A Preoperative Nomogram for Disease Recurrence Following Radical Prostatectomy for Prostate Cancer

Michael W. Kattan, James A. Eastham, Alan M. F. Stapleton, Thomas M. Wheeler, Peter T. Scardino*

Background: Few published studies have combined clinical prognostic factors into risk profiles that can be used to predict the likelihood of recurrence or metastatic progression in patients following treatment of prostate cancer. We developed a nomogram that allows prediction of disease recurrence through use of preoperative clinical factors for patients with clinically localized prostate cancer who are candidates for treatment with a radical prostatectomy. Methods: By use of Cox proportional hazards regression analysis, we modeled the clinical data and disease follow-up for 983 men with clinically localized prostate cancer whom we intended to treat with a radical prostatectomy. Clinical data included pre-treatment serum prostate-specific antigen levels, biopsy Gleason scores, and clinical stage. Treatment failure was recorded when there was clinical evidence of disease recurrence, a rising serum prostate-specific antigen level (two measurements of 0.4 ng/mL or greater and rising), or initiation of adjuvant therapy. Validation was performed on a separate sample of 168 men, also from our institution. Results: Treatment failure (i.e., cancer recurrence) was noted in 196 of the 983 men, and the patients without failure had a median follow-up of 30 months (range, 1–146 months). The 5-year probability of freedom from failure for the cohort was 73% (95% confidence interval = 69%–76%). The predictions from the nomogram appeared accurate and discriminating, with a validation sample area under the receiver operating characteristic curve (i.e., comparison of the predicted probability with the actual outcome) of 0.79. Conclusions: A nomogram has been developed that can be used to predict the 5-year probability of treatment failure among men with clinically localized prostate cancer treated with radical prostatectomy. [J Natl Cancer Inst 1998;90:766–71]

Clinically localized prostate cancer is most often treated with conservative management (i.e., watchful waiting or hormone therapy) (1,2), external beam irradiation (3,4), or radical prostatectomy (5–8) and occasionally with therapeutic interventions such as interstitial radioactive seed implantation (9) or cryotherapy (10). Although several clinical trials are under way, no prospective randomized trials have been published that allow definitive comparison of cancer control rates among alternative treatments (11). Even when such trials are completed, all patients with a clinically localized cancer will not have an equal probability of a successful outcome. There are several established prognostic factors that can be used to predict the risk of recurrence after surgery or radiotherapy, or the risk of metastases or death from cancer after conservative management, including clinical stage (6,12,13), Gleason grade (2,3,14) and serum prostate-specific antigen (PSA) levels (3,7,13). Although these prognostic factors have been combined into a risk profile that can be used to predict the pathologic stage of the cancer (15), we are not aware of published studies that combine these standard clinical factors to predict the likelihood of recurrence or metastatic progression after treatment. We therefore developed a nomogram that predicts (before treatment) the probability of treatment failure, defined as a rising PSA level, following radical prostatectomy for clinically localized prostate cancer.

*Affiliations of authors: M. W. Kattan (Information Technology Program), J. A. Eastham, A. M. F. Stapleton, P. T. Scardino, Matsunaga-Conte Prostate Cancer Research Center and the Scott Department of Urology, and T. M. Wheeler (Department of Pathology), Baylor College of Medicine and The Methodist Hospital, Houston TX.

Correspondence to: Peter T. Scardino, M.D., Scott Department of Urology, Baylor College of Medicine, 6560 Fannin St., ST2100, Houston, TX 77030. E-mail: scardino@bcm.tmc.edu

See “Notes” following “References.”
Materials and Methods

Patients

All 1055 patients admitted to The Methodist Hospital with the intent to treat their clinically localized prostate cancer (clinical stage T1–3a NX M0) with radical retropubic prostatectomy by a single surgeon during the period from June 1983 through December 1996 were potential candidates for this analysis. Each patient was assigned a clinical stage according to the 1992 TNM (i.e., tumor-node-metastasis) classification system (T1, nonpalpable tumor confined to the prostate; T2, confined tumor palpable or visible by imaging; T3a, palpable or visible tumor extending through the capsule of the prostate unilaterally; NX, regional nodal metastases not assessed clinically; M0, no evidence of distant metastases) (12). Pelvic lymph node dissections were performed on all men. Radical prostatectomy was aborted in 24 of the 55 patients who were found to have nodal metastases based on frozen-section analysis during the operation; these men are not excluded from the analysis. However, 55 men initially treated with definitive radiotherapy and one who was treated with cryotherapy 2 years before the radical procedure (16) were excluded from the analysis. No disease follow-up information was available for 16 men, and they were also excluded. The final pathologic stage (12) distribution of the remaining 983 men was as follows: pT1–2 N0, confined to the prostate, (54.2%); pT3a,b N0, microscopic extraprostatic extension, (27.1%); pT3c N0, seminal vesicle involvement, (9.1%); and pT1–3 N+, pelvic lymph node metastasis, (9.6%). Surgical margins were reported as having been positive in 147 (15%) of the patients. Neither pathologic stage nor surgical margins (since they were not known preoperatively) was used as a predictor or outcome variable; this information is provided for comparison with and generalization to other series.

The median age of all patients was 63 years (range, 38–81 years), and 85% of the patients were Caucasian. We selected the following routinely performed clinical variables as predictors of recurrence: pretreatment serum PSA level, primary and secondary Gleason grade in the biopsy specimen, and clinical stage (assigned with use of the TNM system) (12). Pretreatment PSA was measured by the Hybritech tandem-R assay (Hybritech, Inc., San Diego, CA). In 75 patients (7.6%) treated before the PSA assay became available in our institution, no pretreatment PSA level was determined. The Gleason grade and clinical stage of each tumor were assigned by a single pathologist and a single urologist, respectively. In the interest of a parsimonious model, recently developed markers with less demonstrated predictive value (e.g., percent-free PSA) were not included in this analysis. Missing values for PSA (N = 75) and biopsy Gleason grade (N = 16) were imputed with regression models (17) containing all of the predictor variables in order to estimate the value of the missing predictor variable without reference to the outcome (PSA recurrence). Imputing a missing value is generally preferred to deleting a patient’s entire medical record, so that the maximum information is utilized and the bias that may result from a deleted case can be avoided (18). However, for comparison, a dataset consisting of only complete records was modeled as well. The clinical characteristics after imputing appear in Table 1.

Treatment Failure

The time of treatment failure was defined as either the earliest date that the postoperative serum PSA level rose to 0.4 ng/mL or higher (N = 135, confirmed by a second PSA higher than the first by any amount), or the earliest date of clinical evidence of cancer recurrence in patients with an undetectable PSA (N = 2) or no PSA result (N = 2, who developed recurrence before PSA was routinely measured). Patients who were treated with hormonal therapy (N = 8) or radiotherapy (N = 25) after surgery (but before documented recurrence) were treated as failures at the time of second therapy, since we were evaluating the ability of radical prostatectomy alone to cure the patient, and these adjuvant therapies may have masked the appearance of measurable PSA in the serum. Patients for whom radical prostatectomy was aborted because metastatic disease was found to be present in one or more lymph nodes (N = 24) were considered immediate treatment failures.

Validation Sample

A separate sample for validation was composed of 168 patients with prostate cancer who had been treated by any one of five other surgeons at our institution. These were the patients with complete records only and no values were imputed. As with the modeling sample, pretreatment PSA was measured with the Hybritech assay immediately before biopsy (if available) or before radical prostatectomy, and Gleason grading was done by a single pathologist. Each individual surgeon assigned the clinical staging for his/her patients. Patients were accrued from October 1990 through December 1996. All patients from both samples came from our Specialized Program of Research Excellence (SPORE) Prostate Information System database (Baylor College of Medicine).

Statistical Analysis

Estimates of the probability of remaining free from cancer recurrence were calculated with use of the Kaplan–Meier method. Multivariable analysis was conducted with the use of Cox proportional hazards regression. The proportional hazards assumption was verified by tests of correlations with time as well as examination of residual plots. The PSA had a skewed distribution and a suspected nonlinear effect, so it was modeled as a restricted cubic spline (18) of its log. Primary and secondary biopsy Gleason grades, each from 1 to 5, had to be collapsed into low (1–2), moderate (3), and high (4–5) grade categories because of small frequencies at the extreme values (grades 1 and 5). A potential interactive effect was anticipated because of the nature of the Gleason scoring system, so the Gleason scores were combined into 6 categories ≤2+≤2, ≤2+3, ≤3+≤2, ≤3+4, and >3+4 that allowed us to avoid forming categories with small patient counts (Table 1). Similarly, clinical stages T1a and T1b were combined since the number of patients in each was limited (T1a N = 33; T1b N = 50). Although patients in these stages are likely to have different prognoses, these differences would be largely reflected by their different PSA levels and Gleason grades. All decisions with respect to the coding of the nomogram variables were made prior to modeling, as making these decisions afterwards can have deleterious effects on the predictive ability of the model (18). This Cox model was the basis for a nomogram.

Nomogram validation contained three components. First, the nomogram was subjected to bootstrapping, with 200 re-samples, as a means of calculating a relatively unbiased measure of its ability to discriminate among patients, as quantified by the area under the receiver operating characteristic curve (20). With censored data, the receiver operating characteristic calculation (18) is slightly modified from its normal method. Nonetheless, its interpretation is similar. The area under the receiver operating characteristic curve is the probability that, given two randomly drawn patients, the patient who...
recurs first had a higher probability of recurrence. Note that this calculation assumes that the patient with the shorter follow-up recurred. If both patients recur at the same time, or the non-recurrent patient has shorter follow-up, the probability does not apply to that pair of patients. The second validation component was to compare predicted probability of recurrence versus actual recurrence (i.e., nomogram calibration) on the 983 patients, again using 200 bootstrap re-samples to reduce overfit bias, which would overstate the accuracy of the nomogram. Finally, the third validation component was simply to apply the nomogram to the 168 patients not included in the modeling sample. For these patients, their predicted probability of recurrence was compared with actual follow-up, and the area under the receiver operating characteristic curve for these men was calculated. All statistical analyses were performed using S-Plus software (PC Version 3.3, Redmond, WA) with additional functions (called “Design”) (21). All P values resulted from use of two-sided statistical tests.

Results

Of the 983 patients available for analysis, 196 had evidence of treatment failure following radical prostatectomy. For patients without disease recurrence, median follow-up was 30 months (range, 1–146 months). There were 168 (17%) patients with at least 60 months of disease-free follow-up and 19 (2%) patients with at least 120 months of disease-free follow-up. In total, there were 4281 patient-months of follow-up beyond the 5-year point. Overall recurrence-free probability for patients with clinical stage T1–3a NX M0 prostate cancer was 73% (95% confidence interval [CI] = 69%–76%) at 5 years and 68% (95% CI = 62%–73%) at 10 years (Fig. 1). Recurrence beyond the 5-year point has been rare in our series (average annual hazard rate = 0.014/year) (22). No recurrences were observed later than 100 months, but the tail of the curve is retained in Fig. 1 to illustrate follow-up. The PSA, biopsy Gleason grade, and clinical stage were all associated with recurrence (P<.001 for each).

A nomogram incorporating each of these clinical predictors was constructed based on the Cox model and appears in Fig. 2. The nomogram is used by first locating a patient’s position on each predictor variable scale (PSA analysis through clinical stage). Each scale position has corresponding prognostic points (top axis). For example, a PSA of four contributes approximately 37 points; this is determined by comparing the location of the four value on the “PSA” axis to the “Points” scale above and drawing a vertical line between the two axes. The point values for all clinical predictor variables are determined in a similar manner and are summed to arrive at a Total Points value. This value is plotted on the Total Points axis (second from the bottom). A vertical line drawn from the Total Points axis straight down to the 5-year PSA Progression-Free Survival axis will indicate the patient’s probability of remaining free from cancer recurrence for 5 years, assuming he remains alive.

Fig. 1. Kaplan–Meier estimates of disease-free probability with 95% confidence intervals for 983 patients with clinically localized (T1–3a NX M0) prostate cancer treated with radical prostatectomy during the period from June 1983 through December 1996. Numbers below the months indicate patients at risk for recurrence.

Fig. 2. Preoperative nomogram based on 983 patients treated at The Methodist Hospital, Houston, TX, for predicting prostate-specific antigen recurrence after radical prostatectomy.
The nomogram was evaluated for its ability to discriminate among patients’ risk of recurrence. This was measured as the area under the receiver operating characteristic curve for censored data. This area represents the probability that, when two patients are randomly selected, the patient with the worse prognosis (from the nomogram) will recur before the other patient. This measure can range from 0.5 (a coin toss) to 1.0 (perfect ability to discriminate). Using the original 983 patients who were modeled for the nomogram, the area was calculated to be 0.76. This value may be optimistic, since it represents an evaluation of the same patients who were modeled by the nomogram.

To derive an estimate of expected performance of the nomogram with new patients, we performed bootstrapping, a statistical method in which sampling, nomogram building, and nomogram evaluation are repeated a large number of times (23). This approach simulates the presentation of new patients to the nomogram. Through the use of bootstrapping, performance of the nomogram was estimated to be slightly worse, with an area under the receiver operating characteristic curve of 0.74. A decrease in accuracy was expected, since the nomogram had not been fitted to these new patients. However, this small decrease in accuracy indicates that the nomogram should perform with similar accuracy in other similar patient populations.

Fig. 3 illustrates how the predictions from the nomogram compare with actual outcomes for the 983 patients. The x axis is the prediction calculated with use of the nomogram and the y axis is the actual freedom from cancer recurrence for our patients. The dashed line represents the performance of an ideal nomogram, in which predicted outcome perfectly corresponds with actual outcome. Our nomogram performance is plotted as the solid line that connects the dots, corresponding to sub-cohorts (based on predicted risk) within our dataset. Note that, because the dots are relatively close to the dashed line, the predictions calculated with use of our nomogram approximate the actual outcomes. The X’s indicate bootstrap-corrected estimates of the predicted freedom from disease recurrence, which are more appropriate estimates of actual freedom from recurrence. Most of the X’s are close to the dots, indicating that the predictions based on use of the nomogram and modeled data (the dots) are near that expected from use of the new data (the X’s), though there is some regression to the mean at the extremes. The vertical bars in Fig. 3 indicate 95% CIs based on the bootstrap analysis. In general, the performance of the nomogram appears to be within 10% of actual outcome, and possibly slightly more accurate at very high levels of predicted probability.

As a final method of validation, the probability of 5-year recurrence was predicted for the separate sample of 168 patients. Of these men, 12 had disease recurrence. The predictions made with use of the nomogram were compared with actual outcomes, and the area under the receiver operating characteristic curve was calculated and found to be 0.79.

Discussion

The clinical features of prostate cancer have been combined in a nomogram to predict final pathologic stage (15), a problematic end point since some patients with apparently organ-confined cancer will later develop disease recurrence, and many patients with non-organ-confined cancer will remain disease-free (24). Not all patients with extracapsular extension or seminal vesicle involvement are destined to have disease recurrence after radical prostatectomy (5,7,8,12,25,26). Thus, the use of pathologic stage as an end point would limit the utility of a nomogram to accurately predict disease recurrence. Although final pathology has been associated with eventual treatment failure, PSA recurrence is a more appropriate measure of ultimate disease outcome (3,7,13).

Following radical prostatectomy designed to cure the patient of his cancer, the serum PSA should become undetectable (19). Measurable levels of PSA after surgery provide evidence of disease recurrence that may precede detection of local or distant recurrence by many months to years (25). Although clinical experience with elevated serum PSA levels after radical prostatectomy is not yet mature enough to quantify an association with cancer-specific mortality, elevated PSA levels are a reasonable measure of the ability of radical prostatectomy to cure a patient with prostate cancer, provided that the follow-up is long enough. We have utilized serum PSA after radical prostatectomy as an end point for treatment efficacy in an attempt to develop a model that predicts treatment failure. While our definition of recurrence by serum marker (two PSAs equal to or above 0.4 ng/mL and rising) is debatable, we feel it is relatively safe from indicating false positives, which are particularly undesirable for the
patient. Since the cutoff choice may affect the nomogram’s predicted probabilities, the results of our nomogram may be somewhat different than the actual outcome of patients at centers which use a different PSA cutoff rule. Furthermore, using a particular level of PSA as an event suggests that PSA follow-up data are interval-censored (27) (occurring between two time points) rather than right-censored (simply unknown after last negative follow-up), as we have modeled them. Future research should investigate the impact of censorship technique on the probabilities of the nomogram. However, adjuvant treatment decisions are often based on observed PSA recurrences, so that our end point is arguably more useful clinically than the true PSA recurrence time.

Other attempts have been made to predict treatment failure after radical prostatectomy (28,29). Partin et al. (29) were able to identify a group of men at high risk for disease recurrence by combining serum PSA, the grade (Gleason sum) in the radical prostatectomy specimen, and pathologic stage. Such a model may be useful in selecting men who may benefit from adjuvant treatment after radical prostatectomy. It is not useful in a preoperative assessment of the risk of disease recurrence because it relies on the final pathologic features of cancer in the radical prostatectomy specimen. A model by Bauer et al. (30) also relies on final pathologic findings. Their model predicts the relative risk of recurrence rather than a probability of recurrence, which we chose to provide because it seems easier for both physician and patient to interpret the results.

In addition to serving as a prognostic tool, the nomogram in Fig. 2 is useful for interpreting the underlying Cox model. PSA appears to be very influential across its spectrum, though patients with a very high PSA are rarely considered good candidates for surgery. The nomogram also assigns many points for clinical stage T3a and for high-grade cancer, consistent with the clinical expectations of most physicians. Some assignments appear counterintuitive (e.g., T2b>T2c), but these differences reflect variations in coefficient estimates and are not always statistically significant (two-sided P>0.05). Furthermore, it is important to consider possible changes in other variables (e.g., PSA and biopsy Gleason score) when comparing points across levels of a single variable (e.g., clinical stage). The Cox model coefficients, and therefore the resulting nomogram, look very similar when only the complete records (without imputing) are modeled (data not shown).

The nomogram developed here has certain limitations. First, we developed the nomogram in a population of patients treated with radical prostatectomy, so it is only applicable for patients who otherwise appear to be candidates for surgery, rather than all patients diagnosed with prostate cancer. Since a patient and his physician may exert selection bias for a particular treatment (in this case, radical prostatectomy) based on the characteristics of the cancer (clinical stage, grade, and serum PSA level), it would be most appropriate to apply the nomogram as a last step in the decision-making process after the patient has decided upon radical prostatectomy as his treatment choice. The nomogram is not necessarily applicable for changing the mind of the patient who has decided against radical prostatectomy since we do not know his recurrence probability; rather, it is to be used for revisiting the choice of surgery.

Second, the nomogram is not perfectly accurate. The area under the receiver operating characteristic curve on the validation sample was 0.79, while the bootstrap-corrected estimate on the original sample was 0.74, which may be overly conservative in this case. Although the difference between the two may not be statistically significant, it is somewhat odd for the validation sample performance to be higher than even the uncorrected training sample performance (0.76), so true discriminatory power may be closer to 0.74 than 0.79, since the validation sample was small with few recurrences. Also, with respect to accuracy, the CIs at the various predicted probabilities of recurrence (Fig. 3) are somewhat wide, at some levels as much as plus or minus 10%. For the individual patient, this level of error is difficult to interpret since a single patient will either recur or not. One way to apply the nomogram is to say, “Mr. X, if we had 100 men exactly like you, we would expect between [lower confidence limit] and [upper confidence limit] to remain free of their disease for 5 years following radical prostatectomy, assuming they did not die of something else first. And recurrence by PSA after 5 years is rare.” Future validations of the nomogram are necessary to evaluate the discriminatory power of the nomogram by receiver operating characteristic-curve analysis, but the dataset is too small for calibration accuracy assessment.

Future research should address from the societal perspective what should be considered an acceptable probability of recurrence after a given form of treatment for clinically localized prostate cancer. At levels of recurrence observed in our series, treatment appears beneficial for men under age 70, regardless of grade, and for men under 75 with moderate or poorly differentiated cancers and low comorbidities (33). As the probability for recurrence increases, the morbidity of treatment may begin to outweigh the possible increase in life expectancy, resulting in a net decrease in quality-adjusted life-years due to treatment. However, at the individual level, an acceptable probability of recurrence and an assessment of the risks and benefits of surgical treatment of prostate cancer remain the patient’s decision.

In this study, we developed a nomo-
gram that allows one to predict, from the treatment variables clinical stage, Gleason grade, and serum PSA level, the probability of cancer recurrence after radical prostatectomy for localized prostate cancer (clinical stage T1-T3a Nx M0). The nomogram has been constructed by combining readily available preoperative factors and may assist the physician and patient in deciding whether or not radical prostatectomy is an acceptable treatment option. It may also be useful in identifying patients at high risk of disease recurrence who may benefit from neoadjuvant treatment protocols.

References


Notes

Supported in part by a Public Health Service Grant (CA58204), Specialized Program of Research Excellence (SPORE) in prostate cancer, from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services. The authors thank Irene Albright for her careful editorial work on this project.

Manuscript received November 20, 1997; revised March 9, 1998; accepted March 19, 1998.