D5.3 Mapping algorithms from real world outcomes to preference based measures: QoL-AD to EQ-5D

116020 - ROADMAP

**Real world Outcomes across the
AD spectrum for better care: Multi-
modal data Access Platform**

WP5 – Health Economics

Lead contributors	Ines Rombach, Marvi Iftikhar, Alastair Gray, Filipa Landeiro (1 – UOXF)
	Email: ines.rombach@ndorms.ox.ac.uk
Other contributors	John Gallacher (1 – UOXF) Jan Kors (3 – EMC) Ron Handels (4 – UM) Pascal Lecomte (16 – Novartis) Anders Gustavsson (19 – ROCHE) Gurleen Jhuti (19 – ROCHE) Yovanna Castro (19 – ROCHE)

Due date	30/04/2018
Delivery date	17/12/2018
Deliverable type	R
Dissemination level	PU

Description of Work	Version	Date
	V2.0	08/11/2017

Table of contents

Document History

Version	Date	Description
V1.0	04/10/2018	Draft for WP5 review
V1.1	25/10/2018	Implementation of WP5 reviewers' comments, inclusion of validation study
V1.2	08/11/2018	Implementation of the second round of WP5 comments
V1.3	23/11/2018	Implementation of comments from the first internal reviewer
V1.4	30/11/2018	Implementation of comments from the second internal reviewer
V2.0	14/12/2018	Final Version
V2.1	17/12/2018	Format review

Definitions

▪ The following acronyms have been used in this document

- **Actifcare**: Access to Timely Formal Care
- **AD**: Alzheimer's disease
- **CDR**: Clinical dementia rating
- **CI**: Confidence interval
- **CLAD**: Censored least absolute deviations estimator
- **HRQoL**: Health related quality of life
- **IQR**: Interquartile range
- **MAE**: Mean absolute error
- **MMSE**: Mini-Mental State Examination
- **NICE**. National Institute for Health and Care Excellence
- **PwD**: People with dementia
- **QoL-AD**: Quality of life in Alzheimer's disease
- **RMSE**: Root mean square error
- **QALY**: Quality adjusted life year
- **QoL**: Quality of Life
- **SD**: Standard deviation

▪ Partners of the ROADMAP Consortium are referred to herein according to the following codes:

- **AC Immune**. AC Immune SA (Switzerland)
- **AE**. Alzheimer Europe (Luxembourg)
- **AU**. Aarhus Universitet   (Denmark)
- **BIOGEN**. Biogen Idec Limited   (United Kingdom)
- **CBG/MEB**. Aagentschap College ter Beoordeling van Geneesmiddelen (Netherlands)
- **Eli Lilly**. Eli Lilly and Company Ltd   (United Kingdom)
- **EMC**. Erasmus University Rotterdam (Netherlands)
- **GE**. GE Healthcare Ltd   (United Kingdom)
- **HLU**. H. Lundbeck A/S (Denmark)
- **IDIAP JORDI GOL**. Fundació Institut Universitari per a la Recerca a l'Atenció Primària de Salut Jordi Gol i Gurina   (Spain)
- **IXICO**. IXICO Technologies Ltd   (United Kingdom)
- **JPNV**. Janssen Pharmaceutica NV   (Belgium)
- **LSE**. London School of Economics and Political Science   (United Kingdom)
- **LUMC**. Academisch Ziekenhuis Leiden – Leids Universitair Centrum (Netherlands)
- **Memento**. CHU Bordeaux (France)

- **NICE**. National Institute for Health and Care Excellence (United Kingdom)
- **Novartis**. Novartis Pharma AG (Switzerland) – **Project Leader**
- **ROCHE**. F. Hoffmann-La Roche Ltd   (Switzerland)
- **RUG**. Rijksuniversiteit Groningen (Netherlands)
- **SYNAPSE**. Synapse Research Management Partners (Spain)
- **TAKEDA**. Takeda Development Centre Europe LTD (United Kingdom)
- **UCPH**. Københavns Universitet   (Denmark)
- **UEDIN**. University of Edinburgh   (United Kingdom)
- **UGOT**. Goeteborgs Universitet   (Sweden)
- **UM**. Universiteit Maastricht (Netherlands)
- **UOXF**. The Chancellor, Masters and Scholars of the University of Oxford (United Kingdom) – **Coordinator**

- **Consortium.** The ROADMAP Consortium, comprising the above-mentioned legal entities.

- **Consortium Agreement.** Agreement concluded amongst ROADMAP participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.

- **Grant Agreement.** The agreement signed between the beneficiaries and the IMI JU for the undertaking of the ROADMAP project (116020).

- **Project.** The sum of all activities carried out in the framework of the Grant Agreement.

- **Work plan.** Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried out, as specified in Annex I to the Grant Agreement.

Publishable Summary

In this report we present the development of mapping algorithms to enable the estimation of EQ-5D-5L responses and indices from the disease specific Quality of Life in Alzheimer's disease (QoL-AD) questionnaire.

These algorithms can be used by researchers wishing to perform cost-effectiveness analyses in patient populations for whom QoL-AD data are available, but EQ-5D-5L data were not collected.

The mapping algorithms are based on a multinomial regression model, and predict responses to the individual EQ-5D-5L domains first, before converting them to indices using the United Kingdom (UK) value set (response mapping). Conversion to other country specific value sets is possible. Mapping algorithms are available for both self-rated and proxy-rated data.

Stata code to apply the QoL-AD to EQ-5D-5L mapping algorithms will be available to researchers on request.

1. Introduction

Dementia is a progressive neurodegenerative syndrome characterised by cognitive, behavioural, and functional decline. It culminates in memory loss, communication problems, reasoning difficulties, personality changes, and deterioration in the ability to carry out activities of daily living[1,2]. In 2015, dementia was estimated to affect between 4.7% and 7.6% of all those aged over 60 worldwide. The total number of people with dementia worldwide is projected to reach 131 million by 2050[3]. Aside from the significant financial burden associated with this disease, there is also a humanistic burden; the condition impacts on the health related quality of life (HRQoL) of people living with dementia as well as their carers.

Quality of life (QoL) is defined as 'an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns' [4]. HRQoL, a subsection of QoL, reflects 'the individual's perception of the *impact of a health status*, on the ability to perform usual tasks and effects on everyday life, and physical, social and emotional well-being. HRQoL primarily looks at quality of life through the perspective of a person's *health status* and/or impact of a *person's health condition or disability*'[5].

HRQoL instruments are commonly used in medical research to measure patients' disease progression, or the effect of different interventions. Instruments that are aimed at a particular patient population are referred to as disease-specific measures. The Quality of Life Alzheimer's Disease scale (QoL-AD) is a measure commonly used in Alzheimer's disease research.

However, to assess the cost effectiveness of medical treatments and interventions, the National Institute for Health and Care Excellence (NICE, i.e. the body which generates guidelines and recommendations for healthcare in the UK National Health Service) requires the use of a generic QoL measure, namely the EQ-5D[6]. The use of the EQ-5D, and the utilities derived from this instrument, allow comparison of utility across different pathologies, which is important for changing guidance in publicly funded healthcare systems.

Where only disease-specific outcomes have been collected, but cost-effectiveness analyses are planned, mapping algorithms can be used to estimate the EQ-5D outcomes from the data on disease specific instruments.

This study aims to create a mapping algorithm that estimates participants' EQ-5D outcomes based on their QoL-AD data. This will allow utilities to be calculated where QoL-AD has been used but no EQ-5D results are available.

Figure 1 shows the EQ-5D-5L as used in the Access to Timely Formal Care (Actifcare) study. This generic QoL questionnaire has five domains; mobility, self-care, usual activities, pain/discomfort, and anxiety/depression [7]. There are five response levels in the EQ-5D-5L: no problems, slight problems, moderate problems, severe problems, unable to perform activity or extreme pain/ discomfort or extremely depressed or anxious. The EQ-5D-3L is an earlier version of the EQ-5D-5L. The EQ-5D-3L has three levels within each of the five domains: no problems, some problems, extreme problems.

In addition to providing descriptive data on the responses chosen, the health state reported by individuals in their responses can also be converted to a single utility measure or preference weight, for which country-specific value sets exist. These reflect the strength of preference of the general

population with regard to different health states. The UK country specific value set was determined using time-trade-off techniques[8].

The EQ-5D has a maximum value of one, indicating perfect health. Zero is aligned with a health state equal to death and negative values represent HRQoL states considered worse than death.

NICE currently recommends to generate utilities based on the 'cross walk' from the EQ-5D-5L to EQ-5D-3L valuation dataset[9]. The value set of the EQ-5D-3L is approved for use in cost-effectiveness analyses[10].

EQ-5D-5L

Questionnaire 2

Instructions: Under each heading, please tick the ONE box that best describes the participant's health TODAY

MOBILITY

I have no problems in walking about	❑
I have slight problems in walking about	❑
I have moderate problems in walking about	❑
I have severe problems in walking about	❑
I am unable to walk about	❑

SELF-CARE

I have no problems washing or dressing myself	❑
I have slight problems washing or dressing myself	❑
I have moderate problems washing or dressing myself	❑
I have severe problems washing or dressing myself	❑
I am unable to wash or dress myself	❑

USUAL ACTIVITIES *(e.g. work, study, housework, family or leisure activities)*

I have no problems doing my usual activities	❑
I have slight problems doing my usual activities	❑
I have moderate problems doing my usual activities	❑
I have severe problems doing my usual activities	❑
I am unable to do my usual activities	❑

PAIN / DISCOMFORT

I have no pain or discomfort	❑
I have slight pain or discomfort	❑
I have moderate pain or discomfort	❑
I have severe pain or discomfort	❑
I have extreme pain or discomfort	❑

ANXIETY / DEPRESSION

I am not anxious or depressed	❑
I am slightly anxious or depressed	❑
I am moderately anxious or depressed	❑
I am severely anxious or depressed	❑
I am extremely anxious or depressed	❑

Figure 1. EQ-5D-5L questionnaire used in Actifcare study

1.1. QoL-AD

The QoL-AD questionnaire is a disease-specific HRQoL instrument for use in people with dementia. The questionnaire can be completed by the people with dementia themselves, or completed on their behalf by their carers[11].

Figure 2 shows the questionnaire as used in the Actifcare study. It consists of 13 questions each with four possible responses – poor, fair, good and excellent, each equating to 1-4 points respectively. The different domains cover a range of factors, including overall QoL, relationships with family and friends, physical health, memory, and ability to perform household chores and activities. A composite score between 13 and 52 is derived by summing up the answers to the 13 individual questions. Higher scores represent higher QoL.

Some studies have shown low completion rates for the question related to marriage, which arguably cannot be answered by participants who are not married, or are widowed or divorced. Some studies have reported that answers for missing questions were either imputed or not included in their analysis[12-15].

QoL-AD was chosen because it has been used in studies of people with dementia [16-19], and is the recommended disease-specific questionnaire for dementia[20], but has not to our knowledge been mapped to the EQ-5D yet.

The DEMQOL is another instrument often used in dementia research, but techniques to derive utilities from this questionnaire already exist[21]. Alzheimer's Disease Cooperative Study Activities of Daily Living scale is the only dementia specific questionnaire to be mapped so far[22].

Quality of Life-AD

Questionnaire 1

Interviewer: administer according to standard instructions. Circle responses.				
1. How do you feel about your **physical health**? Would you say it's poor, fair, good, or excellent? Circle whichever word you think best describes your physical health right now.	Poor	Fair	Good	Excellent
2. How do you feel about your **energy** level? Do you think it's poor, fair, good, or excellent?	Poor	Fair	Good	Excellent
3. How has your **mood** been lately? Have your spirits been good, or have you been feeling down? Would you rate your mood as poor, fair, good, or excellent?	Poor	Fair	Good	Excellent
4. How about your **living situation**? How do you feel about the place you live now? Would you say it's poor, fair, good or excellent?	Poor	Fair	Good	Excellent
5. How about your **memory**? Would you say it is poor, fair, good, or excellent?	Poor	Fair	Good	Excellent
6. How about your **family** and your relationship with family members? Would you describe it as poor, fair, good, or excellent?	Poor	Fair	Good	Excellent
7. How do you feel about your **marriage**? How is your relationship with (spouse's name). Do you feel it's poor, fair, good, or excellent?	Poor	Fair	Good	Excellent
8. How would you describe your current relationship with your **friends**? Would you say it's poor, fair, good, or excellent?	Poor	Fair	Good	Excellent
9. How do you feel about **yourself**- when you **think of your whole self**, and all the different things about you, would you say it's poor, fair, good, or excellent?	Poor	Fair	Good	Excellent
10. How do you feel about your **ability to do things like chores around the house** or other things you need to do? Would you say it's poor, fair good, or excellent?	Poor	Fair	Good	Excellent
11. How about your **ability to do things for fun**, that you enjoy? Would you say it's poor, fair, good, or excellent?	Poor	Fair	Good	Excellent
12. How do you feel about your current situation with **money**, your financial situation? Do you feel it's poor, fair, good, or excellent?	Poor	Fair	Good	Excellent
13. How would you describe your life as a whole. When you think about your **life as a whole**, everything together, how do you feel about your life? Would you say it's poor, fair good, or excellent?	Poor	Fair	Good	Excellent

Comments: _____

©1996 Rebecca Logsdon, PhD; University of Washington

Figure 2. QoL-AD questionnaire

2. Methods

This mapping study was performed in line with recommendations from the MAPS statement[23].

2.1. Dataset used

This mapping exercise is based on data from the Actifcare study[24]. The Actifcare study is a longitudinal cohort study that aims to develop best practice for access to formal care for dementia patients in the community, and included participants from Germany, the Netherlands, Sweden, Norway, Ireland, UK, Portugal and Italy. Participants were recruited between 2014 and 2016 through memory clinics, general practices, case managers and community mental health teams. Eligible participants had to have a diagnosis of dementia and have a Clinical Dementia Rating (CDR) or of 1 or 2 or Mini-Mental State Examination (MMSE) scores of 24 or below. Eligible participants did have an informal carer but were likely to require formal assistance over the next year[24].

Outcome measures included the EQ-5D-5L and the QoL-AD. Both questionnaires were rated by the people with dementia themselves (self-rated) and by the carers on behalf of the people with dementia (proxy-rated).

Up to three observations were available for each participant, as the study protocol stipulated that data should be collected at baseline, at six and at 12-month follow-up.

External validations of the mapping exercise were performed using data from the 'Diagnostic and economic evaluation of new biomarkers for Alzheimer's disease' (LEARN study)[25].

2.2. Data used for the mapping exercise

In dementia studies, EQ-5D-5L and QoL-AD may be completed by the people with dementia themselves (self-rated) or by their carers (proxy-rated). This study aimed to cover all relevant mapping algorithms that may be required, and the following mapping scenarios were considered:

- Mapping self-rated QoL-AD to self-rated EQ-5D-5L
- Mapping self-rated QoL-AD to proxy-rated EQ-5D-5L
- Mapping proxy-rated QoL-AD to self-rated EQ-5D-5L
- Mapping proxy-rated QoL-AD to proxy-rated EQ-5D-5L

For each of the four scenarios, only observations for which data were available for all five EQ-5D-5L items, and QoL-AD items 1-6 and 8-13 were used.

This mapping exercise does not include item 7 of the QoL-AD. This question is related to marriage, and therefore cannot be answered by all participants. In fact, in our dataset, item 7 was left unanswered for approximately 27% of observations, and the majority of participants who did not answer this item identified as widowed, single, separated or divorced. The total QoL-AD score for participants was calculated by adding the score for each individual question apart from item 7, and then rescaled to range from 13 to 52 points, in line with the full score.

The EQ-5D-5L items were used to generate utility scores using the cross-walk to the EQ-5D-3L based on the UK-specific value set for the UK[26]. The use of the crosswalk over the EQ-5D-5L utility set is currently recommended by NICE, instead of using the current value set developed for the EQ-5D-5L[27,28].

The number of available observations for each of the mapping scenarios is presented in the results' section.

We also ran the different models including item 7, with best results presented in Annex I.

2.3. Statistical models considered for the mapping algorithm

Different statistical models were trialled. Direct mapping describes models whereby the explanatory variables are directly mapped onto the EQ-5D-5L index or utility score. Response models use the explanatory variables to firstly model the responses to the individual items (questions), which are then combined to obtain the EQ-5D-3L index. For the response mapping, each EQ-5D-5L item is modelled separately; hence the response mapping algorithm consists of five statistical models. Details of the different models used, and their model specifications are provided in Table 1.

All models were run as specified in the table, and again by taking into account age of the person with dementia (as a continuous variable) and their sex (as a categorical variable). Only age at baseline was available. The baseline date was used again for the six-month follow-up, but the age was increased by one year for the one-year follow-up assessment.

We avoided using further demographic details as these may not be available to others and thus would limit the use of the chosen model. The 'cluster' option in Stata was used in all models except the censored least absolute deviations estimator (CLAD) model to account for clustering of observations within participants, and generated robust standard errors.

QoL-AD scores were used as the explanatory variables in the model, either as a 'continuous' composite score or as 'categorical' using each individual item.

For example, the model titled 'Direct OLS Categorical' in Table 1 describes a direct mapping model, whereby the EQ-5D utility is used as the continuous outcome variable. The QoL-AD items are used separately and as categorical explanatory variables. Robust standard errors are used.

Age included as a linear explanatory variable was found to be a good fit in the model (compared to quadratic terms, which were not statistically significant).

Table 1. Features of different models used in mapping study

	Model Reference	Outcome variable: EQ-5D-5L			Explanatory variables: QoL-AD		
		Statistical model used	Continuous index	Individual items	Continuous composite score	Items (categorical)	Robust SEs used
Direct mapping	Direct OLS Continuous	OLS	✓		✓		✓
	Direct OLS Categorical	OLS	✓			✓	✓
	Direct Tobit	Tobit	✓			✓	✓
	Direct CLAD	CLAD	✓			✓	
	Direct 2-part	2-part*	✓			✓	✓
Response mapping	Response ologit	ologit		✓		✓	✓
	Response OLS Categorical	OLS		✓		✓	✓
	Response OLS Continuous	OLS		✓	✓		✓
	Response mlogit	mlogit		✓		✓	✓

1st part of the model uses a logistic regression model to predict which participants were in a perfect health state (EQ-5D-5L index = 1). The 2nd part predicts utilities for participants who are not in a perfect health state.

Abbreviations: CLAD - Censored least absolute deviations estimator; OLS – Ordinary least squares; QoL-AD: Quality of Life in Alzheimer's disease; SE – Standard error

2.4.   Calculation of predicted utilities

For the direct mapping, Stata's 'predict' command was used to obtain the predicted EQ-5D-5L index for each observation.

For the two-part model, the index (i.e. utility value) was calculated as follows:

Utility = Pr(PerfectHealth) + (1-Pr(PerfectHealth)) * Y [29]

where Pr(PerfectHealth) is the predicted probability that utility = 1 and Y = predicted utility conditional on imperfect health.

For response mapping using the OLS models, utilities were generated using the predicted responses to the EQ-5D-5L items and then using the crosswalk to the EQ-5D-3L (UK value set).

The mlogit and ologit models produced probabilities of each participant falling into the different item levels on the EQ-5D-5L. These were transformed into probabilities of falling into each of the EQ-5D-3L items using the transition matrix. Subsequently, the UK scoring manual was applied to obtain EQ-5D-3L utilities:

Utility = 1 – Pr(Mobility2)*0.069 – Pr(Mobility3)*0.314 – Pr(SelfCare2)*0.104 – Pr(SelfCare3)*0.214 – Pr(Activities2)*0.036 – Pr(Activities3)*0.094 - Pr(Pain2)*0.123 – Pr(Pain3)*.386 – Pr(Depression2)*0.071 – Pr(Depression3)*0.236 - (1-Pr(Perfect))*0.081 – Pr(N3)*0.269

where Pr(Question2) is the probability of selecting response level 2 (some problems) for said question, and Pr(Question3) is the probability of selecting response level 3 (severe problems) for said question; Pr(N3) is the probability of having severe problems in at least one question.

2.5.   Performance assessment

Model performance was assessed using the root mean square error (RMSE) and the mean absolute error (MAE).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted\ Utility - Observed\ Utility)^2}{N}}$$

$$MAE = \frac{\sum_{i=1}^{N}|Predicted\ Utility_i - Observed\ Utility_i|}{N}$$

Lower RMSEs and MAEs indicate higher accuracy of the predictions of a model.

For the final models, based on the one producing the lowest RMSEs and MAEs, accuracy of the prediction was also compared in different subsets of the dataset based on QoL-AD scores. The purpose of this investigation was to establish if the mapping algorithm performed better for participants with different QoL-AD scores.

Stata/SE version 14.2 was used for all analyses.

2.6.   Validation in a separate dataset

The model for one scenario (proxy-rated QoL-AD mapped to proxy-rated EQ-5D-3L) was validated using data from an external dataset. This dataset was obtained from 'Diagnostic and economic evaluation of new biomarkers for Alzheimer's disease (AD)[25]. This is a prospective cohort study of 241 participants who were followed for 2 years. Eligible patient were suspected of having a primary neurodegenerative disease referred to memory clinics, and had MMSE ≥ 20 and CDR of 0-1 and available proxy. The aim of this study is to evaluate the clinical and economic value of biomarkers.

3. Results

3.1.   Description of the population

Four hundred and fifty one participants were included in the original study, with a possible 1353 observations to be used in this mapping study. After dropping missing items as described above, between 1017 and 1099 observations were available for the different mapping scenarios.

Patient characteristics for all 451 participants included in the Actifcare study were as follows: the mean age at baseline 78 (Standard deviation (SD) 8) for people with dementia, and 66 (SD 13) for their carers. 55% of people with dementia were female, and 66% of carers we female.

The mean Mini-Mental State Examination (MMSE) score was 19 (SD 5), 2%, 78% and 20% respectively had a baseline Clinical Dementia Rating (CDR) of 0.5, 1 and 2, respectively. The baseline EQ-5D index was 0.772 (SD 0.186) and 0.628 (SD 0.222) for self-rated and proxy-rated values, respectively. Baseline QoL-AD values of 36 (SD 6) were reported by people with dementia, and mean baseline proxy-rated scores had a mean value of 31 (SD 6).

Table 2 shows the number of participants/observations used in each of the four scenarios. Information on demographics (age, sex), Mini-Mental State Examination (MMSE) and Clinical Dementia Rating (CDR), mean/median utility based on the EQ-5D and QoL-AD scores are shown for each observation, i.e. up to three data points may be included per participant to reflect data collected at the different time points. This approach best reflects the data used in this mapping exercise, and acknowledges that age, MMSE, CDR, EQ-5D and QoL-AD change between the different follow-up time points.

Considering all observations, the mean age of people with dementia was 77; the average age of carers was 66.

Around 55% of observations in the dataset were from female people with dementia, and roughly two thirds of observations were from female carers.

Mean self-rated QoL-AD scores were 35, while the mean proxy-rated equivalents were around 4 to 5 points lower. Similarly, self-rated EQ-5D utilities had a mean of 0.77 in both scenarios that mapped to the self-rated EQ-5D utilities, while proxy-rated utilities had a mean of 0.60 to 0.62.

A slightly higher proportion of observations related to higher severity of dementia were observed in the proxy-rated QoL-AD mapped to proxy-rated EQ-5D-5L group. This may reflect the more severe participants being unable to complete the questionnaire and therefore being dropped from the other scenarios. Most observations for CDR Ratings of 1 or 2 (mild-moderate dementia).

Table 2. Overview of data included in the mapping study

Demographic Variable	Self-rated QoL-AD → Self-rated EQ-5D	Proxy- rated QoL-AD → Self- rated EQ-5D	Self- rated QoL-AD → Proxy- rated EQ-5D	Proxy- rated QoL-AD → Proxy- rated EQ-5D
Number of participants	427	429	427	437
Number of observations	1020	1019	1017	1099
Excluded observations*	333	334	336	254
PwD Age[1]	77.6 (7.7)	77.5 (7.8)	77.6 (7.7)	77.6 (7.9)
Proxy Age[1]	65.8 (13.3)	66.4 (13.3)	66.3 (13.1)	66.5 (13.3)
PwD Sex (Female)[1]	55.2% (563)	55.3% (563)	55.2% (561)	54.3% (597)
Proxy Sex (Female)[1]	66.8% (278)	67.0%	67.0% (679)	66.8% (732)
MMSE[2]	18.7 (5.3)	18.8 (5.4)	18.7 (5.3)	18.7 (5.4)
CDR 0[3]	0%	0.1%	0%	0%
CDR 0.5[3]	2.9%	2.8%	2.8%	2.7%
CDR 1[3]	69.6%	69.5%	69.3%	66.9%
CDR 2[3]	26.1%	26.4%	26.5%	27.6%
CDR 3[3]	1.5%	1.3%	1.5%	2. 9%
Self-rated QoL-AD mean (SD)	34.9 (6.2)	n/a	34.9 (6.2)	n/a
Self-rated QoL-AD median (range)	35.8 (16.3 – 52.0)	n/a	35.8 (16.3 – 52.0)	n/a
Proxy-rated QoL-AD mean (SD)	n/a	30.3 (6.1)	n/a	30.2 (6.1)
Proxy-rated QoL-AD median (range)	n/a	30.3 (15.2 – 49.8)	n/a	30.3 (15.2 – 49.8)
Self-rated EQ-5D Utility mean (SD)	0.77 (0.21)	0.77 (0.21)	n/a	n/a
Self-rated EQ-5D Utility median (range)	0.81 (-0.26 – 1)	0.81 (-0.26 – 1)	n/a	n/a
Proxy-rated EQ-5D Utility mean (SD)	n/a	n/a	0.62 (0.23)	0.60 (0.24)
Proxy-rated EQ-5D Utility median (range)	n/a	n/a	0.65 (-0.31 – 1)	0.64 (-0.31 – 1)

*Insufficient EQ-5D-5L or QOL-AD data were available for inclusion in the mapping study.

[1]summary data for all observations (each individual participant may have more than one response)

[2]MMSE data unavailable for 7.8%, 8.3%, 8.0%, 13.5% for each scenario respectively. The percentages are based on the population with available MMSE data only.

[3]CDR data is unavailable for 0.9, 0.6%, 0.5%, and 1.9% for each scenario respectively. The percentages are based on the population with available CDR data only

3.2. Description of Observed QoL-AD and Observed EQ-5D

3.2.1. Self-rated QoL-AD vs. self-rated EQ-5D

The correlation between the two questionnaires was 0.49 (Spearman's correlation, 95% confidence interval 0.45-0.54), which is similar to previously documented correlation of 0.48[30].

Figure 3 shows a scatter plot of EQ-5D utilities and QoL-AD score. There is a positive correlation but clear variation can be seen, e.g. for a utility of 1 the observed QoL-AD scores ranged from 26 to 52.
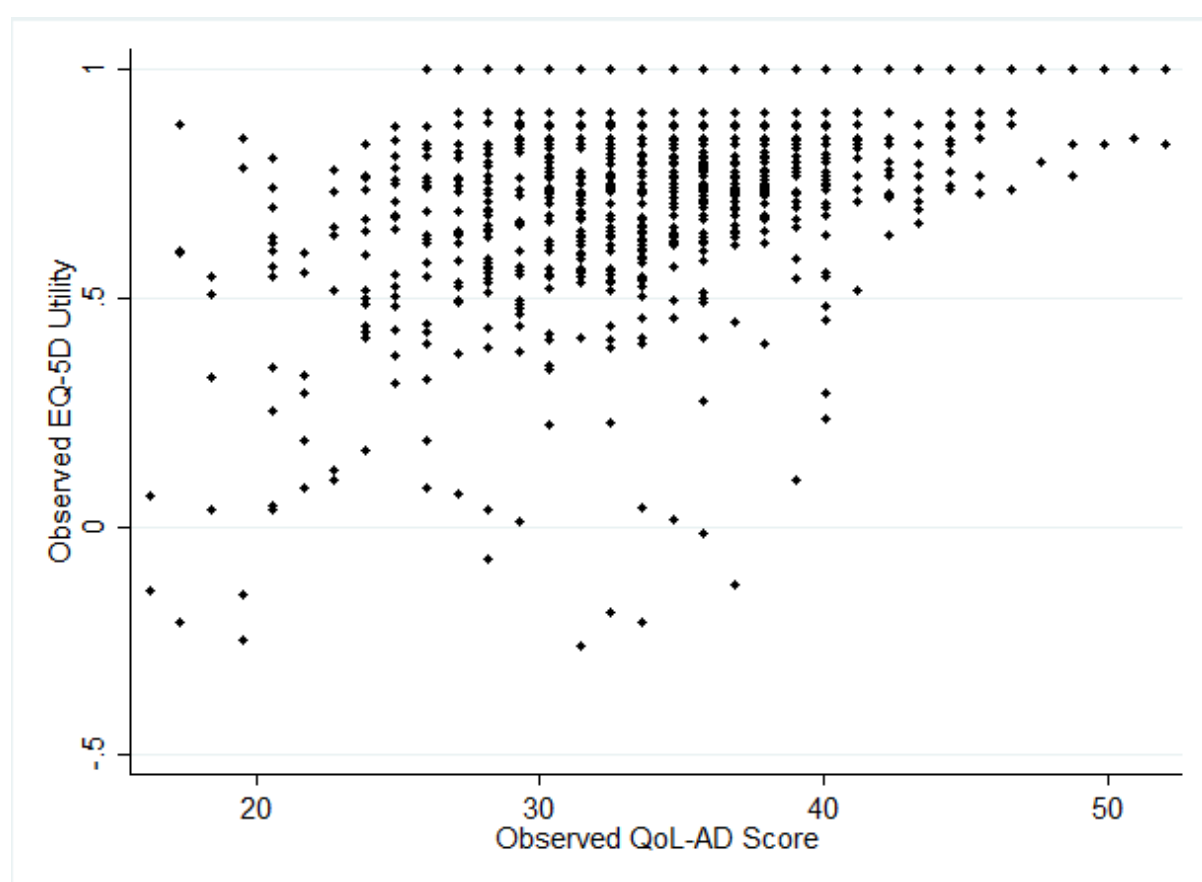

Figure 3. Observed self-rated QoL-AD vs. observed self-rated EQ-5D utilities

For further analysis, we categorised the QoL-AD score as follows:
- 1 = QoL-AD score <20
- 2 = QoL-AD score 20-29
- 3 = QoL-AD score 30-34
- 4 = QoL-AD score 35-39
- 5 = QoL-AD score 40-52

Given that QoL-AD does not correlate with disease severity, as classified by MMSE and CDR, we have chosen the above cut-offs to generate estimates across the range of the QoL-AD. Smaller

intervals were selected for QoL-AD values between 30 to 40 as most of the observations were concentrated in this range.

Table 3 shows that as QoL-AD score increases, so does the mean utility. Lower QoL-AD scores are associated with higher variability (see standard deviations) compared to the higher QoL-AD scores.

Figure 4 is a graphical representation of this data (boxplots). This figure demonstrates the high variability across all the QoL-AD categories, with most variability seen in the first category (QoL-AD < 20).

Table 3. Observed self-rated EQ-5D utility by self-rated QoL-AD category

QoL-AD Category	EQ-5D							
	Obs	Mean	SD	Min	Max	Median	25%	75%
1 (<20)	14	0.318	0.414	-0.247	0.879	0.417	-0.142	0.603
2 (20-29)	190	0.634	0.220	-0.069	1	0.667	0.527	0.791
3 (30-34)	289	0.738	0.197	-0.261	1	0.767	0.636	0.848
4 (35-39)	303	0.832	0.157	-0.127	1	0.837	0.750	1
5 (40-52)	224	0.879	0.133	0.238	1	0.879	0.801	1

Abbreviations: Obs – Observations, SD – Standard Deviation
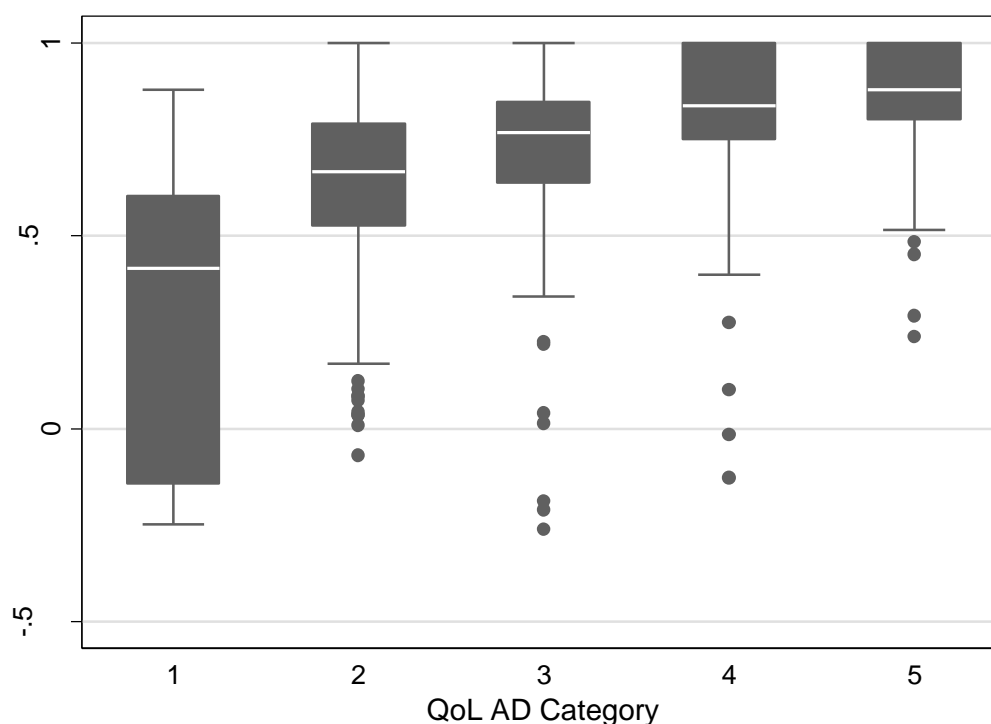

Figure 4. Distribution of observed self-rated EQ-5D utility by self-rated QoL-AD category

The MMSE was not associated with utility values (Table 4 and Figure 5), with very similar utility scores reported across all MMSE categories. The following MMSE cut-offs were used:

- 0 – Normal – MMSE > 26
- 1 – Mild – MMSE 21-26
- 2 – Moderate – MMSE 10-20
- 3 – Severe – MMSE <10

Table 4. Observed self-rated EQ-5D utility by MMSE severity

MMSE Severity	Observed EQ-5D Utility							
	Obs	Mean	SD	Min	Max	Median	25%	75%
0	59	0.723	0.164	0.343	1	0.727	0.635	0.848
1	316	0.791	0.171	-0.016	1	0.809	0.711	0.906
2	522	0.768	0.229	-0.261	1	0.806	0.683	0.906
3	43	0.785	0.234	0.037	1	0.846	0.689	1
Missing	80	0.747	0.221	0.041	1	0.774	0.642	0.906

Abbreviations: Obs – Observations, SD – Standard Deviation
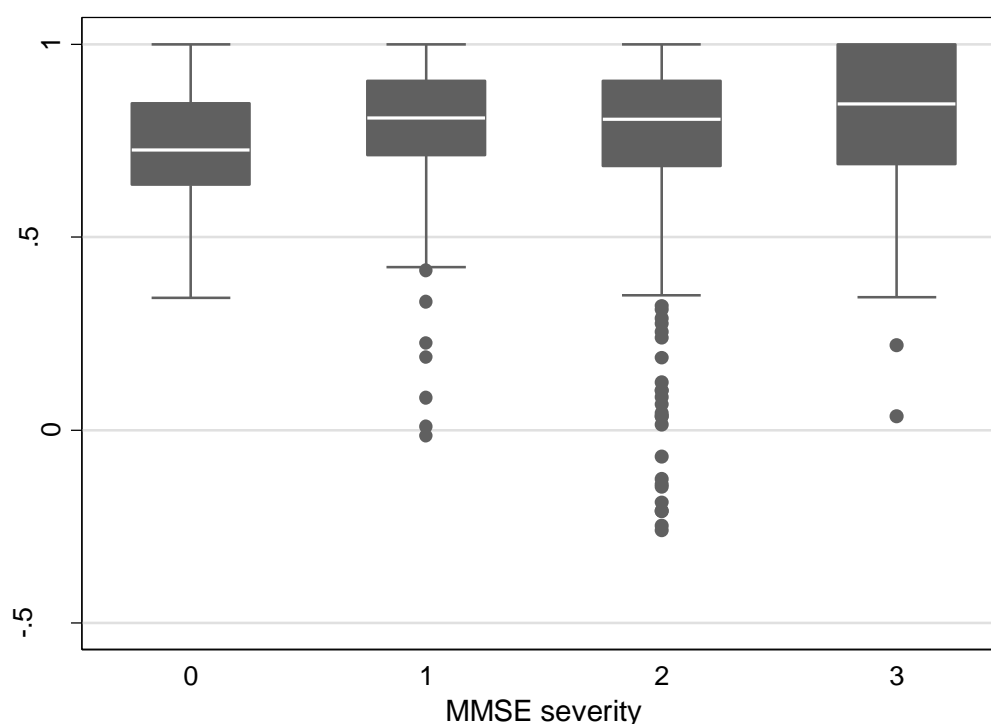
Figure 5. Observed self-rated EQ-5D utilities by MMSE severity

Higher CDR scores were associated with somewhat lower utility scores (Table 5 and Figure 6).

Table 5. Observed self-rated EQ-5D utilities by CDR

CDR Score	Observed EQ-5D Utility							
	Obs	Mean	Std. Dev.	Min	Max	Median	25%	75%
0.5	28	0.856	0.154	0.501	1	0.892	0.754	1
1	704	0.784	0.190	-0.247	1	0.811	0.698	0.906
2	264	0.735	0.249	-0.261	1	0.777	0.647	0.895
3	15	0.685	0.283	-0.127	1	0.786	0.636	0.837
Missing	9	0.743	0.158	0.562	1	0.727	0.635	0.740

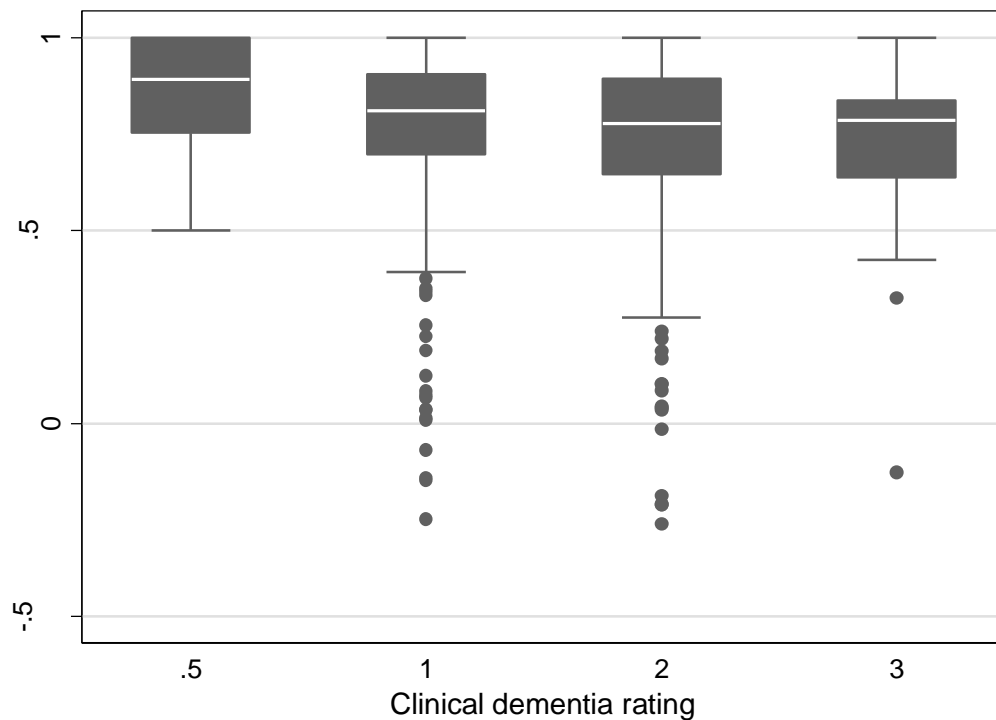*Abbreviations: Obs – Observations, SD – Standard Deviation*

Figure 6. Observed self-rated EQ-5D utilities by CDR

The following tables (Table 6-Table 14) and figures (Figure 7-Figure 18) represent the same results but for each scenario as given by the heading.

3.2.2. Proxy-rated QoL-AD vs. self-rated EQ-5D

The correlation between the two questionnaires was 0.32 (Spearman's correlation, 95% confidence interval 0.27-0.38).
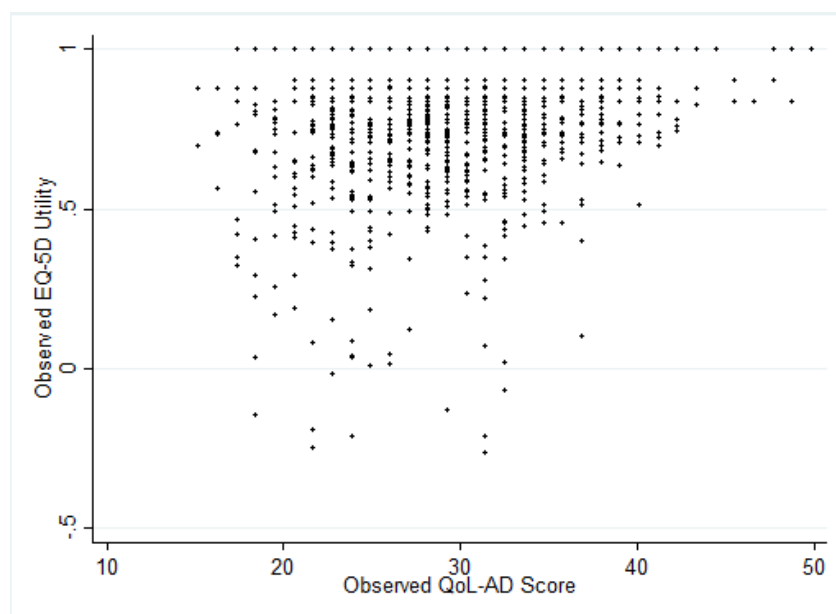

Figure 7. Observed proxy-rated QoL-AD vs. observed self-rated EQ-5D utilities

The correlation between the two questionnaires was 0.49 (Spearman's correlation, 95% confidence interval 0.45-0.54), which is similar to previously documented correlation of 0.48[30].

Table 6. Observed self-rated EQ-5D utility by proxy-rated QoL-AD category

QoL-AD Category	EQ-5D							
	Obs	Mean	SD	Min	Max	Median	25%	75%
1 (<20)	46	0.636	0.379	-0.166	0.848	0.448	0.036	0.704
2 (20-29)	431	0.731	0.216	-0.247	1	0.765	0.635	0.850
3 (30-34)	332	0.791	0.198	-0.261	1	0.836	0.711	0.906
4 (35-39)	141	0.847	0.145	0.101	1	0.848	0.767	1
5 (40-52)	69	0.908	0.109	0.511	1	1	0.837	1

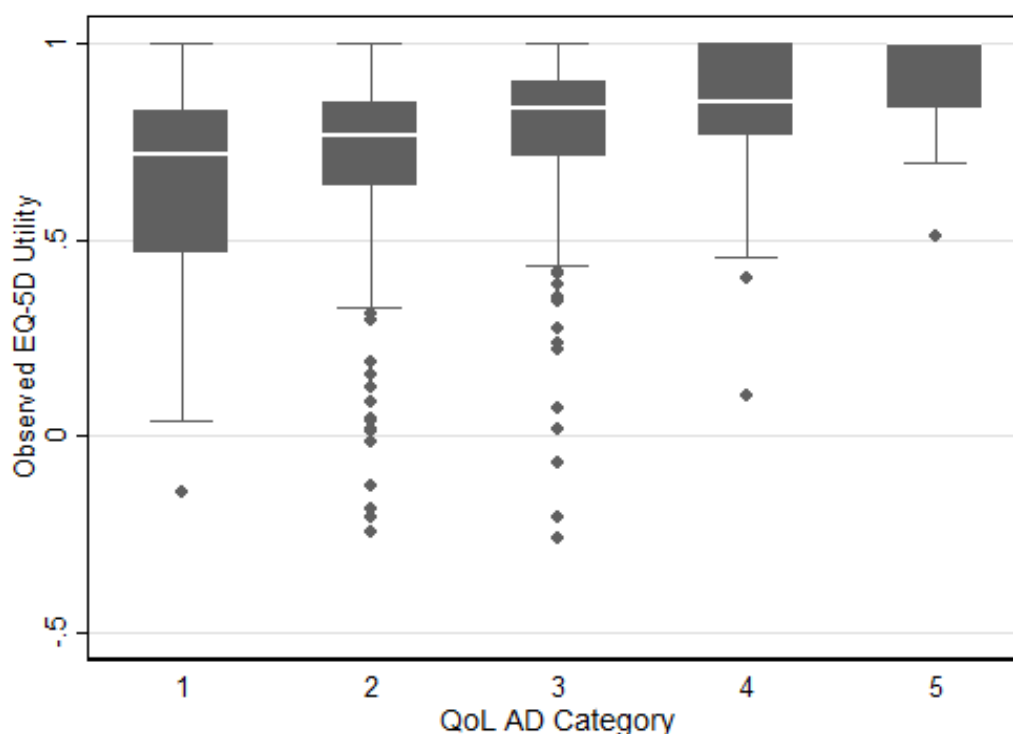*Abbreviations: Obs – Observations, SD – Standard Deviation*

Figure 8. Distribution of observed self-rated EQ-5D utilities by proxy-rated QoL-AD category

Table 7. Observed self-rated EQ-5D utilities by MMSE severity

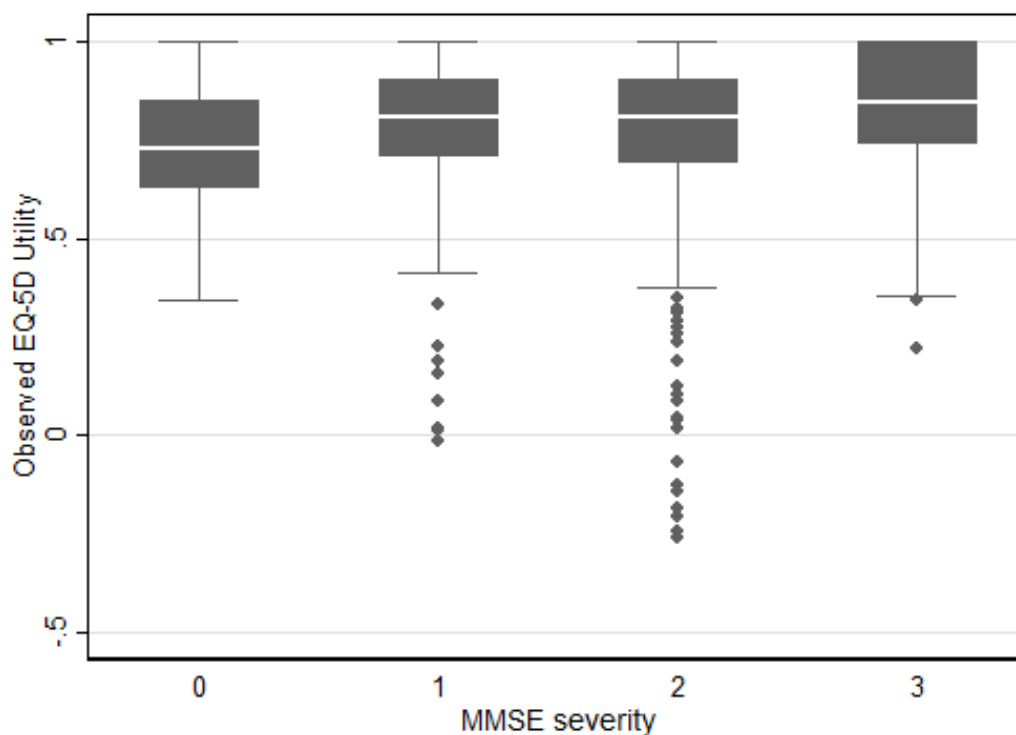| MMSE Severity	Observed EQ-5D Utility							
	Obs	Mean	SD	Min	Max	Median	25%	75%
0	61	0.721	0.164	0.343	1	0.727	0.624	0.848
1	323	0.784	0.180	-0.016	1	0.809	0.708	0.906
2	502	0.774	0.223	-0.261	1	0.809	0.69	0.906
3	48	0.809	0.207	0.221	1	0.847	0.740	1
Missing	85	0.759	0.224	0.041	1	0.778	0.681	1

Abbreviations: Obs – Observations, SD – Standard Deviation

Figure 9. Observed self-rated EQ-5D utilities by MMSE severity

Table 8. Observed self-rated EQ-5D utilities by CDR

CDR Score	Obs	Mean	Std. Dev.	Min	Max	Median	25%	75%
			Observed EQ-5D Utility					
0	1	0.740	n/a	n/a	n/a	n/a	n/a	n/a
0.5	28	0.853	0.158	0.501	1	0.893	0.754	1
1	704	0.786	0.187	-0.247	1	0.814	0.707	0.906
2	267	0.742	0.247	-0.261	1	0.778	0.647	0.906
3	13	0.628	0.307	-0.127	1	0.767	0.425	0.837
Missing	6	0.794	0.175	0.562	1	0.753	0.693	1

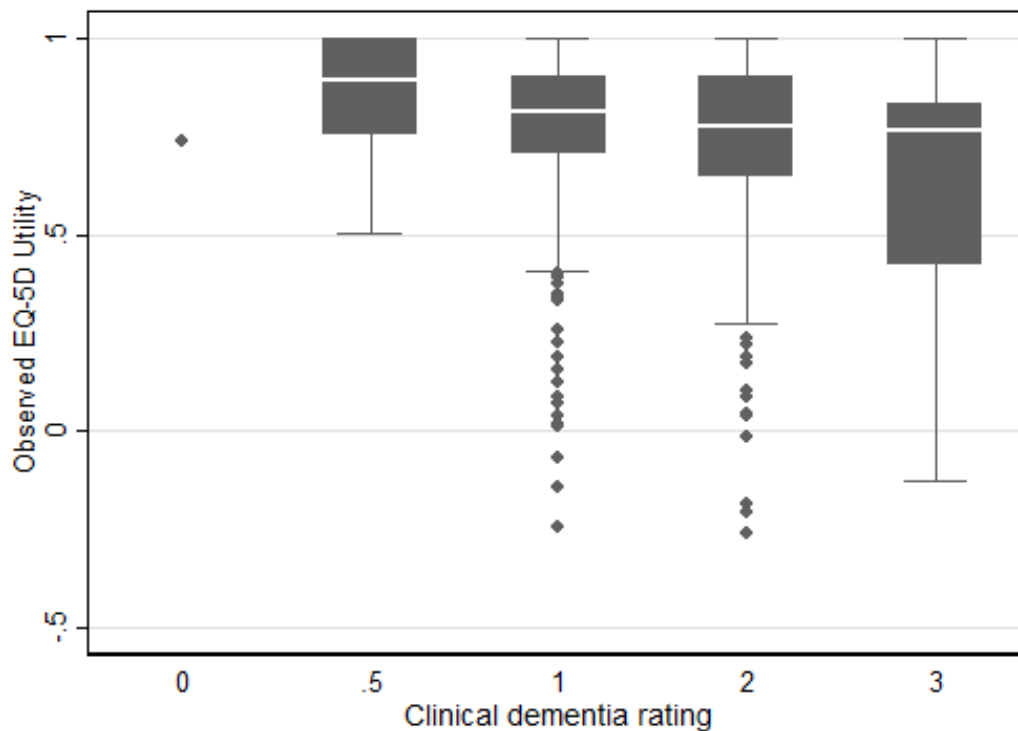*Abbreviations: Obs – Observations, SD – Standard Deviation*

Figure 10. Observed self-rated EQ-5D utilities by CDR

3.2.3. Self-rated QoL-AD vs. proxy-rated EQ-5D

The correlation between the two questionnaires was 0.24 (Spearman's correlation, 95% confidence interval 0.18-0.30).
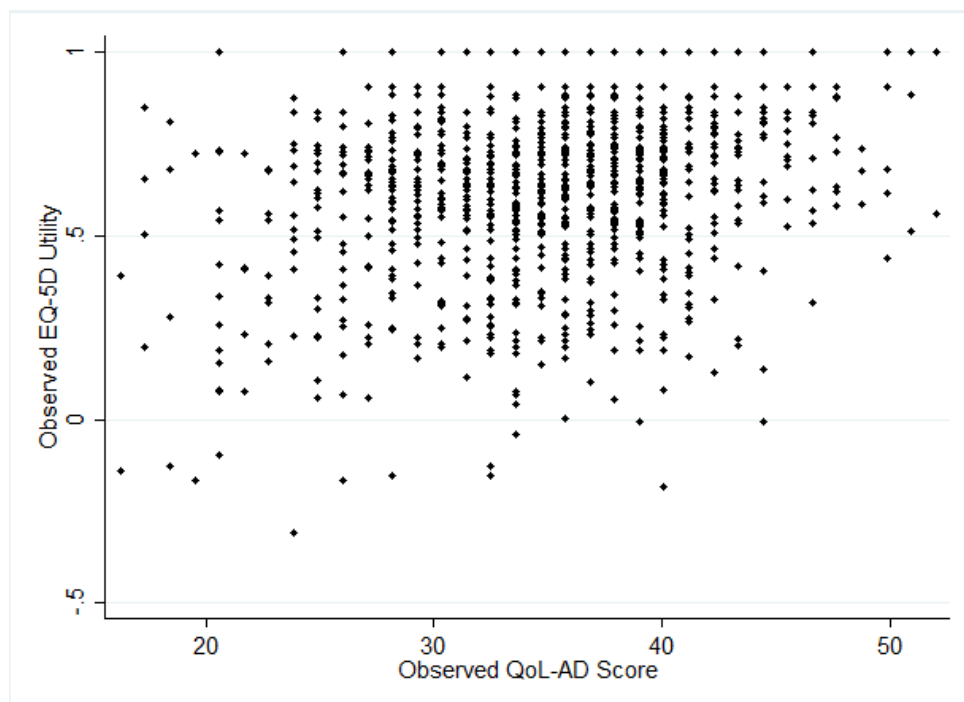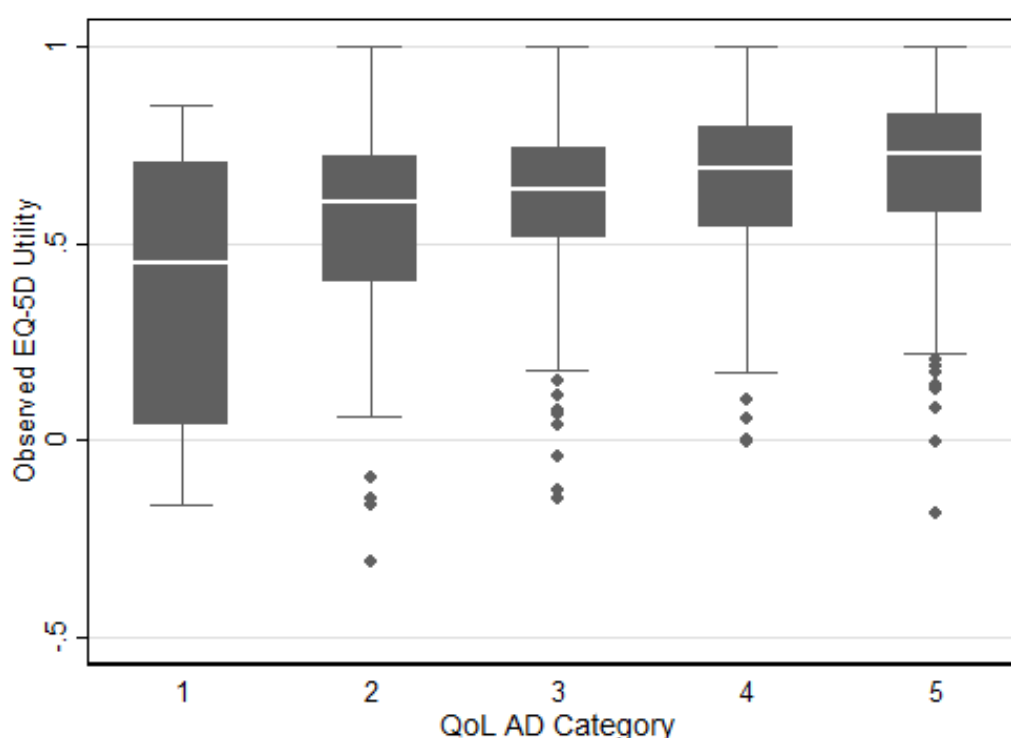

Figure 11. Observed self-rated QoL-AD vs. observed proxy-rated EQ-5D utilities

Table 9. Observed proxy-rated EQ-5D utility by self-rated QoL-AD category

QoL-AD Category	EQ-5D							
	Obs	Mean	SD	Min	Max	Median	25%	75%
1 (<20)	12	0.389	0.379	-0.166	0.848	0.448	0.036	0.704
2 (20-29)	192	0.543	0.242	-0.307	1	0.604	0.400	0.722
3 (30-34)	290	0.604	0.214	-0.151	1	0.637	0.512	0.746
4 (35-39)	298	0.659	0.206	-0.005	1	0.689	0.543	0.796
5 (40-52)	225	0.680	0.220	-0.184	1	0.727	0.580	0.827

Abbreviations: Obs – Observations, SD – Standard Deviation

Figure 12. Distribution of observed proxy-rated EQ-5D utilities by self-rated QoL-AD category

Table 10. Observed proxy-rated EQ-5D utilities by MMSE severity

MMSE Severity	Observed EQ-5D Utility							
	Obs	Mean	SD	Min	Max	Median	25%	75%
0	57	0.665	0.181	0.003	1	0.689	0.567	0.768
1	319	0.663	0.199	0.080	1	0.689	0.548	0.795
2	518	0.607	0.235	-0.166	1	0.644	0.480	0.776
3	42	0.574	0.224	0.069	0.906	0.612	0.404	0.733
Missing	81	0.562	0.282	-0.307	1	0.612	0.413	0.747

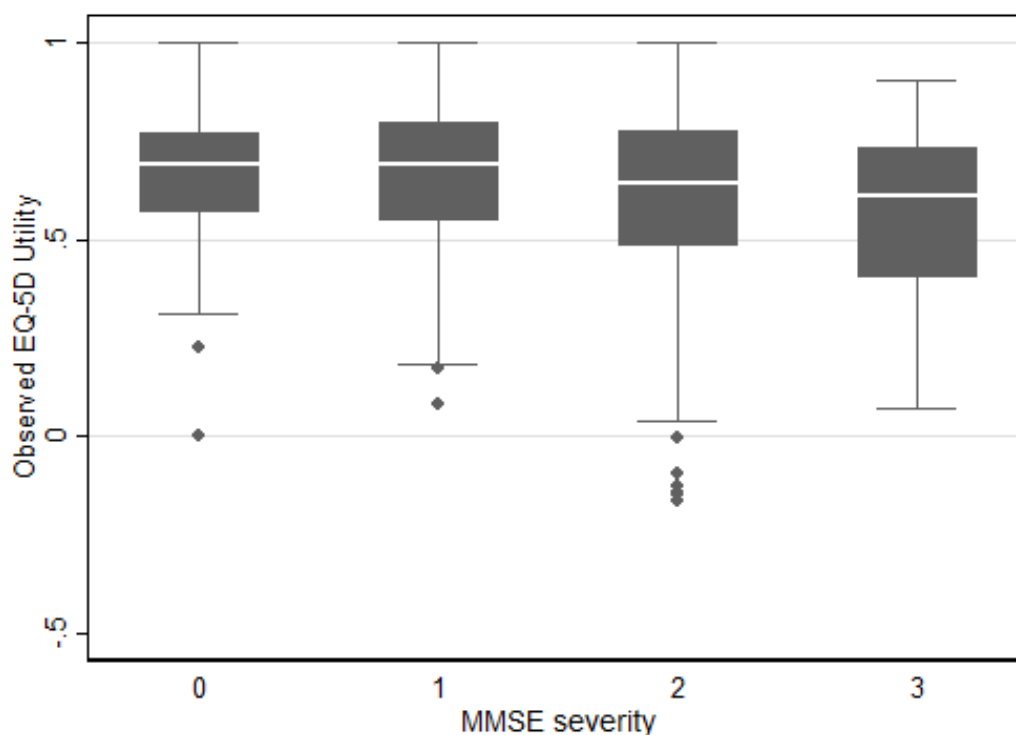*Abbreviations: Obs – Observations, SD – Standard Deviation*

Figure 13. Observed proxy-rated EQ-5D utilities by MMSE severity

Table 11. Observed proxy-rated EQ-5D utilities by CDR

CDR Score	Obs	Mean	SD	Min	Max	Median	25%	75%
				Observed EQ-5D Utility				
0.5	28	0.795	0.150	0.513	1	0.787	0.689	1
1	701	0.668	0.198	-0.166	1	0.695	0.560	0.796
2	268	0.500	0.245	-0.307	1	0.545	0.323	0.672
3	15	0.368	0.249	-0.127	1	0.404	0.187	0.585
Missing	5	0.594	0.328	0.180	1	0.533	0.419	0.837

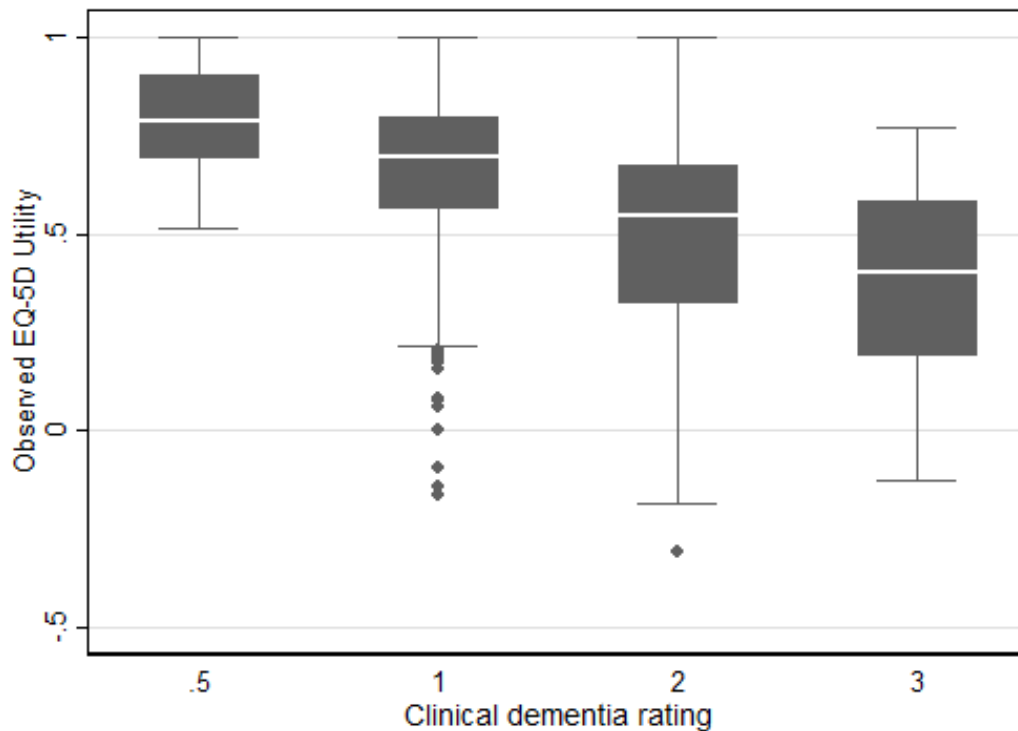*Abbreviations: Obs – Observations, SD – Standard Deviation*

Figure 14. Observed proxy-rated EQ-5D utilities by CDR

3.2.4. Proxy -rated QoL-AD vs. proxy-rated EQ-5D

The correlation between the two questionnaires was 0.48 (Spearman's correlation, 95% confidence interval 0.43-0.52).
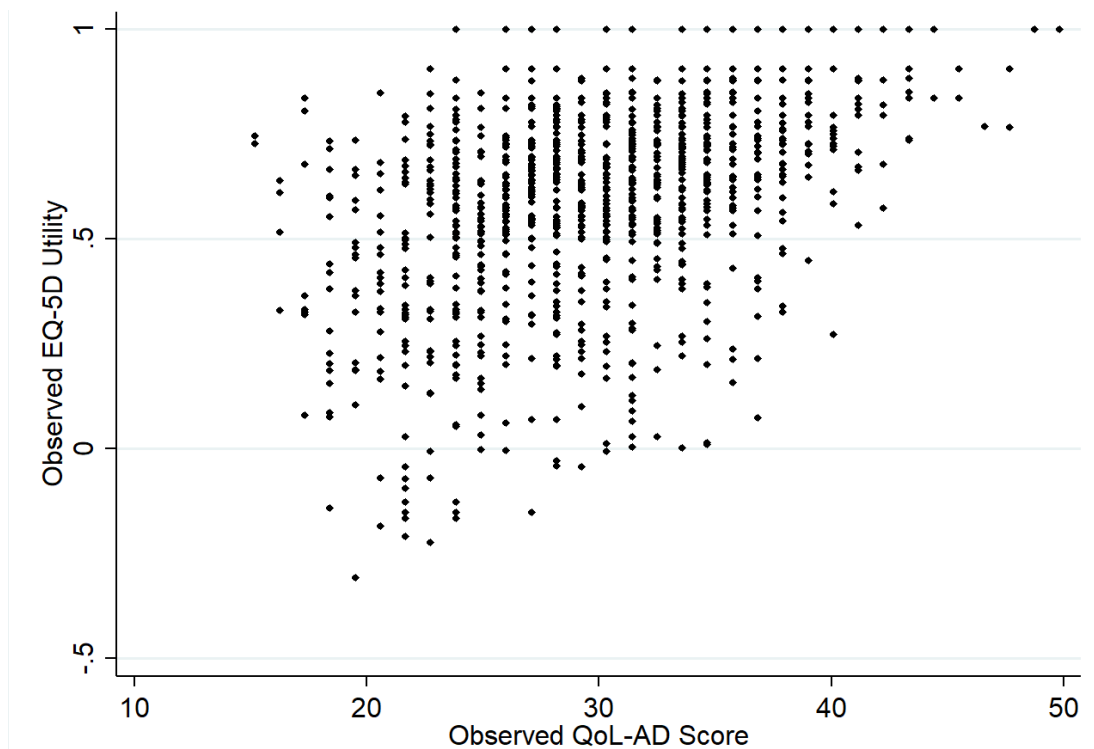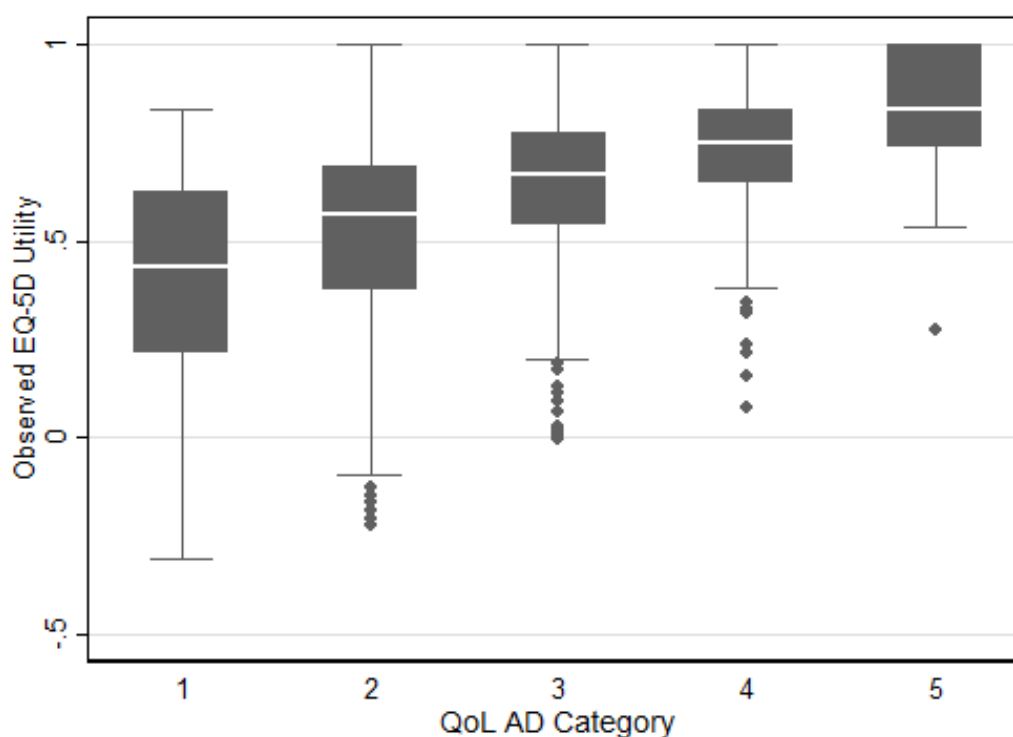

Figure 15. Observed proxy-rated QoL-AD vs. observed proxy-rated EQ-5D utilities

Table 12. Observed proxy-rated EQ-5D utility by proxy-rated QoL-AD category

QoL-AD Category	EQ-5D							
	Obs	Mean	SD	Min	Max	Median	25%	75%
1 (<20)	48	0.418	0.253	-0.307	0.837	0.431	0.217	0.625
2 (20-29)	479	0.522	0.244	-0.223	1	0.567	0.377	0.692
3 (30-34)	352	0.637	0.210	-0.005	1	0.671	0.543	0.776
4 (35-39)	149	0.729	0.182	0.074	1	0.747	0.649	0.837
5 (40-52)	71	0.836	0.144	0.273	1	0.837	0.740	1

Abbreviations: Obs – Observations, SD – Standard Deviation

Figure 16. Distribution of observed proxy-rated EQ-5D utilities by proxy-rated QoL-AD category

Table 13. Observed proxy-rated EQ-5D utilities by MMSE severity

MMSE Severity	Observed EQ-5D Utility							
	Obs	Mean	SD	Min	Max	Median	25%	75%
0	59	0.656	0.204	-0.690	1	0.689	0.552	0.768
1	329	0.656	0.199	-0.005	1	0.673	0.546	0.794
2	512	0.605	0.239	-0.166	1	0.644	0.479	0.777
3	51	0.575	0.221	0.069	0.906	0.612	0.407	0.733
Missing	148	0.465	0.305	-0.307	1	0.528	0.244	0.709

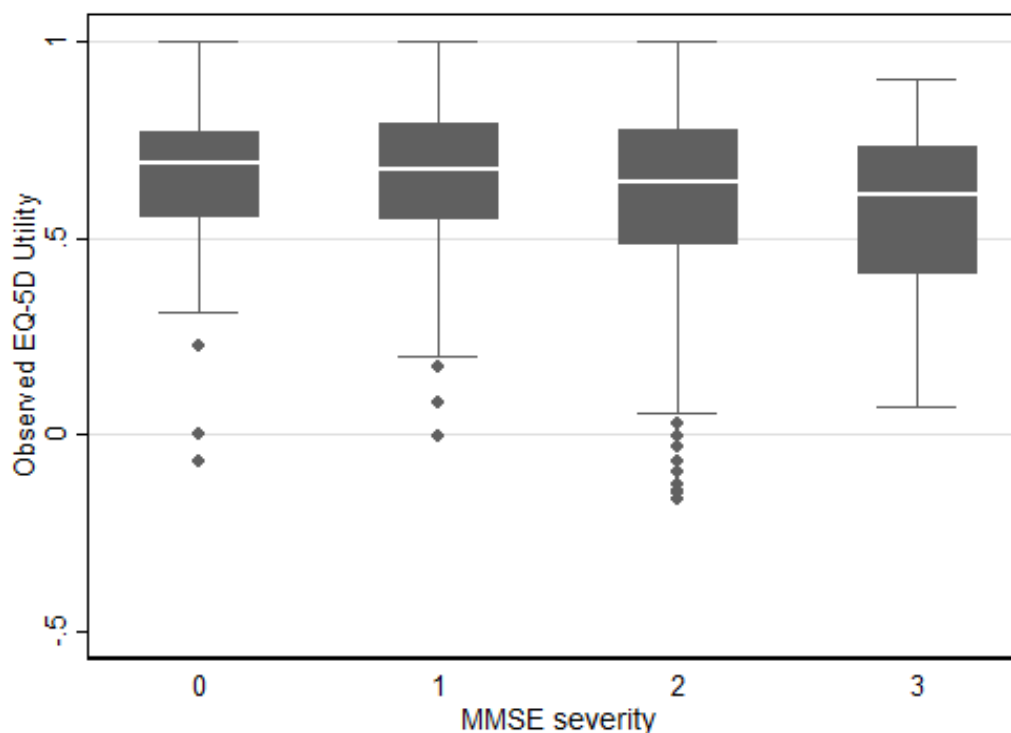*Abbreviations: Obs – Observations, SD – Standard Deviation*

Figure 17. Observed proxy-rated EQ-5D utilities by MMSE severity

Table 14. Observed proxy-rated EQ-5D utilities by CDR

CDR Score	Observed EQ-5D Utility							
	Obs	Mean	SD	Min	Max	Median	25%	75%
0.5	29	0.795	0.146	0.513	1	0.768	0.725	0.906
1	721	0.667	0.197	-0.166	1	0.691	0.558	0.796
2	297	0.478	0.257	-0.307	1	0.518	0.297	0.666
3	31	0.278	0.253	-0.151	0.768	0.233	0.100	0.421
Missing	21	0.389	0.307	-0.223	1	0.385	0.249	0.533

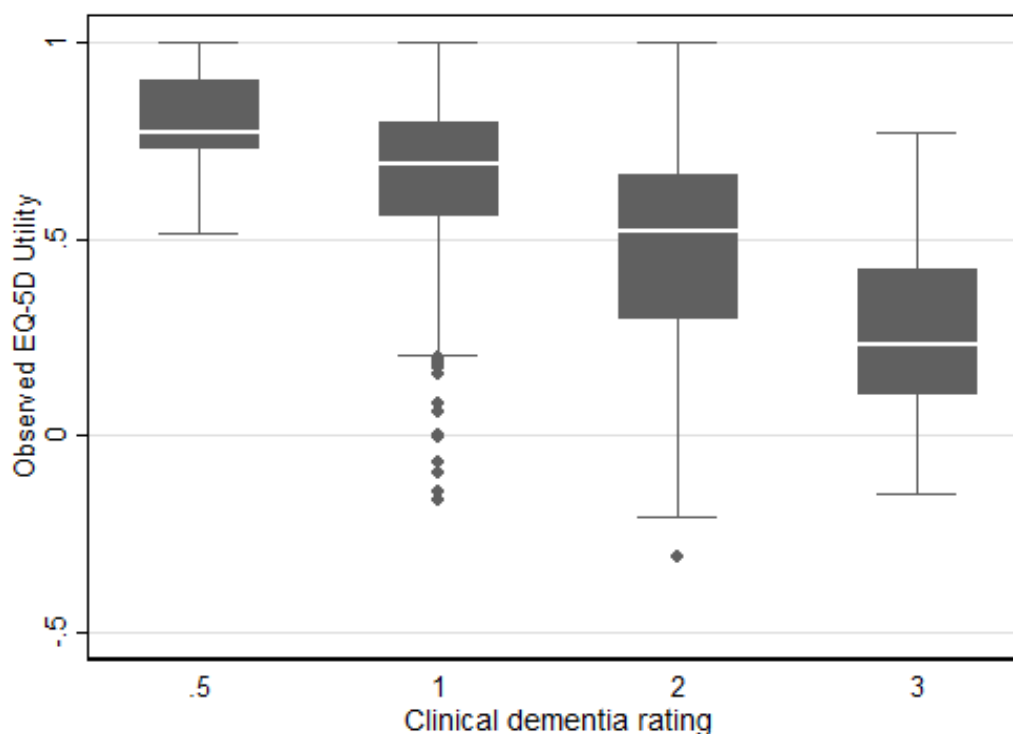*Abbreviations: Obs – Observations, SD – Standard Deviation*

Figure 18. Observed proxy-rated EQ-5D utilities by CDR

3.2.5. Correlation between individual EQ-5D-5L and QoL-AD items

We also considered questions that referred to similar QoL aspects in the QoL-AD and EQ-5D-5L to explore consistency in responses (both self-rated). We present two examples.

Table 15 shows a cross-tabulation between the question on anxiety and depression in the EQ-5D-5L and the question on mood in the QoL-AD. This is an example of questions that appeared to correlate well. Correlation coefficient for the two items was -0.41 (Spearman's correlation, 95% Confidence Interval -0.46 to -0.36). Participants tended to answer 'good' on the QoL-AD if they had answered 'not anxious' or 'slightly anxious' on the EQ-5D-5L. Similarly, participants tended to answer 'poor' on the QoL-AD if they had answered 'severely anxious' or 'extremely anxious' on the EQ-5D-5L. This figure also shows, however, that the 'excellent' category on the QoL-AD was not frequently chosen even when participants reported no difficulties in the EQ-5D-5L questions. This feature was observed across the other questions, too.

Table 16 shows an example where expected good correlation between the questions did not materialise in the observed data. This related to a question regarding physical health on the QoL-AD and mobility on the EQ-5D-5L. Correlation coefficient for the two items was -0.39 [Spearman's correlation, -0.47 to -0.33 95% confidence interval]. Participants who reported that they were 'unable to walk' or had 'severe problems' with mobility reported answers across all QoL-AD categories. There was higher correlation between the questions for participants indicating 'no problems' or 'slight problems' in mobility.

Table 15. Cross-tabulation of EQ-5D-5L item 5 and QoL-AD Item 3

Anxiety/Depression	How has your mood been lately? Have your spirits been good, or have you been feeling down? Would you rate your mood as poor, fair, good, or excellent?				
	Poor	Fair	Good	Excellent	Total
I am not anxious or depressed	14	111	422	64	611
I am slightly anxious or depressed	22	104	139	7	272
I am moderately anxious or depressed	25	53	33	4	115
I am severely anxious or depressed	12	6	1	0	19
I am extremely anxious or depressed	3	0	0	0	3
Total	76	274	595	75	1,020

Table 16. Cross-tabulation of EQ-5D-5L item 1 and QoL-AD item 1

Mobility (EQ-5D)	How do you feel about your physical health? Would you say it's poor, fair, good, or excellent? (QoL-AD)				
	Poor	Fair	Good	Excellent	Total
I have no problems in walking about	17	117	357	91	582
I have slight problems in walking about	18	79	92	18	207
I have moderate problems in walking about	20	90	60	4	174
I have severe problems in walking about	19	20	10	1	50
I am unable to walk about	2	1	4	0	7
Total	76	307	523	114	1,020

3.3.   Comparison of the models

Table 17 shows the performance parameters (RMSE and MAE) for each mapping model for each scenario when age and gender are not included in the models. The lowest RMSE and MAE are highlighted in bold. This table indicates that the response mapping using the mlogit model produced the most accurate results.

RMSE and MAE are lowest for the scenario mapping the self-rated QoL-AD to the self-rated EQ-5D, indicating that EQ-5D scores can be produced with the highest accuracy in this scenario.

Table 17. Assessment of mapping algorithm performance (not taking into account age and gender)

Model	Self-rated QoL-AD → Self-rated EQ-5D		Proxy-rated QoL-AD → Self-rated EQ-5D		Self-rated QoL-AD → Proxy-rated EQ-5D		Proxy-rated QoL-AD → Proxy- rated EQ-5D	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Direct OLS Continuous	0.1812	0.1316	0.1956	0.1428	0.2192	0.1711	0.2145	0.1633
Direct OLS Categorical	0.1625	0.1209	0.1828	0.1314	0.2038	0.1591	0.1939	0.1487
Direct Tobit	0.1662	0.1214	0.1873	0.1329	0.2039	0.1587	0.1943	0.1486
Direct Clad	0.1712	0.1224	0.1879	0.1307	0.2114	0.1590	0.2018	0.1493
Direct 2-part	0.1623	0.1208	0.1827	0.1312	0.2035	0.1592	0.1935	0.1477
Response ologit	0.1634	0.1205	0.1832	0.1322	0.2046	0.1597	0.1950	0.1497
Response OLS Categorical	0.1801	0.1309	0.1992	0.1480	0.2140	0.1692	0.2081	0.1573
Response OLS Continuous	0.1945	0.1426	0.2017	0.1515	0.2201	0.1742	0.2154	0.1650
Response mlogit	**0.1410**	**0.1104**	**0.1611**	**0.1231**	**0.1939**	**0.1530**	**0.1861**	**0.1418**

Table 18 presents the performance parameters when the mapping algorithm includes age and sex. As before, the response mapping using a mlogit model performs best.

Generally, RMSEs and MAEs were lower in the mapping algorithm including age and gender (with the exception of the CLAD model for the model mapping the self-rated QoL-AD to the self-rated EQ-5D, and the OLS model mapping proxy-rated QoL-AD as a continuous variable to predict self-rated EQ-5D utilities), indicating higher accuracy of these models.

For the mlogit response mapping model, prediction accuracy generally improved slightly when age and gender were included in the model. For the mlogit model, the percentage difference in RMSE when taking into account age and sex was -4.4, -2.0, -2.1, -2.3 for the RMSEs and -3.6, -1.9, -2.1, -1.5 for the MAEs for each scenario, respectively, compared to the models excluding age and gender.

Table 18. Assessment of mapping algorithm performance (taking into account age and gender)

Model	Self-rated QoL-AD → Self-rated EQ-5D		Proxy- rated QoL-AD → Self- rated EQ-5D		Self- rated QoL-AD → Proxy-rated EQ-5D		Proxy- rated QoL-AD → Proxy- rated EQ-5D	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Direct OLS Continuous	0.1797	0.1302	0.1937	0.1413	0.2159	0.1686	0.2109	0.1614
Direct OLS Categorical	0.1614	0.1196	0.1809	0.1297	0.2008	0.1563	0.1916	0.1473
Direct Tobit	0.1651	0.1200	0.1851	0.1311	0.2010	0.1560	0.1919	0.1471
Direct Clad	0.1677	0.1195	0.1868	0.1306	0.2082	0.1563	0.2012	0.1497
Direct 2-part	0.1610	0.1192	0.1802	0.1291	0.2007	0.1563	0.1913	0.1465
Response Ologit	0.1623	0.1194	0.1808	0.1304	0.2016	0.1575	0.1923	0.1481
Response OLS Categorical	0.1765	0.1267	0.1986	0.1461	0.2111	0.1666	0.2062	0.1555
Response OLS Continuous	0.1913	0.1386	0.2015	0.1495	0.2171	0.1705	0.2140	0.1633
Response mlogit	**0.1348**	**0.1064**	**0.1579**	**0.1207**	**0.1898**	**0.1498**	**0.1818**	**0.1397**

Some problems were observed when estimating the statistical models. The CLAD model did not always converge. The maximum iterations run were set to 200 (400 for CLAD), and coefficient estimates obtained at this point were used in the mapping algorithm. There were issues with perfect prediction and resulting 'questionable standard errors' for some of the mlogit and ologit models.

For our final model (mlogit, including age and sex, i.e. the mapping algorithm that produced the most accurate results), we compared the RMSE and MAE for different subsets of the QoL-AD score. Cut-offs for the groups were at the median QoL-AD score, splitting the data into halves, and at the 25th and 75th centiles, splitting the data into three subsets. Table 19 shows the results for this investigation.

When the sample was split based on the median QoL-AD score, we found that our RMSE and MAE were lower for the top half for all scenarios apart from the self-rated QoL-AD mapped to proxy-rated EQ-5D.

When the sample was divided based on the interquartile range (IQR), we found that prediction accuracy was best for the upper quartile, then the IQR and worst for the lowest quartile for all but the self-rated QoL-AD mapped to proxy-rated EQ-5D, as above.

For the self-rated QoL-AD mapped to proxy-rated EQ-5D, best prediction accuracy was observed for the interquartile range, followed by the highest quartile, and worst for the lowest quartile.

The results suggest that the model on average predicts more accurately for higher values of QoL-AD.

Table 19. Assessment of final model across different portion of QoL-AD scores

	Self-rated QoL-AD → Self-rated EQ-5D			Proxy- rated QoL-AD → Self-rated EQ-5D			Self- rated QoL-AD → Proxy-rated EQ-5D			Proxy- rated QoL-AD → Proxy- rated EQ-5D		
	Lowest quartile	**IQR**	**Highest quartile**	**Lowest quartile**	**IQR**	**Highest quartile**	**Lowest quartile**	**IQR**	**Highest quartile**	**Lowest quartile**	**IQR**	**Highest quartile**
RSME	0.1570	0.1336	0.1146	0.1792	0.1571	0.1250	0.1925	0.1895	0.1884	0.2101	0.1772	0.1454
MAE	0.1213	0.1072	0.0925	0.1406	0.1185	0.0984	0.1516	0.1481	0.1514	0.1657	0.1366	0.1097
	Top 50%	**Bottom 50%**		**Top 50%**	**Bottom 50%**		**Top 50%**	**Bottom 50%**		**Top 50%**	**Bottom 50%**	
RSME	0.1223	0.1470		0.1250	0.1653		0.1900	0.1895		0.1630	0.2002	
MAE	0.0981	0.1153		0.0984	0.1264		0.1508	0.1487		0.1227	0.1581	

3.4. Predictive accuracy of the mlogit model for the different scenarios

The following graphs (Figure 19 - Figure 22) plot the predicted EQ-5D utilities (calculated via the mapping algorithm) versus the observed EQ-5D utilities for the response mlogit model for each scenario.

The majority of the observed data was at the higher end of the EQ-5D utility values.

In each graph, the dashed line indicates where predicted utilities are expected to lie if the mapping algorithm introduced no error in the predictions.

At lower observed values, over-prediction is seen with predicted values lying above the dashed line. For the higher observed values, overall, over-prediction and under-prediction can both be seen with values laying both above and below the dashed line. For the extreme higher observed values, mostly under-prediction is seen, especially with observed utility of 1 never being predicted a value of 1. This is due to model construct: For an utility score of one to be predicted, a participant would have to predicted outcomes of 'no problems' for all five domains. The mapping algorithm predicts the probability for each participants of falling into each of the five item levels. The model never predicts a probability of 1 for any participants falling into the 'no problems' level. Although probabilities close to 1 are observed, there is always a small probability of other responses being observed. This means that no utilities of 1 can be predicted by the model.

The graphs also show some variation in the utilities predicted by the mapping algorithm for a given observed utility.
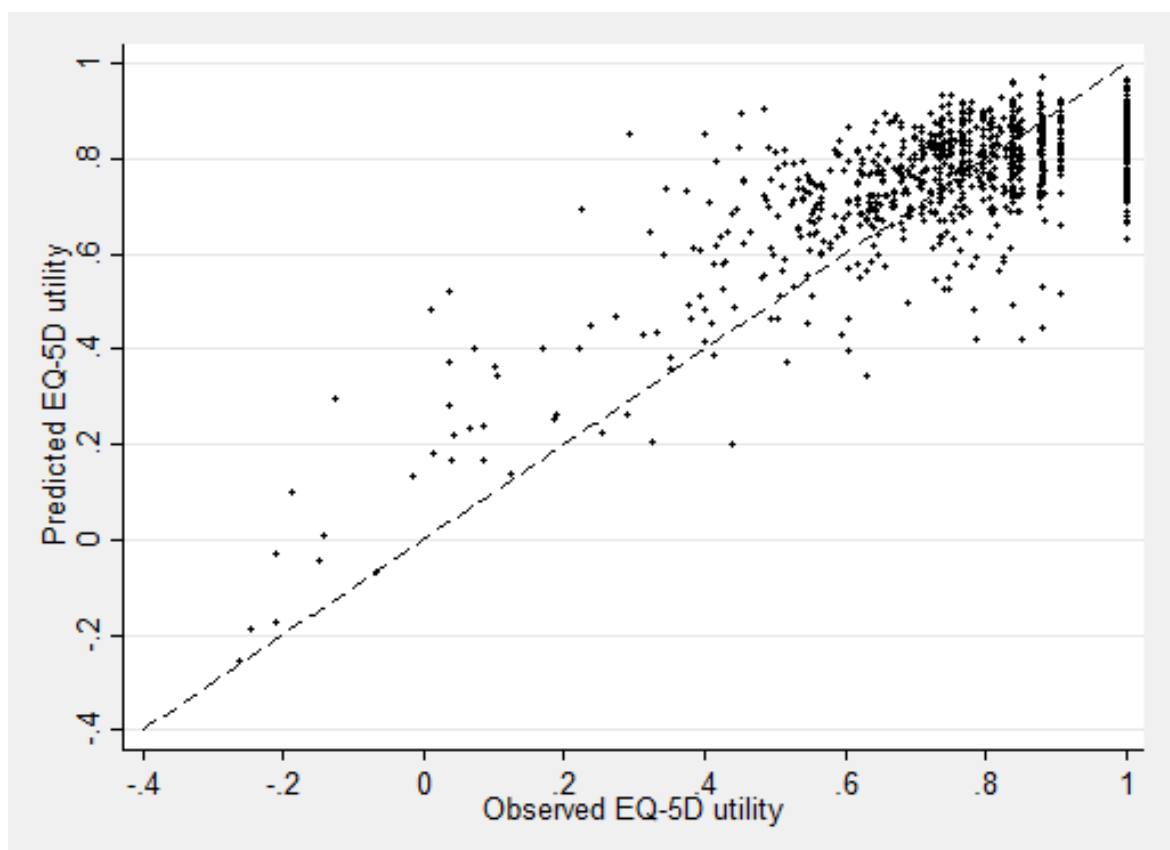
Figure 19. Predicted EQ-5D utilities vs. observed EQ-5D utilities for self-rated QoL mapped to self-rated EQ-5D using mlogit model
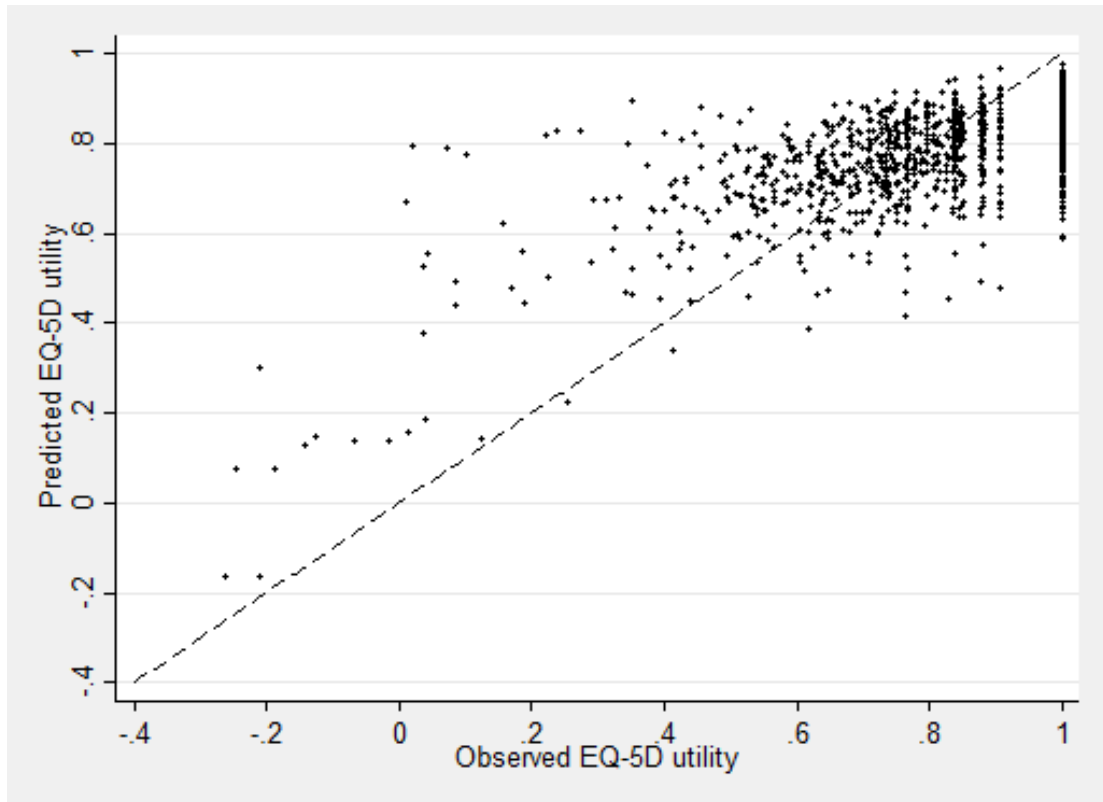
Figure 20. Predicted EQ-5D utilities vs. observed EQ-5D utilities for rroxy-rated QoL-AD mapped to self-rated EQ-5D using mlogit model
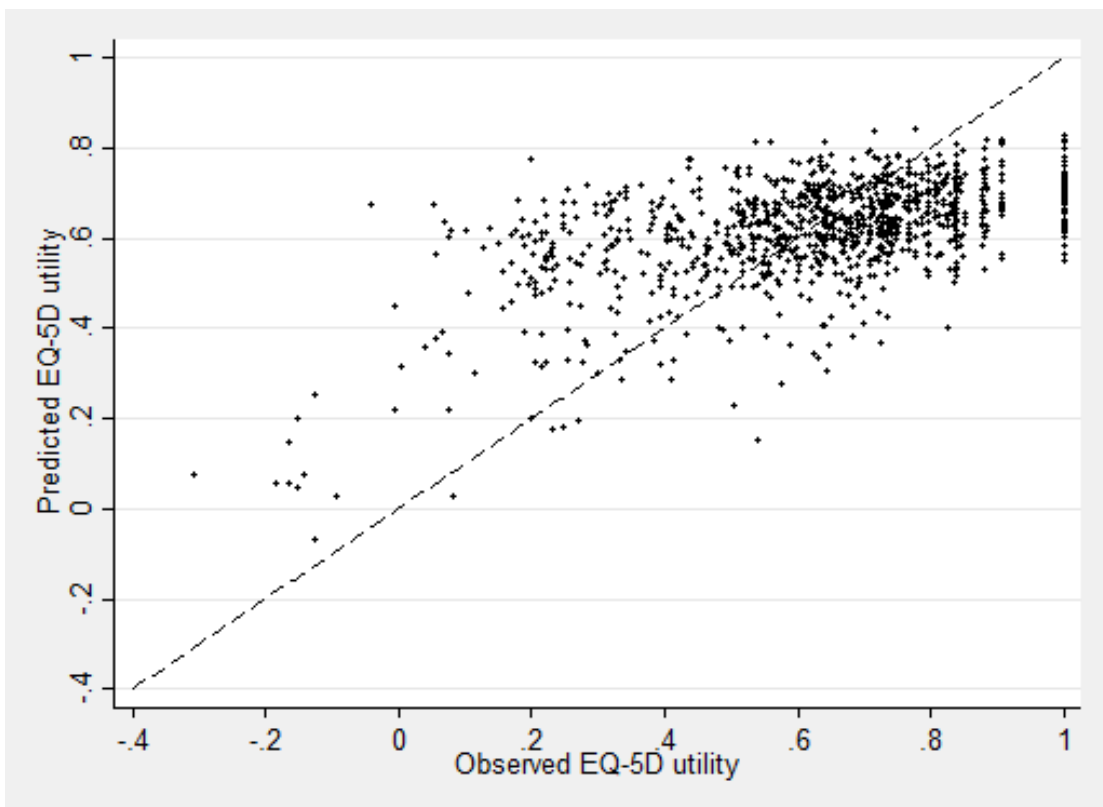

Figure 21. Predicted EQ-5D utilities vs. observed EQ-5D utilities for self-rated QoL-AD mapped to proxy-rated EQ-5D using mlogit model
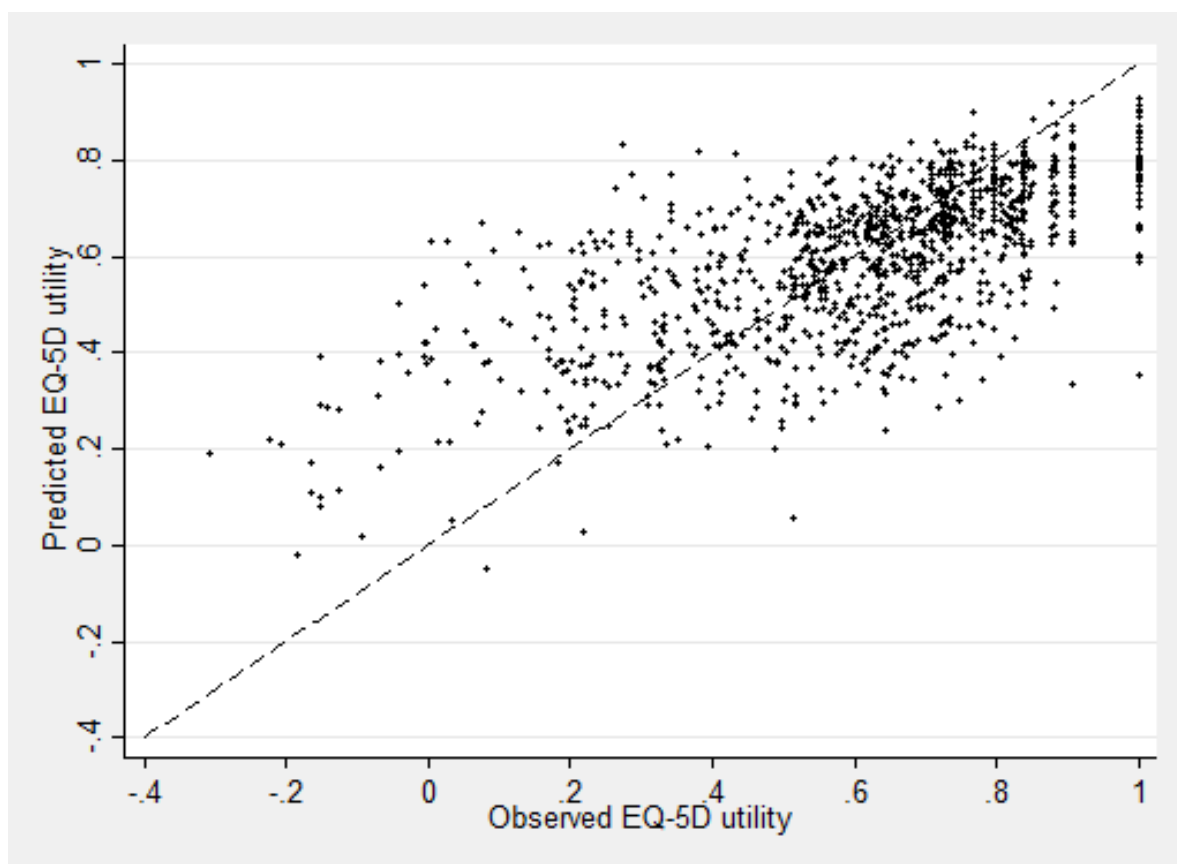
Figure 22. Predicted EQ-5D utilities against observed EQ-5D utilities for proxy-rated QoL-AD mapped to proxy-rated EQ-5D using final model

Table 20 shows the observed and predicted utilities (and standard errors) for the different scenarios overall and for a number of subgroups based on the observed utilities.

For the full sample, the observed means are very similar to the mean utility scores derived from the mapping algorithm for all scenarios.

Generally, utilities are over-estimated for observed utilities below 0.75 for self-rated EQ-5D-5L, and below 0.5 for proxy-rated EQ-5D-5L, while observed utilities above these cut-offs are under-estimated by the mapping algorithm.

Table 20. Means (standard deviaitons) for observed and predicted utility values

Observed EQ-5D range	Self-rated QoL-AD → Self-rated EQ-5D			Proxy-rated QoL-AD → Self-rated EQ-5D			Self-rated QoL-AD → Proxy-rated EQ-5D			Proxy-rated QoL-AD → Proxy-rated EQ-5D		
	N	Observed	Predicted	N	Observed	Predicted	N	Observed	Predicted	N	Observed	Predicted
Full sample	1,020	0.772 (0.209)	0.766 (0.150)	1,019	0.774 (0.207)	0.766 (0.125)	1,017	0.623 (0.227)	0.617 (0.118)	1,099	0.772 (0.209)	0.766 (0.150)
<0.25	30	0.016 (0.148)	0.212* (0.224)	28	0.021 (0.151)	0.387* (0.293)	84	0.1306 (0.129)	0.459* (0.195)	115	0.103 (0.135)	0.386* (0.163)
0.25 to <0.5	53	0.411 (0.063)	0.590* (0.174)	54	0.410 (0.063)	0.652* (0.133)	154	0.384 (0.070)	0.561* (0.116)	173	0.385 (0.068)	0.499* (0.140)
0.5 to <0.75	296	0.656 (0.071)	0.734* (0.099)	291	0.655 (0.071)	0.734* (0.088)	480	0.639 (0.071)	0.625# (0.090)	506	0.637 (0.072)	0.6048# (0.125)
0.75 to <1	414	0.830 (0.046)	0.805# (0.083)	412	0.830 (0.046)	0.793# (0.083)	239	0.823 (0.046)	0.673# (0.070)	248	0.824 (0.046)	0.706# (0.102)
1	227	1 (0)	0.852# (0.061)	234	1 (0)	0.828 # (0.073)	60	1 (0)	0.695# (0.064)	57	1 (0)	0.773# (0.097)

*indicates over-estimation, #indicates under-estimation

Table 21 below creates an overview of the mean and range of EQ-5D utility scores associated with a given QoL-AD score, and present the effect that a change in observed QoL-AD scores have on the mapped utility values.

The same QoL-AD scores can be arrived at through the combination of different item scores (scores are calculated by summation of the individual items), and hence are associated with a range of utilities derived from the mapping algorithm. The QoL-AD values in the table below are based on those observed in the mapping dataset, and therefore are likely combinations of QoL-AD items. Values were chosen to represent the range of the observed QoL-AD scores.

For example, for an observed self-rated QoL-AD score of 30.3, the mean predicted self-rated EQ-5D utility is 0.714. However, depending on the answers to the individual items, predicted self-rated EQ-5D utilities range from 0.357 to 0.848. For an approximate 10 point increase in the self-rated QoL-AD score to 40, the mean predicted self-rated EQ-5D utility is 0.842 (i.e. a 0.128 increase), with predicted self-rated EQ-5D utilities ranging from 0.451 to 0.919.

For the highest QoL-AD score of 52, achieved only as a self-rated score, not a proxy-rated score, the mean predicted EQ-5D utility is 0.965 (minimum 0.961, maximum 0.968) for self-rated EQ-5D scores, and 0.822 (minimum 0.815, maximum 0.830) for proxy-rated EQ-5D scores.

Table 21. Mapped EQ-5D utilities for observed QoL-AD scores

Observed QoL-AD score*	Self-rated QoL-AD → Self-rated EQ-5D Mean (min, max) N**	Proxy-rated QoL-AD → Self-rated EQ-5D Mean (min, max) N**	Self-rated QoL-AD → Proxy-rated EQ-5D Mean (min, max) N**	Proxy-rated QoL-AD → Proxy-rated EQ-5D Mean (min, max) N**
16.3	0.121 (0.008, 0.234) N=2	0.678 (0.643, 0.707) N=4	0.253 (0.078, 0.428) N=2	0.365 (0.240, 0.516) N=4
24.9	0.632 (0.429, 0.780) N=19	0.730 (0.460, 0.900) N=45	0.547 (0.299, 0.686) N=19	0.514 (0.054, 0.7478) N=50
30.3	0.714 (0.357, 0.848) N=51	0.773 (0.520, 0.902) N=56	0.607 (0.277, 0.742) N=51	0.587 (0.214, 0.788) N=59
32.5	0.727 (0.098, 0.895) N=54	0.757 (0.138, 0.906) N=54	0.122 (0.048, 0.731) N=55	0.640 (0.213, 0.764) N=57
40.0	0.842 (0.451, 0.919) N=63	0.873 (0.785, 0.956) N=23	0.636 (0.058, 0.821) N=63	0.778 (0.594, 0.869) N=24
45.5	0.892 (0.820, 0.972) N=14	0.931 (0.891, 0.970) N=2	0.741 (0.633, 0.840) N=14	0.845 (0.820, 0.869) N=2
52.0	0.965 (0.961, 0.968) N=2	n/a	0.822 (0.815, 0.830) N=2	n/a

QoL-AD scores were calculated without item 7 and rescaled to range from 13-52
**N indicates the total number of participant form whom the QoL-AD score was observed*
n/a indicates that the QoL-AD score was not observed in the dataset.

3.5. Predictive accuracy of the mlogit model for the individual items

The mlogit model generates probabilities of an individual falling into each of the five levels of a specific EQ-5D-5L domain, rather than predicting a specific response.

The following graphs show the mean probabilities of a participant falling into each of the 5 levels given the level they were observed to fall into.

For example, the first plot on the first graph shows that on average, participants who indicated no problems with mobility were given a 68% probability of falling into this category (level), approximately 18% probability of falling into category 2 (slight problems), 12% probability of falling into level 3 (moderate problems), and very low probabilities of falling into the levels indicating severe problems and inability to walk about (2% and < 1%, respectively).

The graphs are displayed for all four mapping scenarios.

3.5.1. Predictive accuracy of the mlogit model for the individual items – self-rated QoL-AD mapped to self-rated EQ-5D-5L

Figure 23 shows that the model performs well for predicting outcomes for participants indicating both no problems with mobility, and for those unable to walk. Some level of error is generally observed. While there is a high probability that for with no observed problems with mobility, the correct category will be predicted (probability approximately 70%), there is some probability of the other categories being predicted by the model. For those with slight to severe problems, the model tends to predict that the participants have fewer mobility problems than observed, although more severe mobility problems than observed are also predicted.

Those with observed severe mobility problems, there is a similar probability of no, slight, moderate or severe problems being predicted.

Cases where less severe problems than observed are predicted with lead to an over-prediction in the EQ-5D index (i.e. the predicted index will be higher than that observed), and where the predicted mobility problems are more severe than those observed, the predicted index will be under-predicted, i.e. will be lower than that observed.
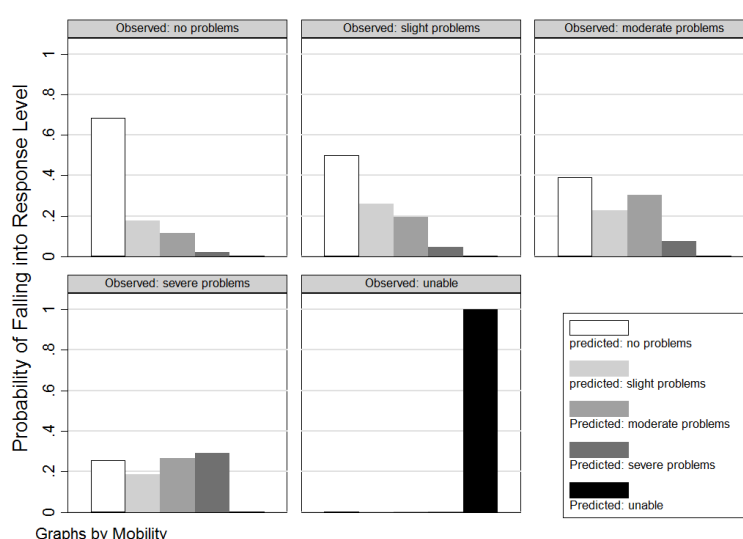
Figure 23. Probability of predicting each response level for a given observed response to EQ-5D-5L item 1.

Figure 24 shows high probabilities for all participants to be placed in the 'no problems' category by the mapping algorithm for the question on self-care.
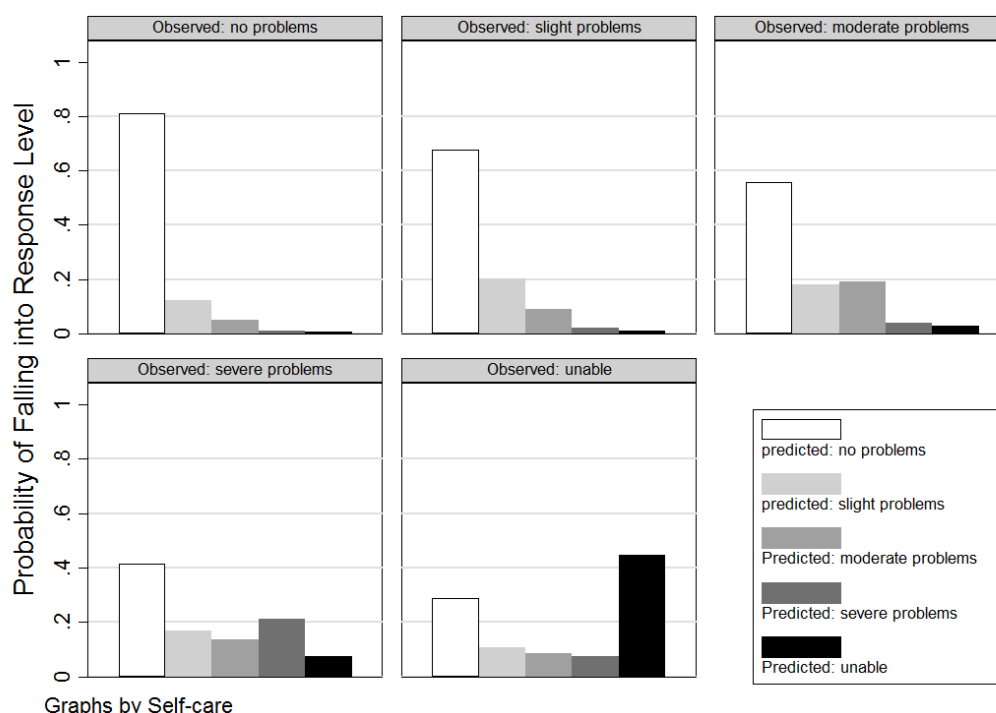
Figure 24. Probability of predicting each response level for a given observed response to EQ-5D-5L item 2

Figure 25 indicates good predictive ability for the usual activities question of the EQ-5D-5L. Participants in the no problems, severe problems and those unable to perform usual activities have high chances of being placed in the correct (i.e. observed) categories. Fewer problems with usual activities than observed are predicted for those with, moderate and severe problems.
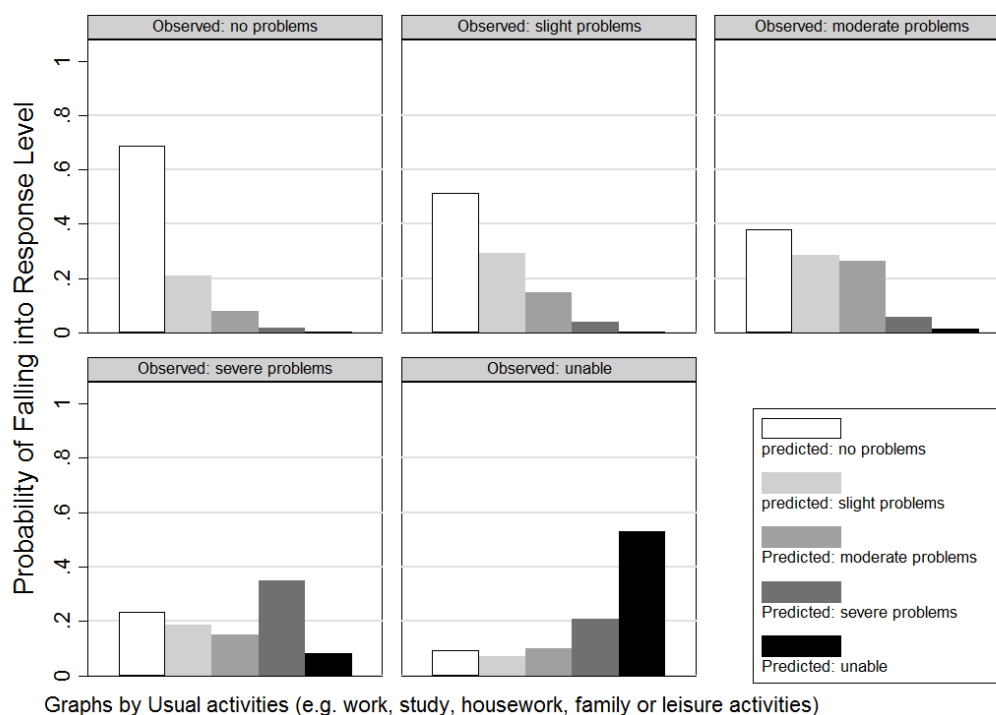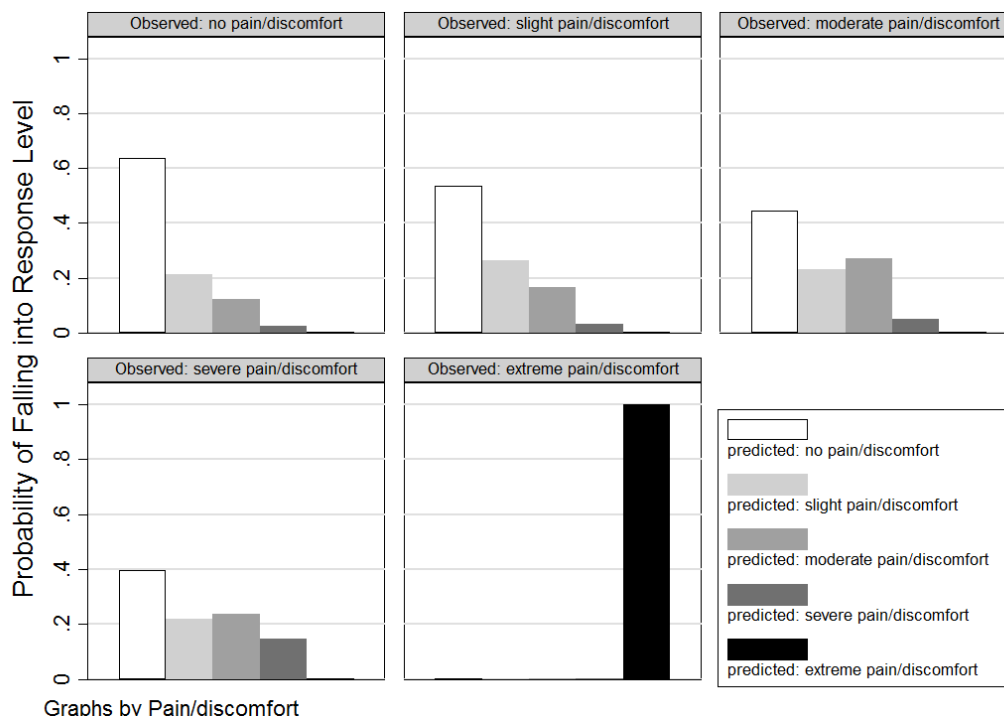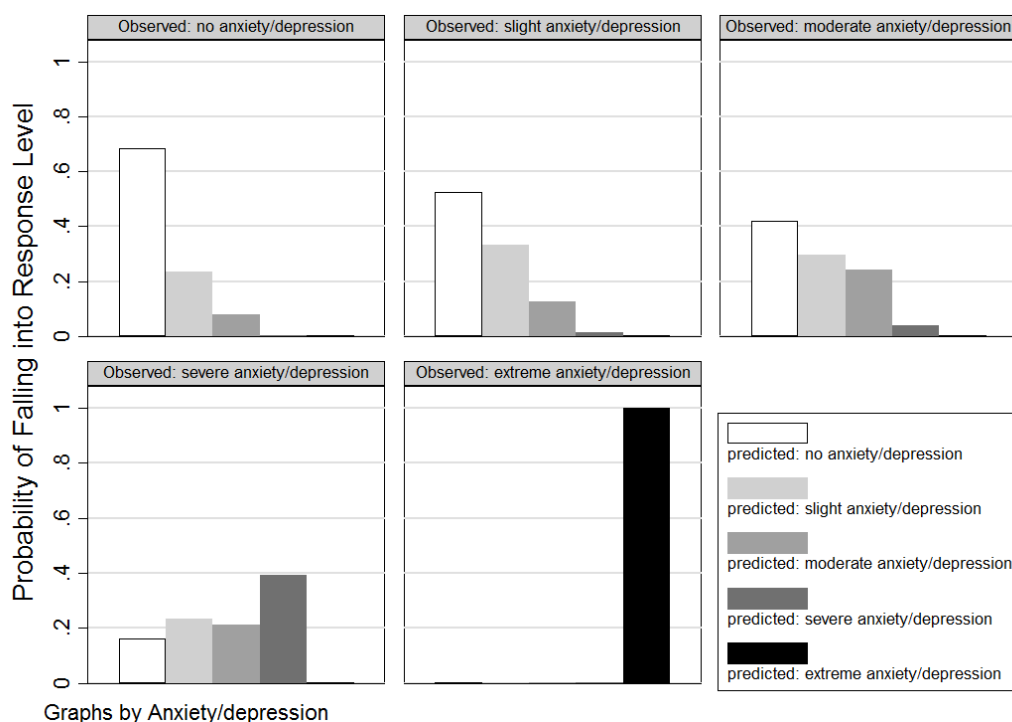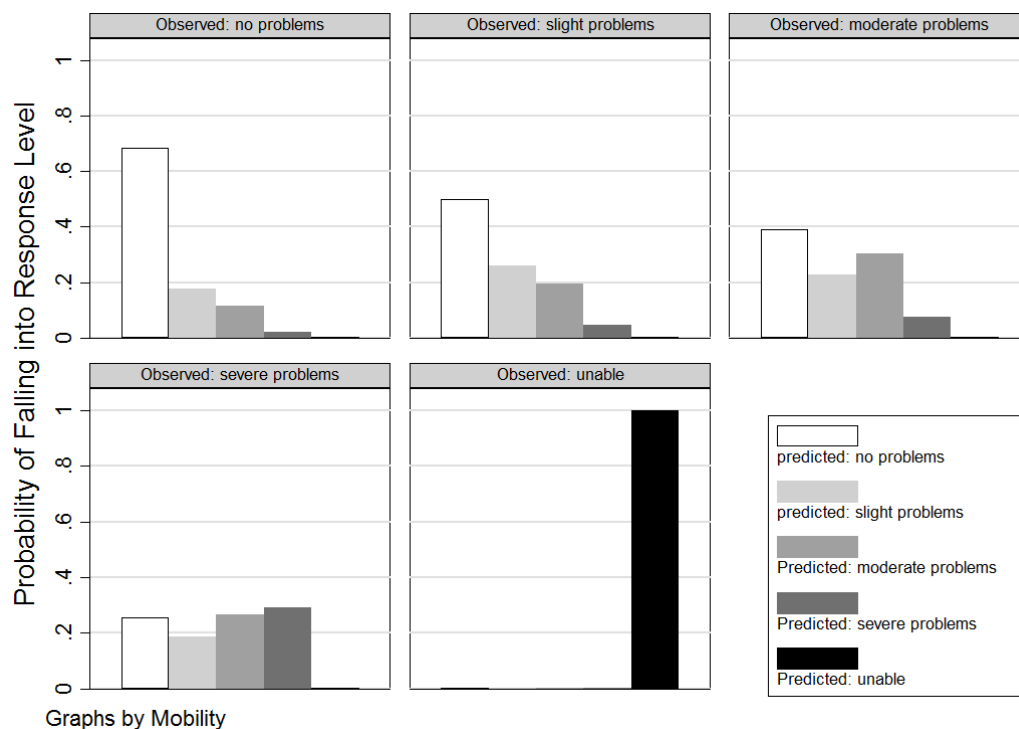

Figure 25. Probability of predicting each response level for a given observed response to EQ-5D-5L item 3

Figure 26 also demonstrates good discriminative ability of the model for those with no and extreme pain and discomfort; the model tends to predict that participants have less pain and discomfort than observed for the other categories.

Figure 26. Probability of predicting each response level for a given observed response to EQ-5D-5L item 4

Figure 27 shows similar pattern to those observed previously for participants with slight, moderate and severe anxiety.

Figure 27. Probability of predicting each response level for a given observed response to EQ-5D-5L item 5

Similar graphs (Figure 28 to Figure 42) for the other mapping scenarios are shown below, with similar patterns being demonstrated.

3.5.2. Predictive accuracy of the mlogit model for the individual items – proxy-rated QoL-AD mapped to self-rated EQ-5D-5L

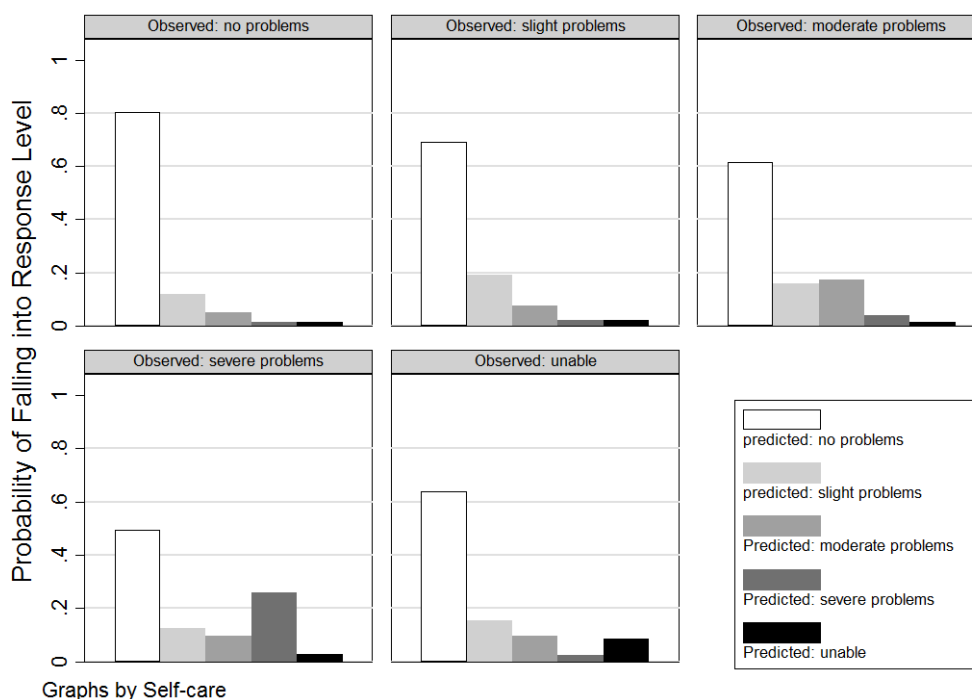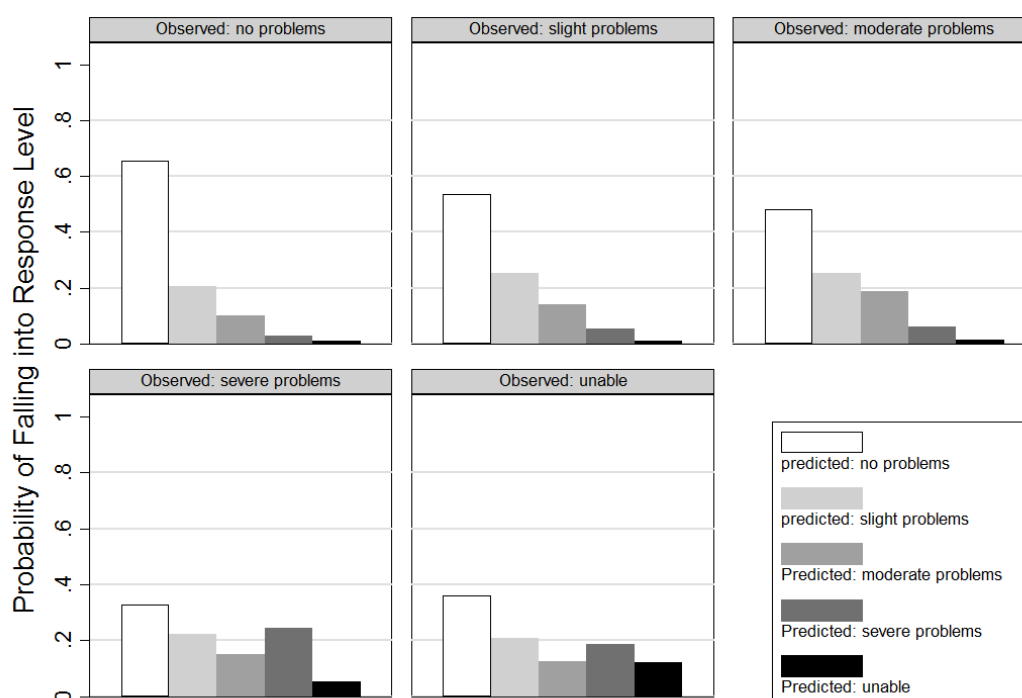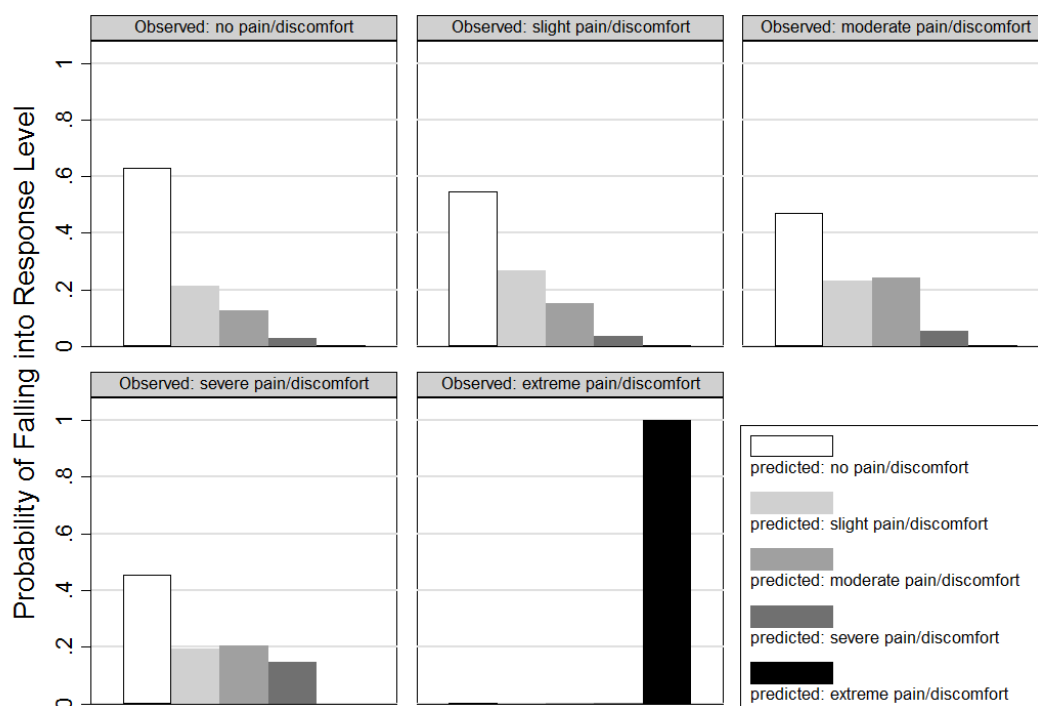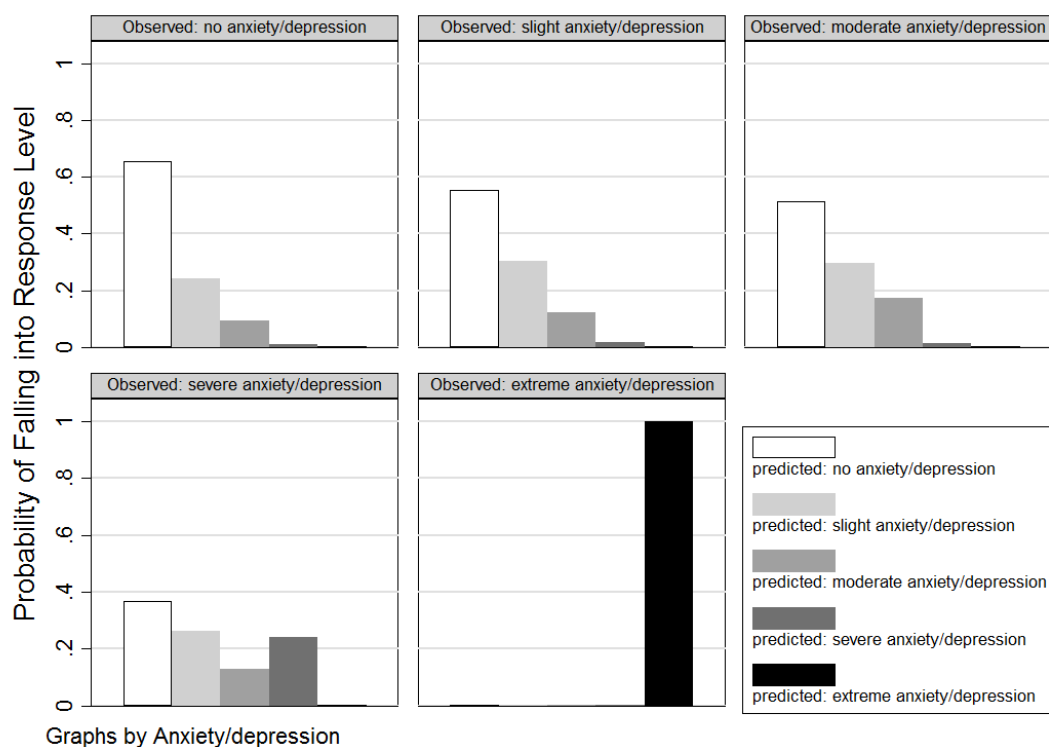**Figure 28. Probability of predicting each response level for a given observed response to EQ-5D-5L item 1**

Figure 29. Probability of predicting each response level for a given observed response to EQ-5D-5L item 2

Figure 30. Probability of predicting each response level for a given observed response to EQ-5D-5L item 3

Figure 31. Probability of predicting each response level for a given observed response to EQ-5D-5L item 4

Figure 32. Probability of predicting each response level for a given observed response to EQ-5D-5L item 5

3.5.3. Predictive accuracy of the mlogit model for the individual items – self-rated QoL-AD mapped to proxy-rated EQ-5D-5L

Figure 33. Probability of predicting each response level for a given observed response to EQ-5D-5L item 1
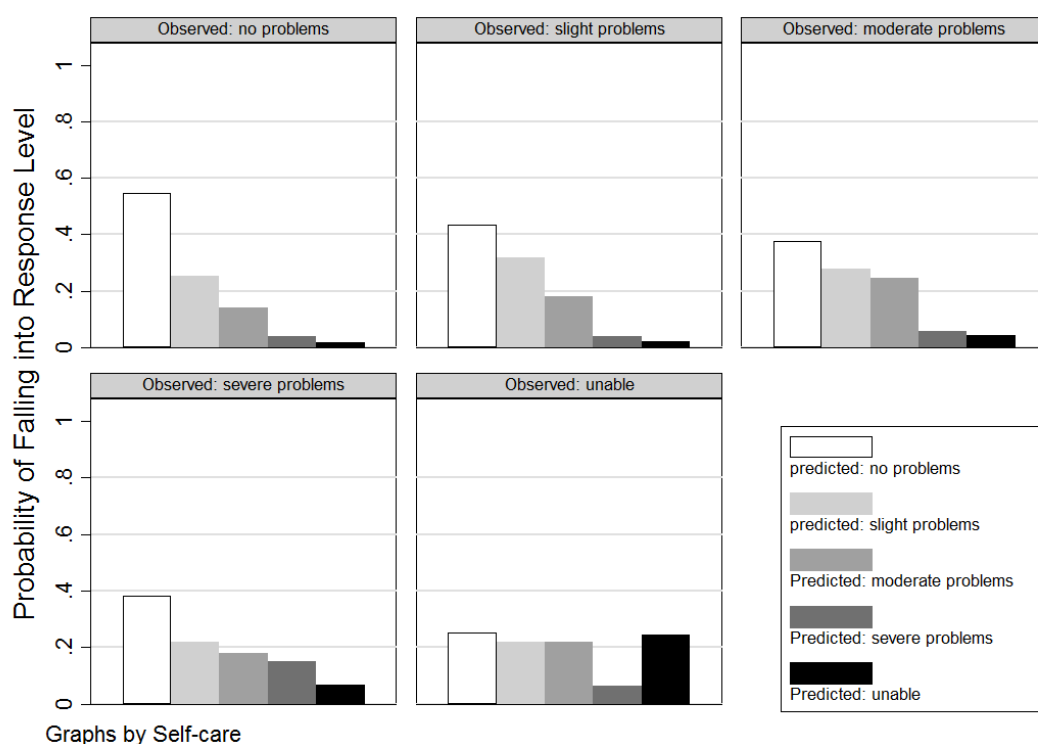
Figure 34. Probability of predicting each response level for a given observed response to EQ-5D-5L item 2
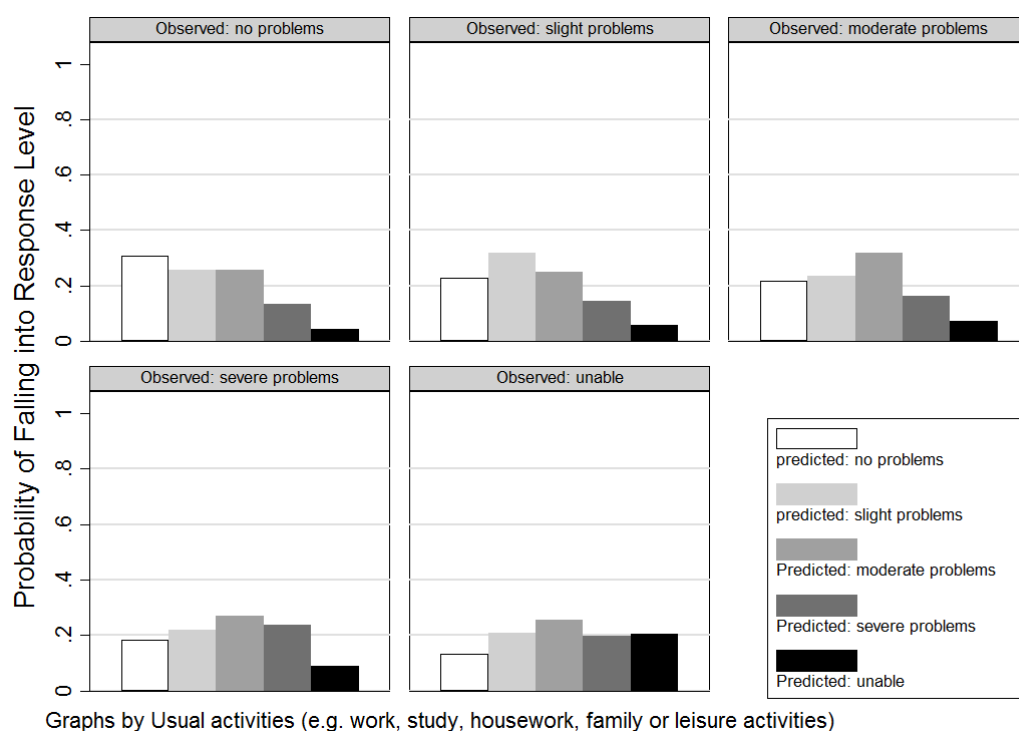

Figure 35. Probability of predicting each response level for a given observed response to EQ-5D-5L item 3

Figure 36. Probability of predicting each response level for a given observed response to EQ-5D-5L item 4

Figure 37. Probability of predicting each response level for a given observed response to EQ-5D-5L item 5

3.5.4. Predictive accuracy of the mlogit model for the individual items – proxy-rated QoL-AD mapped to proxy-rated EQ-5D-5L

Figure 38. Probability of predicting each response level for a given observed response to EQ-5D-5L item 1
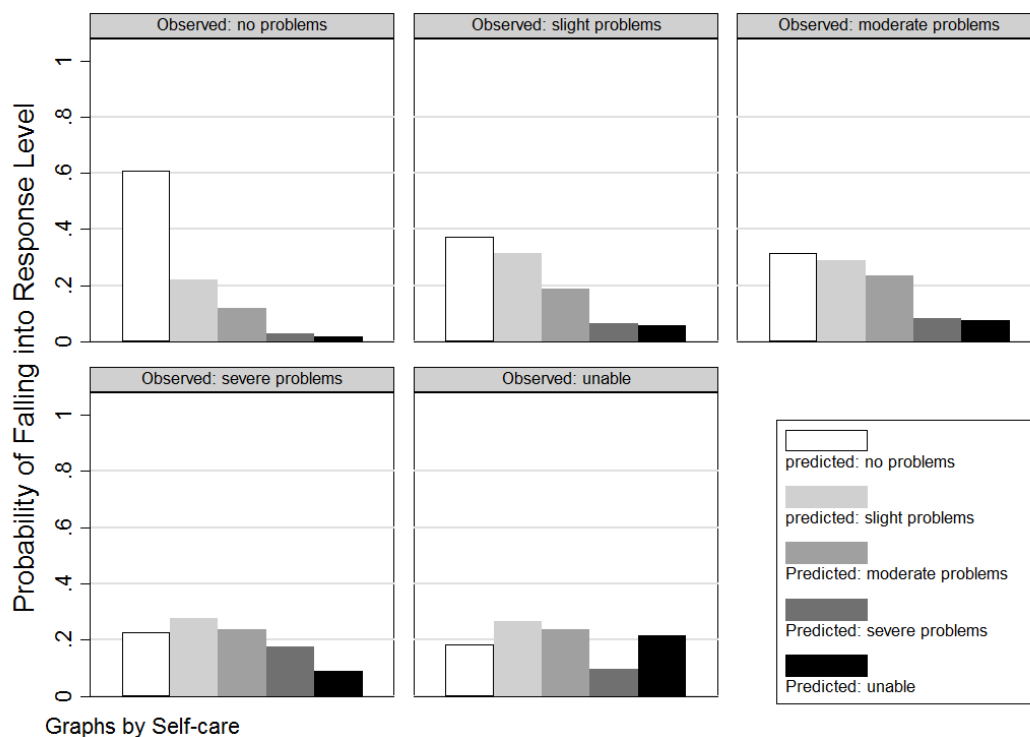
Figure 39. Probability of predicting each response level for a given observed response to EQ-5D-5L item 2
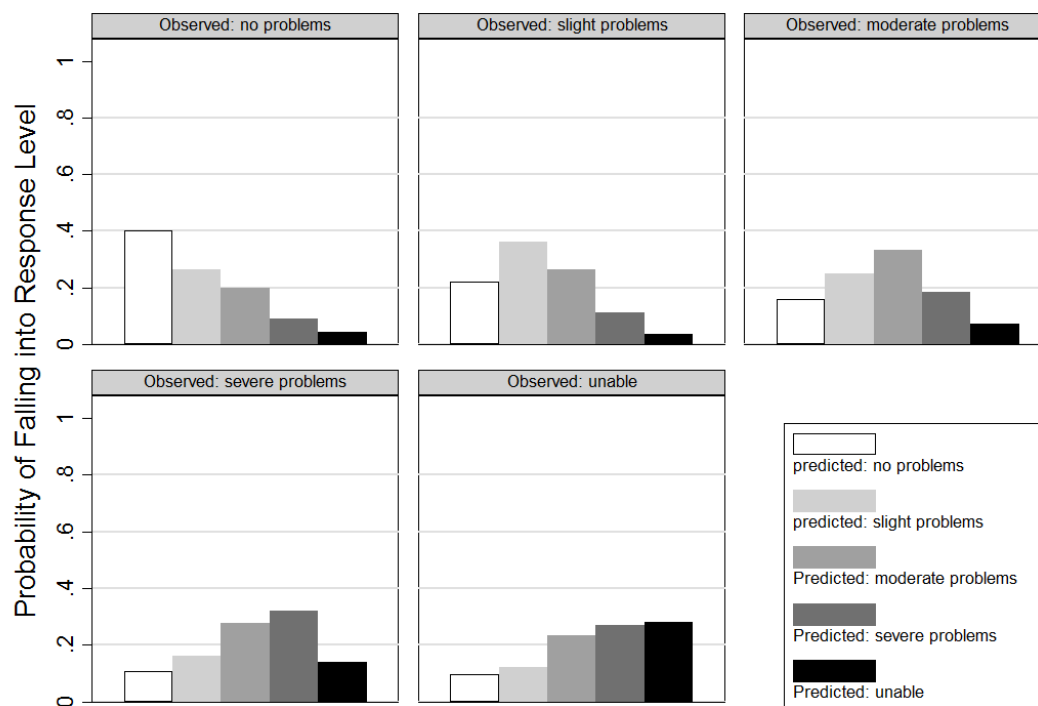
Figure 40. Probability of predicting each response level for a given observed response to EQ-5D-5L item 3
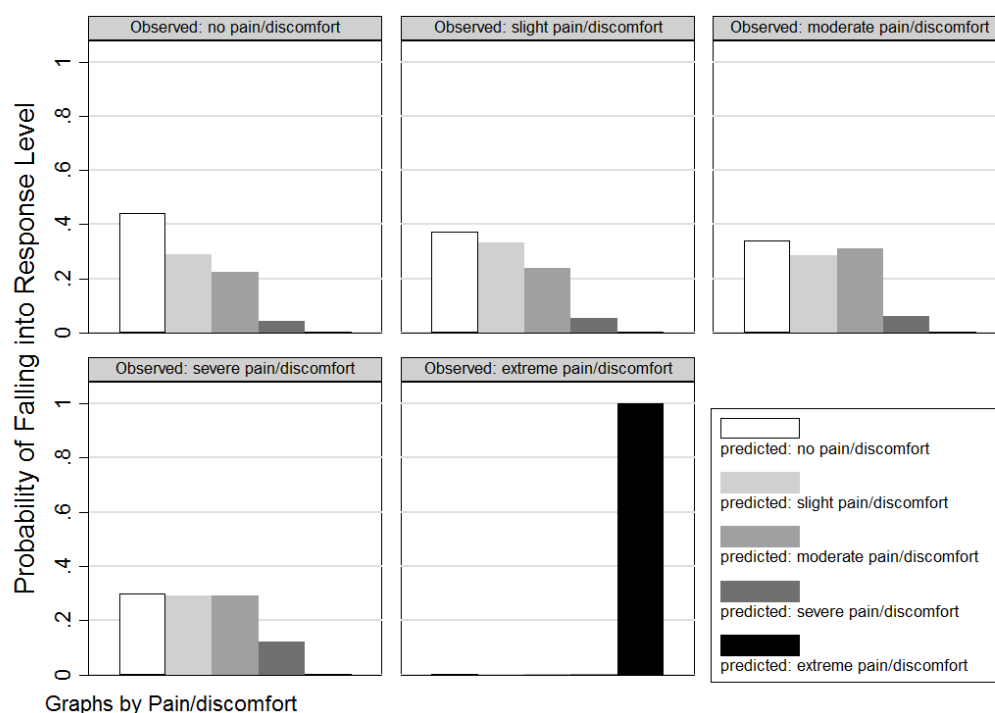

Figure 41. Probability of predicting each response level for a given observed response to EQ-5D-5L item 4
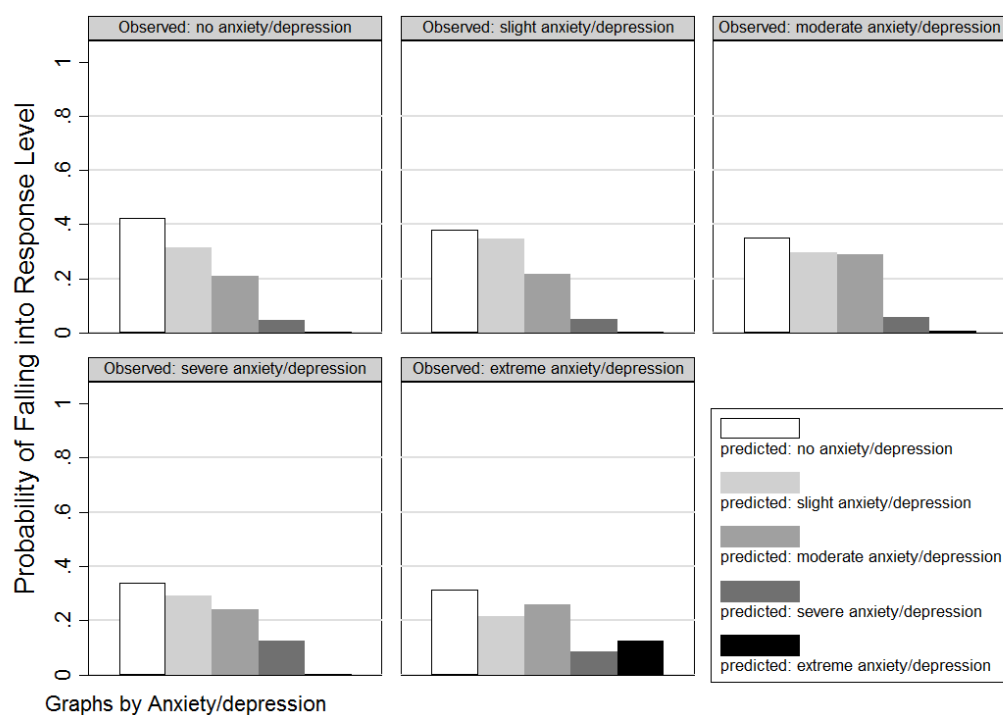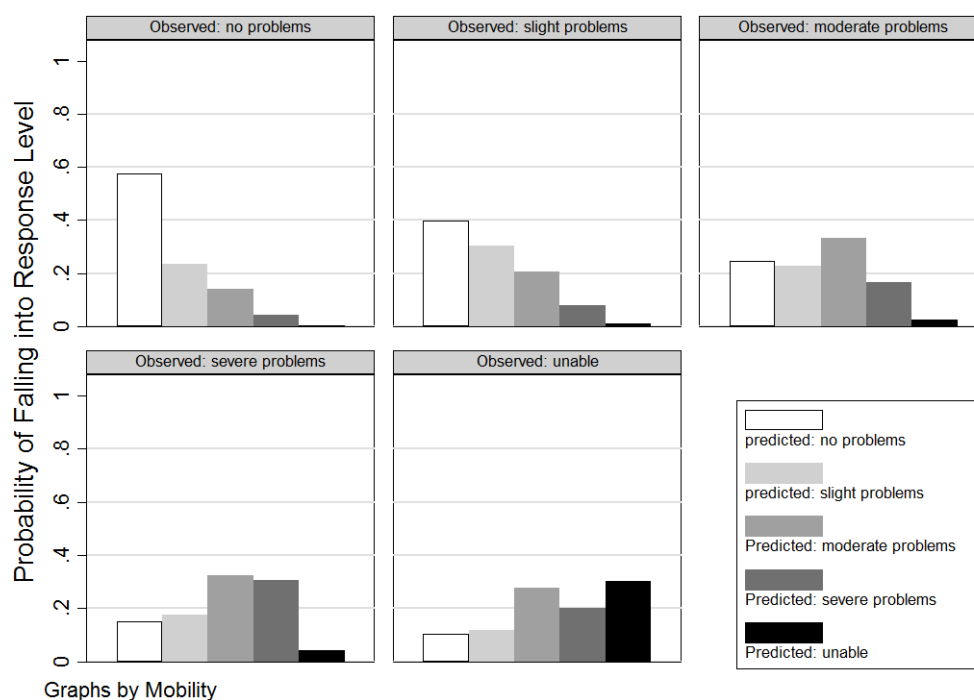
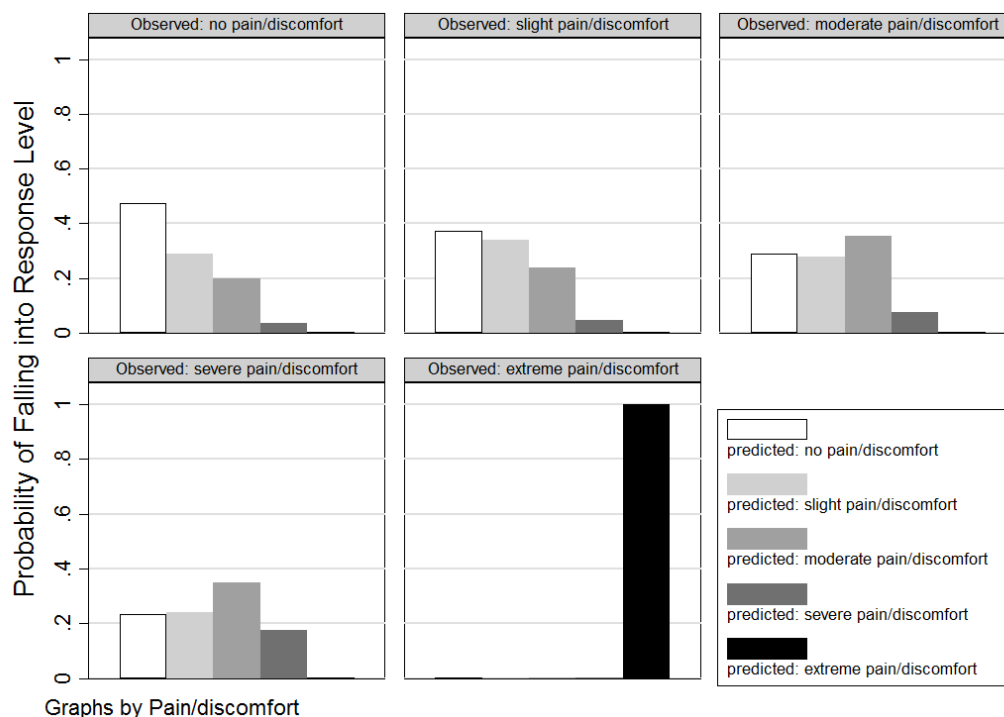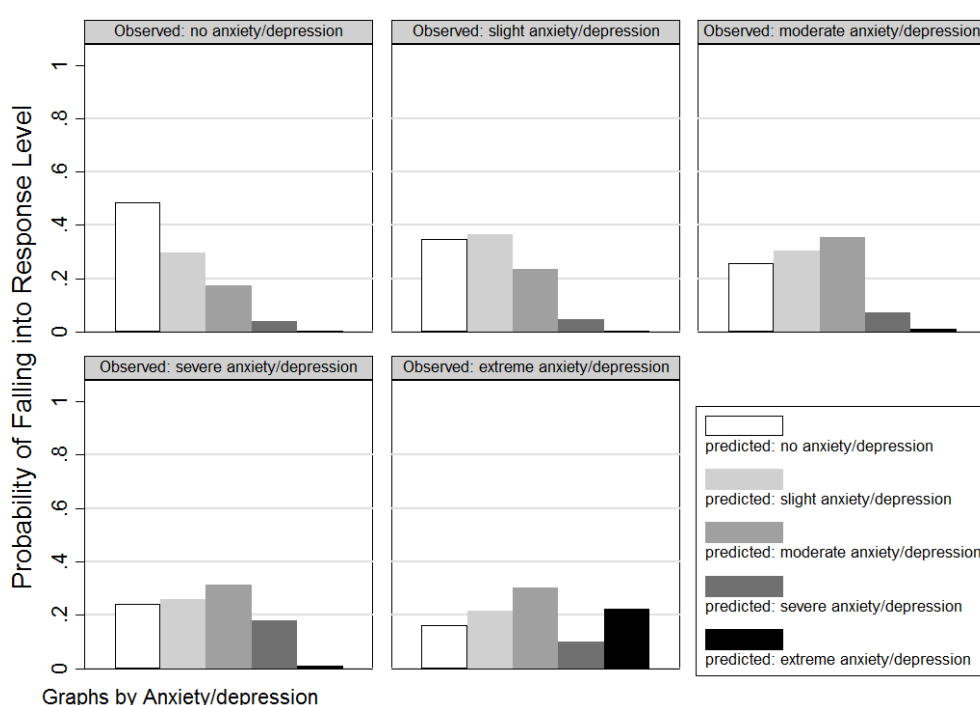*Figure 42. Probability of predicting each response level for a given observed response to EQ-5D-5L item 5*

Stata code to apply the QoL-AD to EQ-5D-5L mapping algorithms and output for the mlogit models will be available to researchers on request.

3.6. External validation study

The external validation dataset used contained 235 participants with up to 4 time points at which both the QoL-AD and EQ-5D-3L were completed by proxies. This has enabled us to validate the model for one of the scenarios (proxy-rated QoL-AD mapped to proxy-rated EQ-5D).

Error! Reference source not found. Table 22 shows the number of participants/observations included in the validation dataset and summaries of demographics (age, sex), Mini-Mental State Examination (MMSE) and Clinical Dementia Rating (CDR), mean/median utility based on the EQ-5D and QoL-AD scores for all observations in the dataset (i.e. the number of observations in this table exceed the number of participants). The QoL-AD and EQ-5D utilities are generally higher in this validation dataset compared to original dataset used for this mapping exercise. Generally, participants in the validation dataset had less severe dementia as measured by MMSE and CDR.

Table 22. Validation Dataset: demographics

Demographic Variable	Proxy- rated QoL-AD → Proxy- rated EQ-5D
Number of participants	235
Number of observations	631
PwD Age [1]	66.7 (9.4)
PwD Sex (Female) [1]	34.2% (216)

Proxy Age [1]	61.8 (10.8)
Proxy Sex (Female) [1]	72.7% (454)
MMSE[*][1]	25.3 (4.4)
CDR 0[**][1]	42 (10.8%)
CDR 0.5[**][1]	202 (51.9%)
CDR 1[**][1]	125 (32.1%)
CDR 2[**][1]	19 (4.9%)
CDR 3[**][1]	1 (<1%)
Proxy-rated QoL-AD mean (SD)[1]	32.0 (5.49)
Proxy-rated QoL-AD median (range)[1]	31.4 (15.2, 52)
Proxy-rated EQ-5D Utility mean (SD)[1]	0.774 (0.216)
Proxy-rated EQ-5D Utility median (range)[1]	0.812 (-0.095, 1)

[1]*summary data for all observations (each individual participant may have more than one response, N=631)*

[*]MMSE data available for 428 observations (32.2% missing) – means reported for available data

[**]CDR data available for 389 observations (38.4% missing) - means reported for available data

Figure 42 shows the observed EQ-5D utilities against the observed QoL-AD scores for the validation dataset.

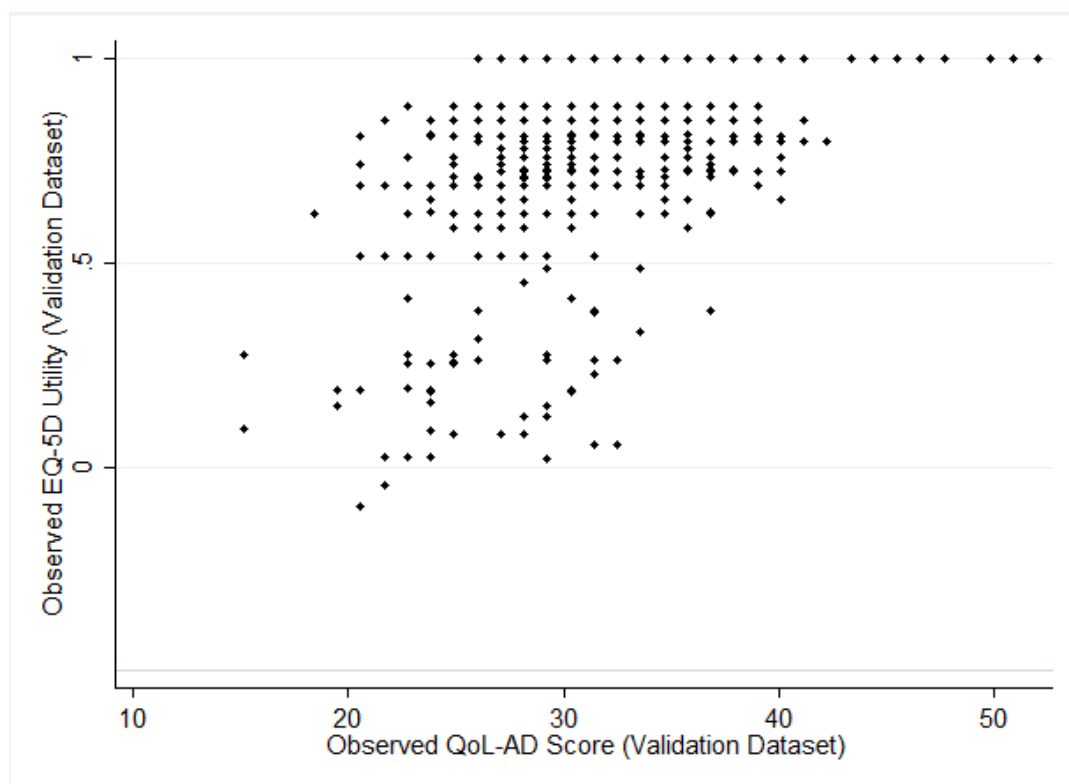The correlation coefficient is 0.56 (Spearman's correlation, 95% Confidence interval 0.50 to 0.61).

Figure 43. Validation dataset: observed proxy-rated EQ-5D and observed proxy-rated QoL-AD

Table 23 shows the EQ-5D observed utilities by QoL-AD category.

Table 23. Observed EQ-5D by QoL-AD category

QoL-AD Category	EQ-5D							
	Obs	Mean	SD	Min	Max	Median	25%	75%
1 (<20)	6	0.246	0.193	0.093	0.62	0.17	0.15	0.274
2 (20-29)	214	0.655	0.238	-0.095	1	0.691	0.516	0.812
3 (30-34)	224	0.800	0.177	0.055	1	0.812	0.725	0.883
4 (35-39)	150	0.887	0.118	0.383	1	0.883	0.796	1
5 (40-52)	37	0.935	0.105	0.656	1	1	0.848	1

Abbreviations: Obs – Observations, SD – Standard Deviation

The following tables (**Error! Reference source not found.**Table 24 and Table 25 show EQ-5D observed utility by MMSE and CDR category. The validation dataset has generally higher mean EQ-5D across all MMSE and CDR categories. The validation dataset has a higher proportion of missing values for MMSE and CDR compared to the original dataset used for mapping.

Table 24. Validation dataset: proxy-rated EQ-5D utility by MMSE severity

MMSE Severity	Observed EQ-5D Utility							
	Obs	Mean	SD	Min	Max	Median	25%	75%
0	222	0.817	0.173	0.020	1	0.814	0.725	1
1	151	0.747	2.215	0.079	1	0.796	0.689	0.883
2	52	0.731	0.255	-0.044	1	0.812	0.673	0.883
3	3	0.633	0.314	0.273	0.848	0.779	0.273	0.848
Missing	203	0.760	0.241	-0.095	1	0.812	.691	1

Abbreviations: Obs – Observations, SD – Standard Deviation

Table 25. Validation dataset: proxy-rated EQ-5D utility by CDR score

CDR Score	Observed EQ-5D Utility							
	Obs	Mean	SD	Min	Max	Median	25%	75%
0	42	0.890	0.111	0.656	1	0.8665	0.796	1
0.5	202	0.809	0.172	0.055	1	0.813	0.725	1
1	125	0.718	0.227	-0.044	1	0.743	0.623	0.850
2	19	0.701	0.227	0.273	1	0.779	0.516	0.883
3	1	0.150	n/a	0.150	0.150	0.150	0.150	0.150
Missing	242	0.762	0.241	-0.095	1	0.812	0.691	1

Abbreviations: Obs – Observations, SD – Standard Deviation

RSME and MAE for this validation study were 0.1981 and 0.1564, respectively, which are about one to 2 percent higher than the errors observed in the main mapping exercise for this scenario.

Table 26 shows that the model performs best for the highest quartile, as observed in the Actifcare dataset.

Table 26. RSME and MAE for different variations of QoL-AD score

	Proxy- rated QoL-AD → Proxy- rated EQ-5D		
	Lowest quartile	**IQR**	**Highest quartile**
RSME	0.2262	0.2030	0.1592
MAE	0.1772	0.1593	0.1330
	Top 50%	**Bottom 50%**	
RSME	0.1784	0.2211	
MAE	0.1429	0.1739	

Figure 44 below shows predicted vs. observed utilities in the validation study, mapping proxy-rated QoL-AD to proxy-rated EQ-5D-5L items.

The graph shows over-estimation of utilities below observed utilities of 0.5, and high amounts of under-estimation for high observed utilities.
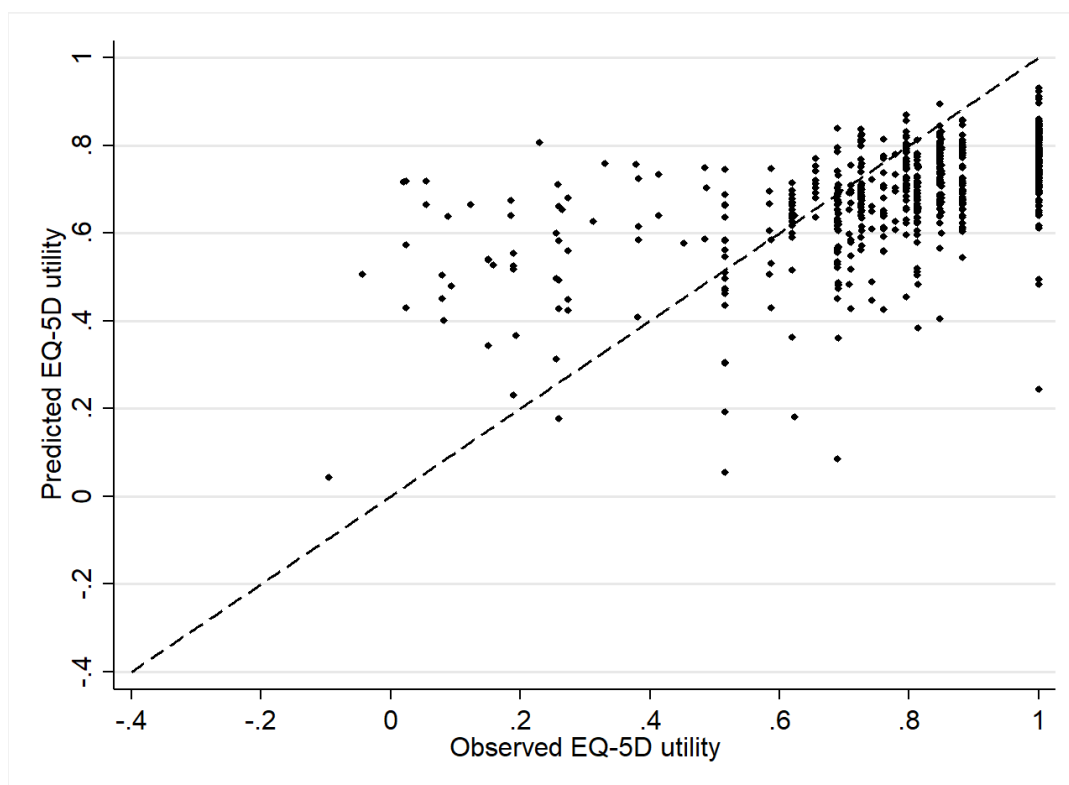
Figure 44. Validation dataset: scatter plot of predicted and observed Utilities

4. Discussion

We have developed mapping algorithms that will allow mapping of QoL-AD to the EQ-5D when reported by either the patient themselves or their proxies. The best model was chosen based on RMSE, while also taking into account practical considerations such as how widely it could be used based on the covariates included.

The RMSEs in our models were between 0.11 and 0.19, which is within ranges observed in a review of mapping studies (0.084 to 0.2)[31]. The MAEs of the models were between 0.106 and 0.150, which is also within the range of MAEs observed in the same review, where they ranged from 0.0011 to 0.19. MAE and RSME for the validation study were 0.1564 and 0.1981 respectively (both within the above ranges).

The recommended mapping algorithm to obtain EQ-5D-5L responses and utilities is a mlogit model that uses data from the QoL-AD items (excluding item 7), as well as the person with dementia's gender and age. Exploration of the association between observed and predicted EQ-5D indices, and the ranges of EQ-5D indices predicted for given QOL-AD scores indicate that the models are plausible with high face validity.

We have chosen to exclude QoL-AD item 7 responses from our models. This question relates to the participant's marriage, and may not be applicable to participants who are single, divorced or widowed. Our dataset and other examples in the literature show that this question is sometimes missed out. This approach means that our recommended mapping algorithm can be used even in scenarios where item 7 is unavailable, and does not require the imputation of such data. This approach also provided us with a larger dataset to work with to establish the model.

We have included age and sex in our recommended mapping algorithm. The RMSE and MAE decreased when we did this, i.e. the prediction accuracy was better when age and gender were included as covariates. As age and gender will commonly be available in studies that seek to implement this mapping algorithm, we do not anticipate problems arising from this added model complexity.

By generating a response-mapping algorithm, researchers will have access to the individual EQ-5D domains, and therefore can establish which domains are driving the observed utilities.

Out of the four scenarios we mapped, the models performed best for when self-rated QoL-AD mapped to the self-rated EQ-5D. This may indicate higher consistency between people with dementia's assessments using both questionnaires. The proxy-rated QoL-AD mapped to the proxy-rated EQ-5D had a higher RMSE, perhaps indicating a lower consistency in carers rating of the patient's QoL using either questionnaire.

By mapping for all different scenarios, we are enabling researchers to base the mapping on the data they have available (either self- or proxy-rated QoL-AD) and map it to self- or proxy-rated EQ-5D utilities as required for their work.

When applied to a validation study, the mapping algorithm for the proxy rated QOL-AD mapped to proxy-rated EQ-5D produced predictions with accuracy for EQ-5D utilities that were similar (though slightly higher) to those in the main dataset, indicating that the mapping algorithm for this scenario can validly be used in other datasets with similar patient

populations. Both datasets contained participants with suspected or known dementia living in the community with informal carers. The validation dataset had a higher proportion of male and older participants with less severe disease. The EQ-5D-3L was used in the validation dataset while the five level version was used in the original dataset.

This study has some limitations. A larger sample size might have improved precision in the estimates. In addition, a larger sample might have included more participants with lower observed QoL-AD and EQ-5D scores, hence providing predictions that are more accurate for participants with lower observed scores. Low number of observations in some of the QoL-AD and EQ-5D-5L categories led to issues around perfect prediction and 'questionable standard errors' for the mlogit and ologit models. These may have also been reduced with a larger sample size.

The mlogit model was found to be the best-fitting model, despite it not accounting for the clear order of the responses of the EQ-5D-5L responses, i.e. moving from participants being 'unable to perform' to 'having no problems'. The ologit model is able to take this into account, but the model assumes proportional odds, i.e. it assumes that the explanatory variables have the same effect on each cumulative split of the dependent variable. The ologit model was found to produce higher RMSEs than the mlogit model, and was therefore not chosen for the recommended mapping algorithm.

Our recommended mapping algorithm over-predicts outcomes for those with below median observed EQ-5D scores, and under-predicts outcomes for those with above median observed EQ-5D scores, as also seen with other studies [29,32]. Generally, our prediction accuracy was found to be best for participants with above median QoL-AD scores. The graphs presented in the results section reflect this by showing a higher variation in the size of the errors for lower observed utilities. The large amount of variability between the predicted and observed utility values is also likely to reflect the medium correlation between the QoL-AD and the EQ-5D utilities.

For both the Actifcare and validation datasets, the results showed that for participants who had a higher score of QoL-AD the RSME and MAE on average was lower compared to those with lower scores of QoL-AD. The MAE in the scenario mapping the self-rated QoL-AD to the proxy-rated EQ-5D was the only exception to this, and the MAE for the interquartile range was marginally lower than that of the highest quartile. Therefore, users of the algorithm will need to be aware of the prediction accuracy of utilities worsening as QoL-AD scores decrease. Generally, the mapping algorithm performs well for predicting outcomes for participants indicating either no problems, and for those unable to perform tasks or extreme problems. The model has lower discriminative ability for predicting outcomes for participants indicating slight, moderate and severe problems. Discriminative ability is also worse for proxy-rated outcomes compared to self-rated outcomes.

The use of additional explanatory variables may have improved the prediction accuracy of our model. However, as additional variables may not always be available in existing datasets, we felt that their inclusion may limit the applications of our mapping algorithm. Also, many other variables commonly collected in this disease area, including the MMSE and CDR have been shown to not correlate well with self-rated QoL scores[33-37], and were therefore were not used in the mapping algorithm.

To estimate EQ-5D utilities, we followed NICE guidance and used the cross-walk to the EQ-5D-3L. This position statement has been confirmed in November 2018[9]

Due to the validation dataset using EQ-5D-3L rather than the five level version, we were unable to assess the diagnostic ability of the model to predict individual responses in the validation.

Our recommended mapping algorithm is a response mapping model that uses the QoL-AD items as explanatory variables. As such, the mapping algorithm can only be used if item-level data are available, and not if researchers only have access to the composite QoL-AD scores. This may limit the implementation of our mapping algorithm somewhat. However, the prediction accuracy was reduced by at least 4% if continuous QoL-AD scores were used, and we therefore do not recommend that approach.

Our mapping algorithm currently does not provide estimates around the uncertainty for the predicted EQ-5D utilities, although these may be helpful when basing cost-effectiveness analyses on the mapped EQ-5D utilities. This is in line with other mapping studies. However, we do provide ranges and mean of predicted utility for different observed QoL-AD scores.

We took the approach of not using an internal validation set, because our sample size is insufficient to consider splitting the dataset. This is in line with guidance on for mapping to health utility states[38]The external validation dataset demonstrated consistency in the results of the mapping algorithm for the scenario that mapped proxy-rated QoL-AD to proxy-rated EQ-5D. High face validity was also demonstrated for the mapped EQ-5D values for all scenarios.

The different scenarios considered in this mapping exercise contained different numbers of observations, depending on the EQ-5D-5L and QoL-AD data available for each scenario. We believe that this was the most pragmatic approach to maximise the number of observations used in each scenario.

All model coefficients and standard errors are available on request; they are not included in appendix as the complexity of the mapping models results in too large an amount of coefficients to present meaningfully.

5. Conclusion and next steps

We have established a mapping algorithm to obtain predictions for both self-rated and proxy-rated EQ-5D utilities based on self-rated or proxy-rated QoL-AD scores.

This algorithm will allow researchers to estimate utilities if the QoL-AD but no EQ-5D-5L scores have been collected. However, for future research, the collection of the EQ-5D-5L alongside disease-specific measures is recommended wherever possible.

6. References

1.   Excellence NIfHaC. Dementia: assessment, management and support for people living with dementia and their carers, 2018.

2.   National Guideline A. National Institute for Health and Care Excellence: Clinical Guidelines. Dementia: Assessment, management and support for people living with dementia and their carers. London: National Institute for Health and Care Excellence (UK)
Copyright (c) National Institute for Health and Care Excellence 2018.; 2018.

3.   Prince M, Wimo A, Guerchet M, Ali G-C, Wu Y-T, Prina M. World Alzheimer Report 2015. The Global Imapct of Dementia. An Analysis of Prevalence, Incidence, Cost and Trends.: Alzeihmer's Disease International, 2015.

4.   Organisation WH. WHOQOL: Measuring Quality of Life. 2018. http://www.who.int/healthinfo/survey/whoqol-qualityoflife/en/ (accessed 29/08/2018.

5.   Schölzel-Dorenbos CJM. Quality of life in dementia: From concept to practice: Radboud University Nijmegen; 2011.

6.   Excellence NIfC. Guide to methods of technology appraisal 2013. 2013. https://www.nice.org.uk/process/pmg9/chapter/the-reference-case (accessed 09 2018).

7.   EQ-5D. EQ-5D Instruments | About EQ-5D. 2017. https://euroqol.org/eq-5d-instruments/2018).

8.   EQ-5D. EQ-5D-3L | Valuation. 2017. https://euroqol.org/eq-5d-instruments/eq-5d-3l-about/valuation/2018).

9.   NICE. Position statement on use of the EQ-5D-5L valuation set for England (updated November 2018). 2018 (accessed 29/11/2018 2018).

10.   Excellence NIfHaC. Position Statement on Use of the EQ-5D-5L Valuation Set. 2017.

11.   Logsdon R, Gibbons L, Susa. M, Teri L. Quality of Life in Alzheimer's Disease: Patient and Caregiver Reports. *Journal of Mental Health and Ageing* 1999; **5**(1).

12.   Bosboom PR, Alfonso H, Eaton J, Almeida OP. Quality of life in Alzheimer's disease: different factors associated with complementary ratings by patients and family carers. *International Psychogeriatrics* 2012; **24**(5): 708-21.

13.   Hoe J, Hancock G, Livingston G, Orrell M. Quality of life of people with dementia in residential care homes. *The British Journal of Psychiatry* 2006; **188**(5): 460-4.

14.   Hoe J, Katona C, Roch B, Livingston G. Use of the QOL-AD for measuring quality of life in people with severe dementia--the LASER-AD study. *Age Ageing* 2005; **34**(2): 130-5.

15.   Thorgrimsen L, Selwood A, Spector A, et al. Whose quality of life is it anyway? The validity and reliability of the Quality of Life-Alzheimer's Disease (QoL-AD) scale. *Alzheimer Dis Assoc Disord* 2003; **17**(4): 201-8.

16.   Bruvik FK, Ulstein ID, Ranhoff AH, Engedal K. The quality of life of people with dementia and their family carers. *Dementia and Geriatric Cognitive Disorders* 2012; **34**(1): 7-14.

17.   McCarney R, Fisher P, Iliffe S, et al. Ginkgo biloba for mild to moderate dementia in a community setting: a pragmatic, randomised, parallel-group, double-blind, placebo-controlled trial. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences* 2008; **23**(12): 1222-30.

18.   Meeuwsen EJ, Melis RJ, Van Der Aa GC, et al. Effectiveness of dementia follow-up care by memory clinics or general practitioners: randomised controlled trial. *Bmj* 2012; **344**: e3086.

19.   Spector A, Thorgrimsen L, Woods B, et al. Efficacy of an evidence-based cognitive stimulation therapy programme for people with dementia: randomised controlled trial. *The British Journal of Psychiatry* 2003; **183**(3): 248-54.

20.   Moniz-Cook E, Vernooij-Dassen M, Woods R, et al. A European consensus on outcome measures for psychosocial intervention research in dementia care. *Aging and Mental Health* 2008; **12**(1): 14-29.

21.   Mulhern B, Rowen D, Brazier J, et al. Development of DEMQOL-U and DEMQOL-

PROXY-U: generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. *Health technology assessment (Winchester, England)* 2013; **17**(5): v-xv, 1-140.

22.   Dakin H, Abel L, Burns R, Yang Y. Review and critical appraisal of studies mapping from quality of life or clinical measures to EQ-5D: an online database and application of the MAPS statement. *Health Qual Life Outcomes* 2018; **16**(1): 31.

23.   Petrou S, Rivero-Arias O, Dakin H, et al. The Maps Reporting Statement for Studies Mapping onto Generic Preference-Based outcome Measures. *Value Health* 2015; **18**(7): A715-6.

24.   Kerpershoek L, De Vugt M, Wolfs C, et al. Access to timely formal dementia care in Europe: protocol of the Actifcare (ACcess to Timely Formal Care) study. *BMC health services research* 2016; **16**(1): 423.

25.   Handels RL, Aalten P, Wolfs CA, et al. Diagnostic and economic evaluation of new biomarkers for Alzheimer's disease: the research protocol of a prospective cohort study. *BMC Neurol* 2012; **12**: 72.

26.   van Hout B, Janssen MF, Feng YS, et al. 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.

27.   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* 2018; **27**(1): 7-22.

28.   Feng Y, Devlin NJ, Shah KK, Mulhern B, van Hout B. New methods for modelling EQ-5D-5L value sets: An application to English data. *Health Econ* 2018; **27**(1): 23-38.

29.   Dakin H, Gray A, Murray D. Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. *Qual Life Res* 2013; **22**(3): 683-94.

30.   Naglie G, Hogan DB, Krahn M, et al. Predictors of patient self-ratings of quality of life in Alzheimer disease: cross-sectional results from the Canadian Alzheimer's Disease Quality of Life Study. *The American Journal of Geriatric Psychiatry* 2011; **19**(10): 881-90.

31.   Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *The European journal of health economics* 2010; **11**(2): 215-25.

32.   Dakin H, Petrou S, Haggard M, Benge S, Williamson I. Mapping analyses to estimate health utilities based on responses to the OM8-30 Otitis Media Questionnaire. *Qual Life Res* 2010; **19**(1): 65-80.

33.   Shearer J, Green C, Ritchie CW, Zajicek JP. Health state values for use in the economic evaluation of treatments for Alzheimer's disease. *Drugs & aging* 2012; **29**(1): 31-43.

34.   Banerjee S, Samsi K, Petrie CD, et al. What do we know about quality of life in dementia? A review of the emerging evidence on the predictive and explanatory value of disease specific measures of health related quality of life in people with dementia. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences* 2009; **24**(1): 15-24.

35.   Sheehan BD, Lall R, Stinton C, et al. Patient and proxy measurement of quality of life among general hospital in-patients with dementia. *Aging & mental health* 2012; **16**(5): 603-7.

36.   Vogel A, Mortensen EL, Hasselbalch SG, Andersen BB, Waldemar G. Patient versus informant reported quality of life in the earliest phases of Alzheimer's disease. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences* 2006; **21**(12): 1132-8.

37.   Lacey L, Bobula J, Rüdell K, Alvir J, Leibman C. Quality of life and utility measurement in a large clinical trial sample of patients with mild to moderate Alzheimer's disease: determinants and level of changes observed. *Value in Health* 2015; **18**(5): 638-45.

38.   Wailoo AJ, Hernandez-Alava M, Manca A, et al. Mapping to Estimate Health-State Utility from Non-Preference-Based Outcome Measures: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value Health* 2017; **20**(1): 18-27.

ANNEXES

ANNEX I. Results for the mapping study including item 7 of the QoL-AD

We also ran the different models for a dataset including item 7 and again by including age and sex. The best model for all datasets was the mlogit.

Table 27 shows that the best dataset for self-rated QoL-AD mapped to self-rated EQ-5D, proxy-rated QoL-AD mapped to self-rated EQ-5D, and self-rated QoL-AD mapped to proxy-rated EQ-5D was the dataset when including item 7 and when taking into account age and sex.

The second best dataset for the latter two was when item 7 was excluded, and age and sex included. For Self-QoL mapped to Self-EQ, however this was the third best with a RSME of 0.1348 and MAE of 0.1064.

For proxy-rated QoL-AD and proxy-rated EQ-5D, the best dataset was when item 7 was excluded and age and sex were taken into account, which is also our final model.

Given that by excluding item 7, we make it possible to include more observations in the mapping work, and that the difference in RMSE is not large, we chose to use this dataset as the best model consistently throughout all scenarios. Age and sex should be available for most if not all people, thus this has also been included.

Table 27. Assessment of best model (mlogit) comparing inclusion of item 7, age and sex

	Lowest RMSE	2nd Lowest RMSE	Lowest MAE	2nd Lowest MAE
Self-rated QoL-AD → Self-rated EQ-5D	Including item 7 with age and sex (0.1335)	Including item 7 without age and sex (0.1344)	Including item 7 With age and sex (0.1052)	Including item 7 without age and sex (0.1061)
Proxy-rated QoL-AD → Self-rated EQ-5D	Including item 7 with age and sex (0.1570)	Excluding item 7 with age and sex (0.1579)	Including item 7 with age and sex (0.1196)	Excluding item 7 with age and sex (0.1207)
Self-rated QoL-AD → Proxy-rated EQ-5D	Including item 7 with age and sex (0.1882)	Excluding item 7 with age and sex (0.1898)	Including item 7 with age and sex (0.1491)	Excluding item 7 with age and sex (0.1498)
Proxy-rated QoL-AD → Proxy-rated EQ-5D	Including item 7 with age and sex (0.1785)	Excluding item 7 with age and sex (0.1818)	Including item 7 with age and sex (0.1369)	Excluding item 7 with age and sex (0.1397)