family well-being, EWB Emotional well-being, FWB Functional well-being, MAE Mean absolute error, RMSE Root mean squared error, MSE Mean square error; AIC Akaike information criterion; BIC Bayesian information criterion, MAPE Mean absolute percentage error

* $p < 0.05$

Table 5 Regression of EQ-5D-3L utility index upon FACT-G scores

Variable	Tobit			CLAD			TPM		
	Model 3	Model 4	Model 5	Model 3	Model 4	Model 5	Model 3	Model 4	Model 5
FACT-G									
PWB	0.022*	0.015*	-0.036*	0.01354*	0.0583*	0.0193*	0.010*	0.055*	0.048*
EWB	0.012*	-0.020*	-0.011*	0.00271*	0.04653*	0.0231*	-0.003*	-0.000	0.001
FWB	0.003*	-0.001*	0.006*	0.00039*	-0.00076*	0.0236*	0.002*	-0.000	0.012
PWB^2		0.000*	-0.000		-0.0011*	-0.0009*		-0.001*	-0.001*
EWB^2		0.001	-0.002		-0.0011*	-0.0017*		-0.000	-0.000
FWB^2		0.000	0.000		0.00003*	0.00007*		0.000	0.000
PWB*EWB			0.004*			0.0023*			0.001
PWB*FWB			-0.000			-0.0005			-0.001
EWB*FWB			-0.000			-0.0006			-0.000
Constant	0.192*	0.549	1.008*	0.5587*	-0.3047	0.1512	0.625*	0.222	0.197
r^2	0.5450	0.5264	0.5616	0.5528	0.5484	0.5690	0.5189	0.5501	0.5904
Adj r^2	0.5393	0.5144	0.5447	0.5472	0.5369	0.5524	0.5129	0.5176	0.5444
Root MSE	0.3606	0.3687	0.3659	0.0700	0.06924	0.0708	0.1079	0.0605	0.0588
AIC	31.5941	35.6972	22.0049	NA	NA	NA	-218.219	-242.7056	-245.1569
BIC	49.0594	63.6417	60.4286	NA	NA	NA	-208.1297	-225.2069	-220.1588
MAE	0.1159	0.1201	0.1181	0.0469	0.0441	0.0423	0.0952	0.1114	0.1109
MSE	0.0172	0.0187	0.0184	0.0048	0.0047	0.0048	0.0115	0.0159	0.0158
MAPE	0.1279	0.1320	0.1284	0.0574	0.0529	0.0506	0.1021	0.1167	0.1160

FACT-G Functional Assessment of Cancer Therapy-General, CLAD Censored least absolute deviations, TPM two-part model, PWB Physical well-being, SFWB Social/family well-being, EWB Emotional well-being, FWB Functional well-being, MAE Mean absolute error, RMSE Root mean squared error, MSE Mean square error; AIC Akaike information criterion; BIC Bayesian information criterion, MAPE Mean absolute percentage error

* $p < 0.05$

The distribution of the observed and predicted utility values were summarized in Table 6. The observed mean EQ-5D-3L utility value was 0.9353, and the median, 90th percentile, and maximum performed the boundary values on account of the ceiling effect of 63.7%. The mean of predicted values based on the OLS model is the same as the observed values, and the GLM and CLAD models tended to be much closer to the observed values. However, the TPM had a poor prediction of the means. The Tobit model had larger mean and median predicted values than the observed ones. Except for Tobit, the median of predicted values was smaller than the observed values. Based on the five regression models, the predicted values exhibited a smaller standard deviation, larger 10th percentile and minimum, except for the Tobit model and TPM, which underestimated and overestimated full health, respectively. As expected, the predicted values by each model had a closer 90th percentile and surpassed the upper bound than observed values, except for the TPM. We also used the Wilcoxon matched-pairs signed-ranks test to examine the distribution between the predicted and observed values in each regression model, which showed statistical significance except for OLS 4, OLS 5, and GLM 3 (all $p < 0.05$). However, the OLS model showed that over-predicting at severe

health states and under-predicting at better health. Similarly, the standard deviation of estimated scores was always much lower for the OLS model than the observed standard deviation. When the linear equated scores were used for model OLS 5, the predicted values were closer to the means of the actual values, with a standard deviation that best matched the observed value and significantly reduced under-prediction for high scores and significantly reduced over-prediction for low scores (Fig. 2).

Table 6 Descriptive summary of EQ-5D-3L utility indices derived from observed and predicted regression models

Model	Mean	SD	Min	P10	Median	P90	Max
Observed	0.9353	0.0997	0.3640	0.7830	1.0000	1.0000	1.0000
OLS3	0.9353	0.0742	0.6390	0.8254	0.9619	1.0001	1.0148
OLS4	0.9353	0.0757	0.5173	0.8390	0.9651	0.9909	1.0086
OLS5	0.9353	0.0787	0.4726	0.8268	0.9627	0.9983	1.0046
GLM3	0.9355	0.0779	0.5990	0.8209	0.9651	1.0015	1.0125
GLM4	0.9355	0.0772	0.4407	0.8489	0.9630	0.9890	1.0087
GLM5	0.9354	0.0798	0.4055	0.8308	0.9640	0.9971	1.0006
Tobit3	1.0323	0.1309	0.5893	0.8265	1.0772	1.1548	1.1903
Tobit4	1.0355	0.1350	0.6154	0.8269	1.0794	1.1709	1.2265
Tobit5	1.0340	0.1409	0.4798	0.8167	1.0682	1.1876	1.2310
CLAD3	0.9510	0.0674	0.6796	0.8507	0.9753	1.0101	1.0225
CLAD4	0.9593	0.0831	0.4307	0.8677	0.9953	1.0003	1.0024
CLAD5	0.9494	0.0837	0.6070	0.8322	0.9872	1.0055	1.0484
TPM3	0.8591	0.0441	0.6612	0.7971	0.8735	0.8972	0.9519
TPM4	0.8408	0.0440	0.4638	0.8137	0.8472	0.8736	0.9558
TPM5	0.8407	0.0439	0.4689	0.8176	0.8465	0.8687	0.9471
OLS (Ev)	0.9353	0.0945	0.4659	0.6703	0.9625	1.0421	1.0429

OLS (Ev) model that applies simple linear equating to the predicted values, OLS Ordinary least squares, GLM Generalized linear models, CLAD Censored least absolute deviations, TPM two-part model, SD standard deviation, Min minimum, Max maximum, P10 10th percentile, P25 25th percentile, P75 75th percentile, P90 90th percentile

Table 7 presents the predicted mean observed and predicted EQ-5D-3L utility values with the statistically significant demographic and clinical characteristic features in the best models from the different regression algorithms. Compared with TPM 3 and Tobit 3, the estimated health utilities from OLS 5, GLM 5, and CLAD 5 were closer to the observed values from the EQ-5D-3L. We also found that all models tended to overpredict the higher top and bottom end of the EQ-5D-3L utility: because a few responders in the FACT-G data set reported severe problems. In addition, the OLS model predictive performance was more accurate than other models under the influence of heteroscedasticity and various misspecification confounding factors. Finally, the predicted EQ-5D-3L utilities for the OLS5 model using simple linear equating were closer to the mean of the actual values.

Table 7 Summary of observed EQ-5D-3L values of different model performance statistics for the Best-performing FACT-G models

Model validation	Observed values	Linear equated values(OLS5)	Predicted values				
			OLS5	GLM5	Tobit3	Clad5	TPM3
Family financial pressure							
Almost no effect	0.9481	0.9567	0.9522	0.9499	1.0657	0.9642	0.8674
Barely acceptable	0.9618	0.9519	0.9485	0.9485	1.0547	0.9628	0.8674
Under pressure	0.9054	0.88772	0.8986	0.9094	0.9644	0.9270	0.8387
Under great pressure	0.8652	0.8622	0.8776	0.8750	0.9166	0.8806	0.8231
ANOVA	F=17.087	F=12.50	F=33.733	F=23.586	F=43.945	F=22.96	F=31.188
<i>P</i>	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001
Significant life events							
No	0.9393	0.9387	0.9380	0.9381	1.0377	0.9518	0.8599
Yes	0.8658	0.8749	0.8877	0.8879	0.9367	0.9057	0.8302
ANOVA	F=6.850	F=5.712	F=5.231	F=4.956	F=7.518	F=3.780	F=7.874
<i>P</i>	<i>P</i> =0.009	<i>P</i> =0.018	<i>P</i> =0.023	<i>P</i> =0.027	<i>P</i> =0.007	<i>P</i> =0.048	<i>P</i> =0.018

OLS Ordinary least squares, GLM Generalized linear models, CLAD Censored least absolute deviations, TPM two-part model

Discussion

The overall objective was to examine the feasibility and effectiveness of mapping algorithms, as well as the conditions under which circumstances they should be considered, and to draw lessons learned for future mapping studies [21]. We developed an algorithm that maps the EQ-5D-3L health utility index of the general cancer population from the FACT-G based on the data collected from a cross-sectional study in China. The current study suggested that the consistency between the predicted and observed EQ-5D-3L utility scores was feasible among five models algorithms. Meanwhile, it also confirmed that the FACT-G no-preference-based scores can estimate EQ-5D-3L health utility scores by using a mapping function for HRQoL, in which the utility scores were generated using the Chinese value set. Our findings suggest that the OLS model is better for predicting EQ-5D-3L utilities compared to other models in terms of goodness of fit and model performance, which is consistent with most existing research findings.

In this study, the coefficients of PWB, EWB, and FWB were significant in most models for all the regression algorithms, whereas, the coefficient of SFWB was not significant in most models. Hence, the SFWB score of the FACT-G was not included in the regression model. We found that the SFWB was not statistically significant and showed weak correlations with the EQ-5D-3L utility index compared to other FACT-G subscale. Previous research showed that mapping studies tended to confirm the predictive ability of health utility more easily when exploring the correlation between the EQ-5D-3L and FACT-G scales [3; 31; 35]. Furthermore, the SFWB may not be entirely captured in the mapping be attributed to the nature of the EQ-5D-3L in that it doesn't cover the social well-being aspect explicitly [16]. Some of the previous study also showed that the SFWB was also not statistically associated with the EQ-5D-3L utility index in regression models, with mapping studies of FACT-P [8], FACT-L [31], FACT-G [18], and FACT-B [33].

In this study, we found that the OLS model had the largest r^2 and adjusted r^2 among the regression models. The r^2 and adjusted r^2 values of all models were larger than 0.5, except for Model 1, which indicated that the model had good explanatory power. The validity of such models in terms of goodness-of-fit and error of prediction, as they are highly variable. Such variability is clearly observed in the present study as well: RMSE varies between 0.0624 for OLS, 0.234 for GLM and to 0.366 for Tobit. Similar pattern observed for MAE. In previous studies, the model's explanatory power ranged from 0.417 to 0.909 in terms of r^2 [13]. The r^2 of model OLS 5 reached 0.623, indicating that the model performed well. The predictive performance of the models is to examine the difference between the predicted and observed values for assessing the mapping algorithm by MAE and RMSE. From reported results (Tables 4 and 5), both CLAD and TPM performed much better than Tobit on the three criteria set by the R2, RMSE and MAE. CLAD also performed better than GLM on both RMSE and MAE.

More than half of the observations in this study reached the EQ-5D-3L ceiling of 1. As a result, various regression methods were employed in an attempt to solve the problem of estimating patients' utility from descriptive HRQOL data, which was hampered by the ceiling effect. Despite the fact that the data had a severe skewness distribution and a ceiling effect, we attempted to solve the problem by using different regression models to estimate the EQ-5D utility. GLM allows the skewed distribution of utility values. Tobit and CLAD models are appropriate for censored or bounded data and may be used if the utility scores exhibit a ceiling effect, indicating that a large proportion of subjects are in full health with a utility score of 1 [32]. In a previous study [21], OLS, Tobit, and CLAD were used, and it was discovered that CLAD performed the best in terms of mean prediction error, while Tobit performed the worst. Despite the fact that 46% of the utility values were at the ceiling level, which should favor CLAD over OLS, the OLS and CLAD produced very similar mapping results [36]. In terms of MAE, MSE, and MAPE, the OLS, GLM, and CLAD models predicted EQ-5D-3L utility quite similarly. Interestingly, the OLS regression model outperformed nearly all goodness-of-fit criteria in predicting EQ-5D-3L utilities, despite the fact that its theoretical assumptions may be incorrect.

Most of the previous literature studies have been published to explore the predictive performance of the model when developing a mapping function to estimate EQ-5D-3L preference-based health utility values from a disease-specific measure of FACT-G for cancer patients in cost-utility analysis. These studies show that mapping models focus more on the selection criteria of model performance and prediction ability. The most appropriate model also depends on the data and the way they are applied. Some mapping studies have reported that the predictive performance of disease-specific measures was achieved by using statistical criteria and standards based on the MAE, MSE, and RMSE criteria to compare the difference between the predicted and the observed values [2; 4; 16]. However, although the values of r^2 , MAE, and RMSE were higher than those of previous studies, the overall predictive ability was not satisfactory. This could be due to differences observed between studies, which could be attributed to differences in the target instrument used and variations in the additional covariates used to predict EQ-5D-3L utility values [34]. We used the mean scores of the FACT-G instead of the patient's prediction when predicting the mean of the EQ-5D-3L utility index [37]. In addition, uncertainties or errors in the economic assessment may affect the accuracy of the utility value, leading to an incorrect estimation of the patient's HRQoL, and further research is needed to assess their impact on the mapping algorithm [2].

Our current study demonstrated that the predictive performance of the FACT-G was effective in the OLS model in the Chinese cancer population. Although the OLS model is a common mapping algorithm and the predicted values are close to the true values, it requires very strict assumptions, namely those of normal distribution and homogeneity of variance. In addition, previous studies have shown that OLS produces a low predictive ability, which will affect its prediction performance. This is similar to previous literature studies that overestimated those with poor health and underestimated those with good health utility values, as shown in Table 6 [13; 38]. Therefore, as reported by Zhang et al., the mapping algorithm of cost-utility analysis predicts the average utility index of the general population, rather than at the individual level [18]. When the data distribution is heavily skewed, it is important to consider the proportion of cancer patients who are in poor health.

Regression-based models generally tends to over-prediction for respondents in poor health and under-prediction for respondents in better health due to regression to the mean. Linear equating can help to alleviate the common problem of under-prediction of high scores and over-prediction of low scores, which forces predicted values to have the same mean and

standard deviation as observed values [34]. The smallest, 10th, and 50th percentile predicted values of the EQ-5D-3L utility values were significantly reduced with linear equating, as shown in Table 6. Given that the majority of observations were in perfect health, it's not surprising that less than half of the EQ-5D-3L value sets had overestimation of scores. Therefore, applying linear equating to predict values reduces biased estimates, resulting in similar variability between estimated and observed values for linear equating models [22]. The linear equivalence model aligns the distribution of the two values' scores on similar scales, corresponding to condition- or disease-specific scores. As a result, the estimated EQ-5D-3L scores should only be used for group-level analysis and should not be applied at the individual level [20; 22; 34]. While linear equating cannot directly estimate EQ-5D-3L item scores, it can be used to select the best model from various regression-based models to estimate predicted values before applying the linear equating method. Before using the linear fitting method, more research is needed to determine the best modeling strategy. In this cross-study, our sample consists of different types of cancer from one center and multiple oncology hospitals, so the sample size is relatively small, which may limit generalizability. The overall EQ-5D-3L utility index for Chinese cancer-patient was estimated at 0.935, which was higher than other utility indexes from the Chinese population. Perhaps most of the cancer patients we examined were in the early stages of their cancer, the quality of life of the patients was not seriously affected. Furthermore, due to a lack of data on disease stage and presentation status classification from all cancer patients, we were unable to assess the potential association between these patient parameters and quality of life scores. To some extent, our model can reflect the purpose and outcome of our research. Further validation of the mapping function should consider larger sample sizes and multiple treatment centers in future studies, which would help to more clearly explain the performance of the mapping function generated and the generalizability of the comparison results. Our finding also suggested that family financial pressure and significant life events might influence the HRQoL for cancer patients. Higher financial pressure and had significant life events were associated with a lower utility index, which may lead to a more severe disease burden that affects the HRQoL of the patient. And what we were finding would determine the underlying relevant factors that impacted the HRQoL or utility values of patients.

There are several limitations to this study. First, the study suffers from a high ceiling effect in the health utility index, which was 62.7% for Chinese cancer patients, leading to a high mean EQ-5D-3L values of 0.9353. The minimum EQ-5D-3L score is 0.364, which is well above the theoretical minimum of -0.149 for the Chinese value set. Although the range of utility values in this study is relatively narrow, Cheung et al. suggested that the range of utility values contributes to the relative performance of the different methods [30]. Concurrently, they developed the Mean Rank Method, a relatively new method that contributed to less over-estimation (under-estimation) of health utility among people in poor (good) health states [30]. Previous research indicated that OLS mapping of utility in a narrow range may be more accurate than in a broad range [29]. The current study has some limitations when compared to previous studies because its utility values are not fully covered. Meanwhile, it may be due to the smaller proportion of people in poorer health and the lack of the negative EQ-5D-3L utility value, which limits the generalizability of outcomes in more severe patients. That is, the study did not take full advantage of the potential range of scores by the Liu et al. algorithm. Recently published studies have suggested that an increasing number of countries are using the EQ-5D-5L tool as a preference-based measure instead of EQ-5D-3L due to its ability to reduce ceiling effects and sensitivity [39]. In addition, the Chinese value set of EQ-5D-5L was published in 2017 [40]. Second, the EQ-5D-3L has a high ceiling effect that may seriously affect the results, limiting the generalizability of the results to more severe patients. Finally, the study collected a relatively small sample and the data consisted of Chinese patients with five different types of cancer. Therefore, future studies need a larger sample size and external data to verify the generalizability of this study.

Conclusion

We developed a FACT-G to EQ-5D-3L mapping algorithm for the economic assessment analysis of cancer patients in China. The algorithm found that the OLS model have a good predictive ability compared to the observed and predictive EQ-5D-3L utility scores among all the regression models. The linear equating model can be used to more accurately estimate EQ-5D-3L utility scores in cost-utility studies. More research is needed to determine the reliability of the estimated value sets. This mapping algorithm may provide policymakers and researchers with references for the economic evaluation of specific health conditions in cost-utility analysis when estimating the health utilities of the Chinese cancer-patients with the Chinese values set.

Abbreviations

FACT-G: Functional Assessment of Cancer Therapy-General; PWB: Physical well-being; FWB: Functional well-being; EWB: Social well-being; HRQoL: health-related quality of life; QoL: Quality of life; EQ-5D-3L: Three-level EuroQol-5-dimension questionnaire; SD: standard deviation; Min: minimum; Max: maximum; PWB: Physical well-being; SFWB: Social/family well-being; EWB: Emotional well-being; FWB: Functional well-being; OLS: Ordinary least squares; GLM: Generalized linear models; CLAD: Censored least absolute deviations; TPM: Two-part model; MAE: Mean absolute error; RMSE: Root mean squared error; Adj r^2 : Adjust R-squared; AIC, Akaike information criterion; BIC, Bayesian information criterion; MSE: Mean squared error; MAPE: Mean absolute percentage error.

Declarations

Ethics approval and consent to participate

The study protocol was approved by the Institutional Review Board of the Cancer Hospital of the Chinese Academy of Medical Sciences (Approval No. 15-071/998). All participants gave their [written] informed consent.

Consent for publication

Not applicable. All results are reported as aggregated data.

Availability of supporting data

The datasets used and/or analyzed during the current study are available.

Competing interests

The authors declare that they have no competing interests.

Authors' Contributions

ZhiKui He and WenJie Liang contributed equally to this study.

Funding

This study was supported by the National Key Public Health Program of China (Cancer Screening Program in Urban China).

Authors' contributions

Li Yang is in charge of the project. Wenjie Liang, Wenjia Xu, and Wenxiu Huang conducted the data collection. Zhikui He drafted the manuscript. Li Yang, Xiaomin Wang, and Kaiyong Huang provided critical comments in revising the manuscript. All authors approved the final version of the manuscript for submission.

Acknowledgements

We sincerely thank all the members of the Cancer Screening Program in Urban China from the National Cancer Center of China, provinces and external expert panels. We are also grateful to the participants for this study.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2018;68(6):394–424.
2. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value Health.* 2013;16(1):202–10.

3. Wang L, Shi JF, Zhu J, Huang HY, Bai YN, Liu GX, Liao XZ, Mao AY, Ren JS, Sun XJ, Guo LW, Fang Y, Zhou Q, Ma HM, Xing XJ, Zhu L, Song BB, Du LB, Mai L, Liu YQ, Ren Y, Lan L, Zhou JY, Qi X, Sun XH, Lou PA, Wu SL, Li N, Zhang K, He J, Dai M. Health-related quality of life and utility scores of patients with breast neoplasms in China: A multicenter cross-sectional survey. *Breast*. 2018;39:53–62. & Health Economic Evaluation Working Group of the Cancer Screening Program in Urban, C.
4. Acaster S, Pinder B, Mukuria C, Copans A. Mapping the EQ-5D index from the cystic fibrosis questionnaire-revised using multiple modelling approaches. *Health Qual Life Outcomes*. 2015;13:33.
5. Shi JF, Kang DJ, Qi SZ, Wu HY, Liu YC, Sun LJ, Li L, Yang Y, Li Q, Feng XX, Zhang LQ, Li J, Li XL, Yang Y, Niyazi M, Xu AD, Liu JH, Xiao Q, Li LK, Wang XZ, Qiao YL. Impact of genital warts on health related quality of life in men and women in mainland China: a multicenter hospital-based cross-sectional study. *BMC Public Health*. 2012;12:153.
6. Dakin H, Gray A, Murray D. Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. *Qual Life Res*. 2013;22(3):683–94.
7. Murasawa H, Sugiyama T, Matsuoka Y, Okabe T, Hino A, Tanaka N, Sugimoto M, Oyama M, Fujimoto K, Horie S, Noto S, Shimozuma K. Health utility and health-related quality of life of Japanese prostate cancer patients according to progression status measured using EQ-5D-5L and FACT-P. *Qual Life Res*. 2019;28(9):2383–91.
8. Wu EQ, Mulani P, Farrell MH, Sleep D. Mapping FACT-P and EORTC QLQ-C30 to patient health status measured by EQ-5D in metastatic hormone-refractory prostate cancer patients. *Value Health*. 2007;10(5):408–14.
9. Liu GG, Wu H, Li M, Gao C, Luo N. Chinese time trade-off values for EQ-5D health states. *Value Health*. 2014;17(5):597–604.
10. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, Silberman M, Yellen SB, Winicour P, Brannon J, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol*. 1993;11(3):570–9.
11. Fang P, Tan KS, Troxel AB, Rengan R, Freedman G, Lin LL. High body mass index is associated with worse quality of life in breast cancer patients receiving radiotherapy. *Breast Cancer Res Treat*. 2013;141(1):125–33.
12. Meregaglia M, Borsoi L, Cairns J, Tarricone R. Mapping health-related quality of life scores from FACT-G, FAACT, and FACIT-F onto preference-based EQ-5D-5L utilities in non-small cell lung cancer cachexia. *Eur J Health Econ*. 2019;20(2):181–93.
13. Diels J, Hamberg P, Ford D, Price PW, Spencer M, Dass RN. Mapping FACT-P to EQ-5D in a large cross-sectional study of metastatic castration-resistant prostate cancer patients. *Qual Life Res*. 2015;24(3):591–8.
14. Askew RL, Swartz RJ, Xing Y, Cantor SB, Ross MI, Gershenwald JE, Palmer JL, Lee JE, Cormier JN. Mapping FACT-melanoma quality-of-life scores to EQ-5D health utility weights. *Value Health*. 2011;14(6):900–6.
15. Gray LA, Wailoo AJ, Hernandez Alava M. Mapping the FACT-B Instrument to EQ-5D-3L in Patients with Breast Cancer Using Adjusted Limited Dependent Variable Mixture Models versus Response Mapping. *Value Health*. 2018;21(12):1399–405.
16. Teckle P, McTaggart-Cowan H, Van der Hoek K, Chia S, Melosky B, Gelmon K, Peacock S. Mapping the FACT-G cancer-specific quality of life instrument to the EQ-5D and SF-6D. *Health Qual Life Outcomes*. 2013;11:203.
17. Teckle P, Peacock S, McTaggart-Cowan H, van der Hoek K, Chia S, Melosky B, Gelmon K. The ability of cancer-specific and generic preference-based instruments to discriminate across clinical and self-reported measures of cancer severities. *Health Qual Life Outcomes*. 2011;9:106.
18. Cheung YB, Thumboo J, Gao F, Ng GY, Pang G, Koo WH, Sethi VK, Wee J, Goh C. Mapping the English and Chinese versions of the Functional Assessment of Cancer Therapy-General to the EQ-5D utility index. *Value Health*. 2009;12(2):371–6.
19. Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International comparisons in valuing EQ-5D health states: a review and analysis. *Value Health*. 2009;12(8):1194–200.
20. Fayers PM, Hays RD. Should linking replace regression when mapping from profile-based measures to preference-based measures? *Value Health*. 2014;17(2):261–5.

21. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ.* 2010;11(2):215–25.
22. Thompson NR, Lapin BR, Katzan IL. Mapping PROMIS Global Health Items to EuroQol (EQ-5D) Utility Scores Using Linear and Equipercentile Equating. *Pharmacoeconomics.* 2017;35(11):1167–76.
23. Lamu AN. Does linear equating improve prediction in mapping? Crosswalking MacNew onto EQ-5D-5L value sets. *Eur J Health Econ.* 2020;21(6):903–15.
24. Chen H, Li N, Ren J, Feng X, Lyu Z, Wei L, Li X, Guo L, Zheng Z, Zou S, Zhang Y, Li J, Zhang K, Chen W, Dai M, He J, & group of Cancer Screening Program in Urban, C. (2019). Participation and yield of a population-based colorectal cancer screening programme in China. *Gut,* 68(8), 1450–1457.
25. Shi J-F, Huang H-Y, Guo L-W, Shi D, Gu X-Y, Liang H, Wang L, Ren J-S, Bai Y-N, Mao AY, Liu G-X, Liao X-Z, Zhang K, He J, Dai M. (2016). Quality-of-life and health utility scores for common cancers in China: a multicentre cross-sectional survey. *The Lancet,* 388.
26. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med.* 2001;33(5):337–43.
27. Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, Hamashima C, Hisashige A, Tamura M. Estimating an EQ-5D population value set: the case of Japan. *Health Econ.* 2002;11(4):341–53.
28. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes.* 2007;5:70.
29. Versteegh MM, Rowen D, Brazier JE, Stolk EA. Mapping onto Eq-5 D for patients in poor health. *Health Qual Life Outcomes.* 2010;8:141.
30. Cheung YB, Tan HX, Luo N, Wee HL, Koh GCH. Mapping the Shah-modified Barthel Index to the Health Utility Index Mark III by the Mean Rank Method. *Qual Life Res.* 2019;28(12):3177–85.
31. Cheung YB, Luo N, Ng R, Lee CF. Mapping the functional assessment of cancer therapy-breast (FACT-B) to the 5-level EuroQoL Group's 5-dimension questionnaire (EQ-5D-5L) utility index in a multi-ethnic Asian population. *Health Qual Life Outcomes.* 2014;12:180.
32. Huang IC, Frangakis C, Atkinson MJ, Willke RJ, Leite WL, Vogel WB, Wu AW. Addressing ceiling effects in health status measures: a comparison of techniques applied to measures for people with HIV disease. *Health Serv Res.* 2008;43(1 Pt 1):327–39.
33. Yang Q, Yu XX, Zhang W, Li H. Mapping function from FACT-B to EQ-5D-5 L using multiple modelling approaches: data from breast cancer patients in China. *Health Qual Life Outcomes.* 2019;17(1):153.
34. Hays RD, Revicki DA, Feeny D, Fayers P, Spritzer KL, Cella D. Using Linear Equating to Map PROMIS(®) Global Health Items and the PROMIS-29 V2.0 Profile Measure to the Health Utilities Index Mark 3. *Pharmacoeconomics.* 2016;34(10):1015–22.
35. Wang Y, Shi J, Du L, Huang H, Wang L, Zhu J, Li H, Bai Y, Liao X, Mao A, Liu G, Ren J, Sun X, Gong J, Zhou Q, Mai L, Zhu L, Xing X, Liu Y, Ren Y, Song B, Lan L, Zhou J, Lou P, Sun X, Qi X, Wu S, Wei W, Zhang K, Dai M, Chen W, He J. Health-related quality of life in patients with esophageal cancer or precancerous lesions assessed by EQ-5D: A multicenter cross-sectional study. *Thorac Cancer.* 2020;11(4):1076–89.
36. Sullivan PW, Ghushchyan V. Mapping the EQ-5D index from the SF-12: US general population preferences in a nationally representative sample. *Med Decis Making.* 2006;26(4):401–9.
37. Ara R, Brazier J. Deriving an algorithm to convert the eight mean SF-36 dimension scores into a mean EQ-5D preference-based score from published studies (where patient level data are not available). *Value Health.* 2008;11(7):1131–43.
38. Pickard AS, Ray S, Ganguli A, Cella D. Comparison of FACT- and EQ-5D-based utility scores in cancer. *Value Health.* 2012;15(2):305–11.
39. Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health.* 2014;17(4):445–53.
40. Luo N, Liu G, Li M, Guan H, Jin X, Rand-Hendriksen K. Estimating an EQ-5D-5L Value Set for China. *Value Health.* 2017;20(4):662–9.

Figures

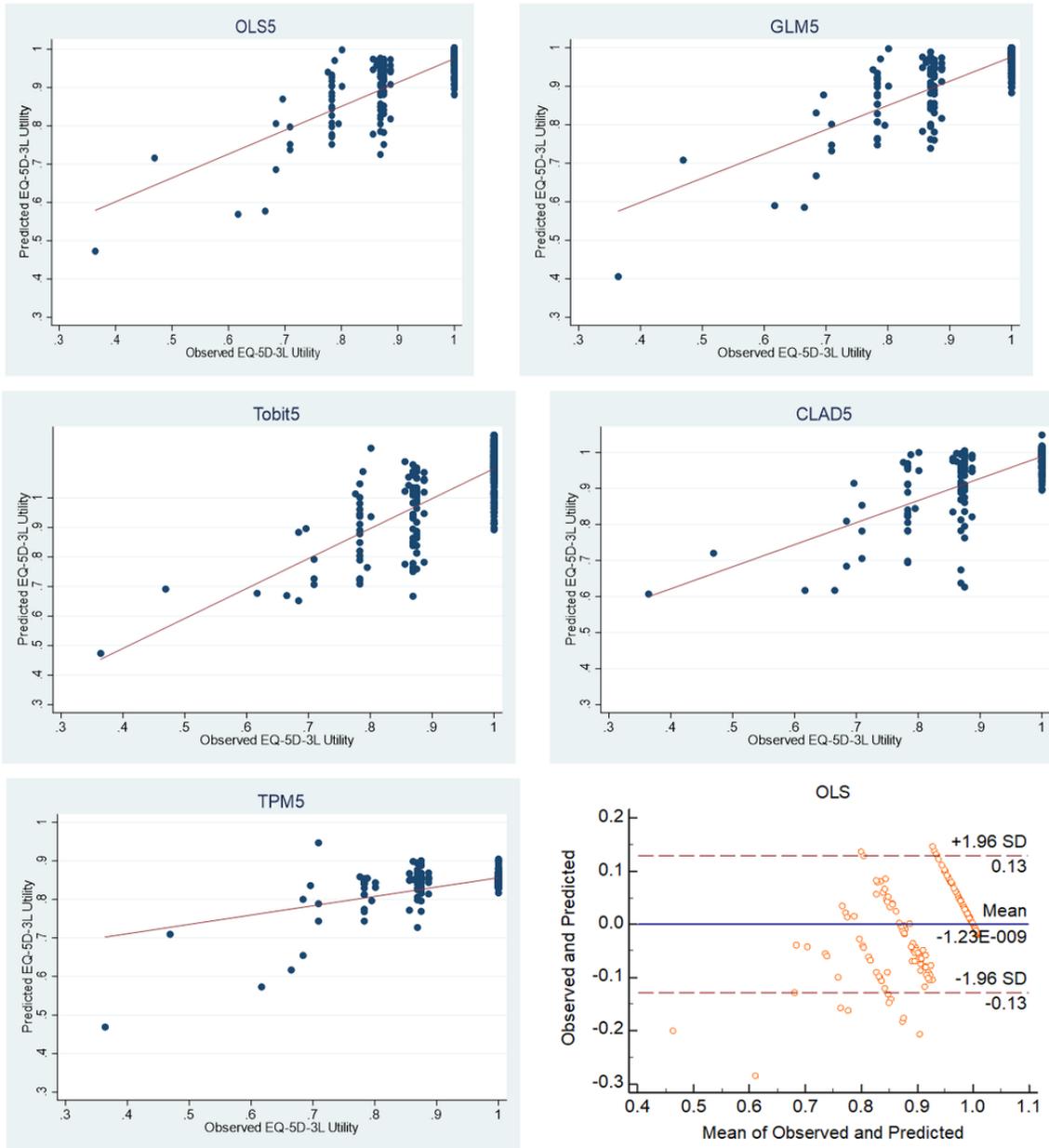


Figure 1

Scatter plots and Bland-Altman plot of observed and predicted EQ-5D-3L utility. OLS Ordinary least squares, GLM Generalized linear models, CLAD Censored least absolute deviations, TPM two-part model. Broken line is the best fitting regression line between observed value and predicted value.

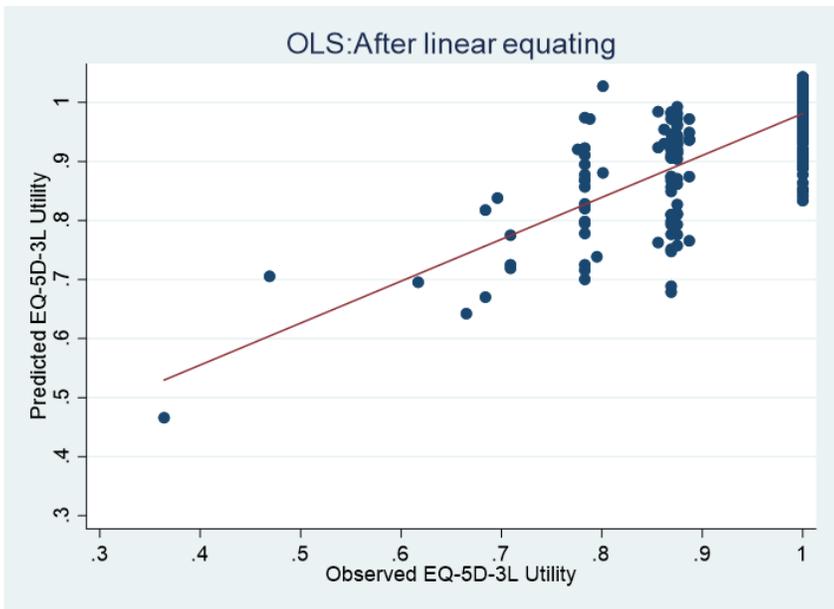
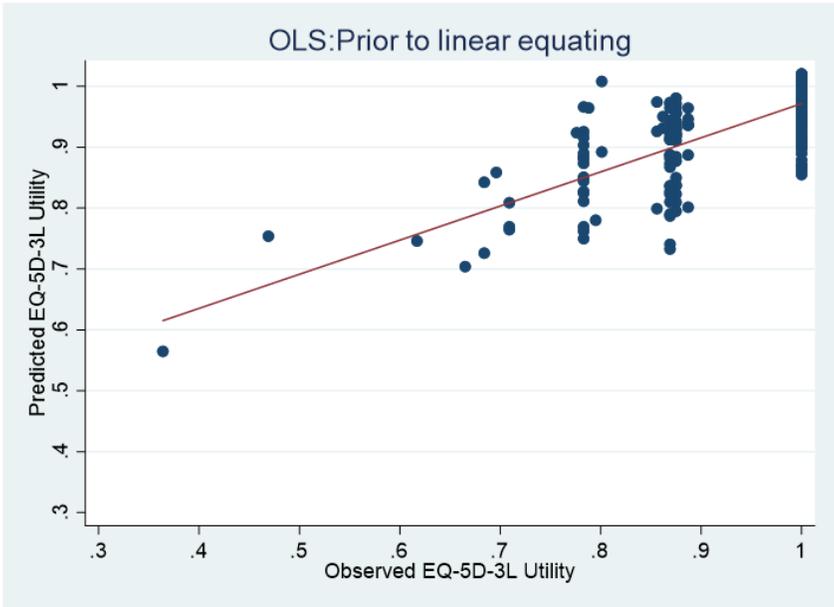


Figure 2

Scatter plots of observed and predicted EQ-5D-3L value sets for the preferred model. Broken line is the best fitting regression line between observed value and predicted value