Article

MDMP&P Policy & Practice

MDM Policy & Practice
1–14

DOI: 10.1177/2381468320971606
journals.sagepub.com/home/mdm
SAGE

Mapping the Kansas City Cardiomyopathy Questionnaire (KCCQ) Onto EQ-5D-3L in Heart Failure Patients: Results for the Japanese and UK Value Sets

Matthias Hunger, Jennifer Eriksson, Stephane A. Regnier, Katsuya Mori, John A. Spertus, and Joaquim Cristino

Background. Health technology assessment bodies in several countries, including Japan and the United Kingdom, recommend mapping techniques to obtain utility scores in clinical trials that do not have a preference-based measure of health. This study sought to develop mapping algorithms to predict EQ-5D-3L scores from the Kansas City Cardiomyopathy Questionnaire (KCCQ) in patients with heart failure (HF). **Methods.** Data from the randomized, double-blind PARADIGM-HF trial were analyzed, and EQ-5D-3L scores were calculated using the Japanese and UK value sets. Several different model specifications were explored to best fit EQ-5D data collected at baseline with KCCQ scores, including ordinary least square regression, two-part, Tobit, and three-part models. Generalized estimating equations models were also fitted to analyze longitudinal EQ-5D data. To validate model predictions, the data set was split into a derivation (n = 4,465) from which the models were developed and a separate sample (n = 1,892) for validation. **Results.** There were only small differences between the different model classes tested. Model performance and predictive power was better for the item-level models than for the models including KCCQ domain scores. R^2 statistics for the item-level models ranged from 0.45 to 0.52. Mean absolute error in the validation sample was 0.10 for the models using the Japanese value set and 0.114 for the UK models. All models showed some underprediction of utility above 0.75 and overprediction of utility below 0.5, but performed well for population-level estimates. **Conclusions.** Using data from a large clinical trial in HF, we found that EQ-5D-3L scores can be estimated from responses to the KCCQ and can facilitate cost-utility analysis from existing HF trials where only the KCCQ was administered. Future validation in other HF populations is warranted.

Keywords
EQ-5D, heart failure, Japan, KCCQ, mapping algorithm, United Kingdom, utility

Date received: March 12, 2020; accepted: September 15, 2020

Heart failure (HF) is a major cardiovascular disorder with a prevalence of >5 per 1,000 in regions of North America, Oceania, and Europe,[1] and rates of 21 per 1,000 after 65 years of age in the United States.[2] In Japan, approximately 1 to 2 million people have HF, which is projected to increase due to the aging of the population and the growing adoption of a Westernized lifestyle.[3,4] In a meta-analysis of 30 studies, 40.2% of HF patients died during a median follow-up of 2.5 years.[5] HF constitutes a high global economic burden estimated

Corresponding Author
Matthias Hunger, Real-World Evidence Strategy & Analytics, ICON Clinical Research GmbH, Konrad-Zuse-Platz 11, 81829 München, Bayern, Germany; Telephone: (49) 89 666 105 121 (Matthias.hunger@iconplc.com).

to cost US$108 billion per annum.[6] Individuals with HF have markedly impaired health-related quality of life (HRQoL) compared with both the general population and those with other chronic diseases.[7] HF is a clinical syndrome caused by structural and/or functional cardiac abnormalities resulting in reduced cardiac output and/or elevated intracardiac pressures at rest or during stress and manifests itself to patients as symptoms of fatigue and breathlessness.[8] An important subset of patients with HF are those with reduced ejection fraction (HFrEF; the left ventricular ejection fraction [LVEF] $<40\%$),[8] which is important given the demographics, comorbidities, response to therapies, and outcomes in this population.[9]

To balance the clinical benefit and costs of treatment, economic evaluations are commonly conducted by assessing the incremental cost per incremental quality-adjusted life years (QALYs) gained. The EQ-5D is a validated generic preference-based questionnaire of HRQoL used to derive health utilities,[10] a measure of preference ranging from 0 (death) to 1 (full health), which can be multiplied by observed survival in economic models to estimate QALYs. The combinations of the EQ-5D health dimensions and their severity levels represents health states that have been valued by the strength of preference to each health state from general population or patient studies, resulting in numerous country-specific value sets for the EQ-5D-3L. It is of importance that economic evaluations are conducted using utilities derived with the value set developed for that particular country due to differences in preferences across countries and cultures.[11] With the Japanese Ministry of Health, Labour and Welfare introducing health technology assessment as of April 2019, where cost-effectiveness is principally assessed as cost per QALY, these considerations warrant a mapping algorithm based on the Japanese value set of the EQ-5D.[12,13]

It is not uncommon, however, that preference-based instruments (such as the EQ-5D) are not administered in clinical trials or observational studies, so as not to overburden the patient or because disease-specific instruments, which are more sensitive to capture health status in a particular disease, are used instead.[14] In such circumstances, countries, such as the UK National Institute for Health and Care Excellence (NICE) and Japan, recommend that an algorithm map the disease-specific instrument onto the EQ-5D for generating utility estimates be used.[12,15] A mapping algorithm allows for the disease-specific instrument to be regressed onto the EQ-5D for estimating utilities. Mapping algorithms can be classified into direct methods, where regression models are used to predict EQ-5D utilities from non-preference-based measures, and response mapping methods, where categorical regression models are used to predict response levels for each of the five EQ-5D domains.[16]

In cardiovascular disease, there are few algorithms mapping a disease-specific HRQoL instrument onto a utility instrument.[17] Chen et al. developed mapping algorithms between the MacNew Heart Disease Quality of Life Questionnaire (MacNew) instrument and six utility instruments, including the EQ-5D and the Short Form 6D (SF-6D).[18] Edlin et al. mapped the Minnesota Living with Heart Failure Questionnaire to the EQ-5D.[17,19] In HF, the Kansas City Cardiomyopathy Questionnaire (KCCQ) is a validated instrument that is gaining increasing use as an endpoint in clinical trials and observational studies,[20] but there are no published algorithms mapping the KCCQ questionnaire to the EQ-5D. To address this gap in the literature, we sought to develop a mapping algorithm from the KCCQ to the EQ-5D-3L that can be used in health technology assessments for HF interventions. Our focus was to develop a mapping algorithm for the Japanese value set to be used in Japan. We also developed a mapping algorithm using the UK value set since this is widely used and could potentially be used in other countries integrating cost-utility analyses in their evaluations of new therapies.

Methods

EQ-5D-3L

The EQ-5D-3L consists of a five-item instrument that also includes a visual analogue scale (VAS). The five items measure patients' perceptions of their "mobility," "self-care," "usual activities," "pain/discomfort," and "anxiety/depression."[10] Each item has three ordinal

Real-World Evidence Strategy & Analytics, ICON plc, Munich, Germany (MH); Real-World Evidence Strategy & Analytics, ICON plc, Stockholm, Sweden (JE); HEOR Excellence, Novartis Pharma AG, Basel, Switzerland (SAR); Market Access and Public Affairs Division, Novartis Pharma K.K, Tokyo, Japan (KM); Saint Luke's Mid America Heart Institute/UMKC, Kansas City, Missouri (JAS); Global Patient Access, Novartis, Basel, Basel-Stadt, Switzerland (JC).
The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Matthias Hunger and Jennifer Eriksson are employed by ICON plc, a research company acting as consultants to pharmaceutical companies. Stephane A. Regnier is an employee and shareholder of Novartis AG. Katsuya Mori is a permanent employee of Novartis Pharmaceuticals Corporation. Joaquim Cristino is a permanent employee and shareholder of the Novartis Pharmaceuticals Corporation. Prof Spertus owns copyright to the KCCQ and provides paid consulting services to Novartis. The author(s) received no financial support for the research, authorship, and/or publication of this article.

responses (1 = no, 2 = moderate, or 3 = extreme problems), defining a total of 243 different health states. Using a scoring algorithm, these health states can be translated into utility values based on valuations by the general population of each country. The resulting utilities are on a scale where 1 represents full health, 0 represents death, and a negative number represents a health state worse than death. In this study, EQ-5D-3L index scores were calculated using the Japanese value set, but coefficients for a mapping algorithm based on the UK value set are also provided as this is a more widely used value set. Detailed methods on the development of each value set, both based on the time tradeoff method, can be found in Tsuchiya et al[21] and Dolan,[22] respectively.

KCCQ

The KCCQ is a self-administered, 23-item questionnaire that quantifies physical limitations, symptom stability, symptoms, self-efficacy, social interference, and HRQoL in patients with HF.[20] Items are summed within each domain and scaled to a score ranging from 0 to 100, where 0 represents the worse symptoms and function and 100 represents the best. In addition, two summary scores can be calculated from the six domains. The Clinical Summary score combines the physical limitation and symptom domains, similar to the New York Heart Association (NYHA), while the Overall Summary score combines the Physical Limitation, Total Symptom, HRQoL, and Social Limitation scores.

Data

Data collected in the PARADIGM-HF trial (Clinical Trials.gov Identifier: NCT01035255) were used in this analysis, details of which have been published elsewhere.[23] The PARADIGM-HF patient population comprised 8,399 adult patients from 47 different countries with HFrEF, NYHA class II-IV, and either a plasma brain natriuretic peptide (BNP) >150 pg/mL or N-terminal pro-brain natriuretic peptide (NT-proBNP) >600 pg/mL or a hospitalization for heart failure within the past 12 months. Patients in the trial were recruited between 2009 and 2012, randomized to receive either enalapril or sacubitril/valsartan, and followed for a median of 27 months. Patients initially had to have an LVEF $\leq 40\%$, but this was changed to $\leq 35\%$ by a protocol amendment after approximately 1,285 patients had been randomized. As a result, there were 7,478 (88.6%) patients with LVEF $\leq 35\%$ and 963 (11.4%) patients with LVEF 35% to 40% randomized. KCCQ and EQ-5D-3L questionnaires were administered at baseline, 4 months, 8 months, 12 months, and annually thereafter through to the final visit.

This analysis used data on patients who completed both the KCCQ and the EQ-5D-3L questionnaires at randomization. Observations with missing values on KCCQ and EQ-5D-3L or any of the relevant baseline characteristics were excluded. To ensure similarity between the estimation sample (PARADIGM-HF) and the inclusion criteria typically being used in clinical trials of patients with heart failure with reduced ejection fraction, patients with LVEF between 35% and 40% were excluded.

From the 8,399 patients included in the primary efficacy population of the PARADIGM-HF trial, 7,623 (91%) completed the KCCQ questionnaire at randomization.[24] Excluding patients with LVEF 35% to 40% and removing observations with missing EQ-5D-3L data or missing covariate data reduced the final estimation sample to 6,357 patients. For the estimation of the mapping algorithms, the final estimation sample was randomly split into a derivation (used to develop the mapping algorithms; $\sim 70\%$ of total sample; $n = 4{,}465$) and a validation sample ($\sim 30\%$; $n = 1{,}892$). The validation sample was used to assess how well the estimated mapping algorithm predicts utility in an independent sample. Clinical and demographic characteristics of the derivation and validation samples are summarized in Table 1.

Model Estimation

A direct mapping method was used, based on regression models where the independent variables were the KCCQ scores (summary or individual domains) and the dependent variable the EQ-5D-3L utility score (derived either from the UK or the Japanese value sets). In the main analysis, the mapping algorithm was estimated from KCCQ and EQ-5D-3L data collected at baseline using ordinary least squares (OLS) regression with robust standard errors.[25] Seven model specifications were fitted for each type of model (OLS, two-part, Tobit, and generalized estimating equations [GEE] models). For some of these model specifications, we applied variable selection methods, where, in each step, a variable was considered for addition to (forward selection) or subtraction from (backward selection) the set of explanatory variables based on the Bayesian information criterion (BIC). In this context, we defined "statistical relevant" variables as variables that improve model fit based on the BIC: Model 1 uses the KCCQs overall score. Models 2 to 5 are

Table 1 Baseline Demographic, Clinical, KCCQ, and EQ-5D-3L Data

	Overall (N = 6,357)	Derivation (n = 4,465)	Validation (n = 1,892)
Age in years			
Mean (SD)	63.51 (11.18)	63.46 (11.38)	63.63 (10.69)
BMI (kg/m^2)			
Mean (SD)	28.34 (5.48)	28.35 (5.55)	28.33 (5.3)
Current smoker			
No	5,404 (85%)	3,778 (84.6%)	1,626 (85.9%)
Yes	953 (15%)	687 (15.4%)	266 (14.1%)
Diabetes			
No	4,170 (65.6%)	2,923 (65.5%)	1,247 (65.9%)
Yes	2,187 (34.4%)	1,542 (34.5%)	645 (34.1%)
Heart rate			
Mean (SD)	72.19 (11.89)	72.15 (11.84)	72.29 (12.01)
Ischemic etiology			
No	2,584 (40.6%)	1,810 (40.5%)	774 (40.9%)
Yes	3,773 (59.4%)	2,655 (59.5%)	1,118 (59.1%)
NT-proBNP (pg/mL)			
Mean (SD)	341.48 (455.57)	345.91 (468.03)	331.05 (424.64)
NYHA class			
I	280 (4.4%)	203 (4.5%)	77 (4.1%)
II	4,442 (69.9%)	3,129 (70.1%)	1,313 (69.4%)
III	1,590 (25%)	1,100 (24.6%)	490 (25.9%)
IV	45 (0.7%)	33 (0.7%)	12 (0.6%)
Previous hospitalization for HF			
No	2,323 (36.5%)	1,645 (36.8%)	678 (35.8%)
Yes	4,034 (63.5%)	2,820 (63.2%)	1,214 (64.2%)
Prior stroke			
No	5,824 (91.6%)	4,080 (91.4%)	1,744 (92.2%)
Yes	533 (8.4%)	385 (8.6%)	148 (7.8%)
Region			
Asia/Pacific and Other	869 (13.7%)	611 (13.7%)	258 (13.6%)
Central Europe	2,234 (35.1%)	1,548 (34.7%)	686 (36.3%)
Latin America	1,082 (17%)	741 (16.6%)	341 (18%)
North America	532 (8.4%)	382 (8.6%)	150 (7.9%)
Western Europe	1,640 (25.8%)	1,183 (26.5%)	457 (24.2%)
Sex			
Female	1,326 (20.9%)	922 (20.6%)	404 (21.4%)
Male	5,031 (79.1%)	3,543 (79.4%)	1,488 (78.6%)
Sodium (mmol/L)			
Mean (SD)	141.44 (3.08)	141.45 (3.03)	141.42 (3.19)
Years since HF diagnosis			
1–5 years	2,411 (37.9%)	1,686 (37.8%)	725 (38.3%)
$\leq$ 1 year	1,873 (29.5%)	1,304 (29.2%)	569 (30.1%)
$>$5 years	2,073 (32.6%)	1,475 (33%)	598 (31.6%)
KCCQ Domain scores, mean (SD)			
Physical limitations	72.85 (22.49)	72.54 (22.55)	73.57 (22.34)
Symptoms	79.56 (19.44)	79.37 (19.58)	80.01 (19.10)
Symptom stability	63.20 (20.91)	63.19 (21.09)	63.21 (20.50)
Quality of life	67.55 (22.35)	67.29 (22.62)	68.18 (21.70)
Self-efficacy	79.34 (19.77)	79.19 (19.84)	79.69 (19.60)
Social limitations	71.87 (25.32)	71.53 (25.40)	72.68 (25.11)
KCCQ-CS score	76.20 (19.20)	75.95 (19.26)	76.79 (19.06)
KCCQ-OS score	72.96 (19.44)	72.68 (19.58)	73.61 (19.09)
EQ-5D-3L utility score, mean (SD)			
Japanese value set	0.772 (0.173)	0.770 (0.173)	0.778 (0.173)
UK value set	0.779 (0.215)	0.776 (0.217)	0.786 (0.210)

BMI, body mass index; HF, heart failure; KCCQ, Kansas City Cardiomyopathy Questionnaire; NT-proBNP, N-terminal pro-brain natriuretic peptide; NYHA, New York Heart Association; SD, standard deviation; UK, United Kingdom.

based on KCCQ domain scores: Model 2 includes all domains regardless of statistical significance. Model 3 includes only statistically relevant KCCQ domain scores; it excluded variables showing an R^2 below 0.05 in a univariate analysis (i.e., an OLS model with only one covariate) or showing a high Pearson correlation of >0.8 with a more predictive KCCQ domain score; afterwards, a backward regression analysis was performed which subsequently removed variables that are not statistically relevant based on the BIC. Model 4 includes KCCQ domain scores and statistically relevant squared terms; it was obtained by applying a forward variable selection method to model 2. Model 5 includes statistically relevant KCCQ domain scores plus statistically relevant demographic and clinical variables (age, sex, region, NYHA class, heart rate, NT-proBNP, sodium, body mass index, diabetes, time since HF diagnosis, ischemic etiology, history of stroke, smoking, and history of hospitalization for HF); it was obtained by applying a forward variable selection method to model 3. Models 6 and 7 are item-level models. Model 6 includes statistically relevant KCCQ item scores and was obtained by applying a forward variable selection model starting from a model with intercept only. Model 7 merges item levels for levels that are shown to be disordered in model 6 (where "disordered" means that regression coefficients do not continuously decrease with increasing limitations).[26]

To address potential bias caused by the nonnormally distributed dependent variable, in particular, the bounded nature of EQ-5D utility scores, resulting in a large spike at 1, two-part models and Tobit models were also explored.[16,26] The two-part model uses logistic regression to predict the probability of whether patients are in perfect health (i.e., have a utility of 1), and a truncated OLS regression model to predict utility values for those not in perfect health (i.e., <1). The results from the two parts of the model are then combined to an overall utility value based on the expected value approach.[26] The Tobit model assumes that there is an underlying latent variable which has a normal distribution and can extend beyond 1.[15,16,26] The mean of this latent variable is modelled as a linear combination of the covariates; the Tobit model assumes that the distribution is censored at 1, taking into account that utility predictions cannot exceed 1. Whereas OLS, two-part, and Tobit models were fitted to cross-sectional data collected at baseline, a fourth mapping algorithm was developed using pooled data collected at baseline, month 4, and month 8. This fourth mapping algorithm was estimated using GEE to account for potential within-subject correlation inherent in such a repeated measurement design.[27]

For the two-part regression model, variable selection was performed independently in each submodel (i.e., the set of variables selected in the logistic regression model to predict the spike at 1 could be different from the set of variables selected in the truncated linear regression model to predict utilities lower than 1). For the GEE models, no variable selection was performed, but the same variables as in the OLS model were included.

Model Performance and Predictive Power

In line with recommendations made by ISPOR and NICE,[15,16] measures of model performance were calculated to compare different model specifications in the derivation sample, while measures of predictive performance were calculated in the validation sample. Model performance calculated in the derivation dataset included the Akaike information criterion (AIC), BIC, and the pseudo-R^2. For the GEE, the "quasi-likelihood under the Independence Model Criterion" (QIC) versions for AIC and BIC were calculated. The pseudo-R^2 was calculated using the formula suggested by Efron:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

where y_i is the observed value for observation i, $\widehat{y_i}$ is the value predicted by the model, and $\bar{y}$ is the overall mean in the sample. For OLS regression, the pseudo-R^2 is identical to the traditional R^2.

Predictive performance calculated in the validation sample included mean absolute error (MAE; i.e., mean absolute difference between estimated and observed utilities) and root mean squared error for each model to compare predicted values with observed values. Mean, standard error (SE), median, and range of observed and predicted values in the validation sample were also calculated.

Some mapping algorithms have reported underprediction of EQ-5D-3L values for mild health states, and overprediction for more severe health states.[28–30] To assess if there was a systematic error in the predictions caused by the severity of the underlying health state, model predictions and observed values are also reported across different levels of observed EQ-5D-3L values in the validation sample, together with MAEs per level (Level 1: <0.25; level 2: 0-25–0.5; level 3: 0.5–0.75; level 4: 0.75–1; levels chosen as per NICE guidance[31]).

Exploratory and Sensitivity Analyses

To investigate the potential issue of overpredicting low utilities, additional exploratory analyses using a three-part model were performed for the selected best-fitting model. In the three-part model, observations were first categorized into three different health states: perfect health (i.e., a utility of 1), severe health (i.e., at least one level 3 in any of the EQ-5D-3L dimension level scores), and moderate health (all remaining observations). Then, a multinomial regression model was fitted to predict the probability of whether responders were in each of the three health states. Afterwards, two truncated linear regression models were applied to predict EQ-5D-3L values for those that were observed to be in moderate or severe health. Similar to the two-part model, results from the three parts of the model were finally combined to an overall utility value based on the expected value approach.

As an exploratory analysis, we also fitted a two-part model where a beta-regression instead of an OLS model was used for the continuous part of the model.[32]

As a sensitivity analysis, the selected best-fitting model was refitted in the subgroup of Asian patients, and in the larger sample of patients enrolled in PARADIGM-HF irrespective of LVEF at baseline (i.e., also including patients with LVEF between 35% and 40%). The best-fitting model was also refit applying a 10-fold cross validation where the full dataset (i.e., combining initial derivation and validation dataset) was partitioned into 10 equally sized segments. Subsequently, 10 iterations of derivation and validation were performed such that within each iteration a different fold of the data was held-out for validation while the remaining nine-folds were used for learning.

All analyses were conducted using the statistical software SAS 9.4.

Results

Baseline characteristics in the overall, derivation, and validation sample are shown in Table 1. Distribution of baseline characteristics was similar as expected given the random split between estimation and validation sample.

Mean KCCQ scale scores and EQ-5D-3L utility scores at baseline in the overall, derivation, and validation sample are shown in Table 1. Mean EQ-5D-3L utility, based on the Japanese value set, was 0.770 (SD: 0.173) in the derivation, and 0.778 (SD: 0.173) in the validation, samples. Mean EQ-5D-3L utility, based on the UK value set, was 0.776 (SD: 0.217) in the derivation and 0.786 (SD: 0.210) in the validation samples. There was a ceiling effect for both value sets in that 30.7% of patients at baseline had a utility value of 1.0 (Figure 1).

Table 2 summarizes measures of predictive power and model fit for OLS models 1 to 7 using the Japanese value set. Overall, all OLS models predicted the overall mean utility in the derivation sample with close precision, although the item-level models 6 and 7 predicted the median utility better than the domain score models. Predicted maximum values ranged from 0.93 for model 1 to 0.98 for model 4, while minimum values ranged from 0.35 for models 2 and 3 to 0.46 for model 4. MAE was highest (i.e., worst) for model 1 and lowest (i.e., best) for the two item-level models 6 and 7. Pseudo-R^2 and AIC also favored item-level models 6 and 7. BIC was best for model 4 and model 7. For all OLS models, MAE was large for patients in poor health (MAE ranging from 0.42 to 0.49 for patients with an EQ-5D score below 0.25; MAE ranging from 0.14 to 0.16 for patients with an EQ-5D score between 0.25 and 0.5). MAE was smaller for patients with observed EQ-5D scores between 0.5 and 1.0 (MAE ranging from 0.08 to 0.11). No patient was predicted to have an EQ-5D score below 0.25. Between 0.3% and 1.7% of patients were predicted to have an EQ-5D score between 0.25 and 0.5; between 34.3% and 41.9% of patients were predicted to have an EQ-5D score between 0.5 and 0.75; between 57.8% and 64.0% of patients were predicted to have an EQ-5D score above 0.75. Table A1 in the appendix shows the comparison between OLS models using the UK value set.

Similarly, item-level models 6 and 7 showed best overall model performance in terms of better MAEs and pseudo-R^2 statistics for the two-part, Tobit, and GEE regression models (data available upon request). Since model 7 ensures that item-level coefficients are always ordered, that is, the item coefficient size increases or decreases by level, model 7 was selected as the best-fitting model in all tested model classes.

Model 7 performance statistics for all regression specifications (i.e., OLS, two-part, Tobit, and GEE), together with results for the exploratory three-part and two-part beta model are presented in Table 3. Overall, there were minor differences in model performance between the different regression models. The comparisons between observed and predicted values by levels of utility show that utilities <0.5 were overpredicted, while utilities >0.75 were slightly underpredicted. This pattern remained, even in the three-part model and was similar to what was observed for the other regression types. The range of predicted values estimated by the three-part model was 0.44 to 0.95 while the observed range was 0.1 to 1 and the range predicted by the

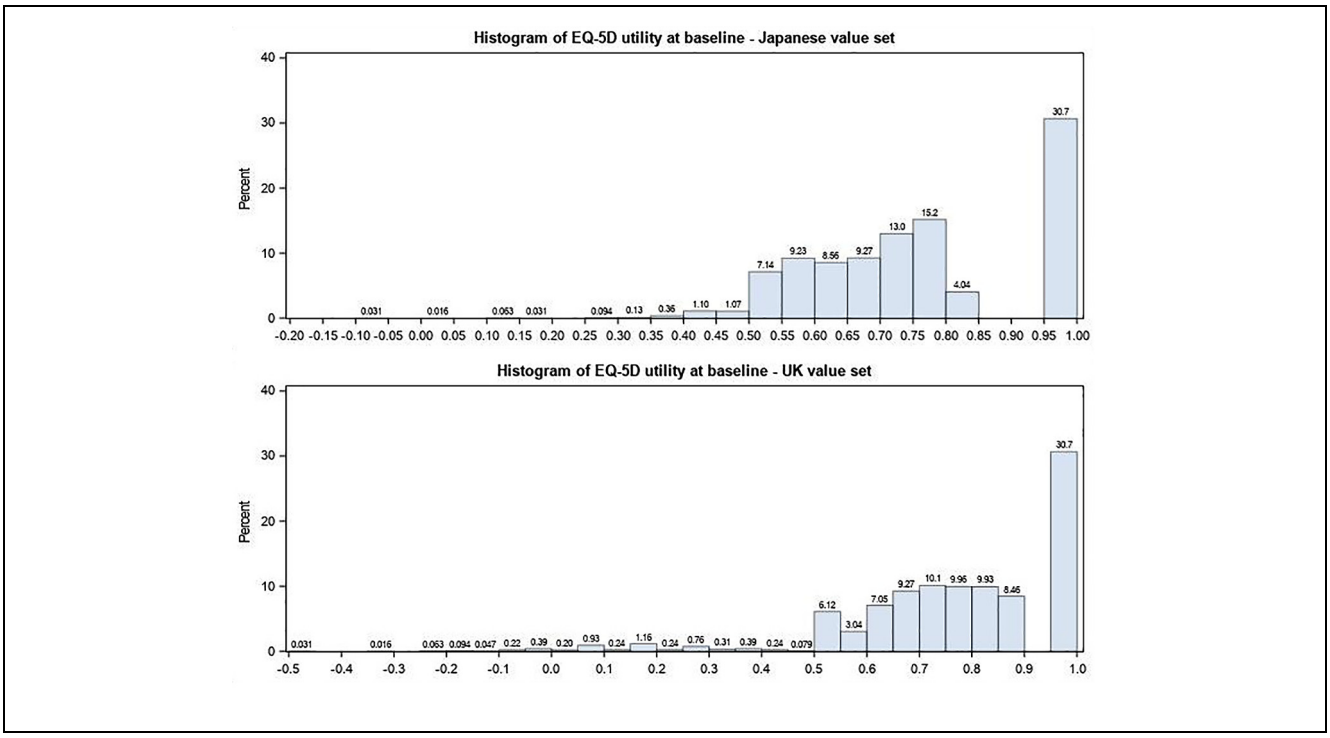

Figure 1 Histogram of EQ-5D utility at baseline.

OLS model was 0.44 to 0.96. Table A2 in the appendix shows the corresponding results for the UK value set.

Table 4 presents the regression coefficients for the best-fitting OLS model 7. A plot of observed versus predicted utility scores in the validation sample (n = 1,892) for OLS model 7 is shown in Figure 2.

Table 5 presents model performance statistics for the sensitivity analyses, where OLS model 7 was fitted to the larger sample including patients with an LVEF between 35% and 40%, and to the subsample of Asian patients, respectively. Overall, model fit statistics for the first sensitivity analysis were very similar to the main analysis. In contrast, there were some differences observed in the Asian patient subgroup analysis. Asian patients had higher utility on average, with a difference close to being clinically relevant (>0.064[33]), and model performance statistics were less favorable than in the main analysis; for example, the MAE was slightly larger (0.115 v. 0.1) and the pseudo-R^2 lower (0.38 v. 0.49).

For the UK value set, GEE model 7 was considered the best-fitting model, as it had a lower MAE than the other models. The regression coefficients for GEE model 7 using the UK value set are shown in Table A3 in the appendix. OLS model 7 using the UK value set is shown in Table A4 in the appendix.

Discussion

To address a common problem in HF studies, where the disease-specific KCCQ is collected and patients' utilities are not, we developed mapping algorithms from the KCCQ onto the EQ-5D-3L. Using a large contemporary trial of HFrEF patients, we used Japanese and the UK value sets to estimate the EQ-5D utilities using a number of alternative statistical methods. All models performed well in the prediction of both mean and median utility, which is what is often the central component of cost-utility analyses. In general, model performance and predictive power was better for the item-level models than for the models including KCCQ domain scores as covariates, with only small differences between the different model classes. Model performance was similar for the UK and Japanese value sets. OLS model 7 was selected as the best-fitting model for the Japanese and the UK value set; while for the UK value set GEE model 7 was similarly good.

To appropriately map disease-specific instruments to preference-based measures, it is required that the two instruments are overlapping with regards to the health domains assessed.[34] If the dimensions assessed by the EQ-5D are not covered by the disease-specific instrument, then

Table 2 Summary of Observed and Predicted Values and Model Performance Statistics per OLS Models (Japanese Value Set)[a]

		OLS 1	OLS 2	OLS 3	OLS 4	OLS 5	OLS 6	OLS 7
	Observed	Total Score	Domain Scores	Significant Domains	Significant Domains and Squared Terms	Significant Domains Plus Demographic and Clinical Terms	Significant Item Levels	Significant Item Levels With Collapsed Unordered Items
Mean (SD)	0.778 (0.173)	0.775 (0.114)	0.775 (0.115)	0.775 (0.115)	0.775 (0.117)	0.774 (0.117)	0.773 (0.119)	0.774 (0.118)
Median	0.768	0.799	0.796	0.797	0.785	0.795	0.781	0.781
Range	0.1–1	0.38–0.93	0.35–0.95	0.35–0.94	0.46–0.98	0.33–0.96	0.42–0.97	0.44–0.96
MAE	—	0.106	0.105	0.105	0.101	0.102	0.099	0.1
RMSE	—	0.126	0.125	0.125	0.124	0.123	0.124	0.124
Pseudo R^2	—	0.456	0.464	0.463	0.478	0.476	0.493	0.490
Adjusted R^2	—	0.456	0.463	0.463	0.477	0.475	0.489	0.488
AIC	—	−13911	−13967	−13963	−14082	−14057	−14162	−14156
BIC	—	−18365	−18389	−18391	−18485	−18441	−18405	−18462

	Observed	OLS 1		OLS 2		OLS 3		OLS 4		OLS 5		OLS 6		OLS 7	
	Mean	Mean	MAE	Mean	MAE	Mean	MAE	Mean	MAE	Mean	MAE	Mean	MAE	Mean	MAE
Level 1: −0.111 to 0.25 (N = 4)	0.15	0.6	0.45	0.59	0.439	0.59	0.44	0.64	0.486	0.57	0.421	0.64	0.484	0.63	0.48
Level 2: 0.25−0.5 (N = 43)	0.43	0.57	0.144	0.57	0.144	0.57	0.144	0.59	0.155	0.57	0.142	0.59	0.159	0.6	0.164
Level 3: 0.5−0.75 (N = 885)	0.64	0.71	0.092	0.71	0.092	0.71	0.092	0.71	0.084	0.71	0.087	0.7	0.08	0.7	0.08
Level 4: 0.75−1.0 (N = 960)	0.92	0.84	0.115	0.84	0.114	0.84	0.114	0.85	0.112	0.85	0.113	0.85	0.113	0.85	0.113

AIC, Akaike information criterion; BIC, Bayesian information criterion; MAE, mean absolute error; OLS, ordinary least squares; RMSE, root mean squared error; SD, standard deviation.

[a]Mapping models were fitted using baseline data.

Table 3 Summary of Observed and Predicted Values and Model Performance Statics per Best-Fitting Model 7 (Japanese Value Set)[a]

	Observed	OLS Model 7		Two-Part Model 7		Tobit Model 7		GEE Model 7		Three-Part Model 7		Two-Part Beta Model 7	
Mean (SD)	0.778 (0.173)	0.774 (0.118)		0.775 (0.117)		0.784 (0.127)		0.778 (0.118)		0.774 (0.117)		0.774 (0.117)	
Median	0.768	0.781		0.781		0.799		0.787		0.777		0.779	
Range	0.1–1	0.44–0.96		0.45–0.95		0.46–0.97		0.42–0.95		0.44–0.95		0.43–0.95	
MAE	—	0.1		0.099		0.1		0.099		0.099		0.099	
RMSE	—	0.124		0.124		0.125		0.121		0.123		0.124	
Pseudo R^2-	—	0.49		0.486		0.478		0.52		0.487		0.486	
AIC	—	−14156		−3089		−73		12932		[b]		−3939	
BIC	—	−18462		−2821		93		13127		[b]		−3666	
	Mean	**Mean**	**MAE**	**Mean**	**MAE**	**Mean**	**MAE**	**Mean**	**MAE**	**Mean**	**MAE**	**Mean**	**MAE**
Level 1: −0.111 to 0.25 (N = 4)	0.15	0.63	0.48	0.62	0.471	0.64	0.483	0.56	0.514	0.63	0.475	0.63	0.472
Level 2: 0.25−0.5 (N = 43)	0.43	0.6	0.164	0.6	0.165	0.6	0.162	0.59	0.155	0.6	0.167	0.6	0.163
Level 3: 0.5−0.75 (N = 885)	0.64	0.7	0.08	0.7	0.08	0.71	0.088	0.7	0.082	0.7	0.079	0.7	0.079
Level 4: 0.75−1.0 (N = 960)	0.92	0.85	0.113	0.85	0.113	0.86	0.107	0.85	0.109	0.85	0.113	0.85	0.113

AIC, Akaike information criterion; BIC, Bayesian information criterion; GEE, generalized estimating equations; MAE. mean absolute error; OLS, ordinary least squares; RMSE, root mean squared error; SD, standard deviation.

[a]GEE model was fitted using data collected at baseline, month 4, and month 8. All other mapping models were fitted using baseline data.

[b]Not presented as it is not easily computed.

the mapping may be compromised. Overall, the dimensions captured by the EQ-5D and the KCCQ were sufficient to enable good estimation for the EQ-5D from the KCCQ and likely reflects that the KCCQ includes items on physical limitations, social limitations, and mental health (for instance the item on enjoyment of life or "feeling discouraged or down") and that HF is a dominating condition in many patients' overall health. The validity of our model selection process is supported by the fact that our best-fitting model includes items covering these three domains; three items from the physical limitation scale and one each from the symptom, social limitation, quality of life, as well as the symptom stability item. This is congruent with previous studies that found the EQ-5D to give most weight to physical functioning.[15,35] Moreover, our mapping models had an R^2 close to 0.5, which is similar to findings reported in previous studies mapping from disease-specific questionnaires to the EQ-5D.[36,37] Goodness-of-fit statistics were also similar to those reported for a published mapping algorithm from the MacNew Heart Disease Quality of Life Questionnaire to the EQ-5D (R^2: 0.54; MAE: 0.113).[18] In a review conducted by Brazier et al., it was found that R^2 statistics for such models typically range from 0.2 to 0.5.[38]

We considered any of the KCCQ items as a potentially relevant covariate in our mapping models, including the symptom stability item and the self-efficacy scale. There is some evidence that the self-efficacy domain measures a different concept than the ways in which HF affects patients' health, and this domain has lower internal consistency than the other KCCQ scales.[20,39] The symptom stability item differs from the other items of the KCCQ in that it evaluates changes in symptoms over time rather than assessing patients' health status in a cross-sectional way. These two domains are not part of any of the KCCQ summary scales. However, we decided to not exclude any of the KCCQ items a priori, and thus included these items in our mapping models if they improved model fit. In fact, the symptom stability item was selected in the final best-fitting model for the Japan value set, as it improved model fit based on the BIC and likely reflects that if patients experience a recent improvement or deterioration in their health status, that this recent change also affects their self-rated utility scores.

Empirical EQ-5D data are known to have several idiosyncrasies, such as the fact that there is an upper bound at 1, which may question the assumption of homoscedastic error terms. Also, these data are often highly skewed and typically exhibit a pronounced ceiling effect, with a substantial number of patients having a

Table 4 Coefficients for the Best-Fitting OLS Model 7 (Japanese Value Set)[a]

Domain Item	Item Level	Coefficient (SD)	95% CI	*P* Value
Intercept				
Intercept		0.9572 (0.0073)	0.9429, 0.9714	<0.001
Physical limitation				
How limited ability to doing gardening, housework or carrying groceries	Extremely limited	−0.0879 (0.0108)	−0.1091, −0.0667	<0.001
	Quite a bit/moderately limited	−0.0649 (0.0069)	−0.0784, −0.0514	<0.001
	Slightly limited	−0.037 (0.0058)	−0.0484, −0.0256	<0.001
	Limited for other reasons or did not do the activity	−0.0744 (0.0126)	−0.0991, −0.0497	<0.001
	Not at all limited	0 (0)	0, 0	<0.001
How limited ability to dressing yourself	Extremely/quite a bit/moderately/slightly limited	−0.0476 (0.0051)	−0.0575, −0.0376	<0.001
	Limited for other reasons or did not do the activity	−0.0509 (0.0213)	−0.0926, −0.0092	0.0168
	Not at all limited	0 (0)	0, 0	<0.001
How limited ability to jogging or hurrying (as if to catch a bus)	Extremely/quite a bit limited	−0.034 (0.0062)	−0.0462, −0.0218	<0.001
	Moderately limited	−0.0161 (0.0061)	−0.028, −0.0042	0.0081
	Limited for other reasons or did not do the activity	−0.0442 (0.0086)	−0.061, −0.0274	<0.001
	Slightly/not at all limited	0 (0)	0, 0	<0.001
Quality of life				
Felt discouraged or down in dumps	All of the time	−0.1194 (0.0179)	−0.1545, −0.0843	<0.001
	Most of the time	−0.0941 (0.0087)	−0.1112, −0.0771	<0.001
	Occasionally	−0.0671 (0.0056)	−0.0781, −0.0561	<0.001
	Rarely felt that way	−0.0305 (0.0051)	−0.0405, −0.0204	<0.001
	Never felt that way	0 (0)	0, 0	<0.001
Social limitation				
How does HF affect lifestyle—visiting family or friends	Extremely limited	−0.083 (0.0127)	−0.1078, −0.0582	<0.001
	Quite a bit limited	−0.0648 (0.0094)	−0.0832, −0.0464	<0.001
	Moderately/slightly limited	−0.0345 (0.0052)	−0.0447, −0.0242	<0.001
	Limited for other reasons or did not do the activity	−0.0614 (0.0118)	−0.0845, −0.0383	<0.001
	Not at all limited	0 (0)	0, 0	<0.001
Symptom burden				
How much has your fatigue bothered you	Extremely bothersome	−0.1076 (0.0174)	−0.1417, −0.0735	<0.001
	Quite a bit/moderately bothersome	−0.0698 (0.0073)	−0.084, −0.0555	<0.001
	Slightly bothersome	−0.0415 (0.0065)	−0.0543, −0.0288	<0.001
	I've had no fatigue	0.0077 (0.0065)	−0.005, 0.0204	0.2372
	Not at all bothersome	0 (0)	0, 0	<0.001
Symptom stability				
Have symptoms of heart failure changed	Much worse	−0.065 (0.0267)	−0.1174, −0.0126	0.015
	Slightly worse/not changed/slightly better	−0.0362 (0.0059)	−0.0477, −0.0247	<0.001
	I've had no symptoms over the last 2 weeks	−0.0083 (0.0067)	−0.0215, 0.0049	0.2181
	Much better	0 (0)	0, 0	<0.001

CI, confidence interval; SD, standard deviation.

[a]Mapping model was fitted using baseline data.

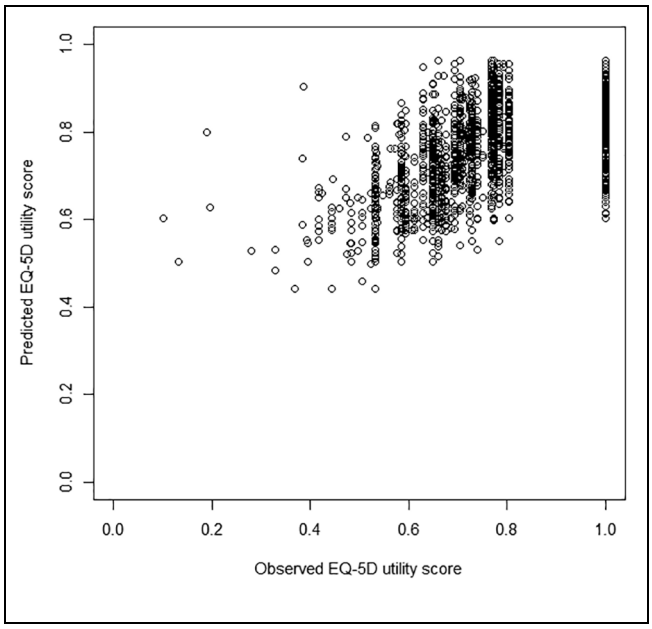

Figure 2 Plot of observed versus predicted EQ-5D utility at baseline in the validation sample (n = 1,892)—Japanese value set.

utility value of 1. Whereas these characteristics may challenge the use of OLS regression from a conceptual point of view, our results indicate that OLS regression models do not perform worse than alternative regression models that more explicitly address specific characteristics of the EQ-5D distribution, such as two-part or Tobit models. This was also observed in other published studies mapping disease-specific measures to the EQ-5D. For example, in Young et al., the OLS model predicted EQ-5D utility from the FACT-G questionnaire better than the Tobit or two-part model.[26] In addition, model diagnostic plots created for the OLS models did not show issues with homoscedasticity and nonnormally distributed error terms, in particular for the selected item-level models. Given that there were only minor differences between model classes in terms of predictive ability, it was decided to favor mapping algorithms based on the less complex OLS model over more complex two-part and Tobit models. This can not only be justified in terms of model parsimony (in that the two-part model typically has about twice the number of regression coefficients than the OLS model) but also regarding future use of the mapping algorithm for other studies and by other researchers.

Table 5 Summary of Observed and Predicted Values and Model Performance Statics for Model 7 (Sensitivity Analyses; Japanese Value Set)

	Population Including Patients With LVEF 35% to 40%				**Subsample of Asian Patients**		
	Observed	**OLS Model 7**			**Observed**	**OLS Model 7**	
Mean (SD)	0.774 (0.174)	0.77 (0.118)			0.832 (0.165)	0.814 (0.098)	
Median	0.741	0.776			0.785	0.822	
Range	0.1–1	0.45–0.96			0.418–1	0.57–0.93	
MAE	—	0.1			—	0.115	
RMSE	—	0.125			—	0.135	
Pseudo R-sq	—	0.484			—	0.379	
AIC	—	−15964			—	−1803	
BIC	—	−20844			—	−2365	
	Mean	**Mean**	**MAE**		**Mean**	**Mean**	**MAE**
Level 1: −0.111 to 0.25 (N = 4)	0.15	0.63	0.48				
Level 2: 0.25−0.5 (N = 56)	0.43	0.59	0.162	Level 2: 0.25−0.5 (N = 1)	0.42	0.69	0.268
Level 3: 0.5−0.75 (N = 1,029)	0.64	0.7	0.08	Level 3: 0.5−0.75 (N = 89)	0.65	0.74	0.112
Level 4: 0.75−1.0 (N = 1,072)	0.92	0.85	0.114	Level 4: 0.75−1.0 (N = 166)	0.93	0.86	0.117

AIC, Akaike information criterion; BIC, Bayesian information criterion; LVEF, left ventricular ejection fraction; MAE, mean absolute error; OLS, ordinary least squares; RMSE, root mean squared error; SD, standard deviation.

Predicted values from the OLS can easily be obtained as a linear combination of covariate values and regression coefficients, whereas more complex calculations would be needed for predictions based on the two-part and Tobit models.

The scoring rules for the KCCQ specify that the physical limitation scale is set to missing if a patient states that he or she was "limited for other reasons or did not do the activity" for at least four out of the six items. Similarly, the social limitation scale is set to missing if a patient was "limited for other reasons or did not do the activity" for three out of the four items in this scale. For the actual single-item score, this response is considered as a separate category, not as a missing value. As a consequence, these patients would be excluded in a complete case analysis using KCCQ domain scores, but not in a complete-case analysis using single-item scores. To ensure that all mapping models tested were fitted to the same patient sample, we excluded patients with missing values in the physical (0.07% of initial sample) or social (2.7% of initial sample) limitation score. An alternative would have been to keep them in the dataset, but to add an additional missing value indicator in the model; this would have allowed applying a domain score mapping model even to patients with missing values in the target sample. Since our best-fitting models are single-item models (where such missing values do not occur) this potential limitation is circumvented.

The mapping algorithm predicted low utilities poorly, which is a common problem with mapping algorithms. Methods to address underprediction were applied to investigate whether predictive accuracy could be improved.[40,41] Despite applying the three-part model, predictions of low utility values were similar to all other regression models. This could possibly be a result of the low frequency of patients with low utilities in the data. Whereas this means that our mapping models are appropriate to predict overall utility in patient populations, similar to the PARADIGM-HF trial, caution is warranted when computing individual predictions for patients in severe health states, or group means for patient populations with substantially more impairments. Nevertheless, for the intended application of these techniques to estimate the cost-utility of new treatments using the KCCQ, the mean values are most important and the failure to estimate the utilities at the extremes are less important.

The mapping algorithm used data from the PARADIGM-HF trial in which patients were enrolled globally. Guidelines for developing mapping algorithms recommend that the target population where the mapping algorithm is to be used should be similar to the sample in which the algorithm was developed.[16] Sensitivity analysis showed that the best-fitting model in the Asian subsample had slightly worse predictive properties than the best-fitting model in the overall population. This could be a result of the reduced sample size as well as the impact of ethnicity. While it is not possible to directly assess the overall impact of ethnicity, it should also be noted that any potential impact of ethnicity in the data used to develop our mapping algorithm cannot be extended to a Japanese population per se, since extremely few Japanese patients were enrolled in the PARADIGM-HF trial.

Strengths of this analysis include the thorough and systematic testing of different model specifications for the mapping algorithm and the large sample size enabling us to develop the mapping algorithm using the derivation sample, and the use of a validation sample to test the predictive accuracy. The best-fitting model was refitted using 10-fold cross-validation, and results suggested that the different splits had little impact on overall predictive ability of the mapping algorithm. To our knowledge, this is the first mapping algorithm of the KCCQ to EQ-5D-3L. Potential weaknesses of this approach include the use of a clinical trial with explicit inclusion and exclusion criteria and validation in a broader clinical population is warranted. In order to use the same dataset across all models fitted, we excluded patients with missing values on KCCQ and EQ-5D-3L or any of the relevant baseline characteristics. It is possible that these patients are different from the ones included in our final datasets; however, the comparison of EQ-5D utility and KCCQ subscale scores between the two groups did not reveal any systematic difference.

In conclusion, we developed a mapping algorithm of KCCQ to EQ-5D-3L for HF patients that may be used when trials have included the KCCQ and there is a need to derive preference-based utility values. OLS model 7 with individual questions was considered the best-fitting model for the Japanese value set, and GEE model 7 for the UK value set. This study facilitates cost-utility analysis of interventions in heart failure in Japan and the UK and may prove useful in future studies of the cost-effectiveness of care in patients with HFrEF.

ORCID iD

Jennifer Eriksson https://orcid.org/0000-0001-5652-969X

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making Policy & Practice* website at https://journals.sagepub.com/home/mpp.

References

1. Moran AE, Forouzanfar MH, Roth GA, et al. The global burden of ischemic heart disease in 1990 and 2010: the Global Burden of Disease 2010 study. *Circulation*. 2014; 129(14):1493–501.
2. Huffman MD, Berry JD, Ning H, et al. Lifetime risk for heart failure among white and black Americans: cardiovascular lifetime risk pooling project. *J Am Coll Cardiol*. 2013;61(14):1510–7.
3. Okura Y, Ramadan MM, Ohno y, et al. Impending epidemic: future projection of heart failure in Japan to the year 2055. *Circ J*. 2008;72(3):489–91.
4. Shiba N, Shimokawa H. Chronic heart failure in Japan: implications of the CHART studies. *Vasc Health Risk Manag*. 2008;4:103-113.
5. Pocock SJ, Ariti CA, McMurray JJV, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J*. 2013;34(19):1404–13.
6. Cook C, Cole G, Asaria P, Jabbour R, Francis DP. The annual global economic burden of heart failure. *Int J Cardiol*. 2014;171(3):368–76.
7. Juenger J, Schellberg D, Kraemer S, et al. Health related quality of life in patients with congestive heart failure: comparison with other chronic diseases and relation to functional variables. *Heart*. 2002;87(3):235–41.
8. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail*. 2016;18(8):891–975.
9. Butler J, Fonarow GC, Zile MR, et al. Developing therapies for heart failure with preserved ejection fraction: current state and future directions. *JACC Heart Fail*. 2014; 2(2):97–112.
10. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72.
11. Knies S, Evers SMAA, Candel MJJM, Severens JL, Ament AJHA. Utilities of the EQ-5D: transferable or not? *Pharmacoeconomics*. 2009;27(9):767–79.
12. Center for Outcomes Research and Economic Evaluation for Health; National Institute of Public Health. Guideline for preparing cost-effectiveness evaluation to the Central Social Insurance Medical Council [cited July 29, 2019]. Available from: https://c2h.niph.go.jp/tools/guideline/guideline_en.pdf
13. Center for Outcomes Research and Economic Evaluation for Health; National Institute of Public Health. Full scale introduction of cost-effectiveness evaluations in Japan. Overview (2/20/2019). Available from: https://c2h.niph.go.jp/tools/system/overview_en.pdf
14. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care*. 1989;27(3 Suppl):S217–S232.
15. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value Health*. 2013;16(1):202–10.
16. Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to estimate health-state utility from non-preference-based outcome measures: an ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health*. 2017; 20(1):18–27.
17. 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(1):31.
18. Chen G, McKie J, Khan MA, Richardson JR. Deriving health utilities from the MacNew Heart Disease Quality of Life Questionnaire. *Eur J Cardiovasc Nurs*. 2015;14(5):405–15.
19. Edlin R, Tsuchiya A, Brazier J. Mapping the Minnesota living with heart failure questionnaire to the EQ-5D index. Report to Novartis; 2002.
20. Green CP, Porter CB, Brensnahan 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(5):1245–55.
21. Tsuchiya A, Ikeda S, Ikegami N, et al. Estimating an EQ-5D population value set: the case of Japan. *Health Econ*. 2002;11(4):341–53.
22. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–108.
23. McMurray JJV, Packer M, Desai AS, et al; PARADIGM-HF Investigators and Committees. Angiotensin–neprilysin inhibition versus enalapril in heart failure. *N Engl J Med*. 2014;371(11):993–1004.
24. Lewis EF, Claggett BL, McMurray JJV, et al. Health-related quality of life outcomes in PARADIGM-HF. *Circ Heart Fail*. 2017;10(8):e003430.
25. White H. Using heteroscedasticity consistent standard errors in the linear regression model. *Econometrica*. 1980; 48:817–38.
26. Young TA, Mukuria C, Rowen D, Brazier JE, Longworth L. Mapping functions in health-related quality of life: mapping from two cancer-specific health-related quality-of-life instruments to EQ-5D-3L. *Med Decis Making*, 2015;35(7): 912–26.
27. Bansback N, Marra C, Tsuchiya A, et al. Using the health assessment questionnaire to estimate preference-based single indices in patients with rheumatoid arthritis. *Arthritis Rheum*. 2007;57(6):963–71.
28. 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.
29. Rivero-Arias O, Ouellet M, Gray A, Wolstenholme J, Rothwell PM, Luengo-Fernandez R. Mapping the modified Rankin scale (mRS) measurement into the generic EuroQol (EQ-5D) health outcome. *Med Decis Making*. 2010;30(3):341–54.
30. Rowen D, Brazier J, Roberts J. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health Qual Life Outcomes*. 2009;7:27.
31. Longworth L, Rowen D.NICE DSU technical support document 10: the use of mapping methods to estimate health state utility values. Report by the Decision Support Unit [cited July 29, 2019]. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD-10-mapping-FINAL.pdf
32. Basu A, Manca A. Regression estimators for generic health-related quality of life and quality-adjusted life years. *Med Decis Making*. 2012;32(1):56–69.
33. Shiroiwa T, Fukuda T, Ikeda S, et al. Japanese population norms for preference-based measures: EQ-5D-3L, EQ-5D-5L, and SF-6D. *Qual Life Res*. 2016;25(3):707–19.
34. Round J, Hawton A. Statistical alchemy: conceptual validity and mapping to generate health state utility values. *Pharmacoecon Open*. 2017;1(4):233–9.
35. Meregaglia M, Borsoi L, Cairns J, Tarricone R. Mapping health-related quality of life scores from FACT-G, FAACT, and FACIT-F onto preference-based EQ-5D-5L utilities in non-small cell lung cancer cachexia. *Eur J Health Econ*. 2019;20:181–93.
36. Askew RL, Swartz RJ, Xing Y, et al. Mapping FACT-melanoma quality-of-life scores to EQ-5D health utility weights. *Value Health*. 2011;14(6):900–6.
37. Teckle P, McTaggart-Cowan H, Van der Hoek K, et al. Mapping the FACT-G cancer-specific quality of life instrument to the EQ-5D and SF-6D. *Health Qual Life Outcomes*. 2013;11:203.
38. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*. 2010;11(2):215–25.
39. Spertus JA, Jones PG. Development and validation of a short version of the Kansas City Cardiomyopathy Questionnaire. *Circ Cardiovasc Qual Outcomes*. 2015;8(5):469–76.
40. Dakin H, Gray A, Murray D. Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. *Qual Life Res*. 2013;22(3):683–94.
41. Versteegh MM, Rowen D, Brazier JE, Stolk EA. Mapping onto Eq-5 D for patients in poor health. *Health Qual Life Outcomes*. 2010.;8:141.