Validity of utilities of patients with esophageal cancer

Peep F. M. Stalmeier
University Medical Centre Nijmegen

Angela G. E. M. de Boer, Mirjam A. G. Sprangers, Hanneke C. J. M. de Haes, Jan J. B. van Lanschot
Academic Medical Centre/University of Amsterdam

Objectives: The convergent validity between utility assessment methods was assessed.
Methods: Investigated were patients with esophageal cancer treated surgically with curative intent. Patients were interviewed in a period from 3 to 12 months after surgical resection. Patients evaluated their actual health and seven other states. Visual analogue scale (VAS) and standard gamble (SG) utilities were obtained for the health states in an interview. Patients also indicated whether or not they preferred death to living in a health state (worse than dead [WTD] preferences).
Results: Fifty patients completed the interview. Convergent validity was excellent at the aggregate and individual level. However, the relation between VAS and SG differed strongly across individuals. On a scale from 0 (dead) to 100 (perfect health), SG scores were lower for patients with WTD preferences (mean difference d = 35; *p* = .002); however, VAS scores did not vary by WTD preferences.
Conclusions: In general, there is good agreement between VAS and SG measures, although patients disagree about how the VAS and SG are related. The standard gamble varied by WTD preferences, however, the VAS did not.

Keywords: Utility assessment, Esophageal cancer

In clinical decision analysis, it is necessary to assess utilities for a relevant set of health states. In many studies, a set of health states is judged by patients or physicians, who may be more or less familiar with these states. The aim of our study was to study convergent validity between three methods to obtain utilities for health states in patients with esophageal cancer. We studied the rank order method, the visual analogue scale (VAS), and the standard gamble (SG). In the remainder of this study, both SG and VAS scores will be called utilities. These utilities were collected as part of a cost-utility analysis in a randomized trial comparing two surgical procedures for patients with esophageal cancer (2).

Convergent validity is usually tested by comparing various methods to assess utilities. We used the rank order, VAS, and SG methods. An interesting case occurs when health states are ranked as worse than dead (WTD). Patients who rank a state as WTD were expected to give lower utilities to that state when compared with patients who consider the state to be better than death. We tested whether utilities are affected by whether or not a respondent gives WTD preferences to one or more health states.

METHODS

Setting

The sample consisted of Dutch patients with esophageal cancer participating in a multicenter randomized clinical trial

113

(2). The trial compared transhiatal and transthoracic resection procedures for the treatment of esophageal cancer. The transhiatal procedure involves a resection through the abdomen and the neck; the transthoracic procedure involves extended lymph node resection in the abdomen and the chest. The latter procedure is more taxing but carries the possibility of better long-term survival.

Procedure

Utilities were collected in a single face-to-face interview. Because it was not possible to interview patients before the resection, utilities were collected in a single interview planned 6 months after the resection. The interview took place in the outpatient department of the university hospitals in Amsterdam and Rotterdam, The Netherlands, when the patient was scheduled for a follow-up visit. The interview was canceled for patients with an unresectable primary tumor or recurrent disease because of possible emotional burden. All interviews were conducted by one of two trained interviewers between February 1997 and July 1999.

Health States

Patients read descriptions of eight health states relevant in decision making concerning resection of esophageal cancer. The descriptions were presented in a random order and are described in Box 1 (1). The health domains described included mobility, stage of disease, pain, hoarseness, tiredness, pneumonia, swallowing and meals, psychological problems, and social support. The health state descriptions were developed by the project team, which consisted of oncology surgeons (n = 3) and psychologists (n = 4) with oncological experience. The eight health states involved (i) own health, (ii) in-hospital after esophagectomy, (iii) in-hospital after esophagectomy complicated by pneumonia, (iv) recovery at home, (v) recurrence-free survival, (vi) local recurrence in neoesophagus, (vii) skeletal metastases, and (viii) unresectable primary tumor. Because the prognosis for these patients is poor with an estimated 5-year survival of only 25 percent, future metastases are a realistic threat to these patients. The unresectable state described a patient who had undergone a surgical procedure that was truncated due to unresectable disease. Durations of the health states were not specified; if the patients had asked about the

---

Box 1. Health State Descriptions: Unlabeled Descriptions Were Shown to Patients

STATE A: home, disease-free
This patient has been operated on for a tumor of the esophagus. He (she) has returned home and is basically able to do everything he (she) could do before the operation. He (she) is still feeling a little tired. He (she) has to take many small meals. Sometimes he (she) is worried that the disease might return. His (her) social contacts are as before the operation.

STATE B: home, recovering
This patient has been operated on for a tumor of the esophagus. He (she) has returned home. Eating is still difficult, and he (she) has difficulty to maintain his (her) weight. He (she) cannot do much yet and is feeling tired. He (she) has got a slight pain. The recovery is not as easy as he (she) thought it would be.

STATE C: hospital, no complications
This patient has been operated on for a tumor of the esophagus some days ago. There were no complications. He (she) has been on the intensive care unit for 2 days. He (she) is on the ward now, but he (she) cannot walk and is in bed all day. He (she) does not have a fever and can breath well but is hoarse. He (she) is dependent on others for almost everything.

STATE D: hospital, pneumonia
This patient has been operated on for a tumor of the esophagus some days ago. After the operation he (she) developed pneumonia. As a result, he (she) has been on the intensive care unit for a week with a fever. He (she) is on the ward now, but he (she) cannot walk and is in bed all day. He (she) is out of breath and hoarse. The operation was more severe than he (she) had expected. He (she) is dependent on others for almost everything.

STATE E: recurrence in neoesophagus
This patient has been operated on for a tumor of the esophagus. The disease has returned in the gastric tube. Eating is difficult; therefore, a stent will be placed in the gastric tube. He (she) is not hungry and is tired. It was a heavy blow for him (her) that the disease has returned, and he (she) is depressed about the future. With regard to social contacts, he (she) is supported by his (her) loved ones.

STATE F: recurrence in bones
This patient has been operated on for a tumor of the esophagus. The disease has returned in his (her) bones. He (she) is in pain, and he (she) will receive radiation therapy. He (she) is not hungry and is tired. It was a heavy blow for him (her) that the disease has returned, and he (she) is depressed about the future. With regard to social contacts, he (she) is supported by his (her) loved ones.

STATE G: unresectable
This patient has been operated on for a tumor of the esophagus. However, during the operation, it became obvious that the tumor was too large and that cure was not possible. This information was a heavy blow for him (her). He (she) is at home now. Eating is difficult; therefore, a stent has been placed in the diseased esophagus. He (she) is not hungry and is tired. He (she) is depressed about the future. With regard to social contacts, he (she) is supported by his (her) loved ones.

---

duration (which never occurred), the duration was defined to be 1 year.

Ranks

The eight states were ranked by the patients from most (rank = 1) to least preferred. Next, patients placed "death within a week" and "perfect health" within this rank order. Thus, the least preferred health state had a rank equal to 10.

Visual Analogue Scale

Subsequently, patients were asked to rate the eight health states by putting a cross on a horizontal line with the end points "worst imaginable health state" and "perfect health." They were asked to rate their own health first. For the purpose of making clear that all remaining health states were hypothetical, the unresectable state was the second state presented. Our patients were well aware that their cancers had not been unresectable; therefore, the unresectable state served this purpose well. The remaining health states were presented in random order.

Standard Gamble

Next, the gambling concept was introduced by means of a practice gamble with financial outcomes. The probability equivalent gamble was used to elicit utilities for the seven hypothetical health states and own health. Patients were confronted with the following choice: "Suppose you have to choose between two options: a gamble with a probability p of perfect health and a probability $(1 - p)$ of dying within 1 week (option A) or living with health state Q_i (option B): which option would you choose?" Q_i was one of the eight health states described above. Cards with the health state descriptions were placed on a probability wheel that was used to visualize the probabilities and options. The presentation order was as described above under the VAS.

The probability at which the patient was indifferent between options A and B was obtained by means of a bracketing procedure that involved forced choices. The first two choices were with p set equal to 0 or 100 percent, in random order. Next, we varied p until the patient expressed indifference between the gamble and the certain option B. In principle, the starting number p was chosen randomly to minimize anchoring effects; however, for the better health states, the starting number p was chosen in the upper end of the range (80–100) to avoid downward biasing. From that starting number, the indifference point was approached by means of a bisection procedure. Once the indifference point was pinpointed within a final range of 5 to 10 percent, the patients were asked to state the indifference point. Negative SG utilities were not assessed, even for the states ranked WTD.

ANALYSES

We assessed convergent validity both at the group level and the individual level. At the group level, convergent validity between the rank order, VAS, and SG utilities was assessed by calculating Pearson's product moment and Spearman's rank correlation coefficients between the mean (averaged across patients) utilities across the eight health states. At the individual level, correlation coefficients between the ranks, VAS, and SG data were calculated for each patient separately and median correlations are reported. A transformation between the averaged VAS and SG scores was calculated by estimating α in the power function: VAS $= 1 - (1 - SG)^\alpha$ by means of nonlinear regression (8). To assess how VAS and SG scores vary across individuals, we calculated Pearson's product moment correlation coefficients of the VAS and SG scores across patients for each of the eight health states separately.

WTD preferences were obtained from the ranking procedure involving dead and perfect health. WTD preferences indicate whether or not patients preferred death to "recurrence in neoesophagus" or "skeletal metastases" in the ranking procedure. We tested whether age varied between those with and without WTD preferences. We compared utilities of patients with and without WTD preferences. Differences were tested with a multivariate analysis of variance and confirmed with nonparametric Mann–Whitney U-tests, because the distributions of the gamble scores were distinctly nonnormal. We normalized SG utilities to a 0–100 scale for ease of comparison with the VAS scores.

RESULTS

Between 1993 and 2002, a total of 221 patients were randomized in the clinical trial. In the interview period between February 1997 and July 1999, ninety-three patients consented to participate in the utility interview. Thirteen (14 percent) patients were excluded because they died before the scheduled interview. An additional twenty-one (23 percent) patients were still alive but were excluded from the interview because of recurrent disease. Of the remaining fifty-nine patients, six (10 percent) ultimately refused participation because of the emotional burden, and three (5 percent) could not be reached. Therefore, fifty of fifty-nine eligible patients were interviewed (85 percent). SG data of two patients were excluded because of cognitive problems or inconsistent results with the SG. Complete VAS and SG records were obtained in forty-five patients.

Of the interviewed patients, most were male (90 percent), married (86 percent), had children (91 percent), and had finished high school (45 percent) or college (20 percent). Their mean age was 63 years (range, 44–79 years), and the average time of the interview was 7 months (range, 3–12 months) after surgery. The interview lasted a mean (SD) of 58 (16) minutes.

Table 1. Mean Ranks, VAS and SG scores, and Their Standard Deviations for Eight Health States

	Rank (mean)	VAS		SG		R(VAS, SG)
		Mean	SD	Mean	SD	
1 own health	2.3 2–4	77	14	97	6	0.56[a]
2 recurrence-free	2.98 2–6	77	11	96	7	0.19
3 recovery at home	4.47 2–7	55	19	92	15	0.14
4 in hospital after esophagectomy	4.60 3–8	54	17	90	15	0.19
5 in hospital after esophagectomy with pneumonia	5.84 4–7	39	16	82	25	−0.25
6 recurrence in neoesophagus	7.60 5–10	18	13	41	31	0.17
7 skeletal metastases	8.22 6–10	16	13	35	30	0.23
8 unresectable primary tumor	8.33 5–10	11	10	34	31	0.27

The range of ranks is given below the mean rank.
[a] $p < .01$, however, this correlation dropped to 0.16 after deletion of one patient with exceptionally low values.

Table 1 presents data on all methods averaged across the forty-five patients with complete records. The mean ranks of perfect health and death were 1 and 9.7, respectively. The top two ranked states have high VAS scores, three states had intermediate VAS scores, and three states had low VAS scores.

The mean rankings of the health states is given in the second column of Table 1. At the group level, the convergent validity between the mean rankings and utilities was excellent: (Pearson's r(rank, VAS) = 1.0, r(rank, SG) = 0.95, r(VAS, SG) = 0.94; Spearman's rank correlations were greater than 0.98. Also at the individual level, convergent validity between the eight utilities obtained with the various methods was good; the median Pearson correlation coefficients were r(rank, VAS) = 0.93, r(rank, SG) = 0.85, and r(VAS, SG) = 0.83. We estimated the power coefficient α in the transformation $VAS = 1 - (1 - SG)^{\alpha}$ at 0.36 (95 percent confidence interval, 0.31–0.40).

The last column in Table 1 presents for each individual health state the correlation between the VAS and SG across patients. The correlation is moderate for the health state "own health" (r = 0.56) and poor for the remaining health states with correlations ranging from −0.25 to 0.27.

There were six patients with WTD preferences. In these six patients, the health states "recurrence in neoesophagus," "skeletal metastases," and "unresectable" were ranked as WTD by 4, 6, and 6 patients, respectively. Age did not vary (t(46) = 0.88; p = .39) between those with (mean age, 63 years; n = 42) and without WTD preferences (mean age, 59 years; n = 6). We compared utilities of patients with or without WTD preferences for "skeletal metastases" or "recurrence in neoesophagus" (Figure 1). For the three worst states, VAS scores did not differ by WTD preferences (mean

difference d = 4; F(1, 46) = 0.31; p = .58); in contrast, SG scores were lower for those with WTD preferences (d = 35; F(1, 43) = 9.22; p = .004). Also for the 3 intermediate states, VAS scores did not vary by WTD preferences (F(1, 46) = 0.62; p = .43); again, SG scores were lower for those with WTD preferences (F(1, 43) = 8.41; p = .006). These results were confirmed by nonparametric tests.

DISCUSSION

We set out to investigate the convergent validity of utilities for health states related to esophageal cancer. We studied patients with esophageal cancer 3–12 months after surgery with curative intent. At the group level, the convergent validity between the mean rankings, the VAS, and the SG utilities was excellent, confirming earlier results. This finding corroborates using these utilities in cost-utility analyses (1).

Because we obtained rankings and VAS and SG scores for each individual, we were also able to assess how these methods converged at the individual level. Again, convergence turned out to be good with median Pearson correlations greater than 0.83, indicating that health states are similarly ranked by the ranking, VAS, and SG instruments.

A general finding is that the relation between utility assessment methods may vary considerably among individuals. Thus, subjects may give high time trade-off (TTO) scores while simultaneously giving low rating scores and vice versa (9). Among VAS, SG, and TTO, correlations of 0.31 to 0.45 have been reported in patient populations (5;6;10). We found a correlation of 0.56 for own health, but this value dropped to 0.16 when one patient with exceptionally low utilities for "own health" was excluded. Our remaining VAS-SG correlations across subjects are much lower (Table 1, last
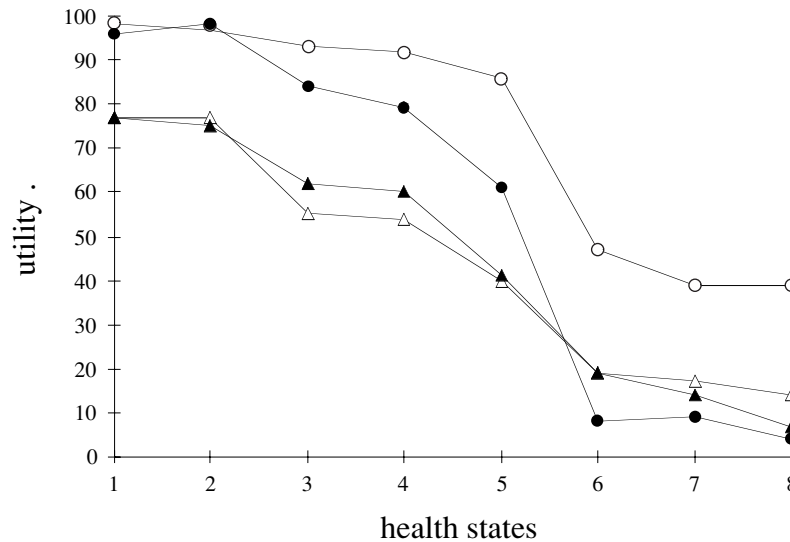
Figure 1. Standard gamble (circles) and visual analogue scale (triangles) utilities grouped by whether skeletal metastases were preferred to death (open symbols) or not (closed symbols). Numbers on the horizontal axis correspond to the health state numbers in Table 1.

column). This finding indicates a lack of convergent validity with respect to the VAS-SG relationship across patients, in other words, the VAS-SG relation differs strongly across subjects.

When average VAS and SG scores are used, the power coefficient α in the VAS-SG transformation VAS $= 1 - (1 - SG)^\alpha$ was estimated at 0.36 (95 percent confidence interval 0.31–0.40), in excellent agreement with the value of 0.37 reported by Krabbe (3) between the EuroQol-VAS and SG. This transformation can be used at the group level but certainly not at the individual level given the variability of the VAS-SG relationship noted above. Nevertheless, even at the group level, such transformations should be cautiously applied. For instance, as a rule, utility scores are relatively lowest for the VAS and highest for the SG, with TTO in between. However, in some patient samples, TTO scores are, on average, higher than SG scores (7;10); therefore, there is evidence that transformations may vary across patient samples (11).

With regard to the WTD preferences, six of forty-eight patients preferred death to "skeletal metastases" or "recurrence in neoesophagus." These six patients with WTD preferences had lower SG utilities. Had we allowed for negative SG utilities, the SG utilities would have been even lower. For the VAS method, however, these six patients with WTD preferences did not have lower VAS scores. The insensitivity of the VAS method to WTD preferences is a drawback, as it suggests that the VAS elicits "quick and dirty" evaluations. In contrast, our data show that SG preferences do vary by WTD preferences, as required by the transitivity axioms of the quality-adjusted life years model. By this criterion, the SG is preferred to the VAS for normative decision making.

One could argue that the VAS need not necessarily vary by WTD preferences. Because the lower end point of the VAS was "worst imaginable health," instead of "dead," presumably those who consider a health state WTD can use a wider range of the scale than those who do not. For instance, on a 0–1 death-perfect health scale, the lowest VAS of those with WTD preferences could be -0.5, whereas it is truly 0 for those without WTD preferences. If this is so, it is not really necessary that those with a WTD preference should have lower scores than those who do not, because the lower VAS end point "worst imaginable health" may represent different utilities for the two groups. Therefore, the present finding needs to be replicated separately for a VAS scale with the lower end point "dead."

Another objection may be that the VAS did not vary by WTD preferences because the VAS scores were already lower from the start. However, for the intermediate states, the VAS did not vary either, even though the SG scores varied by WTD preferences for these intermediate states.

LIMITATIONS

The results above have to be interpreted with the following limitations in mind. First, approximately half of the consenting patients were excluded from the interview, leaving a sample of survivors without recurrent cancer. This group of patients should have the best quality of life, which may affect their health state evaluations. Second, it is unclear how values would have been affected had we interviewed patients before surgery. Thus, the present utilities cannot be generalized to all patients undergoing esophagectomy. Third, the results suggesting that the VAS method is relatively insensitive to WTD preferences, although strongly significant, were

based on subgroup analyses comparing only six with forty-two patients. Fourth, we did not assess negative utilities. With respect to the test of the hypothesis that SG utilities are lower for patients with WTD preferences, these tests were conservative, because the absence of negative utilities decreases differences between those with and without WTD preferences. Fifth, we have not tested the relation between utilities and patient or clinical characteristics. These relations have been tested in reference 1.

POLICY IMPLICATIONS

The present data suggest that the VAS method does not discriminate between those with or without WTDad preferences. Preference-based methods such as the SG may be better suited to assess preferences for bad health states.

CONTACT INFORMATION

Peep F. M. Stalmeier, PhD, Senior Researcher (p.stalmeier@mta.umcn.nl), Department of Medical Technology Assessment, University Medical Centre Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands
Angela G. E. M. de Boer, PhD, (a.g.deboer@amc.uva.nl), Assistant Professor, Coronel Institute for Occupational and Environmental Health, Academic Medical Centre, Meibergdreef 9/K0-110, 1105 AZ Amsterdam, the Netherlands
Mirjam A. G. Sprangers, PhD, Professor (m.sprangers@amc.uva.nl), Department of Medical Psychology, Academic Medical Center, University of Amsterdam, Meibergdreef 15/1105 AZ Amsterdam, The Netherlands
Hanneke C. J. M. de Haes, PhD, Professor (j.c.dehaes@amc.uva.nl), Department of Medical Psychology, University of Amsterdam; Head, Department of Medical Psychology, Academic Medical Centre, Meibergdreef 15, 1105 AZ Amsterdam, the Netherlands
Jan J. B. van Lanschot, MD, PhD, Professor of Surgical Oncology (j.j.vanlanschot@amc.uva.nl), Department of Surgery, University of Amsterdam; Chief of Surgical Oncology, Department of Surgery, Academic Medical Centre, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

REFERENCES

1. de Boer AGEM , Stalmeier PFM, Sprangers MAG, et al. Transhiatal vs extended transthoracic resection in oesophageal carcinoma: Patients' utilities and treatment preferences. *Br J Cancer*. 2002;86:851-857.
2. Hulscher JBF, Van Sandick JW, De Boer AGEM, et al. Extended transthoracic resection compared with limited transhiatal resection for adenocarcinoma of the esophagus and gastric cardia: Results of a randomized study. *N Engl J Med*. 2002;347:1662-1669.
3. Krabbe PFM, Essink-Bot ML, Bonsel GJ. The comparability and reliability of five health-state valuation methods. *Soc Sci Med*. 1997;11:1641-1652.
4. O'Leary JF, Fairclough DL, Jankowski MK, et al. Comparison of time-tradeoff utilities and rating-scale values of cancer-patients and their relatives—evidence for a possible plateau relationship. *Med Decis Making*. 1995,15:132-137.
5. Postulart D, Adang EMM. Response shift and adaptation in chronically ill patients. *Med Decis Making*. 2002;20:186-193.
6. Rutten-van Molken MP, Bakker CH, van Doorslaer EK, et al. Methodological issues of patient utility measurement. Experience from two clinical trials. *Med Care*. 1995;33:922-937.
7. Souchek J, Stacks JR, Brody B, et al. A trial for comparing methods for eliciting treatment preferences from men with advanced prostate cancer. *Med Care*. 2000;38:1040-1050.
8. Torrance GW. Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Econ Plan Sci*. 1976;10:129-136.
9. Tsevat J, Goldman L, Lamas GA, et al. Functional status versus utilities in survivors of myocardial infarction. *Med Care*. 1991;29:1153-1159.
10. Tsevat J, Sherman SN, McElwee JA, et al. The will to live among HIV-infected patients. *Ann Intern Med*. 1999;131:194-198.
11. Yi M, Mrus J, Britto M, et al. The relationship between visual analogue scale scores and utilities for current health. *Med Decis Making*. 2001;21:526.