Mapping for the EQ-5D-5L for Use in Cost-Utility Analysis

by

Jiabi Wen

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Health Policy Research

School of Public Health
University of Alberta

Abstract

Cost-utility analysis (CUA) assesses the cost-effectiveness of health technologies by comparing their costs and health outcomes. The utility is incorporated in the health outcome measures of CUA, and the EQ-5D-5L is one of the most common instruments to estimate utility values. When utility values are not available, mapping from non-preference-based instruments to a preference-based instrument is a popular technique. When CUAs use different preference-based measures, mapping between these measures can transfer the utility values and allow for better comparison across CUAs. However, there are many concerns regarding studies reporting mapped utility values, such as extrapolation issues and the uncertainty of this methodology. The quality of mapping studies has become an important criterion when using them in economic evaluations. The first study of my thesis assessed the reporting quality of mapping studies onto the EQ-5D-5L, especially their completeness of information for CUA applications. The second study developed a novel mapping algorithm from the Edmonton Symptom Assessment System Revised: Renal (ESAS-r: Renal) to the EQ-5D-5L among patients with end-stage renal disease (ESRD).

The first objective of my systematic review was to identify new mapping studies onto the EQ-5D-5L by updating a previous systematic search made by the Health Economics Research Centre (HERC). The second objective was to assess all the EQ-5D-5L mapping studies on their reporting quality, especially the completeness of information for CUA, with the use of two reporting quality checklists. The third objective was to explore whether using reporting quality checklists was associated with improved reporting quality. The review identified 14 new studies since 2018 which were not included in the HERC database. In the assessment of all 39 published studies (including 25 from the HERC database), the overall reporting quality was mostly good.

In several areas I identified problems that would require improvements including 1) estimation of predicted utilities, 2) reporting variances, covariances, and error terms, 3) final model calculation examples, 4) parameter uncertainty, and 5) individual uncertainty. A preliminary comparison showed that the checklists could help to improve the reporting quality of the studies.

The second study of this thesis mapped the ESAS-r: Renal to the EQ-5D-5L in patients with ESRD using data from the Evaluation of Routinely Measured Patient-reported Outcomes in Hemodialysis Care (EMPATHY) trial, a multi-centre clustered randomized-controlled trial of routine measurement of patient-reported outcomes in hemodialysis units in Northern Alberta. Several models were explored in the mapping analysis using data from one study arm, including linear models, censored dependent variable models, mixture models and response mapping. Internal validation was conducted to evaluate the predictive power of the models, and the validation sample was from another arm of the EMPATHY trial. Statistical fit and predictive power were measured by mean absolute error (MAE) and mean squared error (MSE), which, along with theoretical backgrounds, were the selection criteria for the best model. The final sample size for model estimation was 506, after excluding missing records (missing rate: 57.6%). All models produced relatively similar statistical fit and predictive power (Estimation: MAE: 0.056 - 0.120, MSE: 0.007 - 0.028; Validation: MAE: 0.136 - 0.155, MSE: 0.032 - 0.046). All models showed great prediction properties for relatively healthy health states, but poor prediction properties for worse health states. With the consideration of all selection criteria, the generalized estimating equations (Estimation: MAE: 0.120, MSE: 0.027; Validation: MAE:0.140, MSE:0.034) and generalized linear models (Estimation: MAE: 0.116, MSE: 0.028; Validation: MAE: 0.136, MSE: 0.034) on selected ESAS-r: Renal items were considered the best models.

Since the models have not been externally validated, they should be applied in populations with similar patient characteristics as our study sample.

Overall, mapping is a feasible and useful technique to estimate the utility values for conducting CUA. The issues identified in current mapping studies could inform further mapping studies on how to improve reporting quality, especially ensuring the completeness of information for employing mapping algorithms in CUA. The empirical mapping study on ESAS-r: Renal provided mapping algorithms which could be used to predict utility values for patients with ESRD when only ESAS-r: Renal is available.

Preface

This thesis is an original work by Jiabi Wen. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta (Health Research Ethics Board #Pro00077850).

Chapter 2 of this thesis is a collaborative work by me, Xuejing Jin, Jeffrey A. Johnson, and Arto Ohinmaa. JW conceived, designed, and executed the study. XJ, JAJ, and AO supervised the study, provided valuable insights, and assisted with manuscript revision.

Chapter 3 of this thesis is a collaborative work by me, Xuejing Jin, Fatima Al Sayah, Hilary Short, Jeffrey A. Johnson, and Arto Ohinmaa. JW performed data cleaning, analysis and drafted the manuscript. HS and JAJ designed, planned, and implemented the original trial, and HS finalized the trial data. FAS provided valuable insights into the study. XJ, JAJ, and AO supervised the study and assisted with manuscript revision. This work was presented at the 1st EuroQol Early Career Researcher Meeting, Prague, Czech Republic, March 1, 2020.

For my family

Acknowledgements

I would like to give a big thank you to my committee members. I appreciate my supervisor, Dr. Arto Ohinmaa, who has always provided helpful and timely advice and encouragement since I entered the program. I would like to thank the other committee members, Drs. Jeffrey A. Johnson, and Xuejing Jin, who offered valuable comments and feedback on my research directions.

I would like to acknowledge contributions of the following individuals for designing, planning, and implementing the EMPATHY Trial in Alberta Kidney Care - North: Robert Buzinski, Dr. Sara Davison, Paul Duperron (deceased), Dr. Jeffrey A. Johnson, Dr. Scott Klarenbach, Julie Nhan, Hilary Short, Betty A. Wasylynuk.

I also want to thank anyone who provides valuable insights into this thesis. Dr. Manikarnika Kanjilal, Dr. Negar Razavilar, and Angie Chiu were my mentors during my internship (Network of Alberta Health Economists Graduate Internship) at the Institute of Health Economics. They offered me many great suggestions for conducting the two studies of this thesis. Dr. Sentil Senthilselvan, Dr. Katherine Harback, Peiran Yao, and Xiaozhu Chuan helped me a lot with the statistical techniques. Specially, Peiran Yao assisted in proofreading my thesis.

I sincerely appreciate all institutes and agencies that have provided financial support to me, including the Alberta PROMs and EQ-5D Research and Support Unit (APERSU), the School of Public Health at the University of Alberta, the Faculty of Graduate Studies and Research at the University of Alberta, and the Institute of Health Economics. Without their support, I could not be able to pursue this degree.

I would like to thank my parents and my family for always supporting me and believing in me. Their encouragement for my study abroad was unwavering. I also want to thank my aunt and her family, who gave me enormous support and love when I am in Canada. Besides, I want to thank my friends, both in Alberta as well as those far away and across the world, for sharing the happiness and sorrow for the past years.

I must thank all the members from the APERSU and the Alliance for Canadian Health Outcomes Research in Diabetes (ACHORD). In particular, I want to acknowledge Dr. Fatima Al Sayah for her kind and precious advice, support, and guidance over the past two years. I also would like to

thank Sherry Lydynuik and Donna Pressick for their support and help in my every-day research life at APERSU and ACHORD.

Finally, I would like to thank the School of Public Health at the University of Alberta for giving me the platform of studying and making progress. I am thankful to all professors, staff, and peers as they always offered me lots of help and kindness.

Table of Contents

List of Tables

Chapter 2

Chapter 3

List of Figures

Chapter 2

Chapter 3

List of Appendices

Chapter 2

Chapter 3

List of Abbreviations

AD	Anxiety/Depression
AIC	Akaike Information Criterion
ALDVMM	Adjusted Limited Dependent Variable Mixture Model
BETAMIX	Mixture Beta Regression Model
BIC	Bayesian Information Criterion
CUA	Cost-Utility Analysis
CKD	Chronic Kidney Disease
CLAD	Censored Least Absolute Deviations
EMPATHY	Evaluation of routinely Measured PATient-reported outcomes in HemodialYsis care
EQ-5D	EuroQol-Five Dimensions
ESAS-r	Edmonton Symptom Assessment System Revised
ESAS-r: Renal	Edmonton Symptom Assessment System Revised-Renal
ESRD	End-Stage Renal Disease
FEM	Fixed Effect Model
GEE	Generalized Estimating Equations
GLM	Generalized Linear Model
HERC	Health Economics Research Centre
HTA	Health Technology Assessment
HRQL	Health-Related Quality of Life
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MAE	Mean Absolute Error
MAPS	MApping onto Preference-based measures reporting Standards

MO	MObility
MSE	Mean Squared Error
NICE	National Institute for health and Care Excellence
OLS	Ordinary Least Squares
PBM	Preference-Based Measure
PD	Pain/Depression
PROM	Patient-Reported Outcome Measure
QALY	Quality-Adjusted Life Years
REM	Random Effect Model
SC	Self-Care
TPM	Two-Part Model
UA	Usual Activities

Chapter 1. Introduction

1.1 Health Status Measures

An influential definition of health says "health is a state of complete physical, mental and social well-being, and not merely the absence of disease and infirmity", which is provided by the World Health Organization[1]. Health status and health-related quality of life (HRQL) are often used interchangeably, and they refer to the health domain ranging from negatively valued aspects of life, including death, to the more positively values aspects such as role function or happiness[2]. The health status measures typically use psychometric tools to create "constructs" or "domains of content" that have not had a basis in traditional patient assessment[3].

The recent history of health status measurement dates from the early 1970s[4]. The health status measures were motivated to measure the output of the whole health care system and identify changes in the level of health of the population[5]. Over the past years, the health status measures have gained importance because of a desire to measure and reflect the improvements in functional capacity and well-being[2,4]. The traditional physiologic measures provide information to clinicians but are of limited interest to patients, and they only reflect information on morbidity and biological functioning[2,6]. Another important reason to measure HRQL is that patients with the same clinical conditions could have substantially different HRQL[2]. Here is an example given by Guyatt et al, "two patients with the same range of motion and even similar ratings of back pain may have different role function and emotional well-being. Although some patients may continue to work without major depression, others may quit their jobs and have major depression." [2]. The health status measure which measures the health and HRQL from a more patient-focused perspective is known as patient-reported outcome measure (PROM)[7].

1.1.1 Generic Measures and Specific Measures

Health status measures can be classified as generic measures and specific measures. The generic health status measures are "those purport to be broadly applicable across types and severities of disease, across different medical treatments or health interventions, and across demographic and cultural subgroups"[8]. The generic measures are designed for summarizing a spectrum of the concepts of HRQL, and therefore they could be applied to many different impairments, illnesses, patients, and populations[8]. The specific health status measures are "those designed to assess

specific diagnostic groups or patient populations often with the goal of measuring responsiveness or 'clinically important' changes"[8]. The specific measures could be disease-specific (e.g. cancer) or condition-specific (e.g. back pain)[8].

1.1.2 Profile Measures and Index Measures

Another approach classifies the health status measures into profile measures and index measures. The profile measures are instruments that attempt to measure all important aspects of HRQL[9]. A profile measure provides a series of scores, one for each dimension of a patient's health status[9]. The dimensions (or domains) may include general health, physical functioning, social functioning, etc[9]. The index measures summarize the HRQL into a single number by using population preferences or other methods[2]. The preference-based index measures place a value on a specific health condition by reflecting the preferences (utilities) of patients for the treatment process and outcome[2]. The preference-based measure (PBM) could be used in the cost-utility analysis (CUA), and the more detailed introduction was given in Section 1.2.

1.2 Utility and Its Application in Economic Evaluation

Budget constraints evoked the need for evaluation of health technologies to assess their value for money[10]. Economic evaluations could give the answer of whether a health technology is less costly or not, and whether it has more health gains; in other words, its cost-effectiveness[11]. CUA has been the dominant method in economic evaluation[12], and also the recommended method in many countries' economic evaluation guidelines[13,14]. Just as the name suggests, the utility is an important parameter of CUA. In the following sections, the definition and applications of utility were introduced.

1.2.1 Utility Theory

The utility theory currently being used in health economics could be traced back to the 1940s[15], when John von Neumann and Oscar Morgenstern extended the "expected utility theory" as the decision making of a rational individual under undertainty[16]. The expected utility theory was a normative model that prescribed what the rational decision ought to be when an individual is facing uncertain outcomes[15]. The term "utility" tended to be synonymous with preference: the more preferable an outcome, the more utility associated with it[11]. But "utility" is a problematic term, as it did not mean usefulness in the typical use in language, and it also did not mean the

same as what it has meant in traditional economics and philosophy during the nineteenth century or in modern economics[15]. To avoid confusion, it is therefore recommended to use the term "von Neuman and Morgenstern utility" (vNM), but very few studies followed this[15]. In this thesis, the term "utility" represents "vNM utility".

The vNM utility is a cardinal utility[17]. Cardinal utility could tell how much more or less one utility is than another, in contrast to ordinal utility, which merely distinguishes the order of preference within the individual[18]. Applying cardinal utility to health status allows comparison across different individuals[15], and the individual utility could be summed together to elicit the social utility on a specific health status[19]. Finally, the utility for health status is conventionally anchored from 0 to 1, with 1 meaning perfect health, and 0 representing dead[19]. But it is generally accepted that there are some health states considered worse than dead[20], so it is possible to see negative values for utility.

1.2.2 Techniques for Eliciting Utility

Direct Approaches and Indirect Approaches

The PBMs elicit utility through direct or indirect approaches. In the direct approach, respondents state the preference for their current health states or over a variety of hypothetical health states from the perspective of themselves living in those states or someone else living in those states[20]. There are various methods for direct measurements, such as Standard Gamble (SG) and Time Trade-Off (TTO). SG is considered as the "gold standard" of eliciting utilities on health status under uncertainty as it is rooted in von Neumann and Morgenstern utility theory[21,22].

The indirect approach to elicit utility is to use multi-attribute utility measures. Developed from the multi-attribute utility theory[23], the estimation of utility in this approach has two steps: the first is to classify the health status, and the second is to generate the utility value using the pre-scored preference-based formula[20]. The classification system (or descriptive system) is an instrument that uses several domains or dimensions to define health, and these domains or dimensions are the attributes of the multi-attribute utility theory[24].

The advantage of using direct approaches is that every perspective related to health status is considered during elicitation[25]. When using a multi-attribute utility measure, its dimension may omit or have little coverage on some perspectives of health status, or include perspectives that

are not closely related to that health status[25]. However, direct approaches are time-consuming and costly to use[11]. Therefore, the indirect approaches using multi-attribute utility measures to provide values are more popular[11].

The EQ-5D

The EQ-5D is one of the most widely used multi-attribute utility measures[26], and it is a generic PROM measuring HRQL in terms of 5 dimensions: Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression[26,27]. The first version of the EQ-5D has 3 levels (3L) in each dimension for respondents to choose: no problems, some problems, and extreme problems[27]. During the use of the EQ-5D-3L, researchers found that the instrument lacks the ability to measure small changes in health in minor health problems, which is named as ceiling effect[26]. To deal with this problem, the 5-level version of the EQ-5D, the EQ-5D-5L, was developed. This new version has 5 response categories for each dimension: no problems, slight problems, moderate problems, severe problems, and unable to/extreme problems. Based on the EuroQol website (www.euroqol.org), the EQ-5D-3L and -5L have been valued in 34 and 21 countries, respectively. The preliminary EQ-5D-3L Canadian value set was published in 2012[28], and the EQ-5D-5L Canadian valuation study was published in 2016[29].

1.2.3 Quality-Adjusted Life Years and Cost-Utility Analysis

Any type of economic evaluation involves the assessment of costs and health effects of health technologies[11]. When it comes to CUA, health effects are measured by quality-adjusted life years (QALY)[21]. The QALY was designed when there was a need for indicators not only focusing on mortality but also morbidity[5]. QALY is adjusting someone's life expectancy based on the levels of HRQL experienced or predicted over the life-time or part of the life[30]. The QALY is determined as follows: $QALY\ lived\ in\ one\ year\ =\ 1 \times Q$, where Q is the HRQL weight attached to that one year of life[30]. The quality-adjusted life expectancy (QALE) would be the sum of QALYs in each year over the life-time, and if time preference is considered, a discounting term is incorporated to get discounted QALE[30]. The QALY model does not specify what the HRQL weights should be, but since it is used to assist decision-making on appropriate health technologies for groups of individuals, it is appealing to use weights reflecting preferences of the general population[15]. The utility-weighted QALY model is used, and the analysis is called CUA. Compared with other measures of health effects, the strengths of using QALY is that it

allows for a comparison of both quality and length of life on a wide range of therapeutic area, and the utility-weighted QALY can sit within an extra-welfarist framework of resource allocation[11,31].

1.3 Mapping to Obtain Utility Values

While the CUA requires the use of PBMs, such as the EQ-5D-5L as an essential input, many clinical studies do not include these measures as the outcome of trials[11]. It is more common to see clinical studies include disease-specific or non-preference-based generic measures. The limitation of these measures in economic evaluation is that they do not have an associated scoring algorithm to elicit utility values[11,32]. For example, the Edmonton Symptom Assessment System Revised-Renal (ESAS-r: Renal) is a validated disease-specific PROM to capture symptom burdens for patients with end-stage renal disease[33,34], but it does not have a preference-based scoring system. The reason for not using PBMs is that the descriptive system of PBMs may not sensitively capture changes in symptoms or HRQL associated with the disease[35]. Besides, clinical trials often do not plan a concurrent economic evaluation, and in this case, researchers usually do not consider including a PBM in the studies[36]. Therefore, there is a gap to conduct CUA using the outcomes from trials[32]. Mapping is considered as a solution to fill the gap by creating a link between the non-PBM and the PBM[32]. The mapping analysis may also be conducted to convert one PBM to another; this may be done in cases where jurisdictions have requirements for CUA to be undertaken with a specific PBM, such as with National Institute for Health and Care Excellence in the UK[13].

1.3.1 Definition of Mapping

Mapping is an approach involving the estimation of the relationship between a PBM and another PROM using statistical associations (also known as "cross-walking" or estimating exchange rates between instruments)[37]. The degree of overlap in descriptive systems of the PBM and another PROM makes it possible to convert the PROM to PBM with the use of regression techniques[11]. To conduct a mapping analysis, data are obtained from a population where both the PROM and the PBM are administrated[37]. The dataset being estimated for mapping functions is called the estimation sample[38]. Some studies also validate mapping functions in a different dataset, and this dataset is called the validation sample[38]. There are usually five elements of mapping: 1) defining the estimation dataset; 2) model specification; 3) model type (e.g. ordinary

least squares); 4) assessing performance (e.g. goodness of fit, predictive ability); 5) application[38]. Currently, there are three recommendations/checklists on conducting or reporting mapping analyses, which include the best practice recommendation developed by Longworth et al[38], the Mapping onto Preference-based Measures Reporting Standards (MAPS) Reporting Statement[39], and the ISPOR Good Practices Task Force Report on Mapping (ISPOR Good Practices)[14].

1.3.2 Previous Arguments on Mapping

Mapping has been gaining popularity in economic evaluation literature[11]. Up to January 2019, the Health Economics Research Centre (HERC) mapping study database had identified 182 papers conducting novel mapping analysis. It is commonly agreed that mapping meets the needs of using QALYs in appraising health technologies[11,32,37,38,40]. It is also important to note that using utilities estimated from PBMs is advantageous and mapping should be viewed as a second-best solution[38,41]. Previous studies have critiqued that mapping may underestimate individual uncertainty[40], and it might ignore the conceptual issues during the mapping process and when extrapolating the results[41]. To deal with these issues, the three recommendations/checklists recommend that mapping studies report detailed uncertainty information and conduct exploratory analysis to find overlaps in constructs[32,38,39]. But based on a recent reporting quality review, the seventeen papers (all published in 2016) being reviewed have mixed performance in reporting standard errors to reflect uncertainty and the methods of exploratory data analyses[42].

1.3.3 Possible Models Types for Mapping Analyses

The typical characteristics of utility distributions, e.g. skewness and multimodality[43], may need special regression techniques to deal with. Longworth recommendations suggested several options, including linear ordinary least squares, Tobit model, censored least absolute deviation, two-part model, generalized linear model, latent class mixture model, censored mixture model, and multinomial logit model[38]. The recommendations were published in 2013, and several new regression models have been applied in mapping analyses since then. In the remaining sections, a brief methodology review will be provided on the common models being used in mapping studies based on the HERC mapping database[44].

Ordinary Least Squares

Ordinary Least Squares (OLS) model is the most common technique that has been adopted in past mapping studies[42]. The OLS model predicts the linear relationship between the PBM-based utility value (outcome variable) and the summary index or domains/items of another PROM (explanatory variables) and estimates the parameters by the principle of least squares (Eq. 1).

$$\text{Utility} = \beta_0 + \sum_i \beta_i \text{PROM}_i \qquad \text{(Eq. 1)}$$

The linear relationship is simple and easy to interpret, where the deterioration in the quality of life captured by the PROM is linearly associated with the deterioration of the utility values. But there are many assumptions behind the OLS method, including mean independence, normality, and heteroskedasticity[45,46]. Another issue of OLS, which has been pointed out by many studies, is that the mapped utility value may over- or under-shoot the utility boundaries[45,47].

Tobit and Censored Limited Absolute Deviation

The Tobit model and the censored least absolute deviation models are censored dependent variable models and are applied in the mapping studies to solve the boundary issues[48,49]. In the context of health utilities, the upper and lower limits of the utility data are the constraints of the outcome variable. With additional restrictions added to the values hitting the two boundaries, the mapped utility values, therefore, lie within the range (Eq.2-3).

$$\text{Utility}^* = \beta_0 + \sum_i \beta_i \text{PROM}_i \qquad \text{(Eq.2)}$$

$$\text{Utility} = \begin{cases} \text{Utility}^*, & \text{if Utility}^* \text{lies within the range} \\ \text{upper/lower limits}, & \text{if Utility}^* \text{lies out of the range} \end{cases} \qquad \text{(Eq.3)}$$

The interpretation of parameters is the same as in the OLS. The simplicity of OLS is not deteriorated but the accuracy of the estimation of parameters has been increased. The difference between Tobit and CLAD is that Tobit has homoscedasticity and normality assumption but CLAD does not require them[48].

Two-Part Model

The two-part model (TPM) is another regression model for limited dependent variables[50]. Based on the framework of TPM[50], to estimate utility values, the first function estimates the probability of an individual having perfect health ($noted\ as\ \varphi(Utility)$). When the observation has perfect

health, the perfect health utility value is assigned. Otherwise, a linear model (denoted as $g(PROM)$) is used to estimate the utility values. The final estimated utility value is an expected mean:

$$E(Utility) = \varphi(Utility) \times perfect\ health\ utility + (1 - \varphi(Utility)) \times g(PROM) \quad \text{(Eq.4)}$$

Generalized Linear Models

The generalized linear models (GLMs) include a group of models where the outcome variable has a distribution belonging to the exponential family, and the outcome variable is explained by a linear function of explanatory variables[51]. Some distributions in the exponential family are similar to the distribution of utility or disutility (= perfect health utility - utility)[52,53]. The model equation is as follows, where $L(\cdot)$ is the link function for the exponential family.

$$L(Utility/Disutility) = \beta_0 + \sum_i \beta_i \mathrm{PROM}_i \quad \text{(Eq. 5)}$$

Mixture Models

The emerging new models on mapping studies fall into the mixture model category[54,55]. These models have been gaining popularity as they could solve the multimodality issues of the utility distribution[54]. Mixture models assume there are several sub-distributions for utility values. Mixture models estimate the probability of a respondent being in each sub-group and predict the group-specific utility values[54,55]. TPM is a special form of mixture model: one sub-distribution is perfect health, and another is the non-perfect health observations. There is no specific distribution requirement on mixture models, and the distributions commonly used are Tobit-like distribution and beta-binomial distribution[54,55]. Similar to TPM, the final utility predicted by the mixture model is the expectation of sub-group utilities.

Response Mapping

Response mapping is not a specific model type, and it simply means mapping from PROM items to the descriptive system of a PBM. The process of response mapping is independent of the value set of PBM, and therefore any value set could be used to generate utility given the estimated descriptive system[56]. The most common methods in response mapping are logistic models, either multinomial or ordinal[56,57]. Logistic regression models predict the relative probability of a respondents' choice on each level within a dimension. The probability of a health state would be

the product of dimensional probabilities. The final utility is a weighted average of utility values, where the weight is the probability of each health state.

Additional Options

There are some additional options when it comes to specific modelling situations. If the estimation sample is a longitudinal dataset, the cluster effect should be considered. Models for longitudinal data include generalized estimating equations (GEE)[58,59], fixed effect, random effect, and mixed effect models[60,61].

It is possible to directly apply the mean and standard deviation of utility values from the estimation sample to the target sample if two samples share very similar demographics and disease characteristics[32]. However, this is not often used as it is rare to see the patient populations across different trials to have similar outcomes and baseline information.

1.4 Outline of the Thesis

A gap in economic data can exist when preferred PBMs are not included in clinical trials. Given that mapping is the popular solution to this problem[32], it is necessary to ensure that the reporting quality of the mapping analysis is high, and the information for conducting CUA is complete. Chapter 2 is a systematic review of published mapping studies onto the EQ-5D-5L. A reporting quality assessment on these studies was conducted based on the MAPS reporting statement[39], and the ISPOR Good Practices[32]. Strengths and information gaps among the studies were summarized. The findings in Chapter 2 could be a reference to further empirical mapping studies on designing and conducting analysis and reporting results.

Chapter 3 is an empirical mapping study that applied the key findings of Chapter 2. In the context of end-stage renal disease, patients are usually associated with a reduction in life expectancy and quality of life[62]. When a new treatment for end-stage renal disease is available for patients, a CUA would be conducted to decide its value for money. Given that PBMs are not usually included in every study, Chapter 3 aimed to conduct a novel mapping analysis from the ESAS-r: Renal to the EQ-5D-5L following the two guidelines on mapping guidelines[32,39] when applicable.

Finally, Chapter 4 summarized the findings, discussed the limitations, and provided recommendations for future research.

References

1. World Health Organization. Constitution of the world health organization. Published online 1995.
2. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med*. 1993;118(8):622-629.
3. Kazis LE, Anderson JJ, Meenan RF. Effect Sizes for Interpreting Changes in Health Status. *Med Care*. 1989;27(3):S178-S189.
4. Karimi M, Brazier J. Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? *PharmacoEconomics*. 2016;34(7):645-649. doi:10.1007/s40273-016-0389-9
5. Fanshel S, Bush JW. A Health-Status Index and its Application to Health-Services Outcomes. *Oper Res*. 1970;18(6):1021-1066. doi:10.1287/opre.18.6.1021
6. Ware Jr JE. The status of health assessment 1994. *Annu Rev Public Health*. 1995;16(1):327-354.
7. Haywood KL, de Wit M, Staniszewska S, Morel T, Salek S. Developing patient-reported and relevant outcome measures. In: *Patient Involvement in Health Technology Assessment*. Springer; 2017:103-120.
8. Patrick DL, Deyo RA. Generic and Disease-Specific Measures in Assessing Health Status and Quality of Life. *Med Care*. 1989;27(3):S217-S232.
9. Jenkinson C. *Measuring Health And Medical Outcomes*. Routledge; 2013.
10. Harris A, Buxton M, O'Brien B, Rutten F, Drummond M. Using economic evidence in reimbursement decisions for health technologies: experience of 4 countries. *Expert Rev Pharmacoecon Outcomes Res*. 2001;1(1):7-12. doi:10.1586/14737167.1.1.7
11. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Fourth edition. Oxford, United Kingdom: Oxford University Press; 2015.
12. Wong JB. Principles of health economics and application to rheumatic disorders. In: *Rheumatology*. Elsevier; 2015:20-26.
13. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal 2013. National Institute for Health and Care Excellence (NICE)*. NICE; 2013. Accessed April 26, 2020. https://www.nice.org.uk/process/pmg9/chapter/the-reference-case#measuring-and-valuing-health-effects
14. Guidelines for the Economic Evaluation of Health Technologies: Canada. CADTH.ca. Published December 2, 2015. Accessed May 4, 2020. https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada
15. Torrance GW, Feeny D. Utilities and Quality-Adjusted Life Years. *Int J Technol Assess Health Care*. 1989;5(4):559-575. doi:10.1017/S0266462300008461

16. Von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. Princeton University Press; 1944:xviii, 625.

17. Fishburn PC. Retrospective on the Utility Theory of von Neumann and Morgenstern. *J Risk Uncertain*. 1989;2(2):127-157.

18. Bishop RL. Consumer's Surplus and Cardinal Utility. *Q J Econ*. 1943;57(3):421-449. doi:10.2307/1881755

19. Torrance GW. Measurement of health state utilities for economic appraisal: A review. *J Health Econ*. 1986;5(1):1-30. doi:10.1016/0167-6296(86)90020-2

20. Torrance GW, Furlong W, Feeny D. Health utility estimation. *Expert Rev Pharmacoecon Outcomes Res*. 2002;2(2):99-108. doi:10.1586/14737167.2.2.99

21. Gafni A, Birch S, Mehrez A. Economics, health and health economics: HYEs versus QALYs. *J Health Econ*. 1993;12(3):325-339. doi:10.1016/0167-6296(93)90015-7

22. Buckingham K, Devlin N. A theoretical framework for TTO valuations of health. *Health Econ*. 2006;15(10):1149-1154. doi:10.1002/hec.1122

23. Keeney RL, L KR, Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press; 1993.

24. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. *Pharmacoeconomics*. 1995;7(6):490-502.

25. Lipscomb J, Gotay CC, Snyder C. *Outcomes Assessment in Cancer: Measures, Methods and Applications*. Cambridge University Press; 2004.

26. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-1736. doi:10.1007/s11136-011-9903-x

27. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53-72. doi:10.1016/0168-8510(96)00822-6

28. Bansback N, Tsuchiya A, Brazier J, Anis A. Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. *PloS One*. 2012;7(2):e31115. doi:10.1371/journal.pone.0031115

29. Xie F, Pullenayegum E, Gaebel K, et al. A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. *Med Care*. 2016;54(1):98-105. doi:10.1097/MLR.0000000000000447

30. Sassi F. Calculating QALYs, comparing QALY and DALY calculations. *Health Policy Plan*. 2006;21(5):402-408. doi:10.1093/heapol/czl018

31. Round J. Is a QALY still a QALY at the end of life? *J Health Econ*. 2012;31(3):521-527. doi:10.1016/j.jhealeco.2012.01.006

32. 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. doi:10.1016/j.jval.2016.11.006

33. Davison SN, Jhangri GS, Johnson JA. Cross-sectional validity of a modified Edmonton symptom assessment system in dialysis patients: A simple assessment of symptom burden. *Kidney Int*. 2006;69(9):1621-1625. doi:10.1038/sj.ki.5000184

34. Davison SN, Jhangri GS, Johnson JA. Longitudinal validation of a modified Edmonton symptom assessment system (ESAS) in haemodialysis patients. *Nephrol Dial Transplant*. 2006;21(11):3189-3195. doi:10.1093/ndt/gfl380

35. Brazier JE, Rowen D, Mavranezouli I, et al. *Developing and Testing Methods for Deriving Preference-Based Measures of Health from Condition-Specific Measures (and Other Patient-Based Measures of Outcome).* NIHR Journals Library; 2012. Accessed April 26, 2020. http://www.ncbi.nlm.nih.gov/books/NBK100337/

36. Hughes D, Charles J, Dawoud D, et al. Conducting Economic Evaluations Alongside Randomised Trials: Current Methodological Issues and Novel Approaches. *PharmacoEconomics*. 2016;34(5):447-461. doi:10.1007/s40273-015-0371-y

37. 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):215-225. doi:10.1007/s10198-009-0168-z

38. Longworth L, Rowen D. Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value Health*. 2013;16(1):202-210. doi:10.1016/j.jval.2012.10.010

39. Petrou S, Rivero-Arias O, Dakin H, et al. The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. *PharmacoEconomics*. 2015;33(10):993-1011. doi:10.1007/s40273-015-0312-9

40. McCabe C, Edlin R, Meads D, Brown C, Kharroubi S. Constructing Indirect Utility Models: Some Observations on the Principles and Practice of Mapping to Obtain Health State Utilities. *PharmacoEconomics*. 2013;31(8):635-641. doi:10.1007/s40273-013-0071-4

41. Round J, Hawton A. Statistical Alchemy: Conceptual Validity and Mapping to Generate Health State Utility Values. *PharmacoEconomics - Open*. 2017;1(4):233-239. doi:10.1007/s41669-017-0027-2

42. 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):1-9. doi:10.1186/s12955-018-0857-3

43. Mihaylova B, Briggs A, O'Hagan A, Thompson SG. Review of Statistical Methods for Analysing Healthcare Resources and Costs. *Health Econ*. 2011;20(8):897-916. doi:10.1002/hec.1653

44. HERC database of mapping studies, Version 7.0 (last updated: 24th April 2019). http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies.

45. Beck A-JCC, Kieffer JM, Retèl VP, et al. Mapping the EORTC QLQ-C30 and QLQ-H&N35 to the EQ-5D for head and neck cancer: Can disease-specific utilities be obtained? *PLOS ONE*. 2019;14(12):e0226077. doi:10.1371/journal.pone.0226077

46. Ameri H, Yousefi M, Yaseri M, Nahvijou A, Arab M, Akbari Sari A. Mapping the cancer-specific QLQ-C30 onto the generic EQ-5D-5L and SF-6D in colorectal cancer patients. *Expert Rev Pharmacoecon Outcomes Res*. 2019;19(1):89-96. doi:10.1080/14737167.2018.1517046

47. Crott R. Mapping algorithms from QLQ-C30 to EQ-5D utilities: no firm ground to stand on yet. *Expert Rev Pharmacoecon Outcomes Res*. 2014;14(4):569-576. doi:10.1586/14737167.2014.908711

48. Chen S, Khan S. Estimating censored regression models in the presence of nonparametric multiplicative heteroskedasticity. *J Econom*. 2000;98(2):283-316. doi:10.1016/S0304-4076(00)00020-8

49. Amemiya T. Tobit models: A survey. *J Econom*. 1984;24(1):3-61. doi:10.1016/0304-4076(84)90074-5

50. Belotti F, Deb P, Manning WG, Norton EC. Twopm: Two-Part Models. *Stata J*. 2015;15(1):3-20. doi:10.1177/1536867X1501500102

51. Nelder JA, Wedderburn RWM. Generalized Linear Models. *J R Stat Soc Ser Gen*. 1972;135(3):370-384. doi:10.2307/2344614

52. 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):683-694. doi:10.1007/s11136-012-0189-4

53. Gu NY, Botteman MF, Ji X, Bell CF, Carter JA, van Hout B. Mapping of the Insomnia Severity Index and other sleep measures to EuroQol EQ-5D health state utilities. *Health Qual Life Outcomes*. 2011;9:119. doi:10.1186/1477-7525-9-119

54. Hernandez-Alava M, Wailoo A. Fitting Adjusted Limited Dependent Variable Mixture Models to EQ-5D. *The Stata Journal*, 15, Number 3, pp. 737–750.

55. Gray LA, Alava MH. A Command for Fitting Mixture Regression Models for Bounded Dependent Variables Using the Beta Distribution. *Stata J*. 2018;18(1):51-75. doi:10.1177/1536867X1801800105

56. van Hout B, Janssen MF, Feng Y-S, et al. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. *Value Health*. 2012;15(5):708-715. doi:10.1016/j.jval.2012.02.008

57. Yang F, Wong CKH, Luo N, Piercy J, Moon R, Jackson J. Mapping the kidney disease quality of life 36-item short form survey (KDQOL-36) to the EQ-5D-3L and the EQ-5D-5L in patients undergoing dialysis. *Eur J Health Econ*. 2019;20(8):1195-1206. doi:10.1007/s10198-019-01088-5

58. Kay S, Ferreira A. Mapping the 25-item National Eye Institute Visual Functioning Questionnaire (NEI VFQ-25) to EQ-5D Utility Scores. *Ophthalmic Epidemiol*. 2014;21(2):66-78. doi:10.3109/09286586.2014.888456

59. Skaltsa K, Longworth L, Ivanescu C, Phung D, Holmstrom S. Mapping the FACT-P to the Preference-Based EQ-5D Questionnaire in Metastatic Castration-Resistant Prostate Cancer. *Value Health*. 2014;17(2):238-244. doi:10.1016/j.jval.2013.12.005

60. Khan I, Morris S. A non-linear beta-binomial regression model for mapping EORTC QLQ- C30 to the EQ-5D-3L in lung cancer patients: a comparison with existing approaches. *Health Qual Life Outcomes*. 2014;12(1):1-16. doi:10.1186/s12955-014-0163-7

61. Khan I, Morris S, Pashayan N, Matata B, Bashir Z, Maguirre J. Comparing the mapping between EQ-5D-5L, EQ-5D-3L and the EORTC-QLQ-C30 in non-small cell lung cancer patients. *Health Qual Life Outcomes*. 2016;14(1):1-15. doi:10.1186/s12955-016-0455-1

62. Cruz MC, Andrade C, Urrutia M, Draibe S, Nogueira-Martins LA, Sesso R de CC. Quality of life in patients with chronic kidney disease. *Clinics*. 2011;66(6):991-995. doi:10.1590/S1807-59322011000600012

Chapter 2. A Systematic Review of Mapping Studies for the EQ-5D-5L for Use in Cost-Utility Analysis

Abstract

Background

"Mapping" has been a popular technique to fill the gap when utility values are not available for cost-utility analysis (CUA). Reporting of mapping studies needs to convey the information needed by CUA. The Health Economics Research Centre (HERC) has developed a mapping study database. The last HERC review was published in 2018 and the most recent database search update was conducted in January 2019. Although the EQ-5D-5L is a recently developed instrument, it has become one of the most popular indirect preference-based measures in CUA, and the number of studies not included in the HERC database or the last HERC review may be increasing rapidly.

Objectives

In this study, we aimed to 1) identify new mapping studies onto the EQ-5D-5L by conducting an updated systematic search, 2) assess the reporting quality of all mapping studies onto the EQ-5D-5L (including those included in HERC database), especially the completeness of information for CUA, with the use of reporting quality checklists, and 3) explore whether using the checklists in mapping studies can improve reporting quality.

Methods

An updated systematic search was conducted based on the HERC database from January 2018 to May 2020. We used the same search strategies and searched the same databases as the HERC database, except that we also performed searches on Embase. Studies reporting novel algorithms onto the EQ-5D-5L were included. The reporting quality of the study was assessed by criteria (including 31 items) developed from two checklists, the Mapping onto Preference-based Measures Reporting Standards Statement (23 items) and the ISPOR Good Practices Task Force Report on Mapping (12 items). For each item, three levels including "Completed", "Partially completed", and "No" were assessed.

Results

Our review identified 14 new studies not included in the HERC database. Together with the previous studies included in the HERC database, a total of 39 studies were reviewed and assessed. 57% of the studies had a study sample of fewer than 500 respondents. 28% of the studies focused on the cancer population. Ordinary least squares method was the most commonly applied mapping approach (85%). In the quality assessment, 21 items were "Completed" or "Partially completed" with a rate greater than 95%. The 5 poorest performed items were: 1) estimation of predicted utilities, 2) reporting variances, covariances, and error terms, 3) final model calculation example, 4) parameter uncertainty, and 5) individual uncertainty. These five items would also impact the use of mapping algorithms in CUA. Finally, a preliminary comparison showed that the checklists had positive impacts on improving the reporting quality of the study.

Conclusion

This review highlighted the strengths and drawbacks of mapping studies onto the EQ-5D-5L, and this could inform the reporting of further mapping studies.

2.1 Introduction

There has been an increase in the use of cost-utility analysis (CUA) to inform decision making on health care resource allocation[1]. CUA evaluates the value of a health technology by comparing costs and outcomes across the technology options[2], where outcomes are measured by quality-adjusted life years (QALY), which is a composite indicator of quality and quantity of life[3]. The quality is reflected by the preferences on health conditions, which could be estimated by direct approaches (e.g., using standard gamble or time trade-off) or indirect approaches (e.g., through pre-scored indirect preference-based measures (PBM))[4]. The indirect measures are more widely-used than the direct measures as the direct approaches are time-consuming and costly[2].

Clinical studies of new technologies aim to assess clinical safety and efficacy and they usually do not include PBMs, but sometimes include generic or disease-specific non-preference-based measures (non-PBM) to measure patients' health-related quality of life (HRQL)[5,6]. This is because compared with the non-PBMs being used in clinical trials, PBMs may not well capture

changes of symptoms or HRQL that is associated with a specific disease or treatment[7]. Besides, most clinical trials do not plan an associated economic evaluation, and in this case, researchers usually do not consider including a PBM in the study[8]. Since non-PBMs do not have a preference-based scoring system, there remains a gap to do CUA, where the effectiveness data of these trials are not applicable[6]. As such, the CUA may not be possible to conduct. One solution is to develop a mapping algorithm from the non-PBM to the PBM by regression or other possible statistical methods[5]. The mapping algorithm could convert the ratings or scores of the non-PBM to the utility values measured by a PBM. The mapping analysis may also be conducted to convert one PBM to another; this may be done in cases where jurisdictions have requirements for CUA to be undertaken with a specific PBM, such as with National Institute for Health and Care Excellence (NICE) in the UK[9].

The interest in using mapping techniques is growing[10]. Up to January 2019, the Health Economics Research Centre (HERC) mapping study database has identified 182 papers developing novel mapping algorithms onto a PBM from another PROM, and in total 386 algorithms[11]. These algorithms varied in disease populations (134 different diseases or patient groups, such as cancer), source instruments (164 different instruments, such as Functional Assessment of Cancer Therapy-General), and target instruments (11 different instruments, such as the EQ-5D)[11].

The EQ-5D has been most frequently mapped to according to the HERC database[11]. The EQ-5D is an instrument defining health in five dimensions (Mobility, Self-care, Usual Activities, Pain/Discomfort, Anxiety/Depression) and was developed by the EuroQoL group[12]. The first version of the EQ-5D has 3 levels (3L) in each dimension: no problems, some problems, and extreme problems[13]. It has been found that the EQ-5D-3L lacks the ability to measure small changes in health and has ceiling effect (i.e. respondents' scores reach the best possible score of the instrument[14]), and therefore the 5-level version of the EQ-5D, the EQ-5D-5L, was developed[15]. The EQ-5D-5L has 5 response categories for each dimension: no problems, slight problems, moderate problems, severe problems, and unable to/extreme problems. Based on the EuroQol website (www.euroqol.org), the EQ-5D-3L and -5L have been valued in 34 and 21 countries, respectively. Some countries' economic evaluation guidelines stated that the EQ-5D is the preferred measure to estimate HRQL, while some other countries do not state it directly but

recommend using instruments with their country-specific preference sets[9,16]. Therefore, the EQ-5D is one of the most popular PBMs being used in CUA[12,15]. Given that the EQ-5D-5L provides more precise measurements than the previous 3L version, the use of the EQ-5D-5L in the cost-utility analysis is increasing[17].

There are many concerns when using mapping algorithms in economic evaluations. The first is the extrapolation issue of the algorithm. It is believed that the study sample of the mapping analysis should be similar to the population of the economic evaluation when applying the algorithms[18]. Another concern is the uncertainty of the mapping algorithm. Any uncertainty associated with the algorithm and regarding the results should be accounted for in the cost-utility analysis to inform the decision-makers[16]. Therefore, there are many necessary elements to report in the mapping studies. Currently, two reporting quality checklists are available to guide mapping studies. They are the Mapping onto Preference-based Measures Reporting Standards Reporting Statement (MAPS reporting statement)[19] and the ISPOR Good Practices Task Force Report on Mapping (ISPOR Good Practices)[6].

The concerns related to the reporting of studies have raised interesting questions. What are the strengths that previous studies have and should be retained or promoted? What are the common issues of previous studies and how to deal with them to improve reporting quality? This study tried to answer these questions by selecting the EQ-5D-5L as the target measure and systematically reviewing all of the published mapping studies onto the EQ-5D-5L. The reasons for selecting the EQ-5D-5L were as follows. The EQ-5D-5L provides more precise measurements than the 3L version, and the use of the EQ-5D-5L in the cost-utility analysis is increasing[17]. The last HERC review was published in 2018 and only four studies mapping to the EQ-5D-5L were included. The last HERC database search update identified 26 mapping studies onto the EQ-5D-5L up to January 2019. Aside from the significant number of new mapping studies published after the last update of the HERC database, there were already 22 studies never been analyzed. Meanwhile, since the EQ-5D-5L is a relatively new instrument and the mapping studies associated with it were published within the recent 10 years, these papers would be representative of most recently published studies in terms of reporting quality. The MAPS reporting statement (2015) and the ISPOR Good Practices (2017) would be used as references to

check the reporting quality and determine the strengths and weaknesses of the EQ-5D-5L mapping studies.

There were three objectives of this study. The first objective was to identify all mapping studies onto the EQ-5D-5L by updating the systematic search of the HERC mapping study database. The second objective was to assess the quality of reporting of mapping studies, especially the completeness of information for CUA, with the reference of the MAPS reporting statement and the ISPOR Good Practices. The third objective was to explore whether using the checklists could improve reporting quality.

2.2 Methods

Search Strategy

The first objective of this study was to identify all mapping studies onto the EQ-5D-5L. We did this by conducting an updated search based on the HERC mapping study database[11]. The database is an excel spreadsheet listing all of the mapping studies. The version 7.0 is based on searches conducted in January 2019, and an older version (6.0) was published in 2018[1,11]. There were 26 studies using the EQ-5D-5L as the target instrument in version 7.0[11]. In the updated search, we used the same search strategies, but we limited the target instrument to be only the EQ-5D-5L. Search terms included "Mapping" or its synonyms ("Crosswalk", "Transfer to Utility" etc.) and "EQ-5D-5L" or its synonyms ("EuroQoL", "eq5d", etc.) (Appendix 2.1). We searched Medline (via PubMed), Embase, the EuroQoL website (https://euroqol.org/search-for-eq-5d-publications/), the School of Health and Related Research Health Utilities Database (https://www.scharrhud.org/), the Centre for Reviews and Dissemination database (https://www.crd.york.ac.uk/CRDWeb/), and the publications citing 11 systematic reviews or guidelines on mapping studies (via Scopus and Google Scholar) [1,3,10,19–25]. According to the developers of the HERC database, Embase was not searched due to time and resource limits[1], but we included Embase in this updated review. The time frame of the updated search was supposed to be from January 2019 to May 2020. To allow for the time lag during databases indexing the articles, we set our search timeline from July 2018 till May 2020. However, the databases we used could only be searched by year instead of by month, so the final search timeline was from January 2018 to May 2020.

In the updated search, studies that conducted a mapping analysis (including, but not limited to, regression techniques) onto the EQ-5D-5L and reported model coefficients were included. The exclusion criteria were as follows:

- Studies that were developing mapping algorithms onto other PBMs such as the EQ-5D-3L and the SF-6D were excluded.

- Studies that validated an existing mapping algorithm onto the EQ-5D-5L were linked to the original mapping study and would not be counted as another record.

- Studies such as CUA that adopted an existing mapping algorithm were excluded.

- Conference abstracts, systematic reviews, studies generating utilities using valuation techniques such as time trade-off, studies exploring mapping techniques but with the use of simulation data, and frameworks or guidelines on mapping analysis were excluded.

The process of deciding inclusion/exclusion was conducted by a single person (the author of the thesis). Studies included in the updated search together with the studies included in the HERC database of mapping studies (version 7.0) were further assessed.

The second objective was to analyze these studies about their reporting quality, especially the completeness of information for use in CUA. We used a designed data extraction form (Appendix 2.2). Data from each paper were extracted into 12 major categories: basic information of the paper (title, author, etc.), study rationale and objectives, estimation and validation sample information, source and target measures, exploratory analyses, model types, model specifications and estimation, performance and validation, descriptive results, performance comparison and selection, final model report, conclusion, and other information. The data extraction and the following reporting quality assessment were all done by the author of the thesis.

Reporting Quality Assessment

The MAPS reporting statement[19] focuses on reporting standards, and it was published in 2015. It has 23 items and provides recommendations on every section of a paper, i.e. title, abstract, introduction, methods, results, discussion, and other sections[19]. For each item, the MAPS reporting statement provides an example from published studies and therefore making it straightforward to understand. The ISPOR Good Practices, published in 2017, has recommendations on pre-modelling (6 items), modelling (5 items), and reporting (12 items)[6].

The ISPOR Good Practices provides a step-by-step instruction on pre-modelling and modelling considerations, which could well-inform the researchers on how to conduct a mapping analysis formally. But we only used the reporting recommendations in this study. There are some overlaps in the reporting standards between the ISPOR Good Practices and the MAPS reporting statement[1], but the ISPOR Good Practices focuses more on how to fully report the process of model analysis, while the MAPS reporting statement is more about how to make the study more informative to the readers and potential users. The lack of examples also makes following the ISPOR Good Practices on reporting studies not as easy as the MAPS reporting statement.

The reporting quality assessment was conducted for each included study using an assessment form (Appendix 2.3) developed from the two aforementioned checklists. We combined the items which described similar requirements in the two checklists, and some items were divided into sub-items to better reflect the information. There were 31 items in our final assessment form. The form followed the same order as the MAPS reporting statement, starting from the title and the abstract section, and ending with the discussion section and other information sections. The additional items included in the ISPOR Good Practices were inserted into the assessment form with consideration of their orders in the manuscripts. According to the description of each item provided by the MAPS reporting statement[19] and the ISPOR Good Practices[6], there were several key items that are important for the application in CUA, including the estimation of predicted utilities, descriptive information on estimation sample, model coefficients, parameter and individual uncertainty, calculation example, and scope of application. These items were highlighted in the assessment form.

This form was used to assess the achievement of each item. There were three levels, "Completed", "Partially completed" and "No" to rate each level. Some items have multiple requirements, and only if all requirements of an item were achieved, that item could be rated as "Completed".

Assessing the Impacts of the Checklists

The third objective was to explore whether using checklists could improve reporting quality. We conducted a preliminary comparison among the studies stating that they applied the checklists, the studies citing the checklists, and the studies not citing them. The items on which the studies

had common poor performance were identified for the comparison. We hypothesized that studies applying or citing the checklists would perform better on these items.

2.3 Results

General Descriptive Statistics

The updated search included 30 mapping studies onto the EQ-5D-5L after full-text review. Of these 30 studies, 16 studies were already included in the HERC database (version 7.0)[11], and there were 14 new studies. The HERC database (version 7.0)[11] listed 26 mapping studies onto the EQ-5D-5L. Aside from the overlapping ones between the updated search and the database, there were 10 more studies. One study was an internal study report, not published and not peer-reviewed, so we excluded it from our analysis. Therefore, 39 studies were eligible for the final reporting quality assessment (Figure 2.1, PRISMA diagram). There was an increasing trend in the mapping studies onto the EQ-5D-5L over the years, with an exception of 2019 (Figure 2.2).

Research Objectives

34 studies had the main objective to be exploring mapping algorithms between the instruments. Two studies focused on comparing the new methodologies used in mapping[26,27]. One study mapped the EQ-5D-3L to the EQ-5D-5L in order to explore the impact on the results of economic evaluation when transiting from the EQ-5D-3L to the EQ-5D-5L[28]. One study was to calculate the aggregated exchange rates across the health state utility instruments[29]. One study focused on a capability measure and had a comparison with the EQ-5D-5L on their association with depression instruments[30].

Study Sample

Almost two-thirds (57%) of the studies had a sample of fewer than 500 respondents. 17 studies had a study sample of 200-499. 28% of the studies focused on the cancer population (Table 2.1).

Modelling

The majority (85%) of the studies explored OLS in their analysis as a basic method (Table 2.2). Among advanced methodologies, generalized linear model (GLM) (33%), Censored Least Absolute Deviations (CLAD) (28%), beta-binomial regression (28%) and Tobit model (26%) were more popular. Adjusted limited dependent variable mixture model (ALDVMM)[31] and

mixture beta regression model (BETAMIX)[32] were new methods specially developed for mapping studies within the past five years, and there were some studies using them (ALDVMM: 13%, BETAMIX: 8%). Less frequently, robust MM-estimators[26,33–39], logit/probit model[26,34,35,37,40–42], fractional regression model[33,36,38,42], two-part model[43–45], linear equating model[36,44], equipercentile regression[27,46], mean rank method[27,46], multivariate fractional polynomial[34], conditional process analysis[47], Gaussian mixture model[43], extended estimating equations[42], generalized estimating equations[48], copula[28], quantile regression[29], and linear random effect model[49] have also been applied in the 39 studies (Table 2.2). Regarding final models, although there had been considerable critiques of OLS[50–53], it was still being selected as one of the final models in many studies (33%).

The model specifications also varied (Table 2.3). 9 studies only included the source measure scores as independent variables, while others may also consider demographics (59%), clinical items (15%), and polynomial (41%), interaction (33%), or categorical (8%) terms.

Reporting Quality of the Studies

In general, 21 items were "Completed" or "Partially completed" with a rate greater than 95%. The 5 poorest performed items were 1) estimation of predicted utilities, 2) reporting variances, covariances, and error terms, 3) final model calculation example, 4) parameter uncertainty, and 5) individual uncertainty. They had the lowest rates of "Completed" (smaller than 10%) and almost the highest rates of "No". A detailed assessment by section is performed as follows (Table 2.4).

Title, Abstract and Introduction Sections

69% of the studies completed the requirements for the title, i.e. using the term "mapping" or its synonyms, reporting source and target instruments. 28% of the studies were rated "Partially completed". Four studies only used a broad term "EQ-5D", instead of making it clear as "EQ-5D-5L". Four studies which had multiple target instruments did not list the specific target instruments. Two studies did not use "mapping" or its synonyms. 23% of the studies fully completed the abstract requirements, and 77% partially completed. The "Partially completed" were mainly due to no reporting on validation methods (n=18)[54], no reporting on specific

performance indicators of the best models (n=16)[55]. 97% of the studies fully completed the requirements for the study rationale and study objective in the introduction section (Table 2.4).

Methods Section

No study had multiple datasets to consider when estimating the mapping algorithm parameters. There was only one estimation sample for each study and all studies fully completed all requirements in introducing the basic information of their study samples, including study design, setting, collection, etc. Only one study used an external dataset to validate the mapping algorithm, and it met the requirements for reporting the external dataset. Limited reporting of missing data was an issue (missing data: 44% Completed, 23% Partially completed, 33% No) (Table 2.4).

56% of the studies reported completely on source and target measures. 17 studies did not mention whether the higher score indicated better or worse outcomes and this made them rated "Partially completed". Exploring the construct overlaps between the source and target measures is a helpful method to foresee the potential feasibility of developing a mapping algorithm. 51% of the studies used Spearman correlation to do this[35,56], and a few conducted formal factor analyses or principle component analyses (13%)[36,42]. Some studies indicated that they did an exploratory analysis but did not show the results (5%)[45,57] (Table 2.4).

The demonstration of model approaches was detailed in most published studies (97%), but 92% of the studies did not report the methods for calculating predicted utilities. The raw predicted dependent variable may not lie within the feasible utility distribution. For example, the OLS predicted values were not bounded, while the utility data had upper and lower limits. How to estimate the final values needs to be addressed, and only 5% of the studies stated their methods to estimate the predicted values[58]. Lastly, over 80% of the studies fully completed the validation, model performance and selection criteria sections per the requirements (Table 2.4).

Results Section

In the results section, several common deficiencies were found. Many studies did not report the severity distribution of the estimation sample (49% Partially completed, 3% No)[59]. The utility data distribution plot was also frequently not shown (No: 54%). The overall model fit was

demonstrated clearly using the performance indicators, but the conditional fit was not frequently reported (8% Completed and 92% Partially completed) (Table 2.4).

64% of the studies reported the basic information regarding model coefficients (e.g. coefficients, standard error, significance level). However, no study reported the variance-covariance matrix of the coefficients and error terms together, and 6 studies either reported variance-covariance matrix or error terms. These two terms would reflect the variation of the estimation in population- or individual-level. The lack of reporting variation leads to poor performance in reporting parameter uncertainty and individual uncertainty. Only a few (10%) studies fully completed the requirements in parameter uncertainty, and 5% of the studies fully completed the requirements in individual uncertainty (Table 2.4).

Studies also had flaws in presenting a calculation example for the best model, with 5% of the studies giving a specific example in the calculation. Many studies only had model equations or the calculation process (39%), and a few studies provided a user-friendly program to calculate the utility values (13%)[54,60] (Table 2.4).

Face validity was stated either in the results section or discussion section. The statements were not strong enough to fully support face validity in many studies (model performance and face validity: 10% Completed, 46% Partially completed). Ideally, a statement should address whether the sign of coefficients was as expected and if not, also address how to deal with it. But 10 (26%) studies talked more about the effect size of the coefficients and the significance of the correlation between source measure items and target measures[35,61] (Table 2.4).

Discussion Section

In the discussion section, 87% of the studies fully completed the requirements regarding comparison with other studies, and all studies discussed their limitations. The scope of applications was somewhat unclear (scope of applications: 54% Completed, 44% Partially completed). The studies rated "Partially completed" emphasized that external validation was required to validate and test the generalizability of their algorithms. However, this did not directly answer the question of whether a potential user could apply this algorithm to their study when the external validation was not available (Table 2.4).

Completeness of Information for CUA

The items directly associated with the application in CUA were highlighted in Table 2.4. According to the reporting quality assessment, estimation of predicted utilities, reporting on variances, covariances and error terms, parameter uncertainty, individual uncertainty, and calculation example were the 5 poorest-reported items. The rates of "Completed" for these five items were less than 10%. "No" rating surpassed 80% for estimation of predicted utilities, reporting on variances, covariances and error terms, and individual uncertainty. Reporting on the descriptive information on estimation sample (including the severity distribution), basic information on model coefficients, and the scope of the application had "Completed" or "Partially completed" with a rate greater than 95%, but the rates of "Partially completed" for the three items were all greater than 30%. Especially, there were around 50% of the studies not reporting severity distribution when describing the study sample.

Assessing the Impacts of the Checklists

Less than one-third (12, 31%) of the studies applied the MAPS reporting statement, and 20 studies cited the MAPS reporting statement (including those applied the checklist). For the ISPOR Good Practices, 3 studies applied it and 13 studies cited it. No study was found completing all of the items required in the checklist, and only one study that cited the ISPOR Good Practices fulfilled the recommendations at least partially (see Appendix 2.4 for complete extraction records). To explore the potential impacts of the checklists on these studies, we did a preliminary comparison on the items which had "Completed" rates less than 60% among the studies (Table 2.5). Better reporting performance was observed in most items for studies applying/citing the checklists.

2.4 Discussion

This study identified 14 new mapping studies to predict the EQ-5D-5L preference scores. We extended the search time to January 2018 till May 2020 to identify the new studies related to the EQ-5D-5L that were not included in the previous database. Together with the previous studies included in the database, we did a comprehensive review and reporting quality assessment on 39 studies and summarized the common strengths and issues in conducting mapping analyses. Overall, the majority of the requirements were fulfilled completely in most studies, with 21 out

of 31 that had "Completed" and "Partially completed" rates greater than 95%. However, we also identified some issues which would deteriorate the reporting quality of mapping algorithms, and the top 5 issues are directly related to the application in CUA.

There was an increasing trend in the number of published EQ-5D-5L mapping studies, except for 2019. One possible reason for it was that NICE did not recommend the use of the EQ-5D-5L value set for Britain in 2019[62]. But this may also just be a fluctuation of the trend. In the review by Dakin et al, there were fewer publications in certain years with an overall increasing trend[1]. Regardless of the EQ-5D-5L not being recommended in Britain, the new British EQ-5D-5L valuation study is under development, and the use of the EQ-5D-5L is expected to grow in the future.

Dakin et al's review observed a decrease in the number of studies with sample size fewer than 500, which is inconsistent with our study, where 47% of the mapping studies had sample sizes smaller than 500. This might be due to the EQ-5D-5L being a relatively new instrument compared with the EQ-5D-3L. It is possible that not many large-scale trials have administered the EQ-5D-5L.

This review showed that the structures of the current mapping studies were mostly good. The titles and abstracts of these studies were consistent. It is easy for readers to identify these articles and read them. Most items mentioned by the checklists were discussed in the mapping studies, making them more likely to be reliable and valid. Before conducting mapping analysis, many studies used Spearman correlation or other exploratory analyses to explore the relationship between the source and target measures. To deal with the characteristics of the utility data, abundant model types have been tried. Model specifications could include the instrument itself, sociodemographic information, health data, and polynomial, interaction, or categorical terms. This could further address the non-linear issue of the utility distribution and control the clinical and demographic effects on utility.

Based on the review results, we identified several key issues of current mapping studies. Missingness was not often discussed. It was common that the study sample had some missing observations, but studies usually just indirectly stated the number of missing values, but did not discuss the pattern of missing and how the missingness was handled. Without this, the mapping analyses are likely to risk selection bias.

Conditional fit statistics could assess the model fit more clearly, but the use of these statistics in published studies was uncommon. The plot of predicted utility versus observed utility in many studies has shown that poor prediction was common in the lower values of the EQ-5D-5L[60,63], with a systematic overprediction on these utility values. The overall fit statistic is a mean statistic over all of the observations, and it would shrink the effect of over-prediction when calculating for the whole sample. Brazier et al pointed out that using 0.5 as a cut point could see how the model performed in predicting the lower values and higher values of the EQ-5D[21]. The ISPOR Good Practices recommended providing information on fit conditional on disease severity as measured by the clinical outcome measures[6]. Given that the conditional performance statistics are more accurate in describing the model fit, there is a need to apply conditional fit statistics in the selection of final models in the future.

Several poor-performance items are related to the application in CUA, which made the mapping algorithms less user-friendly and provide incomplete information to CUA. Therefore, there is a significant need to improve these items and ensure the completeness of information for CUA. The first is improving the quality of reporting the estimation sample. To use the mapping algorithm in the economic evaluation, the sample of economic evaluation should be similar to the study sample[10]. Both checklists have recommendations on this[6,10,19]. Since many mapping studies were conducted to inform a future economic evaluation, the detailed characteristics of the estimation sample should be provided to allow potential users to compare the populations. These include sociodemographic and disease severity[6,19]. However, the severity distribution was poorly reported among published studies, and the potential users would not be able to know the disease progression information.

Advanced information on model coefficients and uncertainty should also be stated more transparently. Addressing uncertainty is crucial in CUA, and it is usually conducted via sensitivity analyses or scenario analyses[64]. Most CUA models are on the population level, and fewer models are on the individual level. When mapping algorithms were used in CUA to convert utility values, different types of the uncertainty of the mapped utility need to be considered[19]. The population-level uncertainty is reflected in the uncertainty of coefficients, and reporting the variance-covariance matrix would help. The individual-level variability was

reflected in the error term. As such, reporting the error term is necessary. Currently, the reporting of both types of uncertainty was very limited.

The estimation of the predicted utility and the examples illustrating the use of final models were related to the instructions on how to use the models for calculation, but they were not frequently reported. Most of the current studies only provide final model equations in the results section[65,66]. Situations like how to deal with the predicted values out of the range of the utility values should be illustrated. Therefore, it is important to demonstrate how predicted utilities are estimated to the final value in the methods section, and an additional calculation example in the results section could help the users to see whether they understand the algorithms right. ISPOR Good Practices also recommends developing a user-friendly program for calculation[6]. We did find some papers which selected complex models as their final models[54,60] and developed a program to help users for calculation.

Compared with the review done by Dakin et al[1], there were some consistencies in the reporting quality results. In Dakin et al's review on titles and abstracts, studies published between 2014 to 2016 had poorer performance on validation methods (45%) and reporting model performance (59%) ("other methods" was also not frequently reported, but this is not necessary for the mapping study). The studies we reviewed had similar results (validation: 54%, performance statistics: 59%). Dakin et al also did a full-text review of studies published in 2016[1]. The review showed that the study rationale and study objectives of these studies were well-reported. Information on missingness, methods for calculating predicted utilities, reporting standard errors, exploratory data analysis were not provided clearly. We also observed similar results in our study. Meanwhile, we saw a better performance on validation methods illustration and measures of model performance among the studies we reviewed. However, Dakin et al's review did not include a sufficient number of EQ-5D-5L mapping studies (the first EQ-5D-5L mapping study was published in 2014) to allow us to head-to-head compare the quality assessment results on EQ -5D-5L mapping studies.

We also did a preliminary comparison among studies applying, citing, or not citing the three checklists to see the impact of the checklists. For most of the items that were not commonly reported well, better reporting performance was observed in studies citing the checklists. The checklists could somewhat inform the mapping studies on what to report. One thing to highlight

regarding these checklists was that there are discrepancies among them. The ISPOR Good Practices does not emphasize internal validation is necessary but external validation is always preferred. The MAPS reporting statement requires statement and justification if no validation is conducted. Each checklist did have unique items that are important to the reporting quality of mapping studies, such as the requirements on face validity in the MAPS reporting statement and the requirements on conditional fit statistics in the ISPOR Good Practices. To ensure the reporting quality of a mapping study, it is recommended to apply both checklists during the analysis.

The strength of our study was that we applied the two influential mapping studies checklists and did a reporting quality check on all of the published EQ-5D-5L studies. We summarized the strengths of the current mapping studies and pointed out the common weaknesses and the way to improve them. This would be an important message to inform further mapping studies and help them to provide more complete information for the application. We also had a summary of model types that have been used in previous mapping studies, which could be considered for further application.

There were also several limitations in this study. First, we limited our target measure to be just the EQ-5D-5L. There were more mapping studies regarding the EQ-5D-3L, and the mapping studies regarding the EQ-5D-3L may also be informative and representative. Since the process of mapping analysis is independent of what target instrument is being mapped, reviewing the mapping studies onto other instruments, especially the EQ-5D-3L, is an extensive work but could provide a broader view on the reporting performance of the mapping studies. Second, this search was conducted based on a published systematic review and its associated mapping database. The search aimed to find new studies since 2018. If there was any mapping study published before 2018 and not included in the HERC database, we were not able to identify them and include them in our study. Third, this review was conducted by a single person (the author of this thesis). This might increase the likelihood of errors and biases in including/excluding and evaluating the studies. Fourth, we did not use a scoring system for the quality check, which could potentially enable us to do the inferential statistical comparison on the reporting quality of the studies. Currently, we only used descriptive statistics to summarize and compare the results.

2.5 Conclusions

Overall, the quality of reporting is medium to high. However, several issues, especially those related to the application in CUA, including estimation of predicted utilities, reporting variances, covariances, and error terms, reporting final model calculation example, parameter uncertainty, and individual uncertainty, are worth noticing and improving in further studies.

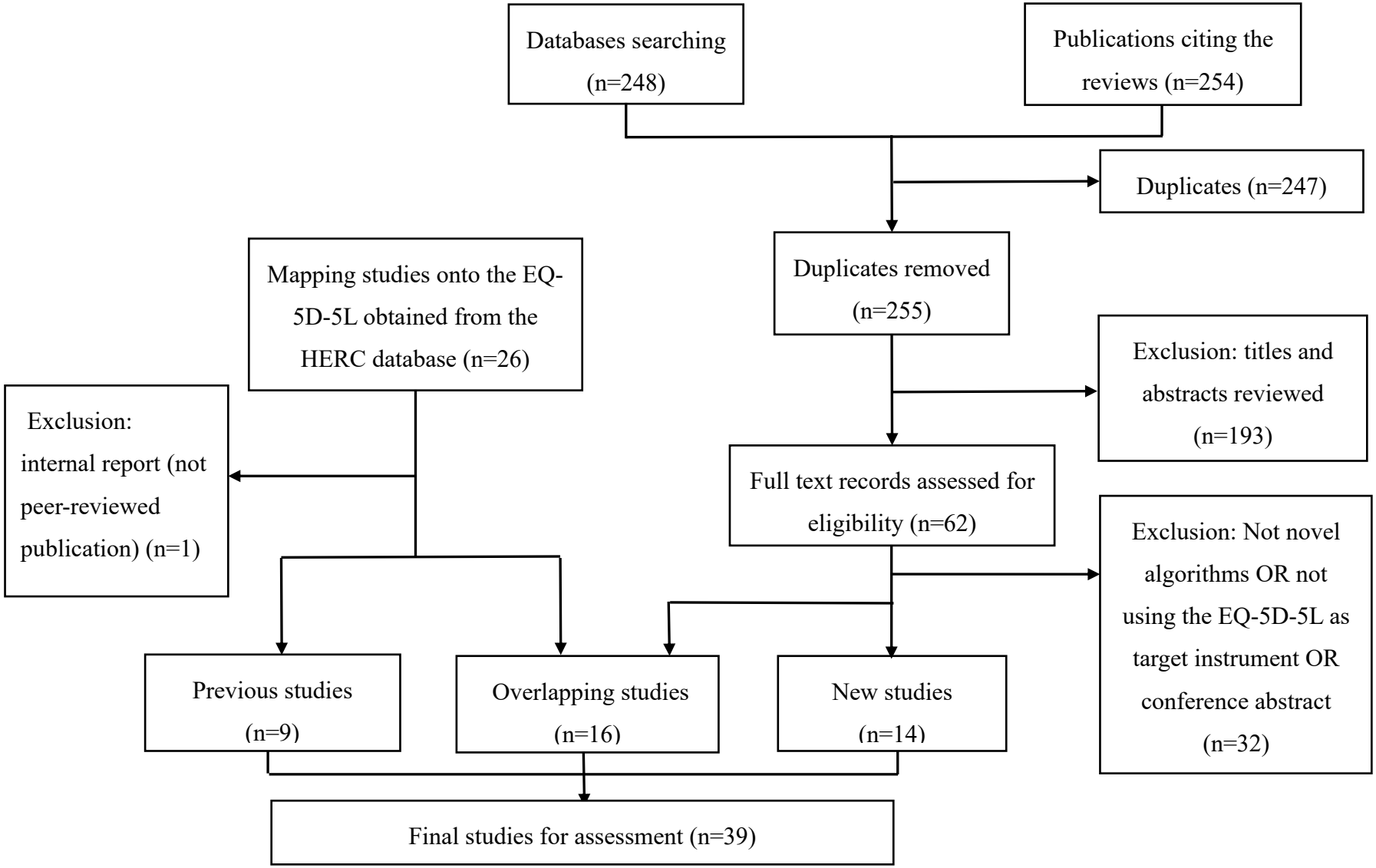


Figure 2.1 PRISMA flow diagram

Note: HERC, Health Economics Research Centre

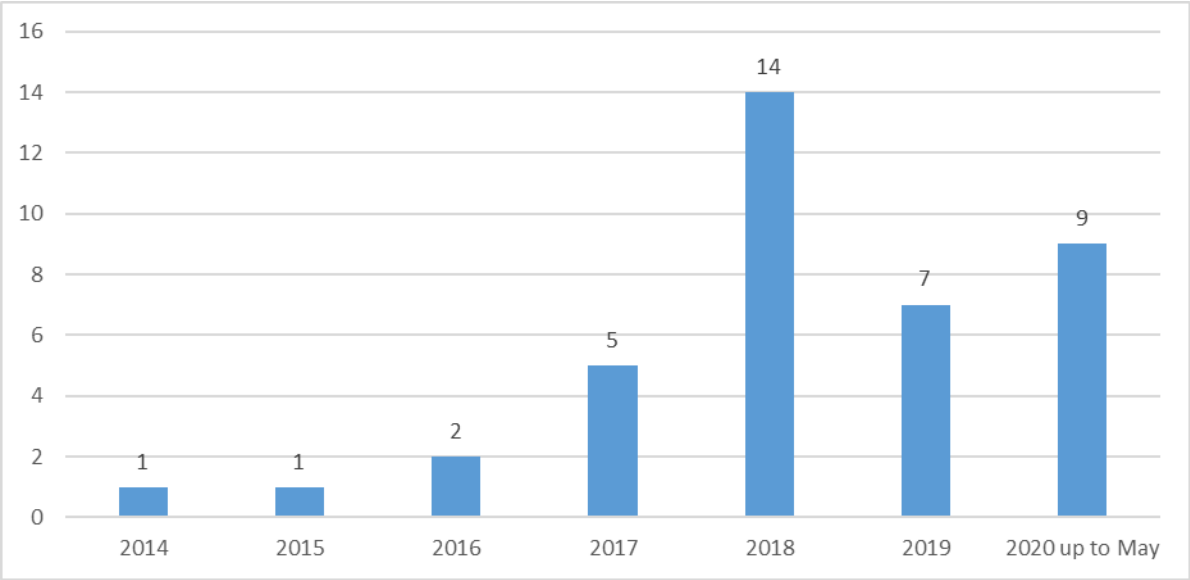


Figure 2.2 Published mapping studies onto the EQ-5D-5L by year of publication

Table 2.1 Study sample characteristics of the published mapping studies onto the EQ-5D-5L (n=39)

Number of observations included in the estimation sample	
0-199	5(13%)
200-499	17(44%)
500-999	12(31%)
1000-4999	3(8%)
5000-10000	2(5%)
Disease area	
cancer	11(28%)
central nervous system	3(8%)
endocrine disorders	1(3%)
eye	1(3%)
general population	3(8%)
heart	2(5%)
mental health and behavioural disorders	4(10%)
musculoskeletal	5(13%)
respiratory system	2(5%)
skin	3(8%)
urogenital	2(5%)
various	2(5%)

Table 2.2 List of model types in mapping studies onto the EQ-5D-5L (n=39)

Model type	No. of studies used this model	No. of studies selected this as final models
Ordinary least square	33(85%)	13(33%)
Generalized linear model	13(33%)	6(15%)
Censored least absolute deviations model	11(28%)	1(3%)
Beta regression	11(28%)	5(13%)
Tobit model	10(26%)	2(5%)
Robust MM-estimators	8(21%)	1(3%)
Logit/Probit	7(18%)	4(10%)
Adjusted limited dependent variable mixture model	5(13%)	1(3%)
Fractional regression model	4(10%)	1(3%)
Mixture beta regression model	3(8%)	1(3%)
Two-part model	3(8%)	1(3%)
Linear equating	3(8%)	0(0%)
Equipercentile regression	2(5%)	0(0%)
Mean rank method	2(5%)	1(3%)
Multivariate fractional polynomial	1(3%)	0(0%)
Conditional process analysis	1(3%)	1(3%)
Gaussian mixture	1(3%)	1(3%)
Extended estimation equation	1(3%)	1(3%)
Generalized estimating equations	1(3%)	1(3%)
Copula	1(3%)	1(3%)
Quantile regression	1(3%)	1(3%)
Linear random effect	1(3%)	0(0%)

Table 2.3 Summary of model specifications of the mapping studies onto the EQ-5D-5L (n=39)

Model specification options	No. of studies
Use of source measure index/dimension	39(100%)
Use of categorical terms	3(8%)
Use of interaction terms	13(33%)
Use of polynomial terms	16(41%)
Use of sociodemographic information	23(59%)
Use of other health data	6(15%)

Table 2.4 Reporting quality assessment results of the mapping studies (n=39)

	Completed	Partially completed	No
MAPS - Title	27(69%)	11(28%)	1(3%)
MAPS - Abstract	9(23%)	30(77%)	0(0%)
MAPS - Introduction: study rationale	38(97%)	1(3%)	0(0%)
MAPS - Introduction: study objective	38(97%)	1(3%)	0(0%)
ISPOR - Methods: candidate dataset description	NA	NA	NA
MAPS, ISPOR - Methods: estimation sample	39(100%)	0(0%)	0(0%)
MAPS - Methods: external validation sample (only 1 study did external validation)	1(100%)	0 (0%)	0(0%)
MAPS - Methods: source and target measures	22(56%)	17(44%)	0(0%)
MAPS, ISPOR - Methods: exploratory data analysis	26(67%)	1(3%)	12(31%)
MAPS, ISPOR - Methods: missing data	17(44%)	9(23%)	13(33%)
MAPS, ISPOR - Methods: modelling approaches	38(97%)	0(0%)	1(3%)
MAPS - Methods: estimation of predicted utilities	2(5%)	1(3%)	36(92%)
MAPS - Methods: validation methods (8 studies did not have validation)	31(100%)	0(0%)	0(0%)
MAPS - Methods: measures of model performance	37(95%)	0(0%)	2(5%)
ISPOR - Methods: Approach to determine the final model (4 studies had only one model)	34(87%)	0(0%)	1(3%)
MAPS - Results: final sample size	33(85%)	5(13%)	1(3%)
MAPS, ISPOR - Results: descriptive information (especially severity distribution)	19(49%)	19(49%)	1(3%)
ISPOR - Results: Utility data distribution plot	21(54%)	0(0%)	18(46%)
ISPOR - Results: Fit statistics, especially conditional fit statistics	3(8%)	36(92%)	0(0%)
ISPOR - Results: Plot on observed and predicted utility values	30(77%)	0(0%)	9(23%)
MAPS - Results: model selection (4 studies had only one model)	34(87%)	0(0%)	1(3%)
MAPS, ISPOR - Results: model coefficients (size, sign, significance, stand error, error term)	25(64%)	12(31%)	2(5%)
ISPOR - Results: model coefficients additional requirements (variance-covariance matrix, error)	0(0%)	6(15%)	33(85%)
ISPOR - Results: Calculation example / user-friendly program	4(10%)	18(46%)	17(44%)
MAPS, ISPOR - Results: uncertainty (parameter)	4(10%)	18(46%)	17(44%)
MAPS, ISPOR - Results: uncertainty (individual)	2(5%)	0(0%)	37(95%)
MAPS - Results: model performance and face validity	4(10%)	35(90%)	0(0%)
MAPS - Discussion: comparisons with previous studies	34(87%)	4(10%)	1(3%)
MAPS - Discussion: study limitations	39(100%)	0(0%)	0(0%)
MAPS - Discussion: scope of applications	21(54%)	17(44%)	1(3%)
MAPS - Other: additional information	37(95%)	0(0%)	2(5%)

Note: MAPS: Mapping onto Preference-based Measures Reporting Standards Reporting Statement[19], ISPOR: ISPOR Good Practices Task Force Report on Mapping[14]. The highlighted items are directly associated with the application in CUA.

Table 2.5 Comparison among studies applying, citing or not citing the checklists on selected items

Selected items from checklists	Among studies applying the checklists	Among studies citing the checklists	Among studies not citing the checklists
MAPS reporting statement	**(n=12)**	**(n=20)**	**(n=19)**
MAPS - Abstract	5(42%)	7(35%)	2(11%)
MAPS - Methods: source and target measures	7(58%)	13(65%)	9(47%)
MAPS - Methods: missing data	8(67%)	11(55%)	6(32%)
MAPS - Methods: exploratory data analysis	5(42%)	5(25%)	0(0%)
MAPS - Methods: estimation of predicted utilities	1(8%)	1(5%)	1(5%)
MAPS - Results: model coefficients	12(100%)	14(70%)	8(42%)
MAPS - Results: uncertainty	0(0%)	3(15%)	3(16%)
MAPS - Results: model performance and face validity	1(8%)	3(15%)	3(16%)
MAPS - Discussion: scope of applications	6(50%)	9(45%)	12(63%)
ISPOR Good Practices	**(n=3)**	**(n=13)**	**(n=26)**
ISPOR - Results: full details on dataset (including severity distribution)	1(33%)	4(31%)	3(12%)
ISPOR - Results: Utility data distribution plot	1(33%)	8(62%)	11(42%)
ISPOR - Results: Fit statistics, especially conditional fit statistics	0(0%)	2(15%)	1(4%)
ISPOR - coefficients, error terms, variances, covariances	0(0%)	0(0%)	0(0%)
ISPOR - Calculation example / user-friendly program	1(33%)	2(10%)	2(11%)
ISPOR - parameter uncertainty	1(33%)	3(23%)	1(4%)
ISPOR - individual uncertainty	0(0%)	2(15%)	0(0%)

MAPS: Mapping onto Preference-based Measures Reporting Standards Reporting Statement[19], ISPOR: ISPOR Good Practices Task Force Report on Mapping[14]

References

1. 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):1-9. doi:10.1186/s12955-018-0857-3
2. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Fourth edition. Oxford, United Kingdom: Oxford University Press; 2015.
3. Mukuria C, Rowen D, Harnan S, et al. An Updated Systematic Review of Studies Mapping (or Cross-Walking) Measures of Health-Related Quality of Life to Generic Preference-Based Measures to Generate Utility Values. *Appl Health Econ Health Policy*. 2019;17(3):295-313. doi:10.1007/s40258-019-00467-6
4. Torrance GW, Furlong W, Feeny D. Health utility estimation. *Expert Rev Pharmacoecon Outcomes Res*. 2002;2(2):99-108. doi:10.1586/14737167.2.2.99
5. Chuang L-H, Whitehead SJ. Mapping for economic evaluation. *Br Med Bull*. 2012;101(1):1-15. doi:10.1093/bmb/ldr049
6. 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. doi:10.1016/j.jval.2016.11.006
7. Brazier JE, Rowen D, Mavranezouli I, et al. *Developing and Testing Methods for Deriving Preference-Based Measures of Health from Condition-Specific Measures (and Other Patient-Based Measures of Outcome).* NIHR Journals Library; 2012. Accessed April 26, 2020. http://www.ncbi.nlm.nih.gov/books/NBK100337/
8. Hughes D, Charles J, Dawoud D, et al. Conducting Economic Evaluations Alongside Randomised Trials: Current Methodological Issues and Novel Approaches. *PharmacoEconomics*. 2016;34(5):447-461. doi:10.1007/s40273-015-0371-y
9. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal 2013. National Institute for Health and Care Excellence (NICE)*. NICE; 2013. Accessed April 26, 2020. https://www.nice.org.uk/process/pmg9/chapter/the-reference-case#measuring-and-valuing-health-effects
10. Longworth L, Rowen D. Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value Health*. 2013;16(1):202-210. doi:10.1016/j.jval.2012.10.010
11. HERC database of mapping studies, Version 7.0 (last updated: 24th April 2019). http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies.
12. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. *Appl Health Econ Health Policy*. 2017;15(2):127-137. doi:10.1007/s40258-017-0310-5

13. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53-72. doi:10.1016/0168-8510(96)00822-6

14. Michalos AC. *Encyclopedia of Quality of Life and Well-Being Research*. Springer Netherlands Dordrecht; 2014.

15. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-1736. doi:10.1007/s11136-011-9903-x

16. Guidelines for the Economic Evaluation of Health Technologies: Canada. CADTH.ca. Published December 2, 2015. Accessed May 4, 2020. https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada

17. 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(6):675-697. doi:10.1007/s40273-018-0623-8

18. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford University Press; 2017.

19. Petrou S, Rivero-Arias O, Dakin H, et al. The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. *PharmacoEconomics*. 2015;33(10):993-1011. doi:10.1007/s40273-015-0312-9

20. 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):683-694. doi:10.1007/s11136-012-0189-4

21. 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):215-225. doi:10.1007/s10198-009-0168-z

22. Stein K, Fry A, Round A, Milne R, Brazier J. What value health? *Appl Health Econ Health Policy*. 2005;4(4):219-228. doi:10.2165/00148365-200504040-00004

23. 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(1):66-89. doi:10.1177/0272989X07309642

24. Round J. Capturing information loss in estimates of uncertainty that arise from mapping algorithms. *Aberd Health Econ Study Group HESG*. Published online 2008.

25. Lin F-J, Longworth L, Pickard AS. Evaluation of content on EQ-5D as compared to disease-specific utility measures. *Qual Life Res*. 2013;22(4):853-874. doi:10.1007/s11136-012-0207-6

26. Kaambwa B, Smith C, Lacey S de, Ratcliffe J. Does Selecting Covariates Using Factor Analysis in Mapping Algorithms Improve Predictive Accuracy? A Case of Predicting EQ-5D-5L and SF-6D Utilities from the Women's Health Questionnaire. *Value Health*. 2018;21(10):1205-1217. doi:10.1016/j.jval.2018.01.020

27. Lee CF, Ng R, Luo N, Cheung YB. Development of Conversion Functions Mapping the FACT-B Total Score to the EQ-5D-5L Utility Value by Three Linking Methods and Comparison with the Ordinary Least Square Method. *Appl Health Econ Health Policy*. 2018;16(5):685-695. doi:10.1007/s40258-018-0404-8

28. Hernández-Alava M, Pudney S. Econometric modelling of multiple self-reports of health states: The switch from EQ-5D-3L to EQ-5D-5L in evaluating drug therapies for rheumatoid arthritis. *J Health Econ*. 2017;55:139-152. doi:10.1016/j.jhealeco.2017.06.013

29. Gamst-Klaussen T, Chen G, Lamu AN, Olsen JA. Health state utility instruments compared: inquiring into nonlinearity across EQ-5D-5L, SF-6D, HUI-3 and 15D. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2016;25(7):1667-1678. doi:10.1007/s11136-015-1212-3

30. Mitchell PM, Al-Janabi H, Byford S, et al. Assessing the validity of the ICECAP-A capability measure for adults with depression. *BMC Psychiatry*. 2017;17(1):46. doi:10.1186/s12888-017-1211-8

31. Hernandez-Alava M, Wailoo A. Fitting Adjusted Limited Dependent Variable Mixture Models to EQ-5D. *The Stata Journal*, 15, Number 3, pp. 737–750.

32. Gray LA, Alava MH. A Command for Fitting Mixture Regression Models for Bounded Dependent Variables Using the Beta Distribution. *Stata J*. 2018;18(1):51-75. doi:10.1177/1536867X1801800105

33. A.n L. Does linear equating improve prediction in mapping? Crosswalking MacNew onto EQ-5D-5L value sets. *J Health Econ*. Published online 2020. doi:10.1007/s10198-020-01183-y

34. Kularatna S, Senanayake S, Chen G, Parsonage W. Mapping the Minnesota living with heart failure questionnaire (MLHFQ) to EQ-5D-5L in patients with heart failure. *Health Qual Life Outcomes*. 2020;18(1):115. doi:10.1186/s12955-020-01368-2

35. Liu T, Li S, Wang M, Sun Q, Chen G. Mapping the Chinese Version of the EORTC QLQ-BR53 Onto the EQ-5D-5L and SF-6D Utility Scores. *The patient*. Published online May 8, 2020. doi:10.1007/s40271-020-00422-x

36. Lamu AN, Chen G, Gamst-Klaussen T, Olsen JA. Do country-specific preference weights matter in the choice of mapping algorithms? The case of mapping the Diabetes-39 onto eight country-specific EQ-5D-5L value sets. *Qual Life Res*. 2018;27(7):1801-1814. doi:10.1007/s11136-018-1840-5

37. Chen G, Garcia-Gordillo MA, Collado-Mateo D, et al. Converting Parkinson-Specific Scores into Health State Utilities to Assess Cost-Utility Analysis. *Patient*. 2018;11(6):665-675. doi:10.1007/s40271-018-0317-5

38. Gamst-Klaussen T, Lamu AN, Chen G, Olsen JA. Assessment of outcome measures for cost-utility analysis in depression: mapping depression scales onto the EQ-5D-5L. *BJPsych Open*. 2018;4(4):160-166. doi:10.1192/bjo.2018.21

39. Chen G, Khan MA, Iezzi A, Ratcliffe J, Richardson J. Mapping between 6 Multiattribute Utility Instruments. *Med Decis Making*. 2016;36(2):160-175. doi:10.1177/0272989X15578127

40. Patton T, Hu H, Luan L, Yang K, Li S-C. Mapping between HAQ-DI and EQ-5D-5L in a Chinese patient population. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2018;27(11):2815-2822. doi:10.1007/s11136-018-1925-1

41. Moore A, Young CA, Hughes DA. Mapping ALSFRS-R and ALSUI to EQ-5D in Patients with Motor Neuron Disease. *Value Health*. 2018;21(11):1322-1329. doi:10.1016/j.jval.2018.05.005

42. Lamu AN, Olsen JA. Testing alternative regression models to predict utilities: mapping the QLQ-C30 onto the EQ-5D-5L and the SF-6D. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2018;27(11):2823-2839. doi:10.1007/s11136-018-1981-6

43. Vilsbøll AW, Kragh N, Hahn-Pedersen J, Jensen CE. Mapping Dermatology Life Quality Index (DLQI) scores to EQ-5D utility scores using data of patients with atopic dermatitis from the National Health and Wellness Study. *Qual Life Res*. Published online 2020. doi:10.1007/s11136-020-02499-1

44. Yang Q, Yu XX, Zhang W, Li H. Mapping function from FACT-B to EQ-5D-5 L using multiple modelling approaches: data from breast cancer patients in China. *Health Qual Life Outcomes*. 2019;17(1):153. doi:10.1186/s12955-019-1224-8

45. Noel CW, Stephens RF, Su JS, et al. Mapping the EORTC QLQ-C30 and QLQ-H&N35, onto EQ-5D-5L and HUI-3 indices in patients with head and neck cancer. *Head Neck*. Published online April 25, 2020. doi:10.1002/hed.26181

46. Wee HL, Yeo KK, Chong KJ, Khoo EYH, Cheung YB. Mean Rank, Equipercentile, and Regression Mapping of World Health Organization Quality of Life Brief (WHOQOL-BREF) to EuroQoL 5 Dimensions 5 Levels (EQ-5D-5L) Utilities. *Med Decis Making*. 2018;38(3):319-333. doi:10.1177/0272989X18756890

47. Lee MS, Yoon J, Kim J, et al. Health-Related Utility of EQ-5D in Korean Adults With Chronic Urticaria: Mapping From Urticaria Outcome Measures. *Allergy Asthma Immunol Res*. 2020;12(4):599-607. doi:10.4168/aair.2020.12.4.599

48. 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(2):181-193. doi:10.1007/s10198-017-0930-6

49. Kaambwa B, Ratcliffe J. Predicting EuroQoL 5 Dimensions 5 Levels (EQ-5D-5L) Utilities from Older People's Quality of Life Brief Questionnaire (OPQoL-Brief) Scores. *Patient*. 2018;11(1):39-54. doi:10.1007/s40271-017-0259-3

50. Payakachat N, Summers KH, Pleil AM, et al. Predicting EQ-5D utility scores from the 25-item National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients with age-related macular degeneration. *Qual Life Res*. 2009;18(7):801-813. doi:10.1007/s11136-009-9499-6

51. Vokó Z, Németh R, Nagyjánosi L, et al. Mapping the Nottingham Health Profile onto the Preference-Based EuroQol-5D Instrument for Patients with Diabetes. *Value Health Reg Issues*. 2014;4:31-36. doi:10.1016/j.vhri.2014.06.002

52. Skaltsa K, Longworth L, Ivanescu C, Phung D, Holmstrom S. Mapping the FACT-P to the Preference-Based EQ-5D Questionnaire in Metastatic Castration-Resistant Prostate Cancer. *Value Health*. 2014;17(2):238-244. doi:10.1016/j.jval.2013.12.005

53. Chuang L-H, Kind P. Converting the SF-12 into the EQ-5D. *PharmacoEconomics*. 2009;27(6):491-505. doi:10.2165/00019053-200927060-00005

54. Vilsboll AW, Kragh N, Hahn-Pedersen J, Jensen CE. Mapping Dermatology Life Quality Index (DLQI) scores to EQ-5D utility scores using data of patients with atopic dermatitis from the National Health and Wellness Study. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. Published online April 15, 2020. doi:10.1007/s11136-020-02499-1

55. Dixon P, Hollingworth W, Sparrow J. Mapping to Quality of Life and Capability Measures in Cataract Surgery Patients: From Cat-PROM5 to EQ-5D-3L, EQ-5D-5L, and ICECAP-O Using Mixture Modelling. *MDM Policy Pract*. 2020;5(1):2381468320915447. doi:10.1177/2381468320915447

56. Wijnen BFM, Mosweu I, Majoie MHJM, et al. A comparison of the responsiveness of EQ-5D-5L and the QOLIE-31P and mapping of QOLIE-31P to EQ-5D-5L in epilepsy. *Eur J Health Econ*. 2018;19(6):861-870. doi:10.1007/s10198-017-0928-0

57. Stephens RF, Noel CW, Su JS, et al. Mapping the University of Washington Quality of life questionnaire onto EQ-5D and HUI-3 indices in patients with head and neck cancer. *Head Neck*. Published online November 24, 2019. doi:10.1002/hed.26031

58. Abdin E, Chong SA, Seow E, Verma S, Tan KB, Subramaniam M. Mapping the Positive and Negative Syndrome Scale scores to EQ-5D-5L and SF-6D utility scores in patients with schizophrenia. *Qual Life Res*. 2019;28(1):177-186. doi:10.1007/s11136-018-2037-7

59. Bilbao A, Martín-Fernández J, García-Pérez L, et al. Mapping WOMAC onto the EQ-5D-5L Utility Index in Patients with Hip or Knee Osteoarthritis. *Value Health*. Published online November 8, 2019. doi:10.1016/j.jval.2019.09.2755

60. Yang F, Wong CKH, Luo N, Piercy J, Moon R, Jackson J. Mapping the kidney disease quality of life 36-item short form survey (KDQOL-36) to the EQ-5D-3L and the EQ-5D-5L in patients undergoing dialysis. *Eur J Health Econ*. 2019;20(8):1195-1206. doi:10.1007/s10198-019-01088-5

61. Su J, Liu T, Li S, Zhao Y, Kuang Y. A mapping study in mainland China: predicting EQ-5D-5L utility scores from the psoriasis disability index. *J Med Econ*. Published online 2020. doi:10.1080/13696998.2020.1748636

62. Position statement on use of the EQ-5D-5L value set for England (updated October 2019) | NICE technology appraisal guidance | NICE guidance | Our programmes | What we do | About. NICE. Accessed June 4, 2020. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l

63. van Hout B, Janssen MF, Feng Y-S, et al. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. *Value Health*. 2012;15(5):708-715. doi:10.1016/j.jval.2012.02.008

64. Briggs A, Sculpher M, Buxton M. Uncertainty in the economic evaluation of health care technologies: The role of sensitivity analysis. *Health Econ*. 1994;3(2):95-104. doi:10.1002/hec.4730030206

65. Ameri H, Yousefi M, Yaseri M, Nahvijou A, Arab M, Akbari Sari A. Mapping the cancer-specific QLQ-C30 onto the generic EQ-5D-5L and SF-6D in colorectal cancer patients. *Expert Rev Pharmacoecon Outcomes Res*. 2019;19(1):89-96. doi:10.1080/14737167.2018.1517046

66. Ameri H, Yousefi M, Yaseri M, Nahvijou A, Arab M, Akbari Sari A. Mapping EORTC-QLQ-C30 and QLQ-CR29 onto EQ-5D-5L in Colorectal Cancer Patients. *J Gastrointest Cancer*. 2020;51(1):196-203. doi:10.1007/s12029-019-00229-6

67. Wong CKH, Cheung PWH, Samartzis D, et al. Mapping the SRS-22r questionnaire onto the EQ-5D-5L utility score in patients with adolescent idiopathic scoliosis. *PLOS ONE*. 2017;12(4):e0175847. doi:10.1371/journal.pone.0175847

68. Peak J, Goranitis I, Day E, Copello A, Freemantle N, Frew E. Predicting health-related quality of life (EQ-5D-5 L) and capability wellbeing (ICECAP-A) in the context of opiate dependence using routine clinical outcome measures: CORE-OM, LDQ and TOP. *Health Qual Life Outcomes*. 2018;16(1). doi:10.1186/s12955-018-0926-7

69. Khan I, Morris S, Pashayan N, Matata B, Bashir Z, Maguirre J. Comparing the mapping between EQ-5D-5L, EQ-5D-3L and the EORTC-QLQ-C30 in non-small cell lung cancer patients. *Health Qual Life Outcomes*. 2016;14(1):1-15. doi:10.1186/s12955-016-0455-1

70. Cheung YB, Luo N, Ng R, Lee CF. Mapping the functional assessment of cancer therapy-breast (FACT-B) to the 5-level EuroQoL Group's 5-dimension questionnaire (EQ-5D-5L) utility index in a multi-ethnic Asian population. *Health Qual Life Outcomes*. 2014;12:180. doi:10.1186/s12955-014-0180-6

71. Collado-Mateo D, Chen G, Garcia-Gordillo MA, et al. "Fibromyalgia and quality of life: mapping the revised fibromyalgia impact questionnaire to the preference-based instruments." *Health Qual Life Outcomes*. 2017;15. doi:10.1186/s12955-017-0690-0

72. Coon C, Bushmakin A, Tatlock S, et al. Evaluation of a crosswalk between the European Quality of Life Five Dimension Five Level and the Menopause-Specific Quality of Life questionnaire. *Climacteric J Int Menopause Soc*. 2018;21(6):566-573. doi:10.1080/13697137.2018.1481381

73. Gray LA, Hernández Alava M, Wailoo AJ. Development of Methods for the Mapping of Utilities Using Mixture Models: Mapping the AQLQ-S to the EQ-5D-5L and the HUI3 in Patients with Asthma. *Value Health J Int Soc Pharmacoeconomics Outcomes Res*. 2018;21(6):748-757. doi:10.1016/j.jval.2017.09.017

74. Gray LA, Wailoo AJ, Hernandez Alava M. Mapping the FACT-B Instrument to EQ-5D-3L in Patients with Breast Cancer Using Adjusted Limited Dependent Variable Mixture

Models versus Response Mapping. *Value Health J Int Soc Pharmacoeconomics Outcomes Res*. 2018;21(12):1399-1405. doi:10.1016/j.jval.2018.06.006

75. Kaambwa B, Chen G, Ratcliffe J, Iezzi A, Maxwell A, Richardson J. Mapping Between the Sydney Asthma Quality of Life Questionnaire (AQLQ-S) and Five Multi-Attribute Utility Instruments (MAUIs). *PharmacoEconomics*. 2017;35(1):111-124. doi:10.1007/s40273-016-0446-4

Chapter 3. Mapping the Edmonton Symptom Assessment System Revised-Renal onto the EQ-5D-5L in Patients with Chronic Kidney Disease

Abstract

Background

The Edmonton Symptom Assessment System Revised-Renal (ESAS-r: Renal) is a disease-specific patient-reported symptom assessment scale focusing on symptoms related to end-stage renal disease (ESRD). It is widely used in Canada and internationally. There is no mapping algorithm from any form of the ESAS to the preference-based scoring system of the EQ-5D-5L to support its use in cost-utility analysis.

Objective

To develop a mapping algorithm from the ESAS-r: Renal to the Canadian EQ-5D-5L index scores.

Methods

We used data from the Evaluation of Routinely Measured Patient-reported Outcomes in Hemodialysis Care (EMPATHY) trial, a multi-centre clustered randomized-controlled trial of routine measurement of patient-reported outcomes in hemodialysis units in Northern Alberta. The EMPATHY trial collected both the ESAS-r: Renal and the EQ-5D-5L for ESRD patients. The model estimation explored direct mapping models which mapped the ESAS-r: Renal items to the EQ-5D-5L index scores directly (linear models, censored dependent variable models, and mixture models) and response mapping models which mapped the ESAS-r: Renal items to the EQ-5D-5L health states (ordinal logistic regression). We performed internal validation to evaluate the power of prediction. Mean absolute error (MAE) and mean squared error (MSE) were calculated to compare the models' statistical fit on the estimation sample and predictive power on the validation sample. The final criteria for the preferred model included theoretical background, statistical fit, and predictive power.

Results

A total of 506 patient records were included for model estimation, after excluding missing records (missing rate: 57.6%). All models produced relatively similar statistical fit and predictive power (Estimation: MAE: 0.056 - 0.120, MSE: 0.007 - 0.028; Validation: MAE: 0.136 - 0.155, MSE: 0.032 - 0.046). All models performed better in terms of prediction properties for relatively healthy health states, but worse for poorer health states. Considering all criteria, the generalized estimating equations (Estimation: MAE: 0.120, MSE: 0.027; Validation: MAE: 0.140, MSE: 0.034) and generalized linear models (Estimation: MAE: 0.116, MSE: 0.028; Validation: MAE: 0.136, MSE: 0.034) on selected ESAS-r: Renal items were considered the best models.

Conclusions

Mapping algorithms from the ESAS-r: Renal to the EQ-5D-5L could be used to predict utility values for patients with ESRD when only ESAS-r: Renal is available. Future research should evaluate the generalizability of these mapping algorithms among ESRD patients.

3.1 Introduction

Chronic kidney disease (CKD) is defined as kidney damage for a period of greater than three months[1]. There are five stages of CKD, classified by the severity of kidney dysfunction (i.e. glomerular filtration rate), and a higher state indicates worse situation[2]. Stage 5 CKD is also known as end-stage renal disease (ESRD), which requires dialysis or kidney transplantation[3]. CKD is prevalent both in Canada and worldwide[4,5]. Data from 2019 Global Kidney Health Atlas showed that approximately 10% of the world's population were living with CKD, and around 0.1% of the world's population had ESRD[6]. The main burdens of symptoms of all stages of CKD include fatigue or lack of energy, feeling drowsy, pain, itchiness, and dry skin[7]. Patients living with CKD, especially those with ESRD, usually have a significant reduction in both life expectancy and health-related quality of life (HRQL)[8].

There are various instruments to identify symptom burdens of the patients with ESRD, such as the Edmonton Symptom Assessment System Revised-Renal (ESAS-r: Renal)[9] and the Palliative Outcome Score-Renal (POS-renal)[10]. These symptom assessment tools identify the extent and severity of symptom burdens of the patients[11], and they are shown to be responsive to changes in

symptoms[12]. Therefore, these measures are commonly used to assess patient outcomes in clinical trials and their symptom management. However, while symptom assessment instruments could capture the health outcomes of ESRD, the use of symptom assessment instruments in economic evaluation is limited by the comparability across different populations, as the results can only be compared with the evaluations that use the same measure. Furthermore, the ESAS-r: Renal does not provide an overall score, making it harder to use if it is selected as the effectiveness measure in economic evaluation.

Cost-utility analysis (CUA), where Quality-Adjusted Life Years (QALY) is the typical outcome, aims for a more universal comparability of evaluations across all technologies. Therefore, CUA remains the base-case analysis in many guidelines for economic evaluations[13,14]. QALY is a weighted sum of life years, and the weight is estimated by preferences (i.e. utilities) for health status. There are two methods to estimate utilities, direct elicitation, or indirect estimation using multi-attribute preference-based measures (PBM). The direct approach is a process of interviewing participants about their preferences for health states using techniques like time trade-off, which is time-consuming and burdensome[15]. For indirect approaches, participants answer the questions in the descriptive system of the PBM, and this will generate a profile or health status which has an associated utility value. The EuroQol-5 Dimension (EQ-5D) is such an instrument which could indirectly estimate utility values[16]. Both the three-level (EQ-5D-3L) and the five-level version (EQ-5D-5L) have become the most widely-used PBMs wordwide[16]. The EQ-5D has also been recommended as an outcome measure for patients with CKD[17].

PBMs are not included in every clinical study on ESRD, either because economic evaluation is not an objective of these studies or symptom measures like the ESAS-r: Renal have been used to better capture the disease-related outcomes. Therefore, there often remains a gap in using the effectiveness data of those trials for economic evaluation. Mapping is a tool to convert scores of non-PBM to utility scores which could be used in CUA using regression techniques[18]. To our knowledge, there is no mapping algorithm from the ESAS-r: Renal to the EQ-5D on patients with ESRD. We, therefore, aimed to develop a mapping algorithm from the ESAS-r: Renal to the EQ-5D-5L following the recommendations from the Mapping onto Preference-based Measures Reporting Standards Reporting Statement (MAPS reporting statement)[19] and the ISPOR Good Practices Task Force Report on Mapping (ISPOR Good Practices)[20].

3.2 Methods

Data

Data were obtained from the Evaluation of Routinely Measured Patient-reported Outcomes in Hemodialysis Care (EMPATHY) trial. It was a multi-centre clustered randomized-controlled trial in hemodialysis units in Northern Alberta. One arm of patients completed both the ESAS-r: Renal and the EQ-5D-5L every two months. In this arm, there were 188 patients, and in total 1193 patient records were collected over the one-year time frame. This was a short panel data, which means there were many individual units (clusters) but few time points[21]. Demographic and clinical information, including age, sex, race, and presence of diabetes were also collected. We used these data for estimating the mapping algorithm parameters. Another arm of patients in the EMPATHY trial voluntarily completed an anonymous outcome survey which had both the ESAS-r: Renal and the EQ-5D-5L every 6 months over the one-year time frame. There were 346 original records. Age and sex were collected in the survey, but race and diabetes were not included. We used these data for validation of the mapping algorithms. We considered this as internal validation, as the patients are from the same randomized-controlled trial and the patient characteristics would be very similar.

The variables considered in the models were ESAS-r: Renal items and demographic information. In the base case, we could not include race and diabetes in the mapping algorithms, as they were not collected in the validation set and we were not able to validate the results if we included them in the estimation models. However, we undertook a sensitivity analysis to see whether these two variables were additionally important to the final mapping algorithms.

Outcome Measures

The EQ-5D-5L is one of the most widely used generic PBMs[16]. It is an HRQL instrument and it has five dimensions: mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD). Each dimension has one question, asking the extent of problems the respondent is experiencing today on five levels (1=no problems, 2=slight problems, 3=moderate problems, 4=severe problems, and 5=extreme problems)[22]. This generates a total of 3125 ($=5^5$) possible health states. The perfect health state is "11111", while the worst health state

is “55555”. In Canada, there is a value set for the EQ-5D-5L based on the time trade-off method, and the estimated EQ-5D-5L utility values ranged from -0.148 to 0.949[23].

The ESAS-r: Renal is a symptom assessment scale used in the ESRD[12]. The original version of the ESAS-r covers 9 symptoms, including pain, tiredness, drowsiness, nausea, lack of appetite, shortness of breath, depression, anxiety, and wellbeing[24]. The renal-specific form has 3 additional questions on symptoms specific to kidney disease, i.e. itching, sleeping, and Restless legs[12]. The response to each question is a scale from 0 to 10, with 0 meaning no problem, and 10 the worst condition. The ESAS-r: Renal used by the EMPATHY trial asked for the average symptom burdens over the past week. Unlike some measures which could calculate a compound score[25,26], the ESAS-r: Renal does not have a single overall score.

There are three items in the ESAS-r: Renal measure similar constructs that the EQ-5D-5L measures. The ESAS-r: Renal pain item corresponds to the EQ-5D-5L PD dimension, and ESAS-r: Renal depression and anxiety items correspond to the EQ-5D-5L AD dimension. We performed a Spearman’s correlation test between the ESAS-r: Renal items and the EQ-5D-5L dimensions and utility scores to have an idea on the degree of overlap in constructs between the two instruments. A Spearman’s correlation of 0-0.3 is considered as negligible correlation; 0.3-0.5, mild correlation; 0.5-0.7, moderate correlation; 0.7-0.9, high correlation; and 0.9 to 1, very high correlation[27].

Estimation Model Types

Utility distribution is complicated and therefore it needs special modelling techniques. Previous evidence showed that utility distributions are bounded, skewed, multimodal, with large spikes on perfect health[20]. The EMPATHY EQ-5D-5L data had similar characteristics, where close to 20% of patient responses were on the level “11111” (Figure 3.1), and the distribution is left-skewed. Traditional linear models estimated by ordinary least square (OLS) may not be appropriate given the distributional characteristics[28,29]. Various methods have been used to deal with the issues of utility data, with two main types of modelling, direct mapping and indirect mapping (response mapping). The direct technique would map the ESAS-r: Renal item scores to the overall EQ-5D-5L utility index. The indirect technique would map the ESAS-r: Renal item scores to each dimension of the EQ-5D-5L descriptive system. For direct mapping, previous studies used regression models whose distribution requirements for the dependent variable are similar to the

utility distribution. Indirect mapping could potentially minimize the distribution issue, as it estimates the overall or dimensional responses.

In this study, we explored both direct and indirect (response) mapping techniques. We considered direct mapping techniques including linear models, censored dependent variable models, mixture models, and response mapping. The linear models include OLS, fixed effect model (FEM), random effect model (REM), generalized estimating equations (GEE), and generalized linear models (GLM). The censored dependent variable models include the Tobit model and Censored Least Absolute Deviations (CLAD) model. The mixture models include mixture beta regression model (BETAMIX) and adjusted limited dependent variable mixture model (ALDVMM). For the response mapping model, we used ordered logistic regression (OLR).

- Linear models: Using linear relationship to explain utility scores from the ESAS-r: Renal is a simple and explicit method. But there is no bound for many linear models, and predicted value out of the range needs to be set to the upper/lower limits. The traditional method to estimate the parameters of linear models is **OLS**. Under assumptions of the linearity, independence, normality, homoscedasticity, and no multicollinearity, the OLS estimator is the Best Linear Unbiased Estimator (BLUE) and Best Unbiased Estimator (BUE)[30]. But using OLS models on utility data is very likely to violate these assumptions, especially the normality and homoscedasticity assumptions[28,29,31,32]. For our dataset, aside from the properties of utility distribution that might violate those assumptions, our dataset was a short panel data and within-individual variation exists[21]. In this case, homoscedasticity would be a very strict assumption. Violating normality and homoscedasticity assumptions would make OLS estimators no longer the BLUE, and moreover, since the estimated standard errors are inaccurate, the hypothesis tests are invalid[30]. To deal with this issue, apart from switching to other model types, using a robust estimation of the variance could at least assure the validity of hypothesis tests under the situation of heteroscedasticity.

During the modelling process, we first started with the traditional OLS and checked the assumptions. If the assumptions were violated, we then considered a robust estimation. Considering that the data were clustered by individual, the cluster-robust variance would be suitable for our data. The Stata option “vce(cluster)” calculates the cluster-robust variance for

regression models[33]. For the other types of models below, we also used the cluster-robust variance to ensure the validity of hypothesis tests when model assumptions were violated or hard to test.

Individual effect and time effect may exist as our dataset was a short panel dataset. OLS model with cluster-robust variance treated the data in a pooled cross-sectional way, which ignores the two effects. One method to deal with this is to directly run a population-averaged model, using **GEE**[34]. The other method is to estimate the individual effect and time effect using **FEM** and **REM**[35]. For time effect, a variable "time" was included in the model. For individual effect, the FEM deals with this by assigning an individual intercept to each subject, and REM allows each subject to have an individual error term. After the regression estimation, the Hausman test could be used to test whether using the fixed or random effect model is preferred[36].

GLM is a generalized form of linear models. Instead of subjecting to the normal distribution, the conditional mean distribution for GLM models could subject to any distribution within the exponential family, while the independent variables still have a linear combination. The error terms in many of the GLMs do not need to conform to the normal distribution and they could be heteroscedastic. The exponential family includes normal distribution, Poisson distribution, Gamma distribution, etc. In many GLM models, the dependent variables are non-negative. To use GLM in the estimation of utility data, disutility values were typically used. Disutility is defined as the upper limit of utility values minus utility values. The disutility values would always be non-negative, while utility may have some negative values. Previous literature suggested Gamma distribution with an identity link, and normal distribution with a log link, are ideal to model the disutility data[33,37]. These two models could partially deal with the bounded issue as the estimated non-negative disutility values could ensure the estimated utility values not greater than the upper limits. The difference between the two models is that the normal model requires equal variances, but the gamma model allows heterogeneity. In our study, we continued to use these two methods. A preliminary comparison was made between the two models to choose one for the final performance assessment.

- **Censored dependent variable models: Tobit model** and **CLAD model** are designed for limited dependent variables[35], such as in the case of bounded utility data[25,38,39]. In the modelling process, the range of the estimated value is pre-designated, and usually is the same as the range

of the value set. The main model equation still has a linear form, but any estimated values out of the range would be replaced by the upper/lower limits. The difference between Tobit and CLAD models is that they use different parameter estimation methods. Tobit uses maximum likelihood estimation, which requires the error terms to be identical and independent, normally distributed with the same variances, while CLAD uses least absolute deviations (derived from the quantile regression), which only requires the error terms to be identically and independently distributed[35]. Therefore, the use of Tobit is more stringent. Previous literature showed that the assumptions of the Tobit model might be violated, and in this case, CLAD provides reliable estimation and could get rid of the issues[25,38,39]. In this paper, both Tobit and CLAD were considered. The assumptions for Tobit were checked, and if violated, we used the cluster-robust variance estimator for Tobit.

- **Mixture models:** Mixture model is able to deal with two or more issues in utility distribution together. The principle of mixture models is assuming there are sub-distributions within the data. To link the sub-distributions together, a logistic or multinomial function is used for classification, and it calculated the probabilities of the observation belonging to each sub-distribution. The mixture model is an advanced form of the two-part model. With sub-distributions, it could solve the multimodality, as each sub-distribution would have a mode. The bounded issue is solved using a similar method as Tobit and CLAD. Lastly, skewness is indirectly solved by controlling the density of each sub-distribution. The final estimated utility value is an expected mean of all sub-distribution estimated values. Two popular mixture models that have been used for mapping studies are **BETAMIX**[40] and **ALDVMM**[41]. The BETAMIX considers several components of beta regression[40], while the ALDVMM considers several Tobit-like distributions[41]. They have user-developed Stata programs that could directly be used for estimation. For our study, we considered both models for the estimation. Many of the assumptions for mixture models are not directly testable[42], and we used cluster-robust variance in case there was any violation in assumptions which we failed to test and identify.

- **Dimensional response mapping:** In response mapping, the scores of the source PROM are mapped to the dimensional responses of the target measure. We used ordinal logistic regression (**OLR**) to map the ESAS-r: Renal items to each of the EQ-5D-5L dimensions. The dependent variable of OLR is an ordered categorical variable, which is suitable for modelling the

dimensional responses of the EQ-5D-5L. These regression models calculate a probability of selecting "k" (k=1, 2, 3, 4, 5) within dimension "y". The product of the probabilities on each dimension is the probability of an EQ-5D-5L health state. For example, the probability of a respondent reporting a health state of "11111" with a given ESAS-r: Renal profile would be the product of the probability of level 1 on MO given ESAS-r: Renal profile, the probability of level 1 on UA given ESAS-r: Renal profile, the probability of level 1 on SC given ESAS-r: Renal profile, the probability of level 1 on PD given ESAS-r: Renal profile, and the probability of level 1 on AD given ESAS-r: Renal profile. The final estimated utility score was an expected average of the utility scores from all possible EQ-5D-5L health states.

Estimation Methods

For each of the estimation methods above, four model specifications were considered:

- all ESAS-r: Renal symptoms (model 1);
- all ESAS-r: Renal symptoms plus age and gender (model 2);
- significant ESAS-r: Renal items selected by backward step-wise method (model 3);
- significant ESAS-r: Renal items and significant demographic variables selected by backward step-wise method (model 4).

In models 3 and 4, we used a backward stepwise selection procedure, which is considered the best selection procedure for the model building when important independent variables are intended to be retained[43]. Pain, depression and anxiety of ESAS-r: Renal were considered important independent variables, as they have the same constructs as PD and AD of the EQ-5D-5L, so they would not be excluded even if they were not statistically significant in the backward step-wise procedure. Similar for age, a recommended demographic information variable to include in the mapping algorithm[20], and we would not exclude it if it was statistically insignificant.

All regression analyses were conducted in Stata version 14 (Stata-Crop, College Station, TX, USA). All estimated utility scores greater than the upper limits of the Canadian value set (0.949) were converted to 0.949, and those smaller than the lower limits (-0.148) were converted to -0.148. For response mapping, Stata could only calculate the probabilities of the selections of

each dimension. The combination of the health states and calculation of utility scores are conducted in Python (Python Software Foundation, DE, USA).

Assessing Model Performance and Selection

We did a preliminary selection within each model type and then systematically assessed all of the selected models. During this preliminary selection process, models which violated the assumptions or had worse fit statistics than alternative models being compared with were excluded. The fit statistics we used were the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Both statistics assess the goodness of fit of the model with a penalty on the increasing number of estimated parameters[44,45]. The penalty term in BIC is larger than in AIC, and therefore BIC tends to support models that are more parsimonious[45]. The smaller the AIC and BIC are, the better the model fit is. After we did the preliminary selection, the preferred models for each type were compared together using fit statistics to see their statistical fit in the estimation sample and the predictive power in the validation sample. Overall and conditional fit statistics were used. Conditional fit statistics were also considered as it could show whether the errors were affected by poor health[46]. Previous literature suggested reporting fit statistics for subsets, such as for EQ-5D $\leqslant$ 0.5 and EQ-5D > 0.5[46]. The fit statistics we chose were mean squared error (MSE) and mean absolute error (MAE), both of which are common among published mapping studies[46]. MSE is the average of the squares of the errors, and here error means the difference between the observed EQ-5D-5L value and the estimated/predicted EQ-5D-5L value. MAE is the average of the absolute value of the errors. AIC and BIC were not considered as model selection statistics, because they are not applicable for response mapping, as the regression models only estimate the dimensions. Some models would have a raw estimated utility score out of the utility range of the Canadian value set, and we calculated the percentage of the out-of-bound estimation as an additional criterion of model fit and prediction power. The final selection of the model involved considerations of the theoretical background of the model, model fit, and prediction power.

3.3 Results

Demographics and Outcome Distributions

The final estimation sample had 506 patient records and the validation sample, 242 patient records, after removing the records that had missing values in the ESAS-r: Renal, the EQ-5D-5L, or demographic information (Table 3.1). Participants in the validation sample were slightly older than the participants in the estimation sample (estimation sample: (mean=64.4, s.d.=14.2), validation sample: (mean=65.5, s.d.=14.3)). The age and gender of the validation sample were comparable with the estimation sample. Based on the estimation sample, the participants in this trial were mostly diabetic. Caucasian was the main ethnicity of the participants, followed by Asian and Aboriginal. Participants in the estimation sample had a relatively better patient-reported HRQL compared with the participants in the validation samples, according to the average EQ-5D-5L index, the proportion of participants in perfect health, the proportion of participants having an EQ-5D-5L index less than 0.5, and the proportion of participants having at least one moderate or severe ESAS-r: Renal symptom (Table 3.1).

Pain, tiredness, drowsiness, and wellbeing were the most frequent and severe symptoms that patients were experiencing, as measured by the ESAS-r: Renal (Figure 3.2). Itchiness and sleep were also serious symptoms for these patients. For these key symptoms, less than 40% of the patients reported no problem ("0" in the scale), and more than 20% of patients reported moderate to severe problems ("5" and above in the scale).

The five dimensions of the EQ-5D-5L had a similar diminishing distribution (Figure 3.3), with decreasing frequencies when going from "no problems" to each of the more severe problem level. Participants had fewer problems in SC compared with other dimensions, with over 60% of the participants having "no problems". MO, UA, and PD were three key dimensions where participants had at least some problems (about 60%). In these dimensions, more participants had "severe problems" or "extreme problems" compared with SC and AD. Overall, the ceiling effect, i.e. respondents reporting "perfect health", was about 20%, and the flooring effect, i.e. respondents reporting "worst health" was almost 0.

Pain in the ESAS-r: Renal was highly correlated with the pain dimension of EQ-5D-5L (ρ = 0.81, $p<0.001$). The ESAS-r: Renal anxiety ($\rho=0.62$, $p<0.001$) and depression ($\rho=0.70$, $p<0.001$)

items were each moderately correlated with the anxiety/depression dimension of EQ-5D-5L (Table 3.2). These three items were also mildly to moderately correlated to the EQ-5D-5L index (pain: ρ=-0.58, depression: ρ= -0.50, anxiety: ρ=-0.44). For the other items that do not have directly the same constructs as the EQ-5D-5L dimensions, mild to moderate correlations were observed with the EQ-5D-5L index (8/9).

Preliminary Selection of Models Within Each Type

- **Linear models**: We started with the **OLS** estimation with all ESAS-r: Renal items. Both assumptions of normality and heteroscedasticity were violated in our dataset (Appendix 3.1). Therefore, a linear model with cluster-robust variance was considered. We then explored **FEM** and **REM** to see if there was any individual effect or time effect. The time effect was not significant, even in crude regression analysis (Appendix 3.2). We then used the model predicted by all ESAS-r: Renal items as an example to illustrate the process of exploring individual effects (Appendix 3.3). We first tested whether FEM and REM were better than the pooled linear regression model, and we did observe the individual effect. The Hausman test showed that the FEM is better than REM for our dataset. Finally, since heteroscedasticity existed in our data, cluster-robust variances were used.

For **GLM**, the dependent variable had various choices of distributions from the exponential family. Appendix 3.4 presents a table of the AIC and BIC statistics of models using gamma distribution with identity link and normal distribution with a log link. Since the likelihood functions were not concave for model specification 1 and 2 using gamma distribution with identity link, we chose normal distribution with a log link. For model specification 3 and 4, based on the fit statistics, gamma distribution with identity link was better than normal distribution with a log link, and therefore we chose gamma distribution with identity link. The normal model still required the equal variance assumption, while the gamma model did not have that requirement. As the equal variance assumption was violated for the normal model (Appendix 3.5), we used robust variance estimation. Besides, considering that our data were individually clustered, we used cluster-robust variance for the 4 GLMs.

In summary, the linear models we explored for further performance comparison were: 1) OLS with robust estimation, 2) FEM with robust estimation, 3) GEE, and 4) selected GLM. OLS with

robust estimation was not as good as models considering individual effect but it was still included, as this is a very basic regression model.

- **Censored dependent variable models**: We did not perform a preliminary comparison between **Tobit** and **CLAD**, as we intended to retain both of them for further assessment. However, we did an assumption check for Tobit, which supported using the cluster-robust variance estimation as the assumptions were violated (Appendix 3.6).

- **Mixture models**: Within mixture models, the number of components, or sub-distributions needed consideration. The estimation functions for the two types of mixture models with three components were not concave. The one-component models were better than the two-component models for **ALDVMM** (Appendix 3.7), and the two-component models were better than the one-component models for **BETAMIX**. In the final assessment, ALDVMM with one component and beta mixture models with two components were considered for further evaluation.

- **Response mapping:** We only used model specification 4 (selected ESAS-r: Renal items and demographic information) for response mapping. This was because the model equations in response mapping were dimensional. It was considered meaningless to include independent variables that did not have potential similar constructs with the dependent variable in the regression model. Specifically, the pain item of the ESAS-r: Renal was an important variable for estimating PD of the EQ-5D-5L, and the anxiety and depression of ESAS-r: Renal were important variables for estimating AD of the EQ-5D-5L.

Final Results and Model Selection

Overall, nine estimation methods were explored, using four different model specifications, and producing 33 models in total. Overall, the model fit in the estimation sample and the predictive power in the validation sample were similar across all models (Table 3.3). The MAE ranged from 0.056 to 0.120 in the estimation sample (models with MAE < 0.110 were highlighted), and 0.138 to 0.155 in the validation sample (models with MAE < 0.141 were highlighted). The MSE ranged from 0.007 to 0.028 in the estimation sample (models with MSE < 0.024 were highlighted), and 0.032 to 0.055 in the validation sample (models with MSE < 0.038 were highlighted). As such, within each statistical criterion, 9-10 models were highlighted. This

indicates that the highlighted models performed better when compared with the rest of the models based on the corresponding statistical criterion.

FEM performed the best in terms of the four statistics, followed by CLAD models, which had relatively small MAE in the estimation and validation sample, and MSE in the validation sample. GEE models tended to perform better when assessed by prediction power, and they fit within the estimation sample was slightly worse than the other highlighted models, but still acceptable. Tobit models, BETAMIX, and response mapping tended to perform better when measured by statistical fit, but their prediction powers were relatively poor. The GLM models used different exponential family distributions across the four model specifications. GLM1 and GLM2 used normal distribution, where the statistical fit was good, but the prediction power was poor. GLM3 and GLM4 used gamma distribution, where the predictive power was ideal, but the statistical fit was not as good as the highlighted models in Table 3.3, but still acceptable. Statistics for the ALDVMM and linear models were mostly fair, with the ALDVMM slightly better than the linear models. Lastly, the percentage of raw estimated values out of the scale range of the Canadian value set was relatively low for most of the models, with OLS and FEM having the most values out of the range (Table 3.3).

Based on the analysis above, we selected FEM (model specification 3), GEE (model specification 3), CLAD (model specification 1), GLM with a gamma distribution (model specification 3), and ALDVMM (model specification 3) for further assessment. We plotted the observed utility values versus predicted values in Figure 3.4 (estimation sample) and Figure 3.5 (validation sample). We further looked at the MAE and MSE for EQ-5D-5L subsets, i.e. EQ-5D-5L < 0.5 and EQ-5D-5L >=0.5. The ALDVMM had a good prediction for poor health (EQ-5D-5L<0.5), as the MAE and MSE statistics were mostly the lowest in the estimation and validation sample (Table 3.4). But this was a result of the systematic underprediction of the EQ-5D-5L values (Figure 3.4). Also, the MAE and MSE statistics of the ALDVMM for the EQ-5D-5L>=0.5 subset were the largest (Table 3.4). The CLAD model had the opposite situation. It predicted well for good health and not well for poor health. In terms of MAE and MSE statistics, CLAD had the lowest statistics at the EQ-5D-5L>=0.5 subset, and the largest statistics at the EQ-5D-5L < 0.5 subset (Table 3.4).

The FEM is a special situation because it accounted for the individual effect. In the estimation sample, it had a perfect prediction. This indicated that there were significant unobserved individual effects which could not be explained by the independent variables. Then in the validation sample, the fixed effect model had a very poor prediction. The estimated values almost lied within 0.4 to 0.8 (Figure 3.5).

GEE and GLM models were the two best models, demonstrating the characteristics of the utility distribution, and having neither consistent underprediction nor overprediction. The MAE and MSE statistics were similar in the estimation and validation sample for the two EQ-5D-5L subsets. Therefore, we selected both GEE and GLM as the final models to predict the EQ-5D-5L from the ESAS-r: Renal (Figure 3.6 and Appendix 3.8). The calculation of predicted EQ-5D-5L utility values for GEE and GLM models was shown in Table 3.5. For example, for a given ESAS-r: Renal profile with no symptoms in every item (all "0"), the GEE-predicted value would equal to 0.899 (= 0.899 – 0.017*0 – 0.009*0 – 0.007*0 – 0.012*0 – 0.014*0 – 0.012*0 + 0.007*0). The GLM-predicted disutility value would equal to 0.026 (= 0.026 + 0.022*0 + 0.011*0 + 0.007*0 + 0.016*0 + 0.015*0), thus the GLM-predicted utility value would be 0.923 (= 0.949 – 0.026). The variance-covariance matrices and error terms were reported in Appendix 3.9.

Sensitivity analyses

We performed sensitivity analyses to check if including diabetes and race as independent variables improved the GLM and GEE models. Diabetes and race were statistically significant in the models, and the new GEE and GLM models had smaller error statistics (MAE: 0.101-0.108, MSE: 0.019-0.021) compared with the GLM3 (MAE: 0.116, MSE: 0.028) and GEE3 (MAE: 0.119, MSE: 0.027), but the difference was small (Appendix 3.10). Another issue was that the predictive power of the models with demographics (e.g. GLM2, GEE2) was not as good as the predictive power of GLM and GEE models without demographics (Table 3.3), especially for GLM2. This suggests that including more demographic variables could help to explain the EQ-5D-5L index score in the estimation set, but this may reduce the generalizability on the validation set.

3.4 Discussion

The ESAS-r: Renal is a simple and validated symptom assessment scale that has been widely used to identify and capture the symptom burdens of patients with ESRD.[3,9] It does not, however, produce an overall summary score that is based on population preferences, and therefore can not be incorporated into QALY calculations and CUA. This study explored mapping algorithms that would allow the estimation of EQ-5D-5L index scores from the ESAS-r: Renal, which could support such applications. We followed established mapping guidelines to conduct our study[19,20], and various methods that have been used in the current mapping studies were considered. We identified two models to predict utility values from ESAS-r: Renal for conducting CUA.

Among the methods we tested, the best models were GEE model and GLM models with gamma distribution and identity link on selected ESAS-r: Renal items. These two selected models had comparable fit statistics and prediction power. GLM provided slightly higher estimated values at the upper end of the utility-scale (relatively healthy states) and lower values at the lower end of the scale. Both models were in linear form, which makes them relatively easy to interpret. Meanwhile, they had less stringent model assumptions. The GEE model estimates the population-averaged result, averaging the utility deterioration estimated by ESAS-r: Renal items. There was no distribution assumption within GEE models. In the GLM model with gamma distribution and identity link, we modelled the disutility data. This combination allows the dependent variable to be consistently non-negative, and the error term could be heterogeneous. One issue with the GEE model was that “Restless legs” had a positive coefficient. “Restless legs” was not a common symptom burden in our dataset, and only a mild correlation was observed between “Restless legs” and EQ-5D-5L utility values. The model coefficient of “Restless legs” was statistically significant but unlikely to be clinically significant as the effect size is very small. Based on the above information, we recommended using the GLM on selected ESAS-r: Renal items as the primary model in future economic evaluations, while the GEE could be used in sensitivity analyses.

Recently, Moskovitz et al explored the possibility of replacing EQ-5D-3L-derived health state utilities with the general form of the ESAS[47]. They did not provide a specific algorithm from the ESAS to the EQ-5D-3L, but they mapped the pain, depression, and anxiety of the ESAS to PD

and AD of the EQ-5D-3L and generated an ESAS-derived EQ-5D utility score. The Spearman's correlations between the ESAS-derived EQ-5D utility scores and the original EQ-5D-3L utility scores were high (0.83-0.91)[47]. Although the instruments they used were not exactly the same as ours, some of their findings and issues were similar to ours. The ESAS/ESAS-r: Renal items that have the same constructs as the EQ-5D, (i.e. pain, anxiety, depression) were moderately to highly correlated to the corresponding EQ-5D domains and the overall indices. Some constructs of the EQ-5D were not covered in the ESAS or the ESAS-r: Renal, i.e. SC, UA, and MO, and these will weaken the ability to predict utilities from the ESAS. In other words, when using the ESAS to predict utility values, the deterioration in utility values induced by having problems in SC/UA/MO was ignored, as the ESAS did not capture the problems in these three dimensions. This led to an overprediction, with predicted utility values persistently greater than the observed values. For patients having more severe problems in these three dimensions, the extent of over-prediction would be greater. Therefore, it was also expected that the ESAS-r: Renal would not predict the EQ-5D-5L very well, especially for the lower-end of the EQ-5D-5L scale.

Overprediction of utility values in the lower-end of the EQ-5D-5L was common in previous studies[38,48]. Aside from lacking enough overlaps in dimensions/constructs discussed above, another reason for this issue was the sample size. Ideally, the observed values should scatter evenly around the regression line, with some observed values greater than the predicted values, and some observed values smaller than the predicted values. Many studies had a relatively small sub-sample with poor health and a large number of observations accumulating with higher values of the EQ-5D. The subjects with higher values of the EQ-5D would have more weight in the estimation function as there were more observations compared to relatively few EQ-5D states with lower values. Therefore, the optimization would favour the pattern among subjects having higher EQ-5D values, and the predicted relationship between the independent variables and dependent variables would be better for observations with higher values of the EQ-5D. Even though there were huge deviations between the predicted and observed values in the lower-end of the EQ-5D, with only a small proportion of such records, it would not affect much in determining the best coefficients estimators.

There were substantial unobserved individual effects among the relationship between the ESAS-r: Renal and the EQ-5D-5L. The perfect performance of FEM in the estimation sample and poor

prediction in the validation sample supported this argument. The FEM used population equations when doing prediction in the validation sample, and the estimated population values had many deviations from the observed EQ-5D-5L utility values. This individual effect could be impacted by the problems in the SC, UA, and MO dimensions which were discussed above, or it could be the difference in the time frame between the ESAS-r: Renal and the EQ-5D-5L. The ESAS-r: Renal used by the EMPATHY trial asked for the average symptom burdens over the past week, while the EQ-5D-5L measured health only in the current day. While the ESAS-r: Renal measured the symptoms in a more persistent way, the utility values predicted by the EQ-5D-5L were more applicable to the immediate situation. It is possible that the average symptoms could not well explain the "immediate health", and this unexplained variability fell into unobserved individual effects.

We did observe common flaws in popular modelling techniques when using them in our dataset. The first was that linear models could not capture the ceiling effect[32,49]. In the estimation, we arbitrarily replaced all estimated values greater than the upper limits with upper limits. The second is the assumption violations in OLS estimation and Tobit models, which is commonly-reported in the previous studies[28,29,32,50]. In this study, homoscedasticity and normality assumptions were also violated for both OLS and Tobit models. We used robust estimation to ensure the validity of hypothesis testing. In this case, the coefficient estimators themselves were not the best estimators to describe the linear- or Tobit-like relationships between the dependent variable and independent variables. The inefficiency of traditional regression techniques was due to the characteristics of the utility data, which were bounded, skewed, and had large spikes in perfect health[20].

There have been various new methods aimed to deal with the issues of the utility distribution, and we tested most of them in our study but did not observe better performance. The response mapping approach seemed to over-fit the data. It had a good prediction in the estimation sample, but it did not perform well in the validation sample. The limited construct overlaps between the ESAS-r: Renal and the EQ-5D-5L made it meaningless to predict MO, SC, and UA dimensions from the ESAS-r: Renal items. The estimated regression models for MO, SC, and UA were hard to extrapolate. Mixture models were designed to concurrently solve the common issues of utility data. In our sample, the ALDVMM model solved the bounded and skewness issue to some

degree. But the systematic underpredictions of the higher values of the EQ-5D were also significant. We only had one component in the ALDVMM model, and the absence of another potential sub-distribution dragged down the estimated values for the observations with higher EQ-5D values. Strictly, for mixture models, there should be two components or more[41]. The one-component setting is similar to a Tobit model, which disobeys the underlying distributional requirements of mixture models. The reason which two-component models are not preferred by AIC and BIC may be due to limited observations to support multiple sub-distributions. Then in the BETAMIX model, two components were used. Again, good statistical fit but poor prediction power was observed, which indicated overfitting issue.

The strength of our study was that we carefully followed the existing mapping guidelines[19,20] and explored all of the commonly-used mapping techniques to develop the mapping algorithm. Therefore, the quality of the analysis and the reporting quality of the study was ensured. For example, the explicit information we gave on the estimation sample could help further studies to identify whether our algorithms are applicable to their population. Second, the two models we recommended were in linear terms, which were relatively user-friendly. One recommended algorithm did not include renal-specific items in the final model. This indicated that this model might be able to predict the EQ-5D-5L from the general form of ESAS, but this had to be validated in other disease samples.

The main limitation of our study is that our dataset includes a lot of missing data, both in the estimation and validation samples. In addition, due to the limitation of the validation sample, we could not consider diabetes and race in the mapping algorithm. These two variables were shown to be statistically significant in the sensitivity analysis. Although the statistical fit of the models with these two variables improved, we were doubtful that the prediction power would also improve.

The algorithms were only internally validated. Therefore, to apply these mapping models in future studies, there were several points that need further attention. Although demographic information was not included in the final model, the population to be applied should be comparable with this study. Here, the estimation sample was formed mostly by older people, and the majority of them had diabetes. The main symptom burdens among the population were pain, drowsiness, tiredness, and wellbeing. Itchiness and sleeping were common but not the most

severe symptoms in our dataset. This was probably due to the good management of ESRD, where the renal-specific symptoms were under control. To assess the generalizability of the mapping models, external validation is needed. Especially, given that we observed an unexpected sign for "Restless legs", using a population where the prevalence of Restless leg symptoms is higher to do a further external validation is important.

3.5 Conclusions

This study has developed a mapping algorithm from the ESAS-r: Renal to the EQ-5D-5L in patients with ESRD. This helps to predict utility values when PBMs are not available for economic evaluation. The algorithm is likely to be robust for the population which is comparable to our estimation sample. The further steps are to evaluate the generalizability of this mapping algorithm to other patient populations and to explore possible mapping algorithms between the generic form of the ESAS and the EQ-5D-5L.

Acknowledgment

I would like to acknowledge the contributions of the following individuals for designing, planning, and implementing the EMPATHY Trial in Alberta Kidney Care - North: Buzinski R, Davison S, Duperron P (deceased), Johnson JA, Klarenbach S, Nhan J, Short H, Wasylynuk BA.

Funding

EMPATHY is a project of the Can-SOLVE CKD Network, supported by the Canadian Institutes of Health Research under Canada's Strategy for Patient-Oriented Research (reference #SCA-145103). This large network supports a large number of research projects, including the EMPATHY Trial, and provides basic infrastructure and resources.

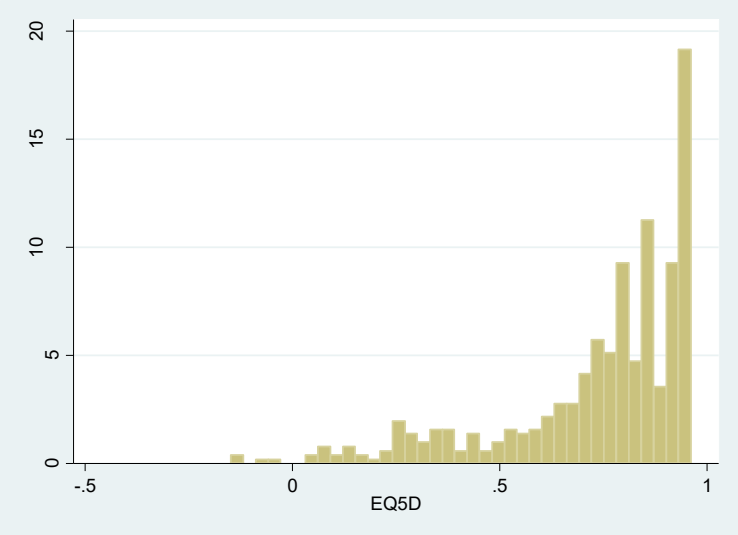


Figure 3.1 EQ-5D-5L utility value distribution in the final estimation sample (patient records=506)

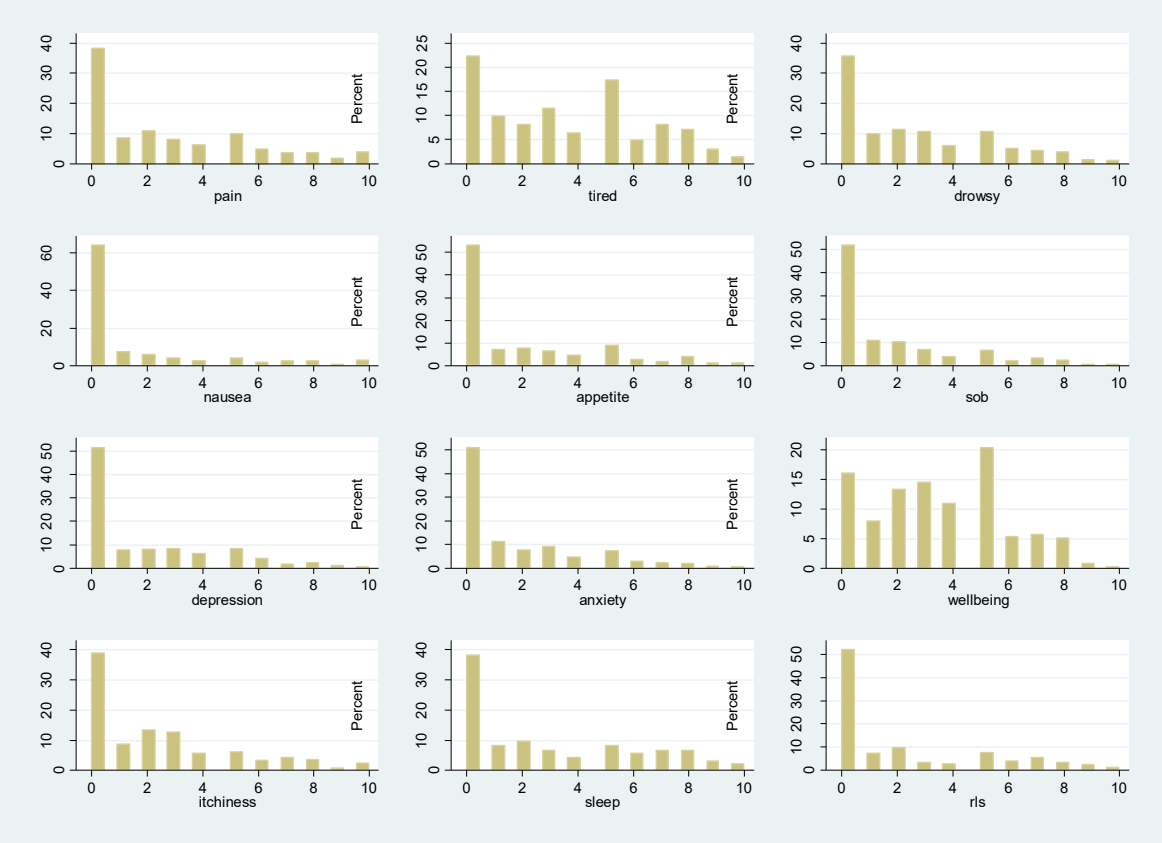


Figure 3.2 The distributions of the ESAS-r: Renal by items in the estimation sample (patients records=506)

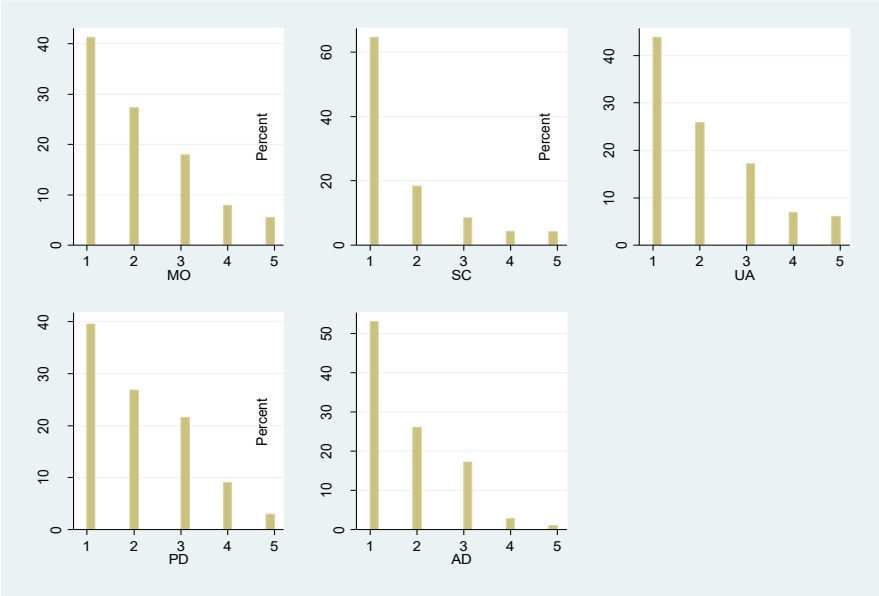
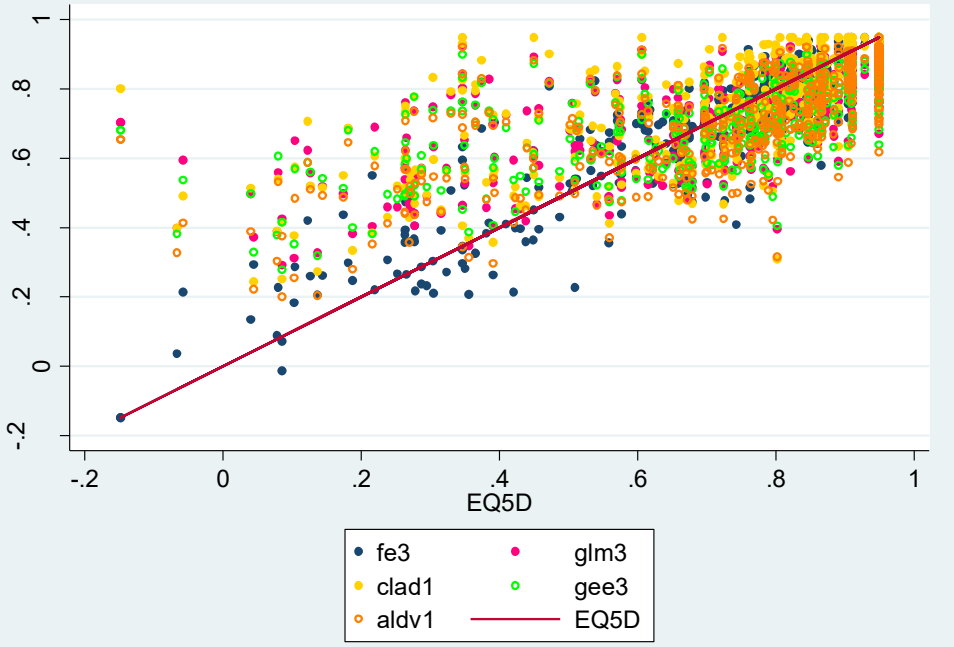
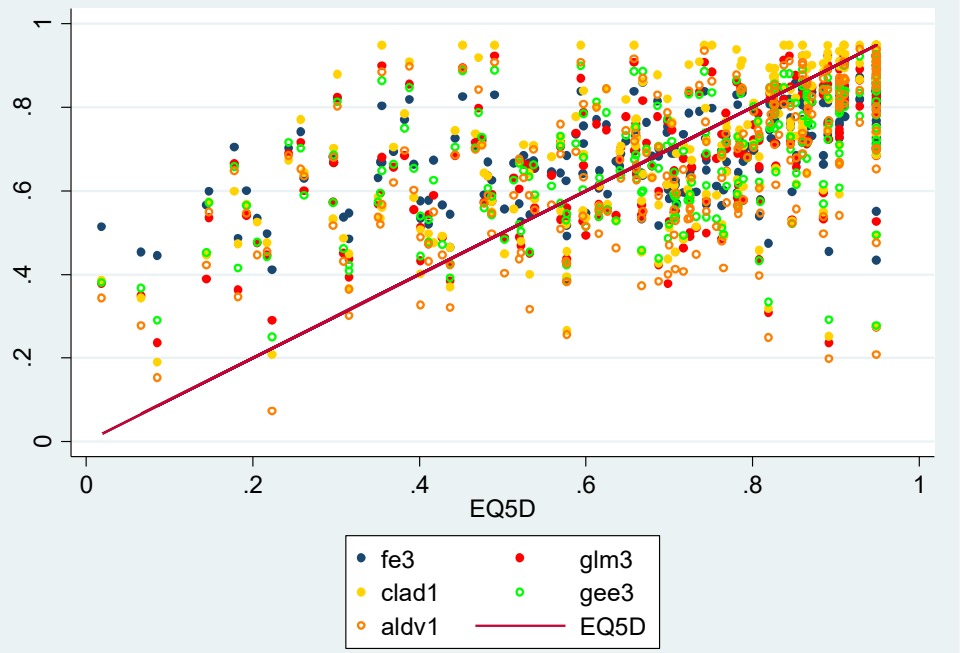


Figure 3.3 The dimensional distributions of the EQ-5D-5L in the estimation sample (patient records=506, MO: mobility, SC: self-care, UA: usual activities, PD: pain/discomfort, AD: anxiety/depression)



Note: fe3: fixed effect model on selected ESAS-r: Renal items, glm3: generalized linear model with gamma distribution and identity link on selected ESAS-r: Renal items, clad1: censored least absolute deviations model on full ESAS-r: Renal items, gee3: generalized estimating equations model on selected ESAS-r: Renal items, aldv1: adjusted limited dependent variable mixture model on full ESAS-r: Renal items

Figure 3.4 Observed EQ-5D (x-axis) versus Predicted EQ-5D (y-axis) in the estimation sample



Note: fe3: fixed effect model on selected ESAS-r: Renal items, glm3: generalized linear model with gamma distribution and identity link on selected ESAS-r: Renal items, clad1: censored least absolute deviations model on full ESAS-r: Renal items, gee3: generalized estimating equations model on selected ESAS-r: Renal items, aldv1: adjusted limited dependent variable mixture model on full ESAS-r: Renal items

Figure 3.5 Observed EQ-5D (x-axis) versus Predicted EQ-5D (y-axis) in the validation sample

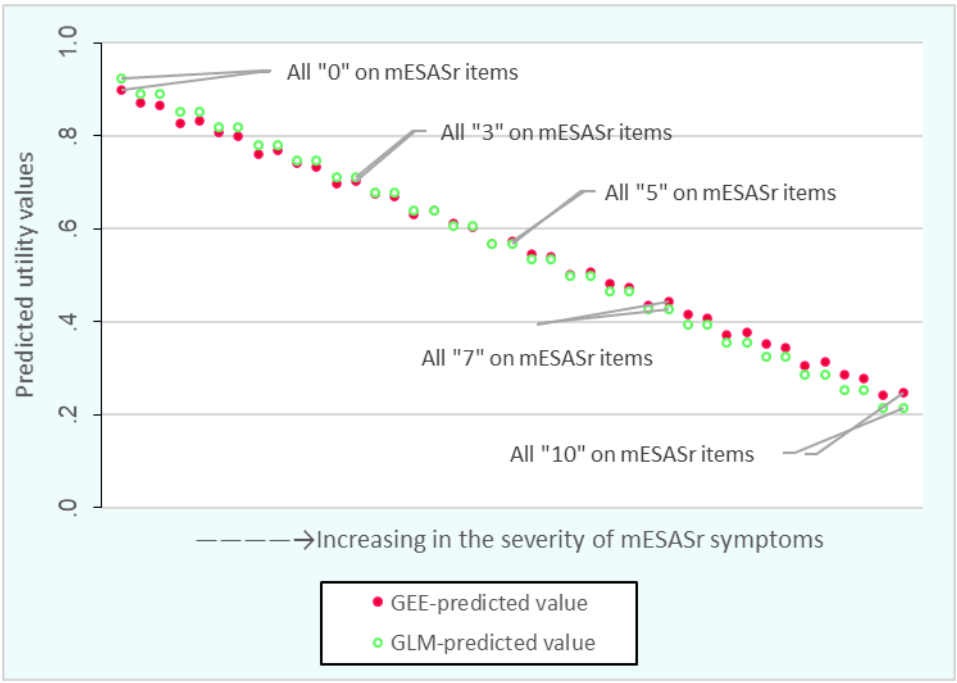


Figure 3.6 Predicted EQ-5D-5L utility values by the generalized linear model and the generalized estimating equations on selected ESAS-r: Renal profiles

Table 3.1 Patient characteristics information

	Estimation set (Patient records=506)	**Validation set (Patient records=242)**
Age - years	64.42 (s.d. = 14.19)	65.48 (s.d.=14.30)
Female sex	45.8 %	47.9%
Diabetes	78.5%	N/A
Ethnicity	Caucasian: 47.2%	
	Asian: 12.4%	
	Aboriginal: 11.66%	
	Black: 6.3%	
	Pacific: 5.3%	
	Indian subcontinent: 4.0%	
	Latin American: 2.3%	
	Mid-eastern: 4.0%	
	Multiracial: 0.40%	
	Unknown: 10.7%	
EQ-5D-5L index score	0.7502 (s.d. = 0.22)	0.6959 (s.d. = 0.22)
EQ-5D-5L in perfect health	19.7%	9.9%
EQ-5D-5L index less than 0.5	13.54%	19.0%
Having at least one moderate or severe symptoms ('>5') in ESAS-r: Renal	53.4%	67.4%

Table 3.2 The Spearman's correlation between the ESAS-r: Renal and the EQ-5D-5L dimensions/utility index

	MO	UA	SC	PD	AD	EQ-5D-5L Utility index
Pain	0.40**	0.37**	0.23**	***0.80****	0.29**	-0.59**
Tiredness	0.38**	0.50**	0.34**	0.47**	0.39**	-0.57**
Drowsiness	0.32**	0.44**	0.29**	0.41**	0.37**	-0.49**
Nausea	0.21**	0.23**	0.12	0.43**	0.28**	-0.36**
Lack of Appetite	0.20**	0.33**	0.18**	0.36**	0.34**	-0.38**
Shortness of Breath	0.32**	0.35**	0.30**	0.32**	0.29**	-0.39**
Depression	0.31**	0.41**	0.19**	0.38**	***0.72****	-0.50**
Anxiety	0.30**	0.37**	0.21**	0.35**	***0.62****	-0.47**
Wellbeing	0.44**	0.50**	0.34**	0.43**	0.44**	-0.57**
Itchiness	0.12*	0.15	0.09	0.16**	0.26**	-0.18**
Sleep	0.23**	0.34**	0.16**	0.42**	0.34**	-0.41**
Restless Legs	0.13*	0.16**	0.10	0.29**	0.26**	-0.27**

Note: Patient records=506, Bold and italic indicates similar constructs. MO: mobility, SC: self-care, UA: usual activities, PD: pain/discomfort, AD: anxiety/depression. **: significant at $p<0.001$, *: significant at $p<0.05$

Table 3.3 Statistical fit and prediction power of the models

	Statistical fit			Prediction power		
Model	MAE	MSE	OFR %	MAE	MSE	OFR %
Ordinary least square 1	0.112	0.025	17 (3.4%)	0.147	0.040	8 (3.3%)
Ordinary least square 2	0.110	0.024	25 (4.9%)	0.153	0.042	15 (6.2%)
Ordinary least square 3	0.112	0.026	18 (3.6%)	0.146	0.039	10 (4.1%)
Ordinary least square 4	0.110	0.024	30 (5.9%)	0.152	0.042	14 (5.8%)
Fixed effect model 1	0.057	0.007	32 (6.3%)	0.139	0.033	0
Fixed effect model 2	0.056	0.007	33 (6.5%)	0.147	0.037	12 (5.0%)
Fixed effect model 3	0.057	0.008	26 (5.1%)	0.139	0.032	0
Fixed effect model 4	0.057	0.008	29 (5.7%)	0.139	0.033	0
Generalized estimating equations 1	0.120	0.027	0	0.139	0.034	0
Generalized estimating equations 2	0.117	0.026	0	0.146	0.038	3 (1.2%)
Generalized estimating equations 3	0.120	0.027	0	0.140	0.034	0
Generalized estimating equations 4	0.117	0.026	7 (1.4%)	0.147	0.038	3 (1.2%)
Tobit model 1	0.112	0.025	0	0.149	0.041	0
Tobit model 2	0.109	0.023	0	0.154	0.043	0
Tobit model 3	0.112	0.025	0	0.148	0.041	0
Tobit model 4	0.110	0.023	0	0.153	0.043	0
Censored least absolute deviation model 1	0.103	0.028	0	0.138	0.037	0
Censored least absolute deviation model 2	0.101	0.026	0	0.145	0.041	0
Censored least absolute deviation model 3	0.104	0.028	0	0.137	0.037	0
Censored least absolute deviation model 4	0.101	0.026	0	0.147	0.042	0
Generalized linear model 1 (normal + log link)	0.116	0.026	1 (0.2%)	0.152	0.046	2 (0.8%)
Generalized linear model 2 (normal + log link)	0.110	0.023	0	0.162	0.055	3 (1.2%)
Generalized linear model 3 (gamma + identity link)	0.116	0.028	0	0.136	0.034	0
Generalized linear model 4 (gamma + identity link)	0.116	0.028	0	0.136	0.034	0
Adjusted limited dependent variable mixture model 1	0.115	0.025	1 (0.2%)	0.147	0.039	0
Adjusted limited dependent variable mixture model 2	0.113	0.024	6 (1.2%)	0.152	0.041	3 (1.2%)
Adjusted limited dependent variable mixture model 3	0.116	0.026	0	0.148	0.039	0
Adjusted limited dependent variable mixture model 4	0.114	0.024	6 (1.2%)	0.151	0.041	1 (0.4%)
Mixture beta regression model 1	0.111	0.025	0	0.152	0.043	0
Mixture beta regression model 2	0.107	0.023	0	0.155	0.046	0
Mixture beta regression model 3	0.112	0.026	0	0.148	0.042	0
Mixture beta regression model 4	0.106	0.024	0	0.154	0.046	0
Ordered logistic regression model 4	0.101	0.021	0	0.150	0.044	0

Note: MAE: mean absolute error, MSE: mean squared error, OFR %: percentage of estimated values out of the range of Canadian value set; Model specifications: 1: full ESAS-r: Renal items; 2: full ESAS-r: Renal items + demographics; 3: selected ESAS-r: Renal items; 4: selected ESAS-r: Renal items + selected demographics

Table 3.4 MAE and MSE statistics in the estimation and validation sample by EQ-5D-5L subset

Performance statistics by model	EO-5D-5L index < 0.5		EO-5D-5L index >= 0.5	
	Estimation	Validation	Estimation	Validation
Fixed Effect model 3				
Mean Absolute Error (MAE)	0.112	0.294	0.048	0.102
Mean Squared Error (MSE)	0.022	0.102	0.005	0.016
Generalized estimating equations 3				
MAE	0.304	0.252	0.088	0.113
MSE	0.118	0.084	0.012	0.023
Generalized linear model 3				
MAE	0.308	0.246	0.083	0.111
MSE	0.125	0.081	0.011	0.023
Censored least absolute deviations 1				
MAE	0.312	0.258	0.068	0.109
MSE	0.133	0.094	0.010	0.024
Adjusted limited dependent variable mixture model1				
MAE	0.265	0.229	0.090	0.128
MSE	0.097	0.075	0.013	0.030

Note: Model specification 1: full ESAS-r: Renal items; Model specification 2: full ESAS-r: Renal items + demographics; Model specification 3: selected ESAS-r: Renal items; Model specification 4: selected ESAS-r: Renal items + selected demographics

Table 3.5 Model coefficients for GEE and GLM model

	GEE (utility)	GLM (disutility)
Pain	-0.017**	0.022**
Tiredness	-0.009*	
Drowsiness		0.011*
Shortness of breath	-0.007	
Depression	-0.012*	0.007
Anxiety	-0.014*	0.016*
Wellbeing	-0.012*	0.015*
Restless legs	0.007*	
_cons	0.899**	0.026**

Note: * means $p < 0.05$, ** means $p<0.001$, gee: generalized estimating equations, glm: generalized linear model

References

1. Levin A, Hemmelgarn B, Culleton B, et al. Guidelines for the management of chronic kidney disease. *CMAJ*. 2008;179(11):1154-1162. doi:10.1503/cmaj.080351
2. Levey AS, Coresh J. Chronic kidney disease. *The Lancet*. 2012;379(9811):165-180. doi:10.1016/S0140-6736(11)60178-5
3. Romagnani P, Remuzzi G, Glassock R, et al. Chronic kidney disease. *Nat Rev Dis Primer*. 2017;3(1):1-24. doi:10.1038/nrdp.2017.88
4. Jha V, Garcia-Garcia G, Iseki K, et al. Chronic kidney disease: global dimension and perspectives. *The Lancet*. 2013;382(9888):260-272. doi:10.1016/S0140-6736(13)60687-X
5. Arora P, Vasa P, Brenner D, et al. Prevalence estimates of chronic kidney disease in Canada: results of a nationally representative survey. *CMAJ*. 2013;185(9):E417-E423. doi:10.1503/cmaj.120833
6. Bello AK, Levin A, Lunney M. Global Kidney Health Atlas: A report by the International Society of Nephrology on the global burden of end-stage kidney disease and capacity for kidney replacement therapy and conservative care across world countries and regions. *Bruss Int Soc Nephrol*. Published online 2019.
7. Almutary H, Bonner A, Douglas C. Symptom Burden in Chronic Kidney Disease: A Review of Recent Literature. *J Ren Care*. 2013;39(3):140-150. doi:10.1111/j.1755-6686.2013.12022.x
8. Cruz MC, Andrade C, Urrutia M, Draibe S, Nogueira-Martins LA, Sesso R de CC. Quality of life in patients with chronic kidney disease. *Clinics*. 2011;66(6):991-995. doi:10.1590/S1807-59322011000600012
9. Davison SN, Jhangri GS, Johnson JA. Cross-sectional validity of a modified Edmonton symptom assessment system in dialysis patients: A simple assessment of symptom burden. *Kidney Int*. 2006;69(9):1621-1625. doi:10.1038/sj.ki.5000184
10. Afshar M, Rebollo-Mesa I, Murphy E, Murtagh FEM, Mamode N. Symptom Burden and Associated Factors in Renal Transplant Patients in the U.K. *J Pain Symptom Manage*. 2012;44(2):229-238. doi:10.1016/j.jpainsymman.2011.08.005
11. Murphy EL, Murtagh FEM, Carey I, Sheerin NS. Understanding symptoms in patients with advanced chronic kidney disease managed without dialysis: use of a short patient-completed assessment tool. *Nephron Clin Pract*. 2009;111(1):c74-80. doi:10.1159/000183177
12. Davison SN, Jhangri GS, Johnson JA. Longitudinal validation of a modified Edmonton symptom assessment system (ESAS) in haemodialysis patients. *Nephrol Dial Transplant*. 2006;21(11):3189-3195. doi:10.1093/ndt/gfl380
13. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal 2013. National Institute for Health and Care Excellence (NICE)*. NICE; 2013. Accessed April 26, 2020.

https://www.nice.org.uk/process/pmg9/chapter/the-reference-case#measuring-and-valuing-health-effects

14. Guidelines for the Economic Evaluation of Health Technologies: Canada. CADTH.ca. Published December 2, 2015. Accessed May 4, 2020. https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada

15. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Fourth edition. Oxford, United Kingdom: Oxford University Press; 2015.

16. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. *Appl Health Econ Health Policy*. 2017;15(2):127-137. doi:10.1007/s40258-017-0310-5

17. Gibbons E, Fitzpatrick R. A structured review of patient-reported outcome measures for people with chronic kidney disease. *Dep Public Health Univ Oxf*. Published online 2010.

18. Longworth L, Rowen D. Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value Health*. 2013;16(1):202-210. doi:10.1016/j.jval.2012.10.010

19. Petrou S, Rivero-Arias O, Dakin H, et al. The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. *PharmacoEconomics*. 2015;33(10):993-1011. doi:10.1007/s40273-015-0312-9

20. 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. doi:10.1016/j.jval.2016.11.006

21. Cameron AC. Panel data methods for microeconometrics using Stata. In: *West Coast Stata Users' Group Meetings 2007*. Vol 13. Stata Users Group; 2007.

22. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-1736. doi:10.1007/s11136-011-9903-x

23. Xie F, Pullenayegum E, Gaebel K, et al. A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. *Med Care*. 2016;54(1):98-105. doi:10.1097/MLR.0000000000000447

24. Hui D, Bruera E. The Edmonton Symptom Assessment System 25 Years Later: Past, Present and Future Developments. *J Pain Symptom Manage*. 2017;53(3):630-643. doi:10.1016/j.jpainsymman.2016.10.370

25. 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(1):1-9. doi:10.1186/1477-7525-7-27

26. Ameri H, Yousefi M, Yaseri M, Nahvijou A, Arab M, Akbari Sari A. Mapping EORTC-QLQ-C30 and QLQ-CR29 onto EQ-5D-5L in Colorectal Cancer Patients. *J Gastrointest Cancer*. 2020;51(1):196-203. doi:10.1007/s12029-019-00229-6

27. Mukaka M. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Med J Med Assoc Malawi*. 2012;24(3):69-71.

28. Vokó Z, Németh R, Nagyjánosi L, et al. Mapping the Nottingham Health Profile onto the Preference-Based EuroQol-5D Instrument for Patients with Diabetes. *Value Health Reg Issues*. 2014;4:31-36. doi:10.1016/j.vhri.2014.06.002

29. Chuang L-H, Kind P. Converting the SF-12 into the EQ-5D. *PharmacoEconomics*. 2009;27(6):491-505. doi:10.2165/00019053-200927060-00005

30. Wooldridge JM. *Introductory Econometrics : A Modern Approach /*. 5th ed. South-Western Cengage Learning,; 2013.

31. Payakachat N, Summers KH, Pleil AM, et al. Predicting EQ-5D utility scores from the 25-item National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients with age-related macular degeneration. *Qual Life Res*. 2009;18(7):801-813. doi:10.1007/s11136-009-9499-6

32. Skaltsa K, Longworth L, Ivanescu C, Phung D, Holmstrom S. Mapping the FACT-P to the Preference-Based EQ-5D Questionnaire in Metastatic Castration-Resistant Prostate Cancer. *Value Health*. 2014;17(2):238-244. doi:10.1016/j.jval.2013.12.005

33. 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):683-694. doi:10.1007/s11136-012-0189-4

34. Hardin JW. Generalized Estimating Equations (GEE). In: *Encyclopedia of Statistics in Behavioral Science*. American Cancer Society; 2005. doi:10.1002/0470013192.bsa250

35. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. MIT Press,; c2002.

36. Hausman JA. Specification Tests in Econometrics. *Econometrica*. 1978;46(6):1251-1271. doi:10.2307/1913827

37. Gu NY, Botteman MF, Ji X, Bell CF, Carter JA, van Hout B. Mapping of the Insomnia Severity Index and other sleep measures to EuroQol EQ-5D health state utilities. *Health Qual Life Outcomes*. 2011;9:119. doi:10.1186/1477-7525-9-119

38. Khan KA, Petrou S, Rivero-Arias O, Walters SJ, Boyle SE. Mapping EQ-5D Utility Scores from the PedsQL™ Generic Core Scales. *PharmacoEconomics*. 2014;32(7):693-706. doi:10.1007/s40273-014-0153-y

39. Kay S, Tolley K, Colayco D, Khalaf K, Anderson P, Globe D. Mapping EQ-5D Utility Scores from the Incontinence Quality of Life Questionnaire among Patients with Neurogenic and Idiopathic Overactive Bladder. *Value Health*. 2013;16(2):394-402. doi:10.1016/j.jval.2012.12.005

40. Gray LA, Alava MH. A Command for Fitting Mixture Regression Models for Bounded Dependent Variables Using the Beta Distribution. *Stata J*. 2018;18(1):51-75. doi:10.1177/1536867X1801800105

41. Hernandez-Alava M, Wailoo A. ALDVMM: Fitting Adjusted Limited Dependent Variable Mixture Models to EQ-5D. *The Stata Journal*, 15, Number 3, pp. 737–750.

42. Wadsworth I, Van Horn ML, Jaki T. A DIAGNOSTIC TOOL FOR CHECKING ASSUMPTIONS OF REGRESSION MIXTURE MODELS. *JP J Biostat*. 2018;15(1):1-20. doi:10.17654/BS015010001

43. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. 2007;26(30):5512-5528. doi:10.1002/sim.3148

44. Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. *Dordr Neth Reidel*. 1986;81.

45. Neath AA, Cavanaugh JE. The Bayesian information criterion: background, derivation, and applications. *WIREs Comput Stat*. 2012;4(2):199-203. doi:10.1002/wics.199

46. 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):215-225. doi:10.1007/s10198-009-0168-z

47. Moskovitz M, Jao K, Su J, et al. Combined cancer patient–reported symptom and health utility tool for routine clinical implementation: a real-world comparison of the ESAS and EQ-5D in multiple cancer sites. *Curr Oncol*. 2019;26(6). doi:10.3747/co.26.5297

48. van Hout B, Janssen MF, Feng Y-S, et al. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. *Value Health*. 2012;15(5):708-715. doi:10.1016/j.jval.2012.02.008

49. Sullivan PW, Ghushchyan VH. EQ-5D Scores for Diabetes-Related Comorbidities. *Value Health*. 2016;19(8):1002-1008. doi:10.1016/j.jval.2016.05.018

50. Kay S, Ferreira A. Mapping the 25-item National Eye Institute Visual Functioning Questionnaire (NEI VFQ-25) to EQ-5D Utility Scores. *Ophthalmic Epidemiol*. 2014;21(2):66-78. doi:10.3109/09286586.2014.888456

Chapter 4. Discussion

4.1 Summary

Economic evaluations have gained importance as a way to improve efficiency when healthcare budgets reach their upper limits and more and more emerging and expensive health technologies are introduced. CUA is recommended as the base-case analysis for economic evaluation in many countries' health technology assessment (HTA) guidelines[1,2]. The utility is a necessary parameter in the CUA. To obtain preferences on health states to estimate utilities, one could use direct measurement techniques, such as TTO, or use instruments having a preference scoring system, such as the EQ-5D-5L. When these utility values from PBMs are not available, mapping is a technique to convert health outcome data from non-PBM into utility values for CUA. Some studies also mapped one PBM to another PBM to allow for the comparison of CUAs that applied different PBMs. Over the years, there have been a great number of publications on new mapping algorithms for PBMs. The mapping study database by the Health Economics Research Centre (HERC) included 182 mapping studies as of January 2019[3].

Though mapping is a feasible method to estimate PBM utilities from other PROMs, this method has drawbacks. Previous studies have critiqued that mapping may underestimate the uncertainty[4], and it might ignore the conceptual issues during the mapping process and when extrapolating the results[5]. Two reporting quality checklists[6,7] have been developed as a guide for authors to report necessary information of the analysis, such as uncertainty and conceptual overlap exploration. The checklists are the Mapping onto Preference-based Measures Reporting Standards (MAPS) reporting statement[6] and the ISPOR Good Practices Task Force Report on Mapping[7]. This thesis assessed the reporting quality of studies which mapped onto the EQ-5D-5L, especially the completeness of information for CUA. The impact of applying checklists on the reporting quality of the study was also explored. The key findings will inform future mapping studies on how to ensure high reporting quality and completeness of information for CUA. These findings were further applied in the development of a novel mapping algorithm in the second part of the thesis.

Chapter 1 introduced a taxonomy of health status measures, utility and its application, and mapping techniques. This chapter started with a general introduction with different health status measures, including generic measures and specific measures, and profile measures and

preference-based measures. The utility theory and the application of utility in QALY and CUA were then introduced. The third part of this chapter included the introduction of mapping, some arguments, and the popular model approaches for mapping.

In Chapter 2, a systematic review and reporting quality assessment of mapping studies for the EQ-5D-5L was reported. The updated search identified 14 new mapping studies, and 39 studies (including the studies in the HERC database) were finally assessed for reporting quality. Various models have been utilized to deal with the special characteristics of the utility distribution in these studies, but the traditional OLS model still remains the base-case analysis model and is also selected as the final model in many mapping studies. Based on the reporting quality assessment, the published studies mostly followed the items on the checklists, which improved their reliability and validity. The five poorest performed items were 1) estimation of predicted utilities, 2) reporting variances, covariances, and error terms 3) final model calculation example, 4) parameter uncertainty, and 5) individual uncertainty. These five items would also impact the use of mapping algorithms in CUA. Finally, a preliminary comparison showed that the checklists have had a positive impact on improving the reporting quality of the study.

In the second study (Chapter 3), mapping algorithms onto the EQ-5D-5L were developed from the Edmonton Symptom Assessment System-Revised: Renal (ESAS-r: Renal) on a population of patients with end-stage renal disease (ESRD). Various model types and four model specifications were tested to determine the best performing model. Two models were selected: the generalized estimating equations (GEE) on selected ESAS-r: Renal symptoms, and the generalized linear model (GLM) using gamma distribution with identity link on selected ESAS-r: Renal symptoms. Both models produced a good fit of the data from the estimation sample and had good predictive power in the validation sample. The GEE model had an unexpected sign in an independent variable, “Restless legs” item from the ESAS-r: Renal. It was recommended to use the GLM model in the base-case analysis of CUA, while the GEE model could be used in the sensitivity analysis. The requirements of the checklists[6,7] were followed in conducting and reporting the analysis, especially for the items which were not frequently reported in previous studies based on our systematic review. The algorithms can be used to predict utility values when the ESAS-r: Renal has been administered for patients with ESRD and the direct utility measures are not

available. The algorithms are likely to be robust for populations comparable to our estimation sample.

4.2 Study Limitations

The first limitation regarding the systematic review is that it was conducted by a single person (the author of this thesis). There was no second reviewer in the including/excluding processes, and the data extraction and quality check were all done by the author of this thesis. This might have increased the likelihood of errors and biases in including/excluding and evaluating the studies.

The reporting quality assessment did not apply any scoring to rate each study. The analytical statistics on reporting performance were mostly descriptive. With scoring, inferential statistical analysis could be conducted. In Dakin et al's reporting quality assessment on the title and abstract sections, they scored the performance with "0" meaning not fulfilled, "1" meaning fulfilled, and "0.5" meaning partially fulfilled in each item, and the sum of the item scores was the final score of the study[8]. They provided the distribution of the final scores and ran a regression to see whether the final scores were significantly associated with the publication year. This analysis answered their study objective of the impact of the MAPS reporting statement on the reporting quality of mapping studies. In our review, the primary objective was to highlight the specific strengths and weaknesses of published EQ-5D-5L mapping studies, so scoring was not developed when applying the two checklists. However, we did a preliminary comparison and observed that the studies which cited any of the checklists had better reporting performance on the least-reported items identified in the review. Scoring each study and using statistical inference could make the conclusions of the study more robust.

It is important to note that the review focused on reporting quality instead of research quality. Studies with high reporting quality are not necessarily associated with high research quality, while studies with high research quality may fail to report everything and may be assessed to have low reporting quality. There are recommendations on how to conduct the mapping analysis. Longworth et al provided some recommendations based on best mapping practice in the overview of mapping to obtain EQ-5D-3L utility for use in the National Institute for Health and Care Excellence HTA[9]. The ISPOR Good Practices also suggests pre-modelling and modelling recommendations[7]. These recommendations could guide investigators in conducting their

analysis. However, it is hard to evaluate the research quality based on these recommendations. These recommendations have not yet been validated, and the level of quality of a study could not be determined by whether it is consistent with these checklists.

In Chapter 3, one significant limitation was the large volume of missing data. Records with missing values in either the ESAS-r: Renal or the EQ-5D-5L were excluded in the study. Mean imputation or other missing data processing methods was not used. These methods were not tried as they would alter the distribution of the ESAS-r: Renal and the EQ-5D-5L, and therefore this may impact the relationship between the two instruments. Another limitation that is related to the data is that diabetes and race were not collected in the validation sample. Therefore, the important clinical and demographic information was not included in the final model. The sensitivity analyses showed that they were significant in GLM and GEE models with regard to the four model specifications in the estimation sample. Although they had a better statistical fit, it was possible that the predictive power of the models including these two variables would not increase. We observed that the GLM and GEE models with demographics (age and sex) fitted better in the estimation sample but predicted worse in the validation sample compared with the models without demographics. Since these two variables were not included in the final model, when extrapolating the mapping algorithm to another population, the patient group should have similar characteristics in the distribution of diabetes and sex.

The “Restless legs” term from the ESAS-r: Renal in the GEE model has an unexpected sign. Controlling other variables, the increase in the severity of “Restless legs” led to an increase in the utility values. The “Restless legs” was not a common symptom burden in the dataset being used, and only a weak correlation was observed between “Restless legs” and the EQ-5D-5L utility index. The coefficient of “Restless legs” was statistically significant but it had a small parameter estimate. Therefore, in further applications for economic evaluation, it may be advisable that using the GLM model in the reference case, while the GEE model could be used in the sensitivity (scenario) analysis. To deal with this limitation, additional external validation could be conducted in a population where “Restless legs” is more prevalent.

The mapping algorithms in Chapter 3 had an over prediction of the utility scores in the lower-end of the EQ-5D-5L range, which is a common limitation of mapping studies[10,11]. The possible causes of this limitation were the lack of construct overlaps between instruments and the small

number of observations with low utility values for estimation. This will deteriorate the validity of applying mapping algorithms in CUA and indicate the limitation of using mapped utility values in CUA.

4.3 Future Applications and Directions

The findings of this study may have a significant impact on future mapping studies and health economic studies related to ESRD patients. The systematic review of the EQ-5D-5L mapping studies pointed out the common strengths and issues in current studies. For future mapping studies, it is important to keep these strengths and also include and provide detailed information on those currently least-reported items.

A more comprehensive systematic review of all mapping studies could be conducted. The review in Chapter 2 identified several problems among mapping studies onto the EQ-5D-5L. Many studies are not applicable to some requirements of the reporting quality checklists, such as recommendations on comparison and selection of candidate estimation sample and recommendations on external validation, while previous mapping studies onto other PBMs may be related to these items. It is impossible to make a conclusion and provide suggestions to future studies regarding these items based on the current review. A broader review which includes studies onto other PBMs could potentially indicate the reporting performance on these items.

There is a need to develop an assessment tool on the research quality of the mapping studies. The quality of reporting does not necessarily represent the research quality. The current recommendations on conducting analysis are not robust to evaluate the research quality. A good research quality assessment tool should indicate the standard of a high-quality study, and the extent of deterioration in quality if specific requirements are not addressed.

The empirical study which mapped the EQ-5D-5L from the ESAS-r: Renal provides a method to estimate utility values when PBMs are not available and allow for economic evaluations on ESRD patients to be conducted. Two algorithms were provided as the best models. This may cause confusion. But according to the Longworth et al recommendations, multiple possible mapping functions could give an indication of the uncertainty associated with the choice of the algorithm[9]. We also justified in Chapter 3 which model should be used in the reference case.

Besides, further external validation is still needed to see whether the algorithms could be used in an ESRD patient population with different patient characteristics.

Another direction related to this empirical mapping study is to explore the possible mapping algorithms between the general (i.e., non-Renal) form of the ESAS-r and the EQ-5D-5L. With the results of the mapping study of this thesis, the ESAS-r: Renal is the first instrument in the ESAS family to have a mapping algorithm. The other forms including the general ESAS-r do not have mapping algorithms. For patients in other clinical areas where other forms of the ESAS-r are commonly used, there is still no algorithm to conduct economic evaluation when PBMs are not available. The GLM model mapping the ESAS-r: Renal to the EQ-5D-5L did not include any ESRD-specific symptoms. Therefore, it would be interesting to test whether the GLM model is applicable to the general ESAS-r.

There were some studies exploring the impact of using mapping algorithms in CUA. It is commonly agreed that using mapped utility values would under-estimate uncertainty[4,9,12]. But one study also pointed out that if the sensitivity analysis showed that the cost-effectiveness result is insensitive to utility values, the issues of mapping studies are not critical to the results of CUA[4]. Further directions regarding mapping studies could focus more on addressing the uncertainty, improving the mapping methodologies, and the impact of using mapped utility values in CUA.

4.4 Conclusion

The review in this thesis highlighted the issues for the current EQ-5D-5L mapping studies based on the mapping studies checklists. This could inform future studies on how to improve the reporting quality of mapping analysis. The novel mapping algorithms developed in this thesis from the ESAS-r: Renal to the EQ-5D-5L could convert the ESAS-r: Renal to the EQ-5D-5L utility index for ESRD patients with similar patient characteristics as the estimation sample of the mapping study.

References

1. Guidelines for the Economic Evaluation of Health Technologies: Canada. CADTH.ca. Published December 2, 2015. Accessed May 4, 2020. https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada
2. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal 2013. National Institute for Health and Care Excellence (NICE)*. NICE; 2013. Accessed April 26, 2020. https://www.nice.org.uk/process/pmg9/chapter/the-reference-case#measuring-and-valuing-health-effects
3. HERC database of mapping studies, Version 7.0 (last updated: 24th April 2019). http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies.
4. McCabe C, Edlin R, Meads D, Brown C, Kharroubi S. Constructing Indirect Utility Models: Some Observations on the Principles and Practice of Mapping to Obtain Health State Utilities. *PharmacoEconomics*. 2013;31(8):635-641. doi:10.1007/s40273-013-0071-4
5. Round J, Hawton A. Statistical Alchemy: Conceptual Validity and Mapping to Generate Health State Utility Values. *PharmacoEconomics - Open*. 2017;1(4):233-239. doi:10.1007/s41669-017-0027-2
6. Petrou S, Rivero-Arias O, Dakin H, et al. The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. *PharmacoEconomics*. 2015;33(10):993-1011. doi:10.1007/s40273-015-0312-9
7. 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. doi:10.1016/j.jval.2016.11.006
8. 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):1-9. doi:10.1186/s12955-018-0857-3
9. Longworth L, Rowen D. Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value Health*. 2013;16(1):202-210. doi:10.1016/j.jval.2012.10.010
10. Khan KA, Petrou S, Rivero-Arias O, Walters SJ, Boyle SE. Mapping EQ-5D Utility Scores from the PedsQL^{TM} Generic Core Scales. *PharmacoEconomics*. 2014;32(7):693-706. doi:10.1007/s40273-014-0153-y
11. van Hout B, Janssen MF, Feng Y-S, et al. Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. *Value Health*. 2012;15(5):708-715. doi:10.1016/j.jval.2012.02.008

12. Yang F, Devlin N, Luo N. Impact of mapped EQ-5D utilities on cost-effectiveness analysis: in the case of dialysis treatments. *Eur J Health Econ*. 2019;20(1):99-105. doi:10.1007/s10198-018-0987-x

References

Abdin, E., Chong, S. A., Seow, E., Verma, S., Tan, K. B., & Subramaniam, M. (2019). Mapping the Positive and Negative Syndrome Scale scores to EQ-5D-5L and SF-6D utility scores in patients with schizophrenia. *Qual Life Res*, *28*(1), 177–186. https://doi.org/10.1007/s11136-018-2037-7

Afshar, M., Rebollo-Mesa, I., Murphy, E., Murtagh, F. E. M., & Mamode, N. (2012). Symptom Burden and Associated Factors in Renal Transplant Patients in the U.K. *Journal of Pain and Symptom Management*, *44*(2), 229–238. https://doi.org/10.1016/j.jpainsymman.2011.08.005

Almutary, H., Bonner, A., & Douglas, C. (2013). Symptom Burden in Chronic Kidney Disease: A Review of Recent Literature. *Journal of Renal Care*, *39*(3), 140–150. https://doi.org/10.1111/j.1755-6686.2013.12022.x

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, *24*(1), 3–61. https://doi.org/10.1016/0304-4076(84)90074-5

Ameri, H., Yousefi, M., Yaseri, M., Nahvijou, A., Arab, M., & Akbari Sari, A. (2019). Mapping the cancer-specific QLQ-C30 onto the generic EQ-5D-5L and SF-6D in colorectal cancer patients. *Expert Rev Pharmacoecon Outcomes Res*, *19*(1), 89–96. https://doi.org/10.1080/14737167.2018.1517046

Ameri, H., Yousefi, M., Yaseri, M., Nahvijou, A., Arab, M., & Akbari Sari, A. (2020). Mapping EORTC-QLQ-C30 and QLQ-CR29 onto EQ-5D-5L in Colorectal Cancer Patients. *J Gastrointest Cancer*, *51*(1), 196–203. https://doi.org/10.1007/s12029-019-00229-6

A.n, L. (2020). Does linear equating improve prediction in mapping? Crosswalking MacNew onto EQ-5D-5L value sets. *Journal of Health Economics*. https://doi.org/10.1007/s10198-020-01183-y

Arora, P., Vasa, P., Brenner, D., Iglar, K., McFarlane, P., Morrison, H., & Badawi, A. (2013). Prevalence estimates of chronic kidney disease in Canada: Results of a nationally representative survey. *CMAJ*, *185*(9), E417–E423. https://doi.org/10.1503/cmaj.120833

Bansback, N., Tsuchiya, A., Brazier, J., & Anis, A. (2012). Canadian valuation of EQ-5D health states: Preliminary value set and considerations for future valuation studies. *PloS One*, *7*(2), e31115. https://doi.org/10.1371/journal.pone.0031115

Beck, A.-J. C. C., Kieffer, J. M., Retèl, V. P., Overveld, L. F. J. van, Takes, R. P., Brekel, M. W. M. van den, Harten, W. H. van, & Stuiver, M. M. (2019). Mapping the EORTC QLQ-C30 and QLQ-H&N35 to the EQ-5D for head and neck cancer: Can disease-specific utilities be obtained? *PLOS ONE*, *14*(12), e0226077. https://doi.org/10.1371/journal.pone.0226077

Bello, A. K., Levin, A., & Lunney, M. (2019). Global Kidney Health Atlas: A report by the International Society of Nephrology on the global burden of end-stage kidney disease and capacity for kidney replacement therapy and conservative care across world countries and regions. *Brussels: International Society of Nephrology*.

Belotti, F., Deb, P., Manning, W. G., & Norton, E. C. (2015). Twopm: Two-Part Models. *The Stata Journal*, *15*(1), 3–20. https://doi.org/10.1177/1536867X1501500102

Bilbao, A., Martín-Fernández, J., García-Pérez, L., Arenaza, J. C., Ariza-Cardiel, G., Ramallo-Fariña, Y., & Ansola, L. (2019). Mapping WOMAC onto the EQ-5D-5L Utility Index in Patients with Hip or Knee Osteoarthritis. *Value in Health*. https://doi.org/10.1016/j.jval.2019.09.2755

Bishop, R. L. (1943). Consumer's Surplus and Cardinal Utility. *The Quarterly Journal of Economics*, *57*(3), 421–449. https://doi.org/10.2307/1881755

Brazier, J. E., Rowen, D., Mavranezouli, I., Tsuchiya, A., Young, T., Yang, Y., Barkham, M., & Ibbotson, R. (2012). Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). In *NIHR Health Technology Assessment programme: Executive Summaries*. NIHR Journals Library. http://www.ncbi.nlm.nih.gov/books/NBK100337/

Brazier, J., Ratcliffe, J., Saloman, J., & Tsuchiya, A. (2017). *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford University Press.

Brazier, John E., Yang, Y., Tsuchiya, A., & Rowen, D. L. (2010). A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *The European Journal of Health Economics*, *11*(2), 215–225. https://doi.org/10.1007/s10198-009-0168-z

Briggs, A., Sculpher, M., & Buxton, M. (1994). Uncertainty in the economic evaluation of health care technologies: The role of sensitivity analysis. *Health Economics*, *3*(2), 95–104. https://doi.org/10.1002/hec.4730030206

Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, *37*(1), 53–72. https://doi.org/10.1016/0168-8510(96)00822-6

Buckingham, K., & Devlin, N. (2006). A theoretical framework for TTO valuations of health. *Health Economics*, *15*(10), 1149–1154. https://doi.org/10.1002/hec.1122

Cameron, A. C. (2007). Panel data methods for microeconometrics using Stata. *West Coast Stata Users' Group Meetings 2007*, *13*.

Chen, G., Garcia-Gordillo, M. A., Collado-Mateo, D., del Pozo-Cruz, B., Adsuar, J. C., Cordero-Ferrera, J. M., Abellán-Perpiñán, J. M., & Sánchez-Martínez, F. I. (2018). Converting Parkinson-Specific Scores into Health State Utilities to Assess Cost-Utility Analysis. *Patient*, *11*(6), 665–675. Scopus. https://doi.org/10.1007/s40271-018-0317-5

Chen, Gang, Khan, M. A., Iezzi, A., Ratcliffe, J., & Richardson, J. (2016). Mapping between 6 Multiattribute Utility Instruments. *Medical Decision Making*, *36*(2), 160–175. https://doi.org/10.1177/0272989X15578127

Chen, S., & Khan, S. (2000). Estimating censored regression models in the presence of nonparametric multiplicative heteroskedasticity. *Journal of Econometrics*, *98*(2), 283–316. https://doi.org/10.1016/S0304-4076(00)00020-8

Cheung, Y. B., Luo, N., Ng, R., & Lee, C. F. (2014). Mapping the functional assessment of cancer therapy-breast (FACT-B) to the 5-level EuroQoL Group's 5-dimension

questionnaire (EQ-5D-5L) utility index in a multi-ethnic Asian population. *Health and Quality of Life Outcomes*, *12*, 180. https://doi.org/10.1186/s12955-014-0180-6

Chuang, L.-H., & Kind, P. (2009). Converting the SF-12 into the EQ-5D. *PharmacoEconomics*, *27*(6), 491–505. https://doi.org/10.2165/00019053-200927060-00005

Chuang, L.-H., & Whitehead, S. J. (2012). Mapping for economic evaluation. *British Medical Bulletin*, *101*(1), 1–15. https://doi.org/10.1093/bmb/ldr049

Collado-Mateo, D., Chen, G., Garcia-Gordillo, M. A., Iezzi, A., Adsuar, J. C., Olivares, P. R., & Gusi, N. (2017). "Fibromyalgia and quality of life: Mapping the revised fibromyalgia impact questionnaire to the preference-based instruments." *Health and Quality of Life Outcomes*, *15*. https://doi.org/10.1186/s12955-017-0690-0

Coon, C., Bushmakin, A., Tatlock, S., Williamson, N., Moffatt, M., Arbuckle, R., & Abraham, L. (2018). Evaluation of a crosswalk between the European Quality of Life Five Dimension Five Level and the Menopause-Specific Quality of Life questionnaire. *Climacteric : The Journal of the International Menopause Society*, *21*(6), 566–573. https://doi.org/10.1080/13697137.2018.1481381

Crott, R. (2014). Mapping algorithms from QLQ-C30 to EQ-5D utilities: No firm ground to stand on yet. *Expert Review of Pharmacoeconomics & Outcomes Research*, *14*(4), 569–576. https://doi.org/10.1586/14737167.2014.908711

Cruz, M. C., Andrade, C., Urrutia, M., Draibe, S., Nogueira-Martins, L. A., & Sesso, R. de C. C. (2011). Quality of life in patients with chronic kidney disease. *Clinics*, *66*(6), 991–995. https://doi.org/10.1590/S1807-59322011000600012

Dakin, H., Abel, L., Burns, R., & Yang, Y. (2018). Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: An online database and application of the MAPS statement. *Health and Quality of Life Outcomes*, *16*(1), 1–9. https://doi.org/10.1186/s12955-018-0857-3

Dakin, H., Gray, A., & Murray, D. (2013). Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. *Quality of Life Research*, *22*(3), 683–694. https://doi.org/10.1007/s11136-012-0189-4

Davison, S. N., Jhangri, G. S., & Johnson, J. A. (2006). Cross-sectional validity of a modified Edmonton symptom assessment system in dialysis patients: A simple assessment of symptom burden. *Kidney International*, *69*(9), 1621–1625. https://doi.org/10.1038/sj.ki.5000184

Davison, Sara N., Jhangri, G. S., & Johnson, J. A. (2006). Longitudinal validation of a modified Edmonton symptom assessment system (ESAS) in haemodialysis patients. *Nephrology Dialysis Transplantation*, *21*(11), 3189–3195. https://doi.org/10.1093/ndt/gfl380

Devlin, N. J., & Brooks, R. (2017). EQ-5D and the EuroQol Group: Past, Present and Future. *Applied Health Economics and Health Policy*, *15*(2), 127–137. https://doi.org/10.1007/s40258-017-0310-5

Dixon, P., Hollingworth, W., & Sparrow, J. (2020). Mapping to Quality of Life and Capability Measures in Cataract Surgery Patients: From Cat-PROM5 to EQ-5D-3L, EQ-5D-5L, and

ICECAP-O Using Mixture Modelling. *MDM Policy & Practice*, *5*(1), 2381468320915447. https://doi.org/10.1177/2381468320915447

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes* (Fourth edition.). Oxford, United Kingdom: Oxford University Press.

Fanshel, S., & Bush, J. W. (1970). A Health-Status Index and its Application to Health-Services Outcomes. *Operations Research*, *18*(6), 1021–1066. https://doi.org/10.1287/opre.18.6.1021

Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-attribute health status classification systems. *Pharmacoeconomics*, *7*(6), 490–502.

Fishburn, P. C. (1989). Retrospective on the Utility Theory of von Neumann and Morgenstern. *Journal of Risk and Uncertainty*, *2*(2), 127–157.

Gafni, A., Birch, S., & Mehrez, A. (1993). Economics, health and health economics: HYEs versus QALYs. *Journal of Health Economics*, *12*(3), 325–339. https://doi.org/10.1016/0167-6296(93)90015-7

Gamst-Klaussen, T., Chen, G., Lamu, A. N., & Olsen, J. A. (2016). Health state utility instruments compared: Inquiring into nonlinearity across EQ-5D-5L, SF-6D, HUI-3 and 15D. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *25*(7), 1667–1678. https://doi.org/10.1007/s11136-015-1212-3

Gamst-Klaussen, T., Lamu, A. N., Chen, G., & Olsen, J. A. (2018). Assessment of outcome measures for cost-utility analysis in depression: Mapping depression scales onto the EQ-5D-5L. *BJPsych Open*, *4*(4), 160–166. https://doi.org/10.1192/bjo.2018.21

Gibbons, E., & Fitzpatrick, R. (2010). A structured review of patient-reported outcome measures for people with chronic kidney disease. *Department of Public Health University of Oxford*.

Gray, L. A., & Alava, M. H. (2018). A Command for Fitting Mixture Regression Models for Bounded Dependent Variables Using the Beta Distribution. *The Stata Journal*, *18*(1), 51–75. https://doi.org/10.1177/1536867X1801800105

Gray, L. A., Hernández Alava, M., & Wailoo, A. J. (2018). Development of Methods for the Mapping of Utilities Using Mixture Models: Mapping the AQLQ-S to the EQ-5D-5L and the HUI3 in Patients with Asthma. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *21*(6), 748–757. https://doi.org/10.1016/j.jval.2017.09.017

Gray, L. A., Wailoo, A. J., & Hernandez Alava, M. (2018). Mapping the FACT-B Instrument to EQ-5D-3L in Patients with Breast Cancer Using Adjusted Limited Dependent Variable Mixture Models versus Response Mapping. *Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *21*(12), 1399–1405. https://doi.org/10.1016/j.jval.2018.06.006

Gu, N. Y., Botteman, M. F., Ji, X., Bell, C. F., Carter, J. A., & van Hout, B. (2011a). Mapping of the Insomnia Severity Index and other sleep measures to EuroQol EQ-5D health state

utilities. *Health and Quality of Life Outcomes*, *9*, 119. https://doi.org/10.1186/1477-7525-9-119

Gu, N. Y., Botteman, M. F., Ji, X., Bell, C. F., Carter, J. A., & van Hout, B. (2011b). Mapping of the Insomnia Severity Index and other sleep measures to EuroQol EQ-5D health state utilities. *Health and Quality of Life Outcomes*, *9*, 119. https://doi.org/10.1186/1477-7525-9-119

Guidelines for the Economic Evaluation of Health Technologies: Canada. (2015, December 2). CADTH.Ca. https://www.cadth.ca/about-cadth/how-we-do-it/methods-and-guidelines/guidelines-for-the-economic-evaluation-of-health-technologies-canada

Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine*, *118*(8), 622–629.

Hardin, J. W. (2005). Generalized Estimating Equations (GEE). In *Encyclopedia of Statistics in Behavioral Science*. American Cancer Society. https://doi.org/10.1002/0470013192.bsa250

Harris, A., Buxton, M., O'Brien, B., Rutten, F., & Drummond, M. (2001). Using economic evidence in reimbursement decisions for health technologies: Experience of 4 countries. *Expert Review of Pharmacoeconomics & Outcomes Research*, *1*(1), 7–12. https://doi.org/10.1586/14737167.1.1.7

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, *46*(6), 1251–1271. JSTOR. https://doi.org/10.2307/1913827

Haywood, K. L., de Wit, M., Staniszewska, S., Morel, T., & Salek, S. (2017). Developing patient-reported and relevant outcome measures. In Patient involvement in health technology assessment (pp. 103–120). Springer.

HERC database of mapping studies, Version 7.0 (last updated: 24th April 2019). (n.d.). http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies.

Herdman, M., Gudex, C., Lloyd, A., Janssen, Mf., Kind, P., Parkin, D., Bonsel, G., & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, *20*(10), 1727–1736. https://doi.org/10.1007/s11136-011-9903-x

Hernández-Alava, M., & Pudney, S. (2017). Econometric modelling of multiple self-reports of health states: The switch from EQ-5D-3L to EQ-5D-5L in evaluating drug therapies for rheumatoid arthritis. *Journal of Health Economics*, *55*, 139–152. https://doi.org/10.1016/j.jhealeco.2017.06.013

Hernandez-Alava, M., & Wailoo, A. (2015). Fitting Adjusted Limited Dependent Variable Mixture Models to EQ-5D. *The Stata Journal*, 15, Number 3, pp. 737–750.

Hughes, D., Charles, J., Dawoud, D., Edwards, R. T., Holmes, E., Jones, C., Parham, P., Plumpton, C., Ridyard, C., Lloyd-Williams, H., Wood, E., & Yeo, S. T. (2016). Conducting Economic Evaluations Alongside Randomised Trials: Current Methodological Issues and Novel Approaches. *PharmacoEconomics*, *34*(5), 447–461. https://doi.org/10.1007/s40273-015-0371-y

Hui, D., & Bruera, E. (2017). The Edmonton Symptom Assessment System 25 Years Later: Past, Present and Future Developments. *Journal of Pain and Symptom Management*, *53*(3), 630–643. https://doi.org/10.1016/j.jpainsymman.2016.10.370

Janssen, M. F., Bonsel, G. J., & Luo, N. (2018). Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive Systems and Value Sets from Seven Countries. *Pharmacoeconomics*, *36*(6), 675–697. https://doi.org/10.1007/s40273-018-0623-8

Jenkinson, C. (2013). *Measuring Health And Medical Outcomes*. Routledge.

Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S., Plattner, B., Saran, R., Wang, A. Y.-M., & Yang, C.-W. (2013). Chronic kidney disease: Global dimension and perspectives. *The Lancet*, *382*(9888), 260–272. https://doi.org/10.1016/S0140-6736(13)60687-X

Kaambwa, B., & Ratcliffe, J. (2018). Predicting EuroQoL 5 Dimensions 5 Levels (EQ-5D-5L) Utilities from Older People's Quality of Life Brief Questionnaire (OPQoL-Brief) Scores. *Patient*, *11*(1), 39–54. Scopus. https://doi.org/10.1007/s40271-017-0259-3

Kaambwa, Billingsley, Chen, G., Ratcliffe, J., Iezzi, A., Maxwell, A., & Richardson, J. (2017). Mapping Between the Sydney Asthma Quality of Life Questionnaire (AQLQ-S) and Five Multi-Attribute Utility Instruments (MAUIs). *PharmacoEconomics*, *35*(1), 111–124. https://doi.org/10.1007/s40273-016-0446-4

Kaambwa, Billingsley, Smith, C., Lacey, S. de, & Ratcliffe, J. (2018). Does Selecting Covariates Using Factor Analysis in Mapping Algorithms Improve Predictive Accuracy? A Case of Predicting EQ-5D-5L and SF-6D Utilities from the Women's Health Questionnaire. *Value in Health*, *21*(10), 1205–1217. https://doi.org/10.1016/j.jval.2018.01.020

Karimi, M., & Brazier, J. (2016). Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? *PharmacoEconomics*, *34*(7), 645–649. https://doi.org/10.1007/s40273-016-0389-9

Kay, S., & Ferreira, A. (2014). Mapping the 25-item National Eye Institute Visual Functioning Questionnaire (NEI VFQ-25) to EQ-5D Utility Scores. *Ophthalmic Epidemiology*, *21*(2), 66–78. https://doi.org/10.3109/09286586.2014.888456

Kay, S., Tolley, K., Colayco, D., Khalaf, K., Anderson, P., & Globe, D. (2013). Mapping EQ-5D Utility Scores from the Incontinence Quality of Life Questionnaire among Patients with Neurogenic and Idiopathic Overactive Bladder. *Value in Health*, *16*(2), 394–402. https://doi.org/10.1016/j.jval.2012.12.005

Kazis, L. E., Anderson, J. J., & Meenan, R. F. (1989). Effect Sizes for Interpreting Changes in Health Status. *Medical Care*, *27*(3), S178–S189. JSTOR.

Keeney, R. L., L, K. R., & Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press.

Khan, I., & Morris, S. (2014). A non-linear beta-binomial regression model for mapping EORTC QLQ- C30 to the EQ-5D-3L in lung cancer patients: A comparison with existing approaches. *Health and Quality of Life Outcomes*, *12*(1), 1–16. https://doi.org/10.1186/s12955-014-0163-7

Khan, I., Morris, S., Pashayan, N., Matata, B., Bashir, Z., & Maguirre, J. (2016). Comparing the mapping between EQ-5D-5L, EQ-5D-3L and the EORTC-QLQ-C30 in non-small cell

lung cancer patients. *Health and Quality of Life Outcomes*, *14*(1), 1–15. https://doi.org/10.1186/s12955-016-0455-1

Khan, K. A., Petrou, S., Rivero-Arias, O., Walters, S. J., & Boyle, S. E. (2014). Mapping EQ-5D Utility Scores from the PedsQL^{TM} Generic Core Scales. *PharmacoEconomics*, *32*(7), 693–706. https://doi.org/10.1007/s40273-014-0153-y

Kularatna, S., Senanayake, S., Chen, G., & Parsonage, W. (2020). Mapping the Minnesota living with heart failure questionnaire (MLHFQ) to EQ-5D-5L in patients with heart failure. *Health and Quality of Life Outcomes*, *18*(1), 115. https://doi.org/10.1186/s12955-020-01368-2

Lamu, Admassu N., & Olsen, J. A. (2018). Testing alternative regression models to predict utilities: Mapping the QLQ-C30 onto the EQ-5D-5L and the SF-6D. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *27*(11), 2823–2839. https://doi.org/10.1007/s11136-018-1981-6

Lamu, A.N., Chen, G., Gamst-Klaussen, T., & Olsen, J. A. (2018). Do country-specific preference weights matter in the choice of mapping algorithms? The case of mapping the Diabetes-39 onto eight country-specific EQ-5D-5L value sets. *Quality of Life Research*, *27*(7), 1801–1814. Scopus. https://doi.org/10.1007/s11136-018-1840-5

Lee, C. F., Ng, R., Luo, N., & Cheung, Y. B. (2018). Development of Conversion Functions Mapping the FACT-B Total Score to the EQ-5D-5L Utility Value by Three Linking Methods and Comparison with the Ordinary Least Square Method. *Applied Health Economics and Health Policy*, *16*(5), 685–695. Scopus. https://doi.org/10.1007/s40258-018-0404-8

Lee, M. S., Yoon, J., Kim, J., Shin, Y. S., Park, H. S., Lee, H. Y., & Ye, Y. M. (2020). Health-Related Utility of EQ-5D in Korean Adults With Chronic Urticaria: Mapping From Urticaria Outcome Measures. *Allergy, Asthma & Immunology Research*, *12*(4), 599–607. https://doi.org/10.4168/aair.2020.12.4.599

Levey, A. S., & Coresh, J. (2012). Chronic kidney disease. *The Lancet*, *379*(9811), 165–180. https://doi.org/10.1016/S0140-6736(11)60178-5

Levin, A., Hemmelgarn, B., Culleton, B., Tobe, S., McFarlane, P., Ruzicka, M., Burns, K., Manns, B., White, C., Madore, F., Moist, L., Klarenbach, S., Barrett, B., Foley, R., Jindal, K., Senior, P., Pannu, N., Shurraw, S., Akbari, A., … Tonelli, M. (2008). Guidelines for the management of chronic kidney disease. *CMAJ*, *179*(11), 1154–1162. https://doi.org/10.1503/cmaj.080351

Lin, F.-J., Longworth, L., & Pickard, A. S. (2013). Evaluation of content on EQ-5D as compared to disease-specific utility measures. *Quality of Life Research*, *22*(4), 853–874. https://doi.org/10.1007/s11136-012-0207-6

Lipscomb, J., Gotay, C. C., & Snyder, C. (2004). *Outcomes Assessment in Cancer: Measures, Methods and Applications*. Cambridge University Press.

Liu, T., Li, S., Wang, M., Sun, Q., & Chen, G. (2020). Mapping the Chinese Version of the EORTC QLQ-BR53 Onto the EQ-5D-5L and SF-6D Utility Scores. *The Patient*. https://doi.org/10.1007/s40271-020-00422-x

Longworth, L., & Rowen, D. (2013). Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value in Health*, *16*(1), 202–210. https://doi.org/10.1016/j.jval.2012.10.010

McCabe, C., Edlin, R., Meads, D., Brown, C., & Kharroubi, S. (2013). Constructing Indirect Utility Models: Some Observations on the Principles and Practice of Mapping to Obtain Health State Utilities. *PharmacoEconomics*, *31*(8), 635–641. https://doi.org/10.1007/s40273-013-0071-4

Meregaglia, M., Borsoi, L., Cairns, J., & Tarricone, R. (2019). 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*, *20*(2), 181–193. https://doi.org/10.1007/s10198-017-0930-6

Michalos, A. C. (2014). *Encyclopedia of quality of life and well-being research*. Springer Netherlands Dordrecht.

Mihaylova, B., Briggs, A., O'Hagan, A., & Thompson, S. G. (2011). Review of Statistical Methods for Analysing Healthcare Resources and Costs. *Health Economics*, *20*(8), 897–916. https://doi.org/10.1002/hec.1653

Mitchell, P. M., Al-Janabi, H., Byford, S., Kuyken, W., Richardson, J., Iezzi, A., & Coast, J. (2017). Assessing the validity of the ICECAP-A capability measure for adults with depression. *BMC Psychiatry*, *17*(1), 46. https://doi.org/10.1186/s12888-017-1211-8

Moore, A., Young, C. A., & Hughes, D. A. (2018). Mapping ALSFRS-R and ALSUI to EQ-5D in Patients with Motor Neuron Disease. *Value in Health*, *21*(11), 1322–1329. Scopus. https://doi.org/10.1016/j.jval.2018.05.005

Mortimer, D., & Segal, L. (2008). Comparing the Incomparable? A Systematic Review of Competing Techniques for Converting Descriptive Measures of Health Status into QALY-Weights. *Medical Decision Making*, *28*(1), 66–89. https://doi.org/10.1177/0272989X07309642

Moskovitz, M., Jao, K., Su, J., Brown, M. C., Naik, H., Eng, L., Wang, T., Kuo, J., Leung, Y., Xu, W., Mittman, N., Moody, L., Barbera, L., Devins, G., Li, M., Howell, D., & Liu, G. (2019). Combined cancer patient–reported symptom and health utility tool for routine clinical implementation: A real-world comparison of the ESAS and EQ-5D in multiple cancer sites. *Current Oncology*, *26*(6). https://doi.org/10.3747/co.26.5297

Mukaka, M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi*, *24*(3), 69–71.

Mukuria, C., Rowen, D., Harnan, S., Rawdin, A., Wong, R., Ara, R., & Brazier, J. (2019). An Updated Systematic Review of Studies Mapping (or Cross-Walking) Measures of Health-Related Quality of Life to Generic Preference-Based Measures to Generate Utility Values. *Appl Health Econ Health Policy*, *17*(3), 295–313. https://doi.org/10.1007/s40258-019-00467-6

Murphy, E. L., Murtagh, F. E. M., Carey, I., & Sheerin, N. S. (2009). Understanding symptoms in patients with advanced chronic kidney disease managed without dialysis: Use of a

short patient-completed assessment tool. *Nephron. Clinical Practice*, *111*(1), c74-80. https://doi.org/10.1159/000183177

Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: Background, derivation, and applications. *WIREs Computational Statistics*, *4*(2), 199–203. https://doi.org/10.1002/wics.199

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384. https://doi.org/10.2307/2344614

National Institute for Health and Care Excellence (NICE). (2013). *Guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence (NICE)*. NICE. https://www.nice.org.uk/process/pmg9/chapter/the-reference-case#measuring-and-valuing-health-effects

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

Patrick, D. L., & Deyo, R. A. (1989). Generic and Disease-Specific Measures in Assessing Health Status and Quality of Life. *Medical Care*, *27*(3), S217–S232. JSTOR.

Patton, T., Hu, H., Luan, L., Yang, K., & Li, S.-C. (2018). Mapping between HAQ-DI and EQ-5D-5L in a Chinese patient population. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *27*(11), 2815–2822. https://doi.org/10.1007/s11136-018-1925-1

Payakachat, N., Summers, K. H., Pleil, A. M., Murawski, M. M., Thomas, J., Jennings, K., & Anderson, J. G. (2009). Predicting EQ-5D utility scores from the 25-item National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients with age-related macular degeneration. *Quality of Life Research*, *18*(7), 801–813. https://doi.org/10.1007/s11136-009-9499-6

Peak, J., Goranitis, I., Day, E., Copello, A., Freemantle, N., & Frew, E. (2018). Predicting health-related quality of life (EQ-5D-5 L) and capability wellbeing (ICECAP-A) in the context of opiate dependence using routine clinical outcome measures: CORE-OM, LDQ and TOP. *Health and Quality of Life Outcomes*, *16*(1). Scopus. https://doi.org/10.1186/s12955-018-0926-7

Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., & Gray, A. (2015). The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. *PharmacoEconomics*, *33*(10), 993–1011. https://doi.org/10.1007/s40273-015-0312-9

Position statement on use of the EQ-5D-5L value set for England (updated October 2019) | NICE technology appraisal guidance | NICE guidance | Our programmes | What we do | About. (n.d.). [CorporatePage]. NICE; NICE. Retrieved June 4, 2020, from https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l

Romagnani, P., Remuzzi, G., Glassock, R., Levin, A., Jager, K. J., Tonelli, M., Massy, Z., Wanner, C., & Anders, H.-J. (2017). Chronic kidney disease. *Nature Reviews Disease Primers*, *3*(1), 1–24. https://doi.org/10.1038/nrdp.2017.88

Round, J. (2008). Capturing information loss in estimates of uncertainty that arise from mapping algorithms. *Aberdeen: Health Economics Study Group (HESG)*.

Round, Jeff. (2012). Is a QALY still a QALY at the end of life? *Journal of Health Economics*, *31*(3), 521–527. https://doi.org/10.1016/j.jhealeco.2012.01.006

Round, Jeff, & Hawton, A. (2017). Statistical Alchemy: Conceptual Validity and Mapping to Generate Health State Utility Values. *PharmacoEconomics - Open*, *1*(4), 233–239. https://doi.org/10.1007/s41669-017-0027-2

Rowen, D., Brazier, J., & Roberts, J. (2009). Mapping SF-36 onto the EQ-5D index: How reliable is the relationship? *Health and Quality of Life Outcomes*, *7*(1), 1–9. https://doi.org/10.1186/1477-7525-7-27

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel*, *81*.

Sassi, F. (2006). Calculating QALYs, comparing QALY and DALY calculations. *Health Policy and Planning*, *21*(5), 402–408. https://doi.org/10.1093/heapol/czl018

Sauerbrei, W., Royston, P., & Binder, H. (2007). Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Statistics in Medicine*, *26*(30), 5512–5528. https://doi.org/10.1002/sim.3148

Skaltsa, K., Longworth, L., Ivanescu, C., Phung, D., & Holmstrom, S. (2014). Mapping the FACT-P to the Preference-Based EQ-5D Questionnaire in Metastatic Castration-Resistant Prostate Cancer. *Value in Health*, *17*(2), 238–244. https://doi.org/10.1016/j.jval.2013.12.005

Stein, K., Fry, A., Round, A., Milne, R., & Brazier, J. (2005). What value health? *Applied Health Economics and Health Policy*, *4*(4), 219–228. https://doi.org/10.2165/00148365-200504040-00004

Stephens, R. F., Noel, C. W., Su, J. S., Xu, W., Krahn, M., Monteiro, E., Goldstein, D. P., Giuliani, M., Hansen, A. R., & de Almeida, J. R. (2019). Mapping the University of Washington Quality of life questionnaire onto EQ-5D and HUI-3 indices in patients with head and neck cancer. *Head Neck*. https://doi.org/10.1002/hed.26031

Su, J., Liu, T., Li, S., Zhao, Y., & Kuang, Y. (2020). A mapping study in mainland China: Predicting EQ-5D-5L utility scores from the psoriasis disability index. *Journal of Medical Economics*. Scopus. https://doi.org/10.1080/13696998.2020.1748636

Sullivan, P. W., & Ghushchyan, V. H. (2016). EQ-5D Scores for Diabetes-Related Comorbidities. *Value in Health*, *19*(8), 1002–1008. https://doi.org/10.1016/j.jval.2016.05.018

Torrance, G. W. (1986). Measurement of health state utilities for economic appraisal: A review. *Journal of Health Economics*, *5*(1), 1–30. https://doi.org/10.1016/0167-6296(86)90020-2

Torrance, G. W., & Feeny, D. (1989). Utilities and Quality-Adjusted Life Years. *International Journal of Technology Assessment in Health Care*, *5*(4), 559–575. https://doi.org/10.1017/S0266462300008461

Torrance, G. W., Furlong, W., & Feeny, D. (2002). Health utility estimation. *Expert Review of Pharmacoeconomics & Outcomes Research*, *2*(2), 99–108. https://doi.org/10.1586/14737167.2.2.99

van Hout, B., Janssen, M. F., Feng, Y.-S., Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P., & Pickard, A. S. (2012). Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. *Value in Health*, *15*(5), 708–715. https://doi.org/10.1016/j.jval.2012.02.008

Vilsbøll, A. W., Kragh, N., Hahn-Pedersen, J., & Jensen, C. E. (2020). Mapping Dermatology Life Quality Index (DLQI) scores to EQ-5D utility scores using data of patients with atopic dermatitis from the National Health and Wellness Study. *Quality of Life Research*. Scopus. https://doi.org/10.1007/s11136-020-02499-1

Vokó, Z., Németh, R., Nagyjánosi, L., Jermendy, G., Winkler, G., Hídvégi, T., Kalotai, Z., & Kaló, Z. (2014). Mapping the Nottingham Health Profile onto the Preference-Based EuroQol-5D Instrument for Patients with Diabetes. *Value in Health Regional Issues*, *4*, 31–36. https://doi.org/10.1016/j.vhri.2014.06.002

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior* (pp. xviii, 625). Princeton University Press.

Wadsworth, I., Van Horn, M. L., & Jaki, T. (2018). A DIAGNOSTIC TOOL FOR CHECKING ASSUMPTIONS OF REGRESSION MIXTURE MODELS. *JP Journal of Biostatistics*, *15*(1), 1–20. https://doi.org/10.17654/BS015010001

Wailoo, A. J., Hernandez-Alava, M., Manca, A., Mejia, A., Ray, J., Crawford, B., Botteman, M., & Busschbach, J. (2017). Mapping to Estimate Health-State Utility from Non–Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value in Health*, *20*(1), 18–27. https://doi.org/10.1016/j.jval.2016.11.006

Ware Jr, J. E. (1995). The status of health assessment 1994. *Annual Review of Public Health*, *16*(1), 327–354.

Wee, H. L., Yeo, K. K., Chong, K. J., Khoo, E. Y. H., & Cheung, Y. B. (2018). Mean Rank, Equipercentile, and Regression Mapping of World Health Organization Quality of Life Brief (WHOQOL-BREF) to EuroQoL 5 Dimensions 5 Levels (EQ-5D-5L) Utilities. *Medical Decision Making*, *38*(3), 319–333. Scopus. https://doi.org/10.1177/0272989X18756890

Wijnen, B. F. M., Mosweu, I., Majoie, M. H. J. M., Ridsdale, L., de Kinderen, R. J. A., Evers, S. M. A. A., & McCrone, P. (2018). A comparison of the responsiveness of EQ-5D-5L and the QOLIE-31P and mapping of QOLIE-31P to EQ-5D-5L in epilepsy. *European Journal of Health Economics*, *19*(6), 861–870. Scopus. https://doi.org/10.1007/s10198-017-0928-0

Wong, C. K. H., Cheung, P. W. H., Samartzis, D., Luk, K. D.-K., Cheung, K. M. C., Lam, C. L. K., & Cheung, J. P. Y. (2017). Mapping the SRS-22r questionnaire onto the EQ-5D-5L utility score in patients with adolescent idiopathic scoliosis. *PLOS ONE*, *12*(4), e0175847. https://doi.org/10.1371/journal.pone.0175847

Wong, J. B. (2015). Principles of health economics and application to rheumatic disorders. In *Rheumatology* (pp. 20–26). Elsevier.

Wooldridge, J. M. (c2002.). *Econometric analysis of cross section and panel data*. MIT Press,.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach /* (5th ed.). South-Western Cengage Learning,.

World Health Organization. (1995). *Constitution of the world health organization*.

Xie, F., Pullenayegum, E., Gaebel, K., Bansback, N., Bryan, S., Ohinmaa, A., Poissant, L., & Johnson, J. A. (2016). A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. *Medical Care*, *54*(1), 98–105. https://doi.org/10.1097/MLR.0000000000000447

Yang, F., Devlin, N., & Luo, N. (2019). Impact of mapped EQ-5D utilities on cost-effectiveness analysis: In the case of dialysis treatments. *Eur J Health Econ*, *20*(1), 99–105. https://doi.org/10.1007/s10198-018-0987-x

Yang, F., Wong, C. K. H., Luo, N., Piercy, J., Moon, R., & Jackson, J. (2019). Mapping the kidney disease quality of life 36-item short form survey (KDQOL-36) to the EQ-5D-3L and the EQ-5D-5L in patients undergoing dialysis. *Eur J Health Econ*, *20*(8), 1195–1206. https://doi.org/10.1007/s10198-019-01088-5

Yang, Q., Yu, X. X., Zhang, W., & Li, H. (2019). Mapping function from FACT-B to EQ-5D-5 L using multiple modelling approaches: Data from breast cancer patients in China. *Health Qual Life Outcomes*, *17*(1), 153. https://doi.org/10.1186/s12955-019-1224-8

Appendices

Appendix 2.1 Search strategies

Searched terms:

"Mapping" related: Mapping, Map, Mapped, Crosswalk, Cross-walk, cross walk, Transfer to utility, Transfer-to-utility, indirect utility

"EQ-5D-5L" related: Eq-5d, Eq-5d-5l, Euroqol, eq5d

EMBASE

# ▲	Searches	Results	Type	Actions	Annotations
1	(map* or cross walk or crosswalk* or cross-walk* or transfer* to utilit* or transfer-to-utilit* or indirect utilit*).tw.	574286	Advanced	Display Results | More	
2	limit 1 to yr="2018 -Current"	87701	Advanced	Display Results | More	
3	(eq5d* or eq-5d* or euroqol).tw.	20124	Advanced	Display Results | More	
4	limit 3 to yr="2018 -Current"	5527	Advanced	Display Results | More	
5	2 and 4	198	Advanced	Display Results | More	

MEDLINE (via PubMed)

Search	Add to builder	Query	Items found
#7	Add	Search **((((((((((((mapping[Text Word]) OR map[Text Word]) OR mapped[Text Word]) OR cross walk* [Text Word]) OR crosswalk*[Text Word]) OR cross-walk*[Text Word]) OR transfer to utilit*[Text Word]) OR transfer-to-utilit*[Text Word]) OR indirect utilit*[Text Word]) AND ("2018/07/01"[PDat] : "3000/12/31"[PDat])))) AND ((((eq5d*[Text Word]) OR euroqol*[Text Word]) OR eq-5d*[Text Word]) AND ("2018/07/01"[Date - Publication] : "3000"[Date - Publication]))**	87
#6	Add	Search **(((eq5d*[Text Word]) OR euroqol*[Text Word]) OR eq-5d*[Text Word]) AND ("2018/07/01"[Date - Publication] : "3000"[Date - Publication])**	2713
#5	Add	Search **((((((((((mapping[Text Word]) OR map[Text Word]) OR mapped[Text Word]) OR cross walk* [Text Word]) OR crosswalk*[Text Word]) OR cross-walk*[Text Word]) OR transfer to utilit*[Text Word]) OR transfer-to-utilit*[Text Word]) OR indirect utilit*[Text Word]) AND ("2018/07/01"[PDat] : "3000/12/31"[PDat]))**	47935

Appendix 2.2 Data extraction form

Title: Authors:

Publication year: Best model:

Whether cited the three checklists?

Abstract:

Introduction – study rationale

Introduction – study objective

Methods – Data: estimation, external validation, missing

Methods – Source and Target measures

Methods – Exploratory analysis (overlaps?)

Methods – Model types, specifications, and estimation (clinical important covariates?)

Methods - Validation

Methods – Performance

Results – Final sample size, descriptive information (utility plot?)

Results – Model performance, assumptions, plots, face validity

Results – Model selection

Results – Model coefficients

Results – Uncertainty

Results – Calculation example

Discussion – Comparison with other studies

Discussion – Limitation

Discussion – Scope

Other information – Source of funding, conflict of interest

Appendix 2.3 Reporting quality assessment form

	Completed	Partially completed	No
MAPS - Title			
MAPS - Abstract			
MAPS - Introduction: study rationale			
MAPS - Introduction: study objective			
ISPOR - Methods: candidate dataset description			
MAPS, ISPOR - Methods: estimation sample			
MAPS - Methods: external validation sample			
MAPS - Methods: source and target measures			
MAPS, ISPOR - Methods: exploratory data analysis			
MAPS, ISPOR - Methods: missing data			
MAPS, ISPOR - Methods: modelling approaches			
MAPS - Methods: estimation of predicted utilities			
MAPS - Methods: validation methods			
MAPS - Methods: measures of model performance			
ISPOR - Methods: Approach to determine the final model			
MAPS - Results: final sample size			
MAPS, ISPOR - Results: descriptive information (especially severity distribution)			
ISPOR - Results: Utility data distribution plot			
ISPOR - Results: Fit statistics, especially conditional fit statistics			
ISPOR - Results: lot on observed and predicted utility values			
MAPS - Results: model selection			
MAPS, ISPOR - Results: model coefficients (size, sign, significance, stand error, error term)			
ISPOR - Results: model coefficients additional requirements (variance-covariance matrix, error)			
ISPOR - Results: Calculation example / user-friendly program			
MAPS, ISPOR - Results: uncertainty (parameter)			
MAPS, ISPOR - Results: uncertainty (individual)			
MAPS - Results: model performance and face validity			
MAPS - Discussion: comparisons with previous studies			
MAPS - Discussion: study limitations			
MAPS - Discussion: scope of applications			
MAPS - Other: additional information			

Note: MAPS: Mapping onto Preference-based Measures Reporting Standards Reporting Statement[19], ISPOR: ISPOR Good Practices Task Force Report on Mapping[14]. The highlighted items are directly associated with the application in CUA.

Appendix 2.4 Complete data extraction

Abbreviations: OLS: ordinary least square, CLAD: Censored least absolute deviations, GLM: generalized linear model, robust MM: robust MM estimation, FRM: fractional regression model, LE: linear equating, CPA: conditional process analysis, BETAMIX: mixture beta regression model, TPM: two-part model, GMM: gaussian mixture model, ALDVMM: adjusted limited dependent variable mixture model, ER: equipercentile regression: MRM: mean rank model, GEE: generalized estimating equations, QR: quantile regression, CM: copula model, EEE: extended estimating equations, LREM: linear random effect model

	Year	Disease type	sample size	Instru -ment	Multiple instruments	Multiple value-sets	Final model type	Model types considered
Ameri [65]	2018	cancer	252	QLQ-C30	2 targets		OLS	OLS; Tobit; CLAD
Ameri [66]	2020	cancer	252	EORTC-QLQ-C30, QLQ-CR29	2 sources		OLS	OLS
Bilbao [59]	2019	musculoskeletal	758	WOMAC			GLM, Beta regression	OLS; Tobit; Beta regression
Dixon [55]	2020	eye	1181				ALDVMM	OLS; ALDVMM
Kularatna [34]	2020	heart	141	MLHFQ	3 targets		Logit/probit, MFP	OLS; CLAD; Beta regression; GLM; logit/probit; MFP; robust MM
Lamu [33]	2020	heart	943	MacNew		2	Beta regression	OLS; Beta regression; GLM; robust MM; FRM; LE
Lee [47]	2020	skin	416	UAS7 UCT	2 sources		CPA	CPA
Liu [35]	2020	cancer	607	EORTC BR	2 targets		Tobit CLAD	OLS; Tobit; CLAD; GLM; logit/probit; robust MM; BETAMIX

Noel [45]	2020	cancer	209	EORTC QLQ C30, EORTC QLQ HN35	2 target, 2 sources		OLS	OLS; TPM
Stephen [57]	2019	cancer	209	UWQol	2 targets	used 5l mapped to 3l value set	OLS	OLS
Su [61]	2020	skin	321	PDI			OLS Tobit GLM	OLS; Tobit; GLM
Vilsboll [54]	2020	skin	1232	DLQI			GMM	OLS; Tobit; TPM; GMM
Yang F [60]	2019	Urogenital	163	KDQOL-36	2 targets	4 value sets	ALDVMM	OLS; ALDVMM; BETAMIX
Yang Q [44]	2019	cancer	446	FACT-B			TPM	OLS; Tobit; TPM; LE
Wong [67]	2017	musculoskeletal	227	SRS-22r		used 5l mapped to 3l value set	OLS	OLS
Wijnen [56]	2018	central nervous system	283	QOLIE-31P		2 value sets	OLS	OLS; CLAD
Wee [46]	2018	general population	658	WHOQOL-BREF			MRM	OLS; ER; MRM
Chen [39]	2016	Various	4461-6415	6D HUI3 15D QWB 8D	5 sources	multi-country but used 1 value set	15D, 6D: OLS, 8D, QWB: robust MM, HUI: GLM	OLS; CLAD; GLM; robust MM
Peak [68]	2018	Mental health and behavioural disorders	83	COREOM, LDQ, TOP	3 sources		OLS	OLS; Tobit
Patton [40]	2018	musculoskeletal	130	HAQ-DI			Logit/probit	Beta regression; logit/probit
Moore [41]	2018	central nervous system	595	ALSFRS-R			OLS	OLS; Tobit; logit/probit
Mitchell [30]	2017	Mental health and behavioural disorders	617	DASS-D K10 K6	3 sources		OLS	OLS

Meregaglia [48]	2019	cancer	332	FACTG FAACT TACIT-F	3 sources		GEE	GEE
Lee [27]	2018	cancer	238	FACT-B			OLS; LE; ER; MRM	OLS; LE; ER; MRM
Lamu [36]	2018	Endocrine disorders	924	D-39		several vale sets	FRM	OLS; CLAD; Beta regression; GLM; robust MM; FRM
Lamu & Olsen [42]	2018	cancer	772	EORTC QLQ-C30	2 targets		Beta regression, EEE	OLS; CLAD; Beta regression; GLM; logit/probit; FRM; EEE
Khan [69]	2016	cancer	985	EORTC QLQ-C30	2 targets	used 5l mapped to 3l value set	Beta regression	Beta regression; ALDVMM
Abdin [58]	2019	Mental health and behavioural disorders	239	PANSS			OLS	OLS; Tobit; CLAD
Chen [37]	2018	central nervous system	228	PDQ8	3 targets		Logit/probit	OLS; GLM; logit/probit; robust MM
Cheung [70]	2014	cancer	238	FACT-B			OLS	OLS; Tobit; CLAD
Collado-Mateo [71]	2017	musculoskeletal	191	FIQR			GLM	OLS; GLM
Coon [72]	2018	Urogenital	352	MENQOL			GLM	OLS; GLM
Gamst-Klaussen [29]	2016	various	7930	6D HUI3 15D	3 sources		QR	QR
Gamst-Klaussen [38]	2018	Mental health and behavioural disorders	917	DASS-21 K-10	2 sources		Beta regression	OLS; Beta regression; GLM; robust MM; FRM
Gray [73]	2018	Respiratory system	856	AOLQS	2 targets		BETAMIX	OLS; ALDVMM; BETAMIX

Hernandez [74]	2017	musculoskeletal	5192	EQ-5D, HAQ and pain on VAS	3 sources	CM	CM
Kaambwa [26]	2018	General population	303	WHQ23	2 targets	Logit/probit	OLS; CLAD; Beta regression; GLM; logit/probit; robust MM
Kaambwa [75]	2017	general population	642	AQLQ-S	5 targets	GLM	OLS; CLAD; Beta regression; GLM
Kaambwa & Ratcliffe [49]	2018	Respiratory system	330	opqol-brief		OLS	OLS; Beta regression; ALDVMM; LREM

MAPS reporting statement[19]: ("*" represents applied the checklist, and "cited" means cited the checklist)

		Title	Abstract	Introduction: study rationale	Introduction: study objective	Methods: estimation sample	Methods: external validation sample	Methods: source and target measures
	Ameri [65]	yes	no implications	yes	yes	yes	NA	partially, high/low score meaning
	Ameri [66]	yes	yes	yes	yes	yes	NA	yes
	Bilbao [59]	yes	yes	yes	yes	yes	NA	yes
*	Dixon [55]	yes	no performance stats	yes	yes	yes	NA	partially, high/low score meaning
cited	Kularatna [34]	yes	no performance stats	yes	yes	yes	NA	yes
*	Lamu [33]	yes	yes, but performance stats only R2	yes	yes	yes	NA	yes

	Lee [47]	partially, "EQ-5D"	no estimation & validation	partially	partially	yes	NA	partially, high/low score meaning
*	Liu [35]	yes	no performance stats	yes	yes	yes	NA	partially, high/low score meaning
*	Noel [45]	yes	no validation	yes	yes	yes	NA	partially, high/low score meaning
*	Stephen [57]	partially, "EQ-5D"	yes	yes	yes	yes	NA	yes
cited	Su [61]	yes	yes	yes	yes	yes	NA	yes
*	Vilsboll [54]	partially, "EQ-5D"	no validation	yes	yes	yes	NA	yes
cited	Yang F [60]	yes	no performance stats	yes	yes	yes	NA	yes
	Yang Q [44]	yes	no validation	yes	yes	yes	NA	partially, high/low score meaning
cited	Wong [67]	yes	not structured	yes	yes	yes	NA	partially, high/low score meaning
	Wijnen [56]	yes	no validation	yes	yes	yes	NA	yes
*	Wee [46]	yes	no performance stats	yes	yes	yes	NA	partially, high/low score meaning
	Chen [39]	partially, "six instruments"	no validation no performance	yes	yes	yes	NA	partially, high/low score meaning
cited	Peak [68]	partially, no "mapping"	yes	yes	yes	yes	NA	yes
	Patton [40]	yes	no performance statistics	yes	yes	yes	NA	partially, high/low score meaning
*	Moore [41]	partially "EQ-5D"	no data source no validation	yes	yes	yes	NA	yes
	Mitchell [30]	no	only method, primary objective	yes	yes	yes	NA	yes
cited	Meregaglia [48]	yes	no performance statistics	yes	yes	yes	NA	partially, high/low score meaning
	Lee [27]	yes	no validation, results not clear	yes	yes	yes	NA	yes
*	Lamu [36]	no population	yes	yes	yes	yes	NA	yes

*	Lamu & Olsen [42]	no population	yes	yes	yes	yes	NA	yes
	Khan [69]	yes	validation not clearly	yes	yes	yes	NA	partially, high/low score meaning
	Abdin [58]	yes	no validation, no performance	yes	yes	yes	NA	yes
	Chen [37]	partially, "health state utilities"	no validation, no performance, no conclusion	yes	yes	yes	NA	partially, high/low score meaning
	Cheung [70]	yes	no performance stats	yes	yes	yes	NA	partially, high/low score meaning
	Collado-Mateo [71]	partial, no target	no validation	yes	yes	yes	NA	partially, high/low score meaning
	Coon [72]	yes	no validation	yes	yes	yes	NA	yes
	Gamst-Klaussen [29]	no mapping	no validation no performance	yes	yes	yes	NA	partially, high/low score meaning
*	Gamst-Klaussen [38]	yes	no validation no performance	yes	yes	yes	NA	partially, high/low score meaning
	Gray [73]	yes	no validation no performance	yes	yes	yes	NA	yes
	Hernandez [74]	no mapping	no validation no results no performance	yes	yes	yes	NA	yes (introduction section)
*	Kaambwa [26]	yes	yes	yes	yes	yes	NA	yes
cited	Kaambwa [75]	partially, "five MAUIs"	no performance statistics	yes	yes	yes	NA	yes
cited	Kaambwa Ratcliffe [49]	yes	no validation	yes	yes	yes	yes	yes

(continued)

Name	Methods: exploratory data analysis	Methods: missing data	Methods: modelling approaches	Methods: estimation of predicted utilities	Methods: validation methods	Methods: measures of model performance	Results: final sample size
Ameri [65]	No	mentioned in results	yes	no	yes	yes	yes
Ameri [66]	No	mentioned in results	yes	no	yes	yes	yes
Bilbao [59]	correlation	yes	yes	no	yes	yes	yes, but not consistent
Dixon [55]	correlation	yes	yes	no	no validation	yes	yes
Kularatna [34]	correlation	no	yes	no	yes	yes	yes
Lamu [33]	yes	yes	yes	no	yes	yes	not clear
Lee [47]	correlation	no	no	no	no validation	no	yes
Liu [35]	correlation	yes	yes	no	yes	yes	yes
Noel [45]	results section univariate, Wilcoxon signed-rank test (but no results)	volume but no solution	yes	no	yes	yes	yes
Stephen [57]	Wilcoxon signed-rank test, but no results	volume but no solution	yes	no	yes	yes	yes
Su [61]	correlation	yes	yes	no	yes	yes	yes
Vilsboll [54]	correlation	yes	yes	no	yes	yes	yes
Yang F [60]	correlation	no	yes	no	yes	yes	yes
Yang Q [44]	correlation	no	yes	no	yes	yes	yes
Wong [67]	previous literature correlation	no	yes	no	yes	yes	yes
Wijnen [56]	correlation	mentioned in results	yes	no	yes	yes	partially, stated in methods
Wee [46]	yes	yes	yes	no	yes	yes	no

Chen [39]	No	partially, illustrated bad-quality data, not really missing data	yes	no	yes	yes	partially, obtained from coefficients table
Peak [68]	No	no	yes	no	yes	yes	yes
Patton [40]	No	stated in results	yes	partially, stated for one model	yes	yes	yes
Moore [41]	No	yes	yes	yes	yes	yes	yes
Mitchell [30]	No	no	yes	no	no validation	no	yes
Meregaglia [48]	correlation	yes	yes	no	yes	yes	yes
Lee [27]	No	yes	yes	no	yes	yes	yes
Lamu [36]	yes	yes	yes	no	yes	yes	yes
Lamu & Olsen [42]	yes	no	yes	no	yes	yes	yes
Khan [69]	No	mentioned in results	yes	no	yes	yes	yes
Abdin [58]	no	no	yes	yes	no validation	yes	yes
Chen [37]	correlation	yes	yes	no	yes	yes	yes
Cheung [70]	no	yes	yes	no	yes	yes	yes
Collado-Mateo [71]	correlation	no	yes	no	yes	yes	yes
Coon [72]	correlation	yes	yes	no	no validation	yes	yes
Gamst-Klaussen [29]	correlation	no	yes	no	no validation	yes	in table
Gamst-Klaussen [38]	yes	no	yes	no	yes	yes	yes
Gray [73]	no	yes	yes	no	no validation	yes	yes
Hernandez [74]	correlation	no	yes	no	no validation	yes	yes

Kaambwa [26]	correlation	yes	yes	no	yes	yes	yes
Kaambwa [75]	correlation	mentioned in results	yes	no	yes	yes	in table
Kaambwa Ratcliffe [49]	correlation	yes	yes	no	yes	yes	yes

(continued) Note: coef: model coefficients, sig: significance, cov: covariance matrix, indivi: individual uncertainty

	Results: descriptive information	**Results: model selection**	**Results: model coefficients**	**Results: uncertainty**	**Results: model performance and face validity**	**Discussion: comparisons with previous studies**	**Discussion: study limitations**	**Discussion: scope of applications**	**Other: additional information**
Ameri [65]	yes	yes	only coef & sig	no	no face validity statement	yes	yes	no	yes
Ameri [66]	yes	yes	only coef & sig	no	yes	yes	yes	partially	yes
Bilbao [59]	yes	yes	yes	no indivi	partial face validity, discussed unexpected coefficients	yes	yes	partially	yes
Dixon [55]	yes	yes	yes	no indivi	no face validity statement	first study	yes	partially	yes
Kularatna [34]	yes	yes	only coef & sig	no	yes	first study	yes	partially	yes
Lamu [33]	yes	yes	yes	no indivi, no cov	no face validity statement	first study	yes	partially	yes
Lee [47]	yes	only one model	yes	no indivi, no cov	partially, only R2, no validation, partial validity, discussed the effect size	first study	yes	yes	yes

Liu [35]	yes	yes	yes	no indivi, no cov	partial face validity, compared effect size with other studies	yes	yes	partially	yes
Noel [45]	yes	yes	yes	no indivi, no cov	partial face validity, discussed why some variables excluded	yes	yes	partially	no
Stephen [57]	yes	yes	yes	no indivi, no cov	partial face validity, discussed why some variables excluded	yes	yes	partially	yes
Su [61]	yes	yes	yes	no indivi, no cov	partial face validity, discussed correlation and the highly significant variables in the model	first study	yes	yes	yes
Vilsboll [54]	yes	yes	yes	no individual variability	yes	first study	yes	yes	yes
Yang F [60]	yes	yes	yes	no cov	no face validity statement	first study	yes	partially	yes
Yang Q [44]	yes	yes	only coef & sig	no but could calculated from CI, no individual variability	partial face validity, discussed why some variables excluded	yes	yes	partially	yes
Wong [67]	yes	yes	only coef & sig		partial face validity, discussed significance of some variables, not stated scale severity	first study	yes	partially	yes
Wijnen [56]	yes	yes	yes	no indivi, no cov	no face validity statement	no, compared performance statistics with not directly related study	yes	yes	yes
Wee [46]	no	yes	no	no	no face validity statement	first study	yes	yes	yes

Chen [39]	yes	yes	yes	no indivi, no cov	no face validity statement	partially, a study mentioned in introduction did not re-discussed in discussion		partially, discussed why transformation is not country-specific	no
Peak [68]	yes	yes	only coef & sig	no	yes	first study	yes	partially	yes
Patton [40]	yes	yes	yes	no indivi, no cov	no face validity statement	yes	yes	partially	yes
Moore [41]	yes	yes	no sig	no indivi, no cov	no face validity statement	first study	yes	partially	yes
Mitchell [30]	yes	only one model	only coef & sig	no	no face validity statement	no	yes	no	yes
Meregaglia [48]	yes	yes	yes	no indivi, no cov	no face validity statement	yes	yes	partially	yes
Lee [27]	yes	yes	no	no	no face validity statement	no	yes	yes	yes
Lamu [36]	yes	yes	yes	no indivi, no cov	partial face validity, discussed correlation and the highly significant variables in the model	yes	yes	yes	yes
Lamu & Olsen [42]	yes	yes	yes	no indivi, no cov	partial face validity, discussed correlation and the highly significant variables in the model	first study	yes	yes	yes
Khan [69]	yes	yes	yes	no indivi, no cov	no face validity statement	first study	yes	yes	yes
Abdin [58]	yes	yes	only coef & sig	no	yes	yes	yes	yes	yes
Chen [37]	yes	yes	yes	no indivi, no cov	no face validity statement	first study	yes	yes	yes

Cheung [70]	yes	yes	only coef & sig	no	no face validity statement	first study	yes	yes	yes
Collado-Mateo [71]	yes	yes	only coef & sig	no	no face validity statement	first study	yes	yes	yes
Coon [72]	yes	no	only coef & sig	no	yes	first study	yes	yes	yes
Gamst-Klaussen [29]	yes	only one model	yes	no individual variability	no face validity statement	yes	yes	yes	yes
Gamst-Klaussen [38]	yes	yes	yes	no indivi, no cov	no face validity statement	yes	yes	yes	yes
Gray [73]	yes	yes	yes	no cov	no face validity statement	yes	yes	yes	yes
Hernandez [74]	partially	yes	yes	no indivi, no cov	no face validity statement	no	yes	yes	yes
Kaambwa [26]	yes	yes	yes	no	no face validity statement	first study	yes	yes	yes
Kaambwa [75]	yes	yes	yes	no indivi, no cov	no face validity statement	first study	yes	yes	yes
Kaambwa Ratcliffe [49]	yes	only one model	yes	no indivi, no cov	no face validity statement	first study	yes	yes	yes

ISPOR Good Practices[14]: ("*" represents applied the checklist, "cited" means cited the checklist)

	Candidate dataset description	**Full details on selected dataset**	**utility data distribution**	**model types**	**overlaps comparison**	**approach to determine**	**fit, conditional fit, plot (predicted value vs observed value)**
Ameri [65]	NA	yes, missing mentioned in results	yes	yes	no	yes	partially, no conditional

	Ameri [66]	NA	yes, missing mentioned in results	no	yes	no	yes	partially, no conditional
cited	Bilbao [59]	NA	no severity #	yes	yes	correlation	yes	no conditional fit but conditional predictions
cited	Dixon [55]	NA	severity # partially	yes	yes	correlation	yes	yes
	Kularatna [34]	NA	no missing	yes	yes	correlation	yes	partially, no conditional
	Lamu [33]	NA	no severity #	no	yes	yes	yes	partially, no conditional
	Lee [47]	NA	no missing	no	no	correlation	NA	only R2, no plot
cited	Liu [35]	NA	yes	no	yes	correlation	yes	partially, no conditional
*	Noel [45]	NA	yes, missing volume but no solution	no	yes	correlation by univariate, Wilcoxon signed-rank test, but no results	yes	partially, no conditional, indicators not fully reported
cited	Stephen [57]	NA	yes, missing volume but no solution	no	yes	Wilcoxon signed-rank test, but no results	yes	partially, no conditional
cited	Su [61]	NA	yes	no	yes	correlation	yes	partially, no conditional
*	Vilsboll [54]	NA	no severity #	yes	yes	correlation	yes	partially, no conditional
cited	Yang F [60]	NA	no severity #, no missing	yes	yes	correlation	yes	partially, no conditional
	Yang Q [44]	NA	no missing	yes	yes	correlation	yes	partially, no conditional
	Wong [67]	NA	no missing	no	yes	no	yes	partially, no conditional
	Wijnen [56]	NA	no severity #, missing mentioned in results	yes	yes	correlation	yes	partially, no conditional
	Wee [46]	NA	no description	no	yes	least square regression	yes	partially, no conditional, no plot
	Chen [39]	NA	no missing	yes	yes	no	yes	partially, no conditional
	Peak [68]	NA	no missing, no severity #	no	yes	no	yes	partially, no conditional
	Patton [40]	NA	no severity #	no	yes	no	yes	partially, no conditional, no plot

cited	Moore [41]	NA	no severity #	yes	yes	no	yes	yes
	Mitchell [30]	NA	no missing, no severity #	no	yes	no	NA	partially, no conditional, no plot
cited	Meregaglia [48]	NA	no severity #	yes	yes	correlation	yes	partially, no conditional
	Lee [27]	NA	no severity #	yes	yes	no	yes	no conditional fit
	Lamu [36]	NA	no severity #	no	yes	yes	yes	partially, no conditional
	Lamu & Olsen [42]	NA	no missing, no severity #	yes	yes	yes	yes	partially, no conditional
	Khan [69]	NA	yes	yes	yes	no	yes	partially, no conditional
cited	Abdin [58]	NA	no missing	yes	yes	no	yes	partially, no conditional, no plot
	Chen [37]	NA	no missing	yes	yes	no	yes	partially, no conditional
	Cheung [70]	NA	no missing	no	yes	no	yes	partially, no conditional, no plot
	Collado-Mateo [71]	NA	no missing	no	yes	no	yes	partially, no conditional, no plot
	Coon [72]	NA	no missing	no	yes	no	no	partially, no conditional, no plot
	Gamst-Klaussen [29]	NA	no missing	yes	yes	correlation	NA	yes
	Gamst-Klaussen [38]	NA	no missing	no	yes	yes	yes	partially, no conditional
cited	Gray [73]	NA	no missing	yes	yes	no	yes	partially, no conditional
	Hernandez [74]	NA	no missing, no severity #	yes	yes	correlation	NA	partially, no conditional
*	Kaambwa [26]	NA	no severity #	no	yes	correlation	yes	partially, no conditional, no plot
	Kaambwa [75]	NA	no missing no severity #	yes	yes	correlation	yes	partially, no conditional
	Kaambwa	NA	no severity	no	yes	correlation	yes	partially, no conditional

Ratcliffe [49]					

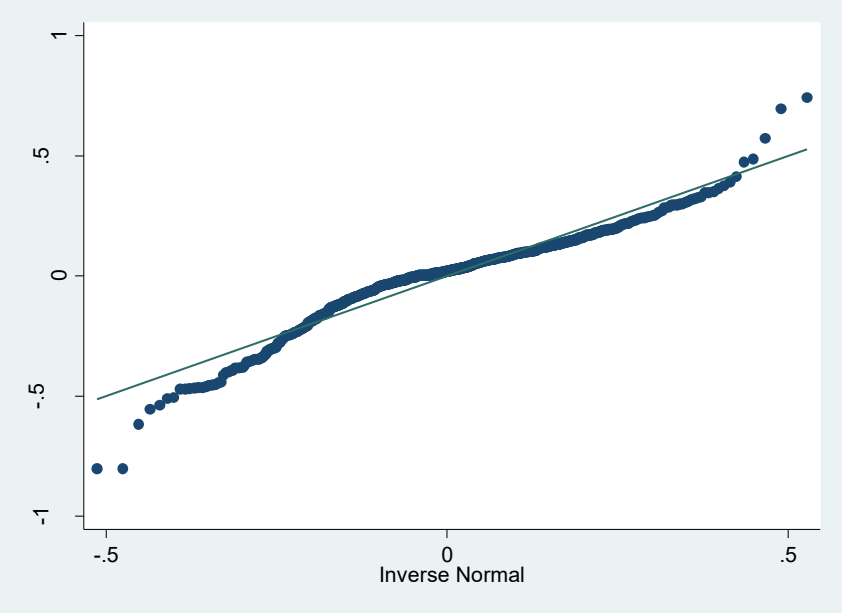
(continued) Note: cov: covariance matrix, CI: confidence interval

Name	Model coefficients, error term, variance-covariance matrix	Calculation example, program	parameter uncertainty, feasible range	individual -level error	validation not routinely required
Ameri [65]	only coefficients	calculation equation	no	no	10-fold
Ameri [66]	only coefficients	calculation equation	no	no	10-fold
Bilbao [59]	no error	calculation equation	yes	no	follow-up
Dixon [55]	no error	no	yes	no	no validation
Kularatna [34]	only coefficients	calculation equation	no	no	3-fold
Lamu [33]	no variance-covariance matrix, no error	calculation process	partially, no cov	no	leave-one-out
Lee [47]	no variance-covariance matrix, no error	no	partially, no cov	no	no validation
Liu [35]	no variance-covariance matrix, no error	calculation equation	partially, no cov	no	2 internal validation
Noel [45]	no variance-covariance matrix, no error	no	partially, no cov	no	10-fold
Stephen [57]	no variance-covariance matrix, no error	no	partially, no cov	no	10-fold
Su [61]	no variance-covariance matrix, no error	calculation equation	partially, no cov	no	hold-out
Vilsboll [54]	no error	program	yes	no	no validation
Yang F [60]	no variance-covariance matrix	program	partially, no cov	yes	10-fold
Yang Q [44]	only coefficients	no	no	no	5-fold
Wong [67]	only coefficients	calculation equation	partially, have 95% CI, no cov	no	10-fold
Wijnen [56]	no variance-covariance matrix, no error	no	partially, no cov	no	split sample by 50%
Wee [46]	no	no	no	no	split validation

Chen [39]	no variance-covariance matrix, no error	no	no	no	split validation
Peak [68]	no variance-covariance matrix, no error	calculation equation	no	no	follow-up
Patton [40]	no variance-covariance matrix, no error	calculation equation, R code	partially, no cov	no	bootstrap sample
Moore [41]	no variance-covariance matrix, no error	calculation equation	partially, no cov	no	split validation (2:1)
Mitchell [30]	only coefficients	no	no	no	no validation
Meregaglia [48]	no variance-covariance matrix, no error	no	partially, no cov	no	split geographically
Lee [27]	only coefficients	conversion table, calculation equation	no	no	follow-up
Lamu [36]	no variance-covariance matrix, no error	no	partially, no cov	no	5-fold
Lamu & Olsen [42]	no variance-covariance matrix, no error	no	partially, no cov	no	random-split
Khan [69]	no variance-covariance matrix, no error	calculation process	partially, no cov	no	random-split
Abdin [58]	only coefficients	calculation equation	no	no	no validation
Chen [37]	no variance-covariance matrix, no error	no	partially, no cov	no	hold-out
Cheung [70]	only coefficients	no	no	no	follow-up
Collado-Mateo [71]	only coefficients	calculation equation	no	no	5-fold
Coon [72]	only coefficients	calculation equation	no	no	no validation
Gamst-Klaussen [29]	no error	yes	yes	no	no validation
Gamst-Klaussen [38]	no variance-covariance matrix, no error	yes	partially, no cov	no	cross-validation
Gray [73]	no variance-covariance matrix	no	partially, no cov	yes	no validation
Hernandez [74]	no variance-covariance matrix, no error	no	no	no	no validation

Kaambwa [26]	no	no	no	no	split-out, k-fold
Kaambwa [75]	no variance-covariance matrix, no error	calculation equation	no	no	split-out, k-fold
Kaambwa Ratcliffe [49]	no variance-covariance matrix, no error	calculation equation	no	no	external validation

Appendix 3.1 The normality and heterogeneity check of OLS estimation



Illustrations: The residual plot does not fit the normal distribution, meaning the normality assumption is very likely to be rejected.

```
. estat hettest, iid

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of EQ5D

         chi2(1)      =    27.03
         Prob > chi2  =   0.0000
```

Illustrations: The null hypothesis is residuals have constant variances, and $p<0.001$ means we have enough evidence to reject the homoscedasticity assumption.

Appendix 3.2 Exploring the time effect in linear models

```
. xtreg EQ5D duramo2 - duramo14, fe r

Fixed-effects (within) regression               Number of obs     =       506
Group variable: id                              Number of groups  =       131

R-sq:                                           Obs per group:
     within  = 0.0224                                         min =         1
     between = 0.0009                                         avg =       3.9
     overall = 0.0063                                         max =        10

                                                F(13,130)         =      1.05
corr(u_i, Xb)  = 0.0013                         Prob > F          =    0.4091

                                    (Std. Err. adjusted for 131 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       EQ5D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    duramo2  |   -.003627   .0339406    -0.11   0.915    -.0707744    .0635204
    duramo3  |  -.0139777   .0228284    -0.61   0.541    -.0591409    .0311855
    duramo4  |   .0118453   .0293201     0.40   0.687     -.046161    .0698516
    duramo5  |  -.0064762   .0251482    -0.26   0.797    -.0562289    .0432765
    duramo6  |  -.0038825   .0318924    -0.12   0.903    -.0669779    .0592129
    duramo7  |  -.0187555   .0275605    -0.68   0.497    -.0732807    .0357696
    duramo8  |  -.0031148   .0487812    -0.06   0.949    -.0996227     .093393
    duramo9  |  -.0001744   .0312598    -0.01   0.996    -.0620181    .0616694
   duramo10  |   .0257853   .0328033     0.79   0.433    -.0391122    .0906828
   duramo11  |  -.0484552    .027074    -1.79   0.076     -.102018    .0051075
   duramo12  |   .0330132   .0357449     0.92   0.357    -.0377038    .1037301
   duramo13  |  -.0234846   .0250595    -0.94   0.350    -.0730618    .0260926
   duramo14  |   .0022456   .0591117     0.04   0.970    -.1146999    .1191911
      _cons  |   .7449027   .0191406    38.92   0.000     .7070353      .78277
-------------+----------------------------------------------------------------
    sigma_u  |  .22107041
    sigma_e  |  .11487752
        rho  |  .78738418   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. test duramo2 duramo3 duramo4 duramo5 duramo6 duramo7 duramo8 duramo9 duramo10 duramo11 duramo12 duramo13 duramo14

 ( 1)  duramo2 = 0
 ( 2)  duramo3 = 0
 ( 3)  duramo4 = 0
 ( 4)  duramo5 = 0
 ( 5)  duramo6 = 0
 ( 6)  duramo7 = 0
 ( 7)  duramo8 = 0
 ( 8)  duramo9 = 0
 ( 9)  duramo10 = 0
 (10)  duramo11 = 0
 (11)  duramo12 = 0
 (12)  duramo13 = 0
 (13)  duramo14 = 0

       F( 13,   130) =    1.05
            Prob > F =    0.4091
```

The above Stata commands ran a crude regression by only including duration month as the independent variable. The joint F test shows that we do not have enough evidence to reject the null hypothesis that there is no significant time effect in the model.

Appendix 3.3 Exploring the individual effect in linear models

```
. xtreg EQ5D pain tired drowsy nausea appetite sob depression anxiety wellbeing itchiness sleep rls, fe

Fixed-effects (within) regression               Number of obs     =       506
Group variable: id                              Number of groups  =       131

R-sq:                                           Obs per group:
     within  = 0.2326                                         min =         1
     between = 0.3697                                         avg =       3.9
     overall = 0.4632                                         max =        10

                                                F(12,363)         =      9.17
corr(u_i, Xb)  = 0.3874                         Prob > F          =    0.0000

------------------------------------------------------------------------------
         EQ5D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
         pain |  -.0127493    .002963    -4.30   0.000    -.0185761   -.0069225
        tired |  -.0060701   .0035997    -1.69   0.093     -.013149    .0010088
       drowsy |   -.003434   .0032134    -1.07   0.286    -.0097532    .0028852
       nausea |   .0053276   .0036081     1.48   0.141    -.0017677    .0124229
     appetite |  -.0030336   .0029418    -1.03   0.303    -.0088187    .0027516
          sob |  -.0045513    .003732    -1.22   0.223    -.0118903    .0027876
   depression |  -.0126642   .0037737    -3.36   0.001    -.0200853   -.0052432
      anxiety |   -.011684   .0034034    -3.43   0.001    -.0183768   -.0049912
    wellbeing |  -.0017995   .0033933    -0.53   0.596    -.0084725    .0048735
    itchiness |   .0000538   .0029767     0.02   0.986    -.0058001    .0059076
        sleep |  -.0019713   .0027037    -0.73   0.466    -.0072882    .0033456
          rls |   .0075385   .0027341     2.76   0.006     .0021618    .0129152
        _cons |   .8474508   .0169078    50.12   0.000     .8142012    .8807004
--------------+---------------------------------------------------------------
      sigma_u |   .1808193
      sigma_e |  .10164259
          rho |    .759889   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(130, 363) = 6.78                     Prob > F = 0.0000

. xtreg EQ5D pain tired drowsy nausea appetite sob depression anxiety wellbeing itchiness sleep rls, mle nolog

Random-effects ML regression                    Number of obs     =       508
Group variable: id                              Number of groups  =       132

Random effects u_i ~ Gaussian                   Obs per group:
                                                             min =         1
                                                             avg =       3.8
                                                             max =        10

                                                LR chi2(12)       =    166.46
Log likelihood  =  302.57171                    Prob > chi2       =    0.0000

------------------------------------------------------------------------------
         EQ5D |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
         pain |  -.0155317   .0027885    -5.57   0.000     -.020997   -.0100663
        tired |  -.0072457   .0034064    -2.13   0.033    -.0139221   -.0005692
       drowsy |  -.0039036   .0030998    -1.26   0.208    -.0099791    .0021718
       nausea |   .0054908   .0032802     1.67   0.094    -.0009383    .0119198
     appetite |  -.0023206   .0027818    -0.83   0.404    -.0077729    .0031316
          sob |  -.0055948   .0033732    -1.66   0.097     -.012206    .0010165
   depression |  -.0125173   .0034691    -3.61   0.000    -.0193166   -.0057179
      anxiety |  -.0133771   .0032193    -4.16   0.000    -.0196869   -.0070673
    wellbeing |  -.0075918   .0032758    -2.32   0.020    -.0140121   -.0011714
    itchiness |   .0015673     .00272     0.58   0.564    -.0037637    .0068984
        sleep |  -.0003488   .0025316    -0.14   0.890    -.0053105     .004613
          rls |   .0066146   .0025546     2.59   0.010     .0016077    .0116216
        _cons |    .869042   .0202632    42.89   0.000     .8293269    .9087571
--------------+---------------------------------------------------------------
    /sigma_u |   .1524488    .012063                      .130548    .1780237
    /sigma_e |   .1014631   .0038146                     .0942553     .109222
          rho |   .6930182   .0397244                     .6113289       .7661
------------------------------------------------------------------------------
LR test of sigma_u=0: chibar2(01) = 180.31             Prob >= chibar2 = 0.000
```

The first regression is a fixed effect model using all ESAS-r: Renal items as independent variables. The last line of the Stata output is an F-test, and the underlying null hypothesis is that

there is no individual effect (i.e. all individual effects, u_i, equal to 0). In this situation, we have F-statistics<0.001, which means we have enough evidence to reject the null hypothesis. Therefore, the fixed effect model is better than the pooled regression.

The second regression is a random effect model using all ESAS-r: Renal items as independent variables. The last line of the Stata output is an F-test, and the underlying null hypothesis is that there is no random effect (i.e. all within-subject variances equal to 0). In this situation, we have F-statistics<0.001, which means we have enough evidence to reject the null hypothesis. Therefore, the random effect model is better than the pooled regression.

```
. quietly xtreg EQ5D pain tired drowsy nausea appetite sob depression anxiety wellbeing itchiness sleep rls, fe

. estimate store fe

. quietly xtreg EQ5D pain tired drowsy nausea appetite sob depression anxiety wellbeing itchiness sleep rls, re

. estimate store re

. hausman fe re
```

	Coefficients			
	(b) fe	(B) re	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
pain	-.0127493	-.0157152	.002966	.0009293
tired	-.0060701	-.0073061	.001236	.0009553
drowsy	-.003434	-.0039367	.0005026	.0005643
nausea	.0053276	.0054925	-.000165	.0013777
appetite	-.0030336	-.002267	-.0007666	.0007916
sob	-.0045513	-.0056693	.001118	.0014831
depression	-.0126642	-.0124904	-.0001738	.0013343
anxiety	-.011684	-.0134722	.0017882	.0009247
wellbeing	-.0017995	-.0079491	.0061496	.0009633
itchiness	.0000538	.001673	-.0016193	.0011207
sleep	-.0019713	-.0002392	-.0017321	.0008448
rls	.0075385	.0065644	.0009741	.0008446

```
                b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                chi2(12) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =       57.93
              Prob>chi2 =      0.0000
```

The null hypothesis for the Hausman test is that the random effect model is more consistent compared with the fixed effect model. The chi-square statistics is less than 0.001, meaning we have enough evidence to reject the null hypothesis, and the fixed effect model is better.

The Hausman test could not be done if the model is estimated by robust variances. The above Hausman test was conducted on a fixed model and random model without robust variances. But in our data, we did observe heterogeneity among fixed effect models. Below test 1 is a test examining heteroskedasticity. The chi-statistics is less than 0.001, meaning there is heteroskedasticity in the residuals. Besides, autocorrelation also existed. Below test 2 is a test

examining autocorrelation in fixed and random effect models. We get F-statistics = 0.025, meaning there is autocorrelation in our data. Therefore, the cluster-robust variance should be used for the fixed model and random effect model. When we compare the standard error estimated with or without robust estimation, they are quite similar both in the fixed effect model and the random effect model. Therefore, the results from the Hausman test are still applicable.

```
. **Test 1: heteroscedasticity

. quietly xtreg EQ5D pain tired drowsy nausea appetite sob depression anxiety wellbeing itchiness sleep rls, fe

. xttest3

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (131)  =   1.8e+35
Prob>chi2 =      0.0000


. **Test 2: auto-correlation

. xtserial pain tired drowsy nausea appetite sob depression anxiety wellbeing itchiness sleep rls

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F(  1,       67) =      5.255
           Prob > F =      0.0250
```

EQ5D	Coef.	Std. Err.	t	P>|t|
pain	-.0127493	.002963	-4.30	0.000
tired	-.0060701	.0035997	-1.69	0.093
drowsy	-.003434	.0032134	-1.07	0.286
nausea	.0053276	.0036081	1.48	0.141
appetite	-.0030336	.0029418	-1.03	0.303
sob	-.0045513	.003732	-1.22	0.223
depression	-.0126642	.0037737	-3.36	0.001
anxiety	-.011684	.0034034	-3.43	0.001
wellbeing	-.0017995	.0033933	-0.53	0.596
itchiness	.0000538	.0029767	0.02	0.986
sleep	-.0019713	.0027037	-0.73	0.466
rls	.0075385	.0027341	2.76	0.006
_cons	.8474508	.0169078	50.12	0.000
sigma_u	.1808193			
sigma_e	.10164259			
rho	.759889	(fraction of variance due t		

F test that all u_i=0: F(130, 363) = 6.78

EQ5D	Coef.	Robust Std. Err.	t	P>|t|
pain	-.0127493	.0034513	-3.69	0.000
tired	-.0060701	.0039874	-1.52	0.130
drowsy	-.003434	.003686	-0.93	0.353
nausea	.0053276	.0049193	1.08	0.281
appetite	-.0030336	.0032008	-0.95	0.345
sob	-.0045513	.004519	-1.01	0.316
depression	-.0126642	.0045629	-2.78	0.006
anxiety	-.011684	.0038378	-3.04	0.003
wellbeing	-.0017995	.0036323	-0.50	0.621
itchiness	.0000538	.0030467	0.02	0.986
sleep	-.0019713	.0026865	-0.73	0.464
rls	.0075385	.0030842	2.44	0.016
_cons	.8474508	.0234895	36.08	0.000
sigma_u	.1808193			
sigma_e	.10164259			
rho	.759889	(fraction of variance due t		

Left: fixed effect model without robust estimation; Right: fixed effect model with robust estimation

EQ5D	Coef.	Std. Err.	z	P>|z|
pain	-.0157152	.0028135	-5.59	0.000
tired	-.0073061	.0034706	-2.11	0.035
drowsy	-.0039367	.0031635	-1.24	0.213
nausea	.0054925	.0033347	1.65	0.100
appetite	-.002267	.0028333	-0.80	0.424
sob	-.0056693	.0034246	-1.66	0.098
depression	-.0124904	.0035299	-3.54	0.000
anxiety	-.0134722	.0032754	-4.11	0.000
wellbeing	-.0079491	.0032537	-2.44	0.015
itchiness	.001673	.0027577	0.61	0.544
sleep	-.0002392	.0025683	-0.09	0.926
rls	.0065644	.0026004	2.52	0.012
_cons	.8708915	.0198499	43.87	0.000
sigma_u	.14586064			
sigma_e	.10150288			
rho	.67373546	(fraction of variance due t		

EQ5D	Coef.	Robust Std. Err.	z	P>|z|
pain	-.0157152	.0031436	-5.00	0.000
tired	-.0073061	.0037157	-1.97	0.049
drowsy	-.0039367	.0035071	-1.12	0.262
nausea	.0054925	.0042608	1.29	0.197
appetite	-.002267	.0030704	-0.74	0.460
sob	-.0056693	.0041086	-1.38	0.168
depression	-.0124904	.0041621	-3.00	0.003
anxiety	-.0134722	.0041008	-3.29	0.001
wellbeing	-.0079491	.0038123	-2.09	0.037
itchiness	.001673	.0028705	0.58	0.560
sleep	-.0002392	.0025809	-0.09	0.926
rls	.0065644	.0028337	2.32	0.021
_cons	.8708915	.021679	40.17	0.000
sigma_u	.14586064			
sigma_e	.10150288			
rho	.67373546	(fraction of variance due t		

Left: random effect model without robust estimation; Right: random effect model with robust estimation

Appendix 3.4 The AIC and BIC statistics for different generalized linear models (GLMs)

```
. estimate stats glm1 glm2 glm3 glm4 glm5

Akaike's information criterion and Bayesian information criterion
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
glm1	506	.	208.7938	13	-391.5875	-336.6425
glm2	506	.	236.8145	15	-443.6289	-380.2309
glm3	506	.	204.2375	9	-390.4751	-352.4363
glm4	506	.	225.0224	6	-438.0448	-412.6856
glm5	506	.	438.929	6	-865.8581	-840.4988

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

Note: generalized linear model (glm)

glm1- Normal + log link, model specification 1 (full ESAS-r: Renal items)

glm2 - Normal + log link, model specification 2 (full ESAS-r: Renal items + age + female)

glm3 - Normal + log link, model specification 3 (selected ESAS-r: Renal items)

glm4 - Normal + log link, model specification 4 (selected ESAS-r: Renal items + selected demographics)

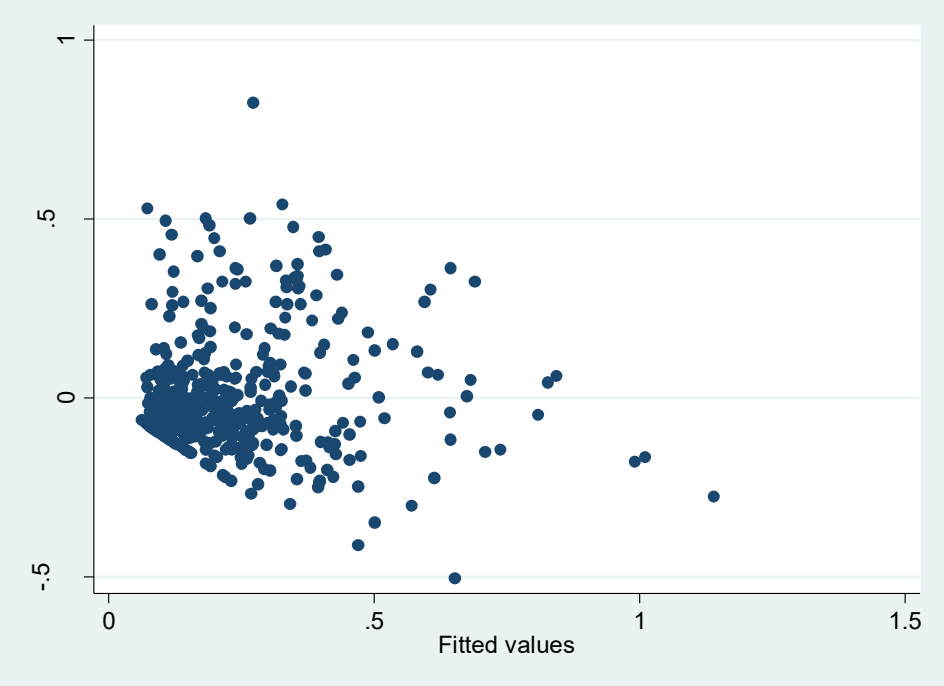
glm5 – Gamma + identity link, model specification 3 & 4 (selected ESAS-r: Renal items/selected ESAS-r: Renal items + selected demographics)

The likelihood functions for model specification 1 and 2 were not concave in gamma + identity link. So, we used forward step-wise to build the model corresponding to model specification 3 and 4.

The AIC and BIC showed that glm5 is better than glm3 and glm4. So, for glm models with model specifications 3 and 4, we used gamma + identity link. Then for glm models with specification 1 and 2, we did not have other options, so we still used normal + log link.

Appendix 3.5 Model assumptions of GLMs

The normal + log link GLM still has the equal variance assumption. We used the residual plot to see if this assumption was valid.



The plot shows that equal variance may not be true as the residuals were more diverse among the large estimated values.

Appendix 3.6 The assumption check for Tobit model

Normality check:

```
. tobcm,p

Conditional moment test against the null of normal errors

                  critical values
    CM        %10       %5        %1
  34.358   7.22987  11.291115 18.698471
```

Illustrations: The conditional moment (CM) is 34.36, which is much greater than the 5% percent level CM, which is 11.29. Therefore, we reject the null hypothesis that the residuals were normally distributed.

Homescadesticity check:

```
. global xlist pain tired drowsy nausea appetite sob depression anxiety wellbeing itchiness sleep rls

. generate y =disutility

. generate dy = disutility > 0

. generate lny = ln(y)
(121 missing values generated)

. quietly summarize lny

. scalar gamma = r(min)

. display "gamma = " gamma
gamma = -3.9373429

. replace lny = gamma - 0.0000001 if lny == .
(121 real changes made)

. quietly tobit lny $xlist, ll

. predict xb1,xb

. matrix btobit = e(b)

. scalar sigma = btobit[1,e(df_m)+2]

. generate threshold = (gamma-xb)/sigma

. generate lambda = normalden(threshold)/normal(threshold)

. quietly generate uifdyeq1 = (lny-xb)/sigma if dy == 1

. quietly generate double gres1 = uifdyeq1

. quietly replace gres1 = -lambda if dy == 0

. quietly generate double gres2 = uifdyeq1^2 -1

. quietly replace gres2 = -threshold*lambda if dy == 0

. quietly generate double gres3 = uifdyeq1^3

. replace gres3 = -(2 + threshold^2)*lambda if dy == 0
(121 real changes made)

. quietly generate double gres4 = uifdyeq1^4 -3

. quietly replace gres4 = -(3*threshold + threshold^3)*lambda if dy == 0

. foreach var in $xlist{
  2. generate score2`var' = gres2*`var'
  3. }

. global scores2 score* score2* gres1 gres2

. generate one = 1

. quietly regress one gres3 gres4 $scores2, noconstant

. display "N R^2 = " e(N)*e(r2) " with p-value = " chi2tail(2,e(N)*e(r2))
N R^2 = 747.48652 with p-value = 4.85e-163
```

Illustrations: the chi-square statistics < 0.001, meaning that there is enough evidence to reject the equal variance assumption.

Appendix 3.7 The performance statistics for mixture models

Adjusted limited dependent variable (aldv) mixture models: The one-component models were better than the two-component models.

(Note: The estimation function for the two-component model for model specification 2 (whole ESAS items + demographics was not concave)

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
C1_aldv1	506	.	211.0822	14	-394.1644	-334.9928
C2_aldv1	506	.	211.0822	15	-392.1644	-328.7663
C1_aldv2	506	.	223.4596	16	-414.9193	-347.2947
C1_aldv3	506	.	205.3559	9	-392.7117	-354.6729
C2_aldv3	506	.	205.3559	10	-390.7117	-348.4464
C1_aldv4	506	.	218.2827	10	-416.5654	-374.3001
C2_aldv4	506	.	218.2827	11	-414.5654	-368.0735

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

Mixture beta regression models: The two-component models were better than the one-component models.

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
C1_beta1	506	.	1386.495	14	-2744.989	-2685.818
C2_beta1	506	.	1528.728	29	-2999.456	-2876.887
C1_beta2	506	.	1394.676	16	-2757.352	-2689.727
C2_beta2	506	.	1545.149	33	-3024.298	-2884.822
C1_beta3	506	.	1380.56	9	-2743.12	-2705.081
C2_beta3	506	.	1520.552	19	-3003.105	-2922.8
C1_beta4	506	.	1388.749	11	-2755.497	-2709.005
C2_beta4	506	.	1491.801	23	-2937.602	-2840.392

Note: N=Obs used in calculating BIC; see **[R] BIC note**.

Appendix 3.8 The selected ESAS-r: Renal profile for plotting Figure 3.5

pain	tired	drowsy	sob	nausea	appetite	depression	anxiety	wellbeing	itchiness	sleep	restless legs
0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9
10	10	10	9	9	9	9	9	9	9	9	9
10	10	10	10	10	10	9	9	9	9	9	9
10	10	10	10	10	10	10	10	10	9	9	9
10	10	10	10	10	10	10	10	10	10	10	10

Appendix 3.9 Uncertainty information of the final models

The generalized linear model (gamma distribution, identity link):

Variance-covariance matrix

Covariance matrix of coefficients of **glm** model

e(V)	disutility pain	drowsy	depression	anxiety	wellbeing	_cons
disutility						
pain	.00001917					
drowsy	-8.185e-07	.00002025				
depression	8.763e-07	-2.707e-06	.00002386			
anxiety	-1.640e-06	-3.852e-06	-7.984e-07	.00004974		
wellbeing	-8.868e-06	-2.173e-06	-.0000163	-.00001514	.00003818	
_cons	-.00001462	-.00001081	8.801e-06	6.736e-06	-.00002723	.00013597

Deviance: 266.23

Dispersion parameter of the gamma distribution: 0.5324

The generalized estimating equations:

Variance-covariance matrix

Covariance matrix of coefficients of **xtgee** model

e(V)	pain	tired	sob	depression	anxiety	wellbeing	rls	_cons
pain	.00001031							
tired	1.069e-06	.00001147						
sob	-1.117e-06	-3.893e-06	.00001943					
depression	2.056e-06	-2.381e-06	.00001018	.00002216				
anxiety	-4.196e-06	-5.878e-07	-3.095e-06	-8.841e-06	.00002377			
wellbeing	-4.421e-06	-1.775e-06	-6.019e-06	-.00001106	6.464e-06	.0000208		
rls	1.999e-07	-2.317e-06	-3.446e-07	1.827e-06	-4.596e-06	-3.258e-07	7.747e-06	
_cons	-.00001597	-.00001249	-.00001137	-.00001334	-4.836e-07	-.00001259	-.00001171	.00029408

Deviance could not be estimated. The generalized estimating equations model is not based on likelihood function and it is hard to define the distribution of the residuals.

Appendix 3.10 Sensitivity analyses

```
. quietly xtgee EQ5D female age diabetes pain tired drowsy nausea appetite sob depression anxiety wellbeing itchiness sleep
> rls Aboriginal Asian Black Caucasian Indiansubcontinent Latin Mideast Multi Pacific if validation == 0, vce(robust)

. test diabetes

 ( 1)  diabetes = 0

           chi2(  1) =    5.60
         Prob > chi2 =    0.0179

. test Asian Black Caucasian Indiansubcontinent Latin Mideast Multi Pacific

 ( 1)  Asian = 0
 ( 2)  Black = 0
 ( 3)  Caucasian = 0
 ( 4)  Indiansubcontinent = 0
 ( 5)  Latin = 0
 ( 6)  Mideast = 0
 ( 7)  Multi = 0
 ( 8)  Pacific = 0

           chi2(  8) =  406.71
         Prob > chi2 =    0.0000
```

This is a generalized estimating equations (GEE) model on the whole ESAS-r: Renal items and demographics. The variable *diabetes* and the variable *race* were significant (p=0.02, p<0.001 respectively). As they were significant in this overall equation, during the backward step-wise procedure, they would still remain significant.

```
. quietly glm disutility age female diabetes pain tired drowsy nausea appetite sob depression anxiety wellbeing itchiness sl
> eep rls Asian Black Caucasian Indiansubcontinent Latin Mideast Multi Pacific if validation == 0, family(normal) link(log)
> vce(cluster id)

. test Asian Black Caucasian Indiansubcontinent Latin Mideast Multi Pacific

 ( 1)  [disutility]Asian = 0
 ( 2)  [disutility]Black = 0
 ( 3)  [disutility]Caucasian = 0
 ( 4)  [disutility]Indiansubcontinent = 0
 ( 5)  [disutility]Latin = 0
 ( 6)  [disutility]Mideast = 0
 ( 7)  [disutility]Multi = 0
 ( 8)  [disutility]Pacific = 0

           chi2(  8) =  121.44
         Prob > chi2 =    0.0000

. test diabetes

 ( 1)  [disutility]diabetes = 0

           chi2(  1) =    3.52
         Prob > chi2 =    0.0607
```

This is a generalized linear model (GLM) on the whole ESAS-r: Renal items and demographics. The variable *diabetes* was partially significant, and the variable *race* was significant (p=0.06, p<0.001 respectively). During the backward step-wise procedure, the variable *diabetes* would become more significant. Both two would remain for the final model.

Overall, for sensitivity analyses, we explored GEE and GLM with 2 model specifications. The first model specification includes the whole ESAS-r: Renal items and demographics, and the other model specification includes selected ESAS-r: Renal items and demographics. The GLM

used normal distribution with log link (the GLM with gamma distribution and identity link did not have a concave parameter estimating function). The table below summarized the mean absolute error (MAE), mean squared error (MSE), and percentage of estimated values out of the range (OFR %).

	MAE	MSE	OFR %
Generalized estimating equations (GEE): all ESAS-r: Renal + all demographics	0.108	0.021	22 (4.3%)
GEE: selected ESAS-r: Renal + selected demographics	0.108	0.021	23 (4.5%)
Generalized linear model (GLM) with normal distribution and log link: all ESAS-r: Renal + all demographics	0.101	0.019	3 (0.6%)
GLM with normal distribution and log link: selected ESAS-r: Renal + selected demographics	0.101	0.020	2 (0.4%)
Final model of the study: GLM with gamma distribution and identity link: selected ESAS-r: Renal items	0.116	0.028	0
Final model of the study: GEE on selected ESAS-r: Renal items	0.119	0.027	0

Note: MAE: mean absolute error; MSE: mean squared error; OFR %: the percentage of out-of-range estimated values.