Chinese Journal of Cancer

Original Article

External validation of nomograms for predicting cancer-specific mortality in penile cancer patients treated with definitive surgery

Yao Zhu[1,2,*], Wei-Jie Gu[1,2,*], Ding-Wei Ye[1,2], Xu-Dong Yao[1,2], Shi-Lin Zhang[1,2], Bo Dai[1,2], Hai-Liang Zhang[1,2] and Yi-Jun Shen[1,2]

Abstract

Using a population-based cancer registry, Thuret *et al*. developed 3 nomograms for estimating cancer-specific mortality in men with penile squamous cell carcinoma. In the initial cohort, only 23.0% of the patients were treated with inguinal lymphadenectomy and had pN stage. To generalize the prediction models in clinical practice, we evaluated the performance of the 3 nomograms in a series of penile cancer patients who were treated with definitive surgery. Clinicopathologic information was obtained from 160 M0 penile cancer patients who underwent primary tumor excision and regional lymphadenectomy between 1990 and 2008. The predicted probabilities of cancer-specific mortality were calculated from 3 nomograms that were based on different disease stage definitions and tumor grade. Discrimination, calibration, and clinical usefulness were assessed to compare model performance. The discrimination ability was similar in nomograms using the TNM classification or American Joint Committee on Cancer staging (Harrell's concordance index = 0.817 and 0.832, respectively), whereas it was inferior for the Surveillance, Epidemiology and End Results staging (Harrell's concordance index = 0.728). Better agreement with the observed cancer-specific mortality was shown for the model consisting of TNM classification and tumor grade, which also achieved favorable clinical net benefit, with a threshold probability in the range of 0 to 42%. The nomogram consisting of TNM classification and tumor grading was shown to have better performance for predicting cancer-specific mortality in penile cancer patients who underwent definitive surgery. Our data support the integration of this model in decision-making and trial design.

Key words Penile neoplasms, neoplasm staging, prognosis, mortality

Over the last decade, prediction models have played an increasingly important role in individualized medicine[1]. Numerous risk stratification scores and nomograms have been developed for the optimal management of cancer. Despite noteworthy advances in other genitourinary malignancies, prognostic tools other than the TNM staging system are quite limited for penile cancer. Using Surveillance, Epidemiology and End Results (SEER) registries, Thuret *et al*.[2] developed 3 nomograms for the prediction of cancer-specific mortality in patients with penile squamous cell carcinoma (**Figure 1**). These promising tools, however, have not been externally validated in an independent case series. Due to distinctive disease characteristics[3], the validity of these nomograms is especially questionable in men from developing countries. Furthermore, only 23% of the patients in Thuret's study underwent inguinal lymphadenectomy and had pN stage[2]. In contrast, recent guidelines from the European Association of Urology and International Consultation on Urological Diseases suggest accurate lymph node (LN) staging in penile cancer, except for low-risk disease (Tis/Ta/T1a)[4,5]. Therefore, assessment of the performance of the nomograms in patients treated with definitive surgery may aid in generalizing the tools for a clinical setting. To accomplish these purposes, we validated Thuret's nomograms in a Chinese series of M0 penile cancer patients who were treated with primary tumor excision and regional lymphadenectomy. The predictive value of the nomograms was evaluated in terms of

Authors' Affiliations: [1]Department of Urology, Fudan University Shanghai Cancer Center, Shanghai 200032, P. R. China; [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, P. R. China.

Corresponding Author: Ding-Wei Ye, Department of Urology, Fudan University Shanghai Cancer Center, Shanghai 200032, P. R. China. Email: dwyeli@163.com.

*Equal contribution

doi: 10.5732/cjc.013.10176

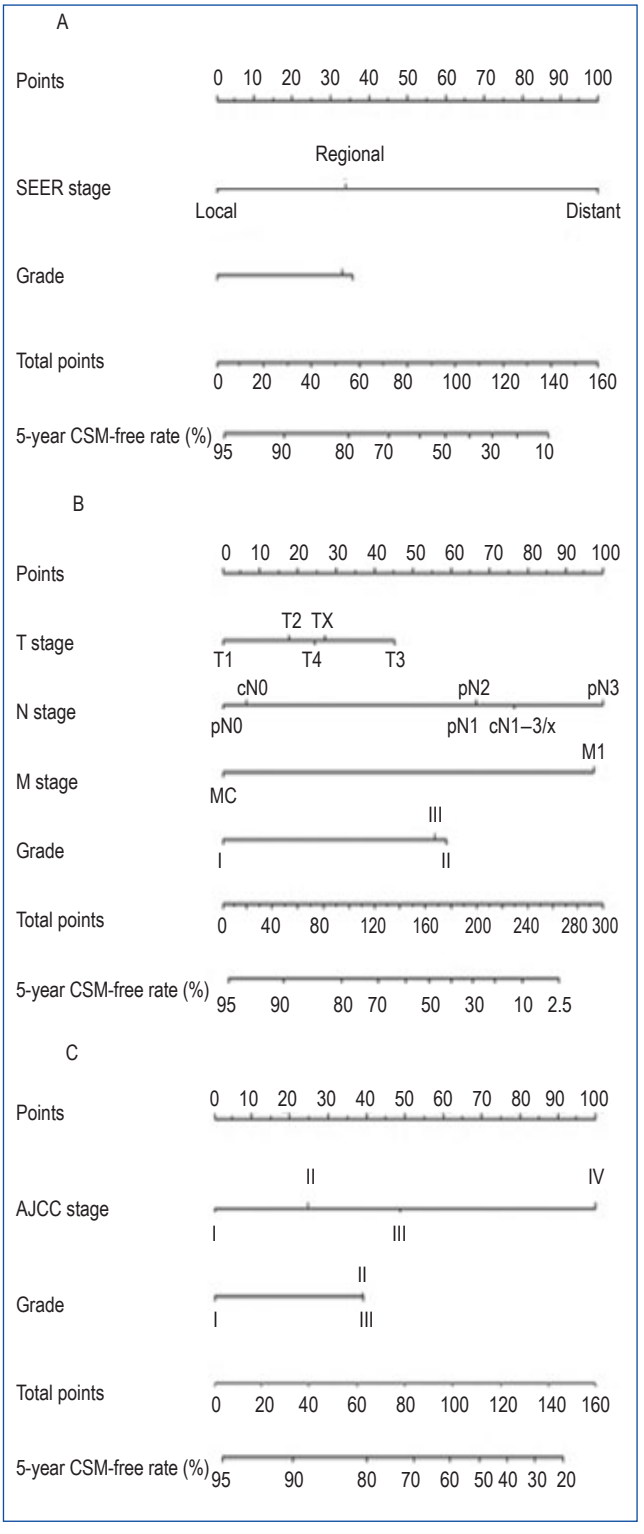

Figure 1. **Nomograms predicting the cancer-specific mortality (CSM)-free rate 5 years after primary tumor excision using Surveillance, Epidemiology and End Result (SEER) staging (A), TNM classification (B), and American Joint Committee on Cancer (AJCC) staging (C) combined with tumor grade (TG).**

discrimination, calibration, and clinical usefulness.

Materials and Methods

Study population

After searching the penile cancer database from Fudan University Shanghai Cancer Center, we identified M0 patients who underwent primary tumor excision and regional lymphadenectomy between 1990 and 2008. All patients had the pathologic diagnosis of penile squamous cell carcinoma, and patients who underwent neoadjuvant chemotherapy or previous groin exploration were excluded. Institutional Review Board approval was obtained.

Once the diagnosis of penile squamous cell carcinoma was confirmed, the patients underwent a clinical staging workup that included physical examination, ilioinguinal CT scan, abdominal ultrasound scan, and chest X-ray. The primary tumor was managed using local excision or partial or total penectomy according to the depth of invasion, size, and patient preference. In our institution, standard, bilateral, radical, inguinal lymphadenectomy was performed on a regular basis, except for Tis/Ta disease[6]. Inguinal LN metastases that were confirmed using biopsy were concurrently resected with the penile lesions, and prophylactic dissection was performed 2 or 6 weeks after removal of the primary disease. Before 2005, the indication of pelvic lymphadenectomy was enlarged pelvic LNs on preoperative cross-sectional images or the involvement of the Cloquet's node in frozen sections. Due to the low negative predictive value of the indication[7], pelvic lymphadenectomy was performed when 1 or more positive inguinal LNs had been found since then. Patients pass the N2 stage received adjuvant chemotherapy or radiotherapy.

External validation of 3 nomograms

Medical records were reviewed to obtain the detailed clinicopathologic data that was needed for the nomograms. To apply Thuret's nomograms in our series, the TNM stage and the American Joint Committee on Cancer (AJCC) stage of the tumors were assigned according the 2002 edition[8]. Tumor grade was classified using the 3-grade Broders scale[9], and the cancer-specific mortality-free survival was defined as the interval between the surgery date and cancer-related death or last follow-up date for censored patients.

As in the initial report, the 5-year cancer-specific mortality-free rate was used for comparisons[2]. Predicted probabilities were calculated from the 3 nomograms, and all 3 nomograms included disease staging and tumor grading. The SEER staging, AJCC staging, and TNM staging were used for disease staging. Therefore, the 3 nomograms were designated as SEER, AJCC, and TNM nomograms in this study.

Statistical analysis

Cancer-specific mortality was estimated using the Kaplan-Meier method. The prognostic accuracy of the nomograms was quantified

using Harrell's concordance index (C-index)[10]. A completely random prediction had a C-index of 0.5, and a perfect rule had a C-index of 1.0. The 95% confidence intervals (CIs) of the C-indexes were calculated using bootstrapping, and 2,000 bootstrap samples, each involving a resampling of the entire dataset of patients with replacement, were assessed. The 95% CIs of pairwise differences between the C-indexes of the prognostic models were similarly estimated. A calibration plot using the val.surv method was used to graphically assess the agreement between the predicted probabilities and observed outcomes. For a prediction model with good calibration, the curve virtually followed a 45-degree slope. Because postoperative mortality risk might influence the decision regarding adjuvant therapies, we performed decision curve analysis to determine the clinical usefulness of the prediction models[11]. The net benefits of the analysis estimates were calculated by summing the benefits and subtracting the weighted harms. A decision curve should be interpreted, and the model with the highest net benefit at a particular threshold probability should be chosen.

For all analyses, a 2-sided P value < 0.05 was considered significant. Statistical analyses were performed using R2.13.0[12].

Results

Patient characteristics

We identified 160 M0 penile cancer patients who were treated with primary tumor excision and regional lymphadenectomy between 1990 and 2008. **Table 1** shows the baseline characteristics of the 160 patients and the SEER registries. The median age of the patients in the validation dataset was 53 years, which was remarkably lower than that of Thuret's cohort. Compared with the development dataset, T2 lesions and G1 tumors were commonly found in our center.

Local excision, partial penectomy, and radical penectomy were performed in 18.1%, 61.3%, and 20.6% of patients, respectively. Thirty-four (21.3%) patients underwent pelvic LN dissection after removal of the inguinal LNs. After a median follow-up of 43 months (range, 6 to 180 months), 32 (20%) patients had cancer-specific events (**Figure 2**). The 5-year cancer-specific mortality-free rate was 78.0% (95% CI = 71.3% to 85.4%) in our group.

The discrimination, calibration, and clinical usefulness of the 3 nomograms

The discrimination ability of the 3 nomograms for cancer-specific mortality, the pairwise differences, and the 95% CI that was calculated using bootstrapping are reported in **Table 2**. For all 3 models, the C-index of the SEER nomogram was the lowest, with significant differences for the pairwise comparisons. No substantial difference in C-index (95% CI = –0.042 to 0.069) between the TNM and AJCC nomograms was evident.

The calibration plot shows that the SEER and AJCC nomograms were likely to overestimate the cancer-specific mortality in a wide range of risks (**Figure 3**). Better agreement with the observed cancer-specific mortality was achieved using the TNM nomogram, especially when the predicted probability was > 40%. In our series, the estimated cancer-specific mortality using the SEER nomogram fell in a small range (0 to 28.5%), and this nomogram showed a narrow predication range in this cohort, which was beyond expectation, even after considering that all metastatic cases that had surgery were excluded from the study.

Table 1. Baseline patient and disease characteristics in our and Thuret's cohorts

Variate	Our cohort (n=160)	Thuret's cohort (n=1,324)	Variate	Our cohort (n=160)	Thuret's cohort (n=1,324)
Age (years)			M category [cases (%)]		
Median	53	68	M0	160 (100)	1,273 (96.1)
Range	20–84	22–102	M1	0 (0)	51 (3.9)
T category [cases (%)]			Grade [cases (%)]		
T1	70 (43.8)	763 (57.6)	G1	83 (51.9)	410 (31.0)
T2	69 (43.1)	334 (25.2)	G2	61 (38.1)	606 (45.8)
T3	17 (10.6)	163 (12.3)	G3	16 (10.0)	308 (23.3)
T4	4 (2.5)	28 (2.1)	SEER stage [cases (%)]		
TX	–	36 (2.7)	Localized	100 (62.5)	729 (55.1)
N category [cases (%)]			Regional	60 (37.5)	515 (38.9)
pN0	100 (62.5)	127 (9.6)	Metastatic	0 (0)	80 (6.0)
pN1	24 (15.0)	58 (4.4)	AJCC stage [cases (%)]		
pN2	24 (15.0)	62 (4.7)	I	49 (30.6)	697 (52.6)
pN3	12 (7.5)	57 (4.3)	II	56 (35.0)	301 (22.7)
cN0	–	948 (71.6)	III	39 (24.4)	189 (14.3)
cN1–3/X	–	72 (5.4)	IV	16 (10.0)	137 (10.3)

pN, pathologic N stage; cN, clinical N stage; SEER, Surveillance, Epidemiology and End Result; AJCC, American Joint Committee on Cancer.

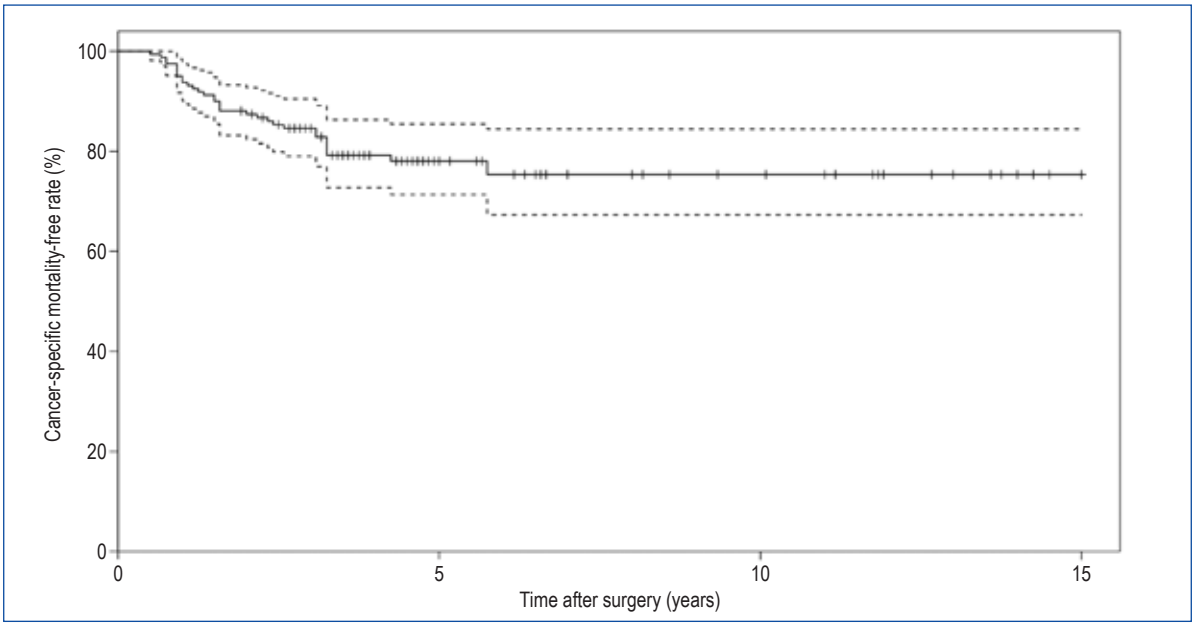

Figure 2. **Kaplan-Meier survival curves of the cancer-specific mortality-free rates in our cohort. The solid curve refers to the cancer-specific mortality-free rates whose 95% confidence intervals are indicated by the dashed lines.**

Table 2. Comparisons of nomogram discrimination

Variable	Cancer-specific mortality	
	C-index	95% confidence interval
Nomogram		
SEER nomogram	0.728	0.645–0.811
TNM nomogram	0.817	0.750–0.878
AJCC nomogram	0.832	0.766–0.892
Differences between nomograms		
TNM nomogram vs. SEER nomogram	0.089	0.052–0.125
AJCC nomogram vs. SEER nomogram	0.104	0.036–0.171
AJCC nomogram vs. TNM nomogram	0.015	–0.042–0.069

Figure 4 illustrates the decision curves of the 3 nomograms. The TNM nomogram showed favorable net benefits with a range of threshold probabilities from 0 to 42%. With a high threshold rate, the AJCC nomogram demonstrated better net benefit.

Discussion

In the current study, we externally validated 3 nomograms for predicting cancer-specific mortality in 160 M0 penile cancer patients who were treated with definitive surgery. The TNM and AJCC nomograms showed good discrimination ability without substantial difference. A better agreement with observed cancer-specific mortality was seen for the model consisting of TNM classification and tumor grading, especially when the predicted probability was > 40%. Because the TNM nomogram has 3 elements in prognostication (T, N, M), it has more risk classifications, especially in high-risk patients (e.g., N+, M+). The TNM nomogram also achieved a favorable net benefit within the threshold probability range of 0 to 42%. Although the AJCC nomogram demonstrated a better net benefit in threshold probability of over 42%, it was likely to underestimate risk. If clinicians use a high threshold (>42%), the weight of overtreatment would be higher than missing high-risk disease. However, in real practice, we rarely use such a high threshold because the outcome of missing high-risk disease is more pronounced in penile cancer.

As a rare disease, few prediction models were developed for estimating the survival outcome of penile cancer. The first nomogram to predict the 5-year cancer-specific mortality-free rate was developed by Kattan *et al*.[13] and had a C-index of 0.747 in the initial

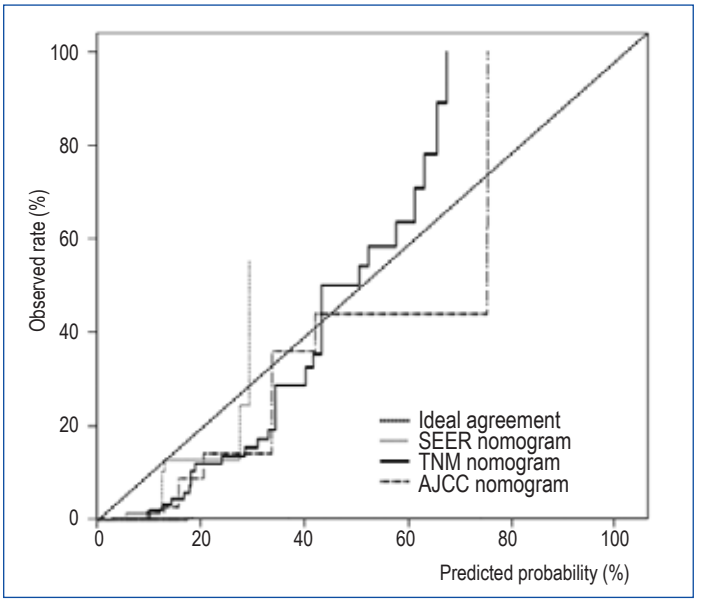

Figure 3. **Calibration of the predicted (*X*-axis) and observed (*Y*-axis) 5-year cancer-specific mortality for the SEER, TNM, and AJCC nomograms.**

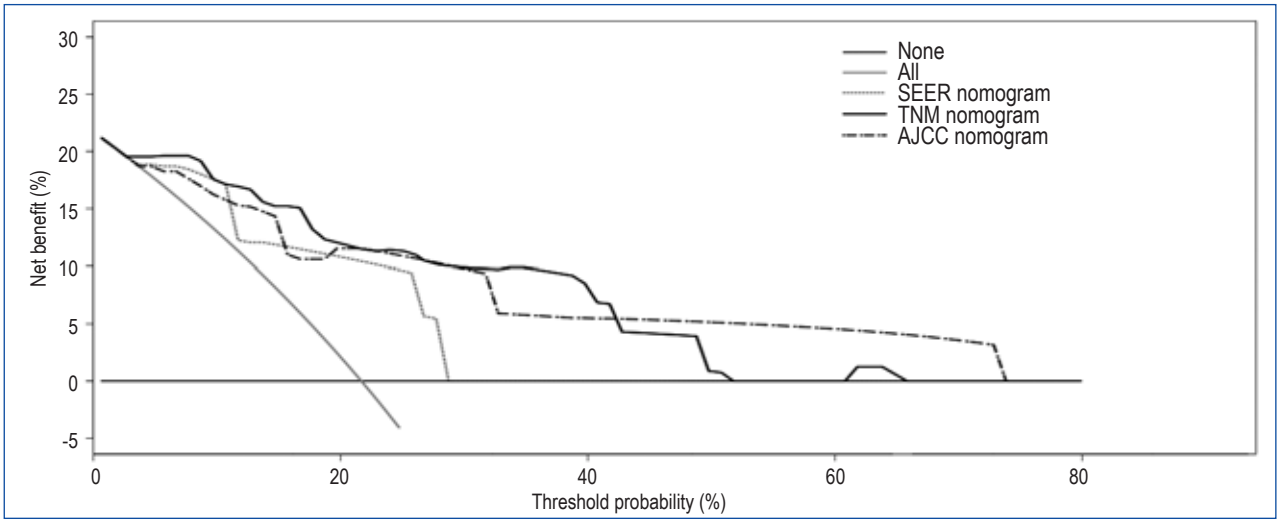

Figure 4. **Decision curves for the predicted probabilities of the SEER, TNM, and AJCC nomograms.**

cohort. Because 7 variables of the model were pathologic features of the primary disease, assessing the model's performance was difficult in centers that applied routine pathology. Furthermore, 58.9% of all 175 patients had not undergone pathologic examination of the regional LNs and were classified as pNx in the analysis. Although it is statistically sound that pNx is the strongest predictor of adverse outcome in their nomogram, doctors may be confused when using the postoperative prediction model in the real world. Using SEER cancer registries, Zini *et al*.[14] constructed a simplified nomogram that achieved a similar C-index (0.738) using only 2 predictors (SEER stage and tumor grade). Until recently, a substantial increase in the discrimination ability was reported for nomograms built by Thuret *et al*.[2]. The AJCC and TNM nomograms achieved C-index values of more than 0.8 in the original report. However, it should be noted that only 23% of the enrolled men underwent inguinal lymphadenectomy, although 39.7% had T2–4 lesion, and 69.0% had a G2–3 tumor. Lack of pN stage may compromise the prognostic value of these nomograms because the literature has clearly shown that clinical examination is inaccurate for nodal staging[4,5]. Secondly, omitting lymphadenectomy in high-risk patients had a negative impact on patient survival[15]. Thus, caution should be taken when using the nomograms in contemporary series that are treated according to the guidelines.

Using a patient population that was treated with definitive surgery, our study overcame the above-mentioned drawbacks. The validation results clearly demonstrated the added benefit of pN stage as a predictor of cancer-specific mortality. The TNM and AJCC nomograms had C-index values better than those of the original report (increase

of C-index = 0.01 and 0.023, respectively). Furthermore, comparison of the discrimination ability of the TNM and AJCC nomograms in 2,000 resamples did not reveal a substantial difference. For all 3 nomograms, the calibration plot shows overestimation of cancer-specific mortality with predicted risk in a range from 0 to 30%. Better survival of our patients was most likely attributed to accurate nodal staging and the therapeutic effect of LN dissection. In our series, better calibration was observed for the predicted probabilities that were calculated using the TNM nomogram. Decision curve analysis plays an important role in assessing a model's clinical usefulness[11]. Although nomograms generate continuous predictive probabilities, a cutoff value is usually needed when making treatment decisions, and threshold probabilities should be chosen to select patients for adjuvant therapies. Head-to-head comparisons of the 3 nomograms illustrated a superior net benefit of the TNM nomogram within the threshold probabilities of 0 to 42%. Compared with the AJCC nomogram, the risk range that favored the TNM nomogram is commonly used in clinical practice.

Accurate estimation of cancer-specific mortality may aid in the selection of candidates for adjuvant therapies to reduce the relapse rate and help to design the follow-up schedule. Pizzocaro *et al*.[16] reported a relapse rate of 45% in 31 patients who were only treated surgically versus 16% in 25 patients who were submitted to adjuvant chemotherapy.

Compared with the original series, our study cohort showed significant differences in clinicopathologic parameters. The median age of our patients was much lower than that of the SEER registries. Unsurprisingly, young patients are more likely to transfer to a tertiary cancer center and recognize the survival benefit of extensive surgery over the complications. Furthermore, the age distribution of our patients might suggest a different etiology for penile cancer in China. Yanagawa *et al*.[17] found that only 11.5% of penile cancers were human papillomavirus (HPV) DNA-positive in a small Japanese series. The mechanism of pathogenesis in Chinese penile cancer patients is currently under investigation in our institution. Our study also included a higher proportion of T2 disease and G1 tumors. Because pathologic review was previously performed for these patients[18], variability caused by multiple investigators was minimized.

The current study is not devoid of limitations. As a retrospective study, the results may be influenced by the heterogeneity of patients and tumors. Although ideal validation should be obtained in a prospective setting, this is hard to perform for a rare disease such as penile cancer. Secondly, the study population was relatively small. However, centralized pathologic evaluation, standardized management, and long median follow-up strengthened our conclusions. Furthermore, the validation results can be especially helpful for patients who are managed under contemporary standard care, which advocates accurate nodal staging in penile cancer.

Conclusions

To our knowledge, we provide the first external validation of the Thuret nomograms in an East Asian population. The nomogram consisting of TNM classification and tumor grading was shown to have better performance for predicting cancer-specific mortality in penile cancer patients who underwent definitive surgery. Our data support the integration of this model in decision-making and trial design.

Received: 2013-10-06; revised: 2013-12-31;
accepted: 2014-02-10.

References

[1] Vickers AJ. Prediction models in cancer care. CA Cancer J Clin, 2011.
[2] Thuret R, Sun M, Abdollah F, et al. Tumor grade improves the prognostic ability of American Joint Committee on Cancer stage in patients with penile carcinoma. J Urol, 2011,185:501–507.
[3] Misra S, Chaturvedi A, Misra NC. Penile carcinoma: a challenge for the developing world. Lancet Oncol, 2004,5:240–247.
[4] Pizzocaro G, Algaba F, Horenblas S, et al. EAU penile cancer guidelines 2009. Eur Urol, 2010,57:1002–1012.
[5] Heyns CF, Fleshner N, Sangar V, et al. Management of the lymph nodes in penile cancer. Urology, 2010,76:S43–57.
[6] Horenblas S. Lymphadenectomy for squamous cell carcinoma of the penis. Part 2: the role and technique of lymph node dissection. BJU Int, 2001,88:473–483.
[7] Zhu Y, Zhang SL, Ye DW, et al. Predicting pelvic lymph node metastases in penile cancer patients: a comparison of computed tomography, Cloquet´s node, and disease burden of inguinal lymph nodes. Onkologie, 2008,31:37–41.
[8] Greene FL, Page DL, Fleming ID, et al. AJCC cancer staging handbook: from the AJCC Cancer Staging Manual. New York: Springer-Verlag, 2002.
[9] Broders A. Practical points on the microscopic grading of carcinoma. NY State J Med, 1932,32:71.
[10] Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag, 2001.
[11] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making, 2006,26:565–574.
[12] R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN, 2011:3–900051.
[13] Kattan MW, Ficarra V, Artibani W, et al. Nomogram predictive of cancer-specific survival in patients undergoing partial or total amputation for squamous cell carcinoma of the penis. J Urol, 2006,175:2103–2108.
[14] Zini L, Cloutier V, Isbarn H, et al. A simple and accurate model for prediction of cancer-specific mortality in patients treated with surgery for primary penile squamous cell carcinoma. Clin Cancer Res, 2009,15:1013–1018.
[15] Theodorescu D, Russo P, Zhang ZF, et al. Outcomes of initial surveillance of invasive squamous cell carcinoma of the penis and

negative nodes. J Urol, 1996,155:1626–1631.
[16] Pizzocaro G, Piva L, Nicolai N. Treatment of lymphatic metastasis of squamous cell carcinoma of the penis: experience at the National Tumor Institute of Milan. Arch Ital Urol Androl, 1996,68:169–172. [in Italian]
[17] Yanagawa N, Osakabe M, Hayashi M, et al. Detection of HPV-DNA, p53 alterations, and methylation in penile squamous cell carcinoma in Japanese men. Pathol Int, 2008,58:477–482.
[18] Zhu Y, Ye DW, Yao XD, et al. New N staging system of penile cancer provides a better reflection of prognosis. J Urol, 2011, 186:518–523.