LEADING ARTICLE

Constructing Indirect Utility Models: Some Observations on the Principles and Practice of Mapping to Obtain Health State Utilities

Christopher McCabe · Richard Edlin ·
David Meads · Chantelle Brown · Samer Kharroubi

Abstract The construction of mapping models is an increasingly popular mechanism for obtaining health state utility data to inform economic evaluations in health care. There is great variation in the sophistication of the methods utilized but to date very little discussion of the appropriate theoretical framework to guide the design and evaluation of these models. In this paper, we argue that recognizing mapping models as a form of indirect health state valuation allows the use of the framework described by Dolan for the measurement of social preferences over health. Using this framework, we identify substantial concerns with the method for valuing health states that is implicit in indirect utility models (IUMs), the conflation of two sets of respondents' values in such models, and the lack of a structured and statistically reasonable approach to choosing which states to value and how many observations per state to require in the estimation dataset. We also identify additional statistical challenges associated with clustering and censoring in the datasets for IUMs, additional to those attributable to the descriptive systems, and a potentially significant problem with the systematic understatement of uncertainty in predictions from IUMs. Whilst recognizing that IUMs appear to meet the needs of reimbursement organizations that use quality-adjusted life years in their appraisal processes, we argue that current proposed quality standards are inadequate and that IUMs are neither robust nor appropriate mechanisms for estimating utilities for use in cost-effectiveness analyses.

C. McCabe (✉)
Capital Health Endowed Research Chair, Faculty of Medicine and Dentistry, University of Alberta, Suite 736 University Terrace, 8303 112 Street, Edmonton, AB T6G 2T4, Canada
e-mail: mccabe1@ualberta.ca

R. Edlin
School of Population Health, University of Auckland, Auckland, New Zealand

D. Meads
Academic Unit of Health Economics, University of Leeds, Leeds, United Kingdom

C. Brown
United BioSource Corporation, London, United Kingdom

S. Kharroubi
Department of Mathematics, University of York, York, United Kingdom

1 Introduction

Health care reimbursement authorities' use of cost-effectiveness analyses to inform their decisions has had implications for the research activities of the disciplines that contribute to health technology assessment. Triallists use preference-based outcomes measures [1], statisticians develop methods for synthesizing evidence of the safety and efficacy [2], and economists are interested in how the value of technologies can best be measured [3]. Value assessment often utilizes preference-based health-related quality-of-life (HRQL) data. For these measures, such as the EQ-5D [4], each state described by the measure can be assigned a standard score measured on a utility scale. However, data on preference-based HRQL measures are not always collected as part of studies establishing clinical effectiveness, and so researchers have sought to link non–preference-based HRQL measures with these utility scales.

For example, a trial of a new treatment for myeloma might collect EORTC QLQ-C30 data [5] to establish clinical effectiveness, but not EQ-5D data. If the health technology assessment for the new treatment required EQ-

5D utility estimates, the trial data would not meet this requirement. However, if another trial dataset had collected both EORTC QLQ-C30 and EQ-5D data [6], then the data might be used to construct a model that predicts the EQ-5D utility scores for each EORTC QLQ-C30 state. This use of 'mapping' or 'cross-walks' has become sufficiently important for the UK National Institute for Health and Clinical Excellence (NICE) to issue technical advice on the process [7]. Authors are beginning to compare the performance of mapping models with directly elicited preference data [8].

Since the aim of 'mapping model' estimates is to act as a proxy for more direct observations, we refer to these models as 'indirect utility models' (IUMs). In contrast, the more standard approach by which observed preference-based measures are valued with a tariff based on elicited utilities is instead referred to as 'direct utility measures' (DUMs). As both IUMs and DUMs are constructed with a view to providing robust and unbiased estimates of utility values for health states, the considerations underlying a gold-standard DUM approach should also be of value when considering a gold-standard IUM approach. Such literature exists, when developing both preference-based quality-of-life measures and the utility scores attached to them [4, 9, 10, 11].

Whilst authors have begun to propose standards for the development of utility 'mapping models' [12], the most recent statements of good practice in the construction of DUMs and IUMs diverge in potentially important ways [11, 12]. Longworth and Rowan [12] propose that the key issue in the assessment of the estimation sample is whether the clinical and demographic characteristics of the estimation sample are similar to the characteristics of the sample to which the IUM will be applied. By contrast, Xie et al. [11] focus on reporting standards when selecting which health states are valued, how many valuations are to be included in the estimation sample, recruitment strategies and survey format. This paper uses the framework proposed by Dolan for developing health state valuation studies [13] (i.e. within DUMs) to consider whether IUM estimation studies are coherent with the fundamental principles of good practice in DUM research. This paper also considers the implications of recognizing IUMs as a method for the measurement and valuation of health for their robust estimation and the critical appraisal of their use in health care resource allocation decision processes.

2 Measuring and Valuing Health

Dolan [13] identifies key questions in the process of measuring the value of health states. As both DUMs and IUMs provide utility values for the same ultimate purpose,

it is helpful to consider each model in relation to these six questions.

2.1 What Is to Be Valued?

There are a number of approaches to defining health for DUMs, including clinical, functional and subjective well-being. Direct measures embody various approaches, and there is no accepted basis on which to prefer any single measure as superior. In the context of IUMs, the choice of the operational definition of health is specified by the non–preference-based measure for which the analyst wishes to estimate utilities.

2.2 How Is it to Be Described?

Within a DUM, the description of health is the HRQL measure to be valued. In a similar way, the description of health when constructing an IUM is the descriptive system of the non–preference-based measure. Despite this, it is important to note that the description of health in the preference-based measure of health used to measure the utility of a specific health state may indirectly impact upon the modelled utilities.

2.3 How Is it to Be Valued?

In DUMs, health state values are normally measured directly, using choice-based methods such as Standard Gamble, Time Trade Off or Discrete Choice Experiments [3]. For IUMs, scores are estimated on utility data obtained by the implicit question 'Which health state in this preference-based measure is equivalent to your health as you have described it in the non–preference-based measure?'. The implications of this difference in the valuation method for the interpretation of the data obtained are discussed in Sect. 3 below.

2.4 Who Is to Value it?

The choice of whose values to use is a pivotal one. By convention, although probably not consensus, the values of the general population are considered to be most appropriate for informing population health care resource allocation decisions [14–16]. For IUMs, the values derived reflect (i) the respondents in the estimation dataset who provide 'equivalent health states' in the non–preference-based and preference-based health state descriptive systems, and (ii) the values of the respondents in the dataset on which the DUM for the preference-based health state measure was estimated. This is discussed in more detail and its implications are explored in Sect. 4 below.

2.5 How Are Values for All Health States to Be Generated?

For DUMs, a bespoke valuation survey is undertaken with careful decisions around the choice of health states to be valued, the number of observations per health state and the socio-economic characteristics of the valuation sample informing the construction of the estimation. Ideally, DUM development studies would also construct a validation dataset with direct utility data for health states not included in the estimation dataset. A second series of decisions concern the specification of the form of the utility function and the appropriate regression techniques to be used after considering the nature of the data.

In contrast, IUMs are estimated on datasets that happen to be available, rather than ones that are constructed for the purpose. Whilst it may often be serendipitous that such datasets exist, this is not sufficient to guarantee that they are fit for purpose. The selection of health states, the number of observations on each and the socio-economic characteristics of the respondents are a product of chance, not design. Section 5 discusses how this might impact upon the process for constructing IUMs, their assessment and the likelihood that they are suitable for use in resource allocation decisions.

2.6 How Are the Valuations to Be Aggregated?

The expected mean utility is generally accepted as the appropriate measure for aggregating health state utility values, when they are intended to inform population health care resource allocation decisions. This applies equally to both types of utility models.

3 How Is Health to Be Valued for IUMs?

IUMs are estimated on mean health state valuation data obtained from a preference-based HRQL measure, such as the EQ-5D. Respondents in the estimation dataset for the IUM provide pairs of health state descriptions, with one health state from each of the HRQL measures. The description from the preference-based HRQL measure is combined with the utility algorithm to identify the expected mean utility for both health states. The implicit valuation method may be described as an equivalence question—respondents are asked to identify the health state descriptions in each instrument that are equivalent, and the descriptive equivalence is assumed to support value equivalence.

The valuation mechanism is a two-stage process: stage 1 identifies the equivalence and stage 2 attaches a utility. For each possible health state, there is only one possible value, which is the mean preference of the respondents in the preference-based HRQL valuation study, given by the utility algorithm. Two respondents choosing the same state in the preference-based measure will be given the same utility value, even if they are not in the same state in the non–preference-based measure.

The valuation mechanism does not involve any explicit question about value, strength of preference or willingness to trade related to the health state being valued. The use of choice-based valuation techniques has been used as an argument for the superiority of certain preference-based measures, including the EQ-5D [17]. IUMs have an uncertain claim to be based upon choice-based preference data.

More importantly, the valuation mechanism does not directly capture any variation in the value attached to any specific health state in the preference-based measure. The use of the mean predicted health state value from the preference-based measure as the input into an IUM systematically understates the variation in the measured value for each state, leading to a systematic understatement of the uncertainty in the predicted utilities. The potential importance of this is discussed further below.

4 Whose Values Should Be Used?

The method of obtaining health state utility values for an IUM relates to the answer to the question 'Whose values should be used?', as the values of the respondents providing the data for the estimation dataset are partially captured in those data. Consider the EORTC QLQ-C30 example referred to earlier; whether an individual experiencing nausea—an EORTC QLQ C30 subscale—considers the symptoms of nausea sufficiently problematic to state that they were unable to perform their usual activities will depend, at least partly, on the disutility they attach to nausea. A similar relationship could credibly exist between nausea or vomiting in the EORTC and anxiety/depression in the EQ-5D. Respondents in the same EORTC state may not be in the same EQ-5D state, and one of the sources of this variation will be differences in the value that respondents attach to the EORTC state. Hence, the preferences of the respondents in the IUM estimation dataset are reflected in the EQ-5D utilities, not just the preferences of respondents to the original EQ-5D valuation study.

In the second stage of the valuation process described above, the values come from the respondents to that valuation survey for the preference-based measure. However, it is only the mean value that is used, and respondent-level variation in health state values is ignored. This means that IUMs provide an arbitrarily constrained account of the uncertainty in the mean health state values for a specific health state.

As IUMs provide a confounded estimate of the values for specific health states and systematically understate the uncertainty in the modelled mean health state values, even well-constructed IUMs will not necessarily provide robust data for health care resource allocation decisions.

5 How Are Values for All Health States to Be Generated?

Brazier and colleagues [3] describe the design considerations for health state preference modelling studies. Whilst there is a limited theoretical framework for the development of statistical health state utility models, there is some consensus in DUM valuation about the use of orthogonal arrays to identify balance across a core set of health states for direct valuation. From here, supplementary states can be selected from across the health state space to achieve an even representation and to allow for the estimation of any likely interaction effects—such as interactions between extreme levels on one or more health state domains [4, 9]. Journal articles and technical papers tend to report the states included in the valuation study and the rationale for these states.

By contrast, IUMs are estimated on convenience datasets. Almost by definition, the health states for both measures reflect the clinical casemix of the population sampled, rather than any conceptual model of the likely relationship between the quality of life measured by the non–preference-based measure and health state utility. Similarly, the number of observations for any particular health state in the estimation dataset reflects an interaction between the casemix of the population sampled and the number of observations. There is no consideration of the minimum number of observations required per state for model estimation to be a statistically meaningful pursuit.

Whilst carefully designed direct health state valuation studies can produce well-populated and relatively well-behaved datasets [9, 10] utility data are increasingly recognized as providing particular statistical challenges even under the best of circumstances [18–21]. Challenges of estimating DUMs associated with censoring and clustering in the descriptive system are likely to be exacerbated for IUMs. Within those datasets available for IUM research, casemix may lead to few or no observations over substantial portions of the health state space, so that the group considered is far more homogeneous than the clinical population at large. Further, the use of mean health state values from the utility algorithm will create additional artificial clustering, since the substantial differences in elicited individual-level utilities are lost. To the degree that variation occurs in the dataset, it can only be driven by variation in the underlying health (casemix) and

differences between respondents in how they locate their health in the space described by the preference-based measure.

The act of pairing states in the two descriptive systems highlights issues associated with the degree of overlap between the descriptive systems. Often the non–preference-based measure will have more domains and /or more levels than the preference-based measure. (This is certainly the case for the EORTC QLQ-C30 and EQ-5D.) Whilst a number of domains are common to both measures, such as Pain and Usual Activities, some domains are only present in one of the measures. Considering domains that are only present in the non–preference-based measure, it is important to consider whether and to what degree they are likely to be correlated with domains that are present in both measures, as well as domains that are present only in the target measure.

Defining independent domains as domains that are present only in the non–preference-based measure and shared domains as those that are present in both measures, the question of whether to include independent domains in an IUM may be important. Correlation may lead to the estimation of significant coefficients on the parameters even though there is no information on the value of different levels of functioning in the utility data. Under these circumstances, parameter estimates for the value of both independent and shared domains of quality of life may be misleading. It is important therefore not to include domains that occur in only one of the descriptive systems, as correlation may lead to spurious but statistically significant parameters in the IUM.

The resolution to this problem may lie in the a priori specification of the likely relationship between the domains of health in the non–preference-based measure and the value of health as described by the preference-based measure. Such a specification might also allow some assessment of the adequacy of the estimation dataset for constructing the proposed model. This requires specification of a 'shared health state space' that the IUM can be expected to value and the presence of data encapsulating interactions between domains and levels that would be expected to impact upon its value.

Given the concerns about the suitability of datasets for estimating IUMs, as well as specifying the expected relationship prior to examining the estimation dataset, sensitivity analyses examining the impact of (i) eliminating states and (ii) eliminating observations on the model parameters may be valuable. Whilst not done frequently, sensitivity analysis eliminating health states has been used before. For example, Kharroubi and McCabe [22] examined the sensitivity of parametric and non-parametric models utility models for the HUI2 using this technique. If a model is insensitive to the specific selection of states that

is used to estimate it, this is interpreted as evidence that the model captures a real underlying relationship rather than one that is unique to the estimation dataset.

It is desirable in statistical model checking to predict different data from those used to fit the model. This is often achieved by reserving some data for model checking whilst using the rest of the data for model fitting. When the distribution of observations over states is relatively uniform, random exclusion would be expected to produce a dataset with a comparable error structure to the full dataset, as all states are equally likely to be affected. This would lead to similar statistical models being estimated on the reduced dataset. However, when the distribution of observations across states is substantially uneven, a random exclusion of observations will impact upon some states more than others. If models estimated on the reduced dataset are substantially different from those estimated on the full dataset, this would be evidence that the models are dependent upon the casemix of the estimation dataset and therefore probably not suitable for predicting utilities for datasets with a different casemix.

6 Using IUMs to Provide Utilities for Cost-Effectiveness Analyses

IUMs are usually constructed to inform cost-effectiveness analyses requiring utilities when only non–preference-based data are available [7]. Such analyses will inform resource allocation decisions, so that IUMs are likely to directly impact upon patient access to care. Given this, there is increasing interest in the appropriate consideration of uncertainty, and the risk of making the wrong decision, in both the academic literature and health policy practice [23, 24].

As described above, the exclusion of respondent variation in utilities means that estimated utilities from IUMs will understate the uncertainty in the predicted utilities. In so doing, it may underestimate the risk of making the wrong decision based upon outputs from the cost-effectiveness analysis. For those decision makers able to reach research-based recommendations, estimates of the value of further research will also be flawed.

One potential solution is for IUMs to be estimated on simulated datasets that retain the uncertainty in the modelled utility values from the direct utility study, whilst allowing this to be combined with the uncertainty in the relationship between the descriptive systems. Thus the direct utility model underlying the utility algorithm for the preference-based measure, including the estimated standard errors, could be used to sample a range of possible utility values for each measured health state. This 'uncertainty supplemented' dataset could then be used to estimate

the IUM. This approach might allow decision makers to fully understand the decision uncertainty to the utility parameters in the model and the value of requiring more research on them.

The criteria for choosing between alternative models of the same relationship should be related to fitness for purpose. Health state utility models are evaluated in terms of goodness of fit and accuracy of model predictions. Here, both the magnitude of the prediction error and bias are important [3, 4, 8–10]. In the context of decision making, the precision of model predictions should also be considered. Models that produce a more tightly defined range of plausible values for a specific health state may be more valuable than models that produce more accurate central estimates but much larger plausible ranges. The failure to consider prediction precision represents a disjuncture between model assessment and the purpose for which models are developed. Whilst the issue may be particularly acute for IUMs, it is an issue for all utility models developed to inform cost-effectiveness analyses.

7 Discussion

The estimation of IUMs is becoming increasingly common, with some degree of respectability in the guarded endorsement of NICE and the implicit endorsement of the EuroQol group [7, 14, 25]. However, there has been limited consideration of their intellectual foundations. In this paper, we propose that it is appropriate to evaluate IUMs as a method for measuring and valuing health-related quality of life for use in resource allocation decisions.

IUMs have the attraction of providing data in an expedited fashion to inform time-sensitive resource allocation decisions. Construction of an IUM is not a particularly resource intensive exercise and typically aims to bridge an evidence gap. Whilst Longworth and Rowen [12] identify them as meeting a need, this implies some faith in their ability to perform the desired function. However, using the framework described by Dolan [13], we have identified substantial concerns with their use. It cannot be assumed that predicted values from an IUM are representative of the values that would have been obtained if the states had been valued directly, as the predictions conflate two sets of preferences. The chance determination of which states are valued in the estimation dataset is likely to impact upon the capacity of the paired data to accurately capture the relationship between the health-related quality-of-life descriptors and health state preferences—a problem that is exacerbated by the serendipitous determination of the number of observations per state. The greater prior likelihood of clustering and censoring the non-bespoke datasets typically used for their estimation has implications for the

choice of regression methods and thus the interpretation of the model outputs. In decision contexts where uncertainty in the estimates of cost effectiveness is an explicit consideration for the appraisal process, IUMs' systematic understatement of the uncertainty in the modelled utilities will lead to both positive and negative reimbursement decisions when funding further research is the more efficient option.

Longworth and Rowan set out to provide recommendations on good practice in the estimation of IUMs, and state that these are summarized in Table 1 of their paper [12]. However, whilst describing the process of estimating these functions, it is unclear whether all of the statements are prescriptive as to what the authors would consider best practice. In contrast, Xie et al. [11] propose a good-practice checklist for the reporting of valuation studies, but they are not prescriptive regarding what represents good practice in the implementation of valuation studies.

Recognizing that our critiques of IUMs are matters of principle that may prove to be empirically inconsequential, we have proposed additional strategies for evaluating IUMs to provide analysts and decision makers with insight into the magnitude of the impact of these problems in individual cases. These are summarized in Appendix 1. We further suggest that those evaluating IUMs look to add some assessment of precision in model predictions to current model evaluation criteria. Further work on characterizing the uncertainty from IUM predictions also appears necessary to support decision makers when choosing between definitive reimbursement decisions and requiring more research.

8 Conclusions

The use of IUMs has increased rapidly over recent years in response to the need of reimbursement authorities to express the value of health care interventions using a consistent scale. An implicit assumption has been that these models successfully estimate the index of interest. Careful consideration of the process of IUM construction using the standard framework for measuring and valuing health casts substantial doubt on this assumption. If, as we have argued, IUMs do not reliably measure preferences over health, their use to allocate limited health care resources is difficult to defend. The use of IUMs as currently constructed carries with it a significant risk of harming rather than promoting population health.

Appendix 1: Recommendations for Construction and Use of IUMs

1. IUM estimates between non–preference-based HRQL instruments and preference-based HRQL values should be driven by theoretically plausible relationships.
2. Relationships between non–preference-based HRQL instruments and preference-based HRQL values should be specified prior to empirical analysis.
3. Cost-effectiveness analysis that use IUM utility estimates should not be considered 'reference case' analyses.
4. Sensitivity analyses should be used to test for the influence of the following factors on the estimated relationships:

(a) The choice of health states used to inform the estimated relationship.
(b) The choice of observations used in the estimation sample.
(c) The choice of IUM.

5. When the estimated relationships are not sensitive to these factors, the utility estimates should be used to estimate the value of information associated with delaying reimbursement and collecting 'reference case' compliant utility data.
6. Should the estimated relationships be found to be sensitive to these factors, the IUM utilities should not be used.

References

1. Collinson FJ, Gregory, W. McCabe, C et al. The STAR TRIAL protocol: a randomized multi-stage II/III study of sunitinib comparing temporary cessation with allowing continuation, at the time of maximal radiological response, in the first line treatment of local advanced/metastatic renal cancer. BMC Cancer. 2012;598:1–11.
2. Welton NJ, Sutton AJ, Cooper NJ, Abrams KR, Ades AE. Evidence synthesis for decision making in health care. Chichester: Wiley; 2012.
3. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press; 2007.
4. Dolan P. Modelling valuations for EuroQol health states. Med Care. 1997;35(11):1095–108.
5. European Organisation for Research and Treatment of Cancer. EORTC QLQ-C30. http://groups.eortc.be/qol/eortc-qlq-c30. Accessed 19 Jun 2013.

6. Guillou PJ, Quirke P, Thorpe H, Walker J, Jayne DG, Smith AM, Heath RM, Brown JM. Short-term endpoints of conventional versus laparoscopic-assisted surgery in patients with colorectal cancer (MRC CLASICC trial): multicentre, randomised controlled trial. Lancet. 2005;365:1718–26.

7. Longworth L, Rowan, D. NICE DSU technical support document 10: the use of mapping methods to estimate health state utility values (2011). http://www.nicedsu.org.uk/TSD%2010%20 mapping%20FINAL.pdf. Accessed 19 Jun 2013.

8. Rowen D, Young T, Brazier J, Gaugris S. Comparison of generic, condition specific and mapped health state utility values for multiple myeloma cancer. Value Health. 2012;15(8):1059–68.

9. Brazier J, Roberts J, Deverill M. The estimation of a preference based measure of health from the SF-36. J Health Econ. 2002;21(2):271–92.

10. McCabe C, Stevens K, Roberts J, Brazier JE. Health state values for the HUI2 descriptive system: results from a UK survey. Health Econ. 2005;14:231–44.

11. Xie F, Gaebel K, Perampaladas K, Doble B, Pullenayegum E. Comparing EQ-5D valuation studies: a systematic review and methodological checklist. Med Decis Mak. Online First March 22, 2013. p. 1–323.

12. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. Value Health. 2013;16:202–10.

13. Dolan P. The measurement of health related quality of life for use in resource allocation decisions in health care. In: Culyer AJ, Newhouse JP, editors. Handbook of health economics. Amsterdam: Elsevier Science BV; 2000.

14. National Institute for Health and Clinical Excellence. Guide to the methods of health technology appraisal (2008). http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf. Accessed 19 Jun 2013.

15. Brazier J, Akerhurst R, Brennan A, et al. Should patients have a greater role in valuing health states? Appl Health Econ Health Policy. 2005;4(4):201-8.

16. Gold MR, Segel JE, Russell LB, Weinsten MC. Cost effectiveness in health and medicine. New York: Oxford University Press, 1996.

17. Brazier JE, Longworth L. NICE DSU technical document 8: an introduction to the measurement and valuation of health for NICE submissions (2011). http://www.nicedsu.org.uk/TSD8%20Introduction%20to%20MVH_final.pdf. Accessed 19 Jun 2013.

18. Huang IC, Frangakis C, Atkinson 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(1 Pt 1):327–39.

19. Craig BM, Oppe M. From a different angle: a novel approach to health valuation. Soc Sci Med. 2010;70(2):169–74.

20. Hernández Alava M, Wailoo A, Ara R. Tails from the Peak District: adjusted censored mixture models of EQ-5D health state utility values (2010). http://eprints.whiterose.ac.uk/11074/1/HEDS_DP_10-08.pdf. Accessed 19 Jun 2013.

21. Austin PC. A comparison of methods for analyzing health related quality-of-life measures. Value Health. 2002;5(4):329–37.

22. Kharroubi SA, McCabe C. Modelling HUI 2 health state preference data using a nonparametric Bayesian method. Med Decis Mak. 2008;28:875–87.

23. National Institute for Health and Clinical Excellence. Patient access schemes and the NICE Patient Access Liaison Unit (2013). http://www.nice.org.uk/aboutnice/howwework/paslu/patientaccessschemesliaisonunit.jsp. Accessed 13 Feb 2013.

24. Claxton K, Palmer S, Longworth L, Bojke L, Griffin S, McKenna C, Soares M, Spackman DE, Youn J. Informing a decision framework for when NICE should recommend the use of health technologies only in the context of an appropriately designed programme of evidence development. Health Technol Assess. 2012;16.

25. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard AS. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15(5):708–15.