Volume 10 • Number 5 • 2007
VALUE IN HEALTH

Mapping FACT-P and EORTC QLQ-C30 to Patient Health Status Measured by EQ-5D in Metastatic Hormone-Refractory Prostate Cancer Patients

Eric Q. Wu, PhD,[1] Parvez Mulani, MS, CPhil,[2] Max H. Farrell, BS,[1] Darryl Sleep, MD, MBBCh, FCS[2]

[1]Analysis Group, Inc., Boston, MA, USA; [2]Abbott Laboratories, Abbott Park, IL, USA

ABSTRACT

Objectives: To construct and validate a prediction model of preference-adjusted health status (EQ-5D) for metastatic hormone-refractory prostate cancer (HRPCA) patients using cancer-specific health-related quality of life (HRQoL) measures.

Methods: Data were obtained from a multicenter, multinational observational study of metastatic HRPCA patients conducted during 2002 to 2004. In addition to clinical and resource utilization, preference-adjusted health status (EQ-5D) and HRQoL (Functional Assessment of Cancer Therapy—Prostate [FACT-P] and European Organization for Research and Treatment of Cancer Quality of Life Questionnaire [EORTC QLQ-C30]) data were collected. Predictive validity of ordinary least square (OLS) and median regressions of various model specifications were tested using cross-validation samples. The selected specification was then further refined and tested for alternative model specifications and restrictions.

Results: OLS regression with both HRQoL measures as individual components and patient demographics was the best-performing model. It explained 58.2% of the observed EQ-5D variation in the validation sample. A model including only the prostate cancer-specific HRQoL measure, FACT-P, explained 53.5% of the observed EQ-5D variation.

Conclusions: The models developed have good predictive validity. These algorithms enable researchers to translate cancer-specific HRQoL measures to preference-adjusted health status in metastatic HRPCA patients. The findings will help perform health status adjustments in cost-utility analyses.

Keywords: EORTC QLQ-C30, EQ-5D, FACT-P, preference-adjusted health status, quality of life.

Introduction

Prostate cancer is a common disease among men in many Western countries. After the age of 40 years, the prevalence of prostate cancer increases significantly, up to 80% by the age of 80 years [1]. Currently, there are approximately 1.9 million patients with prostate cancer in the United States [2]. Although most prostate cancer patients initially respond to hormone-ablation treatment, many go on to develop metastatic hormone-refractory prostate cancer (HRPCA), which eventually causes patient death [3,4]. Each year, approximately 30,000 patients die of HRPCA [5]. Metastatic HRPCA is a disease associated with debilitating symptoms, specifically severe bone pain, reduced life expectancy, and no possibility of cure. In this setting, the focus is usually on palliative therapies that provide symptom relief. Multidimensional health-related quality of life (HRQoL) measures, such as the Functional Assessment of Cancere Therapy—Prostate (FACT-P) and the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30), provide patients with tools to subjectively evaluate their own health status. Nevertheless, these HRQoL tools do not provide information on preference-adjusted health status. This can be captured using tools such as the EuroQoL 5D (EQ-5D). The preference-adjusted health status measures are crucial to developing the cost-utility models used in appraising health-care technologies.

HRPCA trials routinely collect HRQoL information. Nevertheless, preference-adjusted health status data such as EQ-5D values are rarely collected, primarily because of the concerns of patient burden. When patient health status information is not collected, models are needed that can reasonably predict health utilities based on the HRQoL data to study the cost-utility of interventions. Previous research resulted in the creation of health utility measures, in particular the HUI, derived from the Short-Form-36 (SF-36) questionnaire and its abbreviated form, the SF-12 [6–9]. Others have also derived mapping from SF-12 to EQ-5D [9,10]. Such models, which map HRQoL to

Address correspondence to: Eric Q. Wu, Analysis Group, Inc., 111 Huntington Avenue, Tenth Floor, Boston, MA 02199, USA. E-mail: ewu@analysisgroup.com

10.1111/j.1524-4733.2007.00195.x

  1098-3015/07/408 408–414

health utilities, do not exist for the metastatic HRPCA population. Therefore, the objective of this study is to derive a prediction model of metastatic HRPCA patient preference-adjusted health status measured with EQ-5D using patient demographics and patient HRQoL measured by two cancer-specific HRQoL questionnaires: the FACT-P and the EORTC QLQ-C30. The development and test of the model used data from a multinational HRPCA prospective observational study.

EQ-5D is a standardized instrument used to measure preference-adjusted health status; it is cognitively simple and takes only a few minutes to complete. It is applicable to a wide range of health conditions and treatments and provides a simple descriptive profile and a single index value for health status. It is designed to be completed by the respondents themselves and is ideally suited for use in postal surveys, in clinics, and in face-to-face interviews. Recently, EQ-5D preference weights specific to the United States have been developed [11]. Nevertheless, a majority of the literature on the EQ-5D have used UK preference weights. To keep results consistent across study centers in different countries, and comparable to previous research, we applied the UK preference weights in this analysis.

FACT-P comprises the Functional Assessment of Cancer Therapy—General (FACT-G) and a prostate cancer subscale [12]. The FACT-G includes 29 items that measure HRQoL in cancer patients and is completed by the patients themselves. It consists of five subscales measuring physical, functional, social/family, and emotional well-being and satisfaction with the doctor–patient relationship. The FACT-G can be self-administered or used in an interview format. The prostate cancer subscale of the FACT-P includes 12 items specifically designed to measure the HRQoL in prostate cancer patients. The FACT-P is scored by summing the five subscales of the FACT-G plus the prostate cancer subscale to yield a comprehensive HRQoL score [13].

The EORTC QLQ-C30 includes 30 items and is composed of scales that evaluate functioning, symptoms, and overall health. Like the FACT-P, it can be self-administered or used in an interview format within 11 to 12 min [14]. Functioning scales include physical (five items), emotional (four items), cognitive (two items), role (two items), and social (two items) functioning as well as global health status (two items). Symptom scales measure nausea and vomiting (two items), fatigue (three items), and pain (two items), and six single items assessing financial impact and various physical symptoms. Possible scores range from 0 to 100, with higher mean scores on the functional scales representing better functioning and higher mean scores on the symptom scales representing worse symptomatology.

Methods

Study Sample and Data Collection

The study sample came from a multicenter, prospective observational study of patients with metastatic HRPCA conducted between June 5, 2002 and January 8, 2004. Study centers were located in North America, Europe, and Australia. Inclusion criteria were as follows: 19 years or older; histologically documented diagnosis of prostate adenocarcinoma (PCa) with hormone-refractory status. Patients were excluded if they were unable to comply with the requirements, had received an investigational product 4 weeks before the first day of the study, or were eligible for and planned to enter a study with another investigational product at enrollment.

Patient preference-adjusted health status (measured by EQ-5D) and HRQoL (measured by FACT-P and EORTC QLQ-C30) were collected at enrollment and follow-up visits at 3, 6, and 9 months after enrollment. The HRQoL data from this study have been presented previously [15]. Patient demographics and oncological history/disease staging information were collected at enrollment. The demographic information recorded included patient country, race, age, height, and weight as well as insurance status, educational level, marital status, and distance between residence and treatment center. Informed consent was required before patient enrollment.

Model Selection

Three different regression methods were used to build prediction models. First, ordinary least square (OLS) regression was used to construct linear prediction models of EQ-5D. Second, two additional regression methods were included in the model selection to address the potential ceiling and floor effects routinely observed with preference-adjusted health status responses: 1) OLS regression using only observations with an EQ-5D summary score less than 1; and 2) median regression. The median regression method estimates coefficients by minimizing the sum of the absolute value of observed outcome variation from the regression line [16,17]. Even though the median regression model does not explicitly deal with censoring, previous research showed that it is equivalent to censored least absolute deviations regression when censoring occurred in less than 50% of the study sample [18].

Patient HRQoL and demographics were included in the prediction models. To increase the usability of the developed mapping algorithm, we only included the most basic, and commonly collected, demographics. The majority of the patients enrolled were white (98%). Therefore, race was not included as a demographic effect in the models. To include HRQoL measures in the prediction model, for each of the

regression methods identified above, four different sets of explanatory variables were compared to select the best model specification:

1. FACT-P OVERALL score, EORTC QLQ-C30 component scores, demographics, and NO interaction terms between HRQoL measures and demographic variables.
2. FACT-P OVERALL score, EORTC QLQ-C30 component scores, demographics, and WITH interaction terms between HRQoL measures and demographic variables.
3. FACT-P INDIVIDUAL COMPONENT scores, EORTC QLQ-C30 component scores, demographics, and NO interaction terms between HRQoL measures and demographic variables.
4. FACT-P INDIVIDUAL COMPONENT scores, EORTC QLQ-C30 component scores, demographics, WITH interaction terms between HRQoL measures and demographic variables.

The demographic variables included in the model specifications were patient country, age, age-squared, and body-mass index (BMI) computed from the recorded patient height and weight. If the predicted value of EQ-5D fell outside the defined range of [–0.594, 1.000], then it was truncated to the appropriate boundary value.

When assessing prediction model performance, algorithm development data must be distinguished from the data used for model performance assessment. Because a prediction model usually performs better with the data that were used in its development, it is critical to evaluate how well the model works with other data sets. This approach is generally referred to as cross-validation. Following on previous research, we estimated the cross-validation R^2 as the primary indicator of prediction model performance [19–22]. Cross-validation sum of squared errors (SSE), the absolute deviation of the predicted mean, and the first quartile, median, and third quartile of the absolute deviation, were also estimated as secondary indicators of prediction model performance. Selection of the model was based mainly on the primary indicator. The secondary indicators were used to further check the performance of the selected model and as a benchmark when the model selection is not obvious based solely on the primary indicator (i.e., the cross-validation R^2). To calculate the cross-validation model performance indicators, the study sample was first divided into 10 equally sized groups. Each group was used successively to test each model and the remaining 90% of the sample were used to fit the prediction model. The resulting estimated prediction model was then used to estimate the performance of the original 10% of the sample. Finally, the estimated error terms were pooled to estimate the overall performance of the model.

Additional Tests and Refinement of the Selected Model

The best-fit model was further refined to improve performance and simplicity. The modified models were compared with the original model using cross-validation statistics and the final model was selected based primarily on the validation R^2. *F*-tests were also conducted for models with simplified specifications to test whether the simplifications were reasonable. The modified models included the following specifications: 1) excluding EORTC QLQ-C30; 2) using a more coarsely grouped country variable (United States, Canada, Australia, Europe); 3) excluding country effect only; 4) excluding both country effect and BMI; 5) excluding BMI only; 6) adding second order HRQoL terms (i.e., square and interaction terms); and 7) adding BMI squared.

Results

Study Sample Characteristics

The study enrolled 280 patients in seven countries: Australia (n = 40), Canada (n = 48), France (n = 26), Germany (n = 34), Italy (n = 52), the UK (n = 29), and the United States (n = 51). Italy had the highest proportion of patients (18.6%), whereas France had the smallest (9.3%). The majority of the patients enrolled were white (98%); therefore, race was not included as a demographic effect in the models. The study sample had an average age of 72.4 years (SD = 9.0) and an average BMI of 72.4 (SD = 9.0).

As a result of the large number of patient dropouts and deaths during the study, only the baseline EQ-5D and quality of life data were used to develop the mapping algorithm (prediction model). This approach was taken to simplify the model and to avoid making additional assumptions regarding missing values and early patient dropout. Table 1 presents descriptive statistics of EQ-5D and HRQoL measures at the baseline visit.

The average baseline EQ-5D is 0.64 (SD = 0.31), with a median of 0.73 and a skewness of –1.53.

Model Selection

The three regression methods had four different specifications each, for a total of 12 model specifications. The cross-validation results for all prediction models are presented in Table 2. The best-performing model is the OLS model using the component scores of FACT-P and no interaction terms. Its cross-validation R^2 (0.582) is the highest and its cross-validation SSE (0.920) is the lowest among all models. Compared with other estimates, this model's mean and median absolute deviations (0.146 and 0.122, respectively) are among the lowest values. When run on the full sample instead of cross-validation samples, the model produced an R^2 of 0.732. Therefore, it explains 58.2% of

Table 1 Descriptive baseline statistics for health utility and HRQoL variables

Variable	N	Mean	SD
EQ-5D score	276	0.635	0.309
FACT-P variables			
Physical well-being	279	20.9	5.5
Social well-being	276	21.0	4.7
Emotional well-being	279	17.2	4.6
Functional well-being	279	16.6	6.9
Total FACT-G	275	75.4	16.5
FACT-P prostate			
Cancer subscore	280	29.8	7.8
Total FACT-P	275	105.1	22.5
EORTC variables			
Global health	277	59.9	23.2
Physical functioning	278	68.8	25.7
Role functioning	279	65.8	32.9
Emotional functioning	278	73.6	22.8
Cognitive functioning	278	77.0	22.3
Social functioning	278	71.7	28.6
Fatigue	278	41.5	25.9
Nausea and vomiting	278	8.5	17.7
Pain	280	33.9	30.0
Dyspnea	278	22.3	28.7
Insomnia	278	27.2	30.7
Appetite loss	276	19.9	29.6
Constipation	278	23.0	29.3
Diarrhea	277	9.6	21.1
Financial difficulties	276	15.3	25.9

EORTC, European Organization for Research and Treatment of Cancer; FACT-G, Functional Assessment of Cancer Therapy—General; FACT-P, Functional Assessment of Cancer Therapy—Prostate; HRQoL, health-related quality of life.

the observed EQ-5D variation in the validation sample and 73.2% of the variation in the development sample. The predicted EQ-5D values have a mean of 0.62 (SD = 0.27), compared with 0.64 (SD = 0.31) for the observed EQ-5D values. The 25th, 50th, and 75th percentiles of the predicted EQ-5D values are 0.46, 0.68, and 0.82, compared with 0.60, 0.73, and 0.80 for the observed EQ-5D values.

Additional Tests and Refinement of the Selected Model

None of the simplified models with modified specifications generated cross-validation R^2 higher than its original model. The results of testing other specifications are presented in Table 3. The *F*-test results of various restricted models indicated that all of the restrictions significantly decreased the ability of the prediction model to explain the variation of the observed EQ-5D values, with the exception of the model excluding only BMI ($P = 0.04$). This simplified model has a cross-validation R^2 of 0.578, very close to that of the full model. Therefore, researchers should use the original prediction model whenever possible. Nevertheless, when BMI is not collected, a simplified model can be used without significantly sacrificing the prediction power of the model.

A simplified model excluding EORTC QLQ-C30 as an independent variable had a cross-validation R^2 of 0.54, a cross-validation SSE of 1.02, a mean absolute deviation of 0.15, and a median absolute deviation of 0.10. When run on the full sample, instead of the cross-validation samples, the model produced an R^2 of 0.62. This model can be used to predict HRPCA patient EQ-5D when only FACT-P data are collected for HRQoL. Therefore, it may be of particular interest to researchers who only collect FACT-P from HRPCA patients. The weights for the final full linear prediction model and the restricted model excluding EORTC QLQ-C30 are presented in Table 4.

Table 2 Model selection

Specification	Cross-validation *R*-squared	Sum of square errors	Absolute deviation			
			Mean	First quartile	Median	Third quartile
Simple OLS using all patients						
Using the overall FACT-P score						
Not including interaction terms	0.561	0.959	0.149	0.057	0.120	0.207
Including interaction terms	0.283	1.512	0.184	0.074	0.142	0.267
Using individual components of FACT-P						
Not including interaction terms	0.582	0.920	0.146	0.052	0.122	0.214
Including interaction terms	0.259	1.531	0.182	0.061	0.134	0.258
OLS excluding patients with EQ-5D equal to 1 (ceiling effect)						
Using the overall FACT-P score						
Not including interaction terms	0.538	0.984	0.154	0.066	0.124	0.217
Including interaction terms	0.159	1.795	0.195	0.063	0.158	0.275
Using individual components of FACT-P						
Not including interaction terms	0.561	0.946	0.150	0.054	0.124	0.214
Including interaction terms	0.129	1.860	0.197	0.064	0.146	0.269
Median regression using all patients						
Using the overall FACT-P score						
Not including interaction terms	0.573	1.001	0.134	0.039	0.103	0.180
Including interaction terms	0.121	1.903	0.152	0.018	0.109	0.232
Using individual components of FACT-P						
Not including interaction terms	0.555	1.038	0.136	0.046	0.102	0.188
Including interaction terms	0.220	1.626	0.132	0.009	0.081	0.190

FACT-P, Functional Assessment of Cancer Therapy—Prostate; OLS, ordinary least square.

Table 3 Model testing and refinement

Refinement of the original model	Cross-validation *R*-squared	Sum of square errors	Absolute deviation				*P*-value from *F*-test
			Mean	First quartile	Median	Third quartile	
Original model	0.582	0.920	0.146	0.052	0.122	0.214	
Excluding EORTC	0.535	1.023	0.146	0.050	0.103	0.209	<0.01
Grouped country effects (USA, Canada, Australia, Europe)	0.567	0.953	0.145	0.049	0.113	0.199	<0.01
Excluding country effects	0.560	0.967	0.145	0.053	0.118	0.207	<0.01
Excluding country effects, body-mass index	0.561	0.966	0.144	0.051	0.114	0.198	<0.01
Excluding body-mass index	0.578	0.927	0.145	0.051	0.113	0.207	0.04
Full model plus second order HRQoL terms	0.568	0.951	0.143	0.046	0.110	0.212	—
Full model plus the square of BMI	0.580	0.924	0.146	0.050	0.123	0.212	—

EORTC, European Organization for Research and Treatment of Cancer; health-related quality of life.

Discussion

The prediction model of patient preference-adjusted health status constructed and tested in this study had good predictive validity for metastatic HRPCA. It predicted 58.2% of the observed EQ-5D variation in the cross-validation sample, and 73.2% of the variation in the development sample. In comparison, Nichol et al. [6] predicted 50% of the observed variance in HUI scores using SF-36 scores, and Sengupta et al. [7] predicted 47% of the observed variance in HUI using SF-12 scores. Gray et al. [10] compared the performance of their mapping algorithm of SF-12 to EQ-5D with the one developed by Franks et al. [9]. Gray's mapping algorithm had a mean absolute deviation of 0.15 for patients with long-standing illness, whereas Frank's mapping algorithm had a mean absolute deviation of 0.14 for the same study sample. Our study sample had a mean absolute deviation of 0.146, comparable to the other two models (Table 5).

The inclusion criteria of the study are rather broad and future clinical studies will most likely apply similar criteria. Therefore, the study participants are likely to be good representatives of potential population of individuals who might participate in future HRPCA clinical trials.

In the process of selecting models, we found that overall, models with interaction terms of HRQoL and demographic variables performed worse than those without interaction terms. This is likely due to overfitting of the model when interaction terms were included, which made the prediction model fit "too well" for the development sample. This is usually caused by the inclusion of variables that are useful predictors of outcome variables in development data,

Table 4 Linear weights of selected prediction models

Variable	Final full model	Excluding EORTC
	Weight	Weight
Intercept	–0.081	–0.440
FACT-P		
Physical well-being	0.013	0.027
Social well-being	–0.001	–0.003
Emotional well-being	0.007	0.009
Functional well-being	–0.003	0.002
FACT-P subscale	0.003	0.009
Square of FACT-P	—	—
EORTC QLQ-C30*		
Global health	0.000	—
Physical functioning	0.003	—
Role functioning	–0.002	—
Emotional functioning	0.002	—
Cognitive functioning	0.001	—
Social functioning	0.001	—
Fatigue	0.000	—
Nausea and vomiting	0.000	—
Pain	–0.003	—
Dyspnea	0.000	—
Insomnia	0.000	—
Appetite loss	0.000	—
Constipation	–0.001	—
Diarrhea	0.000	—
Financial difficulties	0.000	—
Demographics		
Patient age	0.007	0.009
Square of patient age	0.000	0.000
Body mass index	–0.003	–0.002
Australia	0.022	–0.015
Canada	0.054	0.057
France	–0.019	–0.011
Germany	–0.012	0.027
Italy	–0.139	–0.076
UK	0.034	0.001
USA	0.000	0.000

*All dimensions of EORTC range from 0 to 100; 100 represents the best possible outcome for Global Health, Physical Functioning, Role Functioning, Emotional Functioning, Cognitive Functioning, and Social Functioning, and the worse possible outcome for all other dimensions.
EORTC QLQ-C30, European Organization for Research and Treatment of Cancer Quality of Life Questionnaire; FACT-P, Functional Assessment of Cancer Therapy—Prostate.

Table 5 Sensitivity and specificity

Percentile	Final full model		Excluding EORTC	
	Sensitivity	Specificity	Sensitivity	Specificity
10	0.548	0.904	0.323	0.960
20	0.852	0.765	0.796	0.836
30	0.836	0.700	0.904	0.725
40	0.853	0.655	0.908	0.708
50	0.890	0.674	0.934	0.688
60	0.840	0.703	0.864	0.729
70	0.863	0.644	0.879	0.611
80	0.829	0.638	0.847	0.655
90	0.992	0.184	0.988	0.132

EORTC, European Organization for Research and Treatment of Cancer.

but do not generalize well to other data [23]. At baseline, 13.8% of the study sample had the highest possible EQ-5D score (1.000) and only 0.7% had the lowest possible score (–0.594). Therefore, floor effect was not considered in the study.

For the study of patient HRQoL and preference-adjusted health status, comparison of results across different studies and different diseases is important. As a generic health status measure, EQ-5D has been widely used in research of different diseases and conditions. Therefore, when the actual measure of EQ-5D is not available, the ability to "translate" cancer and prostate cancer-specific HRQoL measures to EQ-5D makes it possible to compare outcomes from HRPCA patients with those from patients with different kinds of cancer or with other diseases. Furthermore, linking HRPCA patient HRQoL to a single-dimension generic preference-adjusted health status measure also enables the calculation of quality of life-adjusted life-year and the analysis of cost-utility. With the advance of cancer treatment, prolonged patient lives, awareness of patient quality of life, and constrained health-care budgets, the ability to estimate cost-utility of interventions and to compare them across different interventions and diseases is becoming ever more important for decision-makers to optimize medical resource allocation. Of course, a translated measure of EQ-5D should be considered a fix, rather than a substitution, for actual health status information. We recommend the collection of EQ-5D in future prospective studies of HRPCA patient when resources allow.

Population average EQ-5D varies by demographics characteristics, economic status, social class, and population base across different countries [11,24]. Our final mapping algorithm in Table 4 showed that HRPCA patients in different countries have different baseline EQ-5D when the UK preference weight is used.

In this study, we decided a priori to use cross-validation R^2 as the primary measure of model predictive validity following previous research [19–22]. Because OLS seeks to minimize the residual sum of squares and median regression seeks to minimize the absolute deviation, normally OLS will outperform median regression using cross-validation *R*-squared as the selection criterion. The purpose of applying median regression as an alternative model is that it may outperform OLS when there is ceiling effect. Judging from cross-validation R^2, OLS is still a better prediction model compared with median regression at the presence of current level of ceiling effect. Nevertheless, as Table 2 showed, the median regression model outperformed OLS on measures of absolute deviation. Given that the unconditional distribution of EQ-5D is skewed, median of absolute deviation could be the preferred measure of predictive validity. In that case, the median model will outperform OLS according to results in Table 2.

A limitation of using regression methods to develop a prediction model is that the prediction precision may be arbitrarily increased because of the fact that predicted values always have the same average as the observed values in the model development sample. A second limitation of the study is the lack of racial diversity of the study sample. Future studies with larger samples and a more diversified racial profile may be conducted to evaluate the prediction model developed in this study. In addition, the mapping algorithm was developed based on the UK preference weight of EQ-5D. Future study may be conducted to develop the mapping algorithm based on the recently developed US preference weight of EQ-5D to yields preferences that are more culturally relevant to the US population. Another limitation of the study is that it did not address how well the predicted EQ-5D measure intervention effect compared with the observed EQ-5D. Future studies with longitudinal measures of EQ-5D should be conducted to assess the mapping algorithm's ability to accurately assess impacts of interventions.

Source of financial support: This research was funded by Abbott Laboratories.

References

1 National Cancer Institute. DCCPS, Surveillance Research Program, Statistical Research and Applications Branch, Surveillance, Epidemiology, and End Results (SEER) Program. Prevalence database: "US Estimated 28-Year L-D Prevalence Counts on 1/1/2003 by Duration." Released April 2006, based on the November 2005 SEER data submission.

2 National Cancer Institute. DCCPS, Surveillance Research Program, Statistical Research and Applications Branch, Surveillance, Epidemiology, and End Results (SEER) Program. Prevalence database: "US Estimated Complete Prevalence Counts on 1/1/2003." Released April 2006, based on the November 2005 SEER data submission.

3 Oh W, Kantoff P. Management of hormone refractory prostate cancer: current standards and future prospects. J Urol 1998;160:1220–9.

4 Hanks GE, Myers CE, Scardino PT. Cancer of the prostate. In: DeVita VT Jr, Hellman S, Rosenberg SA, eds. Cancer Principles and Practice of Oncology. Philadelphia, PA: Lippincott, 1993.

5 Ries LAG, Harkins D, Krapcho M, et al., eds. SEER Cancer Statistics Review. Bethesda, MD: National Cancer Institute, 1975–2003. Available from: http://seer.cancer.gov/csr/1975_2003/ [Accessed October 30, 2006]. Based on November 2005 SEER data submission, posted to the SEER website, 2006.

6 Nichol MB, Sengupta N, Globe DR. Evaluating quality-adjusted life years: estimation of the health utility index (HUI2) from the SF-36. Med Decis Making 2001;21:105–12.

7 Sengupta N, Nichol MB, Wu J, et al. Mapping the SF-12 to the HUI3 and VAS in a managed care population. Med Care 2004;42:927–37.

8 Fryback DG, Lawrence WF, Martin PA, et al. Predicting quality of well-being scores from the SF-36: results from the Beaver Dam Health Outcomes Study. Med Decis Making 1997;17:1–9.
9 Franks P, Lubetkin E, Gold M, et al. Mapping the SF-12 to the EuroQol EQ-5D index in a national U.S. sample. Med Decis Making 2004;24:247–54.
10 Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 response and EQ-5D utility values by response mapping. Med Decis Making 2006;26:18–29.
11 Shaw J, Johnson J, Coons S. US Valuation of the EQ-5D health states development and testing of the D1 valuation model. Med Care 2005;43:203–20.
12 Cella D, Tulsky D, Gray G, et al. The Functional Assessment of Cancer Therapy (FACT) scale: development and validation of the general measure. J Clin Oncol 1993;11:570–9.
13 Esper P, Mo F, Chodak G, et al. Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument. Urology 1997;50:920–8.
14 Aaronson NK, Ahmedzai S, Bergman B, et al. The European organization for research and treatment of cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993;85:365–76.
15 Fishman M, Jacobsen P, Chen N, et al. Quality of life findings from a multicenter, multinational, observational study of patients with metastatic hormone-refractory prostate cancer. J Clin Oncol 2005 ASCO Annual Meeting Proceedings 2005;23:8106.
16 Greene WH. Econometric Analysis (4th ed.). Upper Saddle River, NJ: Prentice Hall, 2000.
17 Rousseeuw PJ. Least median of squares regression. J Am Stat Assoc 1984;79:871–80.
18 Austin PC. A comparison of methods for analyzing health-related quality of life measures. Value Health 2002;5:329–37.
19 Daley J, Jencks S, Draper D, et al. Predicting hospital-associated mortality for Medicare patients: a method for patients with stroke, pneumonia, acute myocardial infarction, and Congestive heart failure. JAMA 1988;260:3617–24.
20 Keeler EB, Kahn KL, Draper D, et al. Changes in sickness at admission following the introduction of the prospective payment system. JAMA 1990; 264:1962–8.
21 Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically ill hospitalized adults. Chest 1991;100:1619–36.
22 Naessens JM, Leibson CL, Krishan I, et al. Contribution of a measure of disease complexity (COMPLEX) to prediction of outcome and charges among hospitalized patients. Mayo Clin Proc 1992; 67:1140–9.
23 Lezzoni LI. Risk Adjustment for Measuring Health Care Outcomes. Ann Arbor, MI: Health Administration Press, 1994.
24 Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. BMJ 1998; 316:736–41.