
Article:
Norman, R., Mercieca-Bebber, R., Rowen, D. orcid.org/0000-0003-3018-5109 et al. (7 more authors) (2019) UK utility weights for the EORTC QLU-C10D. Health Economics, 28 (12). pp. 1385-1401. ISSN 1057-9230

https://doi.org/10.1002/hec.3950

TITLE: UK utility weights for the EORTC QLU-C10D

Richard Norman, PhD. School of Public Health, Curtin University, Perth, Australia; Richard.norman@curtin.edu.au; +61 8 92661327

Rebecca Mercieca-Bebber, PhD. Faculty of Science, School of Psychology, University of Sydney, Sydney, Australia; Faculty of Medicine, Sydney Medical School, Central Clinical School, University of Sydney, Sydney, Australia; NHMRC Clinical Trials Centre, University of Sydney, Sydney, Australia.

Donna Rowen, PhD. School of Health and Related Research, University of Sheffield, Sheffield, UK

John E. Brazier, PhD. School of Health and Related Research, University of Sheffield, Sheffield, UK

David Cella, PhD. Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

A. Simon Pickard, Department of Pharmacy Systems, Outcomes and Policy, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois, USA.

Deborah J Street, PhD. Centre for Health Economics Research and Evaluation (CHERE), University of Technology Sydney, Sydney, Australia

Rosalie Viney, PhD. Centre for Health Economics Research and Evaluation (CHERE), University of Technology Sydney, Sydney, Australia

Dennis Revicki PhD. Evidera Inc, Center for Health Outcomes Research, Bethesda, USA

Madeleine King, PhD. Faculty of Science, School of Psychology, University of Sydney, Sydney, Australia; Faculty of Medicine, Sydney Medical School, Central Clinical School, University of Sydney, Sydney, Australia

On behalf of the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group

ABSTRACT

Background: The EORTC QLU-C10D is a new multi-attribute utility instrument derived from the widely-used cancer-specific quality of life (QoL) questionnaire, EORTC QLQ-C30. It contains ten dimensions (Physical, Role, Social, Emotional Functioning; Pain, Fatigue, Sleep, Appetite, Nausea, Bowel Problems), each with 4 levels.

Objective: To provide United Kingdom (UK) general population utility weights for the QLU-C10D.

Methods: A UK online panel was quota-sampled to align the sample to general population proportions of sex and age (≥18 years). The online valuation survey included a discrete choice experiment (DCE) Each participant was asked to complete 16 choice-pairs, each comprising two QLU-C10D health states plus duration. DCE data were analysed using conditional logistic regression to generate utility weights.

Results: Data from 2,187 respondents who completed at least one choice set were included in the DCE analysis. The final UK QLU-C10D utility weights comprised decrements for each level of each health dimension. For nine of the ten dimensions (all except appetite), the expected monotonic pattern was observed across levels: utility decreased as severity increased. For the final model, consistent monotonicity was achieved by merging inconsistent adjacent levels for appetite. The largest utility decrements were associated with physical functioning and pain. The worst possible health state (the worst level of each dimension) is -0.083, which is considered slightly worse than being dead.

Conclusions: UK-specific utility weights will enable cost-utility analyses (CUA) for the economic evaluation of new oncology therapies and technologies in the UK, where CUA is commonly used to inform resource allocation.

Compliance with Ethical Standards

Funding: The development of the EORTC QLU-C10D research was supported by a National Health and Medical Research Council (Australia) Project Grant (632662). Adaptation of the survey to the UK setting and fieldwork was supported by Abbvie via a subcontract agreement with Evidera (EVA-17629). Associate Professor Norman was supported by a NHMRC Early Career Research Fellowship (1069732). Professor King was supported by the Australian Government through Cancer Australia.

Conflict of Interest: The authors declare they do not have any conflicts of interest in relation to the material reported in this manuscript.

Ethical Approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The need for NHS Research Ethics Committee (REC) approval was formally waived by the Medical Research Council on 19 August 2016 (IRAS project ID 213144).

1. Introduction

Economic evaluation of new health care therapies and technologies is central to decision making in many countries, including the United Kingdom (UK), for example through the National Institute for Health and Care Excellence (NICE) and the Scottish Medicines Consortium (SMC)) (1, 2). Cost-utility analysis (CUA) is a policy-relevant and valuable form of economic evaluation because it quantifies health outcomes on a metric that is applicable across health conditions. In CUA, the quality-adjusted life year (QALY) is a widely used metric of health which allows comparison between different disease areas as it captures changes in both morbidity and mortality. The quality adjustment metric used to calculate QALYs is a utility index (or utility weight), with a maximum of one representing full health, zero representing being dead, and negative values representing health states worse than death (3). The quality adjustment metric is usually generated using preference-based measures (PBMs) such as the EQ-5D (4) or SF-6D (5, 6). These have two components: a descriptive system that systematically describes a comprehensive set of health states (also referred to as a 'health state classification system'), and utility weights generated using a valuation survey that elicits preferences (conventionally from a general population sample) to yield societal preference weights. These utility weights produce a utility score for every health state in the classification system which can then be used to generate QALYs.

PBMs can either be developed *de novo* or by adapting pre-existing health-related quality of life (HRQoL) profile measures (7, 8). The advantage of the latter approach is that it means utility weights can be derived using data that are already collected using an existing measure without additional respondent burden. Such an approach has been applied to the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire-Core 30 (QLQ-C30), a widely-used questionnaire for people with cancer (9). The Multi-Attribute Utility in Cancer (MAUCa) Consortium is a group of international researchers who have initiated a series of country-specific value sets for the QLQ-C30. Previous work by Rowen et al. (10) used factor analysis and Rasch analysis on QLQ-C30 data from a sample of 655 UK-based multiple myeloma patients to derive a descriptive system and then valued a subset of health states using the Time Trade-Off in face-to-face interviews conducted in a UK general population sample of 350. The present study uses a somewhat different descriptive system based on secondary analysis of a larger (n=2616) and more diverse collection of cancer-related datasets, including 13 countries, 15 primary cancer types, localised/regional (n=1037) and recurrent/metastatic stages (n=1579) (11). The resultant descriptive system contains 13 of the QLQ-C30's 30 items across 10 dimensions (Table 1), forms the basis of the EORTC QLU-C10D, and is endorsed by the EORTC as its official PBM (11).

The EORTC QLU-C10D descriptive system has been valued in an Australian general population sample (12, 13). However, preferences for health may differ across countries (for example see Ferreira et al. (14) and Norman et al. (15)). Such differences may reflect underlying cultural, social and work differences across countries, different sociodemographic profiles of valuation samples (for example see Dolan and Roberts (16)), and artefactual differences caused by use of a range of elicitation techniques and study protocols (17). Many international agencies specify preferred methods for health technology assessment submissions to inform resource allocation decisions. Several countries require country-specific utility weights (18). NICE in particular requires UK weights,

and further specifies that the utility weights should be elicited using a representative sample of the general population (1).

It is important any real differences in general population preferences are reflected in utility algorithms, so country-specific weights need to be developed. To avoid artefactual confounding across countries, the MAUCa Consortium derived a standard method for valuing the EORTC QLU-C10D in general population samples, described below. Using this method, we have estimated value sets in Australia (19), Europe (20) and Canada (21). The aim of this current paper is to apply this valuation method in a UK general population sample to produce UK-specific utility weights for the QLU-C10D.

2. Methods

2.1 The QLU-C10D

Table 1 shows the QLU-C10D descriptive system, and explains how the 10 dimensions, each with four levels, map to 13 of the 30 items in the QLQ-C30. The derivation of this health state classification system is described elsewhere (22). It is important to note that the wording of the ten dimensions are based on the original wording of the QLQ-C30. For the items that use more than one QLQ-C30 item (specifically Physical Functioning, Social Functioning, and Bowel Function) the wording of the QLU-C10D levels were carefully selected to ensure translation from the original items, while also reflecting the inter-relatedness of the composite items. For example, the Physical Functioning levels assume that the problems an individual faces with a long walk are at least as bad as for a short walk.

Table 1 The QLU-C10D health state classification system, how it maps to the 13 component items from the QLQ-C30, and the duration attribute included the discrete choice experiment (DCE) valuation survey

Dimension	Level	Stem	Descriptor	QLQ-C30 item scores
Physical Functioning[a,b]	1	You have…	No trouble taking a long walk outside of the house	Item 2 (long walk) = 1
	2		No trouble taking a short walk outside of the house, but at least a little trouble taking a long walk	Item 3 (short walk) = 1 AND Item 2 $\geq$ 2
	3		At least a little trouble taking a short walk outside of the house, and at least a little trouble taking a long walk	Item 3 = 2 AND Item 2 $\geq$ 2
	4		Quite a bit or very much trouble taking a short walk outside the house	Item 3 $\geq$ 3 AND Item 2 $\geq$ 2
Role Functioning	1	You are limited in pursuing your work or other daily activities…	Not at all	Item 6 = 1
	2		A little	Item 6 = 2
	3		Quite a bit	Item 6 = 3
	4		Very much	Item 6 = 4
Social Functioning[a,c]	1	Your physical condition or medical treatment interferes with your social or family life…	Not at all	Items 26 AND 27 = 1
	2		A little	Items 26 OR 27 = 2
	3		Quite a bit	Items 26 OR 27 = 3
	4		Very much	Items 26 OR 27 = 4
Emotional Functioning	1	You feel depressed…	Not at all	Item 24 = 1
	2		A little	Item 24 = 2
	3		Quite a bit	Item 24 = 3
	4		Very much	Item 24 = 4

Pain	1	You have pain…	Not at all	Item 9 = 1
	2		A little	Item 9 = 2
	3		Quite a bit	Item 9 = 3
	4		Very much	Item 9 = 4
Fatigue	1	You feel tired…	Not at all	Item 18 = 1
	2		A little	Item 18 = 2
	3		Quite a bit	Item 18 = 3
	4		Very much	Item 18 = 4
Sleep	1	You have trouble sleeping…	Not at all	Item 11 = 1
	2		A little	Item 11 = 2
	3		Quite a bit	Item 11 = 3
	4		Very much	Item 11 = 4
Appetite	1	You lack appetite…	Not at all	Item 13 = 1
	2		A little	Item 13 = 2
	3		Quite a bit	Item 13 = 3
	4		Very much	Item 13 = 4
Nausea	1	You feel nauseated…	Not at all	Item 14 = 1
	2		A little	Item 14 = 2
	3		Quite a bit	Item 14 = 3
	4		Very much	Item 14 = 4
Bowel Problems[a,c]	1	You…	do not have constipation or diarrhoea at all	Items 16 AND 17 = 1
	2		have a little constipation or diarrhoea	Items 16 OR 17 = 2
	3		have constipation or diarrhoea quite a bit	Items 16 OR 17 = 3
	4		have constipation or diarrhoea very much	Items 16 OR 17 = 4
Duration	1	You will live in this health state for…	1 year, and then die	Not applicable
	2		2 years, and then die	Not applicable
	3		5 years, and then die	Not applicable
	4		10 years, and then die	Not applicable

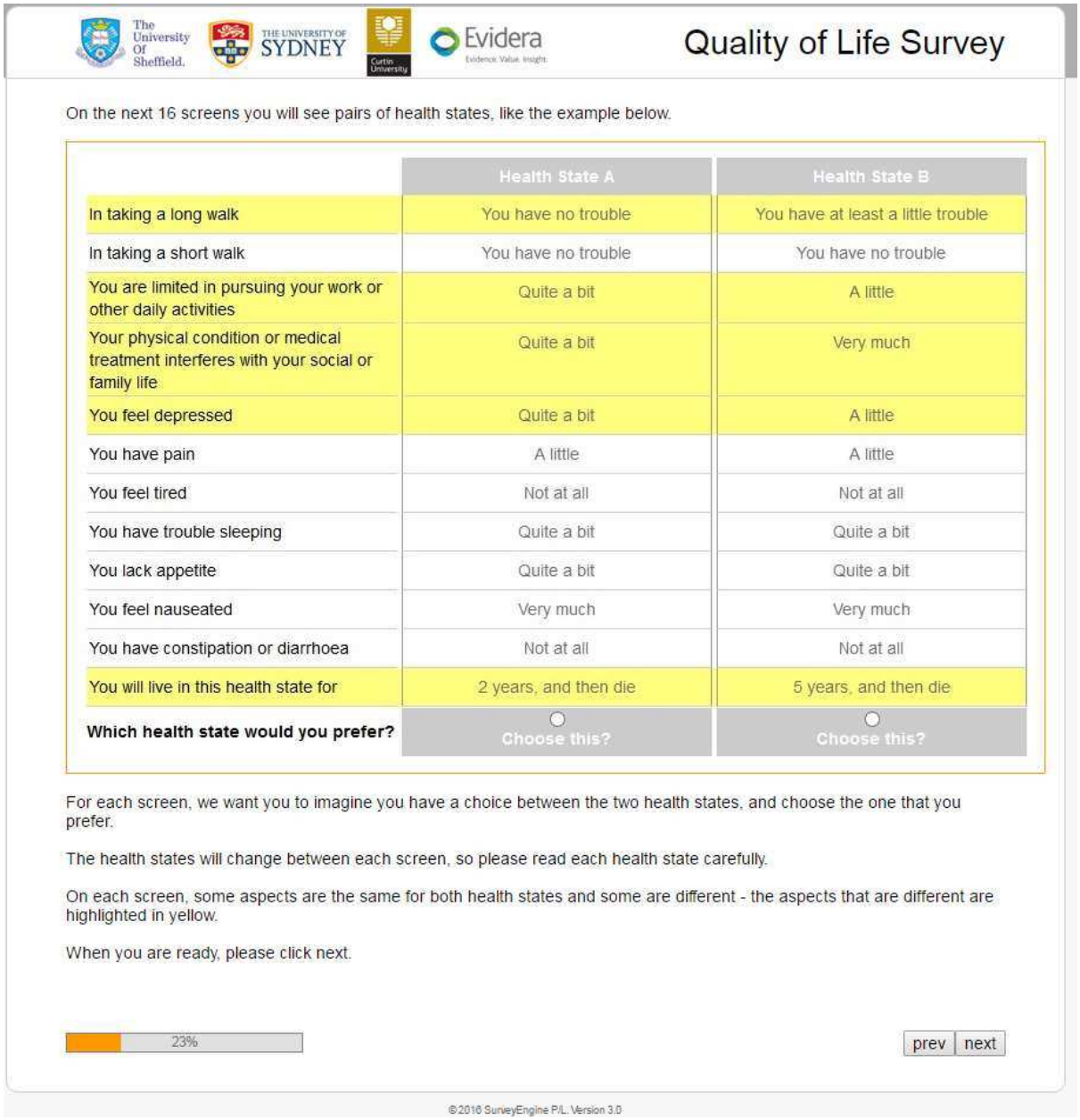
a. Three dimensions of the QLU-C10D each involve two QLQ-C30 items.
b. The Physical Functioning dimension includes 'long walk' and 'short walk' from the QLQ-C30; for the DCE, the levels are determined together, but were presented in the DCE survey separately, as shown in Figure 1.
c. For Social Functioning and Bowel Problems, the QLU-C10D level is determined by the maximum value of the two component items.

2.2 The valuation task

The study used the valuation protocol successfully developed for the Australian valuation of the QLU-C10D (19), with modifications only to the participant demographic section. The DCE component of the valuation survey was informed by previous valuation studies for the EQ-5D-3L and SF-6D instruments (23, 24). The DCE is a technique in which ordinal data is generated by asking respondents to choose between multiple alternatives typically described by a set of dimensions each with a range of possible levels. The inference of cardinal values from ordinal data was first proposed by Thurstone (25). The random utility model, developed by McFadden (26), operationalised this concept. In this, utility is a function of a systematic term and a random term. When faced with a choice, individuals will select the option with the higher utility. McFadden's conditional logit, and its extensions including the mixed logit and generalised multinomial logit (27), have been widely used in health economics generally (28), and in the field of health state valuation specifically (29).

For the QLU-C10D, the DCE task presented pairs of QLU-C10D health states each with a specified duration (life years), which are described as Health States A and B (see Figure 1 for an example).

Figure 1: An Example Choice Set



The QLU-C10D has a substantial number of dimensions, an issue that requires consideration to ensure respondents are able to fully consider the question being posed to them. Previous work explored the feasibility of the task in an online sample, and tested different presentational formats (13). As illustrated in Figure 1, dimensions that differed between Health States A and B were colour coded in yellow. Secondly, the Physical Functioning dimension was adapted to mitigate the impact of having complicated descriptors for levels 2 and 3 (see Table 1). In the DCE task, we presented 'long walk' and 'short walk' as two separate attributes (Figure 1). However, in both the experimental design and data analysis, the Physical Functioning dimension was treated as a single 4-level dimension (online resource 1).

The duration levels were selected balancing a number of considerations. First, we wanted levels that were plausible for most respondents; offering a 90-year old 50 years would be inappropriate. Second, we wanted a broad enough range of years that respondents would have preferences between levels. Third, and in contrast with the previous point, we wanted levels that would not consistently dominate the choice set. The levels used here are similar to both TTO studies, and other DCE studies, in which duration did not dominate decisions.

2.3 Health states valued - DCE design

The design was the same employed in the Australian study (12), and consisted of 960 choice sets. Each set is a paired comparison between options consisting of two health states. Health states were operationalised as 12 attributes in the DCE (i.e. duration, 2 attributes for physical functioning, and the 9 remaining attributes). To further encourage ease of completion and comprehension, we simplified the cognitive task by constraining the number of HRQoL dimensions that differed between health states in any given choice set to four, as was the case done in the preparatory methods experiment (30).

Each respondent was randomly allocated to answer 16 choice sets, but ensuring that each choice set was seen by a similar number of respondents. Each time a choice set was seen, which option was seen as Health State A or B was randomised to mitigate ordering bias. The order of dimensions was randomised for each person to prevent any order effect.

2.4 Survey content

Figure 2 presents the survey outline and ordering of sections on the survey. Beyond the DCE, the survey asked a range of other questions, including the general health question of the SF-36 (31). The sociodemographic questions were worded to allow comparison with normative UK data (see Table 2 for more detail). Respondents were not informed that the health states were cancer-related, as it has been found previously that the mention of disease labels impacts on health state values, probably because prior knowledge or preconceptions of health conditions distort survey participants' values and preferences (32, 33). Respondents completed the QLU-C10D prior to the DCE tasks which familiarised respondents with the classification system.

2.5 Survey implementation and sample recruitment

The content was implemented by SurveyEngine (34), a company specialising in administering choice experiments in an online setting. SurveyEngine and its panel providers comply with the International Code on Market, Opinion and Social Research and Data Analytics (35). Members of the online panel who are residents of the UK were eligible if they were at least 18 years of age. Online panellists received an email invitation to participate, including a hyperlink to the survey, which they could then complete at their convenience. Those who chose to participate by entering the survey were screened based on their responses to sex and age, to ensure quotas of participants matched those of the UK in the most recent census (2011) (36). Recruitment continued until quota sample targets were achieved. Upon completion, participants received a small payment. The survey was active between August-October 2016. Previous work has produced well-behaved and reliable algorithms using 1,000 respondents (23, 24, 37); here to ensure robust estimates, a target sample size of 2,000

was chosen. This is considerably larger than most DCEs in health, and provides an average of 33 observations per choice set, generally considered adequate for analysis (38).

2.6 Statistical analysis

2.6.1 Sample representativeness

Sociodemographic variables were summarized using appropriate descriptive statistics, and where population normative data were available, Chi-squared tests or one-sampled t-tests were used to assess our sample's population representativeness, using SPSS Version 22 (Armonk, NY: IBM Corp).

2.6.2 Utility estimation

Data analysis was undertaken using STATA 13 (39) using an approach previously used to comply with conventional QALY framework restrictions (30, 40-42). As described below, all terms in the utility function contain duration, thus we imposed the zero condition (43, 44). A functional form that satisfied this requirement included the QLU-C10D dimension levels interacted with the duration variable ('*TIME*') (see equations 1 and 2 below). Additionally, we imposed constant proportional time trade off (as is standard in the QALY framework).

The primary analysis was underpinned by Equation 1, in which the utility of option *j* in choice set *s* for survey respondent *i* was assumed to be:

$$U_{isj} = \alpha TIME_{isj} + \beta X'_{isj} TIME_{isj} + \varepsilon_{isj} \quad (1)$$

$$\text{i} = 1, \ldots, \text{I respondents; j} = \text{situations A, B; s} = 1, \ldots, 960 \text{ choice sets}$$

where α was the utility of a life year, and $\mathrm{X}'_{\mathrm{isj}}$ was a set of dummies representing QLU-C10D health state levels presented in option *j*. The error term $\varepsilon_{\mathrm{isj}}$ was assumed to follow a Gumbel distribution. The dead state is valued at zero, as TIME=0, and that removes the two parts of the systematic component of the utility function, as it appears in both. Thus, health states can be considered as worse than dead if the absolute value of $\beta X'_{isj} TIME_{isj}$ exceeds $\alpha TIME_{isj}$.

First, DCE responses were explored using a conditional logit model. Standard errors were adjusted to reflect intra-individual correlation through a clustered sandwich estimator (i.e. STATA's vce (cluster) option). To estimate utility decrements for a QLU-C10D value set, we divided each of the beta terms by alpha (with level one omitted in each dimension to allow identification). Confidence intervals were then estimated around these intervals using the delta method [23].

Model 1 included every move away from Level 1 (no problems) in each dimension within $\mathrm{X}'_{\mathrm{isj}}$. Thus, $\mathrm{X}'_{\mathrm{isj}}$ contained 30 dummies. The items in the QLU-C10D are naturally ordered. So, if we observed non-monotonic ordering in the coefficients for these dummies (i.e. if, as health worsened within each dimension, the absolute size of the coefficient increased), the non-monotonic adjacent levels were combined to estimate a consistent model (Model 2). This restriction has been

standardly imposed in previous studies (45-50). To estimate a value set, it was necessary to transform regression coefficients on to the 0-1 scale. To do this, we generated decrements for each level of each dimension, by dividing the relevant coefficient by the coefficient on duration (the alpha term in Equation 1). Then, to score individual health states, each of the decrements are subtracted from 1. As the decrements on the best level in each dimension are set at 0, this places the health state with level 1 in each dimension at 1. Goodwin and Green note that, for condition specific measures like the QLU-C10D, this is frequently done, but not universal (51). The reason for this assumption is that the QLU-C10D includes a wide variety of dimensions covering functions and symptoms. Thus, while it is condition-specific, we believe it is reasonable to assume that being at the best level in all of its dimensions reflects something close to ideal health.

To explore robustness around the conditional logit estimates, we conducted a series of split-sample validations. Thus, we selected half of the data at random, estimated the conditional logit, and predicted the results for the other half of the data. We repeated this ten times, and estimated the proportion of choice sets in the hold-out correctly predicted. To further explore robustness, we re-estimated the conditional logit adjusting for population non-representativeness in our sample. Specifically, we constructed inverse probability weights based on age, gender, marital status, highest educational attainment, and country of birth (as described in Table 2).

In a secondary approach, we employed a mixed logit (52) which has the advantage that it models preference heterogeneity (53). In Model 3, we assumed that coefficients were drawn from a distribution,

$$U_{isj} = (\alpha + \gamma_i)TIME_{isj} + (\beta + \eta_i)X'_{isj}TIME_{isj} + \varepsilon_{isj} \qquad (2)$$

The α and β terms now estimate population means, and γ_i and η_i are individual-specific variations from them. These latter terms were assumed to follow a multivariate normal distribution (0,∑). However, the standard command limits the number of parameters drawn from a distribution to be no more than 20. This is problematic as we do not have *a priori* reasons for selecting a subset of the coefficients to be drawn from a distribution. Therefore, we employed pseudo-random draws, which allowed all coefficients to vary across individuals (personal communication, Arne Risa Hole, University of Sheffield, 15 June 2015). To compare models in terms of model fit, Akaike information criterion (AIC) and Bayesian information criterion (BIC) estimates are presented.

3. Results

3.1 Sample characteristics and representativeness

Figure 2 shows the respondent flow through the survey, and Table 2 shows the sample sociodemographic characteristics relative to established population norms.

Figure 2. Flow diagram showing number of participants for each section of the survey

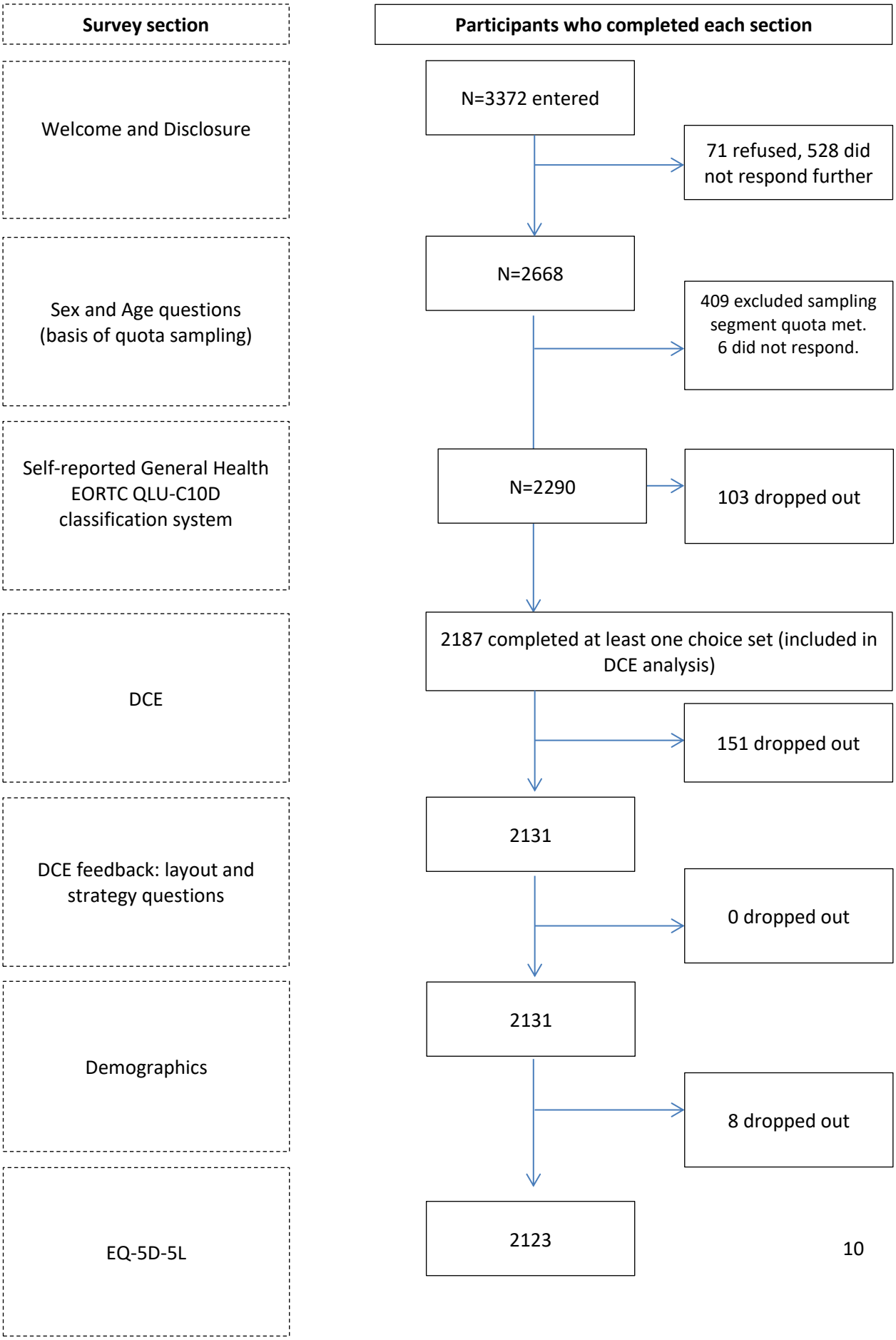


Table 2. Demographic characteristics of the sample compared to UK population norms where available[a]

Question	n Responders	Level	Sample n	Sample proportion	Population proportion	p value[b]
Country of Birth	2128	England	1694	0.796	0.719	<.001
		Scotland	88	0.072	0.082	
		Wales	153	0.041	0.044	
		Northern Ireland	45	0.021	0.029	
		Republic of Ireland	2	0.001	0.007	
		Other English speaking country	51	0.024	0.008	
		Other non-English speaking country	95	0.045	0.111	
Highest Level of Education	2086	No qualifications	109	0.052	0.232	<.001
		Level 1[c]	184	0.088	0.141	
		Level 2[d]	319	0.153	0.152	
		Apprenticeship	37	0.018	0.033	
		Level 3[e]	351	0.168	0.121	
		Level 4[f]	895	0.429	0.270	
		Other[g]	191	0.092	0.051	
Marital status	2128	Single (never married or never registered a same-sex civil partnership)	736	0.346	0.347	<.001
		Married	1017	0.478	0.465	
		In a registered same-sex civil partnership	13	0.006	0.002	
		Separated (but still legally married or still legally in a same-sex civil partnership)	41	0.019	0.027	
		Divorced or formerly in a same-sex civil partnership which is now legally dissolved	221	0.104	0.088	
		Widowed or surviving partner from a same-sex civil partnership	100	0.047	0.070	
Sex	2303	Male	1097	0.476	0.487	0.290
		Female	1206	0.524	0.513	
Age (years)	2302	18-29	459	0.199	0.201	0.746
		30-39	374	0.162	0.165	
		40-49	388	0.169	0.174	
		50-59	376	0.163	0.166	
		60-69	346	0.150	0.139	
		70 or older	359	0.156	0.156	

a Population values obtained from the 2011 UK census, available at https://www.ons.gov.uk/census#censusdataandbackground.

b p values <0.05 suggest sample is not representative of general population. For categorical variables, the chi-squared goodness of fit test was used to compare observed category frequencies to those expected based on population

proportions; for the continuous K10 score, a one-sample t-test compared the observed K10 mean to the population value reported by Slade et al, 2011.
c Level 1 is defined as 1 - 4 O levels / CSEs / GCSEs (any grades), Entry Level, Foundation Diploma, and NVQ Level 1, Foundation GNVQ, Basic Skills.
d Level 2 is defined as 5+ O levels (passes) / CSEs (grade 1) / GCSEs (grades A*- C), School Certificate, 1 A level / 2 - 3 AS levels / VCEs, Higher Diploma and NVQ Level 2, Intermediate GNVQ, City and Guilds Craft, BTEC First / General Diploma, RSA Diploma.
e Level 3 is defined as 2+ A levels / VCEs, 4+ AS levels, Higher School Certificate, Progression / Advanced Diploma and NVQ Level 3, Advanced GNVQ, City and Guilds Advanced Craft, ONC, OND, BTEC National, RSA Advanced Diploma.
f Level 4+ is defined as Degree (for example BA, BSc), Higher degree (for example MA, PhD, PGCE), and Professional qualifications (for example teaching, nursing, accountancy), and Other vocational / work-related qualifications.
g 'other' in our study sample includes all participants who selected the 'other' or 'foreign qualification' options.

Participants were representative of the general population in terms of age and sex (in accordance with our quota sampling). However, the sample was less representative on educational attainment, marital status and country of birth. The study sample was more educated than the general UK population, and more likely to be separated or divorced. A higher proportion of participants in the study sample were born in England or in other English-speaking countries, compared to the general UK population.

3.2 Utility estimates

Table 3 reports the two conditional logit analyses, either with no constraints placed on coefficients (Model 1) or with monotonicity within each dimension imposed (Model 2). This is a fully consistent model where utility decreases or stays the same as health worsens for every level and dimension.

Table 3: Conditional logit: Model 1 (unconstrained) and Model 2 (monotonicity imposed[a])

		Model 1	Model 2
		Coefficient (Robust SE)	Coefficient (Robust SE)
Duration	Linear	0.605 (0.027)***	0.605 (0.027)***
Physical Functioning x duration		-0.038 (0.010)***	
	2		-0.038 (0.010)***
	3	-0.094 (0.011)***	-0.094 (0.011)***
	4	-0.154 (0.010)***	-0.154 (0.010)***
Role Functioning x duration		-0.013 (0.007)*	
	2		-0.013 (0.007)*
	3	-0.047 (0.008)***	-0.046 (0.008)***
	4	-0.067 (0.007)***	-0.066 (0.007)***
Social Functioning x duration		-0.012 (0.007)*	
	2		-0.012 (0.007)*
	3	-0.043 (0.007)***	-0.043 (0.007)***
	4	-0.063 (0.007)***	-0.062 (0.007)***
Emotional Functioning x duration		-0.005 (0.007)	
	2		-0.005 (0.007)
	3	-0.029 (0.008)***	-0.030 (0.008)***
	4	-0.078 (0.007)***	-0.078 (0.007)***
Pain x duration	2	-0.018 (0.007)**	-0.018 (0.007)***
	3	-0.047 (0.008)***	-0.047 (0.007)***
	4	-0.091 (0.007)***	-0.091 (0.007)***

Fatigue x duration	2	-0.022 (0.006)***	-0.022 (0.006)***
	3	-0.029 (0.007)***	-0.029 (0.007)***
	4	-0.035 (0.006)***	-0.035 (0.006)***
Sleep x duration	2	-0.030 (0.006)***	-0.030 (0.006)***
	3	-0.034 (0.007)***	-0.034 (0.007)***
	4	-0.043 (0.006)***	-0.043 (0.006)***
Appetite x duration	2	*-0.022 (0.006)****	-0.017 (0.005)***
	3	*-0.017 (0.007)***	-0.017 (0.005)***
	4	*-0.014 (0.006)***	-0.017 (0.005)***
Nausea x duration	2	-0.029 (0.007)***	-0.029 (0.007)***
	3	-0.036 (0.007)***	-0.036 (0.007)***
	4	-0.057 (0.006)***	-0.057 (0.006)***
Bowel problems x duration	2	-0.032 (0.007)***	-0.032 (0.007)***
	3	-0.037 (0.007)***	-0.037 (0.007)***
	4	-0.051 (0.007)***	-0.051 (0.006)***
Log-likelihood		-18474	-18475
Parameters		31	29
AIC		37011	37009
BIC		37294	37274

a Levels combined to ensure monotonicity within each dimension are noted in italics (required only for Appetite x duration).
As described in methods, the health problems enter as interactions with duration to impose the QALY model on the data
Levels of statistical significance: 1% ***; 5% **; 10% *

Respondents preferred extra years of life, as reflected in the positive and statistically significant coefficient on duration. For each dimension, an increased level of problems was valued negatively (Model 1). Of the 30 coefficients reported for levels of QLU-C10D dimensions interacted with duration all had the expected sign and all were significant at the 10% level, 28 were ordered appropriately, in that as health worsened the absolute size of the coefficient increased. In other words, movement to worse levels of each problem for each dimension was associated with increasing utility decrements (monotonicity), with the exception of the two highest levels of the appetite dimension. In this sample, the effect of appetite on choice was inconsistent and smaller than the effect of the other nine dimensions. Model 2, a fully consistent model, which combines levels 2, 3 and 4 of appetite (denoted *in italics* in the table) reports similar model fit and coefficients to Model 1. The coefficients for the other nine dimensions remain monotonic. As the model fit was not substantially impacted by constraining the model to be monotonic, and the value of monotonicity is apparent given the construction of the levels within each dimension, we therefore selected Model 2 as the preferred model for use in UK-based economic evaluations using the QLU-C10D.

3.3. Robustness

Regarding the hold-out runs, the level of prediction across the ten runs ranged from 63.6% to 64.4%. The second robustness check, which weighted responses to adjust for population non-representativeness on age, gender, highest educational attainment, country of birth, and marital status, suggested the effect of weighting was small. A scatter plot of decrements with and without weighting, showing the anchored coefficients in the derived value sets, is reported in Figure 3.

Figure 3: Scatter Plot of Weighted versus Unweighted Decrements (Model 1)

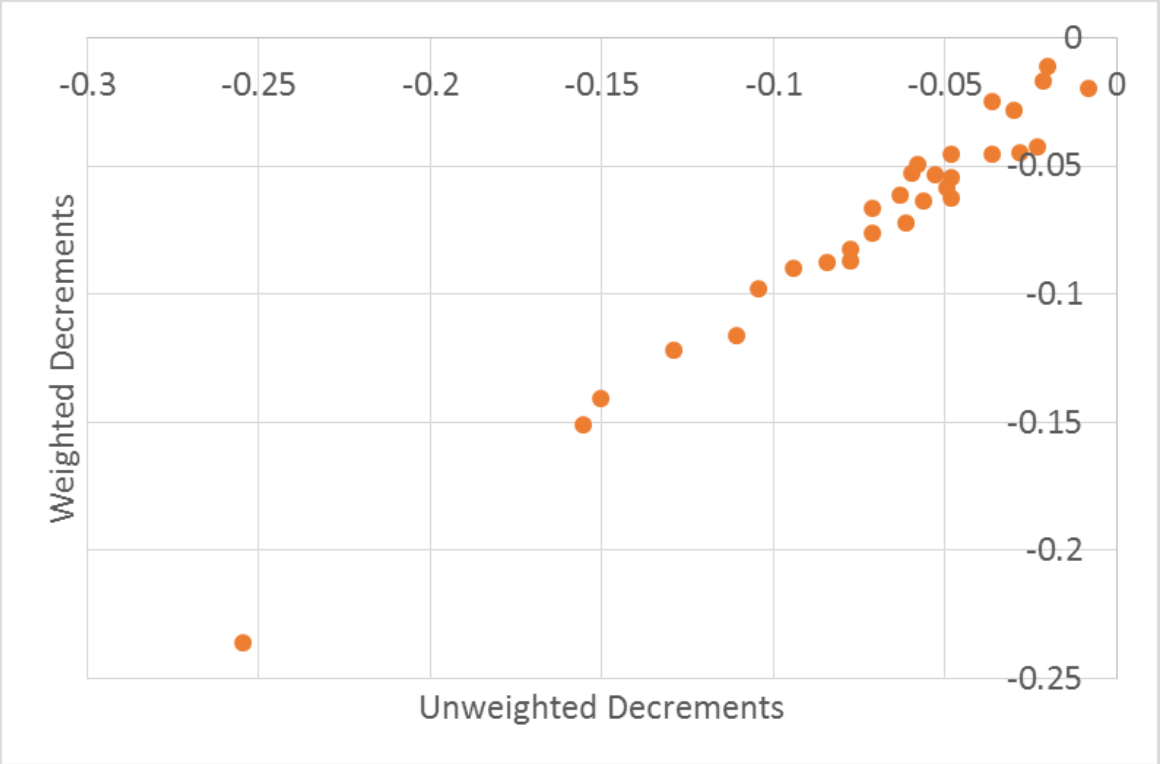


Table 4: Mixed Logit: Model 3

		Model 3	
		Coefficient (Robust SE)	Standard Deviation (SE)
Duration	Linear	1.476 (0.040)***	0.777 (0.022)***
Physical Functioning x duration	2	-0.126 (0.014)***	0.101 (0.041)**
	3	-0.155 (0.015)***	0.157 (0.022)***
	4	-0.235 (0.014)***	0.140 (0.020)***
Role Functioning x duration	2	-0.052 (0.013)***	0.106 (0.026)***
	3	-0.106 (0.014)***	0.165 (0.018)***
	4	-0.120 (0.012)***	0.108 (0.023)***
Social Functioning x duration	2	-0.033 (0.013)**	0.131 (0.025)***
	3	-0.099 (0.013)***	0.077 (0.029)***
	4	-0.166 (0.012)***	0.101 (0.019)***
Emotional Functioning x duration	2	0.022 (0.013)*	0.139 (0.017)***
	3	-0.060 (0.013)***	0.056 (0.020)***

	4	-0.206 (0.013)***	0.174 (0.019)***
Pain x duration	2	-0.024 (0.013)*	0.137 (0.017)***
	3	-0.100 (0.014)***	0.114 (0.020)***
	4	-0.205 (0.014)***	0.271 (0.016)***
Fatigue x duration	2	-0.027 (0.012)**	0.107 (0.021)***
	3	-0.045 (0.013)***	0.055 (0.030)*
	4	-0.083 (0.013)***	0.175 (0.019)***
Sleep x duration	2	-0.021 (0.012)*	0.127 (0.019)***
	3	-0.054 (0.013)***	0.003 (0.033)
	4	-0.079 (0.012)***	0.119 (0.021)***
Appetite x duration	2	-0.035 (0.012)***	0.097 (0.021)***
	3	-0.036 (0.013)***	0.055 (0.030)*
	4	-0.035 (0.012)***	0.120 (0.023)***
Nausea x duration	2	-0.034 (0.012)***	0.126 (0.020)***
	3	-0.073 (0.013)***	0.122 (0.021)***
	4	-0.136 (0.012)***	0.169 (0.017)***
Bowel problems x duration	2	-0.055 (0.012)***	0.139 (0.022)***
	3	-0.086 (0.013)***	0.113 (0.021)***
	4	-0.134 (0.012)***	0.160 (0.018)***
Log-likelihood		-15111	
Parameters		62	
AIC		-30098	
BIC		-29746	

In the mixed logit results (Model 3), the distribution means were generally monotonic, except in two dimensions. Emotional Functioning Level 2 had a positive coefficient but was not statistically significant (p=0.077 relative to Level 1). For Appetite, levels 3 and 4 were non-monotonic, but were much closer than in the conditional logit, varying by only 0.001 between levels 2, 3 and 4 (suggesting respondents did not (on average) discriminate between levels contingent on having any appetite problems.

The value set produced using Model 2 (in Table 3) provide the weights, w_{dl}, for calculating QLU-C10D scores. As is standard, a value of 1 is assigned to individuals whose are at level 1 in all 10 QLU-C10D dimensions. For other states, the utility score is 1 minus the aggregate utility decrement (w_{dl}) for each dimension not at Level 1. Thus, the utility score for a person p, determined by their QLU-C10D level l for each dimension d, is:

$$\text{QLU-C10D}_p = 1 + \sum_{d=1}^{10} w_{dl} \mid \text{QLU-C10D}_{dlp} \tag{3}$$

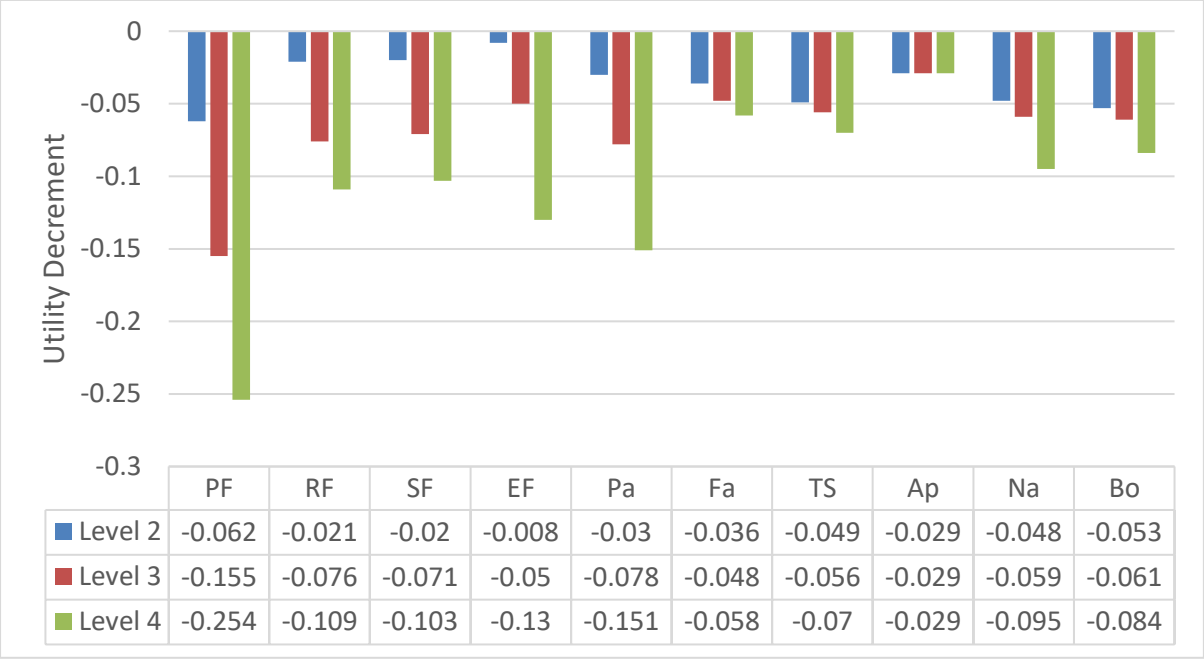
Table 5. The UK QLU-C10D Value Set

		Model 2
Dimension	**Level**	**Utility decrement, w_{dl}**

		(95% CI)
Physical Functioning (PF)	1	0
	2	-0.062 (-0.092,-0.033)
	3	-0.155 (-0.184,-0.125)
	4	-0.254 (-0.281,-0.226)
Role Functioning (RF)	1	0
	2	-0.021 (-0.043,0.000)
	3	-0.076 (-0.098,-0.054)
	4	-0.109 (-0.130,-0.089)
Social Functioning (SF)	1	0
	2	-0.020 (-0.042,0.001)
	3	-0.071 (-0.093,-0.049)
	4	-0.103 (-0.123,-0.084)
Emotional Functioning (EF)	1	0
	2	-0.008 (-0.031,0.014)
	3	-0.050 (-0.073,-0.027)
	4	-0.130 (-0.149,-0.110)
Pain (Pa)	1	0
	2	-0.030 (-0.052,-0.009)
	3	-0.078 (-0.100,-0.056)
	4	-0.151 (-0.170,-0.131)
Fatigue (Fa)	1	0
	2	-0.036 (-0.056,-0.016)
	3	-0.048 (-0.069,-0.027)
	4	-0.058 (-0.077,-0.039)
Sleep (TS)	1	0
	2	-0.049 (-0.069,-0.029)
	3	-0.056 (-0.077,-0.036)
	4	-0.070 (-0.090,-0.051)
Appetite (Ap)	1	0
	2	-0.029 (-0.045,-0.013)
	3	-0.029 (-0.045,-0.013)
	4	-0.029 (-0.045,-0.013)
Nausea (Na)	1	0
	2	-0.048 (-0.068,-0.028)
	3	-0.059 (-0.081,-0.037)
	4	-0.095 (-0.113,-0.076)
Bowel problems (Bo)	1	0
	2	-0.053 (-0.073,-0.033)
	3	-0.061 (-0.083,-0.039)
	4	-0.084 (-0.104,-0.064)

Figure 4: Utility decrements from conditional logit model constrained for monotonicity (Model 2

algorithm)



	PF	RF	SF	EF	Pa	Fa	TS	Ap	Na	Bo
Level 2	-0.062	-0.021	-0.02	-0.008	-0.03	-0.036	-0.049	-0.029	-0.048	-0.053
Level 3	-0.155	-0.076	-0.071	-0.05	-0.078	-0.048	-0.056	-0.029	-0.059	-0.061
Level 4	-0.254	-0.109	-0.103	-0.13	-0.151	-0.058	-0.07	-0.029	-0.095	-0.084

So, to give an example, health state 2221113212 (with each number relating to the level of the dimension as ordered in Table 4 and Figure 4) would be valued as 1-(0.062+0.021+0.02+0.056+0.029+0.053) = 0.759.

4. Discussion

This paper reports UK utility weights for the EORTC QLU-C10D, derived from the UK general population, which can subsequently be used to inform cost-utility analyses of interventions in oncology. This allows direct estimation of utility scores from responses to the QLQ-C30, a widely-used cancer-specific HRQoL questionnaire. Thus, scores can be estimated without additional surveying of respondents, and can also be estimated retrospectively in studies that have not collected utility data. Second, relative to generic measures such as the EQ-5D or SF-6D, it is likely to better capture quality of life in areas most likely to be affected by cancer or cancer treatment, such as fatigue, nausea and bowel problems. Arguably, this should provide a more cancer-sensitive measure of utility. Future research is ongoing to evaluate whether the utilities derived from the QLU-C10D are more sensitive to oncology treatment differences compared with utilities derived from the EQ-5D or other generic utility scales.

The dimensions with the largest utility weights were the generic dimensions (particularly around role and pain). However, cancer-sensitive dimensions were also associated with utility decrements, particularly nausea and bowel problems. Problems with sleep and fatigue had smaller decrements, perhaps because problems with sleep and fatigue are relatively common even in a general population, therefore were less influential in survey respondents' choices between health states than other DCE attributes. The results presented here are similar in range to the Australian QLU-

C10D valuation results (11). In the Australian value set, the value of the worst health state was -0.096, compared with a value of -0.083 in this UK value set.

The unconstrained regression reported in Model 1 is largely monotonic with the exception of the higher levels of appetite. This effect was also seen in the Australian survey where levels 3 and 4 were inconsistent in the unconstrained model (12). As yet, it is uncertain why this pattern occurs. Given the coefficients for levels 2-4 are close to one another and small, it may be that participants considered appetite to be a secondary concern relative to the other attributes, and hence largely ignored it. Alternatively, it may be that lacking appetite is not always considered undesirable among many general population respondents, particularly if they are considering this from a diet and weight-conscious perspective rather than in the context of poor health. It may be that cancer patients, with appetite loss caused by disease or treatment, may have a different perception of its negative impact; this assertion will be tested in a cancer sample, as mentioned above. The correction applied in Model 2 did not substantially impact on model fit or the other utility decrements. We therefore believe Model 2 appropriate for use in cost-utility analyses in the UK, and recommend using the algorithm described in Table 5.

The role of condition-specific preference-based measures such as the EORTC QLU-C10D is debated by health economists (see for example Versteegh *et al* and Brazier *et al* (7, 54)). For use in health technology assessments undertaken to inform resource allocation decisions, their use is often limited to conditions where a generic preference-based measure is not appropriate, sensitive or responsive, or for use in sensitivity analyses (8). This is for reasons of comparability, where health technology assessment is used to inform the allocation of resources across interventions, conditions and populations, and NICE in particular recommend EQ-5D in their reference case (1). Condition-specific preference-based measures do have many advantages in that aspects of their content may seem more relevant to patients, due to their greater focus on dimensions that are important for that condition, and can reduce patient burden if the data is already collected using the parent measure, in this case the EORTC QLQ-C30. Furthermore they enable utility values to be generated directly for datasets where no PBM was included; this may be desirable prospectively when patient burden is an issue, and retrospectively where an economic evaluation was not initially planned. For these reasons they can be used in sensitivity analyses for health technology assessment, and can provide important information that may not be captured by a generic preference-based measure. They also have a role outside of health technology assessment where they can be used to indicate both health and treatment effects (8), and for the purpose of resource allocation within the scope of health systems focused on cancer care (55).

We believe the algorithm reported here is valuable for a number of reasons. It uses a standardised protocol allowing comparison of preferences between countries. It is likely to be sensitive to changes in cancer-specific quality of life, not just around the effect of the disease itself, but of the common side-effects associated with many common treatments. Finally, it is adapted from a widely-used quality of life instrument, so can be applied both retrospectively and prospectively without requiring extra data collection and respondent burden, as noted above.

However, it is important to note some of the possible limitations of the work. First, we have used a partial profile approach, in which only a subset of the dimensions differ in each choice set. However,

we believe that the possible loss of statistical efficiency is largely offset through higher quality of data and lower drop out. For example, recent work demonstrated that the use of partial profiles reduced the use of simplifying heuristics, and drop-out rates (56). Second, arguably online panels may be non-representative of the general population, not only in the sociodemographic variables that we recorded, but more importantly, in ways that may influence their health preferences. We do know that the online completion of these tasks provides similar results to face to face administration (57), but non-representativeness of the sample is a valid concern. However, there are similar concerns around non-representativeness of samples using face to face administration.

The mixed logit results demonstrate that there is considerable heterogeneity in our sample. For economic evaluation that uses mean population preferences, this is arguably not a concern. However, an area where this might be important is in uses of the algorithm concerned with valuing a particular individual's health. For coefficients with a large standard deviation, the certainty with which we can predict the value that individual places on that health state is reduced, and researchers working in that area should be aware of this issue.

It is worthwhile to consider the differences between the EORTC QLU-C10D and the EORTC-8D (10) the existing preference-based measure from the QLQ-C30 with UK preference weights, both in terms of the descriptive system and the valuation algorithm. The two instruments have considerable overlap in dimensions and content, but the QLU-C10D contains two additional dimensions relating to appetite and sleep, which are not directly included in the EORTC-8D. Also, different items were selected to represent the dimensions of role functioning and pain. Of potentially greater importance is the difference in valuation approach. The EORTC-8D study used a Time Trade-Off, yielding a worst state value of 0.291 in comparison to -0.083 for the newer instrument. The lower value in this DCE may reflect the additional (and slightly amended) items, but may also reflect this difference in valuation technique. The impact of this difference is expected to be that, in general, the algorithm reported here will prioritise interventions promoting quality of life improvements, though this will differ depending on the exact change as the impact differs across dimensions. Though these differences are of interest note that the EORTC officially endorses only the EORTC QLU-C10D. Further research will compare the QLU-C10D to existing generic PBMs including EQ-5D to greater understand the impact on QALYs from the choice of measure.

The results presented here allow researchers to use EORTC QLQ-C30 data to populate cost-utility analyses. This is of considerable value given the range of cancer therapies becoming available in the UK and beyond, and the widespread use of the QLQ-C30.

References

1. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. London: NICE; 2013.
2. Scottish Medicines Consortium. Guidance to Manufacturers for Completion of New Product Assessment Form (NPAF). NHS Scotland; 2016.
3. Neumann P, Sanders GD, Russell LB, Siegel JE, Ganiats TG. Cost-Effectiveness in Health and Medicine (2nd edition). New York: Oxford University Press; 2017.
4. Brooks R. EuroQol: the current state of play. Health Policy. 1996;37(1):53-72.
5. Brazier J, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care. 2004;42(9):851-9.
6. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002;21(2):271-92.
7. Brazier JE, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). Health Technol Assess. 2012;16(32):1-114.
8. Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The Role of Condition-Specific Preference-Based Measures in Health Technology Assessment. Pharmacoeconomics. 2017;35(Suppl 1):33-41.
9. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organisation for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute. 1993;85(5):365-76.
10. Rowen D, Brazier J, Young T, Gaugris S, Craig BM, King MT, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2011;14(5):721-31.
11. King MT, Costa DSJ, Aaronson NK, Brazier JE, Cella DF, Fayers PM, et al. Derivation of the health state classification system for the QLU-C10D, an internationally-valid cancer-specific multi-attribute utility instrument derived from the EORTC core quality of life questionnaire, QLQ-C30. Qual Life Res. 2016.
12. King MT, Viney R, Pickard AS, Rowen D, Aaronson NK, Brazier JE, et al. Australian Utility Weights for the EORTC QLU-C10D, a Multi-Attribute Utility Instrument Derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30. Pharmacoeconomics. 2018;36:225-38.
13. Norman R, Viney R, Aaronson NK, Brazier JE, Cella DF, Costa DSJ, et al. Using a discrete choice experiment to value the QLU-C10D: Feasibility and sensitivity to presentation format. Qual Life Res. 2016.
14. Ferreira LN, Ferreira PL, Rowen D, Brazier JE. Do Portuguese and UK health state values differ across valuation methods? Qual Life Res. 2011;20(4):609-19.
15. Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International Comparisons in Valuing EQ-5D Health States: A Review And Analysis. Value Health. 2009;12(8):1194-200.
16. Dolan P, Roberts J. To what extent can we explain time trade-off values from other information about respondents? Soc Sci Med. 2002;54(6):919-29.
17. Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health. 2014;17(4):445-53.
18. Rowen D, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. International Regulations and Recommendations for Utility Data for Health Technology Assessment. Pharmacoeconomics. In Press.
19. King MT, Viney R, Pickard AS, Rowen D, Aaronson NK, Brazier JE, et al. Australian utility weights for the EORTC QLU-C10D, a multi-attribute utility instrument derived from the cancer-specific quality of life questionnaire, EORTC QLQ-C30. Pharmacoeconomics. Provisionally accepted pending minor revisions, 7 Aug 2017.
20. Kemmler G, Gamper E, Nerich V, Norman R, King MT, Holzner B. Comparison of German,

French and Polish utility weights for the EORTC utility instrument QLU-C10D. Qual Life Res. 2016;25(Supp 1):1027.
21. McTaggart-Cowan H, King MT, Chan K, Costa DSJ, Hoch J, Norman R, et al. Estimating the Canadian utility weights for the cancer-specific preference-based instrument, QLU-C10D. Qual Life Res. 2017;26(49).
22. King MT, Costa DS, Aaronson NK, Brazier JE, Cella DF, Fayers PM, et al. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation. 2016;25(3):625-36.
23. Norman R, Viney R, Brazier J, Burgess L, Cronin P, King MT, et al. Valuing SF-6D health states using a Discrete Choice Experiment. Med Decis Making. 2014;34(6):773-86.
24. Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, et al. An Australian discrete choice experiment to value eq-5d health States. Health Econ. 2014;23(6):729-42.
25. Thurstone LL. A law of comparative judgment. Psychol Rev. 1927;34:273-86.
26. McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P, editor. Frontiers in Econometrics. New York: New York Academic Press; 1974. p. pp.105-42.
27. Fiebig D, Keane M, Louviere J, Wasi N. The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. Marketing Science. 2010;29(3):393-421.
28. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883-902.
29. Mulhern B, Norman R, Street DJ, Viney R. One Method, Many Methodological Choices: A Structured Review of Discrete-Choice Experiments for Health State Valuation. Pharmacoeconomics. 2019;37(1):29-43.
30. Norman R, Viney R, Aaronson NK, Brazier JE, Cella DF, Costa DS, et al. Using a discrete choice experiment to value the QLU-C10D: Feasibility and sensitivity to presentation format. Quality of Life Research. 2016;25(3):637-49.
31. Ware JE, Jr., Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. Journal of Clinical Epidemiology. 1998;51(11):903-12.
32. Rowen D, Brazier J, Tsuchiya A, Young T, Ibbotson R. It's all in the name, or is it? The impact of labeling on health state values. Med Decis Making. 2012;32(1):31-40.
33. Green C, Goodwin E, Hawton A. "Naming and Framing": The Impact of Labeling on Health State Values for Multiple Sclerosis. Med Decis Making. 2017;37(6):703-14.
34. P/L S. SurveyEngine Choice Modelling Process & Method Carlton North, Vic: SurveyEngine Pty Ltd; 2015 [Available from: http://surveyengine.com.
35. ICC/ESOMAR. ICC/ESOMAR International Code on Market, Opinion and Social Research and Data Analytics 2016 [Available from: https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ICCESOMAR-International-Code_English.pdf.
36. Office for National Statistics. 2011 UK census 2011 [Available from: https://www.ons.gov.uk/census#censusdataandbackground.
37. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate societal health state utility values. J Health Econ. 2012;31:306-18.
38. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics. 2008;26(8):661-77.
39. StataCorp. Stata Statistical Software: Release 13. College Station, TX.: StataCorp LP; 2013.
40. Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Applied Health Economics and Health Policy. 2013;11(3):287-98.
41. Norman R, Viney R, Brazier JE, Burgess L, Cronin P, King MT, et al. Valuing SF-6D health states using a Discrete Choice Experiment. Medical Decision Making. 2014;34(6):773-86.
42. Viney R, Norman R, King MT, Cronin P, Street DJ, Knox S, et al. Time trade-off derived EQ-5D weights for Australia. Value in Health. 2014;14(6):928-36.

43. Bleichrodt H, Johannesson M. The validity of QALYs: an experimental test of constant proportional tradeoff and utility independence. Med Decis Making. 1997;17(1):21-32.
44. Bleichrodt H, Wakker P, Johannesson M. Characterizing QALYs by Risk Neutrality. Journal of Risk and Uncertainty. 1997;15(2):107-14.
45. Brazier JE, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics. 2002;21:271-92.
46. Mukuria C, Rowen D, Brazier JE, Young TA, Nafees B. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. Value in Health. 2015;18(6):846-55.
47. Mulhern B, Rowen D, Jacoby A, Marson T, Snape D, Hughes D, et al. The development of a QALY measure for epilepsy: NEWQOL-6D. Epilepsy & Behavior. 2012;24(1):36-43.
48. Rowen D, Brazier JE, Young T, Gaugris S, Craig BM, King MT, et al. Deriving a preference-based measure for cancer using the EORTC-QLQC30. Value in Health. 2011;14(5):721-31.
49. Rowen D, Mulhern B, Banerjee S, van Hout B, Young TA, Knapp M, et al. Estimating Preference-Based Single Index Measures for Dementia Using DEMQOL and DEMQOL-Proxy. Value in Health. 2012;15(2):346-56.
50. Stevens K. Valuation of the Child Health Utility 9D Index. Pharmacoeconomics. 2012;30(8):729-47.
51. Goodwin E, Green C. A Systematic Review of the Literature on the Development of Condition-Specific Preference-Based Measures of Health. Appl Health Econ Health Policy. 2016;14(2):161-83.
52. Revelt D, Train K. Customer-specific taste parameters and mixed logit. University of California, Berkeley; 1999 November 23, 1999.
53. Hole A. Fitting mixed logit models by using maximum simulated likelihood. The STATA Journal. 2007;7:388-401.
54. Versteegh MM, Leunis A, Uyl-de Groot CA, Stolk EA. Condition-specific preference-based measures: benefit or burden? Value Health. 2012;15(3):504-13.
55. Dobrez D, Cella D, Pickard AS, Lai JS, Nickolov A. Estimation of patient preference-based utility weights from the functional assessment of cancer therapy - general. Value Health. 2007;10(4):266-72.
56. Jonker MF, Donkers B, De Bekker-Grob EW, Stolk EA. The Effect of Level Overlap and Color Coding on Attribute Non-attendance in Discrete Choice Experiments. Value Health. 2017.
57. Mulhern B, Longworth L, Brazier J, Rowen D, Bansback N, Devlin N, et al. Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. Value Health. 2013;16(1):104-13.

Instructions for calculating EORTC QLU-C10D scores from EORTC QLQ-C30 responses

To calculate the QLU-C10D score from QLQ-C30 responses of a particular patient p, first determine their QLU-C10D level l for each dimension d, following the mapping of QLQ-C30 items to QLU-C10D levels provided in Table 1. A utility score of 1 is assigned to patients whose QLQ-C30 scores indicate they are at level 1 of all 10 dimensions of the QLU-C10D. For all other health states, the utility score is 1 minus each the utility decrement (w_{dl}) for each level down from no problems in each of the 10 QLU-C10D dimensions.

$$\text{QLU-C10D}_p = 1 - \sum_{d=1}^{10} w_{dl} \mid \text{QLU-C10D}_{dlp}$$

STATA code to calculate EORTC QLU-C10D scores from EORTC QLQ-C30 responses, based on the UK utility weights

```
*******************************************************************************
* Example code for converting EORTC QLQ-C30 data into QLU-C10D utility scores
* Written by Richard Norman
*
*
* For further details of the QLU-C10D, see the following papers:
*
* King MT, Costa DSJ, Aaronson NK, Brazier JE, Cella DF, Fayers PM, et al.
Derivation of the health state *classification system for the QLU-C10D, an
internationally-valid cancer-specific multi-attribute utility instrument *derived
from the EORTC core quality of life questionnaire, QLQ-C30. Qual Life Res. 2016.
*King MT, Viney R, Pickard AS, Rowen D, Aaronson NK, Brazier JE, et al. Australian
Utility Weights for the EORTC *QLU-C10D, a Multi-Attribute Utility Instrument
Derived from the Cancer-Specific Quality of Life Questionnaire, *EORTC QLQ-C30.
Pharmacoeconomics. 2018;36:225-38.
*
*
* The utility algorithm reported in this code is based on Model 2 results for an
* UK general population sample, as reported above in this paper
*
* This code is written for STATA users, and notes are added throughout to allow
* conversion to other software as required.
*
* Any questions / comments on the instrument or code should be sent to Richard   *
* Norman (Richard.norman@curtin.edu.au)
*
*******************************************************************************

*******************************************************************************
* Stage 1: Derive of the QLU-C10D dimension levels from the corresponding EORTC
* QLQ-C30 item responses.
*******************************************************************************

*******************************************************************************
* Assumption: For this code to work, it is assumed that the EORTC QLQ-C30 code
* is set up as thirty columns, labelled qlq1-qlq30, each of which can take one
* of four values 1-4, where 1 = "Not at all", 2 = "A little", 3 = "Quite a
* bit", and 4 = "Very much". To derive the QLU-C10D, we only need 13 of these
* items, as described in Table 3 of King et al.(2016) and Table 1 of King et al,
* Pharmacoeconomics DOI 10.1007/s40273-017-0582-5.
*
```

```
* Seven of QLU-C10D items are single items from the EORTC QLQ-C30, and three
* (PF, SF, BO) are composite, combined as described below.
*******************************************************************************

* Generate the QLU-C10D  level for Physical Functioning from QLQ-C30 item 2
* (long walk) and item 3 (short walk)

gen pf=.
replace pf=1 if qlq2==1
replace pf=2 if qlq2>1
replace pf=3 if qlq3>1
replace pf=4 if qlq3>2

*PF note: Level 1 is no problems in a long walk (and logically also in a short
* walk). Level 2 is any problem taking a long walk but no problem taking a short *
walk. Level 3 is a little problem with
 a short walk (and a logically a little * or more problem with a long walk), and
level 4 is higher level problems (quite * a bit or very much) with both short and
long walk.

* Generate the QLU-C10D level for Role Functioning from QLQ-C30 item 6
* (work and daily activities)
gen rf=.
replace rf=1 if qlq6==1
replace rf=2 if qlq6==2
replace rf=3 if qlq6==3
replace rf=4 if qlq6==4

* Generate the QLU-C10D  level for Social Functioning from QLQ-C30 item 26
* (family life) and item 27 (social activities)
gen sf=.
replace sf=1 if qlq26==1 & qlq27==1
replace sf=2 if qlq26==2 | qlq27==2
replace sf=3 if qlq26==3 | qlq27==3
replace sf=4 if qlq26==4 | qlq27==4

* SF note: SF in the QLU-C10D effectively uses the maximum value of either
* qlq26 or qlq27. So if qlq26 is 4 and qlq27 is 1, the utility dimension is at
* level 4.  The vertical bar in the code signifies OR.

* Generate the QLU-C10D level for Emotional Functioning from QLQ-C30 item 24
* (depressed)
gen ef=.
replace ef=1 if qlq24==1
replace ef=2 if qlq24==2
replace ef=3 if qlq24==3
replace ef=4 if qlq24==4

* Generate the QLU-C10D level for Pain from QLQ-C30 item 9 (pain)
gen pa=.
replace pa=1 if qlq9==1
replace pa=2 if qlq9==2
replace pa=3 if qlq9==3
replace pa=4 if qlq9==4

* Generate the QLU-C10D level for Fatigue from QLQ-C30 item 18 (tired)
gen fa=.
replace fa=1 if qlq18==1
replace fa=2 if qlq18==2
replace fa=3 if qlq18==3
replace fa=4 if qlq18==4

* Generate the QLU-C10D level for Sleep from QLQ-C30 item 11 (trouble sleeping)
gen sl=.
replace sl=1 if qlq11==1
```

```
replace sl=2 if qlq11==2
replace sl=3 if qlq11==3
replace sl=4 if qlq11==4

* Generate the QLU-C10D level for Appetite from QLQ-C30 item 13 (lack appetite)
gen ap=.
replace ap=1 if qlq13==1
replace ap=2 if qlq13==2
replace ap=3 if qlq13==3
replace ap=4 if qlq13==4

* Generate the QLU-C10D level for Nausea from QLQ-C30 item 14 (nauseated)
gen na=.
replace na=1 if qlq14==1
replace na=2 if qlq14==2
replace na=3 if qlq14==3
replace na=4 if qlq14==4

* Generate the QLU-C10D level for Bowel Problems from QLQ-C30 item 16 (constipated)
and item 17 (diarrhea)
gen bo=.
replace bo=1 if qlq16==1 & qlq17==1
replace bo=2 if qlq16==2 | qlq17==2
replace bo=3 if qlq16==3 | qlq17==3
replace bo=4 if qlq16==4 | qlq17==4

* BO note: BO in the QLU-C10D effectively uses the maximum value of either
* qlq16 or qlq17. So if qlq16 is 4 and qlq17 is 1, the utility dimension is at
* level 4.


*******************************************************************************
* Stage 2: Generate utility decrements and sum to estimate utility scores      *
* The utility decrements (e.g. pfdec is the utility decrement for Physical     *
* Functioning dimension) are derived from Table 4 of the current paper under   *
* review, and are derived from a conditional logit constrained to be monotonic *
* within each dimension.                                                       *
*******************************************************************************

gen pfdec=0
replace pfdec=-0.062 if pf==2
replace pfdec=-0.155 if pf==3
replace pfdec=-0.254 if pf==4

gen rfdec=0
replace rfdec=-0.021 if rf==2
replace rfdec=-0.076 if rf==3
replace rfdec=-0.109 if rf==4

gen sfdec=0
replace sfdec=-0.020 if sf==2
replace sfdec=-0.071 if sf==3
replace sfdec=-0.103 if sf==4

gen efdec=0
replace efdec=-0.008 if ef==2
replace efdec=-0.050 if ef==3
replace efdec=-0.130 if ef==4

gen padec=0
replace padec=-0.030 if pa==2
replace padec=-0.078 if pa==3
replace padec=-0.151 if pa==4

gen fadec=0
```

```
replace fadec=-0.036 if fa==2
replace fadec=-0.048 if fa==3
replace fadec=-0.058 if fa==4

gen sldec=0
replace sldec=-0.049 if sl==2
replace sldec=-0.056 if sl==3
replace sldec=-0.070 if sl==4

gen apdec=0
replace apdec=-0.029 if ap==2
replace apdec=-0.029 if ap==3
replace apdec=-0.029 if ap==4

gen nadec=0
replace nadec=-0.048 if na==2
replace nadec=-0.059 if na==3
replace nadec=-0.095 if na==4

gen bodec=0
replace bodec=-0.053 if bo==2
replace bodec=-0.061 if bo==3
replace bodec=-0.084 if bo==4

* Generate the QLU-C10D utility score
gen qluc10d = 1+pfdec+rfdec+sfdec+efdec+padec+fadec+sldec+apdec+nadec+bodec

replace qluc10d=. if (qlq2==. | qlq3==. | qlq6==. | qlq26==. | qlq27==.
| qlq24==. | qlq9==. | qlq18==. | qlq11==. | qlq13==. | qlq14==. |
qlq16==. | qlq17==.)

*******************************************************************************
* The new variable qluc10d is a utility score where full health (i.e. level 1
* in each of the utility levels) is scored at 1, and states worse than dead are
*scored <0.
*These data can now be used to
* construct quality-adjusted life years (QALYs) for cost-utility analysis.
*******************************************************************************
```

SPSS code to calculate EORTC QLU-C10D scores from EORTC QLQ-C30 responses, based on the Australian utility weights

```
*******************************************************************************
* Example code for converting EORTC QLQ-C30 data into QLU-C10D utility scores  *
* Adapted for SPSS by Dan Costa                *
*
*
* For further details of the QLU-C10D, see the following papers:             *
*
* King MT, Costa DSJ, Aaronson NK, Brazier JE, Cella DF, Fayers PM, et al.
Derivation of the health state *classification system for the QLU-C10D, an
internationally-valid cancer-specific multi-attribute utility instrument *derived
from the EORTC core quality of life questionnaire, QLQ-C30. Qual Life Res. 2016.
*King MT, Viney R, Pickard AS, Rowen D, Aaronson NK, Brazier JE, et al. Australian
Utility Weights for the EORTC *QLU-C10D, a Multi-Attribute Utility Instrument
Derived from the Cancer-Specific Quality of Life Questionnaire, *EORTC QLQ-C30.
Pharmacoeconomics. 2018;36:225-38.
*
* The utility algorithm reported in this code is based on Model 2 results for an
* UK general population sample, as reported above in this paper
*
* This code is written for SPSS users, and notes are added throughout to allow*
* conversion to other software as required.
*
```

```
* Any questions / comments on the instrument or code should be sent to Daniel Costa
daniel.costa@sydney.edu.au         *
*
*************************************************************************



*************************************************************************
* Stage 1: Derive of the QLU-C10D dimension levels from the corresponding EORTC*
* QLQ-C30 item responses.                    *
*************************************************************************

*************************************************************************
* Assumption: For this code to work, it is assumed that the EORTC QLQ-C30 code *
* is set up as thirty columns, labelled qlq1-qlq30, each of which can take one *
* of four values 1-4, where 1 = "Not at all", 2 = "A little", 3 = ?Quite a     *
* bit?, and 4 = ?Very much". To derive the QLU-C10D, we only need 13 of these  *
* items, as described in Table 3 of King et al.                    *
* Seven of QLU_C10D items are single items from the EORTC QLQ-C30, and three   *
* (PF, SF, BO) are composite, combined as described below.                    *
*************************************************************************

* Generate the QLU-C10D  level for Physical Functioning from QLQ-C30 item 2
* (long walk) and item 3 (short walk).
compute pf=$sysmis.
if qlq2=1 pf=1.
if qlq2>1 pf=2.
if qlq3>1 pf=3.
if qlq3>2 pf=4.
exe.


*PF note: Level 1 is no problems in a long walk (and logically also in a short
* walk). Level 2 is any problem taking a long walk but no problem taking a short
* walk. Level 3 is a little problem with a short walk (and a logically a little
* or more problem with a long walk), and level 4 is higher level problems (quite
* a bit or very much) with both short and long walk.

* Generate the QLU-C10D level for Role Functioning from QLQ-C30 item 6
* (work and daily activities).
compute rf=$sysmis.
if qlq6=1 rf=1.
if qlq6=2 rf=2.
if qlq6=3 rf=3.
if qlq6=4 rf=4.
exe.


* Generate the QLU-C10D level for Social Functioning from QLQ-C30 item 26
* (family life) and item 27 (social activities).
compute sf=$sysmis.
if qlq26=1 & qlq27=1 sf=1.
if qlq26=2 | qlq27=2 sf=2.
if qlq26=3 | qlq27=3 sf=3.
if qlq26=4 | qlq27=4 sf=4.
exe.

* SF note: SF in the QLU-C10D effectively uses the maximum value of either
* qlq26 or qlq27. So if qlq26 is 4 and qlq27 is 1, the utility dimension is at
* level 4.  The vertical bar in the code signifies OR.

* Generate the QLU-C10D level for Emotional Functioning from QLQ-C30 item 24
* (depressed).
compute ef=$sysmis.
if qlq24=1 ef=1.
```

```
if qlq24=2 ef=2.
if qlq24=3 ef=3.
if qlq24=4 ef=4.
exe.

* Generate the QLU-C10D level for Pain from QLQ-C30 item 9 (pain).
compute pa=$sysmis.
if qlq9=1 pa=1.
if qlq9=2 pa=2.
if qlq9=3 pa=3.
if qlq9=4 pa=4.
exe.

* Generate the QLU-C10D level for Fatigue from QLQ-C30 item 18 (tired).
compute fa=$sysmis.
if qlq18=1 fa=1.
if qlq18=2 fa=2.
if qlq18=3 fa=3.
if qlq18=4 fa=4.
exe.

* Generate the QLU-C10D level for Sleep from QLQ-C30 item 11 (trouble sleeping).
compute sl=$sysmis.
if qlq11=1 sl=1.
if qlq11=2 sl=2.
if qlq11=3 sl=3.
if qlq11=4 sl=4.
exe.

* Generate the QLU-C10D level for Appetite from QLQ-C30 item 13 (lack appetite).
compute ap=$sysmis.
if qlq13=1 ap=1.
if qlq13=2 ap=2.
if qlq13=3 ap=3.
if qlq13=4 ap=4.
exe.

* Generate the QLU-C10D level for Nausea from QLQ-C30 item 14 (nauseated).
compute na=$sysmis.
if qlq14=1 na=1.
if qlq14=2 na=2.
if qlq14=3 na=3.
if qlq14=4 na=4.
exe.

* Generate the QLU-C10D level for Bowel Problems from QLQ-C30 item 16 (constipated)
and item 17 (diarrhea).
compute bo=$sysmis.
if qlq16=1 & qlq17=1 bo=1.
if qlq16=2 | qlq17=2 bo=2.
if qlq16=3 | qlq17=3 bo=3.
if qlq16=4 | qlq17=4 bo=4.
exe.

* BO note: BO in the QLU-C10D effectively uses the maximum value of either
* qlq16 or qlq17. So if qlq16 is 4 and qlq17 is 1, the utility dimension is at
* level 4. The vertical bar in the code signifies OR.


*************************************************************************
* Stage 2: Generate utility decrements and sum to estimate utility scores      *
* The utility decrements (e.g. pfdec is the utility decrement for Physical     *
* Functioning dimension) are derived from Table 4 of the current paper under   *
* review, and are derived from a conditional logit constrained to be monotonic *
* within each dimension.                                                       *
```

```
*******************************************************************************

compute pfdec=0.
if pf=2 pfdec=-0.062.
if pf=3 pfdec=-0.155.
if pf=4 pfdec=-0.254.
exe.

compute rfdec=0.
if rf=2 rfdec=-0.021.
if rf=3 rfdec=-0.076.
if rf=4 rfdec=-0.109.
exe.

compute sfdec=0.
if sf=2 sfdec=-0.020.
if sf=3 sfdec=-0.071.
if sf=4 sfdec=-0.103.
exe.

compute efdec=0.
if ef=2 efdec=-0.008.
if ef=3 efdec=-0.050.
if ef=4 efdec=-0.130.
exe.

compute padec=0.
if pa=2 padec=-0.030.
if pa=3 padec=-0.078.
if pa=4 padec=-0.151.
exe.

compute fadec=0.
if fa=2 fadec=-0.036.
if fa=3 fadec=-0.048.
if fa=4 fadec=-0.058.
exe.

compute sldec=0.
if sl=2 sldec=-0.049.
if sl=3 sldec=-0.056.
if sl=4 sldec=-0.070.
exe.

compute apdec=0.
if ap=2 apdec=-0.029.
if ap=3 apdec=-0.029.
if ap=4 apdec=-0.029.
exe.

compute nadec=0.
if na=2 nadec=-0.048.
if na=3 nadec=-0.059.
if na=4 nadec=-0.095.
exe.

compute bodec=0.
if bo=2 bodec=-0.053.
if bo=3 bodec=-0.061.
if bo=4 bodec=-0.084.
exe.

* Generate the QLU-C10D utility score
compute qluc10d = 1+pfdec+rfdec+sfdec+efdec+padec+fadec+sldec+apdec+nadec+bodec.
```

```
if (sysmis(qlq2) or sysmis(qlq3) or sysmis(qlq6) or sysmis(qlq26) or sysmis(qlq27)
or sysmis(qlq24) or sysmis(qlq9) or sysmis(qlq18) or sysmis(qlq11) or sysmis(qlq13)
or sysmis(qlq14) or sysmis(qlq16) or sysmis(qlq17)) qluc10d=$sysmis.
exe.

* Show all decrement values to 3 decimal places.
formats pfdec rfdec sfdec efdec padec fadec sldec apdec nadec bodec qluc10d (F6.3).

*****************************************************************************
* The new variable qluc10d is a utility score where full health (i.e. level 1
* in each of the utility levels) is scored at 1, and states worse than dead are
*scored <0.
*These data can now be used to
* construct quality-adjusted life years (QALYs) for cost-utility analysis.
*****************************************************************************
```