UCLA
UCLA Previously Published Works

Title

Predicting the EQ-5D utilities from the Kansas City Cardiomyopathy Questionnaire in patients with heart failure

Permalink

https://escholarship.org/uc/item/4f00s5vt

Journal

European Heart Journal - Quality of Care and Clinical Outcomes, 7(4)

ISSN

2058-5225

Authors

Thomas, Merrill
Jones, Philip G
Cohen, David J
et al.

Publication Date

2021-07-21

DOI

10.1093/ehjqcco/qcab014

Peer reviewed

Predicting the EQ-5D utilities from the Kansas City Cardiomyopathy Questionnaire in patients with heart failure

Merrill Thomas [1,2], **Philip G. Jones**[1,2]**, David J. Cohen**[2]**, Arnold V. Suzanne**[1,2]**,
Elizabeth A. Magnuson**[1]**, Kaijun Wang**[1]**, Vinod H. Thourani**[3]**, Gregg C. Fonarow** [4]**,
Alexander T. Sandhu**[5]**, and John A. Spertus**[1,2]*

[1]Cardiovascular Research, Saint Luke's Mid America Heart Institute, 4401 Wornall Road, Kansas City, MO 64111, USA; [2]Department of Biomedical and Health Informatics, University of Missouri-Kansas City, Kansas City, MO, USA; [3]Department of Cardiovascular Surgery, Marcus Valve Center, Piedmont Heart Institute, 95 Collier Road Northwest, Suite 5015, Atlanta, GA 30309, USA; [4]Department of Internal Medicine, Division of Cardiology, Ahmanson-UCLA Cardiomyopathy Center, Ronald Reagan UCLA Medical Center, 100 UCLA Medical Plaza, Suite 630, Los Angeles, CA 90095, USA; and [5]Division of Cardiology, Department of Medicine, Stanford University, 300 Pasteur Dr, Stanford, CA 94305, USA

Introduction	Evaluation of health status benefits, cost-effectiveness, and value of new heart failure therapies is critical for supporting their use. The Kansas City Cardiomyopathy Questionnaire (KCCQ) measures patients' heart failure-specific health status but does not provide utilities needed for cost-effectiveness analyses. We mapped the KCCQ scores to EQ-5D scores so that estimates of societal-based utilities can be generated to support economic analyses.
Methods	Using data from two US cohort studies, we developed models for predicting EQ-5D utilities (3L and 5L versions) from the KCCQ (23- and 12-item versions). In addition to predicting scores directly, we considered predicting the five EQ-5D health state items and deriving utilities from the predicted responses, allowing different countries' health state valuations to be used. Model validation was performed internally via bootstrap and externally using data from two clinical trials. Model performance was assessed using R^2, mean prediction error, mean absolute prediction error, and calibration of observed vs. predicted values.
Results	The EQ-5D-3L models were developed from 1000 health status assessments in 547 patients with heart failure and reduced ejection fraction (HFrEF), while the EQ-5D-5L model was developed from 3925 patients with HFrEF. For both versions, models predicting individual EQ-5D items performed as well as those predicting utilities directly. The selected models for the 3L had internally validated R^2 of 48.4–50.5% and 33.7–45.6% on external validation. The 5L version had validated R^2 of 57.7%.
Conclusion	Mappings from the KCCQ to the EQ-5D can yield the estimates of societal-based utilities to support cost-effectiveness analyses when EQ-5D data are not available.
Keywords	Health status • Health economics • Utilities

Introduction

Comprehensive assessments of the benefits of novel therapies entail explicitly identifying the effect of treatment on both patients' survival and their health status: their symptoms, function, and quality of life. Once a treatment is demonstrated to improve outcomes relative to an alternative treatment, an analysis of its cost-effectiveness is often done to establish its value. Conducting such analyses requires

---

distilling patients' health status to a utility, which is a number between 0 and 1, where 0 represents death and 1 perfect health. The utility is then multiplied by survival to generate quality-adjusted life years from which the incremental cost-effectiveness ratio of the new treatment relative to the standard can be calculated.[1] While utilities can theoretically be calculated by direct elicitation from patients using the standard gamble or time trade-off methods,[2–5] current approaches seek to use society's values of health derived from health status assessments.[6,7] The most common means of assessing health state utilities is by administering the EQ-5D questionnaire, for which responses to the items can be mapped to a country's specific utility weights.[8,9] If EQ-5D data are unavailable, algorithms for obtaining estimates of EQ-5D utility scores from both generic and disease-specific health status assessments are being developed.[10]

In some studies, the EQ-5D is not captured. Instead, disease-specific health status instruments, such as the Kansas City Cardiomyopathy Questionnaire (KCCQ), may be used because they are more specific to the patient population and more sensitive to changes due to the cardiovascular treatment. In such cases, cost-effectiveness analyses may require use of results from prior studies to approximate population-average utilities, although important differences often exist in the heart failure populations between different studies. On the other hand, if a disease-specific health status measure is available, it could theoretically be used to provide an estimate of utilities in the population of interest.

Heart failure is a chronic condition for which there is high morbidity and mortality and for which patients' health status can be severely impaired. Consequently, the development of new therapies is an active area of research and development. Many clinical trials have used the KCCQ[11,12] to quantify the impact of new therapies on health status. The KCCQ is a disease-specific patient-reported outcome (PRO) with established validity, reliability, responsiveness, and interpretability.[13–16] The KCCQ directly measures outcomes recently endorsed by the Food and Drug Administration in their guidelines for endpoints of drug development in heart failure[17] and has been qualified as a clinical outcome assessment for approval and labelling of novel treatments.[18,19] While the KCCQ measures patients' health status, there are no currently available methods to map it to utilities, and therefore it cannot be used to support cost-effectiveness analyses. To overcome this limitation, we sought to develop algorithms for mapping the KCCQ to EQ-5D utility scores so that estimates of societal-based utilities could be generated to support economic analyses.

Methods

EQ-5D

The EQ-5D is a standardized instrument for assessing generic health status and consists of a descriptive system and a visual analog scale.[8,9] The descriptive system addresses five domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Patients rank their state of health in each domain on an ordinal scale ranging from no problems to extreme problems. Responses to these items can be converted to a utility score by using valuations obtained from the general population.[6,7] For utilities, a score of 0 represents death and a score of 1.0 represents perfect health; negative scores are possible and indicate a state perceived to be worse than death. Importantly, different countries and cultures may

apply different valuations to different health states. Accordingly, the same items are used in different countries, and country-specific utilities are obtained by mapping the item responses to that country's utility weights.

The original version of the EQ-5D, known as the EQ-5D-3L, provided three response options for each of the five health state items.[20–22] A newer version, the EQ-5D-5L, retains the same five domains as the 3L but expanded the number of response options to five for each item. The 5L has improved discriminatory power, validity, and sensitivity as compared with the 3L, while reducing ceiling effects.[20–22] Algorithms for deriving societal utilities are version-specific. However, for countries for which 5L-based utility scores have yet to be derived, crosswalks are available that map EQ-5D-5L responses to 3L-based scores.[23]

Kansas City Cardiomyopathy Questionnaire

The KCCQ is a 23-item self- or interview-administered disease-specific PRO that quantifies six domains of patients' heart failure-related health status: their physical limitations, symptom stability, symptom frequency, symptom burden, quality of life, and social limitations.[24] Scores generated for each domain range from 0 to 100 with higher scores indicating better health status.[24] The overall summary (OS) score integrates the physical limitations, symptom frequency, symptom burden, quality of life, and social limitation scores; the clinical summary score integrates physical limitations, symptom frequency, and symptom burden.[24] The original KCCQ-23 has also been reduced to a 12-item instrument (KCCQ-12) for ease of implementation.[13] The KCCQ-12 includes symptom frequency, physical and social limitations, and quality of life domains and can also be used generate clinical and OS scores.[13] The KCCQ has been extensively validated and shown to be reproducible and sensitive to clinical change in multiple heart failure aetiologies including heart failure with reduced and preserved ejection fractions and valvular heart disease.[13–15,24–26] In addition, among patients with heart failure, KCCQ scores are strongly and independently associated with survival, heart failure admissions, and costs.[27–29]

Data sources

To support both versions of the EQ-5D, for development of the mapping algorithms, we utilized data from the KCCQ Interpretability Study (KCCQINT) for the 3L version, and from the CHAnge the Management of Patients with HF (CHAMP-HF) study for the 5L. KCCQINT is an observational cohort study of 547 patients with heart failure and reduced ejection fraction (HFrEF) recruited from 14 heart failure clinics in the USA between 2001 and 2002.[25] Patients completed the KCCQ and EQ-5D-3L at baseline and 6 (±2) weeks later. All participants provided informed consent, and Institutional Review Board (IRB) approval was obtained at each study site. CHAMP-HF is a multicentre, prospective registry of 4839 outpatients with HFrEF (left ventricular ejection fraction ≤40%) enrolled from 140 centres in the USA between 2015 and 2017.[30] The KCCQ and EQ-5D-5L were collected at baseline and 3, 6, 9, 12, 18, and 24 months. For model development, we used patients' 6-month assessments. All study participants provided written informed consent, and each study centre obtained site-specific institutional review board approval. Novartis Pharmaceuticals Corporation (East Hanover, NJ) sponsored CHAMP-HF and shared the CHAMP-HF data but did not fund these analyses, nor review the paper prior to submission.

For the EQ-5D-3L, we externally validated our models using data from Heart Failure: A Controlled Trial Investigating Outcomes of Exercise Training (HF-ACTION) and Placement of Aortic Transcatheter Valves 2 (PARTNER 2A). HF-ACTION is a randomized controlled trial of 2331 patients with HFrEF recruited from 82 centres in the USA, Canada, and France between 2003 and 2007.[31] Patients completed the KCCQ and EQ-5D at baseline and 3, 6, 9, 12, 24, and 36 months. HF-ACTION data

were accessed through the National Heart, Lung, and Blood Institute with approval from the institutional review board at Saint Luke's Mid America Heart Institute, which granted a waiver of informed consent. PARTNER-2A is a randomized controlled trial of 2032 intermediate-risk patients with severe symptomatic aortic stenosis recruited from 57 centres in the USA and Canada between 2011 and 2013.[32] PARTNER-2A was selected for further validation given the different phenotype (valve disease) and older age of the patients. The KCCQ and EQ-5D were collected at baseline and 1 month, 1 year, and 2 years. For the purpose of this study, we used the baseline and 1-year data. For the EQ-5D-5L, as we did not have access to any other studies' data using this instrument, we performed model validation using 12-month assessments from CHAMP-HF.

Statistical analysis

We employed two general approaches to predict EQ-5D scores from the KCCQ. In the first approach, we predicted utility scores directly, using US TTO-based valuations.[33] Within this approach, we considered two alternative models of EQ-5D scores: 1) a linear regression model and 2) a two-part regression model, using logistic regression for the probability of EQ-5D utility being equal to 1.0 and linear regression for scores <1.0. The two-part model was included to account for the 'spike' of scores at 1.0, representing patients who reported perfect health status.

In the second approach, we developed separate prediction models for each of the five EQ-5D health state domains (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) using proportional odds regression. These models yield predicted probabilities of each possible response to each of the five EQ-5D items, as a function of the KCCQ. The EQ-5D index score was then calculated by applying the EQ-5D scoring algorithm (using the US TTO value set) to the predicted probabilities of responses for each item, treating these probabilities as 'fractional points' for the item responses. A key advantage of this approach is that, by directly predicting EQ-5D health states from the KCCQ, predictions are not tied to any specific country's valuations, and thus can be applied to different value sets by applying the corresponding EQ-5D scoring algorithm to the predicted responses.

For all models described above, we evaluated different configurations of the KCCQ scores or items as predictor variables. These included 1) a model including only the KCCQ OS score, 2) a model including all individual KCCQ domain scores, and 3) a model including all individual KCCQ items. Nonlinear effects were examined using restricted cubic splines for scores and quadratic terms for individual items. Models were developed for both the 23- and 12-item versions of the KCCQ.

For the EQ-5D-3L models, we used both baseline and 6-week assessments from KCCQINT for model development, using generalized estimating equations to account for the correlation of repeated measurements within patients. We evaluated interactions of predictor effects with assessment time and found no significant differences ($P > 0.18$ across all EQ-5D items and all KCCQ scores) and, therefore, pooled both assessments in estimating predictor coefficients.

For the EQ-5D item-specific predictions, we utilized cumulative link models. We examined various link functions (logit, probit, log–log, complementary log–log), comparing model fit using AIC statistics, and used score tests and graphical methods to evaluate the parallel slopes assumption. In general, a logit link (corresponding to a proportional odds model) performed as well or better than other options, with no substantial departures from proportionality.

Under the assumption that the relationship between KCCQ scores and EQ-5D outcomes would be independent of other patient characteristics, we based our models solely on the KCCQ without incorporation of other covariates. To test this assumption, in the KCCQINT study, we augmented our models with patient demographic characteristics (age, sex, race) and clinical co-morbidities (stroke, arthritis, lung disease, diabetes, peripheral artery disease). We found that none of these factors were significantly associated with EQ-5D scores after accounting for participants' KCCQ scores.

In the KCCQINT study, some KCCQ subscales were missing due to skipped items, the most common being the Social Limitation scale ($n = 33$). To account for missing data, we used multiple imputation methods. The imputation model included the 23 individual KCCQ items and the 5 individual EQ-5D items at both baseline and 6 weeks, along with the EQ-5D Visual Analog Scale and patients' age and sex as auxiliary covariates. Twenty randomly imputed data sets were generated; models were built on each imputed data separately and results were pooled to obtain final beta weights.

In CHAMP-HF, of the 4839 enrolled patients, 3925 completed the 6-month assessment. Of those who completed the assessment, there were no skipped items for either the KCCQ or EQ-5D.

Model performance was assessed using R^2, mean prediction error, mean absolute prediction error, and calibration plots of observed vs. predicted values. Internal validation was performed using bootstrap methods to evaluate the risk of overfitting. External validation was performed in HF-ACTION and PARTNER 2A.

Results

EQ-5D-3L

Study population

The 547 patients with HFrEF in the KCCQINT analytic cohort had a median age of 60 years (Interquartile Range (IQR) 52, 71), 23.8% were women, and the median baseline ejection fraction was 25% (IQR 20, 30; *Table 1*). The median KCCQ clinical summary and OS scores were 66.7 (IQR 46.9, 85.4) and 61.5 (41.7, 80.5), respectively. Out of the 1094 total possible assessments (547 patients times 2 assessments), data were missing on 94 assessments (8.6%) for the following reasons: incomplete baseline EQ-5D ($n = 15$), incomplete 6-week assessment due to death ($n = 5$), hospitalization ($n = 5$), refusal ($n = 13$), loss to follow-up ($n = 39$), incomplete 6-week KCCQ ($n = 1$), or incomplete 6-week EQ-5D ($n = 16$). As such, our derivation cohort included 1000 health status assessments. The mean EQ-5D-3L index score was 0.75 (SD 0.17), and the median was 0.78 (interquartile range 0.69–0.84). Fourteen percent of participants had an EQ-5D-3L index score of 1.0. The distribution of participants' EQ-5D-3L index scores is presented in *Figure 1*.

Model derivation

Model performance characteristics are summarized in Supplementary material online, *Table S1*. For all three prediction methods (linear model of EQ-5D index, two-part model of index, proportional odds models of EQ-5D items), models based solely on the KCCQ OS score yielded nearly identical bootstrap-corrected R^2 values of ~50%. In all cases, expanding the predictor set to include all KCCQ domain scores or all individual KCCQ items resulted in slightly improved apparent fit, but after bootstrap correction, the model fit was no better or slightly worse than that generated by the OS score alone. Mean prediction errors were all near zero, and calibration slopes were near 1.0, indicating good fit and calibration to the data. Adding co-morbidities to the model did not contribute significantly

Table I Baseline characteristics of patients in the Kansas City Cardiomyopathy Questionnaire interpretability study

Demographics	
Age (years)	60.0 (52.0, 71.0)
Missing	1
Female sex	130 (23.8%)
Race or ethnicity	
African American	143 (26.3%)
Latino	14 (2.6%)
Asian	6 (1.1%)
Caucasian	370 (68.0%)
Multiracial	1 (0.2%)
Others	10 (1.8%)
Missing	3
Co-morbidities	
Hypertension	308 (56.3%)
Hyperlipidemia	264 (48.3%)
Diabetes	183 (33.5%)
Prior myocardial infarction	218 (39.9%)
Prior percutaneous coronary intervention	74 (13.5%)
Prior coronary artery bypass graft	150 (27.4%)
Prior valve surgery	27 (4.9%)
COPD/asthma	100 (18.3%)
Renal failure	31 (5.7%)
Baseline ejection fraction (%)	25.0 (20.0, 30.0)
Missing	1
Health status	
NYHA class	
I	59 (10.8%)
II	227 (41.6%)
III	231 (42.3%)
IV	29 (5.3%)
Missing	1
KCCQ physical limitation score	62.5 (41.7, 83.3)
Missing	1
KCCQ symptom stability score	50.0 (50.0, 50.0)
Missing	3
KCCQ symptom frequency score	70.8 (47.9, 87.5)
Missing	3
KCCQ symptom burden score	75.0 (50.0, 91.7)
KCCQ quality of life score	58.3 (33.3, 75.0)
KCCQ social limitation score	56.3 (31.3, 83.3)
Missing	16
KCCQ clinical summary score	66.7 (46.9, 85.4)
KCCQ overall summary score	61.5 (41.7, 80.5)
EQ-5D visual analog scale	62.0 (50.0, 80.0)
Missing	30
EQ-5D utility index	0.778 (0.689, 0.844)
Missing	15

Continuous variables presented as median (interquartile range) and categorical variables presented as proportions.
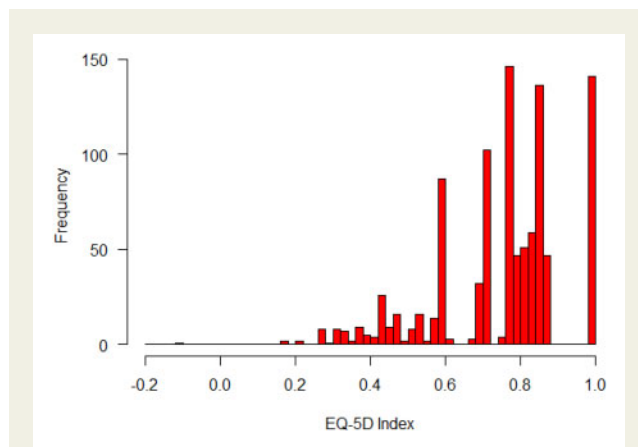COPD, chronic obstructive pulmonary disease; NYHA, New York Heart Association.

Figure I Distribution of EQ-5D index scores.

to model performance; apparent R^2 for the model with additional covariates was 51% compared with 50% with the KCCQ alone (Supplementary material online, *Table S2*).

External validation

We performed external validation of the two-part model based on the KCCQ OS score alone and the item-specific model in HF-ACTION and PARTNER-2A. Baseline characteristics for these cohorts are presented in Supplementary material online, *Table S3*, and results from the external validation are summarized in Supplementary material online, *Table S4*. In the HF-ACTION cohort, predictive accuracy was slightly lower, particularly at baseline ($R^2 =$ 0.34), but with R^2 values in the 40–45% range over follow-up. Predicted utilities tended to underestimate observed utilities slightly, on average, with a slightly greater bias at baseline (-0.034 for the two-part model, -0.040 for the item-specific model) that decreased over follow-up (-0.014 and -0.025 at 24 months, respectively). In the PARTNER 2A cohort, predictive accuracy was better at baseline compared with HF-ACTION ($R^2 \sim$40% in PARTNER vs. 34% in HF-ACTION). At 1-year Follow-up, predictive accuracy was similar in the TAVR group of PARTNER 2A and HF-ACTION but worse for the surgical replacement cohort, potentially due to residual impacts of surgery, of PARTNER 2A (R^2 44% in PARTNER after TAVR vs. 36% after surgery). Mean EQ-5D was underpredicted by about 0.03 at baseline and overpredicted by 0.02–0.03 at follow-up in the PARTNER 2A cohort.

The results for model derivation and validation using the KCCQ-12 resulted in similar performance ($R^2 \sim$47% vs. 50%) and are presented in Supplementary material online, *Tables S5* and *S6*.

EQ-5D-5L

Study population

Among the participants in CHAMP-HF, 3925 were included, with a median age of 67 years (IQR 59, 75), of whom 29.4% were women and the median ejection fraction of the entire cohort was 30% (IQR 23.5, 36.0; *Table 2*). The mean KCCQ-12 score at 6 months was 71.0 ± 22.3 and the mean EQ-5D-5L index score was 0.7 ± 0.24 with 27.4% of participants having an EQ-5D-5L index score of 1.0.

Table 2 Baseline characteristics of patients in the CHAnge the Management of Patients with HF study

	CHAMP-HF (*n* = 3925)
Demographics	
Age (years)	67 (59–75)
Female sex	1152 (29.4%)
Hispanic or Latino ethnicity	746 (19.0%)
Race or ethnicity	
Black/African American	663 (16.9%)
White	2931 (74.7%)
Others	331 (8.4%)
Co-morbidities	
Hypertension	3291 (83.8%)
Hyperlipidemia	3107 (79.2%)
Diabetes	1644 (41.9%)
Prior myocardial infarction	1354 (34.5%)
Renal failure or dysfunction	811 (20.7%)
Baseline ejection fraction (%)	30.0 (23.5–36.0)
Health status	
NYHA class	
I	417 (10.8%)
II	2278 (59.2%)
III	1096 (28.5%)
IV	59 (1.5%)
KCCQ physical limitation score	66.7 (41.7–87.5)
KCCQ symptom score	75.0 (54.2–91.7)
KCCQ quality of life score	62.5 (37.5–87.5)
KCCQ social limitation score	75.0 (50.0–91.7)
KCCQ overall summary score	67.7 (47.9–84.4)
EQ-5D utility index	0.82 (0.73–0.88)

Continuous variables presented as median (interquartile range) and categorical variables presented as proportions.
NYHA, New York Heart Association.

Model derivation

Model performance characteristics are summarized in Supplementary material online, *Table S7*. For all three prediction methods (linear model of EQ-5D index, two-part model of index, proportional odds models of individual EQ-5D items), models including individual KCCQ items performed best followed by models including KCCQ domain scores, and then models including only the OS score. Models including individual KCCQ items resulted in nearly identical bootstrap-corrected R^2 values of ~57% across all three modelling strategies. Mean prediction errors and calibration slopes again indicated good fit and calibration to the data.

Model validation

Models for predicting the EQ-5D-5L were validated using health assessment date in CHAMP-HF at 12 months, and results are summarized in Supplementary material online, *Table S8*. Predictive accuracy was similar in the validation cohort with R^2 values ranging 49–56%. Mean EQ-5D was slightly overpredicted at 12 months (0.001–0.004 for the two-part model, 0.007–0.013 for the item-specific model).

Final selected models

We selected models predicting individual EQ-5D health state items as the final models. These models performed as well as models predicting EQ-5D utility scores directly. In addition, by directly predicting EQ-5D health states from the KCCQ, predictions are not tied to any specific country's valuations, and thus, valuations from different countries can be obtained from the KCCQ. The performance and validation characteristics of these models are summarized in *Table 3* and *Figure 2*.

Using predicted utilities: statistical considerations

The results of external validation in various data sets suggest that the proposed KCCQ to EQ-5D mappings have good accuracy for predicting population-average utilities, which are necessary for cost-effectiveness analyses. However, it is important to note that predicted utilities *for individual patients* will be, by the nature of prediction models, shrunken away from their actual utilities and towards the population average (since no prediction model is perfect). An important implication of this for analyses is that the variability of predicted utilities is smaller than that of actual utilities; any analysis that relies on more than just the population-average utility and simply treats the individual predictions as actual will underestimate the true variability. If the predicted utility is the dependent variable in an analysis, this will produce unconservative results (e.g. confidence intervals and *P*-values that are too small). Considering this, we compared the variability of predicted and actual utilities in our various validation cohorts; results are provided in Supplementary material online, *Table S9*. While there is variation from study to study, these results suggest that the variance of actual utilities may be ~3 times greater than that of the predicted values for the EQ-5D-3L mappings (using either the KCCQ-23 or the KCCQ-12), and roughly 1.7 times greater for EQ-5D-5L predictions. These inflation factors could be applied in certain analyses to correct for underestimation of variability. For example, a confidence interval for the EQ-5D-3L population-average utility could be estimated by first obtaining a confidence interval using usual methods, treating the predicted values as actual, but then adjusting the resulting confidence interval by multiplying its half-width by the square root of 3.

Discussion

The evaluation of the cost-effectiveness of novel therapies is an increasingly important step in identifying the value of new treatments. A critical requirement in performing such analyses is to have utility weights available that can be used to adjust the survival estimates for quality of life. A number of heart failure clinical trials and many observational studies used for comparative effectiveness research may only have health status assessments from the KCCQ and thus lack a direct or indirect measure of individual patients' utilities. Underscoring the need for approaches to estimate cost-effectiveness when utilities have not been collected, the National Institute for Health and Clinical Excellence emphasized the importance of such methods in their 2013 guidelines supporting mapping the EQ-5D to health-related quality of life measures when utilities are not available.[34]

Table 3   Kansas City Cardiomyopathy Questionnaire to EQ-5D mapping: final selected models

Mapping	Model development				Model validation			
	Data source	N	R^2 (%)	Mean prediction error	Data source	N	R^2 (%)	Mean prediction error
KCCQ-23 to EQ-5D-3L	KCCQINT, 0 and 6w	1000	50	0.00	HF-ACTION, 12m	1604	42	-0.03
					PARTNER-2A, SAVR, 12m	678	36	0.02
					PARTNER-2A, TAVR, 12m	791	44	0.02
KCCQ-12 to EQ-5D-3L	KCCQINT, 0 and 6w	1000	48	-0.01	HF-ACTION, 12m	1604	40	-0.03
					PARTNER-2A, SAVR, 12m	678	34	0.01
					PARTNER-2A, TAVR, 12m	791	42	0.01
					CHAMP-HF, 12m	3522	46	-0.03
KCCQ-12 to EQ-5D-5L	CHAMP-HF, 6m	3925	58	-0.00	CHAMP-HF, 12m	3522	56	0.01
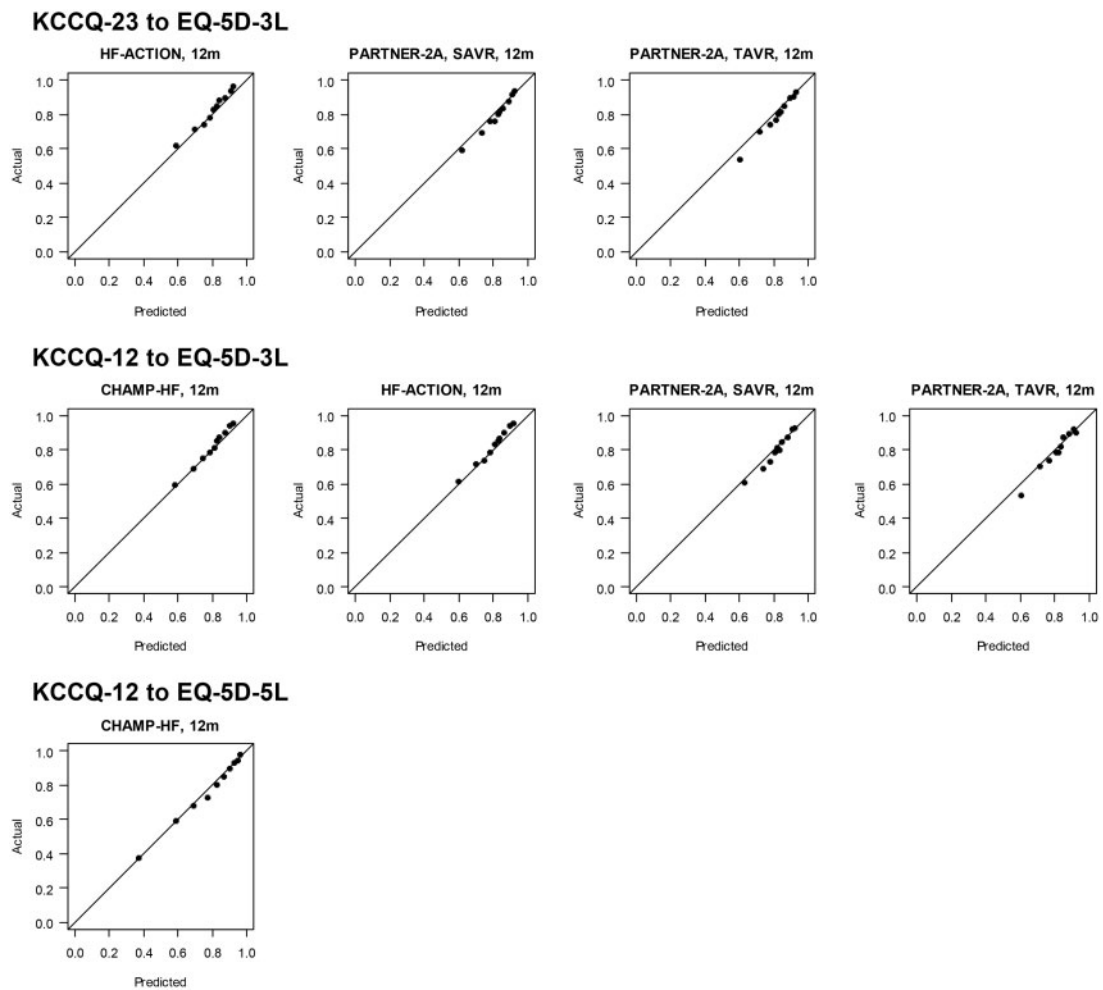
Figure 2 External validation of models.

To facilitate cost-effectiveness analyses when only the KCCQ is available, we created a series of models to map KCCQ scores and individual item responses to the EQ-5D. We found that approximately 50% of the variability in EQ-5D scores could be predicted based on KCCQ scores alone. Even more importantly, given that cost-effectiveness analyses focus on population-level estimates, rather than

patient-level prediction, the mean predicted errors were small and would be applied to both groups in an economic analysis, thereby limiting potential biases. The KCCQ OS score tended to be the best predictor of EQ-5D-3L scores across a range of modelling strategies. In contrast, models using individual KCCQ items were the best predictors of EQ-5D-5L scores. In both models, we would recommend using the models that predict the distribution of responses to the individual EQ-5D items so that different country's weightings could be used to estimate the cost-effectiveness of treatment within that country.

This methodological work adds to the growing body of literature in which non-preference-based measures of quality of life or symptoms are mapped to utilities. To date, at least 391 mapping studies have been performed, with 207 of these mapping to utilities derived from the EQ-5D.[10] Almost all studies performed direct mapping to predict EQ-5D utilities, and approximately one-fifth conducted response mapping to predict responses to each EQ-5D domain.[10] The most common analytic approach used in mapping studies is the ordinary least squares method of linear regression.[10,35,36] While mapping algorithms have been developed for a wide range of disease categories, only nine have been developed for the derivation of utility scores in patients with in cardiovascular disease.[10]

Our work extends this body of literature by mapping responses from a disease-specific PRO for heart failure to EQ-5D data for the derivation of health state utilities. To the best of our knowledge, this is the first study to develop a mapping algorithm for the conversion of KCCQ scores to utility weights that also provided validation data, supported the 23- and 12-item versions of the KCCQ and created algorithms for both the 3L and 5L versions of the EQ-5D. The only previous studies to estimate utility weights in the heart failure population are an unpublished study using the Minnesota Living with Heart Failure Questionnaire (MLWHF) and a second using the MLWHF plus other patient characteristics and demographics.[10,36,37] Our study extends this prior work in several important ways. First, we developed self-contained algorithms to map the KCCQ alone to the EQ-5D, which has the advantage of being more widely applicable as these algorithms do not depend on the collection of additional data. The validity of this approach was supported by finding no significant associations with other patient characteristics and EQ-5D-3L utilities after inclusion of the KCCQ OS score. Second, we developed algorithms for the mapping of KCCQ scores to each of the five individual EQ-5D domains. This allows valuations from different countries to be obtained from the KCCQ, which is important for the value of heart failure therapies to be evaluated by different county's regulatory boards. Third, mapping the KCCQ itself is important as it is commonly collected in clinical trials and is more sensitive to changes in patients' disease status compared with the MLWHF.[24] The sensitivity of the KCCQ is also important as evidence suggests that the validity of predicted utilities may be more dependent on the sensitivity of the measure than on the technique used for derivation.[35] Lastly, we developed models to predict both the 3L and 5L versions of the EQ-5D.

While our study offers unique advantages compared with others, there are also some similarities with prior studies. First, we found the R^2 for our models to be within the acceptable range reported in the literature.[35,36] Importantly, while it might be hypothesized that generic health status measures would have stronger correlations with utilities than disease-specific measures, we found R^2 values that were quite comparable to those reported for generic measures in other studies, which may reflect that heart failure is often the most limiting condition for patients. Second, similar to findings in a previous study mapping the Seattle Angina Questionnaire to the EQ-5D, the 2-part model did not perform better than simple linear regression even though the EQ-5D has a ceiling effect.[38] This may relate to the finding that only 14% of EQ-5D assessments in this study demonstrated a ceiling effect which was lower than that found in the prior SAQ mapping study. Lastly, similar to prior studies, the addition of covariates to our models did not add to model performance.[38,39]

While there are advantages and disadvantages to each model we developed in this analysis, we believe the proportional odds model estimating individual responses of the EQ-5D is particularly useful. By directly estimating the predicting EQ-5D health states from the KCCQ, any specific country's or culture's value sets can be applied to the predicted responses. As heart failure trials often enrol patients in different countries, this approach of mapping the KCCQ to the EQ-5D is particularly useful for the evaluation of cost-effectiveness from different country perspectives. Statistical code for all models is available by contacting John Spertus at spertusj@umkc.edu.

Limitations

The findings from our study should be interpreted in light of the following potential limitations. First, data were obtained from participants in outpatient registries, who may differ from inpatients and from the general heart failure population by their willingness to participate. However, our algorithms are intended for use in similar registries or clinical trials, and it is important that the population used in the mapping process be similar to the population in which the algorithm is to be applied.[36] Second, while missing data were relatively infrequent and we used multiple imputation to mitigate biases, it is possible that non-ignorable missingness persists (e.g. patients with worse quality of life may have been less likely to complete assessments). However, we believe that the likelihood of this affecting our results is fairly low since to do so would imply that the relationship between KCCQ and EQ-5D is different among those with vs. without missing data. Third, we mapped a disease-specific PRO to a preference-based generic utility scale, which assumes that there is overlap between what is being measured by the two tools. This assumption has been questioned when mapping a narrow area of quality of life to the entire domain of health-related quality of life.[35] However, a literature review found the explanatory power of this type of mapping to be similar to the explanatory power of generic to generic mappings, and the explanatory power of our models falls within this same range.[35] Finally, while we de-coupled predictions from specific countries' valuations, we have assumed that the item-specific beta weights derived based on this US outpatient study apply to other countries as well.

Conclusions

While direct collection of EQ-5D data from patients for the derivation of utility scores to be used in cost-effectiveness analysis is still the preferred option, validated mapping algorithms from other non-preference-based health-related quality of life measures are useful

when data from the EQ-5D are not available. In our work, we derived mapping algorithms for the KCCQ to both EQ-5D-derived utilities and to the individual EQ-5D responses. As the KCCQ is frequently collected in the clinical trials evaluating new heart failure therapies, the algorithms presented here could be used by researchers performing economic evaluations based on data from these trials. As such, the code for our algorithms will be made freely available.

Supplementary material

Supplementary material is available at *European Heart Journal – Quality of Care and Clinical Outcomes* online.

Funding

Data availability statement

The data from the KCCQINT underlying this article will be shared on reasonable request to the corresponding author.

The data from HF-ACTION underlying this article are available from the National Heart, Lung, and Blood Institute by request using the following portal: https://biolincc.nhlbi.nih.gov/home/.

The data from PARTNER2A in this article were provided by Edwards Lifesciences by permission. Data will be shared on request to the corresponding author with the permission of Edwards Lifesciences.

The data from CHAMP-HF in this article were provided by Novartis by permission. Data will be shared on request to the corresponding author with the permission of Novartis

References

1. Weinstein MF. *Clinical Decision Analysis.* Philadelphia: WB Saunders and Company; 1980.
2. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics* 1995;**7**:490–502.
3. Torrance GW, Furlong W, Feeny D, Boyle M. Multi-attribute preference functions. Health Utilities Index. *Pharmacoeconomics* 1995;**7**:503–520.
4. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ* 1986;**5**:1–30.
5. Von Neumann JM, *Theory of Games and Economic Behaviour.* Princeton, NJ: Princeton University Press; 1944.
6. Gold MS, Russel LB, Weinstein MC, *Cost-Effectiveness in Health and Medicine.* Oxford: Oxford University Press; 1996.
7. Gafni A. The standard gamble method: what is being measured and how it is interpreted. *Health Serv Res* 1994;**29**:207–224.
8. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;**33**:337–343.
9. Pattanaphesaj J, Thavorncharoensap M. Measurement properties of the EQ-5D-5L compared to EQ-5D-3L in the Thai diabetes patients. *Health Qual Life Outcomes* 2015;**13**:14.
10. Dakin H, Abel L, Burns R, Yang Y. 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 Qual Life Outcomes* 2018;**16**:31.
11. McMurray JJ, Packer M, Desai AS, Gong J, Lefkowitz MP, Rizkala AR et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. *N Engl J Med* 2014;**371**:993–1004.
12. McMurray JJV, Solomon SD, Inzucchi SE, Kober L, Kosiborod MN, Martinez FA et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. *N Engl J Med* 2019;**381**:1995–2008.
13. Spertus JA, Jones PG. Development and validation of a short version of the Kansas City Cardiomyopathy Questionnaire. *Circ Cardiovasc Qual Outcomes* 2015;**8**:469–476.
14. Pokharel Y, Khariton Y, Tang Y, Nassif ME, Chan PS, Arnold SV et al. Association of serial Kansas City Cardiomyopathy Questionnaire assessments with death and hospitalization in patients with heart failure with preserved and reduced ejection fraction: a secondary analysis of 2 randomized clinical trials. *JAMA Cardiol* 2017;**2**:1315–1321.
15. Joseph SM, Novak E, Arnold SV, Jones PG, Khattak H, Platts AE et al. Comparable performance of the Kansas City Cardiomyopathy Questionnaire in patients with heart failure with preserved and reduced ejection fraction. *Circ Heart Fail* 2013;**6**:1139–1146.
16. Spertus JA, Jones PG, Sandhu AT, Arnold SV. Interpreting the Kansas City Cardiomyopathy Questionnaire in clinical trials and clinical care: JACC state-of-the-art review. *J Am Coll Cardiol* 2020;**76**:2379–2390.
17. U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). *Treatment for Heart Failure: Endpoints for Drug Development Guidance for Industry*; 2019. https://www.fda.gov/media/128372/download.
18. U.S. Food & Drug Administration. *Center for Devices and Radiological Health*; 2020. https://www.fda.gov/about-fda/fda-organization/center-devices-and-radiological-health.
19. Clinical Outcome Assessments (COA), Qualification Submissions Office of Cardiology Hematology, Endocrinology, and Nephrology (OCEHM), Division of Cardiovascular and Nephrology (DCN). *DDT COA #000084: Kansas City Cardiomyopathy Questionnaire (KCCQ)*; 2020.
20. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res* 2013;**22**:1717–1727.
21. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *Pharmacoeconomics* 2018;**36**:675–697.
22. Agborsangaya CB, Lahtinen M, Cooke T, Johnson JA. Comparing the EQ-5D 3L and 5L: measurement properties and association with chronic conditions and multimorbidity in the general population. *Health Qual Life Outcomes* 2014;**12**:74.
23. EuroQol Research Foundation. *Implications for CEA EQ5D.* https://euroqol.org/eq-5d-instruments/3l-vs-5l/comparing-eq-5d-3l-and-eq-5d-5l-descriptive-system/.
24. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *J Am Coll Cardiol* 2000;**35**:1245–1255.
25. Spertus J, Peterson E, Conard MW, Heidenreich PA, Krumholz HM, Jones P et al. Monitoring clinical changes in patients with heart failure: a comparison of methods. *Am Heart J* 2005;**150**:707–715.
26. Arnold S, Spertus J, Lei Y, Allen K, Chhatriwalla A, Leon M et al. Use of the Kansas City Cardiomyopathy Questionnaire for monitoring health status in patients with aortic stenosis. *Circ Heart Fail* 2013;**6**:61–67.

27. Chan PS, Soto G, Jones PG, Nallamothu BK, Zhang Z, Weintraub WS *et al.* Patient health status and costs in heart failure: insights from the eplerenone post-acute myocardial infarction heart failure efficacy and survival study (EPHESUS). *Circulation* 2009;**119**:398–407.

28. Heidenreich PA, Spertus JA, Jones PG, Weintraub WS, Rumsfeld JS, Rathore SS *et al.* Health status identifies heart failure outpatients at risk for hospitalization or death. *J Am Coll Cardiol* 2006;**47**:752–756.

29. Kosiborod M, Soto GE, Jones PG, Krumholz HM, Weintraub WS, Deedwania P *et al.* Identifying heart failure patients at high risk for near-term cardiovascular events with serial health status assessments. *Circulation* 2007;**115**:1975–1981.

30. DeVore AD, Thomas L, Albert NM, Butler J, Hernandez AF, Patterson JH *et al.* Change the management of patients with heart failure: rationale and design of the CHAMP-HF registry. *Am Heart J* 2017;**189**:177–183.

31. O'Connor CM, Whellan DJ, Lee KL, Keteyian SJ, Cooper LS, Ellis SJ, HF-ACTION Investigators *et al.* Efficacy and safety of exercise training in patients with chronic heart failure: HF-ACTION randomized controlled trial. *JAMA* 2009;**301**:1439–1450.

32. Leon MB, Smith CR, Mack MJ, Makkar RR, Svensson LG, Kodali SK *et al.*; PARTNER 2 Investigators. Transcatheter or surgical aortic-valve replacement in intermediate-risk patients. *N Engl J Med* 2016;**374**:1609–1620.

33. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care* 2005;**43**:203–220.

34. National Institute for Health and Clinical Excellence (NICE). Guide to the Methods of Technology Appraisal 2013; 2013.

35. Mortimer D, Segal L. Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. *Med Decis Making* 2008;**28**:66–89.

36. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ* 2010;**11**:215–225.

37. Calvert MJ, Freemantle N, Yao G, Cleland JG, Billingham L, Daubert JC *et al.*; CARE-HF investigators. Cost-effectiveness of cardiac resynchronization therapy: results from the CARE-HF trial. *Eur Heart J* 2005;**26**:2681–2688.

38. Wijeysundera HC, Tomlinson G, Norris CM, Ghali WA, Ko DT, Krahn MD. Predicting EQ-5D utility scores from the Seattle Angina Questionnaire in coronary artery disease: a mapping algorithm using a Bayesian framework. *Med Decis Making* 2011;**31**:481–493.

39. Goldsmith KA, Dyer MT, Buxton MJ, Sharples LD. Mapping of the EQ-5D index from clinical outcome measures and demographic variables in patients with coronary heart disease. *Health Qual Life Outcomes* 2010;**8**:54.