

Improving the mapping of condition-specific health-related quality of life onto SF-6D score

Yingsi Yang, Ph.D.¹, M. Y. Wong, Ph.D.¹,
Cindy L. K. Lam, M.D.², Carlos K. H. Wong, Ph.D.*²

¹ *Department of Mathematics, The Hong Kong University of Science and Technology,
Hong Kong, China*

² *Department of Family Medicine and Primary Care, The University of Hong Kong,
Hong Kong, China*

Abstract

Background This study sought to improve the predicative performance and goodness-of-fit of mapping models, as part of indirect valuation, by introducing cubic spline smoothing to map a group of health-related quality of life (HRQOL) measures onto a preference-based measure.

Methods This study was a secondary analysis of a cross-sectional health survey data assessing the HRQOL for patients with colorectal neoplasms. Mapping functions of condition-specific Functional Assessment of Cancer Therapy – Colorectal (FACT-C) onto preference-based SF-6D measure were developed by using a data set of 553 Chinese subjects with different stages of

*Address correspondence to: Carlos K. H. Wong, Department of Family Medicine and Primary Care, The University of Hong Kong, 3/F, Ap Lei Chau Clinic, 161 Ap Lei Chau Main Street, Ap Lei Chau, Hong Kong. Tel: (+852) 2518-5688. Fax: (+852) 2814-7475. Email: carlos@hku.hk. Financial support for this study was provided in part by Small Project Funding (Project code 200907176135) from CRCG of the University of Hong Kong and Health and Health Service Research Fund (HHSRF #08090851) of Food and Health Bureau, HKSAR. The funding agreement ensured the authors independence in designing the study, interpreting the data, writing, and publishing the report.

colorectal neoplasm. The missing values of FACT-C were imputed by using multiple imputation. Then three widely applicable models – ordinary least square (OLS), Tobit and two-part models, were employed for the mapping function after applying the cubic spline smoothing on the data. For the evaluation of the effectiveness of cubic spline smoothing and multiple imputation, the goodness-of-fit and prediction performance of each model were compared.

Results Analyses showed that the models fitted with transformed data from cubic spline smoothing offered better performance in goodness-of-fit and prediction than the models fitted with the original data. The values of R^2 were improved by over 10%, and the root mean square error and the mean absolute error were both reduced. The best goodness-of-fit and performance were achieved by OLS model using transformed data from cubic spline smoothing.

Conclusions Cubic spline smoothing and multiple imputation were recommended for the mapping of HRQOL measures onto the preference-based measure. Among the three mapping models, the simple-to-use OLS model had the best performance.

Keywords: Quality of Life; Colorectal Cancer; FACT-C; SF-6D; Mapping Models; Cubic Spline

1 Introduction

Economic evaluation including cost-utility analysis is frequently used to compare health interventions. The vital implications of economic evaluation are to inform resource allocation of heavily-demanded health service system and advise the optimal medical decision to health policy makers. One of the major outcome measures from cost-utility analysis is the incremental cost-effectiveness ratios (ICER), representing the difference in costs between two interventions divided by the difference in their respective effectiveness in quality-adjusted life-years (QALYs). As highly recommended by National Institute for Health and Care Excellence in United Kingdom [1], the QALYs are calculated by summing up the product of utility score of each health state and its length of stay. The preference-based scores in the form of EQ-5D, HUI or SF-6D are the most common multi-attribute classification systems estimated through the application of regression-based mapping [2, 3]. Despite the potential loss of information, regression-based mapping is commonly used as a part of indirect valuation method to transform the scoring metric from the source health-related quality of life (HRQOL) measures to the target preference-based score, which theoretically ranges in the zero–one (dead–full health) scale. Unlike direct valuation methods such as standard gamble and time-trade off, the derivation of preference-based scores using regression-based mapping is particularly convenient and simple to apply [2]. However, to an extent, the mapping is still being challenged for methodological considerations regarding low level of predictive ability and high degrees of residual errors.

In an attempt to estimate the utility scores for the appraisal of interventions related to colorectal cancer, the functional assessment of cancer therapy-colorectal (FACT-C) measure is the commonly used condition-specific HRQOL measure. Owing to relevance of disease condition, the FACT-C measure had higher sensitivity and responsiveness to change over time in all HRQOL aspects except for the social well-being compared to generic measure [4]. However, the FACT-C, in current form of non-preference-based condition-specific measure, was not feasible for the use in

economic evaluation.

This study concerns the mapping from the condition-specific FACT-C measure onto the preference-based SF-6D. In previous work [5], such a mapping was established using the linear regression model. The aim of this study is to improve the mapping using the cubic spline smoothing technique, and to evaluate our improvements of prediction performance and goodness-of-fit to commonly used models, including ordinary least square, Tobit and two-part models. The data we used in this study is the colorectal neoplasm (CRN) health survey data collected from a Chinese population with CRN. Demographic and clinical characteristics were also incorporated in the models to improve their prediction performance.

2 Methods

2.1 Subject

The study was a secondary analysis of health survey data on a sample of 553 patients with CRN (colorectal polyps or cancer) at a colorectal specialist outpatient clinic in Hong Kong between October 2009 and July 2010. A survey was utilized to collect the HRQOL, self-reported socio-demographic characteristics and medical record information, including cancer stages (based on the staging system of American Joint Committee on Cancer [6]), primary tumor site, presence of stoma and current treatment. A summary of the demographic and clinical characteristics of the patients are shown in Table 1. Among all the patients in the study, the average age is 63.2 years with standard deviation of 11.3. The percentage of females is 42.0%. The proportions range from 13.0% to 20.6% for different stages of colorectal neoplasm. A condition-specific FACT-C and generic preference-based SF-6D instrument were administered by trained interviewers via face-to-face or telephone interviews. Details of subject recruitment and procedures were listed in previous studies [4, 5, 7, 8, 9, 10, 11, 12].

Table 1: Demographic and clinical characteristics of study subjects.

	Total (n=553)
Age (Year, mean \pm SD)	63.2 \pm 11.3
Sex, n(%)	
Male	321 (58.0)
Female	232 (42.0)
Stage of colorectal neoplasm, n(%)	
Low-risk polyp	93 (16.8)
High-risk polyp	72 (13.0)
Stage I	83 (15.0)
Stage II	101 (18.3)
Stage III	114 (20.6)
Stage IV	82 (14.8)
Unknown	8 (1.4)
Duration of diagnosis (Month, mean \pm SD)	46.6 \pm 55.8
Treatment status, n(%)*	
Palliative	63 (16.4)
Adjuvant	26 (6.8)
No	296 (76.9)
Stoma, n(%)*	
Yes	51 (13.2)
No	334 (86.8)

* Colorectal cancer patients only (n=385).

2.2 Instruments

The FACT-C, developed by Center on Outcomes, Research and Education [13], is a 36-item condition-specific HRQOL instrument that measures five major subscales: 7-item physical well-being (PWB), 7-item social well-being (SWB), 6-item emotional well-being (EWB), 7-item functional well-being (FWB), and 9-item colorectal cancer subscale (CCS). All items are scored using a five-point Likert scale (0=“Not at all”, 1=“A little”, 3=“Somewhat”, 4=“Quite a bit” and 5=“Very much”). Higher scores in subscales indicate better HRQOL. Psychometric properties, regarding validity and reliability, of the Traditional Chinese version 4 of the FACT-C has been reported previously [4, 8, 11].

The SF-6D is one of the widely used generic preference-based measures with

a multi-attribute classification system for calculating the utility score of dead–full-health scale anchored on 0–1. The SF-6D classification system consists of six dimensions, physical functioning, role functioning, social functioning, pain, mental health and vitality, with three to five response levels each. Response combination of six dimensions was then converted into a SF-6D score following the preference weights derived from the Hong Kong general population [14, 15]. In Hong Kong preference scoring algorithm, the utility scores range from 0.315 (lowest possible state) to 1.000 (Perfect health state) with higher scores indicating better HRQOL.

2.3 Data Analysis

Our data analysis procedure contain three steps. First the missing data is imputed from the original FACT-C questionnaires. The complete FACT-C data is then transformed using the cubic spline function. Finally the mapping of FACT-C onto SF-6D is built based on the ordinary least square (OLS) method. We provide detailed explanations of each step as follows.

Data Manipulation In the original questionnaires of the FACT-C, the sample size of the complete dataset was only 226 out of 553 in the overall dataset. i.e., more than 50% observations in the entire dataset had missing values. We first adopted the scoring guidelines by Functional Assessment of Chronic Illness Therapy (FACIT) [16] to handle the missing values, i.e., prorating the subscale scores in the case with more than 50% of the items answered (e.g., a minimum of 4 out of 7 items, 4 out of 6 items, etc). Among the adjusted subscales, there are still 9 missing out of 553 observations. These 9 missings were simply deleted in the previous study [5] since the percentage of the missing is small. In this study, the scores of the subscales were further weighted by the number of the questions in each subscale. Besides following the FACIT scoring guidelines, we also imputed the missing answers in the original questionnaire by the overall distribution of that particular answer. This is based on the fact that we observed no pattern of the missing data, therefore assumed that the

data is missing at random. An imputed dataset with ten imputations was created using multivariate imputation by chained equations [17, 18].

Cubic Spline Transformation The cubic spline was first developed in the field of numerical analysis as an interpolation technique. It is a sufficiently smooth cubic polynomial function that is piecewise-defined. The function curve is not only smooth within the segments divided by the user-defined knots, but also sufficiently smooth at the places where the polynomial pieces connect. The cubic smoothing technique is also applied in regression modeling to capture the nonlinear effects and improve the explanatory power of a linear model with the spline transformed data [19, 20]. In our case, we employed the cubic spline transformation on the scores of the five subscale in FACT-C. A cubic B-spline is a piecewise cubic curve split at a series of distinct knots, with the property that the derivatives of the curve are continuous at the knots up to the degree of two. With the transformed data, we still use linear models for the mapping function.

Model Specification For the mapping of FACT-C onto SF-6D, we employed the OLS method since it is the most widely applied and straightforward. According to previous work [5], ceiling effect is observed for SF-6D since the questionnaire does not provide enough distinguishability for the patients exhibiting a high quality of life. Therefore, in addition to OLS, we also considered the Tobit model [21] and two-part model [22] for comparison, both of which were designed to handle the ceiling effect in SF-6D. In the Tobit model, it assumes that there is a latent response Y_i^* satisfying $Y_i^* = X_i\beta + \epsilon_i^*$ with $\epsilon_i^* \sim N(0, \sigma^2)$, and the observed response Y_i satisfies $Y_i = Y_i^*$ if $Y_i^* \leq 1$, and $Y_i = 1$ for otherwise. The two-part model was defined as $\text{logit}(\Pr(Y_i = 1|X_i)) = X_i\alpha$ and $E(Y_i|X_i, Y_i < 1) = X_i\beta$ with the overall model $E(Y|X) = [X\beta + \exp(X\alpha)]/[1 + \exp(X\alpha)]$. All three models were fitted with the main effect, as well as the adjustment of the demographics and clinical variables given in Table 1. Both multiple imputation and spline transformation were applied to the data before fitting the model.

Model Comparison Some commonly used criteria were employed to assess the performance of each model. We used R^2 as a statistical measure of how well the regression line approximates the real data points. We also included adjusted R^2 , which takes account of the inflation of R^2 when more variables are added into the model. In addition, the Akaike information criterion (AIC) [23] and Bayesian information criterion (BIC) [24] provided a measure of the relative quality and goodness of fit of the models. The predictive ability of the models were compared by using the root mean square error (RMSE) and the mean absolute error (MAE), both of which are based on the differences between the predicted and observed SF-6D scores for each individual observation. In addition, the number of observations with the absolute error (AE) greater than the threshold of 0.05 and 0.1 and the corresponding proportions in the full sample were calculated respectively. Since the two-part model only gave the parameter estimation and model prediction for the part with SF-6D scores less than one, it is not appropriate to compare with the other models using the criteria based on only one part of the model. Therefore, only the AIC and BIC were available for the two-part model. All the values of these measures, i.e., R^2 , adjusted R^2 , AIC, BIC, RMSE, MAE, and AE, were presented as the average of results based on the ten imputations.

All regressions and other statistical analysis were conducted by using SAS (version 9.3, by SAS Institute Inc., Cary, NC, USA). The multiple imputation was performed by using the MICE package (version 2.18) in R (version 3.0.1) [25].

3 Results

3.1 Cubic Spline Smoothing

In order to smooth the mapping function and search for a better fitted curve with higher explanatory power, we considered the spline transformation on the main effects, i.e., the values of the subscales PWB, SWB, EWB, FWB and CCS. The spline transformation we chose is the cubic B-spline with four knots. Our cubic B-spline provides sufficient degrees of freedom to better approximate the pattern displayed by the data.

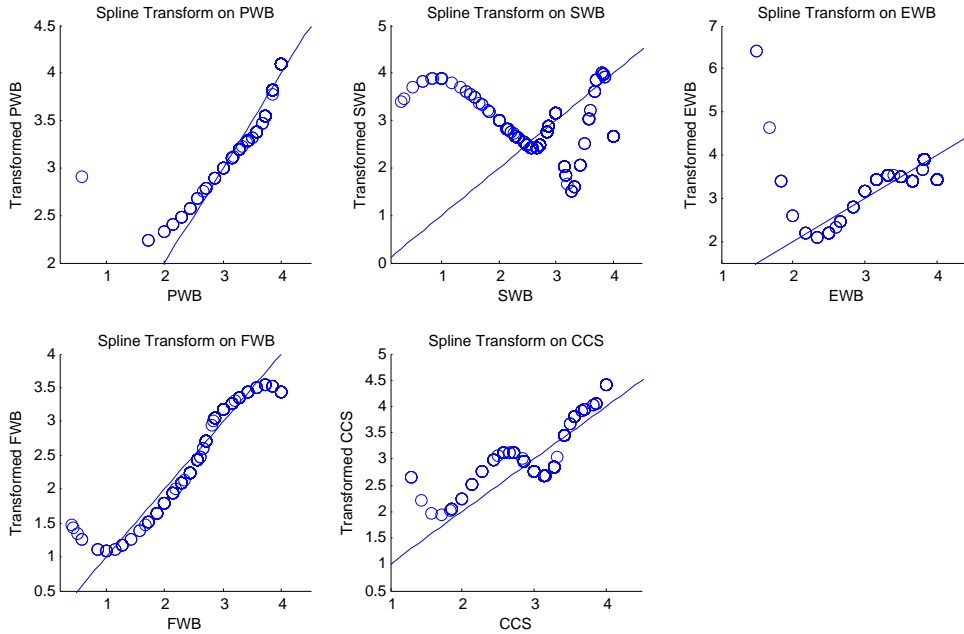


Figure 1: Cubic spline transformation on the five subscales of FACT-C.

The spline transformations on each of the five subscales of FACT-C are all converged, and the resulting transformation functions are shown in Figure 1. From the graphs of the transformation function, it is noticed that none of the spline transformations is an identical transformation. It shows that the cubic spline transformation is necessary for improving the model fitting.

3.2 Parameter Estimation and Prediction of the Regression Models for the Mapping Function

The mapping function onto SF-6D was modeled with three methods, i.e., the ordinary least square method (OLS), the Tobit model, and two-part model as mentioned in Section 2.3. The models were fitted with the data imputed by multiple imputation. Both the original and spline-transformed data of the five main effects were used in the model fittings for comparison. Table 2 shows the results of the parameter estimations for the three regression models. Table 3 shows the summary of the SF-6D scores predicted by the fitted mapping functions. Note that the results of the predictions from the two-part model are not included in Table 3. This is because the two-part model only provided the predicted SF-6D scores given by the linear part, in which the observations with SF-6D score equal to one were excluded.

As shown in the lower part of Table 2, for the models using the spline transformed data, the five main-effect terms from the FACT-C subscales scores were all significant in the models. The positive values of the coefficients for the five main-effect indicated that an increase of the score for any of the subscales would lead to increase in the score of SF-6D and vice versa. Among all the demographic and clinical characteristics covariates, the effects of “Female” in “Sex” and the subgroup “Low Risk Polyps” in “Stage of colorectal neoplasm” were also significant. This implied that the female patients had significant lower health state than male, and all the groups with different stage of colorectal neoplasm had similar scores of SF-6D except the group of “Low Risk Polyps”. Meanwhile, only the PWB, EWB, FWB and CCS in the FACT-C subscales and the effect of “Sex” were significant in the models using the data without spline transformed as shown in the upper part of Table 2. All the interaction terms were not significant in the models with or without the transformed data.

In general, the scores of SF-6D predicted by the three models were similar in terms of the summary statistics such as mean, standard deviation, median, minimum and maximum values in Table 3. The minimum values of the predicted SF-6D

Improving the Mapping of FACT-C onto SF-6D

Table 2: Parameter estimations for OLS, Tobit and two-part models with no or cubic spline transformation.

	OLS		Tobit Model		Two-Part Model [†]	
	Coeff	95% CI	Coeff	95% CI	Coeff	95% CI
Multiple Imputation and No Spline Transformation						
FACT-C						
PWB	0.1046*	(0.0826, 0.1266)	0.1040*	(0.0817, 0.1264)	0.1039*	(0.0816, 0.1262)
SWB	0.0075	(-0.0054, 0.0204)	0.0090	(-0.0042, 0.0222)	0.0071	(-0.0061, 0.0203)
EWB	0.0223*	(0.0021, 0.0424)	0.0202	(-0.0003, 0.0407)	0.0212*	(0.0006, 0.0418)
FWB	0.0657*	(0.0499, 0.0815)	0.0692*	(0.0531, 0.0854)	0.0624*	(0.0464, 0.0785)
CCS	0.0371*	(0.0162, 0.0580)	0.0404*	(0.0190, 0.0617)	0.0359*	(0.0147, 0.0571)
Characteristics						
Sex [‡]						
Female	-0.0194*	(-0.0353, -0.0035)	-0.0217*	(-0.0379, -0.0055)	-0.0173*	(-0.0335, -0.0011)
Stage [§]						
Low Risk	0.0590	(-0.0400, 0.1578)	0.0463	(-0.0556, 0.1483)	0.0643	(-0.0056, 0.1029)
High Risk	0.0253	(-0.0736, 0.1242)	0.0106	(-0.0913, 0.1124)	0.0311	(-0.0086, 0.0708)
Stage I	0.0223	(-0.0185, 0.0631)	0.0220	(-0.0197, 0.0637)	0.0280	(-0.0143, 0.0704)
Stage II	0.0336	(-0.0067, 0.0739)	0.0321	(-0.0090, 0.0733)	0.0361	(-0.0057, 0.0780)
Stage III	0.0273	(-0.0120, 0.0666)	0.0243	(-0.0158, 0.0645)	0.0315	(-0.0096, 0.0725)
Constant	0.0535	(-0.0817, 0.1887)	0.0640	(-0.0773, 0.2052)	0.0560	(-0.0283, 0.1403)
Multiple Imputation and Cubic Spline Transformation						
FACT-C						
PWB	0.1179*	(0.0996, 0.1361)	0.1188*	(0.1003, 0.1374)	0.1189*	(0.1008, 0.1369)
SWB	0.0171*	(0.0060, 0.0282)	0.0180*	(0.0067, 0.0293)	0.0223*	(0.0114, 0.0332)
EWB	0.0226*	(0.0070, 0.0381)	0.0218*	(0.0059, 0.0376)	0.0240*	(0.0088, 0.0391)
FWB	0.0787*	(0.0651, 0.0924)	0.0802*	(0.0663, 0.0940)	0.0750*	(0.0615, 0.0884)
CCS	0.0383*	(0.0212, 0.0553)	0.0428*	(0.0253, 0.0603)	0.0404*	(0.0233, 0.0576)
Characteristics						
Sex [‡]						
Female	-0.0158*	(-0.0302, -0.0014)	-0.0172*	(-0.0319, -0.0025)	-0.0142*	(-0.0286, -0.0002)
Stage [§]						
Low Risk	0.0442*	(0.0088, 0.0797)	0.0132	(-0.0794, 0.1058)	0.0442*	(0.0095, 0.0790)
High Risk	-0.0030	(-0.0925, 0.0865)	-0.0134	(-0.1059, 0.0791)	0.0177	(-0.0177, 0.0531)
Stage I	0.0131	(-0.0245, 0.0507)	0.0130	(-0.0254, 0.0514)	0.0136	(-0.0236, 0.0509)
Stage II	0.0252	(-0.0120, 0.0624)	0.0239	(-0.0141, 0.0619)	0.0214	(-0.0153, 0.0581)
Stage III	0.0146	(-0.0218, 0.0509)	0.0122	(-0.0249, 0.0493)	0.0124	(-0.0233, 0.0482)
Constant	-0.1071*	(-0.1958, -0.0183)	-0.0584	(-0.1933, 0.0765)	-0.1097*	(-0.1973, -0.0222)

* Significant with $p < 0.05$.

† The estimation of the parameters is for the linear part of the two-part model.

‡ The reference group of the effect "Sex" is "Male".

§ The reference group of the effect "Stage of colorectal neoplasm" is "Stage IV".

Table 3: Comparisons of fitted SF-6D scores amongst regression models with no or cubic spline transformation.

	No Spline Transform		Cubic Spline Transform	
	OLS	Tobit	OLS	Tobit
Mean	0.8238	0.8259	0.8238	0.8258
SD	0.1047	0.1068	0.1104	0.1125
Median	0.8473	0.8496	0.8455	0.8476
Min	0.3115	0.3108	0.4635	0.4599
Max	1.0127	1.0000	1.0302	1.0000

Note: OLS=Ordinary least square; SD=Standard deviation

scores are slightly larger for the models with cubic spline transformation than the ones without transformed data. The maximum values predicted by the Tobit model are all equal to one since the Tobit model has an assumption on the upper bound according to its model definition.

3.3 Model Selection and Comparison

Spline Transformation The goodness-of-fit statistics and prediction performance of the three models, i.e., OLS, Tobit and two-part models, with and without the cubic spline transformation are compared in Table 4. From the table, it is shown that the values of R^2 and adjusted R^2 for OLS were increased by 11.16% and 11.78% respectively after the spline transformation. The values of AIC and BIC decreased by 5% to 10% with the spline transformed data. Similarly, the Tobit and two-part models also lead to reduced AIC and BIC. The changes in R^2 /adjusted- R^2 and AIC/BIC imply that the spline transformation led to a better fitted model with more power in explaining the data. The RMSE and MAE for all three models were also decreased when compared with the models without spline transformation. This suggested that the models fitted with the transformed data had a stronger predictive ability. From both aspects of goodness-of-fit and prediction performance, the cubic spline transformation contributed to the improvement of the models. The cubic

spline transformation enabled the model to capture the trend of the data in a more detailed way, yet only using a polynomial function without interaction terms. The model with such a function is also simple and easy to explain in terms of the implied causal relationship. In general, the cubic spline transformation improves the overall performance of all three models.

Table 4: Goodness-of-fit statistics and prediction performance of mapping models with no or cubic spline transformation.

	No Spline Transform			Cubic Spline Transform		
	OLS	Tobit	Two-Part	OLS	Tobit	Two-Part
Goodness-of-Fit Statistics						
R^2	0.5867	N/A	N/A	0.6522	N/A	N/A
Adjusted R^2	0.5739	N/A	N/A	0.6415	N/A	N/A
AIC	-2100	-956	-874	-2196	-1044	-956
BIC	-2586	-879	-736	-2676	-965	-818
Prediction Performance						
RMSE	0.0878	0.0879	N/A	0.0806	0.0806	N/A
MAE	0.0673	0.0669	N/A	0.0611	0.0610	N/A
AE > 0.05	283.3 (51.23%)	283.3 (51.23%)	N/A	257.4 (46.55%)	283.3 (51.23%)	N/A
AE > 0.10	131.1 (23.71%)	131.1 (23.71%)	N/A	110.8 (20.04%)	131.1 (23.71%)	N/A

Note: OLS=Ordinary least square; AIC=Akaike information criterion; BIC=Bayesian information criterion
 RMSE=Root mean square error; MAE=Mean absolute error; AE=Absolute error

OLS, Tobit and Two-Part Models From Table 4, it can be seen that the OLS performed better in the goodness-of-fit than the other two models both with and without the spline transformation. The tobit model had the MAE slightly smaller than that of OLS for the case without the spline transform. However, for both RMSE and the proportions of AE greater than 0.05 and 0.10, OLS model still had a better prediction performance than the Tobit and two-part models. Therefore, OLS model is preferred in general, not only because of its simplicity and intuitiveness, but also due to its advantages in goodness-of-fit and predictive ability. Note that in Table 3, the predicted SF-6D scores with OLS model may exceed one, the upper bound of the SF-6D score, and become difficult to explain. In the case when the

OLS model is used for prediction purposes, the predicted SF-6D scores greater than one should be clamped to one. Based on our data, this led to prediction performance that is slightly better in the proportions of AE greater than 0.05 and 0.10, but very similar in RMSE and MAE, when compared with the Tobit model. Hence, in the case when the individual prediction is the main target, the Tobit or two-part models could be an useful alternative, despite their possible trade-off in the goodness-of-fit and prediction performance.

Table 5: Goodness-of-fit statistics and prediction performance of the mapping models using cubic spline transformed data imputed by following the scoring guidelines by FACIT.

	OLS	Tobit	Two-Part
Goodness-of-Fit Statistics			
R^2	0.6497	N/A	N/A
Adjusted R^2	0.6390	N/A	N/A
AIC	-2156	-1031	-955
BIC	-2627	-954	-817
Prediction Performance			
RMSE	0.0816	0.0801	N/A
MAE	0.0605	0.0603	N/A
AE > 0.05	245 (45.20%)	250 (45.62%)	N/A
AE > 0.10	114 (21.03%)	113 (20.62%)	N/A

Note: OLS=Ordinary least square; AIC=Akaike information criterion;
 BIC=Bayesian information criterion; RMSE=Root mean square error;
 MAE=Mean absolute error; AE=Absolute error

Multiple Imputation Tables 5 and 6 present the results of the models using data with cubic spline transformation imputed by following the scoring guidelines by FACIT. Comparing to the results with multiple imputations in Tables 3 and 4, the prediction of the SF-6D scores obtained by the three models were similar in both cases. This implies that the multiple imputation was performed in a valid basis that was consistent with the distribution of the observed data. Furthermore, the OLS model with transformed data from multiple imputation performed slightly better both in goodness-of-fit and predictive ability than the OLS model with the original

Table 6: Comparisons of fitted SF-6D scores amongst the mapping models using cubic spline transformed data imputed by following the scoring guidelines by FACIT.

	OLS	Tobit
Mean	0.8243	0.8259
SD	0.1088	0.1120
Median	0.8488	0.8486
Min	0.4508	0.4449
Max	1.0254	1.0000

Note: OLS=Ordinary least square; SD=Standard deviation

data. Therefore, in order to achieve valid inference and good model performance, multiple imputation should be considered if the proportion of missing values is large in the data set. On the other hand, in the case of only a small percentage of missing values, following the scoring guidelines [16], i.e., averaging the scores in each subscale, can be a simple alternative to handle the missing values.

4 Discussions

In our study, we used three models and employed the cubic spline transformation and multiple imputation on missing values to improve the mapping function and perform the data analysis. Among all the results, the OLS model with cubic spline transformation and multiple imputation performed the best in general, with all R^2 and adjusted R^2 values beyond 60%. Note that in terms of explanatory power, this result outperforms the majority (86.7%) of mapping models from condition-specific measure to a generic preference-based measures reported by Brazier et al. [2]. Moreover, the mapping function also has some other advantages compared to the best model shown in the previous study [5] performed on the same set of data. In the current OLS model with imputed data and smoothing technique, the goodness-of-fit was improved by an increase in R^2 and adjusted R^2 , and a decrease in the RMSE and MAE; the terms of the covariates were further simplified with only the main effects, but no interaction term.

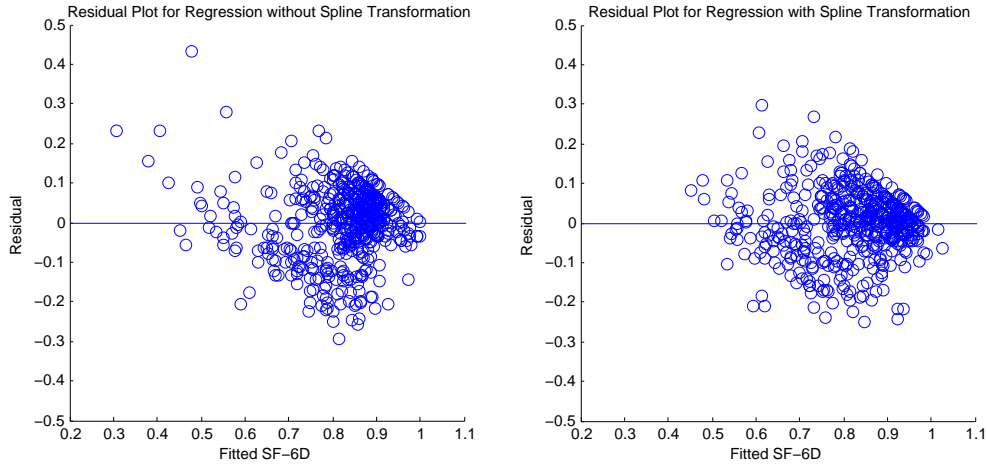


Figure 2: Residual plots for the models with no or cubic spline transformation.

The good performance in goodness-of-fit and predictive ability achieved by the proposed OLS model is due to the better consistency with the model assumptions, i.e., the residuals are independently normally distributed with mean zero and constant variance. The residual plot of the OLS model with the transformed data is shown on the rightside of Figure 2. The Tobit and two-part model handled the problem of the ceiling effect by treating the predicted SF-6D scores with a truncated model, i.e. clamping the response value by the upper bound one. However, truncated model is not suitable for modeling the SF-6D scores in our case, because by definition the SF-6D score follows a natural range of 0-1, rather than being truncated from a larger ranged value. In our study, the value of AIC is -2196 for OLS model, but -1044 and -956 for the Tobit and two-part models, respectively. It implies that the OLS model is better fitted than the Tobit and two-part models. The same conclusion was also drawn by the values of BIC.

Figure 3 shows the scatter plots of mean observed and predicted SF-6D preference-based scores obtained by OLS and Tobit models. The patterns of over-estimation were observed for both OLS and Tobit models when the observed SF-6D scores were lower than 0.8. A comparison of the mean error, RMSE and MAE of the OLS and Tobit models using the data with and without spline transformation is shown in Table 7. By comparing the mean error of the models in different ranges of the observed

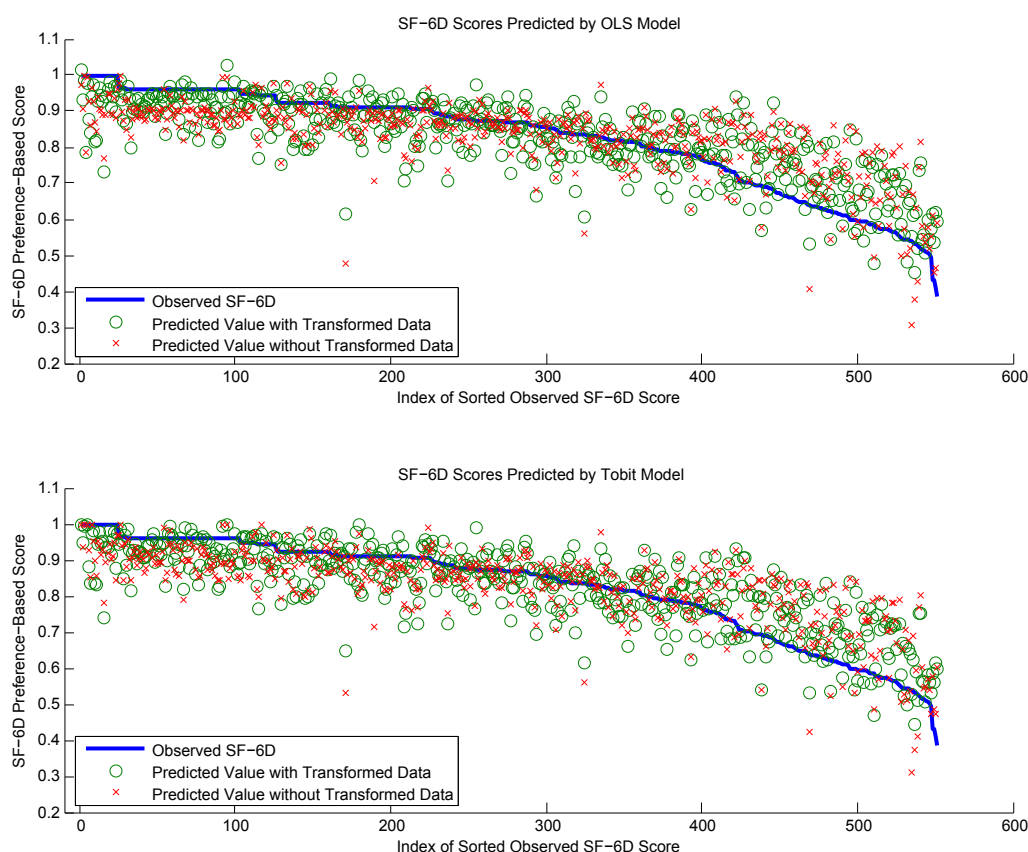


Figure 3: Scatter plot of the observed and predicted SF-6D scores in the descending order of the observed scores.

SF-6D scores, it is also noticed that a tendency of over-estimation was observed in each model when the SF-6D scores were lower than 0.8, as observed in Figure 3. This agrees with the result in a previous mapping paper where the over-estimation exists when the observed SF-6D scores were lower than 0.8 [5]. However, the models using the data with spline smoothing had smaller magnitude of over-estimation for observed value of SF-6D less than 0.8, and smaller RMSE and MAE both in general and in most of the ranges of the observed SF-6D scores. It implies that the cubic spline smoothing can help in improving the effect of over-estimation and reducing the RMSE and MAE.

With the spline smoothing method, models for the mapping function were greatly improved. In our study, we chose the cubic spline with four knots for the spline transformation. There are other options in selecting the degree and the number of pieces of the polynomial function obtained by the transformation. The cubic spline

Table 7: Mean error, RMSE, and MAE of the SF-6D scores predicted by OLS and Tobit models using the data with and without spline smoothing

	No Spline Transform		Cubic Spline Transform	
	OLS	Tobit	OLS	Tobit
Mean Error				
0.385-0.500 (n=5)	-0.1040	-0.0922	-0.1577	-0.1480
0.501-0.600 (n=51)	-0.0855	-0.0886	-0.0790	-0.0806
0.601-0.700 (n=63)	-0.1052	-0.0941	-0.0744	-0.0669
0.701-0.800 (n=68)	-0.0462	-0.0426	-0.0293	-0.0286
0.801-0.900 (n=139)	0.0114	0.0088	0.0150	0.0119
0.901-1.000 (n=226)	0.6217	0.0514	0.0423	0.0374
Whole Range (n=551)	0.0000	-0.0017	0.0000	-0.0017
RMSE				
0.385-0.500 (n=5)	0.1274	0.1232	0.1670	0.1610
0.501-0.600 (n=51)	0.1388	0.1405	0.1148	0.1183
0.601-0.700 (n=63)	0.1319	0.1246	0.1049	0.1001
0.701-0.800 (n=68)	0.0833	0.0840	0.0840	0.0857
0.801-0.900 (n=139)	0.0531	0.0542	0.0611	0.0601
0.901-1.000 (n=226)	0.1003	0.0742	0.0716	0.0683
Whole Range (n=551)	0.0878	0.0879	0.0806	0.0806
MAE				
0.385-0.500 (n=5)	0.1040	0.1002	0.1577	0.1480
0.501-0.600 (n=51)	0.1139	0.1166	0.0945	0.0986
0.601-0.700 (n=63)	0.1190	0.1105	0.0883	0.0830
0.701-0.800 (n=68)	0.0673	0.0647	0.0654	0.0668
0.801-0.900 (n=139)	0.0375	0.0394	0.0456	0.0453
0.901-1.000 (n=226)	0.0666	0.0580	0.0518	0.0505
Whole Range (n=551)	0.0673	0.0669	0.0611	0.0610

Note: OLS=Ordinary least square; RMSE=Root mean square error;
MAE=Mean absolute error

is the most common choice since it provides enough flexibility in describing the data while still not too complicated in the expression. Other settings may also be applied based on different cases.

We performed the analysis using the data with ten multiple imputations. We also experimented with different numbers of imputations, which all led to similar results. The performance of the models is neither sensitive to the number of multiple imputations nor the distributional assumption of the imputation [26]. As a common practice, the number of imputations is usually set between five and ten.

There are also other models and methods which are applicable for fitting the mapping function and improving the results obtained by linear model without any data manipulation. The censored least absolute deviations model (CLAD) is one of the alternatives which is commonly considered [27, 28]. It also models the data in a similar way as Tobit model to handle the ceiling effect with a truncated upper bound, but the optimization criterion is based on the absolute error instead of the squared error. Such an optimization criterion is applicable when the assumptions for applying least square error are not valid, i.e., the residuals are not independently normally distributed with mean zero and constant variance. However, in the residual plots shown in Figure 2, the residuals in our models were evenly spread around the horizontal line of zero without a certain pattern. It implies that the residuals were independently and normally distributed with mean zero and a constant variance. Therefore, we did not include CLAD model in our study, since the residuals satisfied the assumptions in an acceptable degree after the spline transformation as shown in Figure 2.

The limitation of our study is mainly on the generalizability of our mapping function. Our mapping function obtained using the data collected from the Chinese CRN population may not be applicable to other patient groups. For the patients with other kinds of cancer, the mapping function should be validated with the sample of the specific cancer before fitting on the data sets. Furthermore, the SF-6D scores we used were based on Hong Kong preferences weights [14]. The proposed

mapping function in this paper should be applied with caution on the data from populations other than Chinese. However, our methodology for improving the model performance can be applied similarly to other patient groups in general.

5 Conclusions

The mapping of FACT-C subscale scores onto SF-6D preference-based scores for Chinese patients were effectively improved by different models and data manipulation techniques. The cubic spline method can capture the nonlinear effects of the FACT-C subscales and increase the goodness-of-fit and prediction performance. The mostly recommended model is OLS model, not only in the sense of model performance but also for its simplicity and intuitive explanation. Other than OLS model which is commonly used in practice, Tobit model and two-part model can be used as alternative for more precise prediction on the upper bound of SF-6D scores but without too much loss in goodness-of-fit and predictive ability. The problem of missing value in original questionnaire of FACT-C can be handled by multiple imputation or by simple averaging following the scoring guidelines. The latter one is simple and easy to apply, but the former one is better grounded theoretically, and has advantages when assessing the model performance.

Further studies can be focused on improving the mapping of condition-specific health-related quality of life other than FACT-C onto SF-6D scores, and the questionnaires for non-Chinese populations. Other smoothing methods can also be investigated for the mapping function.

Acknowledgements

We thank Ms. Mansy Sham and Dr. Lei Yang for helping with proofreading and editing the manuscript.

Funding for this study was provided by Small Project Funding (Project code 200907176135) from CRCG of the University of Hong Kong, and Health and Health Service Research Fund (HHSRF # 08090851) of Food and Health Bureau, HKSAR.

References

- [1] National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal; 2013. Available from: <http://publications.nice.org.uk/pmg9>.
- [2] 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. *The European Journal of Health Economics*. 2010;11(2):215–225.
- [3] Mortimer D, Segal L. Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. *Medical Decision Making*. 2008;28(1):66–89.
- [4] Wong CK, Lam CL, Law WL, Poon JT, Kwong DL, Tsang J, et al. Condition-specific measure was more responsive than generic measure in colorectal cancer: all but social domains. *Journal of Clinical Epidemiology*. 2013;(5):557–565.
- [5] Wong CK, Lam CL, Rowen D, McGhee SM, Ma KP, Law WL, et al. Mapping the Functional Assessment of Cancer Therapy-General or -Colorectal to SF-6D in Chinese Patients with Colorectal Neoplasm. *Value in Health*. 2012;15(3).
- [6] Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals of surgical oncology*. 2010;17(6):1471–1474.
- [7] Wong CK, Lam CL, Poon JT, McGhee SM, Law WL, Kwong DL, et al. Direct medical costs of care for Chinese patients with colorectal neoplasia: A health

- care service provider perspective. *Journal of Evaluation in Clinical Practice*. 2012;18(6):1203–1210.
- [8] Wong CK, Lam CL, Law WL, Poon JT, Chan P, Kwong DL, et al. Validity and reliability study on traditional Chinese FACT-C in Chinese patients with colorectal neoplasm. *Journal of Evaluation in Clinical Practice*. 2012;18(6):1186–1195.
- [9] Wong CK, Lam CL, Poon JT, Kwong DL. Clinical Correlates of Health Preference and Generic Health-Related Quality of Life in Patients with Colorectal Neoplasms. *PloS one*. 2013;8(3):e58341.
- [10] Wong CKH, Lam CL, Wan Y, Rowen D. 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*. 2013;16(2):373–384.
- [11] Wong CK, Lam CL, Mulhern B, Law WL, Poon JT, Kwong DL, et al. Measurement invariance of the Functional Assessment of Cancer Therapy Colorectal quality-of-life instrument among modes of administration. *Quality of Life Research*. 2013;22(6):1415–1426.
- [12] Wong CK, Mulhern B, Wan YF, Lam CL. Responsiveness was Similar Between Direct and Mapped SF-6D in Colorectal Cancer Patients Who Declined. *Journal of Clinical Epidemiology*. 2014;67(2):219–227.
- [13] Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system: properties, applications, and interpretation. *Health and Quality of Life Outcomes*. 2003;1(1):79.
- [14] Lam CL, Brazier J, McGhee SM. Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. *Value in Health*. 2008;11(2):295–303.

- [15] McGhee SM, Brazier J, Lam CL, Chin WL, Chau J, Cheung A, et al. Quality-adjusted life years: population-specific measurement of the quality component. *Hong Kong Medical Journal*. 2011;17(6 Supplement 6):17.
- [16] Cella D. *FACIT Manual: Manual of the Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System*. 4th ed. Center on Outcomes, Research and Education; 1997.
- [17] Brand JP. *Development, Implementation and Evaluation of Multiple Imputation Strategies for the Statistical Analysis of Incomplete Data Sets*. Erasmus University Rotterdam; 1999.
- [18] Van Buuren S, Brand JP, Groothuis-Oudshoorn C, Rubin DB. Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*. 2006;76(12):1049–1064.
- [19] Durrleman S, Simon R. Flexible regression models with cubic splines. *Statistics in Medicine*. 1989;8(5):551–561.
- [20] Poirier DJ. Piecewise regression using cubic splines. *Journal of the American Statistical Association*. 1973;68(343):515–524.
- [21] Austin PC, Escobar M, Kopec JA. The use of the Tobit model for analyzing measures of health status. *Quality of Life Research*. 2000;9(8):901–910.
- [22] Huang I, Frangakis C, Atkinson MJ, Willke RJ, Leite WL, Vogel WB, et al. Addressing ceiling effects in health status measures: A comparison of techniques applied to measures for people with HIV disease. *Health Services Research*. 2008;43(1p1):327–339.
- [23] Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 1974;19(6):716–723.
- [24] Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978;6(2):461–464.

- [25] Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*. 2011;45(3).
- [26] Schafer JL. Multiple imputation: A primer. *Statistical Methods in Medical Research*. 1999;8(1):3–15.
- [27] Austin PC. A comparison of methods for analyzing health-related quality-of-life measures. *Value in Health*. 2002;5(4):329–337.
- [28] Pullenayegum EM, Tarride JE, Xie F, Goeree R, Gerstein HC, O'Reilly D. Analysis of health utility data when some subjects attain the upper bound of 1: Are Tobit and CLAD models appropriate? *Value in Health*. 2010;13(4):487–494.