


ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Preference-Based Assessments

The Functional Assessment of Cancer Therapy Eight Dimension (FACT-8D), a Multi-Attribute Utility Instrument Derived From the Cancer-Specific FACT-General (FACT-G) Quality of Life Questionnaire: Development and Australian Value Set

Madeleine T. King, PhD, Richard Norman, PhD, Rebecca Mercieca-Bebber, PhD, Daniel S.J. Costa, PhD, Helen McTaggart-Cowan, PhD, Stuart Peacock, PhD, Monika Janda, PhD, Fabiola Müller, PhD, Rosalie Viney, PhD, Alan Simon Pickard, PhD, David Cella, PhD, on behalf of the Multi-Attribute Utility in Cancer Consortium

ABSTRACT

Objectives: To develop a cancer-specific multi-attribute utility instrument derived from the Functional Assessment of Cancer Therapy - General (FACT-G) health-related quality of life (HRQL) questionnaire.

Methods: We derived a descriptive system based on a subset of the 27-item FACT-G. Item selection was informed by psychometric analyses of existing FACT-G data (n = 6912) and by patient input (n = 82). We then conducted an online valuation survey, with participants recruited via an Australian general population online panel. A discrete choice experiment (DCE) was used, with attributes being the HRQL dimensions of the descriptive system and survival duration, and 16 choice-pairs per participant. Utility decrements were estimated with conditional logit and mixed logit modeling.

Results: Eight HRQL dimensions were included in the descriptive system: pain, fatigue, nausea, sleep, work, social support, sadness, and future health worry; each with 5 levels. Of 1737 panel members who accessed the valuation survey, 1644 (95%) completed 1 or more DCE choice-pairs and were included in analyses. Utility decrements were generally monotonic; within each dimension, poorer HRQL levels generally had larger utility decrements. The largest utility decrements were for the highest levels of pain (-0.40) and nausea (-0.28). The worst health state had a utility of -0.54, considerably worse than dead.

Conclusions: A descriptive system and preference-based scoring approach were developed for the FACT-8D, a new cancer-specific multi-attribute utility instrument derived from the FACT-G. The Australian value set is the first of a series of country-specific value sets planned that can facilitate cost-utility analyses based on items from the FACT-G and related FACIT questionnaires containing FACT-G items.

Keywords: condition-specific, health-related quality of life, multi-attribute utility, preference-based, QALY, quality of life, quality-adjusted life-year, utility, value set.

VALUE HEALTH. 2021; 24(6):862–873

Introduction

Multi-attribute utility instruments (MAUIs) are used in cost-utility analysis (CUA) through estimation of the quality-adjusted life-year (QALY). MAUIs comprise a descriptive system that covers the relevant dimensions of health-related quality of life (HRQL), with a scoring algorithm (derived using a preference-based approach) that provides a value set (or utilities) covering each health state described by the MAUI on a scale that is anchored on 0 representing "dead" and 1 representing "full health," suitable for estimating QALYs.

MAUIs can be derived from existing HRQL profile measures,[1] allowing utilities to be generated prospectively and retrospectively from data collected with the source HRQL instrument. This approach has been used for the generic SF-36 and SF-12 via the SF-6D utility algorithm,[2,3] and for several condition-specific measures.[4] The Multi-Attribute Utility in Cancer Consortium (MAUCa) has applied this approach to the European Organisation for Research and Treatment of Cancer (EORTC) core HRQL questionnaire, QLQ-C30, producing the EORTC QLU-C10D.[5–7] This article describes the development of a MAUI from the widely used Functional Assessment of Cancer Therapy – General (FACT-G)

questionnaire,[8] including the development of a descriptive system, valuation methodology, and the Australian value set.

Methods

This research was conceived, designed, and conducted by the MAUCa Consortium. Human Research Ethics Committee approval was provided by the University of Sydney (No. 13207). The methods are similar to those developed previously for the EORTC QLU-C10D[5-7] and are described briefly in this article.

Descriptive System

The FACT-G's 27 items are organized into 4 well-being dimensions (see Appendix Table A in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007). Our aim was to select a suitable subset for the new MAUI descriptive system, with at least 1 item from each FACT-G dimension. Our analyses focused on individual FACT-G items to determine which best represented a particular FACT-G dimension, and which could be excluded due to poor psychometric performance or redundancy. We applied 9 prespecified criteria[5]: (1) item fit, (2) disordered item response thresholds, (3) spread of item thresholds across the latent variable, (4) differential item function (DIF) by sex and cancer site, (5) local dependence, (6) ceiling/floor effects, (7) item sensitivity to early versus late stage cancer, (8) responsiveness to change due to treatment, (9) patient input about the relative importance of items. Criteria 1-8 were informed by secondary analysis of existing FACT-G data sets. For criteria 1-5, we first assessed the FACT-G dimension structure with confirmatory factor analysis, then applied Rasch analysis within dimensions. Items exhibiting significant floor or ceiling effects were considered poor candidates (criterion 6). Criteria 7-8 were assessed by calculating effect sizes: mean difference/standard deviation (SD) and mean change from baseline to on-treatment/SD(change), respectively. For criterion 9, we surveyed a sample of cancer patient recruited from 4 Australian hospitals (1 metropolitan, 3 rural). See Appendix Report A in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007 for the patient importance survey methods and questionnaire content. The resulting subset of FACT-G items, representing a range of HRQL dimensions, established the descriptive system, the framework for eliciting general population values from which to estimate a value set.

Valuation Method

We used a discrete choice experiment (DCE) approach similar to those used in previous value set estimation projects,[6,9-14] noting its proven feasibility for multi-attribute utility measures with a relatively large number of dimensions.[7] The valuation task involved choosing between 2 health states (a "choice pair"), with each health state described by levels of HRQL dimensions in the descriptive system and a specified survival duration (life-years). This task captured the trade-offs people make between HRQL and survival, as required for CUA.

DCE Design

The DCE contained 9 attributes: the 8 HRQL dimensions of the FACT-8D and duration. We created an experimental design comprising 100 choice sets in which 5 of the attributes differed between the 2 health state options in each choice set, while 4 attributed were the same across the 2 options. We imposed this overlap because we decided that the cognitive challenge of considering 9 dimensions simultaneously was too arduous, hence likely to cause respondents to adopt simple decision heuristics. Forcing overlap necessarily worsens statistical efficiency of the design (relative to a design that does not), but we decided that it was necessary to balance statistical and respondent efficiency and that that level of overlap was a pragmatic trade-off between the two. We focused on C-efficiency rather than the more widely used D-efficiency. Under C-efficiency, we are interested in the accuracy of the ratios of coefficients rather than in the coefficients themselves. Given the analysis described below, estimates ratios of coefficients to the coefficient on time, it was appropriate to follow this approach. To achieve these design requirements, we generated a large number of random choice pairs and kept only those in which exactly 5 dimensions differed. We continued this process until we had 10 000 choice pairs that met this criterion. We then used *Ngene* to select the 100 choice pairs to be used in the DCE, using a modified Fedorov algorithm to optimize C-efficiency, with duration used as the denominator. Small non-zero priors were used to indicate that levels within each dimension were logically ordered; Appendix Table B in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007 contains the final design. The final design was not strictly orthogonal, largely because of the forced overlap. Each respondent was allocated a random selection of 16 of the 100 choice sets. Which option was seen as situation A or B was randomized within each choice set to mitigate ordering bias. The dimensions were always presented in the same order, because we found previously that dimension order did not systematically bias utility weights.[15]

Valuation Survey and Sample

The valuation survey included several components, in this order: welcome/disclosure; sex and age (for screening and quota sampling); self-reported health (SF-36 general health question,[16] FACT-G general population version [FACT-GP][17]); the DCE; respondent perception of the difficulty and clarity of the DCE choice task and strategies used; sociodemographic variables and self-reported mental health (Kessler-10).[18]

A company specializing in choice experiments, SurveyEngine,[19] managed sample recruitment via an Australian online panel (Pureprofile), survey administration, and data collection. The target population was the Australian adult general population (≥18 years). Quota sampling by age and sex was used to achieve population representativeness according to census data.[20] We determined that a target sample size of 1600 would provide adequate precision for our parameter estimates, based on previous experience that reliable algorithms were achieved from similar experimental designs and modeling approaches with data from 1000 respondents.[13,14,21] With 1600 participants each completing 16 choice sets, there would be an average of 256 observations per choice set, generally considered adequate for robust DCE analysis.[22]

Data Analysis

We used descriptive statistics to summarize sample characteristics, perceived difficulty, and clarity of the DCE task and choice strategies. We assessed the sample's representativeness relative to the Australian population with chi-square tests against best available normative data.

Utility Estimation

The DCE data were analyzed in STATA.[23] using a functional form consistent with standard QALY model restrictions.[7,13,21,24,25] The QALY model requires that all health states have zero utility at dead[26,27]; a functional form that satisfied this requirement included the HRQL levels interacted with the TIME variable (representing survival duration in equations 1 and 2). As the QALY

model assumes constant proportional time trade-off, the relationship between utility and TIME was considered linear. In this functional form, the impact of moving away from level 1 (no problems) in each HRQL dimension was characterized through the 2-factor interaction term with TIME. Thus in the resultant utility algorithm, the impact of each level worse than "no problems" represents a decrement away from full health (value of 1).

We analyzed the DCE data in 2 ways. The primary analysis used a conditional logit model (equation 1). Here, the utility of option j in choice set s for survey respondent i was assumed to be:

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

$i = 1, \ldots, I$ *respondents*; $j =$ situations A, B; $s = 1, \ldots,$ 100 choice sets

where α was the utility associated with a life-year in full health, X'_{isj} was a vector of dummy variables representing the levels of the health state presented in option j, and β was the corresponding vector of utility weights associated with each level in each dimension within X'_{isj}, for each life-year. We assumed a Gumbel distribution for the error term ε_{isj}. Given repeated choice sets per respondent, a clustered sandwich estimator adjusted the standard errors to allow for intra-individual correlation, implemented by STATA's vce (cluster) option. To estimate utility decrements for each move away from level 1 (no problems) in each HRQL dimension, we divided each β term by α. To estimate confidence intervals around these ratios, we used STATA's wtp command, using the delta method.[28]

Model 1 included every move away from level 1 in each dimension. Thus, X'_{isj} contained 32 terms (8 dimensions $\times$ [5-1] levels per dimension). If non-monotonicity was observed among levels within a dimension, the non-monotonic levels were combined in model 2, as done previously.[3,6,29-33]

The secondary analysis (model 3, equation 2) used a mixed logit,[34] which assumed that coefficients were drawn from a normal distribution, allowing for preference heterogeneity among individuals.

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

Thus, α and the vector of βs now represent population mean preferences, while γ_i and η_i are individual deviations around those mean preferences. We assumed these deviations had multivariate normal distributions $(0, \sum)$. We used the mixlogit STATA command[28] to estimate α, the vector of β s, and the standard deviations of γ and the vector of ηs, with 1 adjustment. The standard command limits the number of random parameters to 20; we used scrambled Halton draws to circumvention this restriction, allowing all 33 coefficients to be drawn from a distribution.[35]

We informally compared models in terms of model fit and parsimony using the log-likelihood, Akaike Information Criteria, and Bayesian Information Criteria, noting our models were not nested.

Results

FACT-8D Descriptive System

We obtained 24 data sets with FACT-G responses and suitable covariate data from 6912 patients in total (see Appendix Table C in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007). All commonly occurring cancers were represented, with 61% of patients having localized/regional disease and 39% recurrent/metastatic (Appendix Table D in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007). Common treatments were represented; 52% related to monotherapy (22% chemotherapy, 3% radiotherapy, 27% surgery), 40% multiple therapies, and 7% no therapies (see Appendix Table E in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007). These data were pooled for Rasch analyses (criteria 1-5, Appendix Report A), item response frequencies (criterion 6, Appendix Figure A in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007), and sensitivity to disease stage (criterion 7, Table 1). Ten of the 24 data sets contained both baseline and on-treatment observations for >30 patients (total n = 881); these provided estimates of responsiveness by study (see Appendix Table F in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007). Characteristics of the 92 patients surveyed for criterion 9 are presented in Appendix Table G in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007. Table 1 summarizes results of all analyses for all criteria.

In the physical well-being (PWB) dimension, item #6_feel-ill was excluded due to local dependency with nausea. Fatigue (#1_lack energy) performed well on sensitivity and reasonably on responsiveness, had good spread, and has been identified a key cancer symptom by patients.[36,37] Item #4_pain also performed reasonably well on responsiveness, sensitivity, and spread and has also been identified by patients as a key cancer symptom.[36,37] A third patient-identified key symptom was nausea,[36,37] which may be caused by cancer and treatments for cancer such as chemotherapy and radiotherapy. Nausea performed well on responsiveness and was deemed particularly important for this cancer-specific MAUI as it is rarely included in generic MAUIs; #2_nausea was therefore included despite a ceiling effect and some threshold disorder.

In the social well-being (SWB) dimension, all items had disordered thresholds, so this criterion was discounted. Although #13_partner was rated of high importance by patients in our substudy, it lacked sensitivity and exhibited DIF by sex, so it was excluded. #8_closefriends lacked sensitivity and responsiveness, exhibited DIF by sex, and was least important to patients, so it was excluded. #12commfam exhibited DIF by stage, lacked responsiveness, and was relatively unimportant to patients, so it was excluded. #9_suppfam had reasonable spread and was rated of high importance by patients. It cross-loaded with #10suppfriends, and to allow for patients without family, items #9 and #10 were included as a combined pair to represent SWB.

In the emotional well-being (EWB) dimension, all but 1 item had disordered thresholds, so this criterion was discounted. #16_hope was excluded due to ceiling effects and poor responsiveness; #18_worry-dying lacked importance and responsiveness. #19_worry-worse was rated highest in importance and was the most sensitive, so it was included. #14_sad had the widest spread, providing good coverage of the latent variable. Further, because sadness may be a normal response to a cancer diagnosis[38] and pervasive sadness is an expression of depression,[39] we decided that it provided important coverage of cancer-specific mental health, and had included depression in the EORTC QLU-C10D.[5]

In the functional well-being dimension (FWB), there was local dependency for 5/7 items, and generally comparable sensitivity and responsiveness. Three global items (#22_enjoy, #25_fun, #26_QOL) were excluded because these items are not sufficiently specific for inclusion in a MAUI. Two items (#20_able-to-work, #21_work-fulfilling) were locally dependent; because #20 was rated most important by patients, it was included. #24_sleep was included, despite its misfit to Rasch model, because it is an important symptom in cancer and was included in QLQ-C10D.[5]

Table 1. Summary of results to determine which items to include in the utility descriptive system. Items in italics were selected for the FACT-8D descriptive system.

FACT-G Domain		CFA[†]	Nine criteria used to assess items for exclusion/inclusion in the utility descriptive system*										
Item #	Item stem code		Rasch-based criteria (1-5)					6. Ceiling/ Floor**	7. Sens-itivity	8. Responsive-ness		9. Importance[‖‖]	
		Load	Fit[‡]	DT[§]	Spread[‖]	DIF[¶]	LD[#]		Effect size[††]	Large[‡‡]	Mod[§§]	Most	Top 3
Physical well-being													
1	*Lack energy*	-	-	-	−0.88, 2.65	-	-	-	*0.49*	*1*	*3*	-	-
2	*Nausea*	-	-	*x*	−1.12, 0.26	-	*x*	-	*0.31*	*2*	*3*	-	-
3	Famneeds	-	-	-	−0.69, 0.58	-	-	-	0.36	0	3	1	3
4	*Pain*	-	-	-	−0.71, 0.69	-	-	-	*0.28*	*1*	*3*	-	-
5	Side-effects	-	-	-	−0.21, 0.63	-	-	-	0.42	2	2	2	2
6	Feel ill	-	-	-	−1.06, 0.63	-	x	-	0.35	0	3	4	1
7	Bed	-	-	-	−1.47, 0.47	-	-	-	0.30	0	3	3	4
Social/Family well-being													
8	Closefriends	-	-	x	−0.53, 1.90	x	-	-	0.00	0	2	6	6
9	*Suppfam*	*Cross*	-	*x*	−1.25, 0.89	-	*x*	-	*-0.14*	*0*	*2*	*2*	*1*
10	*Suppfriend*	-	-	*x*	−1.88, 1.62	*x*	*x*	-	*-0.16*	*0*	*4*	*4*	*4*
11	Acceptfam	-	-	x	−2.65, 1.19	-	-	-	0.16	0	2	3	3
12	Commfam	-	-	x	−0.11, 0.90	x	-	-	0.20	0	2	5	5
13	Partner	-	-	x	−1.00, 0.63	x	-	-	-0.04	0	3	1	2
Emotional well-being													
14	*Sad*	-	-	*x*	−2.00, 1.88	-	-	-	*0.28*	*0*	*1*	*3*	*3*
15	Coping	Cross	-	x	−0.81, 1.13	-	-	-	0.32	0	2	2	1
16	Hope	-	-	x	−1.82, -0.40	-	-	Ceiling	0.44	0	1	5	6
17	Nervous	-	-	-	−1.06, 1.02	-	-	-	0.28	0	0	6	4
18	Worry_dying	-	-	x	−0.81, 0.75	-	x	-	0.41	0	0	4	5
19	*Worry_worse*	-	-	*x*	−0.46, 1.84	-	*x*	-	*0.46*	*0*	*2*	*1*	*2*
Functional well-being													
20	*Able to work*	-	-	-	−0.46, 1.09	-	*x*	-	*0.36*	*0*	*3*	*1*	*1*
21	Work-fulfilling	-	-	-	−0.44, 1.38	-	x	-	0.32	0	2	6	7
22	Enjoy	-	-	-	−1.28, 1.01	-	x	-	0.33	0	3	2	2
23	Accept	Cross	-	x	−2.59, 0.44	-	-	-	0.26	0	0	3	6
24	*Sleep*	-	*x*	-	−0.53, 1.21	-	-	-	*0.17*	*0*	*2*	*4*	*5*
25	Fun	-	-	x	−1.25, 1.44	-	x	-	0.32	1	2	5	4
26	QOL	-	-	x	−1.44, 1.38	x	x	-	0.41	0	2	6	3

*Criteria used to determine which items to include in the utility descriptive system: (1) fit of items to the Rasch model, by FACT-G domain; (2) disordered response thresholds (DT); (3) spread of item thresholds across the latent variable; (4) differential item function (DIF); (5) local dependence (LD); (6) floor and ceiling effects; (7) sensitivity to differences between early and late stage cancer; (8) responsiveness to change due to treatment; (9) patient opinion about relative importance of items within domains. A dash (-) indicates no problem with this item for this criterion.

[†]Confirmatory factor analysis (CFA).

[‡]Fit: A cross (x) indicates poor fit to the Rasch model (analyzed by item groupings confirmed in CFA).

[§]DT: A cross (x) indicates 1 or more disordered thresholds (DT) for this item.

[‖]Spread: We inspected the item maps and present each item's lowest and highest response threshold as summary statistics.

[¶]DIF: A cross (x) indicates the presence of differential item functioning (DIF) between early versus late stage disease.

[#]LD: A cross (x) indicates local dependence (LD) between 2 items.

**"Ceiling" indicates a ceiling effect.

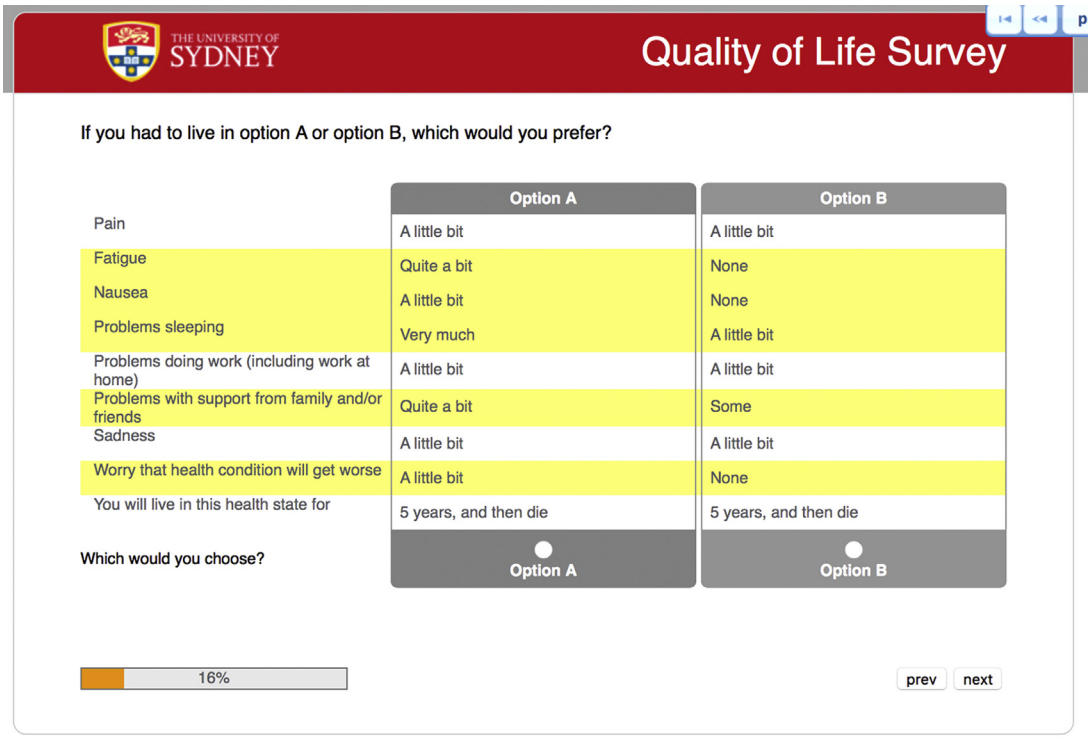
[††]Effect size for sensitivity was calculated as the mean of late-stage patients minus the mean of early-stage patients divided by the pooled standard deviation for these 2 groups; a positive value indicates better outcomes in early-stage patients.

[‡‡]"Large" refers to an effect size greater than 0.5 or less than -0.5; the number in this column indicates the number of data sets with and effect size for this item satisfying the above criteria.

[§§]"Mod" (moderate) refers to an effect size between 0.2 and 0.5 or -0.2 and -0.5; the number in this column indicates the number of data sets with and effect size for this item satisfying the above criteria.

[‖‖]The numbers in these columns indicate the rank of the item in terms of frequency it was chosen as most important or in the top 3 within its domain. The 3 cancer symptoms (#1 fatigue, #2 nausea, #4 pain) were not included in the Patient Importance Substudy.

Figure 1. An example choice set from the discrete choice experiment valuation task.



Valuation Survey

We initially used the original wording of FACT-G items in the DCE health states. We piloted this in 209 respondents; 77% found it clear/very clear, but 45% found it difficult/very difficult to choose between pairs of health states (see Appendix Table H in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007). The utility estimates from this pilot revealed several instances (more than in previous studies using this methods) in which the utility decrements between levels were non-monotonic, and in some cases had positive values (see Appendix Figure B in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007). The mix of positively and negatively framed items was the only feature that distinguished the FACT-8D DCE from similar previously successful DCEs.[7,13,14,24] We therefore revised the wording of the DCE health states by reframing the 3 positively worded dimensions (work, sleep, support) with negative framing (Fig. 1). We then commenced the main valuation study. We reviewed initial data from the next 200 respondents, and as the patterns of utility decrements more closely conformed with our previous experience, we continued recruitment.

Sample Characteristics and Representativeness

For the main valuation study, 2643 panel members initially opted in, 2330 of whom progressed to complete the age and sex questions, of whom 1737 were within age and sex quota, and 1644/1737 (95%) completed at least 1 DCE choice set and were included in the valuation analysis (Fig. 2). The sample differed statistically from the general population in all measured characteristics except age and sex, but differences were generally small (≤5% for every level), with 2 exceptions; the sample contained proportionally more educated people and people with mental health problems (Table 2).[40,41]

Respondents' Perceptions of the DCE Valuation task and time to complete

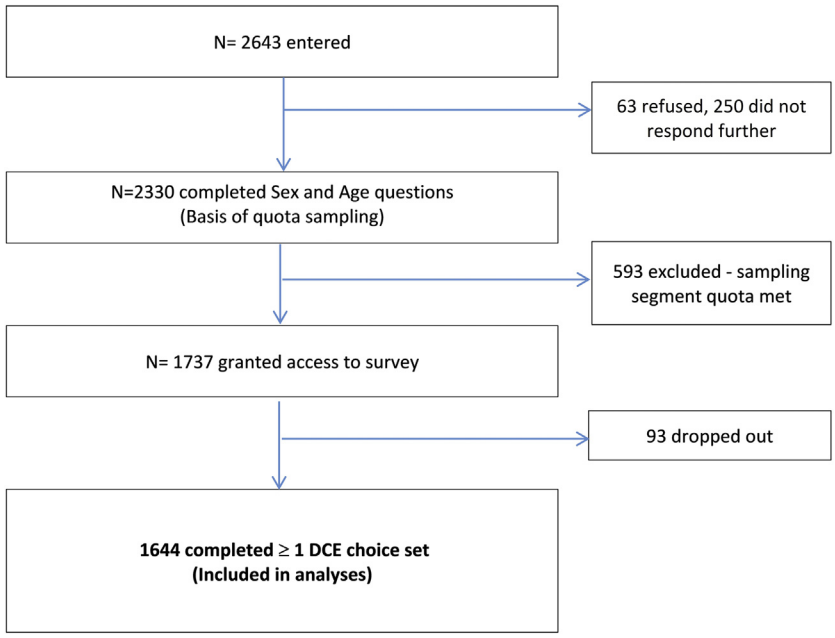
The 1562 participants who answered these questions responded similarly to the 209 in the pilot on the clarity and difficulty questions (see Appendix Table H, Figs. C, D in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007), so the modification to the DCE presentation did not materially influence these perceptions. However, there was a statistically significant difference in choice strategy ($P < .01$, see Table H, Fig. E in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007), with a greater proportion of those in the pilot focusing on just a few aspects of the health states, while those in the main survey focused on the aspects that were highlighted in yellow.

The mean time to complete the DCE valuation task was 6 minutes and 45 seconds (6'45"), and the median was 4'30". The first choice-set took longest, with times sequentially reducing as participants became familiar with the task (see Fig. F in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007).

Utility Estimates

Conditional logit results (Table 3) showed respondents valued additional years of life, and movements away from "no problems" in HRQL dimensions were generally associated with negative valuations. In model 1, level 2 of fatigue and sadness and levels 2 and 3 of sleep were positive but very small, and not statistically significantly different from zero. Utility decrements were monotonic for *pain* and *nausea*, but other dimensions had some small non-monotonicities. These issues were addressed in model 2, by constraining the coefficients for levels 4 and 5 of *fatigue*, levels 4 and 5 of *sleep*, levels 2 and 3 of *problems doing work (including work at home)*, levels 2 and 3 of *problems with support from family and friends*, and levels 2 and 3 of *worry my health will get worse*, to

Figure 2. Respondent flow and sample sizes for each component of the valuation survey.



have the same coefficient. The effect of this restriction on log-likelihood was small (24 points), with Akaike Information Criteria and Bayesian Information Criteria disagreeing about the preferred model. Figure 3 shows the utility decrements for each level of each dimension from model 2 with corresponding 95% confidence intervals; corresponding numerical values are given in Appendix Table I in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007.

In the mixed logit results (model 3, Table J in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007), the mean coefficients showed a similar pattern to that in the conditional logit (Fig. F in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007). Extra years of life in full health were generally highly valued (large, statistically significant mean coefficient), although considerable between-respondent heterogeneity was revealed in a relatively large, significant standard deviation coefficient. Of the 32 mean coefficients involving FACT-8D levels interacted with duration, 28 had the expected negative coefficients, and those that did not were not significantly different from zero at the 5% level. The standard deviation coefficients were generally significant; of the 33 standard deviations estimated, 23 were significant at the 5% level. The log-likelihood was considerably better than for the conditional logit analysis, suggesting that the assumption of preference homogeneity assumed in the conditional logit should be relaxed.

FACT-8D Utility Algorithm

The utility algorithm and scoring instructions for calculating an Australian value set (ie, weights for all FACT-8D health states) is provided in Appendix Report B in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007, with STATA and SPSS syntax.

Discussion

This article reports the development of the FACT-8D, a MAUI derived from the widely used cancer-specific HRQL profile measure, the FACT-G. Eight HRQL dimensions were included in the FACT-8D descriptive system: pain, fatigue, nausea, sleep, work, social support, sadness, and future health worry, each with 5 levels. This descriptive system includes 32 768 possible health states, capturing a wide combination of common impacts of cancer on quality of life. The DCE method used to elicit the Australian general population's valuations of those health states yielded utility decrements that were generally monotonic within each dimension; that is, poorer HRQL levels generally had larger utility decrements. The largest utility decrements were for the highest levels of pain and nausea, both common symptoms caused by cancer, and in the case of nausea, also caused by some treatments for cancer. The worst possible health ("pits") state had a utility of -0.54, considerably worse than dead (value = 0). This is comparable to the pits state value of the Australian EQ-5D(3L) DCE-based value set (–0.516),[14] and considerably lower than that of the Australian EORTC QLU-C10D value set (–0.095).[6]

The Australian FACT-8D value set provides another MAUI tool for Australian economic evaluations, particularly those in which a FACIT questionnaire has been used to assess HRQL. Some national regulatory agencies require country-specific value sets for decisions about national healthcare spending.[42,43] This study also provides the prerequisites for developing those value sets: the FACT-8D descriptive system and DCE valuation method. Country-specific valuation studies are underway in Canada, Japan, the United States, and the United Kingdom. Additional methodological extensions may also be useful, particularly to address the limitation inherent in building a utility generator (such as the FACT-8D or EORTC QLU-C10D) from an existing HRQL questionnaire, and hence a predefined descriptive system. Although this approach provides efficiencies, it is founded on the assumption of the content validity of the parent measure, which can change over time as new treatment options emerge. This may be solved by "bolt-on" dimensions to fill gaps in the FACT-8D descriptive system to capture treatment-related symptoms for emerging treatments such as immunotherapy. Bolt-on domains could also be used to make the FACT-8D more sensitive to symptoms of specific cancer sites, such as abdominal symptoms of ovarian cancer and

Table 2. Sociodemographic characteristics and self-reported health of the valuation survey sample (n = 1644) compared to those of the Australian general population.

Question	No. responders	Level	Number	Sample proportion or mean*	Population value*,†	*P* value‡
Sex	1644	Male	804	0.49	0.49	0.738‖
		Female	840	0.51	0.51	
Age (years)	1644	18-29	353	0.21	0.22	0.352‖
		30-39	289	0.18	0.18	
		40-49	287	0.17	0.18	
		50-59	266	0.16	0.16	
		60-69	213	0.13	0.13	
		70 or older	236	0.14	0.12	
General Health Question	1549	Excellent	156	0.10	0.10	<0.001
		Very good	467	0.30	0.35	
		Good	556	0.36	0.37	
		Fair	280	0.18	0.15	
		Poor	90	0.06	0.03	
Mental health§	1549	Kessler-10		*x=18.40*	*μ=14.50*	<0.001
Marital status	1557	Married (registered)	715	0.46	0.49	0.004
		Separated	49	0.03	0.03	
		Divorced	160	0.10	0.10	
		Widowed	71	0.05	0.06	
		Never married	562	0.36	0.32	
Highest level education	1557	Year 11 or below	245	0.16	0.28	<0.001
		Year 12	286	0.18	0.17	
		Trade or certificate I-IV	259	0.17	0.24	
		Diploma	251	0.16	0.09	
		Bachelor's degree	379	0.24	0.19	
		Higher degree	137	0.09	0.04	
Aboriginal or Torres Strait	1557	Yes	128	0.08	0.03	<0.001
Islander Status		No	1429	0.92	0.97	
Country of birth	1557	Australia	1185	0.76	0.79	<0.001
		English speaking, but NOT Australia	225	0.14	0.10	
		Not English speaking	147	0.09	0.11	

*Sample or population percentage prevalence reported in regular text; mean values reported in italics.
†Australian population sex and age distribution.[40] Australian population prevalence of the following variables: Aboriginal or Torres Strait Islander status, highest level of education, general health question, marital status, and country of birth were derived from the Household, Income, and Labour Dynamics in Australia Survey (HILDA, Wave 10), limited to those aged 18 years and older.
‡The chi-squared goodness-of-fit test was used to compare observed category frequencies to those expected based on Australian population proportions. Observed means (general health and Kessler-10) were compared to corresponding Australian population mean values using one-sample *t*-tests.
§Kessler-10 Australian population values were derived from the 2007 Australian National Health Survey.[41] Possible score range: 10–50. Low distress scores: 10–15, moderate: 16–21, high: 22–30, and very high: 31–50.
‖Indicates sample data is not significantly different from the Australian general population.

thoracic symptoms of lung cancer. The bolt-on approach has been demonstrated for the EQ-5D[44]—for example, to fill the psychosocial gap,[45] vision-specific gaps,[46] and even for cultural adaptation.[47]

The development of a MAUI from the FACT-G has important advantages for incorporating HRQL considerations into cancer treatment funding decisions. First, it allows quantification of utility for use in economic evaluation (specifically CUA) from responses to the FACT-G, a widely used cancer-specific HRQL questionnaire that is also included at the core of a large suite of FACIT questionnaires.[48] Although mapping algorithms are available to score utilities from FACT-G responses via mapping to generic MAUIs,[49] our valuation methodology is considered theoretically and empirically stronger because it complies with the Checklist for Reporting valuATion StudiEs (CREATE).[50] Second, the FACT-8D scoring algorithm can be applied retrospectively to any study that has used the FACT-G or a related FACIT questionnaire to assess HRQL. Third, the FACT-8D captures

Table 3. Conditional logit: parameter estimates (standard errors) for model 1 (unconstrained) and model 2 (monotonicity imposed*).

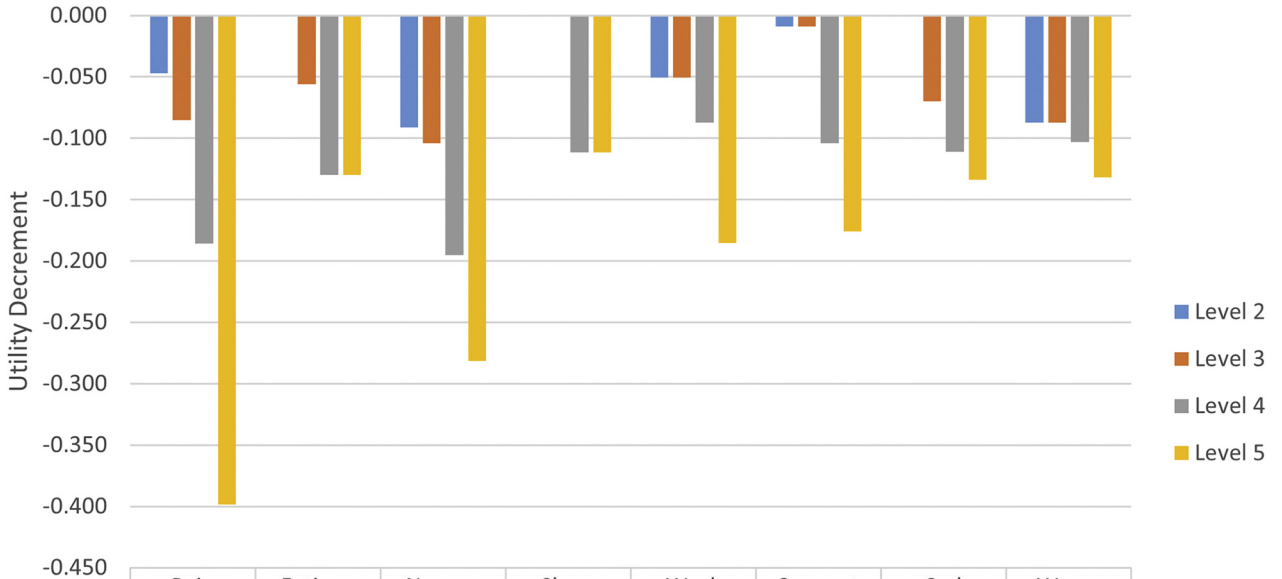
Dimension	Level	Unconstrained (Model 1)	Level	Monotonic (Model 2)
Survival duration (life years)	-	0.385 (0.014)†	-	0.394 (0.013)†
Pain	2	–0.014 (0.007)‡	2	–0.018 (0.007)§
	3	–0.03 (0.008)†	3	–0.034 (0.008)†
	4	–0.064 (0.008)†	4	–0.073 (0.008)†
	5	–0.162 (0.007)†	5	–0.157 (0.007)†
Fatigue (lack of energy)	2	*0.001 (0.008)*	2	0
	3	–0.024 (0.006)†	3	–0.022 (0.005)†
	4	*–0.053 (0.007)*†	4, 5	–0.051 (0.005)†
	5	*–0.050 (0.007)*†		
Nausea	2	–0.036 (0.007)†	2	–0.036 (0.006)†
	3	–0.038 (0.006)†	3	–0.041 (0.006)†
	4	–0.085 (0.006)†	4	–0.077 (0.006)†
	5	–0.110 (0.008)†	5	–0.111 (0.007)†
Problems sleeping	2	*0.001 (0.006)*	2	0
	3	*0.011 (0.007)*	3	0
	4	*–0.052 (0.007)*†	4,5	–0.044 (0.005)†
	5	*–0.019 (0.007)*†		
Problems doing work (including work at home)	2	*–0.027 (0.008)*†	2,3	–0.02 (0.005)†
	3	*–0.018 (0.006)*†		
	4	–0.032 (0.007)†	4	–0.034 (0.006)†
	5	–0.071 (0.006)†	5	–0.073 (0.006)†
Problems with support from my family and/or friends	2	*–0.017 (0.007)*§	2,3	–0.004 (0.005)
	3	*0.008 (0.006)*		
	4	–0.038 (0.006)†	4	–0.041 (0.006)†
	5	–0.064 (0.006)†	5	–0.069 (0.006)†
Sadness	2	*0.006 (0.007)*	2	0
	3	–0.028 (0.008)†	3	–0.028 (0.006)†
	4	–0.039 (0.006)‡	4	–0.044 (0.006)‡
	5	–0.053 (0.007)†	5	–0.053 (0.007)†
Worry my health will get worse	2	*–0.042 (0.006)*†	2,3	–0.034 (0.006)†
	3	*–0.026 (0.007)*†		
	4	–0.048 (0.008)‡	4	–0.04 (0.007)‡
	5	–0.05 (0.007)†	5	–0.052 (0.007)†
	Log-likelihood	–14971.275		–14995.282
	df	33		24
	AIC	30008.55		30038.56
	BIC	30300.07		30250.58

Levels of statistical significance: †1%; ‡10%; §5%.
AIC indicates Akaike information criterion; BIC, Bayesian information criterion.
*The coefficient for each level of each QOL domain was estimated as the interaction of that level with duration. Levels combined to ensure monotonicity within each dimension are noted in italics.

dimensions reflecting symptoms and impacts of cancer and its treatments that are not included in generic instruments, specifically nausea, fatigue, sleep problems, and worry about future health. The main drivers of utility in the Australian valuations were pain and nausea, common symptoms of cancer. It will be interesting to see whether these findings are replicated in Canada, the United Kingdom, and the United States, where valuation studies are currently underway.

Cost-utility analysis represents a major part of the reimbursement process in many countries.[42,43,51-53] In Australia, government guidelines for preparing submissions to the federal Pharmaceutical Benefits Advisory Committee favor direct estimation of utilities over mapping and do not mandate a particular MAUI but prefer Australian-based preference weights[52] (pages 37, 77). Submissions for cancer interventions frequently present FACT-G or FACIT data. Therefore, the value set presented here will

Figure 3. Australian FACT-8D utility decrements by dimension and level (derived from model 2 conditional logit, monotonicity imposed).



	Pain	Fatigue	Nausea	Sleep	Work	Support	Sad	Worry
Level 2	-0.047	0.000	-0.091	0.000	-0.051	-0.009	0.000	-0.087
Level 3	-0.085	-0.056	-0.104	0.000	-0.051	-0.009	-0.070	-0.087
Level 4	-0.186	-0.130	-0.195	-0.112	-0.087	-0.104	-0.111	-0.103
Level 5	-0.398	-0.130	-0.282	-0.112	-0.185	-0.176	-0.134	-0.132

aid Australian resource allocation decisions. In the United Kingdom, while the EQ-5D is generally preferred by the National Institute for Health and Care Excellence, a case can be made for condition-specific MAUIs if supported by empirical evidence. This highlights the need for head-to-head comparisons of the sensitivity of the FACT-8D and the EQ-5D using data from studies that have used both questionnaires in studies.

The study had strengths and limitations. The development of the descriptive system was psychometrically thorough, based on methods used previously by the MAUCa Consortium to develop the EORTC QLU-C10D.[5] The use of a large pooled data set representing a wide range of cancers and treatments (although not immunotherapy), and the inclusion of both oncologist and patient opinion, supports the clinical validity and applicability of the descriptive system. We used a DCE approach similar to one that we had previously established as feasible,[7] and modeling approaches appropriate to our data structure and analysis purpose.[21] As we had done for the EORTC QLU-C10D, we simplified the choice task by not asking respondents to trade across all dimensions at once, so the experimental design was not strictly orthogonal design; we decided this was the right balance between statistical and respondent efficiency, given the cognitive complexity of the task. The FACT-8D DCE presented a challenge we had not previously encountered[7,13,14,24]; when we piloted the DCE with the FACT-G's original positive framing of the work, sleep, and support items, credible preference weights could not be derived. We solved this by changing the polarity of these items, as discussed in the following paragraph. Unfortunately, we did not take steps to prevent people completing the DCE on very small screen sizes, such as cell phones, and we did not collect data on devices used; this should be done in future online DCEs. Patterns in the time taken to complete the DCE choice sets suggest most participants genuinely engaged in the valuation task. The valuation survey sample was large, with quota sampling achieving population representativeness for age and sex. Our sample was generally representative of the general Australian population, although respondents tended to be better educated and have poorer mental health.

The initial pronounced non-monotonicity in the pilot data posed a challenge we had not previously encountered when using DCE methods for utility estimation.[7,13,14,24] The mix of positively and negatively framed items distinguished the FACT-8D DCE from our previously successful DCEs. We therefore revised the wording of the DCE health states by reframing the 3 positively worded dimensions (work, sleep, support) as problem statements to help respondents with the choice task. The subsequent patterns of utility decrements then conformed with our previous experience. This then posed the challenge of how to map the corresponding positively framed FACT-G responses to utility decrements in the FACT-8D utility scoring algorithm. We opted for direct mapping using the original FACT-G item wording when calculating utility decrements (Table 1, Appendix Report B in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.01.007). This approach acknowledges that reframing was used solely for the purpose of making the DCE valuation task more feasible for participants (by decreasing the cognitive complexity), and that negatively and positively framed versions of these items would not necessarily yield mirror image results. Our rationale for choosing this option is that it aligns with our original aim, that is, to develop a preference-based means of generating utility scores from FACT-G responses. However, we acknowledge that this pragmatic solution is somewhat perplexing, as it is unclear how it might affect the applicability of the resulting preference weights to the standard FACT-G items. Better solutions may be found in

future research, particularly if similar problems arise in other valuation DCEs with positively and negatively framed items. An advantage of this pragmatic solution is that it allows utility values to be generated from the wealth of existing FACT-G data available, thus enabling these data sets to inform economic evaluation and other analyses. Future research will examine the psychometric properties of the FACT-8D.

Our DCE was conducted as an online self-complete exercise, with participants recruited via an online panel. This creates 2 potential sources of bias: (1) mode of administration (ie, online self-complete or in-person interviewer-assisted); and (2) online panels may not be representative of the general population. Mulhern et al (2013) investigated both these issues for DCE-type valuation surveys,[54] finding that online respondents were more highly educated than the general population (consistent with our findings) and the in-person sample. They also found that in-person interviewer-assisted completion took significantly longer than online self-completion, but there were no systematic differences between online and interviewer-assisted in responses to the valuation questions, and this was held when demographic differences between the 2 groups were controlled. Online panels enable geographically diverse samples, matched on targeted variables, to be recruited in a timely and economical manner. The extent to which nonrepresentativeness on sociodemographic variables as a limitation is as yet unknown; we will explore this in future modeling of FACT-8D valuation data pooled across countries.

Health economists continue to debate the appropriateness of using disease-specific utility weights for CUA. Generic utility instruments addressing more general health issues, such as the EQ-5D, are more typically used in CUA, because the metrics for these are comparable across health conditions and interventions. However, the capacity of generic utility instruments to capture all the issues that are clinically relevant in the context of cancer has been the subject of ongoing debate.[55] The FACT-8D provides utility weights for CUA of oncology interventions, and arguably may be more sensitive to differences in HRQL than generic instruments, but empirical research is needed to assess this claim.

Conditional logit and mixed logit models resulted in similar mean utility decrements away from level 1 in each dimension. Our preferred algorithm (model 2) is based on a conditional logit, which we selected because: (1) mean response is usually most relevant for economic evaluation, with heterogeneity a tangential issue; (2) the mixed logit model reported small positive coefficients for level 2 on 3 dimensions (Social Functioning, Emotional Functioning, Fatigue) and would therefore require further post hoc adjustment to enforce the monotonic structure of the instrument; (3) uncertainty about the appropriate distributional assumptions needed in mixed logit modeling.

The mental health dimension in the FACT-8D, represented by *Sadness* and *Worry that my health condition will get worse*, was given relatively less weight than in the Australian DCE-based valuation algorithms for the EQ-5D-3L[14] and EQ-5D-5L,[24] which placed *Anxiety/Depression* as 1 of the 3 most important of the 5 dimensions. For the cancer-specific EORTC QLU-C10D, the mental health dimension, represented by *Depression*, ranked as fourth most important of 10 dimensions in the Australian DCE-based valuation algorithm.[6] Depression and anxiety disorders are the most common mental disorders in Australia[56]; these terms may therefore evoke more serious health states than those conjured by the words "sadness" or "worry" for Australian survey participants. The difference in relative weighting of the mental health dimensions between algorithms probably reflects the different representations of mental health and the number and type of dimensions being compared within the DCE.

Because the FACT-8D is a new MAUI in the field, there is a need to identify how well it performs relative to existing instruments and whether there are particular patient populations where it performs well or poorly. As an early example, Herdman et al (2020) explored the validity and responsiveness of the FACT-8D in a specific cancer population (relapsed/refractory mantle cell lymphoma).[57] They noted that although the FACT-8D showed good convergent validity and responsiveness, the EQ-5D-5L showed superior known groups validity. These findings may not generalize to other data sets and cancer patient populations, and we support further validation work to identify the usefulness of the FACT-8D, beyond the practical consideration that it can be used in situations where preference-accompanied measures (such as the EQ-5D) have not been collected.

Conclusions and Future Directions

This research facilitates CUA for oncology interventions that include the FACT-G measure (including the many FACIT measures that embed the FACT-G items), particularly when a separate utility instrument was not included. In future clinical studies, using the FACT-G together with the FACT-8D algorithm will provide both granular HRQL dimension scores and preference-based utility index scores, with efficiencies for both patients (reduced burden) and researchers (reduced research costs), as the need to administer separate quality of life and utility measures is avoided. The value set reported here will facilitate Australian resource allocation decisions that are informed by CUA.

Future work will develop FACT-8D value sets for other countries using a standardized valuation protocol, explore international comparability of results, and review the performance of the FACT-8D against generic utility instruments. The suitability of the FACT-8D and other MAUIs such as the EORTC QLU-C10D and EQ-5D to capture the effects of modern treatments such as immunotherapy requires future research, and will continue to do so as cancer treatments evolve. Development of additional "bolt-on" dimensions could be used to fill gaps in the FACT-8D descriptive system for specific cancer sites and emerging treatments; these could be derived from FACT site-, treatment- and symptom-specific modules. The FACT-8D presents a starting point for such future research directions, and an opportunity to evaluate a broadly cancer-specific measure that can be applied across different cancer sites and treatments.

Compliance With Ethical Standards

All procedures 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 study was approved by the University of Sydney Human Research Ethics Committee, approval number 2012/2444. Informed consent was obtained from all individual participants included in the study.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2021.01.007.

Article and Author Information

Accepted for Publication: January 4, 2021

Published Online: May 7, 2021

doi: https://doi.org/10.1016/j.jval.2021.01.007

Author Affiliations: The University of Sydney, Faculty of Science, School of Psychology, Sydney, NSW, Australia (King, Mercieca-Bebber, Costa, Müller); Curtin University - Perth City Campus, and Department of Health Policy and Management, Bentley Campus, Perth, ACT, Australia (Norman); The University of Sydney, Faculty of Medicine and Health, NHMRC Clinical Trials Centre, Sydney, NSW, Australia (Mercieca-Bebber); Pain Management Research Institute, Saint Leonards, NSW, Australia and The University of Sydney, Sydney Medical School, Sydney, NSW, Australia (Costa); Canadian Centre for Applied Research in Cancer Control, Vancouver, BC, Canada and British Columbia Cancer Agency, Vancouver, BC, Canada (McTaggart-Cowan, Peacock); Simon Fraser University, Faculty of Health Sciences, Burnaby, BC, Canada (McTaggart-Cowan); Queensland University of Technology, School of Public Health, Institute of Health and Biomedical Innovation, Brisbane, QLD, Australia (Janda); Amsterdam University Medical Centres, Department of Medical Psychology, Amsterdam Public Health Research Institute, Amsterdam, Noord-Holland, NL (Müller); University of Technology Sydney, Centre for Health Economics Research and Evaluation, Sydney, NSW, Australia (Viney); University of Illinois at Chicago, Department of Pharmacy Systems, Outcomes and Policy, Chicago, IL, USA (Pickard); Northwestern University Feinberg School of Medicine, Department of Medical Social Sciences, Chicago, IL, USA (Cella).

Correspondence: Madeleine King, School of Psychology, Griffith Taylor Building (A19), The University of Sydney, NSW 2006, Australia. Email: madeleine.king@sydney.edu.au

Author Contributions: *Concept and design*: King, Norman, McTaggart-Cowan, Peacock, Janda, Viney, Pickard, Cella
Acquisition of data: Norman, Costa, Cella
Analysis and interpretation of data: King, Norman, Mercieca-Bebber, Costa, McTaggart-Cowan, Peacock, Janda, Müller, Viney, Pickard, Cella
Drafting of the manuscript: King, Norman, Mercieca-Bebber, Costa
Critical revision of the paper for important intellectual content: King, Norman, Mercieca-Bebber, Costa, McTaggart-Cowan, Peacock, Janda, Müller, Viney, Pickard, Cella
Statistical analysis: King, Norman, Mercieca-Bebber, Costa, Müller
Provision of study materials or patients: Norman, Mercieca-Bebber, Costa, Cella
Obtaining funding: King, Peacock, Janda, Viney, Pickard, Cella
Administrative, technical, or logistic support: King, Norman, Mercieca-Bebber, Costa, Müller.
Supervision: King

Conflict of Interest Disclosures: Drs King, Norman, Mercieca-Bebber, Costa, and Janda reported receiving grants from the Australian National Health and Medical Research Council during the conduct of the study. Dr Norman is an editor for *Value in Health* and had no role in the peer review process of this article. Dr Viney is a member of the Scientific Executive of the EuroQol Foundation and reported receiving grants from National Health and Medical Research Council during the conduct of the study; and grants and travel support from the EuroQoL Foundation outside the submitted work. Dr Cella reported receiving personal fees from FACIT.org outside the submitted work. No other disclosures were reported.

Funding/Support: This work was supported by Project Grant 632662 from the Australian National Health and Medical Research Council.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: Members of the Multi-Attribute Utility in Cancer (MAUCa) Consortium include: N. Aaronson, J. Brazier, D. Cella, D.S.J. Costa, P. Fayers, P. Grimison, M. Janda, G. Kemmler, M.T. King (Chair), H. McTaggart-Cowan, R. Mercieca-Bebber, R. Norman, S. Peacock, A.S. Pickard, D. Rowen, G. Velikova, R. Viney, D. Street, and T. Young. Margaret-Ann Tait provided valuable production assistance during submission of this article.

REFERENCES

1. Brazier JE, Rowen D, Mavranezouli I, 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 Technology Assessment (Winchester, England)*. 2012;16(32):1–114.
2. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care*. 2004;42(9):851–859.
3. Brazier JE, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21:271–292.
4. 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.
5. King MT, Costa DS, Aaronson NK, et al. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. *Qual Life Res*. 2016;25(3):625–636.
6. King MT, Viney R, Pickard AS, 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(2):225–238.
7. Norman R, Viney R, Aaronson NK, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. *Qual Life Res*. 2016;25(3):637–649.
8. Cella D, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol*. 1993;11(3):570–579.
9. Gamper EM, King MT, Norman R, et al. EORTC QLU-C10D value sets for Austria, Italy, and Poland. *Qual Life Res*. 2020;29(9):2485–2495.
10. Kemmler G, King M, Norman R, et al. German value sets for the EORTC QLU-C10D, a cancer-specific utility instrument based on the EORTC QLQ-C30. *Qual Life Res*. 2019;28:3197–3211.
11. McTaggart-Cowan H, King M, Norman R, et al. The EORTC QLU-C10D: the Canadian valuation study and algorithm to derive cancer-specific utilities from the EORTC QLQ-C30. *MDM Policy Pract*. 2019;4(1):2381468319842532.
12. Norman R, Mercieca-Bebber R, Rowen D, et al. U.K. utility weights for the EORTC QLU-C10D. *Health Econ*. 2019;28(12):1385–1401.
13. Norman R, Viney R, Brazier JE, et al. Valuing SF-6D health states using a discrete choice experiment. *Med Decis Making*. 2014;34(6):773–786.
14. Viney R, Norman R, Brazier JE, et al. An Australian discrete choice experiment to value EQ-5D health states. *Health Econ*. 2014;23(6):729–742.
15. Norman R, Kemmler G, Viney R, et al. Order of presentation of dimensions does not systematically bias utility weights from a discrete choice experiment. *Value Health*. 2016;19(8):1033–1038.
16. Ware Jr JE, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J Clin Epidemiol*. 1998;51(11):903–912.
17. Brucker PS, Yost K, Cashy J, et al. General population and cancer patient norms for the Functional Assessment of Cancer Therapy-General (FACT-G). [Erratum appears in *Eval Health Prof*. 2005;28(3):370]. *Eval Health Prof*. 2005;28(2):192–211.
18. Kessler RC, Andrews G, Colpe LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med*. 2002;32(6):959–976.
19. SurveyEngine P/L. SurveyEngine choice modelling process and method. Carlton North, Vic: SurveyEngine Pty Ltd; 2015. http://surveyengine.com.
20. Australian Bureau of Statistics. Australian demographic statistics: March 2013 Cat. No. 3101.02013. http://www.abs.gov.au/ausstats/abs@.nsf/lookup/3101.0Media%20Release1Mar%202013. Accessed February 15, 2018.
21. Bansback N, Brazier JE, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate societal health state utility values. *J Health Econ*. 2012;31:306–318.
22. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics*. 2008;26(8):661–677.
23. StataCorp. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP; 2013.
24. Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. *Appl Health Econ Health Policy*. 2013;11(3):287–298.
25. Viney R, Norman R, King MT, et al. Time trade-off derived EQ-5D weights for Australia. *Value Health*. 2011;14(6):928–936.
26. 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.
27. Bleichrodt H, Wakker P, Johannesson M. Characterizing QALYs by risk neutrality. *J Risk Uncertain*. 1997;15(2):107–114.
28. Hole A. Fitting mixed logit models by using maximum simulated likelihood. *Stata J*. 2007;7:388–401.
29. Mukuria C, Rowen D, Brazier JE, et al. Deriving a preference-based measure for myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015;18(6):846–855.
30. Mulhern B, Rowen D, Jacoby A, et al. The development of a QALY measure for epilepsy: NEWQOL-6D. *Epilepsy Behav*. 2012;24(1):36–43.
31. Rowen D, Brazier JE, Young T, et al. Deriving a preference-based measure for cancer using the EORTC-QLQC30. *Value Health*. 2011;14(5):721–731.
32. Rowen D, Mulhern B, Banerjee S, et al. Estimating preference-based single index measures for dementia using DEMQOL and DEMQOL-Proxy. *Value Health*. 2012;15(2):346–356.

33. Stevens K. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics*. 2012;30(8):729–747.
34. Revelt D, Train K. Customer-specific taste parameters and mixed logit. University of California, Berkeley; November 23, 1999.
35. Kolenikov S. Scrambled Halton sequences in Mata. *Stata J*. 2012;12(1):29–44.
36. Cella D, Rosenbloom SK, Beaumont JL, et al. Development and validation of eleven symptom indexes to evaluate response to chemotherapy for advanced cancer. *J Natl Compr Canc Netw*. 2011;9(3):13–24.
37. Yanez B, Pearman T, Lis CG, et al. The FACT-G7: a rapid version of the functional assessment of cancer therapy-general (FACT-G) for monitoring symptoms and concerns in oncology practice and research. *Ann Oncol*. 2013;24(4):1073–1078.
38. Greenberg DB. Barriers to the treatment of depression in cancer patients. *J Natl Cancer Inst Monogr*. 2004;32:127–135.
39. Simon GE, Von Korff M. Medical co-morbidity and validity of DSM-IV depression criteria. *Psychol Med*. 2006;36(1):27–36.
40. Australian Bureau of Statistics. 2011 Australian Census (March 2013). http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Mar%202013?OpenDocument. Accessed September 4, 2019.
41. Slade T, Grove R, Burgess P. Kessler Psychological Distress Scale: normative data from the 2007 Australian National Survey of Mental Health and Well-being. *Australian & New Zealand Journal of Psychiatry*. 2011;45(4):308–316.
42. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies. https://cadth.ca/dv/guidelines-economic-evaluation-health-technologies-canada-4th-edition. Accessed December 15, 2019.
43. National Institute for Health and Clinical Excellence. *Guide to the Methods of Technology Appraisal*. London: NICE; 2013.
44. Finch AP, Brazier JE, Mukuria C. Selecting bolt-on dimensions for the EQ-5D: examining their contribution to health-related quality of life. *Value Health*. 2019;22(1):50–61.
45. Chen G, Olsen JA. Filling the psycho-social gap in the EQ-5D: the empirical support for four bolt-on dimensions. *Qual Life Res*. 2020;09:09.
46. Gandhi M, Ang M, Teo K, et al. A vision "bolt-on" increases the responsiveness of EQ-5D: preliminary evidence from a study of cataract surgery. *Eur J Health Econ*. 2020;21(4):501–511.
47. Kangwanrattanakul K, Gross CR, Sunantiwat M, Thavorncharoensap M. Exploration of a cultural-adaptation of the EQ-5D for Thai population: a "bolt-on" experiment. *Qual Life Res*. 2019;28(5):1207–1215.
48. Cella D. *Manual of the Functional Assessment of Chronic Illness Therapy (FACIT) scales* (version 4). Evanston, IL: Northwestern University; 1997.
49. McTaggart-Cowan H, Teckle P, Peacock S. Mapping utilities from cancer-specific health-related quality of life instruments: a review of the literature. *Expert Rev Pharmacoecon Outcomes Res*. 2013;13(6):753–765.
50. Xie F, Pickard AS, Krabbe PF, et al. A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). *Pharmacoeconomics*. 2015;33(8):867–877.
51. Scottish Medicines Consortium. *Guidance to manufacturers for completion of new product assessment form (NPAF)*; 2016. Available at: https://www.scottishmedicines.org.uk/media/5498/working-with-smc-updated-october-2020.pdf.
52. Pharmaceutical Benefits Advisory Committee. *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). Version 5.0, September 2016*. Canberra: Department of Health; 2016. https://pbac.pbs.gov.au.
53. French National Authority for Health. Choices in methods for economic evaluation: a methodological guide. http://www.has-sante.fr/portail/jcms/r_1499251/en/choices-in-methods-for-economic-evaluation.
54. Mulhern B, Longworth L, Brazier J, et al. Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. *Value Health*. 2013;16(1):104–113.
55. Wailoo A, Davis S, Tosh J. *The incorporation of health benefits in cost utility analysis using the EQ-5D*. Decision Support Unit, School of Health and Related Research, University of Sheffield. 2010.
56. Reavley NJ, Jorm AF, Cvetkovski S, Mackinnon AJ. National depression and anxiety indices for Australia. *Aust N Z J Psychiatr*. 2011;45(9):780–787.
57. Herdman M, Kerr C, Pavesi M, et al. Testing the validity and responsiveness of a new cancer-specific health utility measure (FACT-8D) in relapsed/refractory mantle cell lymphoma, and comparison to EQ-5D-5L. *J Patient Rep Outcomes*. 2020;4(1):22.

The FACT-8D, a multi-attribute utility instrument derived from the cancer-specific quality of life questionnaire, FACT-G: development and Australian value set

Supplementary appendix

Contents

Table A FACT-G items: order of appearance in FACT-G (#), FACIT item code, item stem wording, item stem code and multi-item scale to which they belong.

Item #	FACIT code	Item stem wording	Item stem code	Scale
1	GP1	I have a lack of energy	energy	PWB
2	GP2	I have nausea	nausea	PWB
3	GP3	Because of my physical condition, I have trouble meeting the needs of my family	famneeds	PWB
4	GP4	I have pain	pain	PWB
5	GP5	I am bothered by side effects of treatment	side effects	PWB
6	GP6	I feel ill	ill	PWB
7	GP7	I am forced to spend time in bed	bed	PWB
8	GS1	I feel close to my friends	closefriends	SWB
9	GS2	I get emotional support from my family	suppfam	SWB
10	GS3	I get support from my friends	suppfriend	SWB
11	GS4	My family has accepted my illness	acceptfam	SWB
12	GS5	I am satisfied with family communication about my illness	commfam	SWB
13	GS6	I feel close to my partner (or the person who is my main support)	partner	SWB
-		I am satisfied with my sex life*	sex	SWB
14	GE1	I feel sad	sad	EWB
15	GE2	I am satisfied with how I am coping with my illness	coping	EWB
16	GE3	I am losing hope in the fight against my illness	hope	EWB
17	GE4	I feel nervous	nervous	EWB
18	GE5	I worry about dying	worry_dying	EWB
19	GE6	I worry that my condition will get worse	worry_worse	EWB
20	GF1	I am able to work (include work at home)	able-to-work	FWB
21	GF2	My work (include work at home) is fulfilling	work-fulfilling	FWB
22	GF3	I am able to enjoy my life	enjoy	FWB
23	GF4	I have accepted my illness	accept	FWB
24	GF5	I am sleeping well	sleep	FWB
25	GF6	I am enjoying the things I usually do for fun	fun	FWB
26	GF7	I am content with the quality of my life right now	qol	FWB

PWB = physical well-being; SWB = social well-being; EWB = emotional well-being; FWB = functional well-being

* Note that because respondents are invited to opt out of responding to this item, it was not included in any analysis reported in this summary.

Table B DCE design matrix

	Health State A									Health State B								
Design row	Pain	Fatigue	Nausea	Problems Sleeping	Problems doing work (including work at home)	Problems with support from family / friends	Sadness	Worry health condition will get worse	Duration	Pain	Fatigue	Nausea	Problems Sleeping	Problems doing work (including work at home)	Problems with support from family / friends	Sadness	Worry health condition will get worse	Duration
1	2	3	2	4	2	5	1	1	10	1	3	1	5	2	4	1	4	10
2	1	2	4	4	2	5	2	5	10	4	2	1	4	2	3	1	4	10
3	5	5	3	3	3	4	1	1	10	5	3	1	3	1	4	4	2	10
4	4	1	1	5	2	5	3	3	10	5	5	3	5	2	3	4	3	10
5	5	3	2	2	1	5	1	2	10	5	1	5	1	1	1	3	2	10
6	4	5	3	1	4	4	4	3	10	4	1	2	5	5	4	4	2	10
7	2	1	1	4	3	5	4	5	10	3	1	1	4	5	4	4	3	2
8	2	3	1	3	4	3	2	1	1	1	5	1	5	4	1	2	1	10
9	1	1	5	3	1	1	1	4	10	2	1	4	3	4	3	1	4	2
10	3	5	5	5	1	2	3	5	10	3	4	2	5	1	1	2	4	10
11	2	3	5	1	4	1	1	5	1	2	5	3	2	2	1	1	5	10
12	2	5	2	1	3	2	1	1	10	4	4	1	2	5	2	1	1	10
13	1	4	4	4	5	2	1	1	2	1	2	4	4	4	1	1	5	10
14	3	4	5	2	1	5	3	1	1	3	2	5	4	1	4	4	1	10
15	5	2	3	5	3	2	5	1	2	5	1	1	5	5	3	5	1	10
16	3	4	1	5	3	1	2	2	10	5	5	4	5	2	4	2	2	10
17	3	5	3	5	4	5	1	2	10	3	3	1	5	2	2	1	1	10
18	1	3	2	1	3	5	1	1	2	5	4	1	1	3	5	1	2	10
19	5	5	4	1	4	4	5	5	1	5	2	4	2	3	1	2	5	1
20	1	5	5	3	5	3	3	4	2	1	5	4	3	5	5	2	3	10
21	3	1	5	4	2	1	5	3	10	3	1	3	3	1	5	5	3	1
22	2	3	1	1	2	1	3	5	10	2	4	1	3	2	1	5	2	1
23	3	1	1	4	2	1	4	1	10	4	1	4	2	2	1	1	1	2
24	1	2	3	3	5	3	1	2	10	1	2	1	5	1	3	1	5	1

	Health State A									Health State B								
Design row	Pain	Fatigue	Nausea	Problems Sleeping	Problems doing work (including work at home)	Problems with support from family / friends	Sadness	Worry health condition will get worse	Duration	Pain	Fatigue	Nausea	Problems Sleeping	Problems doing work (including work at home)	Problems with support from family / friends	Sadness	Worry health condition will get worse	Duration
25	1	3	4	1	5	3	3	4	10	1	2	1	4	5	1	2	4	10
26	4	5	5	4	4	1	4	2	1	4	2	4	2	5	1	4	2	10
27	5	1	1	2	5	4	3	3	10	1	1	5	2	4	3	3	3	2
28	1	1	2	4	5	2	5	3	10	1	3	2	4	4	2	1	4	1
29	5	3	3	3	4	1	1	4	10	5	3	1	2	1	5	1	4	2
30	2	5	4	3	1	3	2	5	10	5	4	4	4	1	3	3	5	1
31	2	1	4	5	1	2	3	3	10	1	1	5	3	4	1	3	3	10
32	2	2	2	5	3	2	3	5	1	1	1	2	5	1	3	3	5	10
33	4	4	1	2	3	1	1	2	1	4	4	2	3	2	1	4	1	1
34	3	5	1	1	4	2	5	3	2	3	4	1	1	5	2	1	4	10
35	5	5	1	1	1	5	4	1	10	4	2	1	1	1	4	1	1	2
36	2	2	4	3	3	5	4	1	1	2	5	1	5	1	1	4	1	1
37	4	1	3	5	3	2	1	3	1	5	2	5	5	3	2	4	2	1
38	3	1	3	1	4	1	5	3	1	1	5	3	4	1	1	4	3	1
39	1	2	4	1	5	2	1	4	10	5	5	4	3	3	2	1	3	10
40	2	1	5	5	2	4	1	2	2	2	1	5	2	2	3	2	3	10
41	1	2	4	2	5	3	1	3	1	1	3	2	4	5	4	1	1	1
42	5	4	2	5	3	3	3	2	2	1	4	5	4	3	3	5	2	1
43	4	3	3	1	1	3	5	4	1	4	1	1	1	1	4	5	5	5
44	1	3	4	2	5	4	1	4	1	1	5	4	1	3	3	3	4	1
45	3	5	5	1	5	5	2	2	1	4	5	1	4	2	5	2	4	1
46	5	4	3	1	2	5	4	3	1	5	2	3	3	2	5	5	1	5
47	5	4	2	3	5	1	4	1	1	5	3	1	3	1	5	4	5	1
48	1	1	3	4	1	1	4	2	1	2	2	3	4	3	1	3	5	1
49	1	1	3	2	4	2	1	2	10	2	1	2	2	2	2	3	2	2

	Health State A									Health State B								
Design row	Pain	Fatigue	Nausea	Problems Sleeping	Problems doing work (including work at home)	Problems with support from family / friends	Sadness	Worry health condition will get worse	Duration	Pain	Fatigue	Nausea	Problems Sleeping	Problems doing work (including work at home)	Problems with support from family / friends	Sadness	Worry health condition will get worse	Duration
50	2	3	3	3	4	3	3	1	10	4	5	3	3	4	4	4	1	2
51	4	4	2	5	5	1	1	5	5	4	4	2	5	4	3	2	2	1
52	1	1	5	1	3	5	3	3	1	2	5	3	4	3	2	3	3	1
53	3	3	5	4	5	3	1	3	5	1	3	4	2	5	4	1	3	1
54	5	5	1	2	1	1	1	3	10	5	1	1	2	5	4	2	2	10
55	1	5	3	5	2	5	2	3	1	5	4	3	5	3	3	1	3	1
56	1	2	2	5	2	5	5	1	2	1	4	3	2	2	1	5	1	10
57	3	3	4	1	1	1	2	4	1	1	2	3	1	4	1	1	4	1
58	2	4	5	3	1	4	1	5	1	2	1	4	4	1	4	1	1	10
59	3	3	1	2	5	3	3	1	1	5	3	3	2	1	4	5	1	1
60	3	4	4	5	3	2	3	1	1	3	5	1	5	3	5	3	3	5
61	4	5	4	4	4	5	5	4	5	4	5	1	2	3	3	5	4	1
62	1	5	2	3	2	3	4	1	5	5	2	1	3	2	3	4	2	1
63	5	1	1	5	4	3	5	5	1	1	3	3	1	4	3	5	5	5
64	4	1	2	2	5	3	1	5	2	4	1	1	1	5	1	2	1	2
65	1	4	3	2	1	5	5	2	10	3	2	4	2	1	5	4	3	10
66	1	3	3	1	5	1	4	1	10	1	1	3	1	2	3	2	1	2
67	3	1	1	1	5	2	4	1	1	3	5	1	5	4	1	1	1	1
68	2	2	1	2	4	4	5	1	10	2	2	2	5	1	4	5	2	1
69	3	1	1	3	4	4	2	5	1	3	5	1	2	4	1	5	5	5
70	2	4	3	2	1	2	1	1	1	3	4	3	4	1	2	3	5	5
71	2	3	2	1	4	3	2	2	1	3	3	5	4	3	3	2	2	10
72	4	1	1	1	2	5	1	5	10	4	2	3	1	4	4	1	5	1
73	2	4	5	5	1	5	4	5	1	2	3	5	5	5	2	4	1	10
74	5	2	5	3	5	1	1	5	5	5	5	4	3	1	4	1	1	5

	Health State A									Health State B								
Design row	Pain	Fatigue	Nausea	Problems Sleeping	Problems doing work (including work at home)	Problems with support from family / friends	Sadness	Worry health condition will get worse	Duration	Pain	Fatigue	Nausea	Problems Sleeping	Problems doing work (including work at home)	Problems with support from family / friends	Sadness	Worry health condition will get worse	Duration
75	4	3	2	1	1	1	5	1	10	4	2	4	1	1	1	1	4	1
76	5	3	5	5	3	3	3	1	1	5	4	1	2	3	5	4	1	1
77	2	4	2	5	2	4	2	2	5	2	1	1	2	2	3	2	1	5
78	3	5	1	5	2	4	4	4	1	5	5	1	3	5	1	4	5	1
79	1	1	3	4	1	4	2	4	2	5	1	3	1	1	5	2	2	10
80	4	5	5	4	2	3	3	2	5	2	5	4	5	1	1	3	2	5
81	3	5	1	4	4	4	1	2	10	3	2	1	3	3	4	5	4	10
82	4	2	3	3	2	2	2	3	1	4	2	3	3	3	3	4	2	5
83	3	1	1	1	4	2	5	5	10	3	3	1	2	3	2	4	4	10
84	5	4	3	3	1	1	1	4	5	3	2	1	5	1	1	1	2	5
85	5	2	3	4	1	5	3	1	5	3	2	3	2	1	1	3	4	1
86	1	3	2	4	4	5	2	3	1	2	5	2	4	2	5	2	4	2
87	5	3	4	5	1	1	2	3	10	4	1	1	5	1	1	1	3	1
88	5	5	4	1	5	5	3	5	5	5	2	3	1	1	4	1	5	5
89	5	5	3	3	2	2	1	2	2	2	3	3	3	4	2	3	1	2
90	5	1	2	2	3	1	3	2	2	3	1	2	2	1	3	1	1	2
91	5	4	1	1	1	4	1	4	2	1	4	1	1	4	1	4	4	10
92	5	5	5	2	2	3	3	3	2	5	4	5	5	5	2	3	4	2
93	2	3	3	4	5	5	5	4	1	1	3	5	3	3	2	5	4	1
94	5	4	2	5	3	1	3	1	5	5	4	5	4	5	3	3	1	1
95	5	4	3	4	4	1	1	1	2	5	4	5	5	4	5	5	1	5
96	1	4	3	5	1	2	4	3	5	1	4	4	5	1	1	3	2	2
97	4	1	2	3	4	2	2	2	5	2	1	3	5	3	5	2	2	5
98	2	3	5	5	2	5	3	2	1	1	3	3	5	2	4	4	5	1
99	1	4	3	1	4	4	3	2	5	5	4	4	4	3	2	3	2	5
100	5	4	1	1	2	1	5	2	5	5	5	3	2	1	1	5	2	10

Table C Characteristics of the 24 datasets and the number of observations each contributed to various psychometric analyses

Study label	Study name	Primary cancer and stage	Country	Treatment				Obs/ patient	n	N
				Chemo	Radio	Surgery	Other		CFA/ Rasch	Respon-sivenes s
CRC	CanChange	Colorectal, mixed stages	Australia	863	220	1892	0	1	1843	0
LACE	Laparoscopic Approach to Cancer of the Endometrium	Endometrial, early stage	Australia	0	0	334	0	1	255	0
PTS	Pulling Through Study	Breast, mixed stages	Australia	114	190	249	65	1	191	0
TD	PhD (Tracey Di Sipio)	Breast, mixed stages	Australia	159	276	323	171	1	255	0
AOCS	Australian Ovarian Cancer Study	Ovarian, mixed stages	Australia	361	14	0	74	2+	704	353
AVCBT	Audiovisual Computer-Based Testing study	Mixed sites and stages	USA	551	88	490	103	2+	568	31
BioQOL	Bilingual Intercultural Oncology Quality of Life project	Mixed sites and stages	USA	427	354	1017	201	1	1252	0
FACIT-GI	Goal Interference study	Mixed sites and stages	USA	169	65	33	12	2+	21	33
FACT-Br	FACT-Br validation	Breast, unknown stage	USA	120	0	0	0	2+	0	68
FACT-Leu	FACT Leukemia module development	Leukemia, early stage	USA	38	0	9	0	2+	36	32
FHNSI	FACT Head and Neck Symptom Index validation	Head and neck, late stage	USA	82	44	27	0	2+	63	66
FKSI	FACT Kidney Symptom Index development	Kidney, mixed stage	USA	6	9	0	0	2+	10	8
Pegfil	Pegfilgrastim	Mixed sites and stages	USA	14	0	0	0	1	4	0
QScore	QScore validation	Mixed sites and stages	USA	573	136	624	102	2+	435	116
Rush CS	Rush clinical significance	Mixed sites and stages	USA	59	19	0	25	2+	54	65

Thrombo	Thrombocytopenia	Mixed sites and stages	USA	75	13	38	9	2+	63	56
Compass	Computerized Phone Activating Self Survey	Breast, mixed stages	USA	18	0	5	36	1	46	0
Cytokine Fatigue	Cytokine Fatigue study	Mixed sites and stages	USA	0	0	2	13	1	39	0
FHSI	FACT Hepatobiliary Symptom Index-8 validation	Mixed sites, late stage	USA	19	2	22	0	1	29	8
Lung study	Lung Study	Lung, late stage	USA	230	0	0	0	1	161	0
Lymphoma	Lymphoma Study	Lymphoma, mixed stages	USA	48	8	4	0	2+	25	59
NCCN	NCCN Study	Mixed sites, late stage	USA	533	0	0	0	1	447	0
Prostate	Prostate Pathways	Prostate, mixed stages	USA	92	84	99	80	1	108	0
Vancouver	BC Cancer Breast, Colorectal, and Lung Quality of Life Study	Mixed sites and stages	Canada	231	47	3	88	1	303	0

Table D Pooled data used for Stage 1 psychometric analyses: Frequency of observations by primary cancer site and stage for the pooled dataset used to assess Criteria 1-7 (n=6912 patients)

	Stages I-III		Stage IV/recurrent/metastatic	
	Frequency	**Percent**	**Frequency**	**Percent**
Brain	0	0.00	49	0.71
Breast	979	14.16	558	8.07
Colorectal	1917	27.73	467	6.76
Genito-urinary	6	0.09	37	0.54
Gynaecological	499	7.22	568	8.22
Head and neck	243	3.52	166	2.40
Leukaemia	36	0.52	0	0.00
Liver/bile/pancreas	0	0.00	69	1.00
Lung	280	4.05	495	7.16
Malignant lymphoma	17	0.25	69	1.00
Malignant melanoma	2	0.03	4	0.06
Oesophagus/stomach	17	0.25	1	0.01
Prostate	154	2.23	111	1.61
Sarcoma	2	0.03	4	0.06
Testicular	13	0.19	7	0.10
Other	61	0.88	81	1.17
Total	**4226**	**61.1%**	**2686**	**38.9%**

Table E Pooled data used in Stage 1 to assess Criteria 1-7: Frequency of observations for each treatment (n=6912 observations, 1 per patient)

	Frequency	**Percent**
No treatment	474	6.86
Radiotherapy	208	3.01
Hormonal therapy	136	1.97
Surgery	1848	26.74
Chemotherapy	1503	21.74
Analgesics	1	0.01
Chemotherapy/Radiotherapy	179	2.59
Hormonal therapy/Radiotherapy	24	0.35
Other treatment	58	0.84
Surgery/Radiotherapy	260	3.76
Surgery/Chemotherapy	1317	19.05
Surgery/Chemotherapy/Radiotherapy	337	4.88
Surgery/Hormonal therapy	193	2.79
Chemotherapy/Hormonal therapy	36	0.52
Surgery/Chemotherapy/Hormonal therapy	76	1.10
Surgery/Hormonal therapy/Radiotherapy	123	1.78
Chemotherapy/Hormonal therapy/Radiotherapy	6	0.09
Surgery/Chemotherapy/Hormonal therapy/Radiotherapy	133	1.92
Total	**6912**	**100%**

Table F Criterion 8: Responsiveness to change (effect size[a]), by item and QLQ-C30 domain scale, for each study[b]

Item #	Item stem code	1	2	3	4	5	6	7	8	9	10	# ≥0.50	# ≥0.2 and <0.50	Top*
	N	353	31	33	68	32	66	59	116	65	56			
1	energy	-0.33	0.25	-0.31	-0.17	-0.03	0.57	0	-0.05	-0.18	0.05	1	3	1
2	nausea	-0.65	-0.30	-0.56	-0.18	0.07	0	0.07	-0.22	-0.22	0.08	2	3	4
3	famneeds	-0.22	-0.25	-0.04	0.02	0.08	0.30	0.05	-0.17	-0.14	0.18	0	3	2
4	pain	-0.08	-0.03	0	-0.12	-0.14	0.24	0.02	0.10	-0.62	0.09	1	1	1
5	sideeffects	-0.53	-0.24	0.11	-0.33	-0.10	0.56	0.13	-0.16	-0.20	-0.11	2	2	1
6	ill	-0.41	-0.16	-0.10	-0.09	0.2	0.42	0.22	-0.03	-0.11	-0.14	0	4	1
7	bed	-0.24	-0.14	-0.13	-0.29	0.25	0.40	0.04	0.01	-0.23	-0.15	0	5	1
8	closefriends	-0.07	-0.08	0.13	-0.19	0.28	0.11	-0.25	0.02	-0.19	0.03	0	2	5
9	suppfam	0.05	0.12	0.22	-0.17	0.2	-0.02	-0.17	-0.20	0.03	0.18	0	2	1
10	suppfriend	0	0.27	0.40	-0.16	0.24	0.07	-0.18	-0.13	-0.31	0.09	0	4	1
11	acceptfam	-0.07	0.15	0.32	0	0.2	-0.2	-0.19	-0.30	0.16	0.19	0	2	3
12	commfam	-0.01	0.14	0.25	-0.13	0.15	-0.02	0.05	0.06	-0.04	0.16	0	1	0
13	partner	0	0.28	0.09	-0.12	0.14	0	0	-0.20	0.33	0.08	0	3	2
14	sad	-0.10	0.14	0	-0.08	-0.03	-0.08	-0.19	0.07	-0.39	0.07	0	1	2
15	coping	-0.16	0.14	0.08	-0.08	0.16	0.2	-0.22	-0.12	-0.08	0.23	0	2	4
16	hope	-0.09	-0.14	-0.08	-0.10	0.03	0.10	0.19	0.20	-0.25	-0.04	0	1	3
17	nervous	-0.09	0	-0.07	-0.02	0.09	0.02	0	0.14	-0.08	-0.03	0	0	1
18	dying	-0.09	0	0.02	-0.03	-0.03	-0.06	0	0.09	0	0.03	0	0	0
19	worse	-0.09	-0.23	-0.07	0.05	0	-0.04	-0.18	0.19	0.01	-0.08	0	1	2

20	work	-0.32	-0.13	-0.17	-0.15	0.24	0.39	-0.22	0.06	-0.17	0.08	0	3	4
21	fulfil	-0.19	0.13	-0.15	-0.21	0.17	0.32	-0.18	0.19	0.06	0.10	0	2	3
22	enjoy	-0.26	0.15	-0.03	-0.03	0.30	0.35	-0.11	-0.12	-0.02	0.20	0	3	2
23	accept	-0.17	0	0.09	0.06	0.16	0.17	-0.06	-0.15	0.1	0.08	0	0	0
24	sleep	-0.16	-0.08	0.07	-0.16	0.22	0.27	-0.04	0.05	-0.03	-0.07	0	2	0
25	fun	-0.15	-0.08	-0.33	-0.09	0.34	0.50	-0.03	-0.13	-0.14	0.15	1	2	2
26	qol	-0.31	0	-0.21	-0.12	0.05	0.16	-0.07	-0.05	-0.08	0.10	0	2	0

[a] Effect size for responsiveness was calculated as mean change from baseline to on-treatment divided by standard deviation of change.

[b] Studies: 1 = AOCS; 2 = AVCBT; 3 = FACIT-GI; 4 = FACT-Br; 5 = FACT-Leukemia; 6 = FHNSI; 7 = Lymphoma; 8 = QScore; 9 = Rush CS; 10 = Thrombo

Light blue cells: $0.2 \leq |\text{effect size}| < 0.5$

Dark blue cells: $|\text{effect size}| \geq 0.5$

≥ 0.5 indicates the number of data sets for which the item had $|\text{effect size}| \geq 0.5$

> 0.2 indicates the number of data sets for which the item had $0.2 < |\text{effect size}| < 0.5$

* Top = number of times the item was the most responsive

Table G Characteristics of patients surveyed to obtain patient input on the relative importance of items: survey development (qualitative phase, n=10), quantitative survey (n=82, used to assess Criterion 9)

		Qualitative (n=10)	Quantitative (n=82)
Age	Mean (SD)	56.7 (14.5)	61.7 (9.7)
Sex	Female	2	46
	Male	8	33
	Not recorded	0	2
Primary cancer site	Colorectal	1	20
	Breast	0	14
	Ovarian	0	11
	Lung	1	11
	Prostate	1	6
	Oesophagus/stomach	3	0
	Pancreatic	3	0
	Other	1	18
	Not recorded	0	2
Stage	Localised	4	26
	Metastasised	4	36
	Unknown	0	12
	Not recorded	2	8
Treatment	Surgery	5	48
	Chemotherapy	9	73
	Radiotherapy	4	23
	Hormone therapy	2	14
Time since diagnosis	< 6 months	6	25
	6-12 months	1	15
	12-24 months	0	14
	2-5 years	0	14
	5-10 years	2	5
	> 10 years	0	6
	Not recorded	1	3
Current treatment	Chemotherapy only	4	62
	Chemotherapy and radiotherapy	2	0
	Hormone therapy only	1	2
	Chemotherapy and hormone	0	1
	Other	2	5
	Not recorded	1	12

Table H Respondents' perceptions of the discrete choice experiment valuation task and their choice strategies

How clear was the presentation of the health states?					
	Pilot		**Main study**		**Test statistic**
Response	**Frequency**	**Percent**	**Frequency**	**Percent**	
Very unclear	2	1.0%	28	1.8%	
Unclear	7	3.4%	48	3.1%	
Neither clear nor unclear	40	19.1%	254	16.3%	
Clear	109	52.2%	789	50.5%	$\chi2 = 2.86$
Very clear	51	24.4%	443	28.4%	df = 4
Total	**209**	**100%**	**1562**	**100%**	*p-value* = 0.58
How difficult was it to choose between the pairs of health states on each screen?					
	Pilot		**Main study**		**Test statistic**
Response	**Frequency**	**Percent**	**Frequency**	**Percent**	
Very difficult	15	7.2%	84	5.4%	
Difficult	78	37.3%	572	36.6%	
Neither easy nor difficult	65	31.1%	495	31.7%	
Easy	42	20.1%	313	20.0%	$\chi2 = 2.29$
Very easy	9	4.3%	98	6.3%	df = 4
Total	**209**	**100%**	**1562**	**100%**	*p-value* = 0.68
Did you have a strategy for choosing between the pairs of health states on each screen?					
	Pilot		**Main study**		**Test statistic**
Response	**Frequency**	**Percent**	**Frequency**	**Percent**	
I did not have a strategy	17	8.1%	177	11.3%	
I focused on just a few aspects of the health states	54	25.8%	269	17.2%	
I focused on the aspects that were highlighted in yellow	29	13.9%	339	21.7%	
I considered most of the aspects	49	23.4%	305	19.5%	
I considered all of the aspects	51	24.4%	431	27.6%	$\chi2 = 18.61$
Other	9	4.3%	41	2.6%	df = 5
Total	**209**	**100%**	**1562**	**100%**	*p-value* < .01

Note: As test statistic, chi-square ($\chi2$) test of homogeneity was applied.

Table I Utility decrements used in the FACT-8D utility algorithm[a]

Dimesion	Level	Utility decrement (95% CI)
Pain	2	0.047 (0.012,0.081)
	3	0.085 (0.047,0.124)
	4	0.186 (0.151,0.221)
	5	0.398 (0.368,0.428)
Fatigue (lack of energy)	3	0.056 (0.031,0.082)
	4,5	0.13 (0.107,0.153)
Nausea	2	0.091 (0.062,0.121)
	3	0.104 (0.077,0.13)
	4	0.195 (0.168,0.222)
	5	0.282 (0.248,0.316)
Problems sleeping	4,5	0.112 (0.089,0.135)
Problems doing work (including work at home)	2,3	0.051 (0.024,0.077)
	4	0.087 (0.056,0.117)
	5	0.185 (0.16,0.211)
Problems with support from my family and/or friends	2,3	0.009 (-0.015,0.033)
	4	0.104 (0.074,0.134)
	5	0.176 (0.148,0.204)
Sadness	3	0.07 (0.041,0.099)
	4	0.111 (0.084,0.137)
	5	0.134 (0.103,0.164)
Worry my health will get worse	2,3	0.087 (0.061,0.112)
	4	0.103 (0.069,0.136)
	5	0.132 (0.1,0.164)

[a]From Model 2, conditional logit, monotonicity imposed

Table J: Mixed logit (Model 3)

		Model 3	
Mean		**Mean (Robust SE)**	**SD (Robust SE)**
Duration	**Linear**	0.616 (0.019)*	0.304 (0.011)*
Pain x Duration	2	-0.014 (0.009)	-0.033 (0.023)
	3	-0.058 (0.011)*	-0.076 (0.021)*
	4	-0.092 (0.010)*	-0.011 (0.027)
	5	-0.251 (0.010)*	0.170 (0.012)*
Fatigue x Duration	2	-0.010 (0.010)	-0.031 (0.024)
	3	-0.043 (0.008)*	-0.031 (0.016)
	4	-0.075 (0.010)*	-0.010 (0.018)
	5	-0.078 (0.009)*	-0.008 (0.020)
Nausea x Duration	2	-0.056 (0.008)*	0.006 (0.019)
	3	-0.057 (0.008)*	-0.022 (0.017)
	4	-0.106 (0.008)*	-0.048 (0.017)*
	5	-0.144 (0.010)*	-0.034 (0.027)
Sleep x Duration	2	0.010 (0.008)	0.043 (0.018)*
	3	0.024 (0.010)*	0.028 (0.019)
	4	-0.069 (0.009)*	-0.068 (0.015)*
	5	-0.023 (0.009)*	0.069 (0.017)*
Work x Duration	2	-0.020 (0.010)	0.016 (0.032)
	3	-0.022 (0.009)*	0.003 (0.014)
	4	-0.036 (0.009)*	-0.051 (0.016)*
	5	-0.101 (0.008)*	0.032 (0.015)*
Support x Duration	2	-0.022 (0.009)*	0.028 (0.019)
	3	0.011 (0.008)	-0.016 (0.030)
	4	-0.059 (0.009)*	-0.002 (0.023)
	5	-0.093 (0.008)*	0.078 (0.014)*
Sadness x Duration	2	0.013 (0.009)	-0.060 (0.020)*
	3	-0.042 (0.010)*	0.000 (0.021)
	4	-0.060 (0.008)*	-0.008 (0.015)
	5	-0.092 (0.009)*	0.049 (0.016)*
Worry x Duration	2	-0.056 (0.008)*	0.012 (0.014)
	3	-0.045 (0.009)*	0.026 (0.020)
	4	-0.069 (0.010)*	-0.022 (0.016)
	5	-0.071 (0.009)*	-0.036 (0.018)*

Note. * indicates significance, i.e. the 95% Confidence Interval does not entail 0.

Figure A Criterion 6: Item response frequencies for the 26 items of the FACT-G included in the Stage 1 analysis

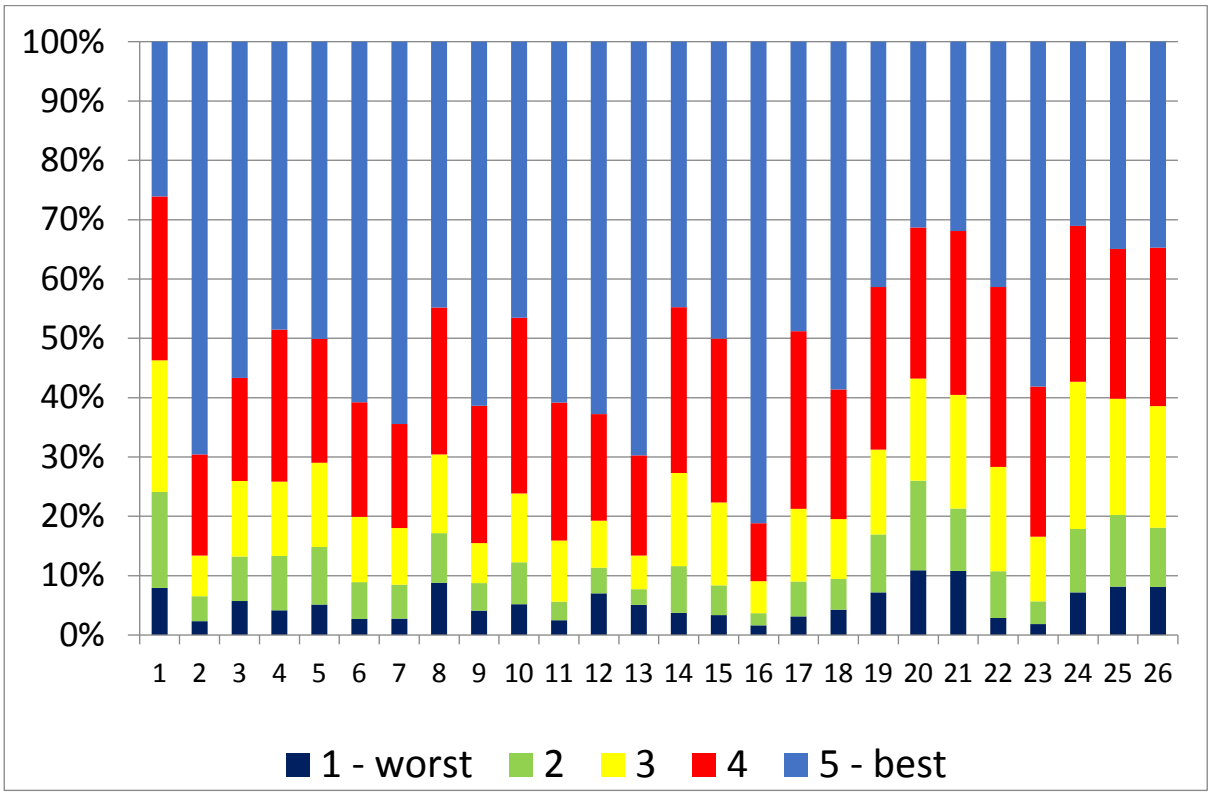


Figure B FACT-8D utility decrements from the pilot survey (n= 209)

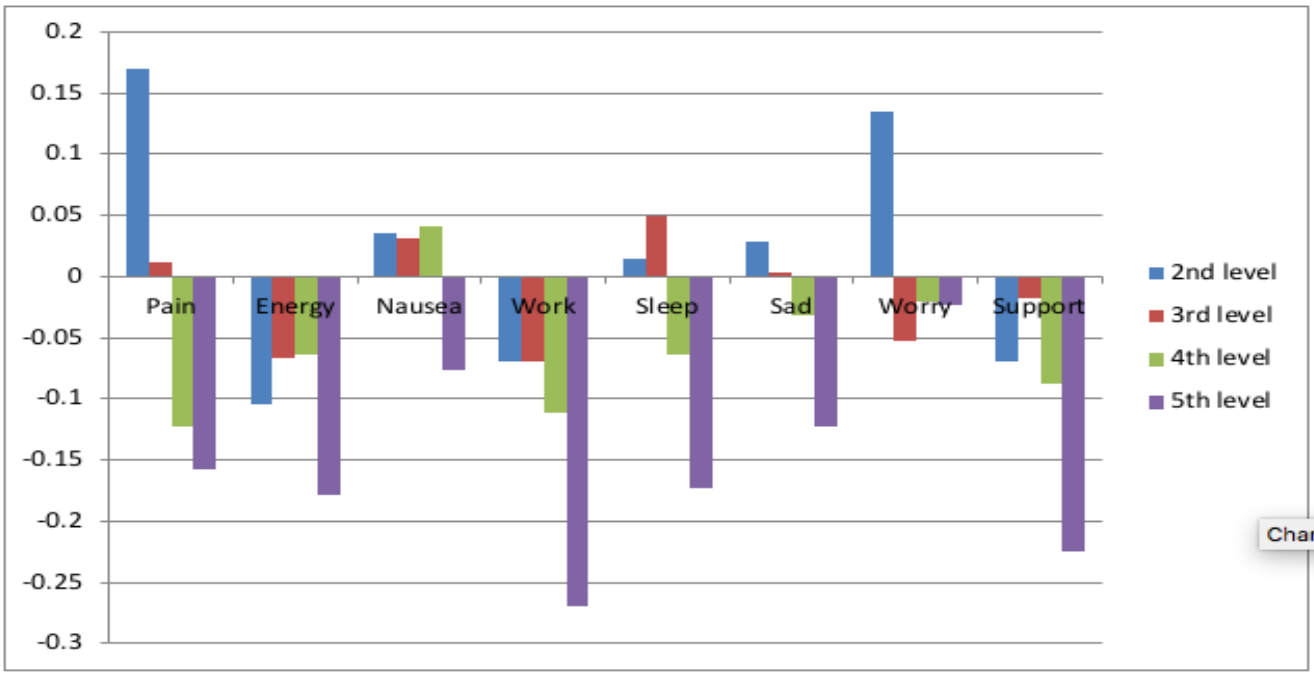


Figure C: Participants' perceptions of clarity of the DCE choice task: pilot (n=209) and main study (n=1562)

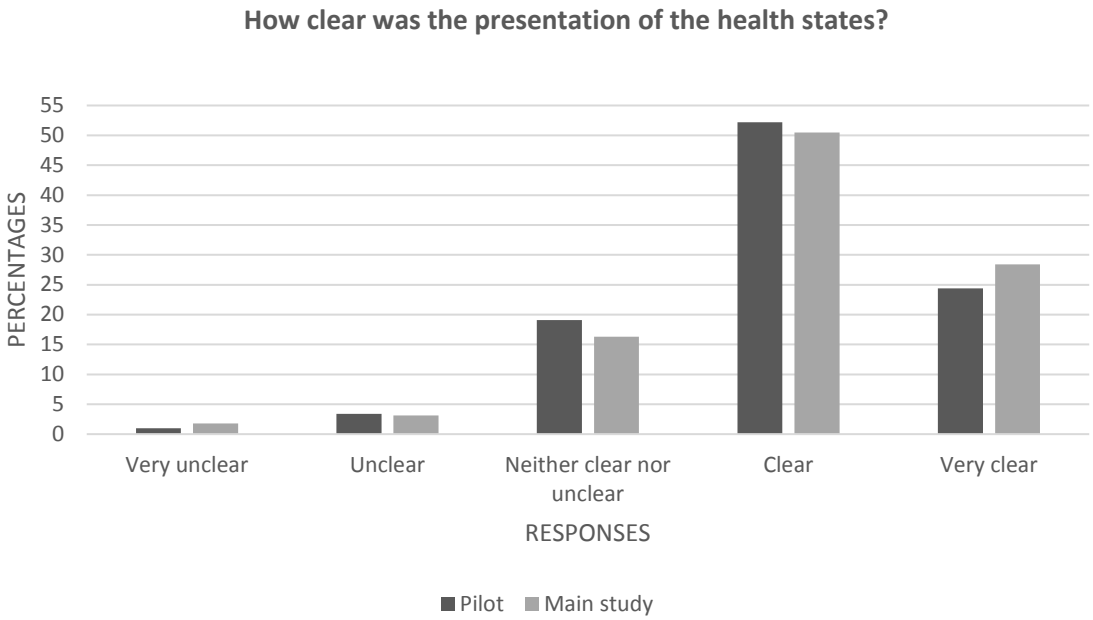


Figure D: Participants' perceptions of the difficulty of the DCE choice task: pilot (n=209) and main study (n=1562)

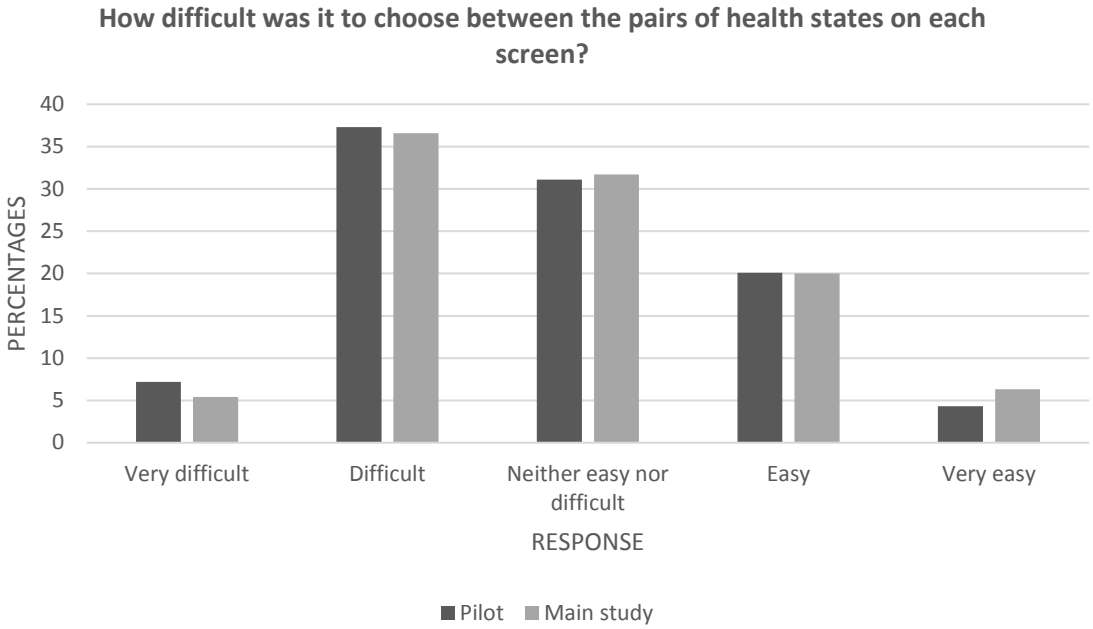
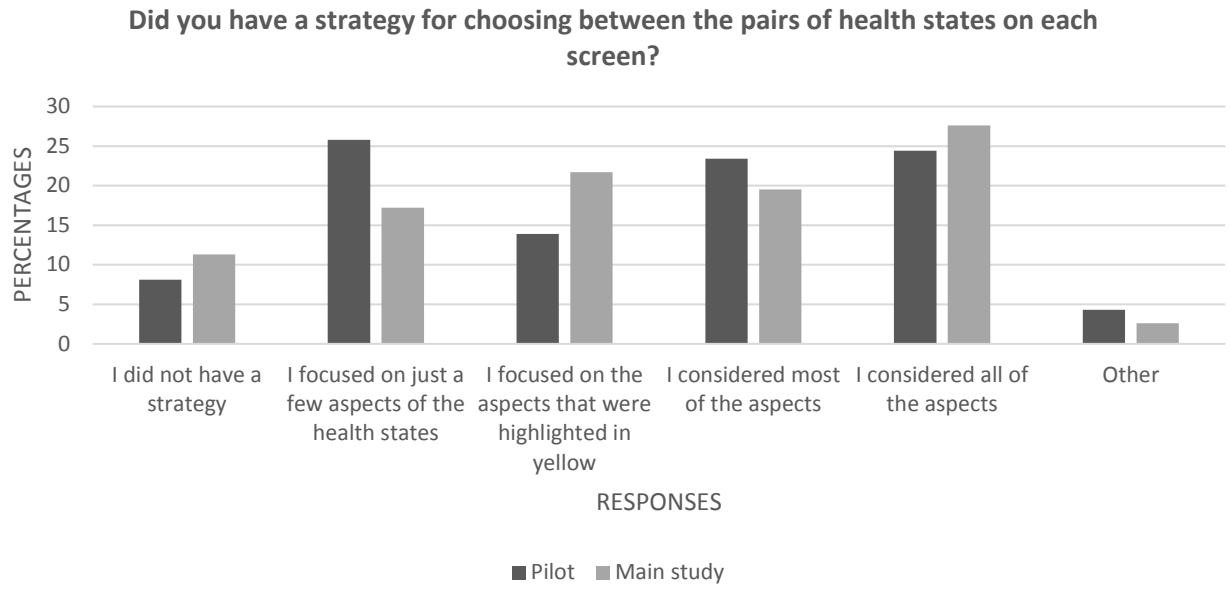


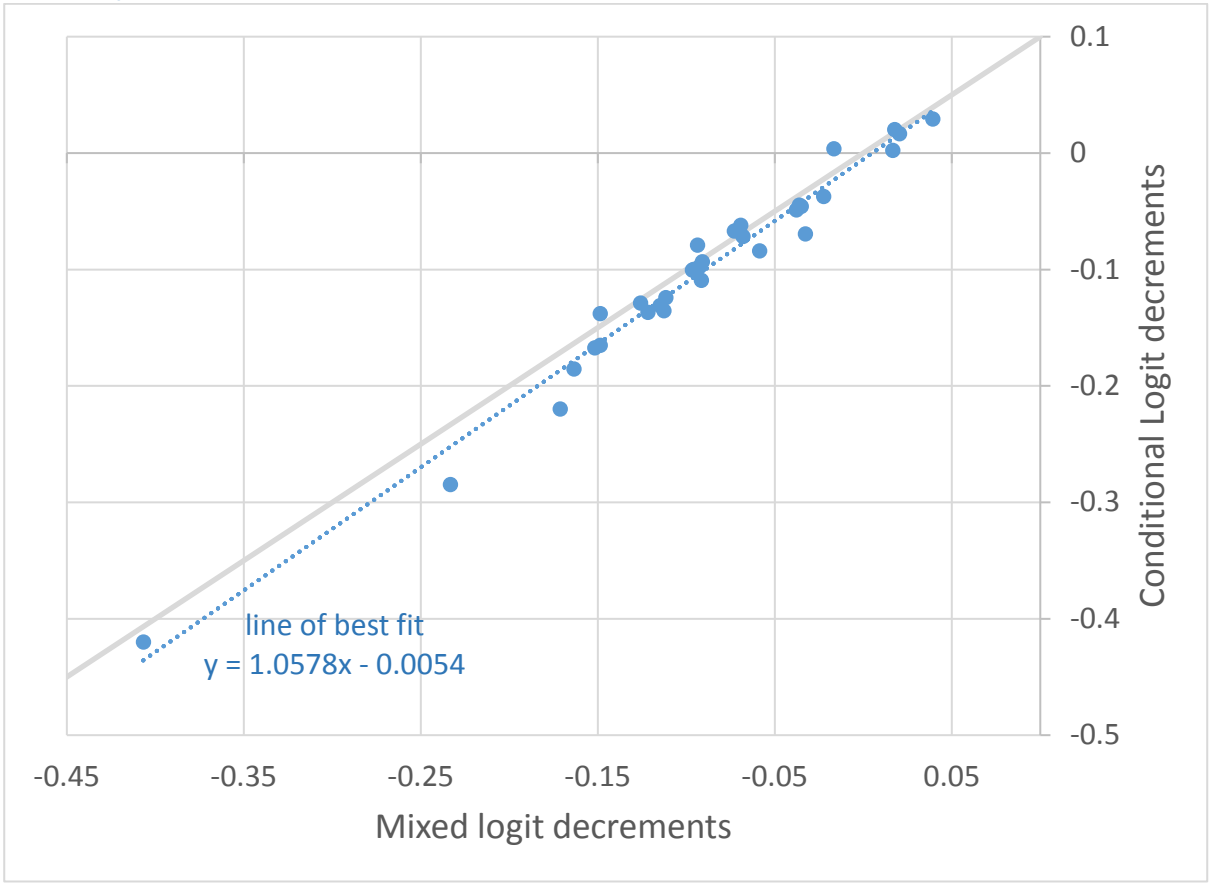
Figure E: Participants' perceived choice strategies in the DCE choice task: pilot (n=209) and main study (n=1562)



Coding of free text from 41 respondents in the main study who chose 'Other' strategy	Frequency	Percent
Length of survival	8	32.0%
Length of survival + additional aspect	7	28.0%
Pain	2	8.0%
Focused on aspects highlighted in yellow	2	8.0%
Other (including own health state and other symptoms)	6	24.0%
Total	**25**	**100%**

Note. Of those 41 participants who selected the response 'Other' when asked which strategy they applied for choosing between the pairs of health states, 16 participants did not provide a valid or informative answer, resulting in 25 valid responses. Length of survival time was considered by the majority of participants (n = 15) when choosing between the health states. Note that 1 of those participants preferred a shorter life span.

Figure F: Scatter plot of utility decrements generated by conditional logit (Model 1) and mixed logit (Model 3) with line of best fit (dotted) and line of equality (solid)



Appendix Report A - Rasch Analysis (Criteria 1-5)

Table of Contents

Rasch analysis

Once dimensional structure was established by confirmatory factor analysis, Rasch analysis was conducted on the items as part of the criteria for selection in the health state classification system. A random sub-sample of approximately 250 was selected for Rasch analysis using an in-built function in RUMM2020. This was done because fit statistics in Rasch analysis are sensitive to sample size, and 250 is within the recommended range to allow appropriate interpretation of fit statistics.

Criteria:

(1) Model fit: overall fit of the data to the Rasch model was indicated by a non-significant (Bonferroni-adjusted) chi-squared statistic; item and person fit was indicated by fit residual standard deviations less than 1.5;
(2) Response format: this was assessed by examining whether item response thresholds were ordered, as indicated by a threshold map;
(3) spread of item thresholds across the latent variable;
(4) Invariance across groups (differential item functioning, DIF): invariance across the different levels of a factor was indicated by a non-significant (Bonferroni-adjusted) two-way anova (level x class interval) with expected score as the dependent variable; gender and stage were the factors tested (DIF by primary site was not analysed because of the large number of categories); category probability curves were examined to aid interpretation;
(5) Local dependency; residual correlation coefficients were used to determine if pairs of items formed sub-factors (or "sub-tests"), while controlling for item difficulty; residual correlations of approximately.3 or more above the mean inter-item correlation were considered high.

Based on the results of the confirmatory factor analyses, Rasch analysis was performed on the following sets of items:

(a) 1-7 (physical well-being)
(b) 8-13 (social well-being)
(c) 14-19 (emotional well-being)
(d) 20-26 (functional well-being)

The following section describes the results of the Rasch analyses. For each domain, a short summary of results is provided first, followed by a table summarising the key statistics (fit statistics, presence of DIF, local dependency and disordered thresholds), then finally a more detailed presentation of results (nature of DIF and disordered thresholds).

Physical well-being

Summary: Item 2 (nausea) exhibited a disordered threshold and local dependence with item 6 (ill). Although overall item fit just failed to meet criterion to be considered good, no individual item exhibited poor fit. No DIF was observed.

Table 1. Item criteria from Rasch – Physical well-being

	Item		Location	Item fit	DIF[a]	LD[b]	DT[c]
Items 1-7							
Model fit *p* = .01	1	energy	0.80	0.07	-	-	-
Item fit = 1.52	2	nausea	-0.56	-0.34	-	6	x
Person fit = 0.96	3	famneeds	0.06	-1.13	-	-	-
	4	pain	0.10	2.23	-	-	-
	5	sideeffects	0.24	0.69	-	-	-
	6	ill	-0.21	-1.92	-	2	-
	7	bed	-0.44	-2.07	-	-	-

[a] DIF = differential item functioning; cell contains the name of the variable for which DIF was observed

[b] LD = local dependency; cell contains the number of the item with which there is a correlation

[c] DT = disordered threshold; x indicated the presence of a disordered threshold

"-" indicates the absence of DIF, LD or DT

Figure 1. Thresholds – Physical well-being

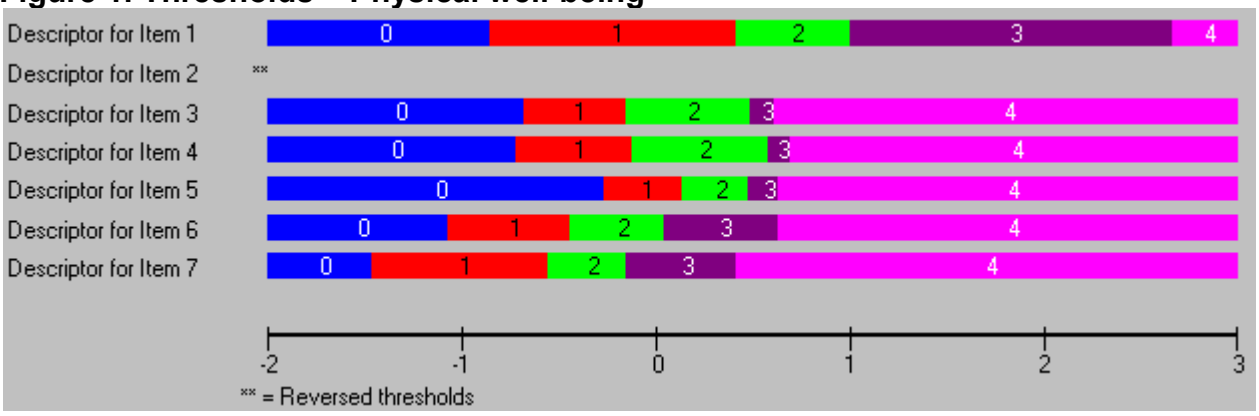
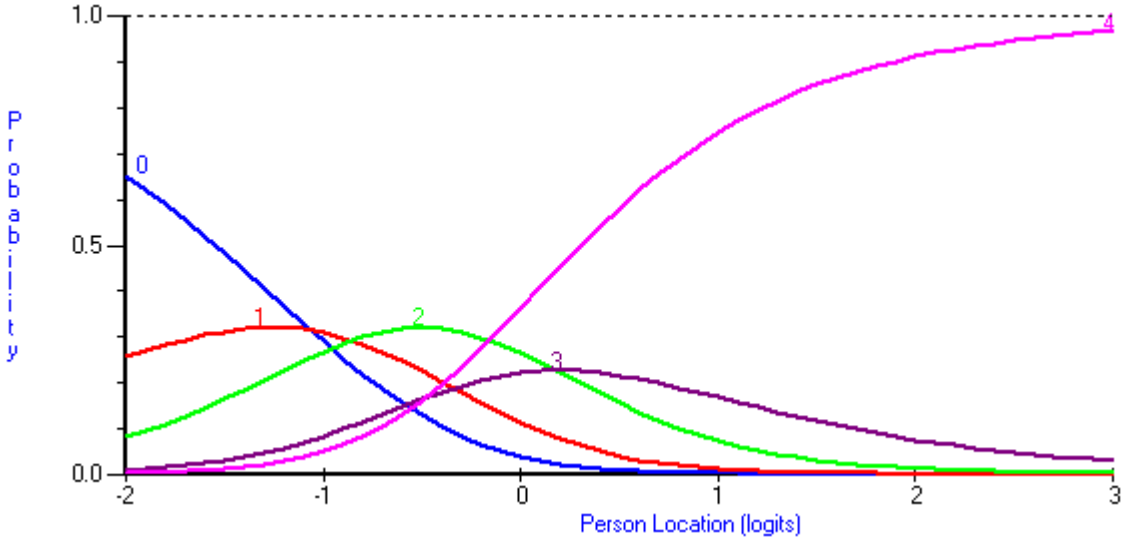


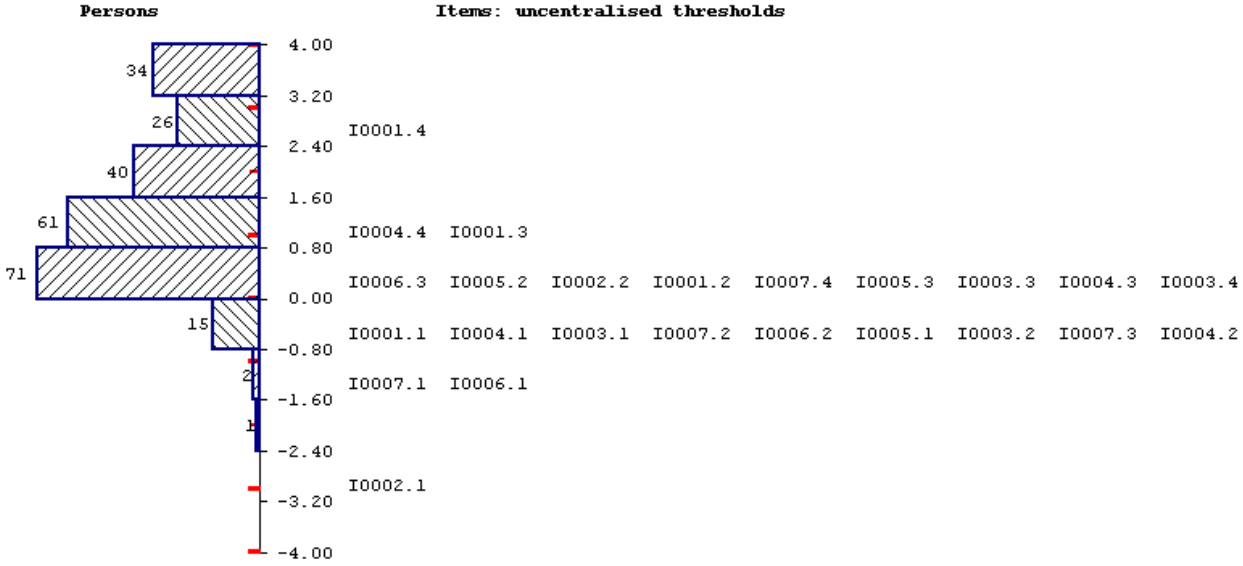
Figure 2. Physical well-being – disorder in item 2



Recode item 2 as 01112 (combined middle three categories)
Fit
Model fit p = .002
Item fit = 1.52
Person fit = 0.94

This recode fixed the disorder in item 2. Overall item and person fit changed very little.

Figure 3. Item map, Physical well-being



Social well-being

Summary: Four of the six items exhibit DIF, items 9 and 10 exhibit local dependency and all items exhibit disordered thresholds. Overall fit to the Rasch model was good.

Table 2. Item criteria from Rasch –Social well-being

	Item		Location	Item fit	DIF[a]	LD[b]	DT[c]
Items 8-13							
Model fit *p* = .031	8	closefriends	0.381	0.545	sex	-	x
Item fit = 0.73	9	suppfam	-0.089	-1.564	-	10	x
Person fit = 1.09	10	suppfriend	-0.03	-0.104	stage	9	x
	11	acceptfam	-0.376	0.252	-	-	x
	12	commfam	0.215	-0.435	stage	-	x
	13	partner	-0.101	-0.294	sex	-	x

* Statistics for model with item 5 rescored with middle two categories pooled

[a] DIF = differential item functioning; cell contains the name of the variable for which DIF was observed

[b] LD = local dependency; cell contains the number of the item with which there is a correlation

[c] DT = disordered threshold; x indicated the presence of a disordered threshold

"-" indicates the absence of DIF, LD or DT

Figure 4. Social well-being – disorder in item 8

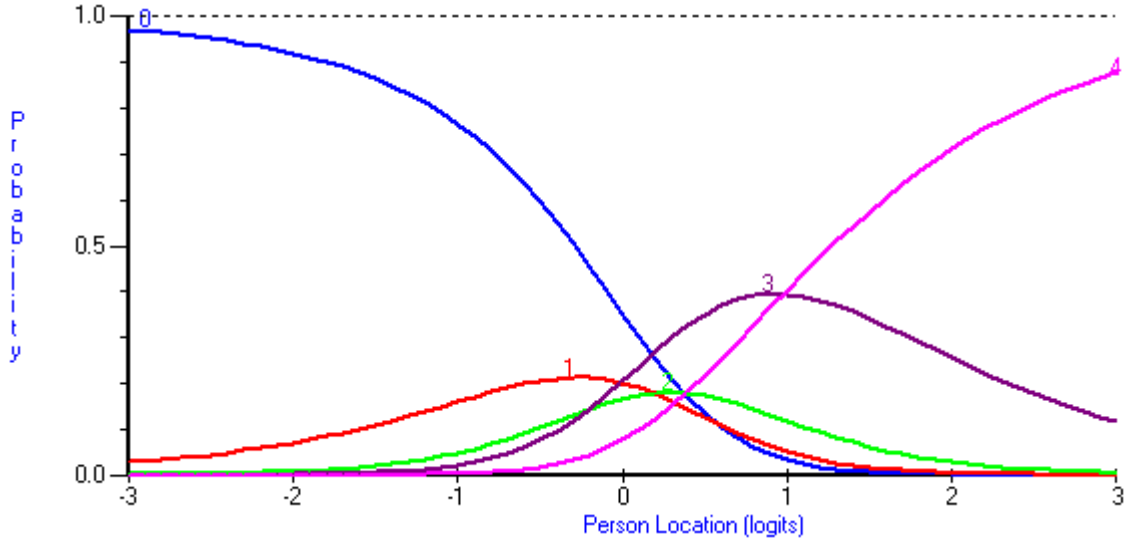


Figure 5. Social well-being – disorder in item 9

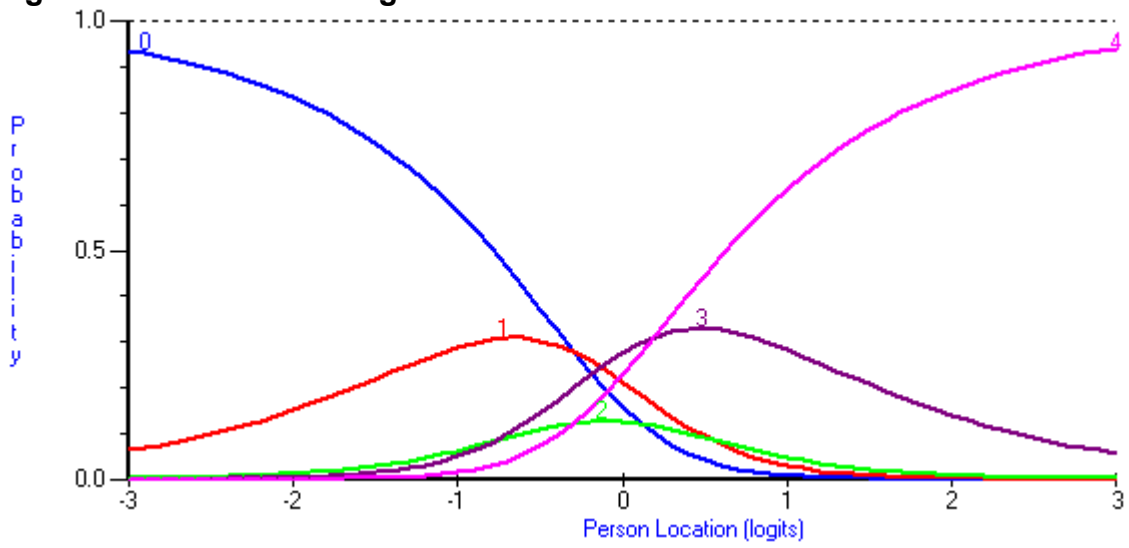


Figure 6. Social well-being – disorder in item 10

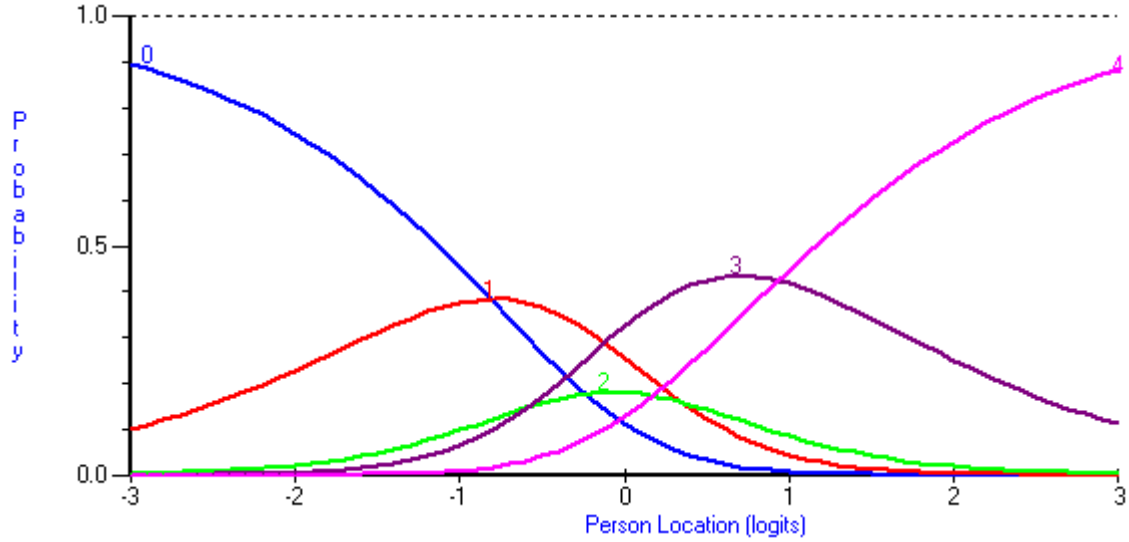


Figure 7. Social well-being – disorder in item 11

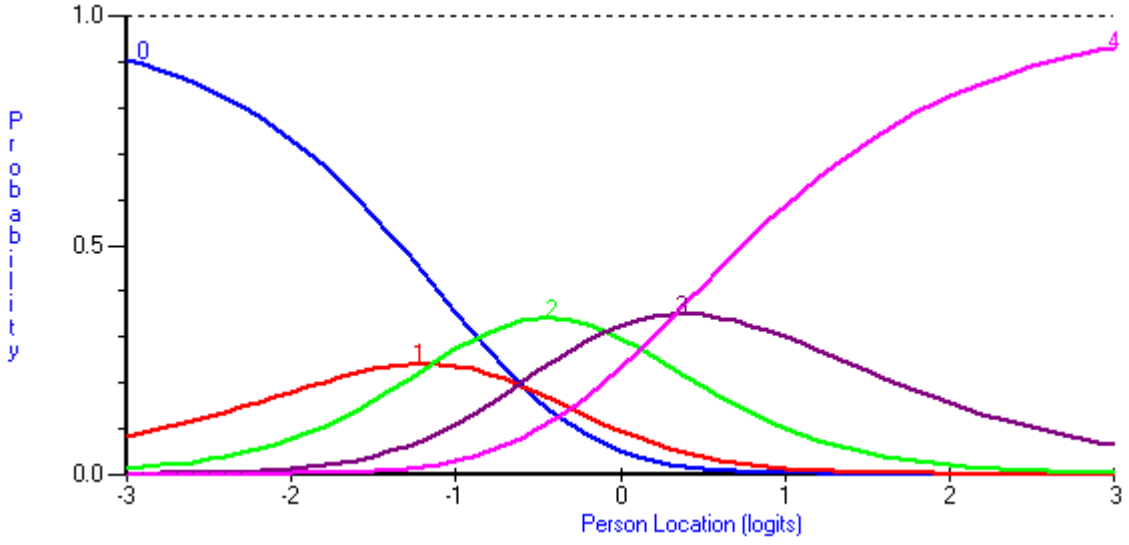


Figure 8. Social well-being – disorder in item 12

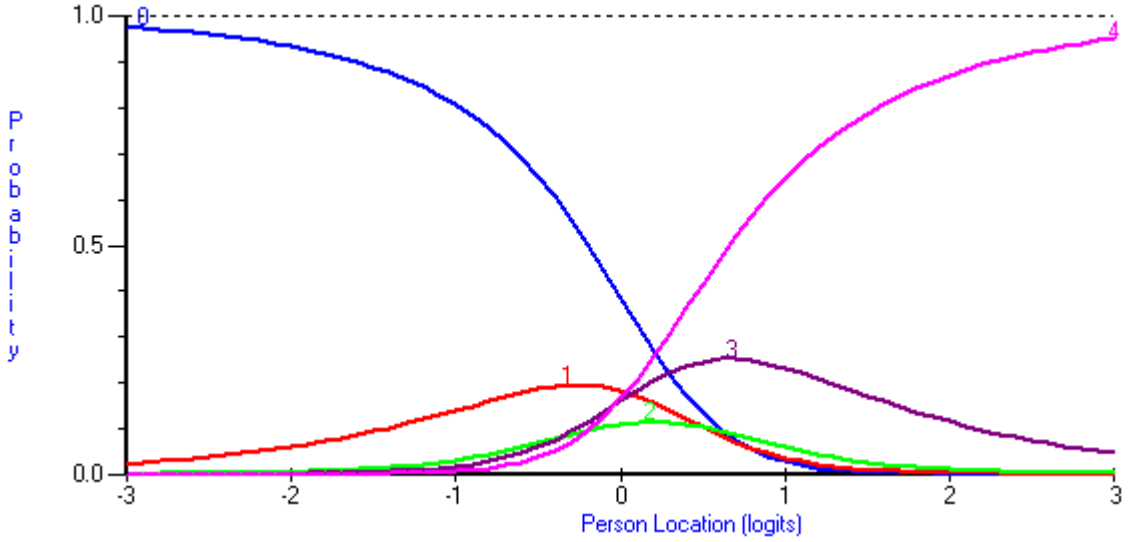
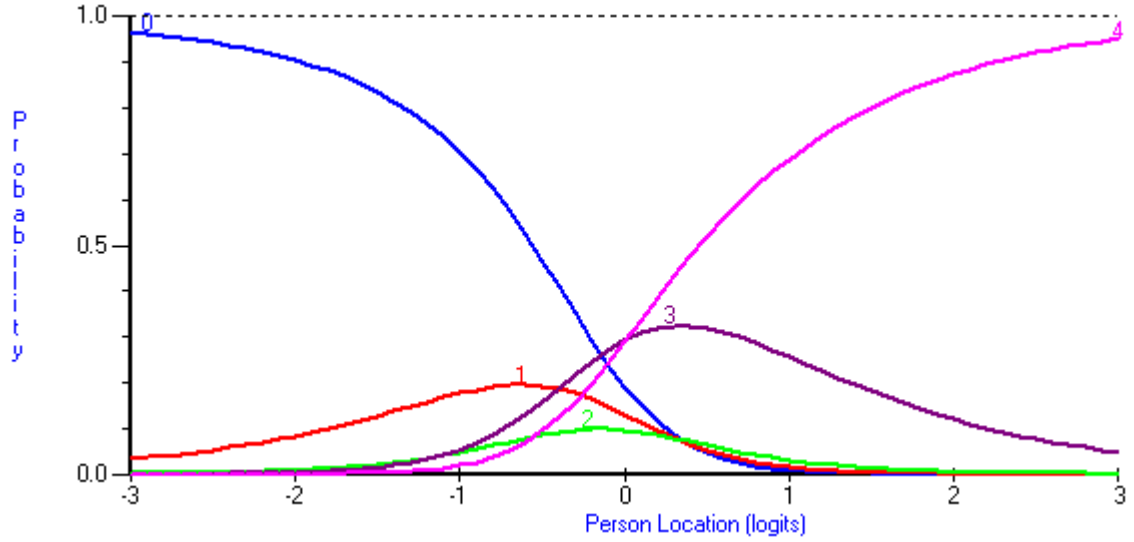


Figure 9. Social well-being – disorder in item 13



Combining the middle three response categories for all of the items in this domain may fix the disorder.

Recode all items (8-13) as 01112 (combined middle three categories)

Fit
Model fit p = .50
Item fit = 1.01
Person fit = 1.31

This recode fixed all disordered thresholds, and overall fit is still good.

Figure 10. Thresholds – Social well-being (recoded)

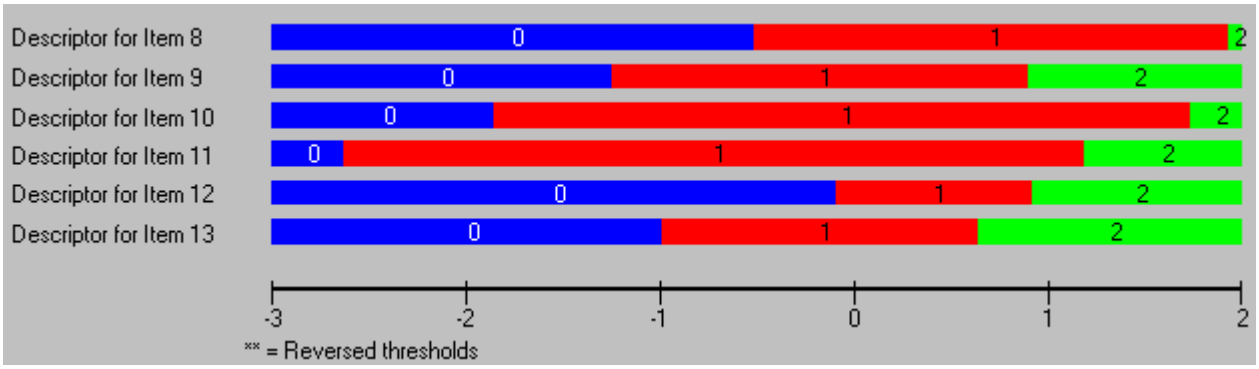
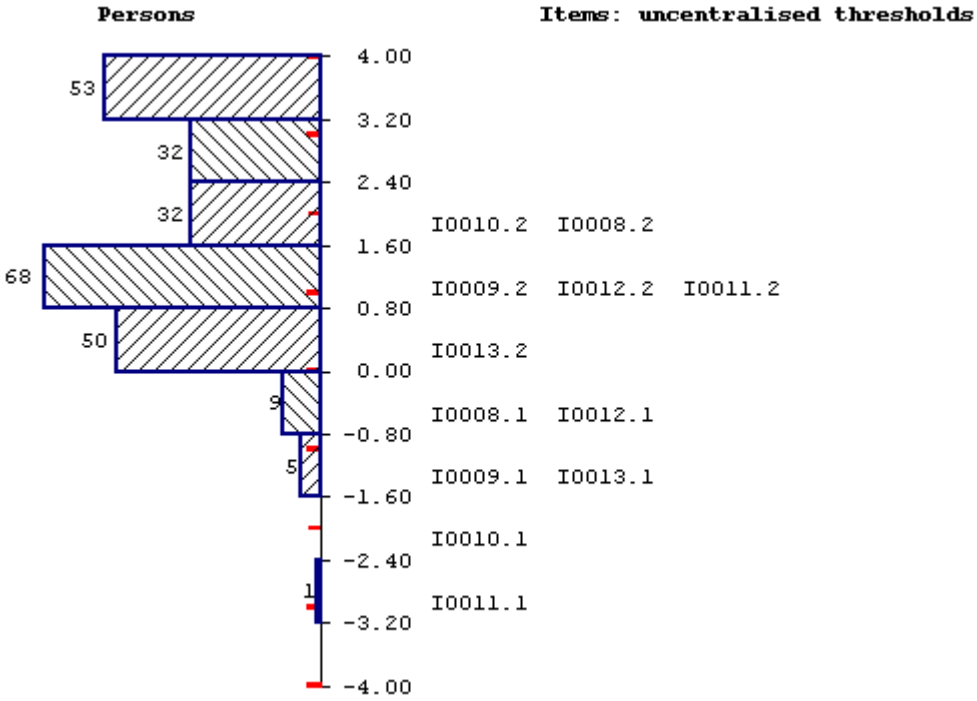


Figure 11. Item map, Social well-being



DIF

When the items were recoded as described above, the following instance of DIF were observed: item 10 by stage, item 12 by stage and item 13 by sex. These are shown graphically below.

Figure 12. DIF by stage (blue = early, red = late) for item 10 (suppfriend)

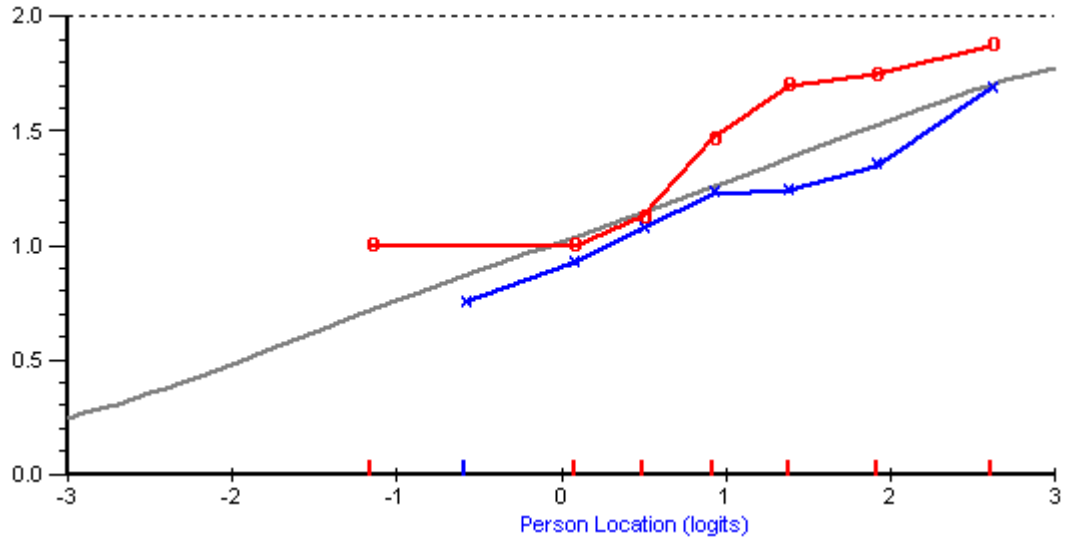


Figure 13. DIF by stage (blue = early, red = late) for item 12 (commfam)

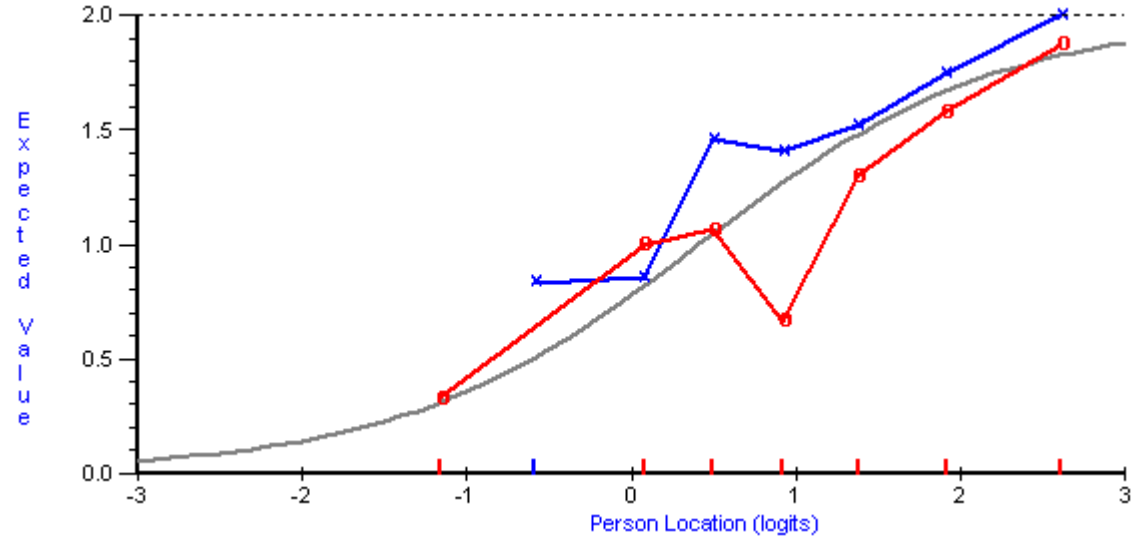
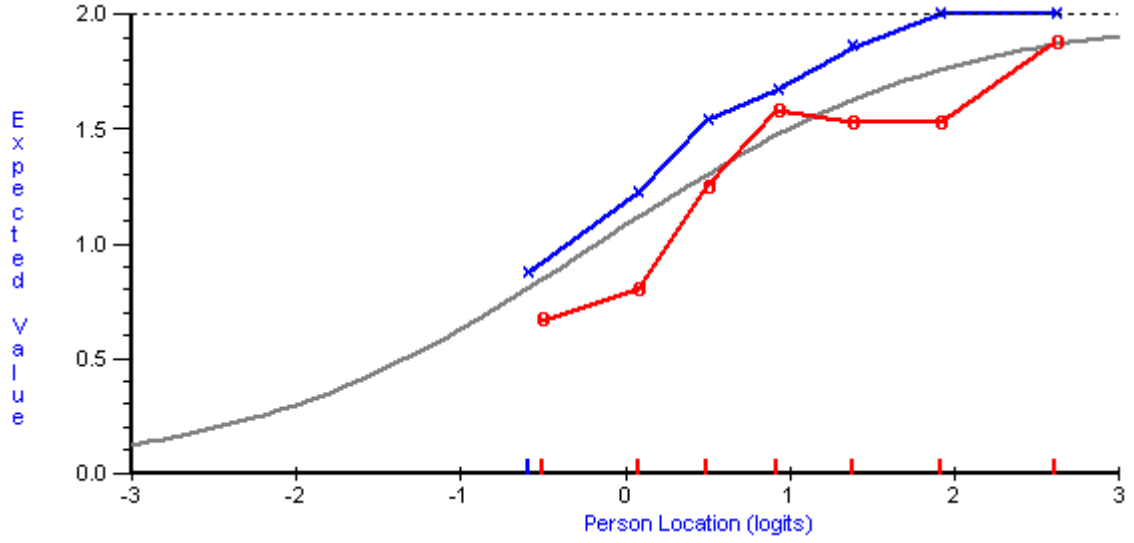


Figure 14. DIF by sex (blue = male, red = female) for item 13 (partner)



Emotional well-being

Summary: Model, item and person fit were good, although all but one of the item thresholds exhibited disorder. There was local dependence between items 18 and 19, and not item exhibited DIF.

Table 3. Item criteria from Rasch –Emotional well-being

	Item		Location	Item fit	DIF[a]	LD[b]	DT[c]
Items 14-19							
Model fit *p* = .017	14	sad	0.095	-0.445	-	-	x
Item fit = 1.44	15	coping	0.194	2.37	-	-	x
Person fit = 1.05	16	hope	-0.5	-0.148	-	-	x
	17	nervous	-0.167	0.535	-	-	-
	18	dying	-0.075	-2.075	-	19	x
	19	worse	0.454	0.07	-	18	x

* Statistics for model with item 5 rescored with middle two categories pooled

[a] DIF = differential item functioning; cell contains the name of the variable for which DIF was observed

[b] LD = local dependency; cell contains the number of the item with which there is a correlation

[c] DT = disordered threshold; x indicated the presence of a disordered threshold

"-" indicates the absence of DIF, LD or DT

Figure 15. Emotional well-being – disorder in item 14

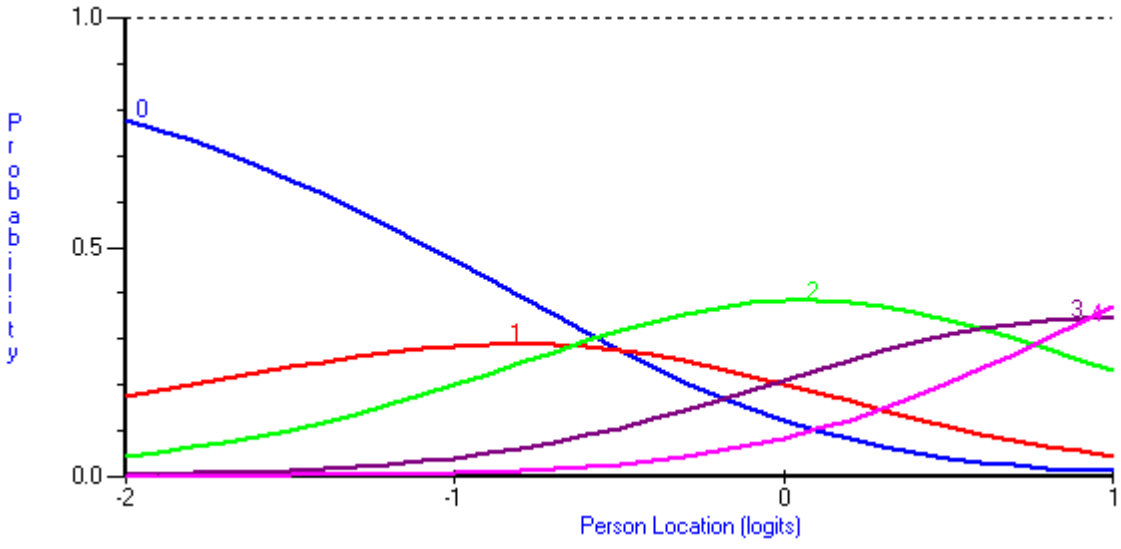


Figure 16. Emotional well-being – disorder in item 15

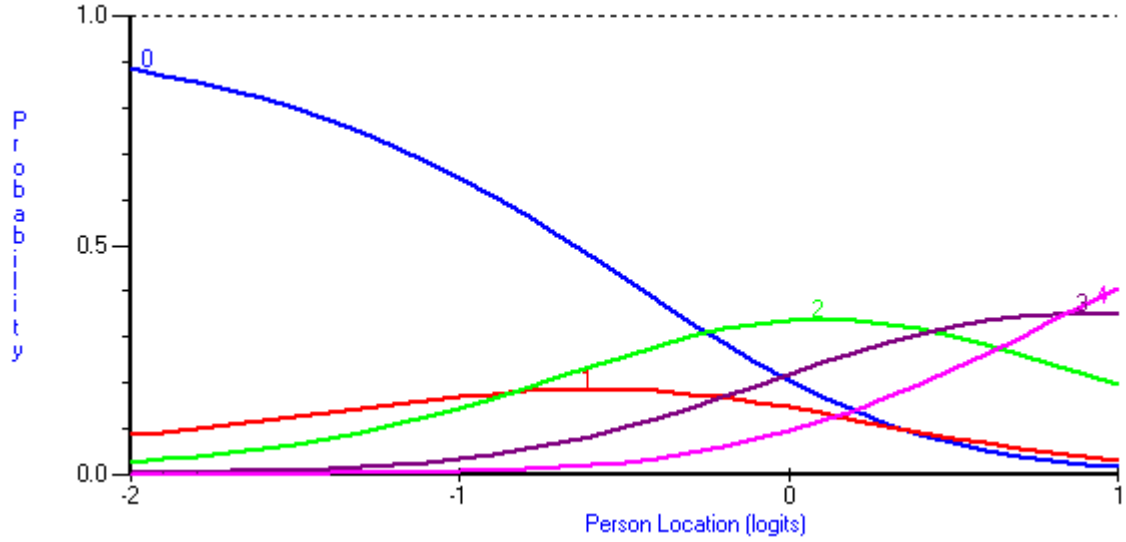


Figure 17. Emotional well-being – disorder in item 16

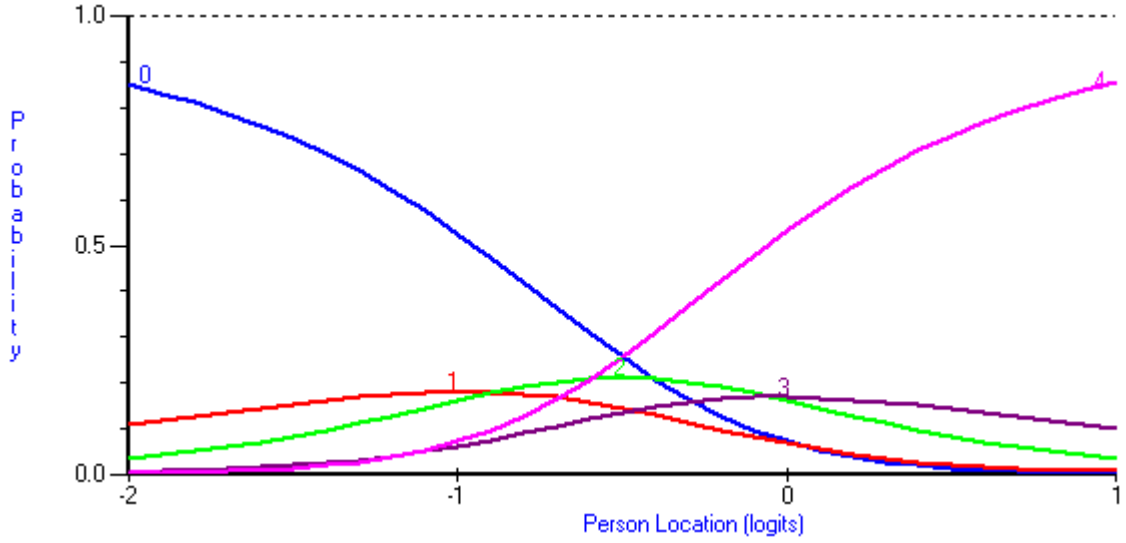


Figure 18. Emotional well-being – disorder in item 18

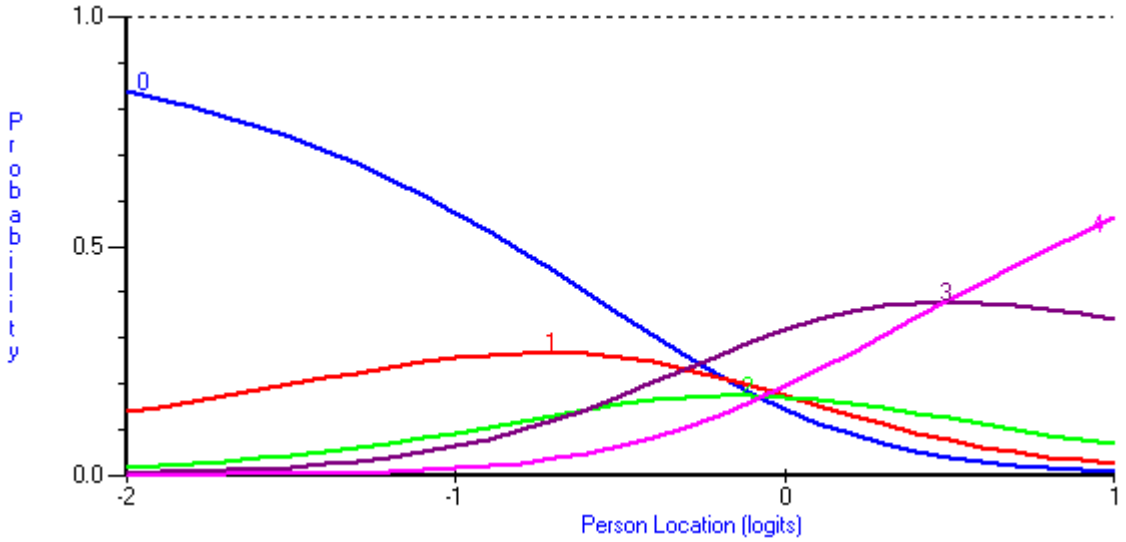
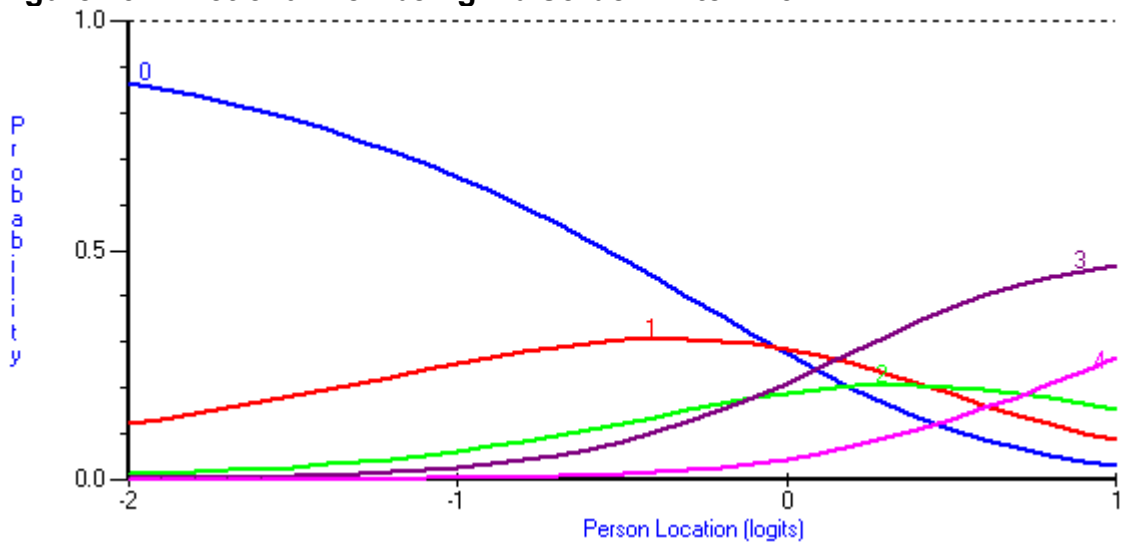


Figure 19. Emotional well-being – disorder in item 19



Recode items 14 and 16 as 01112; items 15, 18 and 19 as 01123

Fit
Model fit p = .01
Item fit = 1.21
Person fit = 1.11

This recode fixed all disordered thresholds, and overall fit is still good.

Figure 20. Thresholds – Emotional well-being (recoded)

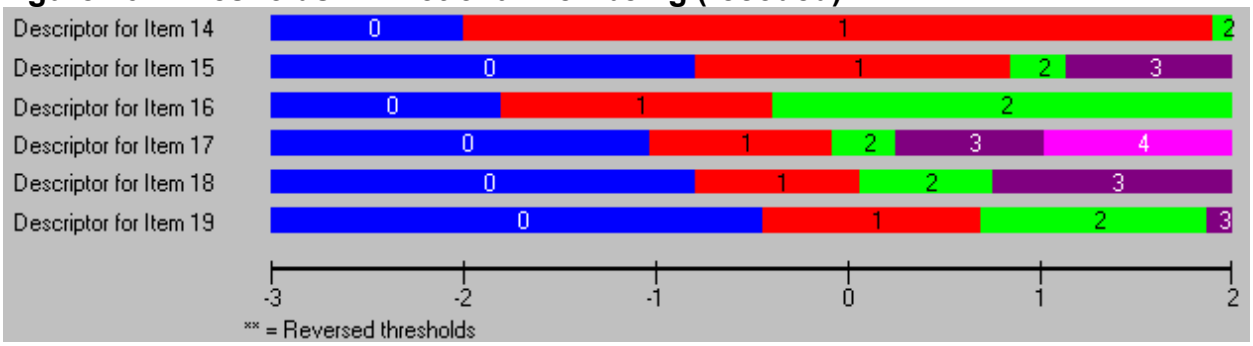
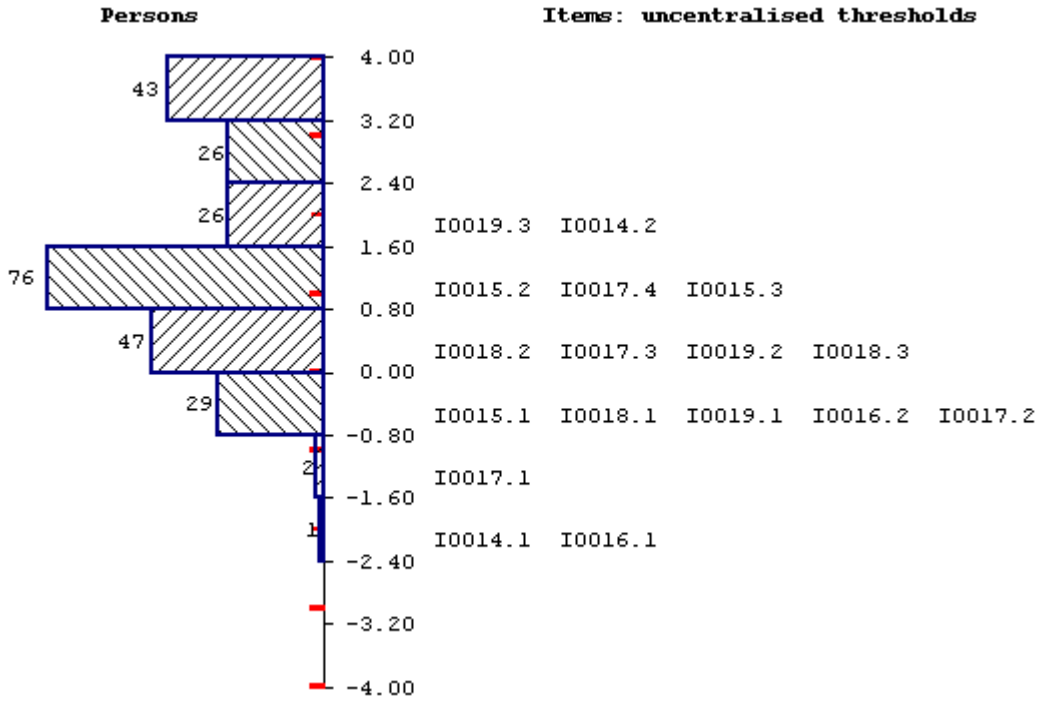


Figure 21. Item map, Emotional well-being



Functional well-being

Summary: Model and item were poor, and person fit were good. Item 24 (sleep) exhibited misfit. There was local dependence between items 20 and 21, and between items 22, 25 and 26, and three items exhibited disordered thresholds. Item 26 exhibited DIF.

Table 4. Item criteria from Rasch – Functional well-being

	Item		Location	Item fit	DIF[a]	LD[b]	DT[c]
Items 20-26							
Model fit *p* = .000	20	work	0.316	0.832	-	21	-
Item fit = 2.19	21	fulfil	0.229	0.434	-	20	-
Person fit = 1.10	22	enjoy	-0.427	-2.583	-	25, 26	-
	23	accept	-0.591	1.357	-	-	x
	24	sleep	0.134	3.703	-	-	-
	25	fun	0.239	-2.167	-	22, 26	x
	26	qol	0.101	-1.029	stage	22, 25	x

* Statistics for model with item 5 rescored with middle two categories pooled
[a] DIF = differential item functioning; cell contains the name of the variable for which DIF was observed
[b] LD = local dependency; cell contains the number of the item with which there is a correlation
[c] DT = disordered threshold; x indicated the presence of a disordered threshold
"-" indicates the absence of DIF, LD or DT

Figure 22. Functional well-being – disorder in item 23

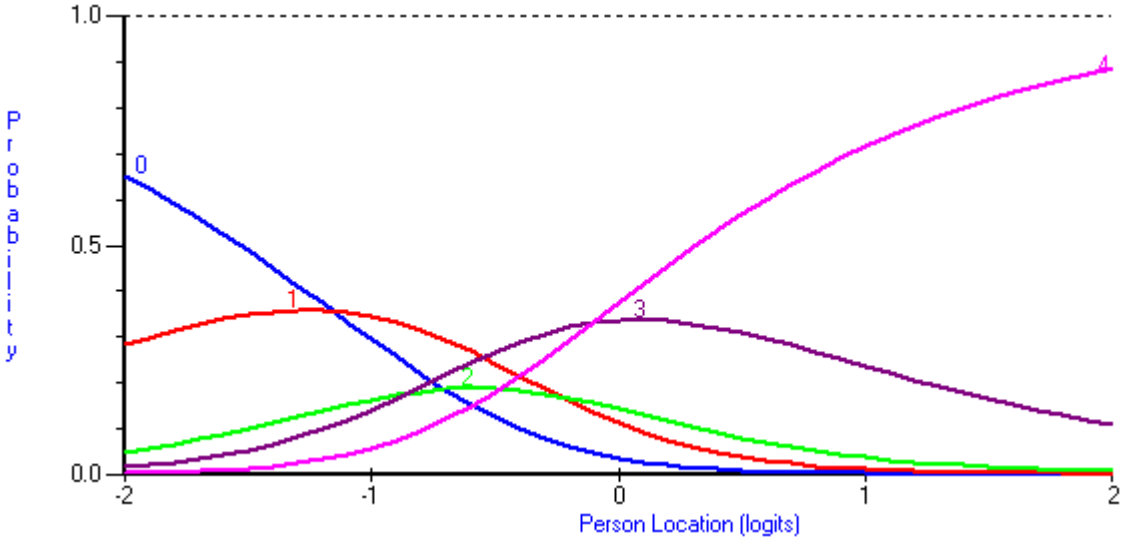


Figure 23. Functional well-being – disorder in item 25

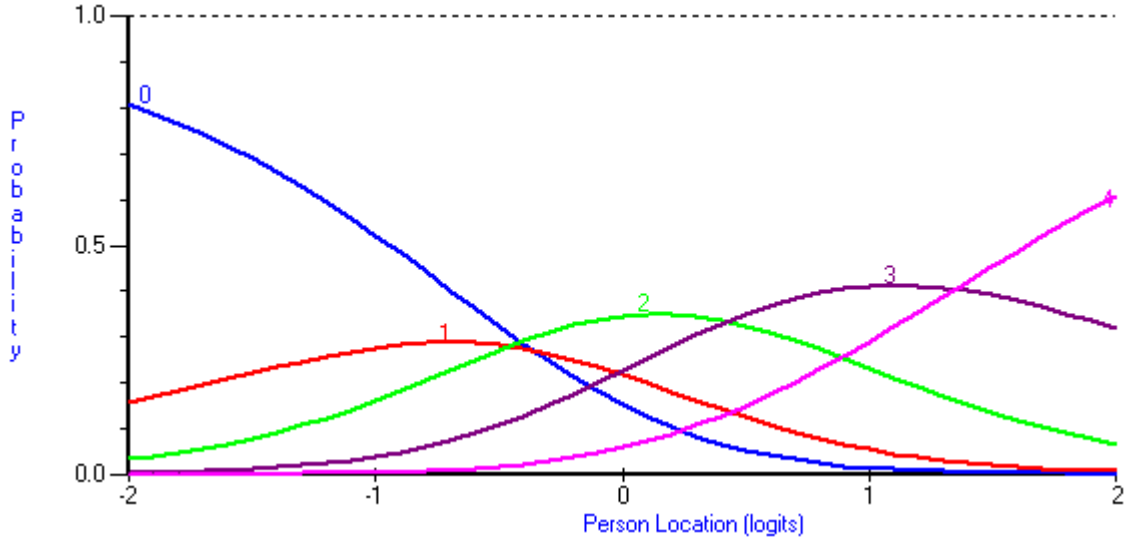
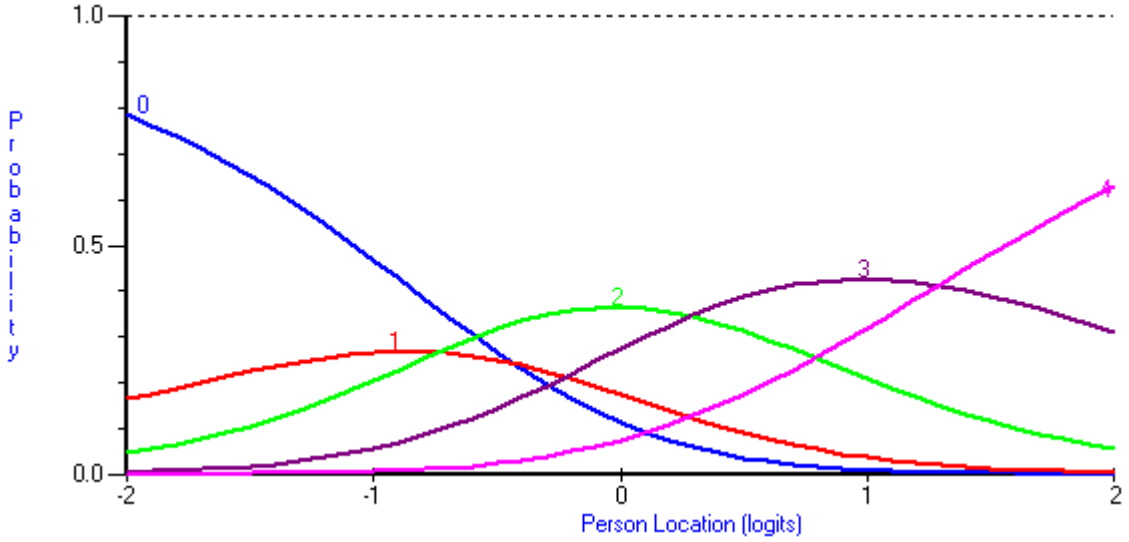


Figure 24. Functional well-being – disorder in item 26



Recode items 23 as 01112, and items 25 and 26 as 01123 + remove item 24

Fit
Model fit p = .317
Item fit = 1.487
Person fit = 1.191

This recode fixed all disordered thresholds, and with the removal of item 24 overall fit is good.

Figure 25. Thresholds – Functional well-being (recoded)

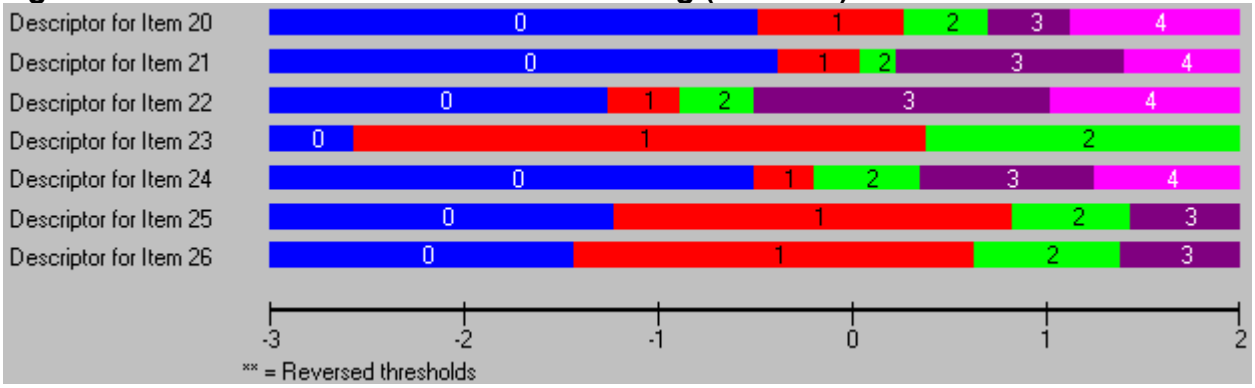


Figure 26. Item map, Functional well-being

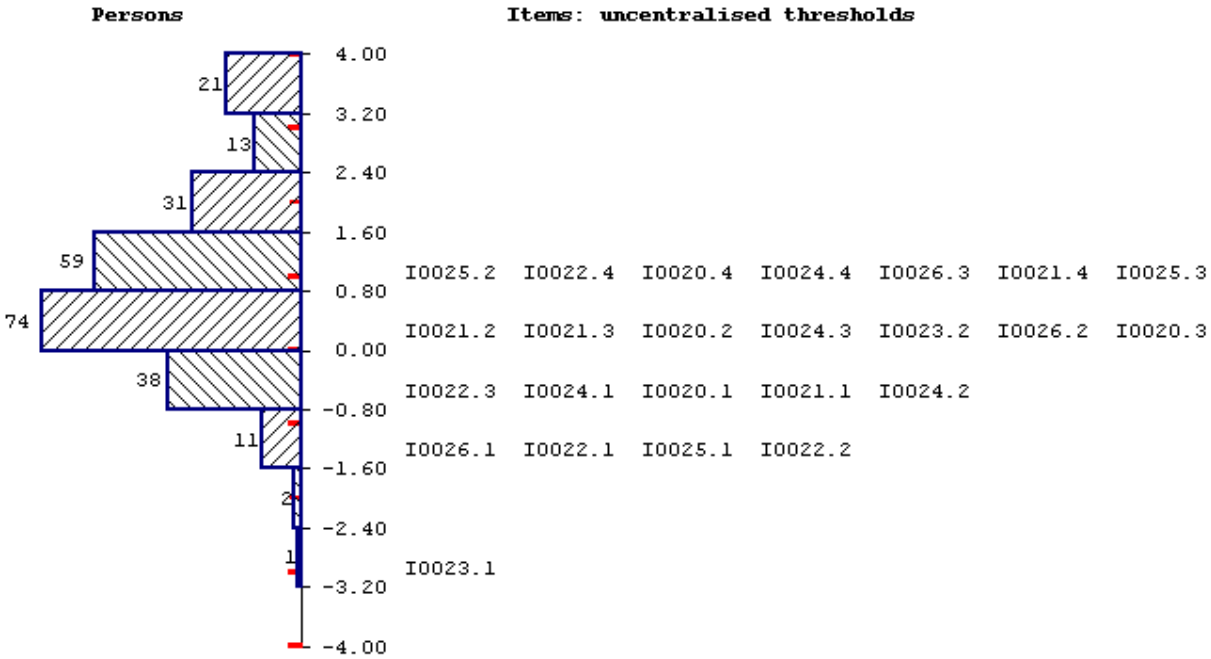
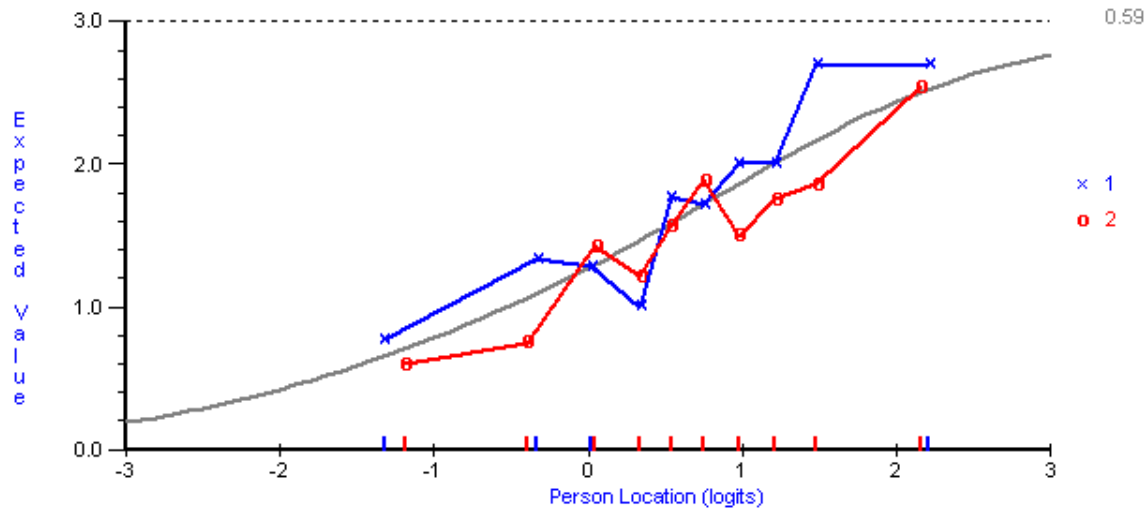


Figure 27. DIF by stage (blue = early, red = late) for item 26 (qol)



Appendix Report B FACT-8D utility algorithm and scoring instructions

This appendix contains instructions for calculating FACT-8D utility scores from FACT-G responses, whether collected from the FACT-G or any related FACIT questionnaire containing FACT-G items.

STATA and SPSS syntax to implement this utility scoring algorithm with Australian utility decrements is also provided.

Australian utility decrements are included in Table 1 and the STATA and SPSS syntax, but the general principles hold for utility sets from other countries.

Instructions

For any patient *p* who has provided responses to the nine FACT-G items in Table 1 (whether via completion of the FACT-G or any related FACIT questionnaires containing these items), that patient's FACT-8D utility score is calculated as follows.
First, determine the corresponding level *l* for each dimension *d* and the associated utility decrement (w_{dl}), following the mapping of FACT-G items levels to Australian FACT-8D utility decrements in Table 1 (estimated from the Australian general population).

FACT-8D scoring algorithm

A utility score of 1 is assigned to patients whose FACT-G scores indicate they are at level 1 of all 8 dimensions of the FACT-8D. 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 8 FACT-8D dimensions.

$$\text{FACT-8D}_p = 1 - \sum_{d=1}^{8} w_{dl} \mid \text{FACT-8D}_{dlp}$$

For example, a health state with *quite a bit of pain*, *somewhat lacking energy, not at all able to work*, *feeling a little bit sad*, getting *very much emotional support from family* and *a little support from friends, sleeping very (much) well*, *no nausea*, and *not at all worried that condition will worsen*, would be valued at 1 minus the decrements for *Pain* level 4, *Fatigue* level 3, and *Work* level 5 = 1 – 0.186 – 0.056 – 0.185) = 0.573. The best possible health state has a value of 1, and the worst possible state has a value of -0.54 (1 – 0.398 – 0.130 – 0.282 – 0.112 – 0.185 – 0.176 – 0.134 – 0.132).
STATA and SPSS syntax to implement this utility scoring algorithm is provided below.

Table 1 FACT-8D descriptive system: how the dimensions and levels map to the 9 component FACT-G items, and associated Australian utility decrements

FACT-8D Dimension (*d*) *FACT-G question*	**FACT-G item**	**FACT-G item level** (*l*) **and associated utility decrement** (w_{dl})				
		Level 1 **BEST**	**Level 2**	**Level 3**	**Level 4**	**Level 5** **WORST**
Pain *I have pain*	GP4	Not at all 0	A little bit -0.047	Somewhat -0.085	Quite a bit -0.186	Very much -0.398
Fatigue (lack of energy) *I have a lack of energy*	GP1	Not at all 0	A little bit 0	Somewhat -0.056	Quite a bit -0.130	Very much -0.130
Nausea *I have nausea*	GP2	Not at all 0	A little bit -0.091	Somewhat -0.104	Quite a bit -0.195	Very much -0.282
Sleep *I am sleeping well*	GF5	Very much 0	Quite a bit 0	Somewhat 0	A little bit -0.112	Not at all -0.112
Work *I am able to work (include work at home)*	GF1	Very much 0	Quite a bit -0.051	Somewhat -0.051	A little bit -0.087	Not at all -0.185
Support* *I get emotional support from my family and support from my friends*	GS2, GS3	Very much 0	Quite a bit -0.009	Somewhat -0.009	A little bit -0.104	Not at all -0.176
Sadness *I feel sad*	GE1	Not at all 0	A little bit 0	Somewhat -0.070	Quite a bit -0.111	Very much -0.134
Worry my health will get worse *I worry that my condition will get worse*	GE6	Not at all 0	A little bit -0.087	Somewhat -0.087	Quite a bit -0.103	Very much -0.132

*For the Support dimension, take the better of the two items.

STATA syntax

STATA code to calculate FACT-8D utility scores from FACT-G responses using Australian utility set

Written by Richard Norman richard.norman@curtin.edu.au
9th December 2019

```
* This code is designed to convert FACT-G responses into FACT-8D utility weights.
* It uses the Australian DCE-derived weights developed by [removed to prevent
* unblinding during peer-review].
* It is based on the assumption that the underlying data are coded between 0 and 4
* where 0 means 'Not at all', 1 means 'A little bit', 2 means 'Somewhat',
```

```
* 3 means 'Quite a bit', and 4 means 'Very much'. The coding of the variables
* is clustered by domain, so Physical Well-Being items are labelled GP1-GP7,
* Social / Family Well-Being items are labelled GS1-GS7, Emotional Well-Being items
* are labelled GE1-GE6, and Functional Well-Being are labelled GF1-GF7.

gen pai = gp4
gen fat = gp1
gen nau = gp2
gen sle = 4-gf5
gen wrk = 4-gf1
gen sup = 4 - max(gs2,gs3)
gen sad = ge1
gen wor = ge6

gen paidec=.
replace paidec=0 if pai==0
replace paidec=-0.047 if pai==1
replace paidec=-0.085 if pai==2
replace paidec=-0.186 if pai==3
replace paidec=-0.398 if pai==4

gen fatdec=.
replace fatdec=0 if fat==0
replace fatdec=0 if fat==1
replace fatdec=-0.056 if fat==2
replace fatdec=-0.130 if fat==3
replace fatdec=-0.130 if fat==4

gen naudec=.
replace naudec=0 if nau==0
replace naudec=-0.091 if nau==1
replace naudec=-0.104 if nau==2
replace naudec=-0.195 if nau==3
replace naudec=-0.282 if nau==4

gen sledec=.
replace sledec=0 if sle==0
replace sledec=0 if sle==1
replace sledec=0 if sle==2
replace sledec=-0.112 if sle==3
replace sledec=-0.112 if sle==4

gen wrkdec=.
replace wrkdec=0 if wrk==0
replace wrkdec=-0.051 if wrk==1
replace wrkdec=-0.051 if wrk==2
replace wrkdec=-0.087 if wrk==3
replace wrkdec=-0.185 if wrk==4

gen supdec=.
```

```
replace supdec=0 if sup==0
replace supdec=-0.009 if sup==1
replace supdec=-0.009 if sup==2
replace supdec=-0.104 if sup==3
replace supdec=-0.176 if sup==4

gen saddec=.
replace saddec=0 if sad==0
replace saddec=0 if sad==1
replace saddec=0.070 if sad==2
replace saddec=-0.111 if sad==3
replace saddec=-0.134 if sad==4

gen wordec=.
replace wordec=0 if wor==0
replace wordec=-0.087 if wor==1
replace wordec=-0.087 if wor==2
replace wordec=-0.103 if wor==3
replace wordec=-0.132 if wor==4

gen fact8d = 1 + paidec + fatdec + naudec + sledec + wrkdec + supdec + saddec +
wordec
```

SPSS syntax

SPSS code to calculate FACT-8D utility scores from FACT-G responses using Australian utility set

Written by Daniel Costa daniel.costa@sydney.edu.au
16th December 2020

```
* Encoding: UTF-8.
* Encoding: .

* This code is designed to convert FACT-G responses into FACT-8D utility weights.
* It uses the Australian DCE-derived weights developed by [removed to prevent
* unblinding during peer-review].
* It is based on the assumption that the underlying data are coded between 0 and 4
* where 0 means 'Not at all', 1 means 'A little bit', 2 means 'Somewhat',
* 3 means 'Quite a bit', and 4 means 'Very much'. The coding of the variables
* is clustered by domain, so Physical Well-Being items are labelled GP1-GP7,
* Social / Family Well-Being items are labelled GS1-GS7, Emotional Well-Being items
* are labelled GE1-GE6, and Functional Well-Being are labelled GF1-GF7.

compute pai = gp4.
compute fat = gp1.
compute nau = gp2.
compute sle = 4-gf5.
compute wrk = 4-gf1.
compute sup = 4 - max(gs2,gs3).
compute sad = ge1.
compute wor = ge6.
```

```
exe.

compute paidec=$sysmis.
if pai=0 paidec=0.
if pai=1 paidec=-0.047.
if pai=2 paidec=-0.085.
if pai=3  paidec=-0.186.
if pai=4  paidec=-0.398.

compute fatdec= $sysmis.
if fat=0 fatdec=0.
if fat=1 fatdec=0.
if fat=2 fatdec=-0.056.
if fat=3 fatdec=-0.130.
if fat=4 fatdec=-0.130.

compute naudec=$sysmis.
if nau=0 naudec=0.
if nau=1 naudec=-0.091.
if nau=2 naudec=-0.104.
if nau=3 naudec=-0.195.
if nau=4 naudec=-0.282.

compute sledec=$sysmis.
if sle=0 sledec=0.
if sle=1 sledec=0.
if sle=2 sledec=0.
if sle=3 sledec=-0.112.
if sle=4 sledec=-0.112.

compute wrkdec=$sysmis.
if wrk=0 wrkdec=0.
if wrk=1 wrkdec=-0.051.
if wrk=2 wrkdec=-0.051.
if wrk=3 wrkdec=-0.087.
if wrk=4 wrkdec=-0.185.

compute supdec=$sysmis.
if sup=0 supdec=0.
if sup=1 supdec=-0.009.
if sup=2 supdec=-0.009.
if sup=3 supdec=-0.104.
if sup=4 supdec=-0.176.

compute saddec=$sysmis.
if sad=0 saddec=0.
if sad=1 saddec=0.
if sad=2 saddec=0.070.
if sad=3 saddec=-0.111.
if sad=4 saddec=-0.134.
```

```
compute wordec=$sysmis.
if wor=0 wordec=0.
if wor=1 wordec=-0.087.
if wor=2 wordec=-0.087.
if wor=3 wordec=-0.103.
if wor=4 wordec=-0.132.

compute fact8d = 1 + paidec + fatdec + naudec + sledec + wrkdec + supdec + saddec +
wordec.
exe.
```