



Published in final edited form as:

Cancer. 2009 March 1; 115(5): 1005–1010. doi:10.1002/cncr.24083.

Preoperative and Postoperative Nomograms Incorporating Surgeon Experience for Clinically Localized Prostate Cancer

Michael W. Kattan, PhD¹, Andrew J. Vickers, PhD², Changhong Yu, MS¹, Fernando J. Bianco, MD³, Angel M. Cronin, MS², James A. Eastham, MD³, Eric A. Klein, MD⁴, Alwyn M. Reuther, MS⁴, Jose Edson Pontes, MD⁵, and Peter T. Scardino, MD³

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

²Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, New York

³Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, New York

⁴Glickman Urologic and Kidney Institute, Cleveland Clinic, Cleveland, Ohio

⁵Department of Urology, Wayne State University, Detroit, Michigan

Abstract

BACKGROUND—Accurate preoperative and postoperative risk assessment has been critical for counseling patients regarding radical prostatectomy for clinically localized prostate cancer. In addition to other treatment modalities, neoadjuvant or adjuvant therapies have been considered. The growing literature suggested that the experience of the surgeon may affect the risk of prostate cancer recurrence. The purpose of this study was to develop and internally validate nomograms to predict the probability of recurrence, both preoperatively and postoperatively, with adjustment for standard parameters plus surgeon experience.

METHODS—The study cohort included 7724 eligible prostate cancer patients treated with radical prostatectomy by 1 of 72 surgeons. For each patient, surgeon experience was coded as the total number of cases conducted by the surgeon before the patient's operation. Multivariable Cox proportional hazards regression models were developed to predict recurrence. Discrimination and calibration of the models was assessed following bootstrapping methods, and the models were presented as nomograms.

RESULTS—In this combined series, the 10-year probability of recurrence was 23.9%. The nomograms were quite discriminating (preoperative concordance index, 0.767; postoperative concordance index, 0.812). Calibration appeared to be very good for each. Surgeon experience seemed to have a quite modest effect, especially postoperatively.

CONCLUSIONS—Nomograms have been developed that consider the surgeon's experience as a predictor. The tools appeared to predict reasonably well but were somewhat little improved with the addition of surgeon experience as a predictor variable.

Keywords

prostate cancer; surgeon experience; recurrence; predictive value; nomogram

© 2009 American Cancer Society.

Corresponding author: Peter T. Scardino, MD, Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY 10021; Fax: (212) 988-0874; scardinp@mskcc.org.

Conflict of Interest Disclosures

Jose Edson Pontes is on the speakers bureau for Astra Zeneca.

Previous studies suggest that the number of cases performed by a surgeon is predictive of prostate cancer recurrence even after controlling for the individual patient characteristics.¹⁻³ This finding would suggest that, when counseling a patient as to his risk of recurrence, an adjustment should be made for the experience of the surgeon. Such an adjustment would improve the predictive accuracy of a tool, especially when applied across surgeons who have a broad range with respect to experience.

We used individual patient data from 4 institutions to develop preoperative and postoperative models to predict biochemical recurrence after radical prostatectomy. Beyond the usual preoperative and postoperative patient characteristics, we adjusted for the number of cases performed before the present case to reflect surgeon experience. These models were internally validated with use of bootstrapping methods and cross-validated among institutions. Moreover, the predictive accuracy was evaluated with and without adjustment for surgeon experience to characterize the improvement associated with this adjustment.

METHODS

Sources of Data and Study Cohort

The study cohort comprises 9376 patients with clinically localized prostate cancer treated by open radical retropubic prostatectomy between 1987 and 2003. Patient data were obtained from 4 institutions: Memorial Sloan-Kettering Cancer Center, Baylor College of Medicine, Wayne State University, and the Cleveland Clinic. All data were de-identified before analysis. Patients with clinical stage T1a or T1b disease (N = 165), who received neoadjuvant therapy (n = 1316), adjuvant therapy (n = 85), or who had missing data for either surgeon (n = 144) or prostate specific antigen (PSA; n = 66), were excluded, leaving 7724 patients eligible for analysis. All information was obtained with appropriate Institutional Review Board waivers.

Eligible patients were treated by 1 of 72 surgeons, all of whom saw patients only at the study institutions while on staff. Surgeons who performed their initial radical prostatectomies at a prior institution were asked to provide their prior experience. Approximately half (38) of the surgeons performed radical prostatectomy only at a study institution, and the majority of the rest (22) performed fewer than 20 radical prostatectomies before treating their first patient at their current institution. Thus, we have data on all or almost all of the study surgeons' patients throughout their careers to date.

Outcomes

Patient follow-up was conducted according to accepted clinical practice at each institution. In general, this consisted of serum PSA measurements every 3 months to 4 months during the first postoperative year, semiannually the second year, and annually thereafter. Cancer recurrence was defined as a PSA level >0.4 ng/mL and rising. A secondary treatment initiated for a detectable and rising PSA ≥ 0.4 ng/mL was also considered an event.

Statistical Methods

For each patient, surgeon experience was coded as the number of radical prostatectomies performed by the surgeon before the patient's operation. This number reflects total prior experience over the surgeon's entire career, including operations conducted at former institutions, and those for patients ineligible for analysis. Therefore, surgeon experience differs for each patient treated by a particular surgeon. Only a single billing surgeon was recorded for each operation: operations in which a surgeon assisted, such as during fellowship training, were not counted toward prior experience for the assisting surgeon.

Estimates of the probability of remaining free of progression were calculated by the Kaplan-Meier method. Patients were censored if they were lost to follow-up or died from causes other than prostate cancer. Multivariable analysis was performed with Cox proportional hazards regression. Continuous variables were modeled with restricted cubic splines to relax linearity assumptions. For model validation, we assessed both discrimination and calibration. Discrimination refers to the ability of the nomogram to rank patients by their risk, such that patients with a higher predicted risk of treatment failure should be more likely to recur. Discrimination was measured using the c-index, which is similar to an area under the receiver operating characteristic curve, and is applicable to time-to-event data. We used the method of Harrell et al⁴ to compute the c-index for each model. Model c-indices were compared by quantifying the difference in bootstrap-corrected concordance indices of models with and without surgeon experience. A large number of bootstrapping resamples ($B = 2000$) was drawn, ensuring the reliability of the P value that measured the extent of improvement in the concordance index if surgeon experience was included as a predictor. Calibration refers to the accuracy of the nomogram and is assessed by a visual inspection of the plots of predicted probability of progression versus actual outcome. All statistical analysis was performed using *S-Plus* software (*S-plus 2000*; Insightful Corp, Redmond, Wash) with additional functions (called “Design”) added.⁵

RESULTS

Descriptive statistics appear in Table 1. Whereas most surgeons (57%) had performed <50 cases, many surgeons (22%) had performed 100 cases.

Time to recurrence appears in Figure 1. There were 1314 recurrences. Median follow-up for patients without recurrence was 3.9 years. Only a small proportion of patients died without recurrence, with a 5-year overall survival probability of 81%. This finding suggests that adjustment for competing risk would have a negligible effect.

The preoperative nomogram predicting a 10-year freedom from biochemical recurrence appears in Figure 2. Surgeon experience alters the predicted probability to an appreciable extent, after holding other preoperative features of the patient constant. The concordance index for this model is 0.767. This value remains high (0.764) for a model that lacks surgeon experience ($P = .113$ from a paired permutation test). The calibration curve for the preoperative nomogram appears in Figure 3, along with the calibration of the model that lacks surgeon experience. The nomogram seems to be well calibrated.

Figure 4 illustrates the postoperative nomogram that predicts a 10-year freedom from recurrence. Of interest, the impact of surgeon experience is slightly less in this model. The concordance index for this nomogram is 0.812, while the value is 0.811 when surgeon experience is removed ($P = .145$ from a paired permutation test). Figure 5 illustrates the calibration of the nomogram, which is very good. Also shown in Figure 5 is the calibration for the model that lacks surgeon experience. In both models, we included year of surgery as a predictor to adjust its impact on the recurrence of prostate cancer. However, we hid year of surgery from the nomograms to reflect a contemporary patient prediction.

In Figure 6, we provide a scatterplot of the 10-fold cross validated predicted probabilities for the pairs of pre-operative (Fig. 6A) and postoperative (Fig. 6B) models. We arbitrarily added lines to indicate $\pm 10\%$ difference in predicted probabilities. Particularly for the postoperative nomogram, very few patients would have a difference of 10% after adjusting for surgeon experience. However, 10% is arbitrary, and it is difficult to state what magnitude of a difference might affect a patient's decision.

DISCUSSION

In consideration of the growing literature that surgeon experience matters,¹⁻³ we have constructed new nomograms to directly incorporate this variable into predictions of biochemical recurrence. We found that adjustment for surgeon experience in the predictions resulted in quite minor improvements in discrimination of the preoperative and postoperative nomograms. In some cases, surgeon experience seems to have a clinically significant impact on a patient's predicted probability of biochemical recurrence. However, calibration of the nomograms was not materially affected by consideration of surgeon experience.

It is important to note that the present study is not concerned with the scientific question of whether a surgeon improves over time. Prior literature has addressed that topic.¹⁻³ The focus of the present study is on the practical question of whether a patient's prognosis should be further adjusted, whether preoperatively or postoperatively, by the number of cases that the surgeon has performed. These are different questions that require different types of analyses.

It is unclear exactly what would drive the improved prognosis associated with surgeon experience. It is possible that higher volume surgeons have different case mix than lower volume surgeons.⁶ However, given the variables we adjusted for in the nomograms, it is not entirely obvious what other factors may be used by higher volume surgeons in deciding upon which patients to operate. Nonetheless, incorporating surgeon experience in the nomogram needs to be studied, which is the focus of this analysis.

There are several limitations to our study, which is retrospective. Perfect predictive accuracy was not achieved. Scientific progress is made as incremental improvements occur. Future studies need to make further improvements. A potential limitation in our study was that surgeon experience may be confounded with follow-up assessment frequency. However, our prior sensitivity analysis³ provided no evidence of this confounding. In additional sensitivity analyses, we cross-validated the models by predicting outcomes for each institution after it was omitted from the development data set. The concordance indices ranged from 0.758 to 0.740 (for preoperative) and 0.789 to 0.840 (for postoperative) for predictions at each institution, indicating good nomogram stability across institutions.

Our outcome, biochemical recurrence, is not the ideal or sole outcome for treatment decision making. However, virtually all patients who develop metastasis first develop PSA recurrence. Furthermore, one would not recommend a patient receive a treatment if it had 100% chance of PSA failure. There is likely a threshold below which a treatment would not be recommended, even based on PSA failure. In addition, much of our treatment guidelines are presently based on PSA recurrence as an outcome measure, and PSA recurrence currently triggers second treatments. These treatments impact the patient's quality of life. Finally, patients who develop PSA recurrence, are on average, upset by this.⁷ Having said this, another limitation is that we were unable to distinguish Gleason 4 + 3 from 3 + 4 patients because 2 institutions only had Gleason sum available. Furthermore, our nomograms assume a PSA recurrence definition of 0.4 ng/mL and rising, based on the analysis by Stephenson et al.⁸ Our nomograms may not be accurate in other PSA recurrence definition settings.

In conclusion, we have developed new preoperative and postoperative nomograms that include surgeon experience as predictors. For the majority of patients, incorporating surgeon experience will not greatly affect the predicted probabilities. More accurate tools remain needed.

References

1. Klein EA, Bianco FJ, Serio AM, et al. Surgeon experience is strongly associated with biochemical recurrence after radical prostatectomy for all preoperative risk categories. *J Urol*. 2008; 179:2212–2216. [PubMed: 18423716]
2. Vickers AJ, Bianco FJ, Gonen M, et al. Effects of pathologic stage on the learning curve for radical prostatectomy: evidence that recurrence in organ-confined cancer is largely related to inadequate surgical technique. *Eur Urol*. 2008; 53:960–966. [PubMed: 18207316]
3. Vickers AJ, Bianco FJ, Serio AM, et al. The surgical learning curve for prostate cancer control after radical prostatectomy. *J Natl Cancer Inst*. 2007; 99:1171–1177. [PubMed: 17652279]
4. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982; 247:2543–2546. [PubMed: 7069920]
5. 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. [PubMed: 8668867]
6. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med*. 2002; 137:511–520. [PubMed: 12230353]
7. Ames SC, Tan WW, Ames GE, et al. Quality of life of men with biochemical recurrence of prostate cancer. *J Psychosoc Oncol*. 2008; 26:17–34. [PubMed: 18285298]
8. Stephenson A, Kattan MW, Eastham J, et al. Defining biochemical recurrence of prostate cancer after radical prostatectomy: a proposal for a standardized definition. *J Clin Oncol*. 2006; 24:3973–3978. [PubMed: 16921049]

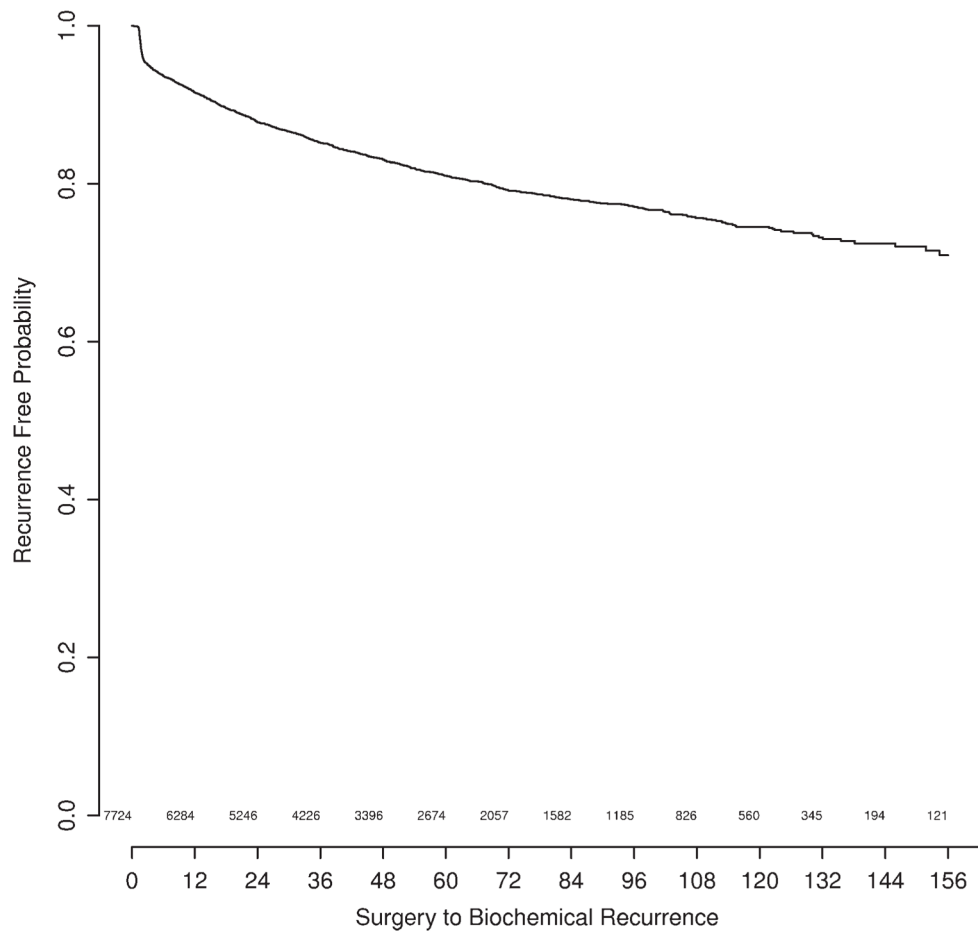


FIGURE 1.
Freedom from biochemical recurrence of prostate cancer.

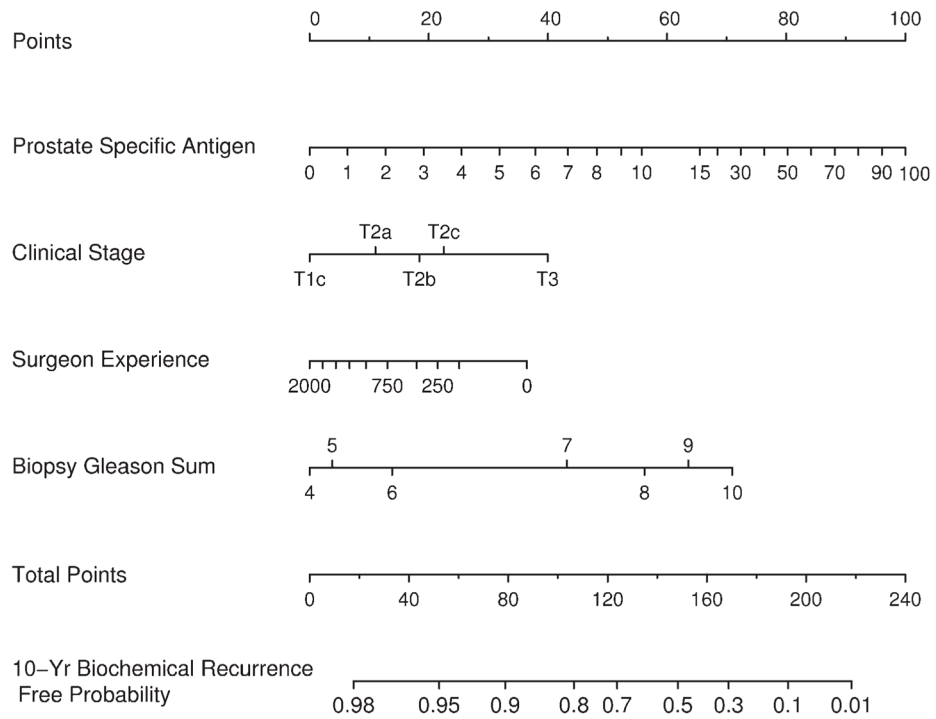


FIGURE 2. Nomogram predicting 10-year freedom from biochemical recurrence of prostate cancer based on preoperative information.

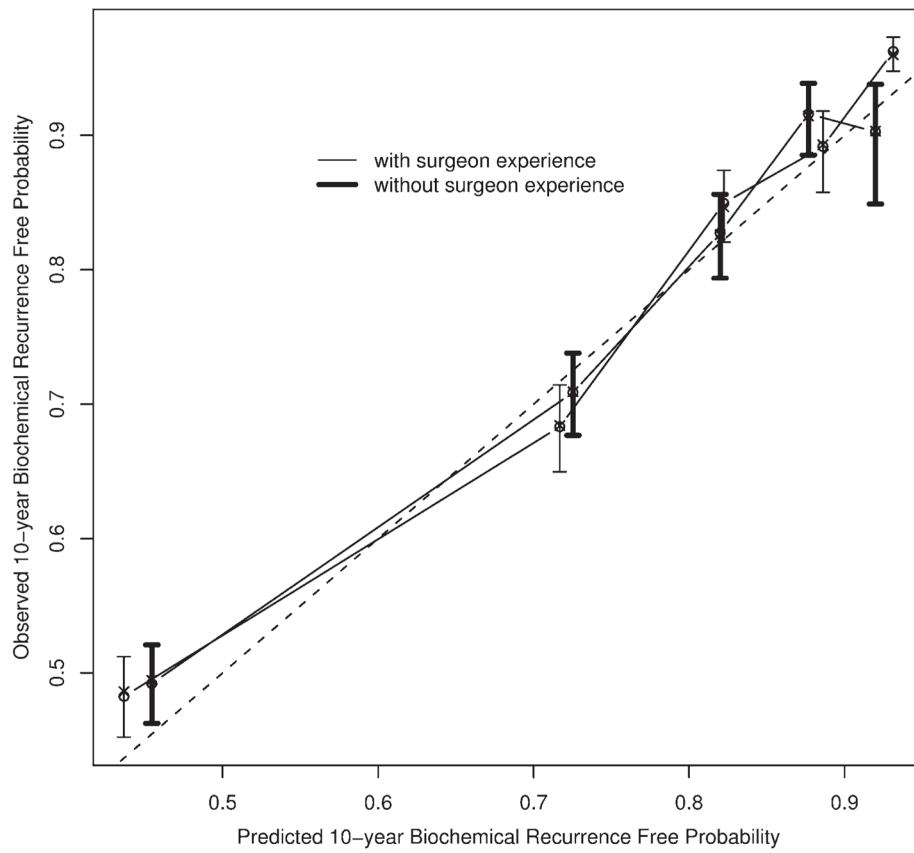


FIGURE 3.
Calibration curves for preoperative nomograms.

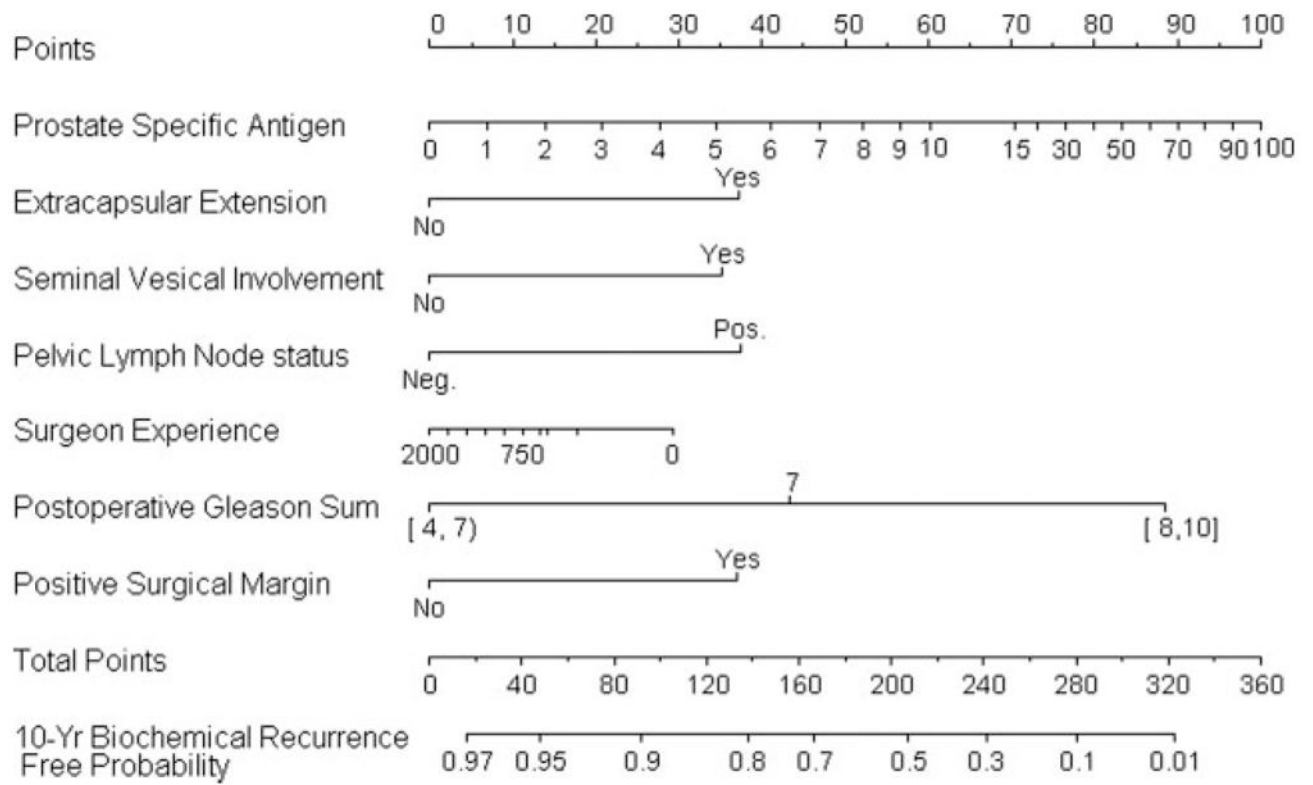


FIGURE 4. Nomogram predicting 10-year freedom from biochemical recurrence of prostate cancer based on postoperative information.

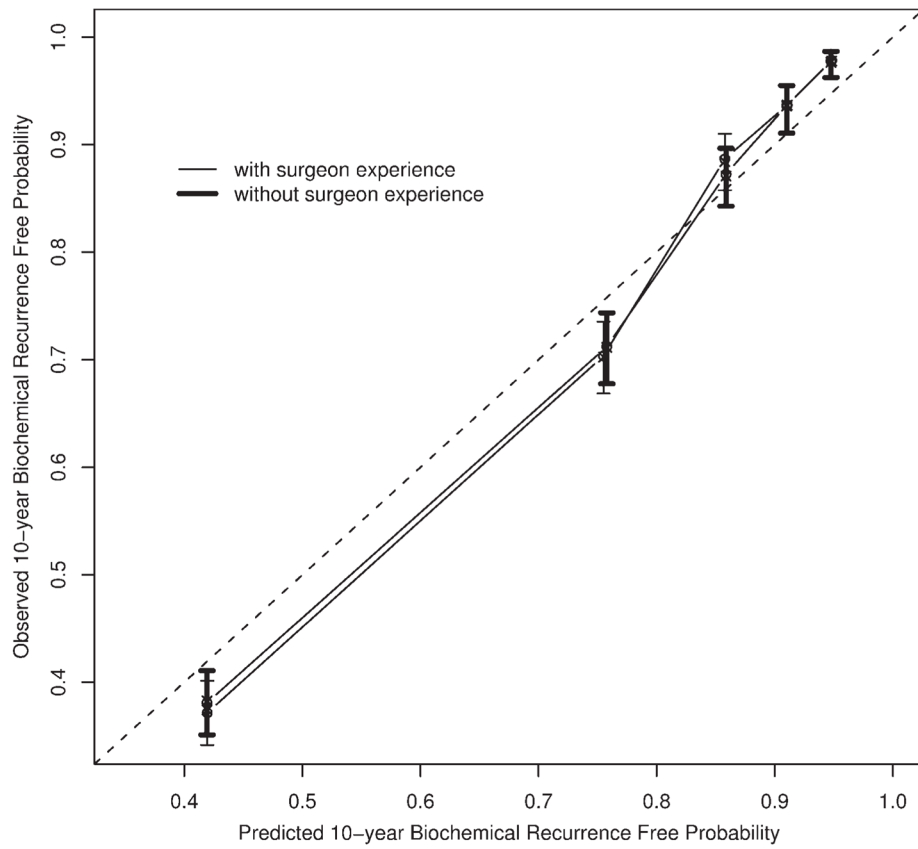


FIGURE 5.
Calibration curves for postoperative nomograms.

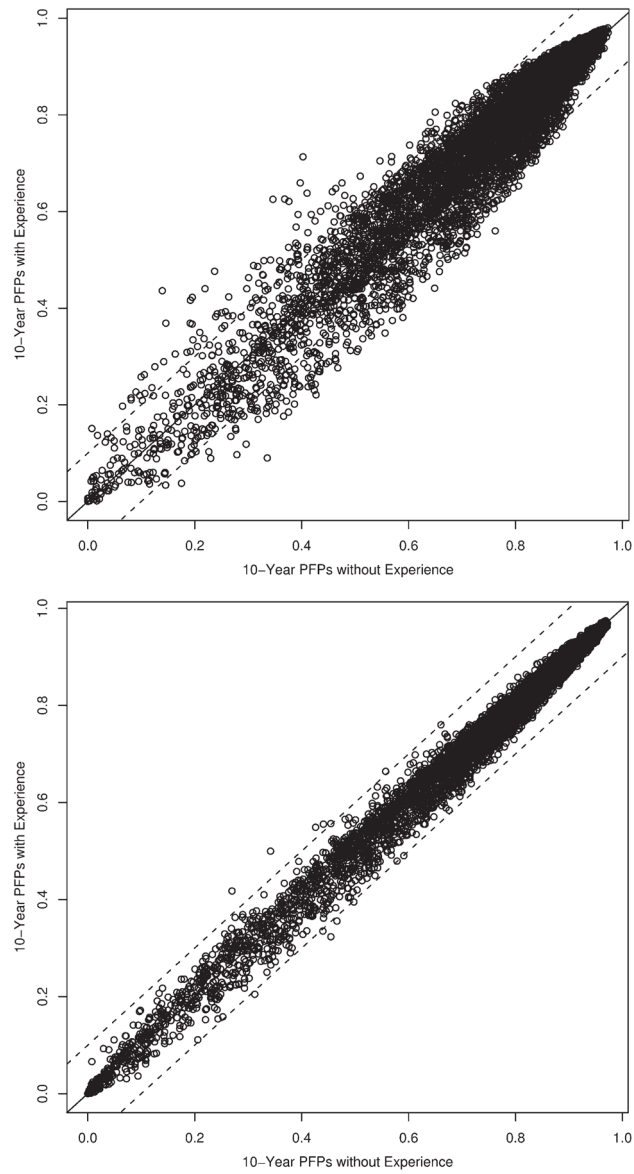


FIGURE 6. Comparison of model predicted probabilities. (*Top*) Preoperative. (*Bottom*) Postoperative. PFPs, progression-free probability.

Table 1

Descriptive Statistics

Variable	No.	%
Clinical stage		
T1c	3560	46.1
T2a	2259	29.2
T2b	819	10.6
T2c	960	12.4
T3	126	1.6
Biopsy Gleason sum		
4–6	5249	68.0
7	2043	26.5
8–10	432	5.6
Pathology Gleason sum		
4–6	3370	43.6
7	3847	49.8
8–10	507	6.6
Extracapsular extension (yes)	2359	30.5
Seminal vesicle involvement (yes)	746	9.7
Pelvic lymph node status (positive)	302	3.9
PSA (ng/mL)		
Minimum	0.5	
1 st Quartile	5.0	
Median	6.8	
Mean	9.4	
3 rd Quartile	10.1	
Maximum	228.6	
Surgery experience (N)		
Minimum	0.0	
1 st Quartile	82.8	
Median	273.0	
Mean	453.7	
3 rd Quartile	655.0	
Maximum	1978.0	
Year of surgery		
Minimum	1987	
1 st Quartile	1994	
Median	1998	
Mean	1997	
3 rd Quartile	2001	
Maximum	2003	