Bladder Cancer

Precystectomy Nomogram for Prediction of Advanced Bladder Cancer Stage

Pierre I. Karakiewicz\textsuperscript{a,*}, Shahrokh F. Shariat\textsuperscript{b}, Ganesh S. Palapattu\textsuperscript{c}, Paul Perrotte\textsuperscript{a}, Yair Lotan\textsuperscript{b}, Craig G. Rogers\textsuperscript{5}, Gilad E. Amiel\textsuperscript{d}, Amnon Vazin\textsuperscript{d}, Amit Gupta\textsuperscript{b}, Patrick J. Bastian\textsuperscript{c}, Arthur I. Sagalowsky\textsuperscript{b}, Mark Schoenberg\textsuperscript{c}, Seth P. Lerner\textsuperscript{d}

\textsuperscript{a}Cancer Prognostics and Health Outcomes Unit, University of Montreal, Montreal, Quebec, Canada
\textsuperscript{b}Department of Urology, University of Texas Southwestern Medical Center, Dallas, Texas, USA
\textsuperscript{c}The James Buchanan Brady Urological Institute, The Johns Hopkins Hospital, Baltimore, Maryland, USA
\textsuperscript{d}The Scott Department of Urology, Baylor College of Medicine, Houston, Texas, USA

1. Introduction

Accurate stage assignment is critical in treating bladder cancer. In most instances, nonmuscle-invasive transitional cell carcinoma is amenable to transurethral resection (TUR) and/or intravesical therapy. For most patients with localized, muscle-invasive bladder cancer, however, radical cystectomy represents the mainstay treatment modality. Unfortunately, pathologic stage is frequently under-
estimated with TUR variables. In one series, 30% of patients with clinical stage T1 disease were understaged at TUR [1]. When muscularis propria was not included in the specimen, the rate of understaging increased to 60% [1]. Clinical stage assignment was also shown to vary when agreement was assessed between local versus central pathology review [2]; central pathology down-staged 53% of specimens from T1 to Ta. The rate of clinical understaging occurs in 27% to 62% of cases [3–6].

These findings suggest that precystectomy pathologic stage predictions are not perfectly accurate and that predictions based on TUR T stage could be improved. Based on these considerations, we investigated whether multivariate models could generate more accurate stage predictions than any TUR staging variable in isolation. To test this, we developed two multivariate nomograms with the intent of accurately predicting advanced pathologic T stage and/or presence of nodal metastases at cystectomy.

### 2. Material and methods

#### 2.1. Patient population

All studies were undertaken with the approval and oversight of the Institutional Review Board for the Protection of Human Subjects at each institution. A total of 958 consecutive patients who underwent radical cystectomy and pelvic lymphadenectomy with curative intent to treat their bladder cancer by select surgeons at The Department of Urology, University of Texas Southwestern, Dallas, Texas; The James Buchanan Brady Urological Institute, The Johns Hopkins Hospital, Baltimore, Maryland; and The Scott Department of Urology, Baylor College of Medicine, Houston, Texas, between 11 March 1984 and 24 January 2003 and who had data available were potential candidates for this analysis. For each patient, comprehensive clinical and pathologic information were collected and entered into an institutional review board-approved database. Multiple internal and external data reviews and quality checks were performed to assure the accuracy and completeness of data elements. The indications for radical cystectomy were tumour invasion into the muscularis propria or prostatic stroma or Ta, T1, or carcinoma in situ (CIS) refractory to TUR with intravesical chemotherapy and/or immunotherapy. No patient had distant metastatic disease at the time of cystectomy. TUR stage was assigned by the operative surgeon according to the 2002 TNM system. Of the 958 patients, 232 were excluded for the following reasons: pathology other than transitional cell carcinoma in 135, missing clinical T stage in 112, missing clinical grade in 30, missing concomitant CIS in 93, missing pathologic T stage in 78, missing pathologic N stage in 33, and unknown neo-adjuvant chemotherapy (NACHT) status in 38. Therefore, 726 evaluable patients were included in the analyses of the primary outcomes of interest: presence of pT3–4 stage and presence of pN1–3 stage.

#### 2.2. Pathology

Staff pathologists from each institution with expertise in genitourinary pathology examined all specimens according to institutional protocols. Multiple, well-oriented quadrant sections from the tumour, adjacent and distant bladder wall, ureters, and urethra were processed. Pelvic lymph node dissections were examined grossly, and all lymphoid tissue was submitted for histological examination. The 2002 TNM classification was used for pathologic staging, and the 1973 WHO classification was used for pathologic grading. Patients who had surgery before 2002 had their pathologic stage updated to reflect the 2002 TNM staging system. To ensure validity of the data extraction of pathologic outcomes, two clinicians read the pathology reports of 219 consecutive patients, while blinded to patient clinical parameters and the findings of the other reviewer. Interreader reliability measured using the intraclass correlation coefficient was greater than 0.95 for all pathologic parameters.

#### 2.3. Statistical analyses

Univariate and multivariate logistic regression models addressed two separate outcomes: stage pT3–4 at radical cystectomy and stage pN1–3 at radical cystectomy. The predictors consisted of age, gender, T stage at TUR (TUR stage), tumour grade at TUR (TUR grade), concomitant CIS at TUR, and delivery of NACHT. A backward step-down selection process was applied to the multivariate models, which contained all predictors, to arrive at the most informative and the most parsimonious model [7]. Multivariate logistic regression coefficients were used to generate predictive nomograms [7]. Accuracy of these nomograms was quantified with receiver operator characteristic area under the curve (AUC). An AUC value of 1.0 indicates perfect predictions, while 0.5 is equivalent to a toss of a coin. Internal validation was performed using 200 bootstrap resamples [8]. Calibration plots were generated to explore nomogram performance. All statistical tests were performed with S-Plus Professional software (MathSoft, Inc., Seattle, Washington, USA), and statistical significance was set at $p < 0.05$.

### 3. Results

The descriptive variables of the 726 evaluable patients are shown in Table 1. Mean age at cystectomy was 64.6 yr (median: 66, range: 34–89) and 593 patients were male (81.7%). At TUR, 96 (13.2%) had Ta or Tis disease, 173 (23.8%) had T1 disease, 375 (51.7%) had T2 stage, and 82 (11.3%) had stages T3 or T4. Ninety-one percent of patients had TUR grade 3 cancers, and 40.5% had concomitant CIS. Thirty-eight patients (5.2%) received NACHT. Table 2 shows the cross-tabulation between clinical and pathologic stages. Overall, 304 patients (41.9%) had pT3–4 stages at cystectomy. Presence of pN1–3 stages was recorded in 23.8%.
Agreement between TUR and cystectomy stage was recorded in 259 (35.7%) patients (Table 2). Of all patients, 302 (41.6%) had lower stage at TUR than at cystectomy and were classified as understaged at TUR. Conversely, stage reduction, which implies lower stage at cystectomy than at TUR, was noted in 153 (21.1%) patients. Of 173 patients with nodal metastases at cystectomy, clinical T1 or lower disease was found in 40 (23.1%), clinical T2 disease in 101 (57.7%), and clinical T3 and T4 stages in 32 (18.5%).

Table 3 shows the univariate and multivariate logistic regression models for prediction of pT3–4 disease. In univariate analyses, TUR stage, TUR grade, neo-adjuvant chemotherapy, and concomitant CIS at TUR were significant predictors of pT3–4 disease. In multivariate analysis, only TUR stage remained a significant predictor of pT3–4 disease.
grade, concomitant CIS (preoperative CIS), and NACHT were significantly associated with pT3–4 stage. Of these predictors, TUR stage (71.4%), concomitant CIS (61.3%), and TUR grade (54.4%) had the highest bootstrap-corrected predictive accuracy. After inclusion of all variables in the full multivariate model, age, TUR stage, TUR grade, and concomitant CIS were independently associated with pT3–4 stage. The bootstrap-corrected accuracy of the full multivariate model was 75.4%. Backward variable elimination, performed to increase parsimony while maintaining at least 75.4% accuracy, yielded a reduced model without gender and NACHT, with 75.7% accuracy.

Table 4 shows the univariate and multivariate logistic regression analyses for prediction of pathologic N1–3 disease. In univariate analyses, TUR stage, TUR grade, and concomitant CIS were the only variables associated with pN1–3 stage. Univariate bootstrap-corrected predictive accuracy estimates were 61.0% for TUR stage, 54.3% for concomitant CIS, and 53.8% for TUR grade. After inclusion of all variables in the full multivariate model, only TUR stage and grade remained statistically significantly associated with pN1–3 stage. Bootstrap-corrected accuracy of the full multivariate model was 63.3%. Backward variable elimination, performed to increase parsimony while maintaining maximum accuracy, yielded a reduced model with only TUR stage and grade. Its accuracy was 63.1%. Accounting for the effect of institution and that of study period had a trivial effect on predictive accuracy, which ranged from −0.6% to +0.1%.

Fig. 1A shows the reduced model nomogram and the calibration plot for prediction of pT3–4 stage at cystectomy. The calibration plot demonstrates virtually ideal predictions (45° line). Fig. 1B shows the reduced model nomogram and the respective calibration plot for prediction of pN1–3 stage at cystectomy. The calibration plots of this nomogram demonstrate departures from ideal predictions in the low probability range.

### Table 3 – Univariate and multivariate logistic regression analyses predicting the probability of pT3–4 stage at cystectomy with precystectomy variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate model odds ratio (p)</th>
<th>Individual variable predictive accuracy (%)</th>
<th>Multivariate full model odds ratio (p)</th>
<th>Multivariate reduced model odds ratio (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.0 (0.09)</td>
<td>54.2</td>
<td>1.02 (0.002)</td>
<td>1.02 (0.002)</td>
</tr>
<tr>
<td>Male gender</td>
<td>0.9 (0.8)</td>
<td>50.8</td>
<td>1.04 (0.9)</td>
<td>–</td>
</tr>
<tr>
<td>TUR stage</td>
<td>(&lt;0.001)</td>
<td>71.4</td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
</tr>
<tr>
<td>T2 vs. Ta, TIS, T1</td>
<td>4.0 (&lt;0.001)</td>
<td>3.2 (&lt;0.001)</td>
<td>23.1 (&lt;0.001)</td>
<td>21.8 (&lt;0.001)</td>
</tr>
<tr>
<td>T3 or higher vs. Ta, TIS, T1</td>
<td>22.1 (&lt;0.001)</td>
<td>3.2 (&lt;0.001)</td>
<td>23.1 (&lt;0.001)</td>
<td>21.8 (&lt;0.001)</td>
</tr>
<tr>
<td>TUR grade</td>
<td>0.4 (&lt;0.001)</td>
<td>61.3</td>
<td>0.6 (0.003)</td>
<td>0.6 (0.003)</td>
</tr>
<tr>
<td>Concomitant CIS</td>
<td>2.2 (0.019)</td>
<td>51.9</td>
<td>0.8 (0.5)</td>
<td>–</td>
</tr>
<tr>
<td>Model predictive accuracy (%)</td>
<td></td>
<td>75.4</td>
<td>75.7</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4 – Univariate and multivariate logistic regression analyses predicting pathologic stages N1–3 with precystectomy variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate model odds ratio (p)</th>
<th>Individual variable predictive accuracy (%)</th>
<th>Multivariate full model odds ratio (p)</th>
<th>Multivariate reduced model odds ratio (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.9 (0.3)</td>
<td>51.7</td>
<td>1.0 (0.6)</td>
<td>–</td>
</tr>
<tr>
<td>Male gender</td>
<td>1.1 (0.5)</td>
<td>50.1</td>
<td>1.2 (0.5)</td>
<td>–</td>
</tr>
<tr>
<td>TUR stage</td>
<td>(&lt;0.001)</td>
<td>61.0</td>
<td>(&lt;0.001)</td>
<td>(&lt;0.001)</td>
</tr>
<tr>
<td>T2 vs. Ta, TIS, T1</td>
<td>2.1 (&lt;0.001)</td>
<td>1.9 (0.003)</td>
<td>3.7 (&lt;0.001)</td>
<td>3.6 (&lt;0.001)</td>
</tr>
<tr>
<td>T3 or higher vs. Ta, TIS, T1</td>
<td>3.7 (&lt;0.001)</td>
<td>3.4 (0.006)</td>
<td>3.4 (0.006)</td>
<td>3.3 (0.008)</td>
</tr>
<tr>
<td>TUR grade</td>
<td>(0.004)</td>
<td>53.8</td>
<td>(0.006)</td>
<td>0.008</td>
</tr>
<tr>
<td>3 vs. 1 and 2</td>
<td>3.5 (0.004)</td>
<td>3.4 (0.006)</td>
<td>3.3 (0.008)</td>
<td></td>
</tr>
<tr>
<td>Concomitant CIS</td>
<td>0.7 (0.03)</td>
<td>54.3</td>
<td>0.9 (0.4)</td>
<td>–</td>
</tr>
<tr>
<td>Neo-adjuvant chemotherapy</td>
<td>0.9 (1.0)</td>
<td>50.7</td>
<td>0.6 (0.2)</td>
<td>–</td>
</tr>
<tr>
<td>Model predictive accuracy (%)</td>
<td></td>
<td>63.3</td>
<td>63.1</td>
<td></td>
</tr>
</tbody>
</table>
4. Discussion

Clinical staging is the major determinant governing therapeutic decision-making. We found that discrepancy between clinical and pathologic stage is common in patients who undergo radical cystectomy. Our data indicated agreement between TUR and pathologic stage in 35.7% of patients. Understaging was recorded in 42%. Notably, 32% of patients with TUR stage T2 or less had pT3–4 and/or pN1–3 stages at cystectomy. Nodal metastases at cystectomy were found in 24% of patients.

Many reasons account for the discrepancy between TUR and cystectomy stage, such as sampling error due to incompleteness of the TUR, delay in the interval from TUR to radical cystectomy, and poor sensitivity of preoperative staging tools [9,10]. Pathologic interpretation of specimens plays a central role [11–13]. Thus, there are numerous difficulties with predictions of cystectomy T stage.
using TUR stage. Prediction of nodal metastases prior to cystectomy is fraught with even greater difficulties [14]. To address the difficulty with accurate pT and pN stage assignment at cystectomy, we developed two nomograms that predict the probability of advanced T stage and of nodal metastases.

Our univariate models of pT$^{3-4}$ disease demonstrated that TUR T stage had 71.4% accuracy and represented the most informative single predictor of pT$^{3-4}$. The addition of all other available predictors resulted in a 4.0% accuracy gain, which confirmed that multivariate models are more accurate than most informative single predictors. Backward elimination of the least informative variables resulted in a reduced regression model, which was 4.3% more accurate than TUR T stage.

Nomograms predicting pT$^{3-4}$ at cystectomy demonstrate the effect of precystectomy variables on the probability of pT$^{3-4}$. Advancing TUR stage was directly proportional to increasing probability of pT$^{3-4}$. Presence of CIS at TUR reduced the probability of pT$^{3-4}$ stage. CIS at TUR, a harbinger of unfavourable disease, may represent the need to proceed with an early cystectomy [3].

TUR T stage represented the most informative predictor of N$^{1-3}$ disease, with a predictive accuracy of 61.0%. Addition of all variables increased accuracy by 2.3%. Backward variable elimination reduced the full model to two variables: TUR T stage and TUR grade. Their contribution resulted in the most parsimonious model, with 63.1% accuracy (i.e., 2.1% higher than with T stage alone). Within the full N$^{1-3}$ model, the same direction of the effect remained as in the models predicting pT$^{3-4}$ stage, where increasing T stage and high TUR grade were associated with increasing probability of N$^{1-3}$ disease.

Although, we have proven that pT$^{3-4}$ and pN$^{1-3}$ can be predicted more accurately with multiple variables than with a single variable, our nomograms are not perfectly accurate. The pT$^{3-4}$ nomogram provides a 4.0–4.3% gain versus TUR T stage. The pN$^{1-3}$ nomogram provides 2.1–2.3% gain versus TUR T stage. Thus, the pT$^{3-4}$ nomogram could improve the classification of pT$^{3-4}$ disease in 40–43 per 1000 additional patients versus TUR T stage alone. Similarly, the pN$^{1-3}$ nomogram could improve the classification of pN$^{1-3}$ disease in 21–23 per 1000 additional patients versus TUR T stage alone. Despite these improvements, the maximal accuracy of pT$^{3-4}$ predictions was 75.7%. Thus, 24.3% of patients would still be misclassified. Even more importantly, despite the improvement in pN$^{1-3}$ predictions, the maximum accuracy of 63.3% implies that 36.7% of patients would still be misclassified.

Suboptimal predictive accuracy is a common problem in bladder cancer. For example, Sylvester et al. [15] reported a coding scheme designed to predict the probability of recurrence and progression in patients with noninvasive bladder cancer; their recurrence model was 66% accurate and their progression model was 74–75% accurate. Shariat et al. [16] also addressed prediction of recurrence in patients with nonmuscle invasive TCC, reporting 75% accuracy without biomarkers and 81% with NMP22. Thus, novel disease markers are clearly needed to improve the current ability to predict bladder cancer stage and biology.

The above findings indicate that, like other predictive models, the two proposed nomograms are limited by suboptimal accuracy [15–19]. Although the nomograms are not perfect, should they be used? And if not, what is the alternative? As indicated by journal citations [15–19], suboptimal predictive and prognostic models are reported and used, based on the contention that in absence of perfectly accurate models it is better to predict with a systematic rule, with known accuracy and performance characteristics, than to predict in a less accurate and possibly haphazard fashion. In a direct comparison, Spechtt et al. [20] confirmed the contention that even imperfect nomograms (72%) are better than expert-clinician ratings (54%).

Thus, despite their lack of optimal accuracy, our nomograms may qualify as stepping stones towards structured, standardized, methodological, and objective risk stratification prior to cystectomy, upon which more accurate tools could be developed in the future. These multivariate diagnostic algorithms are capable of providing an individual probability of advanced T and N stages at cystectomy, with known accuracy and performance characteristics. Moreover, nomograms offer the possibility of assessing the increment in predictive accuracy, which is related to the addition of a candidate marker to a group of established predictors such as stage and grade, and could be used for testing of novel biomarkers or other promising diagnostic tools. This method is superior to multivariate p-value assessment of the statistical significance of a novel marker, as highly significant multivariate p-values may be deceptively associated with predictors that do not add to predictive accuracy [21]. Nomograms offer the advantage of being readily amenable to bootstrapping, which simulates the use of the diagnostic model in new patients and reduces overfit bias [8].

When a newly developed rule is applied to future patients, bootstrapping provides a more accurate estimate of expected predictive accuracy [8]. This
computer-intensive resampling method has been shown to represent the most efficient internal validation technique. It replicates the process of sample generation from an underlying population, by drawing samples with replacement from the original dataset. Two hundred bootstrap resamples simulate the application of a newly developed rule to 200 new samples of the same size as the original sample and with similar baseline characteristics. Bootstrapping is superior to other statistical alternatives, such as split sample or cross-validation techniques. Despite its superiority, bootstrapping is not a substitute of an external validation cohort, which represents the ultimate test of a newly developed rule. Finally, the performance characteristics of nomograms are amenable to graphical exploration with calibration plots. These allow the clinical users to become familiar with strength and weaknesses of the predicted probabilities, according to their distribution.

Aside from these strengths and weaknesses of our models, several additional points warrant mention. Our models are not applicable to patients who were pretreated with radiotherapy or to those harbouring pathologies other than transitional cell carcinoma. The proposed nomograms do not incorporate certain predictors of bladder cancer aggressiveness, such as time from initial diagnosis, number of recurrences, or response to intravesical therapy. Although, these are clinically important, they were not universally available in our databases, which were combined between three institutions. Finally, our nomograms were derived from patients treated with cystectomy at three North American referral centres. Thus, the indications for cystectomy reflect the T stage and associated pathologic findings at TUR that prompted cystectomy at these particular centres. Alternative cystectomy indications may change the relation between TUR stage and pathologic stage. Thus, external validation of our findings should be performed at centres where cystectomy indications differ from ours.

This study demonstrated that the combined contribution of staging variables results in more accurate predictions than the most informative predictor in isolation. Our nomograms accurately predict the probability of pT3–4 stage and provide estimates of the probability of pN1–3 disease with reasonable accuracy. Despite departures from ideal predictions, our nomograms represent the first attempt at defining objective, systematic, standardized, multivariate models capable of providing individual cystectomy stage predictions, with known accuracy and established performance characteristics.

Acknowledgements

PI Karakiewicz is partially supported by Fonds de la Recherche en Santé du Québec, the CHUM Foundation, the Department of Surgery and Les Urologues Associés du CHUM. SF Shariat is supported by the Austrian Program for Advanced Research and Treatment. GE Amiel and A Vazina are supported by the Lynn and Deanna Monical Fellowship in Urologic Oncology and Minimally Invasive Surgery.

References

Editorial Comment
Michael Kattan
kattanm@ccf.org

Karakiewicz et al. provide another nice illustration of the improvement in predictive accuracy that continuous models, such as nomograms, have over risk groups, such as staging systems. The authors appropriately acknowledge that nomograms are not perfect, just better than our typical methods of risk stratification.

Fig. 1 outlines the typical options a clinician has when a patient asks for a prediction. Among the options, nomograms will tend to offer superior predictive accuracy. Denying the ability to predict in the individual patient is faulty reasoning if the clinician truly desires to provide an accurate probability for the patient [1]. Furthermore, using knowledge and experience, in general, yields a prediction that is less accurate than that provided by a statistical equation [2]. And finally, equations have proven themselves in many direct comparisons against risk groups [3]. Why quote an overall average to all patients, which does not discriminate among them at all?

Of course, this argument assumes that the clinician desires to deliver an accurate prediction, and if there is not a legitimate treatment decision...
at stake, this assumption may not hold; the clinician may simply want to provide hope in a situation with no alternatives. Another assumption is that the patient is numerate and can understand a statement such as “if we had 100 patients just like you, 40 would experience disease recurrence within 5 years.” Researchers are actively studying approaches to communicating with patients who cannot interpret this statement. But clinicians should not use lack of patient understanding as an excuse for not providing the most accurate predictions presently available.

References


Editorial Comment

A. Heidenreich, Köln, Germany
axel.heidenreich@uk-koeln.de

Nomograms to predict locally advanced tumors prior to radical cancer surgery are useful if they improve currently available staging systems. Although the nomogram presented by Karakiewicz et al. might be helpful to already preoperatively identify patients being at risk for locally advanced bladder cancer, there are some concerns to me which have to be addressed more carefully before this nomogram can be transferred into daily clinical routine. The increase in accuracy (2–4%) is very modest and in absolute sense (AUC = 63 to 75%) far from satisfactory. About 30% of all patients will still be understaged with the nomogram which might be attributed to several factors:

- Quality of TUR-B is one major factor contributing to the validity of the prediction of the recurrence rate and the final pathohistological stage at radical cystectomy. As has been shown by a recent EORTC trial there might be variations in the correct staging of bladder tumors varying from 5–70% among different institutions [1]. For the generation of the nomogram, retrospective results of 3 different institutions and many more surgeons have been assessed so that numerous surgical strategies might have been applied over the 20 years time interval.
- Repeat transurethral resection of the bladder should be performed in all patients with multifocal or muscle invasive bladder cancer. Repeat resection by itself will improve the clinical staging system without the need of a nomogram as has been demonstrated recently [2]. Differences in the resection technique might explain the high rate of understaging in patients with muscle invasive bladder cancer in the series presented.
- Nowadays, many patients undergo staging CT scans of the abdomen and pelvis prior to radical cystectomy making the identification of locally advanced bladder cancer more easy and appropriate as in earlier studies with either no or first and second generation CT scans [3]. The lack of adequate imaging techniques might explain the unusual high frequency of pT3 cancer in the presented series.
- Although the authors are to be congratulated for their efforts to establish a pre-cystectomy nomogram to predict locally advanced disease, the authors should attempt to repeat their efforts to develop a valid and clinically useful nomogram approaching a modern series of patients in different institutions also considering standardized primary and repeat TUR techniques and modern imaging modalities.

References