Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments

Louise Longworth, BA (Hons), MSc, PhD1,*, Donna Rowen, BA (Hons), MSc, PhD2

1Health Economics Research Group, Brunel University, Uxbridge, Middlesex, UK; 2School of Health and Related Research, University of Sheffield, Sheffield, UK

Abstract

Quality-adjusted life-years (QALYs) are widely used as an outcome for the economic evaluation of health interventions. However, preference-based measures used to obtain health-related utility values to produce QALY estimates are not always included in key clinical studies. Furthermore, organizations responsible for reviewing or producing health technology assessments (HTAs) may have preferred instruments for obtaining utility estimates for QALY calculations. Where data using a preference-based measure or the preferred instrument have not been collected, it may be possible to “map” or “crosswalk” from other measures of health outcomes. The aims of this study were 1) to provide an overview of how mapping is currently used as reported in the published literature and in an HTA policymaking context, specifically at the National Institute for Health and Clinical Excellence in the United Kingdom, and 2) to comment on best current practice on the use of mapping for HTA more generally. The review of the National Institute for Health and Clinical Excellence guidance found that mapping has been used since first established but that reporting of the models used to map has been poor. Recommendations for mapping in HTA include an explicit consideration of the generalizability of the mapping function to the target sample, reporting of standard econometric and statistical tests including the degree of error in the mapping model across subsets of the range of utility values, and validation of the model(s). Mapping can provide a route for linking outcomes data collected in a trial or observational study to the specific preferred instrument for obtaining utility values. In most cases, however, it is still advantageous to directly collect data by using the preferred utility-based instrument and mapping should usually be viewed as a “second-best” solution.

Keywords: health-related quality of life, methodology, quality-adjusted life-years, utility assessment.

Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

The use of “mapping” techniques or “crosswalks” to link the outcomes from different measures of health or health-related quality of life (HRQoL) have become increasingly popular within health economics, particularly for the purpose of conducting cost-utility analyses of health technologies or other health interventions [1]. These methods have been particularly useful when the studies used to estimate the effectiveness of the interventions have not collected health-related utility data. In addition, where a health technology assessment (HTA) or reimbursement agency has recommended the use of a specific preference-based measure, the mapping approach can be used to link to health-related utility values for that instrument by using outcomes obtained from other measures.

This is the situation in the United Kingdom, where the National Institute for Health and Clinical Excellence (NICE) has stated that its economic evaluations should be presented by using an incremental cost per quality-adjusted life-year (QALY) framework [2]. It has also indicated that preferably the estimates of health-related utility used in QALY calculations should have been obtained by using the EuroQol five-dimension (EQ-5D) questionnaire. This recommendation has arisen from the evidence that estimates of health state utility differ according to the instrument and/or elicitation method used. Therefore, in the context of NICE, there is a need to map to, or predict, EQ-5D data, even where other measures of health-related utility have been collected.

There are various options available to the analyst when considering the use of mapping for the estimation of health state utilities for HTA. What measures to map to and which measures to map from? What form should the model underpinning the algorithm take? Which statistical methods should be used? How should results be tested, validated, and reported? This is evident from the considerable variation in the approaches reported in the growing body of literature on mapping studies for cost-utility analyses. There are also currently no clear guidelines on the best methods and standards available. Throughout the article, mapping to the EQ-5D is provided as an example; however, the issues and methods discussed can equally apply to other instruments for measuring health-related utility. The references to the EQ-5D reflect that this instrument is the most commonly mapped to within this area [1] and also reflects that the study was originally conceived to support NICE and its stakeholders when using mapping techniques [3].

The aims of this article were threefold: 1) to provide an overview of how mapping is currently used in an HTA

* Address correspondence to: Louise Longworth, Health Economics Research Group, Brunel University, Uxbridge, Middlesex UB8 3PH, UK.
E-mail: Louise.longworth@brunel.ac.uk.
1098-3015/$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jval.2012.10.010
policy-making context, specifically at NICE in the United Kingdom; 2) to provide an overview of the broader literature on mapping; and 3) to provide guidance on best practices where this is possible based on the findings of the two reviews.

Methods

What Is Meant by “Mapping”?

In mathematical terms, “mapping” refers to correspondence by which each element of a given set has associated with it one or more elements of a second set [4]. In the context presented here, we define “mapping” as the development and use of a model or algorithm to predict health-related utility values by using data from other measures of health outcomes. While the term “mapping” is used in many other health-related contexts, we are primarily interested in its use to obtain EQ-SD values. The focus is on empirical mapping as this is consistent with NICE recommendations [2]. Other forms of mapping based on parametric approaches or qualitative analyses are not considered further. Furthermore, we distinguish “mapping” studies from studies directly eliciting health-related utility values by asking participants to directly value their own health state, and from studies in which a description of a hypothetical health state generated from a vignette or other instrument is valued. In the context of NICE in the United Kingdom, this latter approach of valuing health states described by other instruments (specifically validated patient-reported outcome measures) is recognized as a possibility when the EQ-5D has been demonstrated to be inappropriate for the particular condition of interest [5].

In this context of mapping, the data used to predict EQ-SD values could consist of condition-specific quality-of-life measures (such as the asthma quality-of-life questionnaire [6]), other generic quality-of-life questionnaires (such as the short form 36 health survey (SF-36) [7]), clinical indicators of disease severity (such as the Canadian Cardiovascular Society Score for angina [8] or the Psoriasis Area and Severity Index [9]), sociodemographic variables, or a combination of these. Data on the “target” preference-based measure (in this case, the EQ-5D) and the “source” predictive measure(s) (the measures of health outcomes that will be used to map to the preference-based measure) must be collected within a separate estimation sample. From these sample data, models can be developed to estimate the relationship between the target measure and other measures of health outcomes. Where the target measure is a multiattribute instrument, the data can be mapped to either the EQ-5D utility value or the EQ-5D dimension responses. A set of predicted EQ-5D values can then be obtained by applying the statistical model to data contained in the study of effectiveness. The predicted EQ-5D values can then be analyzed using standard methods for trial-based analyses or summarized for each health state within an economic model. For example, to map SF-36 to the EQ-5D, an estimation sample is used to estimate models regressing the EQ-5D on SF-36 variables. The model is used as a prediction equation to predict the EQ-5D in the clinical trial or other study of interest that contains SF-36 data, and these predicted values can then be used in an economic model. Alternatively, if a published function mapping SF-36 to the EQ-5D exists, this may be used rather than estimating the mapping function by using an estimation sample.

In its simplest form, mapping can be considered equivalent to taking the mean value for a given health state. For example, consider the case of a condition categorized into two health states: stable disease and progressive disease. If EQ-5D data and the health state category were collected for a sample of patients, we could estimate the mean utility value for patients at each of the two stages of disease. These mean values could then be assigned to patients in a trial in which the stage of disease is recorded. However, simply using a mean value (and distribution where reported) for a similar broadly defined health state from another data set or the reported literature can mask the variation between patients.

Whether the mapping approach will offer an advantage over simply using mean values from an external data set will, in part, depend on the structure of the economic model being used to reflect the decision problem. If the model has a simple structure with few health states, then reliable estimates of the mean (and variance) of the EQ-5D values associated with those health states may suffice. Where there are multiple predictors of health status that can be measured and reflected in the decision model, however, then the mapping approach can predict the EQ-5D value more accurately. For example, if the health states in a model are defined according to a 20-point measure of disease severity, it may not be possible to obtain values for each of these 20 levels of severity from a sufficient number of patients. However—providing there is a predictable relationship between the EQ-5D and the severity measure—the relationship between the measures can be estimated on the basis of all the data to provide estimates for each of the 20 health states. Mapping also enables the EQ-5D data to be linked directly back to data collected within the clinical trial(s) used to inform the estimates of cost-effectiveness.

Overview of Mapping from the Health Economics Literature

An overview of the methods used for mapping to estimate health-related utility values is provided. It draws on a previously conducted review of published studies reporting empirical results of mapping exercises [1]. The overview considers the following elements of mapping in turn:

1. Defining the “estimation” data set
2. Model specification
3. Model type
4. Assessing performance
5. Application

Each of these stages will be discussed in turn below with reference to the published literature on mapping. Although specific reference is given to the EQ-5D, the approaches described below could apply to other preference-based measures of HRQOL. Identification of best practice is stated where possible.

Review of the Use of Mapping at NICE

The use of mapping in a policy-making context will be presented with specific reference to NICE in the United Kingdom. The 2008 NICE Guide to the Methods of Technology Appraisal recommends that HRQOL data be collected directly from patients by using the EQ-5D and that consideration be given to “mapping” or “cross-walking” from other HRQOL measures to the EQ-5D when such data are not available [2]. Although the 2008 edition was the first of the NICE methods guides to suggest mapping as a potential solution for an absence of health-related utility data, previous methods guides recommend self-assessment of health status by patients [10] and data from validated generic preference-based measures [11] but did not offer guidance to the analyst on how to conduct cost-utility analyses if such data had not been collected within clinical studies.

Two published reviews of NICE Technology Appraisals are available that report on the methods used to obtain the health-related utility values included in its HTAs [12,13]. These reviews report the methods of eliciting the values (e.g., time trade-off and standard gamble), the instruments used to collect information on changing health status (e.g., the EQ-5D questionnaire), and
Results

An Overview of Mapping

Mapping involves three key stages. First, a separate “estimation” data set is required that contains the data that you are mapping from, the “source,” and the “target” preference-based measure. Second, regression methods are used to “map” this data onto either the index score or the classification system of the target measure. Third, the regression results are applied to the trialse(s) or observational study data set to estimate preference-based scores for the target measure at either the mean or observational level. Ideally, a validation stage should also be applied, whereby the regression results are validated against another data set.

The Estimation Sample

The generation of the mapping function involves the estimation of the statistical relationship between the target measure and the predictive measure(s) by using an estimation data set. The first step in the mapping approach is to obtain the estimation sample. As this assumes that the statistical relationship is the same across the estimation and trial data sets, the choice of the “estimation sample” is crucially important.

The estimation sample is the group of people, usually patients, who will complete the preferred instrument (in this example, the EQ-5D) to report their own health and from whom data on the “source” measures will also be obtained. To be confident about the generalizability of the mapping function to the target sample, the clinical and demographic characteristics of people in the estimation sample should be as similar as possible to the characteristics of the sample to which the mapping algorithm will be applied. All covariates used in the mapping function should be overlapping in distribution for the estimation and target samples. It is recommended that all variables included within the target source (e.g., the main clinical trial(s) used to inform clinical effectiveness within the economic model) that are thought likely to have an impact on target EQ-5D values should be included in the estimation sample. If no existing data set is available that includes both the source and target measures, it may also be possible to use an estimation sample to map from the source to an intermediary measure and a second estimation data set to map from the intermediary measure onto the target measure. These estimates should be treated with caution as the process generates intermediary measure estimates that are then used to generate estimates, meaning that mapping twice is likely to increase the error and uncertainty around the predicted EQ-5D values. If this approach is taken, the uncertainty should be fully accounted for within the economic analysis.

Model Specification

The model specification can take a number of forms depending on which best suits the data and the decision problem at hand. If the target preference-based measure is a multiattribute utility measure (i.e., has multiple dimensions such as the EQ-5D), the dependent variable could either be the utility index value or the responses to the dimensions of health described by the instrument. The explanatory variables should be those that best predict the preference-based values for health states included in the economic analysis. If the mapping function is intended to have wider use by other researchers, consideration will need to be given to whether the explanatory variables are routinely collected in other studies. The model form will depend on the data: additive models are currently most commonly used; however, alternative model specifications have been used in the literature.

A recent review of mapping studies included a systematic search of the literature supplemented by unpublished studies (identified by contacting researchers) in early 2007 and reports on 30 studies covering 119 mapping models [1]. The review of the literature found that the most common target measure was the EQ-5D (15 studies). The most commonly used source measures in the literature were SF-36 (7 studies) and short-form 12 (SF-12) (6 studies). The most common model specification involved the use of a preference-based index (usually for the EQ-5D) as the dependent variable and dimension or item scores (usually from the SF instruments) as independent variables. Articles also examined model specifications including squared terms and interaction terms to explore possible nonlinear relationships between the target and source measures. The review found that these had little impact, but it is likely that this differs by source and target measures, patient group, and patient severity. The review found that the inclusion of nonhealth variables such as sociodemographics made some improvement in the accuracy of the mapping function. Table 1 provides a summary of different model specifications for the mapping function.

The recent review of mapping studies found that explanatory power using $R^2$ was often low for models that involved mapping a condition-specific measure onto a generic preference-based measure and errors were often larger than for models mapping a generic measure onto a generic preference-based measure [1]. This may occur because of limited conceptual overlap as important dimensions in the condition-specific measure may not appear in the generic measure and vice versa.

The estimation of the mapping regression relies on statistical dependence between the target measure (e.g., the EQ-5D) and the source measures, and the avoidance of omitted variables. If the source measures have little conceptual overlap with the dimensions of the EQ-5D, the regression model may suffer from omitted...
variable bias and have poor explanatory power and large prediction errors. This can undermine the model, and the uncertainty around the predicted values may be substantial. Where the EQ-5D is shown to not adequately capture the impact of the condition or treatment, it may be necessary to consider using an alternative approach to utility estimation.

The selection of explanatory variables should be based on a combination of judgment based on prior knowledge of the clinical relationships between variables, and standard statistical and econometric techniques. Consideration should be given to the variables that are expected to have an impact on the EQ-5D values of people with the condition of interest. This can be based on patient and clinical opinion obtained directly or reported in the literature. Decision rules for the inclusion of variables could be specified a priori, such as levels of statistical significance and the signs of the coefficients matching prior stated beliefs. Correlation should be used to examine the relationship between source and target measures, and if there is poor correlation, this indicates that the mapping function will perform poorly (see [21] for an example of this). Akaike’s information criterion ([22]) and the Bayesian information criterion ([23]) can be used to inform the choice of model specification (see, e.g., [24].) Other tests should also be used to enable the researcher to define a robust model, such as examining the extent to which the model suffers from misspecification (e.g., Ramsey RESET test [25]), omitted variables and heteroscedasticity (e.g., the Park test [26]), or nonnormality in the errors (e.g., the Jarque-Bera test [27]; see Brazier et al. [28] for an explanation of its usage in panel data).

The severity of the condition reflected in the source measure should also be captured by the target measure. If the source

---

Table 1 – Overview of mapping process and recommendations.

| Estimation sample | Clinical and demographic characteristics in the estimation sample should be similar to the characteristics of the “target” sample to which the mapping algorithm will be applied. Covariates used in the mapping function should be overlapping in distribution for the estimation and target samples. Variables included within the target sample thought likely to have an impact on EQ-5D values should be included in the estimation sample. If no data set is available that includes the explanatory variables and the EQ-5D data, data collection of EQ-5D data will be required to estimate the mapping regression. |
| Dependent variables | EQ-5D index, EQ-5D dimension levels |
| Independent variables | Condition-specific measure: overall score, summary scores, item-level scores, item-level dummies, interaction terms, squared terms, cubic terms. Generic measures: overall score, summary scores, item-level scores, item-level dummies, interaction terms, squared terms, cubic terms. Clinical measures: overall score, summary score, categorical dummies. Sociodemographic variables. Other relevant health data. |
| Model selection and specification | Use prior knowledge of clinical relationships. Use standard statistical techniques to examine the data prior to mapping estimation (e.g., frequency tables and correlations). Fully describe the data set used to estimate the regression model including both range of EQ-5D and plots showing EQ-5D distribution. Fully describe the range of EQ-5D predicted values used in the cost-effectiveness model. |
| Performance | Goodness of fit: Statistical significance, sign and size of coefficients $R^2$ and adjusted $R^2$. Information criterion of AIC and BIC. Further tests of model fit such as Ramsey RESET test, Park test, Jarque-Bera test. Plots to examine whether model assumptions are valid. Predictive ability: Root mean squared error (RMSE) and mean squared error (MSE). RMSE, MSE, mean error, mean absolute error by subset of severity range of the EQ-5D and/or predictive measure(s). Plots of observed and predicted EQ-5D scores. |
| Validation | Application and assessment of mapping algorithm when applied to a validation sample. Validation sample can be a separate patient sample to the estimation data set or the data set used to estimate the mapping function can be randomly separated into estimation and validation samples. |
| Uncertainty | Uncertainty in health-related utility values should be incorporated into economic analyses. Multiple possible mapping functions can be used to produce utility values for sensitivity analyses. |

AIC, Akaike’s information criterion; BIC, Bayesian information criterion; EQ-5D, EuroQol five-dimensional.
measure cannot capture the same health problems captured by the levels of severity or impairment described by the target measure (e.g., the EQ-5D), it will not be able to accurately predict these levels. The extent to which this is a problem will depend on the severity range of the target and source measures and the severity range of the estimation and trial data set. The data set used to estimate the mapping regression should be fully described including both the range of predicted EQ-5D values and graphical plots showing the distribution of the predicted EQ-5D data. The range of predicted EQ-5D values used in the cost-effectiveness model should also be fully described to provide information of whether the predicted EQ-5D data have involved extrapolation.

Model Type

The appropriate model type differs depending on the data set and how it is applied. For multiattribute systems such as the EQ-5D, it also depends on whether the aim is to predict the index value or whether it is to predict responses to each of the dimensions of health described by the instrument. As specified earlier, the properties of the regression data set should be clearly outlined. These properties should then be used to inform model selection and a justification provided explaining why the selected regression model was chosen. This is further discussed below with particular reference to the commonly used EQ-5D.

Mapping to EQ-5D Index Values

The model type used to map source measures onto the EQ-5D should take into account the distribution of EQ-5D values in the estimation data set. The EQ-5D has been shown to exhibit ceiling effects, meaning that typically EQ-5D data sets have a substantial proportion of people reporting full health with an EQ-5D value at 1. Although the distribution of EQ-5D utility values varies by patient group and study, often a bimodal or trimodal distribution is observed, with one peak around full health, one peak for moderate states, and a further peak for more severe states. The recent mapping review [1] found that the most common estimation technique was ordinary least squares (OLS), yet linear regressions may not always accurately predict the EQ-5D distribution for high and low EQ-5D values [1,16].

Some of the standard model specifications have been shown to predict fewer values toward the extremes of the utility scale, even where they are evident in the observed source data. OLS has been criticized in particular as being inappropriate for regressions mapping onto the EQ-5D due to the bounded nature of EQ-5D values as by definition people cannot have an EQ-5D value higher than 1, which represents “full health.” In addition, the standard UK value set has a lowest possible value of −0.594. The OLS model does not restrict the range of values and therefore may lead to implausible predicted values outside the existing range of EQ-5D values. Researchers have explored alternative types of models to overcome the theoretical limitations of OLS models for the analysis of EQ-5D values, including tobit [16,29,30] and censored least absolute deviation (CLAD) [16,29–31]. The results of this research have been mixed, with some concluding that CLAD provides an improvement in model performance compared with OLS [16,29,30], others stating that the improvement of CLAD over OLS is small [18], and the review of mapping studies finding that the use of tobit and CLAD had little impact. Most of the models are based on mean values, apart from CLAD, which is a median model. The choice between the use of mean and median values requires normative judgments as well as statistical considerations. Health valuation for economic evaluation for decision making has been mainly based on mean models [20,32–34].

The choice and application of alternative models is an area of recent and ongoing research, and a large number of models have been recently explored in the mapping literature. This includes the use of models to address the EQ-5D distribution including a generalized linear model (GLM) [35], a latent class model [31], a two-part and two step model (TP/TS) or two-step model [31,35,36], and a random effects censored mixture model [24]. The first part of the TP/TS model uses a logit regression to estimate the probability that an individual (at the observational level) is in full health and the second part estimates EQ-5D utilities for individuals who are not in full health using OLS [31,35,36], a GLM [35], or a log-transformed EQ-5D index (TP/TS-L model) [31] (although note that these articles combine the two models differently to produce predicted utilities). One article addresses overprediction for severe health states by estimating separate regressions for mild and severe states and using cutoff points on the source measure to identify which model should be used to predict EQ-5D values at the observational level [37].

The results from this recent research have been mixed. The studies estimating these models found that TP/TS models and GLMs do not seem to offer an improvement on OLS in terms of performance. The possibility of negative utility values must be considered when using a TP/TS model or a GLM. The study using the TP/TS-L model reported that all values in its data set were positive [31]. The study including the GLM used gamma and log-normal distributions (where the dependent variable was expressed as a disutility) and found that the models using the gamma distribution did not converge [35]. This study also found that OLS had superior performance to both the GLM and the TP/TS model [35]. Another study found that OLS regression was more accurate than the CLAD model, the multinomial logit model, and the TP/TS model at estimating the group mean, yet the accuracy deteriorated in older and less healthy subgroups and for these the TP/TS model performed better [36]. The latent class model can handle data where there are more than two “classes” in the data, and so it is more flexible to deal with the trimodal distribution of the EQ-5D at estimating the group mean values. One study [31] found that the latent class model and the TP/TS-L model performed better than OLS, CLAD, and a TP/TS model using OLS in the second stage. An adjusted censored mixture model has been used to deal with the bimodal or trimodal distribution of EQ-5D values, and although high errors were observed, the authors concluded that the method offers a vast improvement in performance in comparison to OLS and tobit based on other selection criteria [24]. Further research using the latent class model, the TP/TS-L model, and random effects censored mixture model is encouraged, especially for smaller patient data sets as existing research has been conducted on relatively large data sets, which may not be typical for the data sets used to estimate mapping functions for NICE submissions [31,24].

Mapping to the EQ-5D Dimension Responses

Although EQ-5D utility value sets are usually treated as continuous, in practice they take a limited number of discrete values. An alternative approach is to map to the descriptive system of the measure, which enables the value set to be applied separately and therefore may better reflect the distribution of values that would have been obtained if collected directly. This approach has several benefits. It can facilitate the use of the mapping function in different countries or jurisdictions as value sets from other countries can be applied to predictions from the mapping exercise. This can reduce the burden on companies required to produce HTAs for multiple reimbursement agencies in different countries as mapping to EQ-5D utility values would require the estimation of separate mapping functions for each country-specific value set. In addition, it is likely to reduce the
variation between measures of having to produce alternative mapping functions for each specific country or jurisdiction. The most commonly used approach to mapping to the EQ-5D dimensions has been through the use of logistic regression. Some articles reported using a multinomial logit model to estimate separate mapping functions to predict the level of each dimension of the EQ-5D, and then applying the standard published value sets to obtain utilities [36,38]. Articles comparing this approach to other approaches, however, found that it did not offer an improvement [16,36].

Assessing Performance

Measures of explanatory power such as \( R^2 \) report how well the mapping function explains the variation in utilities in the estimation data set. Although this is a useful indicator of performance, it does not show whether the mapping function is equally appropriate across the entire range of utilities that can be produced by the target measure. If the aim of mapping is to estimate EQ-5D data when such data are unavailable from the primary source of effectiveness, the accuracy of predictions is a key aspect of performance. Mean absolute error (mean absolute difference between estimated and observed utilities) and root mean squared error both indicate the "error" in the estimates in the data set used to estimate the regression, and smaller errors are preferred. While these errors are not necessarily representative of the errors in the estimates when the results are applied in the separate data set, they can provide some indication of how large the errors are expected to be.

Some mapping studies have reported underprediction for EQ-5D values representing mild health states and overprediction for more severe states (see, e.g., [16,24,39]), with better prediction for mild and moderate states. In the literature, however, surprisingly few studies report error across subset range, meaning that the true extent of the problem cannot be determined. Errors should be reported across subsets of the EQ-5D range (e.g., EQ-5D \( 0 \leq 0.25, 0.25 \leq 0.5, 0.5 \leq 0.75, 0.75 \leq 1 \)), and a plot of observed and predicted values should be used. These are useful for indicating whether there is systematic bias in the predictions and whether heteroscedasticity is present. If there is systematic bias in the predictions, consideration will need to be given to how it impacts on the results of the cost-effectiveness analysis: for example, if the bias could systematically favor one intervention over another. In addition, errors reported across subsets of the range of the predictive measure(s) (see [24], e.g.) can inform application of the mapping algorithm in the trial data set and reporting this should be considered.

Some studies divide their estimation data set into two samples: an estimation sample and a validation sample (e.g., [39–42]). The mapping function is estimated on the estimation data set, and its performance is examined by using the validation sample. This has the advantage that it assesses the mapping function by its prime purpose; however, it reduces the sample size of the estimation sample. A randomly allocated split of the data should enable the analyst to assess how well the algorithm predicts the EQ-5D values for the validation sample. If predictive ability is poor when assessed on the basis of a nonrandom split of the data, it may not be possible for the analyst to judge whether the poor performance is due to the functional form of the model or a lack of generalizability to a systematically different population. The reduced precision in the coefficients of the mapping function from the reduction in sample size may be overcome by reestimating the mapping model by using the full data set once the specification of the model has been assessed by using the split-sample approach. If the division of the estimation data set into two is truly random, the model is expected to perform well; however, this does not necessarily indicate that it will perform similarly when applied to the trial data if the characteristics of the sample are different from those of the estimation and validation samples.

Application—Reflecting Measurement Error and Heterogeneity

There are at least three sources of measurement error when using mapping to estimate the EQ-5D or other health-related utility data from preference-based measures. These include heterogeneity in the measurement of the utility index values (e.g., as reflected in the standard errors from the general population valuation surveys of the preference-based measure). In addition, it includes the measurement error arising from the source and target measures (e.g., the self-reported description of own health on the EQ-5D, and the measurement error associated with other measures of health outcomes used to link the source and target measures). In addition, there is also the error arising from the estimation of the relationship between the source and target measures. While all possible sources of measurement error are not always fully reflected when presented, it is important to bear in mind that mapping can contribute to this latter further source of error as a result of the mapped values being predicted rather than directly reported. Furthermore, some researchers have shown that the confidence intervals around the predicted values as a result of mapping tend to be narrower than the confidence intervals around directly observed values, reflecting a reduction in the heterogeneity of the utility values [40,41].

As a minimum, the parameter uncertainty in the estimated regression analysis should be taken into account by using the variance-covariance matrices in the probabilistic sensitivity analysis (see [43] for an overview). If there are multiple possible mapping functions, these can be applied in sensitivity analyses to give an indication of the uncertainty associated with the choice of the algorithm. Alternative algorithms, however, capturing plausible forms of the relationship between the predicted EQ-5D values and alternative explanatory variables may not be available. Further research is needed to establish the best ways of capturing all the uncertainty in the mapped EQ-5D values. Until then analysts and decision makers should be aware that uncertainty and heterogeneity around mean mapped values may be underestimated.

Application—Use of Previously Developed Mapping Functions

Generating a de novo mapping function gives the analyst control over the inclusion and exclusion criteria for the estimation sample, and therefore influence over the generalizability of the mapping function to the target population. However, existing mapping functions may be available in the literature to the analyst. In these circumstances, we recommend that careful consideration be given to the generalizability of the mapping function to the target population, including the range of disease severity over which the function was estimated and the potential for systematic differences in the populations that could have an impact on the EQ-5D values. Most of the considerations outlined above would also apply to the review and use of published algorithms. There may be circumstances in which all the variables included in the published algorithm are not available to the analyst in the data set. Applying these algorithms is still theoretically possible by applying mean values to these variables; however, this reduces the granularity in the resulting estimates.

The Use of Mapping in NICE Technology Appraisals

Mapping has been most commonly used in NICE submissions where health-related utility data have not been directly collected within the
clinical trials of the treatments under consideration. Mapping techniques, however, have also been used to incorporate health-related utility data collected directly within the main clinical trial of interest into economic models, where the structure of the model is driven by other outcome measures. For example, an economic model may have been constructed to define health states by using a clinical measure of disease severity. In this case, mapping techniques can be used to explain the relationship between the two measures and to estimate the utility value (or distribution of values) associated with a health state defined by the clinical measure. An alternative approach would be to simply estimate the mean and variance for each of the health states described by the model from the data collected. For example, in the case of treatment for rheumatoid arthritis, the Health Assessment Questionnaire [44] is a commonly used measure of clinical outcomes. Several studies have sought to explain the relationship between health-related utility and Health Assessment Questionnaire scores by using mapping-type methods (see [24] for a recent overview). It is possible to use this approach even when utility data have been collected directly within the primary source(s) for clinical effectiveness, as a means of incorporating the data within the economic model. Care should be taken, however, so that the estimated mapping function accurately reflects the observed data from the trial when applied to the model.

The NICE Methods guidance refers to predicting EQ-5D data as the target variable but is not restrictive about the source measures from which the EQ-5D data are predicted. Reference is given to measures included in the clinical trials [2]; however, this is not restrictive and other types of studies may be more appropriate in some circumstances; for example, if the events of interest are rarely observed within a trial setting.

Other explicit recommendations made by NICE include that the mapping should be based on empirical data. This means that both the EQ-5D and the measure(s) used to map from are administered in a sample of people to generate empirical data, rather than researchers attempting to map on the basis of judgment (or other expert opinion) alone or the face value of the measures. Also, the adequacy of the mapping function should be demonstrated and validated, and the statistical properties of the function should be described [2]. The guidance is not prescriptive in stating which statistical or other tests should be undertaken; however, the most appropriate statistical and econometric tests have been discussed above.

Of the two reviews of the health-related utility data, the early review focused on independent assessment reports produced for the Technology Appraisals Programme up to May 2003 [12]. The authors report two clear cases of mapping in appraisals using empirical data: one where data from the Health Assessment Questionnaire were mapped onto the EQ-5D and another where data from the Child Health Assessment Questionnaire were mapped onto the EQ-5D. The methods used to undertake the mapping were reported to be limited or not reported at all. Stein et al. [12] identified a further five appraisals where health states had been mapped to preference-based instruments by using opinion rather than empirical data. In two cases, the mapping was conducted on the basis of clinical opinion, in one case it was based on opinions of the HTA analysts, and no details were provided in the remaining two cases. All but one of the HTAs mapped health states onto the EQ-5D; the other HTA mapped health states onto the Index of Health-Related Quality of Life.

The second published review of health-related utility data included in NICE included 46 appraisals conducted from the time of the implementation of the 2004 Methods Guide up to the time that the current Methods Guide was introduced in 2008 [15]. It included the independent assessment reports and the evidence dossiers submitted by the sponsors of technologies. Thirty-nine appraisals included cost-utility analyses; when including both independent and sponsor submissions, this accounted for 71 individual cost-utility analyses submitted to NICE. The review found that the use of mapping had increased since the previous review to over a quarter of submissions over the period of the review (n = 19). Empirical data were used to generate the mapping mechanism in 16 submissions, 1 was based on expert opinion, and the methods used in the other 2 submissions were unclear. Six of the submissions used published mapping algorithms in their analyses, and a further appraisal used an existing, but unpublished, algorithm. The majority of submissions included analyses that mapped onto the EQ-5D (n = 14). Other end points for the mapping process included the health utilities index (n = 2), short-form 6 dimension (derived from SF-36) (n = 1), and patients’ time trade-off values for their own health (n = 2). In most cases, health-related utility data were mapped from condition-specific measures (n = 14); the remainder mapped from generic HRQOL measures (n = 2) or nonstandardized vignettes of health states (n = 1), or the details were unclear (n = 2).

The updated review found that although the 2008 Guide explicitly allowed mapped estimates of health-related utility to be included in submissions, its use had actually fallen. Of the 44 appraisals, only 4 included mapping to estimate health-related utility data. All four appraisals based the mapping algorithm on empirical data. They were based on previously developed mapping algorithms that were publicly available as fully published studies (n = 2), in abstract form (n = 1), or from a previous NICE HTA report (n = 1). All four HTAs mapped from a condition-specific measure of either HRQOL or disease severity. Half of these analyses mapped data onto EQ-5D values, and the other half mapped onto patients’ time trade-off values of their own health. The submissions contained little information about the statistical properties of the mapping algorithms; however, they did provide references to the original documents that described how the algorithms were developed.

The evidence from NICE Technology Appraisal submissions to date suggests that although mapping has been used in submissions from the very early beginnings of NICE, the level of detail with which the mapping algorithms and analyses have been presented in the documentation has been generally poor, with few details of the statistical performance of the mapping algorithms being presented to the NICE Technology Appraisal Committee. In addition, much of the source data has included clinical measures of health or condition-specific measures of HRQOL. This is in contrast to the majority of the published literature that identified that maps from other generic measures of HRQOL (such as the SF-36) to the EQ-5D.

Conclusions

The main advantage of mapping in this context is that it can facilitate the estimation of health-related utility for cost-utility analysis where the utility data have not been directly collected, either at all or using the preferred utility instrument. It is particularly useful for organizations such as NICE that need a consistent approach to measuring and valuing health outcomes. In most cases, however, it will still be preferable to collect utility data directly rather than to introduce additional uncertainty into the estimates as a result of having to estimate the utility values on the basis of responses to other measures of health outcomes. This enables the data collected to directly reflect the impact of treatment on overall HRQOL, rather than just on the variables used to estimate the mapping algorithm. For example, if the mapping algorithm includes only a clinical measure, the mapping function may not reflect the impact of other effects of treatment that are not captured by the clinical measure. In addition, uncertainty and errors around the estimates can affect the accuracy of the EQ-5D utilities when used in economic
There may be, however, exceptions where other data sources are most appropriate, for example, where the trials are small or do not capture significant numbers of events that are expected to have an impact on HRQOL.

Much of the interest in mapping to date has focused on the EQ-5D as the target measure for mapping algorithms. Mapping to the EQ-5D should be used only when the EQ-5D is appropriate for that patient group and condition. All generic measures and the EQ-5D in particular may not be appropriate for all patient groups and conditions, and alternative methods such as the use of condition-specific preference-based measures may be considered to be more appropriate under these circumstances.

The review of NICE guidance has shown that there has been a decline in the practice of using researcher or clinical opinion to map between measures; however, the reporting of mapping studies is still poor in NICE submissions. Most of the mapping studies that have been included in NICE submissions have mapped from condition-specific measures of quality of life or clinical indicators of disease severity. The literature search for the recent mapping review [1] demonstrated that there was little published evidence examining the suitability of mapping in patient data sets. Since that review was conducted, however, mapping studies estimated by using patient data sets have been increasingly used and published.

We undertook an updated literature search by using the same search strategy as the recent review and found 31 studies meeting the inclusion criteria after an initial title sift. The large number of studies that were identified signals the recent popularity of mapping, and many of these articles offer methodological developments to approaches undertaken prior to 2007. The development and use of mapping algorithms for use in HTA is a developing area of methodological and applied research. Recent developments include approaches such as mapping between preference-based measures using general population visual analogue scale values for both measures valued alongside each other [46]. Recent developments in associated areas that may be informative for the mapping literature include mapping between Rasch scores and utility scores [47], the use of Gaussian processes [48], and single equation and two-part beta regression models estimated by using maximum-likelihood, quasi-likelihood, and Bayesian Markov-chain Monte Carlo methods [49].

One study suggests that the performance of different models varies at the overall and subgroup levels [36], two studies found no significant difference between mapped and observed QALY estimation to inform model selection and specification. The target sample is sufficiently represented. Standard statistical sample to include a broader range of people, provided that the some circumstances, it may be appropriate for the estimation with the target sample for the mapping analysis, although in most cases, mapping should be considered at best a second-best solution to directly collected utility data. If there is no overlap in content between the measures of interest, mapping is unlikely to be able to appropriately capture the relationship to estimate health-related utility. Alternative methods for estimating health-related utility data should be considered in these circumstances.

Acknowledgments

The production of this document was funded by the National Institute for Health and Clinical Excellence through its Decision Support Unit. We are grateful to John Brazier, Ben van Hout, Andrea Manca, Simon Pickard, Jennifer Priaulx, and Allan Wailoo for their very helpful comments on an earlier draft. The views and any errors or omission expressed in this document are those of the authors only.

Source of financial support: The article reports the outputs of a project funded by the Decision Support Unit of the National Institute for Health and Clinical Excellence.

REFERENCES


