Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mapping FACT-Hep to SF-6Dv2 utility values in patients with primary liver cancer: a comparison of conventional regression and machine learning approaches

Yujin Zhang
Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital
Wen Li
University of Electronic Science and Technology of China
Xianxi Liu
Chengdu Medical College
Longlin Jiang
Chengdu University of Traditional Chinese Medicine
Rong Hu
Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital
Lixia Zhang
Sichuan Cancer Hospital
Yang Qing
yangqingsc@163.com

Sichuan Cancer Hospital

Research Article

Keywords: primary liver cancer, FACT-Hep, SF-6Dv2, health utility, mapping, machine learning, health-related quality of life

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11292208/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose

To develop and internally validate mapping algorithms for predicting Short Form Six-Dimension version 2 (SF-6Dv2) utility values from the Functional Assessment of Cancer Therapy-Hepatobiliary (FACT-Hep) in patients with primary liver cancer, and to compare conventional regression and machine-learning approaches.

Methodology

A multicenter cross-sectional study included 608 patients from two hospitals in Sichuan Province, China. FACT-Hep, SF-6Dv2, demographic, and clinical data were collected. Conceptual overlap and multicollinearity were examined before model development. Five predictor specifications were evaluated using ordinary least squares (OLS), Tobit, ordered probit (Oprobit), ridge regression, support vector regression (SVR), and gradient boosting regression (GBR). The two best specifications from each method were retained for fivefold cross-validation and random 70/30 holdout validation. Model selection was based on prediction-error and agreement measures, average rank value (ARV), and graphical assessment.

Results

The mean observed SF-6Dv2 utility was 0.7332 (SD 0.2306), ranging from − 0.2270 to 1.0000. Oprobit5 performed best among the conventional models in the initial analysis, whereas GBR5 showed the strongest apparent performance among machine-learning models. After integrating fivefold cross-validation and random holdout validation, GBR4 achieved the lowest integrated ARV (3.05), followed by GBR5 (3.50), SVR4 (4.00), Oprobit5 (4.15), and SVR3 (4.25). Graphical analyses showed closer agreement in the central utility range but regression toward the mean at the extremes.

Conclusion

FACT-Hep can be mapped to SF-6Dv2 utility values with acceptable predictive performance. GBR4 showed the best overall performance after integrating the two internal validation procedures, while Oprobit5 was the strongest conventional alternative. Independent external validation is needed before wider application.

Background

Primary liver cancer remains a major global public-health burden. Hepatocellular carcinoma (HCC), which accounts for approximately 75–85% of primary liver cancers, is particularly important in China, where it ranks among the most frequently diagnosed malignancies and remains a leading cause of cancer-related mortality[1, 2]. The latest Chinese HCC guideline highlights the considerable national disease burden and the increasing complexity of contemporary treatment, including surgical resection, ablation, transarterial therapies, radiotherapy, molecularly targeted therapy, and immune-based systemic treatment[2] .

With advances in HCC treatment, clinical benefit can no longer be characterized solely by overall survival or progression-free survival. Patients may experience substantial physical symptoms, impaired functioning, fatigue, emotional distress, and treatment-related toxicities, making health-related quality of life (HRQoL) an increasingly important patient-centered endpoint[3–6]. A recent systematic review of systemic therapies for HCC confirmed the increasing incorporation of HRQoL into modern HCC trials, while an AASLD scoping review emphasized that patient-reported outcomes are associated with disease severity and prognosis and should receive greater attention in both research and clinical practice[4, 5].

The Functional Assessment of Cancer Therapy-Hepatobiliary (FACT-Hep) is a disease-specific HRQoL instrument developed for patients with hepatobiliary malignancies[7]. It combines the FACT-General core with hepatobiliary-specific concerns. FACT-Hep has demonstrated validity in hepatobiliary cancer populations and has also shown prognostic relevance in HCC[7–9]. A future study has shown that quality-of-life indices, such as FACT-Hep, can distinguish between different degrees of harm and offer prognostic information for patients with HCC independently of other indicators[8].

FACT-Hep is a non-preference-based measure, so it cannot directly provide the health-state utility values needed for quality-adjusted life-year (QALY) estimation. Cost-effectiveness analysis needs preference-based utility functions to combine both survival and HRQoL into a single indicator[10]. SF-6Dv2 is an updated preference-based health-state classification system based on SF-36v2[11]. A mainland Chinese SF-6Dv2 value set has been developed using time trade-off and discrete-choice methods to enhance its applicability in health-economic evaluations of the Chinese population [12].

Not all preference-based indicators are routinely collected with disease-specific HRQoL questionnaires in practice. As no direct utility data is available, a statistical method for mapping is employed to derive health utility from non-preference indicators[13–15]. The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Good Practices Task Force suggests that, in a mapping study, determine the conceptual overlap of the source and target instruments, explore alternative model specifications, assess predictive accuracy, and carry out appropriate validation[13].

Several mapping studies have used cancer-specific FACT instruments to estimate SF-6D utilities. FACT-G and FACT-C have been mapped to SF-6D in Chinese patients with colorectal neoplasms[16, 17], while FACT-B has been mapped to SF-6Dv2 in breast cancer[18]. More recently, mapping algorithms have also been developed from FACT-H&N and FACT-L to SF-6D in patients with thyroid and lung cancer[19, 20]. However, no suitable algorithm is currently available to convert FACT-Hep scores into SF-6Dv2 utilities for patients with primary liver cancer. To address this gap, the present study aimed to develop and internally validate a FACT-Hep-to-SF-6Dv2 mapping algorithm in a Chinese multicenter sample.

Machine-learning approaches have recently attracted interest in mapping because they can accommodate nonlinear relationships and complex interactions without requiring a fixed parametric functional form[21, 22]. However, machine learning should not automatically be assumed to outperform regression-based approaches. A 2025 systematic review found only modest average improvements in goodness-of-fit for machine-learning methods and emphasized the importance of interpretation, validation, and transportability[22].

Evidence specifically mapping FACT-Hep to SF-6Dv2 in patients with primary liver cancer remains scarce. Therefore, the study aimed to develop and internally validate FACT-Hep to SF-6Dv2 mapping algorithms in patients with primary liver cancer from Southwest China. Conventional regression methods and machine-learning algorithms were compared using multiple predictor specifications, two internal-validation strategies, and graphical assessments of distributional and individual-level agreement.

Methods

Study design and participants

This study followed the Mapping to Preference-Based Measures Reporting Standards (MAPS) checklist pro-posed for instrumental mapping [15, 23] and the Inter-national Society for Pharmacoeconomics and Outcomes Research (ISPOR) reporting standards on mapping [13], as well as the systematic review of mapping research in the annual report of the National Institute for Health and Care Excellence (NICE) [24].

This multicenter cross-sectional study was conducted at two tertiary Grade A hospitals in Sichuan Province, China, between July 2025 and March 2026. A total of 608 patients with primary liver cancer were recruited. The inclusion criteria were as follows: (1) patients diagnosed with primary liver cancer confirmed by imaging examination or pathological diagnosis; (2) aged 18–80 years; (3) clear consciousness with basic reading and communication abilities; (4) willingness to participate in

this study and written informed consent. The exclusion criteria included: (1) presence of other severe organic diseases; (2) mental disorders, cognitive impairment or significant communication difficulties; (3) currently participating in other clinical studies. All participants provided written informed consent prior to the survey. This study was approved by the Ethics Committee of Sichuan Provincial People's Hospital (Approval No. 2024-689-1).

Research tool

FACT-Hep

FACT-Hep is a disease-specific HRQoL instrument developed for hepatobiliary malignancies[7]. It comprises the 27-item FACT-G, covering physical well-being (PWB), social/family well-being (SWB), emotional well-being (EWB), and functional well-being (FWB), and an 18-item hepatobiliary cancer subscale, yielding 45 items in total. Each item is rated on a 5-point Likert scale from 0 (“not at all”) to 4 (“very much”), with negatively worded items reverse-scored as appropriate. The total score ranges from 0 to 180, with higher scores indicating better HRQoL. The Simplified Chinese version has demonstrated satisfactory reliability, validity, and responsiveness in Chinese patients with hepatocellular carcinoma [25, 26].

SF-6Dv2

SF-6Dv2 is a preference-based measure comprising six dimensions: physical functioning (PF), role limitation (RL), social functioning (SF), pain (PN), mental health (MH), and vitality (VT)[11]. Each dimension has five severity levels, except pain, which has six levels, with higher levels indicating greater health impairment. Health states were converted into utility values using the Chinese SF-6Dv2 value set[12]. A value of 1 represents full health, 0 represents a health state equivalent to death, and negative values represent states considered worse than death. The Chinese TTO-based value set allows utility values as low as approximately-0.277.

Data collection

FACT-Hep and SF-6Dv2 questionnaires were administered to all participants, together with the collection of demographic and clinical information. Trained researchers conducted data collection using standardized procedures to ensure data quality.

Conceptual overlap and multicollinearity

Before model development, the distributions of FACT-Hep and SF-6Dv2 scores were examined using descriptive statistics. Conceptual overlap between FACT-Hep and SF-6Dv2 was assessed using Pearson’s correlation coefficients. Correlations were calculated between FACT-Hep scores and the SF-6Dv2 utility index, as well as between the FACT-Hep domains and the six SF-6Dv2 dimensions. The absolute magnitude of the correlation coefficient (|r|) was interpreted as very weak (< 0.20), weak (0.20–0.39), moderate (0.40–0.59), strong (0.60–0.79), or very strong (≥ 0.80) [27, 28]. The direction and magnitude of these correlations were used to evaluate the conceptual overlap between the source and target measures.

Multicollinearity among the FACT-Hep domain predictors was assessed using variance inflation factors (VIFs). A VIF < 5 was considered to indicate the absence of problematic multicollinearity[29]. Conceptual overlap and multicollinearity were examined before model estimation to assess the suitability of FACT-Hep as a source measure for mapping onto SF-6Dv2 utilities, in accordance with recommendations for health utility mapping studies[13].

Mapping approaches

Six mapping approaches were evaluated, comprising three conventional regression methods (OLS, Tobit, and Oprobit) and three machine-learning methods (ridge regression, SVR, and GBR). OLS directly modeled SF-6Dv2 utility values, whereas Tobit regression was used to account for their bounded distribution. Oprobit was implemented as a response-mapping approach by separately modeling the six SF-6Dv2 dimensions and converting the predicted response-level probabilities into expected utilities using the Chinese SF-6Dv2 value set [13, 20]. Ridge regression, SVR, and GBR were used to investigate potentially complex relationships between FACT-Hep predictors and SF-6Dv2 utilities. Ridge regression uses L2 regularization to reduce model instability, SVR captures nonlinear relationships through kernel-based learning, and GBR sequentially combines regression trees to model nonlinearities and interactions [21, 22]. For GBR, 1,000 trees were fitted with an interaction depth of 3, a learning rate of 0.01, a minimum of 10 observations per terminal node, and a bag fraction of 0.8. All approaches were evaluated using the same predefined predictor specifications.

Model specifications

Five a priori predictor specifications were constructed.

Model 1

FACT-Hep total score.

Model 2

FACT-Hep total score plus its squared term.

Model 3

PWB + SWB + EWB + FWB +HCS domain scores.

Model 4

Model 3 plus the squared terms of PWB, SWB, EWB, FWB, and HCS.

Model 5

Model 4 plus age and sex.

The sequential specifications were designed to investigate whether domain-level information, nonlinear relationships, and demographic covariates improved utility prediction. Similar sequential specifications have been used in previous FACT-based mapping studies.

Model-performance assessment

For the conventional models, predictive performance was evaluated using root mean squared error (RMSE), mean absolute error (MAE), Lin's concordance correlation coefficient (CCC), the proportion of observations with absolute error greater than 0.05 (AE > 0.05), the proportion with absolute error greater than 0.10 (AE > 0.10), Akaike information criterion (AIC), and Bayesian information criterion (BIC).Lower RMSE, MAE, AE > 0.05, and AE > 0.10 indicate better predictive accuracy, whereas higher CCC indicates stronger agreement between predicted and observed utility values. AIC and BIC were used to compare alternative specifications within comparable likelihood-based model families rather than to directly compare models based on fundamentally different likelihood structures, consistent with previous recommendations.

Based on the preferred direction of each model family, performance indicators were ranked, and an average rank value (ARV) was calculated. Lower ARV showed better overall results.

Candidate-Model Selection and Internal Validation

Five predictor specifications were tested for each of the six model types, and 30 initial models were obtained. The two best-performing specifications in each approach were selected, and 12 candidate models were retained for internal validation. First, five-fold cross-validation was conducted by dividing the data into five roughly equal subsets; each subset served as the validation set once, and the other four were used for model training.

The same 12 models were also tested with a random 70/30 development-validation split. Out-of-fold or holdout predictions were evaluated by RMSE, MAE, CCC, AE > 0.05, and AE > 0.10. In summary, the performance of both sets of validation procedures was evaluated based on the five indicators, and a total of ten ranks were assigned. Lower RMSE, MAE, AE > 0.05, and AE > 0.10 and higher CCC received better ranks; the mean of the ten ranks was defined as the integrated ARV, and a lower value indicated better overall performance. One representative model from each model family was then selected for graphical comparison using violin plots, cumulative distribution functions (CDFs) and Bland-Altman plots with LOESS curves. All statistical analysis was performed in R 4.3.2.

Results

Participant characteristics

A total of 608 patients with primary liver cancer were included in the study. As shown in Table 1, 76.8% (n = 467) of the participants were male. The study population was predominantly older, with 46.2% (n = 281) aged > 60 years and 37.3% (n = 227) aged 51–60 years. Stage III was the most common TNM stage (35.0%, n = 213), followed by stage IV (30.3%, n = 184). Most participants were of Han ethnicity (93.4%, n = 568), and 51.8% (n = 315) had a disease duration of < 6 months.

Table 1
Participant demographic and clinical characteristics

Demographic profiles	**N (%)**
Gender	
Male	467 (76.8%)
Female	141 (23.2%)
Age	
18–30	5 (0.8%)
31–40	33 (5.4%)
41–50	62 (10.2%)
51–60	227 (37.3%)
>60	281 (46.2%)
TNM stage	
I	96 (15.8%)
II	115 (18.9%)
III	213 (35%)
IV	184 (30.3%)
Education	
Primary school and below	261 (42.9%)
Middle School	241 (39.6%)
High School	53 (8.7%)
Undergraduate/Diploma	53 (8.7%)
Course of the disease	
<6 months	315 (51.8%)
6 months < duration ≤ 1 year	165 (27.1%)
1 year < duration ≤ 5 years	99 (16.3%)
> 5 years	29 (4.8%)
Ethnicity	
Han ethnicity	568 (93.4%)
Others	40 (6.6%)
Health insurance	
Self-pay	19 (3.1%)
National health insurance	585(96.2%)
Other	4 (0.7%)

Figure 1 shows the distributions of FACT-Hep total scores and SF-6Dv2 utility values. FACT-Hep scores were concentrated toward the higher end of the scale, with the greatest density around 150 points. In contrast, SF-6Dv2 utility values showed a non-normal and multimodal distribution, with most observations concentrated in the higher utility range. The observed SF-6Dv2 utility values ranged from − 0.227 to 1.000, with a mean of 0.733 (SD = 0.231) and a median of 0.756. Overall, the two measures showed distinct distributional patterns, supporting the need to consider flexible modeling approaches when mapping FACT-Hep scores onto SF-6Dv2 utilities.

Conceptual overlap and multicollinearity

Figure 2 presents the conceptual overlap between FACT-Hep scores and SF-6Dv2 utility and dimensions, together with the assessment of multicollinearity among the five domain-level predictors. Overall, higher FACT-Hep scores were associated with higher SF-6Dv2 utility, whereas correlations with the six SF-6Dv2 dimension levels were predominantly negative, reflecting the opposite scoring directions of the two instruments. The FACT-Hep total score showed a moderate positive correlation with SF-6Dv2 utility ($r = 0.42$, $P < 0.001$). Among the FACT-Hep scores, PWB and TOI showed the highest correlations with utility (both $r = 0.52$), followed by FWB ($r = 0.45$), and HCS ($r = 0.39$), whereas SWB showed little correlation with utility ($r = -0.05$). At the dimension level, the strongest correlation was observed between TOI and SF-6Dv2 social functioning ($r = -0.58$). The VIFs for PWB, SWB, EWB, FWB, and HCS were 1.67, 1.22, 1.11, 1.70, and 1.96, respectively. All values were well below the prespecified threshold of 5, indicating no evidence of problematic multicollinearity among the domain-level predictors.

Model performance

A total of 30 mapping models were developed using six modeling approaches and five predictor specifications. As shown in Tables 2 and 3, models incorporating domain-level FACT-Hep information generally performed better than those based on the total score alone. Among the conventional approaches, Models 4 and 5 performed best within the OLS and Tobit families. OLS5 ranked first among the OLS models (RMSE = 0.1621, MAE = 0.1178, CCC = 0.672, ARV = 1.4), while Tobit5 ranked first among the Tobit models (RMSE = 0.1628, MAE = 0.1189, CCC = 0.682, ARV = 1.2). Oprobit5 was the highest-ranked Oprobit specification (RMSE = 0.1889, MAE = 0.1212, CCC = 0.458, ARV = 1.1), although its predictive accuracy and agreement were less favorable than those of OLS5 and Tobit5.

Table 2
Model performance of conventional regression methods for mapping FACT-Hep to SF-6D v2 utility score

Model	RMSE	MAE	CCC	AE > 0.05(%)	AE > 0.10(%)	AIC	BIC	ARV
OLS1	0.2095	0.1543	0.296	82.24	61.02	-169.43	-156.20	4.9
OLS2	0.2081	0.1522	0.312	82.73	59.21	-175.55	-157.91	4.1
OLS3	0.1809	0.1318	0.554	77.30	48.68	-339.56	-308.69	3.0
OLS4	0.1625	0.1188	0.669	71.22	45.23	-460.28	-407.36	1.6
OLS5	0.1621	0.1178	0.672	71.22	43.26	-459.51	-397.76	1.4
Tobit1	0.2098	0.1530	0.313	82.07	60.36	-1.86	11.37	4.7
Tobit2	0.2084	0.1513	0.326	82.89	60.20	-4.93	12.71	4.3
Tobit3	0.1814	0.1305	0.570	74.51	47.53	-173.35	-142.48	3.0
Tobit4	0.1632	0.1198	0.680	73.03	44.57	-288.56	-235.64	1.8
Tobit5	0.1628	0.1189	0.682	73.03	42.93	-288.77	-227.03	1.2
Oprobit1	0.2223	0.1485	0.186	76.32	53.78	8727.48	8855.37	4.7
Oprobit2	0.2214	0.1476	0.198	76.97	56.09	8696.71	8851.07	4.3
Oprobit3	0.2015	0.1314	0.361	70.39	43.42	8053.04	8286.78	3.0
Oprobit4	0.1892	0.1223	0.456	69.57	38.65	7629.85	7995.89	1.9
Oprobit5	0.1889	0.1212	0.458	67.93	37.83	7590.24	8009.21	1.1

Table 3
Model performance of machine-learning methods for mapping FACT-Hep to SF-6D v2 utility score

Model	RMSE	MAE	CCC	AE > 0.05(%)	AE > 0.10(%)	R^2	MAPE	ICC	ARV
GBR1	0.1893	0.1407	0.467	80.10	58.72	0.3249	56.603	0.467	4.5
GBR2	0.1893	0.1407	0.467	80.10	58.72	0.3249	56.603	0.467	4.5
GBR3	0.1155	0.0806	0.847	53.13	25.49	0.7489	33.449	0.848	2.4
GBR4	0.1155	0.0806	0.847	53.13	25.49	0.7489	33.449	0.848	2.4
GBR5	0.1145	0.0794	0.849	54.93	25.16	0.7531	29.275	0.849	1.3
Ridge1	0.2095	0.1543	0.296	82.24	61.02	0.1737	65.357	0.296	4.9
Ridge2	0.2081	0.1522	0.312	82.73	59.21	0.1847	62.023	0.312	4.1
Ridge3	0.1792	0.1303	0.557	76.81	48.52	0.3950	56.296	0.557	3.0
Ridge4	0.1617	0.1176	0.669	70.72	44.41	0.5073	48.634	0.670	2.0
Ridge5	0.1613	0.1165	0.671	70.39	42.60	0.5099	48.002	0.672	1.0
SVR1	0.2092	0.1467	0.283	82.89	56.91	0.1761	68.328	0.283	4.3
SVR2	0.2118	0.1469	0.264	80.76	56.41	0.1555	68.379	0.264	4.8
SVR3	0.1599	0.1132	0.646	70.89	41.61	0.5182	50.893	0.647	1.0
SVR4	0.1788	0.1233	0.522	73.68	46.05	0.3978	59.369	0.523	2.1
SVR5	0.1790	0.1235	0.521	73.68	46.22	0.3966	59.435	0.521	2.9

Among the machine-learning approaches, GBR5 showed the best apparent performance (RMSE = 0.1145, MAE = 0.0794, CCC = 0.849, R^2 = 0.7531, ICC = 0.849, ARV = 1.3). GBR3 and GBR4 yielded identical performance across the evaluated metrics (RMSE = 0.1155, MAE = 0.0806, CCC = 0.847). Because the additional squared terms included in GBR4 did not improve predictive performance, GBR3 was retained in preference to GBR4 on the basis of model parsimony. Ridge5 and SVR3 ranked first within the ridge and SVR families, respectively. Based on the within-method rankings, with parsimony used to resolve the tie between GBR3 and GBR4, 12 candidate models were retained for internal validation: OLS4, OLS5, Tobit4, Tobit5, Oprobit4, Oprobit5, GBR3, GBR5, Ridge4, Ridge5, SVR3, and SVR4.**Internal Validation**

The 12 candidate models were evaluated using fivefold cross-validation and random 70/30 holdout validation (Table 4). In fivefold cross-validation, GBR3 achieved the lowest RMSE (0.1495), while GBR5 had the lowest MAE (0.0988); their CCC values were 0.740 and 0.736, respectively. In the random holdout validation, GBR5 showed the most favorable overall error profile, with RMSE = 0.1551, MAE = 0.1009, and CCC = 0.726, while GBR3 yielded RMSE = 0.1574, MAE = 0.1047, and CCC = 0.721. When the ten metric ranks from both validation procedures were integrated, GBR5 achieved the lowest ARV (1.40), followed by GBR3 (1.60). The next-ranked models were SVR3 (5.45), Ridge5 (5.55), and Ridge4 (5.70). Thus, the two GBR specifications were clearly the most stable candidates across the two internal validation procedures.

Table 4
Results of model validation analysis: goodness of fit assessed through fivefold cross-validation and random sampling validation

	Fivefold-cross validation					Random sample validation					
	RMSE	**MAE**	**CCC**	**AE > 0.05(%)**	**AE > 0.1(%)**	**RMSE**	**MAE**	**CCC**	**AE > 0.05(%)**	**AE > 0.1(%)**	**ARV**
OLS4	0.1665	0.1216	0.655	73.684	46.382	0.1726	0.1219	0.653	71.585	44.262	7.1
OLS5	0.1670	0.1211	0.653	72.862	45.066	0.1749	0.1238	0.646	72.131	42.623	7.4
Tobit4	0.1674	0.1229	0.665	74.671	44.737	0.1750	0.1246	0.659	73.770	48.087	8
Tobit5	0.1679	0.1223	0.664	74.671	44.901	0.1782	0.1275	0.651	75.956	46.995	8.85
Oprobit4	0.1902	0.1235	0.451	69.737	39.309	0.1996	0.1261	0.438	73.770	38.798	8.55
Oprobit5	0.1905	0.1226	0.450	68.586	38.816	0.2017	0.1264	0.430	69.945	39.344	8.05
GBR3	0.1495	0.0990	0.740	55.592	29.934	0.1574	0.1047	0.721	59.563	32.787	1.6
GBR5	0.1501	0.0988	0.736	55.757	30.428	0.1551	0.1009	0.726	56.284	32.240	1.4
Ridge4	0.1657	0.1202	0.654	72.697	45.559	0.1709	0.1198	0.654	71.038	42.623	5.7
Ridge5	0.1661	0.1196	0.653	71.875	44.408	0.1732	0.1211	0.646	69.945	41.530	5.55
SVR3	0.1655	0.1176	0.622	71.711	43.257	0.1744	0.1196	0.612	71.038	42.077	5.45
SVR4	0.1815	0.1255	0.507	73.191	45.395	0.1950	0.1305	0.472	72.131	49.180	10.35

Distributional comparison

Numerical performance was supplemented by an analysis of the distribution and agreement between observed and predicted SF-6Dv2 utilities. As shown in Table 5 and Fig. 3, the observed utility values had a wider distribution (mean = 0.7332, SD = 0.2306; range − 0.2270 to 1.0000) than most of the model's predictions. GBR5 closely matched the observed mean (0.7332) and had a relatively wide predicted distribution (SD = 0.1839; range − 0.1538 to 0.9938). GBR3 also had a mean of 0.7332 and an SD of 0.1854. By contrast, Oprobit5 produced a higher and more compressed distribution (mean = 0.7973, SD = 0.0931), while OLS and Tobit predictions extended above 1.0. Ridge predictions were capped at an upper bound of 1.0 in the descriptive results, and SVR3 had a slightly higher mean of 0.7493.

Table 5
Descriptive summary of SF-6D(v2) utility index derived from observed and predicted values of best fitting

Model	Mean	SD	Minimum	P10	Median	P90	Maximum
sf6d	0.7332	0.2306	-0.2270	0.4525	0.7560	0.9620	1.0000
OLS4	0.7332	0.1635	0.1861	0.4943	0.7718	0.9037	1.0969
OLS5	0.7332	0.1640	0.1772	0.4950	0.7697	0.9035	1.1010
Tobit4	0.7411	0.1735	0.1661	0.4907	0.7768	0.9250	1.1343
Tobit5	0.7412	0.1743	0.1550	0.4842	0.7785	0.9262	1.1402
Oprobit4	0.7974	0.0923	0.4860	0.6569	0.8174	0.8969	0.9662
Oprobit5	0.7973	0.0931	0.4806	0.6586	0.8156	0.8983	0.9673
GBR3	0.7332	0.1854	-0.1840	0.4671	0.7633	0.9254	0.9760
GBR5	0.7332	0.1839	-0.1538	0.4678	0.7666	0.9237	0.9938
Ridge4	0.7320	0.1613	0.1862	0.4943	0.7718	0.9037	1.0000
Ridge5	0.7319	0.1616	0.1774	0.4950	0.7701	0.9033	1.0000
SVR3	0.7493	0.1378	0.2715	0.5455	0.7857	0.8892	1.0036
SVR4	0.7576	0.1151	0.4116	0.5923	0.7792	0.8767	1.0373

As shown in Fig. 4, the CDF curves of the predicted distributions are closest to the observed distribution in the central utility range and diverge more significantly in the tails. As shown in Fig. 5, the Bland-Altman plots with LOESS curves all exhibited a regression-to-the-mean pattern; that is, poorer observed health states were generally overpredicted, and higher observed utilities were generally underpredicted. Among the shown models, GBR5 had a relatively good distribution agreement but still showed prediction errors at the edges.

Considering the above, GBR5 was selected as the preferred mapping model because it had the lowest integrated ARV (1.40) in five-fold cross-validation and random 70/30 holdout validation, and showed good agreement with the observed utility distribution. GBR3 was a close alternative with an integrated ARV of 1.60. Oprobit5 was still the top-ranked Oprobit specification in the initial within-method screening, but its internal-validation performance was lower than that of the leading machine-learning models.Table 5 Descriptive summary of SF-6D(v2) utility index derived from observed and predicted values of best fitting models.

Discussion

This study created and internally validated algorithms to map FACT-Hep onto SF-6Dv2 utilities for patients with primary liver cancer. Thirty models were initially estimated using 6 types of models and 5 sets of predictors, and 12 models were selected for further 5-fold cross validation and random 70/30 validation. Test results indicated that GBR5 had the best overall internal-validation performance (ARV = 1.40) with GBR3 following closely after (ARV = 1.60). The results of Models 3–5 compared with those of total-score specifications suggest it is helpful to retain the domain-level information from PWB, SWB, EWB, FWB and the hepatobiliary cancer subscale (HCS) for prediction of utility. Note, however, that the improvement should not be solely credited to HCS since the several predictors were introduced in one fell swoop across specifications. Another Chinese mapping study also found models which consider subscale information and nonlinear terms may be more useful than summary score models[30].

The observed associations of FACT-Hep and SF-6Dv2 also support the mapping hypothesis. Physical and functional attributes had a relatively large overlap, but associations in some social-emotional areas were weaker. It is not surprising that FACT-Hep focuses on HRQoL related to hepatobiliary cancer, and SF-6Dv2 is a general preference-based instrument. Recently, Chinese data have shown that SF-6Dv2 can distinguish between different types of cancer and clinical features in patients with liver cancer [31].

Oprobit5 was the top performer in the Oprobit family during the initial screening, but its apparent RMSE (0.1889) and CCC (0.458) were lower than those of OLS5 and Tobit5. Furthermore, Oprobit4 and Oprobit5 performed relatively poorly in the internal validation, with integrated ARVs of 8.55 and 8.05, respectively. Therefore, response mapping is still a conceptually appealing way because it predicts ordered SF-6Dv2 dimension responses before deriving utilities, but it was not the most accurate method for the current dataset. This supports the previous results that the relative advantages of the direct and response-mapping approaches vary depending on the source instrument, target utility measure and data distribution. [30, 32].

GBR was the best one after the internal validation. GBR3 and GBR5 were significantly better ranked than the other candidate models, and GBR5 had a low prediction error and the highest overall validation rank. The difference between GBR3 and GBR5 was relatively small, and adding the squared domain terms to the age and sex terms only improved the performance marginally over the domain-level specification; however, GBR5 demonstrated better stability in the random holdout set. Gradient Boosting can handle non-linear relationships and interactions among the data, which may be suitable for the irregular and multi-modal distribution of SF-6Dv2 utilities in this sample. Machine-learning methods have also shown good predictive performance in other Chinese cancer mapping studies. [33]. Nevertheless, the close performance of GBR3 and GBR5 indicates that additional model complexity should be justified by reproducible validation gains rather than apparent fit alone.Another important finding was the reduced ability of the models to reproduce extreme utility values. Predicted distributions were generally narrower than the observed distribution, with a tendency to overestimate poorer health states and underestimate better health states. Such regression toward the mean is a recurrent challenge in utility mapping and suggests that model evaluation should not rely solely on average prediction errors. Studies using SF-6Dv2 have therefore increasingly combined RMSE and MAE with cross-validation and distributional or agreement-based assessments [34, 35].

The main contribution of this study is the development of a mapping framework specifically linking FACT-Hep with SF-6Dv2 in primary liver cancer. Rather than selecting the model with the best apparent fit, we used a staged process involving candidate screening, two internal validation procedures, integrated ARV ranking, and graphical assessment. The results identify GBR5 as the preferred model because it achieved the lowest integrated ARV across the two validation procedures, while GBR3 provided a similarly competitive and more parsimonious alternative. This approach reduces reliance on any single performance statistic and provides a more balanced basis for model selection. The study also has practical implications. SF-6Dv2 is increasingly being used to generate utilities for economic evaluation, including in Chinese cancer populations [31, 36].

If FACT-Hep has been collected but preference-based utility data are unavailable, the proposed algorithm can still provide SF-6Dv2 estimates for QALY calculation and cost-utility analysis. This is especially true for retrospective datasets and clinical studies that do not contain preference-based indicators. Mapped utilities should supplement, rather than substitute for, directly measured utilities when direct measurement is feasible.

This study has several strengths. FACT-Hep and SF-6Dv2 were collected simultaneously at multiple centres, and a mapping system based on patient-level paired observations was developed. Six traditional and machine-learning methods were compared with the same predictor specifications. In addition, five-fold cross-validation, random sample validation, multiple prediction indicators and graphs were used to assess the performance of the model more comprehensively than a single goodness-of-fit index in model selection.

The following are some deficiencies. First, although this was a multi-center study, both hospitals were in Sichuan Province, and the generalizability of the algorithm to other areas has not been verified. Second, five-fold cross-validation and random sample validation were both carried out on the same dataset, and thus constituted internal rather than external validation. Third, prediction remained less accurate at the extremes of the SF-6Dv2 distribution, particularly for poorer health states.

In the future, GBR4 will be validated in multiple contexts independently, and its effectiveness at various times during the disease and treatment process will also be explored. Other clinical predictors and alternative nonlinear or ensemble methods can also be used to improve the accuracy of prediction for serious illness. Given that utility values may differ among populations and value sets, the applicability of the algorithm outside the Chinese SF-6Dv2 context should also be examined carefully [37, 38].

Conclusion

FACT-Hep can be mapped onto SF-6Dv2 with acceptable predictive performance. After integrating the two internal validation procedures, GBR5 showed the best overall performance. The algorithm may support utility estimation when SF-6Dv2 data are unavailable, but independent external validation is required before broader use in health economic evaluation.

Declarations

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests

The authors have no relevant financial or non-financial interests to disclose

Author contributions

The conceptualization was performed by Qing Yang. The data curation was performed by Yujin Zhang, Rong Hu, Longlin Jiang, Wen Li and Xianxi Liu. The formal analysis was performed by Yujin Zhang and Wen Li. The methodology was performed by Qing Yang and Lixia Zhang. The project administration was performed by Qing Yang, Yujin Zhang and Wen Li. The original draft writing was performed by Yujin Zhang and Wen Li. The draft review and editing were performed by Qing Yang and Lixia Zhang. All authors read and approved the final manuscript.

Ethics approval

This study was approved by the Ethics Committee of Sichuan Provincial People's Hospital (Approval No. 2024-689-1).

Consent to participate

All participants have signed informed consent forms, and this manuscript does not contain any identifiable patient information.

Consent to publish

All authors have read the final manuscript, fully agreed with all contents of this work, and consented to its submission and formal publication in this journal.

References

1. Sung, H., Filho, A. M., Laversanne, M., Ferlay, J., Siegel, R. L., Soerjomataram, I., et al. (2026). Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. *C Ca: A Cancer Journal For Clinicians*, *76*(4), e70090. 10.3322/caac.70090
2. Zhou, J., Sun, H., Wang, Z., Cong, W., Zeng, M., Zhou, W., et al. (2025). China Liver Cancer Guidelines for the Diagnosis and Treatment of Hepatocellular Carcinoma (2024 Edition). *Liver Cancer*, *14*(6), 779–835. 10.1159/000546574
3. Vogel, A., Chan, S. L., Dawson, L. A., Kelley, R. K., Llovet, J. M., Meyer, T., et al. (2025). Hepatocellular carcinoma: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Annals Of Oncology*, *36*(5), 491–506. 10.1016/j.annonc.2025.02.006
4. Jayabalan, D., Dhakal, S., Raguragavan, A., Saxena, A., Jeffrey, G. P., Calzadilla-Bertot, L. (2025). Hepatocellular Carcinoma and Health-Related Quality of Life: A Systematic Review of Outcomes From Systemic Therapies. Int J Hepatol, 2025:1083642. 10.1002/hep.32313
5. Serper, M., Parikh, N. D., Thiele, G., Ovchinsky, N., Mehta, S., Kuo, A., et al. (2022). Patient-reported outcomes in HCC: A scoping review by the Practice Metrics Committee of the American Association for the Study of Liver Diseases. *Hepatology*, *76*(1), 251–274. 10.1002/hep.32313
6. Finn, R. S., Kudo, M., Merle, P., Meyer, T., Qin, S., Ikeda, M., et al. (2025). Health-related quality of life analysis of the LEAP-002 study of lenvatinib plus pembrolizumab versus lenvatinib as first-line treatment for advanced hepatocellular carcinoma. *ESMO Open*, *10*(6), 105065. 10.1016/j.esmoop.2025.105065
7. Heffernan, N., Cella, D., Webster, K., Odom, L., Martone, M., Passik, S., et al. (2002). Measuring health-related quality of life in patients with hepatobiliary cancers: the functional assessment of cancer therapy-hepatobiliary questionnaire. *Journal Of Clinical Oncology*, *20*(9), 2229–2239. 10.1200/JCO.2002.07.093
8. Kim, G. A., Kim, H. I., Chang, S., An, J., Lee, D., Lee, H. C., et al. (2019). A Prospective Evaluation of the Reliability and Utility of Quality of Life Measures in Patients With Hepatocellular Carcinoma. *American Journal Of Clinical Oncology*, *42*(7), 555–563. 10.1097/COC.0000000000000555
9. Gmür, A., Kolly, P., Knöpfli, M., & Dufour, J. F. (2018). FACT-Hep increases the accuracy of survival prediction in HCC patients when added to ECOG Performance Status. *Liver International : Official Journal Of The International Association For The Study Of The Liver*, *38*(8), 1468–1474. 10.1111/liv.13711
10. Brazier, J., Ara, R., Azzabi, I., Busschbach, J., Chevrou-Séverac, H., Crawford, B., et al. (2019). Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value In Health : The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*, *22*(3), 267–275. 10.1016/j.jval.2019.01.004
11. Brazier, J. E., Mulhern, B. J., Bjorner, J. B., Gandek, B., Rowen, D., Alonso, J. (2020). Developing a New Version of the SF-6D Health State Classification System From the SF-36v2: SF-6Dv2. Med Care, 2020;58(6):557 – 65. 10.1097/MLR.0000000000001325
12. Wu, J., Xie, S., He, X., Chen, G., Bai, G., Feng, D., et al. (2021). Valuation of SF-6Dv2 Health States in China Using Time Trade-off and Discrete-Choice Experiment with a Duration Dimension. *Pharmacoeconomics*, *39*(5), 521–535. 10.1007/s40273-020-00997-1
13. Wailoo, A. J., Hernandez-Alava, M., Manca, A., Mejia, A., Ray, J., & Crawford, B., et a.l (2017). Mapping to Estimate Health-State Utility from Non-Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value In Health : The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*, *20*(1), 18–27. 10.1016/j.jval.2016.11.006
14. Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., et al. (2015). Preferred Reporting Items for Studies Mapping onto Preference-Based Outcome Measures: The MAPS Statement. *Pharmacoeconomics*, *33*(10), 985–991. 10.1186/s12955-015-0305-6
15. Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., et al. (2015). The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration.

Pharmacoeconomics, *33*(10), 993–1011. 10.1007/s40273-015-0312-9

16. Wong, C. K., Lam, C. L., Rowen, D., McGhee, S. M., Ma, K. P., & Law, W. L. (2012). Mapping the Functional Assessment of Cancer Therapy-general or -Colorectal to SF-6D in Chinese patients with colorectal neoplasm. *Value In Health : The Journal Of The International Society For Pharmacoeconomics And Outcomes Research*, *15*(3), 495–503. 10.1016/j.jval.2011.12.009
17. Teckle, P., McTaggart-Cowan, H., Van der Hoek, K., Chia, S., Melosky, B., Gelmon, K., et al. (2013). Mapping the FACT-G cancer-specific quality of life instrument to the EQ-5D and SF-6D. *Health And Quality Of Life Outcomes*, *11*, 203. 10.1186/1477-7525-11-203
18. Nahvijou, A., Safari, H., Yousefi, M., Rajabi, M., Arab-Zozani, M., & Ameri, H. (2021). Mapping the cancer-specific FACT-B onto the generic SF-6Dv2. *Breast Cancer*, *28*(1), 130–136. 10.1007/s12282-020-01141-9
19. Yang, Q., Huang, D., Jiang, L., Tang, Y., & Zeng, D. (2023). Obtaining SF-6D utilities from FACT-H&N in thyroid carcinoma patients: development and results from a mapping study. *Front Endocrinol (Lausanne)*, *14*, 1160882. 10.3389/fendo.2023.1160882
20. Yang, Q., Jiang, L. L., Li, Y. F., & Huang, D. (2023). Prediction of the SF-6D utility score from Lung cancer FACT-L: a mapping study in China. *Health And Quality Of Life Outcomes*, *21*(1), 122. 10.1186/s12955-024-02237-y
21. Aghdaee, M., Parkinson, B., Sinha, K., Gu, Y., Sharma, R., Olin, E., et al. (2022). An examination of machine learning to map non-preference based patient reported outcome measures to health state utility values. *Health Economics*, *31*(8), 1525–1557. 10.1002/hec.4503
22. Hong, T., Xie, S., Liu, X., Wu, J., & Chen, G. (2025). Do Machine Learning Approaches Perform Better Than Regression Models in Mapping Studies? *A Systematic Review Value Health*, *28*(5), 800–811. 10.1016/j.jval.2024.12.010
23. Dakin, H., Abel, L., Burns, R., & Yang, Y. (2018). Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. *Health And Quality Of Life Outcomes*, *16*(1), 31. 10.1186/s12955-018-0857-3
24. Longworth, L., Yang, Y., Young, T., Mulhern, B., Hernández Alava, M., Mukuria, C., et al. (2014). Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technology Assessment*, *18*(9), 1–224. 10.3310/hta18090
25. Zhu, Z. C., Lang, Q. B., Chen, Z., Li, D. T., & Ling, C. Q. (2008). Evaluation of Chinese version of the Functional Assessment of Cancer Therapy-Hepatobiliary questionnaire. *Zhong Xi Yi Jie He Xue Bao*, *6*(4), 341–345. 10.3736/jcim20080403
26. Wang, Y., Ding, S., Meng, Q., Cun, Y., Li, G., & Chen, Y. (2026). A Longitudinal Study on Psychometric Properties of the Simplified Chinese Version of FACT-Hep for Patients with Hepatocellular Carcinoma by Classical Test Theory and Generalizability Theory. *Cancer Manag Res*, *18*, 547225. 10.2147/CMAR.S547225
27. Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia And Analgesia*, *126*(5), 1763–1768. 10.1213/ANE.0000000000002864
28. Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal : The Journal Of Medical Association Of Malawi*, *24*(3), 69–71.
29. Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean J Anesthesiol*, *72*(6), 558–569. 10.4097/kja.19087
30. Chen, Z., Yang, L., & Zhang, Y. (2025). Mapping KDQOL-36 Onto EQ-5D-5L and SF-6Dv2 in Patients Undergoing Dialysis in China. *Value Health Reg Issues*, *48*, 101103.
31. Tao, Y., Cao, Y., Xu, L., Shi, J., Leng, L., Yang, H., et al. (2025). Health utility scores of six common cancers in China measured by SF-6Dv2. *Health And Quality Of Life Outcomes*, *23*(1), 5. 10.1186/s12955-025-02332-8
32. Zhao, Z., Shi, Z., Dou, L., Li, C., & Li, S. (2026). Mapping Health State Utility from Disease-Specific Measures in Spinal Muscular Atrophy and Paroxysmal Nocturnal Hemoglobinuria. *Pharmacoeconomics*, *44*(2), 187–205. 10.1007/s40273-025-01549-1

33. Jiang, L., Li, K., Lu, S., Hong, Z., Wang, Y., Xie, Q., et al. (2025). Mapping the EORTC QLQ-C30 and QLQ-LC13 to the SF-6D utility index in patients with lung cancer using machine learning and traditional regression methods. *Health And Quality Of Life Outcomes*, *23*(1), 66. 10.1186/s12955-025-02394-8
34. Guo, W., Xie, S., Wang, D., & Wu, J. (2024). Mapping IWQOL-Lite onto EQ-5D-5L and SF-6Dv2 among overweight and obese population in China. *Quality Of Life Research*, *33*(3), 817–829. 10.1007/s11136-023-03568-x
35. Fang, H., Hong, T., Liu, X., Luo, C., Hou, Y., & Xie, S. (2025). Mapping the ADDQoL to the EQ-5D-5L and SF-6Dv2 among Chinese patients with type 2 diabetes mellitus. *Health And Quality Of Life Outcomes*, *23*(1), 46. 10.1186/s12955-025-02371-1
36. Li, M., Fang, B., Gu, H., & Jiang, Y. (2024). EQ-5D-5L and SF-6Dv2 health utilities scores of diffuse large B-cell lymphoma patients in China. *Health And Quality Of Life Outcomes*, *22*(1), 80. 10.1186/s12955-024-02297-0
37. Touré, M., & Poder, T. G. (2024). Differences in health utilities between cancer patients and the general population: The case of Quebec using the SF-6Dv2. *Social Science And Medicine*, *351*, 117001. 10.1016/j.socscimed.2024.117001
38. Mukuria, C., Rowen, D., Mulhern, B., McDool, E., Kharroubi, S., Bjorner, J. B., et al. (2025). The Short Form 6 Dimensions (SF-6D): Development and Evolution. *Applied Health Economics And Health Policy*, *23*(1), 19–33. 10.1007/s40258-024-00919-8

Figures

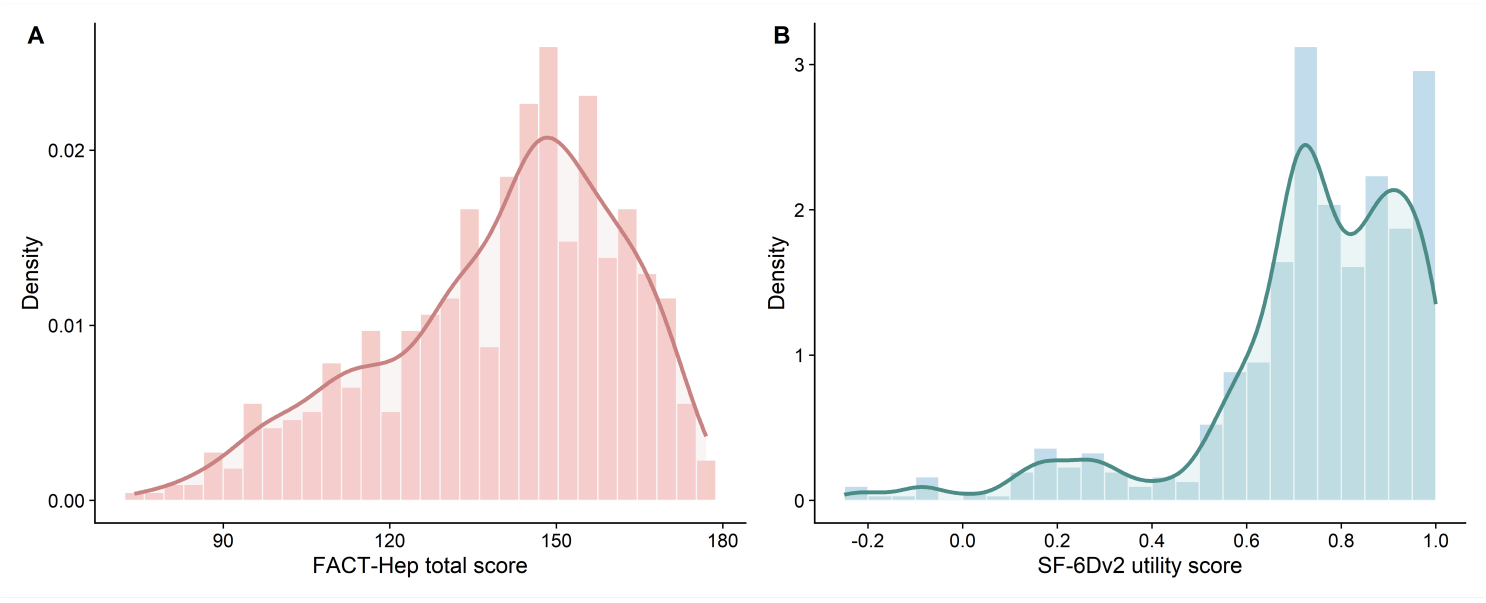


Figure 1

Distribution of FACT-Hep total scores and SF-6Dv2 utility values.

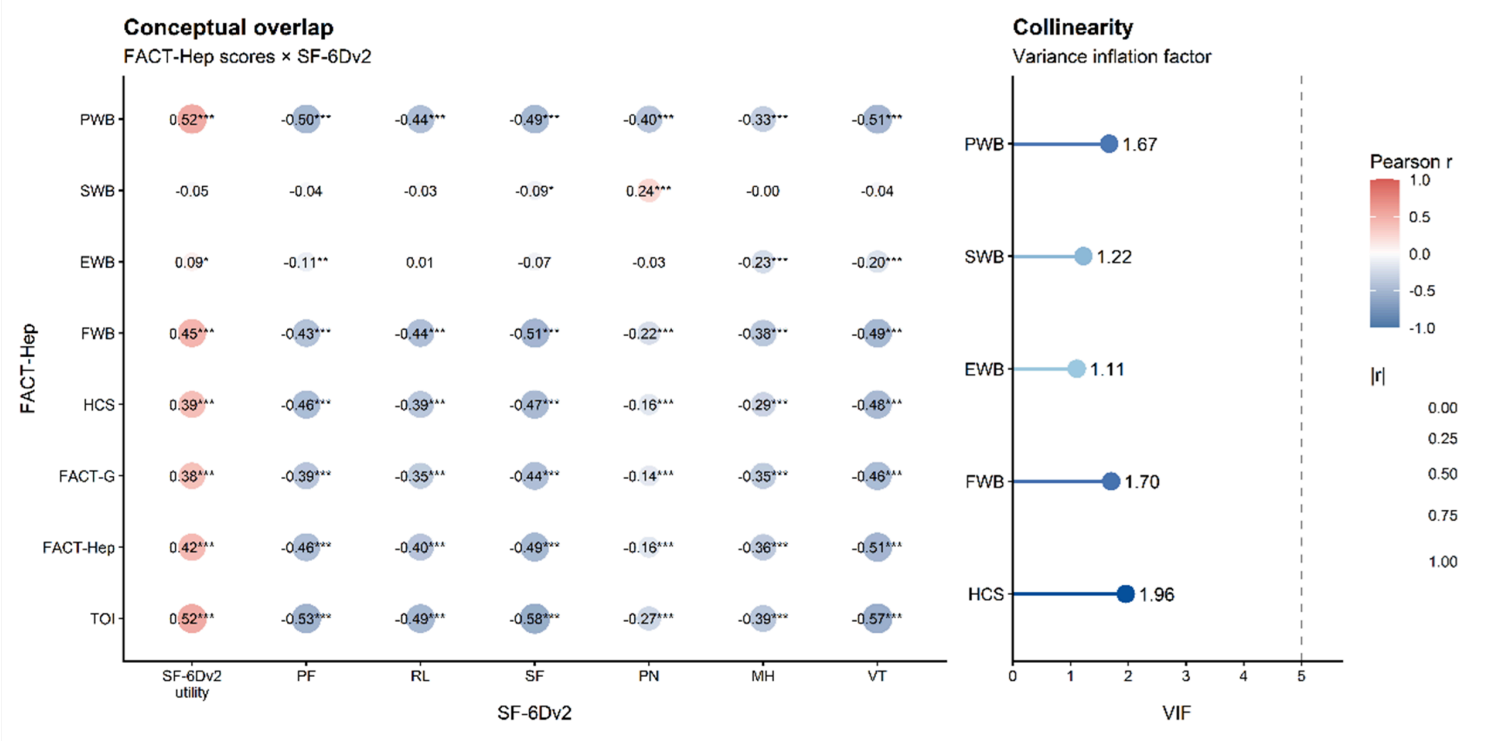


Figure 2

Conceptual overlap between FACT-Hep domains and SF-6D v2 dimensions and variance inflation factors for FACT-Hep predictors

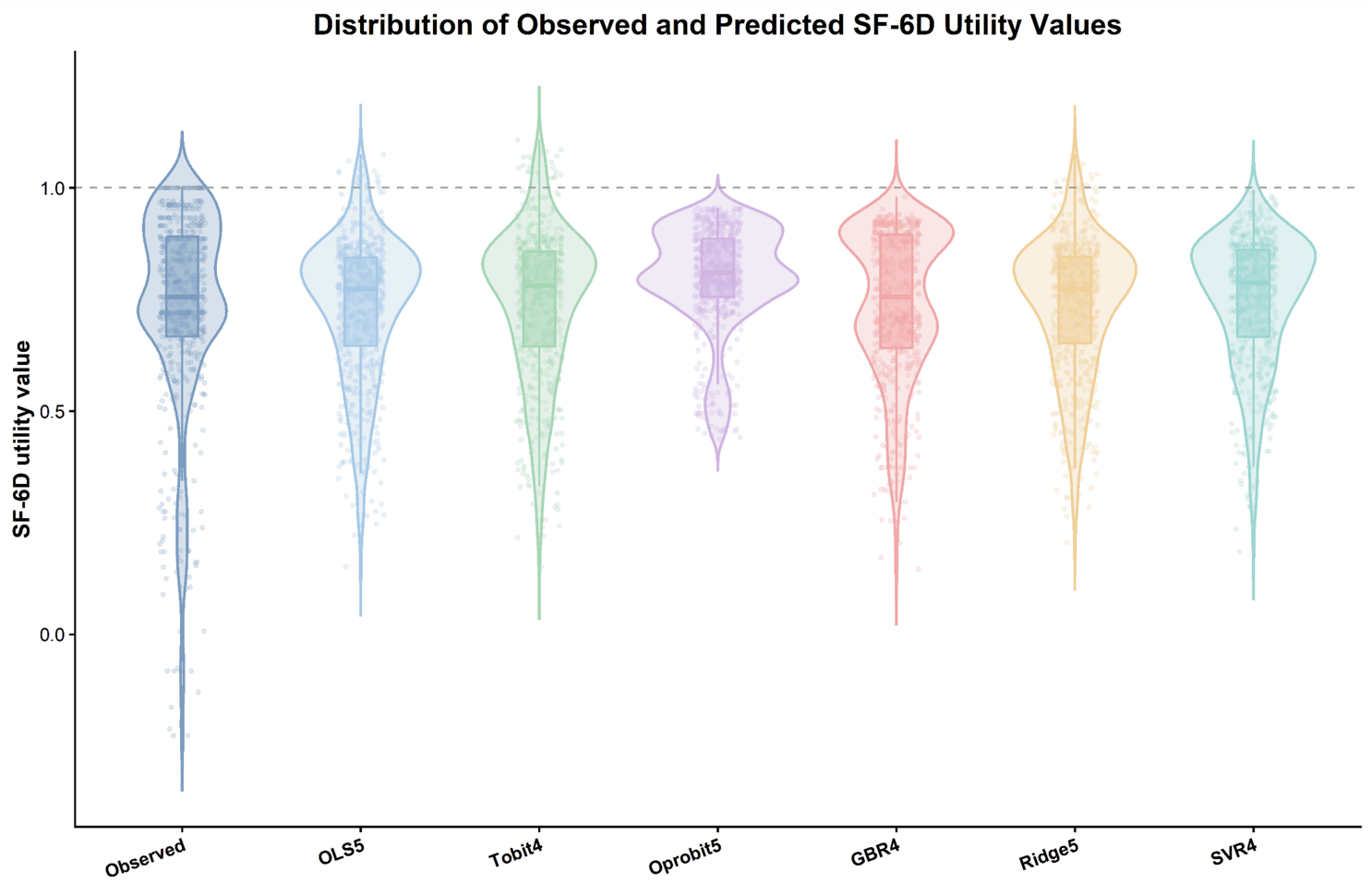


Figure 3

Distribution of observed and predicted SF-6D v2 utility values across selected mapping models

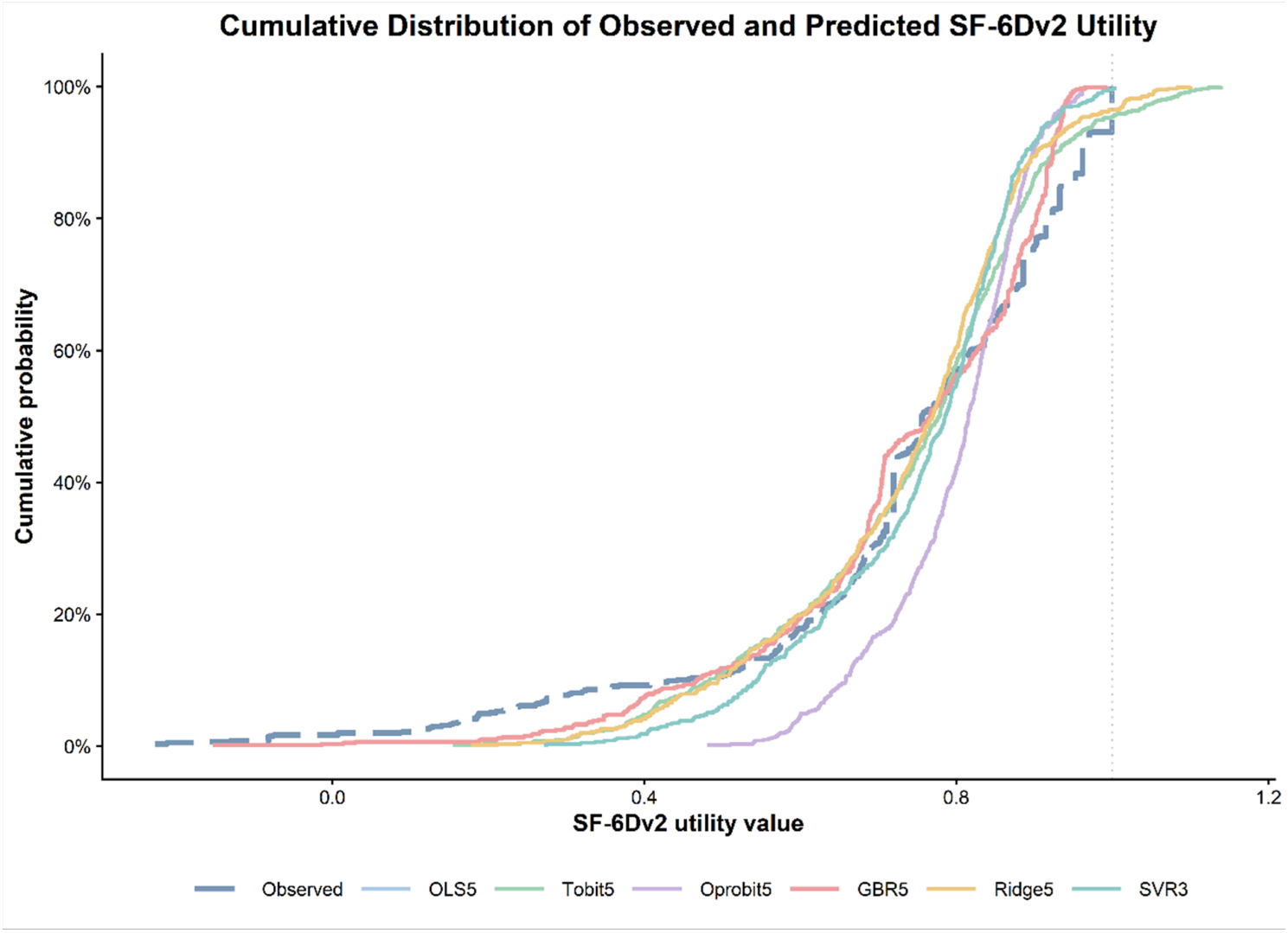


Figure 4

Cumulative distributions of observed and predicted SF-6D v2 utility values across selected mapping models

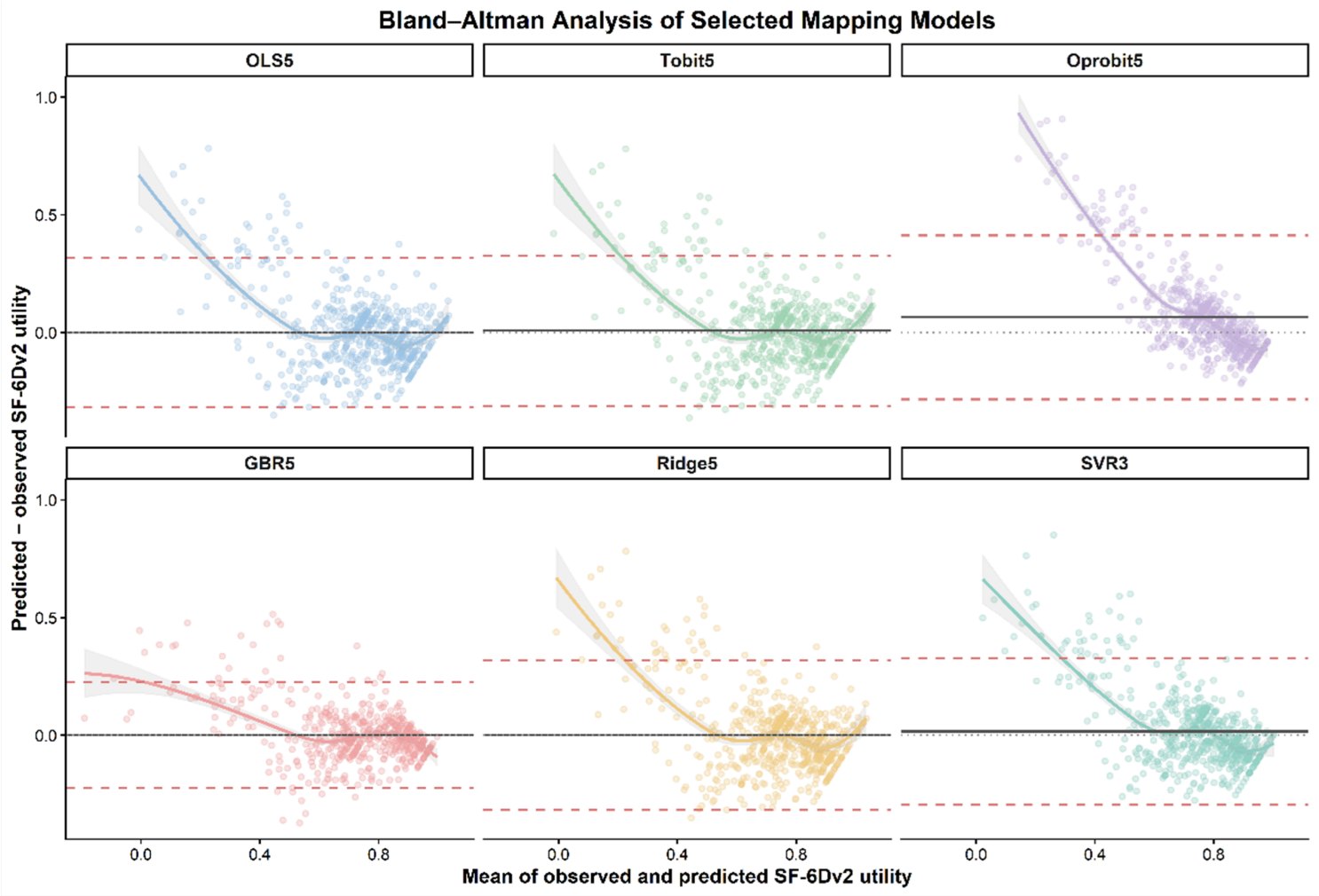


Figure 5

Bland–Altman plots with LOESS curves for agreement between observed and predicted SF-6D (v2) utility values