Article:

Wailoo, Allan, Hernandez Alava, Monica, Manca, Andrea et al. (2017) Use of Mapping to Estimate Utility Values from Non-Preference-Based Outcome Measures for Cost per QALY Economic Analysis:Good Research Practices Task Force. Value in Health. pp. 18-27. ISSN: 1524-4733

https://doi.org/10.1016/j.jval.2016.11.006

Use of Mapping to Estimate Utility Values from Non-Preference-Based Outcome Measures for Cost per QALY Economic Analysis: Good Research Practices Task Force

Allan Wailoo[1], Monica Hernandez Alava[1], Andrea Manca[2], Aurelio Mejia[3], Joshua Ray[4], Bruce Crawford[5], Marc Botteman[6], Jan Busschbach[7].

[1] *School of Health and Related Research, University of Sheffield, UK*
[2] *Centre for Health Economics, University of York, York, England, UK*
[3] *Instituto de Evaluación Tecnológica en Salud, Bogotá, Colombia.*
[4] *F. Hoffmann-La Roche, Basel, Switzerland*
[5] *IMS health, Tokyo, Japan.*
[6] *Pharmerit International, Bethesda, US*
[7] *Section Medical Psychology, Erasmus Medical Center, Rotterdam, The Netherlands.*

Draft 19th May 2016

Total words: Current 8189

Abstract:

1. BACKGROUND. THE ISPOR TASK FORCE PROCESS

ISPOR to insert details here

2. INTRODUCTION

The assessment of health-related quality of life is critical in the evaluation of health care technologies and services, and in regulatory and reimbursement decisions. "Preference-based measures" (PBMs) play a central role in these evaluations. They allow patients to describe the impact of ill health and have an associated "utility" score (or tariff) for each of those health state descriptions where a value of 1 represents full health, 0 represents the value of dead, and negative values (if defined by the PBM) represent states worse than death. These utility scores can then be used for the calculation of Quality Adjusted Life Years (QALYs), which are an outcome metric for health benefit used in many health economic evaluations.

The most widely-used PBMs are *generic*: applicable to a wide range of diseases, patients and interventions. Examples include the EQ-5D[1], SF-6D[2], a derivative of the SF-36 instrument, and the Health Utilities Index (HUI)[3]. Many national guidelines for economic evaluation suggest or require the use of these generic instruments, such as England and Wales[4] , Spain[5], France[6], Thailand, Finland, Sweden, Poland, New Zealand, Canada, Colombia and The Netherlands. Some recommend the use of a particular instrument, usually the EQ-5D[7].

In many situations, clinical studies do not include a PBM. Often they will include one or more of the many patient-reported outcome measures (PROMs) which are not full PBMs because they do not have an associated, preference-based scoring system. Thus they do not permit construction of a QALY measure. Studies typically will also include physical outcomes (not patient-reported) which are measured "objectively", that is, without the interpretation of or report by the patient. In the absence of a PBM outcome, researchers will need to derive the "missing" PBM in order to estimate QALYs from these studies. In these circumstances the question is whether it is possible, and how, to predict the value that a PBM would have taken had this been collected, given what we know about the observed clinical outcome(s) and allowing for the mediating effect of the individual characteristics of study participants. "Mapping" attempts to answer this question and, in so doing, bridges the gap that often exists between available evidence on the effect of a health technology in one metric and the requirement for decision makers to express it in a different one (QALYs). It can also be used to provide a means of converting outcomes in one PBM to a different PBM.

"Mapping" makes use of another dataset, which may be observational rather than experimental. This dataset must have the same outcomes that are measured in the relevant clinical study/studies, and the patients' responses to a standard PBM instrument. This external dataset is used to estimate a statistical relationship between the two types of outcome measure. Combining the estimated statistical relationship together with the outcome data from the trial allows an estimate of the effect of the treatment in health utility terms and subsequently may be used to calculate QALYs. The practice of fitting a statistical model to health utility data has variously been referred to as ''mapping,'' ''cross-walking'' and ‘‘transfer to utility’’[8]. “Mapping” has entered into common usage so is used throughout this report.

In the context of economic evaluation, the evidence gap which gives rise to the need for mapping is commonly encountered. For example, Kearns *et al* (2013)[9] reviewed 79 recent NICE Technology Appraisals and found that mapping models were used in almost a quarter of cases. These included mapping from the Psoriasis Area Severity Index (PASI) in patients with psoriasis, from the Functional Assessment of Cancer Therapy – General (FACT-G) in patients with cervical cancer, and from the Patient Assessment of Constipation – Symptoms (PAC-SYM) and Patient Assessment of Constipation – Quality of Life (PAC-QOL) in women with chronic constipation, *inter alia*. The need for mapping may arise because of a failure to include a PBM in the relevant clinical studies (as described above), or because those studies are not sufficient alone to provide the utility information to estimate cost-effectiveness. There could be a requirement for extrapolation beyond the range of health states observed in clinical studies or a requirement to synthesise evidence from several clinical studies, not all of which include evidence on PBMs. Thus, mapping is an issue both for economic evaluation alongside trial data analysis without PBMs as well as for many economic modelling studies. And because studies that have been conducted historically will remain part of the evidence base as comparators for the evaluation of new technologies, mapping is likely to remain a requirement for some time, even when good practices for utility estimation are followed in contemporary clinical studies[10].

The current practice of mapping includes substantial variation in methods which are known to lead to differences in cost-effectiveness estimates[11,12]. The purpose of this Task Force report is to set out Good Research Practices that are relevant for the conduct of mapping studies for use in all types of QALY-based economic evaluation. The recommendations also have broader relevance to all situations where analysts wish to estimate preference-based outcomes as a function of any other variables, for example, where utilities are used as measures of provider performance[13]. Recommendations cover all areas of mapping practice: the selection of datasets for the mapping estimation, model selection and performance assessment, reporting standards, and the use of results including the appropriate reflection of variability and uncertainty. Such recommendations are critical in the face of inconsistent current practices, substantial variation in results between approaches and the risk of bias in several methods. Whilst other recommendations have been made[14,15], this document is unique because it takes an international perspective, is comprehensive in its coverage of the aspects of mapping practice, and reflects the current state of the art.

3. PRE-MODELLING CONSIDERATIONS

Prior to undertaking a statistical analysis for the purpose of mapping, the analyst must consider a number of different factors relating to the proposed and potential uses of the mapping itself. These uses create requirements for the dataset(s) in which the statistical analyses will be undertaken and tested.

Mapping is almost always undertaken with some pre-defined purpose and in many of those cases this is to inform a specific cost-effectiveness analysis (CEA). Clear understanding of the evidence gap to be addressed requires an understanding of relevant existing utility evidence, the requirements of the decision-making body that will assess the results of the analysis and the CEA in which the results are to be used. These factors help to inform the analytical choices which ensure unbiased estimates in the cost-effectiveness study. There will be requirements to appropriately reflect uncertainty and, additionally in some situations, the variability of estimates (for example, if simulating individual patients in a cost-effectiveness model).

The needs of the CEA will help guide the analyst's choice of methods and datasets that can be expected to perform appropriately for these specific needs. Where the analysis is to be used to populate a decision analytic model, one needs to consider what health states are reflected in that model – how are they defined and how do those definitions relate to both the clinical outcome measure or measures of relevance and the target PBM? If there is little overlap between the clinical outcomes and the PBM then mapping is unlikely to be successful. A descriptive comparison of the content of the different outcome measures, including the suggested PBM, is a useful starting point. This will highlight the specific facets of health each instrument measures. It is not a requirement for the PBM and clinical outcomes to address the same symptoms or functional (dis)abilities in order for mapping to be an appropriate approach but they do need to measure the same underlying concepts.

Many models, such as transition state models, will typically define a relatively small number of discrete health states. Other situations may require a combination of health states that can be derived in part from a mapping study and in part from other evidence. For example, the model may differentiate health states based on a disease outcome measure and the therapy patients are receiving, or the adverse events they experience, or their comorbidities. Mapping and other existing evidence can provide a range of options for addressing these evidence gaps.

Mapping outcomes to the utilities of a PBM is usually done with regression analyses. At one end of the spectrum, there are rare occasions where regression models can be avoided entirely simply by taking the mean and variance of the utility value for patients with the relevant health criteria. This simple approach is entirely legitimate if there is a single summary measure of disease to explain utility with no additional covariates that are considered important and there are sufficient observations of patients within each category. However, it should be noted that this may limit the generalisability of the mapping to other CEAs where these conditions do not hold.

Regression type analyses do become a requirement once additional covariate and/or extrapolation outside the range of the observed data are required, as is often the case. This might be because there are multiple disease specific outcome measures that reflect different dimensions of disease that collectively are used to estimate health utility. Or it could be because the analyst wishes to incorporate the effect of socio-demographics on health utility. For instance, age is likely to be a relevant variable in many situations as it will be related to health and quality of life. Another reason to consider regression models for mapping is the possibility of the need to extrapolate beyond the range of disease severity observed in the data. Whilst extrapolation beyond the range of the data is best avoided in any situation, this is not always feasible. Mapping studies are frequently based on datasets that do not include the full range of patient disease severity, particularly when these datasets are from randomised controlled trials with exclusion criteria for comorbidities and other aspects of severity. This contrasts with the needs of decision models, particularly those for patients with chronic conditions, which may model patients' lifetimes and thus span the entire feasible spectrum of disease.

It is well established that some methods for such regression analyses exhibit bias, the extent of which is in part dependent on the target utility measure. More details are provided in section 4, but it can be noted at this point that bias is typically greatest at the extremes of disease severity – for patients in severe ill- health these approaches overestimate their true health utility and for those in good health they underestimate health utility[16]. With this in mind the analyst must assess the requirements of the CEA. For instance, what is the range of disease to be addressed by the decision model? This judgment should not only be made against the characteristics of candidate patients at the point in the patient pathway where the technology of interest is being assessed (model baseline), but should be informed by the range of future health states to be covered in the model. Since this may cover a long term extrapolation encompassing patients experiencing diverse pathways including disease progression, therapy response and disease remission, a very wide range of disease severity can sometimes be covered.

Similar considerations influence the requirements for datasets in which the mapping function is to be estimated. Additional requirements are that, obviously, candidate datasets must come from studies of individuals completing both the relevant clinical outcome measure(s) and the target PBM simultaneously. There is no reason why randomised studies would be more desirable for mapping studies. Indeed, as alluded to above, randomised studies often have less diverse patients than other study types in terms of disease severity because of strict inclusion and exclusion criteria and limited follow up. Observational studies may be more likely to be drawn from representative patient groups, have larger sample sizes and can be conducted at relatively low cost. Where there is more than one candidate dataset then consideration should be given to the additional data fields the different studies include which may facilitate more precise estimates of the target PBM as well as the sample size, generalisability of the patient population and any potential biases in the study designs. However, this needs to be balanced with the use of those values in subsequent CEAs. The availability of information on respondents' age, for example, is likely to improve model fit and ought to be incorporated into a CEA. Datasets may be combined where common covariates exist and differences between patients and study designs are not expected to influence the relationship between covariates and PBM.

Uncertainty in the estimates should be minimized. This is facilitated in part by the use of datasets with larger numbers of observations and by avoiding extrapolation beyond the range of the data when feasible. Matching the range of disease severity in the dataset with the population of the CEA is important, but the range of other patient characteristics used as covariates in the mapping model are also relevant here.

Finally, the analyst needs to be aware of any potential biases in the dataset. Biases in this situation refers to those factors which influence a patient's reported health utility other than through an impact on the clinical outcome measure(s) used as explanatory variables. For instance, in some situations the types of therapies patients are receiving may exert some bias, for example, where those therapies are associated with adverse events unrelated to the clinical outcome being measured in the mapping dataset.

Summary of pre-modelling recommendations

1. Consider the use or potential uses of the mapping:
   a. Is it for use in a cohort decision model, patient level model or trial-based cost-effectiveness analysis?
   b. What are the health states that require utility estimates from the mapping and how do they relate to the PBM?
   c. What is the range of disease severity for which utility values are required?
2. Provide a descriptive account of the clinical explanatory variable, the dependent PBM and the extent to which they overlap.
3. Assess if a regression-based mapping is required.
   a. How many health states require estimates of utility?
   b. Are there additional covariates of importance?
   c. Are there sufficient observations within each category?
4. Identify if more than one dataset is potentially available for estimation. Compare the characteristics of candidate datasets.
5. To what extent does the distribution of patient characteristics in the sample datasets reflect those that are the subject of the cost effectiveness analysis? In particular, are all extremes of disease

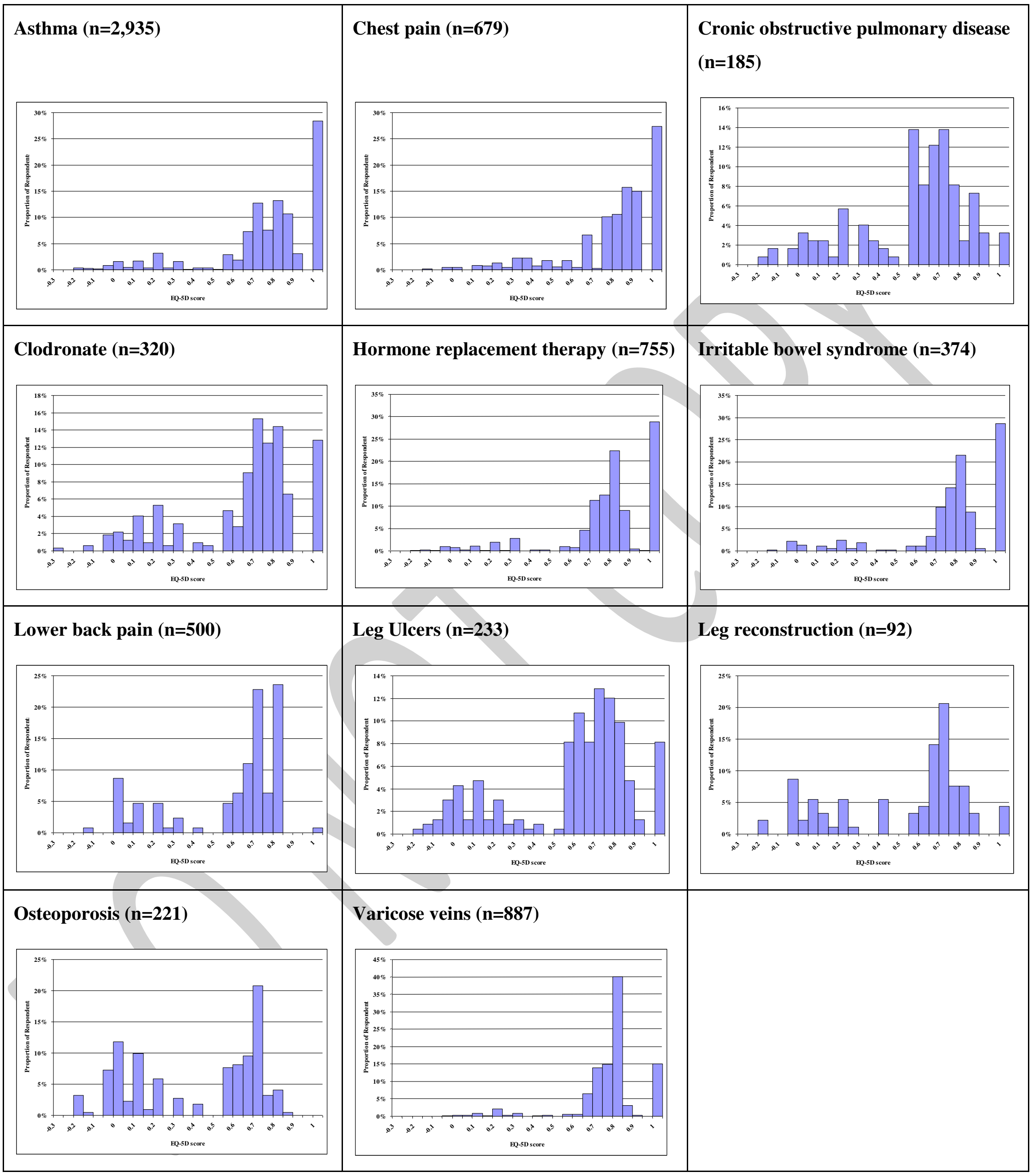
4. MODELLING AND DATA ANALYSIS

Selection of the statistical model

Utility measures tend to exhibit a number of non-normal distributional characteristics. These measures can be considered a type of limited dependent variable at both the top and bottom of their ranges: by definition a value of 1 is the maximum value that can be achieved and is considered equal to "full health". There is a lower limit which varies by instrument, sometimes referred to as the "pits" state. Note that these limits in utilities are not the same as "censoring".

Additional aspects of the distribution of utilities that influence the statistical model choice are the presence of large spikes in the distribution (typically at the "full health" upper bound), skewness, multimodality and gaps in the range of feasible values. Figure 1 displays examples of the distribution of EQ5D-3L from a range of different disease areas. The extent to which these features are present varies according to the instrument and scoring algorithm of the PBM that is the target for the mapping study, and the nature of the patient group. The presence of any of these features makes the application of simple statistical regression methods challenging and this is compounded when several of these features are simultaneously present.

Figure 1. The distribution of EQ5D-3L across different disease areas



There is considerable evidence that these distributional features result in systematic bias when linear regression methods are used to analyze the EQ-5D-3L instrument, the most commonly studied patient reported outcome in the mapping literature[17,18,19]. Similar findings have been shown to apply to models like the Tobit[19] (designed to deal with limited dependent variables), two-part models[20] (which attempt to address the mass of observations seen at full health) and censored least absolute deviations models[21,22]. A common finding in those reports is that expected health utility associated with mild health states is underestimated whilst utility for more severe health states is overestimated. When mapping studies with these biases are used in economic evaluations, clinically effective therapies appear less cost-effective than they truly are. Studies have shown that the magnitude of this bias is not trivial[11,12].

Recent work compares the performance of different statistical methods for mapping. One set of methods estimate the summary utility score directly. Amongst these direct methods, there is some empirical evidence to support the performance of two approaches: the limited dependent variable mixture model approach[19,12] and the beta-based regression approaches[17,23].

Alternatively, indirect methods estimate utilities as part of a two-stage procedure[24]. These methods have also demonstrated improvements over standard methods in some settings[12,25,26,27]. In the first stage, a so-called "response mapping" model uses a series of (either dependent or independent) separate regression functions to estimate the level on each separate domain of the descriptive system of the target PBM. Models suitable for ordered categorical data should be used for this first stage and the correlation between dimension responses incorporated[27] It is then straightforward to calculate the expected utility score as stage 2 of the procedure based on the probabilities assigned to each of the health states in the descriptive system and their associated utilities. This separation allows the analyst to apply any utility tariff to the models estimated in stage 1, according to their requirements. However, it should be noted that the appropriateness of the model and its fit is specific to the tariff in which it has been tested. Furthermore, response mapping models require sufficient observations in each of the levels of the descriptive system. Without this, the model(s) cannot be estimated.

We do not advocate any specific set of methods as the performance of different methods will vary according to the characteristics of the target utility measure, the disease and patient population in question, the nature of the explanatory clinical variables and the form of intended use in the CEA. We therefore suggest that it is wise to use a model type for which there is existing empirical evidence of good performance, and that respects the key features of the target utility measure, particularly the limited range of feasible utility values that can be taken in order to avoid problems in implementing results in a cost-effectiveness model.

Obviously, mapping does require analysts to adhere to good practice for statistical analysis in general. Below, we highlight some aspects of good practice that relate in particular to mapping. For instance, a plot of the distribution of the target utility measure provides a starting point for considering potentially appropriate modelling methods for direct analysis of the utility index. Analysts should use models that have theoretical plausibility, whose key assumptions hold, and that have a body of existing empirical evidence supporting their validity in the mapping literature. The use of models that do not meet these criteria requires additional justification and the results should be subject to additional scrutiny. This additional justification can be in the form of evidence that demonstrates that the mapping does not suffer from bias in the particular application, or that the nature of that bias is not an issue given the use of the mapping in CEA. For example, if the analyst intends to populate a cohort decision model where only a small number of health states are defined and these health states are not located at the extremes of poor/good health, then bias from the mapping may have a negligible effect on estimated cost-effectiveness. However, it is difficult to assess the impact of any potential bias *a priori*.

In most situations it will be extremely important to utilize mapping methods that meet the criteria set out above. This is because the extent and impact of biased estimates on cost-effectiveness will be significant and predictions outside the feasible utility range could be made. For example, model-based CEAs where health states are at the extremes of disease severity, individual patient simulation models, or analyses based on individual level data such as CEAs conducted alongside a single clinical trial will all be at risk of substantially biased cost-effectiveness estimates if inappropriate mapping methods are applied.

We note that some model types will require iterative estimation methods. It is imperative that the analyst ensures proper convergence of the estimation algorithm, whether undertaken in a classical[28] or Bayesian[29] framework.

It is also typical for candidate datasets to comprise multiple observations from the same individuals over time. In general one should seek to make use of all observations. Multilevel models can be used to reflect the correlations between these observations. At a minimum, clustered standard errors should be calculated. Where there are reasons to believe that there has been a break in the relationship between the covariates and the PBM then separate models should be estimated and the stability of the parameters tested.

The selection of covariates

In most situations the dataset in which the mapping is to be performed will contain information on a range of potential explanatory variables. The primary decision for the analyst concerns the choice of non-preference-based measure that will serve as the key link between the clinical effectiveness data and the preference-based one. In many situations the non-preference-based measure will be obvious because it will be the primary outcome measure used in clinical studies, or the sole quality of life instrument amongst the secondary outcomes. However, often those measures are formed of individual questions, which in turn can be reported either as dimension scores or a single summary score. Typically, there will be greater explanatory power from a regression model that uses disaggregated information from an outcome measure as explanatory variables. However, not only does this increase the number of explanatory variables but it may not provide the link to clinical evidence in a form that is widely usable (see, for example, Longworth et al[30] who modelled the 36 individual question responses to the EORTC instrument). This can be illustrated using the example of Rheumatoid Arthritis (RA). Typically, cost-effectiveness studies make use of the Health Assessment Questionnaire (HAQ) mapped to a preference-based instrument[31]. The HAQ is a summary score of functional impairment that ranges from 0-3 derived from 8 sub-sections each of which is comprised of 2 or 3 individual questions. Whilst the analyst may find a better performing model if using the individual item or dimension scores as explanatory variables, as opposed to the single 0-3 summary score, this should not be the sole criteria for covariate choice (see, for example, Bansback et al[32]). Where the mapping function is to be used to estimate health utility from individual questions or component scores, as might be the case in an economic evaluation conducted alongside a clinical trial, such an approach will be useful. However, decision models that synthesize data from several clinical studies will typically rely on the published results which will report only the summary score.

In other settings the analyst may have a choice of one or more disease specific outcomes. In Ankylosing Spondylitis (AS) for example, clinical studies typically report both BASDAI and BASFI outcomes measures of disease activity and functional impairment. The conceptual overlap with a preference-based instrument may be improved by the inclusion of multiple instruments and, hence, model fit.

Covariates can also be sociodemographic, disease characteristics and treatments. It is good practice to include covariates in order to avoid mis-specification of the model (resulting in the effects of the omitted variable being allocated to the error term and biased estimates for the coefficients). This remains the case even though the economic evaluation may not be designed to directly use each of these explanatory variables. The analyst can still use the mapping and simply set the value of the explanatory variable to that appropriate to their setting. This is preferable to omitting the explanatory variable. Of course, judgment is required here in order to avoid the inclusion of covariates that are highly correlated in the interest of developing a parsimonious mapping model.

Covariates should be theoretically justified *a priori* and reported in a manner that permits analysts to use results whether the covariate in question is used directly in their specific CEA or not. For instance, for most uses of mapping functions in CEA, the inclusion of age as a covariate is required and should be retained in preferred models even if not statistically significant. This allows any effect of ageing, independent of that which is captured as part of the clinical outcome measure(s), to be properly reflected. Where the mapping is intended for use in a CEA alongside a trial, covariates common to both the mapping dataset and the trial can be used to improve the generalizability of one to the other.

Summary of statistical modelling recommendations

1. Consider whether the cost-effectiveness analysis requires a formal regression based mapping model approach, or if it is suitable to take the mean value for sub-samples of patients.
2. If regression is required then model selection should be based on:
    a. Consideration of the most straightforward statistical model type whose assumptions are compatible with the target utility instrument. Use a plot of the distribution of the utility data to help inform that choice.
    b. Existing empirical evidence of the performance of different methods. There is no reason for this to be restricted to evidence from any specific disease area.
    c. The type of cost-effectiveness analysis where the mapping will be used and the extent to which biased estimates will affect the results.
3. For response mapping, models should be selected that respect the ordered nature of the categorical data in the descriptive system. Expected values should be calculated analytically.
4. Selection of the preferred mapping model is an iterative process that should conform to good practice common to all regression analyses.
5. Covariates should be theoretically justified *a priori*. Exclusion of covariates, even if they are not to be used in the cost-effectiveness model, risks mis-specification.

5. REPORTING OF MAPPING STUDIES

Mapping studies often form an important element of evidence submitted to decision-making Health Technology Assessment (HTA), pricing or reimbursement authorities. The findings must, therefore, be reported in a manner that allows a full assessment of the quality and relevance of the mapping by those that do not have access to the individual level data. In addition to this transparency requirement, it will be helpful to other analysts that sufficient information is reported to use the results in their own CEAs.

The dataset

Where more than one dataset could feasibly be used for mapping, provide a qualitative account of the selection rationale, at a minimum. The characteristics of the sample used in the estimation dataset must be provided fully. All variables should be described in terms of a measure of central tendency and distribution. Special attention should be given to the full distribution of patient observations at the extremes of disease severity, as described by the disease specific measures to be used as explanatory variables. This gives an indication of the extent to which the sample overlaps with the patients that are the focus of any CEA and, therefore, the extent of extrapolation required beyond the observed data.

Full information must be provided about the methods for sampling patients, both in the study as a whole and those sub-samples selected for use in the mapping study.

Many studies will include multiple observations from the same individuals over time. In this situation, it is important to report the pattern of those multiple, longitudinal observations and any features of the patients that change over those observations. For instance, if the follow-up period is substantial, then age is an important variable that will vary substantially from baseline. The number of available observations will differ according to the combination of covariates selected and this can lead to substantial differences between any final analysis and the description of the entire study sample. This also has implications for the ability to compare between models using measures of fit or penalised likelihood statistics.

Justification of statistical model type

As outlined above, there are numerous statistical challenges inherent in the analysis of utility data arising from its distributional features. The analyst should seek to select and justify their choice of method(s) *a priori* with reference to existing literature that has tested alternative methods using the target preference-based measure in question, examination of the distributional features in the estimation dataset, and the proposed use of the mapping function in any future cost-effectiveness study.

An algebraic description of the model is transparent, concise, unambiguous and ensures results can be used correctly by any competent analyst. Non-standard models, that have not been described elsewhere, must always contain such a description. An example of a predicted value from the mapping regression for some set of covariates should be reported. In some publications, additional software that calculates predictions for user defined inputs has been provided[25,33].

Justification for covariates used and how specified

Datasets used for mapping will typically offer the analysts a broad range of potential explanatory variables. These cover disease specific outcome measures, which often may be scored either as multiple components or summary index scores, of which there may be more than one, clinical measures, symptom specific information and demographics *inter alia*. A theoretical justification should be given for the inclusion of all variables within the set to be examined in the statistical analyses. It is instructive to provide an account of the dimensions of quality of life covered in the disease specific outcome(s) and contrast them with those covered by the target utility-based measure.

The methods used to move from a potentially large set of explanatory variables to a preferred model that is likely to include a smaller number, and in a particular form, must be detailed. There are many ways in which such regression models can be determined[9].

Model selection and performance

Theoretical justification for the selection of model type(s) should be provided drawing on previous literature and the specific features of the mapping to be performed, with a particular focus on the target utility measure. Regression models make assumptions which should be explicitly acknowledged and tested or assessed for plausibility. The proposed use of the mapping, if known, should also be discussed. Relevant aspects include the range of disease for which the results will be used, the manner in which uncertainty is to be considered and whether the analysis requires only expected utility values conditional on covariates (as is typically the case in a cohort decision model) or if simulated data is required (as in a trial-based analysis or patient level simulation model).

Results must be reported in a manner that provides transparency: readers of the results must be made aware of the process of selecting a preferred model(s) from the set of feasible ones and they must be provided with sufficient information to judge the validity of that process. This means that they need to be able to fully assess the performance of the preferred model(s) (and will require details on at least some aspects of performance of the less preferred models). Judgements are required at each stage of the model building process: reporting needs to highlight these judgements and their rationale. Sufficient information should be supplied to allow readers to be able to use the results of the mapping model in future cost–effectiveness studies.

One aspect of performance that is particularly important is model fit – the extent to which modelled values coincide with those observed in the data. Movement to a preferred model should not mechanistically follow some rule-based on overall fit. Specific judgement will be required and this will be context specific; for example, whether or not to include a particular covariate. Detailed information on model fit is required, however, for the final preferred model(s). Summary measures of fit like the R^2 are of very limited value here, particularly when presented in isolation, and provide little information of the validity of the mapping for use in subsequent CEA. The degree of between patient variability is inherently high in quality of life data, given the (warranted) subjective nature of quality of life. This results in relatively low R^2 statistics. Penalised likelihood statistics, such as the Akaike Information Criteria and Bayesian Information Criteria (AIC/BIC), provide a more appropriate means for comparisons of specifications within model types. Other summary measures of fit such as the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) have typically been applied in the mapping literature. These measures have their origins in the field of forecasting. It should, therefore, be recognised that these measures can appear very insensitive when applied in the mapping field because of the limited range of the dependent utility variable and the degree of variability inherent in patient outcome data. Any measure of model fit should be reported both for entire sample and for specific data ranges, defined in terms of the clinical explanatory variable(s). A plot of mean predicted and mean observed utility values conditional on the clinical variable helps to identify the existence and location of any systematic bias (see, for example, Wailoo et al.[33]) and where that bias occurs.

The fit of a model should not be assessed solely by reference to the point estimates of the predicted values compared to the data. It should also consider the uncertainty around those predictions and the model outputs once patient variability is included, as described below.

Reporting of results

All coefficient values must be reported to a sufficient number of decimal places to permit accurate estimation. Rescaling and centering covariates around their sample mean can facilitate this. Uncertainty in the estimated coefficients and associated correlation is imperative to allow the reflection of parameter uncertainty in the CEA – the covariance matrix should therefore be routinely reported[34] to allow probabilistic sensitivity analysis (PSA) to be undertaken. In addition to parameter uncertainty, the use of a mapping function to impute data at the individual level (for example, when conducting an analysis alongside a clinical trial) requires that the individual level variation is also reflected. In real world data, it is obvious that individuals with identical observable characteristics do not report identical health utility values. If mapping regression models are used simply to impute the same conditional expected value for these individuals, that individual level unexplained variability has been ignored and misrepresents both the clinical study and the results of the mapping. Information on the assumed degree and form of this variability is contained in the mapping regression error term(s) distribution and can be used as the basis for simulation methods that reflect this. Therefore, it is also essential that details of the error terms are reported routinely. With the availability of on-line materials, published mapping studies have no reason not to include these important items of information.

The guidance above relating to model selection suggests that one ought not select a model that is capable of producing estimates that lie outside the feasible range for the utility scale. But if such a model has been selected then when sampling from the mapping function, either for uncertainty or variability analysis, the frequency with which these samples lie outside the feasible range must be reported. It must also be reported how such unfeasible values were subsequently used or amended in the CEA. When a mapping is produced without any specific CEA in mind, it can still be useful to report the results of a simulated dataset from the model. This can help inform future CEAs and also forms a means of comparing the distribution of the data simulated from the model to the distribution of the original data (and can thus be used as part of the model selection process).

Empirical Validation

As with other statistical models, validation of the mapping model is relevant. Much of the guidance reported here is based on this requirement. The description of the dataset and the decision problem in which it is to be used, the process of model building and the performance of the final preferred model – each of these elements provides information on validation. To what extent can we have confidence that the model's predictions are accurate within the relevant patient group and to what extent might they be relevant in other similar patient groups? Existing UK guidelines on mapping recommend empirical validation[14] in this respect, described as estimation of the model in two datasets, either from two separate studies (external validation) or from splitting a single dataset (internal validation), though numerous other methods can be used for internal validation (for example, using bootstrapping-based approaches). In many situations, these empirical validation techniques will simply not be an option because there is only one candidate dataset of insufficient sample size to contemplate splitting.

Where any of these validation methods could feasibly be undertaken, there remains uncertainty about which of the available range of methods are most appropriate in the mapping setting and the additional value of the information these analyses provide. Sample splitting imposes the additional penalty of reduced sample size for estimation. For these reasons, we believe it would be premature to recommend empirical validation be conducted for all mapping studies. This is consistent with approaches undertaken for other regression-derived inputs to CEA.

Validation of alternative *methodological* approaches to the analysis of utility data can be achieved through repeated head-to-head testing in real-world and simulated datasets from different disease areas. However, routine multi-sample validation methods are not required for standard applied mapping studies because of the limitations noted above.

Summary of reporting standards recommendations

1. Describe relevant differences between datasets that are candidates for mapping estimation
2. Give full details of the selected dataset. Describe how the study was run and patients were sampled. Provide baseline and follow-up characteristics including the distribution of patients' disease severity. Missingness in the longitudinal pattern of responses should be described.
3. Plot the distribution of the utility data.
4. Justify the type of model(s) selected with reference to the characteristics of the target utility distribution and the proposed use of the mapping function.
5. Compare the dimensions of health covered by the target utility instrument and those covered by the explanatory clinical measure(s).
6. Describe the approach to determining the final model. Include tests conducted and judgements made.
7. Summary measures of fit are of limited value for the total sample. Provide information on fit conditional on disease severity as measured by the clinical outcome measure(s). A plot of mean predicted versus mean observed utility conditional on the clinical variable(s) should be included.
8. Coefficient values, error term(s) distributions(s), variances and covariances are required.
9. Provide an example predicted value for some set of covariates. Consider providing a program that calculates predictions for user defined inputs.
10. Parameter uncertainty in a mapping regression should be reflected using standard methods for Probabilistic Sensitivity Analysis (PSA). Assessment of model suitability for use in cost-effectiveness analysis should also consider the distribution of utility values for PSA, with particular focus on whether these lie outside the feasible utility range for the PBM.
11. When imputing data from a mapping function individual level variability should be incorporated using simulation methods and information about the distribution of the error term(s). These simulated data can be compared to the raw observed data, including an assessment of the range of values compared to the feasible range for the PBM.
12. Re-estimation of mapping results in a separate dataset is not routinely required.

6. THE USE OF RESULTS FROM MAPPING MODELS.

Selection of a mapping model for a cost-effectiveness study

Analysts may often need to select an existing mapping, perhaps from the published literature, to populate their cost-effectiveness model. In some situations, there may be no existing mapping that matches the population of interest. This might be that the precise characteristics of the patients do not match in terms of demographics, stage or severity of disease. In other situations it may be a more fundamental disparity such as the mapping being based on patients with a different disease. For example, the EORTC QLQ30 is a PROM used with patients with any type of tumour. Mappings have been estimated based on samples of patients with breast cancer[35]. Judgements about the suitability of a mapping study in a CEA should be based on an assessment of the differences between the patients or diseases in question. Are these differences likely to make the relationship between the mapping covariates and the target PBM non-generalizable?

Predicted values

The primary use of mapping for economic evaluation is to predict the mean health state utility value for a set of explanatory variables: in other words, the expected value conditional on covariates. If the guidance presented here has been followed, then a full understanding of the model specification and the estimated coefficients will have been provided and it will be obvious how to derive the required expected values. It may also be helpful for the mapping study to report the expected utility value and standard error for a given set of covariates for future reference. Some published studies go further and provide pre-programmed spreadsheet calculators as supplementary files[25,33].

Variability

A full specification of the statistical model and its estimated results, including error term(s) distribution(s), provides the required information to allow an analyst to reflect individual level variability. At its simplest, this may comprise a single normally distributed error term with mean zero and variance as reported. It is, therefore, straightforward to sample from the relevant conditional distribution to reflect variability around any required health state/patient characteristics.

Uncertainty

PSA is the standard accepted method for reflecting parameter uncertainty in health economic models. Monte Carlo simulation can be used to sample from the relevant joint distribution for regression model inputs, including mapping studies, provided the model specification, coefficient estimates and variance-covariances are reported.

7. CONCLUSIONS

Whilst the inclusion of appropriate preference-based measures in clinical studies is always recommended (see ISPOR Good Practice Guide Wolowacz et al[10] for guidance on this issue), this will not always be feasible or sufficient for the needs of economic evaluation. Mapping is, therefore, needed to allow analysts to bridge the gap between clinical evidence and the evidence required for economic evaluation. Provided that mapping analyses are undertaken appropriately, reported transparently and their results used appropriately, decision makers can be confident in the validity of estimates obtained in this manner.

8. REFERENCES

[1] The EuroQol Group (1990). EuroQol-a new facility for the measurement of health-related quality of life. Health Policy 16(3):199-208.

[2] Brazier, JE, Roberts, JR,. The estimation of a preference-based index from the SF-12. Medical Care, 2004;42(9):851-859.

[3] Horsman, J., Furlong, W.,Feeny, D., Torrance, D. The health utilities index (HUI®): concepts, measurement properties, and applications, Health Qual Life Outcomes. 2003; 1: 54.

[4] NICE Guide to the Methods of Technology Appraisal 2013

[5] Lopez-Bastida J, Oliva J, Antoñanzas F, Garcia-Altes A, Gisbert R, Mar J, Puig-Junoy J. Spanish recommendations on economic evaluation of health technologies. Eur J Health Econ. (2010);11:513–20.

[6] Haute Autorité de santé (HAS) http://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf 2013.

[7] ISPOR. Pharmacoeconomic Guidelines Around The World . Available at: http://www.ispor.org/PEguidelines/index.asp (accessed 1st Feb 2016)

[8] Mortimer D, Segal L, Sturm J. Can we derive an ‘‘exchange rate’’ between descriptive and preference-based outcome measures for stroke? Results from the transfer to utility (TTU) technique. Health Qual Life Outcomes. 2009;7:33.

[9] Kearns, B. Ara, R., Wailoo, A., Manca, A., Hernandez Alava, M., Abrams, K., Campbell, M. (2013) “Good Practice Guidelines for the use of Statistical Regression Models in Economic Evaluations”, Pharmacoeconomics, DOI 10.1007/s40273-013-0069-y

[10] Wolowacz, S. et al Good Research Practices for Collecting Health-State Utility Values for Economic Models in Clinical Studies. forthcoming

[11] Pennington, B., Davis, S. (2014) “Mapping from the Health Assessment Questionnaire to the EQ-5D: The Impact of Different Algorithms on Cost-Effectiveness Results”, Value in Health, Vol.17(8):762-71.

[12] Hernandez Alava, M., Wailoo, A., Wolfe, F., Michaud, K. (2014) “A comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes”, Medical Decision Making, Vol: 34:919–930.

[13] Devlin, N., Parkin, D. & Browne, J. (2009). Using the EQ-5D as a performance measurement tool in the NHS (Report No. 09/03). London, UK: Department of Economics, City University London.

[14] Longworth, L., Rowen, D. NICE DSU Technical Support Document 10: The use of mapping methods to estimate health state utility values. 2011.

[15] Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, Gray A. Preferred reporting items for studies mapping onto preference-based outcome measures: The MAPS statement. Health and Quality of Life Outcomes (2015) 13:106

[16] 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 2009;11:215–25.

[17] Basu, A. and A. Manca (2012). "Regression estimators for generic health-related quality of life and quality-adjusted life years." Med Decis Making 32(1): 56-69.

[18] Khan, Kamran A., et al. "Mapping between the Roland Morris Questionnaire and generic preference-based measures." Value in Health 17.6 (2014): 686-695.

[19] Hernández Alava M, Wailoo AJ, Ara R. Tails from the Peak District: Adjusted Limited Dependent Variable Mixture Models of EQ-5D Health State Utility Values. Value Health 2012; 15: 550-561.

[20] Acaster S, Pinder B, Mukuria C, et al. Mapping the EQ-5D index from the cystic fibrosis questionnaire-revised using multiple modelling approaches, Health and Quality of Life Outcomes 2015;13:33.

[21] Pullenayegum EM, Tarride J, Xie F, et al. Analysis of health utility data when some subjects attain the upper bound of 1: are Tobit and CLAD models appropriate? Value Health 2010;13:487–94.

[22] Huang I, Frangakis C, Atinson MJ, et al. Addressing ceiling effects in health status measures: a comparison of techniques applied to measures for people with HIV disease. Health Serv Res 2008;43:327–39.

[23] Hunger M, Baumert J, Holle R. Analysis of SF-6D index data: is beta regression appropriate? Value in Health 2011;14(5) : 759-67.

[24] Tsuchiya, A., Brazier, J., McColl, E., Parkin, D. (2002) “Deriving preference-based single indices from non-preference-based condition specific instruments: Coverting AQLQ into EQ5D indices”, University of Sheffield, Health Economics and Decision Science Discussion paper 02/01.

[25] Wailoo, A., Hernández, M., Philips, C., Brophy, S., Siebert, S. (2015) “Modelling health state utility values in Ankylosing Spondylitis: comparisons of direct and indirect methods”, Value in Health, Vol.18(4):425-431.

[26] Gray A, Rivero-Arias O, Clarke P. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. Med Decis Making. 2006;26(1):18–29

[27] Conigliani, C., Manca, A., Tancredi, A. (2015) “Prediction of patient-reported outcome measures via multivariate ordered probit models”, Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol.

[28] Cramer, J.S., 1986, Econometric applications of maximum likelihood methods (Cambridge University Press, New York, NY).

[29] Gelman, Andrew, and Kenneth Shirley. "Inference from simulations and monitoring convergence." Handbook of Markov Chain Monte Carlo (2011): 163-174.

[30] Longworth L, Yang Y, Young T, Mulhern B, Hernández Alava M, Mukuria C, et al.Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. Health Technol Assess 2014;18(9)

[31] Tosh J, Brennan A, Wailoo A, Bansback N. The Sheffield rheumatoid arthritis health economic model. Rheumatology (2011) 50 Suppl 4:iv26-iv31

[32] Bansback, N., Marra, C., Tsuchiya, A., et al. (2007) "Using the health Assessment Questionnaire to Estimate Preference-based single indices in patients with Rheumatoid Arthritis", Arthritis and Rheumatism, Vol.57(6):963-71.

[33] Wailoo, A., Hernandez Alava, M. Escobar Martinez A. (2014) Modelling the relationship between the WOMAC Osteoarthritis Index and EQ-5D, Health and Quality of Life Outcomes, 12:37, DOI: 10.1186/1477-7525-12-37

[34] Ara, R., Wailoo, A.J. NICE DSU Technical Support Document 12: The use of health state utility values in decision models. 2011.

[35] Crott R, Briggs A. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. Eur J Health Econ (2010) 11:427–434