White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/215778/

Version: Published Version

Article:
King, M.T., Revicki, D.A., Norman, R. et al. (2024) United States value set for the functional assessment of cancer therapy-general eight dimensions (FACT-8D), a cancer-specific preference-based quality of life instrument. PharmacoEconomics - Open, 8 (1). pp. 49-63. ISSN: 2509-4262

https://doi.org/10.1007/s41669-023-00448-5

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

PharmacoEconomics - Open (2024) 8:49–63
https://doi.org/10.1007/s41669-023-00448-5

ORIGINAL RESEARCH ARTICLE

United States Value Set for the Functional Assessment of Cancer Therapy-General Eight Dimensions (FACT-8D), a Cancer-Specific Preference-Based Quality of Life Instrument

Madeleine T. King[1] · D. A. Revicki[2] · R. Norman[3] · F. Müller[4] · R.C. Viney[5] · A. S. Pickard[6] · D. Cella[7] · J. W. Shaw[8] on behalf of the MAUCa Consortium

Accepted: 11 October 2023 / Published online: 7 December 2023
© The Author(s) 2023

Abstract

Objectives To develop a value set reflecting the United States (US) general population's preferences for health states described by the Functional Assessment of Cancer Therapy (FACT) eight-dimensions preference-based multi-attribute utility instrument (FACT-8D), derived from the FACT-General cancer-specific health-related quality-of-life (HRQL) questionnaire.

Methods A US online panel was quota-sampled to achieve a general population sample representative by sex, age (≥ 18 years), race and ethnicity. A discrete choice experiment (DCE) was used to value health states. The valuation task involved choosing between pairs of health states (choice-sets) described by varying levels of the FACT-8D HRQL dimensions and survival (life-years). The DCE included 100 choice-sets; each respondent was randomly allocated 16 choice-sets. Data were analysed using conditional logit regression parameterized to fit the quality-adjusted life-year framework, weighted for socio-odemographic variables that were non-representative of the US general population. Preference weights were calculated as the ratio of HRQL-level coefficients to the survival coefficient.

Results 2562 panel members opted in, 2462 (96%) completed at least one choice-set and 2357 (92%) completed 16 choice-sets. Pain and nausea were associated with the largest utility weights, work and sleep had more moderate utility weights, and sadness, worry and support had the smallest utility weights. Within dimensions, more severe HRQL levels were generally associated with larger weights. A preference-weighting algorithm to estimate US utilities from responses to the FACT-General questionnaire was generated. The worst health state's value was −0.33.

Conclusions This value set provides US population utilities for health states defined by the FACT-8D for use in evaluating oncology treatments.

1 Introduction

Advances in cancer treatments often come with substantial costs. The value of new therapies is determined by considering health outcomes relative to healthcare costs [1]. Various health technology assessment (HTA) valuation methodologies exist [2]. Cost–utility analysis (CUA) quantifies health outcomes on a scale applicable across health conditions: the quality-adjusted life-year (QALY) [3]. In CUA, survival time is quality-adjusted using health-related quality of life (HRQL) to produce QALYs. In many countries, CUA is a preferred HTA method and the QALYs is a preferred health outcome measure [2].

In the USA, HTA is not utilized at the national level. Rather, private and public health insurers conduct internal HTAs to inform decisions about coverage and copays. Some states have HTA programs, e.g., California and Washington. The Institute for Clinical and Economic Review (ICER) [4] is an independent non-profit research organization that aims to inform pricing using a value assessment framework that incorporates clinical effectiveness, QALYs, other potential benefits and disadvantages, contextual considerations and incremental cost-effectiveness [5].

Utility is the HRQL weighting metric used to calculate QALYs. It has two anchors: full health, anchored at 1 (the maximum value), and being dead, anchored at zero [3]. Health states valued below zero are regarded as worse than being dead. To determine all possible values on the utility metric requires a valuation study, involving what is being valued (e.g. the health states relevant to the interventions under consideration), whose preferences are elicited (e.g. patients with the health condition, the general public), the

Extended author information available on the last page of the article

Key Points for Decision Makers

The Functional Assessment of Cancer Therapy–General (FACT-G), a patient-reported questionnaire, is commonly used to assess quality of life in oncology trials in the US and elsewhere.

This paper describes the development of a preference-weighted algorithm which reflects the values of the US general population and enables direct calculation of health utilities from FACT-G data and the many Functional Assessment of Chronic Illness Therapy (FACIT) questionnaires that contain the FACT-G items.

Given the widespread use of the FACIT questionnaires, this algorithm can facilitate incorporation of common symptoms and impacts of cancer into health technology assessment in the USA.

valuation technique used to elicit preferences and how it is administered [3]. Traditional valuation techniques include standard gamble, time trade-off and rating scales [3, 7]. Discrete choice experiments (DCEs) have been used more recently [8]. HRQL questionnaires that yield utility scores by combining dimensions of HRQL into a single index score using preference weights derived from a valuation study are called preference-based measures [9], as distinct from profile measures which simply summarize self-reported responses as separate HRQL dimension scores [9]. Multi-attribute utility instruments (MAUIs) are preference-based measures that assess HRQL with a relatively small number of questions, each describing a HRQL dimension, and generate a preference-based utility score for each health state defined by the MAUI via a preference-weighted algorithm [10].

The Functional Assessment of Cancer Therapy Eight Dimensions (FACT-8D) [11] is a MAUI derived from the cancer-specific HRQL profile measure, the Functional Assessment of Cancer Therapy General (FACT-G) [12]. The FACT-G is widely used in oncology clinical trials [13], either as a stand-alone questionnaire or included within many of the Functional Assessment of Chronic Illness Therapy (FACIT) questionnaires. The FACT-8D enables quantification of utility scores from FACT-G data and captures dimensions reflecting symptoms and impacts of cancer and its treatments that are not typically included in generic preference-based measures, specifically nausea, fatigue, sleep problems and worry about future health. While mapping algorithms are available to derive utilities from FACT-G responses via mapping to generic MAUIs [14], the FACT-8D is considered theoretically and empirically stronger because it comprises a descriptive system and a valuation method that complies with the Checklist for Reporting Valuation Studies of Multi-Attribute Utility-Based Instruments [10].

For use in HTA, QALYs should reflect the societal preferences of the country the HTA addresses. To date, FACT-8D value sets have been published for Australia [11] and Canada [15]. The aim of this paper was to develop a FACT-8D value set to inform HTA in the USA.

2 Methods

This research was conceived, designed and conducted by the Multi-Attribute Utility in Cancer (MAUCa) Consortium. The University of Sydney Human Research Ethics Committee approved MAUCa's program of research (No. 13207). The US valuation study was deemed exempt from US Institutional Review Board (IRB) by the Advarra IRB (CA209-466C, Pro00032061).

2.1 Survey Sampling and Implementation

A cross-sectional population-based survey conducted between 18 February 2019 and 9 April 2019 collected sociodemographic and health status data, with a discrete choice experiment (DCE) included as the valuation component. SurveyEngine, a company specializing in online choice experiments, managed sample recruitment (via a US online panel), survey administration and data collection. SurveyEngine and its panel provider complied with the International Code on Market, Opinion and Social Research and Data Analytics [16]. Members of the online panel were eligible if 18 years or older and able to read English. Online panellists received an email invitation, including a hyperlink to the survey. Any who attempted to enter the survey via mobile phones were excluded, as the DCE was too complex for a small screen. Consent was sought from those who successfully entered the survey, and those who consented were screened for quota sampling to ensure that the age, sex, race and ethnicity distributions of the sample approximated those of the US general population, per US 2010 Census data. Those who completed the survey were awarded panel points (approximate value 1 USD).

2.2 FACT-8D Dimensions and Levels

The FACT-8D has eight dimensions: pain, fatigue, nausea, sleep, work, support, sadness and worry, derived from nine FACT-G items. Table 1 shows how the FACT-8D dimensions and levels map to the corresponding FACT-G source items. Note that all the FACT-G source items have

Table 1 Mapping between the FACT-8D descriptive system (dimensions and levels), FACT-G items, and attributes and levels in the discrete choice experiment

FACT-8D dimension	DCE attribute wording	FACT-8D level	Descriptor	FACT-G item[a] scores[b]
Pain	Pain	0	Not at all	GP4 = 0
		1	A little bit	GP4 = 1
		2	Somewhat	GP4 = 2
		3	Quite a bit	GP4 = 3
		4	Very much	GP4 = 4
Fatigue	Fatigue	0	Not at all	GP1 = 0
		1	A little bit	GP1 = 1
		2	Somewhat	GP1 = 2
		3	Quite a bit	GP1 = 3
		4	Very much	GP1 = 4
Nausea	Nausea	0	Not at all	GP2 = 0
		1	A little bit	GP2 = 1
		2	Somewhat	GP2 = 2
		3	Quite a bit	GP2 = 3
		4	Very much	GP2 = 4
Sleep[c]	Problems sleeping	0	Not at all	GF5 = 4
		1	A little bit	GF5 = 3
		2	Somewhat	GF5 = 2
		3	Quite a bit	GF5 = 1
		4	Very much	GF5 = 0
Work[c]	Problems doing work, including work at home	0	Not at all	GF1 = 4
		1	A little bit	GF1 = 3
		2	Somewhat	GF1 = 2
		3	Quite a bit	GF1 = 1
		4	Very much	GF1 = 0
Support[c,d]	Problems with support from my family and/or friends	0	Not at all	max(GS2, GS3) = 4
		1	A little bit	max(GS2, GS3) = 3
		2	Somewhat	max(GS2, GS3) = 2
		3	Quite a bit	max(GS2, GS3) = 1
		4	Very much	max(GS2, GS3) = 0
Sadness	Sadness	0	Not at all	GE1 = 0
		1	A little bit	GE1 = 1
		2	Somewhat	GE1 = 2
		3	Quite a bit	GE1 = 3
		4	Very much	GE1 = 4
Worry my health will get worse	Worry my health will get worse	0	Not at all	GE6 = 0
		1	A little bit	GE6 = 1
		2	Somewhat	GE6 = 2
		3	Quite a bit	GE6 = 3
		4	Very much	GE6 = 4

[a]Nine FACT-G items are included in the FACT-8D: GP4 (I have pain), GP1 (I have a lack of energy), GP2 (I have nausea), GF5 (I am sleeping well), GF1 (I am able to work including work at home), GS2 (I get emotional support from my family) and GS3 (I get support from my friends), GE1 (I feel sad) and GE6 (I worry that my condition will get worse)

[b]FACT-G item scores correspond to the following response options: 0 (not at all), 1 (a little bit), 2 (somewhat), 3 (quite a bit) and 4 (very much)

[c]Because the FACT-G items that determine the FACT-8D dimensions sleep, work and support are positively framed, reverse scoring is required so that FACT-8D Level 0 represents the best score and Level 4 represents the worst score across all dimensions

[d]The FACT-8D support dimension contains two items; the FACT-8D level allocated is the maximum score of the FACT-G items GS2 and GS3, i.e., the best level of support, whether from family or friends

five response options: not at all (0), a little bit (1), somewhat (2), quite a bit (3) and very much (4). Note also that the FACT-G items for pain, fatigue, nausea, sadness and worry are all negatively framed, so a higher score indicates more symptoms, sadness or worry. Because the remaining three FACT-G items that determine the FACT-8D dimensions sleep, work and support are positively framed, reverse scoring is required so that FACT-8D Level 0 represents the best score and Level 4 represents the worst score across all dimensions. Note also that the FACT-8D support dimension contains two items; the FACT-8D level allocated is the best level of support, whether from family or friends. The FACT-8D describes over 390,000 possible health states (5^8 = 390,625).

2.3 Discrete Choice Experiment

A DCE was used to generate preference data that were used to estimate the parameters of the US FACT-8D preference-weighting algorithm. Following methods previously developed for FACT-8D valuation [11], the DCE contained nine attributes: the eight FACT-8D dimensions and survival duration. Table 1 shows how the levels in the DCE mapped to the FACT-8D descriptive system and corresponding FACT-G source items. In the DCE, survival duration had four levels: 1, 2, 5 and 10 years. These levels were suggested by the oncologist members of the MAUCa Consortium, based on the rationale that 1, 2, 5 and 10 years are common survival goal posts for patients and clinicians, and are commonly used as time-points for survival endpoints in clinical trials, with the shorter durations (1 years, 2 years) applying to advanced cancers and the longer durations applying to early stages of cancer (5 years, 10 years).

Designing the experiment for the DCE involved selecting pairs of health states (choice-sets) that optimized statistical efficiency in estimating the utility model parameters. We selected a C-efficient approach as our data analysis focused on the ratios of coefficients in the conditional logit model, as this was the most appropriate focus for our analysis purpose (Sect. 2.5). The DCE experimental design comprised 100 choice-sets that optimized statistical efficiency in estimating the utility model parameters. Each choice-set comprised two FACT-8D health states, each described by nine attributes (eight HRQL dimensions and duration). We simplified the cognitive task by constraining the number of attributes that differed between health states in any choice-set to five, in line with the typical number of attributes in DCEs used to develop preference-based measure value sets [17]. We decided to vary four HRQL dimensions and duration, and used a method devised by Bleimer to determine which to vary in each choice-set [18]. We generated random choice-sets, keeping only those with exactly five dimensions differing, until we had 10,000 choice-sets that met this criterion. We then used *Ngene* version 1.3 [19, 20], a software for designing experiments, to select 100 choice-sets that optimized the experimental design's C-efficiency using a modified Fedorov algorithm with duration as the denominator [21]. Small non-zero priors were used to indicate that levels within each dimension were logically ordered. Table A (Supplement 1) contains the final experimental design.

The valuation task required participants to consider 16 pairs of hypothetical health states (i.e. 16 choice-sets), described as 'Option A' and 'Option B' (Fig. 1), and for each choice-set, select the health state they would prefer to live in until death. Dimensions that differed between Options A and B were highlighted in yellow, a presentation format preferred by participants in our previous DCE valuation methods experiment [22].

There were two levels of randomization in the DCE component of the survey: (1) each respondent was allocated 16 randomly selected choice-sets (without replacement) from the 100 in the DCE design, and (2) which option was seen as Option A or Option B was randomized within each choice-set. The nine DCE attributes were always presented in the same order, as previous work showed that order does not systematically bias preference weights [23].

2.4 Other Survey Content

The survey included sociodemographic characteristics, the FACT-G [12] and self-reported general health (assessed by a single question commonly used in national surveys in the USA [24]). The order of survey components is shown in Fig. 2. After completing the DCE component, participants were asked four fixed-format questions about the difficulty and clarity of the valuation task and strategy used to choose between health states (Appendix A, Supplement 2).

2.5 Statistical Analyses

Descriptive statistics summarized sample demographics, self-reported general health and participant feedback on the DCE valuation task. Sample representativeness was assessed against US population reference data for demographics and self-reported general health using chi-square tests.

DCE data quality was assessed by tallying how many respondents chose either all A's or all B's across the choice-sets and by considering the time respondents took to complete the survey. We divided respondents into deciles based on total survey time, ran a conditional logit model on the DCE data in each decile and then graphed the Pseudo-R^2 and the number of statistically significant coefficients for

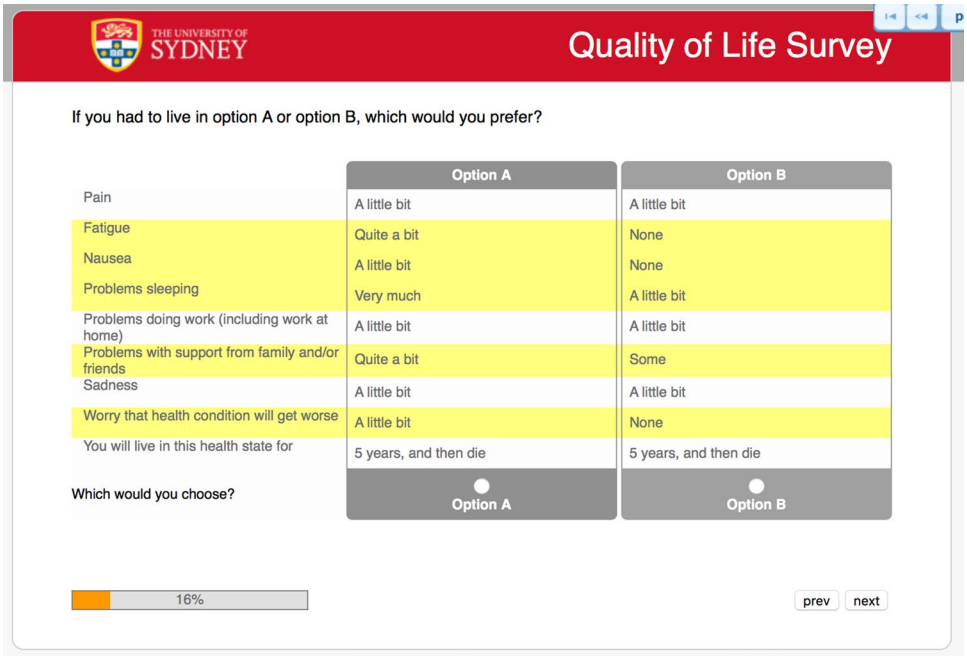


Fig. 1 An example of a choice set in the US FACT-8D discrete choice experiment. Each choice-set contained two hypothetical health states, Option A and Option B. Each health state was described in terms of levels of the eight FACT-8D quality of life domains (e.g. a little bit of pain, quite a bit of fatigue, etc.) and survival duration (e.g. you will live in this health state for 5 years, and then die). The study participant was asked to indicate which health state would they prefer

each decile, interpreting low values on either indicator as suggesting relatively low-quality data.

The DCE data were analysed with STATA 13.0 [25], using a functional form used previously to estimate utilities from DCE data consistent with standard QALY model restrictions [23, 26–29]. The QALY model requires that all health states have zero utility at death, i.e., 'the zero condition' [30, 31]. A functional form that satisfied this requirement included the interaction between the FACT-8D levels and the *TIME* variable (Eqs. 1 and 2). Therefore, as *TIME* tended to zero, the systematic component of the utility function tended to zero. Another typical requirement of the QALY model is constant proportional time trade-off [30], therefore the relationship between utility and *TIME* (life years) was assumed linear.

A useful feature of this functional form is that the impact of deviating from Level 1 (no problems) in each dimension was characterized through a two-factor interaction term with duration (the experimental design allowed for these interactions). This enabled a preference-weighting algorithm in which the effect of each level of each dimension could be included as a decrement from full health.

The DCE data were analysed in two ways, reflecting different approaches to modelling heterogeneity (Eqs. 1 and 2). The primary analysis (Eq. 1) used conditional logit models in which the utility of option j in choice-set s for survey respondent i was assumed to be:

$$U_{isj} = \alpha TIME_{isj} + \beta X_{isj}\prime TIME_{isj} + \varepsilon_{isj}$$
$$i = 1, \ldots, I \text{ respondents}; j = \text{ health state options A, B};$$
$$s = 1, \ldots, 100 \text{ choice} - \text{sets}. \quad (1)$$

where α is the utility associated with a life year, $X_{isj}\prime$ is a vector of dummy variables representing the levels of the FACT-8D health state presented in option j and β is the corresponding vector of preference weights associated with each level in each dimension within $X_{isj}\prime$, for each life year. The error term ε_{isj} was assumed to have a Gumbel distribution. To adjust the standard errors to allow for intra-individual correlation, a clustered sandwich estimator was used via STATA's *vce* (cluster) option. To estimate preference weights for each deviation from Level 1 (no problems) in each FACT-8D dimension, we divided each of the β terms by α [26], and used the delta method to estimate standard errors for these ratios [32].

Two conditional logit models were estimated. Model 1 included every decrement from the best level (i.e., Level 1, no problems) in each dimension (Eq. 1). Thus, $X_{isj}\prime$ contained 32 terms (8 dimensions × [5-1] levels within each). Non-monotonicity in such models typically reflects noise, with the non-monotonic parameter estimates being not statistically different from each other [33]. Model 2 followed the same general form as Equation 1 but imposed a restriction of monotonicity across levels of each dimension by combining non-monotonic levels. Model 2 therefore included a reduced number of estimates in β (the vector of preference weights). The MAUCa consortium has used this approach previously for the European Organisation for Research and Treatment

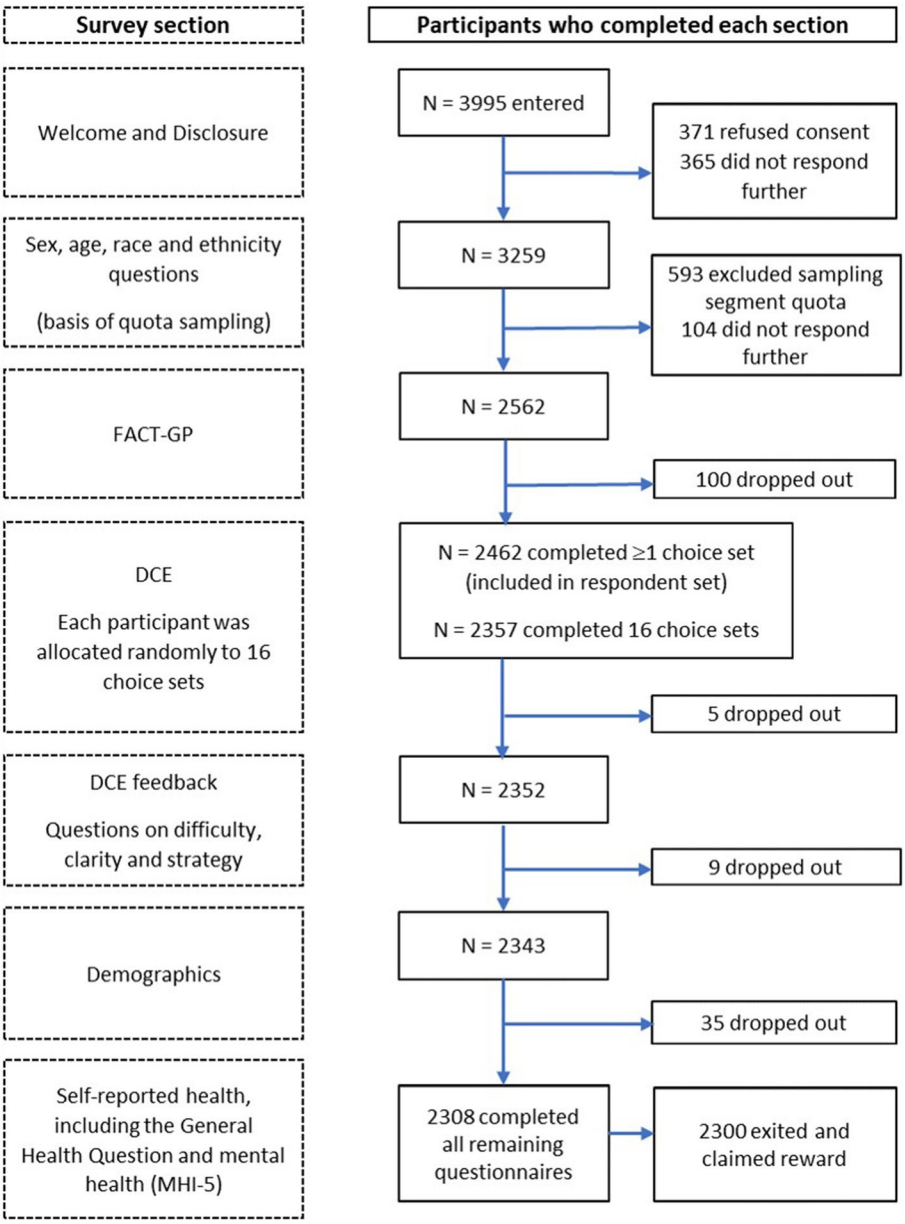


Fig. 2 Valuation survey flow chart. This figure shows the number of survey participants who completed each section of the valuation survey, and the numbers who were excluded or dropped out. Abbreviation: DCE, discrete choice experiment

of Cancer (EORTC) Quality Of Life Utility-Core 10 Dimensions (QLU-C10D) [34–42] and the FACT-8D [11, 15]. The impact of constraining coefficients was assessed with change in the model pseudo-R^2; ideally, the imposition of monotonicity would not reduce model fit markedly.

The secondary analysis (Eq. 2, Model 3) used mixed logit modelling [43] which assumed that coefficients were randomly drawn from a distribution, allowing for preference heterogeneity among individuals (i.e. random coefficients).

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

Thus, α and the vector of βs represent population mean preferences, while γ_i and η_i are individual deviations around those mean preferences. These deviations were assumed to be distributed multivariate normal (0, $\sum$). We used the *mixlogit* STATA command [44] to estimate α, the vector of βs and the standard deviations of γ and the vector of ηs, with the following adjustment. The standard procedures limit the number of parameters drawn from a distribution to 20; to allow all 33 coefficients (including duration) to be drawn from distributions, we used pseudo-random draws.

For variables that deviated from the US general population by ≥ 2.0% in any category, iterative proportional fitting, or raking, weights were included in DCE models [45]. Raking was implemented using the *ipfweight* command in STATA 13.0, with observations with missing demographic data assigned a weight of 1. Variance inflation due

to weighting was assessed by calculating the percentage increase in the standard errors of the unweighted versus weighted Model 1 coefficients.

3 Results

3.1 Sample Characteristics

Figure 2 shows the number of participants that completed each section of the survey. Of 3995 respondents who opted into the survey, 18% (736 out of 3995) opted out after reading about the survey. Of 3259 who consented, 18% (593 out of 3259) were excluded because sampling quotas were met, and 3% (104 out of 3259) did not respond further. Of the remaining 2562 respondents, 96% (2462 out of 2562) completed at least one choice-set, and 92% (2357 out of 2562) completed all 16 choice-sets.

Data from the 2462 participants who completed at least one choice-set were included in analyses to estimate the US FACT-8D valuation set (the DCE analysis sample). Table 2 summarizes the characteristics of these 2462 participants compared with the US general population; these respondents were representative of the general US population for the four quota sampling variables: age ($p = 0.74$), sex ($p = 0.47$), race ($p = 0.99$) and ethnicity ($p = 0.90$). However, people with high school education or less were under-represented by 19%, and people with degrees were over-represented by 19% ($p < 0.001$). Marital status was generally comparable, although never-married people were slightly over-represented (by 2.6%), and widowed people were slightly under-represented (by 2.1%; $p < 0.001$). Self-reported general health did not differ significantly from national data ($p = 0.62$). Education and marital status were included in weighted models.

3.2 Data Quality Metrics

In total, 32 people gave either all A's ($n = 16$) or all B's ($n = 16$) across the choice-sets, 1.3% of the DCE analysis sample. The median time to complete the survey was 13.1 min (interquartile range 9.3–18.7 min). The plots of conditional logit (Model 1, unweighted) model fit and number of statistically significant coefficients by completion time decile showed that the fastest 30% of completers had the poorest model fit and least number of statistically significant coefficients (Fig. A, Supplement 2).

3.3 Respondent Feedback on DCE Task

Of the 2352 participants who answered these questions, 39% rated the difficulty of this survey as 'about the same' compared with other surveys they had done, while 41% rated it as 'harder' (Supplement 3, Table 1a, Fig. 1). Most participants (77%) rated the presentation of the health states as 'clear' or 'very clear' (Supplement 3, Table 1b, Fig. 2). While 41% of participants found it 'very difficult' or 'difficult' to choose between pairs of health states, 29% found it 'easy' or 'very easy' (Supplement 3, Table 1c, Fig. 3). When asked about the strategy used by participants to choose between the pairs of health states, 24% indicated they considered all aspects of the health state, 21% reported focusing on the yellow highlighted aspects, 20% said they focused on just a few aspects and 12% indicated they did not use a strategy. The remaining 2% ($n = 55$) reported using another strategy (Supplement 3, Table 1d, Fig. 4); 33 of these participants provided a brief description (Supplement 3, Table 2). Among these, survival was the dominant theme (mentioned by $n = 20$ out of 33, 61%), with pain and family/support also mentioned by some ($n = 9$ out of 33, 27%).

3.4 Preference Modelling

Estimates of the coefficients in the four pre-specified conditional logit models [(Model 1-unconstrained and Model 2-monotonicity imposed) × (unweighted and weighted)] are reported in Table B (Supplement 2). Imposing monotonicity ('ordering') and weighting generally had little effect on preference weight estimates (Fig. 3). Across all four models, duration had a large and positive coefficient, showing that study participants generally valued longer life. The HRQL coefficients were all negative except Level 2 of fatigue and sadness in Model 1 (unweighted and weighted); these were set to 0 in Model 2. The HRQL dimensions with the largest decrements were pain and nausea, suggesting they had the greatest effect on respondents' preferences for FACT-8D health states. The remaining dimensions had more moderate impacts on preferences.

In total, 11 non-monotonicities were observed among the coefficients in in Model 1 in the unweighted analysis and 12 when weighted (marked in italics in Table B, Supplement 2); all but one of these were between adjacent levels in each unweighted and weighted analysis, with the majority between Levels 2 and 3 ['a little bit' versus 'somewhat', four non-monotonicities in unweighted (4 out of 11, 36%) and weighted (4 out of 11, 36%) analyses], and between Levels 4 and 5 ('quite a bit' versus 'very much', a further four non-monotonicities in each unweighted and weighted analysis). The mean difference between non-monotonic coefficients was 0.0107 (median 0.0070, minimum 0.0002, and maximum 0.0375). Re-estimation with these disordered levels combined (Model 2) reduced the pseudo-R^2 from 0.135 to 0.133 in unweighted analyses and from 0.129 to 0.127 in weighted analyses; imposing these constraints had modest impact on model fit. Standard errors of Model 1 coefficients

Table 2 Self-reported sociodemographic characteristics and health of the US valuation survey sample (n = 2462 participants who completed at least 1 of the 16 discrete choice experiment choice sets) compared with the US general population

Characteristics	Category	Sample (n)	Sample (%)	US pop-ulation (%)	Test statistic[a]	p-value
Sex[b]	Male	1164	47.3%	48.0 %	$X^2 = 0.51$	0.47
	Female	1298	52.7%	52.0 %		
Age[b]	18–29 years	532	21.6 %	22.1 %	$X^2 = 2.73$	0.74
	30–39 years	400	16.2 %	17.1 %		
	40–49 years	458	18.6 %	18.6 %		
	50–59 years	456	18.5 %	17.9 %		
	60–69 years	308	12.5 %	12.5 %		
	70 years or older	308	12.5 %	11.9 %		
Race[c]	American Indian or Alaska Native	21	0.9 %	0.9 %	$X^2 = 0.19$	0.999
	Asian	117	4.8 %	4.8 %		
	Black or African American	316	12.8 %	12.6 %		
	Native Hawaiian or Pacific Islander + Other/Multiple	229	9.3 %	9.3 %		
	White	1779	72.3 %	72.4 %		
Ethnicity[c]	Hispanic or Latino	399	16.2 %	16.3 %	$X^2 = 0.02$	0.90
	Not Hispanic or Latino	2063	83.8 %	83.7 %		
Education[d]	< High school education	69	2.9 %	11.7 %	$X^2 = 519.60$	< 0.001
	High school graduate	441	18.8 %	29.0 %		
	Some college, no degree	443	18.9 %	19.1 %		
	Associate's degree, occupational	154	6.6 %	4.0 %		
	Associate's degree, academic	117	5.0 %	5.5 %		
	Bachelor's degree	652	27.8 %	19.5 %		
	Master's degree	337	14.4 %	8.2 %		
	Professional degree	60	2.6 %	1.3 %		
	Doctoral degree	70	3.0 %	1.6 %		
	Missing	119	–			
Marital status[e]	Married	1220	52.1 %	52.4 %	$X^2 = 22.67$	< 0.001
	Divorced	227	9.7 %	9.8 %		
	Widowed	86	3.7 %	5.8%		
	Never married	810	34.6 %	32.0 %		
	Missing	119	–			
General health question (GHQ)[f]	Excellent [5]	445	19.0 %	19.6%	$X^2 = 2.62$	0.62
	Very good [4]	817	34.9 %	34.6%		
	Good [3]	711	30.3 %	30.3%		
	Fair [2]	300	12.8 %	12.1%		
	Poor [1]	70	3.0 %	3.4%		
	Missing	119	–			

[a]The chi-square goodness-of-fit test was used to compare observed category frequencies with those expected based on population proportions

[b]2010 (April) USA Census data from https://www2.census.gov/programs-surveys/popest/datasets/2000-2010/intercensal/national/

[c]2010 USA Census data from https://www.census.gov/data/datasets/2010/demo/popest/modified-race-data-2010.html; race: the category 'Other/Multiple' of the sample entails participants who selected 'Other race (please specify)' and participants who selected multiple races (e.g. 'White' and 'Asian').

[d]2016 USA Census data from: https://www.census.gov/data/tables/2016/demo/education-attainment/cps-detailed-tables.html; education: the category '< High school education' of the sample includes: No education, 1st-4th grade, 5th-6th grade, 7th-8th grade, 9th grade, 10th grade, 11th grade; completion of 12th grade corresponds to the category 'High school graduate'

[e]2017 USA Census data from https://www.census.gov/data/tables/time-series/demo/families/ marital.html; marital status: the category 'Married' of the sample entails participants who reported to be in a registered marriage, married (de facto) or to be separated

[f]The question asked was 'In general would you say that your health is', with five response options shown in this row. US reference data were provided by the USA 2005–2006 National Health Measurement Study (n = 3841) as reported in Maglinte et al., Journal of Clinical Epidemiology, 2012; 65: 497-502, Table 2

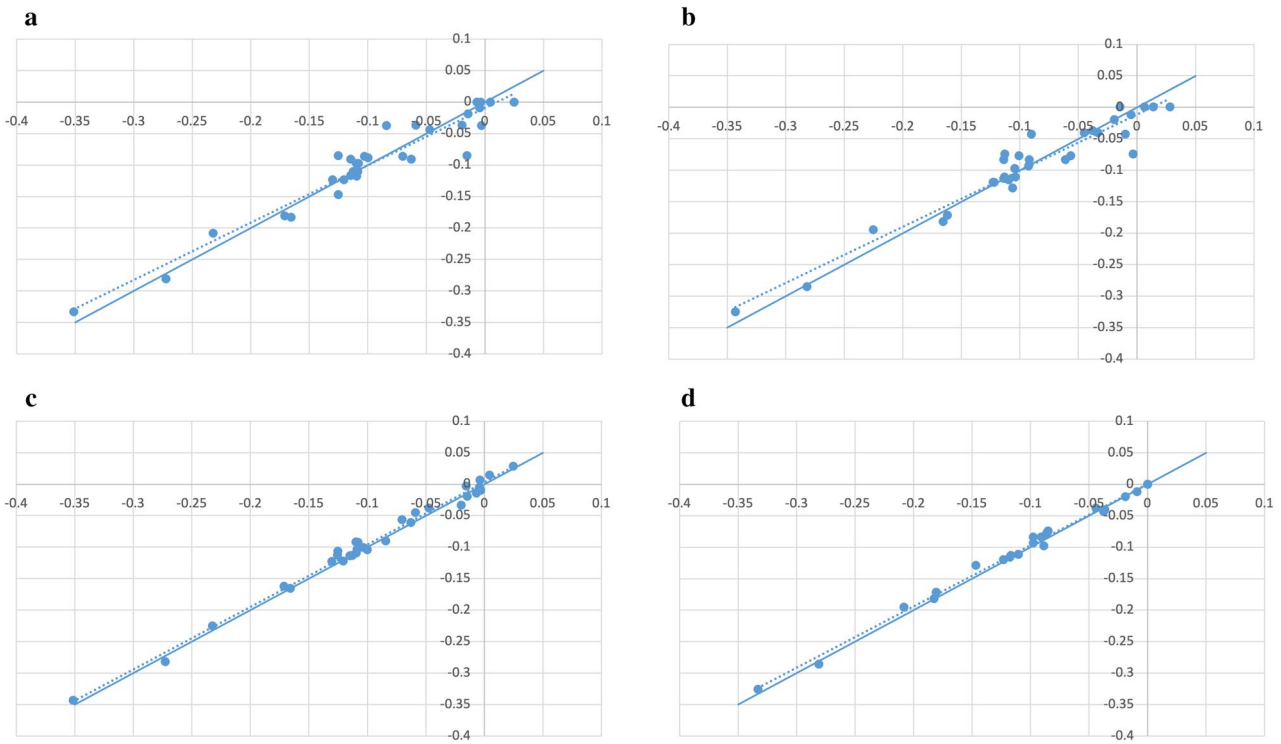
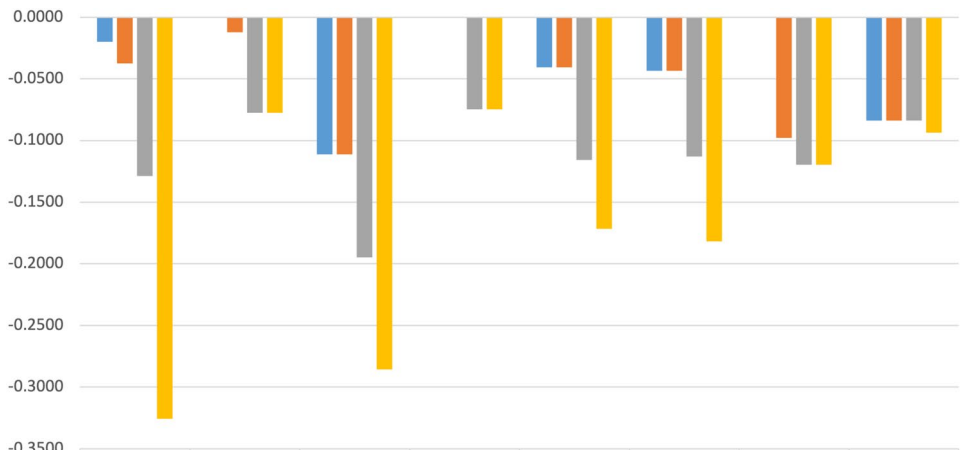


Fig. 3 Scatterplots showing the effect of imposing monotonicity and weighting on preference weight estimates. Panel **A** effect of ordering (unweighted): Model 1 versus Model 2 (both unweighted). Panel **B** effect of ordering (weighted): Model 1 versus Model 2 (both weighted). Panel **C** effect of weighting (unconstrained), Model 1 unweighted versus weighted. Panel **D** effect of weighting (ordered), Model 2 unweighted versus weighted. In each panel, the solid line is the line of equality, the dotted line is the line of best fit, the first model named above is the horizontal axis and the second model named is the vertical axis



	Pain	Fatigue	Nausea	Sleep	Work	Support	Sadness	Worry
Level 1	-0.0200	0.0000	-0.1113	0.0000	-0.0407	-0.0433	0.0000	-0.0838
Level 2	-0.0376	-0.0120	-0.1113	0.0000	-0.0407	-0.0433	-0.0977	-0.0838
Level 3	-0.1288	-0.0776	-0.1949	-0.0745	-0.1157	-0.1131	-0.1197	-0.0838
Level 4	-0.3256	-0.0776	-0.2857	-0.0745	-0.1716	-0.1819	-0.1197	-0.0937

Fig. 4 FACT-8D preference weights by dimension and level (Model 2 conditional logit, monotonicity imposed, weighted). Preference weights indicate the amount the FACT-8D score is reduced for each level of each dimension in the FACT-8D scoring algorithm

in weighted analyses were on average 19% larger (minimum 12%, median 20%, and maximum 29%), but in only two cases did this variance inflation reduce the level of statistical significance (pain level 3 from 5% to 10%; problems working level 2 from 1% to 5%) and in one case increased it (problems working level 3 from not significant to 10%).

In the unweighted mixed logit results (Model 3, Table C, Supplement 2), the mean of the random coefficients

showed a similar pattern to the fixed coefficients from the unweighted unconstrained conditional logit model. Extra years of life in full health were generally highly valued (large, statistically significant mean coefficient). Of the 32 mean coefficients involving FACT-8D levels interacted with duration, 26 had the expected negative coefficients. The standard deviation coefficients were generally significant; of the 33 standard deviations estimated, 26 were significant at the 5% level, revealing considerable between-respondent heterogeneity. The preference weights from the mixed logit were generally smaller than those from the conditional logit model (Fig. B, Supplement 2).

For the purposes of economic evaluation, we recommend using the preference weights derived from the weighted constrained conditional logit model; these reflect the views of the general population and ensure the monotonic nature of the FACT-8D dimensions are reflected in the US scoring algorithm. These are presented in Fig. 4, Table D (Supplement 4), and in the scoring instructions (Appendix B, Supplement 4) and corresponding STATA/SPSS code in plain text (Supplement 5). To illustrate this scoring, if a patient's FACT-G responses indicate that patient is at Level 1 for all FACT-8D dimensions other than fatigue and worry, where they are at Level 3, their health state would be valued at $1 - 0.0120 - 0.0838 = 0.9042$. The worst health state is estimated as −0.3302.

4 Discussion

This study provided a value set representing the preferences of the US general population for health states described by the FACT-8D, a MAUI derived from the FACT-G, a widely used cancer-specific HRQL questionnaire [13]. In this value set, pain and nausea (both common symptoms of cancer) are the largest drivers of FACT-8D utility scores, followed by problems with social support and work, and then sadness and worry about future health, with sleep and fatigue having the smallest impact but still statistically significant at higher levels. The preference-weighted algorithm enables direct calculation of health utilities from FACT-G data including the many FACIT questionnaires that contain the FACT-G items. Given the widespread use of the FACIT questionnaires, this algorithm has potential to facilitate incorporation of common symptoms and impacts of cancer into HTA in the USA.

Adding this new value set to HTA toolkits prompts several questions, including how and why it differs from other value sets and what the consequences are. Factors that can affect value sets include the valuation method (e.g. time trade-off or DCE, plus specific method details), the health states being valued (the kind and number of dimensions, items and levels) and the respondent population (general population versus patients, different countries and diseases). Valuation studies conducted by the MAUCa consortium used a common valuation protocol and utility modelling approach in general population samples with quota sampling and raking to achieve population-representative value sets. These methodological consistencies provide relatively well-controlled comparisons between countries. The FACT-8D preference weights from the US valuation study were generally similar to those generated with the same valuation protocol in Australia and Canada [11, 15]; these similarities illustrate that, when the same health states are valued with the same method in countries with similar cultures, similar results ensue. The EuroQol group also found similar health-state values for the 5-level EuroQol-5 Dimension (EQ-5D-5L) across the USA, Canada, England and the Netherlands using a common DCE valuation protocol [46], but we could not compare these directly with our results due to different modelling approaches.

Head-to-head comparisons of the FACT-8D with generic MAUIs will inform the debate over the ability of generic and disease-specific instruments to capture treatment effects [47, 48]. We are aware of only one such study to date [49]; it utilized data from a randomized trial in which 250 patients with non-Hodgkin's lymphoma completed the FACT-lymphoma (FACT-Lym) and the EQ-5D-5L. Using the Australian DCE-based value set for the FACT-8D and an English value set based on a time-trade-off/DCE hybrid model for the EQ-5D-5L, it found that the FACT-8D showed equally good convergent validity and similar levels of responsiveness, but was outperformed by the EQ-5D-5L in terms of known groups' validity [49]. Similar analyses in other cancer types and treatments would address the generalizability of these findings. However, such psychometric analyses provide limited insight into how different MAUIs might perform in cost–utility analyses. This is better achieved by head-to-head comparison of the QALYs generated by two MAUIs and their impact on cost–utility analyses, as illustrated in three comparisons of the EORTC QLU-C10D with the EuroQol-5 Dimension (EQ-5D) [50–52]. Jansen et al. analysed data from a randomized controlled trial (RCT) of an exercise program to improve swallowing and communication following laryngectomy [50]. Here, the base-case EQ-5D analysis suggested that the intervention was more effective in terms of QALYs and had a slightly lower cost per QALY than control care; in contrast, the EORTC QLU-C10D sensitivity analysis suggested that the intervention was slightly less effective and had a higher cost per QALY. Shaw et al. analysed data from five phase 3 RCTs of nivolumab with or without ipilimumab for the treatment of solid tumours [51]. Similar to Jansen et al., they found that the EQ-5D-3L yielded higher incremental

QALY gains than the QLU-C10D for 4 of the 5 trials. This was confirmed by Kim et al., utilizing an Australian cost–utility analysis of one of the four trials analysed by Shaw et al. Here the EQ-5D-3L yielded slightly higher QALYs gained than the QLU-C10D, lower incremental cost-effectiveness ratio and similar cost-effectiveness acceptability curves [52]. In all three studies, the EQ-5D generated higher QALYs gained than the QLU-C10D. As such evidence accumulates, the variability and generalizability of findings across clinical contexts and utility measures will be revealed. We should also try to understand why differences arise, whether measures capture the full range of treatment effects and which is most fit for the purpose of economic evaluation in a particular decision-making setting and jurisdiction.

The FACT-8D allows utilities and hence QALYs to be estimated without the administration of an additional health preference measure. However, there are potential barriers to its use as the sole preference-based measure in HTA in the USA. The US Second Panel on Cost-Effectiveness in Health and Medicine recommended generic preference-based measures for reference cases to enhance comparability across studies, but emphasized that the measure used should be 'fit for purpose', i.e., able to detect clinically important differences and changes in health across the interventions under consideration [6]. A potential limitation of the FACT-8D is content validity, as its content was taken from the FACT-G, developed as a HRQL profile measure over 30 years ago, so it may not capture the effects of contemporary cancer treatments such as skin reactions and diarrhoea caused by targeted therapies and immunotherapy [53, 54]. However, this criticism applies equally to established generic preference-based measures, including the EuroQol measures. The FACT-8D content is weighted towards psychosocial issues, which may make it appropriate for psychosocial interventions [55]. When designing studies, choice of measure should be matched to the patient population, intervention and relevant regulatory agencies' HTA requirements. Depending on context, even when a generic measure is preferred for the reference case, the FACT-8D could serve as an adjunct measure to improve robustness of evidence in HTA [56].

Various aspects of model choice are relevant when developing a value set for a MAUI. A systematic review of DCE studies used in health state valuation published between June 2018 to November 2022 found that, of 65 included studies, 48 (74%) used conditional logit models, 24 (37%) used mixed logit and 9 (14%) used latent-class logit models. We chose a conditional logit model as our primary analysis, and to generate the US value set, because its parameterization suited our purpose, i.e., for public decision-making, in which mean population preferences are the parameters of interest, needed for the utility algorithm. The distributions of individual preferences were not relevant for our purpose; the conditional logit model was therefore more parsimonious than the mixed logit model in this context. We favoured monotonic preference weights within each HRQL domain, as these are consistent with increasing impact of worsening HRQL levels. For the US value set, we needed to impose some constraints to achieve monotonicity. Finally, we favoured a weighted model, as this adjusted for observed non-representativeness of our sample and therefore yielded utility weights that represent the preferences of the US general population. For completeness, we reported four model fit statistics for each model. The inclusion of survival duration in the DCE enabled a functional form for utilities that was consistent with standard QALY model restrictions. This was first proposed by Bansback et al. in 2012 [26] and has been used widely since, as illustrated in Wang et al.'s systematic review, in which 29 out of 65 (45%) included a duration attribute [57].

The strengths and limitations of this study are discussed in the next three paragraphs. First, we used a DCE approach previously established as feasible in Australia and Canada [11, 15]. This included simplifying the DCE choice task by not asking respondents to trade across all dimensions at once, which also meant the experimental design was not strictly orthogonal design; we felt 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 encountered with the EORTC QLU-C10D because the FACT-G items had mixed framing (some positively framed, some negatively framed), while all the QLQ-C30 items are negatively framed. When we piloted the FACT-8D DCE with the FACT-G's verbatim positive framing of the work, sleep and support items, credible preference weights could not be derived [11]. We solved this by reversing the polarity of these items, as discussed elsewhere [11, 15].

Second, we used modelling approaches appropriate to our data structure and analysis purpose [26]. Our approach to DCE modelling, including constraint for monotonicity, has been used across all MAUCa studies, and is theoretically and empirically stronger than mapping [14]. We acknowledge the relatively large number of non-monotonicities; about one-third of the 32 potential non-monotonicities between adjacent levels were observed. About one-third of the observed non-monotonicities were between Levels 2 and 3 ('a little bit' versus 'somewhat'); the distinction between these two levels in the DCE may not have been as clear as in the FACT-G self-report measure, where the intended gradation of the five response categories is explicit in the questionnaire layout. That explanation may not hold for the non-monotinicities between Levels 4 and 5 ('quite a bit' versus 'very much'), which accounted for a further one-third of the observed non-monotonicities, as these descriptors are clearly

different. Some of these may reflect a genuine lack of graduation in preferences in moving from one level to the next. Although non-monotonicities affected all HRQL dimensions except pain, they were generally very small and had little impact on model fit. Our choice of monotonic main-effects model for calculating utility is readily accessible for a range of end-users, clinically interpretable and consistent with the FACIT conceptual model.

Third, the study had a number of strengths and some limitations in terms of sampling and survey administration. We restricted survey participation to individuals with devices having larger screen sizes. 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, sex, race and ethnicity. However, recruitment was restricted to participants who were able to read English, which may have introduced some bias. We assessed population representativeness across a range of demographic characteristic and adjusted for non-representativeness using raking, which has benefits in terms of bringing the sample in line with pre-specified population characteristics, but costs in terms of inflating variances, particularly when the sample and population differ by a large amount [45]. Even though the standard errors of weighted model coefficients were inflated by weighting, in only two cases did this reduce the level of statistical significance, and the preference weights derived using raked or unraked responses were very similar. The survey sample was derived from a large internet panel; we have previously found poor mental health was over-represented in online panels [11, 38, 58]. Although we assessed mental health in this survey, we were unable to find valid US data against which to assess representativeness. The influence of DCE respondent sociodemographic characteristics on preference weights will be assessed in future analysis of pooled data from international valuations of the FACT- 8D. Non-representativeness in unmeasured variables is always a covert threat to generalizability. Although online panels provide convenient economical sampling frames, potential selection biases deserve further research [59].

5 Conclusions

The US FACT-8D utility algorithm provides a new option for obtaining utility values and QALYs from oncology trials that have used the FACT-G or other FACIT questionnaires in which it is embedded. Further review is needed of the FACT-8D's performance relative to generic utility measures across a range of clinical settings. For this purpose, it should be used alongside an accepted generic measure.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41669-023-00448-5.

Acknowledgements Multi-Attribute Utility in Cancer (MAUCa) Consortium members Neil Aaronson, John Brazier, David Cella, Daniel Costa, Peter Fayers, Peter Grimison, Monika Janda, Georg Kemmler, Madeleine King (Chair), Nan Luo, Helen McTaggart-Cowan, Rebecca Mercieca-Bebber, Richard Norman, Dennis Revicki (in memorandum), Stuart Peacock, Simon Pickard, Donna Rowen, Galina Velikova, Rosalie Viney, Deborah Street and Tracey Young. Margaret-Ann Tait provided valuable production assistance during submission of this manuscript.

Declarations

Funding/Support This study was funded by Bristol Myers Squibb via a consultant agreement with Evidera (EVA-20314-001). The development of the FACT-8D descriptive system and valuation methodology was funded by an Australian National Health and Medical Research Council Project Grant (PG632662).

Conflict of Interest Disclosures DC is president of FACIT.org, which owns and distributes the FACT and FACIT questionnaires. FM reported personal fees from Evidera during the conduct of the study. SP does healthcare consulting with Second City Outcomes Research and Maths in Health. JS reports being an employee and shareholder of Bristol Myers Squibb. MK and RN report grants from Bristol Myers Squibb via Evidera during the conduction of the study. At the time of submission of this paper, RV, RN, SP and JS were members of the EuroQol Group; RV and JS were members of the EuroQol Group's Scientific Executive; and RV chaired the Scientific Executive.

Role of the Funder/Sponsor The funder had no role in the design and conduction of the study; collection, management, analysis, and interpretation of the data; or preparation and decision to submit the manuscript for publication. The funder did review the final manuscript and approved it for publication with no substantive changes required.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics Approval This research was conceived, designed and conducted by the Multi-Attribute Utility in Cancer (MAUCa) Consortium; the University of Sydney Human Research Ethics Committee approved the MAUCa Consortium's program of research (No. 13207). The US valuation study was deemed exempt from US Institutional Review Board (IRB) by the Advarra IRB (CA209-466C, Pro00032061).

Consent to Participate Not applicable.

Consent for Publication (from patients/participants) Not applicable.

Code Availability: Not applicable.

Author Contributions MK, DC, RN, SP, DR and RV contributed towards the study concept and design; MK and RN contributed towards acquisition of data; MK, DC, FM, RN, SP, DR, JS and RV contributed towards analysis and interpretation of data; MK, DR, RN and JS contributed towards drafting of the manuscript; MK, DC, FM, RN, SP, JS and RV provided critical revision of paper for important intellectual content; RN conducted statistical analysis; MK and JS contributed towards obtaining funding; FM provided administrative, technical or logistic support; MK provided overall study supervision. DR was

unable to approve the final version for submission, as he died prior to submission; his family approved his inclusion as an author.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc/4.0/.

References

1. Schnipper LE, Davidson NE, Wollins DS, Tyne C, Blayney DW, Blum D, et al. American Society of Clinical Oncology Statement: a conceptual framework to assess the value of cancer treatment options. J Clin Oncol. 2015;33(23):2563–77.
2. Wang Y, Qiu T, Zhou J, Francois C, Toumi M. Which criteria are considered and how are they evaluated in health technology assessments? A review of methodological guidelines used in Western and Asian countries. Appl Health Econ Health Policy. 2021;19(3):281–304.
3. Weinstein M, Torrance G, McGuire A. QALYs: the basics. Value Health. 2009;11:S5–9.
4. Pizzi LT. The Institute for Clinical and Economic Review and its growing influence on the US healthcare. Am Health Drug Benefits. 2016;9(1):9–10.
5. ICER. Institute for Clinical and Economic Review 2020-2023 Value Assessment Framework. 2021. https://icer.org/wp-content/uploads/2022/01/ICER_2020_2023_VAF_120821.pdf.
6. Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. JAMA. 2016;316(10):1093–103.
7. Green C, Brazier J, Deverill M. Valuing health-related quality of life. A review of health state valuation techniques. Pharmacoeconomics. 2000;17(2):151–65.
8. Bahrampour M, Byrnes J, Norman R, Scuffham PA, Downes M. Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods. Eur J Health Econ. 2020;21(7):983–92.
9. Mayo NE. Dictionary of quality of life and health outcomes measurement: Internatiional Society of Quality of Life Research (ISOQOL); 2015, p. 179.
10. Xie F, Pickard AS, Krabbe PF, Revicki D, Viney R, Devlin N, et al. A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). Pharmacoeconomics. 2015;33(8):867–77.
11. King MT, Norman R, Mercieca-Bebber R, Costa DSJ, McTaggart-Cowan H, Peacock S, et al. 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. Value Health. 2021;24(6):862–73.
12. Cella D, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol. 1993;11(3):570–9.
13. Smith AB, Cocks K, Parry D, Taylor M. Reporting of health-related quality of life (HRQOL) data in oncology trials: a comparison of the European Organization for Research and Treatment of Cancer Quality of Life (EORTC QLQ-C30) and the Functional Assessment of Cancer Therapy-General (FACT-G). Qual Life Res. 2014;23(3):971–6.
14. 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):753–65.
15. McTaggart-Cowan HKM, Norman R, Costa DSJ, Pickard AS, Viney R, Peacock SJ. The FACT-8D, a new cancer-specific utility algorithm based on the Functional Assessment of Cancer Therapies-General (FACT-G): a Canadian valuation study. Health Qual Life Outcomes. 2022;20:97.
16. ICC/ESOMAR. The International Code on Market, Opinion and Social Research and Data Analytics 2016. https://iccwbo.org/publication/iccesomar-international-code-market-opinion-social-research-data-analytics/. Accessed 7 October 2022.
17. Mulhern B, Norman R, Street DJ, Viney R. One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. Pharmacoeconomics. 2019;37(1):29–43.
18. Bliemer M. Constraining the number of attributes that differ between health states in any given choice-set. Personal communication, recipient Richard Norman. 2016.
19. ChoiceMetrics. Ngene Software website. [cited 2022 June 27]. Available from: www.choice-metrics.com. Accessed 7 October 2022.
20. Rose JM, Collins AT, Bliemer MC, Hensher DA. Ngene 1.0 stated choice experiment design software. 2009.
21. Cook RD, Nachtrheim CJ. A comparison of algorithms for constructing exact D-Optimal designs. Technometrics. 1980;22:315–24.
22. Norman R, Viney R, Aaronson NK, Brazier JE, Cella DF, Costa DSJ, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Qual Life Res. 2016;25(3):637–49.
23. Norman R, Kemmler G, Viney R, Pickard AS, Gamper EM, Holzner B, et al. Order of presentation of dimensions does not systematically bias utility weights from a discrete choice experiment. Value Health. 2016;19(8):1033–8.
24. Maglinte GA, Hays RD, Kaplan RM. US general population norms for telephone administration of the SF-36v2. J Clin Epidemiol. 2012;65(5):497–502.
25. StataCorp,. Stata Statistical Software: Release 13. College Station: StataCorp LP.; 2013.
26. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate societal health state utility values. J Health Econ. 2012;31:306–18.
27. 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):287–98.
28. Norman R, Viney R, Brazier JE, Burgess L, Cronin P, King MT, et al. Valuing SF-6D health states using a Discrete Choice Experiment. Med Decis Making. 2014;34(6):773–86.
29. Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, et al. An Australian discrete choice experiment to value eq-5d health States. Health Econ. 2014;23(6):729–42.

30. 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.
31. Bleichrodt N, Wakker P, Johannesson M. Characterizing QALYs by risk neutrality. J Risk Uncertain. 1997;15(2):107–14.
32. Hole AR. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Econ. 2007;16:827–40.
33. Brazier J, Roberts J, Deverill MT. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002;21(2):271–92.
34. Finch A, Gamper E, Norman R, Viney R, Holzner B, King M, et al. Estimation of an EORTC QLU-C10D value set for Spain using a discrete choice experiment. Pharmacoeconomics. 2021;39:1085–98.
35. Gamper EM, King MT, Norman R, Efficace F, Cottone F, Holzner B, et al. EORTC QLU-C10D value sets for Austria, Italy, and Poland. Qual Life Res. 2020;29(9):2485–95.
36. Jansen F, Verdonck-de Leeuw I, Gamper E, Norman R, Holzner B, King M, et al. Dutch utility weights for the EORTC cancer-specific utility instrument: the Dutch EORTC QLU-C10D. Qual Life Res. 2021;30:2009–19.
37. Kemmler G, King M, Norman R, Viney R, Gamper E, Holzner B. German value sets for the EORTC QLU-C10D, a cancer-specific utility instrument based on the EORTC QLQ-C30. Qual Life Res. 2019;28:3197–211.
38. King MT, Viney R, Simon Pickard A, Rowen D, Aaronson NK, Brazier JE, et al. Australian utility weights for the EORTC QLU-C10D, a multi-attribute utility instrument derived from the cancer-specific quality of life questionnaire, EORTC QLQ-C30. Pharmacoeconomics. 2018;36(2):225–38.
39. McTaggart-Cowan H, King M, Norman R, Costa DSJ, Pickard S, Regier DA, 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.
40. Nerich V, Gamper EM, Norman R, King M, Holzner B, Viney R, et al. French value-set of the QLU-C10D, a cancer-specific utility measure derived from the QLQ-C30. Appl Health Econ Health Policy. 2021;19(2):191–202.
41. Norman R, Mercieca-Bebber R, Rowen D, Brazier JE, Cella D, Pickard AS, et al. U.K. utility weights for the EORTC QLU-C10D. Health Econ. 2019;28(12):1385–401.
42. Revicki DA, King MT, Viney R, Pickard AS, Mercieca-Bebber R, Shaw JW, et al. United States utility algorithm for the EORTC QLU-C10D, a multiattribute utility instrument based on a cancer-specific quality-of-life instrument. Med Decis Making. 2021;41(4):485–501.
43. Revelt D, Train KE. Customer-specific taste parameters and mixed logit. Berkeley: University of California; 1999.
44. Hole AR. Fitting mixed logit models by using maximum simulated likelihood. STATA J. 2007;7:388–401.
45. Deming WE, Stephan FF. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann Math Stat. 1940;11(4):427–44.
46. Krabbe PFMP, Devlin NJP, Stolk EAP, Shah KKM, Oppe MP, van Hout BP, et al. Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. Med Care. 2014;52(11):935–43.
47. Rowen D, Brazier J, Ara R, Azzabi ZI. The role of condition-specific preference-based measures in health technology assessment. Pharmacoeconomics. 2017;35(Suppl 1):33–41.
48. Versteegh MM, Leunis A, Uyl-de Groot CA, Stolk EA. Condition-specific preference-based measures: benefit or burden? Value Health. 2012;15(3):504–13.
49. Herdman M, Kerr C, Pavesi M, Garside J, Lloyd A, Cubi-Molla P, 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.
50. Jansen F, Coupe VMH, Eerenstein SEJ, Cnossen IC, van Uden-Kraan CF, de Bree R, et al. Cost-utility and cost-effectiveness of a guided self-help head and neck exercise program for patients treated with total laryngectomy: results of a multi-center randomized controlled trial. Oral Oncol. 2021;117: 105306.
51. Shaw JW, Bennett B, Trigg A, DeRosa M, Taylor F, Kiff C, et al. A comparison of generic and condition-specific preference-based measures using data from nivolumab trials: EQ-5D-3L, mapping to the EQ-5D-5L, and European Organisation for Research and Treatment of Cancer Quality of Life Utility Measure-Core 10 Dimensions. Value Health. 2021;24(11):1651–9.
52. Kim H, Cook G, Goodall S, Liew D. Comparison of EQ-5D-3L with QLU-C10D in metastatic melanoma using cost-utility analysis. Pharmacoeconom Open. 2021;5(3):459–67.
53. National_Cancer_Institute. Immunotherapy Side Effects. 2022. https://www.cancer.gov/about-cancer/treatment/types/immunotherapy/side-effects. Accessed 7 October 2022.
54. National_Cancer_Institute. Targeted Therapy to Treat Cancer. 2022. https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies Accessed 7 October 2022.
55. Gabriel I, Creedy D, Coyne E. A systematic review of psychosocial interventions to improve quality of life of people with cancer and their family caregivers. Nurs Open. 2020;7(5):1299–312.
56. Feeny D, Furlong W, Torrance GW. Commentary. In praise of studies that use more than one generic preference-based measure. Int J Technol Assess Health Care. 2019;35(4):257–62.
57. Wang H, Rowen DL, Brazier JE, Jiang L. Discrete choice experiments in health state valuation: a systematic review of progress and new trends. Appl Health Econ Health Policy. 2023;21(3):405–18.
58. Smith A, King M, Butow P, Grimison P, Olver I. Reduced quality of life in internet-panel controls. Epidemiology. 2013;24(3):462–3.
59. Bethlehem J. Selection bias in web surveys. Int Stat Rev. 2010;78(2):161–88.

Authors and Affiliations

Madeleine T. King[1] · D. A. Revicki[2] · R. Norman[3] · F. Müller[4] · R.C. Viney[5] · A. S. Pickard[6] · D. Cella[7] · J. W. Shaw[8] on behalf of the MAUCa Consortium

✉ Madeleine T. King
madeleine.king@sydney.edu.au

[1] School of Psychology, Faculty of Science, University of Sydney, Sydney, NSW, Australia

[2] Revicki Outcomes Research Consulting, Sarasota, FL, USA

[3] School of Population Health, Curtin University, Perth, WA, Australia

[4] Medical Psychology, Amsterdam UMC Location University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands

[5] Centre for Health Economics Research & Evaluation, Faculty of Health, University of Technology Sydney, Sydney, NSW, Australia

[6] Department of Pharmacy Systems, Outcomes and Policy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA

[7] Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

[8] Patient-Reported Outcomes Assessment, Global Health Economics and Outcomes Research, Bristol Myers Squibb, Lawrenceville, NJ, USA