Dixon, P., Dakin, H., & Wordsworth, S. (2016). Generic and disease-specific estimates of quality of life in macular degeneration: Mapping the MacDQoL onto the EQ-5D-3L. *Quality of Life Research*, *25*(4), 935-945. https://doi.org/10.1007/s11136-015-1145-x

Generic and disease-specific estimates of quality of life in macular degeneration: Mapping the MacDQoL onto the EQ-5D-3L

Padraig Dixon[1*], Helen Dakin[2], Sarah Wordsworth[2],

1 School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, United Kingdom

2 Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Old Road Campus, Oxford, OX3 7LF, United Kingdom

* Corresponding author

Email of corresponding author: padraig.dixon@bristol.ac.uk

Telephone number of corresponding author: 00 44 117 928 7319

Keywords: MacDQoL, mapping algorithm, EuroQoL EQ-5D, age-related macular degeneration.

Abstract

Purpose: The Macular Degeneration Quality of Life (MacDQoL) instrument is a validated condition-specific measure of quality of life in patients with macular degeneration. This paper presents the first mapping algorithm to predict EQ-5D from responses to the MacDQoL instrument.

Methods: Responses to the MacDQoL and EQ-5D-3L instruments from 482 patients were collected from the IVAN multicentre trial of two alternative drug treatments for neovascular age-related macular degeneration. Regression specifications were estimated using OLS, censored least absolute deviation, Tobit and two-part models. Their predictive performance was assessed using mean squared error. An internal validation sample based on a random selection of 25% of patients was used to assess the performance of the model estimated on the remaining 75% of patients.

Results: A two-part model had the best predictive performance on the full sample. The covariates of this model include responses and weighted impact scores for all 23 condition specific domains of the MacDQoL, and responses to a general MacDQoL quality of life question. The selected models were successful at predicting means and standard deviations of target populations, but prediction is weaker at the upper and lower extremes of the EQ-5D-3L distribution.

Conclusion: The mapping algorithms provide a means of predicting EQ-5D-3L index scores from MacDQoL scores, and could facilitate cost-effectiveness analyses when the latter but not the former are available to researchers. Further validation of the performance of the algorithms using external data would provide a means of establishing the robustness of the algorithms.

Introduction

There is growing interest from health economists and quality-of-life researchers in understanding the relationship between condition-specific and generic measures of patient outcome, especially where the former but not the latter are collected within clinical trials and other study contexts [1-3]. Condition-specific measures of health-related quality of life (HRQoL) may offer greater sensitivity than generic measures [4]. However, for health care commissioners, generic measures are useful in providing comparable information on effectiveness and cost-effectiveness across different interventions and patient groups.

There is evidence that validity, sensitivity and responsiveness of generic measures differs between diseases [5]. Recent studies have suggested that the responsiveness of generic HRQoL measures in visual disorders may be limited [2; 5] although the evidence base in this area is small. Our paper examines the relationship between generic and condition-specific measures of HRQoL in macular degeneration. Age-related macular degeneration is a progressive chronic eye disease. It is a leading cause of irreversible blindness and visual disability, accounting for approximately 50% of all vision impairments or blind registrations in developed countries [6]. Prevalence is expected to increase significantly in the developing world with demographic change leading to aging populations [7]. In the UK, 39,700 new cases of neovascular macular degeneration are estimated to occur each year, with an estimated prevalence of 263,000 cases nationally [8]. Estimated prevalence at different age groupings in the UK ranges from 0.1% of men and women aged 50-54 to 24.7% of men aged over 90 and 25.6% of women [8]. Given the large number of patients with the condition, it is important to explore the HRQoL improvements

associated with various treatments for the condition and assess how incremental cost-effectiveness compares with interventions for other conditions in order to assess where and how limited healthcare resources should be best deployed.

The MacDQoL (**Mac**ular **D**egeneration **Q**uality of **L**ife) [9] is a macular degeneration-specific outcome measure. The MacDQoL is considered acceptable to patients and is related to several vision measures [10]. Its test-retest reliability has been confirmed [11], as have its metric properties in UK, French, German, Italian and American populations [12]. Translations or adaptations are available in 15 languages [9]. There is currently no mechanism to estimate EQ-5D values from responses to the MacDQoL instrument. This is a potential limitation of the instrument as health state preference values are required to inform cost-effectiveness analyses in technology appraisals and drug reimbursement decisions. While direct elicitation studies require recruitment of a new sample of patients or citizens answering cognitively challenging questions, mapping studies can be conducted on existing datasets. The objective of our paper was to use trial data on patients with age-related macular degeneration to produce a set of coefficients that can be used to reliably transform patients' responses on the MacDQoL questionnaire to EQ-5D-3L index scores. It is understood to the first example of MacDQoL being mapped onto a generic utility instrument. This paper presents mapping algorithms that predict generic EQ-5D-3L index scores from the MacDQoL.

Methods

The data were derived from the IVAN trial [13; 14],a UK factorial randomised controlled trial which assessed ranibizumab versus bevacizumab for the treatment of macular degeneration. Participants aged 50 and above had active but previously untreated neovascular age-related macular degeneration in the study eye, and were randomised to receive ranibizumab or bevacizumab, and either discontinuous treatment or continuous monthly injections for two years. IVAN participants completed both MacDQoL and EQ-5D-3L at three time points.

HRQoL instruments

The MacDQoL consists of two general overview questions and 23 specific questions or domains. The two general items (DQOL-I and DQOL-II) are scored individually and measure generic present quality of life (DQOL-I), and macular degeneration-specific quality of life (DQOL-II) [10]. Each of the 23 specific domains contains an 'impact' question[1] and an 'importance' question[2]. The use of an importance-weighted impact allows for an overall weighted impact of each item to be calculated as the product of an impact rating and an importance rating, with a range from -9 (maximum negative impact of macular degeneration) to +3 (maximum positive impact of macular degeneration). The average weighted impact score is calculated from the sum of weighted ratings specific domains divided by the number of applicable domains, and also has a range from -9 to +3. The use of both individualised impact and importance responses reflects how MacDQoL was designed to capture the quality of life impacts

[1] Impact response categories are answered in response to questions such as 'If I did not have macular degeneration, friendships and social life would be…'. Typical categories are 'very much better' [scoring -3], 'much better' [scoring -2], 'a little better' [scoring -1], 'the same' [scoring 0], and 'worse' [scoring +1].

[2] Importance response categories are answered in response to questions such as 'My friendships and social life are..' with response categories 'very important' [scoring 3], 'important' [scoring 2], 'somewhat important' [scoring 1], 'not at all important' [scoring 0].

of macular degeneration, rather than simply measuring only the functional impact of the condition of visual performance. A more extended discussion of the instrument is presented in [10] and in the online supplementary material.

EQ-5D-3L is a standardised, generic measure of health status that provides a descriptive profile of health states, and a single index value for health status [15]. It is the preferred utility instrument of the National Institute for Health and Care Excellence (NICE) [16]. The index value is anchored on a maximum value of 1 for perfect health, 0 to represent death, and negative values that reflect states worse than death. It contains 5 dimensions: mobility; self-care; usual activities; pain and discomfort; and anxiety and depression. The three level version (EQ-5D-3L) measures these dimensions by three categories, corresponding to no problems, some problems, and extreme problems.

Data The 610 adults participating in IVAN self-completed large print EQ-5D-3L questionnaires in UK English (with assistance from research nurses where necessary) at baseline, at 3, 12 and 24 months, at study exit and after serious adverse events. EQ-5D-3L responses were valued using the UK time trade-off tariff [17].

MacDQoL was administered in English by telephone at 3, 12 and 24 months during the IVAN trial. Data from all time points were eligible for inclusion in the analysis. Four domains of the MacDQoL instrument allowed people to respond ‘Not applicable’ (or similar), instead of answering the impact/importance questions: work, personal relationships, family life, and holidays. These values were recoded to zero (to indicate no impact on quality of life) so as to be retained for analysis in for some of the estimated models; this is described in more detail in the online appendix

(Supplementary Material 1). Observations with missing data on any other MacDQoL or any of the five EQ-5D-3L index questions were excluded from the analysis.

The MacDQoL instrument was due to be administered 14 days after EQ-5D-3L to reduce patient burden but on 1,008 occasions (88% of all instances for which paired data on the date of EQ-5D-3L and MacDQoL are available), it was completed more than 14 days after the EQ-5D, with a mean difference of 41 days later than protocol. Only observations where the MacDQoL was completed no more than 90 days after EQ-5D-3L completion, and no more than 1 day before EQ-5D-3L was due to be administered, were used in our mapping analysis in order to ensure consistency between the EQ-5D-3L and the MacDQoL instruments. Sensitivity analyses varying this cut-off, and further details of this issue, are reported in the online appendix (Supplementary Material 1). After excluding patients who did not complete both the EQ-5D-3L and MacDQoL within the specified timeframe, and the 20 observations with missing data on one or more EQ-5D-3L or MacDQoL questions, the regression analyses were estimated with up to 817 observations from 462 patients (70% of those randomised) out of 858 MacDQoL responses (95% of all completed MacDQoL questionnaires included in the analysis).

Statistical methods

The objective of the analysis was to estimate mean EQ-5D-3L values from participant responses to the MacDQoL instrument. To assess the degree of overlap between the two instruments, we first estimated Spearman's rho (rank correlation coefficients) showing the correlations between EQ-5D-3L and DQOL-I, DQOL-II and the MacDQoL using the `-spearman-` Stata command.

Regression models using different specifications and different estimators were applied to the data in order to identify models that predicted EQ-5D-3L successfully,

as determined by mean squared error (MSE) and mean absolute error (MAE). There is no clear guidance as to which of MSE and MAE is preferable [18]. MSE was used as the primary measure here because it penalises deviations from the mean more heavily than does the MAE. Four estimators that have been used in previous mapping studies [3] were applied and their predictive performance compared: ordinary least squares (OLS); Tobit; centred least absolute deviations (CLAD); and two-part models.

The starting point for this analysis was the estimation of linear OLS models. Linear models are widely used in mapping studies [3], and offer a useful basis for comparison with more complex models. They also retain the possibility of being predictively successful in their own right. OLS predictions of EQ-5D-3L values >1 were recoded to 1. A criticism of OLS models is that they do not reflect the natural 'ceiling' of EQ-5D-3L values at 1, and require this ex-post recoding adjustment. Alternative estimators that account directly for the upper limit on EQ-5D-3L were therefore also assessed.

Tobit models [19] allow for continuous dependent variables that have constrained ranges, and therefore can allow for the censoring of EQ-5D-3L at 1. The model estimates, using maximum likelihood, a continuous distribution for values of EQ-5D-3L index scores below the ceiling value of 1, while placing a positive (discrete) probability value on the censored outcome of EQ-5D=1. The rationale for applying Tobit is to allow the estimator to reflect the natural ceiling of EQ-5D-3L at perfect health (when EQ-5D-3L=1) rather than to model the censoring to which Tobit models are usually applied [19]. Pullenayegum et al [20] note that Tobit (and CLAD – see below) models may be biased, when the intention is to perform economic evaluation, if it is assumed that 'true' utility extends beyond 1. However, any potential bias of an estimator is of secondary concern in the context of this study, where we are aiming to

accurately predict mean EQ-5D-3L index scores, rather than to produce, for example, unbiased estimates of covariates that are believed to be causally associated with quality of life.

We also estimated the CLAD models [21] using the –`clad-` command in Stata [22]. CLAD models, which minimise absolute deviations from the median, are similar to Tobit models in their capacity to account for the upper limit in EQ-5D-3L as a dependent variable in regressions, but differ from Tobit models in being consistent even if error terms are not normal and having standard errors that are robust to heteroskedasticity.

The final set of estimators comprised two-part models. These models account for the upper limit of EQ-5D-3L by separately modelling individuals at the upper limit of EQ-5D, and modelling those individuals below this limit. Unlike Tobit, CLAD and OLS models, the two-part model uses two different estimators to produce a single estimate. Specifically, two-part models estimate in the binary first 'part' a logistic regression to predict the probability of having an EQ-5D-3L score equal to exactly 1 (Probability(Utility = 1)), and in the second part use an OLS model to predict utility (U) for those individuals with less than perfect health (i.e. EQ-5D-3L scores <1). The second part of the two-part model therefore regresses the EQ-5D score on the same or a closely related set of covariates as used in the first part on those patients with less than perfect health. For the OLS part of the two-part model model, predictions of EQ-5D-3L of greater than 1 were again recoded to 1. Predictions are produced by combining the predictions of the first-part logit model with those of the second-part OLS model by weighting the OLS prediction by the probability of having imperfect health using the following formula:

$$\text{Predicted Utility} = \text{Probability(Utility=1)} + (1 - \text{Probability(Utility=1)})*U$$

189 Seemingly unrelated regression (SUR), implemented with the –`suest-` command in
190 Stata, was used to adjust the standard errors of the two-part model to allow for
191 correlation between the error terms of the two parts of the model.

192 For each estimator, we explored the performance of a number of different model
193 specifications with using different scores or responses from MacDQoL. We explored
194 some models in which each MacDQoL domain was coded using one ordinal variable
195 measuring response level. We also evaluated whether weighted impact scores (the
196 product of impact and importance) improved EQ-5D-3L predictions. However, the
197 latter two specifications treat the MacDQoL responses as though they were
198 continuous, implicitly assuming that the levels between response levels are evenly
199 spaced. We therefore also explored the impact of using dummy variables for different
200 levels of the ordinal response variables, which recognises that the variable has a
201 categorical nature. Further specifications considered the effect on predictive
202 accuracy of recoding or excluding questions with an 'N/A' response.

203 MacDQoL impact variables had limited numbers of responses in some of the most
204 extreme categories, which meant that econometric models including dummy
205 variables for all response levels dropped variables. To address this issue, responses
206 were recoded for models using dummy variables so that the most extreme responses
207 were collapsed into adjacent categories: the 'Very much better' and 'Much better'
208 categories were merged, as were the 'Worse' and 'The same categories'. Only 14
209 MacDQoL responses (1.6%) indicated that macular degeneration *improved* quality of
210 life in any domain. Estimating models without merging these categories therefore
211 tended to produce collinearity and could not be estimated (see online supplementary
212 material). It may be worth investigating this issue in the future if a much larger
213 dataset becomes available for analysis. The recoded domain responses used in the
214 analysis therefore had three levels.

We excluded from consideration model specifications that did not have convergent likelihood, which excluded large parts of the sample (because of low levels of data on one variable) and/or which dropped variables because of perfect collinearity. Coefficients with $p>0.05$ and/or with counterintuitive signs were retained in model specifications given that the primary objective of the estimated models was to maximise prediction accuracy.

Twenty two different model specifications were evaluated for each of the four different estimators, producing up to 88 MSE values. A list of combinations of model specification and estimator is presented in Tables 4 and 5 of the online appendix. The appropriateness of model specification was determined by reference to predictive accuracy. The predictive accuracy of the chosen specification were also assessed by examining Q-Q plots of actual versus predicted EQ-5D index scores, scatter plots of actual versus predicted scores, and simple distributional comparisons (see online supplementary material).

All models were estimated using Stata version 13.1 (Statacorp: College Station, Texas) using clustering to adjust standard errors to allow for repeated observations made on individual patients.

Validation

Ideally, the prediction accuracy of a mapping algorithm should be tested on an external dataset to assess whether the model is 'over-fitted' to noise in the data rather than the underlying relationship, and to evaluate how the model will perform in other datasets. External data were not available in this case, so an internal validation sample was used [1]. This was conducted by using the (pseudo)-random number generating function in Stata to select a 25% 'validation' sample (n=115) leaving an

239 ‘estimation’ sample with the remaining 75% of data (n=367). This percentage split is
240 typical of that used in other studies [23].

Results

Summary statistics

Participants in the IVAN trial had a mean age of 77.7 years, of whom 60% were female. Participants needed a best corrected distance of visual acuity of at least 25 letters on the Early Treatment Diabetic Retinopathy Study chart to be eligible for inclusion, and the mean at baseline was 61.4 letters [13]. Some 14% of individuals randomised had angina, 19% dyspnoea and less than 10% had experienced one or both of myocardial infarction or stroke; 64% were current or past smokers [13]. Table 1 presents descriptive statistics for the key variables used in the analysis for this paper for the 482 patients with no missing data on EQ-5D-3L and MacDQoL domains used in regression analysis. Further information on the demographic characteristics of participants in the IVAN trial is available in [13].

Table 1 Summary statistics of HRQoL instruments

	Full sample (N=860 observations from 482 patients)	Estimation sample (N=655 observations from 367 patients)	Validation sample (N=205 observations from 115 patients)
EQ-5D: mean (SD)[a]	0.8417 (0.1960)	0.8459 (0.1912)	0.8282 (0.2108)
	Min: -0.181 Max: 1	Min: -0.181 Max: 1	Min: -0.077 Max: 1
MacDQoL weighted impact: mean (SD)	-1.8801 (1.7902)	-1.9670 (1.8352)	-1.6005 (1.61003)
	Min: -9 Max: 0.1363	Min:-9 Max:0.1363	Min: -7.318 Max: 0
DQOL-I: mean (SD)	1.3934 (0.9947)	1.4167 (1.0115)	1.3186 (0.9371)
	Min: -3 Max: 3	Min: -3 Max: 3	Min:-1 Max: 3
DQOL-II: mean (SD)	-1.4639 (1.0393)	-1.4984 (1.0113)	-1.3529 (1.1198)
	Min: -3 Max: 1	Min: -3 Max: 1	Min: -3 Max: 1

a – SD: Standard deviation.

Some 21% of responses to the DQOL-I question (macular degeneration-specific quality of life) indicated no negative impact of macular degeneration on HRQoL. EQ-5D-3L and DQOL-I, DQOL-II and the MacDQoL average weighted impact measure were highly correlated under Spearman's rho, and the null of independence was

strongly rejected in all cases ($p<0.01$). Lower EQ-5D-3L scores were correlated with lower MacDQoL estimates of quality of life. There was a small cluster of observations with low EQ-5D-3L scores: 0.58% of observations were below 0, and 3.95% were below 0.5.

Mapping MacDQoL onto EQ-5D-3L

The best-performing model differed depending on whether predictive accuracy was assessed using the estimation sample or the validation sample (Supplementary material 1, Tables 4 and 5).

The model having the lowest MSE in the estimation sample (Model 4 – see Supplementary material for numbering) was a two-part model that included dummy variables for all 23 MacDQoL domains and continuous interactions of impact and importance for all of these domains in the first (logit) part of the model, while the second (OLS) part of the model included the same set of covariates, plus the generic quality of life measure DQOL-I (treated as a continuous variable).

The model with the lowest MSE in the validation sample (Model 5) was an OLS model estimated on all 23 MacDQoL domains treated as dummy variables, and continuous interactions of impact and importance for all of these domains. This is the same set of covariates as the first part of the two-part model selected in the estimation sample.

When these models were estimated on the full sample of 817 observations and predictions generated, the two-part Model 4 had a better predictive performance than the OLS Model 5. Generally, it is more appropriate to select models based on their performance in the validation sample to avoid over-fitting and typically the model that performs best in the validation sample also performs best on the full sample. However, in this case, results differ markedly because the validation sample has

lower mean EQ-5D-3L scores and prediction errors are especially bad for such patients. In the validation sample, 9.28% of observations had EQ-5D-3L <0.6, compared with 6.42% in the estimation sample. Differences in the distribution of EQ-5D-3L between the two samples may have affected the choice of model and estimator.

Table 2 Summary comparison of observed and predicted EQ-5D-3L on full, estimation and validation samples

Variable	N[a]	Mean	Standard deviation	Minimum	Maximum	MSE[d]	MAE[d]
Observed data							
Observed EQ-5D-3L – full regression sample	817	0.8441	0.1937	-0.181	1		
Observed EQ-5D-3L – estimation sample	623	0.8488	0.1886	-0.181	1		
Observed EQ-5D-3L – validation sample	194	0.8290	0.2091	-0.077	1		
Predictions – two-part Model 4[b]							
Predicted EQ-5D-3L from two-part model	817	0.8481	0.0844	0.4037	1	**0.0293**	**0.1269**
Predicted EQ-5D-3L from two-part model – estimation sample	623	0.8523	0.0944	0.2089	1	**0.0256**	**0.1178**
Predicted EQ-5D-3L from two-part – validation sample	194	0.8668	0.0924	0.4511	1	0.0500	**0.1631**
Predictions - OLS Model 5[c]							
Predicted EQ-5D-3L from OLS model – full sample	817	0.8436	0.0785	0.4837	1	0.0310	0.1329
Predicted EQ-5D-3L from OLS model – estimation sample	623	0.8480	0.0880	0.3317	1	0.0273	0.1245
Predicted EQ-5D-3L from OLS model – validation sample	194	0.8619	0.0884	0.4235	1	**0.0494**	0.1642

a – Number of observations: Of the 482 patients included in any regression analysis, 20 were excluded from the OLS and two-part models shown here because of missing data in EQ-5D-3L scores and MacDQoL domains variables used in regressions. This reduced the number of available observations by 43 from 860 to 817 for the largest regression samples.
b - Specification of first part of Model 4: A logistic regression to predict EQ-5D=1 where covariates comprised all 23 MacDQoL domain responses as dummy variables, and all 23 weighted impact variable values as continuous variables. Specification of second part of Model 4: An OLS regression of EQ-5D-3L values for those patients with EQ-5D index scores<1, using the same explanatory variables the first part, plus the DQOL-I variable as a continuous variable.
c - Specification of Model 5: An OLS regression of EQ-5D-3L values in which explanatory variables comprised all 23 MacDQoL domains treated as dummy variables, and continuous interactions of impact and importance for all of these domains.
d - Values shown in bold are the lowest in their respective samples of those estimated and displayed in the table.

We therefore took the conservative approach of presenting both models: i.e. those informed by predictive accuracy of models in both the estimation sample (Model 4) and the validation sample (Model 5). Variance-covariance matrices are available as online appendix (Supplementary Material 2) and can be used to estimate standard errors around predicted values when the mapping algorithm is applied to other samples.

Table 3 Coefficients to predict EQ-5D-3L from MacDQoL estimated using the full sample (n=817)

Variable		Two-part Model 4		OLS Model 5
		Logit coefficients (SE)	OLS coefficients (SE)	Coefficients (SE)
Household tasks[a]	A little better	-0.0035 (0.3125)	-0.0353 (0.0281)	-0.0200 (0.0239)
	Much better or very much better	0.2789 (0.4947)	-0.0223 (0.0419)	-0.0092 (0.0366)
Personal affairs[a]	A little better	-0.7794 (0.3088)	0.0148 (0.0285)	-0.0333 (0.026)
	Much better or very much better	-1.0097 (0.5231)	0.0469 (0.0427)	-0.0107 (0.0413)
Shopping[a]	A little better	0.4939 (0.2855)	-0.0113 (0.0253)	0.0256 (0.023)
	Much better or very much better	0.2746 (0.4576)	-0.0544 (0.0383)	-0.0022 (0.0337)
Work[a]	A little better	-0.2048 (0.7607)	0.0168 (0.0844)	-0.0003 (0.0583)
	Much better or very much better	-1.5117 (1.3659)	0.0189 (0.0948)	-0.0610 (0.08)
Relationships[a]	A little better	0.2794 (0.5866)	-0.0462 (0.0664)	-0.0283 (0.0495)
	Much better or very much better	0.6153 (1.0143)	-0.0321 (0.1257)	0.0128 (0.0902)
Family life[a]	A little better	-0.5902 (0.4357)	-0.0655 (0.0374)	-0.063 (0.0323)
	Much better or very much better	-1.0378 (0.7979)	-0.1384 (0.0726)	-0.1320 (0.0685)
Friendships[a]	A little better	-0.137 (0.346)	-0.0014 (0.0279)	-0.0078 (0.0265)
	Much better or very much better	0.6822 (0.6614)	-0.044 (0.0495)	0.0105 (0.0481)
Physical appearance[a]	A little better	-0.4769 (0.4034)	0.0245 (0.0328)	-0.0045 (0.0291)
	Much better or very much better	-1.0209 (0.8434)	0.0102 (0.0678)	-0.0381 (0.0566)
Physical activity[a]	A little more	-0.2113 (0.3348)	0.0543 (0.0316)	0.0159 (0.0274)
	Much more or very much more	-0.9505 (0.555)	0.0316 (0.0541)	-0.0291 (0.0488)
Out and about[a]	A little better	-0.7453 (0.3237)	-0.0325 (0.0313)	-0.0567 (0.0282)
	Much better or very much better	-1.5539 (0.5259)	0.0177 (0.0463)	-0.0819 (0.0484)
Holidays[a]	A little better	0.1084 (0.2988)	0.015 (0.0281)	0.0186 (0.0231)
	Much better or very much better	-1.035 (0.4759)	0 (0.0401)	-0.0198 (0.0339)
Leisure and hobbies[a]	A little more	0.0764 (0.2922)	-0.021 (0.0295)	-0.0029 (0.025)
	Much more or very much more	0.1379 (0.4243)	-0.0296 (0.0402)	-0.0176 (0.0347)
Self-confidence[a]	A little better	0.5343 (0.3112)	-0.0033 (0.0271)	0.0171 (0.0235)
	Much better or very much better	1.2402 (0.5675)	-0.0053 (0.0477)	0.0538 (0.0404)
Motivation[a]	A little better	0.1446 (0.3171)	0.0535 (0.0263)	0.0363 (0.0239)
	Much better or very much better	0.4464 (0.5721)	0.0581 (0.049)	0.0514 (0.0438)
	A little better	-0.1992 (0.4257)	0.0740 (0.0389)	0.0454 (0.0366)

Reaction of others[a]	Much better or very much better	0.32 (0.7181)	0.0354 (0.0776)	0.061 (0.073)
Feelings about the future[a]	A little better	0.1088 (0.2805)	-0.0314 (0.0246)	-0.0098 (0.0235)
	Much better or very much better	0.0033 (0.3993)	0.004 (0.038)	0.0063 (0.0323)
Financial situation[a]	A little better	0.0984 (0.7161)	-0.0346 (0.0441)	-0.0164 (0.047)
	Much better or very much better	-0.207 (1.6711)	-0.1145 (0.0996)	-0.1115 (0.1009)
Independence[a]	A little better	0.3108 (0.3318)	0.0263 (0.0281)	0.0275 (0.0274)
	Much better or very much better	0.3307 (0.5383)	0.0746 (0.0523)	0.0517 (0.0461)
Doing things for others[a]	A little better	-0.2579 (0.2987)	0.0457 (0.029)	0.0024 (0.0254)
	Much better or very much better	-0.2808 (0.5)	-0.0079 (0.0495)	-0.0312 (0.0419)
Experiencing mishaps or losing things[a]	A little better	0.234 (0.3211)	-0.0607 (0.0299)	-0.0171 (0.0277)
	Much better or very much better	0.9399 (0.6205)	-0.0512 (0.0684)	0.0115 (0.0594)
Enjoyment of meals[a]	A little more	-0.2301 (0.3928)	0.0181 (0.0254)	0.0067 (0.0292)
	Much more or very much more	1.0081 (0.647)	0.0185 (0.0606)	0.0656 (0.0576)
Time taken to do things[a]	A little less	-0.3971 (0.244)	0.0033 (0.0298)	-0.0264 (0.0223)
	Much less or very much less	-0.1084 (0.4568)	0.0401 (0.0419)	0.0091 (0.0359)
Enjoyment of nature[a]	A little more	-0.0565 (0.3128)	0.0027 (0.0251)	0.0001 (0.0231)
	Much more or very much more	-1.0278 (0.4949)	0.0114 (0.0313)	-0.038 (0.0358)
Impact/importance interaction	Household tasks	-0.0378 (0.0816)	-0.0102 (0.0071)	-0.0102 (0.0063)
	Personal affairs	-0.1396 (0.0753)	0.0024 (0.0066)	-0.0022 (0.0062)
	Shopping	0.056 (0.0679)	-0.0049 (0.0058)	0.0016 (0.005)
	Work	-0.3265 (0.228)	0.0084 (0.0221)	-0.0116 (0.0161)
	Relationships	-0.0583 (0.1677)	-0.0129 (0.0204)	-0.0106 (0.014)
	Family life	-0.1342 (0.1198)	-0.0196 (0.0122)	-0.0146 (0.0103)
	Friendships	0.0432 (0.0971)	-0.0116 (0.008)	-0.0028 (0.0076)
	Physical appearance	-0.0846 (0.1306)	0.0022 (0.0107)	-0.0015 (0.0089)
	Physical activities	-0.0975 (0.0946)	-0.0006 (0.0096)	-0.0045 (0.0086)
	Out and about	-0.2568 (0.0768)	0.0074 (0.008)	-0.0105 (0.0079)
	Holiday	-0.1742 (0.0903)	-0.0038 (0.0088)	-0.0098 (0.007)
	Leisure	0.0469 (0.0628)	-0.002 (0.0055)	-0.0006 (0.0048)
	Self-confidence	0.2509 (0.0887)	-0.0016 (0.0075)	0.0107 (0.0067)
	Motivation	-0.0764 (0.088)	0.0153 (0.0079)	0.0043 (0.0073)
	Reaction of others	0.2273 (0.1279)	0.0221 (0.017)	0.0312 (0.0159)
	Feelings about the future	-0.0212 (0.0582)	-0.0032 (0.0058)	-0.0036 (0.0046)
	Financial situation	0.1248 (0.2599)	-0.0151 (0.015)	-0.0084 (0.0158)
	Independence	0.1857 (0.0734)	0.0111 (0.0073)	0.0162 (0.0068)

	Do for others	-0.1099 (0.0792)	-0.0013 (0.0075)	-0.0072 (0.0066)
	Mishaps	0.0835 (0.0957)	-0.0055 (0.0116)	0.0008 (0.0101)
	Enjoyment of meals	0.1699 (0.0946)	0.0106 (0.0119)	0.0163 (0.011)
	Time taken to do things	0.0215 (0.0775)	0.0028 (0.0072)	0.0022 (0.0065)
	Enjoyment of nature	-0.0596 (0.0746)	0.0036 (0.0052)	-0.0005 (0.0055)
DQOL-I			0.0358 (0.0105)	
Constant		0.2991 (0.1705)	0.6701 (0.0264)	0.8781 (0.0146)
Pseudo R^2 [b]		0.1363		
Adjusted R^2 [b]			0.0450	0.0940
MSE (full sample)			0.0293	0.0310
MAE (full sample)			0.1269	0.1329

a: All 23 items in the MacDQoL instrument begin with 'If I did not have macular degeneration…' The respondent is then asked to consider the impact of the condition on the activities described above and rate it on a five-point scale (worse, same, a little better, much better, or very much better). The two extreme categories have been combined with the second most extreme. The reference response to these questions is 'Same or worse' – meaning that the variables should be interpreted against a patient response that the presence of macular degeneration has no impact on that specific domain or their actions in that domain would be worse without macular degeneration.

b:R^2 values are those associated with the respective regressions used to generate these coefficients – pseudo R^2 in the case of the logit regression and adjusted R^2 in the case of the OLS regressions.

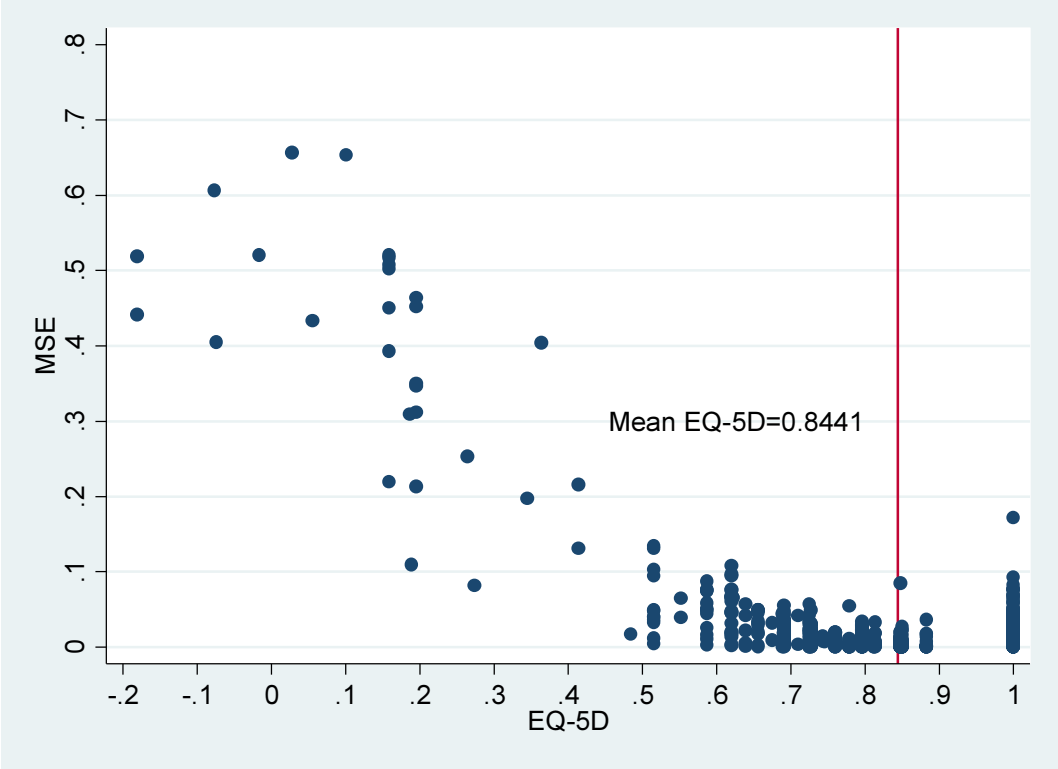
For the OLS model, EQ-5D-3L values can be calculated from the coefficients in this table by taking the constant term (0.8781) and adding the coefficients for the relevant item responses (e.g. adding -0.02 if the patient would find household tasks a little better if they did not have macular degeneration), then, for each domain, adding on the product of the coefficient for the impact/importance interaction (e.g.-0.0378 for household tasks), that patient's impact score on that domain (e.g. 2 if they would find household tasks a little better) and their importance score (e.g. 3 if household tasks are very important to them). For both OLS and two-part models, the impact scores used to calculate weighted impact should be Worse=0, The same=1, A little better=2, Much better=3, Very much better=4, and for Importance: 0=Not at all important, 1= Somewhat important, 2=Important, 3=Very important. EQ-5D-3L values can be calculated in a similar way for the two-part model. The procedure for the OLS component of the two-part model (U) is as described for the OLS model but using coefficients from the second column of data in the table above. The log odds of having perfect health (utility = 1) can be calculated by taking the constant term (0.2991) and adding on the coefficients for the relevant impact item responses and the product of the coefficient for weighted impact scores, that patient's impact score and their importance score. Log-odds should then be converted to probabilities by taking the exponent of the log odds to calculate the odds of a utility of 1 and dividing the odds by 1+odds to give Probability(Utility=1). Predicted utility can then be calculated as:
Predicted Utility = Probability(Utility=1) + (1 – Probability(Utility=1))*U

The purpose of the regressions is to predict EQ-5D-3L index scores rather than estimate the magnitude and sign of coefficients, but the following example shows how these coefficients may be interpreted. The 'Personal affairs' question asks the respondent to state 'If I did not have [macular degeneration], I could handle my personal affairs…'. Using the recoded variables deployed in the analysis, the coefficient of the OLS regression on the 'Better or very much better' dummy of the 'Personal Affairs' question is -0.0107, which can be interpreted as suggesting that

patients who could manage their personal affairs much better or very much better would have approximately 1% lower EQ-5D-3L scores than patients for whom macular degeneration had no impact on management of their personal affairs. The coefficients cannot be given a causal interpretation, and in any event there is no clear pattern amongst the coefficients that indicate which domains of MacDQoL might have a particular influence on EQ-5D index scores. In the OLS models, coefficients including 'family life' impacts have p-values <0.05, and likewise the 'personal affairs' impact domain in the logit model. Several coefficients (e.g. shopping on the two-part logit model) have counterintuitive signs, suggesting that EQ-5D-3L scores would be improved if macular degeneration had a larger impact on that domain. However, all variables were kept in the model to avoid reducing prediction accuracy by omitting parts of the source instrument.

Prediction errors were high at the extremes of the EQ-5D-3L distribution, particularly the lower end, while errors were smaller near mean EQ-5D-3L: MSE is <0.1 for values within ±0.15 of mean EQ-5D-3L (Figure 1). This suggests that the accuracy of prediction, as measured by MSE, is smallest when predicting mean EQ-5D-3L compared to other quantiles of EQ-5D-3L. This suggests that both algorithms are better suited to prediction of mean EQ-5D-3L from MacDQoL than for values far from mean EQ-5D-3L index scores.

368 **Fig. 1 MSE at different levels of EQ-5D-3L – full sample, two-part Model 4**



369

Discussion

This paper has presented the methods and results of a mapping algorithm that generates EQ-5D-3L values from responses to the condition-specific MacDQoL instrument. The mapping algorithm can be used to estimate EQ-5D-3L index values in circumstances where only MacDQoL data is available – this circumstance may arise in study designs where the latter but not the former has been collected by researchers. Although MacDQoL is not widely used at present, this may change with increasing research on eye disease. Estimated models are successful at predicting the mean EQ-5D-3L for this specific sample, but may be less successful for predicting individuals' utility (particularly for those with low or very high utility). We found that the predictive performance of the studied models is weaker for patients with EQ-5D-3L values <0.6 than for values≥0.6. A similar trend has been encountered in other literature concerned with eye disease and HRQoL [24]. For cost-effectiveness analysis, accurate prediction of sample means may be more important than the scores of individuals at particular parts of the utility distribution [25].

As this is the first study mapping between these instruments, it was not possible to compare the predictive performance against other published studies, and it is important to test the algorithm presented here on other datasets when they become available. A review of mapping studies in other areas [26] found MAE estimates to range between 0.0011 to 0.19, which encompasses the MAEs found in this study for the selected models (0.1269 for the two-part model and 0.1329 for the OLS model). The only other study mapping predicting EQ-5D-3L for patients with macular degeneration identified in a recent systematic review [3] reported a root mean

square error of 0.2163 for its recommended model when mapping from the National Eye Institute Vision Function Questionnaire [27]. The square of this term is 0.0468, which is somewhat larger than the MSEs for preferred models in the different samples (Models 4 and 5) of 0.0293 and 0.0310, although comparisons are affected by the different samples and instruments involved.

Our best performing models included some coefficients that were non-significant and/or had counterintuitive signs. This has also been observed in previous studies, including those with much larger samples [28]. Such counterintuitive signs could cause the mapping algorithm to predict a reduction in EQ-5D when the patient's MacDQoL profile has improved and could reflect collinearity within our sample or genuine opposition between domains. We suggest that it is appropriate to include the whole of the source instrument for completeness, and because omitting inconsistent or non-significant coefficients can reduce prediction accuracy [28]. Selecting models based on information criteria may have given a more parsimonious model, but are not appropriate for selecting mapping models, where prediction accuracy is the main concern [1; 29; 30].

Participants in IVAN had mean baseline EQ-5D-3L index scores of 0.85, which is higher than estimates of UK age-adjusted mean population norms [31], potentially because of the trial inclusion criteria and the requirement for monthly visits [32]. This is a limitation of the study if the relationship between the two instruments varies with patients' general health. If slope coefficients were to change when estimated on a population in worse health, this would affect the level of QALYs estimated, but not necessarily affect estimates of incremental QALYs or incremental cost-effectiveness. Further validation of the mapping models on an external dataset is required to assess prediction accuracy in other populations.

The choice of estimator differed depending on whether model selection was based on MSE in the validation sample or MSE in the estimation sample. This could be explained by the greater proportion of individuals with low EQ-5D-3L scores in the validation sample. An external validation dataset would provide a more robust test of the model's overall predictive accuracy, and may offer evidence regarding the applicability of the algorithm to other types of patient population.

The sample size of this study is a limitation. A larger sample size would improve the precision of estimates, and may address the differences between validation and estimation samples as well as the small number of instances of counter-intuitive signs of particular variables in the preferred model specification. Larger sample sizes may also support the application of other potentially relevant approaches such as response mapping [33]. The number of observations available used in this analysis (817) is below the median (1,167) and mean (6,069) number of observations in a recent review [3] of mapping studies.

MacDQoL, like other 'DQoL' instruments, asks patients to rate the importance of each domain, as well as rating the impact of their condition. We found that the product of importance and impact affected EQ-5D utility over and above the effect of impact scores. An unexplored issue is whether it is reasonable to map from a disease-specific instrument that uses these patient-reported measures of importance to a generic instrument (EQ-5D) that uses valuations of preferences between health states provided by members of the general public. We note that NICE guidance for technology appraisal [16] recommends the use of a generic instrument valued using public (not patient) preferences, but does not specify the characteristics that ought to be possessed by disease-specific instruments from which EQ-5D index values are obtained via mapping algorithms.

Conclusion

Our paper presents the first set of algorithms to map from MacDQoL to health state values, which will facilitate cost-effectiveness studies in this area. The models are reasonably simple, in that any dataset with complete responses to MacDQoL can be used to predict EQ-5D, with no additional data required. The methods described are also likely to be applicable to similar analyses in other disease areas. The models presented in this paper can be used to estimate mean EQ-5D-3L values in other samples. However, our models have been evaluated only on patients with neovascular age-related macular degeneration with MacDQoL average weighted impact scores between -9 and 0.14, with a median value of -1.36, and future work is required to assess whether our models would perform as well in other patient groups.

Acknowledgements and compliance with ethical standards

Acknowledgements: The IVAN trial (Trial registration: ISRCTN92166560) was funded by the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme (project number 07/36/01). The views and opinions expressed are those of the authors and do not necessarily reflect those of the HTA programme, NIHR, the UK National Health Service or the Department of Health. The NIHR had no role in the design, conduct or reporting of this mapping study.

We are grateful to all IVAN trial participants and to the IVAN research team. We are extremely grateful to the IVAN site staff, who collected the EQ-5D-3L data and also the Clinical Trials Evaluation Unit (University of Bristol) who made all the phone calls to collect the MacDQoL data, then managed and cleaned the data for the analyses for this paper.

We thank two anonymous reviewers for helpful comments.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Written, informed consent was obtained from all individual participants included in the study.

References

1. Longworth, L., & Rowen, D. (2013). Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. Value in Health, 16(1), 202-210.
2. Tosh, J., Brazier, J., Evans, P., & Longworth, L. (2012). A review of generic preference-based measures of health-related quality of life in visual disorders. Value in Health, 15(100883818), 118-127.
3. Dakin, H. (2013). Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. Health and Quality of Life Outcomes, 11, 6.
4. Brazier, J., & Dixon, S. (1995). The use of condition specific outcome measures in economic appraisal. Health Economics, 4(4), 255-264.
5. Longworth, L., Yang, Y., Young, T., Mulhern, B., Hernández Alava, M., Mukuria, C., Rowen, D., Tosh, J., Tsuchiya, A., Evans, P., Devianee Keetharuth, A., & Brazier, J. (2014). 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, 18(9).
6. Usha, C., Jennifer, E., & Philip, J. R. (2010). Age related macular degeneration. BMJ, 340.
7. Woo, J. H., Sanjay, S., & Au Eong, K.-G. (2009). The epidemiology of age-related macular degeneration in the Indian subcontinent. Acta Ophthalmologica, 87(3), 262-269.
8. Owen, C. G., Jarrar, Z., Wormald, R., Cook, D. G., Fletcher, A. E., & Rudnicka, A. R. (2012). The estimated prevalence and incidence of late stage age related macular degeneration in the UK. British Journal of Ophthalmology.
9. Mitchell, J., & Bradley, C. (2004). Design of an individualised measure of the impact of macular disease on quality of life (the MacDQoL). Quality of Life Research, 13(6), 1163-1175.
10. Mitchell, J., Wolffsohn, J., Woodcock, A., Anderson, S., McMillan, C., Rubinstein, M., Amoaku, W., & Bradley, C. (2005). + Psychometric evaluation of the MacDQoL individualised measure of the impact of macular degeneration on quality of life. Health and quality of life outcomes, 3(1), 25.
11. Mitchell, J., Wolffsohn, J., Woodcock, A., Anderson, S. J., Ffytche, T., Rubinstein, M., Amoaku, W., & Bradley, C. (2008). The MacDQoL Individualized Measure of the Impact of Macular Degeneration on Quality of Life: Reliability and Responsiveness. American Journal of Ophthalmology, 146(3), 447-454.e442.
12. G., B., Mesbah, M., & Bradley, C. (2011). Metric properties of the MacDQoL, individualized macular-disease-specific quality of life instrument, and newly identified subscales in French, German, Italian, and American populations. Value in Health, 1(1524-4733 (Electronic)), 10.
13. Chakravarthy, U., Harding, S. P., Rogers, C. A., Downes, S. M., Lotery, A. J., Wordsworth, S., & Reeves, B. C. (2012). Ranibizumab versus Bevacizumab to Treat Neovascular Age-related Macular Degeneration: One-Year Findings from the IVAN Randomized Trial. Ophthalmology, 119(7), 1399-1411.
14. Chakravarthy, U., Harding, S. P., Rogers, C. A., Downes, S. M., Lotery, A. J., Culliford, L. A., & Reeves, B. C. (2013). Alternative treatments to inhibit VEGF in age-related choroidal neovascularisation: 2-year findings of the IVAN randomised controlled trial. The Lancet.
15. van Reenen, M., & Oppe, M. (2015). EQ-5D-3L User Guide: Basic information on how to use the EQ-5D-3L instrument Version 5.1. Rotterdam.
16. NICE. (2013). Guide to the methods of technology appraisal. Manchester.

17. Dolan, P. (1997). Modeling Valuations for EuroQol Health States. Medical Care, 35(11), 12.
18. Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., & Gray, A. (2015). Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. Quality of Life Research, 1-7.
19. Wooldridge, J. M. (2009). Introductory Econometrics: A Modern Approach: A Modern Approach: South Western, Cengage Learning.
20. Pullenayegum, E. M., Tarride, J.-E., Xie, F., Goeree, R., Gerstein, H. C., & O'Reilly, D. (2010). Analysis of Health Utility Data When Some Subjects Attain the Upper Bound of 1: Are Tobit and CLAD Models Appropriate? Value in Health, 13(4), 487-494.
21. Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. Journal of Econometrics, 25(3), 303-325.
22. Jolliffe D, Krushelnytskyy B, & Semykina, A. (2000). Censored least absolute deviations estimator: CLAD. Stata Technical Bulletin, STB-58(November).
23. Longworth, L., & Rowen, D. (2011). NICE DSU Technical Support Document 10: The Use of Mapping Methods to Estimate Health State Utility Values. Decision Support Unit, ScHARR, University of Sheffield.
24. Browne, C., Brazier, J., Carlton, J., Alavi, Y., & Jofre-Bonet, M. (2012). Estimating quality-adjusted life years from patient-reported visual functioning. Eye, 26(10), 1295-1301.
25. Sengupta, N., Nichol, M. B., Wu, J., & Globe, D. (2004). Mapping the SF-12 to the HUI3 and VAS in a Managed Care Population. Medical Care, 42(9), 927-937.
26. Brazier, J., Yang, Y., Tsuchiya, A., & Rowen, D. (2010). A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. The European Journal of Health Economics, 11(2), 215-225.
27. Payakachat, N., Summers, K., Pleil, A., Murawski, M., Thomas, J., Jennings, K., & Anderson, J. (2009). Predicting EQ-5D utility scores from the 25-item National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients with age-related macular degeneration. Quality of Life Research, 18(9210257, bqm), 801-813.
28. Dakin, H., Gray, A., & Murray, D. (2013). Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. Quality of Life Research, 22(3), 683-694.
29. Brazier, J., Connell, J., Papaioannou, D., Mukuria, C., Mulhern, B., Peasgood, T., Lloyd Jones, M., Paisley, S., O'Cathain, A., Barkham, M., Knapp, M., Byford, S., Gilbody, S., & Parry, G. (2014). A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol Assess, 18(34).
30. Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., & Gray, A. (2015). Preferred Reporting Items for Studies Mapping onto Preference-Based Outcome Measures: The MAPS Statement. PharmacoEconomics, 1-7.
31. Ara, R., & Brazier, J. E. (2010). Populating an Economic Model with Health State Utility Values: Moving toward Better Practice. Value in Health, 13(5), 509-518.
32. IVAN Study Team. (2011). IVAN Study Protocol V7.0: NIHR Technology Assessment Programme.

33. Gray, A. M., Rivero-Arias, O., & Clarke, P. M. (2006). Estimating the Association between SF-12 Responses and EQ-5D Utility Values by Response Mapping. Medical Decision Making, 26(1), 18-29.