

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/22972>

Please be advised that this information was generated on 2018-07-07 and may be subject to change.

### CRITICIZING CURRENT SURGICAL POLICIES IN TESTICULAR CANCER USING THE JUDGEMENT OF THE CLINICIANS THEMSELVES

EW Steyerberg, PhB Marshall, HJ Keizer, JDF Habbema, Department of Public Health, Erasmus University, Rotterdam and University Hospital, Leiden, The Netherlands.

Metastatic testicular cancer can be cured in the majority of patients by chemotherapy. After chemotherapy, residual masses of the initial metastases may remain, which may effectively be removed by a surgical resection. The selection of patients varies widely for this procedure, with resection rates between 25% and 80%. This variation is largely explained by differences in management of patients with small residual masses (0-20mm). In this analysis we therefore aimed to evaluate the potential benefit or harm caused by resection of small masses.

A decision analysis model was constructed for the strategies 'resection' and 'follow-up'. The outcome considered was the 5-year survival rate. With each strategy, one of three histologies may be present in the residual mass. First, fully benign tissue (necrosis) has the same prognosis, whether resected or not. Second, mature teratoma is potentially malignant and may grow, causing additional risks if not resected. Third, the presence of cancer cells indicates that the patient was not completely cured, and that additional chemotherapy is indicated. Prognostic estimates for follow-up of mature teratoma or cancer are not observable, since resection is the only reliable way to determine the histological diagnosis. We therefore elicited these estimates from 7 expert clinicians. The clinicians came from centers representing the whole spectrum of currently used selection policies. The prognostic estimates were averaged, and a plausible range was constructed for each estimate.

When using average estimates, a 4.4% higher 5-year survival rate was expected with resection of masses 11-20mm in size. For very small masses (0-10mm), a 2.9% higher 5-year survival rate was expected. The benefit of resection was rather robust when the estimates were varied over their plausible ranges. The minimum benefits according to the individual clinicians' estimates were 2.2% and 0.8% respectively.

In conclusion, the clinicians' own judgements indicate a substantial benefit of resection, even in very small residual masses. Current surgical policies for residual masses therefore need to be reconsidered.

### SEQUENCE EFFECTS IN TRADEOFFS: ARE GOOD YEARS AFTER BAD YEARS BETTER THAN THE REVERSE?

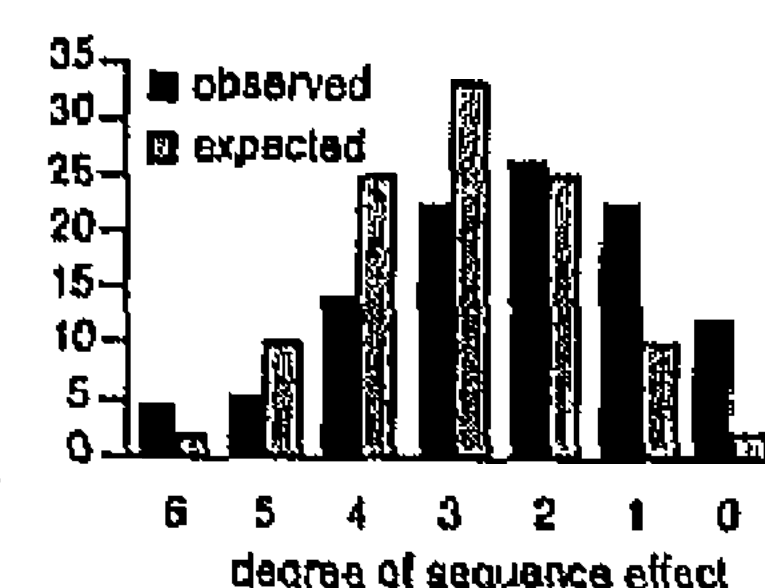
PFM Krabbe, ML Essink-Bor, GJ Bonsel, Department of Public Health, Faculty of Medicine, Erasmus University Rotterdam, and Department of Clinical Epidemiology and Biostatistics, Academic Medical Centre Amsterdam, The Netherlands.

**Purpose.** We studied whether a different sequence of health states in a health profile yields different valuations. **Methods.** The empirical task was part of a large standardized experiment involving 104 students. Thirteen health states were valued by two different modes of a modification of the TTO method.

The standard mode for TTO determined indifference between 10 years in suboptimal state A versus X years in "perfect health" followed by 10-X years in the "worst imaginable health state" (the task for the respondents was to assess X). The experimental, reverse mode for TTO offered the same tradeoff task, but here Y years in the "worst imaginable health state" were followed by 10-Y years in "perfect health". If sequence effects were absent, X should equal 10-Y. "Worst imaginable health state" rather than "dead" was chosen as a reference outcome which enable us to construct health profiles and to avoid negative TTO values for health states worse than death. **Results.** A small but distinct overall effect of sequence was detected for the bad health states. Apparently, respondents prefer starting with the bad years ending up healthy rather than the reverse, despite a clear instruction that health would return to normal after 10 years.

Detailed analysis at the individual level showed that a proportion of the respondents could be classified as sensitive for the sequence of the events in time, either preferring best years first, or the reverse: "happy end" respondents.

**Conclusions.** We found a sequence effect in our study: X was valued less than 10-Y for the bad health states. The sequence effect in our study contradicts conventional time preference (discounting), which in particular affects remote years in full health. These results suggest that a lifetime health profile may not simply be regarded as a chain of separately valued QALY periods. Furthermore, the results are conflicting with the constant proportional tradeoff assumption of the TTO method.



### MEASURING STAKEHOLDER PREFERENCES FOR SCHIZOPHRENIA OUTCOMES

Martha Shumway, University of California, San Francisco, California.

Understanding stakeholder preferences is essential for identification of effective treatments for schizophrenia, a severe and chronic psychiatric disorder with multiple, conflicting outcomes. However, measuring preferences for schizophrenia outcomes poses special challenges. First, several stakeholder groups are involved in schizophrenia treatment, including patients, patient's families, clinicians, and members of the general public. Second, patients--whose preferences are most central--often have psychiatric symptoms which limit their ability to express their preferences.

Three studies examined the suitability of 4 preference assessment methods (Category Rating, Time Trade-off, Paired Comparison and Direct Importance Rating) for evaluating schizophrenia outcomes. In the first study, 21 clinicians evaluated all 4 methods in focus groups. All methods were more difficult to use than anticipated. Time Trade-off was significantly more difficult than other methods. The methods also yielded different rankings of 7 key outcomes. The second study examined the effects of 2 presentation formats and 2 time frames on clinician's Time Trade-off ratings of schizophrenic health states. Format and time frame did affect ratings, but did not affect task comprehension or ease of use. In the third study, 20 persons with schizophrenia evaluated the Category Rating, Paired Comparison and Direct Importance Rating methods in individual interviews. They favored the Direct Importance method which took the least time, however the Category Rating and Paired Comparison methods also appeared comprehensible and acceptable.

These studies suggest that standard preference assessment methods are suitable for measuring stakeholder preferences for schizophrenia outcomes. However, both clinicians and patients found the methods challenging and felt they would have difficulty evaluating a large number of health states. The Time Trade-off method appears less suitable than other methods, possibly due to the chronic, but highly variable, course of schizophrenic illness.

### "STATE" VERSUS "PROCESS" DESCRIPTIVE FORMATS: EFFECTS ON PREFERENCES.

PFM Krabbe, JDF Habbema, PC Levendag, CA Meeuwis, PIM Schmitz, Center for Clinical Decision Sciences, Department of Public Health, Faculty of Medicine, Erasmus University Rotterdam, and Dr. Daniel den Hoed Cancer Center, The Netherlands.

**Purpose.** Utilities incorporated in computations for quality-adjusted life years (QALYs) are conventionally elicited by presenting concise descriptive formats to respondents on the stable end-state after a particular treatment. We compared such a "state" description based approach with a "process" procedure which contains detailed information of the entire treatment regime.

**Methods.** Laryngeal cancer, the object of our study, is treated by radiation therapy (RT) or by surgical removal of the larynx including the vocal cords (S). Two different descriptive formats for both RT and S were used. "State" is a brief description with relevant information about the stable end-state after RT or S. The "Process" description included detailed information about: tumor recurrence rates, 5-year survival rates, grade of mutilation, physical impairment and side-effects.

The time-tradeoff method was used to elicit utilities for the two treatments for both descriptive formats. Two groups of respondents were selected to participate: former laryngeal cancer patients (n=24) and a control group (n=24). One half of each group valued the State or the Process descriptions. Utilities elicited by the State descriptions were incorporated, which is the conventionally practice, into a decision-tree for the computation of QALYs. Process based utilities were used directly.

**Results.** The two descriptive formats gave different results. Former patients and controls exhibited largely similar patterns. Utilities elicited by Process were on average almost similar for RT and S. For State descriptions, average utilities were higher for RT than for S. An interesting finding among the former patients was that in general those who preferred RT resp. S were also treated by RT resp. S in the past.

**Conclusions.** Different utilities were elicited for State and Process. Several reasons may be responsible for this. In particular the type and amount of information may affect individual assessments. The common practice of the formalized QALY methodology to incorporate utilities elicited by State-like descriptions may be a too restrictive procedure for the respondents. On the other hand, performing evaluations based on Process-like descriptions demands without any doubt a considerable cognitive task of participants. Research into the causes of the differences in results between the "State" and the "Process" approach is required.

### ARE METHODS FOR CALCULATING QALYs IN COST-EFFECTIVENESS ANALYSES IMPROVING?

PJ Neumann and DE Zinner, Harvard School of Public Health, Boston, Massachusetts, and Massachusetts Institute of Technology, Cambridge, Massachusetts.

**Objectives.** The objectives of this paper are: 1) to investigate variations among cost-effectiveness analyses in the published literature in their methods used to calculate QALYs; and 2) to examine whether these methods have been improving over time.

**Data and Methods.** Using a MEDLINE database search, we identified 86 original cost-effectiveness analyses, published between 1975 and 1995, which used QALYs as the measure of effectiveness. For each study, we recorded: the health state classification system employed; the measurement technique used as the basis of valuing health states; and whose preferences were used in the calculation. Two readers evaluated each study independently, then met to reach consensus. We also examined whether methods adhered to "good practices" as defined by the U.S. Panel on Cost-Effectiveness in Health and Medicine, and whether methods have been improving over time.

**Results.** 20% of studies used a pre-existing health state classification scale (e.g., Health Utilities Index). Most studies (59%) assigned quality-of-life weights to health states particular to the condition under investigation (e.g., minor toxicity with chemotherapy); 13% of studies employed a single "general disability" state. The source of the preferences were mostly those of the study authors themselves (45%), followed by community (21%) clinician (15%), and patient (13%) preferences, respectively. In 8% of studies, the source was not explicitly stated. A variety of measurement techniques for valuing health states were used, including author's judgment (45%), time tradeoff (17%), category or rating scale (17%), and standard gamble (4%). In 15% of cases, the technique could not be determined. Few studies adhered to good practices as defined by the Panel, and there was little evidence that methods have been improving over time.

**Conclusions.** The results illustrate extensive variation in the manner in which QALYs have been calculated, and suggest the need for more scientific rigor and consistency with respect to the methods used.

### THE MCMASTER HEALTH UTILITY INDEX (HUI) AND THE EUROQOL-5D ASSESSED IN PATIENTS WITH PERIPHERAL ARTERIAL DISEASE IN THE UNITED STATES AND THE NETHERLANDS.

JL Bosch, EEE van Wilck, PL Baum, MC Donaldson, JJAM van den Dungen, MGM Hunink, University of Groningen and University Hospital Groningen, the Netherlands, Harvard School of Public Health and Harvard Medical School, Boston, Massachusetts.

**Purpose:** To assess the relationship between the McMaster HUI (HUI) and the EQ-5D in the USA and the Netherlands, and to compare health profiles of patients with claudication (i.e. mildest stage of peripheral arterial disease) and a major amputation (most severe stage) based on responses to the McMaster HUI (HUI) and EQ-5D.

**Methods:** The McMaster HUI (HUI) and EQ-5D were completed by 112 patients with peripheral arterial disease (Dutch: n=73, American: n=39; claudication: n=38, major amputation: n=14, critical ischemia: n=6, minor amputation n=13, after revascularization: n=41). The relationship between the two indices was assessed with a correlation coefficient (r<sup>2</sup>). The modes of the responses to the various dimensions of the health profiles were compared.

**Results:** Although the mean McMaster HUI (HUI) and EQ-5D values were significantly different (0.70 (SD 0.22) versus 0.54 (SD 0.26) respectively), the values correlated moderately well (r<sup>2</sup> 0.41). Since the McMaster HUI (HUI) formula is based on SG utilities and the EQ-5D index on RS values, we transformed the EQ-5D with Torrance's transformation. This improved the relationship between the two indices only slightly (r<sup>2</sup> 0.44). No differences in outcomes or the correlation coefficient between the two countries were demonstrated. The mean McMaster HUI (HUI) and mean EQ-5D were significantly different in the claudication group, 0.76 (SD 0.15) versus 0.51 (SD 0.21) respectively, and in the amputation group 0.61 (SD 0.19) versus 0.43 (SD 0.24) respectively (p < 0.05). The mean values of the McMaster HUI (HUI) differed significantly for claudication compared with amputation (p < 0.05), whereas this was not the case for the mean EQ-5D values. In addition, the health profiles for claudication and amputation based on the McMaster HUI (HUI) responses demonstrated differences in the dimensions mobility and pain, whereas the EQ-5D values yielded equivalent health profiles for these two health states.

**Conclusion:** Although the McMaster HUI (HUI) was developed in North America and the EQ-5D in Europe this affected neither the outcomes nor the relationship between the indices across countries. The EQ-5D yielded lower values than the McMaster HUI (HUI), which was not explained by transforming the RS-based EQ-5D index to a SG utility. Finally, the McMaster HUI (HUI) was more discriminative between different stages of peripheral arterial disease than the EQ-5D.