Published in final edited form as:

Gynecol Oncol. 2013 February ; 128(2): 175–180. doi:10.1016/j.ygyno.2012.10.024.

Comparison Of Methods To Estimate Health State Utilities For Ovarian Cancer Using Quality of Life Data: A Gynecologic Oncology Group Study

Lisa M. Hess, PhD¹, William E. Brady, MS², Laura J. Havrilesky, MD³, David E. Cohn, MD⁴, Bradley J. Monk, MD⁵, Lari Wenzel, PhD⁶, and David Cella, PhD⁷

¹Indiana University School of Medicine; Dept. of Public Health & Obstetrics and Gynecology; Indianapolis, IN 46202

²Gynecologic Oncology Group Statistical & Data Center; Roswell Park Cancer Institute; Buffalo, NY 14263

³Duke University School of Medicine; Dept. of Obstetrics and Gynecology; Durham, NC 27710

⁴Ohio State University School of Medicine; Dept. of Obstetrics and Gynecology; Columbus, OH 43026

⁵Creighton University School of Medicine; Dept. of Obstetrics and Gynecology; Phoenix, AZ 85013

⁶University of California Irvine; Dept. of Medicine and Public Health; Irvine, CA 92697

⁷Northwestern University Feinberg School of Medicine; Dept. of Medical Social Sciences; Chicago, IL 60611

Abstract

Background—Cost-effectiveness/cost-utility analyses are increasingly needed to inform decisions about care. Algorithms have been developed using the Functional Assessment of Cancer Therapy (FACT) quality of life instrument to estimate utility weights for cost analyses. This study was designed to compare these algorithms in the setting of ovarian cancer.

Methods—GOG-0152 was a 550-patient randomized phase III trial of interval cytoreduction, and GOG-0172 was a 415-patient randomized phase III trial comparing intravenous versus intraperitoneal therapy among women with advanced ovarian cancer. QOL data were collected via the FACT at four time points in each study. Two published mapping algorithms (Cheung and Dobrez) and a linear transformation method were applied to these data. The agreement between measures was assessed by the concordance correlation coefficient (r_{CCC}), and paired t-tests were used to compare means.

Results—While agreement between the estimation algorithms was good (ranged from 0.72 to 0.81), there were statistically significant ($p < 0.001$) and clinically meaningful differences between

© 2012 Elsevier Inc. All rights reserved.

Corresponding Author: Lisa M. Hess PhD, Indiana University School of Medicine, 714 N. Senate Ave., EF250, Indianapolis, IN 46202, Phone: 317-274-3148, Fax: 317-274-3443, lmhess@iupui.edu.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

the scores: mean scores were higher with Dobrez than with Cheung or the linear transformation method.. Scores were also statistically significantly different ($p<0.001$) between studies.

Conclusions—In the absence of prospectively collected utility data, the use of mapping algorithms is feasible, however, the optimal algorithm is not clear. There were significant differences between studies, which highlights the need for validation of these algorithms in specific settings. If cost analyses incorporate mapping algorithms to obtain utility estimates, investigators should take the variability into account.

Keywords

health utilities; ovarian cancer; quality of life; methodology; comparative effectiveness research

INTRODUCTION

In response to U.S. Congressional and Public Health Service interest in comparative effectiveness research (CER), there has been a substantial increase in funding to conduct research that compares the “real world” value of standard and experimental treatments for diseases that impact the nation’s health. One common approach to comparing the value of competing treatments is to conduct cost-effectiveness analyses (CEAs) or cost-utility analyses (CUA), which explicitly incorporate quality of life into the value equation. Such analyses are enhanced by the ability to place a value on the health status of patients receiving the treatments being compared.

In the U.S., the cost of cancer is escalating, leading to increased research to identify the treatment strategies that are associated with the most value to patients and payers. One limitation in such analyses is the lack of utility weight data specific to the treatment regimens and outcomes being compared. This is particularly true for the less common cancers, such as ovarian cancer, and unless prospective data are collected within the trial, the utility weights associated with outcomes related to experimental regimens are unknown.

Utility weights have been catalogued and identified for many chronic diseases, including various cancers. It is well established that utility weights can vary depending on the respondent and method of utility elicitation; moreover cancer is not one disease, but rather a set of very heterogeneous diseases. Therefore, it is not unexpected that utility weights differ by the site of cancer origin, and have additionally been found to vary based on the clinical presentation within each disease site, such as stage at diagnosis and cell type (i.e. epithelial cell, germ cell, stromal cell), which reflects the heterogeneity within the term ‘cancer’ and within each type of cancer [1,2].

In the U.S., the most common instrument used to assess quality of life in ovarian cancer trials is the Functional Assessment of Cancer Therapy (FACT), which is part of the body of Functional Assessment in Chronic Illness (www.facit.org) patient reported outcomes system [3].

Cost-utility analyses (CUAs) can consider quality of life (QOL) outcomes by calculating a quality-adjusted life year (QALY) by multiplying time in years (e.g. survival time) by a utility weight specific to the condition or treatment of interest. With the increasing cost of cancer care, the accuracy and availability of CEA or CUA are essential to decisions about programmatic and individual research or health care spending. Although survival and quality of life data are frequently captured in cancer clinical trials, only rarely are appropriate data collected to calculate a QALY. As a result, a number of algorithms and estimation methods have come into use by which FACT data may be used to estimate the associated utility weight for patient outcomes [4,5]. The ability to use existing quality of life

data to predict utility values would allow researchers to apply data that have already been collected in large clinical trials to the performance of accurate cost-effectiveness analyses and other comparative effectiveness research. However, if the QALYs derived using the algorithms differ, the conclusions of a cost-effectiveness analysis are also likely to vary regarding the optimal treatment strategy depending on the method applied to the data.

While some investigators have developed FACT-based algorithms for other specific common cancers [7,8], there is no known algorithm specific to ovarian cancer utility weight estimation. As stated earlier, cancer is a term for a very heterogeneous set of diseases, therefore, it is important that disease-specific analyses are conducted when evaluating the value of outcomes for any condition, and utility weights specific to that disease are needed for CUA.

CUAs of treatment for ovarian cancer have been required to estimate utility weights for analysis due to the lack of existing population-based utility weight data for ovarian cancer. The methods that have been used in these studies have primarily included various methods of utility estimation from quality of life data collected by the FACT [6,9,10] or estimations based on other conditions [11]. However, due to the lack of information about how the estimation methods compare, the outcomes could be biased due to the lack of reliable utility weights for inclusion in CUAs. Several other studies have obtained utility values from ovarian cancer patients using the EQ-5D or time trade off (TTO) and standard gamble methods for hypothetical states, but these data are from very small samples (less than 100) and have yet to be applied to CUAs [12–14]. Additional details of the methods and theory of the standard gamble and TTO methods are available elsewhere [21].

There are three known approaches to value health as measured by the FACT-General (FACT-G) scale, making it possible to directly compare treatments for CER [4–6]. Dobrez, et al.[4] developed an algorithm to estimate a utility weight based on the association between the FACT and TTO valuation of the patient's own health in a retrospective study of breast cancer (n=250), prostate cancer (n=180) cell lung cancer (n = 146), head and neck cancer (n = 164), and non-Hodgkin's lymphoma (n = 148) patients. The TTO valuation differs from other utility estimates in that it does not elicit a societal valuation of the patient's health state as recommended for CUA [21] but rather elicits the valuation directly from the patient. A second study by Cheung and colleagues mapped the FACT to a validated health utility instrument, the EQ-5D, using the preferences of 558 patients with predominately breast (37.1%), head and neck (18.6%), and colorectal cancers (10.9%) in China [5]. The EQ-5D (similar to other health utility instruments such as the HUI) elicits the health state directly from the patient, but the utility weight associated with that health state is based on societal valuation and is considered a standard methodology for CUA. Others have linearly transformed the 0–156 range of possible FACT scores to the 0–1.0 range of a health utility within CUAs [6,10]. While this particular method is not based on underlying theories or methodologies, it nevertheless has been applied to at least two ovarian cancer CUAs due to the absence of available utility weight data. To date there is no information as to which of these methods is better, or whether they would produce similar results if applied to the same clinical trial data within a specific disease setting. It is unknown how these algorithms or methods perform when applied to FACT data collected from patients with ovarian cancer, since 0% and 6% of the participants, respectively, in the Dobrez and Cheung studies to develop these algorithms, had a diagnosis of a gynecologic cancer, and the FACT linear transformation method has yet to be tested.

This study was designed to apply these three utility estimation methods to data collected from large phase III trials of ovarian cancer. The goal was to better understand the relationship between QOL as measured by the FACT tool and health utility estimates if

CUAs continue to use FACT data in their estimation of the quality-adjusted effectiveness of treatments for ovarian cancer. This retrospective analysis was designed to better understand the implications of using these algorithms in gynecologic cancer cost-effectiveness research.

METHODS

GOG Studies

GOG-0152 was a 550-patient, randomized, phase III trial comparing interval cytoreduction *versus* no interval cytoreduction in patients with advanced ovarian cancer after initial surgery and combination chemotherapy with intravenous (IV) paclitaxel plus IV cisplatin [15,16]; 424 of the 550 patients were randomized after completing postoperative chemotherapy. GOG-0172 was a 415-patient randomized, phase III trial comparing IV paclitaxel plus IV cisplatin *versus* intraperitoneal (IP) cisplatin plus IV paclitaxel in patients with stage III ovarian cancer [17]. The demographic and disease characteristics of study participants are included in Table 1. The patient demographics were very similar, in that the eligibility criteria had similar requirements; however, there are important differences in the amount of residual disease following surgery that were permitted on each study. GOG-0152 required that patients be suboptimally debulked (at least 1 cm of residual disease was required for study eligibility), whereas GOG-0172 only enrolled patients that were optimally debulked, with no more than 1cm of residual disease.

The FACT-G was administered in both studies. GOG-0152 used an older version of the FACT-G than did GOG-0172; there were five items in the emotional well-being subscale in GOG-0152, while there were six in GOG-0172 (which matches the number used in the Cheung algorithm as described below). In the calculation of this subscale for GOG-0152, the mean across the five items was then multiplied by six. The FACT was administered at four time points in each study. In GOG-0152 it was administered after three cycles of chemotherapy—prior to randomization, at the sixth cycle of chemotherapy, and six and twelve months after the start of treatment, and in GOG-0172 it was administered prior to randomization, prior to the fourth cycle of chemotherapy, and 3–6 weeks and 12 months after treatment.

Utility Weight Estimates

At least 50% of the questions on a subscale had to be answered in order to calculate the subscale score, and at least 80% of the total questions had to be answered in order to calculate the total score. The utility score estimate developed by Cheung, et al.(5) is calculated as follows:

$$Utility_{Cheung} = 0.238 + 0.014 \times GP + 0.006 \times GE + 0.008 \times GF$$

where *GP*, *GE*, and *GF* are the physical, emotional, and functional well-being subscales, respectively. These subscales have 7, 6, and 7 items, respectively, so the range for the possible values of the Cheung utility score is 0.238 to 0.998.

The utility score developed by Dobrez, et al.[4] is calculated as follows:

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

where $Q1$ = physical well-being: lack of energy, $Q2$ = physical well-being: feel sick, $Q3$ = functional well-being: able to work, and $Q4$ = functional well-being: able to enjoy life, with all questions ordered from 0 indicating worst quality of life to 4 indicating the best, and with the numbers in sub-brackets above indicating “equal to.” The range for the possible values of the Dobrez utility score is 0.4556 to 1.0.

A utility weight based on the full FACT-O was also calculated using a linear transformation. The maximum total score on the FACT-O is 156 (=39 questions $\times$ 4 points/question), so the $Utility_{FACT-O_156} = \text{FACT-O total score}/156$, and the possible range for this utility value is 0 to 1.0.

Statistical Analysis

Summary statistics, histograms, and cumulative distribution plots were conducted for each utility estimation method. The agreement between measures was assessed by the concordance correlation coefficient (CCC or r_{CCC}), [18] which assesses differences in means (μ_C and μ_D) and variances (σ_C^2 and σ_D^2) between the scores and the Pearson correlation (ρ , which measures the degree of linear association between the two measures). The CCC ranges from -1 to 1 and is equal to 1 (indicating perfect agreement) if and only if: the means for the two scores are the same ($\mu_C = \mu_D$), and the variances for the two scores are the same ($\sigma_C^2 = \sigma_D^2$), and $\rho = 1$. The CCC is defined as:

$$\rho_{CCC} = \frac{2\sigma_{C,D}}{\sigma_C^2 + \sigma_D^2 + (\mu_C - \mu_D)^2},$$

where $\sigma_{C,D}$ is the covariance between the two scores. The sample CCC is calculated as:

$$CCC = r_{CCC} = \frac{2s_{C,D}}{s_C^2 + s_D^2 + (\bar{x}_C - \bar{x}_D)^2}$$

where $s_{C,D}$ is the sample covariance between the two scores and s_C^2 and s_D^2 are the sample variances for the Cheung and Dobrez scores, and $\bar{x}_C$ and $\bar{x}_D$ are the sample means. The calculation of the standard error for the CCC and the resulting calculation of confidence intervals is described in Lin, 1989 [18]. Paired t-tests were used to compare means for the two scores.

Repeated measures mixed models were used to examine differences between the utility scores across the four time points and to test whether these differences were significantly different between levels of these other factors: protocol, performance status, stage, response (GOG-0152 only). Three covariance structures for the repeated measures (AR[1], compound

symmetric, and unstructured) were assessed, and the Akaike Information Criteria (AIC) was used to choose the structure. Based on the results, a compound symmetric covariance structure was used for all analyses.

RESULTS

FACT data were available from 746 (89%) patients enrolled to GOG-0152/GOG-0172 for analysis of all three estimation methods at time point 1 (baseline). Over time, the number of participants with FACT data declined due to study withdrawal, incomplete study forms, or due to morbidity or mortality. At time point 4, data were available from 569 patients, 76% of those who had data available for all three methods at baseline.

The distribution of the estimated utility scores using the Cheung method and linear transformation were highly continuous, while the Dobrez score was not; there are only 48 possible values of the Dobrez score. Figure 1 shows histograms of each utility score, and Figure 2 shows the cumulative distribution of each method to estimate utility weights for both protocols across all time points. These figures demonstrate how the distribution of scores using the Dobrez method differs from both the Cheung and linear transformation methods. The mean values by study protocol for each method are presented in Table 2. The difference in mean utility estimation methods ranged from 0.018 to 0.083 at the various time points. There were statistically significant ($p < 0.001$) pairwise differences between the three scores at each time point in both protocols, and the 95% confidence intervals for the CCCs did not include 1, i.e., were all significantly different from 1 (Table 3).

Table 4 presents mean differences (across the four time points) between factors (protocol, performance status, stage, and response) for each difference in utility scores. The differences between estimation methods across factors ranged from 0.02 to 0.11. The mean differences between methods were significantly different by study protocol. The Dobrez algorithm estimated a higher utility than the other methods in both studies, but the difference was greater in the GOG-0172 study than in the GOG-0152 study. There were no statistically significant differences in the performance of the utility estimates by tumor response or by stage of disease at diagnosis. In the GOG-172 study, differences between the Dobrez algorithm and the other two methods increased with increasing performance status score (i.e., with decreasing performance status).

CONCLUSIONS

While there was substantial agreement (as measured by CCCs) between the utility weights, the CCCs were statistically significantly different from 1, and the utility means were statistically significantly different from each other. This suggests that CUAs using these estimation methods algorithms could come to very different conclusions, depending on the method selected. For example, the overall cost and effectiveness of IP therapy (undiscounted) was \$39,861 and 5.16 QALYs, compared with IV therapy at a cost of \$18,822 and 4.59 QALYs with an incremental cost effectiveness ratio (ICER) of \$37,454/QALY[6]. Hypothetically, if a difference of 0.08 in utility weight estimates were to be applied to this study's estimate for the IP arm (the IV arm was not weighted with a utility value), QALYs would be 4.75 for IP therapy, resulting in a new ICER in excess of \$130,000, much exceeding the willingness-to-pay threshold. Alternatively, if the error in utility weights were an underestimation, the analysis may have shown IP therapy to be a cost-saving alternative compared to IV therapy. The method and approach for utility estimation must be used with caution if decisions for care are to be based on CUAs using these approaches.

While the Dobrez algorithm was less continuous (due to fewer possible scores) than the other two methods, it is not clear if this would translate to a less accurate representation of utility. It is possible due to the fact that only four of the 27 items on the FACT were used by the Dobrez method that higher scores were obtained as compared to the other methods. However, the accuracy of the Cheung or linear methods compared to standard instruments (e.g. EQ-5D, Health Utility Index or HUI) remains unknown in the setting of ovarian cancer. Other researchers have compared these algorithms to the EQ-5D across a diverse set of cancers, and similar to the current study, also found significant differences between methods [20].

A limitation of this work is that neither GOG-152 nor GOG-172 collected prospective utility data. Therefore, there is insufficient evidence at this time to recommend any one method over another due to the fact that these estimations could not be compared to a standard methodology that elicits societal valuations of health, as recommended for CUA [19]. Many instruments that elicit patient-reported outcomes have associated societal utility values and are available for incorporation into clinical trials (e.g. EQ-5D, HUI). At a minimum, estimation methods must be validated against these accepted societal-based preference instruments in the target population prior to implementation in CUAs. There is a need to conduct a prospective study collecting both FACT and a standard societal utility-based instrument so that these algorithms can be assessed in comparison to accepted methodologies. If researchers choose to incorporate mapping algorithms to obtain utility estimates for CUA in the setting of ovarian cancer, it is necessary to take into account the variability and differences in estimates (as well as 95% confidence interval of the estimates) for sensitivity analyses, depending on the algorithm selected. Interpretations of CUA have the potential to vary given the significant difference between these mapping algorithms, and it should be noted that these methods have yet to be prospectively compared with a standard utility instrument in this population. In the interim, survival or life years gained may be the only reliable option for CEAs in this population for retrospective studies. Future clinical trials that include patient-reported outcomes should strongly consider incorporating a process of prospective collection of utility data with the EQ-5D, HUI or other standard instrument for CUAs.

Acknowledgments

This study was supported by National Cancer Institute grants to the Gynecologic Oncology Group Administrative Office (CA 27469), the Gynecologic Oncology Group Statistical and Data Center (CA 37517). The following institutions participated in this study: University of Alabama at Birmingham, Duke University Medical Center, Abington Memorial Hospital, Walter Reed Army Medical Center, Wayne State University, University of Minnesota Medical School, Colorado Gynecologic Oncology Group, UCCC, University of Mississippi Medical Center, Colorado Foundation for Medical Care, University of California Medical Center at Los Angeles, University of Washington Medical Center, University of Pennsylvania Cancer Center, Milton S. Hershey School of Medicine of the Pennsylvania State University, University of Cincinnati College of Medicine, University of North Carolina School of Medicine, University of Iowa Hospitals and Clinics, University of Texas Health Science Center at Dallas, Indiana University Cancer Center, Wake Forest University School of Medicine, Albany Medical College, University of California Medical Center at Irvine, Tufts-New England Medical Center, Rush University Medical Center, State University of New York Downstate Medical Center, University of Kentucky, Cleveland Clinic Foundation, State University of New York at Stony Brook, Washington University School of Medicine, Cooper Hospital/University Medical Center, Columbus Cancer Council, M.D. Anderson Cancer Center, University of Massachusetts Memorial Medical Center, Fox Chase Cancer Center, Medical University of South Carolina, Women's Cancer Center, University of Oklahoma Health Science Center, University of Virginia Health Science Center, Tacoma General Hospital, Thomas Jefferson University Hospital, Mayo Clinic, Tampa Bay Cancer Consortium, Gynecologic Oncology Network/Brody School of Medicine, Ellis Fischel Cancer Center.

References

1. Sturza J. A review and meta-analysis of utility values for lung cancer. *Medical decision making: an international journal of the Society for Med Dec Making*. 2010; 30:685–93.

2. Brennan, V.; Wolowacz, S. A systematic review of breast cancer utility weights. ISPOR; 13th Annual Meeting; 2008. Available at: <http://www.rtihs.org/request/index.cfm?fuseaction=display&PID=10711>
3. Hess, LM.; Stehman, FK. A systematic review. International journal of gynecological cancer: official journal of the International Gynecological Cancer Society; 2012. State of the Science in Ovarian Cancer Quality of Life Research.
4. Dobrez D, Cella D, Pickard AS, Lai JS, Nickolov A. Estimation of patient preference-based utility weights from the functional assessment of cancer therapy - general. Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research. 2007; 10:266–72. [PubMed: 17645681]
5. Cheung YB, Thumboo J, Gao F, Ng GY, Pang G, Koo WH, 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. [PubMed: 18783392]
6. Bristow RE, Santillan A, Salani R, Diaz-Montes TP, Giuntoli RL 2nd, Meisner BC, et al. Intraperitoneal cisplatin and paclitaxel versus intravenous carboplatin and paclitaxel chemotherapy for Stage III ovarian cancer: a cost-effectiveness analysis. Gynecol Oncol. 2007; 106:476–81. [PubMed: 17688927]
7. Kind P, Macran S. Eliciting social preference weights for Functional Assessment of Cancer Therapy-Lung health states. Pharmacoeconomics. 2005; 23:1143–53. [PubMed: 16277549]
8. Wu EQ, Mulani P, Farrell MH, Sleep D. Mapping FACT-P and EORTC QLQ-C30 to patient health status measured by EQ-5D in metastatic hormone-refractory prostate cancer patients. Value Health. 2007; 10:408–14. [PubMed: 17888106]
9. Havrilesky LJ, Pokrzywinski R, Revicki D, Higgins RV, Nycum LR, Kohler MF, et al. Cost-effectiveness of combination versus sequential docetaxel and carboplatin for the treatment of platinum-sensitive, recurrent ovarian cancer. Cancer. 2012; 118:386–91. [PubMed: 21598242]
10. Havrilesky LJ, Secord AA, Darcy KM, Armstrong DK, Kulasingam S. Cost effectiveness of intraperitoneal compared with intravenous chemotherapy for women with optimally resected stage III ovarian cancer: a Gynecologic Oncology Group study. J Clin Oncol. 2008; 26:4144–50. [PubMed: 18757328]
11. Lesnock JL, Farris C, Krivak TC, Smith KJ, Markman M. Consolidation paclitaxel is more cost-effective than bevacizumab following upfront treatment of advanced epithelial ovarian cancer. Gynecol Oncol. 2011; 122:473–8. [PubMed: 21665250]
12. Havrilesky LJ, Broadwater G, Davis DM, Nolte KC, Barnett JC, Myers ER, et al. Determination of quality of life-related utilities for health states relevant to ovarian cancer diagnosis and treatment. Gynecol Oncol. 2009; 113:216–20. [PubMed: 19217148]
13. Hess LM, Malone DC, Skrepnek GH, Reed PG, Armstrong E, Coons SJ. Oncologist preferences for health states associated with the treatment of advanced ovarian cancer. Appl Health Econ Health Policy. 2010; 8:217–23. [PubMed: 20578777]
14. Hess LM, Malone DC, Reed PG, Skrepnek GH, Weihs KJ. Preferences of patients and oncologists for health states associated with the treatment of advanced ovarian cancer. Health Outcomes Res Med. 2010; 1:e51–e9.
15. Rose PG, Nerenstone S, Brady MF, Clarke-Pearson D, Olt G, Rubin SC, et al. Secondary surgical cytoreduction for advanced ovarian carcinoma. N Engl J Med. 2004; 351:2489–97. [PubMed: 15590951]
16. van der Burg ME, van Lent M, Buyse M, Kobierska A, Colombo N, Favalli G, et al. The effect of debulking surgery after induction chemotherapy on the prognosis in advanced epithelial ovarian cancer. Gynecological Cancer Cooperative Group of the European Organization for Research and Treatment of Cancer. New Engl J Med. 1995; 332:629–34. [PubMed: 7845426]
17. Armstrong DK, Bundy B, Wenzel L, Huang HQ, Baergen R, Lele S, et al. Intraperitoneal cisplatin and paclitaxel in ovarian cancer. New Engl J Med. 2006; 354:34–43. [PubMed: 16394300]
18. Lin LI. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989; 45:255–68. [PubMed: 2720055]
19. Drummond M. Introducing economic and quality of life measurements into clinical studies. Ann Med. 2001; 33:344–9. [PubMed: 11491193]

20. Pickard AS, Ray S, Ganguli A, Cella D. Comparison of FACT- and EQ-5D-based utility scores in cancer. *Value Health*. 2012; 15:305–11. [PubMed: 22433762]
21. Drummond, MF.; Sculpher, MJ.; Torrance, GW.; O'Brian, BJ. *Methods for the economic evaluation of health care programmes*. Oxford University Press; USA: 2005.

\$watermark-text

\$watermark-text

\$watermark-text

Research Highlights

- Cost analyses of ovarian cancer treatment could be biased if estimation methods are used to measure quality-adjusted life years.
- Comparisons of several utility-estimation methods found significant differences ($p < 0.001$) between the utility values from these methods.
- There's a need to validate utility estimation methods before they can be recommended for cost analyses in ovarian cancer.

\$watermark-text

\$watermark-text

\$watermark-text

Figure 1. Histograms of each utility weight estimation method (both protocols, all time points): Dobrez (left); Cheung (center); and the FACT linear transformation (right)

Figure 2. Cumulative distribution of utility weights as measured by the Cheung algorithm, Dobrez algorithm, and the FACT linear transformation method (both protocols, all time points)

Table 1

Characteristics of study participants

Enrolled Treatment group	GOG-0152 N=550		GOG-0172 N=415	
	Secondary surgery	Chemotherapy alone	Intravenous chemotherapy	Intraperitoneal chemotherapy
Eligible, randomized	N=216	N=208	N=210	N=205
Age - median (range)	58.1 (25–82)	57.0 (27–82)	57.5 (31–84)	57.7 (25–84)
Performance status – n (%)				
0	83 (38%)	83 (40%)	90 (43%)	91 (44%)
1	119 (55%)	108 (52%)	112 (53%)	99 (48%)
2	14 (6%)	17 (8%)	8 (4%)	15 (7%)
Cell type – n (%)				
Serous	165 (76%)	159 (76%)	170 (81%)	158 (77%)
Endometrioid	17 (8%)	11 (5%)	12 (6%)	17 (8%)
Clear Cell	4 (2%)	3 (1%)	9 (4%)	11 (5%)
Mixed	20 (9%)	17 (8%)	11 (5%)	14 (7%)
Other/unspecified	10 (5%)	18 (8%)	8 (4%)	5 (2%)
Stage – n (%)				
III	200 (93%)	200 (96%)	210 (100%)	205 (100%)
IV	16 (7%)	8 (4%)	0	0
Residual disease – n (%)				
None visible	0	0	74 (36%)	78 (38%)
1cm	0	0	135 (64%)	127 (62%)
1–2 cm	27 (12%)	26 (12%)	0	0
2.1–5cm	92 (43%)	91 (44%)	0	0
>5cm	97 (45%)	91 (44%)	0	0

\$watermark-text

\$watermark-text

\$watermark-text

Table 2

Utility scores by study protocol and time point

		GOG-0152				GOG-0172			
	Time point	N	Mean	Min	Max	N	Mean	Min	Max
Cheung Algorithm	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 Algorithm	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 Transformation	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

Table 3

Measures of agreement between estimation methods

Study	Time point	N	Comparison	r_{ccc} (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
			C vs D	0.79 (0.75, 0.83)	0.84
	2	229	C vs F	0.90 (0.88, 0.92)	0.94
			D vs F	0.66 (0.60, 0.71)	0.78
			C vs D	0.79 (0.74, 0.82)	0.80
			C vs F	0.89 (0.87, 0.91)	0.94
	3	333	D vs F	0.64 (0.58, 0.69)	0.73
			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
			C vs D	0.74 (0.69, 0.78)	0.83
	2	313	C vs F	0.93 (0.92, 0.95)	0.94
			D vs F	0.64 (0.57, 0.69)	0.76
			C vs D	0.81 (0.77, 0.84)	0.85
			C vs F	0.93 (0.91, 0.94)	0.94
	3	323	D vs F	0.69 (0.63, 0.74)	0.78
			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

Table 4

Mean (SE) differences in utility scores across all time points, by factors

Factor	Cheung minus Dobrez	Cheung minus FACT linear transformation	Dobrez minus FACT linear transformation
Protocol			
GOG-0152	−0.0278 (0.00258)	0.0365 (0.00156)	0.0641 (0.00300)
GOG-0172	−0.0424 (0.00260)	0.0250 (0.00157)	0.0676 (0.00302)
<i>p</i> -value	<0.001	<0.001	<0.001
Performance Status			
GOG-0152			
0	−0.0274 (0.00396)	0.0357 (0.00239)	0.0630 (0.00455)
1	−0.0288 (0.00347)	0.0359 (0.00209)	0.0642 (0.00399)
2	−0.0226 (0.00958)	0.0463 (0.00577)	0.0693 (0.01098)
<i>p</i> -value	0.821	0.212	0.867
GOG-0172			
0	−0.0359 (0.00394)	0.0252 (0.00244)	0.0613 (0.00464)
1	−0.0440 (0.00370)	0.0251 (0.00228)	0.0692 (0.00436)
2	−0.0839 (0.01204)	0.0223 (0.00733)	0.1074 (0.01409)
<i>p</i> -value	<0.001	0.931	0.007
Stage (GOG-0152 only)			
3	−0.0277 (0.00258)	0.0362 (0.00156)	0.0636 (0.00297)
4	−0.0292 (0.01079)	0.0433 (0.00654)	0.0719 (0.01242)
<i>p</i> -value	0.897	0.286	0.519
Tumor Response to Treatment (GOG-0152 only)			
Complete response	−0.0167 (0.00567)	0.0386 (0.00349)	0.0553 (0.00650)
Partial response	−0.0285 (0.00431)	0.0352 (0.00265)	0.0638 (0.00493)
Stable disease	−0.0265 (0.00549)	0.0344 (0.00337)	0.0609 (0.00628)
<i>p</i> -value	0.242	0.659	0.580