Postoperative Nomogram for Disease Recurrence After Radical Prostatectomy for Prostate Cancer

Michael W. Kattan, Thomas M. Wheeler, and Peter T. Scardino

**Purpose:** Although models exist that place patients into discrete groups at various risks for disease recurrence after surgery for prostate cancer, we know of no previously published work that combines pathologic factors to predict an individual's probability of disease recurrence. Because clinical stage and biopsy Gleason grade only approximate pathologic stage and Gleason grade in the prostatectomy specimen, prediction of prognosis should be more accurate when postoperative information is added to preoperative variables. Therefore, we developed a postoperative nomogram that allows more accurate prediction of probability for disease recurrence for patients who have received radical prostatectomy as treatment for prostate cancer, compared with the preoperative nomogram we previously published.

**Patients and Methods:** By Cox proportional hazards regression analysis, we modeled the clinical and pathologic data and disease follow-up for 996 men with clinical stage T1a–T3c NXM0 prostate cancer who were treated with radical prostatectomy by a single surgeon at our institution. Prognostic variables included pretreatment serum prostate-specific antigen level, specimen Gleason sum, prostatic capsular invasion, surgical margin status, seminal vesicle invasion, and lymph node status. Treatment failure was recorded when there was either 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 this set of men and a separate sample of 322 men from five other surgeons' practices from our institution.

**Results:** Cancer recurrence was noted in 189 of the 996 men, and the recurrence-free group had a median follow-up period of 37 months (range, 1 to 168 months). The 7-year recurrence-free probability for the cohort was 73% (95% confidence interval, 68% to 76%). The predictions from the nomogram appeared to be accurate and discriminating, with a validation sample area under the receiver operating characteristic curve (ie, a comparison of the predicted probability with the actual outcome) of 0.89.

**Conclusion:** A postoperative nomogram has been developed that can be used to predict the 7-year probability of disease recurrence among men treated with radical prostatectomy.


The most common definitive therapy for the treatment of clinically localized prostate cancer is radical prostatectomy. Unfortunately, approximately one third of men treated with radical prostatectomy later experience progression of their disease. Typically, the first indication that the disease has progressed is a detectable level of serum prostate-specific antigen (PSA) measured months or years after surgery. Early identification, before detectable PSA is measured, of men likely to ultimately experience disease progression would be useful in considering early adjuvant therapy. Accurate identification of the risk of disease recurrence would also be particularly useful in clinical trials to assure comparability of treatment and control groups or to identify appropriate candidates for investigational treatment, such as gene therapy. The purpose of this study was to develop a nomogram, based on the follow-up of men treated at our institution, that would predict the probability that a man will experience progression of prostate cancer after radical prostatectomy. The nomogram uses data that are routinely collected and available immediately postoperatively, so that it can be applied before the PSA level begins to rise in most men. To develop this nomogram, we used methods similar to those that we used previously to construct a preoperative nomogram based on clinical stage, Gleason grade, and serum PSA levels. Because clinical stage and Gleason grade in a biopsy sample only approximate pathologic stage and Gleason grade in the radical prostatectomy specimen, we anticipated that a nomogram that incorporated these pathologic variables would provide more accurate prediction than our previous nomogram, which utilized only those clinical variables known before definitive treatment.

From the Departments of Epidemiology and Biostatistics and Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY; Department of Pathology, Baylor College of Medicine; and The Methodist Hospital, Houston, TX.

Submitted July 2, 1998; accepted January 19, 1999.

Supported in part by a Specialized Program of Research Excellence (SPORE) grant in prostate cancer (CA58204) from the National Cancer Institute.

A PalmPilot version of this nomogram is available free of charge at http://nomograms.org.

Address reprint requests to Peter T. Scardino, MD, c/o Brenna Nichols, Urology Service, Department of Surgery, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021; email nicholsb@mskcc.org.

© 1999 by American Society of Clinical Oncology.

0732-183X/99/1705-1499
PATTERNS AND METHODS

Patients

All 1,145 patients who were treated with radical retropubic prostatectomy by a single surgeon during the period from June 1983 through June 1997 were potential candidates for this analysis. Pelvic lymph node dissections were performed on all men. Radical prostatectomy was aborted in 32 of the 58 patients who had nodal metastases determined on frozen-section analysis during the operation; these 32 men were excluded from the analysis. Also excluded were men treated with definitive radiotherapy (N = 56), hormone therapy (N = 43), cryotherapy (N = 3), or other radiotherapy (N = 3) before the radical procedure. No disease follow-up information was available for 12 men, and they were also excluded. This left 996 men for analysis. The clinical procedure. No disease follow-up information was available for 12 men, and they were also excluded. This left 996 men for analysis. The clinical stages of study subjects were as follows: T1a (3.2%), T1b (4.3%), T1c (16.5%), T2a (27.1%), T2b (24.1%), T2c (18.5%), T3a (5.4%), T3b (0.1%), and T3c (0.89%). The final pathologic stages, determined by the study pathologist after each surgical specimen was sectioned serially at 5 mm intervals3 were distributed as follows: pT2N0, confined to the prostate (55.8%); pT3aN0, prostatic capsular invasion (PCI), either focal or established (27.2%); pT3cN0, seminal vesicle involvement (9.1%); and pT2-3N1, pelvic lymph node metastasis (7.1%). Surgical margins were positive (ink touching cancer cells at the edge of the specimen) in 143 (14%) of the patients.6

The level of PCI, with respect to the stroma of the prostate, prostatic capsule, and periprostatic soft tissue, was classified as listed in Table 1. Seminal vesicle involvement was defined as cancer within the muscular coat of the seminal vesicle, not simply tumor in the fat adjacent to the seminal vesicle.6

The median age of all patients was 63 years (range, 38 to 81 years), and 88% of the patients were Caucasian. For predictors of recurrence, we selected pretreatment serum PSA level, in addition to the following routinely performed pathologic tests, as variables: Gleason sum in the surgical specimen, PCI, surgical margin status, seminal vesicle invasion, and lymph node status. Biopsy Gleason grade and clinical stage were not included as predictor variables because they are both preoperative estimates of their pathologic counterparts, which are included as predictors.

Table 1. Classification of PCI, With Respect to the Stroma of the Prostate, Prostatic Capsule, and Periprostatic Soft Tissue

<table>
<thead>
<tr>
<th>Level</th>
<th>Designation*</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>L0</td>
<td>None</td>
<td>Tumor confined to prostatic stroma within the boundary of normal prostatic acini.</td>
</tr>
<tr>
<td>L1</td>
<td>None</td>
<td>Tumor confined to prostatic stroma, but outside the boundary of normal prostatic acini.</td>
</tr>
<tr>
<td>L2</td>
<td>Capsular invasion</td>
<td>Tumor invading into but not completely through the prostatic capsule. A superiorly and at the apex where the ‘‘capsule’’ does not exist, the distinction between L1 and L2 is somewhat arbitrary.</td>
</tr>
<tr>
<td>L3F</td>
<td>Focal</td>
<td>Tumor outside the prostate to a depth of less than one high-power field on no more than two separate sections.</td>
</tr>
<tr>
<td>L3E</td>
<td>Established</td>
<td>Any amount of extraprostatic tumor more than L3F.</td>
</tr>
</tbody>
</table>

*Designation refers to terms along the “Prostate Capsular Invasion” axis of the nomogram (Fig 2).

Pretreatment PSA level was measured by the Hybritech Tandem-R assay (Hybritech, Inc., San Diego, CA). In 64 patients (6.4%) who were treated before the PSA assay became available in our institution, no pretreatment PSA level was determined. All prostates were totally embedded and sectioned by the whole-mount technique. A single pathologist (T.M.W.) measured the pathologic variables. In the interest of creating a parsimonious model, recently developed markers, or those with less demonstrated predictive value that we do not routinely measure on every patient (eg, percentage of free PSA, DNA ploidy, total tumor volume) were not included in this analysis. Missing values for PSA (N = 64), PCI (N = 9), Gleason sum (N = 4), seminal vesicle invasion (N = 3), and lymph node status (N = 3) were imputed with regression models containing all of the predictor variables 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.8 However, for comparison, a data set consisting of only complete records was modeled as well. The descriptive statistics of all predictor variables after imputing appear in Table 2.
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 \( (N = 124) \), confirmed by a second PSA measurement that was higher than the first by any amount, or the earliest date of clinical evidence of cancer recurrence in patients with an undetectable PSA level \( (N = 4) \) or no PSA result \( (N = 27) \) who developed recurrence before PSA detection was routinely measured. Patients who were treated with hormone therapy \( (N = 6) \) or radiotherapy \( (N = 26) \) after surgery but before documented disease recurrence were treated as failures at the time of second therapy, because we were interested in predicting who would eventually need a second treatment course for their cancer, and adjuvant therapy may mask the appearance of measurable PSA in the serum. An additional two men, one of whom was treated before PSA detection was available as a clinical test, were reported as dead of prostate cancer with no available documentation to support evidence or time of recurrence prior to death, and these patients were considered treatment failures as of the date of death.

Validation Sample

A separate sample for validation was composed of 322 patients with prostate cancer who had been treated by any one of five other surgeons at our institution. These patients had complete records only, and no values were imputed. As with the modeling sample, pretreatment PSA level was measured with the Hybritech assay immediately before biopsy (if available) or before radical prostatectomy, and pathologic variables were measured by a single pathologist. Each individual surgeon assigned the clinical staging for his/her patients. Patients were accrued from October 1990 through June 1997. 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 recurrence were calculated with the Kaplan-Meier method. Multivariable analysis was conducted with Cox proportional hazards regression. The proportional hazards assumption was verified by tests of correlations with time and examination of residual plots. PSA status had a skewed distribution and suspected nonlinear effect, so it was modeled as a restricted cubic spline\(^8\) of its log. Similarly, Gleason sum was suspected to be nonlinear and also modeled with a restricted cubic spline function. Prostate cancer within the confines of the glandular prostate (group L0) or in the prostatic stroma but beyond the limit of the normal acini (group L1) had to be combined as "None" since no patients in either group experienced recurrence, which would prohibit convergence of the Cox algorithm. All decisions with respect to the coding of the nomogram variables were made prior to modeling, because making these decisions afterward can have deleterious effects on the predictive ability of the model.\(^8\) This Cox model was the basis for a nomogram, and our modeling and validation procedure is similar to that used previously.\(^1\)

Nomogram validation contained three components. First, the nomogram was subjected to bootstrapping, with 200 "resamples," 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.\(^9,10\) With censored data, the receiver operating characteristic calculation\(^8\) 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 relapses first had a higher probability of disease recurrence. Note that this calculation assumes that the patient with the shorter follow-up relapses. If both patients relapse at the same time, or the nonrelapsing patient has a shorter follow-up, the probability does not apply to that pair of patients. The second validation component was to compare the predicted probability of disease recurrence versus actual recurrence (ie, nomogram calibration) on the 996 patients, again using 200 bootstrap resamples 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 322 patients not included in the modeling sample. For these patients, their predicted probability of disease recurrence was compared with actual follow-up results, 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 4.0; Redmond WA) with additional functions (called Design)\(^10\) added. All \( P \) values resulted from the use of two-sided statistical tests.

RESULTS

Of the 996 patients available for analysis, 189 had evidence of recurrence of prostate cancer after radical prostatectomy. For patients without disease recurrence, the median follow-up period was 37 months (range, 1 to 168 months). There were 222 patients with at least 60 months’ disease-free follow-up, 109 with 84 months’ disease-free follow-up, and 31 patients with at least 120 months’ disease-free follow-up. Overall recurrence-free probability for these patients with clinical stage T1a-T3c N0-1 M0 prostate cancer was 75% (95% confidence interval [CI], 72% to 79%) at 5 years, 73% (95% CI, 68% to 76%) at 7 years, and 71% (95% CI, 66% to 75%) at 10 years (Fig 1). Disease recurrence beyond the 7-year point is rare in our series.\(^11\) No recurrences were observed later than 97 months, but the tail of the curve is retained in Fig 1 to illustrate follow-up. In the multivariable model, all variables were associated with recurrence \( (P < .01 \) for each).

A nomogram incorporating each of these clinical predictors was constructed based on the Cox model is shown in Figure 2. The nomogram is used by first locating a patient’s position on each predictor variable scale (PSA value through lymph node status). Each scale position has corresponding prognostic points (top axis). For example, a PSA value of 4 contributes approximately 78 points; this is determined by comparing the location of the 4 value on the “PSA” axis to the “Points” scale above and drawing a vertical line between the two axes. The point values for all 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 84-Month PSA Progression-Free Survival axis will indicate the patient’s probability of remaining free from cancer recurrence for 7 years, assuming he remains alive.
The nomogram was evaluated for its ability to discriminate among patients' risk of disease 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 relapse before the other patient. This measure can range from 0.5 (a coin toss) to 1.0 (perfect ability to discriminate). Using the original 996 patients who were modeled for the nomogram, the area was calculated to be 0.88. This value may be optimistic, because it represents an evaluation of the same patients who were modeled by the nomogram.

To derive an estimate of expected performance of the nomogram against new patients, we performed bootstrapping, a statistical method in which sampling, nomogram building, and nomogram evaluation are repeated a large number of times. This approach simulates the presentation of new patients to the nomogram. With the use of bootstrapping, performance of the nomogram was essentially unchanged, with an area under the receiver operating characteristic curve of 0.88. A decrease in accuracy was expected. However, no finding of any decrease suggests that the nomogram should provide similar accuracy when applied to additional, similar patient populations.

Figure 3 illustrates how the predictions from the nomogram compare with actual outcomes for the 996 patients. The x-axis represents the prediction calculated with use of the nomogram, and the y-axis represents 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 subcohorts (based on predicted risk) within our data set. 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 very close to the dots, indicating that the predictions based on use of the nomogram and modeled data (the dots) are near those expected from use of the new data (the X’s). The vertical bars in Fig 3 indicate 95% confidence intervals 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 7-year recurrence was predicted for the separate sample of 322 patients. Of these 322, 20 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 0.89.
DISCUSSION

Traditionally, the judgment of which patients are at high risk for failure after radical prostatectomy has been based largely on final pathologic stage. Final pathologic stage alone is a problematic variable for judging high-risk disease because some patients with apparently organ-confined cancer will later develop disease recurrence, and many patients with non–organ-confined cancer will remain disease-free. Not all patients with PCI or seminal vesicle involvement are destined to have disease recurrence after radical prostatectomy. Thus the use of individual pathologic features appears to be insufficient to estimate risk of recurrence; a method of combining them is needed.

After radical prostatectomy designed to cure the patient of his cancer, the serum PSA level should become undetectable. 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.

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 period is long enough. We have utilized serum PSA values after radical prostatectomy as end points for treatment efficacy in an attempt to develop a model that predicts treatment failure. Although our definition of recurrence by serum marker (two PSA values equal to or above 0.4 ng/mL and rising) is debatable, we feel that it is relatively safe from indicating false positives, which are particularly undesirable for the patient. Because the cutoff choice would affect the nomogram’s predicted probabilities, the results of our nomogram may be somewhat different from the actual outcome of patients at centers that 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 \(^{20}\) (occurring between two time points) rather than right-censored (simply unknown after the 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 made on the basis of observed PSA recurrences, so our end point is arguably more useful clinically than the true PSA recurrence time.

In light of this information, we sought to combine aspects of final pathology in a manner that can predict PSA recurrence. Others have advocated such an approach. In 1995, Partin and colleagues\(^{21}\) published a model that was derived using 216 men with clinical stage T2b and T2c prostate cancer who were treated by a single urologist. The model utilized pretreatment serum PSA value, radical prostatectomy Gleason score, and pathologic stage as specimen-confined or non-specimen-confined. Their model computed log relative risk and categorized patients into low-, intermediate-, and high-risk groups. In a validation cohort of 214 patients treated by one of three different urologists at two institutions, Partin et al were able to stratify those patients as well, based on their Kaplan-Meier PSA recurrence-free survival rates, although no statistical testing of strata differences was performed. Bauer et al\(^{22}\) recently emulated the approach of Partin et al with 378 patients but added race as a predictor variable and widened the cohort to include clinical stages T1a through T2c. Another difference with the Bauer et al model was the cutoffs used to distinguish the risk groups (Partin et al used relative risk in excess of 4 and 5.75 to indicate moderate and high risk, respectively, whereas Bauer used 10 and 30 for the same risk groups). The validation cohort of Bauer et al of 99 men indicated a difference in survival rates between the low- and high-risk groups but no difference between intermediate risk and either low or high risk. In another recent study, Bauer et al\(^{23}\) added biomarkers p53, Ki-67, and bcl-2 to the relative-risk calculation.

Our approach extends the previous work of Bauer et al in some important ways. Most notably, our nomogram provides the patient with his probability of disease recurrence instead of a relative risk because we believe that the patient is more likely to comprehend the probability. Although the relative risk informs the patient of his risk of disease recurrence relative to another patient with certain characteristics, the actual probability should facilitate decision-making for the patient. The disadvantage of the probability approach we are using over the relative-risk approach used previously is that when reporting a probability, the point in time must be specified. Too early a time point (eg, probability of disease recurrence within 2 years) loses clinical usefulness by being inconclusive. Too late a time point has the disadvantage of potentially being estimated when few patients in the series are at risk, so that the level of precision may be low. In the study presented here, we chose a time point of 84 months in an attempt to balance these concerns. For patients with PSA levels that are persistently undetectable between 60 and 84 months, disease recurrence, as estimated by PSA detection is very rare after 84 months, which provides support for judging whether surgery is effective; however, we still had 109 patients at risk for
recurrence in our model at 84 months and 2,712 patient-months of follow-up beyond the 84-month point, suggesting that our estimate of the probability at that time point will remain reasonably stable.

Another manner in which our approach differs from those previously published lies in our methods of validation and assessment. The previous work by Partin et al and Bauer et al illustrates the extreme difficulty in validating a survival model. Using validation cohorts, both investigative groups produced Kaplan-Meier estimates for the risk strata; however, probably because of small sizes of the cohorts, neither study was able to report all pairwise significant differences among the strata (ie, each strata being different from each other strata). We enhanced the efficiency of validation and assessment in two ways. First, we utilized bootstrapping so that each patient could legitimately be used for both model development and model assessment. This more fully utilizes the data set at hand than does the approach of dividing the data set into strata. Second, we report an overall measure of the ability of our model to discriminate among the individual patient’s risk of relapse. In this manner, we avoid having to form strata that combine patients who are at varying levels of risk into the same risk group. Instead, our discrimination measure (area under the receiver operating characteristic curve for censored data) compares each pair of patients and quantifies the degree to which the model was able to rank those patients. Moreover, we were able to bootstrap the discrimination measure to obtain a reasonable estimate of expected discrimination ability on future data. As two final points of difference with previous studies, we chose to include patients with clinical stage T3b and T3c disease and utilized relatively large data sets (N = 996 for derivation and N = 322 for validation).

Although prognostic nomograms based on clinical, rather than pathologic, factors could be developed for a variety of different forms of definitive treatment (including conservative management or “watchful waiting”), a nomogram that incorporates pathologic factors known only after the prostate is removed may have several important uses in clinical trials, in addition to potentially comforting the patient who is at low probability of recurrence. First, it should be useful in identifying patients who are appropriate for a clinical trial. Use of the nomogram assists the clinician in determining the patient’s baseline probability of disease recurrence and providing this information to the patient, which should facilitate the decision-making about whether adjuvant therapy is necessary and worth the side effects. Second, as an extension of the first use, the nomogram can potentially assist the clinician in quantifying the expected benefit relative to the baseline risk. A patient at very low risk for recurrence may not have much to gain from the introduction of a novel adjuvant treatment program. In conjunction with the expected efficacy of adjuvant therapy, the nomogram allows quantification of this potential net gain. This is useful even after a clinical trial demonstrates superiority of one treatment over another, because the degree of benefit could be highly variable among patients who are at different baseline risks. Third, the nomogram can be used to verify the effectiveness of randomization. Treatment arms should have very similar average baseline risks. Fourth, the nomogram may make it possible to reduce the sample sizes of clinical trials for adjuvant therapies. A typical multivariate analysis consumes several degrees of freedom to adjust for the potential effects of confounding variables. In other words, part of the sample-size requirement for a new trial is associated with estimating the effect of the new therapy, and part is associated with adjusting for the effects of the patient’s baseline variables. By collapsing the effects of several baseline variables into an overall recurrence risk (which consumes fewer degrees of freedom than do the individual components), a smaller sample may be required because of a potentially smaller demand placed on the trial data to be able to adjust for baseline differences in the treatment arms. Fifth, a uniform method of patient description would help to facilitate comparisons across studies. Typical studies report univariable tables of each baseline variable that do not illustrate potential differences in their joint distribution, which would be considered in the nomogram.

Other possible uses of the nomogram include facilitation of the search for a new marker of eventual recurrence after surgery for prostate cancer. Analogous to the clinical trial use above, the sample-size requirements to evaluate whether a new marker contributes to the prognostic ability of existing markers may be reduced. In the nomogram, the ability of the previous markers is collapsed into an overall risk measure that may require a smaller sample size for adjustment, which in turn would reduce the overall sample-size requirement and thus the number of patients who need to have their new markers measured. Another major use of the nomogram is related to the desire to provide cost-effective treatment for society. By quantifying the expected benefit that a patient is to receive from a potential treatment and incorporating its cost, a calculation is facilitated as to whether a treatment’s expected benefit is worth its expected cost. The purpose here is not to deny the treatment to the patient but instead to decide whether the treatment is cost-effective from society’s point of view (ie, whether it should be reimbursable).

In addition to serving as a prognostic tool, the nomogram in Fig 2 is useful for interpreting the underlying Cox model.
For example, it appears that PSA is very influential across its spectrum. Also, the nomogram assigns points for the levels of PCI, consistent with degree of tumor spread. Similarly, positive margins, seminal vesicle invasion, and positive lymph nodes each increase the number of points the patient receives toward recurrence. However, the point assignment for Gleason sum appears to be counterintuitive (eg, sum = 3 worse than sum = 4 worse than sum = 5), but these differences reflect variations in coefficient estimates and are not statistically significant (two-sided P > .05). Furthermore, it is important to consider possible changes in other variables (eg, PSA) when comparing points across levels of a single variable (eg, seminal vesicle invasion), because patients who differ on one axis are likely to differ on another axis and not be held constant, as one might assume when comparing horizontal distances across axes. 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, the nomogram is not perfectly accurate. The area under the receiver operating characteristic curve on the validation sample was 0.89, whereas the bootstrap-corrected estimate on the original sample was 0.88. Thus, in 11% to 12% of patient pairs, the patient with the better prognosis actually relapsed first. If there are indicators of the presence of cancer cells outside the field of resection other than PSA, grade, and pathologic stage, more accurate nomograms must await the discovery and characterization of these markers (such as p53 or p27 abnormalities or increased apoptotic index). Also, with respect to accuracy, the confidence intervals at the various predicted probabilities of recurrence (Fig 3) are somewhat wide—at some levels as much as ±10%. For the individual patient, this level of error is difficult to interpret because a single patient will either relapse 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 7 years, assuming they did not die of something else first. And recurrence by PSA after 7 years is rare.” Future nomograms with mature data sufficient for predicting beyond 7 years will be beneficial.

Third, we modeled the data from a single surgeon, and all data came from the same institution. Most of the patients were Caucasian, although other researchers have found no effect of race in multivariable recurrence models prior to variable selection after controlling for fewer pathologic criteria ($P = .083, 30 P = .054, 23$ not shown$^{22}$). Although the validation was performed on data that had been obtained from different surgeons and accrued more recently than the data in the nomogram, there may be commonalities among them, because all patients were treated by surgeons in a single department of urology with resident physicians rotating. Fourth, a single expert pathologist performed all pathologic assessments, using a uniform method of tissue processing and analysis. The accuracy of the nomogram in the wider medical community assumes comparable grading and staging accuracy by other pathologists. Fifth, the applicability of the nomogram assumes that the probability of cancer control after radical prostatectomy is similar when surgeons at other institutions perform the surgery. In fact, there may be substantial variations in outcome, partially due to technical aspects of the operation, such as the extent of resection of tissue around the prostate. Future validations of the nomogram are necessary to evaluate the degree of this limitation. Finally, the nomogram is not applicable to patients given prostate cancer treatment (such as androgen ablation or irradiation therapy) before radical prostatectomy, since these manipulations may alter the histologic appearance or apparent extent of the cancer$^{25}$ without altering PSA progression. Our validation data set was useful for measuring the discriminatory power of the nomogram by receiver operating characteristic curve analysis, but the data set is too small for calibration accuracy assessment.

In this study, we developed a postoperative nomogram that allows one to predict, from the serum PSA level, degree of PCI, specimen Gleason sum, surgical margin status, seminal vesicle invasion, and lymph node status, the probability of cancer recurrence after radical prostatectomy for prostate cancer. This nomogram is not useful preoperatively, because the prostate must be removed before the nomogram can be used. For preoperative risk assessment, we have developed a preoperative nomogram.$^1$ The postoperative nomogram offers greater accuracy (area under the receiver-operating characteristic curve, 0.89) than the preoperative nomogram (area under the receiver-operating characteristic curve, 0.74), because of the wealth of information obtained upon prostate removal.

The postoperative nomogram has been constructed by combining readily available factors and may assist the physician and patient in deciding whether or not adjuvant therapy is an acceptable treatment option. It may also be useful in the design of adjuvant treatment protocols. The use of nomograms combining multiple prognostic variables could favorably influence the efficiency of clinical trials.

ACKNOWLEDGMENT

The authors thank Irene Albright and Linda Higgins of the Scott Department of Urology at the Baylor College of Medicine for their careful editorial work on this project. The authors also thank Carie Seale-Hawkins, also of Baylor, for her dedicated database management efforts over the past decade.
REFERENCES

7. Harrell FE: Transcan: An S-Plus function. Program available from statlib@lib.stat.cmu.edu. Send e-mail `send transcan from S`., 1995
10. Harrell FE: Design: S-Plus function for biostatistical/epidemiologic modeling, testing, estimation, validation, graphics, prediction, and typesetting by storing enhanced model design attributes in the fit. 1994. Programs available from statlib@lib.stat.cmu.edu