JJCO Japanese Journal of Clinical Oncology

Jpn J Clin Oncol 2013;43(12)1255–1260
doi:10.1093/jjco/hyt154
Advance Access Publication 21 October 2013

External Validation of Preoperative Nomograms Predicting Biochemical Recurrence after Radical Prostatectomy

Ayako Tanaka[1], Makoto Ohori[1,*], Lakin Paul[2], Changhong Yu[2], Michael W. Kattan[2], Yoshio Ohno[1] and Masaaki Tachibana[1]

[1]Department of Urology, Tokyo Medical University, Tokyo, Japan and [2]Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA

*For reprints and all correspondence: Makoto Ohori, Department of Urology, Tokyo Medical University, 6-7-1 Nishishinjuku, Shinjuku-ku, Tokyo, Japan. E-mail: mak_ohori@yahoo.co.jp

Received June 17, 2013; accepted September 16, 2013

Objective: Preoperative nomograms can accurately predict the rate of biochemical recurrence after radical prostatectomy. Although these nomograms were shown to be valid in several external validation cohorts of Caucasian patients, they have not been validated in non-Caucasian patients from Asian countries. We therefore validated these preoperative nomograms in a Japanese cohort, using different cutoff values of prostate-specific antigen concentrations for biochemical recurrence.

Methods: We analyzed 637 patients who underwent radical prostatectomy for clinically localized prostate cancer at the Tokyo Medical University Hospital between February 2000 and January 2011. We evaluated two prostate-specific antigen cutoff values for biochemical recurrence, 0.2 and 0.4 ng/ml. Using *c*-index and calibration plots, we validated the previously developed Kattan and Stephenson nomograms.

Results: Overall, the mean 5-year non-biochemical recurrence rate was 72 $\pm$ 4%. Using a prostate-specific antigen cutoff values of 0.2 and 0.4 ng/ml, the *c*-indices for the Kattan nomogram were 0.714 and 0.733. Similarly, using a prostate-specific antigen cutoff values of 0.2 and 0.4 ng/ml, the *c*-indices for the Stephenson nomograms were 0.717 and 0.671. The calibration plots showed that the predictive value of the Stephenson nomogram at a prostate-specific antigen cutoff of 0.2 ng/ml was close to the actual outcomes compared with other combinations of nomograms and prostate-specific antigen cutoff levels.

Conclusions: Because the *c*-indices of both nomograms were generally high, these nomograms can be applied to our cohort. The addition of biopsy information did not markedly improve the *c*-index but resulted in good calibration, indicating that the Stephenson nomogram may be a better fit for our patient cohort.

Key words: prostate cancer – preoperative nomograms – prostatectomy – biochemical recurrence – Japanese

INTRODUCTION

Prostate cancer is a male-specific cancer most frequently detected in western countries, and its incidence has been increasing in Japan (1–3). Prostate cancer was the ninth most frequent cause of cancer-related deaths in males, but its incidence rate was fourth and has been predicted to be second in 2020 (1–3). The gold standard of treatment in Japan, European countries, and the USA is radical prostatectomy (RP). However, approximately one-third of the patients who undergo RP, later experience prostate-specific antigen (PSA)

recurrence; therefore, it is important to identify factors predictive of recurrence.

The first nomogram for the preoperative prediction of PSA recurrence after RP (4) utilized standard preoperative factors, including clinical T stage, Gleason score of biopsy samples and serum PSA concentration. This nomogram was later updated by the addition of biopsy features, including the number of positive and negative biopsy cores (5).

Although nomograms may provide the most accurate predictions of recurrence, they must be validated in external groups of patients with characteristics different from the original dataset. Although both nomograms have been validated in patients from western countries, most of whom were Caucasians (6–11), the increase in RP in Asian countries including Japan makes the validation and/or development of nomograms important in predicting patient outcomes. To our knowledge, however, no validation study of these nomograms has been performed in non-Caucasian patients from Asian countries. We therefore sought to validate these preoperative nomograms in a Japanese cohort, using different cutoff levels of PSA to define biochemical recurrence (BCR).

PATIENTS AND METHODS

Patient Selection

The external validation population consisted of 637 patients who underwent RP for T1-3N0M0 prostate cancer at the Tokyo Medical University Hospital between February 2000 and January 2011. Patients who had neoadjuvant hormonal therapy ($n = 102$) or high-intensity focused ultrasound ($n = 1$) were excluded from the present study. Clinical T sub-stages were determined by digital rectal examination, and it was assigned according to the 2002 TNM staging system, and pathology outcomes, such as pathological stage, the status of surgical margins and Gleason score in RP specimens, were obtained from the official pathology reports.

Data Collection from Previous Nomograms

We used the nomograms on the Memorial Sloan-Kettering Cancer Center website to obtain the prediction values (12). The preoperative nomograms developed by Kattan et al. (4) and Stephenson et al. (5) were used to calculate the probability that a patient would be free from BCR 5 years after RP. Pretreatment disease variables used in the Kattan nomogram include standard factors, such as serum PSA concentration, primary and secondary Gleason grade on prostate biopsy and clinical T stage. The Stephenson nomogram included these standard factors, as well as the number of positive and negative biopsy cores.

Statistical Analysis

The time of BCR was defined as the earliest time after RP at which the postoperative serum PSA concentration rose to ≥ 0.2 or ≥ 0.4 ng/ml. The probability of remaining BCR free was calculated by the Kaplan–Meier method. Five patients were treated with adjuvant radiotherapy before the evidence of PSA recurrence after RP, whose PSA recurrence day was made into the day of operation. Factors predictive of BCR were calculated using uni- and multivariate analyses in a Cox proportional hazards model.

The validation procedure consisted of the calculation of nomogram predictions for each patient and the comparison of these results with actual outcomes, as determined by Harrell's concordance index (*c*-index) (13). Accuracy of calibration was determined by plotting predicted versus actual probabilities of freedom from BCR in groups of patients.

Statistical analyses were performed using STATA IC 11 (StataCorp., College Station, TX, USA), S-plus (Insightful Corporation, Seattle, WA, USA) or open-source R statistical software (R Development Core Team 2008) with Design package added. Statistical significance was defined as a two-sided P value of <0.05.

RESULTS

Table 1 shows the clinical and pathological characteristics of the 637 patients. At the time of RP, the median age was 66 years (range 46–81 years). More than 70% of the men had non-palpable tumors (T1c). The median preoperative PSA concentration level was 7.4 ng/ml (range 1.1–89.0 ng/ml). In the analysis of prostatectomy specimens, the rate of positive margins was high, i.e. 46%.

Using a PSA cutoff of 0.2 ng/ml, 169 (27%) patients experienced BCR after RP, after a median follow-up of 44 months (range 1–136 months). The mean 5-year non-BCR rate was $72 \pm 4\%$ (Fig. 1).

Univariate analysis showed that all preoperative factors were significantly predictive of BCR (Table 2). In multivariate analysis, serum PSA concentration ($P = 0.003$) and the number of negative biopsy cores ($P = 0.002$) remained significant, and primary Gleason grade was close to significant ($P = 0.054$). Similar findings in both uni- and multivariate analyses were observed using a PSA cutoff of 0.4 ng/ml.

In validating the Kattan preoperative nomogram, we found that the *c*-indices using 0.2 and 0.4 ng/ml as PSA cutoff values for recurrence were 0.714 and 0.733, respectively. Using a PSA concentration of 0.2 ng/ml, the Kattan nomogram markedly underestimated the actual outcomes (Fig. 2). The predictive ability of the Kattan nomogram appeared to be better when a PSA cutoff of 0.4 ng/ml was used, although there was a trend toward overestimation in the range between 0.6 and 0.8 (Fig. 3).

Similarly, we validated the Stephenson preoperative nomogram. The *c*-indices using PSA cutoff values of 0.2 and 0.4 ng/ml were 0.717 and 0.671, respectively. When a PSA cutoff of 0.2 ng/ml was used, the predictions made by the Stephenson preoperative nomogram were close to the actual outcomes (Fig. 4). At PSA 0.4 ng/ml, however, the nomogram markedly overestimated the actual outcomes (Fig. 5).

Table 1. Baseline demographic and clinical characteristics of the 637 patients

Variable	Number	%
Age (median)	46–81 (66)	
Preoperative PSA level, ng/ml (median)	1.1–89.0 (7.4)	
0.1–4.0	26	4
4.1–10.0	414	65
10.1–20.0	136	21
20.1–50.0	54	9
50.1–	7	1
Biopsy Gleason score		
2–6	200	31
7 (3 + 4)	151	24
7 (4 + 3)	130	20
8–10	156	25
Clinical T stage		
T1c	471	74
T2a	82	13
T2b	58	9
T2c	16	2
T3a	10	2
Total number of biopsy cores (median)	2–26 (15)	
Number of positive biopsy cores (median)	1–12 (7)	
Percent positive biopsy cores		
<34%	420	66
34–50%	75	12
>50%	142	22
Number of negative biopsy cores (median)	0–23 (12)	
Pathological stage		
Organ confined disease, pT2	390	61
Extraprostatic extension, pT3a	189	30
Seminal vesicle invasion, pT3b	48	7
Lymph node metastasis, pN+	10	2
Positive surgical margins	293	46

PSA, prostate-specific antigen.

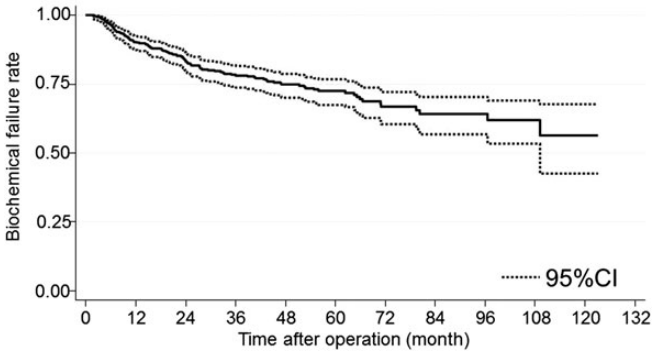

Figure 1. The Kaplan–Meier curve of biochemical recurrence (BCR), defined as prostate-specific antigen (PSA) ≥0.2 ng/ml after radical prostatectomy (RP).

DISCUSSION

The Kattan nomogram, a preoperative nomogram widely used in urologic clinics throughout the world, was first described in 1998 (4). Since its predictive value and that of the Stephenson nomogram are used to decide the optimal form of treatment for individual patients, they must first be validated to determine whether the predicted outcomes are close to the actual outcomes before their actual use. The validity of these nomograms has been confirmed in several external validation cohorts, consisting mostly of Caucasians, from the USA, Europe and Australia (6–11). To the best of our knowledge, however, there have been no validation studies in non-Caucasians from Asian countries.

Since these nomograms were accurate in Caucasian populations, with *c*-indices between 0.67 and 0.83 (6–11), many physicians considered that they were sufficiently accurate for clinical use. Using the Kattan nomogram, we observed *c*-indices of 0.714 and 0.733 as defining BCR for PSA cutoff values of 0.2 and 0.4 ng/ml, respectively, suggesting that this nomogram is reasonably accurate as a predictive model in our population.

A *c*-index of the original Kattan nomogram was high, 0.79. In the study by Kattan et al., a single expert pathologist graded the prostate biopsies and a single physician performed all digital rectal examinations and operated on all patients. In our study, however, prostate biopsies were evaluated by several pathologists and clinical T stage was decided by several urologists. An evaluation of 1701 men from the Cancer of the Prostate Strategic Urological Research Endeavor (CAPSURE) dataset reported a *c*-index of 0.68, probably because of evaluation by several specialists, not only one (6).

The differences in *c*-indices may also be due to the differences in clinical and pathological features of our patients compared with patients in previous studies. Of our patients, 74% had non-palpable cancer, compared with 15% in the study by Kattan et al. and 52% in the study by Stephenson et al. Moreover, the percentages of patients with the Gleason scores of 8 to 10 in biopsy specimens were 25, 10 and 6%, respectively. Naito et al. (14) developed the nomogram to predict pathological stage using a large series of Japanese cohort and compared it to the 2001 Partin tables (15–17). The cohorts of Naito et al. and Partin et al. differed in terms of the area under the curve in a receiver operating characteristics analysis for predicting organ-confined prostate cancer. This difference reflected not only ethnic and institutional differences between cohorts, but also the upward shift in the Gleason score because of its assessment according to the 2005 ISUP consensus (18). In the present study, we could not analyze how the modification of the 2005 ISUP affected on our cohort so that it may also be factor to have a lower *c*-index. Further, Naito et al. pointed out that the differences in the rate of positive

Table 2. Cox proportional hazard models of factors predicting biochemical failure using a PSA cutoff value of 0.2 ng/ml

Variables	Univariate		Multivariate	
	Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
Preoperative PSA, ng/ml	1.02 (1.01–1.04)	<0.001	1.01 (1.00–1.03)	0.003
Biopsy primary Gleason grade	1.55 (1.19–2.00)	0.001	1.31 (0.99–1.72)	0.054
Biopsy secondary Gleason grade	1.33 (1.04–1.70)	0.019	1.17 (0.90–1.52)	0.216
Clinical T stage	1.31 (1.11–1.55)	0.001	1.03 (0.85–1.26)	0.703
Number of positive biopsy cores	1.08 (1.01–1.16)	0.025	0.93 (0.84–1.03)	0.174
Number of negative biopsy cores	0.88 (0.84–0.93)	<0.001	0.89 (0.82–0.95)	0.002

CI, confidence interval.

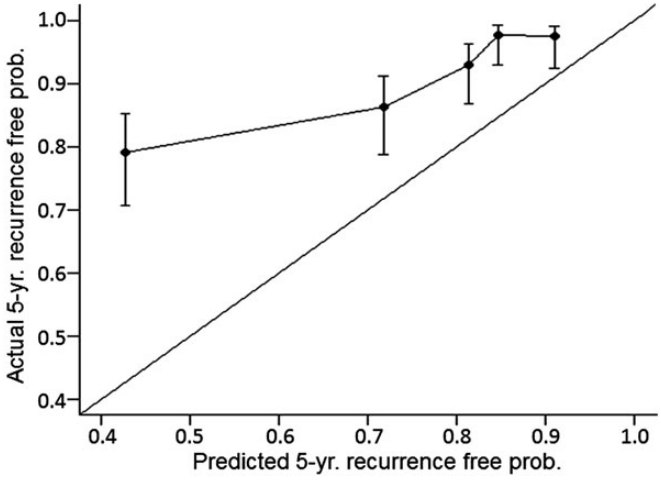

Figure 2. Calibration curve for the Kattan preoperative nomogram in our patient population, with biochemical failure defined as PSA ≥0.2 ng/ml.

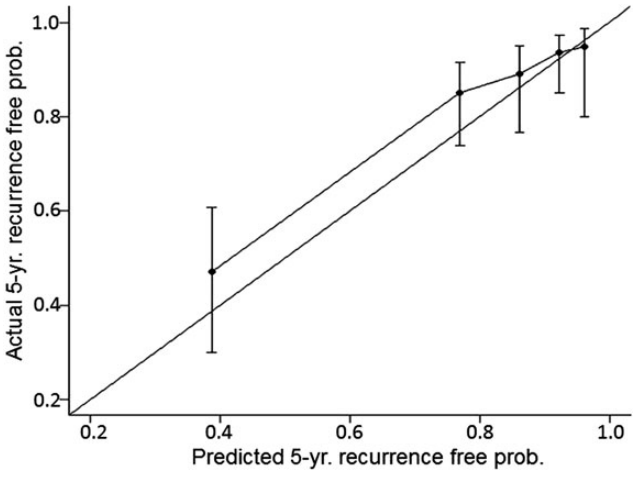

Figure 4. Calibration curve for the Stephenson preoperative nomogram in our patient population, with biochemical failure defined as PSA ≥0.2 ng/ml.

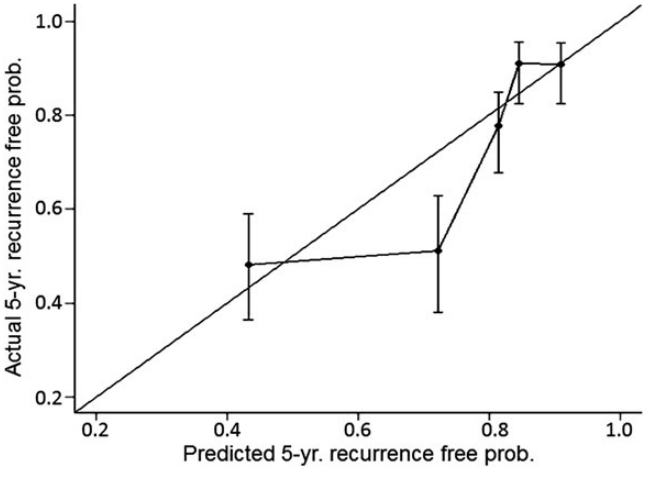

Figure 3. Calibration curve for the Kattan preoperative nomogram in our patient population, with biochemical failure defined as PSA ≥0.4 ng/ml.

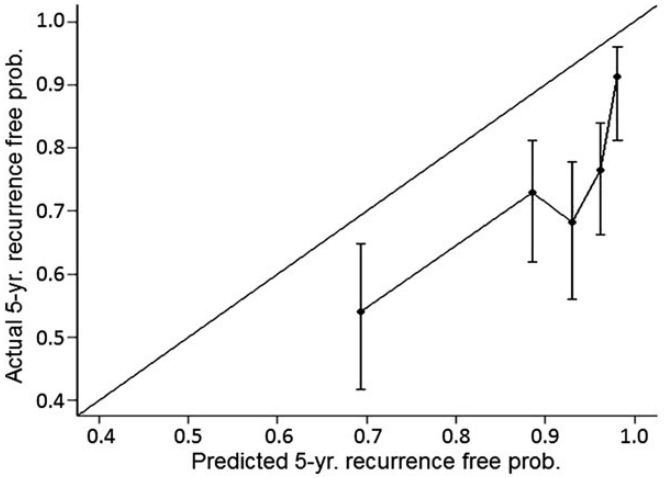

Figure 5. Calibration curve for the Stephenson preoperative nomogram in our patient population, with biochemical failure defined as PSA ≥0.4 ng/ml.

surgical margins, which may represent the level of surgical skill may result in the different outcomes. The rate of positive surgical margin in the present study was high, 46%, which was >37 and 15% in the studies by Naito et al. and Kattan et al., respectively. It is possible that the presence of positive margins may result in a higher BCR even if there are no evidence of extraprostatic extension at the sites of surgical margins (19) so that it may affect the ability to predict BCR. Also, this high rate of positive margins may make the true rate of extracapsular extension lower so that it may also obscure the ability to predict the pathological stage (14).

Since outcomes differed by endpoint, the definition of BCR could be altered by choosing a different PSA cutoff level. Because the previous nomograms used PSA of 0.4 ng/ml as a cutoff and we have used PSA of 0.2 ng/ml in daily practice, we questioned how PSA cutoff levels affect validation. We found that, although the *c*-indices that represent the descriptive ability of the models were similar, the calibration plots were markedly different. An analysis of data from more than 30 community-based urology practices using these two PSA cutoff values reported that the *c*-indices at 0.2 and 0.4 ng/ml were 0.68 and 0.71, respectively, suggesting that changing the definition of BCR did not change the *c*-index (6). That study, however, failed to show the calibrations. In contrast, we confirmed that calibrations between predicted and actual outcomes are mandatory for actual use of these nomograms in individual patients.

The descriptive ability of the predictive models may be improved by including new markers. We initially expected that the addition of biopsy information to the standard variables would significantly enhance the prediction of BCR against the previous report (20). We found the similar result that the inclusion of the number of negative or positive biopsy cores did not enhance the predictive ability of the nomogram in our cohort, although the number of negative biopsy cores was an independent predictor of BCR. However, the inclusion of biopsy information may help to make the predicted values closer to the actual outcomes (Fig. 4). We therefore felt that the Stephenson nomogram may be a better fit for our patient cohort. Nevertheless, this may or may not be true for other Japanese cohort, so that it is better for each institution to validate the previous nomograms. In the literatures, many investigators indicated that the inclusion of magnetic resonance imaging (MRI) information in addition to the standard clinical information may help to improve the prediction of BCR (21–23), while the interpretation of MRI findings is sometime subjective and is not easy. Thus, we need to make an effort to involve the imaging study such as MRI into a nomogram and to search a new marker to predict BCR, which should be readily available at clinic as well.

CONCLUSION

We found that the preoperative nomograms devised by Kattan et al. and Stephenson et al. are useful in predicting BCR in Japanese patients who were treated with RP. However, it is mandatory to calibrate the nomograms to know the tendency of under- or over-estimation before we provide the actual predictions to the patients.

Conflict of interest statement

None declared.

References

1. Parkin DM, Pisani P, Ferlay J. Global cancer statistics. *CA Cancer J Clin* 1999;49:33–64.
2. National Cancer Center. Cancer statistics in Japan'11. http://ganjoho.jp/public/statistics/backnumber/2011_en.html. (September 2013 accessed).
3. The Japanese Urological Association: Guideline of prostate cancer 2012: Kanehara Publishers Inc. 2012: 12–3. (in Japanese).
4. Kattan MW, Eastham JA, Stapleton AMF, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 1998;90:766–71.
5. Stephenson AJ, Scardino PT, Eastham JA, et al. Preoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy. *J Natl Cancer Inst* 2006;98:715–7.
6. Greene KL, Meng MV, Elkin EP, et al. Validation of the Kattan preoperative nomogram for prostate cancer recurrence using a community based cohort: results from cancer of the prostate strategic urological research endeavor (CAPSURE). *J Urol* 2004;171:2255–9.
7. Isbarn H, Karakiewicz PI, Walz J, et al. External validation of a preoperative nomogram for prediction of the risk of recurrence after radical prostatectomy. *Int J Radiat Oncol Biol Phys* 2010;77:788–92.
8. Thanigasalam R, Rasiah KK, Stricker PD, et al. Stage migration in localized prostate cancer has no effect on the post-radical prostatectomy Kattan nomogram. *BJU Int* 2009;105:642–7.
9. Kattan MW, Vickers AJ, Yu C, et al. Preoperative and postoperative nomogram incorporating surgeon experience for clinically localized prostate cancer. *Cancer* 2009;115:1005–10.
10. Roupret M, Hupertan V, Comperat E, et al. Cross-cultural validation of prognostic tool: example of the Kattan preoperative nomogram as a predictor of prostate cancer recurrence after radical prostatectomy. *BJU Int* 2009;104:813–18.
11. Graefen M, Karakiewicz PI, Cagiannos I, et al. International validation of a preoperative nomogram for prostate cancer recurrence after radical prostatectomy. *J Clin Oncol* 2002;20:3206–12.
12. Memorial Sloan-Kettering Cancer Center. http://www.mskcc.org/cancer-care/adult/prostate/prediction-tools (September 2013 accessed)
13. Harrell FE, Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982;247:2543–6.
14. Naito S, Kuroiwa K, Kinukawa N, et al. Validation of Partin tables and development of a preoperative nomogram for Japanese patients with clinically localized prostate cancer using 2005 International Society of Urological Pathology consensus on Gleason grading: data from the Clinicopathological Research Group for Localized Prostate Cancer. *J Urol* 2008;180:904–9.
15. 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;58:843–8.
16. Partin AW, Yoo J, Carter HB, et al. The use of prostate specific antigen, clinical stage and Gleason score to predict pathological stage in men with localized prostate cancer. *J Urol* 1993;150:110–4.
17. Partin AW, Kattan MW, Subong EN, et al. Combination of prostate-specific antigen, clinical stage, and Gleason score to predict pathological stage of localized prostate cancer. A multi-institutional update. *JAMA* 1997;277:1445–51.
18. Epstein JI, Allsbrook WC, Jr, Amin MB, et al. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason grading of prostatic carcinoma. *Am J Surg Pathol* 2005;29:1228–42.

19. Shuford MD, Cookson MS, Chang SS, et al. Adverse prognostic significance of capsular incision with radical retropubic prostatectomy. *J Urol* 2004;172:119–23.
20. Graefen M, Ohori M, Karakiewicz PI, et al. Assessment of the enhancement in predictive accuracy provided by systematic biopsy in predicting outcome for clinically localized prostate cancer. *J Urol* 2004;171:200–3.
21. D'Amico AV, Whittington R, Malkowicz SB, et al. Combination of the preoperative PSA level, biopsy Gleason score, percentage of positive biopsies, and MRI T-stage to predict early PSA failure in men with clinically localized prostate cancer. *Urology* 2000;55: 572–77.
22. Fuchsjäger MH, Shukla-Dave A, Hricak H, et al. Magnetic resonance imaging in the prediction of biochemical recurrence of prostate cancer after radical prostatectomy. *BJU Int* 2009;104:315–20.
23. Muglia VF, Westphalen AC, Wang ZJ, Kurhanewicz J, Carroll PR, Coakley FV. Endorectal MRI of prostate cancer: incremental prognostic importance of gross locally advanced disease. *AJR* 2011;197:1369–74.