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 EORTC QLQ-C30 into EQ-5D-5L index for Korean patients with gastric cancer

Doyeon Jin
Ewha Womans University
Sangyoon Shin
Ewha Womans University
Bum Jun Kim
Hallym University Sacred Heart Hospital, Hallym University College of Medicine
Minkyu Jung
Yonsei University College of Medicine
Jeonghoon Ahn
ahnjeonghoon@ewha.ac.kr
Ewha Womans University
Dae Young Zang
Hallym University Sacred Heart Hospital, Hallym University College of Medicine

Research Article

Keywords: EQ-5D-5L, EORTC QLQ-C30, Direct mapping, Gastric cancer

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11332116/v1

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

Additional Declarations: No competing interests reported.

Mapping EORTC QLQ-C30 into EQ-5D-5L index for Korean patients with gastric cancer

Doyeon Jin[1] · Sangyoon Shin[1] · Bum Jun Kim[2] · Minkyu Jung[3] · Jeonghoon Ahn[1,*] · Dae Young Zang[2,*]

[1] Department of Health Convergence, Ewha Womans University, Seoul, Republic of Korea

[2] Division of Oncology-Hematology, Department of Internal Medicine, Hallym University Sacred Heart Hospital, Hallym University College of Medicine, Anyang-si, Republic of Korea

[3] Division of Medical Oncology, Yonsei Cancer Center, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

* Corresponding author:
Jeonghoon Ahn, Ehawyeodae-gil 52, Ewha Womans University, Seoul, 03760, Republic of Korea. ahnjeonghoon@ewha.ac.kr
Daeyoung Zang, 22, Gwanpyeong-ro 170beon-gil, Dongan-gu, Anyang-si, Gyeonggi-do 14068, Republic of Korea. fhdzang@hallym.or.kr

Abstract

Purpose This study aimed to develop and validate a mapping algorithm to derive EQ-5D-5L utility values from EORTC QLQ-C30 responses in Korean patients with gastric cancer.

Methods Secondary data from a multicenter cohort of 399 patients across 24 hospitals in Korea were analyzed, with utility calculated using the Korean EQ-5D-5L valuation set. Five regression approaches (OLS, TPM, ALDVMM, CLAD, and Beta regression) were applied under two specifications: Model 1, without control variables, and Model 2, with age and sex as control variables. Models were evaluated using RMSE and MAE, with 5-fold cross-validation and time point-specific external validation (6, 12, and 24 months).

Results Among the 399 patients (mean age, 61.42 ± 9.36 years; male, 71.9%), the mean baseline EQ-5D-5L utility was 0.883 ± 0.115, with a ceiling effect observed in 39.35% of patients. TPM showed the best predictive performance among the five regression approaches in both Model 1 (R^2=0.546, RMSE=0.077, MAE=0.060) and Model 2 (R^2=0.549, RMSE=0.077, MAE=0.060), consistent with 5-fold cross-validation results. In time point-specific external validation, however, the model with the lowest MSE varied by time point.

Conclusions This study offers a validated mapping algorithm for converting EORTC QLQ-C30 scores into EQ-5D-5L utility values in Korean gastric cancer patients, supporting its application in clinical research and healthcare policymaking.

Keywords EQ-5D-5L · EORTC QLQ-C30 · Direct mapping · Gastric cancer

Plain English summary

Health-related quality of life (HRQoL) of gastric cancer patients becomes important when deciding which treatments and support services should be covered, because these decisions weigh benefits to patients against costs. Many gastric cancer studies in Korea collected a cancer-specific quality of life questionnaire (the EORTC QLQ-C30), which can not produce a utility score used for economic evaluation and health policy decisions. Therefore, the way to convert QLQ-C30 scores into EQ-5D-5L utility values for Korean gastric patients is necessary. Our research team built and tested statistical formulas that translate QLQ-C30 responses into EQ-5D-5L utility scores,

using data from 399 gastric cancer patients treated at 24 hospitals across Korea. Five different statistical approaches were compared with and without adjustment for the patient's age and sex. A two-part model gave the most accurate predictions overall, explaining roughly half of the variation in patients' actual utility scores with an average error of about 0.06 on a scale where most patients scored near 0.88. It means researchers can now reuse QLQ-C30 data from Korean gastric patients to estimate utility scores for economic evaluation.

Abbreviations

AIC: Akaike Information Criterion

ALDVMM: Adjusted Limited Dependent Variable Mixture Model

BIC: Bayesian Information Criterion

CLAD: Censored Least Absolute Deviation

EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30

EORTC QLQ-STO22: European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Stomach 22

EQ-5D-5L: EuroQol 5-Dimension 5-Level

EQ-VAS: EuroQol Visual Analogue Scale

FACT-G: Functional Assessment of Cancer Therapy-General

FACT-Ga: Functional Assessment of Cancer Therapy-Gastric

HRQoL: Health-Related Quality of Life

MAE: Mean Absolute Error

MSE: Mean Squared Error

OLS: Ordinary Least Squares

QALY: Quality-Adjusted Life Year

QLQ-BR23: Quality of Life Questionnaire-Breast Cancer 23

QLQ-LC13: Quality of Life Questionnaire-Lung Cancer 13

RMSE: Root Mean Squared Error

TPM: Two-Part Model

Introduction

Gastric cancer remains one of the leading causes of cancer-related mortality worldwide, accounting for approximately 650,000 deaths annually [1]. In Korea, gastric cancer is major leading causes of cancer death, with an estimated prevalence of 350,000 individuals as of 2023 [2]. As survival outcomes of gastric cancer have improved by novel drugs and therapeutic advances in multidisciplinary cancer care [3]. Evaluation of cancer treatment has increasingly incorporated health-related quality of life (HRQoL) alongside survival. Particularly for gastric cancer patients, who often experience prolonged recovery and long-term treatment sequelae, evaluation of HRQoL of the patients became important indicator for efficient allocation of healthcare resources [4].

HRQoL is commonly assessed using either generic or disease-specific instruments. The EuroQol 5-Dimension (EQ-5D), a generic preference-based measure, is widely used in cost-utility analysis (CUA) because value sets are available for converting health states into utility scores. Although the EQ-5D-3L has been used extensively for the utility scores, the EQ-5D-5L was developed with five response levels per dimension to reduce ceiling effects and improve descriptive richness and validity of three-level version [5]. The expanded system has been reported to improve sensitivity and discriminatory power by reducing measurement error. However, it may still be less responsive to cancer-specific symptoms and treatment-related toxicities.

To evaluate HRQoL of cancer patients, several instruments have been developed. For patients with gastric cancer specifically, disease-specific instruments such as the EORTC QLQ-STO22 and the Functional Assessment of Cancer Therapy-Gastric (FACT-Ga) module have also been developed to capture gastric-cancer-related symptoms, including dysphagia, reflux, and eating restrictions, that are not addressed by generic cancer questionnaires [6]. The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30) and Functional Assessment of Cancer Therapy - General (FACT-G) assess HRQoL across cancer types in general. The EORTC QLQ-C30 has been translated into more than 120 languages and has demonstrated strong content validity and clinical relevance across a wide range of cancer types and treatment settings, including gastric cancer [7]. The EORTC QLQ-C30 consists of five functional scales, three symptom scales, a global health status/QoL scale, and six single items. A core questionnaire is supplemented by cancer-site-specific modules such as QLQ-BR23 for breast cancer and the QLQ-LC13 for lung cancer, capturing symptoms and treatment-related side effects not covered by the core. Because the EORTC QLQ-C30 isn't a preference-based instrument, its domain scores are difficult to provide societal preferences and can't be applied to derive quality-adjusted life years (QALYs) [8].

Mapping algorithms address this limitation by converting disease-specific scores into preference-based utility values, allowing HRQoL data collected with different instruments to be expressed on a common metric. Beyond enabling QALY estimation, this also facilitates comparison of findings across studies that use different quality-of-life instruments and supports pooling of data in meta-analyses [9, 10]. Such algorithms have been reported for breast cancer [11], head and neck cancer [10], and colorectal cancer [8], reflecting growing recognition that mapping algorithms should be developed for specific cancer types given differences in symptom profiles and score distributions. In gastric cancer, however, the primer mapping study in this population used EQ-5D-3L [12], and a Korean study mapping the QLQ-C30 to utility values included a mixed cancer population without gastric cancer-specific analysis [13].

A further consideration is that utility values derived from EQ-5D-5L are sensitive to country-specific value sets, and preferences for health states vary across cultural and demographic contexts [14, 15]. As of 2022, only 25 countries had developed EQ-5D-5L value sets using varying estimation methods, limiting cross-national comparability [16, 17]. Korea has established its own value set using a national preference-based model [18]. However, most existing mapping studies have relied on European or multinational cohort data [19], which do not adequately reflect the clinical environment and patient characteristics of the Korean population.

Therefore, this study aimed to develop and validate a mapping algorithm to derive EQ-5D-5L utility values from EORTC QLQ-C30 responses in Korean patients with gastric cancer.

Methods

Sample and data collection

This study used secondary data from a randomized, multicenter phase III clinical trial evaluating optimal duration of adjuvant oxaliplatin-based chemotherapy in patients with stage II–III gastric cancer, conducted across 24 hospitals under the Korean Cancer Study Group (KCSG) (trial registration: NCT04787354). HRQoL was assessed longitudinally at baseline, 6, 12, and 24 months using the EQ-5D-5L, EQ-VAS, and EORTC QLQ-C30. Data used in this analysis was collected between September 2021 and July 2024. Because mapping algorithms were developed using baseline data, this analysis had a cross-sectional design nested within the longitudinal cohort. Of the 399 patients assessed at baseline, 270 (67.7%) were followed up to 6 months, 168 (42.1%) to 12 months, and 70 (17.5%) to 24 months.

The survey and demographic data were derived from the clinical trial. The inclusion criteria of clinical trial were as follows: (1) histologically or cytologically confirmed adenocarcinoma of the stomach or gastroesophageal junction; (2) pathological stage II or III disease according to the American Joint Committee on Cancer (AJCC), 8th edition; (3) prior curative D2 gastrectomy with R0 resection; (4) age ≥19 years; (5) Eastern Cooperative Oncology Group (ECOG) performance status of 0–2; and (6) adequate bone marrow, renal, and hepatic function. The exclusion criteria included: (1) evidence of metastatic disease; (2) prior chemotherapy or radiotherapy for gastric cancer; (3) unresolved postoperative complications; (4) a history of another malignancy within the preceding 3 years; (5) clinically significant cardiac disease; (6) active viral infection; and (7) other clinically significant, uncontrolled comorbid conditions that could compromise study participation, per investigator judgment. Full eligibility criteria are provided in Supplementary Table S1.

A total of 399 patients met these criteria and were included in the analytic sample. This study used data collected as part of a randomized, multicenter phase III clinical trial (KCSG ST21-05; ClinicalTrials.gov identifier: NCT04787354), for which the study protocol—including HRQoL assessment using the EQ-5D-5L, EQ-VAS, and EORTC QLQ-C30—was approved by the Institutional Review Board of Hallym University Medical Center (approval no. NCT04787354), with written informed consent obtained from all participants. This study was exempted from full review by the Institutional Review Board of Ewha Womans University (approval no. ewha-202503-0036-01), as it involved minimal risk and no collection of identifiable information.

EORTC QLQ-C30

The independent variable was derived from the EORTC QLQ-C30, a 30-item cancer-specific instrument comprising five functional scales (physical, role, emotional, cognitive, social), three symptom scales (fatigue, nausea/vomiting, pain), a global health status/QoL scale, and six single items (dyspnea, insomnia, appetite loss, constipation, diarrhea, financial difficulties). Items were rated on 4-point Likert scales, except the global health status/QoL scale, which used a 7-point scale; raw scores were linearly transformed to a 0–100 scale following the EORTC QLQ-C30 Scoring Manual (3rd edition) [20].

EQ-5D-5L

The outcome variable was HRQoL utility measured with the EQ-5D-5L, a generic preference-based instrument comprising five dimensions—mobility, self-care, usual activities, pain/discomfort, and anxiety/depression—each rated on five severity levels. Utility values were derived using the Korean EQ-5D-5L value set based on the N4 model of Kim et al. [18], which applies an additional decrement when any dimension is rated at severity level 4 or above:

$$\begin{aligned}\text{Utility} = 1 - (&0.096 + 0.046 \times \text{M2} + 0.058 \times \text{M3} + 0.133 \times \text{M4} + 0.251 \times \text{M5} + 0.032 \times \text{S2} + 0.050 \times \text{S3} \\ &+ 0.078 \times \text{S4} + 0.122 \times \text{S5} + 0.021 \times \text{U2} + 0.051 \times \text{U3} + 0.100 \times \text{U4} + 0.175 \times \text{U5} \\ &+ 0.042 \times \text{P2} + 0.053 \times \text{P3} + 0.166 \times \text{P4} + 0.207 \times \text{P5} + 0.033 \times \text{A2} + 0.046 \times \text{A3} \\ &+ 0.102 \times \text{A4} + 0.137 \times \text{A5} + 0.078 \times \text{N4}\end{aligned}$$

where M, S, U, P, and A denote the mobility, self-care, usual activities, pain/discomfort, and anxiety/depression dimensions (subscripts 2–5 indicate severity level relative to level 1), and N4 = 1 if any dimension is rated at level 4 or 5.

Data analysis

Descriptive statistics and correlation analysis

Descriptive statistics for EQ-5D-5L utility and QLQ-C30 scores were calculated at baseline and at 6, 12, and 24 months to characterize longitudinal HRQoL trends. Pearson correlation was used to examine associations between EQ-5D-5L utility and the QLQ-C30 summary score and global health status/QoL scale; Spearman correlation was used to explore associations between individual QLQ-C30 subscales and EQ-5D-5L dimensions, and among QLQ-C30 subscales themselves.

Model selection strategy

Because no established mapping methodology exists for gastric cancer, model selection was guided by ISPOR mapping recommendations [21] together with prior mapping studies in other cancer types. Prior to model selection, the distributional characteristics of EQ-5D-5L utility were examined: (1) ceiling effects, defined as the proportion of respondents reporting no problems across all EQ-5D-5L dimensions (utility = 1.0); (2) multicollinearity among independent variables, addressed by combining or excluding highly correlated variables; and (3) multimodality of the utility distribution.

Regression model development

OLS regression was applied first as a baseline model, consistent with its common use in mapping studies. Given the anticipated ceiling effect and the bounded, non-normal distribution of EQ-5D-5L utility, four additional models were applied: a two-part model (TPM) and an adjusted limited dependent variable mixture model (ALDVMM) to address ceiling effects and multimodality, and censored least absolute deviations (CLAD) regression and beta regression to accommodate non-normal, heteroscedastic residuals and the [0,1]-bounded utility distribution. Rationale for model selection according to distributional characteristics is summarized in Supplementary Table S2.

The general form of the mapping model is expressed as:

$$\text{Model 1 (QLQ} - \text{C30 only): EQ} - \text{5D} - \text{5L} = f(QLQ - C30)$$

$$\text{Model 2 (QLQ} - \text{C30 plus age and sex): EQ} - \text{5D} - \text{5L} = f(QLQ - C30,\ Age,\ Sex)$$

where *f* denotes each of the five regression approaches (OLS, TPM, ALDVMM, CLAD, beta regression) applied in this study. Age and sex were included as covariates given their established association with HRQoL and their inclusion in prior QLQ-C30–to–EQ-5D mapping studies. The detailed description for each model specification is provided in Supplementary Material M1.

Model evaluation and validation

Model performance was evaluated using mean absolute error (MAE), root mean square error (RMSE), R^2, Akaike information criterion (AIC), and Bayesian information criterion (BIC). Internal validity was assessed via 5-fold cross-validation, with the sample randomly partitioned into five subsets, each serving once as the validation set. The best-performing model based on cross-validation was selected as the final mapping algorithm, and its external validity was subsequently evaluated using time-point–specific validation at 6, 12, and 24 months. To conduct subgroup analysis, two variables, surgery type and AJCC stage, which indicate disease severity were compared based on the magnitude of the between-group difference in baseline EQ-5D-5L utility. As surgery type showed larger difference, it was selected for subgroup analysis.

All statistical analyses were performed using R Software (version 4.3.3; http://www.R-project.org) [22] and Stata (version 18.5; Stata Corp LP, College Station, TX, USA) [23].

Results

Descriptive characteristics

Of the 399 patients included in the analytic sample, as shown in Table 1, 287 (71.93%) were male and 112 (28.17%) were female, with a mean age of 61.42 ± 9.36 years. The largest age group was patients in their 60s (43.11%), followed by those in their 50s (27.32%), 70s (18.05%), 40s (7.77%), 30s (2.01%), and 80 years or older (1.75%); no patients in their 20s were included. Most patients underwent subtotal gastrectomy (70.18%; total gastrectomy, 29.82%), and the majority were classified as AJCC stage II or IIIA (78.45%; stage IIIB or IIIC, 21.55%).

Table 1 Descriptive characteristics

	N (%)
Gender	
Male	287(71.93)
Female	112(28.17)
Age	
20-29	0(0.00)
30-39	8(2.01)
40-49	31(7.77)
50-59	109(27.32)
60-69	172(43.11)
70-79	72(18.05)
≥ 80	7(1.75)
Surgery	
Subtotal gastrectomy	280(70.18)
Total gastrectomy	119(29.82)
AJCC stage	
II or IIIA	313(78.45)
IIIB or IIIC	86(21.55)

Baseline HRQoL and distributional characteristics

Table 2 presents the distribution of EQ-5D-5L, EQ-VAS, and QLQ-C30 scores at baseline. The mean EQ-5D-5L utility was 0.883 ± 0.115 (median 0.862), and the mean EQ-VAS score was 0.739 ± 0.167. The mean QLQ-C30 summary score was 85.58 ± 11.69, and the global health status/QoL score was 65.06 ± 20.95. Among functional scales, cognitive functioning was highest (90.14 ± 13.64), followed by emotional (86.68 ± 17.20), physical (85.70 ± 14.39), role (85.59 ± 19.39), and social functioning (83.63 ± 20.98). Among symptom scales, nausea/vomiting was lowest (7.60 ± 14.54), while fatigue was highest (23.98 ± 19.48); pain, dyspnea, constipation, financial difficulties, insomnia, diarrhea, and appetite loss fell between these values.

A ceiling effect was observed, with 157 patients (39.35%) reporting no problems across all EQ-5D-5L dimensions (utility = 1.0). The utility distribution was also multimodal, with a density peak at utility = 1.0 and at least one additional mode; Hartigan's dip test [24] rejected the null hypothesis of unimodality ($p < 0.001$) (Figure 1**).**

Table 2 Distribution of EQ-5D-5L, EQ-VAS and EORTC QLQ-C30 scores at baseline

		Mean(SD)	Median(Q1-Q3)
EQ-5D-5L		0.883 (0.115)	0.862(0.816-1.000)
EQ-VAS		73.932 (16.659)	80.0(60.0, 88.5)
EORTC QLQ-C30	Summary score	85.58 (11.69)	87.78(78.46, 94.64)
	Global health status/QoL	65.06 (20.95)	66.67(50.00, 83.33)
	Physical functioning	85.70 (14.39)	86.67(80.00, 93.33)

Role functioning	85.59 (19.39)	100.00(66.67, 100.00)
Emotional functioning	86.68 (17.20)	91.67(75.00, 100.00)
Cognitive functioning	90.14 (13.64)	100.00(83.33, 100.00)
Social functioning	83.63 (20.98)	100.00(66.67, 100.00)
Fatigue	23.98 (19.48)	22.22(11.11, 33.33)
Nausea and vomiting	7.60 (14.54)	0.00(0.00, 16.67)
Pain	10.07 (16.06)	0.00(0.00, 16.67)
Dyspnea	10.28 (19.13)	0.00(0.00, 33.33)
Insomnia	15.54 (23.94)	0.00(0.00, 33.33)
Appetite loss	21.89 (27.04)	0.00(0.00, 33.33)
Constipation	12.53 (20.84)	0.00(0.00, 33.33)
Diarrhea	17.29 (23.01)	0.00(0.00, 33.33)
Financial difficulties	15.46 (21.74)	0.00(0.00, 33.33)

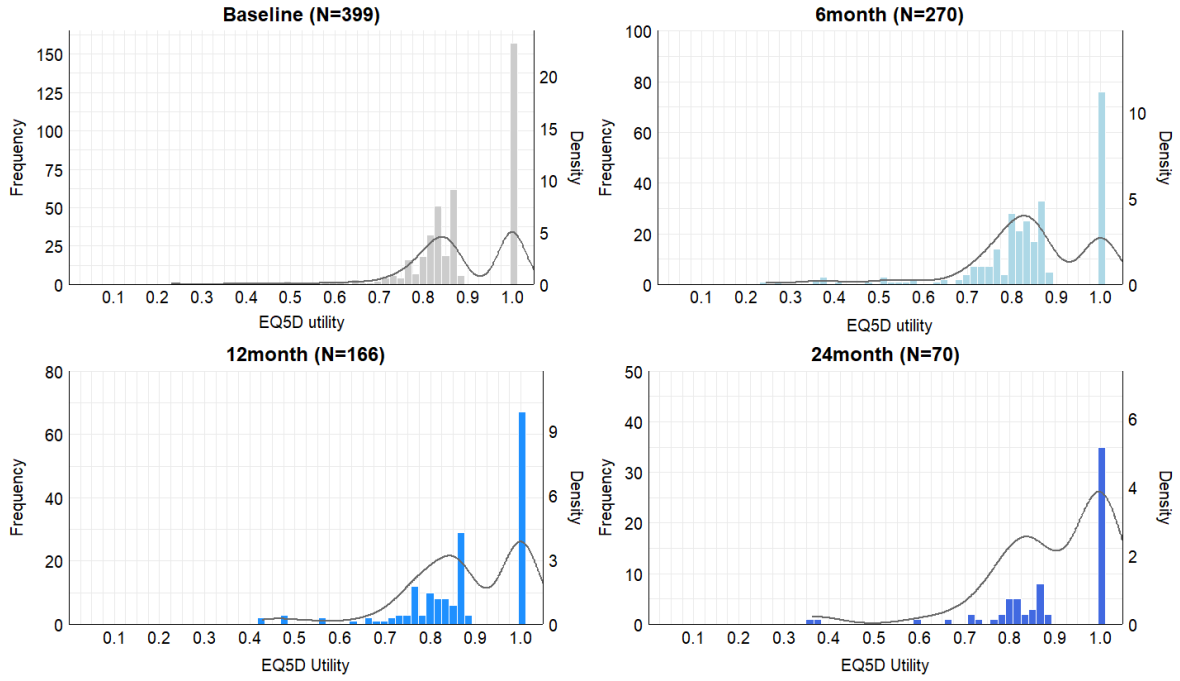


Fig. 1 Distribution of EQ-5D-5L utility values

Correlation between HRQoL measures

EQ-5D-5L utility was significantly and positively correlated with all other HRQoL measures ($p < 0.001$), most strongly with the QLQ-C30 summary score ($r = 0.630$), followed by the global health status/QoL scale ($r = 0.492$) and EQ-VAS ($r = 0.471$).

At the dimension level, all five EQ-5D-5L dimensions were significantly and negatively correlated with global health status/QoL, most strongly for usual activities ($r = -0.433$, $p < 0.001$). Anxiety/depression showed the strongest association with any single QLQ-C30 functional scale, correlating most strongly with emotional functioning ($r = -0.702$, $p < 0.001$); mobility correlated most strongly with physical functioning ($r = -0.392$, $p < 0.001$), usual activities with role functioning ($r = -0.513$, $p < 0.001$), and pain/discomfort with the pain scale ($r = 0.539$, $p < 0.001$). Among symptom scales, fatigue showed the strongest associations with mobility, usual activities, and anxiety/depression ($r = 0.288–0.454$, $p < 0.001$), as detailed in Supplementary Table S3.

Multicollinearity among QLQ-C30 subscales was examined via Spearman correlation and variance inflation

factors (VIF). Functional scales were most strongly intercorrelated between emotional and cognitive functioning ($r = 0.684$, $p < 0.001$), and symptom scales between fatigue and appetite loss ($r = 0.565$, $p < 0.001$). All VIF values were below 3, under the conventional threshold of 4 [25], so no independent variables were combined or excluded due to multicollinearity (Supplementary Table S4.).

Mapping EORTC QLQ-C30 to EQ-5D-5L

Table 3 summarizes the predictive performance of the five regression approaches—OLS, TPM, ALDVMM, CLAD, and beta regression—under both Model 1 (QLQ-C30 only) and Model 2 (QLQ-C30 plus age and sex). The performance of models was described by R^2, adjusted R^2, RMSE, MAE, AIC, and BIC. For each specification, the best-performing approach was selected based on the lowest RMSE and MAE. TPM showed the best overall performance under both Model 1 ($R^2 = 0.546$, adjusted $R^2 = 0.509$; RMSE = 0.077; MAE = 0.060) and Model 2 ($R^2 = 0.549$, adjusted $R^2 = 0.507$; RMSE = 0.077; MAE = 0.060), consistently outperforming OLS, ALDVMM, CLAD, and beta regression in that order.

This ranking was confirmed by 5-fold cross-validation (Table 4). TPM again showed the best performance under Model 1 ($R^2 = 0.481$, adjusted $R^2 = 0.461$; RMSE = 0.083; MAE = 0.064) and Model 2 (RMSE = 0.084; MAE = 0.065), narrowly ahead of OLS, ALDVMM, and beta regression, with CLAD showing the weakest fit under both specifications (Model 1: $R^2 = 0.428$; Model 2: $R^2 = 0.394$). Based on its consistent superiority across the full-sample and cross-validated comparisons, TPM was selected as the final mapping algorithm under both Model 1 and Model 2. Figure 2 shows the agreement between predicted and observed EQ-5D-5L utility values for the final TPM model under both specifications.

Table 3 Performance of regression methods for mapping the EORTC QLQ-C30 to the EQ-5D-5L utility scores

	R^2(adjusted R^2)	RMSE	MAE	AIC	BIC
Model 1(QLQ-C30 only)					
OLS	0.530(0.512)	0.079	0.063	-863.769	-795.956
TPM	0.546(0.509)	0.077	0.060	-851.115	-731.446
ALDVMM(k=1)	0.518(0.498)	0.080	0.064	0.064	-343.103
ALDVMM(k=2)	0.504(0.484)	0.081	0.063	-535.689	-396.076
CLAD	0.502(0.482)	0.081	0.061	-	-
Beta	0.370(0.345)	0.084	0.071	-1399.219	-1331.407
Model 2(QLQ-C30 plus age and sex)					
OLS	0.533 (0.512)	0.078	0.062	-861.830	-786.039
TPM	0.549 (0.507)	0.077	0.060	-846.162	-710.538
ALDVMM(k=1)	0.520(0.500)	0.079	0.064	-408.581	-332.790
ALDVMM(k=2)	0.497(0.473)	0.081	0.064	-550.692	-395.123
CLAD	0.498(0.474)	0.081	0.060	-	-
Beta	0.374 (0.349)	0.083	0.070	-1400.275	-1324.484

Table 4 Performance of fivefold cross-validation of regression models

	R^2(adjusted R^2)	RMSE	MAE	AIC	BIC
Model 1(QLQ-C30 only)					
OLS	0.468 (0.448)	0.084	0.066	-688.740	-624.721
TPM	0.481 (0.461)	0.083	0.064	-827.880	-768.045
ALDVMM(k=1)	0.459 (0.438)	0.084	0.068	-328.939	-264.920
ALDVMM(k=2)	0.462 (0.438)	0.088	0.067	-292.939	-161.135
CLAD	0.428 (0.405)	0.087	0.066	-	-
Beta	0.440 (0.418)	0.086	0.073	-797.477	-737.643

Model 2(QLQ-C30 plus age and sex)					
OLS	0.458 (0.434)	0.084	0.066	-687.513	-615.963
TPM	0.464 (0.440)	0.084	0.065	-810.984	-743.171
ALDVMM(k=1)	0.469 (0.446)	0.084	0.067	-323.672	-248.356
ALDVMM(k=2)	0.464 (0.440)	0.085	0.066	-281.672	-127.273
CLAD	0.394 (0.367)	0.089	0.068	-	-
Beta	0.457 (0.433)	0.084	0.072	-806.220	-738.407

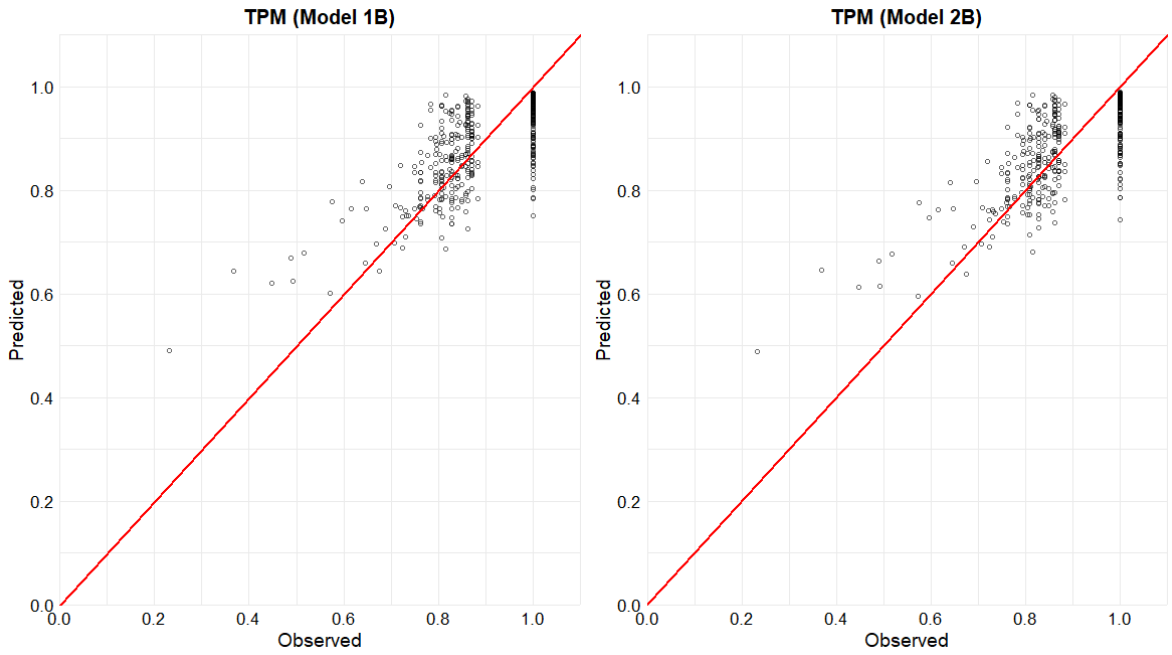


Fig. 2 Predicted EQ-5D-5L utility values versus observed values for the TPM model under Model 1 and Model 2

Table 5 presents the full regression coefficients for the final TPM model under Model 1 and Model 2. Under Model 1, physical functioning, role functioning, emotional functioning, social functioning, nausea/vomiting, and appetite loss were significant predictors of EQ-5D-5L utility, whereas global health status/QoL was not (β = 0.0004, $p > 0.05$). Among the functional scales, physical functioning showed the largest coefficient (β = 0.0025, $p < 0.001$), and among the symptom scales, nausea/vomiting ($\beta = -0.0006$, $p = 0.028$) and appetite loss($\beta = 0.0006$, $p = 0.001$) showed the largest coefficients. The same set of QLQ-C30 predictors remained significant under Model 2, with physical functioning again showing the largest coefficient (β = 0.0027, $p < 0.001$) and nausea/vomiting the largest among symptom scales ($\beta = -0.0006$, $p = 0.035$). Neither sex nor age reached statistical significance in Model 2; female patients had, on average, a 0.0140 higher predicted utility than male patients, and predicted utility increased by 0.0004 per additional year of age.

Table 5 Coefficient estimates of TPM

	TPM(Model 1)		TPM(Model 2)	
	Coefficient	*p*	Coefficient	*p*
constant	0.4026	<0.001	0.3506	<0.001
PF	0.0025	<0.001	0.0027	<0.001
RF	0.0006	0.025	0.0006	0.031
EF	0.0008	0.013	0.0008	0.008
CF	-0.0003		-0.0002	

SF	0.0008	0.001	0.0008	0.002
FA	0.0004		0.0004	
NV	-0.0006	0.028	-0.0006	0.035
PA	-0.0003		-0.0002	
DY	0.0003		0.0004	
SL	0.0003		0.0003	
AP	0.0006	0.001	0.0005	0.004
CO	0.0001		0.0000	
DI	0.0000		0.0000	
FI	-0.0003		-0.0003	
QL	0.0004		0.0004	
Sex	-	-	0.0140	
Age	-	-	0.0004	

Final mapping algorithm

TPM is a two-part model that combines (1) the estimated probability of reporting perfect health (utility = 1), modeled via logistic regression, and (2) the estimated conditional mean utility among patients not in perfect health (utility < 1), modeled via OLS regression. The final predicted value is calculated as a probability-weighted average of these two components:

$$\hat{Y} = \hat{P}(Y = 1) \times 1 + (1 - \hat{P}(Y = 1)) \times \hat{Y}_{Y<1}$$

Model 1 (QLQ-C30 only):

$$\begin{aligned}\hat{P}(Y = 1) = 1/(1 + exp \\ - (-9.0833 + 0.0304 \times PF + 0.0133 \times RF + 0.0541 \times EF - 0.0122 \times CF \\ + 0.0011 \times SF - 0.0164 \times FA + 0.0133 \times NV - 0.0321 \times PA - 0.0056 \times DY \\ - 0.0112 \times SL - 0.0032 \times AP - 0.0016 \times CO - 0.0057 \times DI + 0.0034 \times FI \\ + 0.0252 \times QL))\end{aligned}$$

$$\begin{aligned}\hat{Y}_{Y<1} = 0.4026 + 0.0025 \times PF + 0.0006 \times RF + 0.0008 \times EF - 0.0003 \times CF + 0.0008 \times SF \\ + 0.0004 \times FA - 0.0006 \times NV - 0.0003 \times PA + 0.0003 \times DY + 0.0003 \times SL \\ + 0.0006 \times AP + 0.0001 \times CO + 0.0000 \times DI - 0.0003 \times FI + 0.0004 \times QL\end{aligned}$$

Model 2 (QLQ-C30 plus age and sex):

$$\begin{aligned}\hat{P}(Y = 1) = 1/(1 + exp \\ - (-9.0833 + 0.0304 \times PF + 0.0133 \times RF + 0.0541 \times EF - 0.0122 \times CF \\ + 0.0011 \times SF - 0.0164 \times FA + 0.0133 \times NV - 0.0321 \times PA - 0.0056 \times DY \\ - 0.0112 \times SL - 0.0032 \times AP - 0.0016 \times CO - 0.0057 \times DI + 0.0034 \times FI \\ + 0.0252 \times QL - 0.3744 \times Female + 0.0054 \times Age))\end{aligned}$$

$$\begin{aligned}\hat{Y}_{Y<1} = 0.4026 + 0.0025 \times PF + 0.0006 \times RF + 0.0008 \times EF - 0.0003 \times CF + 0.0008 \times SF \\ + 0.0004 \times FA - 0.0006 \times NV - 0.0003 \times PA + 0.0003 \times DY + 0.0003 \times SL \\ + 0.0006 \times AP + 0.0001 \times CO + 0.0000 \times DI - 0.0003 \times FI + 0.0004 \times QL \\ + 0.0140 \times Female + 0.0004 \times Age\end{aligned}$$

where Female is coded 1 for female patients and 0 for male patients.

External validation

Model predictions were evaluated at 6, 12, and 24 months using mean squared error (MSE, calculated as the sum

of squared bias and variance between predicted and observed EQ-5D-5L utility values), with lower MSE indicating better predictive accuracy. All five regression approaches were compared at each time point under both Model 1 and Model 2 (Table 6). Under Model 1, the model with the lowest MSE varied by time point: OLS performed best at 6 months (MSE = 0.0083) and 24 months (MSE = 0.0077), while TPM performed best at 12 months (MSE = 0.0078). A similar pattern was observed under Model 2, where OLS performed best at 6 months (MSE = 0.0082) and TPM performed best at both 12 months (MSE = 0.0077) and 24 months (MSE = 0.0076).

As TPM showed the most consistent overall performance across the full-sample and cross-validated comparisons, its predictive accuracy over time was further examined separately using RMSE and MAE, the metrics used throughout the rest of the analysis (Table 7). Predictive accuracy for TPM declined modestly relative to the baseline model at all three follow-up time points. Under Model 1, RMSE was lowest at 12 months (0.088) and highest at 6 months (0.093), while MAE was lowest at 24 months (0.061), indicating that the time point with the lowest error varied by metric. Under Model 2, both RMSE and MAE were lowest at 24 months (0.087 and 0.060, respectively), showing more consistent improvement over time than Model 1.

Table 6 External validation of models at 6, 12, and 24 months

Time point	Model type	Method	Bias	Variance	MSE
6 months	Model 1	OLS	0.0188	0.0112	0.0083
		TPM	0.0180	0.0113	0.0087
		ALDVMM(k=1)	0.0206	0.0114	0.0086
		ALDVMM(k=2)	0.0249	0.0084	0.0091
		CLAD	0.0312	0.012	0.009
		Beta	0.0231	0.0168	0.0102
	Model 2	OLS	0.0179	0.0114	0.0082
		TPM	0.0168	0.0117	0.0086
		ALDVMM(k=1)	0.0194	0.0117	0.0086
		ALDVMM(k=2)	0.0260	0.0771	0.0095
		CLAD	0.0322	0.0126	0.0094
		Beta	0.0220	0.0173	0.0103
12 months	Model 1	OLS	0.0199	0.0066	0.0080
		TPM	0.0202	0.0071	0.0078
		ALDVMM(k=1)	0.0234	0.0067	0.0083
		ALDVMM(k=2)	0.0229	0.0055	0.0083
		CLAD	0.0353	0.0070	0.0093
		Beta	0.0250	0.0064	0.0092
	Model 2	OLS	0.0192	0.0066	0.0080
		TPM	0.0195	0.0072	0.0077
		ALDVMM(k=1)	0.0228	0.0068	0.0084
		ALDVMM(k=2)	0.0222	0.0049	0.0089
		CLAD	0.0350	0.0070	0.0095
		Beta	0.0250	0.0065	0.0093
24 months	Model 1	OLS	0.0162	0.0075	0.0077
		TPM	0.0166	0.0081	0.0079
		ALDVMM(k=1)	0.0649	0.0075	0.0090
		ALDVMM(k=2)	0.0656	0.0059	0.0091
		CLAD	0.0328	0.0085	0.0085
		Beta	0.0148	0.0070	0.0084
	Model 2	OLS	0.0145	0.0075	0.0077
		TPM	0.0147	0.0083	0.0076
		ALDVMM(k=1)	0.0637	0.0076	0.0087

	ALDVMM(k=2)	0.0621	0.0057	0.0087
	CLAD	0.0312	0.0084	0.0083
	Beta	0.0141	0.0067	0.0082

Table 7 Predictive accuracy of the TPM model at 6, 12, and 24 months

Time point	Model type	RMSE	MAE
6 months	Model 1	0.093	0.070
	Model 2	0.093	0.071
12 months	Model 1	0.088	0.065
	Model 2	0.088	0.064
24 months	Model 1	0.089	0.061
	Model 2	0.087	0.060

Subgroup analysis

Baseline EQ-5D-5L utility differed modestly, though not statistically significantly, between candidate severity indicators: by surgery type (subtotal gastrectomy, 0.888; total gastrectomy, 0.872) and by AJCC stage (stage II/IIIA, 0.883; stage IIIB/IIIC, 0.886). As the difference was larger for surgery type, subgroup analysis was conducted using surgery type. Both TPM specifications performed better among patients who underwent subtotal gastrectomy (n = 280; Model 1: RMSE = 0.076, MAE = 0.059; Model 2: RMSE = 0.076, MAE = 0.058) than among those who underwent total gastrectomy (n = 119; Model 1: RMSE = 0.080, MAE = 0.063; Model 2: RMSE = 0.079, MAE = 0.063), as shown in Table 8. Predictive accuracy in the subtotal gastrectomy subgroup was closely comparable to that of the full sample (RMSE = 0.077, MAE = 0.060).

Table 8 Predictive performance of the TPM model by surgery type

Surgery type	Model type	RMSE	MAE
Subtotal gastrectomy	Model 1	0.076	0.059
	Model 2	0.076	0.058
Total gastrectomy	Model 1	0.080	0.063
	Model 2	0.079	0.063

Discussion

This study suggests an algorithm mapping EORTC QLQ-C30 to EQ-5D-5L specifically for a gastric cancer patients in Korean. Prior studies in this population relied on EQ-5D-3L [12] or on mixed-cancer cohorts without a gastric-specific analysis [13], neither of which reflects current valuation practice or the clinical profile of Korean gastric cancer patients. By using data from a nationwide, 24-hospital randomized trial and providing an explicit algorithm for both an unadjusted and covariate-adjusted specification, this study offers a directly applicable tool for deriving EQ-5D-5L utilities in future Korean gastric cancer research.

The mean baseline utility observed here (0.883) was higher than utility values reported for more heterogeneous or higher-morbidity gastric cancer samples elsewhere. The utility was reported as the 0.550 among a mixed-severity Greek cohort [12], but is broadly consistent with values reported for post-surgical, early-stage gastric cancer patients in Korea (0.77–0.85) [26]. This might be reflected the trial's eligibility criteria, which restricted enrollment to patients who had undergone curative D2 gastrectomy with R0 resection for stage II–III disease, and is consistent with the observed ceiling effect (39.35% reporting full health).

The ceiling effect and multimodal utility distribution could be reason why TPM outperformed the other four regression approaches. By separately modeling the probability of reporting perfect health and the conditional distribution of utility among patients with any impairment, TPM directly accommodates this bimodal structure, whereas single-equation approaches such as OLS and beta regression assume a continuous, unimodal distribution that does not fit the data well. Beta regression's comparatively weak performance (R^2 = 0.370–0.374) suggests that, without an explicit mixture or two-part structure, its flexibility in modeling the [0,1] range was insufficient to capture the concentration of values at the ceiling.

Neither age nor sex was a statistically significant predictor in Model 2, in contrast to mapping studies in other cancer types that have reported sex-based HRQoL disparities [27]. This may partly reflect the relative clinical homogeneity of this trial population, all of whom underwent curative-intent surgery with good performance status (ECOG 0–2), which may narrow demographic differences in HRQoL that are more apparent in samples with broader disease severity. External validation also revealed a nuance not always reported in mapping studies: although TPM was the most consistent performer overall, OLS occasionally achieved lower error at specific follow-up time points, indicating that baseline-derived model rankings may not be fully stable over time.

This study has several strengths. First, it used prospectively collected, multicenter randomized trial data spanning 24 hospitals. The data from prospectively collected participants can provide minimized recall bias and the multicenter design can more precisely represent Korean patients with gastric cancer. Second, a comparison of five regression approaches under two model allowed the specification allowed that the TPM had best performance with transparent evidence. Finally, both 5-fold cross-validation and temporal external validation were performed. Cross-validation assesses whether a model's fit reflects a genuine relationship between predictors and the outcome, rather than the specific composition of the estimation sample, while temporal external validation—applying the baseline-derived model to independent follow-up data collected at later time points—provides a more rigorous test of predictive stability under real-world conditions [28]. By combining both approaches, this study offers a more comprehensive assessment of the mapping algorithm's robustness than internal validation alone would allow.

This study also has several limitations. First, our study applied direct mapping rather than indirect mapping, predicting the response level for each EQ-5D-5L dimension and then applying the value set to the predicted response profile. Prior studies have shown that indirect mapping can achieve better predictive performance than direct mapping while offering greater flexibility across country-specific value sets [29]. However, direct mapping was used exclusively in this study, consistent with the more common approach in applied

pharmacoeconomic practice [30]. Future studies should examine the feasibility of indirect mapping in Korean gastric cancer populations and directly compare its performance with the direct mapping approach used here.

Second, the included study population was focused on patients with stage II or III gastric cancer. It would be difficult to apply the results of the study to several gastric cancer patients; patients with stage IV disease, those in the pre-surgical treatment phase, and those receiving non-surgical treatment. Consequently, the mapping algorithm developed here is most directly applicable to postoperative patients with mid-stage disease, and it is need to be cautious when applying it to patients with more advanced or unresectable gastric cancer.

Third, although temporal analyses were conducted to assess model stability, TPM did not consistently outperform the other regression approaches at every follow-up time point; OLS achieved lower prediction error at some time points. This suggests that TPM may not fully capture the specific distributional characteristics of utility at each stage of follow-up. When applying the algorithm developed in this study, time-point-specific characteristics of the target population should be considered, and future research could explore models that explicitly account for longitudinal changes in the predictor–outcome relationship.

Conclusions

This study provides a validated mapping algorithm for converting EORTC QLQ-C30 scores into EQ-5D-5L utility values in Korean patients with gastric cancer, with TPM offering the most consistent predictive performance. The algorithm developed here supports its application in future clinical research and health economic evaluation when direct EQ-5D-5L measurement is unavailable.

Statements and Declarations

Acknowledgements This article is based on the first author's master's thesis submitted to Ewha Womans University in partial fulfillment of the requirements for the degree of Master of Convergence Health, 2025.

Funding This research was supported by a grant of Patient-Centered Clinical Research Coordinating Center (PACEN) funded by the Ministry of Health & Welfare, Republic of Korea (grant number: RS-2020-KH095267).

Competing interests The authors declare no competing interests.

Author contributions All authors participated in the study design, data acquisition, and interpretation of results. Doyeon Jin conducted the data analysis and drafted the manuscript. Sangyoon Shin and Jeonghoon Ahn were involved in the critical revision of the manuscript for important intellectual content. Bum Jun Kim, Minkyu Jung and Dae Young Zang made substantial contributions to data collection.

Ethics approval and consent to participate This study was performed in line with the principles of the Declaration of Helsinki. This study used secondary data from a randomized, multicenter phase III clinical trial (KCSG ST21-05; ClinicalTrials.gov identifier: NCT04787354), for which ethical approval and written informed consent from all participants were obtained by the original trial investigators. This secondary analysis was exempted from full review by the Institutional Review Board of Ewha Womans University (approval no. ewha-202503-0036-01), as it involved minimal risk and no collection of identifiable information.

Consent for publication Not applicable. All results are reported as aggregated data.

Data availability The datasets are available from the corresponding author on reasonable request.

References

[1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3), 229–263. https://doi.org/10.3322/caac.21834

[2] Park, E. H., Jung, K. W., Park, N. J., Kang, M. J., Yun, E. H., Kim, H. J., Kim, J. E., Kong, H. J., Choi, K. S., & Yang, H. K. (2025). Cancer statistics in Korea: Incidence, mortality, survival, and prevalence in 2022. *Cancer Research and Treatment*, 57(2), 312–330. https://doi.org/10.4143/crt.2025.264

[3] Puhr, H. C., Reiter, T. J., Preusser, M., Prager, G. W., & Ilhan-Mutlu, A. (2023). Recent advances in the systemic treatment of localized gastroesophageal cancer. *Cancers*, 15(6), 1900. https://doi.org/10.3390/cancers15061900

[4] Tannock, I. F., Pe, M. L., Booth, C. M., Brundage, M., Cherny, N. I., Coens, C., Eisenhauer, E. A., Geissler, J., Giesinger, J. M., Gyawali, B., Oosting, S. F., Pond, G. R., Reijneveld, J. C., Soto-Perez-de-Celis, E., Tregear, M., & van der Graaf, W. T. A. (2025). Importance of responder criteria for reporting health-related quality-of-life data in clinical trials for advanced cancer. *The Lancet Oncology*, 26(9), e499–e507. https://doi.org/10.1016/S1470-2045(25)00288-8

[5] Jo, M. W., Ahn, J., Kim, S. H., Shin, S. J., Park, J. Y., Ock, M., Son, W. S., & Kim, S. O. (2014). *Estimation of Korean adults' utility for EQ-5D-5L health states* [Research report]. National Evidence-based Healthcare Collaborating Agency.

[6] Woo, A., Fu, T., Popovic, M., Chow, E., Cella, D., Wong, C. S., Lam, H., Pulenzas, N., Lechner, B., Vuong, S., Ganesh, V., & Bottomley, A. (2016). Comparison of the EORTC STO-22 and the FACT-Ga quality of life questionnaires for patients with gastric cancer. *Annals of Palliative Medicine*, 5(1), 13–21. https://doi.org/10.3978/j.issn.2224-5820.2016.01.02

[7] Cocks, K., Wells, J. R., Johnson, C., Schmidt, H., Koller, M., Oerlemans, S., Aaronson, N., Yost, K., McQuellon, R., Ganz, P. A., Basch, E., Coens, C., Musoro, J., Bottomley, A., & Fung, C. (2023). Content validity of the EORTC quality of life questionnaire QLQ-C30 for use in cancer. *European Journal of Cancer*, 178, 128–138. https://doi.org/10.1016/j.ejca.2022.10.026

[8] Ameri, H., Yousefi, M., Yaseri, M., Nahvijou, A., Arab, M., & Akbari Sari, A. (2019). Mapping EORTC-QLQ-C30 and QLQ-CR29 onto EQ-5D-5L in colorectal cancer patients. *Journal of Gastrointestinal Cancer*, 51(1), 196–203. https://doi.org/10.1007/s12029-019-00229-6

[9] Khan, I., Morris, S., Pashayan, N., Matata, B., Bashir, Z., & Maguirre, J. (2016). Comparing the mapping between EQ-5D-5L, EQ-5D-3L and the EORTC-QLQ-C30 in non-small cell lung cancer patients. *Health and Quality of Life Outcomes*, 14, 60. https://doi.org/10.1186/s12955-016-0455-1

[10] Noel, C. W., Stephens, R. F., Su, J. S., Xu, W., Krahn, M., Monteiro, E., Goldstein, D. P., Giuliani, M., Hansen, A. R., & de Almeida, J. R. (2020). Mapping the EORTC QLQ-C30 and QLQ-H&N35, onto EQ-5D-5L and HUI-3 indices in patients with head and neck cancer. *Head & Neck*, 42(9), 2277–2286. https://doi.org/10.1002/hed.26181

[11] Liu, T., Li, S., Wang, M., Sun, Q., & Chen, G. (2020). Mapping the Chinese version of the EORTC QLQ-BR53 onto the EQ-5D-5L and SF-6D utility scores. *The Patient*, 13(5), 537–555. https://doi.org/10.1007/s40271-020-00422-x

[12] Kontodimopoulos, N., Aletras, V. H., Paliouras, D., & Niakas, D. (2009). Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. *Value in Health*, 12(8), 1151–1157. https://doi.org/10.1111/j.1524-4733.2009.00569.x

[13] Kim, S. H., Jo, M. W., Kim, H. J., & Ahn, J. H. (2012). Mapping EORTC QLQ-C30 onto EQ-5D for the assessment of cancer patients. *Health and Quality of Life Outcomes*, 10, 151. https://doi.org/10.1186/1477-7525-10-151

[14] Kim, T. J., Oh, K. T., Joo, E. K., Lee, H. S., Kim, T. H., Jeon, J. B., Jung, S. S., Yoo, D. H., & Bae, S. C. (2002). Health-related quality of life in Korean patients with ankylosing spondylitis. *Journal of the Korean Rheumatism Association*, 9(4, Suppl.).

[15] Scalone, L., Cortesi, P. A., Ciampichini, R., Belisari, A., D'Angiolella, L. S., Cesana, G., & Mantovani, L. G. (2013). Italian population-based values of EQ-5D health states. *Value in Health*, 16(5), 814–822. https://doi.org/10.1016/j.jval.2013.04.008

[16] Devlin, N., Pickard, S., & Busschbach, J. (2022). The development of the EQ-5D-5L and its value sets. In N. Devlin et al. (Eds.), *Value Sets for EQ-5D-5L: A Compendium, Comparative Review & User Guide* (pp. 1–12). Springer.

[17] Janssen, M. F., Bonsel, G. J., & Luo, N. (2018). Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *PharmacoEconomics*, 36(6), 675–697. https://doi.org/10.1007/s40273-018-0623-8

[18] Kim, S. H., Ahn, J., Ock, M., Shin, S., Park, J., Luo, N., & Jo, M. W. (2016). The EQ-5D-5L valuation study in Korea. *Quality of Life Research*, 25(7), 1845–1852. https://doi.org/10.1007/s11136-015-1205-2

[19] Roudijk, B., Janssen, B., & Olsen, J. A. (2022). How do EQ-5D-5L value sets differ? In N. Devlin et al. (Eds.), *Value Sets for EQ-5D-5L: A Compendium, Comparative Review & User Guide* (pp. 235–258). Springer.

[20] Fayers, P., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., & Bottomley, A. (2001). *EORTC QLQ-C30 scoring manual* (3rd ed.). European Organisation for Research and Treatment of Cancer.

[21] Wailoo, A. J., Hernandez-Alava, M., Manca, A., Mejia, A., Ray, J., Crawford, B., Botteman, M., & Busschbach, J. (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*, 20(1), 18–27. https://doi.org/10.1016/j.jval.2016.11.006

[22] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[23] StataCorp. (2024). Stata Statistical Software: Release 18. College Station, TX: StataCorp LLC.

[24] Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, 13(1), 70–84. https://doi.org/10.1214/aos/1176346577

[25] Van Steen, K., Curran, D., Kramer, J., Molenberghs, G., Van Vreckem, A., Bottomley, A., & Sylvester, R. (2002). Multicollinearity in prognostic factor analyses using the EORTC QLQ-C30: Identification and impact on model selection. *Statistics in Medicine*, 21(24), 3865–3884. https://doi.org/10.1002/sim.1361

[26] Lee, H. J., Ock, M., Kim, K. P., & Jo, M. W. (2018). Estimation of population-based utility weights for gastric cancer-related health states. *Patient Preference and Adherence*, 12, 909–918. https://doi.org/10.2147/PPA.S151946

[27] Mantziari, S., Elliott, J. A., Markar, S. R., et al.; ENSURE Study Group. (2024). Sex-related differences in

oncologic outcomes, operative complications and health-related quality of life after curative-intent oesophageal cancer treatment: Multicentre retrospective analysis. *BJS Open*, 8(2), zrae026. https://doi.org/10.1093/bjsopen/zrae026

[28] Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ*, 350, g7594. https://doi.org/10.1136/bmj.g7594

[29] Hagiwara, Y., Shiroiwa, T., Taira, N., Kawahara, T., Konomura, K., Noto, S., Fukuda, T., & Shimozuma, K. (2020). Mapping EORTC QLQ-C30 and FACT-G onto EQ-5D-5L index for patients with cancer. *Health and Quality of Life Outcomes*, 18(1), 354. https://doi.org/10.1186/s12955-020-01611-w

[30] National Evidence-based Healthcare Collaborating Agency. (2021). *Guidelines for economic evaluation of pharmaceuticals in Korea*. NECA.