

RESEARCH ARTICLE

External Validation of a Clinical Scoring System for Hematuria

Seung Bae Lee¹, Hyung Suk Kim², Myong Kim², Ja Hyeon Ku^{2*}

Abstract

Background: The aim of this study was to evaluate the accuracy of a new scoring system in Korean patients with hematuria at high risk of bladder cancer. **Materials and Methods:** A total of 319 consecutive patients presenting with painless hematuria without a history of bladder cancer were analyzed, from the period of August 2012 to February 2014. All patients underwent clinical examination, and 22 patients with incomplete data were excluded from the final validation data set. The scoring system included four clinical parameters: age ($\geq 50 = 2$ vs. $< 50 = 1$), gender (male = 2 vs. female = 1), history of smoking (smoker/ex-smoker = 4 vs. non-smoker = 2) and nature of the hematuria (gross = 6 vs. microscopic = 2). **Results:** The area under the receiver-operating characteristic curve (95% confidence interval) of the scoring system was 0.718 (0.655-0.777). The calibration plot demonstrated a slight underestimation of bladder cancer probability, but the model had reasonable calibration. Decision curve analysis revealed that the use of model was associated with net benefit gains over the treat-all strategy. The scoring system performed well across a wide range of threshold probabilities (15%-45%). **Conclusions:** The scoring system developed is a highly accurate predictive tool for patients with hematuria. Although further improvements are needed, utilization of this system may assist primary care physicians and other healthcare practitioners in determining a patient's risk of bladder cancer.

Keywords: Hematuria - bladder cancer - scoring system - risk stratification - validation

Asian Pac J Cancer Prev, 15 (16), 6819-6822

Introduction

Hematuria is the most common symptom presenting in patients with bladder cancer. Since screening is not currently recommended, most patients are diagnosed after presenting with hematuria to their clinician. Roughly 4% of the patients with microscopic hematuria and up to 40% of the patients with gross hematuria could be harboring malignancy (Grossfeld et al., 2001). However, the prevalence of urinary tract cancer in the general population is low, translating to a low prior probability of disease and its lack of screening. Moreover, although American Urological Association (AUA) best practice policy recommendations and the most recent AUA guideline include urine testing, imaging, and cystoscopy (Grossfeld et al., 2001; Davis et al., 2012), these recommendations do not perform well in identifying which patients were most likely to have malignant tumors. Therefore, alternative criteria are needed for better identification of patients who truly require further evaluation to effectively and efficiently manage patients with hematuria.

It has been demonstrated that an individual patient's risk for bladder cancer depends on several factors, including age, tobacco use and chemical exposures (Nielsen et al., 2007). However, few models have been developed to quantify individual risk of having bladder cancer after presenting with hematuria (Nielsen et al.,

2007; Davis et al., 2012). Furthermore, as the existing models rely on nuclear matrix protein-22 (NMP22) (Lotan et al., 2009) or immunocytology (Cha et al., 2012a), they are not practical for a daily clinical environment. Recently, Tan et al. (2013) developed a new scoring system using four clinical variables (age, gender, smoking status, and nature of the hematuria) to stratify patients with hematuria into high or low risk of having bladder cancer.

The aim of this study was to evaluate the accuracy of this scoring system in Korean patients with hematuria.

Materials and Methods

This study was carried out after obtaining approval from the institutional review board of Seoul National University Hospital. Patients visiting our institution from August 2012 to February 2014 were included in this study, and 319 consecutive patients presenting with painless hematuria without a history of bladder cancer were analyzed. Patient information such as age, gender, smoking history, and nature of the hematuria were recorded. All patients underwent clinical examination, including CT urography and cystourethroscopy, with biopsy of any suspicious lesions. Microscopic hematuria was defined as three or more erythrocytes visible per high-power field under white-light microscopy. Mid-stream urine specimens were collected, immediately processed,

¹Department of Urology, ¹Seoul National University Boramae Hospital, ²Seoul National University Hospital, Seoul, Korea *For correspondence: kuuro70@snu.ac.kr

and examined cytologically. Cytologic examination was performed by trained personnel. A total of 22 patients were excluded from the final validation data set due to incomplete data. Consequently, 297 patients were available for final analyses. The demographic data for the external validation cohort is shown in Table 1.

The scoring system proposed by Tan et al (Tan et al., 2013) to predict the probability that patients presenting with hematuria had bladder cancer included four clinical parameters: age ($\geq 50=2$ vs $<50=1$), gender (male=2 vs female=1), history of smoking (smoker/ex-smoker=4 vs non-smoker=2) and nature of the hematuria (gross=6 vs microscopic=2). The total score ranged from 6 to 14.

The overall predictive values of the scoring system were compared with several criteria. Discrimination refers to the ability of the risk prediction model to distinguish those with the event from those without the event. Our measure of discrimination was area under the receiver-operating characteristic (ROC) curve. A score of 1 suggests that the model can perfectly discriminate between the patients who have bladder cancer and those who do not, while a score of 0.5 indicates that the model has no discriminative ability. The total area under the ROC curve estimates were internally validated using 500 bootstrap samples. Calibration refers to how closely the predicted probabilities reflect the actual risk. General calibration was assessed using a calibration plot. The relationship between the model-derived and actuarial outcome was graphically explored within the calibration plots to explore performance of the model. The validation was performed using 200 bootstrap resamples to decrease overfit bias. Finally, decision curve analysis (DCA) was used to explore the clinical value of the model (Vickers and Elkin, 2006). DCA is a method for evaluating the clinical net benefit of prediction models; one sums the benefits (true positives) and subtracts the harms (false positives).

For all statistical analyses, a two-sided $p < 0.05$ was regarded as significant. Models, statistics, and figures were

prepared using R 2.13.2 (<http://www.cran.r-project.org>).

Results

The bootstrap-corrected predictive accuracy of the model was good (Figure 1). The area under the ROC curve (95% confidence interval) of the scoring system was 0.718 (0.655-0.777). In our validation cohort, a score of 12 was found to be the best score threshold (Table 2). By using the cut-off value of 12, the scoring system had a sensitivity of 69.6% and a specificity of 69.7%. The positive predictive and negative predictive values were 45.5% and 86.4%, respectively.

To assess the agreement between the predicted and actual outcomes, a calibration curve was generated (Figure 2). The dashed diagonal line represents the performance of an ideal model, where the predicted outcome would correspond perfectly with the actual outcome. The performance of the current model is plotted as the solid line. The calibration plot demonstrated a slight underestimation of bladder cancer probability, but the model was reasonably calibrated. Although the calibration curve did not perfectly match the line of identity (the line at a 45° angle), the deviation was pictorially minimal, as the solid line was close to the dashed line of the ideal model.

Figure 3 presents the result of the DCA. DCA revealed that the use of the model was associated with net benefit gains over the treat-all strategy. The scoring system performed well across a wide range of threshold probabilities (15%-45%).

Table 1. Patients Characteristics

Clinical Factors	Points	No. (%) of Bladder Cancer Case	No. of Total Patients
Age (years)	≥ 50	72 (27.5%)	262
	< 50	7 (20.0%)	35
Gender	Male	67 (36.6%)	183
	Female	12 (10.5%)	114
Smoker status	Smoker/ex-smoker	62 (36.7%)	169
	Non-smoker	17 (13.3%)	128
Hematuria	Gross	68 (31.9%)	213
	Microscopic	11 (13.1%)	84

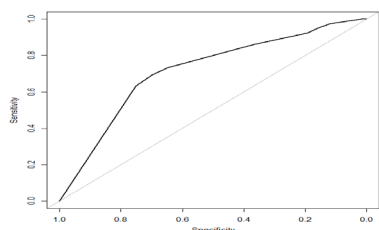


Figure 1. Receiver Operating Characteristics Curve. Area Under the Curve (95%CI)=0.718 (0.655-0.777)

Table 2. Model Performance of the Clinical Scoring System for Hematuria

Score Threshold	Sensitivity	Specificity	PPV	NPV
6.5	1.000	0.018	0.270	1.000
7.5	0.975	0.119	0.286	0.929
8.5	0.949	0.161	0.291	0.897
9.5	0.924	0.193	0.293	0.875
10.5	0.861	0.367	0.330	0.879
11.5	0.734	0.647	0.430	0.870
12.5	0.696	0.697	0.455	0.864
13.5	0.633	0.752	0.481	0.850

*PPV:Positive Predictive value; NPV: Negative Predictive value

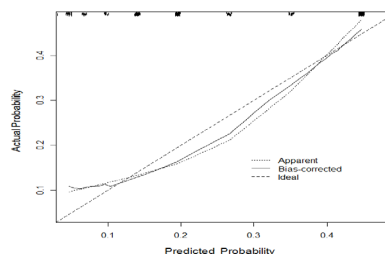


Figure 2. Calibration Plot. The calibration curve demonstrates the relation between the predicted and observed bladder cancer rates. The diagonal line represents the performance of an ideal model. The solid line represents the actual model performance that compares the predicted and observed bladder cancer rates (using 200 bootstrap samples). Points estimated below the diagonal line correspond to over prediction, whereas points situated above the diagonal line correspond to under prediction

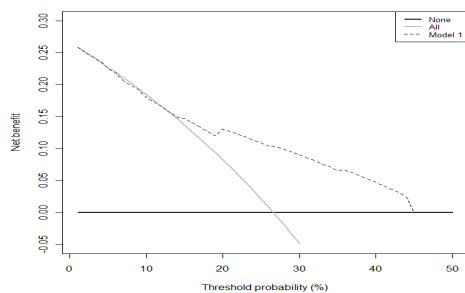


Figure 3. Decision Curve Analysis. In decision curve analysis, the y-axis measures net benefit, calculated by summing the benefits (true positives) and subtracting the harms (false positives). The straight line represents the assumption that all patients will have bladder cancer, and the horizontal line represents the assumption that no patients will have bladder cancer. The dotted line indicates the net benefit of using the new model

Discussion

In 2012, bladder cancer accounted for 7% of new cancer diagnoses and 3% of cancer deaths in American men (Siegel et al., 2012). Approximately 25% of the patients with bladder cancer present with muscle-invasive diseases at diagnosis. Early detection and timely treatment of the cancer are critical, since delays in diagnosis and treatment have been shown to adversely affect survival in patients with bladder cancer (Gore et al., 2009). Traditional screening tools (e.g., cystoscopy) are not practical for population-based screening because they are invasive and not cost-efficient (Botteman et al., 2003). In addition, the procedure has been associated with pain or discomfort in about a third of cases (Van Der Aa et al., 2008). Due to concerns about the number of unnecessary diagnostic evaluations as well as the low prevalence of bladder cancer in the general population (Sutton, 1990), screening is not currently recommended.

Presenting with the clinical sign of hematuria is not trivial, as most patients with bladder cancer are only diagnosed after the development of this symptom. Hematuria is a common urological presentation, estimated to constitute 4-20% of all urological visits (Mariani et al., 1989). In the adult population, it can have different causes, such as urinary tract infections, urolithiasis, benign prostatic hyperplasia, urologic malignancies and secondary tumor invasion from other site (ex. Cervical cancer). In gynecologic section, it was demonstrated that hematuria can be used as a screening test to detect urinary bladder mucosal infiltration of cervical cancer (Chuttiangtum et al., 2012). It has been reported that Urological cancer is found in approximately 5% of the patients presenting with microscopic hematuria, and in around 20% of the patients with gross hematuria (Khadra et al., 2000; Cohen and Brown, 2003). Therefore, screening for bladder cancer may have different performance characteristics in selected high-risk populations. In other areas of urology, e.g. patients with raised prostate-specific antigen levels, sophisticated algorithms and nomograms have been developed that stratify the patient's risk of cancer. With this information available, some patients avoid a prostate biopsy whilst other urological conditions can be identified and managed.

However, little research has been carried out on the investigation of hematuria (Khadra et al., 2000; Lotan et al., 2009; Cha et al., 2012a). Using age, gender, race, smoking history, hematuria, and NMP22 findings, Lotan et al (Lotan et al., 2009) developed a logistic regression model-based nomogram to predict the presence of bladder cancer in high-risk patients. External validation of this nomogram gave an AUC of more than 80%. Cha et al (Cha et al., 2012a) developed another nomogram with an extremely high level of accuracy (bootstrap-corrected AUC: 90.8%) in patients with painless hematuria. This nomogram incorporated immunocytology into a multivariable prediction model for the detection of bladder cancer. Cha et al. (2012b) also developed a highly accurate, well-calibrated nomogram to predict the individual risk of bladder cancer for a patient presenting with asymptomatic hematuria. In this nomogram, age, gender, smoking history, degree of hematuria, and urine cytology define the risk. Although these models have been shown to be highly accurate, they are difficult to apply in a daily clinical environment. Recently, it has been tried in women with hematuria to evaluate the impact of reproductive factors (menopausal status, parity, age at first delivery and age at the last delivery) on the prediction of the risk for developing bladder cancer (Yavuzcan et al., 2013).

Tan's scoring system demonstrated adequate discrimination in the first study (area under the ROC curve=80.4%) (Tan et al., 2013). Nevertheless, this model needs to be validated in a secondary dataset. Such external validation is essential, as a predictive tool is only useful if it is both accurate and generally applicable. The aim of the present study was to externally validate the previously developed model in the prediction of bladder cancer in a patient cohort with hematuria. We report the first external validation of this model in a contemporary single institution cohort of patients. Good predictability was obtained by the clinical scoring system in this study; the accuracy of the model was 71.8%. The correspondence seen between the actual and ideal model predictions by the calibration plot suggested good calibration of the scoring system in the validation cohort. Furthermore, the model demonstrated a meaningful net benefit gain. In addition, we observed a score of 12 as the best score threshold in our validation cohort. Tan et al. (2013) reported that a threshold figure of 10 gives the scoring system a sensitivity of 90% and a specificity of 55.7%. Using the suggested cut-off value of 10 in our cohort, the scoring system had a sensitivity of 86.1% but a specificity of only 36.7%, while the cut-off value of 12 yielded a sensitivity of 69.6% and a specificity of 69.7%.

In general, model performance tends to be lower when external validation is performed, for several reasons. First, the construction of a valid model relies on the appropriate selection of variables for the analysis. For example, smoking history was categorized as "current", "history of" or "never". Quantification of tobacco use could have improved the accuracy of the model. Second, undoubtedly, it is possible that inclusion of other variables might have improved the performance of the model. Therefore, addition of occupational or other chemical exposures

would also likely improve the model's predictive accuracy. Third, in Tan's scoring system, the cut-off age of 50 years was chosen. Although many studies have considered ages above this to be a risk factor for bladder cancer (Messing et al., 1987; Alishahi et al., 2002; Madeb and Messing, 2008), Shariat et al (Shariat et al., 2010) demonstrated that the incidence of bladder cancer increases more significantly after 65 years of age. Finally, combinations of molecular markers may improve the prediction of bladder cancer in patients with hematuria. The ability of biomarker data to improve model predictions has already been confirmed (Shariat et al., 2008). However, the limited availability of data on such molecular markers might restrict the widespread application of such models. In addition, since racial variation in the expression of molecular markers may also be present, validation of the models would need to be performed in research including different races.

The study was limited by several factors. Data management was retrospectively assessed for analysis, which is the main drawback. Because 22 patients without complete clinical information or whose information was not available had to be excluded, a possible selection bias may have been introduced. Finally, this study set came from a single institution and included a relatively small number of patients. However, bootstrap corrected predictive accuracy reduces overfit bias, offers the possibility of internal validation without sample size limitations and provides the most bias-free estimates of discrimination properties.

In conclusion, the scoring system proved to be a highly accurate predictive tool for patients with hematuria. This routine checkup for age, gender, history of smoking, and nature of the hematuria is universally applicable, adds no additional costs, and is not a time-consuming process. These results encourage the application of this scoring system for predicting bladder cancer. Although further improvements are needed, utilization of this scoring system may assist primary care physicians and other healthcare practitioners in determining a patient's risk of bladder cancer.

References

- Alishahi S, Byrne D, Goodman CM, Baxby K (2002). AUDIT SECTION-Haematuria investigations based on a standard protocol: Emphasis on the diagnosis of urological malignancy. *J Roy Coll Sug Edin*, **47**, 422-7.
- Botteman MF, Pashos CL, Redaelli A, Laskin B, Hauser R (2003). The health economics of bladder cancer: a comprehensive review of the published literature. *Pharmacoeconomics*, **21**, 1315-30.
- Cha EK, Tirsar L-A, Schwentner C, et al (2012a). Immunocytology is a strong predictor of bladder cancer presence in patients with painless hematuria: a multicentre study. *Eur Urol*, **61**, 185-92.
- Cha EK, Tirsar L-A, Schwentner C, et al (2012b). Accurate risk assessment of patients with asymptomatic hematuria for the presence of bladder cancer. *World J Urol*, **30**, 847-52.
- Chuttiangtum A, Udomthavornsuk B, Chumworathayi B (2012). Hematuria screening test for urinary bladder mucosal infiltration in cervical cancer. *Asian Pac J Cancer Prev*, **13**, 4931-33.
- Cohen RA, Brown RS (2003). Microscopic hematuria. *New England J Med*, **348**, 2330-8.
- Davis R, Jones JS, Barocas DA, et al (2012). Diagnosis, evaluation and follow-up of asymptomatic microhematuria (AMH) in adults: AUA guideline. *J Urol*, **188**, 2473-81.
- Gore JL, Lai J, Setodji CM, Litwin MS, Saigal CS (2009). Mortality increases when radical cystectomy is delayed more than 12 weeks. *Cancer*, **115**, 988-96.
- Grossfeld GD, Litwin MS, Wolf JS, et al (2001). Evaluation of asymptomatic microscopic hematuria in adults: the American Urological Association best practice policy--part I: definition, detection, prevalence, and etiology. *Urology*, **57**, 599-603.
- Khadra MH, Pickard RS, Charlton M, Powell PH, Neal DE (2000). A prospective analysis of 1,930 patients with hematuria to evaluate current diagnostic practice. *J Urol*, **163**, 524-7.
- Lotan Y, Capitanio U, Shariat SF, Hutterer GC, Karakiewicz PI (2009). Impact of clinical factors, including a point of care nuclear matrix protein 22 assay and cytology, on bladder cancer detection. *BJU Int*, **103**, 1368-74.
- Madeb R, Messing EM (2008). Long-term outcome of home dipstick testing for hematuria. *World J Urol*, **26**, 19-24.
- Mariani A, Mariani M, Macchioni C, et al (1989). The significance of adult hematuria: 1,000 hematuria evaluations including a risk-benefit and cost-effectiveness analysis. *J Urol*, **141**, 350-5.
- Messing E, Young T, Hunt V, Emoto SE, Wehbie JM (1987). The significance of asymptomatic microhematuria in men 50 or more years old: findings of a home screening study using urinary dipsticks. *J Urol*, **137**, 919-22.
- Nielsen ME, Shariat SF, Karakiewicz PI, et al (2007). Advanced age is associated with poorer bladder cancer-specific survival in patients treated with radical cystectomy. *Eur Urol*, **51**, 699-708.
- Shariat SF, Sfakianos JP, Droller MJ, et al (2010). The effect of age and gender on bladder cancer: a critical review of the literature. *BJU Int*, **105**, 300-8.
- Shariat SF, Walz J, Roehrborn CG, et al (2008). External validation of a biomarker-based preoperative nomogram predicts biochemical recurrence after radical prostatectomy. *J Clin Oncol*, **26**, 1526-31.
- Siegel R, DeSantis C, Virgo K, et al (2012). Cancer treatment and survivorship statistics, 2012. *CA-Cancer J Clin*, **62**, 220-41.
- Sutton JM (1990). Evaluation of hematuria in adults. *JAMA*, **263**, 2475-80.
- Tan GH, Shah SA, Ann HS, et al (2013). Stratifying patients with haematuria into high or low risk groups for bladder cancer: a novel clinical scoring system. *Asian Pac J Cancer Prev*, **14**, 6327-30.
- Van Der Aa MN, Steyerberg EW, Sen EF, et al (2008). Patients' perceived burden of cystoscopic and urinary surveillance of bladder cancer: a randomized comparison. *BJU Int*, **101**, 1106-10.
- Vickers AJ, Elkin EB (2006). Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*, **26**, 565-74.
- Yavuzcan A, Caglar M, Kayikci MA, et al (2013). Can reproductive characteristics predict bladder cancer in women with haematuria? *Asian Pac J Cancer Prev*, **14**, 5107-10.