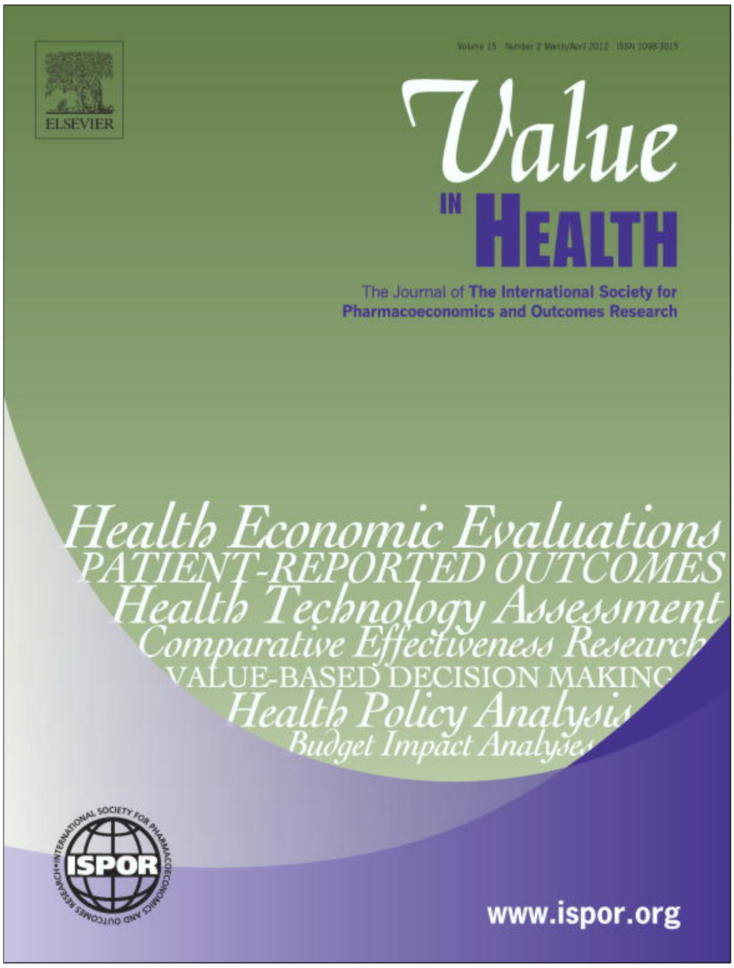
Volume 15 Number 2 March/April 2012 ISSN 1098-3015
ELSEVIER
Value in Health
The Journal of The International Society for Pharmacoeconomics and Outcomes Research
Health Economic Evaluations
PATIENT-REPORTED OUTCOMES
Health Technology Assessment
Comparative Effectiveness Research
VALUE-BASED DECISION MAKING
Health Policy Analysis
Budget Impact Analyses
ISPOR
www.ispor.org

VALUE IN HEALTH 15 (2012) 305–311

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Patient Reported Outcomes

Comparison of FACT- and EQ-5D–Based Utility Scores in Cancer

A. Simon Pickard, PhD[1,], Saurabh Ray, PhD[2], Arijit Ganguli, PhD[2], David Cella, PhD[3]*

[1]Second City Outcomes Research, Oak Park, IL, USA; [2]Global Health Economics and Outcomes Research, Abbott Laboratories, Abbott Park, IL, USA; [3]Northwestern University, Feinberg School of Medicine, Chicago, IL, USA

A B S T R A C T

Objective: Although utility-based algorithms have been developed for the Functional Assessment of Cancer Therapy (FACT), their properties are not well known compared with those of generic utility measures such as the EQ-5D. Our objective was to compare EQ-5D and FACT preference-based scores in cancer patients. **Methods:** A retrospective analysis was conducted on cross-sectional data collected from 472 cancer patients who completed both FACT-General and the EQ-5D. Preference-based scores were calculated by using published scoring functions for the EQ-5D (Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108; Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care 2005;43:203–20) and FACT (Dobrez D, Cella D, Pickard AS, et al. Estimation of patient preference-based utility weights from the Functional Assessment of Cancer Therapy-General. Value Health 2007;10:266–72; Kind P, Macran S. Eliciting social preference weights for Functional Assessment of Cancer Therapy-Lung health states. Pharmacoeconomics 2005;23:1143–53; Cheung YB, Thumboo J, Gao F, et al. Mapping the English and Chinese versions of the Functional Assessment of Cancer Therapy-General to the EQ-5D utility index. Value Health 2009;12:371–6). Scores were compared on the basis of clinical severity by using Eastern Cooperative Oncology Group performance status ratings by physicians and patients. Relative efficiency of each scoring function was examined by using ratios of *F* statistics. **Results:** Mean scores for the overall cohort were lowest when using Kind and Macran's FACT UK societal algorithm (0.55, SD 0.09) and highest when using Dobrez et al.'s FACT US patient algorithm (0.83, SD 0.08). Mean difference scores associated with clinical severity, when extrapolated to quality-adjusted life-years (QALYs), had a range of 0.18 QALYs gained using FACT (Kind and Macran) to 0.45 QALYs gained using the EQ-5D (Dolan). However, relative efficiencies suggested that FACT (Kind and Macran) scores may provide greater statistical power to detect significant differences based on clinical severity. **Conclusions:** We found important differences in utilities scores estimated by each algorithm, with FACT-based algorithms tending to underestimate the QALY benefit compared with algorithms based on the EQ-5D. These differences highlight some of the challenges in using disease-specific preference-based measures for decision making despite potentially more relevant disease-specific content.

Keywords: cancer, EQ-5D, health-related quality of life, health state utilities, utility assessment.

Introduction

The ability to generate utility scores as an outcome in oncology trials is essential to the conduct of cost-utility analyses. Utility scores enable the calculation of quality-adjusted life-years (QALYs) [1], a metric that adjusts time in a health state by the desirability or preference for that health state to evaluate the value and/or cost-effectiveness of treatments for cancer. The most widely used utility measures are generic measures, particularly the EQ-5D [2]. Numerous national value sets are available to score the EQ-5D, with the most highly cited societal value sets being derived from general population in the United Kingdom [3] and the United States [4]. The EQ-5D has generally demonstrated validity and reliability in studies of cancer [5]. There are often instances, however, when utility scores are desired but no preference-based measure is used in a study, or the generic measure may lack responsiveness to meaningful changes in health-related quality of life. The former issue has fostered research focused on the mapping of non–preference-based disease-specific measures to generic measures that generate utility scores [6].

Preference-based scoring functions for cancer-specific measures include several for the Functional Assessment of Cancer Therapy (FACT) [7], a well-established family of cancer-specific measures. Originally developed by using methods based on psychometric theory, the FACT scoring involves the summation of ordinal-level responses to items belonging to each scale. This approach contrasts with the "utility" approach to health measurement, where a summary score is derived by applying a utility function or set of preference weights assigned to the levels and dimensions of the measure.

The preference-based algorithms published for the FACT system have varied in their methods and rater perspective. Dobrez

* *Address correspondence to:* A. Simon Pickard, Second City Outcomes Research, 414 N. Ridgeland Avenue, Oak Park, IL 60302, USA.
E-mail: simon.pickard@secondcityoutcomesresearch.com.

1098-3015/$36.00 – see front matter Copyright © 2012, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.
doi:10.1016/j.jval.2011.11.029

Table 1 – Utility-based algorithms.

Author, year [reference]	Measure	Domains in measure	Items/domains in algorithm	Source of utilities	Range of possible values
Dolan, 1997 [3]	EQ-5D (3-level)	Mobility, usual activities, self-care, pain/discomfort, anxiety/depression	Mobility, usual activities, self-care, pain/discomfort, anxiety/depression	Societal (UK)	−0.59 to 1.0
Shaw et al., 2005 [4]	EQ-5D (3-level)	Mobility, usual activities, self-care, pain/discomfort, anxiety/depression	Mobility, usual activities, self-care, pain/discomfort, anxiety/depression	Societal (US)	−0.109 to 1.0
Dobrez et al., 2007 [8]	FACT-G	Physical well-being (PWB), emotional well-being (EWB), functional well-being (FWB), social well-being (SWB)	Two items: PWB, Two items: FWB	Cancer patients (US)	0.50–1.04
Kind and Macran, 2005 [9]	FACT – Lung (FACT-L, version 4)	PWB, EWB, FWB, SWB, Lung component – symptoms (LCS)	PWB, EWB, FWB, SWB, LCS	Societal (UK)	0.18–0.70
Cheung et al., 2009 [11]	FACT-G (version 4)	PWB, EWB, FWB, SWB	Scale scores for PWB, EWB, FWB	Mapped from Dolan 1997 (UK societal)	0.238–0.982

FACT-G, Functional Assessment of Cancer Therapy-General.

et al. [8] estimated a set of value weights for selected items from FACT-General based on time trade-off (TTO) scores directly elicited for own health from US cancer patients. Kind and Macran [9] derived a set of societal preference weights by directly eliciting visual analogue scale (VAS) ratings from the general population in the United Kingdom for a descriptive health classifier system based on FACT-Lung (FACT-L). Cheung and colleagues mapped both English and Chinese versions of FACT to the EQ-5D utility scores and derived a single mapping function for both languages based on three of the four summary scores from FACT.

Given the many differences between the EQ-5D and FACT-based scoring functions, we would expect them to generate different scores. The extent to which the scoring functions differ, however, has not been well documented or contrasted. Because different decision makers have different needs, they may wish to not only consider the value that each approach can contribute to decision making but also understand how the metrics differ relative to each other, particularly with respect to statistical efficiency and QALY calculations. Thus, our objective was to examine preference-based scores generated by EQ-5D and FACT scoring functions to better understand the strengths and limitations of each approach in valuing health.

Methods

Data

A retrospective analysis was conducted on a data set where patients completed both the EQ-5D and FACT. The cohort consisted of 534 cancer patients who participated in a US-based multicenter symptom scale validation study, which has been previously described [10]. Patients completed both instruments on the same day. All patients had advanced cancer classified as relating to any 1 of 11 tumor sites: bladder, brain, breast, colorectal, head/neck, hepatobiliary/pancreas, kidney, lung, lymphoma, ovary, or prostate. Approximately equal proportions of male and female patients were recruited for the non–gender-specific cancers.

Patients were recruited from six sites that were geographically representative of National Comprehensive Cancer Network member institutions, an alliance of National Cancer Institute–approved comprehensive cancer centers: Duke University Medical Center, Fred Hutchinson Cancer Research Center, Dana Farber Cancer Institute, H. Lee Moffitt Cancer Center and Research Institute, and the Robert H. Lurie Comprehensive Cancer Center at Northwestern University. In addition, patients were recruited by members of the Cancer Health Alliance of Metropolitan Chicago, a coalition of four community support agencies serving the Chicago metropolitan area. The Cancer Health Alliance of Metropolitan Chicago organizations provide social, emotional, and informational support services to cancer patients free of charge and are unaffiliated with a medical center or university. Each Cancer Health Alliance of Metropolitan Chicago agency serves different geographical and sociodemographic cancer patient populations.

Measures/algorithms

We compared the scores generated by five algorithms/scoring functions, two of them based on the EQ-5D and three based on responses to FACT (Table 1). They are described in greater detail below.

EQ-5D

The EQ-5D descriptive system consists of five dimensions (Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression), each with three levels (3L) of health [2]. The EQ-5D also includes a 20-cm VAS, which asks the respondents to rate their health today from 0 (worst imaginable health) to 100 (best imaginable health). A preference-based score is calculated from responses to the health state descriptive system that is typically interpreted along a continuum where 1 represents best possible health and 0 represents dead, with some health states being worse than dead (<0). Participants were asked to complete the standard US English language EQ-5D (3L) and the VAS. The US and UK English language versions of the EQ-5D are nearly identical, differing only in the instruction to place a tick (UK) or checkmark (US) in the box that best describes your own health state today.

The EQ-5D preference-based scores were calculated by using the algorithms developed by Dolan [3] from the general population in the United Kingdom and by Shaw and colleagues [4] for the United States. A ceiling effect is observed in milder health conditions with the EQ-5D health classifier system [10,12], a limitation that applies equally to both algorithms. The scores generated by the Dolan UK value set range from −0.59 (for health state vector 33333, which represents the worst health state) to 1.0 (for health state vector 11111, which represents full health), while the scores for the US value set from Shaw and colleagues cover a smaller range of scale, from −0.109 to 1.0.

FACT

Participants completed the FACT questionnaire by using a version specific to their cancer diagnosis [7] The general subscales common to all versions (FACT-G) include physical well-being, social/family well-being, emotional well-being, and functional well-being. Preference-based scores were generated by using the FACT-G–based algorithm by Dobrez and colleagues [8], a preference-based algorithm for the FACT-L by Kind and Macran [9], and Cheung et al.'s [11] scoring function that predicted EQ-5D scores from FACT-based summary scores. Although the latter is not a preference-based algorithm, we included it for comparative purposes.

The algorithm by Dobrez and colleagues is based on four items from FACT-G that were selected on the basis of criteria that included the extent of item correlation with TTO scores, Eastern Cooperative Oncology Group (ECOG) performance status, the significance of item coefficients, and testing using item-response theory–based models. The modeling was conducted by using TTO utilities directly obtained from 1433 cancer patients with 1 of 10 different diagnoses. The final additive model included two items from physical well-being and two items from functional well-being subscales. The scale's scores ranged from 0.50 to 1.04.

The VAS-based social preference weights for FACT-L by Kind and Macran was based on a postal survey of 433 respondents from the UK general population. A reconfigured descriptive system was created for FACT-L by consulting with experts and using factor analysis. The final model included 10 items, with at least one from each of the four FACT-G scales and one item from the additional concerns subscale "short of breath." The summary index–based scores ranged from a minimum of 0.111 to a maximum of 0.703. As in this study, the algorithm may be modified to apply more broadly to all cancer patients with the assumed response "not at all" to the "short of breath" item if patients completed FACT-G rather than FACT-L, resulting in a scale score ranging from 0.182 to 0.703.

The function for mapping FACT-G scores to EQ-5D index-based scores by Cheung et al. [11] was based on a survey of 558 cancer patients in Singapore. Regression models mapped the English and Chinese versions of FACT-G scale scores for physical, emotional, and functional well-being to the EQ-5D index-based score. It was concluded that a single mapping function could be used for both language versions.

ECOG

Performance status was evaluated by using the ECOG classification system [13]. The ECOG scale is used by physicians and researchers to determine progression of disease, impact of the disease on daily activities, and guide appropriate treatment and prognosis. Performance status is based on a score that ranges from 0 (fully active) to 5 (dead). Although ECOG performance status is typically assessed by clinicians, in this study patients were also asked to rate their ECOG status.

Analysis

The distribution of scores generated by each utility algorithm was described by using summary statistics including min, max, mean, median, SD, and range of scores. Ceiling and floor effects were examined among subjects whose responses were entirely at the top or the bottom of the descriptive system associated with each algorithm. Correlations between EQ-5D and FACT preference-based scores and ECOG performance status were assessed by using Spearman's rho (r_s) or Pearson's correlation coefficient (r) when appropriate. Strength of correlation was interpreted as follows: absent (<0.2), weak (0.2–0.34), moderate (0.35–0.5), and strong (>0.5) [14].

For insight into the discriminative ability of each algorithm, the cohort was stratified by ECOG subgroups and mean scores reported in terms of absolute difference and magnitude of difference (effect size) between subgroups. Effect size was calculated on the basis of difference in mean scores for adjacent ECOG subgroup divided by a pooled SD that weighted by sample size of each subgroup. For each preference-based scoring function, one-way ANOVA was used to test for statistically significant differences in scores according to clinical severity using the ECOG status. To compare the relative statistical efficiency of the algorithms, relative efficiency (RE) ratios were calculated on the basis of ratios of F statistics [15]. A higher RE implies greater power to detect a statistically significant difference between groups.

Table 2 – Respondent characteristics, all cancer patients (n = 472).

Characteristic	
Age (y), mean (SD)	57 (13)
Gender (female), n (%)	216 (46)
Type of cancer, n (%)	
Bladder	28 (6)
Brain	40 (10)
Breast	41 (9)
Head/neck	46 (10)
Hepatobiliary	48 (10)
Kidney	48 (10)
Lung	44 (9)
Lymphoma	40 (9)
Ovarian	40 (9)
Prostate	44 (9)
Colorectal	47 (10)
ECOG level (physician-rated), n (%)	
0	131 (28)
1	253 (54)
2	73 (16)
3	15 (3)
ECOG level (patient-rated), n (%)	
0	108 (23)
1	227 (48)
2	118 (25)
3	19 (4)

ECOG, Eastern Cooperative Oncology Group.

Results

Among respondents with complete data (n = 472), the mean age was 57 (SD 12) years and 54% were males (Table 2). The ECOG level was rated by all patients (n = 534) and by physicians for 91% of the patients (n = 472), that is, 9% were missing. Compared with physician ratings, a significantly greater proportion of patients rated their ECOG status as more severe ($P < 0.001$; Wilcoxon signed ranks test).

The mean scores for the overall cohort generated by each algorithm ranged from 0.553 (SD 0.086) for Kind and Macran's FACT UK societal algorithm to 0.827 (SD 0.084) for the FACT US patient algorithm (Table 3). If all scoring functions had the same range and maximum scores, these results would imply that Kind and Macran's algorithm would provide the greatest opportunity for scores to improve. The upper, lower, and range of the scale scores, however, differ between algorithms. The range of observed values was largest for EQ-5D UK societal [1,14], more than double the range of observed scores for FACT UK societal by Cheung et al. (0.53), FACT UK societal by Kind and Macran (0.52), and Dobrez et al.'s FACT US patient perspective (0.50). The compressed range of FACT-based scoring algorithms restrict the extent to which utility scores can change compared with the EQ-5D.

Table 3 – Mean utility scores using EQ-5D and FACT-G algorithms, all cancer patients (n = 472).

Score	Mean	SD	Median (IQR)	Min	Max
EQ-5D UK societal—Dolan	0.719	0.227	0.727 (0.620–0.850)	−0.14	1.00
EQ-5D US societal—Shaw et al.	0.785	0.154	0.804 (0.708–0.854)	0.21	1.00
FACT US patient—Dobrez et al.	0.827	0.084	0.852 (0.757–0.886)	0.54	1.04
FACT UK societal—Kind and Macran	0.553	0.086	0.564 (0.497–0.619)	0.28	0.70
FACT UK societal mapped—Cheung et al.	0.749	0.113	0.760 (0.676–0.838)	0.42	0.95

IQR, interquartile range; min, minimum observed value; max, maximum observed value.

Ceiling effects were negligible for scoring functions based on the FACT descriptive system. When stratifying the 17% of patients at the top of the EQ-5D descriptive system by ECOG subgroup, a ceiling effect was observed. Among the 17% of patients who self-reported full health on the EQ-5D, 57% were not rated as fully active by physicians, that is, ECOG level 1, 2, or 3, and 44% of patients rated themselves as not fully active. No floor effect was detected on either measure.

In comparing mean scores across ECOG subgroups, the difference was the largest for EQ-5D UK societal scores (0.46), while the smallest difference was observed for FACT US patient scores (0.13) (Table 4). Kind and Macran's algorithm and Cheng et al.'s scoring function had the highest REs (RE = 2.37 and 2.64, respectively, using the EQ-5D UK societal as the reference when stratified by patient-rated ECOG subgroups) (Table 4). When the same analysis was performed by using physician-rated rather than patient-rated ECOG subgroups, similar results were observed in terms of the relative performance of each scoring function (Table 5). The *F* statistics indicated that greater discriminative ability was observed when subjects were stratified into ECOG subgroups that were patient rated rather than physician rated.

All the preference-based scores were moderately to strongly correlated, ranging from $r = 0.48$ to $r = 0.99$ (Table 6). Kind and Macran's and Cheung et al.'s FACT-based scores were strongly correlated with EQ-5D scores ($r_s = 0.67–0.69$). Stronger correlations were observed between utility scores and patient ECOG ratings ($r = 0.44–0.58$) than between utility scores and physician ECOG ratings ($r = 0.31–0.42$). Only moderate correlation was observed between physician- and patient-rated ECOG status ($r_s = 0.43$).

Discussion

This comparison of preference-based scores derived from the EQ-5D and FACT descriptive systems found support for the construct validity of each utility-based algorithm. Mean utility scores increased monotonically with better clinical outcomes based on the ECOG status, and moderate to strong intermeasure correlations were observed. The utility scores generated by each algorithm for the overall cohort and for ECOG performance status–based patient subgroups generally differed to an extent that would be considered meaningful [16]. The differences in mean scores between ECOG subgroups for all algorithms tended to be larger (effect sizes >0.80) with more functional impairment (Table 5), which indicates a consistent pattern of substantial decline in preferences for health states with more severe levels of functioning.

In comparing the mean scores of ECOG-based groups, the largest difference in mean scores was obtained by using the EQ-5D UK population-based algorithm by Dolan. If the results in Tables 4 and

Table 4 – Utility scores by patient-rated ECOG status.

Algorithm	ECOG status	n	Min	Max	Median	Mean	SD	MD by ECOG status	Effect size	*F* statistic (RE ratio)
EQ-5D UK societal—Dolan	0	108	0.26	1	0.85	0.87	0.15	0.46*		34.9 (1.0)
	1	227	−0.14	1	0.73	0.72	0.21	0.15	0.83	
	2	118	0.08	1	0.69	0.63	0.21	0.09	0.43	
	3	19	−0.07	0.8	0.52	0.41	0.28	0.22	0.90	
EQ-5D US societal— Shaw et al.	0	108	0.51	1	0.84	0.89	0.11	0.31*		39.2 (1.1)
	1	227	0.21	1	0.8	0.78	0.14	0.11	0.88	
	2	118	0.33	1	0.77	0.72	0.14	0.06	0.43	
	3	19	0.26	0.83	0.6	0.58	0.18	0.14	0.88	
FACT US patient—Dobrez et al.	0	108	0.65	1.04	0.89	0.89	0.08	0.13*		38.0 (1.1)
	1	227	0.65	1	0.85	0.83	0.07	0.06	0.80	
	2	118	0.6	0.94	0.78	0.78	0.08	0.04	0.67	
	3	19	0.54	0.93	0.76	0.76	0.09	0.02	0.24	
FACT UK societal—Kind and Macran	0	108	0.46	0.70	0.64	0.63	0.05	0.18*		82.6 (2.4)
	1	227	0.31	0.69	0.57	0.56	0.07	0.07	1.17	
	2	118	0.29	0.65	0.49	0.49	0.07	0.07	1.00	
	3	19	0.28	0.56	0.47	0.45	0.08	0.04	0.53	
FACT UK societal—Cheung et al.	0	108	0.63	0.95	0.86	0.85	0.07	0.25*		92.1 (2.6)
	1	227	0.50	0.93	0.77	0.76	0.09	0.09	1.05	
	2	118	0.43	0.85	0.67	0.66	0.10	0.10	1.04	
	3	19	0.42	0.75	0.60	0.60	0.10	0.06	0.66	

FACT, Functional Assessment of Cancer Therapy; RE, relative efficiency ratio using EQ-5D UK societal as reference; MD, mean difference; ECOG, Eastern Cooperative Oncology Group.

* Difference in mean scores between ECOG levels 0 and 3.

Table 5 – Utility scores by physician-rated ECOG status.

Algorithm	ECOG status	n	Min	Max	Median	Mean	SD	MD by ECOG status	Effect size	F statistic (RE ratio)
EQ-5D UK societal—Dolan	0	131	−0.02	1.00	0.85	0.80	0.19	0.45*		19.5 (1.1)
	1	253	−0.07	1.00	0.73	0.72	0.22	0.08	0.39	
	2	73	−0.14	1.00	0.69	0.63	0.24	0.09	0.41	
	3	15	0.09	0.69	0.52	0.45	0.20	0.18	0.83	
EQ-5D US societal—Shaw et al.	0	131	0.31	1.00	0.84	0.84	0.13	0.24*		22.4 (1.3)
	1	253	0.26	1.00	0.80	0.79	0.15	0.06	0.40	
	2	73	0.21	1.00	0.77	0.72	0.16	0.07	0.45	
	3	15	0.40	0.78	0.60	0.60	0.12	0.12	0.87	
FACT US patient—Dobrez et al.	0	131	0.65	1.00	0.85	0.86	0.08	0.14*		17.9 (1.0)
	1	253	0.60	1.04	0.85	0.82	0.08	0.04	0.47	
	2	73	0.65	1.00	0.80	0.80	0.08	0.03	0.32	
	3	15	0.54	0.89	0.71	0.73	0.08	0.07	0.81	
FACT UK societal—Kind and Macran	0	131	0.37	0.70	0.60	0.59	0.07	0.18*		37.3 (2.1)
	1	253	0.35	0.70	0.56	0.55	0.08	0.04	0.54	
	2	73	0.29	0.69	0.52	0.51	0.09	0.05	0.54	
	3	15	0.28	0.51	0.42	0.41	0.07	0.09	1.19	
FACT UK societal—Cheung et al.	0	131	0.51	0.95	0.82	0.80	0.10	0.23*		36.7 (2.1)
	1	253	0.44	0.94	0.76	0.75	0.10	0.05	0.54	
	2	73	0.42	0.92	0.72	0.70	0.12	0.05	0.46	
	3	15	0.47	0.69	0.56	0.57	0.07	0.12	1.35	

FACT, Functional Assessment of Cancer Therapy; RE, relative efficiency ratio (EQ-5D UK societal as reference); MD, mean difference; ECOG, Eastern Cooperative Oncology Group.
* Difference in mean scores between ECOG levels 0 and 3.

5 were extrapolated to a scenario that estimates a QALY benefit based on changes in ECOG status, then FACT-based algorithms would tend to underestimate the QALY benefit compared with algorithms based on the EQ-5D. For instance, using the mean scores in Table 5, an improvement in ECOG status from 3 to 0 produces a benefit of 0.45 QALYs using the Dolan EQ-5D UK societal algorithm, compared with only 0.18 QALYs using Kind and Macran's FACT UK societal algorithm. Both Kind and Macran's and Cheung et al.'s FACT-based algorithms, however, demonstrated greater relative statistical efficiency than did the other algorithms. The greater REs are a function of the relatively smaller variance associated with those algorithms. This suggests that these algorithms may provide greater power to detect differences between groups or over time in cohorts with small sample sizes.

Results supported the expectation that patient-based preference scores would provide higher estimates of health-related quality of life than would scores derived from societal preference weighted algorithms. Past studies have found that the source of preferences can influence values assigned to health states, that is, patient versus societal [17,18]. There are, however, many possible explanations for the differences in utility scores generated by each algorithm, including the attributes represented in the health state classifier, the number of levels or categories per health state dimension, the method used to elicit utilities (e.g., VAS or time trade-off), the study design/methodology, choice of utility function/modeling technique, and criteria to select the preferred statistical model. Although the conventional QALY model would dictate that health states be valued on a scale where dead must be 0 and 1 represent perfect health [1], there are differences in the calibration of measures. Studies have shown that generic utility measures such as the EQ-5D, SF-36 and health utilities index generate different QALY estimates for the same group of patients [12,19–21]. Previous studies that compared EQ-5D UK and US values sets reported differences in utility scores of the general population [22,23]. Furthermore, the FACT US patient value set was derived from TTO scores while the FACT UK societal value set was derived from VAS scores, which are scaled differently [24]. Unlike the EQ-5D, Kind and Macran's and Dobrez's FACT-based algo-

Table 6 – Correlation between preference-based scores.

	EQ-5D UK—Dolan	EQ-5D US—Shaw et al.	FACT—Dobrez et al.	FACT—Kind and Macran	FACT UK—Cheung et al.	ECOG (physician)
EQ-5D US—Shaw et al.	0.99					
FACT—Dobrez et al.	0.48	0.48				
FACT—Kind and Macran	0.68	0.68	0.69			
FACT—Cheung et al.	0.67	0.67	0.69	0.91		
ECOG (physician)	−0.34	−0.34	−0.29	−0.38	−0.37	
ECOG (patient)	−0.49	−0.49	−0.43	−0.61	−0.62	0.43

All $P < 0.01$.
ECOG, Eastern Cooperative Oncology Group; FACT, Functional Assessment of Cancer Therapy.

rithms do not generate scores that include dead (0) and full health (1), which could be considered criticisms of these models. On the other hand, methodologies used to estimate EQ-5D value sets have been criticized for the transformation of preferences for worse-than-death health states before estimation [25].

There are strengths and limitations to each of the scoring functions examined in this study. The scores based on EQ-5D UK societal and US societal value sets were similar, and the differences in the observed scores may be attributed in part to the range of each scoring function. An advantage of using the EQ-5D in cancer is that the utility scores appear to span a broad range of values and suffer no floor effects. Healthier cancer patients, however, are subject to ceiling effects. Kind and Macran's FACT UK societal value set demonstrated the best ability to discriminate among patients stratified by ECOG performance status, despite a limited range of scale scores with a ceiling utility score of 0.70 for best health possible. The discriminative ability appears to be related to the "cornerstone" approach used to estimate utility weights for each item, where disutility was estimated only for the worse level of each dimension and then equal intervals were assigned to intermediate levels of response, which resulted in an algorithm where it is possible to generate most of the numeric values along the scale continuum. Dobrez et al.'s FACT US patient algorithm consisted of only four items and also suffered from a limited range of scale, but it had an RE comparable to that of the EQ-5D scoring functions.

It is important to note that the use of an algorithm to generate patient-derived utility scores generally serve a purpose different from that served by societal preferences or QALY calculations, and may be more likely or appropriate to be applied to clinical decision making rather than for application in cost-utility analysis that informs resource allocation across health care. This is not because patient utilities lack validity per se, but because they may produce an inadvertent bias against patients in resource allocation decisions because of the "QALY trap" [26]. Specifically, because patient-derived utility scores tend to be higher than utility scores obtained from societal weights, there is less potential to capture the improvement associated with interventions that may help cancer patients.

Several additional issues are relevant to this comparison of algorithms. The study cohort consisted of patients with 11 different types of cancer and at various stages of cancer, and this may have limited generalizability to certain subgroups and cancers. This study did not utilize the full potential range of scores derived for the FACT-L UK societal algorithm [9], which included one disease-specific item that was not completed by most of the cohort, and so we assumed "no problems" as a response to that item. Thus, there is greater potential upside to the FACT-L algorithm than that observed in this study if FACT-L responses rather than FACT-G responses are available. Although the Cheung et al. algorithm performed favorably compared with the EQ-5D, this is a misleading finding. The mapping function is a biased estimator of EQ-5D scores, predicting lower scores for the better health states and higher scores for the poorer health states and smaller SDs for all states. Thus, it has better discriminative power by construction and should not be recommended in place of the measure that it attempts to predict.

The physician-based ECOG ratings had lower correlations with the preference-based scores than did patient-rated ECOG responses. This result was not surprising as the same rater, the patient, provided the ECOG rating and the health utility assessment. There is a substantial literature that acknowledges that health-care providers and significant others can provide valid rating of health, and agreement is acceptable on the more observable aspects of health and functioning [27,28]. It is unclear whether the patient or clinician rating is more desirable for the purpose of our study, because both were useful as separate measures of severity of the patients.

In conclusion, we found meaningful differences in index-based utilities scores predicted by EQ-5D UK and US societal and FACT UK societal and FACT US patient algorithms. Each algorithm has its strengths and limitations, and the preferred choice of utility algorithm will depend on the intended application. The present study, however, illustrates one of the major impediments to the uptake of disease-specific preference-based scoring functions. Unlike the EQ-5D, a generic preference-based measure of health that facilitates a standard currency for informing decisions in health systems using cost-utility analysis, alternative preference-based scoring functions create challenges for decision makers in their interpretation despite potentially more relevant disease-specific content. There is still much opportunity for the development of algorithms based on descriptive systems from disease-specific measures such as FACT that improve upon the valuation approaches and sampling methods that have been used to this point.

Source of financial support: This study was financially supported by Abbott Laboratories.

REFERENCES

[1] Weinstein MC, Torrance G, McGuire A. QALYs: the basics. Value Health 2009;12(Suppl. 1):S5–9.
[2] Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;33:337–43.
[3] Dolan P. Modeling valuations for EuroQol health states. Med Care 1997; 35:1095–108.
[4] Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care 2005;43:203–20.
[5] Pickard AS, Wilke CT, Lin HW, Lloyd A. Health utilities using the EQ-5D in studies of cancer. Pharmacoeconomics 2007;25:365–84.
[6] Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. Eur J Health Econ 2010;11:215–25.
[7] Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. Health Qual Life Outcomes 2003;1:79.
[8] Dobrez D, Cella D, Pickard AS, et al. Estimation of patient preference-based utility weights from the Functional Assessment of Cancer Therapy-General. Value Health 2007;10:266–72.
[9] Kind P, Macran S. Eliciting social preference weights for Functional Assessment of Cancer Therapy-Lung health states. Pharmacoeconomics 2005;23:1143–53.
[10] Pickard AS, De Leon MC, Kohlmann T, et al. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Med Care 2007;45:259–63.
[11] Cheung YB, Thumboo J, Gao F, et al. Mapping the English and Chinese versions of the Functional Assessment of Cancer Therapy-General to the EQ-5D utility index. Value Health 2009;12:371–6.
[12] Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 2004;13: 873–84.
[13] Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5: 649–55.
[14] Juniper EF, Guyatt GH, Jaeshke R. How to develop and validate a new health-related quality of life instrument. In: Spilker B, ed., Quality of Life and Pharmacoeconomics in Clinical Trials. Philadelphia, PA: Lippincott-Raven Publishers, 1996:chap 6.
[15] Hays RD, Anderson R, Revicki D. Psychometric considerations in evaluating health-related quality of life measures. Qual Life Res 1993; 2:441–9.
[16] Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes 2007;5:70.
[17] Ubel PA, Richardson J, Menzel P. Societal value, the person trade-off, and the dilemma of whose values to measure for cost-effectiveness analysis. Health Econ 2000;9:127–36.
[18] Boyd NF, Sutherland HJ, Heasman KZ, et al. Whose utilities for decision analysis? Med Decis Making 1990;10:58–67.
[19] Marra CA, Marion SA, Guh DP, et al. Not all "quality-adjusted life years" are equal. J Clin Epidemiol 2007;60:616–24.
[20] Pickard AS, Johnson JA, Feeny DH. Responsiveness of generic health-related quality of life measures in stroke. Qual Life Res 2005; 14:207–19.

[21] Pickard AS, Wang Z, Walton SM, Lee TA. Are decisions using cost-utility analyses robust to choice of SF-36/SF-12 preference-based algorithm? Health Qual Life Outcomes 2005;3:11.
[22] Nan L, Johnson JA, Shaw JW, Coons SJ. A comparison of EQ-5D index scores derived from the US and UK population-based scoring functions. Med Decis Making 2007;27:321–6.
[23] Shrive FM, Ghali WA, Johnson JA, et al. Use of the U.S. and U.K. scoring algorithm for the EuroQol-5D in an economic evaluation of cardiac care. Med Care 2007;45:269–73.
[24] Brazier J, Green C, McCabe C, Stevens K. Use of visual analog scales in economic evaluation. Exp Rev Pharmacoecon Outcomes Res 2003;3:293–302.
[25] Shaw JW, Pickard AS, Yu S, et al. A median model for predicting United States population-based EQ-5D health state preferences. Value Health 2010;13:278–88.
[26] Ubel PA, Nord E, Gold M, et al. Improving value measurement in cost-effectiveness analysis. Med Care 2000;38:892–901.
[27] Sneeuw KC, Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. J Clin Epidemiol 2002;55:1130–43.
[28] Sprangers MA, Sneeuw KC. Are healthcare providers adequate raters of patients' quality of life—perhaps more than we think? Acta Oncol 2000;39:5–8.