Leeds Institute of Health Sciences
FACULTY OF MEDICINE AND HEALTH
UNIVERSITY OF LEEDS

AUHE Working Paper 2020/03

Limited evidence for reference dependent preferences in EQ-5D-3L valuation

Edward J.D. Webb*† John L. O'Dwyer† David Meads† Paul Kind† Penny Wright‡

February 2020

This working paper should cited as: Webb EJD, O'Dwyer JL, Meads D, Kind P and Wright P. Limited evidence for reference dependent preferences in EQ-5D-3L valuation. Leeds: University of Leeds, Academic Unit of Health Economics; 2020. Available from: medicinehealth.leeds.ac.uk/downloads/download/170/auhe_resources

*Corresponding author: e.j.d.webb@leeds.ac.uk, Worsley Building, Clarendon Way, Leeds, LS2 9NL, UK
†Academic Unit of Health Economics, Leeds Institute of Health Sciences, University of Leeds
‡Leeds Institute of Medical Research at St James's, University of Leeds, Leeds

DISCLAIMER

The series enables staff and student researchers based at or affiliated with the AUHE to make recent work and work in progress available to a wider audience. The work and ideas reported here may not always represent the final position and as such may sometimes need to be treated as work in progress. The material and views expressed in the series are solely those of the authors and should not be quoted without their permission.

Additional information

Funding

This work was supported by the Medical Research Council and the National Institute for Health Research (NIHR). Grant title: HQLC Dementia Carers Instrument Development: DECIDE (MR/M025179/1).

Ethical approval

Ethical approval was received from the Research Ethics Committee, Faculty of Medicine and Health, University of Leeds (approval reference MREC16-085). Consent was obtained electronically from survey participants.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgements

We would like to thank Janet Oyebode, Carol Brayne, Hareth Al-Janabi, Alan Tennant, Linda Clare and Zoe Hoare as well as survey participants.

Abstract

Background: The study aimed to test whether preferences for EQ-5D-3L health states depend on the reference point of an individual's own health state and whether carers' preferences depend on the reference point of the person they care for.

Methods: A sample of the UK general population and a sample of carers completed a discrete choice experiment (DCE) valuing the EQ-5D-3L instrument. Participants completed EQ-5D-3L for themselves and carers also completed a proxy version for the person they care for. DCE responses were analysed using mixed logit models. Valuations were transformed from a latent scale to a full health=1, dead=0 scale using responses from a visual analogue scale exercise.

Results: In contrast to previous findings, robust evidence for reference dependence was found only in the general population for mobility. On the 1-0 scale, it was found that only small effects in QALY terms lie within the 95% confidence interval of the reference dependence model parameters. For carers, no significant reference dependence on the health state of the people they care for was observed.

Conclusion: Limited evidence was found of reference dependence on either own health or the health of a person being cared for when valuing EQ-5D.

Keywords: Reference dependence; loss aversion; valuation; discrete choice experiment; EQ-5D

JEL Codes: I0, D01

1 Introduction

EQ-5D is a commonly used measure of health-related quality of life (HRQoL) and is a tool that often performs a critical role in health policy and funding decisions (National Institute for Health and Care Excellence (NICE), 2013). It measures health along 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, with each dimension having either 3 or 5 levels, depending on which version of the instrument is used. Individuals' levels on each dimension together form a health state, denoted for example 12132, with a total of either 243 states for EQ-5D-3L, the 3-level version, or 3125 states for EQ-5D-5L, the 5-level version. Values can be assigned to each health state on a scale anchored at 1, equivalent to full health, i.e. the state 11111, and 0, equivalent to death. From that it is possible to calculate quality-adjusted life years (QALYs). Given the crucial part it plays in decisions with wide-ranging financial and welfare consequences, it is vital to understand the process by which individuals value an EQ-5D state.

There exists a large body of evidence that individuals' preferences and decision making systematically deviate from the canonical model of utility theory (Kagel & Roth, 2016; Plott & Smith, 2008; Marzilli Ericson & Fuster, 2014). Deviations are also found when assessing the utility of a health state. Individuals will often (though by no means always) adjust to new conditions, but systematically fail to forecast that they will do so (Dolan & Kahneman, 2008). Thus the utility assigned to a health state worse than an individual's own may be lower than a person currently in said state, either due to lack of direct experience of the health state, loss aversion, or a failure to anticipate how their lifestyle could adjust to new circumstances, or by some other mechanism.

Such dependence of preferences on the reference point of an individual's own health state is a violation of the assumptions of canonical economic utility on which health state valuation methods are grounded. Jonker, Attema, Donkers, Stolk, and Versteegh (2016) provide the only evidence of which we are aware for reference dependent preferences when valuing EQ-5D-5L. The preference weights associated with levels of the instrument were lower for respondents whose own health was above that level than for those whose health was at or below it. This study performs an exercise with a similar goal using the 3 level version of the instrument.

A difference in the valuation of health states by the general public and people actually experiencing them opens up a debate about which perspective should be used when allocating healthcare resources. Reference dependence and regarding states as worse if they are below one's current health date implies many treatments could be overvalued, as they don't provide as much benefit as the general public presumes. On the other hand, it can be argued that the general public's preferences should be used to allocate public resources In addition, using the preferences of the public introduces a Rawlsian veil of ignorance, allocating resources to conditions prior to discovering what health problems one will experience in later life. Yet these arguments hold less water if the general public's preferences are to a large extent based on inaccurate forecasts and lack of information on how it would be to experience health states worse than one's own. As the amount of reference dependence could vary according to health dimension and condition, this could also lead to distortions in resource allocation.

The issues above are of great practical, theoretical and philosophical importance, however they are beyond the scope of this paper and we do not take a position on them. Yet, regardless of what position is taken, the debate should be informed by empirical evidence, and it is this which this paper contributes.

In the current study, individuals completed a discrete choice experiment (DCE) survey to value EQ-5D-3L. DCEs are an increasingly popular method of creating EQ-5D value sets (Devlin, Shah, Feng, Mulhern, & Hout, 2018; Mulhern, Bansback, Hole, & Tsuchiya, 2017; Robinson, Spencer, Pinto-Prades, & Covey, 2017; Mulhern et al., 2014; Ramos-Goñi et al., 2017; Ramos-Goni et al., 2013) and involve participants making a series of choices as to which of two EQ-5D states they consider to be better. From the trade-offs they make, the preference weight assigned to each level of each dimension may be estimated. The data was collected as part of a project which, amongst other outcomes, compared the valuation of health states by carers and the general population. Participants reported their own EQ-5D-3L state, and carers also completed a proxy EQ-5D-3L instrument giving the health state of the person they care for. The study makes a number of contributions. Firstly, although it is far from being an exact replication of Jonker et al.'s study, it is an attempt to find the same phenomenon. Examining the robustness of previously published results is important in the light of concerns about publication bias and the number of "significant" results that are in fact type 1 errors in academic literature (Open Science Collaboration, 2015; Camerer et al., 2016). It also examines whether the results are extendable to using a different version of the EQ-5D instrument in a different population.

In addition this study has several unique features. The data comes not only from the general population, but also from carers. This is an intriguing population in which to study reference dependence, as most will have a close relationship (for example spouse or parent) with someone who is in a relatively low health state, and are more than likely to be in a higher health state themselves. Due to this situation, it raises the possibility that another person's health is a salient reference point. The proxy EQ-5D-3L responses from carers made it possible to test whether the health state of someone close to an individual can form a reference point.

2 Methods

Ethical approval was granted by the research ethics committee of the Faculty of Medicine and Health, University of Leeds (approval reference MREC16-085).

2.1 Survey design

In the DCE, participants valued the EQ-5D-3L instrument. They were shown two EQ-5D-3L health states and asked to indicate which, in their opinion, was better. An example question is shown in Figure 1a. The choiceDes package for R was created to create a D-efficient survey design with 10 questions per respondent and 4 survey versions. Choices which presented a dominant option were excluded. No restrictions were placed on possible combinations of levels, as there is no widely accepted, rigorous definition as to what constitutes an "unrealistic" EQ-5D-3L state.

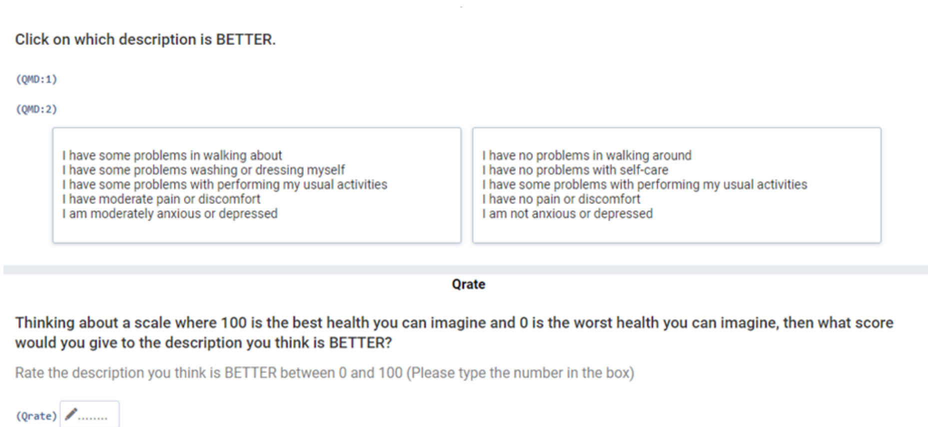
Participants also rated the following health states using an online implementation of the Visual Analogue Scale (VAS): 11111, 33333, and death, all on the same screen. They used a mouse to drag a box containing a description of the health state up and down the scale, with a line from the box to the scale indicating its position, and the numerical value currently selected also indicated in the corner of the box. Participants also rated their own health that day using the VAS on a separate screen. The implementation of VAS is shown in Figure 1b.

All participants completed the EQ-5D-3L instrument for themselves, and those who were carers completed a proxy for the person they are caring for. Finally, all participants answered questions about themselves and their health, and carers completed some additional questions on their caring experiences.

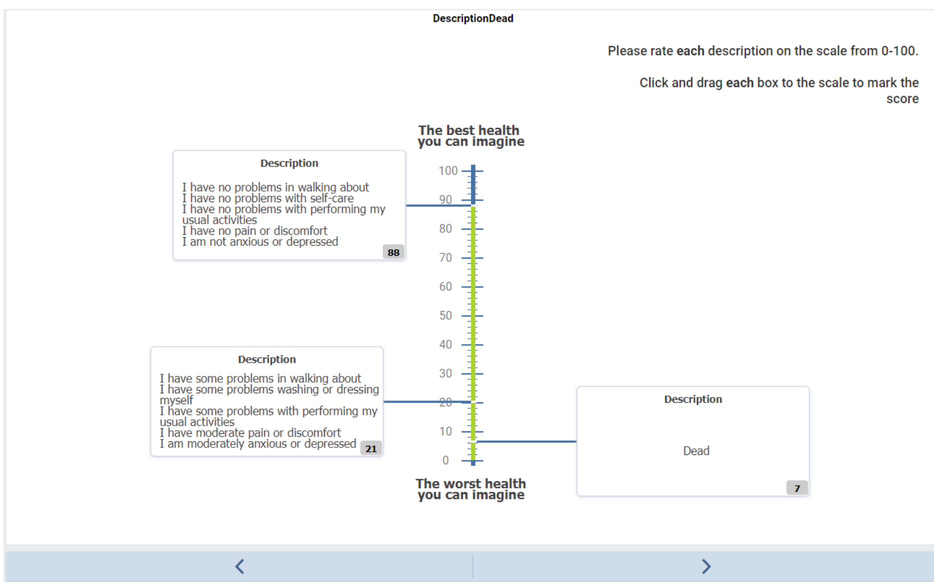
2.2 Recruitment

Respondents were recruited from an online panel managed by a survey and market research company. Respondents "straightlining" DCE responses, i.e. always selecting either the left or right hand side option were excluded. Recruitment was conducted in two waves. In the first, carers for people with dementia and carers for people with other conditions were over sampled to fulfil the requirements of a separate research project. In the second wave, there was no oversampling of these groups.

To construct a general population sample, the proportions of carers for people with dementia and other carers in wave 2 were calculated, and a corresponding proportion of each group in wave 1 sampled at random. These samples, together with non-carers from wave 1, were then combined with the wave 2 sample. A carer sample was obtained by combining all carers from both waves.



(a) Discrete choice experiment task



(b) Visual analogue scale task

Figure 1: Example survey tasks

2.3 Statistical analysis

For the baseline model with no reference dependence, let the utility individual i receives from alternative k in choice task j be

$$u_{ijk} = 1 - \beta_i X_{jk} + \varepsilon_{ijk}. \tag{1}$$

Here X_{jk} is a vector of dummy variables indicating the levels of each health dimension in k and β_i is a vector of 10 parameters, one for levels 2 and 3 of each health dimension indicating the decrement in utility from being in that level compared to level 1. ε_{ijk} is an extreme value distributed error term.

To add reference dependence, following Jonker et al., the utility function is altered to be

$$u_{ijk} = 1 + \beta_i \left(1 + \gamma_i Z_{ijk}\right) X_{jk} + \varepsilon_{ijk}. \tag{2}$$

Z_{ijk} is a vector of dummy variables which for each health dimension take the value 1 if the level in state k is higher or equal to i's own health and take the value 0 if it is lower. γ_i is a vector of reference dependence parameters. Thus the decrement in utility individuals ascribe to levels higher or equal to their own current level is $\beta\left(1+\gamma\right)$, and the decrement ascribed to levels worse than their own is β. The reference dependent parameters are always allowed to vary across dimensions, but may be restricted to be the same for both levels 2 and 3 of each dimension ("one parameter models") or unrestricted ("two parameter models"). For the carer sample, the vector Z_{ijk} of dummy variables referred to the proxy EQ-5D completed on behalf of the person they care for.

Mixed logit (MIXL) models were estimated with coefficients assumed to have a normal distribution, so that individual i's utility decrement for level $\ell \in \{2,3\}$ of dimension d is $\beta_{ild} \sim N\left(\beta_{ld}, \sigma_{ld}^2\right)$. For reference dependence, this is altered to become $\beta_{ild} \sim N\left(\beta_{ld}\left(1+\gamma_{ld} z_{ild}\right), \sigma_{ld}^2\left(1+\sigma_{\gamma_{ld}}\right)^2\right)$. Baseline, one parameter and two parameter models were estimated for the general population sample. For the carer sample, a baseline model was estimated as well as one and two parameter models with dependence on the health of the person they care for. As a robustness check, models were also estimated separately for carers with people for dementia and carers for people with other conditions, and coefficients compared using Welch's t-test. As a further robustness check, models were re-estimated with only participants who reported caring for an individual for at least 10 hours per day, as the large amount of time spent caring for their needs could make their health a more salient reference point. They were also re-estimated with only respondents who cared for a partner, as that group was considered most likely to co-habit and thus could see the cared for person's health as more salient.

Reference dependence parameters from Jonker et al. are on potentially very different latent scales. To enable comparisons, the γ coefficients in table V of Jonker et al. were rescaled by adjusting the value of the EQ-5D "pits" state of 55555 so that it equals the value of the EQ-5D-3L "pits" state of 33333 obtained from the current article's models. For further details of the anchoring method, see Webb, O'Dwyer, Meads, Kind, and Wright (In press).

DCE results are on a latent scale with no units, so VAS responses were used to anchor results to the scale with full health valued at 1 and death valued at 0 used for QALY calculation. Individual i's anchored VAS valuation of 33333 is given by $\widetilde{VAS}_{i33333} = {}^{(VAS_{i33333} - VAS_{idead})}/_{(VAS_{i11111} - VAS_{idead})}$, where VAS_{is} is i's unanchored valuation of state s. Individual i's latent scale valuation of 33333 is

$$u_{i33333} = 1 + \beta_{i33333}\left(1 - \gamma_{i33333} z_{i33333}\right) \tag{3}$$

where Z_{i33333} is a vector of dummies indicating if an individual/person cared for occupies level 3 on each dimension. Equating this to the individual's anchored valuation of 33333 shows that multiplying utility decrements by a factor

$$\alpha_i = \frac{\widetilde{VAS}_{i33333} - 1}{\beta_{i33333}\left(1 - \gamma_{i33333} z_{i33333}\right)} \tag{4}$$

anchors them to the full health = 1, dead = 0 scale.

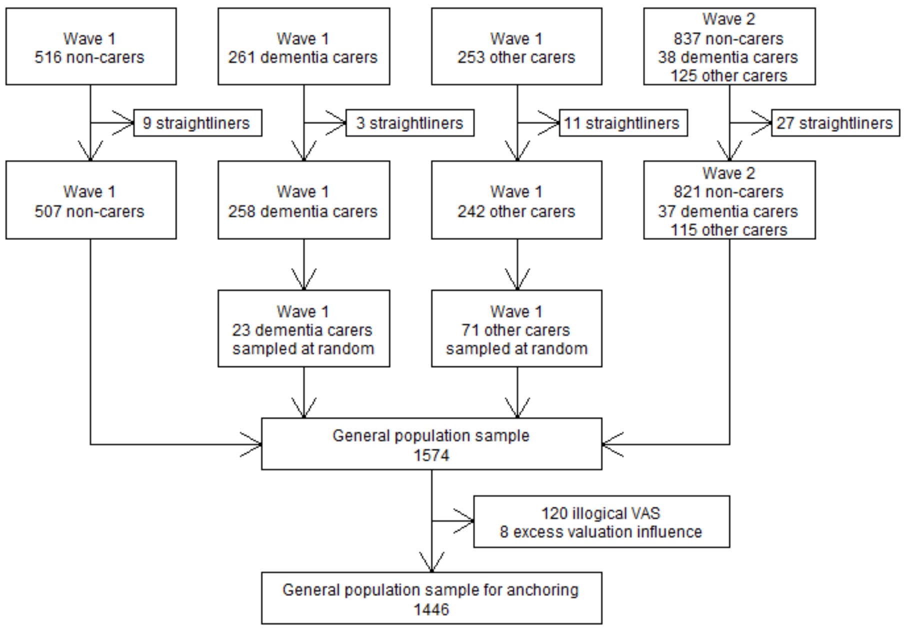
As utility decrements are estimated as population means, there are $2^5 = 32$ possible different latent scale valuations of 33333, depending on whether individuals occupy level 3 of each dimension or not. Thus (up to) 32 mean values of α were calculated for each of these groups, and an overall population mean calculated by taking a weighted average according to group size.

To be able to rescale valuations, respondents must give logical VAS responses, i.e. $VAS_{11111} > VAS_{dead}$. Thus participants giving illogical responses were excluded from the samples for anchored analysis. In addition, exploratory analysis of the data revealed that a small number of respondents with extremely low anchored VAS valuations of 33333 had a large impact on anchored utility decrements. Thus both general population and carer samples were filtered in the following way: A multinomial logit (MNL) baseline model was run, and anchored coefficients calculated. A series of MNL models were then run excluding each participant in turn and anchored coefficients were calculated as a proportion of coefficients with the full sample, allowing the mean influence a single individual has on results to be found. The overall mean influence of all respondents was calculated, and individuals who were more than two standard deviations away from this were designated as having an excess influence on valuation and excluded from analysis. After these exclusions, baseline and two parameter models were run for both the general population and carer samples. Results were then anchored to the full health = 1, dead = 0 scale.

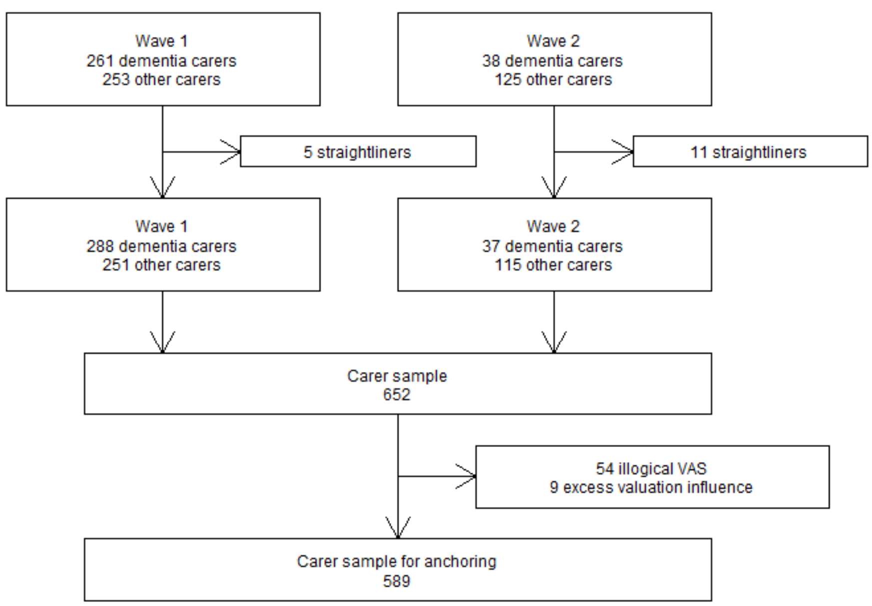
Models were estimated using simulated maximum likelihood implemented in the CMC Choice Modelling Centre Code for R version 1.1 (CMC, 2017). There were t-tests used to judge coefficient statistical significance, with significance judged at the 5% level after adjusting for multiple testing using Holm's sequential Bonferroni correction (Holm, 1979). All analysis was carried out in R version 3.3.1.

3 Results

In total 1030 respondents were recruited in the first wave and 1000 were recruited in the second. Figure 2 demonstrates how the final samples were arrived at of 1574 for the general population (1446 for anchored results) and 652 for carers (589 for anchored results).



(a) General population sample



(b) Carer sample

Figure 2: Sample selection process

Table 1 summarises respondents' demographics. The general population and carer samples are similar in terms of mean age (45.4 vs. 46.7) and number of female respondents (52.7% vs. 55.7%). General population respondents were slightly less likely to be employed (54.4% vs. 59.5%), but slightly more likely to be retired (20.5% vs. 16.7%) or a student (6.8% vs. 2.91%). The table also shows that samples for latent scale and anchored analysis are very similar.

Table 1: Demographic information and visual analogue scale responses

		General population (for anchoring)	General population	Carers (for anchoring)	Carers
Age (variance)		45.4 (290)	45.8 (292)	46. (226)	47.6 (225)
Female (%)		829 (52.7)	765 (52.9)	363 (55.7)	329 (55.9)
Occupation (%)	Employed/Self-employed	857 (54.4)	773 (53.5)	388 (59.5)	347 (58.9)
	Retired	322 (20.5)	310 (21.4)	109 (16.7)	104 (17.7)
	Housework	104 (6.61)	93 (6.43)	60 (9.20)	52 (8.83)
	Student	107 (6.80)	102 (7.05)	19 (2.91)	18 (3.06)
	Seeking work	89 (5.65)	82 (5.67)	23 (3.53)	21 (3.57)
	Other/prefer not to say	95 (6.04)	86 (5.95)	53 (8.13)	47 (7.98)
Visual analogue scale responses (variance)	Own health	73.5 (372)	74 (359)	70.8 (388)	71 (390)
	Person cared for's health			47 (516)	45.8 (504)
	11111	89.5 (315)	92.4 (170)	86.5 (331)	89.1 (231)
	33333	21.6 (408)	19.3 (289)	23.4 (433)	20.3 (284)
	Dead	10.6 (464)	6.72 (186)	12.6 (517)	8.14 (230)
N		1574	1446	652	589

Table 2 summarises EQ-5D responses. Both general population and carer samples generally reported being in good health. The percentage of general population respondents in level 1 ranged from 91.2% for self-care to 59% for pain/discomfort, whereas with carers it ranged from 86.5% for self-care to 48.9% for anxiety/depression. The carer sample generally reported slightly worse health. Few respondents reported being in level 3 on any dimension, with the greatest proportion observed for anxiety/depression (6.99% general population, 7.82% carers). The health of people being cared for is considerably worse, as for each dimension fewer than a third were reported to be in level 1. Many more level 3 responses were observed, with the greatest proportion (26.7%) seen for usual activities. Responses are very similar between samples for latent scale and anchored analysis.

Table 2: EQ-5D responses

		Gen. pop.	Gen. pop. (anchor)	Carer	Carer (anchor)	Person cared for	Person cared for (anchor)
Mobility (%)	Level 1	1264 (80.3)	1165 (80.6)	480 (73.6)	434 (73.7)	143 (21.9)	124 (21.1)
	Level 2	303 (19.3)	276 (19.1)	169 (25.9)	152 (25.8)	467 (71.6)	428 (72.7)
	Level 3	7 (0.445)	5 (0.346)	3 (0.460)	3 (0.509)	42 (6.44)	37 6.28)
Self-care (%)	Level 1	1435 (91.2)	1330 (92.0)	564 (86.5)	517 (87.8)	187 (28.7)	169 (28.7)
	Level 2	135 (8.58)	113 (7.81)	84 (12.9)	69 (11.7)	359 (55.1)	320 (54.3)
	Level 3	4 (0.254)	3 (0.207)	4 (0.613)	3 (0.509)	106 (16.3)	100 (17.0)
Usual activities (%)	Level 1	1225 (77.8)	1134 (78.4)	463 (71.0)	427 (72.5)	82 (12.6)	65 (11.0)
	Level 2	329 (20.9)	293 (20.3)	180 (27.6)	153 (26.0)	406 (62.3)	367 62.3)
	Level 3	20 (1.27)	19 (1.31)	9 (1.38)	9 (1.53)	164 (25.2)	157 (26.7)
Pain/ discomfort (%)	Level 1	928 (59.0)	856 (59.2)	331 (50.8)	305 (51.8)	155 (23.8)	139 (23.6)
	Level 2	578 (36.7)	530 (36.7)	273 (41.9)	244 (41.4)	389 (59.7)	352 (59.8)
	Level 3	68 (4.32)	60 (4.15)	48 (7.36)	40 (6.79)	108 (16.6)	98 (16.6)
Anxiety/ depression (%)	Level 1	951 (60.4)	888 (61.4)	319 (48.9)	301 (51.1)	212 (32.5)	194 (32.9)
	Level 2	513 (32.6)	463 (32.0)	282 (43.3)	243 (41.3)	347 (53.2)	310 (52.6)
	Level 3	110 (6.99)	95 (6.57)	51 (7.82)	45 (7.64)	93 (14.3)	85 (14.4)
N		1574	1446	652	589	652	589

Table 3 gives the results of latent scale general population models. For the baseline model, coefficients are statistically significant, and level 3 coefficients are lower than level 2 ones. For the one parameter model, the point estimates of all reference dependence coefficients are negative, as expected (i.e. utility decrements associated with a given level are greater for individuals whose current health is above that level). However, only the coefficients for mobility and pain/discomfort are statistically significant. With the two parameter model, all reference dependence coefficients are negative, with the exception of usual activities level 2, however only mobility level 2 achieves statistical significance.

Table 3: Latent scale results for general population sample

	Baseline	CI	One param.	CI	Two param.	CI
MO2	-0.561*	[-0.635, -0.487]	-0.670*	[-0.746, -0.593]	-0.674*	[-0.751, -0.596]
MO3	-2.53*	[-2.71, -2.36]	-2.54*	[-2.71, -2.38]	-2.56*	[-2.74, -2.38]
SC2	-0.427*	[-0.501, -0.353]	-0.436*	[-0.512, -0.360]	-0.435*	[-0.511, -0.358]
SC3	-1.01*	[-1.13, -0.881]	-1.01*	[-1.13, -0.885]	-1.01*	[-1.14, -0.886]
UA2	-0.308*	[-0.394, -0.222]	-0.312*	[-0.399, -0.225]	-0.301*	[-0.392, -0.211]
UA3	-0.732*	[-0.831, -0.632]	-0.743*	[-0.842, -0.645]	-0.754*	[-0.854, -0.654]
PD2	-0.507*	[-0.577, -0.438]	-0.566*	[-0.643, -0.488]	-0.569*	[-0.651, -0.488]
PD3	-1.49*	[-1.61, -1.37]	-1.51*	[-1.62, -1.39]	-1.51*	[-1.63, -1.39]
AD2	-0.451*	[-0.519, -0.384]	-0.473*	[-0.542, -0.403]	-0.460*	[-0.536, -0.385]
AD3	-1.20*	[-1.30, -1.09]	-1.22*	[-1.32, -1.12]	-1.23*	[-1.33, -1.12]
γ MO/MO2			-0.888*	[-1.15, -0.620]	-0.868*	[-1.14, -0.593]
γ SC/SC2			-0.236	[-0.675, 0.204]	-0.182	[-0.656, 0.292]
γ UA/UA2			-0.0824	[-0.512, 0.347]	0.215	[-0.354, 0.785]
γ PD/PD2			-0.225*	[-0.395, -0.0544]	-0.231	[-0.434, -0.0284]
γ AD/AD2			-0.147	[-0.333, 0.0394]	-0.0415	[-0.302, 0.219]
γ MO3					-0.861	[-1.67, -0.0496]
γ SC3					-0.652	[-1.40, 0.0970]
γ UA3					-0.685	[-1.45, 0.0778]
γ PD3					-0.206	[-0.554, 0.141]
γ AD3					-0.313	[-0.656, 0.0307]

Note. CI=95% confidence interval; N=1574; *=statistical significance at 5% level after Holm's sequential Bonferroni correction (Holm, 1979)

Table 4 lists the latent scale results for carers. In the baseline model, all coefficients are negative with level 3 coefficients lower than for level 2, and all are statistically significant with the exception of usual activities level 2. No reference dependence parameters in either the one or two parameter models were significant. In the one parameter model, the pain/discomfort and anxiety/depression coefficients had an unexpected sign, whereas for the two parameter model, the usual activities level 2, pain/discomfort level 3 and anxiety/depression level 3 parameters were positive.

Table 5 shows results from the general population sample on the full health=1, dead=0 scale. In both the one and two parameter models, all main (i.e. non-reference dependence) parameters are significant, with level 3 utility decrements of greater magnitude than those for level 2. Reference dependence parameters are all negative for both models, with the exception of usual activities level 2 in the two parameter model, however only mobility in the one parameter model is statistically significant.

Anchored results for the carer sample are shown in Table 6. All main coefficients are of the expected sign and magnitude and are significant, apart from usual activities level 2, which is positive but insignificant. No reference dependence parameters are significant. In the one parameter model, four out of five are negative, the exception being pain/discomfort, and with the two parameter model, half are positive and half negative.

Table 4: Latent scale results for carer sample (reference dependence on health of person cared for)

	Baseline	CI	One param.	CI	Two param.	CI
MO2	-0.413*	[-0.635, -0.487]	-0.440*	[-0.554, -0.325]	-0.431*	[-0.568, -0.294]
MO3	-1.94*	[-2.71, -2.36]	-2.00*	[-2.24, -1.76]	-2.01*	[-2.25, -1.76]
SC2	-0.276*	[-0.501, -0.353]	-0.311*	[-0.432, -0.190]	-0.329*	[-0.455, -0.203]
SC3	-0.688*	[-1.13, -0.881]	-0.719*	[-0.882, -0.556]	-0.716*	[-0.887, -0.546]
UA2	-0.0745	[-0.394, -0.222]	-0.0771	[-0.247, 0.0930]	-0.0125	[-0.0533, 0.0283]
UA3	-0.406*	[-0.831, -0.632]	-0.422*	[-0.575, -0.270]	-0.417*	[-0.536, -0.299]
PD2	-0.419*	[-0.577, -0.438]	-0.417*	[-0.548, -0.286]	-0.481*	[-0.613, -0.348]
PD3	-1.28*	[-1.61, -1.37]	-1.29*	[-1.47, -1.11]	-1.26*	[-1.44, -1.08]
AD2	-0.380*	[-0.519, -0.384]	-0.391*	[-0.502, -0.280]	-0.375*	[-0.500, -0.251]
AD3	-1.02*	[-1.30, -1.09]	-1.05*	[-1.19, -0.904]	-1.05*	[-1.19, -0.903]
γ MO/MO2			-0.109	[-0.336, 0.118]	-0.0472	[-0.492, 0.397]
γ SC/SC2			-0.218	[-0.586, 0.149]	-0.338	[-0.804, 0.129]
γ UA/UA2			-0.282	[-1.22, 0.658]	9.25	[-24.5, 43.0]
γ PD/PD2			0.0599	[-0.297, 0.417]	-0.253	[-0.608, 0.103]
γ AD/AD2			0.0016	[-0.345, 0.348]	0.0837	[-0.430, 0.597]
γ MO3					-0.0971	[-0.441, 0.247]
γ SC3					-0.0725	[-0.730, 0.585]
γ UA3					-0.59	[-1.43, 0.249]
γ PD3					0.437	[-0.187, 1.06]
γ AD3					-0.0943	[-0.760, 0.572]

Note. CI=95% confidence interval; N=652; *=statistical significance at 5% level after Holm's sequential Bonferroni correction (Holm, 1979)

Figure 3 illustrates the comparison between the current article's results and those found by Jonker et al. For the general population sample, Jonker et al.'s point estimates of reference dependence parameters lie outside the 95% confidence intervals (CIs) of the current study's estimated for mobility, pain/discomfort and anxiety/depression, but not for self-care or usual activities. With the carer sample, Jonker et al.'s point estimates lie within the current study's CIs for all dimensions apart from pain/discomfort.

4 Discussion

In contrast to previous research, limited reference dependence in individuals' preferences for EQ-5D health states is found here. Reference dependence was only reliably found for mobility in the general population, with the parameter for pain/discomfort only significant in the one parameter model. Otherwise, point estimates of the relevant parameters were not significant, although with the general population models they nearly always had the expected sign.

There are various possible causes for the difference in results. The first is that when valuing the 3L version of EQ-5D, in contrast to the 5L version, individuals exhibit little or no reference dependence except with respect to mobility. With fewer levels, the 3L instrument is less finely grained, with fewer people in less than

Table 5: Results on full health=1, dead=0 scale for general population sample

	One param.	CI	Two param.	CI
MO2	-0.0838*	[-0.0938, -0.0738]	-0.0840*	[-0.0942, -0.0739]
MO3	-0.325*	[-0.348, -0.301]	-0.326*	[-0.350, -0.302]
SC2	-0.0571*	[-0.0669, -0.0473]	-0.0570*	[-0.0668, -0.0471]
SC3	-0.130*	[-0.146, -0.113]	-0.130*	[-0.147, -0.113]
UA2	-0.0407*	[-0.0515, -0.0298]	-0.0389*	[-0.0502, -0.0275]
UA3	-0.0986*	[-0.112, -0.0856]	-0.0997*	[-0.113, -0.0865]
PD2	-0.0740*	[-0.0840, -0.0641]	-0.0741*	[-0.0846, -0.0637]
PD3	-0.197*	[-0.212, -0.182]	-0.198*	[-0.214, -0.182]
AD2	-0.0604*	[-0.0696, -0.0512]	-0.0584*	[-0.0682, -0.0487]
AD3	-0.155*	[-0.169, -0.142]	-0.157*	[-0.171, -0.143]
γ MO/MO2	-0.101*	[-0.136, -0.0670]	-0.0992*	[-0.133, -0.0657]
γ SC/SC2	-0.0251	[-0.0839, 0.0337]	-0.0221	[-0.0836, 0.0394]
γ UA/UA2	-0.000621	[-0.0534, 0.0522]	0.0294	[-0.0382, 0.0971]
γ PD/PD2	-0.0223	[-0.0427, -0.00185]	-0.0213	[-0.0455, 0.00287]
γ AD/AD2	-0.0243	[-0.0476, -0.00110]	-0.0101	[-0.0417, 0.0215]
γ MO3			-0.125	[-0.214, -0.0354]
γ SC3			-0.102	[-0.208, 0.00449]
γ UA3			-0.0907	[-0.186, 0.00456]
γ PD3			-0.0266	[-0.0700, 0.0168]
γ AD3			-0.0463	[-0.0921, -0.000468]

Note. CI=95% confidence interval; N=1446; *=statistical significance at 5% level after Holm's sequential Bonferroni correction (Holm, 1979)

full health. The lowest levels are also considered worse than in the 5L version (Devlin et al., 2018; Dolan, 1997). Given these differences, it should not be assumed that individuals value both instruments in the same way. Reference dependence is a common but by no means universal finding when measuring preferences, and depends on the circumstances under which they are measured (see e.g. List, 2003; Plott & Zeiler, 2005; Isoni, Loomes, & Sugden, 2011).

Assuming individuals' preferences for EQ-5D-3L health states exhibit little or no reference dependence, whether or not this is a beneficial feature or not depends on one's point of view. It is a positive finding if reference dependence is regarded as a bias. If individuals' true preference relations follow the axioms of canonical utility theory, then deviations from them are mistakes. On the other hand, if reference dependence is in fact a true aspect of individuals' preferences, rather than an artefact of imperfect responses to surveys, it could be considered a drawback of EQ-5D-3L that it fails to capture it. This is a matter of active debate (see e.g. Sugden, 2015) and is beyond the scope of the current article. Another explanation for the differences in results is that the two samples are not directly comparable. Jonker et al. used Dutch respondents, whereas this study used British ones. It is possible that some populations display reference dependence and some do not, and a recommendation for national valuation survey is that they investigate the presence of reference dependence in their population.

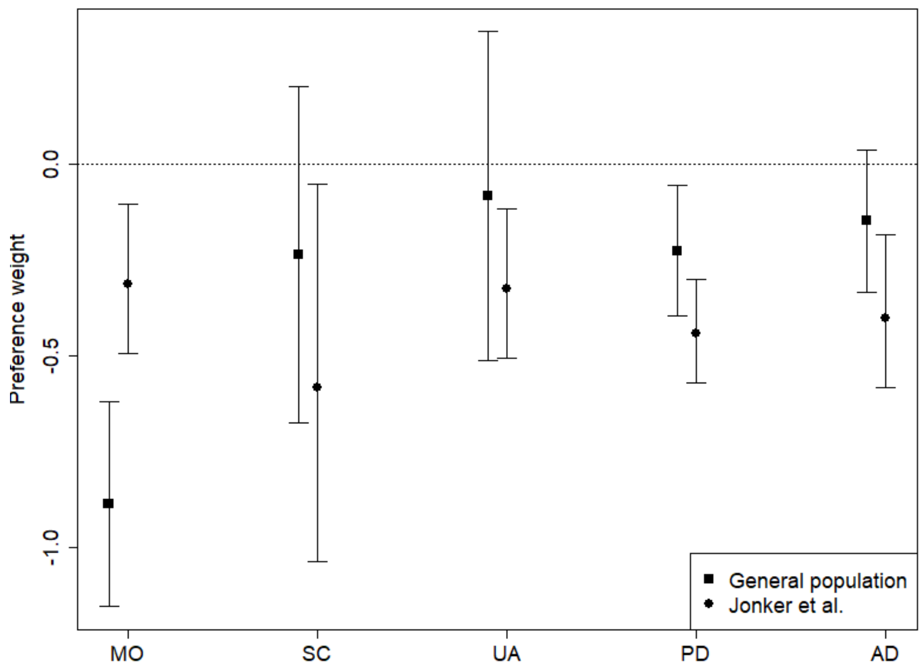
Table 6: Results on full health=1, dead=0 scale for carer sample (reference dependence on health of person cared for)

	One param.	CI	Two param.	CI
MO2	-0.0693*	[-0.0876, -0.0510]	-0.0678*	[-0.0894, -0.0462]
MO3	-0.320*	[-0.358, -0.283]	-0.315*	[-0.353, -0.276]
SC2	-0.0528*	[-0.0726, -0.0330]	-0.0565*	[-0.0766, -0.0364]
SC3	-0.116*	[-0.142, -0.0892]	-0.111*	[-0.138, -0.0845]
UA2	-0.023	[-0.0461, 0.000193]	-0.0194	[-0.0441, 0.00527]
UA3	-0.0755*	[-0.0995, -0.0515]	-0.0755*	[-0.0996, -0.0513]
PD2	-0.0754*	[-0.0987, -0.0522]	-0.0851*	[-0.106, -0.0640]
PD3	-0.216*	[-0.248, -0.184]	-0.207*	[-0.236, -0.177]
AD2	-0.0644*	[-0.0829, -0.0458]	-0.0629*	[-0.0828, -0.0431]
AD3	-0.172*	[-0.199, -0.145]	-0.168*	[-0.193, -0.143]
γ MO/MO2	-0.00787	[-0.0465, 0.0308]	-0.00389	[-0.0729, 0.0651]
γ SC/SC2	-0.0433	[-0.101, 0.0139]	-0.0753	[-0.141, -0.00963]
γ UA/UA2	-0.0305	[-0.127, 0.0661]	0.0513	[-0.232, 0.335]
γ PD/PD2	0.0139	[-0.0446, 0.0725]	-0.0342	[-0.0830, 0.0146]
γ AD/AD2	-0.000619	[-0.0517, 0.0505]	-0.000121	[-0.0723, 0.0721]
γ MO3			0.00604	[-0.0514, 0.0635]
γ SC3			0.02	[-0.0823, 0.122]
γ UA3			-0.0653	[-0.187, 0.0566]
γ PD3			0.0934	[-0.0143, 0.201]
γ AD3			0.016	[-0.0766, 0.109]

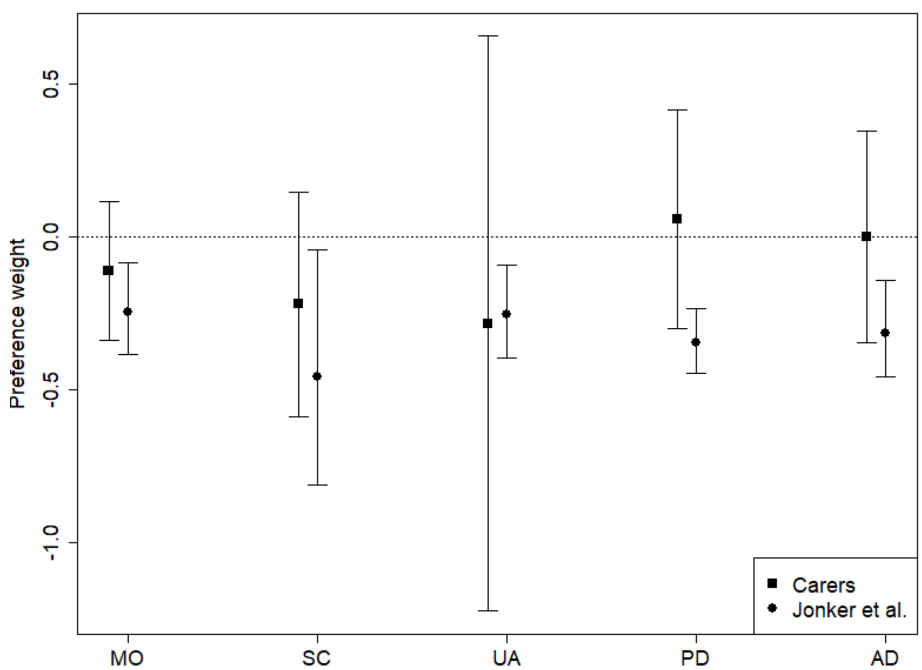
Note. CI=95% confidence interval; N=589; *=statistical significance at 5% level after Holm's sequential Bonferroni correction (Holm, 1979)

An alternative explanation for the lack of robust evidence for reference dependence is that the DCE did not have sufficient statistical power to detect it. This is especially so given that the survey was not specifically designed to examine reference dependence. Figure 3 allows a comparison of the effect sizes seen in the two studies. It may be seen that after calibration of the two sets of results, the current paper has a greater effect size (i.e. more negative parameters) for mobility, and a lower effect size for the other four dimensions. For self-care and usual activities, effects as large as Jonker et al.'s cannot be ruled out, however for pain/discomfort and anxiety/depression, Jonker et al.'s point estimates lie outside the 95% CIs of the current study.

Comparing the two sets of results is inevitably imperfect, especially given they are for different versions of the EQ-5D instrument. Another way of examining effect size is to exploit the results anchored to the full health=1, dead=0 scale. Using the results of the one parameter model in Table 5, it is possible to calculate what the maximum differences in QALY valuation due to reference dependence that still lie within the 95% CI are. These are generally quite small for the four statistically insignificant dimensions, with the largest difference within a 95% CI being only 0.011 for self-care level 3. No maximum difference for level 2 of any of the four dimensions exceeds 0.005. Thus even were the survey to have had greater power to detect statistically



(a) General population sample



(b) Carer sample

Figure 3: Comparison of reference dependence parameter results with Jonker et al.
Note. Error bars show 95% confidence intervals for the current study and 95% credible intervals for Jonker et al.

significant reference dependence effects, they would probably be of limited meaningful size.

There is another advantage of using anchored results. Reference dependence is essentially comparing the preferences of two groups, those whose health is better than a given state, and those whose health is as good or worse. Comparisons between different groups on a latent scale can be problematic, as preference and scale are confounded (Vass, Wright, Burton, & Payne, 2017; Hess & Rose, 2012; Hess & Train, 2017). It may be that reference dependence parameters in a model are in fact picking up differences in scale, rather than differences in preference weights. Being in a lower health state is associated with being older, and thus potentially making more errors in responding to an online survey. Thus it is a strength of the study that its results are verified using valuations anchored to a 0-1 scale.

So far, the discussion has focused on reference dependence on one's own health. However, it is also possible that valuation of health states also depends on the health of individuals who are close to them. With the carer sample, proxy EQ-5D data is available for the health of an individual that respondents have a close relationship with. This makes it possible to test for the possibility that socially close individuals can also form a salient reference point for preferences over health states. However, it was found that individuals' decisions had no significant dependence on the health state of the person they cared for. Nevertheless, it would be interesting to pursue this avenue of research further. Figure 3 shows that when comparing this study's results to those of Jonker et al., the point estimates of the latter are contained within the 95% CIs of the former for every dimension but pain/discomfort, implying that reasonably large effects may not have been detected by the current study.

Given that Jonker et al. found significant reference dependence on one's own health with EQ-5D-5L, it would be useful to perform a similar exercise with carers using EQ-5D-5L. Another possibility is examining the interdependence of health-state preference of other people in close relationships such as couples or parents and children.

This study has several strengths. It has a large sample size, both for the general public and carers. However, it should be noted that Johnson et al. (2013) found the benefit in terms of statistical power of adding more respondents to a DCE is relatively small for sample sizes above 300. It is a strength that the study presents results not only on a latent scale, but on a full health=1, dead=0 scale. This both avoids problems with scale heterogeneity and allows for an estimation of how large an effect reference dependence can have on QALY calculation. Finally, it is a strength that it extends the concept of reference dependence from dependence on one's own health to the health of an individual with whom one has a close relationship.

This study also has several limitations. The data was not collected with the primary aim of examining reference dependence, and thus the survey was not designed to optimise the probability of observing it. It is difficult to compare the previous findings of Jonker et al. as the re-calibration performed here has several issues, such as the use of different survey instruments. Very few respondents reported being in level 3 on any dimension, making robust estimation of two parameter models difficult. There were also relatively small proportions of respondents in level 2 for each dimension. Finally, there is no direct measure of how close a relationship carers have with the person they care for, making it difficult to assess the plausibility of the person they care for's health being a salient reference point.

5 Conclusion

This study ran a DCE in which participants' preferences for EQ-5D-3L health states were measured. It found limited evidence that those preferences were dependent on the reference point of the participants' own health states, apart from on the mobility dimension for the general population. However, reference dependence remains an important topic, and one on which further research is needed to determine the robustness of the phenomenon. EQ-5D is a commonly used instrument and valuations of its health states are used to make policy decisions with wide-ranging financial and health impacts. Whether reference dependence represents a bias or not is a debate we do not wish to enter in to. However, the debate needs to be informed by greater evidence.

List of abbreviations

CI	Confidence interval
DCE	Discrete choice experiment
HRQoL	Health-related quality of life
MIXL	Mixed logit
MNL	Multinomial logit
QALY	Quality adjusted life year
VAS	Visual analogue scale
MO	Mobility
SC	Self-care
UA	Usual activities
PD	Pain or discomfort
AD	Anxiety or depression

References

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., ... Chan, T. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433-1436.

CMC. (2017). *CMC choice modelling code for R.* Retrieved from `www.cmc.leeds.ac.uk`

Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B., & Hout, B. (2018). Valuing health-related quality of life: An EQ-5D-5L value set for england. *Health economics*.

Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical care*, *35*(11), 1095-1108.

Dolan, P., & Kahneman, D. (2008). Interpretations of utility and their implications for the valuation of health. *The Economic Journal*, *118*(525), 215-234.

Hess, S., & Rose, J. M. (2012). Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation*, *39*(6), 1225-1239.

Hess, S., & Train, K. (2017). Correlation and scale in mixed logit models. *Journal of Choice Modelling*, *23*, 1-8.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.

Isoni, A., Loomes, G., & Sugden, R. (2011). The willingness to pay—willingness to accept gap, the endowment effect, subject misconceptions, and experimental procedures for eliciting valuations: Comment. *American Economic Review*, *101*(2), 991-1011.

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., ... Bridges, J. F. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health*, *16*(1), 3-13.

Jonker, M. F., Attema, A. E., Donkers, B., Stolk, E. A., & Versteegh, M. M. (2016). Are health state valuations from the general public biased? A test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. *Health economics*.

Kagel, J. H., & Roth, A. E. (2016). *The handbook of experimental economics, volume 2: The handbook of experimental economics.* Princeton University Press.

List, J. A. (2003). Does market experience eliminate market anomalies? *The Quarterly Journal of Economics*, *118*(1), 41-71.

Marzilli Ericson, K. M., & Fuster, A. (2014). The endowment effect. *Annu. Rev. Econ.*, *6*(1), 555-579.

Mulhern, B., Bansback, N., Brazier, J., Buckingham, K., Cairns, J., Devlin, N., ... Longworth, L. (2014). The feasibility of the DCE-TTO for deriving health-state values for EQ-5D-5L. In *Preparatory study for the revaluation of the EQ-5D tariff: methodology report.*

Mulhern, B., Bansback, N., Hole, A. R., & Tsuchiya, A. (2017). Using discrete choice experiments with duration to model EQ-5D-5L health state preferences: testing experimental design strategies. *Medical Decision Making*, *37*(3), 285-297.

National Institute for Health and Care Excellence (NICE). (2013). *Guide to the methods of technology appraisal 2013 process and methods* (Tech. Rep. No. PMG9). Retrieved from https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pd

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.

Plott, C. R., & Smith, V. L. (2008). *Handbook of experimental economics results* (Vol. 1). Elsevier.

Plott, C. R., & Zeiler, K. (2005). The willingness to pay–willingness to accept gap. *The American Economic Review*, *95*(3), 530-545.

Ramos-Goni, J. M., Rivero-Arias, O., Errea, M., Stolk, E. A., Herdman, M., & Cabasés, J. M. (2013). Dealing with the health state 'dead'when using discrete choice experiments to obtain values for EQ-5D-5L heath states. *The European Journal of Health Economics*, *14*(1), 33-42.

Ramos-Goñi, J. M., Pinto-Prades, J. L., Oppe, M., Cabasés, J. M., Serrano-Aguilar, P., & Rivero-Arias, O. (2017). Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Medical care*, *55*(7), e51-e58.

Robinson, A., Spencer, A. E., Pinto-Prades, J. L., & Covey, J. A. (2017). Exploring differences between TTO and DCE in the valuation of health states. *Medical Decision Making*, *37*(3), 273-284.

Sugden, R. (2015). Looking for a psychology for the inner rational agent. *Social Theory and Practice*, *41*(4), 579-598.

Vass, C. M., Wright, S., Burton, M., & Payne, K. (2017). Scale heterogeneity in healthcare discrete choice experiments: A primer. *The Patient-Patient-Centered Outcomes Research*, 1-7.

Webb, E. J., O'Dwyer, J. L., Meads, D., Kind, P., & Wright, P. (In press). Transforming discrete choice experiment latent scale values for EQ-5D-3L using the visual analogue scale. *Economic Journal of Health Economics*.

Appendix

Table A.1: Latent scale results for separate baseline models for dementia and other carers

	Dementia carers	CI	Other carers	CI	*p* (Welch's *t*-test)
MO2	-0.545*	[-0.726, -0.363]	-0.348*	[-0.726, -0.363]	0.088
MO3	-2.39*	[-2.84, -1.93]	-1.71*	[-2.84, -1.93]	0.013
SC2	-0.333*	[-0.504, -0.162]	-0.243*	[-0.504, -0.162]	0.417
SC3	-0.940*	[-1.23, -0.654]	-0.552*	[-1.23, -0.654]	0.03
UA2	0.0916	[-0.115, 0.298]	-0.208*	[-0.115, 0.298]	0.03
UA3	-0.385*	[-0.615, -0.156]	-0.45*	[-0.615, -0.156]	0.67
PD2	-0.452*	[-0.617, -0.287]	-0.413*	[-0.617, -0.287]	0.717
PD3	-1.53*	[-1.84, -1.22]	-1.13*	[-1.84, -1.22]	0.039
AD2	-0.410*	[-0.581, -0.238]	-0.373*	[-0.581, -0.238]	0.737
AD3	-1.11*	[-1.35, -0.863]	-1.02*	[-1.35, -0.863]	0.548
N	295		357		

Note. CI=95% confidence interval; *=statistical significance at 5% level after Holm's sequential Bonferroni correction (Holm, 1979)

Table A.2: Latent scale results for separate one parameter models for dementia and other carers

	Dementia carers	CI	Other carers	CI	p (Welch's t-test)
MO2	-0.565*	[-0.768, -0.362]	-0.363*	[-0.768, -0.362]	0.118
MO3	-2.37*	[-2.80, -1.94]	-1.75*	[-2.80, -1.94]	0.019
SC2	-0.374*	[-0.568, -0.180]	-0.27*	[-0.568, -0.180]	0.415
SC3	-0.962*	[-1.24, -0.679]	-0.573*	[-1.24, -0.679]	0.03
UA2	0.0979	[-0.138, 0.334]	-0.219	[-0.138, 0.334]	0.036
UA3	-0.413*	[-0.644, -0.182]	-0.466*	[-0.644, -0.182]	0.731
PD2	-0.455*	[-0.655, -0.254]	-0.391*	[-0.655, -0.254]	0.638
PD3	-1.48*	[-1.78, -1.17]	-1.13*	[-1.78, -1.17]	0.073
AD2	-0.422*	[-0.603, -0.242]	-0.364*	[-0.603, -0.242]	0.617
AD3	-1.11*	[-1.36, -0.868]	-1.02*	[-1.36, -0.868]	0.563
γ MO	-0.135	[-0.459, 0.188]	-0.0451	[-0.459, 0.188]	0.725
γ SC	-0.262	[-0.740, 0.215]	-0.17	[-0.740, 0.215]	0.809
γ UA	-0.658	[-1.78, 0.462]	-0.0904	[-1.78, 0.462]	0.407
γ PD	0.0255	[-0.501, 0.552]	0.126	[-0.501, 0.552]	0.788
γ AD	-0.107	[-0.656, 0.442]	0.0982	[-0.656, 0.442]	0.587
N	295		357		

Note. CI=95% confidence interval; *=statistical significance at 5% level after Holm's sequential Bonferroni correction (Holm, 1979)

Table A.3: Latent scale results for separate two parameter models for dementia and other carers

	Dementia carers	CI	Other carers	CI	p (Welch's t-test)
MO2	-0.542*	[-0.775, -0.310]	-0.356*	[-0.775, -0.310]	0.211
MO3	-2.40*	[-2.84, -1.97]	-1.76*	[-2.84, -1.97]	0.016
SC2	-0.388*	[-0.592, -0.183]	-0.287*	[-0.592, -0.183]	0.455
SC3	-0.967*	[-1.26, -0.676]	-0.565*	[-1.26, -0.676]	0.028
UA2	0.0911	[-0.149, 0.332]	-0.183	[-0.149, 0.332]	0.089
UA3	-0.416*	[-0.660, -0.172]	-0.474*	[-0.660, -0.172]	0.714
PD2	-0.513*	[-0.720, -0.306]	-0.458*	[-0.720, -0.306]	0.69
PD3	-1.46*	[-1.76, -1.15]	-1.11*	[-1.76, -1.15]	0.072
AD2	-0.414*	[-0.611, -0.216]	-0.341*	[-0.611, -0.216]	0.58
AD3	-1.13*	[-1.39, -0.870]	-1.03*	[-1.39, -0.870]	0.527
γ MO2	-0.0053	[-0.590, 0.579]	0.0038	[-0.590, 0.579]	0.985
γ SC2	-0.322	[-0.977, 0.334]	-0.321	[-0.977, 0.334]	0.999
γ UA2	-0.436	[-3.55, 2.68]	0.355	[-3.55, 2.68]	0.658
γ PD2	-0.28	[-0.812, 0.251]	-0.186	[-0.812, 0.251]	0.797
γ AD2	-0.0473	[-0.801, 0.706]	0.252	[-0.801, 0.706]	0.583
γ MO3	-0.138	[-0.640, 0.365]	-0.0721	[-0.640, 0.365]	0.852
γ SC3	-0.242	[-1.23, 0.745]	0.0126	[-1.23, 0.745]	0.718
γ UA3	-0.704	[-1.99, 0.580]	-0.586	[-1.99, 0.580]	0.892
γ PD3	0.389	[-0.381, 1.16]	0.385	[-0.381, 1.16]	0.996
γ AD3	-0.0316	[-1.13, 1.07]	0.005	[-1.13, 1.07]	0.957
N	295		357		

Note. CI=95% confidence interval; *=statistical significance at 5% level after Holm's sequential Bonferroni correction (Holm, 1979)

Table A.4: Latent scale results for carers who care for at least 10 hours per day

	One param.	CI	Two param.	CI
MO2	-0.490*	[-0.760, -0.220]	-0.584*	[-0.858, -0.310]
MO3	-2.01*	[-2.44, -1.58]	-2.18*	[-2.67, -1.68]
SC2	-0.198	[-0.489, 0.0935]	-0.246	[-0.521, 0.0281]
SC3	-0.646*	[-0.963, -0.329]	-0.622*	[-0.957, -0.286]
UA2	0.14	[-0.0740, 0.355]	0.0786	[-0.230, 0.388]
UA3	-0.316*	[-0.534, -0.0974]	-0.395	[-0.688, -0.102]
PD2	-0.285	[-0.493, -0.0769]	-0.309	[-0.573, -0.0451]
PD3	-1.10*	[-1.40, -0.798]	-1.20*	[-1.53, -0.858]
AD2	-0.427*	[-0.645, -0.209]	-0.391*	[-0.627, -0.154]
AD3	-1.03*	[-1.32, -0.740]	-1.10*	[-1.37, -0.822]
γ MO/MO2	-0.0195	[-0.610, 0.571]	-0.384	[-1.01, 0.239]
γ SC/SC2	-0.58	[-1.61, 0.446]	-1.06	[-2.17, 0.0586]
γ UA/UA2	-1.66	[-3.12, -0.211]	-4.08	[-18.4, 10.2]
γ PD/PD2	0.559	[-0.486, 1.60]	0.49	[-1.03, 2.01]
γ AD/AD2	-0.0988	[-0.678, 0.481]	0.297	[-0.763, 1.36]
γ MO3			0.0411	[-0.495, 0.577]
γ SC3			0.0966	[-1.09, 1.29]
γ UA3			-1.37	[-3.11, 0.373]
γ PD3			0.413	[-0.695, 1.52]
γ AD3			-0.306	[-1.15, 0.542]

Note. CI=95% confidence interval; N=199; *=statistical significance at 5% level after Holm's sequential Bonferroni correction (Holm, 1979)

Table A.5: Latent scale results for carers who care for a partner

	One param.	CI	Two param.	CI
MO2	-0.368*	[-0.573, -0.162]	-0.291	[-0.554, -0.0285]
MO3	-2.04*	[-2.51, -1.56]	-2.10*	[-2.71, -1.49]
SC2	-0.312	[-0.568, -0.0560]	-0.332	[-0.621, -0.0432]
SC3	-0.636*	[-0.994, -0.278]	-0.666*	[-1.06, -0.276]
UA2	0.174	[-0.132, 0.481]	-0.0054	[-0.0179, 0.00714]
UA3	-0.369	[-0.629, -0.108]	-0.457	[-0.757, -0.157]
PD2	-0.326	[-0.573, -0.0783]	-0.303	[-0.582, -0.0247]
PD3	-1.22*	[-1.53, -0.915]	-1.28*	[-1.62, -0.931]
AD2	-0.276	[-0.519, -0.0342]	-0.183	[-0.510, 0.145]
AD3	-1.06*	[-1.32, -0.799]	-1.12*	[-1.43, -0.819]
γ MO/MO2	0.0123	[-0.490, 0.515]	0.551	[-1.15, 2.25]
γ SC/SC2	-1.08	[-2.11, -0.0499]	-1.05	[-2.08, -0.0289]
γ UA/UA2	-1.26	[-2.71, 0.200]	26.1	[-28.3, 80.6]
γ PD/PD2	-0.229	[-0.990, 0.532]	0.141	[-1.28, 1.56]
γ AD/AD2	0.0447	[-1.69, 1.78]	1.22	[-2.95, 5.39]
γ MO3			-0.229	[-0.715, 0.258]
γ SC3			-1.22	[-2.51, 0.0740]
γ UA3			-0.775	[-2.65, 1.10]
γ PD3			-0.586	[-1.64, 0.471]
γ AD3			-1.69	[-4.79, 1.40]

Note. CI=95% confidence interval; N=199; *=statistical significance at 5% level after Holm's sequential Bonferroni correction (Holm, 1979)