

Agreement between prostate cancer patients and their clinicians about utilities and attribute importance

Arthur S. Elstein PhD,* Gretchen B. Chapman PhD,† Joan S. Chmiel PhD,‡ Sara J. Knight PhD,§ Cheeling Chan MS,** Robert B. Nadler MD,†† Timothy M. Kuzel MD,‡‡ Amy K. Siston PhD*** and Charles L. Bennett MD PhD†††

*Professor Emeritus, Department of Medical Education, University of Illinois at Chicago, Chicago, IL, USA, †Associate Professor, Department of Psychology, Rutgers University, Piscataway, NJ, USA, ‡Professor, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA, §Research Health Science Specialist, San Francisco VA Medical Center and Assistant Adjunct Professor, University of California at San Francisco, San Francisco, CA, USA, **Statistical Analysis, Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA, ††Associate Professor of Urologic Surgery, Head Section of Endourology, Laparoscopy and Stone Disease, Northwestern University, Feinberg School of Medicine, Chicago, IL, USA, †††Associate Professor of Medicine and Associate Director of Clinical Affairs, Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Chicago, IL, USA, ****Assistant Professor, Department of Psychiatry and Behavioural Sciences, Northwestern University, Chicago, IL, USA, and ††††Associate Director, Midwest Center for Health Services Research and Policy Studies, VA Chicago Healthcare System; Professor of Medicine, Division of Hematology/Oncology, Feinberg School of Medicine, Northwestern University; Co-director, Cancer Control Program of the Robert H. Lurie Comprehensive Cancer Centre of Northwestern University; Center for Healthcare Studies, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

Abstract

Correspondence

Arthur S. Elstein
338 Brookside Dr.
Wilmette
IL 60091
USA
E-mail: aelstein@uic.edu

Accepted for publication

19 January 2004

Keywords: preferences, substituted judgement, utility assessment, values

Supported in part by grant IIR 95-120 to the last author from the Department of Veterans Affairs, Division of Health Services Research & Development.

Purpose To examine the agreement between prostate cancer patients' utilities for selected health states and their rankings of the importance of six attributes of the health states and the clinicians' judgements of what would be in the patients' best interests.

Method Patients with newly diagnosed localized prostate cancer individually completed a time trade-off utility assessment shortly after being diagnosed. The health states evaluated were constructed from a multi-attribute utility model that incorporated six aspects of living with the disease and outcomes of treatment. Each patient assessed his current health state and three hypothetical states that might occur in the future, and provided rankings of the importance of the six attributes. The clinicians caring for each patient independently provided their views of what utilities and importance rankings would be in the patient's best interest.

Results The across-participant correlations between patients' and clinicians' utilities were very low and not statistically significant. Across-participant correlations between patient and clinician importance rankings for the six attributes were also low. Across-health state and across-attribute correlations between utilities or importance rankings were highly variable across patient-clinician pairs.

Conclusion In the clinical settings studied, there is not a strong relationship between valuations of current and possible future health states by patients with newly diagnosed prostate cancer and their clinicians. Implications of these results for substituted judgement, when clinicians advise their patients or recommend a treatment strategy, are discussed.

Introduction

Involving patients in decisions about their own care is widely advocated,^{1–4} particularly when all available treatments carry significant downside risks and trade-offs are involved. A decision analytic model of shared decision-making asks patients to assess the value of a reasonable range of potential outcomes of alternative treatment strategies. These utilities can then be combined with estimates of the probabilities of various outcomes, contingent on the patient's overall health and the treatment strategy selected, to calculate an expected utility for each treatment alternative.^{5–7} The approach focuses the patient's attention on the possible outcomes of treatment, not on the treatment alternatives themselves.

Involving patients in shared decision-making is especially relevant for localized prostate cancer. Despite a recent randomized clinical trial comparing outcomes of radical prostatectomy and watchful waiting,⁸ the optimal management strategy remains controversial. Because of the trade-offs involved and the downside risks associated with all treatments,⁹ treatment choice is sensitive to risk attitude and the utilities for post-treatment complications.^{10–15} Utilities based on grouped data may not be suitable for individual clinical decision-making and those of individual patients should be elicited to optimize treatment choice.¹⁶

Several investigations have studied differences between patients and clinicians in their assessments of hypothetical health states. For example, Boyd *et al.*¹⁷ reported differences in utilities for colostomy between five groups: surgeons and oncologists, patients whose rectal cancer had been treated by either abdomino-perineal resection with colostomy or radiotherapy without colostomy, and two groups of healthy volun-

teers. There were significant differences between groups, regardless of the method of utility assessment. The mean utilities of the physicians and the patients with colostomy were closer to each other than to the other groups' utilities, but no *post-hoc* comparisons were conducted to determine whether the differences between them were statistically significant or not. Further, this study did not use correlational methods to explore the relationships between physicians' ratings and those of specific patients. An extensive review of this topic is in Stiggelbout.¹⁸

The most common methods of utility assessment, standard gamble and time trade-off, are cognitively complex and usually require a trained person to administer the interview schedule. Patients have to evaluate unfamiliar and possibly anxiety-provoking descriptions,¹⁹ and the methods may be burdensome, especially if the patients are worried about their illness and their prospects.^{20–22} Some patients may prefer to have their physician make the treatment decision on their behalf. Clinicians would then be called upon to substitute their judgement of what is in the patient's best interest.

The question then arises, if we could know what the patient would have preferred, how closely does the surrogate's judgement match the patient's preferences? Studies of this question have typically examined treatment preferences, especially for life-sustaining treatments, in hypothetical situations.^{23–28} The accuracy of substituted judgements of patients' preferences has ranged between 50 and 80%. In some studies, the level of agreement is significantly better than chance, in others not.²⁹

From a decision-theoretic standpoint, however, asking about treatment preferences combines two assessments (probabilities and utilities of possible outcomes) that are analytically

distinct. It may be that the complexity of this task is a major cause of the inaccuracy observed between surrogates and patients. If assessment concentrated on patients' judgements of the attractiveness of outcomes, better agreement might be obtained. Empirical evidence suggests that clinicians are limited in their understanding of patients' preferences, even when attending physicians are queried.³⁰

This study uses correlational methods to examine the relationships between clinicians' and patients' valuations of a set of health states associated with newly diagnosed localized prostate cancer.

We ask two research questions:

1. What are the correlations between utilities of four health states provided by patients and by clinicians on behalf of these patients?
2. What is the correlation between patients' and clinicians' rankings of importance of six attributes in a multi-attribute model?

Materials and methods

Patients

Patients with newly diagnosed localized prostate cancer were recruited in five Veterans Administration Medical Centers. Patients were approached individually in the clinic while waiting for an appointment and asked if they would be willing to participate in a survey of how men with prostate cancer feel about different health states. They were told that the information would be confidential and that they could end the interview at any time. Patient participation rates were not recorded, but in a previous similar cross-sectional study at these sites, 96% of the patients approached at baseline agreed to participate. The assessment was ordinarily done on the first or second clinic visit after being told of the diagnosis, and before the commencement of treatment or expectant management. The research protocol was approved by the IRB at each site and all patients signed an informed consent form before the interview.

Procedure for assessing patients

Based on previous work on assessing the quality of life in prostate cancer patients,^{31–37} we constructed a multi-attribute model of health states associated with prostate cancer. The attributes in this model were: pain, mood, sexual function, bladder and bowel function, fatigue and energy, and appetite. For each attribute, three levels of function were defined (high, moderate, and low), and these levels were used to construct verbal descriptions of three clinically realistic health states. These health states (A, B and C) are presented in Table 1. A fourth personalized health state (D) was constructed by the patient from the components to represent, as closely as possible, his current health.

We measured the patient's preferences for the four health states using the time trade-off (TTO) method.^{6,38–40} The questions were worded in an impersonal format we have used in previous research.²⁰ Each patient was asked to imagine that he has two friends, Smith and Jones. Mr Smith's health fits the description of one of the hypothetical health states and he will live for 10 years. Mr Jones has perfect health but will live somewhat less than 10 years. The patient was asked, 'If you had to be one of these two people, who would you rather be?' The hypothetical health states were evaluated in one of two orders, A–B–C and C–B–A. After these assessments, the patients were asked to pick the statements for each attribute that best described their health over the past month. These selections were used to construct a customized description of the patient's current health.

The patient then ranked the importance of each attribute in the model by imagining that he was in health state C and then deciding which attribute he would choose to change from the worst level to the best level, if only one could be changed. The attribute selected was ranked first in importance and the ranking continued, patients choosing which attribute they would change second, third, etc., until all attributes had been ranked.

Finally, TTO was used to evaluate the patient's customized health state.

Table 1 Health state descriptions

Attribute	State A	State B	State C
Pain	Mr Smith has very little or no pain, and it is easily controlled by medication	Mr Smith has a bearable amount of pain and it is moderately well controlled by medication	Mr Smith has a great deal of pain much of the time, and it is not well controlled by medication, or the side effects of the medication to control pain are very unpleasant
Mood	He hardly ever feels tense, worried, irritable, sad, or depressed (only once or twice a month or even less)	He feels tense, worried, irritable, sad, or depressed sometimes (only once or twice a week)	He feels tense, worried, irritable, sad or depressed much of the time (almost every day)
Sexual function	His ability to have sex and enjoy it has been affected very little by his condition	His ability to have sex and enjoy it has been affected a fair amount by his condition	His ability to have sex and enjoy it has been affected very much by his condition
Bladder and bowel function	He rarely has difficulties or problems with urinating or bowel function (only once or twice a month or even less)	He has occasional difficulties or problems with urinating or bowel function (only once or twice a week)	He has frequent difficulties or problems with urinating or bowel function (almost every day)
Fatigue/energy	He is able to do most of his usual activities nearly all of the time. He is not overly tired and his energy is pretty good	He has some difficulty doing his usual activities. He does less than before, and he is tired quite a bit of the time. He needs some assistance with some daily activities (e.g. dressing, washing, using the toilet)	He has a lot of trouble doing most usual activities, both at work and at home. He needs a lot of assistance with many daily activities (e.g. dressing, washing, using the toilet). He is very tired much of the time and he spends a lot of time resting
Appetite	Usually good	Sometimes poor	Usually poor

Procedure for assessing clinicians

We assumed that identifying the clinician who was working most closely with each patient at the time of the interview would provide the best measure of patient–clinician concordance on utilities and attributes. Accordingly, to the extent possible, we obtained these rankings from clinicians so identified, either by the patient or by the medical record. It was not realistic for clinicians in these settings to complete a lengthy interview during clinic hours. Therefore, a condensed assessment form was developed that each clinician could complete individually. On the day the patient was interviewed, it was placed in the mailbox of the clinician currently following the patient. The recipient was strongly encouraged to return the form within 1 week. The form provided the patient's name and asked

the clinician to respond using their judgement about what would be in the patient's best interest. This procedure enabled us to collect the clinician's rating in close temporal proximity to the patient rating, and allowed the rater to complete the form in a time and place conducive to making thoughtful judgements.

The form explained that we were conducting a survey of how patients with prostate cancer and their caregivers evaluate a number of health states the patients are experiencing or might experience in the future. The form provided the patient's name and asked the clinician to answer the questions using their best judgement about what would be in the patient's best interest. They were not asked to estimate or guess the patient's responses to the equivalent questions. We believe that the form of the question better matches the clinician's task when asked to make

a decision on a patient's behalf, i.e. to represent what would be in that patient's best interest.

TTO assessment

Each health state and the definition of perfect health were presented in parallel columns. The clinician was asked to mark on a 10-cm line (representing 10 years, the same time frame as used for patients) the number of years of perfect health that in their judgement would be equivalent to 10 years in each health state for that patient. The patient's description of his current health (state D) was entered into the form by the data collector. The states were always assessed in the order A–B–C–D. The utility of each state was determined by measuring the point of the X on the line and rounding to the nearest half-centimetre.

Importance rankings

The clinician was asked to imagine the patient is in health state C, and that he can select a medicine that will change this health state to perfect health on only one aspect. Which aspect would you advise him to change first? next? etc. The clinician was asked to rank the aspects 1–6, in the order in which they should be changed. A list of the attributes was then provided.

Results

Sample

A total of 127 patients and their clinicians were assessed. For complete data, each patient–clinician pair provided 20 data points at the assessment: four TTO utilities from the patient and four from the clinician, and six attribute ranks from the patient and from the clinician. Patient–clinician pairs were eliminated from the analysis if they were missing three or more of the 20 data points. As a result, seven patient–clinician pairs were eliminated from the sample, leaving $n = 120$. Patient–clinician pairs missing one or two of the 20 data points were retained in the analysis and the missing values were replaced with item means. This substitution allowed all the correlations to be based on the same sample size.

Table 2a Mean (SD) TTO judgements given by patients and clinicians for four health states

Health state	Patients	Clinicians
A	0.76 (0.23)	0.91 (0.11)
B	0.51 (0.30)	0.75 (0.17)
C	0.31 (0.30)	0.47 (0.26)
D	0.73 (0.25)	0.82 (0.20)

States A, B and C are hypothetical and are defined in Table 1. State D is the patient's personalized (current) health.

Table 2b ANOVA results

Effect	d.f.	F	MSE
Role	1,119	43*	0.14
Health state	3,357	306*	0.03
Role × health state	3,337	7.98*	0.03

* $P < 0.0001$. Role = patient or clinician. Health states are described in Table 1.

Utility judgements

Table 2a shows the means of the patients' and clinicians' utilities for the three hypothetical health states and current health (D). On average, both groups order the hypothetical states correctly, in the sense that the mean utility of $A > B > C$.

Sixty-seven of 120 patients (=56%) and 94 of 120 clinicians (=78%) correctly ordered the states as $A > B > C$. There were frequent ties (e.g. $A = B$ or $B = C$). Only one clinician and six patients ordered states incorrectly ($A < B$ or $B < C$). The clinicians, on average, rated each state higher than the patients.

Results of a 2 (role: patient vs. clinician) × 4 (health state) repeated measures ANOVA on the judgements are given in Table 2b. The main effects for role and health state indicate that clinicians gave higher values than did patients, and the health states were ranked appropriately, $A > B > C$, with state D receiving a utility rating close to that of A or B. Finally, the role-by-health state interaction means that the discrepancy between clinician and patient ratings was larger for states B and C than it was for states A and D.

The correlations between clinician and patient utilities present a very different picture. Table 3

Table 3 Pearson correlations between patients' and clinicians' utilities for four health states ($n = 120$)

Health state	
A	-0.04
B	-0.00
C	-0.06
D	-0.09

shows the Pearson r correlation coefficients across patient–clinician pairs for four health states. That is, four correlations were computed, one for each health state, each based on 120 paired observations. All are slightly negative and not significantly different from zero.

Table 4 shows the mean Pearson correlations computed across health states for each patient–clinician pair. That is, 120 correlations were computed, one for each patient–clinician pair, each based on four paired observations (the four health states). These correlations measure how well the pairs agree on judging the four health states. They capture patient–clinician consistency in within-subject, across health-state variation. A Fisher Z transformation was applied to the correlation of each patient–clinician pair to normalize the distribution of correlation coefficients. A t -test was performed on these transformed coefficients to test whether the average correlation was greater than 0. The mean transformed Z -score was transformed back into a correlation coefficient for Table 4. The mean and median correlations are notably larger than those in Table 3, reflecting high agreement between patient–clinician pairs on ordering the four health states. The correlations reported are Pearson correlations (not rank-order correlations); thus in order to obtain a perfect correla-

Table 4 Pearson correlations across health state utilities for 120 patient–clinician pairs

Mean ^a	Median	Range	Inter-quartile range	t -test ^b
0.74	0.72	-0.85–1.00	0.58–0.91	13.16*

^aMean values were calculated after Fisher Z transformation of the correlation coefficients. The means presented here have been transformed back to correlation coefficients.

^bTesting whether the mean correlation coefficient differs from zero.
* $P < 0.0001$.

tion of 1.0, a patient and his clinician would have to agree on both the ordering of the four health states and on the relative distance between them. Although the mean correlations were high, correlations varied greatly across patient–clinician pairs and 5% (6/120) were negative.

Attribute ranks

The importance rankings of the attributes in the multi-attribute model offer another way to explore patient–clinician agreement. To assess agreement on the importance rankings, we computed correlations across pairs of raters and within each pair of raters for all six attributes. We first correlated the ranks assigned to each attribute by each patient–clinician pair. Six correlations were computed (one for each health attribute), each based on 120 observations. Table 5 shows these results. Four of the six correlations reported are statistically significant, but none is large enough to permit accurately substituting the clinicians' attribute rankings for the patients.

We next computed a Spearman rank-order correlation across health attributes for each patient–clinician pair. A total of 120 correlations were calculated, one for each patient–clinician pair, each based on six observations (the six health attributes). These correlations were subjected to a Fisher Z transformation, and a t -test then performed on these transformed coefficients to test whether the average correlation was greater than 0. The mean and median Z scores were transformed back into correlation coefficients for Table 6. The mean and median

Table 5 Correlations across patients' and clinicians' ranks for six health attributes ($n = 120$)

Health attributes	r
Pain	0.20*
Mood	0.10
Sex	0.33*
Bowel	0.33*
Energy	0.10
Appetite	0.21*

* $P < 0.05$.

Table 6 Rank-order correlations across health state attributes for 120 patient–clinician pairs

Mean ^a	Median	Range	Inter-quartile range	<i>t</i> -test ^b
0.58	0.52	–0.94–1.00	0.40–0.70	7.67*

^aMean values were taken after Fisher Z transformation of the correlation coefficients. The mean values presented here have been transformed back to correlation coefficients.

^bTesting whether the mean correlation coefficient differs from zero.
* $P < 0.0002$.

correlations are greater than the correlations in Table 5, showing that patient–clinician pairs generally agreed on the ordering of the importance of the six health attributes. However, the correlation varied greatly across pairs of raters and 14% (17/120) were negative. The correlations reported in Table 6 are Spearman rank-order correlations (not Pearson correlations); thus in order to obtain a perfect correlation of 1.0, a patient and his clinician would only have to agree on the ordering of the six health attributes.

Tables 4 and 6 suggest that patient–clinician agreement about utility of health states is greater than the parallel agreement about importance of attributes. A *t*-test comparing the mean correlation in Table 4 with the analogous mean correlation in Table 6 was significant ($t(119) = 4.92$, $P < 0.0001$).

Discussion

The mean utility judgements of the clinicians are higher than the patients' means for all four health states. Patients' judgements of current health (experienced utility) and their anticipated utility of future possible states are lower than clinicians' judgements. Previous studies typically find that the general public rates most disease states lower than patients with the disease. For example, people who do not have chronic renal disease and are not on dialysis rate that state much lower than do patients with renal disease.⁴¹ The usual explanation for this finding is that the general public underestimates the capacity to adapt to chronic illness, so that state seems much worse to them than it does to patients. However, it has also been shown that physicians often rate their patients' quality of life as better than the

patients themselves do.⁴² This discrepancy can be plausibly explained by hypothesizing that the two groups of non-patients employ different reference points: from the viewpoint of the general public, the patient's current health state is worse than their own, while clinicians may judge the patient's health with reference to how bad it may ultimately be. Compared with that reference point, any of the health states used in our study looks better to clinicians than to patients.

In terms of across-pair correlations, agreement about utility judgements between clinicians and their patients was quite low (Table 3). In the settings of this study, clinicians were generally poor predictors of the preferences of individual patients. Thus, if a prostate cancer patient in the VA setting were to decline to provide utility assessments for shared decision-making and were to ask his clinician to provide estimates on his behalf, it is unlikely that the clinician's numbers would even roughly approximate what the patient would have said. While the data on mean utilities in Table 2 suggest that it would be possible to develop a correction factor for the clinicians' ratings that would predict patients' utilities fairly accurately, the correlational data in Table 3 show that this is not possible.

Further, the results for hypothetical states and the current health state are quite comparable, suggesting that the results of previous studies, in which treatment preferences were judged for hypothetical conditions, should not be discounted on the grounds that only judgements about hypothetical preferences were involved. Agreement between clinician–patient pairs about the ranking of the attributes in the model was somewhat better (Table 5).

Much higher patient–clinician agreement was found when correlations are computed across health states or attributes than when they are computed within health states or attributes across patient–clinician pairs. This indicates that clinicians are much better at predicting whether a patient will value one health state more than another than they are at predicting whether a patient will value a particular health state more or less than another patient will. As the health states were constructed so that state A dominates

state B which in turn dominates state C, ranking health states is much simpler than ranking the importance of their underlying attributes. Clinicians could reasonably assume that their patients' utility judgements will rank order $A > B > C$, and then all that remains is to predict how much better A is than B (relative to B vs. C), and to evaluate the patient's current health. Similarly, clinicians are better at predicting whether a patient will view one attribute as more important than another than they are at predicting whether a patient views an attribute as more important than another patient does.

This study is fundamentally concerned with what might occur when patients decline to examine and discuss their preferences for treatment outcomes and leave these assessments in the hands of the clinicians caring for them. The problem is identified in the medical ethics literature as substituted judgement. When clinicians advise patients in this fashion, how closely does the advice match with what the patient would have said, had he been involved in the decision? What are the results when a clinician, for whatever reasons, advises a patient with prostate cancer about how the possible outcomes of treatment might affect treatment choice?

In this study, there is no consistent relationship between clinical judgements of the patient's valuations and what the patients report their values and utilities to be. The low correlations may be partly due to unreliability of measurement on both sides of the pair.

It can be argued that the prostate cancer patients in these settings will be better off if a clinician's judgements are substituted for the patient's own evaluations. This position argues that all things considered, the clinician may well have a better idea of what is in the patient's best interest than does the patient himself. However, many clinicians and patients would find this position unreasonably paternalistic, given that the patients are competent adults. Certainly the movement in the past two decades towards patient empowerment and shared decision-making has sought to find ways to involve patients more in decision-making and not to rely solely on professional authority.

Empirical studies of the accuracy of substituted judgements have shown that surrogate accuracy is usually in the range of 50–80% concordance, too imperfect to insure that substituted judgements could be used with confidence in making decisions for individual patients. Some investigators have found that surrogates' predictions more closely resembled their own treatment preferences than the preferences of the individuals whose preferences they were trying to predict.^{25,43,44} In the face of low agreement about treatment preferences, they recommended focusing the assessment on goals of treatment and preferred quality of life, not on the available options. The current study shows that altering the task to focus on outcomes instead of on-treatment preferences leads to similar findings, with very low to moderate levels of agreement between clinicians and patients. Our study is consistent with the literature in pointing out that accurate substituted judgement remains an ideal of clinical care rather than an everyday reality.

Limitations

This study has four important limitations.

First, the procedures for utility elicitation and importance rankings were not identical for both groups. Some variation of scores due to differences in method was probably introduced. The clinicians' importance rankings may to some extent reflect what they know to be clinically modifiable. Nevertheless, we do not believe that these variations are the main sources of the low correlations obtained in this study, as our results are quite consistent with other studies.^{23,28,29}

Secondly, at each site there were fewer clinicians than patients. Inevitably each clinician rated more than one patient, and so the clinician–patient pairs are not completely independent. Indeed, one cannot imagine a clinical setting in which this feature could be provided. For reasons of confidentiality, the identities of the clinicians and patients who provided the ratings were stripped from the data files used in our analyses, and we cannot accurately adjust for the fact that each clinician in the study (a) contributed to more than one pair of observations

and (b) very likely contributed a different number than any other clinician rater. The literature on proxy judgements of patient quality of life suggests that proxy experience with the patient and with the health states being judged contributes to higher agreement between proxy and patient judgements.^{45–49} Thus, the low rank-order correlations between clinician–patient pairs may be due in part to the fact that newly diagnosed prostate cancer patients are relatively not well known by the caregivers.

Thirdly, in the settings of this study, direct patient care is provided mainly by residents and nurses, while attending physicians have a supervisory role. The correlations might well be higher in a practice setting where a panel of patients is followed by a single attending clinician or by a small team that works together closely.

Fourthly, our results do not necessarily imply that substituting a clinician's judgements for the patient's would lead to a change in treatment strategy for these patients. Although previous research has shown that treatment choice for localized prostate cancer should be utility-driven and take a patient's preferences into account, we cannot say that the differences found between clinician's and patient's utility assessments would have led to different treatment choices, either in a decision analysis or in a less formal decision-making model. That determination awaits further research.

Given these limitations, we believe the results should be interpreted cautiously. Nevertheless, they do suggest that clinicians who are called upon to exercise substituted judgement should not assume that their judgements of a patient's best interests would correspond closely with a patient's stated preference judgements, even if they are well informed about the patient's clinical condition and co-morbidities. At least in the circumstances of this study, there is imperfect agreement between what clinicians judge is in the patient's best interest and what the patients report they value.

Acknowledgements

We thank the anonymous patients and clinicians who participated in this study, the directors of

the clinics, and the data collectors at the various clinical sites. Three reviewers provided helpful comments on a previous draft of this paper.

References

- O'Connor AM, Rostom A, Fiset V *et al.* Decision aids for patients facing health treatment or screening decisions: systematic review. *British Medical Journal*, 1999; **312**: 731–734.
- Kassirer JP. Incorporating patients' preferences into medical decisions. *New England Journal of Medicine*, 1994; **330**: 1895–1896.
- Chan EC, Sulmasy DP. What should men know about prostate-specific antigen screening before giving informed consent? *American Journal of Medicine*, 1998; **105**: 354–355.
- Davis JW, Kuban DA, Lynch DF, Schellhammer PF. Quality of life after treatment for localized prostate cancer: differences based on treatment modality. *Journal of Urology*, 2001; **166**: 947–952.
- Weinstein MC, Fineberg HV, Elstein AS *et al.* *Clinical Decision Analysis*. Philadelphia: Saunders, 1980.
- Hunink M, Glasziou P, Siegel J *et al.* *Decision Making in Health and Medicine: Integrating Evidence and Values*. New York: Cambridge University Press, 2001.
- Rouse DJ, Owen J. Decision analysis. *Clinical Obstetrics and Gynecology*, 1998; **41**: 282–295.
- Holmberg L, Bill-Axelsson A, Helgesen F *et al.* A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer. *New England Journal of Medicine*, 2002; **347**: 781–789.
- Steineck G, Helgesen F, Adolfsson J *et al.* Quality of life after radical prostatectomy or watchful waiting. *New England Journal of Medicine*, 2002; **347**: 790–796.
- Chodak GW. Comparing treatments for localized prostate cancer: persisting uncertainty. *JAMA*, 1998; **280**: 1008–1010.
- Fleming C, Wasson JH, Albertsen PC *et al.* A decision analysis of alternative treatment strategies for clinically localized prostate cancer. *JAMA*, 1993; **269**: 2650–2658.
- Beck JR, Kattan MW, Miles BJ. A critique of the decision analysis for clinically localized prostate cancer. *Journal of Urology*, 1994; **152**: 1894–1899.
- Kattan MW, Cowen ME, Miles BJ. A decision analysis for treatment of clinically localized prostate cancer. *Journal of General Internal Medicine*, 1997; **12**: 299–305.
- Cantor SB, Spann SJ, Volk RJ, Cardenas MP, Warren MM. Prostate cancer screening: a decision analysis. *Journal of Family Practice*, 1995; **41**: 33–41.
- Krahn MD, Mahoney JE, Eckman MH, Trachtenberg J, Pauker SG, Detsky AS. Screening for prostate

- cancer: a decision analytic view. *JAMA*, 1994; **272**: 773–780.
- 16 Cowen ME, Miles BJ, Cahill DF, Giesler RB, Beck JR, Kattan MW. The danger of applying group-level utilities in decision analyses of the treatment of localized prostate cancer in individual patients. *Medical Decision Making*, 1998; **18**: 376–380.
 - 17 Boyd NF, Sutherland HJ, Heasman KZ, Tritchler DL, Cummings BJ. Whose utilities for decision analysis? *Medical Decision Making*, 1990; **10**: 58–67.
 - 18 Stiggelbout AM. Assessing patients' preferences. In: Chapman GB, Sonnenberg FA (eds) *Decision Making in Health Care: Theory, Psychology, and Applications*. New York: Cambridge University Press, 2000: 289–312.
 - 19 Goldstein MK, Tsevat J. Applying utility assessment at the bedside. In: Chapman GB, Sonnenberg FA (eds) *Decision Making in Health Care: Theory, Psychology, and Applications*. New York: Cambridge University Press, 2000: 313–333.
 - 20 Chapman GB, Elstein AS, Sharifi R, Kuzel T, Andrews A, Bennett CL. Prostate cancer patients' utility for health states: how it looks depends on where you stand. *Medical Decision Making*, 1998; **18**: 278–286.
 - 21 Clarke AE, Goldstein MK, Michelson D, Garber AM, Lenert LA. The effect of assessment method and respondent population on utilities elicited for Gaucher's disease. *Quality of Life Research*, 1997; **6**: 169–184.
 - 22 Jimison HB, Sher PP. Advances in presenting health information to patients. In: Chapman GB, Sonnenberg FA (eds) *Decision Making in Health Care: Theory, Psychology, and Applications*. New York: Cambridge University Press, 2000: 334–361.
 - 23 Coppola KM, Ditto PH, Danks JH, Smucker WD. Accuracy of primary care and hospital-based physicians' predictions of elderly outpatients' treatment preferences with and without advance directives. *Archives of Internal Medicine*, 2001; **161**: 431–440.
 - 24 Ditto PH, Danks JH, Smucker WD *et al.* Advance directives as acts of communication: a randomized controlled trial. *Archives of Internal Medicine*, 2001; **161**: 421–430.
 - 25 Fagerlin A, Ditto PH, Danks JH, Houts RM, Smucker WD. Projection in surrogate decisions about life-sustaining medical treatments. *Health Psychology*, 2001; **20**: 166–175.
 - 26 Sulmasy DP, Haller K, Terry PB. More talk, less paper: predicting the accuracy of substituted judgments. *American Journal of Medicine*, 1994; **96**: 432–438.
 - 27 Sulmasy DP, Terry PB, Weisman CS *et al.* The accuracy of substituted judgments in patients with terminal diagnoses. *Annals of Internal Medicine*, 1998; **128**: 621–629.
 - 28 Seckler AB, Meier DE, Mulvihill M, Paris BE. Substituted judgment: how accurate are proxy predictions? *Annals of Internal Medicine*, 1991; **115**: 92–98.
 - 29 Suhl J, Simons P, Reedy T, Garrick T. Myth of substituted judgment: surrogate decision making regarding life support. *Archives of Internal Medicine*, 1998; **105**: 90–96.
 - 30 Fischer GS, Tulsy JA, Rose MR, Siminoff LA, Arnold RM. Patient knowledge and physician predictions of treatment preferences after discussion of advance directives. *Journal of General Internal Medicine*, 1998; **13**: 447–454.
 - 31 da Silva FC, Reis E, Costa T, Denis L. Quality of life in patients with prostatic cancer: a feasibility study. *Cancer*, 1993; **71**(Suppl. 3): 1138–1142.
 - 32 Cassileth BR, Soloway MS, Vogelzang NJ *et al.* Quality of life and psychosocial status in stage D prostate cancer patients. *Quality of Life Research*, 1992; **1**: 323–329.
 - 33 Fossa SD, Aaronson NK, Newling D *et al.* Quality of life and treatment of hormone resistant metastatic prostatic cancer. The EORTC Genito-Urinary Group. *European Journal of Cancer*, 1990; **26**: 1133–1136.
 - 34 Litwin MS, Hays RD, Fink A *et al.* Quality-of-life outcomes in men treated for localized prostate cancer. *JAMA*, 1995; **273**: 129–135.
 - 35 Herr HW, Kornblith AB, Ofman UA. A comparison of the quality of life of patients with metastatic prostate cancer who received or did not receive hormonal therapy. *Cancer*, 1993; **71**(Suppl.): 1143–1150.
 - 36 Pedersen KV, Carlsson P, Rahmqvist M *et al.* Quality of life after radical retropubic prostatectomy for carcinoma of the prostate. *European Urology*, 1993; **4**: 7–11.
 - 37 Bennett CL, Matchar D, Crawford D. Cost-effectiveness models of flutamide: are they helpful to policy-makers? *Cancer*, 1996; **9**: 1854–1861.
 - 38 Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Services Research*, 1972; **7**: 118–133.
 - 39 Froberg DG, Kane RL. Methodology for measuring health-state preferences – II: Scaling methods. *Journal of Clinical Epidemiology*, 1989; **42**: 459–471.
 - 40 Chapman GB, Elstein AS. Utility assessment: methods and research. In: Bennett CL, Stinson TJ (eds) *Cancer Policy: Research and Methods*. Boston: Kluwer, 1998: 13–23.
 - 41 Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. *Journal of Chronic Diseases*, 1978; **31**: 697–704.
 - 42 Suarez-Almazor ME, Conner-Spady B, Kendall CJ, Russell AS, Skeith K. Lack of congruence in the ratings of patients' health status by patients and their physicians. *Medical Decision Making*, 2001; **21**: 113–121.

- 43 Schneiderman LJ, Kaplan RM, Pearlman RA, Teetzel H. Do physicians' own preferences for life-sustaining treatment influence their perceptions of patients' preferences? *Journal of Clinical Ethics*, 1993; **4**: 28–33.
- 44 Schneiderman LJ, Kaplan RM, Rosenberg E, Teetzel H. Do physicians' own preferences for life-sustaining treatment influence their perceptions of patients' preferences? A second look. *Cambridge Quarterly of Healthcare Ethics*, 1997; **6**: 131–137.
- 45 Au E, Loprinzi CL, Dhodapkar M *et al.* Regular use of a verbal pain scale improves the understanding of oncology inpatient pain intensity. *Journal of Clinical Oncology*, 1994; **12**: 2751–2755.
- 46 Clipp EC, George LK. Patients with cancer and their spouse caregivers: perceptions of the illness experience. *Cancer*, 1992; **69**: 1074–1079.
- 47 Knight SJ, Chmiel JS, Sharp LK *et al.* Spouse ratings of quality of life in patients with metastatic prostate cancer of lower socioeconomic status: an assessment of feasibility, reliability, and validity. *Urology*, 2001; **57**: 275–280.
- 48 Sneeuw KC, Aaronson NK, Osoba D *et al.* The use of significant others as proxy raters of the quality of life of patients with brain cancer. *Medical Care*, 1997; **35**: 490–506.
- 49 Sneeuw KC, Aaronson NK, Sprangers MA, Detmar SB, Wever LD, Schornagel JH. Value of caregiver ratings in evaluating the quality of life of patients with cancer. *Journal of Clinical Oncology*, 1997; **15**: 1206–1217.