

Nomogram Incorporating PSA Level to Predict Cancer-Specific Survival for Men With Clinically Localized Prostate Cancer Managed Without Curative Intent

Michael W. Kattan, PhD¹
 Jack Cuzick, PhD²
 Gabrielle Fisher, PhD²
 Daniel M. Berney, MD³
 Tim Oliver, MD⁴
 Christopher S. Foster, MD⁵
 Henrik Møller, MD⁶
 Victor Reuter, MD⁷
 Paul Fearn⁸
 James Eastham, MD⁸
 Peter T. Scardino, MD⁸
 and the Transatlantic Prostate Group

¹ Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio.

² Cancer Research UK Centre for Epidemiology, Mathematics and Statistics, Wolfson Institute of Preventive Medicine, St. Bartholomew's Medical School, Queen Mary, University of London, London, UK.

³ Department of Histopathology, St. Bartholomew's Hospital, London, UK.

⁴ Department of Medical Oncology, St. Bartholomew's Hospital, London, UK.

⁵ Department of Cellular Pathology and Molecular Genetics, Liverpool University Hospital, Liverpool, UK.

⁶ King's College, Thames Cancer Registry, London, UK.

⁷ Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York.

⁸ Department of Urology, Memorial Sloan-Kettering Cancer Center, New York, New York.

See editorial on pages 1–3, this issue.

Address for reprints: Peter T. Scardino, MD, Chairman, Department of Surgery, Memorial Sloan-Kettering Cancer Center, 1275 York Ave.,

BACKGROUND. The prognosis of men with clinically localized prostate cancer is highly variable, and it is difficult to counsel a man who may be considering avoiding, or delaying, aggressive therapy. After collecting data on a large cohort of men who received no initial active prostate cancer therapy, the aim was to develop, and to internally validate, a nomogram for prediction of disease-specific survival.

METHODS. Working with 6 cancer registries within England and numerous hospitals in the region, a population-based cohort of men diagnosed with prostate cancer between 1990 and 1996 was constructed. All men had baseline serum prostate-specific antigen (PSA) measurements, centralized pathologic grading, and centralized review of clinical stage assignment. Based on the clinical and pathologic data from 1911 men, a statistical model was developed and validated that served as the basis for the nomogram. The discrimination and calibration of the nomogram were assessed with use of one-third of the men, who were omitted from modeling and used as a test sample.

RESULTS. The median age of the included men was 70.4 years. The 25th and 75th percentiles of PSA were 7.3 and 32.6 ng/mL respectively, and the median was 15.4 ng/mL. Forty-two percent of the men had high-grade disease. The nomogram predicted well, with a concordance index of 0.73, and had good calibration.

CONCLUSIONS. An accurate tool was developed for predicting the probability that a man with clinically localized prostate cancer will survive his disease for 120 months if the cancer is not treated with curative intent immediately. The tool should be helpful for patient counseling and clinical trial design. *Cancer* 2008;112:69–74. © 2007 American Cancer Society.

KEYWORDS: prostate cancer, watchful waiting, curative intent, nomogram.

It is difficult to counsel men who are interested in delaying or avoiding aggressive treatment of clinically localized prostate cancer because their outcomes are so highly variable. Whereas most such men would likely not die of their cancer, some would, and these groups are difficult to distinguish at the time of diagnosis. As such, the men who opt for no definitive treatment of their prostate cancer must deal with vast uncertainty, and may later experience regret about their treatment choice.

New York, NY 10021; Fax: (216) 539-4731;
 E-mail: scardinp@mskcc.org

Received February 23, 2007; revision received
 May 29, 2007; accepted June 25, 2007.

Outcome prediction models for watchful waiting are rare relative to those available for aggressive therapies such as surgery or radiation therapy. Albertsen et al.^{1,2} developed a series of graphs that illustrate disease-specific survival for men based on age, tumor grade, and comorbidity. The accuracy of these graphs is unclear (ie, their discrimination and calibration): they are only for men aged 55–75 and they do not incorporate prostate-specific antigen (PSA), an important predictor of disease progression after surgery or radiation.^{3,4} Tewari et al.⁵ produced prediction tables that can be applied to men treated with watchful waiting. However, the calibration of these predictions is unknown, and only 467 men in their dataset were treated with watchful waiting.

The purpose of this study was to determine whether an analysis of a large and validated dataset of men with prostate cancer who received no initial definitive therapy would provide discriminatory parameters from which a nomogram for prediction of disease-specific survival could be generated.

MATERIALS AND METHODS

Study Population and Data Collection

This was a population-based study in which potential cases were identified from 6 Cancer Registries in England. Within each registry, collaborating hospitals were sought and cases from these hospitals were reviewed. National approval was obtained from the Northern Multi-Centre Research Ethics Committee, followed by Local Ethics Committee approval at each of the collaborating Hospital Trusts. Details regarding the construction of this cohort may be found in Cuzick et al.⁶

The inclusion criteria were histologic diagnosis of prostate cancer by transurethral resection of the prostate (TURP) or needle biopsy between 1990 and 1996 (inclusively) and a baseline serum PSA measurement ≤ 100 ng/mL available before or within 6 months after diagnosis. PSAs measured after TURP or the start of hormonal therapy were declared ineligible. We selected for men with clinically localized disease by excluding all cases with clear evidence of metastatic disease (by bone scan, x-ray, radiograph, computed tomography [CT] scan, magnetic resonance imaging [MRI], bone biopsy, lymph node biopsy, or pelvic lymph node dissection) or clinical indications of metastatic disease (including pathologic fracture, soft tissue metastases, spinal compression, bone pain, or T4 clinical tumor stage) at the time, or within 6 months, of diagnosis. We also selected for men without significant competing risks

by excluding those 76 years of age or older at the date of diagnosis, those who died from any cause within 6 months of diagnosis, those with other cancer diagnoses associated with high risk of death, and those who had less than 6 months follow-up after diagnosis. To reduce dataset bias from patients who had treatment decision and scheduling delays rather than true 'watchful waiting,' we excluded patients who had a radical prostatectomy or radiation therapy within 6 months after diagnosis.

Registry data collection officers and trained medical staff conducted on-site medical record reviews at each of 51 Hospital Trusts. Medical records were abstracted in the following order to minimize the detailed review of ineligible cases: First, the reviewer confirmed the date of birth and date of diagnosis of prostate cancer. Next, the reviewer searched for all PSA test results and stopped abstraction of records if no PSA test result was found before or within 6 months after histologic diagnosis. Then the reviewer obtained information regarding digital rectal examination and clinical stage, surgical, radiation, or medical treatment for prostate cancer, diagnoses of other cancers, and endpoints such as positive image studies, subsequent prostate, lymph node, or metastasis site biopsies, clinical indications of metastasis, and the date and cause of death, if noted.

Clinical staging was centrally reviewed and, where unstated, was assigned by a urologist based on documented findings of the digital rectal examination. In approximately 25% of cases no information was available and in a further 16% of cases stage could not be assigned. In both circumstances the clinical stage was statistically imputed (see Statistical Methods).

Criteria for Gleason grading were established by review of test samples among the pathology subcommittee of the investigating team. The criteria were defined and applied according to standard sources.⁷ Original histology materials from the diagnostic procedure, whether needle biopsy or TURP, were requested, collected, and centrally reviewed by a member of the pathology subcommittee to confirm the diagnosis and to assign a Gleason grade for each patient. Difficult cases were reviewed by 2 reference pathologists. When no Gleason grade could be assigned these grades were statistically imputed (see Statistical Methods). Percent cancer was defined as positive chippings over total chippings for TURP specimens ($\times 100$) and length of cancer over total core length for biopsies ($\times 100$). Outcomes were determined through medical records and cancer registry data. Data on cause and date of death were obtained and in December 2004 the cancer registries were

queried to obtain the most contemporary survival data. Deaths were classified into 2 categories according to WHO criteria: death from prostate cancer and death from other causes. Patients still alive at last follow-up were censored at that date.

Statistical Methods

Disease-specific death was estimated using the cumulative incidence method. Missing values for biopsy Gleason grade, clinical stage, and percent cancer were imputed by the method of chained equations, which uses a regression model for each variable with missing data, using all other variables as predictors, in an iterative fashion until all missing values are predicted. Baseline PSA values were required per our protocol because incorporating PSA in the prediction model was the major novelty of this study. Imputation was preferred to omitting incomplete records due to the bias potentially generated.⁸ Five imputations on the training dataset were performed. Patients with a missing method of diagnosis or who died of unknown causes were omitted from modeling given the small number of these patients. This provided a sample size of 1911, which we split randomly into training (2 of 3, N = 1310) and test (1 of 3, N = 601) sets. Cox proportional hazards regression was used for multivariable analysis. Ordinal and continuous variables were first fitted using restricted cubic splines to relax linearity assumptions. No variable selection was performed. The Cox model, arbitrarily fit to the first imputed dataset, was the basis for the nomogram.

Nomogram validation comprised 2 activities. First, discrimination was quantified with the concordance index. Similar to the area under the receiver operating characteristic curve, but appropriate for censored data, the concordance index provided the probability that, in a randomly selected pair of patients in which 1 patient dies before the other, the patient who died first had the worse predicted outcome from the nomogram. We evaluated this index in the test set of patients using 5 different models fit to the imputations (ie, 5 different concordance indices were calculated).

Second, calibration was assessed. This was done by grouping patients with respect to their nomogram-predicted probabilities and then comparing the mean of the group with the observed Kaplan-Meier estimate of disease-specific survival. Again, the test set was used for this activity and 5 different calibration curves were constructed, corresponding to the 5 different models. All analyses were performed using S-Plus 2000 Professional software (Statistical

TABLE 1
Descriptive Statistics

Variable	No.	%
Clinical classification		
T1	367	19
T2	536	28
T3	241	13
Not available	767	40
Biopsy Gleason grade		
<4 + <4	590	31
3 + 4	265	14
4 + 3	169	9
4 + 4	213	11
>4 or >4	151	8
Not available	523	27
Method of diagnosis		
Needle biopsy	1035	54
TURP	876	46
Year of diagnosis		
1990	30	2
1991	77	4
1992	145	8
1993	260	14
1994	393	21
1995	480	25
1996	526	28
Early hormones		
No	1289	67
Yes	622	33
Serum PSA, ng/mL		
Minimum	0.3	
First quartile	7.3	
Median	15.4	
Mean	23.8	
Third quartile	32.6	
Maximum	100.0	
Age at diagnosis		
Minimum	44.3	
First quartile	66.4	
Median	70.4	
Mean	69.2	
Third quartile	73.2	
Maximum	76.0	
Percentage cancer		
Minimum	0.0	
First quartile	9.3	
Median	30.8	
Mean	39.7	
Third quartile	67.4	
Maximum	100.0	
Not available	522	

TURP indicates transurethral resection of the prostate; PSA, prostate-specific antigen.

Sciences, Seattle, Wash) with the Mice, Design, and Hmisc libraries added.⁹

RESULTS

The descriptive statistics appear in Table 1. Table 2 presents PSA and age by clinical stage and Gleason

TABLE 2
Median PSA and Age by Clinical Stage and Gleason Grade

Variable		Clinical classification		
		T1	T2	T3
<4 + <4	PSA, ng/mL	7.8	11.3	20.1
	Age, y	70.1	70.0	71.5
3 + 4	PSA, ng/mL	12.5	20.0	40.1
	Age, y	70.5	71.0	72.7
4 + 3	PSA, ng/mL	19.6	25.0	48.6
	Age, y	72.1	70.3	71.3
4 + 4	PSA, ng/mL	29.0	19.6	29.6
	Age, y	69.4	70.9	70.2
>4 or >4	PSA, ng/mL	18.1	24.2	30.1
	Age, y	68.1	69.6	71.6

PSA indicates prostate-specific antigen.

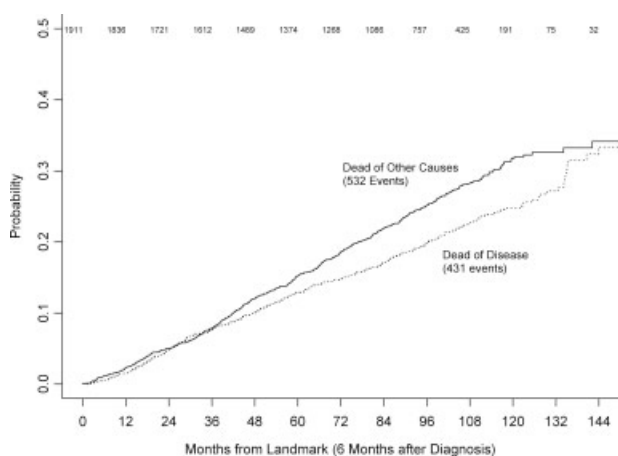


FIGURE 1. Probabilities of cause-specific death are shown. Figures at top indicate number of patients at risk.

grade. At the end of last follow-up, 431 men had died of prostate cancer and 532 had died of other causes. Figure 1 illustrates the cumulative incidence of death by cause.

A Cox model was fitted using the following predictors: clinical stage, biopsy Gleason grade, method of diagnosis, percent cancer, baseline PSA, age at diagnosis, and whether hormones were given at time of diagnosis. The hazard ratios for the 5 models appear in Table 3. A nomogram illustrating the Cox model based on the first imputed dataset appears in Figure 2. The nomogram is discriminating, with concordance indices of 0.73 to 0.74 based on 5 different models for the 5 different imputed datasets. The calibration of the 5 models appears to be very good (see Fig. 3).

TABLE 3
Hazard Ratios for the 5 Models

	Model number				
	1	2	3	4	5
Clinical stage (vs T2)					
T1	0.61	0.71	0.73	0.60	0.67
T3	0.90	1.03	1.03	0.94	1.11
Biopsy Gleason grade (vs <4 + <4)					
3 + 4	2.29	1.95	2.03	2.18	2.33
4 + 3	1.35	2.10	1.69	1.42	1.94
4 + 4	1.70	1.59	1.66	1.37	2.41
>4 or >4	2.75	2.43	2.15	2.64	2.50
Method of diagnosis (vs TURP)					
Needle biopsy	0.87	0.89	0.92	0.88	0.85
Early hormones (vs no)					
Yes	1.27	1.38	1.30	1.26	1.30
Percentage cancer, 25% increase	1.22	1.27	1.26	1.23	1.18
Serum PSA, 10 ng/mL increase	1.05	1.04	1.04	1.05	1.06
Age, 10-y increase	1.46	1.49	1.46	1.46	1.43

TURP indicates transurethral resection of the prostate; PSA, prostate-specific antigen.

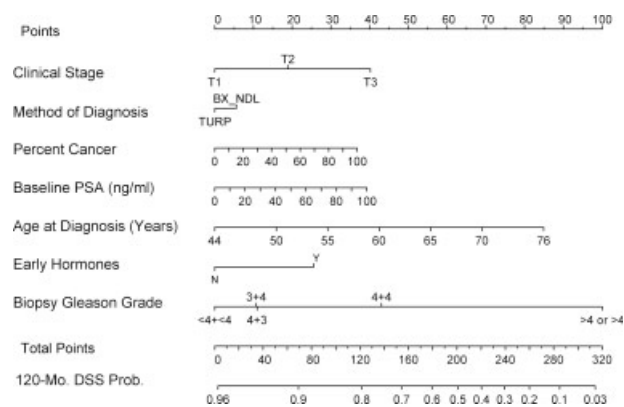


FIGURE 2. In this nomogram for disease-specific survival, BX_NDL = needle biopsy. This tool is not applicable to a man with a prior cancer diagnosis. “Early hormones” were administered within 6 months of diagnosis.

DISCUSSION

Men considering avoiding potentially curative therapy for their clinically localized prostate cancer are in critical need of outcome prediction models. Relatively few such models presently exist. We developed and internally validated the first nomogram for disease-specific survival in men managed without curative intent. In this cohort, 24% of men died of prostate cancer within 10 years. This proportion of deaths from disease is similar to that observed in early gold seed studies,¹⁰ suggesting that this treatment was not substantially more efficacious than watchful waiting. We have chosen to present our

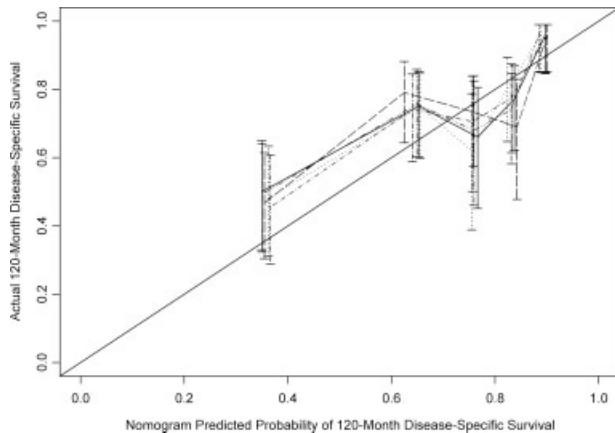


FIGURE 3. Calibration for 5 different models on test dataset is depicted in this nomogram. Horizontal axis is nomogram prediction of probability of disease-specific survival. Vertical axis is actual disease-specific survival estimated at 120 months with the Kaplan-Meier method. Vertical bars represent 95% confidence intervals (CI).

model as a nomogram due to predictive accuracy superiority over traditional risk groups and subset analysis.¹¹ Indeed, this nomogram, with a concordance index of 0.73, and very good calibration, appears to be accurate. External validation in another dataset is not yet available, although highly desirable.

We included in the nomogram a variable to indicate whether the patient had received hormones within 6 months of his diagnosis (“Early Hormones”). We believe this is useful for a few reasons. First, 33% of the men had started on hormones within 6 months, and 6 months after diagnosis is the appropriate timepoint for application of this nomogram. Thus, in a practical setting some men will have started hormones within 6 months, and the user of the nomogram should indicate this when computing the patient’s score. Note that PSA measurements after the start of hormonal therapy or TURP should not be considered when selecting the baseline value. Second, early hormones is predictive of disease-specific survival. Interestingly, it suggests a worse prognosis. Rather than concluding that hormones are harmful, a more likely interpretation is that they are a surrogate for more advanced disease (eg, symptomatic or obstructive). Therefore, ignoring the finding that some men received early hormones (ie, not including this variable in the nomogram) weakens predictive accuracy. And removing these men from the dataset (ie, making a nomogram only for those who did not receive early hormones) only loses information and applicability. Therefore, we have retained this variable in the

nomogram. However, it is important to realize that the nomogram is not designed to be used to decide whether to administer early hormones. Instead, whether they have been administered should simply be indicated on the nomogram when calculating the probability.

Although successful, there are important limitations to this study. We constructed our cohort retrospectively and this undoubtedly introduced a degree of error and bias. Men had to have a valid PSA value to be included in this study, raising the potential for selection bias effect. However, PSA was being introduced hospital-by-hospital at the time of this study. We were not able to obtain 27% of the diagnostic tissue for grading and unable to assign a clinical stage in 40% of the men, and these missing values therefore had to be imputed. The men in this cohort did not have their cancer detected through a screening program. Instead, the majority were detected by TURP. However, statistical analysis for method of diagnosis suggests that this has very little effect after adjustment for other prognostic features (Fig. 2). Nonetheless, contemporary screening programs may have a larger effect. Furthermore, whereas PSA testing detects cancers earlier, it is unclear whether these cancers portend a survival advantage after controlling for the variables in our nomogram.

Our analysis suggests that biopsy Gleason grade has the largest effect on disease-specific survival. Age at diagnosis also has a large impact, followed by serum PSA and clinical stage. It is unclear why PSA did not have a stronger effect after adjusting for the other predictors. The re-review of the Gleason grades possibly strengthened its predictive ability.

In conclusion, we developed and internally validated a nomogram for predicting the probability that a human will survive his prostate cancer if he does not have it treated immediately. This tool should be useful for individual patient counseling and clinical trial designs.

REFERENCES

1. Albertsen P, Hanley J, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA*. 2005;293:2095–2101.
2. Albertsen JL, Hanley JA, Gleason DF. Competing risk analysis of men aged 55 to 74 years at diagnosis managed conservatively for clinically localized prostate cancer. *JAMA*. 1998;280:975.
3. Partin AW, Mangold LA, Lamm DM, Walsh PC, Epstein JI, Pearson JD. Contemporary update of prostate cancer staging nomograms (Partin Tables) for the new millennium. *Urology*. 2001;58:843–848.

4. Shipley WU, Thames HD, Sandler HM, et al. Radiation therapy for clinically localized prostate cancer: a multi-institutional pooled analysis. *JAMA*. 1999;281:1598–1604.
5. Tewari A, Johnson CC, Divine G, et al. Long-term survival probability in men with clinically localized prostate cancer: a case-control, propensity modeling study stratified by race, age, treatment and comorbidities. *J Urol*. 2004;171:1513–1519.
6. Cuzick J, Fisher G, Kattan M, et al. Long-term outcome among men with conservatively treated localized prostate cancer. *Br J Cancer*. 2006;95:1186–1194.
7. Deshmukh N, Foster C. Grading Prostate Cancer. Philadelphia: WB Saunders; 1997.
8. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387.
9. Harrell FE Jr. Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer; 2001.
10. Eastham JA, Groshen SL, Kattan MW, et al. Fifteen-year survival and recurrence rates after radiotherapy for localized prostate cancer. *J Clin Oncol*. 1997;15:3214–3222.
11. Kattan MW. Nomograms are superior to staging and risk grouping systems for identifying high-risk patients: preoperative application in prostate cancer. *Curr Opin Urol*. 2003;13:111–116.