COMPARISON OF METHODS TO ESTIMATE HEALTH STATE UTILITIES USING QUALITY OF LIFE DATA

Lisa M Hess PhD[1]; William E Brady MS[2]; Laura J Havrilesky MD[3]; David E Cohn MD[4]; Bradley Monk MD[5]; Lari Wenzel PhD[6]; David Cella PhD[7]

[1]Indiana University School of Medicine, Departments of Public Health & Obstetrics and Gynecology, Indianapolis USA
[2]Gynecologic Oncology Group Statistical and Data Center, Roswell Park Cancer Institute, Buffalo USA; [3]Duke University School of Medicine, Department of Obstetrics and Gynecology, Durham USA
[4]Ohio State University School of Medicine, Department of Obstetrics and Gynecology, Columbus USA; [5]Creighton University School of Medicine, Department of Obstetrics and Gynecology, Phoenix USA
[6]University of California-Irvine School of Medicine, Irvine USA; [7]Northwestern University School of Medicine, Department of Medical Social Sciences, Chicago USA

Background

In an era of health care cost containment, cost-effectiveness analyses (CEA) are increasingly needed to inform decisions about care. Cancer clinical trials in the US often collect quality of life (QOL) data using the FACT instruments, but only rarely include utility instruments for CEA. Two algorithms (Cheung and Dobrez) have been developed using the FACT to map QOL to a utility value. Other researchers have looked at the total QOL FACT and linearly transformed it to a 0-1.0 scale. It is unknown how they compare when applied to a gynecologic cancer population, information which is necessary if CEA are to use FACT data in their estimation of the quality-adjusted effectiveness of treatments in this population.

Methods

GOG-0152 was a 550-patient randomized phase III trial comparing interval cytoreduction versus no interval cytoreduction and GOG-0172 was a 415-patient randomized phase III trial comparing intravenous (IV) cisplatin plus IV paclitaxel versus intraperitoneal cisplatin plus IV paclitaxel among women with advanced ovarian cancer. QOL data were collected at four time points in each study. The mapping algorithms and data transformation equations were applied to these data (Figure 1). The possible range for the Cheung scale is 0.238 to 0.998, the Dobrez scale is 0.4556 to 1.0, and the linearly transformed score possible range is 1.0 to 1.0. The agreement between measures was assessed by the concordance correlation coefficient (ρ_{CCC}) and paired t-tests were used to compare means between the estimated utility values.

$$Utility_{Cheung} = 0.238 + 0.014\times PWB + 0.006\times EWB + 0.008\times FWB$$

$$Utility_{Dobrez} = 1 + (-0.2222\times Q1_{[0,1]} - 0.1137\times Q1_{[2,3]}) + (-0.1537\times Q2_{[0]}) + (-0.0431\times Q3_{[0,1]}) + (-0.1254\times Q4_{[0,1]} - 0.0641\times Q4_{[2]} - 0.0345\times Q4_{[3]})$$

$$Utility_{Linear} = \text{total FACT QOL score}/156$$

Figure 1. Utility score calculations

PWB=physical well being; EWB=emotional well being; FWB=functional well being; Q1=lack of energy (PWB); Q2=feel sick (PWB) ; Q3=able to work (FWB); Q4=able to enjoy life (FWB)

Results

Results using the Cheung and FACT scores were highly continuous, while the Dobrez score was not; there are only 48 possible values of the Dobrez score. The mean values by study protocol for each method are presented in Table 1.

Table 1. Utility scores by study protocol and time point

		GOG-0152				GOG-0172			
	Time point	N	Mean	Min	Max	N	Mean	Min	Max
Cheung Utility	1	373	0.79	0.33	1.00	397	0.73	0.31	1.00
	2	345	0.78	0.34	1.00	320	0.72	0.38	0.99
	3	342	0.84	0.44	1.00	330	0.76	0.34	1.00
	4	302	0.84	0.42	1.00	276	0.84	0.38	1.00
	ALL	1362	0.81	0.33	1.00	1323	0.76	0.31	1.00
Dobrez Utility	1	363	0.82	0.46	1.00	385	0.79	0.46	1.00
	2	344	0.81	0.46	1.00	313	0.77	0.46	1.00
	3	334	0.86	0.61	1.00	323	0.79	0.46	1.00
	4	301	0.87	0.46	1.00	273	0.87	0.46	1.00
	ALL	1342	0.84	0.46	1.00	1294	0.80	0.46	1.00
FACT Linear Utility	1	371	0.75	0.34	0.96	396	0.70	0.22	0.99
	2	342	0.74	0.38	0.99	320	0.70	0.38	0.96
	3	340	0.81	0.35	1.00	330	0.74	0.34	0.99
	4	301	0.81	0.39	0.99	276	0.81	0.40	1.00
	ALL	1354	0.77	0.34	1.00	1322	0.73	0.22	1.00

Concordance (ρ_{CCC}) between estimation methods ranged from 0.60 to 0.93 at the various time points in the study (Table 2).

Table 2. Agreement measures

Study	Time point	N	Comparison	Pccc (95%CI)	Pearson correlation
GOG-0152	1	362	C vs D	0.77 (0.72, 0.80)	0.82
			C vs F	0.89 (0.87, 0.91)	0.94
			D vs F	0.62 (0.56, 0.67)	0.76
	2	229	C vs D	0.79 (0.75, 0.83)	0.84
			C vs F	0.90 (0.88, 0.92)	0.94
			D vs F	0.66 (0.60, 0.71)	0.78
	3	333	C vs D	0.79 (0.74, 0.82)	0.80
			C vs F	0.89 (0.87, 0.91)	0.94
			D vs F	0.64 (0.58, 0.69)	0.73
	4	297	C vs D	0.81 (0.77, 0.84)	0.84
			C vs F	0.91 (0.88, 0.92)	0.95
			D vs F	0.69 (0.63, 0.74)	0.79
GOG-0172	1	384	C vs D	0.72 (0.67, 0.76)	0.81
			C vs F	0.92 (0.90, 0.93)	0.94
			D vs F	0.60 (0.55, 0.65)	0.76
	2	313	C vs D	0.74 (0.69, 0.78)	0.83
			C vs F	0.93 (0.92, 0.95)	0.94
			D vs F	0.64 (0.57, 0.69)	0.76
	3	323	C vs D	0.81 (0.77, 0.84)	0.85
			C vs F	0.93 (0.91, 0.94)	0.94
			D vs F	0.69 (0.63, 0.74)	0.78
	4	272	C vs D	0.79 (0.74, 0.83)	0.81
			C vs F	0.90 (0.88, 0.92)	0.93
			D vs F	0.63 (0.56, 0.69)	0.71

D=Dobrez algorithm
C=Cheung algorithm
F=FACT linear transformation

Results, continued

However, there were statistically significant (p<0.001) differences between the scores at all time points in both protocols: mean scores were slightly higher with Dobrez than with Cheung , and Cheung was higher than the linearly transformed score (Figure 2). Differences in performance status, response to treatment, and stage of disease did not help to explain the differences in utility scores.

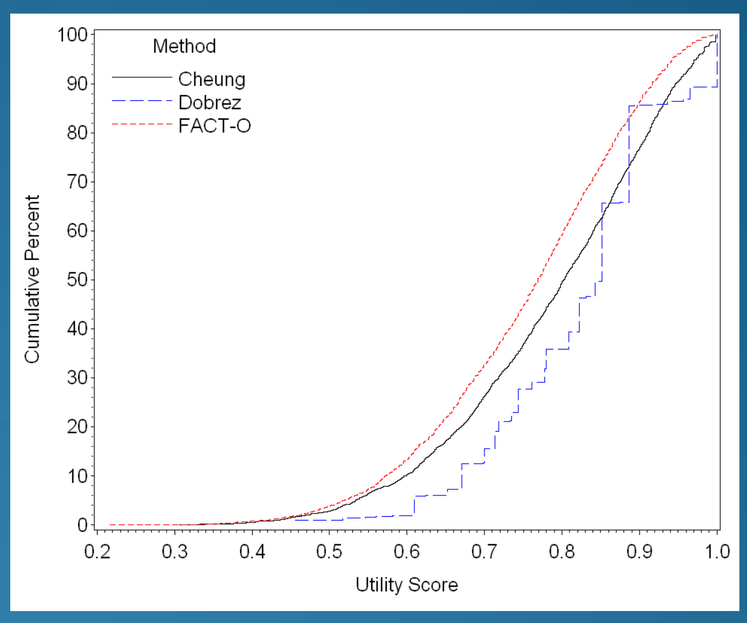

Figure 2. Cumulative distribution of utility scores

Conclusions

If CEAs incorporate mapping algorithms to obtain utility estimates in the absence of prospectively collected utility data in the setting of ovarian cancer, investigators should take into account the variability and differences in estimates depending on the algorithm selected. Interpretations of CEA have the potential to vary given the significant difference between these two algorithms. Future work is needed to assess if these algorithms obtain estimates that are comparable to values obtained by standard utility instruments in this population.

References:

Cheung Y-B, et al. Mapping the English and Chinese Versions of the Functional Assessment of Cancer Therapy–General to the EQ-5D Utility Index. *Value in Health*. 2009;**12**:371–376.

Dobrez D, et al. Estimation of Patient Preference-Based Utility Weights from the Functional Assessment of Cancer Therapy—General. *Value in Health*. 2007;**10**:266-272.