
Article:
Pennington, B. orcid.org/0000-0002-1002-022X, Hernandez-Alava, M., Pudney, S. et al. (1 more author) (2018) The Impact of Moving from EQ-5D-3L to -5L in NICE Technology Appraisals. PharmacoEconomics. ISSN 1170-7690

https://doi.org/10.1007/s40273-018-0701-y

Title

The impact of moving from EQ-5D-3L to 5L in NICE Technology Appraisals

Authors

Becky Pennington[1]

Monica Hernandez-Alava[1]

Stephen Pudney[1]

Allan Wailoo[1]

Affiliations and addresses

1. School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Corresponding author

Becky Pennington. Email address: b.pennington@sheffield.ac.uk, phone: 0114 222 0745

ORCID: 0000-0002-1002-022X

Abstract

Background

The EuroQol-Five Dimension (EQ-5D) is the National Institute for Health and Care Excellence (NICE)'s preferred measure of health-related quality of life (QoL) in adults. The 3-level (3L) value set is currently recommended for use, but the 5-level (5L) is increasingly being used in practice. We aimed to explore the impact of moving from 3L to 5L in NICE appraisals.

Methods

We adapted our existing mapping for use with health state utility values derived from a population where the original distribution of utilities was unknown. We used this mapping to estimate 5L utilities for 21comparisons of interventions from models used in NICE technology appraisal decision-making, covering a range of disease areas.

Results

All utilities increased using 5L, and the differences between highest and lowest utilities decreased. In 10 oncology comparisons, using 5L generally increased the incremental quality-adjusted life years (QALYs) as the benefit from improving survival increased. In 4 non-oncology comparisons where the intervention improved QoL only, the incremental QALYs decreased as the benefit of improving QoL was reduced. In 7 non-oncology comparisons where interventions improved survival and QoL, there was a trade-off between increasing the benefit from survival and decreasing the benefit from improving QoL.

Conclusion

3L and 5L lead to substantially different estimates of incremental QALYs and cost-effectiveness. The direction and magnitude of the change is not consistent across case studies. Using 5L instead of 3L may lead to different reimbursement decisions. NICE will face inconsistencies in decision-making if it uses 3L and 5L concurrently.

Key Points for Decision Makers

- Our mapping can be used to convert mean 3L utilities into 5L utilities for use in economic models
- Using 5L instead of 3L in economic models slightly decreases ICERs for interventions that improve survival, but can substantially increase ICERs for interventions that only or primarily reduce morbidity
- Using 5L instead of 3L will lead to different decisions about cost-effectiveness and this cannot be rectified by applying a simple adjustment to the ICERs below which interventions are cost-effective

1 Introduction

The EQ-5D is the National Institute for Health and Care Excellence (NICE)'s preferred measure of health-related quality of life in adults[1]. NICE currently states that the EQ-5D 3 Level (3L) value set should be used and not the EQ-5D 5 Level (5L) value set, although a review of this statement is planned for 2018[2]. It has been claimed that the descriptive system of 5L is superior to 3L[3] and other studies suggest that 5L has increased sensitivity and precision[4, 3]. Given the increasing use of 5L in practice, it is important to determine the impact of moving from 3L to 5L in economic evaluation. To do this, we use statistical mapping to convert existing evidence compiled on a 3L basis to the new 5L basis.

We have previously developed methods for mapping between 3L and 5L and demonstrated that they do not generally give the same results when used in cost-effectiveness analyses conducted alongside clinical trials[5]. In that publication, we showed that 5L shifts mean scores up the utility scale towards full health and compresses them into a smaller range, compared to 3L. In 13 comparisons from nine case studies based on direct observation of 3L at the individual level, we found that mapping to 5L increased incremental cost-effectiveness ratios (ICERs) for most interventions, with one notable exception where the intervention substantially improved survival[5]. There is therefore a need to further explore the comparison between interventions that improve quality of life only and interventions that improve survival. Case studies from the NICE Technology Appraisal (TA) programme are of particular interest, given their relevance to decision making and reimbursement in England, and NICE's requirement for such work to inform their position on the use of EQ-5D[2].

In our previous work, the mapping model was used to convert observed 3L health descriptions to 5L utilities, but it can also be used to convert 3L utilities to 5L utilities [5] in cases where the underlying 3L health descriptions are not available.

Our aim was to explore the impact of moving from 3L to 5L in NICE TAs and how that impact might differ across the range of disease areas considered by NICE's TA programme. Since economic evaluation for most TAs involves economic models using mean utilities for health states, we needed

to apply our method for mapping mean utilities. The analysis presented here is in two parts. First we develop a strategy for using our mapping model to convert mean 3L utilities to 5L. We then use this strategy to explore the impact of using mapped 5L utilities instead of 3L utilities in a range of economic models that have been used in NICE's TA programme.

2 Methods

2.1 Mapping

Data

Our mapping method uses econometric models estimated from either of two datasets: the EuroQol Group international coordinated study (EQG) (n=3551) and the US-based FORWARD databank (n=5205). The EQG dataset contains 3L and 5L responses for a sample drawn from eight broadly defined disease groups and a healthy population. The FORWARD dataset samples patients with rheumatoid disease and contains 3L and 5L responses and responses to many other general and rheumatoid disease specific questions. The 3L and 5L responses are converted into utilities using the standard value sets for the UK / England[6, 7]. The 3L and 5L value sets have been discussed elsewhere, where analysis found that the distribution of value sets differed, with 5L values higher than 3L values, a smaller range of 5L values and less difference between adjacent states in 5L[8]. Further information on the FORWARD and EQG datasets is provided elsewhere[5, 9].

Our mapping jointly models responses to the five domains of 3L and 5L. The model is very flexible to allow the 3L and 5L responses within each domain to be strongly correlated, but the strength of the relationship to vary by severity of illness, by using one of five different copulas (functions which describe the dependence between variables). The model also allows the relationship between 3L and 5L to vary between domains, and accommodates common factors linking response behaviour across domains. It allows for the effect of age and sex without assuming they influence 3L and 5L in the same way [10]. Within and out-of-sample testing of the mapping model found no significant concerns and the biases often seen in mapping studies applying simplistic methods were not present[11]. Because we are modelling responses to the descriptive systems, the mapping is not biased by the health or preferences of people within the datasets from which the mapping was derived.

The 3L and 5L utility value scales are non-linear, so it is not appropriate to map the 3L response to the most likely 5L response, and then to value that response using 5L – such a procedure would introduce bias. Instead, when individual responses are available, the model is used to generate the probability of each possible configuration of 5L responses for each individual's 3L response; then applies the 5L utility scale to each of those responses; and then weights each possible score according to the probability of that response to give the predicted (conditional mean) 5L utility score for each individual.

In cases where only the 3L utility and not the response to each domain is known there may be multiple 3L responses that correspond to this utility (since the valuation of 3L response combinations are not necessarily unique). Alternatively, the 3L utility may not correspond to any of the utilities for the 243 health states, for example if another measure has been mapped to 3L and an approximate utility has been estimated, or where the 3L utility is the mean of a sample of individual patient utilities. In these situations, the mapping can consider a range around the 3L utility under consideration and search within this range for valid 3L utilities. The user must specify the range around the 3L utility in which the model should search for valid 3L utilities – this range is termed the 'bandwidth'. If there is only one valid 3L value within the bandwidth, the model uses this as a valid and unique 3L utility. If there are multiple 3L utilities within the bandwidth, the model assigns a weight to each utility based on its distance from the input utility, such that more weight is given to closer values. Each of these nearby 3L values is then mapped to 5L as described above, and the mapped 5L utility is calculated as the weighted average. A full technical description of the mapping model, with guidance on its use, is provided elsewhere [10, 12].

Analysis

In cost-effectiveness models, often the utility used for a population health outcome is the mean of a set of utilities from an unknown range, and the mapping then needs to be used appropriately to reflect this. If we knew the true distribution of the 3L utilities that form the mean, we could map each individual 3L utility to 5L and then calculate the mean for 5L. Information on the true distribution will rarely be available so instead we use weighted averaging of 5L utilities calculated by mapping

every valid 3L utility within a range of the utility used in the economic model. The size of this range is defined by the user-specified bandwidth. A smaller bandwidth averages the 5L predictions corresponding to a more localised part of the 3L utility scale; a larger bandwidth averages over a broader range of the 3L utility scale. To develop a strategy for choosing the bandwidth, we analysed the EQG and FORWARD datasets to compare actual 5L utilities with values mapped from mean 3L utilities, exploring the effect of varying bandwidth sizes. Using the more detailed and homogenous patients in the FORWARD dataset, we carried out the analysis separately for each of eleven subgroups, categorised by patient global assessment of disease severity.

2.2 Case studies

Data

We selected 20 case studies of models used in NICE TA decision making representing the range and type of technologies and disease areas typical of the TA programme. Each of the economic models used in the case studies was based predominantly, but not exclusively on 3L utilities (since it is often not possible to find EQ-5D values for all utility inputs in a model, limiting to exclusively 3L would have dramatically reduced the number of potential case studies). For all cases, final guidance had been produced, ICERs were reported in the guidance, and the role these ICERs played in developing recommendations was clear. In selecting case studies, we initially reviewed a list of models critiqued or developed by two Assessment Groups, and selected case studies that met the above criteria. We then discussed the selected case studies with NICE staff, who advised other indications that should be considered, based on their experience of working on a number of appraisals, and so we reviewed a list of past TAs to identify potential suitable topics. We considered that our approach reflected a pragmatic cross section of the types of interventions considered by NICE, and sufficiently large to draw general conclusions. A comparison of the distribution of our case studies and all NICE TAs from TA001 to TA516 across International Classification of Diseases-10 classifications is shown in Figure 1. We negotiated access to the models via NICE and the relevant Assessment Group and gained access and permission to use all of the case studies we requested.

Seventeen of the 20 cases were Single Technology Appraisals (STAs). We focussed on the technology under appraisal and the primary comparator reported in the final guidance. Two cases were Multiple Technology Appraisals (MTAs): in one case only one comparison was considered (the new technology dominated other comparators); in the other case we considered three interventions, drawing pairwise comparisons between the least effective and next least effective, and the latter with the most effective. Additionally, we considered one case study from the Highly Specialized Technology (HST) programme. Therefore, we considered 21 comparisons.

Analysis

In each model, we mapped 3L utilities to 5L utilities using the Stata command EQ5DMAP, for a range of bandwidth values depending upon the 3L utility. We only mapped utilities which were stated as being derived from 3L data, so values from other preference-based measures, expert opinion, or unclear sources were not mapped (these were generally more minor aspects of the analyses). We used both EQG and FORWARD-based versions of the mapping. We used the most general version of the mapping model, which uses the best fitting copula in each domain[9].

We calculated total and incremental quality-adjusted life years (QALYs) for the intervention and comparator using mapped 5L utilities, and then calculated ICERs. We analysed the change in incremental QALYs and ICERs for comparisons in three categories:

- Oncology interventions that improved survival (10)
- Non-oncology interventions that improved survival (4)
- Non-oncology interventions which did not improve survival (7).

We explored differences within and across categories, to understand general tendencies in terms of the settings in which mapping to 5L may increase QALY gains and those in which QALY gains may decrease. We compared the ICERs to the range of maximum acceptable ICERs specified in the NICE Methods Guide[1] and conducted scenario analyses to explore the impact of mapping disutilities.

3 Results

3.1 Mapping

The mean directly-observed 5L utilities in the EQG and FORWARD datasets (0.712 and 0.778) were compared to 5L utilities produced by mapping the mean 3L utility (0.628 and 0.681), using varying bandwidths to perform the mapping. We found that small bandwidths (less than 0.1) returned 5L estimates very close to those observed (largest difference 0.023 for EQG and 0.028 for FORWARD). For EQG, the mapping slightly overestimated the mean 5L utility as the bandwidth increased towards 0.3 (largest difference 0.047), and then decreased towards the actual utility at a bandwidth of approximately 0.4. Above the (very large) bandwidth value of 0.4, the mapped mean EQG utility declined, falling further below the actual mean (largest difference 0.196 at a bandwidth of 1). This is primarily because, with a larger bandwidth, the mapping can include utilities much lower than the mean 3L utility, but cannot include utilities much higher as they are capped at 1. For FORWARD, the mapped utilities remained close to the actual mean as the bandwidth increased towards 0.32 (largest difference 0.025). Above 0.32 the estimated mean utility declined further below the actual mean, for the same reason as EQG (largest difference 0.253 at a bandwidth of 1).

To determine appropriate bandwidths for a range of mean 3L values, we performed the same analysis as described above for eleven severity groups of the FORWARD data, categorised by the Rheumatoid Arthritis Patient Global Assessment. The patient global assessment is included in the American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) Rheumatoid Arthritis remission criteria[13]. The version used in the FORWARD is the 21 point numerical rating scale version of a Visual Analogue Scale. The inclusion of this variable in the FORWARD dataset permitted us to mimic analyses for datasets with different average levels of illness severity while retaining much of the individual variability in EQ-5D scores typically observed. There is substantial variation in EQ-5D utilities within each severity group, so choice of bandwidth has an effect. The detailed results are shown in Figure 2 and suggest the following bandwidth strategy, which we have followed in our analysis of the case studies.

(1) For mapping a 3L utility in the gap between full health (1.0) and the next feasible 3L value (0.883), the bandwidth should be just large enough to reach 1.

(2) For 3L utilities in the range 0.883 to 0.7 (inclusive), smaller bandwidths are preferable and should not exceed 0.1. In our application we use a bandwidth of 0.05.

(3) For 3L utilities below 0.7 down to 0.6, a bandwidth of 0.2 is recommended.

(4) For 3L utilities below 0.6, a larger bandwidth is appropriate; we recommend the value 0.4.

3.2 Case studies

In each case study we replaced the 3L utilities which were used in the original economic models by 5L utilities mapped from them. The results are summarised by Figure 3 in terms of the changes to estimates of incremental QALYs and by Figure 4 in terms of the corresponding ICERs. Three case studies are omitted from Figure 3 and Figure 4 as their results are confidential.

The switch to 5L increased utilities in all cases with one exception (where the 3L utility was 1), using both the EQG and FORWARD datasets for mapping to 5L. We observed a general tendency for the increase in utility to be larger for EQG-based mapping than for FORWARD-based mapping if the 3L utility was high (around 0.8 or higher), but smaller if the 3L utility was low (0.8 or below). Consequently, the choice of reference dataset used for mapping has some bearing on the results, and it may affect studies based on populations with different mean utilities in different ways. Since all utilities increased when mapped to 5L, the total QALYs for intervention and comparator groups in each case study also increased. However, the difference between the lowest and highest utility in each case study decreased after mapping to 5L. This is a consequence of the more compressed range of the 5L value set and is consistent with previous results reported in Hernandez et al[5]. It means that the incremental QALY gain will increase for an intervention which derives all its benefit from improving survival, but decrease for an intervention which derives all its benefit from improving quality of life. The interventions in the four non-oncology comparisons which did not improve survival are examples of the latter type of intervention, and Figure 3 shows that all of their incremental QALY gains

decreased using 5L. The mean change in incremental QALYs was -29.0% for EQG (range: -41.6%, -16.2%) and -44.9% for FORWARD (range: -55.2%, -24.4%).

The interventions in the other 17 comparisons derived their incremental QALY gain from a combination of improving survival and improving quality of life. This meant that in some cases the QALY gains increased, and in other cases the QALY gains decreased (see Figure 3).

For the ten comparisons for oncology interventions, the mean change in incremental QALYs was 8.3% for EQG (range: 0.5%, 15.2%) and 7.3% for FORWARD (range: -7.3%, 16.2%). The incremental QALY gain decreased using 5L in only one case study, and then only when FORWARD-based mapping was used. In that case study, utilities were defined by pre- and post-progression and were higher for the intervention than comparator group. Patients receiving the intervention spent longer in the pre-progression state than patients receiving the comparator, but less time in the post-progression state. Using 5L, the utilities for the pre- and post-progression states all increased, but the increase was greater for post-progression than for pre-progression states. Although the switch to 5L caused an increase in the QALY gain from spending more time in pre-progression and delaying death, with FORWARD-based mapping, it was offset by a decrease in the QALY gain from keeping people in pre-progression instead of post-progression.

In all other comparisons from the oncology case studies, the increases in QALY gain from improving survival outweighed the decrease in QALY gain from delaying progression. Two oncology case studies modelled utility by time to death rather than pre- and post-progression: in each case the intervention increased the time spent in the health state furthest from death and patients on intervention and comparator spent approximately the same time in the other health states. The magnitude of the increase in incremental QALYs in these cases therefore depended only on the magnitude of the increase in the utility of the health state furthest from death.

For the seven comparisons for non-oncology interventions which improved survival, the mean change in incremental QALYs was -12.0% for EQG (range: -38.8%, 5.0%) and -19.8% for FORWARD (range: -53.5%, 6.6%). The use of mapped 5L utilities slightly increased the incremental QALYs for

two case studies: one in circulatory system disorders and one in musculoskeletal disease. The incremental QALYs decreased or only marginally changed (less than 0.002) for the other four case studies, which considered mental and behavioural disorders, nervous system diseases, musculoskeletal diseases, infections and metabolic disorders. As shown in Figure 3, there was no relationship between the size of 3L incremental QALYs and the direction of change. It is therefore not possible to draw general conclusions on the basis of disease area or of 3L incremental QALYs.

In most case studies, we did not map the disutilities associated with adverse events because the source of the disutility was not EQ-5D or was not reported. Where we did map adverse events, the incidence of adverse events was higher for comparator than intervention groups, and mapping the disutility to 5L decreased its size, so the incremental benefit of the intervention was reduced. To further explore the impact of disutilities, we considered one case study which included disutilities for disease-related exacerbations, adverse events and caregiver burden. Leaving these disutilities unchanged but mapping the primary health states led to decreases in QALYs gained of 72% (EQG) and 64% (FORWARD). Including all disutilities further decreased health gain to 61% (EQG) and 47% (FORWARD).

Figure 4 shows changes in ICERs. These ranged from a decrease of 13.23% to an increase of 7.89% for oncology interventions that improved survival (mean decrease of 8.61% EQG and 5.57% FORWARD); from a decrease of 6.22% to an increase of 115.16 % for non-oncology interventions that improved survival (mean increase of 13.02% EQG and 28.02% FORWARD); and from an increase of 22.14% to an increase of 123.44% for non-oncology interventions that did not improve survival (mean increase of 53.5% EQG and 108.3% FORWARD).

4 Discussion

Mapping from 3L to 5L shifts the distribution of utilities upwards, towards full health and compresses them into a smaller range. This means that, for an intervention that only improves survival and does not improve quality of life, there is a tendency for the incremental QALYs to increase (and ICERs to decrease) when 5L is used. The increase in QALYs will generally be larger using FORWARD than

using EQG. For an intervention that only improves quality of life by delaying or avoiding disease progression and does not improve survival, incremental QALYs will decrease when 5L is used. The decrease in QALYs will usually be larger when using FORWARD rather than EQG based mapping. Many interventions improve both quality of life and survival, so there is a trade-off between the gain in incremental QALYs from improving survival and the decrease in incremental QALYs from reducing the benefit of delayed progression.

Our findings echo previous research findings that 3L and 5L lead to different estimates of cost-effectiveness, and that switching to 5L typically increases ICERs for interventions that do not extend life[5]. Our research expands on this by applying mapping in the context of model-based rather than individual-based direct evaluation. We also consider a wider range of case studies of interventions, including 17 that extend life. Other research has found that compared to 5L, 3L overestimates health problems and therefore underestimates utility scores[3], consistent with our findings that utility values increase using 5L. The increased sensitivity and precision of 5L compared to 3L[3], and in particular the reduction in ceiling effects[4], may be expected to result in greater changes in utility using 5L, and hence higher QALY gains, but instead we found that 5L reduced the difference between best and worst health states. Janssen et al's comparison of 3L and 5L in seven countries found that while 3L had better discriminatory power between healthy versus disease states, 5L had better discriminatory power between mild and moderate/severe states[3]. In the examples considered in our case studies, the utilities for the health states were closer to the healthy and mild states than severe states in the study by Janssen et al, so it is unsurprising that the difference between the health state utility values decreased.

We demonstrate that the use of 5L may slightly decrease the ICERs for life-extending interventions (although they can increase) and can substantially increase the ICERs for interventions which only or primarily improve quality of life. Due to variation in the magnitude and direction of change in ICERs, it is not possible to define a simple adjustment that could be applied to the range of maximum acceptable ICERs that is used for determining whether interventions are cost-effective when 5L is used and that would be consistent with decisions based on 3L.

We have seen that the ICERs for a few interventions are below a £20,000-£30,000 range using 3L, and above this range using 5L, demonstrating that using 3L and 5L concurrently may lead to inconsistencies in decision-making[1]. Policy decisions about 3L versus 5L will impact real patients and their treating clinicians.

Our study only included case studies that primarily used EQ-5D, and only mapped the inputs in those case studies that clearly used EQ-5D. Although EQ-5D is the stated preferred measure of health-related quality of life in NICE's reference case[1], it is not necessarily used in all appraisals. It could be argued that since non-EQ-5D utilities in NICE appraisals are being treated as if they are EQ-5D that they should also be mapped to 5L. Previous research has explored the differences between 3L and other preference-based measures of health [14-18], but no such studies exist for 5L. In situations where utilities (or more commonly disutilities for adverse events) were elicited from experts, if the experts are assumed to have estimated these based on the 3L utility system, then arguably the utilities should be mapped. Alternatively, if the experts are assumed to have estimated these based on understanding the concept of utility scores without referring to specific quality of life measures, then arguably the values should not be mapped.

Our approach to selecting case studies was pragmatic rather than systematic – in addition to representing a range of disease areas and using EQ-5D, selection was influenced by date of publication and suggestions from the project team. The decision to stop after 20 case studies was taken once the results for 20 had been summarised, and the project team felt that conclusions were unlikely to change with the addition of more case studies. In the comparison in Figure 2, our case studies tend to generally reflect the areas covered by NICE TAs. Neoplasms is the classification with the most appraisals – this is higher in our analysis than in all appraisals considered by NICE as we wanted to include case studies with a range of QALY and survival gains. Furthermore, the proportion of NICE TAs which are oncology has increased in recent years, and we wanted to reflect current and future practice. Figure 2 appears to indicate that we do not consider any case studies of eye conditions, but one of our case studies considers diabetic macular oedema, which is classified as an endocrine disorder, but the economic modelling is based around visual acuity. Our analysis does not

contain any case studies in respiratory conditions whereas there are several NICE TAs in this area, but we note that several of these are in influenza where EQ-5D was not used. We have a lower proportion of musculoskeletal conditions and a higher proportion of mental and behaviour disorders, but consider that our case studies do reflect the range of disease considered in NICE's TA programme. However, it is important to note that we do not make general conclusions about the relationship between disease area or incremental QALYs using 3L and the magnitude or direction of change using 5L and instead report the impact of interventions that improve survival and/or reduce morbidity. Therefore, our findings are not dependent upon the selected case studies, and as such our case study selection process is unlikely to have biased our results. Our findings and conclusions apply to any scenario where an intervention improves survival, reduces morbidity or both.

Of course, all results are subject to the underlying data used to map the two EQ-5D variants. There are several limitations to both datasets[9] and a programme of work is being undertaken by NICE and the Department of Health and Social Care in the UK to address these uncertainties[5].

5 Conclusion

5L and 3L lead to very different assessments of cost effectiveness. Consistent with analyses conducted using case studies based on economic evaluations alongside trials we find that the impact on NICE TAs of moving from 3L to 5L could be profound. The ICERs for interventions that improve survival can decrease using 5L, but by a relatively small degree. However, it is not possible to predict whether the counter effect of decreased QALY gain from any additional morbidity effect will decrease or increase ICERs. Appraisals of technologies that only improve morbidity can see very large increases in ICERs using 5L, in some cases these more than double.

The 5L is not simply an extended version of the 3L. Differences in both descriptive systems and valuation methods mean the two cannot be treated as if they were interchangeable. A move to 5L will lead to different decisions than would have been the case under 3L. There is no simple proportional adjustment, such as changing the range of ICERs below which interventions are considered to be cost-effective, that could rectify this conundrum. Future changes to NICE policy need to be aware of this information in order to ensure decision making is consistent, fair and reflects scientific state of the art.

If in the future NICE determines that 5L should be used instead of 3L, our mapping can be used to convert 3L utilities to 5L utilities in economic models.

Data Availability Statement

The models used in the current study are not publically available as they formed part of companies' submissions to NICE.

Acknowledgements

The authors wish to thank the FORWARD databank, their patient participants and directors Kaleb Michaud and Fred Wolfe. The authors wish to thank the companies who gave permission for their models to be considered as case studies in this report. The authors wish to acknowledge the contributions of Rosie Lovett, Jacoline Bouvy, Jo Richardson, Janet Robertson, and Sophie Cooper at NICE who suggested case studies for inclusion, facilitated access to models and critically reviewed the analysis.

Author contributions

Monica Hernandez-Alava and Stephen Pudney developed the mapping algorithm and adapted it for use with mean utility scores. Becky Pennington used the algorithm to map utility scores for the case studies, adapted the models to use the 5L utilities and analysed the results. Allan Wailoo conceived the idea for the analysis and provided suggestions and critique of the approach. Becky Pennington drafted the first version of the manuscript, which all authors reviewed and made changes to. All authors reviewed and approved the final manuscript.

Conflicts of Interest

Becky Pennington reports no conflicts of interest. Monica Hernandez-Alava reports no conflicts of interest. Stephen Pudney reports no conflicts of interest. Allan Wailoo reports no conflicts of interest.

References

1. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. 2013. https://www.nice.org.uk/process/pmg9/chapter/the-reference-case. Accessed 13 December 2017.
2. National Institute for Health and Care Excellence. Position statement on use of the EQ-5D-5L valuation set. 2017. https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisal-guidance/eq5d5l_nice_position_statement.pdf.
3. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive Systems and Value Sets from Seven Countries. Pharmacoeconomics. 2018. doi:10.1007/s40273-018-0623-8.
4. Buchholz I, Janssen MF, Kohlmann T, Feng YS. A Systematic Review of Studies Comparing the Measurement Properties of the Three-Level and Five-Level Versions of the EQ-5D. Pharmacoeconomics. 2018. doi:10.1007/s40273-018-0642-5.
5. Hernandez Alava M, Wailoo A, Grimm S, Pudney S, Gomes M, Sadique Z et al. EQ-5D-5L versus EQ-5D-3L: The Impact on Cost-Effectiveness in the United Kingdom. Value in Health. 2017. doi:10.1016/j.jval.2017.09.004.
6. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Econ. 2017. doi:10.1002/hec.3564.
7. Dolan P. Modeling Valuations for EuroQol Health States. Medical Care. 1997;35(11):1095-108.
8. Mulhern B, Feng Y, Shah K, Janssen MF, Herdman M, van Hout B et al. Comparing the UK EQ-5D-3L and English EQ-5D-5L Value Sets. Pharmacoeconomics. 2018. doi:10.1007/s40273-018-0628-3.
9. Wailoo A, Hernandez Alava M, Grimm S, Pudney S, Gomes M, Sadique Z et al. Comparing the EQ-5D-3L and 5L versions. What are the implications for cost effectiveness estimates? 2017. http://scharr.dept.shef.ac.uk/nicedsu/methods-development/eq-5d-5l/.
10. Hernandez-Alava M, Pudney S. Econometric modelling of multiple self-reports of health states: The switch from EQ-5D-3L to EQ-5D-5L in evaluating drug therapies for rheumatoid arthritis. J Health Econ. 2017;55:139-52. doi:10.1016/j.jhealeco.2017.06.013.
11. Hernandez Alava M, Wailoo A, Pudney S. Methods for mapping between the EQ-5D-5L and the 3L for technology appraisal. . 2017. https://scharr.dept.shef.ac.uk/nicedsu/methods-development/eq-5d-5l/.
12. Hernandez-Alava M PS. eq5dmap: a command for mapping from 3-level to 5-level EQ-5D. The STATA Journal.
13. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J et al. American College of Rheumatology/European League against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Ann Rheum Dis. 2011;70(3):404-13. doi:10.1136/ard.2011.149765.
14. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 2004;13(9):873-84. doi:10.1002/hec.866.
15. Gamst-Klaussen T, Chen G, Lamu AN, Olsen JA. Health state utility instruments compared: inquiring into nonlinearity across EQ-5D-5L, SF-6D, HUI-3 and 15D. Qual Life Res. 2016;25(7):1667-78. doi:10.1007/s11136-015-1212-3.
16. Konerding U, Moock J, Kohlmann T. The classification systems of the EQ-5D, the HUI II and the SF-6D: what do they have in common? Qual Life Res. 2009;18(9):1249-61. doi:10.1007/s11136-009-9525-8.
17. Longworth L, Yang Y, Young T, Mulhern B, Hernandez 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):1-224. doi:10.3310/hta18090.
18. Wee HL, Machin D, Loke WC, Li SC, Cheung YB, Luo N et al. Assessing differences in utility scores: a comparison of four widely used preference-based instruments. Value Health. 2007;10(4):256-65. doi:10.1111/j.1524-4733.2007.00174.x.

Figure captions and legends

Figure 1: Distribution of case studies and NICE appraisals across International Classification of Disease-10 categories

NICE: National Institute for Health and Care Excellence

Figure 2: FORWARD data classified by severity: actual and mapped means of 5L utility, using different bandwidths for mapping from 3L mean utility

3L: EuroQol-5 Dimension-3 Level. FORWARD: National Data Bank for Rheumatic Diseases.

Figure 3: Incremental Quality Adjusted Life Years using 3L and 5L

3L: EuroQol-5 Dimension-3 Level. 5L: EuroQol-5 Dimension-5 Level. EQG: EuroQol Group coordinated study. FORWARD: National Data Bank for Rheumatic Diseases.

Figure 4: Incremental Cost-Effectiveness Ratios using 3L and 5L

3L: EuroQol-5 Dimension-3 Level. 5L: EuroQol-5 Dimension-5 Level. EQG: EuroQol Group coordinated study. FORWARD: National Data Bank for Rheumatic Diseases.