OPEN

Bladder Cancer

JU Insight

THE JOURNAL of UROLOGY®
www.auajournals.org/journal/juro

Predicting Response to Intravesical Bacillus Calmette-Guérin in High-Risk Nonmuscle-Invasive Bladder Cancer Using an Artificial Intelligence–Powered Pathology Assay: Development and Validation in an International 12-Center Cohort

Yair Lotan, Viswesh Krishna, Waleed M. Abuzeid, et al.

Correspondence: Stephen B. Williams (stbwilli@utmb.edu).

Full-length article available at https://doi.org/10.1097/JU.0000000000004278.

Study Need and Importance: A significant proportion of patients with high-risk nonmuscle-invasive bladder cancer (HR-NMIBC) will experience poor outcome despite optimal management. Biomarkers may allow urologists to identify patients most likely to recur or progress, helping guide clinical decision-making. We developed and validated artificial intelligence–based histologic assays that extract interpretable features from digitized pathology images of HR-NMIBC derived from routine transurethral resection of bladder tumor to predict the risk of recurrence, progression, cystectomy, and development of bacillus Calmette-Guérin (BCG)–unresponsive disease.

What We Found: Cases (n = 944) of HR-NMIBC were accrued across 12 institutions internationally and were used in development and external validation cohorts (development: 303, validation: 641, median follow-up: 36 months). In the validation cohort, "high recurrence risk" cases had inferior high-grade recurrence-free survival vs "low recurrence risk" cases (HR, 2.08, $P < .0001$; part A of Figure). "High progression risk" patients had poorer progression-free survival (HR, 3.87, $P < .001$; part B of Figure) and higher risk of cystectomy (HR, 3.35, $P < .0001$) than "low progression risk" cases. Cases harboring the BCG-unresponsive disease signature had a shorter time to development of BCG-unresponsive disease than cases without the signature (HR, 2.31, $P < .0001$). Artificial intelligence assays provided predictive information beyond clinicopathologic factors.

Limitations: Notable limitations include the retrospective study design and heterogeneity in BCG regimens.

Interpretation for Patient Care: In an age of BCG shortages, these biomarkers could help identify patients least likely to benefit from BCG. Determination of high progression risk may aid in counseling regarding cystectomy. These biomarkers may also facilitate optimal selection of patients for intravesical BCG vs alternative chemotherapy or novel regimens.

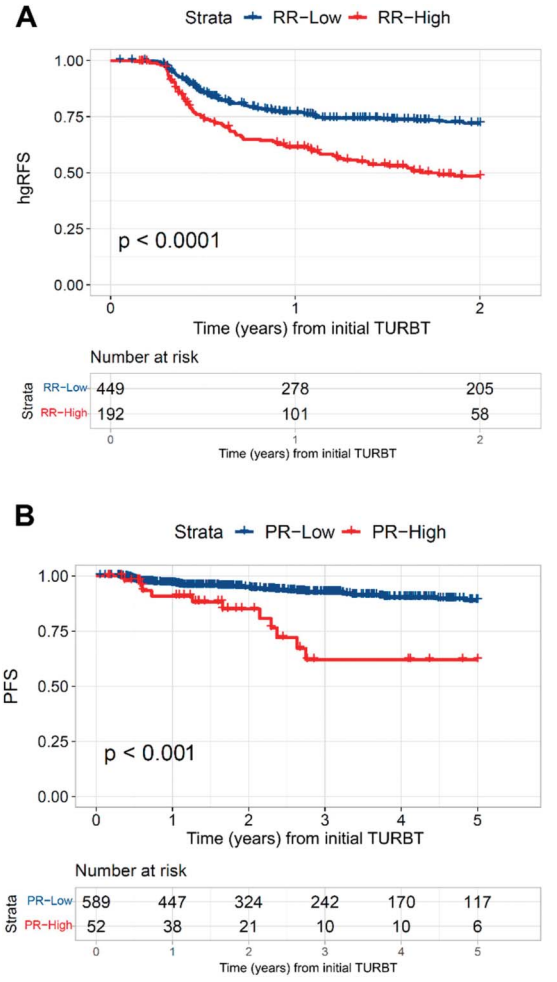

Figure. A, Recurrence risk (RR) model identifies patients at high (RR-High) and low (RR-Low) risk of high-grade recurrence (HR, 2.08; 95% CI, 1.80-2.40; $P < .0001$). High-grade recurrence-free survival (hgRFS) is censored at 2 years. B, The progression risk (PR) model identifies patients at high (PR-High) and low (PR-Low) risk of progression to muscle invasion (HR, 3.87; 95% CI, 2.75-5.44; $P < .001$). Progression-free survival (PFS) was censored at 5 years. TURBT indicates transurethral resection of bladder tumor.

THE JOURNAL OF UROLOGY®

https://doi.org/10.1097/JU.0000000000004278
Vol. 213, 192-204, February 2025
Printed in U.S.A.

THE JOURNAL of UROLOGY®
www.auajournals.org/journal/juro

OPEN

Predicting Response to Intravesical Bacillus Calmette-Guérin in High-Risk Nonmuscle-Invasive Bladder Cancer Using an Artificial Intelligence–Powered Pathology Assay: Development and Validation in an International 12-Center Cohort

Yair Lotan,[1]* Viswesh Krishna,[2]* Waleed M. Abuzeid,[2]* Bryn Launer,[3] Sam S. Chang,[3] Vrishab Krishna,[2] Siddhant Shingi,[2] Jennifer B. Gordetsky,[4] Thomas Gerald,[1] Solomon Woldu,[1] Eugene Shkolyar,[5] Dickon Hayne,[6] Andrew Redfern,[7] Lisa Spalding,[6] Courtney Stewart,[8] Eduardo Eyzaguirre,[9] Shamsunnahar Imtiaz,[10] Vikram M. Narayan,[10] Vignesh T. Packiam,[11] Michael A. O'Donnell,[11] Roger Li,[12] Loic Baekelandt,[13] Steven Joniau,[13] Tahlita Zuiverloon,[14] Mario I. Fernandez,[15] Marcela Schultz,[16] Patrick J. Hensley,[17] Derek Allison,[18] John A. Taylor,[19] Ameer Hamza,[20] Ashish Kamat,[21] Vivek Nimgaonkar,[2] Snehal Sonawane,[2] Daniel L. Miller,[2] Drew Watson,[2] Damir Vrabac,[2†] Anirudh Joshi,[2†] Jay B. Shah,[5†] and Stephen B. Williams[8†]

[1]*Department of Urology, University of Texas Southwestern Medical Center, Dallas, Texas*
[2]*Valar Labs, Palo Alto, California*
[3]*Department of Urology, Vanderbilt University Medical Center, Nashville, Tennessee*
[4]*Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee*
[5]*Department of Urology, Stanford Medicine, Palo Alto, California*
[6]*Department of Urology, University of Western Australia, Perth, Australia*
[7]*Department of Medical Oncology, University of Western Australia, Perth, Australia*
[8]*Department of Urology, University of Texas Medical Branch, Galveston, Texas*
[9]*Department of Pathology, University of Texas Medical Branch, Galveston, Texas*
[10]*Department of Urology, Emory University School of Medicine, Atlanta, Georgia*
[11]*Department of Urology, University of Iowa, Iowa City, Iowa*
[12]*Department of Urology, Moffitt Cancer Center, Tampa, Florida*
[13]*Department of Urology, UZ Leuven, Leuven, Belgium*
[14]*Department of Urology, Erasmus University Medical Center, Rotterdam, The Netherlands*
[15]*Department of Urology, Clínica Alemana Universidad del Desarrollo, Santiago, Chile*
[16]*Department of Pathology, Clínica Alemana Universidad del Desarrollo, Santiago, Chile*
[17]*Department of Urology, University of Kentucky College of Medicine, Lexington, Kentucky*
[18]*Department of Pathology, University of Kentucky College of Medicine, Lexington, Kentucky*
[19]*Department of Urology, University of Kansas Medical Center, Kansas City, Kansas*
[20]*Department of Pathology, University of Kansas Medical Center, Kansas City, Kansas*
[21]*Department of Urology, MD Anderson Cancer Center, Houston, Texas*

Purpose: There are few markers to identify those likely to recur or progress after treatment with intravesical bacillus Calmette-Guérin (BCG). We developed and validated artificial intelligence (AI)–based histologic assays that extract interpretable features from transurethral resection of bladder tumor digitized pathology images to predict risk of recurrence, progression, development of BCG-unresponsive disease, and cystectomy.

Materials and Methods: Pre-BCG resection-derived whole-slide images and clinical data were obtained for high-risk nonmuscle-invasive bladder cancer cases treated with BCG from 12 centers and were analyzed through a segmentation and feature extraction pipeline. Features associated with clinical outcomes were defined and tested on independent development and validation cohorts. Cases were classified into high or low risk for recurrence, progression, BCG-unresponsive disease, and cystectomy.

Results: Nine hundred forty-four cases (development: 303, validation: 641, median follow-up: 36 months) representative of the intended use population were included

THE JOURNAL OF UROLOGY®

https://doi.org/10.1097/JU.0000000000004278
Vol. 213, 192-204, February 2025
Printed in U.S.A.

(high-grade Ta: 34.1%, high-grade T1: 54.8%; carcinoma in situ only: 11.1%, any carcinoma in situ: 31.4%). In the validation cohort, "high recurrence risk" cases had inferior high-grade recurrence-free survival vs "low recurrence risk" cases (HR, 2.08, $P < .0001$). "High progression risk" patients had poorer progression-free survival (HR, 3.87, $P < .001$) and higher risk of cystectomy (HR, 3.35, $P < .001$) than "low progression risk" patients. Cases harboring the BCG-unresponsive disease signature had a shorter time to development of BCG-unresponsive disease than cases without the signature (HR, 2.31, $P < .0001$). AI assays provided predictive information beyond clinicopathologic factors.

Conclusions: We developed and validated AI-based histologic assays that identify high-risk nonmuscle-invasive bladder cancer cases at higher risk of recurrence, progression, BCG-unresponsive disease, and cystectomy, potentially aiding clinical decision making.

Key Words: artificial intelligence, bladder cancer, progression, recurrence

NONMUSCLE-INVASIVE bladder cancer (NMIBC) represents 70% of bladder cancers. Per the National Comprehensive Cancer Network, AUA, and European Association of Urology (EAU) management guidelines, standard-of-care treatment of high-risk NMIBC (HR-NMIBC) involves transurethral resection of bladder tumor (TURBT) followed by intravesical bacillus Calmette-Guérin (BCG) or radical cystectomy in those harboring the highest-risk features.[1-4]

Despite treatment, HR-NMIBC is associated with 5-year recurrence and progression rates of approximately 50% and 20%, respectively.[1] BCG intolerance drives treatment discontinuation in 8% of patients.[5] Coupled with BCG supply shortages and ongoing evaluation of alternative first-line therapies, the identification of patients most likely to benefit from BCG is critical.[6,7] There are limited means of predicting risk of recurrence or progression after TURBT and BCG therapy in HR-NMIBC. Meanwhile, artificial intelligence (AI) has demonstrated the capability to quantify morphological features in pathology and harness these data to prognosticate outcomes and predict treatment response in other cancers.[8,9]

In this multicenter study, we describe the development and validation of a deep learning–based AI model, termed the computational histology AI (CHAI) platform, that analyzes digital whole-slide images (WSIs) derived from routine pre-treatment TURBT specimens of BCG-naïve HR-NMIBC and identifies patients at high or low risk of post-BCG high-grade recurrence, progression, development of BCG-unresponsive disease (BUD), and cystectomy.

METHODS

Study Population

Institutional review board approval was obtained at each of the 12 participating academic centers. Methodology and findings were reported per REMARK guidelines.[10] Clinicopathologic data from BCG-naïve patients (18 years and older) with primary or recurrent HR-NMIBC per AUA criteria diagnosed between 2000 and 2022 were acquired.[2] Patients underwent initial TURBT ± restaging TURBT, followed by intravesical BCG. The initial TURBT-derived

Submitted March 6, 2024; accepted September 30, 2024; published October 9, 2024.

Funding/Support: Valar Labs provided support for whole-slide digital imaging of archived pathology slides when requested.

Conflict of Interest Disclosures: Dr Lotan—consultant (Valar Labs, Pacific Edge, Photocure, Astrazeneca, Vessi Medical, Nucelix, Merck, Engene, CAPs Medical, C2I Genomics, Fergene, Abbvie, Ambu, Seattle Genetics, Verity Pharmaceuticals, Urogen, Stimit, Nanorobot, Convergent Genomics, Aura Biosciences, Nonagen, Pfizer, Phinomics, CG Oncology, Virtuoso Surgical, Xcures, Uroviu, On Target Lab, Promis Diagnostics, Uroessentials); Dr Chang—consultant (Valar Labs); Dr Gordetsky—consultant (Janssen); Dr Shkolyar—consultant (Johnson & Johnson); Dr Hayne—consultant (Merck, Pfizer, Urogen, BMS), clinical trial support (Valar Labs, Medical Developments International, Astrazeneca); Dr Packiam—consultant (Valar Labs); Dr Fernandez—speaker bureau (Grunenthal); advisory board, (Janssen, Tecnofarma); Dr Shah—consultant (Valar Labs); Dr Williams—consultant (Valar Labs, Janssen, Merck, Provepharm, Photocure), advisory board (BJUI).

Author Contributions:

Conception and design: Redfern, Joshi, Vrabac, Miller, Shah, Gordetsky, Baekelandt, O'Donnell, Shingi, Williams, Krishna, Nimgaonkar, Abuzeid, Lotan.

Drafting the manuscript: Joshi, Vrabac, Sonawane, Williams, Zuiverloon, Krishna, Abuzeid.

Data analysis and interpretation: Joshi, Kamat, Launer, Vrabac, Miller, Hayne, Watson, Eyzaguirre, Shah, Gordetsky, Taylor, Baekelandt, O'Donnell, Hensley, Li, Chang, Shingi, Sonawane, Woldu, Williams, Zuiverloon, Gerald, Packiam, Narayan, Krishna, Nimgaonkar, Abuzeid, Lotan.

Data acquisition: Hamza, Redfern, Joshi, Launer, Stewart, Vrabac, Allison, Hayne, Watson, Eyzaguirre, Shkolyar, Shah, Gordetsky, Taylor, Spalding, Baekelandt, Schultz, Fernandez, Hensley, Li, Imtiaz, Shingi, Woldu, Williams, Joniau, Gerald, Narayan, Krishna, Lotan.

Critical revision of the manuscript for scientific and factual content: Hamza, Redfern, Joshi, Kamat, Launer, Stewart, Vrabac, Miller, Allison, Hayne, Watson, Eyzaguirre, Shkolyar, Shah, Gordetsky, Taylor, Spalding, Baekelandt, Schultz, Fernandez, O'Donnell, Hensley, Li, Chang, Imtiaz, Shingi, Williams, Joniau, Zuiverloon, Gerald, Packiam, Narayan, Krishna, Nimgaonkar, Abuzeid, Lotan.

Supervision: Hamza, Joshi, Vrabac, Hayne, Watson, Eyzaguirre, Shkolyar, Shah, Gordetsky, Taylor, Schultz, O'Donnell, Hensley, Li, Chang, Imtiaz, Shingi, Sonawane, Woldu, Williams, Narayan, Krishna, Abuzeid, Lotan.

Statistical analysis: Joshi, Launer, Vrabac, Baekelandt, Shingi, Williams, Gerald, Packiam, Krishna, Nimgaonkar, Abuzeid.

Other: Redfern, Kamat, Stewart, Miller, Allison, Spalding, Fernandez, Joniau, Zuiverloon.

Data Availability Statement: The data sets generated and analyzed during the current study are not available.

Corresponding Author: Stephen B. Williams, MD, MBA, MS, Division of Urology, The University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555 (stbwilli@utmb.edu).

* Co-first authors.

† Co-senior authors.

pathologic specimen was evaluated by board-certified genitourinary pathologists at each center. Hematoxylin and eosin (H&E) slides ($\geq$1 per case) representative of the pathologic diagnosis were selected by genitourinary pathologists and converted to digital WSIs at 40$\times$ magnification. WSIs were centrally reviewed by a pathologist for quality and diagnostic accuracy. Exclusion criteria included inadequate pathologic tissue; induction/maintenance treatment with non-BCG intravesical agents after TURBT; and presence of muscle-invasive, node-positive, or metastatic disease.

The study population was stratified into a development cohort (5 centers) for model development and an external validation (EV) cohort (7 centers).[11]

Clinical and Pathologic Features

Clinicopathologic data and BCG treatment details were acquired. When possible, patients were treated with 6 weekly doses of BCG as an induction course, followed by 3 weekly maintenance doses at 3 months and then every 6 months for 3 years. Adequate BCG was defined per the International Bladder Cancer Group (IBCG) and Food and Drug Administration (FDA) as receipt of at least 5 of 6 doses of an induction course followed by either (1) $\geq$ 1 maintenance course (2/3 doses) or (2) $\geq$ 2 of 6 doses of a reinduction course.[12] Disease recurrence was established through cystoscopy with pathologic confirmation. Generally, cystoscopy was performed approximately every 3 months for 2 years and then every 6 months through 5 years. Clinical outcomes were recorded.

Study End Points

The primary end points were high-grade recurrence-free survival (hgRFS) and progression-free survival (PFS). Progression was defined by muscle invasion ($\geq$T2 stage).[13] Secondary end points were cystectomy-free survival (CFS) and time to BUD. Recurrence, progression, and cystectomy end points and time to follow-up were calculated from time of initial TURBT. BUD was defined per IBCG and FDA as (1) persistent/recurrent CIS alone or with recurrent Ta/T1 disease within 12 months of completion of adequate BCG therapy or (2) recurrent high-grade Ta/T1 disease within 6 months of completion of adequate BCG therapy or (3) T1 high-grade disease at first evaluation after a course of induction BCG.[12] For all time-to-event end points, the start time was the date of initial TURBT. Patients were censored at the earliest of "date of death," "date of last follow-up," or "2 years from initial TURBT" (hgRFS and BUD end points) and "5 years from initial TURBT" (PFS and CFS end points).

AI Model Development

The first component of model development was image feature extraction performed by cell and tissue models after cell and tissue segmentation.[14] Deep learning segmentation models were trained with 80:20 development:internal validation splits using a data set of 370,000 nuclei (62 million μm^2 of tissue area) from the development cohort. AUC was evaluated to assess segmentation performance. To train these models, board-certified pathologists—each with 7 to 8 years of experience—segmented nuclei and tissue regions from initial TURBT-derived WSIs in the development cohort (Figure 1). Slide visualization software (QuPath) was used for segmentation and classification annotations. To fine-tune immune cell detection, H&E slides with paired immunohistochemistry (CD68) were used to identify ground-truth cell classifications. Using AI model predictions, 600 features measuring tumor microenvironment characteristics were calculated, including features indicative of tumor aggressiveness, nuclear and cellular pleomorphism, mitotic activity, and immune infiltration.

The second component was predictive model development. Cox proportional hazards (CPH) regression was used to identify key image features strongly associated with each clinical outcome. Separate signatures based on the resulting continuous linear predictor from the CPH regression models were developed for hgRFS, PFS, and time to BUD. WSIs from each case were analyzed for predictive model development. Cutoffs for each linear predictor were selected to allow dichotomization into high-risk and low-risk groups for hgRFS (RR-high, RR-low), PFS (PR-high, PR-low), and presence of BUD signature (BUD+, BUD−). Positive and negative groups of patients were defined as 30:70, 10:90, and 30:70 based on the target of a 2-fold difference in each clinical end point between high-risk and low-risk groups. The PFS model was used for identification of high/low risk of cystectomy. For additional model development details, see Supplementary Materials (https://www.jurology.com).

AI Model Validation

The predefined models for recurrence, progression, and BUD and associated cutoffs were locked before external validation. Each case in the validation cohort consisted of $\geq$ 1 WSI derived from the initial TURBT and correlative clinical data. The outcomes of cases identified as high-risk and low-risk of high-grade recurrence, progression, development of BUD, and cystectomy were compared.

Statistical Analysis

Analyses were performed to identify clinical differences between the development and independent EV cohorts. Medians and IQRs for each continuous clinical characteristic were provided along with P values derived from (Welch) unequal variance t tests for the equality of means. Counts and proportions for binary clinical characteristics and P values derived from 2-sample tests for equality of proportions with continuity correction were provided.

The performance of each predefined AI model was assessed using multivariable CPH regression models and corresponding likelihood ratio (LR) χ^2 statistics to test the null hypothesis that the HR for each (binary) AI model was not equal to 1. A prespecified P value $< .05$ was considered significant. Multivariable CPH regression analyses and LR tests were performed to test the hypothesis that the AI model provided predictive value beyond standard clinical covariates and the European Organization for Research and Treatment of Cancer (EORTC) 2016 and EAU 2021 risk models.[15,16] No adjustment for multiplicity of testing the recurrence-free survival (RFS), PFS, time to BCG unresponsiveness, and CFS end points was made. Plots of the nonparametric maximum likelihood Kaplan-Meier estimates of survival stratified by the predefined low-risk and high-risk groups and associated log-rank tests for each clinical end point are provided. RFS and PFS end points were censored at 2 and 5 years,

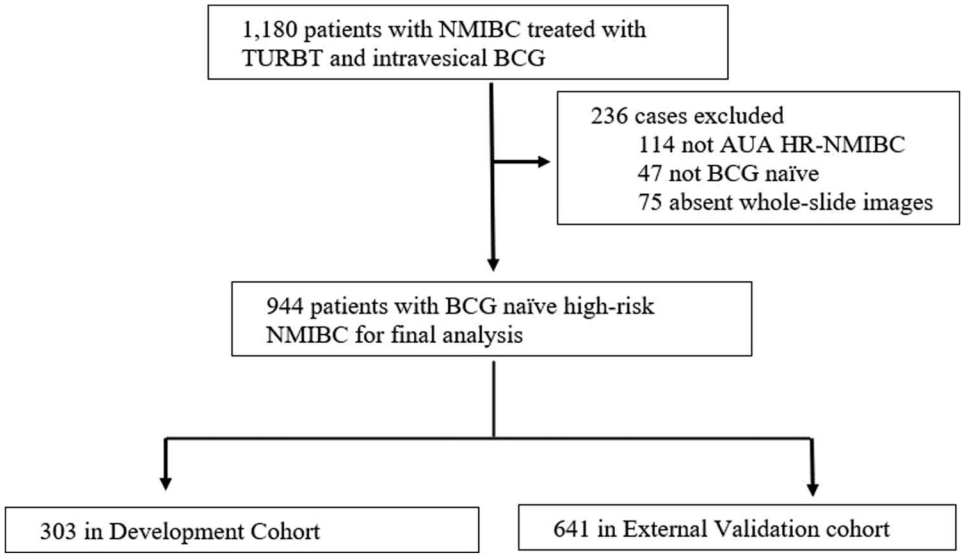

Figure 1. Allocation of an international, multi-institutional study population across developmental and validation cohorts. Nine hundred forty-four patients treated for high-risk (HR) nonmuscle-invasive bladder cancer (NMIBC) with transurethral resection of bladder tumor (TURBT) and intravesical bacillus Calmette-Guérin (BCG) across 12 institutions were allocated across 2 cohorts: development cohort (5 centers, 303 patients) and an international external validation cohort (7 centers, 641 patients).

respectively. In the EV cohort, missing data were imputed to allow for application of the EORTC 2016 (number of tumors and history of recurrence) and EAU 2021 risk models (history of recurrence, size of tumor, and focality) based on additive regression, bootstrapping, and predictive mean matching. Complete case analyses using

Table. *Patient Characteristics Across the Development and Validation Cohorts*

	Development cohort (n = 303)	External validation cohort (n = 641)	Development vs validation cohort (*P* value)
Age, median (IQR), y	71 (64-79)	70 (64-77)	.29
Sex, No. (% male)	248 (81.8)	519 (81.0)	.75
Race, No. (%)			**< .001**
White	271 (89.4)	440 (68.6)	
Black	9 (2.9)	48 (7.5)	
Hispanic	5 (1.7)	51 (8.0)	
Asian	10 (3.3)	11 (1.7)	
Other	2 (0.7)	4 (0.6)	
Unknown	6 (2.0)	87 (13.6)	
Never smoker, No. (%)	93 (30.7)	205 (32.0)	.77
Disease status, No. (%)			**< .001**
Primary	226 (74.6)	408 (63.7)	
Recurrent	19 (6.3)	108 (16.8)	
Unknown	58 (19.1)	125 (19.5)	
Re-TURBT, No. (%)	130 (42.9)	261 (40.7)	.52
T stage, No. (%)			**.02**
Ta	122 (40.3)	199 (31.1)	
Tis	27 (8.9)	78 (12.2)	
T1	154 (50.8)	362 (56.7)	
Focality, No. (%)			.61
Multifocal	139 (45.9)	301 (47.0)	
Unknown	19 (6.3)	48 (7.5)	
CIS, No. (%)	102 (33.7)	194 (30.3)	.29
Variant histology, No. (%)	30 (9.9)	29 (4.5)	**< .001**
Adequate BCG, No. (%)	202 (66.7)	426 (66.5)	.95
BCG unresponsive, No. (%)	48 (15.8)	101 (15.8)	.97
High-grade recurrence at 2 y, No. (%)	88 (29.0)	202 (31.5)	.44
Progression to muscle invasion at 5 y, No. (%)	23 (7.6)	51 (8.0)	.85
Progression per IBCG at 5 y, No. (%)	42 (13.9)	84 (13.1)	.75
Radical cystectomy, No. (%)	40 (13.2)	78 (12.2)	.65

Abbreviations: BCG, bacillus Calmette-Guérin; CIS, carcinoma in situ; IBCG, International Bladder Cancer Group; TURBT, transurethral resection of bladder tumor.
P values represent univariate comparisons between the development cohort and external validation cohort. In the external validation cohort, T stage distribution was pTis alone (12.2%), pTa (31.1%), and pT1 (56.7%) with concomitant CIS present in 30.3% of cases. Adequate BCG was administered in 66.3% of cases, and 15.8% had BCG-unresponsive disease. High-grade recurrence, progression, and cystectomy rates were 31.5%, 13.9%, and 13.1%, respectively. Groups were compared using Wilcoxon rank-sum test, Fisher exact test, and Pearson χ^2 test. Bolded text represents statistically significant *P* values ($P < .05$) when comparing the development and validation cohorts.

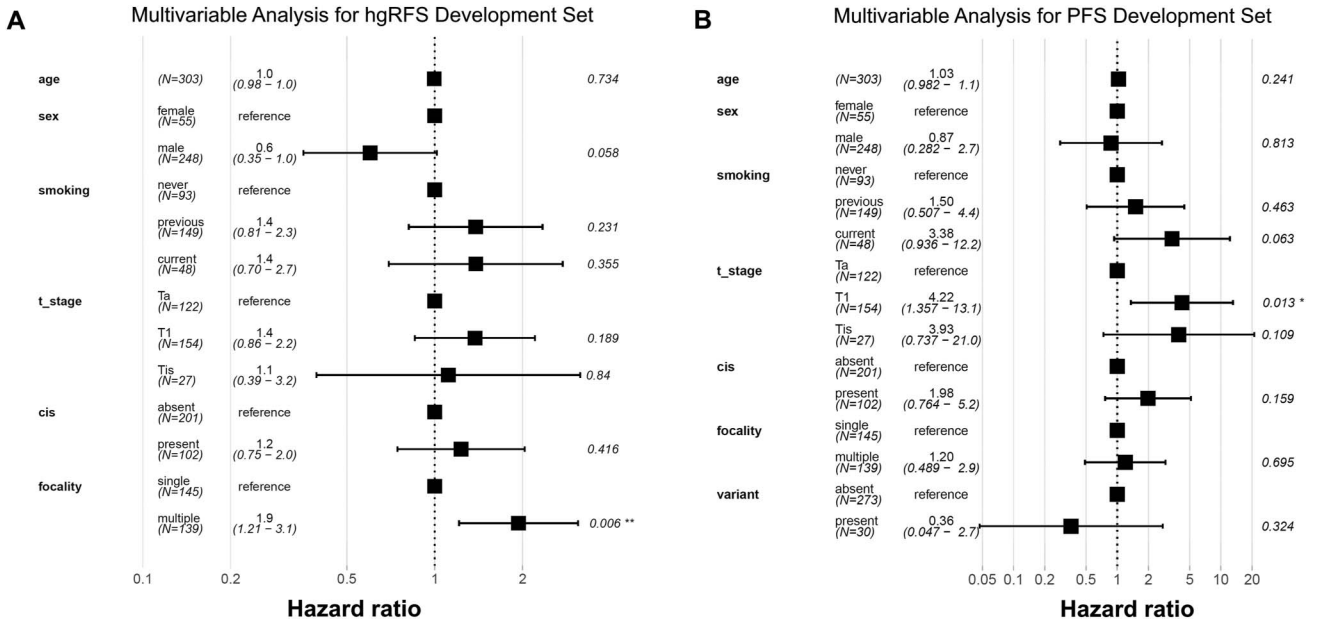

Figure 2. Multivariable Cox proportional hazards regression analysis to determine the strength of association of clinicopathologic factors with high-grade recurrence-free survival (hgRFS) and progression-free survival (PFS) in the development cohort and the relative contribution of these factors to survival risk vs the artificial intelligence model. A, Focality, specifically the presence of multiple tumors on initial transurethral resection of bladder tumor, was significantly associated with hgRFS (HR, 1.9; 95% CI, 1.21-3.1; P = .006). B, Presence of T1 stage disease at initial transurethral resection of bladder tumor was associated with PFS (HR, 4.22; 95% CI, 1.36-13.1; P = .01).

casewise deletion were performed for comparative purposes. See Supplementary Material (https://www.jurology.com) for additional analysis details.

Statistical analysis was performed using R v4.2.0 (R Foundation, Vienna, Austria).

RESULTS

Cohort Characteristics

A total of 1180 BCG-naïve patients with primary or recurrent NMIBC treated with TURBT and BCG were identified across 12 centers. After application of inclusion/exclusion criteria, 944 cases were analyzed (development cohort, n = 303; EV cohort, n = 641; Figure 1).

The development and EV cohorts had median follow-up times of 45.8 and 32.0 months, respectively. Clinicopathologic features are summarized in the Table. The EV cohort differed statistically from the development cohort in racial distribution, fewer primary and more recurrent NMIBC, less Ta disease, and reduced variant histology.

AI Models Benefit From the Addition of Clinicopathologic Factors

To identify factors that could improve AI model performance (Figure 2), multivariable CPH regression and LR χ^2 statistics were performed to determine the association of clinicopathologic factors with hgRFS and PFS in the development cohort. Multifocality was significantly associated with hgRFS (LR P value = .006) and T1 stage associated with PFS (LR P value = .01; Figure 2, A and B) suggesting these factors may be highly prognostic of hgRFS and PFS. No other clinical factors were significantly associated with clinical end points.

To increase prediction accuracy, the AI model for hgRFS incorporated multifocality and the PFS model incorporated T1 status. Using the χ^2 proportion test for clinical parameters, the AI model (χ^2 16.9) contributed 75.1% of the survival risk determination for hgRFS, whereas focality (χ^2 5.6) contributed 24.9% of survival risk calculation. For PFS risk, the AI model (χ^2 4.6) contributed 54.8% of model performance and presence of T1 disease provided 45.2% of performance. The CPH regression models for hgRFS and PFS were used to identify features associated with BUD and risk of cystectomy, respectively.

In the development cohort, the EAU 2021 and EORTC 2016 risk models showed poor prediction of PFS and hgRFS, respectively, as previously reported.[17,18] Consequently, these risk models were not incorporated into the AI models.

Performance of AI Models for hgRFS, PFS, BUD, and CFS End Points in the External Validation Cohort

The AI models for hgRFS, PFS, BUD, and CFS were tested in the EV cohort. The EORTC 2016 recurrence risk (RR) model (EORTC score 1-6) was not

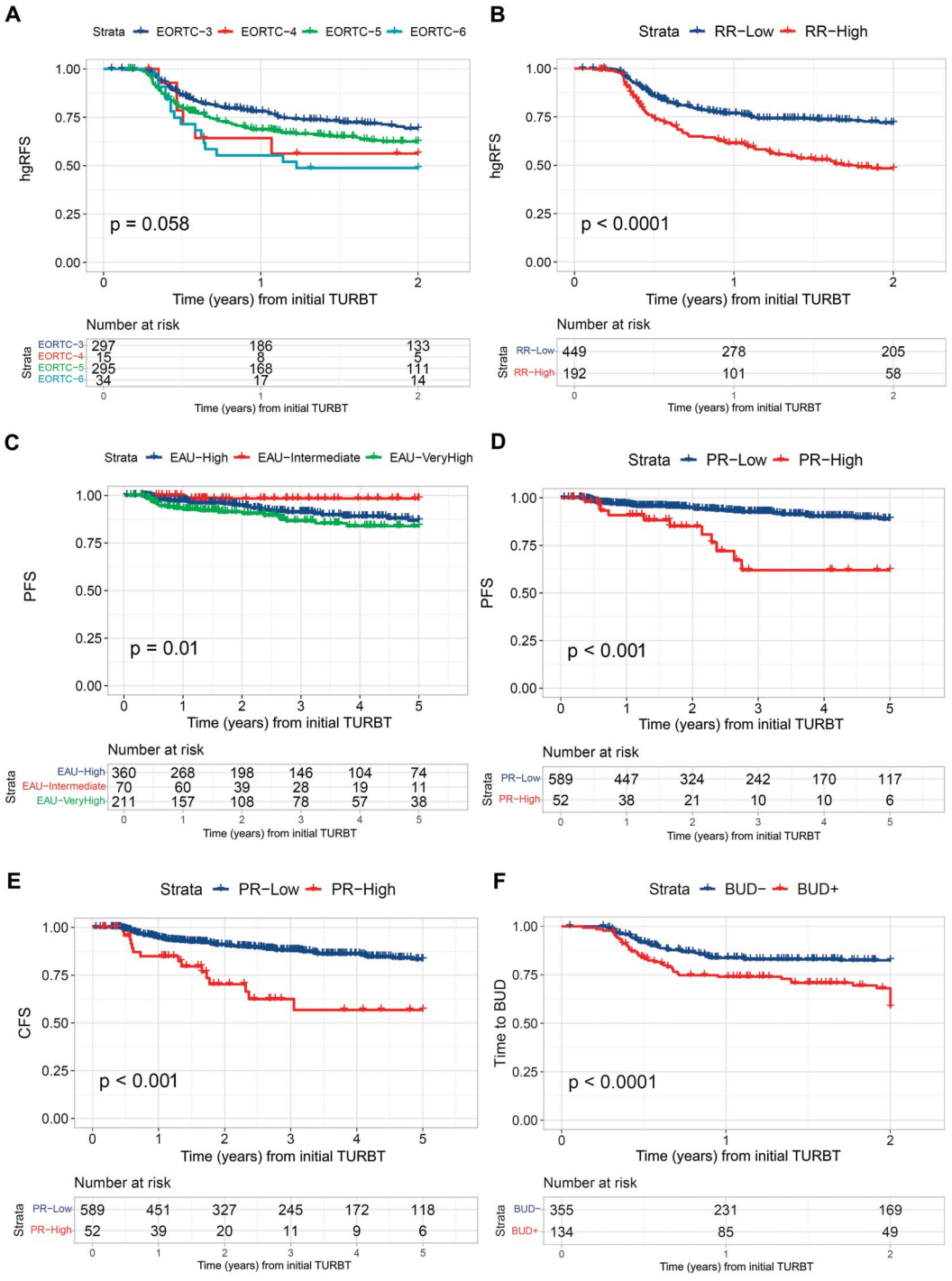

Figure 3. Kaplan-Meier curves demonstrating stratification of patients in the validation cohort into high-risk and low-risk groups for key clinical outcomes by clinicopathologic risk scoring model or by artificial intelligence–enabled model. A, European Organization for Research and Treatment of Cancer (EORTC) 2016 risk does not stratify patients for recurrence risk (RR; baseline: EORTC 3; EORTC 4 HR 1.66, 95% CI 1.09-2.54; EORTC 5 HR 1.35, 95% CI 1.16-1.57; EORTC 6 HR 2.02, 95% CI 1.53-2.65; likelihood ratio $P = .058$). Model instability is suggested by the survival plot for EORTC group 4 crossing over that for group 5. The HRs for recurrence-free survival (RFS) for the ordinal EORTC groups were not monotonically increasing, suggesting a lack of trend. B, The RR model identifies patients at high (RR-high) and low (RR-low) risk of high-grade (hg) recurrence (HR 2.08, 95% CI 1.80-2.40; $P < .0001$). C, The European Association of Urology (EAU) nonmuscle-invasive bladder cancer 2021 risk scoring tool identifies patients at very high, high, and intermediate risk of progression (reference: EAU "high"; EAU "intermediate" HR 0.17, 95% CI 0.06-0.49; EAU "very high" HR 1.51, 95% CI 1.13-2.01; likelihood ratio $P = .01$). D, Comparatively, the progression risk (PR) model identifies patients at high (PR-high) and low (PR-low) risk of progression to muscle invasion with greater separation of risk groups (HR 3.87, 95% CI 2.75-5.44; $P < .001$). E, The PR model was able to identify patients at high and low risk of cystectomy (HR 3.35, 95% CI 2.51-4.47; $P < .001$). F, Presence of the bacillus Calmette-Guérin–unresponsive disease (BUD) histologic signature (BUD+) portends a shorter time to development of BUD than absence of the signature (BUD−) among cases receiving adequate bacillus Calmette-Guérin (HR 2.31, 95% CI 1.89-2.82; $P < .0001$). CFS indicates cystectomy-free survival; PFS, progression-free survival; TURBT, transurethral resection of bladder tumor.

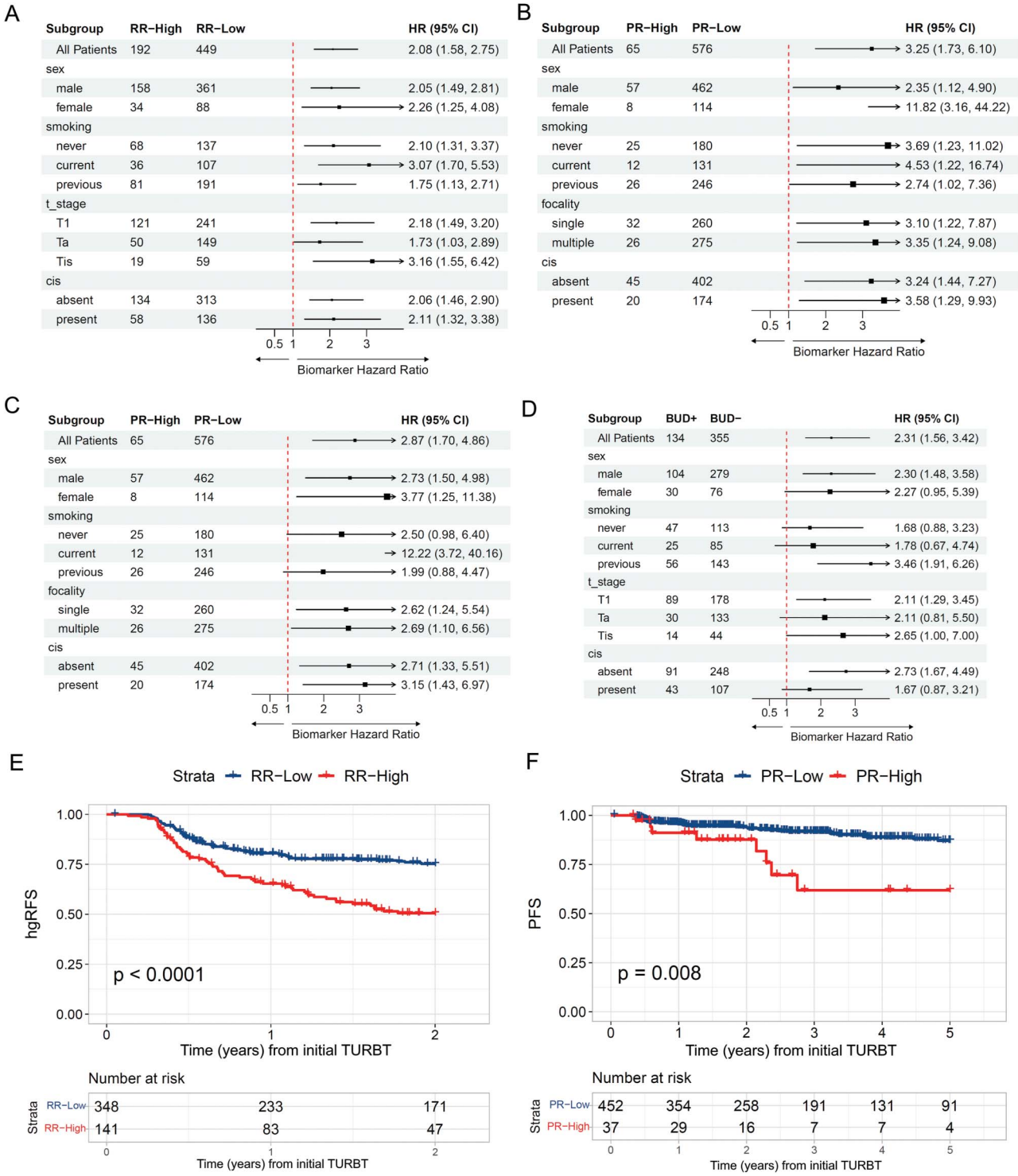

Figure 4. Artificial intelligence model stratifies outcomes across patient subgroups in the validation cohort. A, Recurrence risk (RR) artificial intelligence model performance is demonstrated across different population subgroups with HRs of 1.73 to 3.16 between RR-high and RR-low groups. B, Progression risk (PR) artificial intelligence model performance is shown across population subgroups with HRs of 2.35 to 11.82. C, Risk of cystectomy as determined by the PR model maintains performance across subgroups with HRs of 1.99 to 12.22. D, Time to bacillus Calmette-Guérin–unresponsive disease (BUD) model performance is sustained across subgroups with HRs of 1.67 to 3.46. E, Application of the RR artificial intelligence model to the subgroup of patients treated with Food and Drug Administration–defined adequate bacillus Calmette-Guérin demonstrates continued discriminatory performance between RR-high and RR-low cases (HR 2.23, 95% CI 1.89-2.64; $P < .0001$). F, In the subgroup of patients treated with adequate bacillus Calmette-Guérin, PR-high cases showed significantly poorer progression-free survival (PFS) than PR-low cases (HR 3.24, 95% CI 2.19-4.79; $P = .008$). Note that the sample size reported for the adequate bacillus Calmette-Guérin cohort exceeds the number of cases reported in the external validation cohort for Table because cases that received adequate bacillus Calmette-Guérin and those that had Food and Drug Administration–defined BUD were included. cis indicates carcinoma in situ; hgRFS, high-grade recurrence-free survival; TURBT, transurethral resection of bladder tumor.

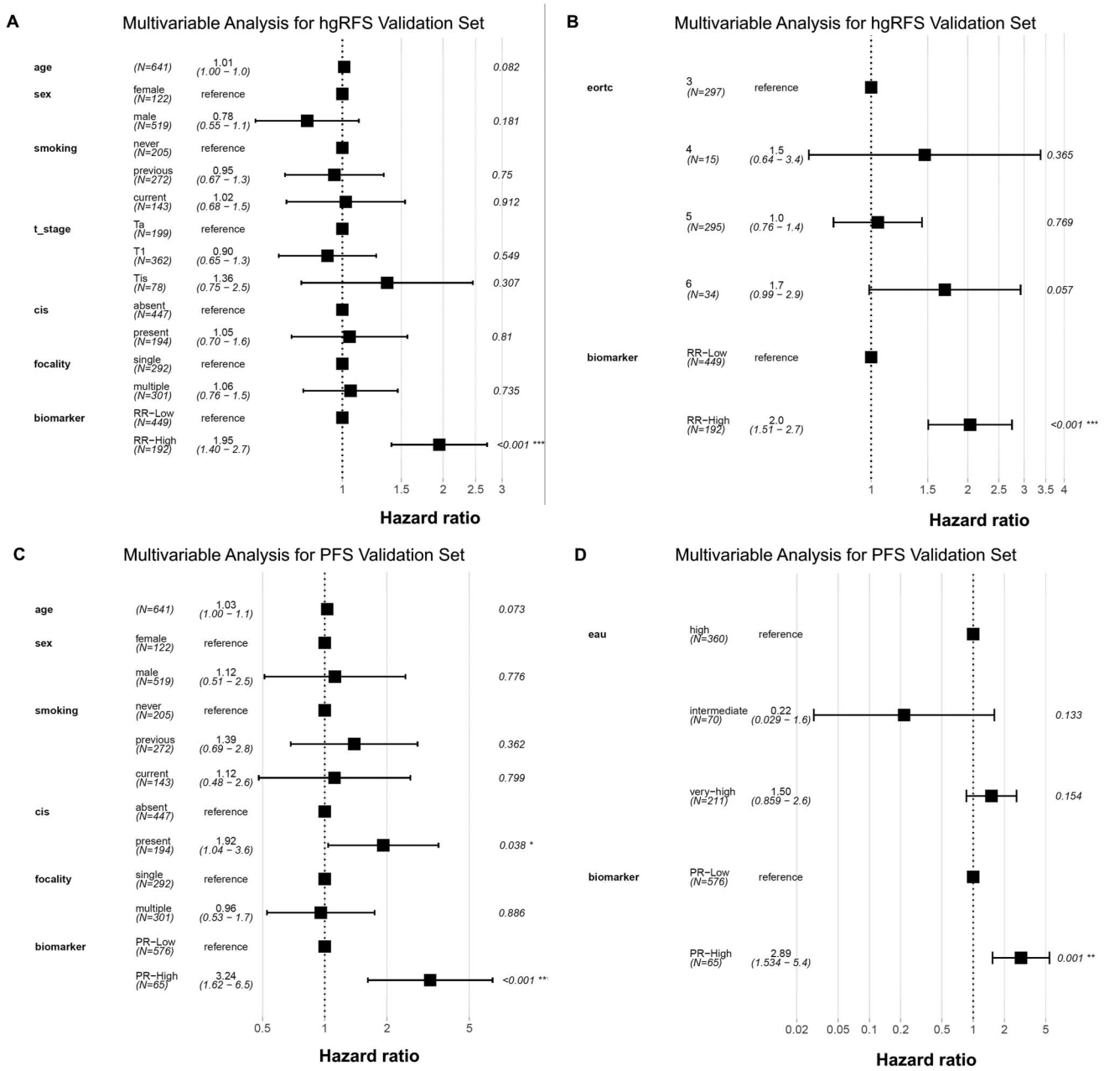

Figure 5. Artificial intelligence models demonstrate predictive performance beyond and independently of clinicopathologic factors in the validation cohort. The recurrence risk (RR) artificial intelligence model identified cases at high risk of high-grade recurrence-free survival (hgRFS) independently of individual clinicopathologic features (HR 1.95, 95% CI 1.40-2.70; $P < .001$; A) and of composite measures including the European Organization for Research and Treatment of Cancer (EORTC) 2016 recurrence score (HR 2.0, 95% CI 1.51-2.70; $P < .001$; B). The progression risk (PR) artificial intelligence model identified patients at high risk of progression-free survival (PFS) independently of individual clinicopathologic risk factors (HR 3.24, 95% CI 1.62-6.50; $P < .001$; C) and composite measures including the European Association of Urology (EAU) 2021 risk scoring model for progression (HR 2.89, 95% CI 1.53-5.40; $P < .001$; D). cis indicates carcinoma in situ.

significantly associated with hgRFS in the EV cohort (LR $P = .058$; Figure 3, A). The HRs for RFS for the ordinal EORTC groups were not monotonically increasing, suggesting a lack of trend. By contrast, using the AI-based model, patients classified as RR-high had significantly poorer hgRFS than RR-low cases (HR, 2.08; 95% CI, 1.80-2.40; $P < .0001$). The RR model identified approximately 6-fold more patients at high risk of recurrence vs EORTC group 6 (Figure 3, B). RR in RR-high vs RR-low cases was 25% vs 14%, 39% vs 23%, and 52% vs 28% at 6, 12, and 24 months, respectively.

The EAU NMIBC 2021 progression risk model stratified patients for PFS in the EV cohort (LR $P = .01$; Figure 3, C) but was outperformed by the PR model. PR-high patients had worse PFS than

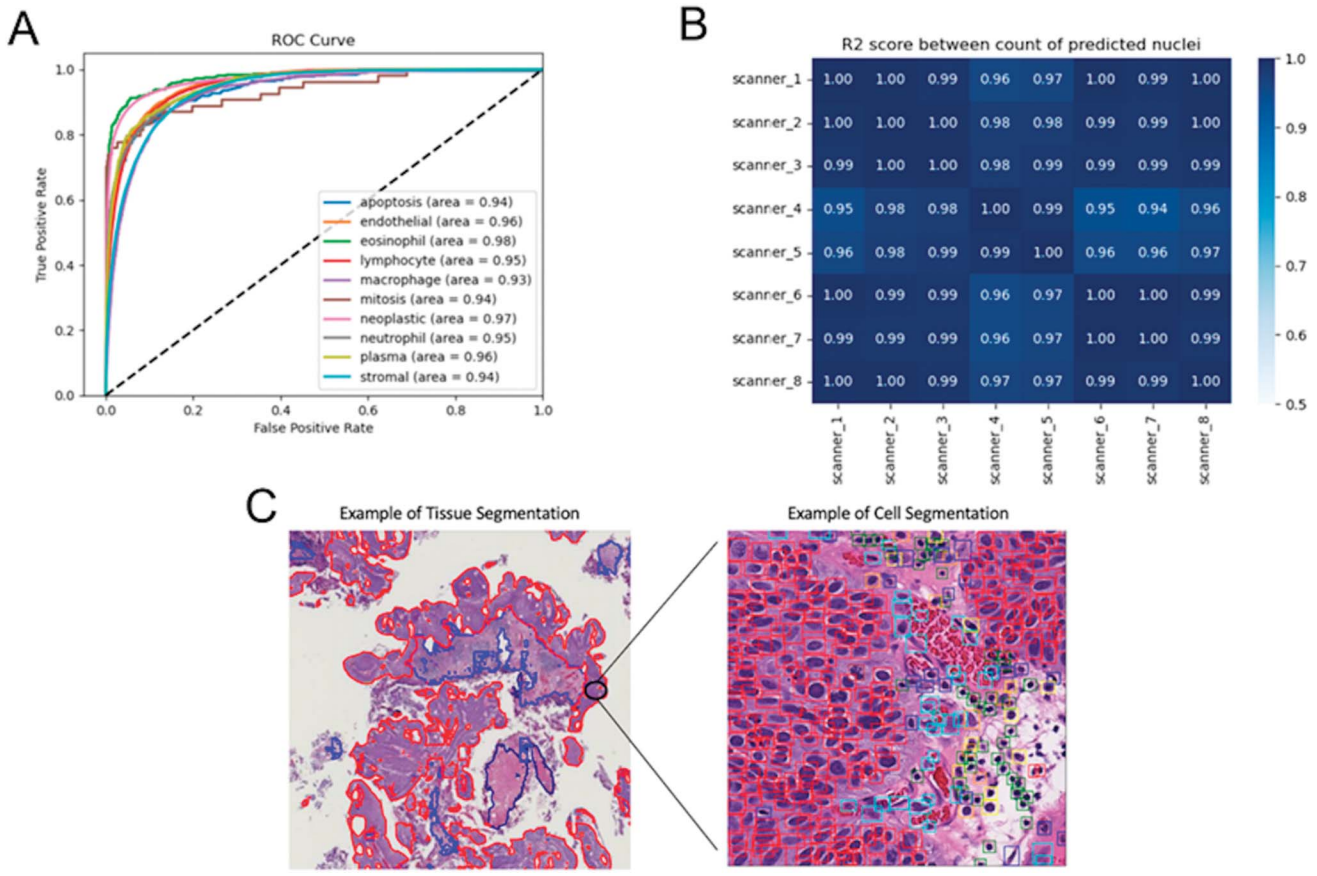

Figure 6. Artificial intelligence–derived segmentation accurately identifies specific cell and tissue types to generate clinically relevant morphologic features. A, Area under the receiver operating characteristic curve of the cell segmentation model for different cell types (using pathologist annotation as the ground truth) derived from the development cohort. B, Correlations of different tissue area predictions by the model when a set of 23 slides derived from single-center data included in the validation cohort were scanned with 8 different scanner models. Whole-slide images were obtained using Leica Aperio AT2, CS2, and GT450 models (Leica, Weltzlar, Germany); 3D Histech Pannoramic 250 and 1000 models (3D Histech, Budapest, Hungary); and Hamamatsu S20 and S360 models (Hamamatsu Photonics, Shizuoka, Japan). C, Representative histologic image from a case in the validation cohort illustrating the tissue segmentation model on a transurethral resection of bladder tumor sample.

PR-low (HR, 3.87; 95% CI, 2.75-5.44; $P < .001$; Figure 3, D). PR-high patients had a higher risk of cystectomy vs PR-low patients (HR, 3.35; 95% CI, 2.51-4.47; $P < .001$; Figure 3, E). BUD+ cases had shorter time to development of BUD than BUD− cases (HR, 2.31; 95% CI, 1.89-2.82; $P < .0001$; Figure 3, F). Similar results were demonstrated between complete case analysis and analysis of cases incorporating imputed data. AI model performance was maintained after adjusting for site-specific variations. Adjusted HRs for the hgRFS, PFS, CFS, and BUD models were 2.10, 3.93, 3.05, and 2.21, respectively.

AI Models Are Significantly Associated With Clinical Outcomes in Clinical Subgroups

To evaluate performance in populations encountered in clinical practice, AI models were applied to patient subgroups stratified by clinicopathologic features (Figure 4, A-D). Recurrence and progression models maintained performance with HRs of 1.73 to 3.16 and 2.35 to 11.82, respectively.

In cases receiving adequate BCG, the RR model continued to identify patients at high/low risk of recurrence (HR, 2.23; 95% CI, 1.89-2.64; $P < .0001$; Figure 4, E) and the PR model differentiated cases at high/low risk of progression (HR, 3.24; 95% CI, 2.19-4.79; $P = .008$; Figure 4, F) suggesting that receipt of adequate BCG does not affect performance.

AI Models Demonstrate Predictive Performance Beyond and Independently of Clinicopathologic Factors

Multivariable CPH regression for hgRFS was performed comparing the (full) main-effects model with the AI predictor for RFS and clinicopathologic features vs the (reduced) main-effects model excluding the AI predictor. A P value for the corresponding LR χ^2 statistic for the (binary) AI model was considered significant. The RR model identified cases at high/low risk of hgRFS independently of individual clinicopathologic features (HR, 1.95; 95% CI, 1.40-2.70; LR $P < .001$). Similarly, the AI model for hgRFS

provided predictive information beyond the EORTC 2016 model (HR, 2.0; 95% CI, 1.51-2.70; LR $P < .001$; Figure 5, A and B).

The PR model identified patients at high/low risk of PFS independently of individual clinicopathologic risk factors (HR, 3.24; 95% CI, 1.62-6.50; LR $P < .001$) and the EAU 2021 progression model (HR, 2.89; 95% CI, 1.53-5.4; LR $P < .001$; Figure 5, C and D). Neither the EORTC 2016 nor EAU 2021 model independently discriminated patients at high risk of recurrence or progression, respectively.

AI Models Are Robust to Slide Scanner Model and Magnification

At the cell and tissue levels, relative to the ground truth of pathologist annotation, the image feature extraction AI model generated AUC values of 0.98 and 0.99 for segmentation and classification tasks, respectively (Figure 6, A). Robustness of the image feature extraction model was evaluated by correlating model predictions across different scanner and magnification configurations. The AI model demonstrated high correlation at 20× and 40× magnification ($r = 0.99$, $P < .001$) and across 8 digital scanner models ($P < .001$; Figure 6, B). This generalizability was robust for predicted nuclei counts, and high correlation across configurations was demonstrated for individual cell types. Segmentation of a representative WSI is illustrated (Figure 6, C).

DISCUSSION

In this study, we developed and validated AI-based histologic assays that analyzed pretreatment TURBT-derived WSIs from a multicenter cohort of HR-NMIBC spanning 12 international academic centers. These assays accurately identified cases with an elevated risk of recurrence, progression, development of BUD, and cystectomy.

The identification of patients with HR-NMIBC unlikely to respond favorably to intravesical BCG is an unmet need that may facilitate more effective therapy selection and clinical trial enrollment. To date, there is no accurate, reproducible, and convenient means to identify patients more likely to recur or progress with BCG therapy.[19] Indeed, clinical risk prediction tools including the EORTC and Club Urologico Espanol de Tratamiento Oncologico (CUETO) risk calculators significantly overestimate the risks of recurrence and progression in BCG-treated patients likely because of their development cohorts that lacked BCG-treated patients in the former and used an abbreviated BCG maintenance course in the latter.[18,20-22] In bladder cancer, AI-based WSI analysis has been used to grade disease, to predict the likelihood of positive lymph nodes from analysis of the primary tumor histology, and to prognosticate overall survival.[23-25]

This is the first study of 4 AI-driven histologic assays developed on the CHAI platform for determining risk of (1) recurrence, (2) progression, (3) BUD, and (4) cystectomy in HR-NMIBC. To mitigate the risk of overfitting, we describe the external validation of model performance. The AI-enabled assays identified patients with a 2.1, 3.9, 2.3, and 3.4-fold higher risk of recurrence, progression, BUD, and cystectomy, respectively.

Both the RR and PR AI-enabled models stratified patients into high-risk and low-risk groups independently of individual clinicopathologic features and of EORTC and EAU risk models. Indeed, the EORTC 2016 and EAU 2021 models demonstrated poorer predictive ability in identifying patients at high risk of recurrence and progression, respectively, than the RR and PR AI-enabled assays. In the EAU model, this poorer predictive ability may be because the development cohort for the model excluded cases treated with intravesical BCG potentially compromising prognostic capability in BCG-treated patients.[15]

The RR and PR models integrated multifocality and presence of T1 disease, respectively, as clinical features that enhanced performance beyond purely AI-based histologic analysis. Nevertheless, most of the predictive performance was derived from AI-based histologic analysis. The resulting assays sustained performance across patient subgroups including smoking history, T stage, focality, and presence of CIS. Critically, the RR and PR models demonstrated no performance degradation (HR, 2.2 and 3.2, respectively) in a subgroup receiving adequate BCG therapy despite the known improvement in survival outcomes with adequate BCG.[26] This is important because existing risk calculators such as EAU 2021 overestimate progression risk in patients receiving adequate BCG.[22]

The CHAI platform leverages existing clinical workflows by analyzing digitized H&E-stained diagnostic slides from pre-BCG TURBT specimens. By better understanding an individual's risk of recurrence, progression, developing BUD, and cystectomy—before the initiation of intravesical BCG therapy—clinicians are armed with nuanced data to help guide treatment selection and patient counseling regarding alternative treatment approaches, such as intravesical chemotherapy.[27] For example, presence of the BUD biomarker signature or stratification into the PR-high group may prompt clinicians to try efficacious salvage treatments in patients who recur on BCG and who decline radical cystectomy.[28,29] This may also have important implications when novel therapies or therapeutic combinations become available because the risk/benefit balance could be influenced by the

availability of tools to identify cases less likely to benefit from BCG alone. Pretreatment risk stratification may have implications for tumor surveillance regimens. The robustness of the CHAI platform to differences in scanner models widely used across academia and private practice enhances potential generalizability.

Our study has limitations. Although the CHAI platform was validated on a large international cohort intended to reproduce real-world heterogeneity in disease presentation and treatment, optimal validation would use a prospective clinical trial to mitigate confounding and bias. There was notable heterogeneity in the duration of BCG treatment with 66% receiving FDA/IBCG-defined "adequate BCG." Although this may be representative of real-world challenges because of BCG intolerance or drug supply shortages, heterogeneity in treatment may confound disease outcomes. However, the cohort size likely mitigated some of the potential confounding as further evidenced by the stability of model performance between the development and validation cohorts. Not all patients who were pT1 underwent re-TURBT, potentially confounding oncologic outcomes in this subgroup.

CONCLUSIONS

An AI-based approach was used to identify digital pathology signatures of adverse oncologic outcomes, and the resulting AI-based histology assays developed and validated across an international cohort of HR-NMIBC cases. The assays analyze pre-BCG TURBT specimens and can identify patients with HR-NMIBC who are at elevated risk of recurrence, progression, development of BUD, and cystectomy independently of clinicopathologic features. The CHAI platform may facilitate patient counseling and treatment selection in patients with HR-NMIBC.

ACKNOWLEDGMENTS

We thank all participants of the study. Valar Labs was involved in funding portions of this study related to pathologic slide digitization and clinical data acquisition. Brian Liao, BS (University of Texas Medical Branch); Chia-Sui Kao, MD (Cleveland Clinic); Dattatraya Patil, MBBS (Emory University); Anand Rajan, MD (University of Iowa); Stephanie Chen, MD (University of Iowa); Joon Kyung Kim, MD (University of Kentucky); Ekin Tiu, BS (Valar Labs); and Louis J. Vaickus, MD, PhD (Valar Labs and Dartmouth-Hitchcock Medical Center), contributed to data collection or analysis.

REFERENCES

1. Lenis AT, Lec PM, Chamie K, Mshs MD. Bladder cancer: a review. *JAMA*. 2020;324(19):1980-1991. doi:10.1001/jama.2020.17598
2. Chang SS, Boorjian SA, Chou R, et al. Diagnosis and treatment of non-muscle invasive bladder cancer: AUA/SUO guideline. *J Urol*. 2016;196(4):1021-1029. doi:10.1016/j.juro.2016.06.049
3. Babjuk M, Burger M, Capoun O, et al. European Association of Urology guidelines on non-muscle-invasive bladder cancer (Ta, T1, and carcinoma in situ). *Eur Urol*. 2022;81(1):75-94. doi:10.1016/j.eururo.2021.08.010
4. Flaig TW, Spiess PE, Abern M, et al. NCCN Guidelines® Insights: bladder cancer, version 2.2022. *J Natl Compr Canc Netw*. 2022;20(8):866-878. doi:10.6004/jnccn.2022.0041
5. Oddens J, Brausi M, Sylvester R, et al. Final results of an EORTC-GU cancers group randomized study of maintenance bacillus Calmette-Guerin in intermediate- and high-risk Ta, T1 papillary carcinoma of the urinary bladder: one-third dose versus full dose and 1 year versus 3 years of maintenance. *Eur Urol*. 2013;63(3):462-472. doi:10.1016/j.eururo.2012.10.039
6. Harvey M, Chislett B, Perera M, Lawrentschuk N, Bolton D, Jack G. Critical shortage in BCG immunotherapy: how did we get here and where will it take us?. *Urol Oncol*. 2022;40(1):1-3. doi:10.1016/j.urolonc.2021.09.022
7. Kates M, Chu X, Hahn N, et al. Background and update for ECOG-ACRIN EA8212: a randomized phase 3 trial of intravesical Bacillus Calmette-Guerin (BCG) versus intravesical docetaxel and gemcitabine treatment in BCG-naive high-grade non-muscle-invasive bladder cancer (BRIDGE). *Eur Urol Focus*. 2023;9(4):561-563. doi:10.1016/j.euf.2023.06.006
8. Vrabac D, Smit A, Rojansky R, et al. DLBCL-Morph: morphological features computed using deep learning for an annotated digital DLBCL image set. *Sci Data*. 2021;8(1):135. doi:10.1038/s41597-021-00915-w
9. Nimgaonkar V, Krishna V, Krishna V, et al. Development of an artificial intelligence-derived histologic signature associated with adjuvant gemcitabine treatment outcomes in pancreatic cancer. *Cell Rep Med*. 2023;4(4):101013. doi:10.1016/j.xcrm.2023.101013
10. McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer*. 2005;93(4):387-391. doi:10.1038/sj.bjc.6602678
11. Taylor JM, Ankerst DP, Andridge RR. Validation of biomarker-based risk prediction models. *Clin Cancer Res*. 2008;14(19):5977-5983. doi:10.1158/1078-0432.CCR-07-4534
12. Kamat AM, Apolo AB, Babjuk M, et al. Definitions, end points, and clinical trial designs for bladder cancer: recommendations from the society for immunotherapy of cancer and the International Bladder Cancer Group. *J Clin Oncol*. 2023;41(35):5437-5447. doi:10.1200/JCO.23.00307
13. Lamm D, Persad R, Brausi M, et al. Defining progression in nonmuscle invasive bladder cancer: it is time for a new, standard definition. *J Urol*. 2014;191(1):20-27. doi:10.1016/j.juro.2013.07.102
14. Goode A, Gilbert B, Harkes J, Jukic D, Satyanarayanan M. OpenSlide: a vendor-neutral software foundation for digital pathology. *J Pathol Inform*. 2013;4(1):27. doi:10.4103/2153-3539.119005
15. Sylvester RJ, Rodriguez O, Hernandez V, et al. European Association of Urology (EAU) prognostic factor risk groups for non-muscle-invasive bladder cancer (NMIBC) incorporating the WHO 2004/2016 and WHO 1973 classification systems for grade: an update from the EAU NMIBC guidelines panel. *Eur Urol*. 2021;79(4):480-488. doi:10.1016/j.eururo.2020.12.033
16. Cambier S, Sylvester RJ, Collette L, et al. EORTC nomograms and risk groups for predicting recurrence, progression, and disease-specific and overall survival in non-muscle-invasive stage Ta-T1 urothelial bladder cancer patients treated with 1-3 years of maintenance Bacillus Calmette-Guerin. *Eur Urol*. 2016;69(1):60-69. doi:10.1016/j.eururo.2015.06.045

17. Krajewski W, Aumatell J, Subiela JD, et al. Accuracy of the CUETO, EORTC 2016 and EAU 2021 scoring models and risk stratification tables to predict outcomes in high-grade non-muscle-invasive urothelial bladder cancer. *Urol Oncol.* 2022;40(11):491.e11-491.e19. doi:10.1016/j.urolonc.2022.06.008

18. Xylinas E, Kent M, Kluth L, et al. Accuracy of the EORTC risk tables and of the CUETO scoring model to predict outcomes in non-muscle-invasive urothelial carcinoma of the bladder. *Br J Cancer.* 2013;109(6):1460-1466. doi:10.1038/bjc.2013.372

19. Zuiverloon TC, Nieuweboer AJ, Vekony H, Kirkels WJ, Bangma CH, Zwarthoff EC. Markers predicting response to bacillus Calmette-Guerin immunotherapy in high-risk bladder cancer patients: a systematic review. *Eur Urol.* 2012;61(1):128-145. doi:10.1016/j.eururo.2011.09.026

20. Vedder MM, Marquez M, de Bekker-Grob EW, et al. Risk prediction scores for recurrence and progression of non-muscle invasive bladder cancer: an international validation in primary tumours. *PLoS One.* 2014;9(6):e96849. doi:10.1371/journal.pone.0096849

21. Jobczyk M, Stawiski K, Fendler W, Rozanski W. Validation of EORTC, CUETO, and EAU risk stratification in prediction of recurrence, progression, and death of patients with initially non-muscle-invasive bladder cancer (NMIBC): a cohort analysis. *Cancer Med.* 2020;9(11):4014-4025. doi:10.1002/cam4.3007

22. Lobo N, Hensley PJ, Bree KK, et al. Updated European Association of Urology (EAU) prognostic factor risk groups overestimate the risk of progression in patients with non-muscle-invasive bladder cancer treated with Bacillus Calmette-Guerin. *Eur Urol Oncol.* 2022;5(1):84-91. doi:10.1016/j.euo.2021.11.006

23. Barrios W, Abdollahi B, Goyal M, et al. Bladder cancer prognosis using deep neural networks and histopathology images. *J Pathol Inform.* 2022;13:100135. doi:10.1016/j.jpi.2022.100135

24. Mundhada A, Sundaram S, Swaminathan R, Lawrence DC, Govindarajan S, Makaram N. Differentiation of urothelial carcinoma in histopathology images using deep learning and visualization. *J Pathol Inform.* 2023;14:100155. doi:10.1016/j.jpi.2022.100155

25. Harmon SA, Sanford TH, Brown GT, et al. Multiresolution application of artificial intelligence in digital pathology for prediction of positive lymph nodes from primary tumors in bladder cancer. *JCO Clin Cancer Inform.* 2020;4:367-382. doi:10.1200/CCI.19.00155

26. Subiela JD, Krajewski W, Gonzalez-Padilla DA, et al. Unlocking the potential of adequate Bacillus Calmette-Guerin immunotherapy in very-high-risk non-muscle-invasive bladder carcinoma: a multicenter analysis of oncological outcomes and risk dynamics. *Eur Urol Oncol.* Published online February 13, 2024. doi:10.1016/j.euo.2024.01.017

27. McElree IM, Steinberg RL, Mott SL, O'Donnell MA, Packiam VT. Comparison of sequential intravesical gemcitabine and docetaxel vs Bacillus Calmette-Guerin for the treatment of patients with high-risk non-muscle-invasive bladder cancer. *JAMA Netw Open.* 2023;6(2):e230849. doi:10.1001/jamanetworkopen.2023.0849

28. Racioppi M, Di Gianfrancesco L, Ragonese M, Palermo G, Sacco E, Bassi PF. Electromotive drug administration (EMDA) of mitomycin C as first-line salvage therapy in high risk "BCG failure" non muscle invasive bladder cancer: 3 years follow-up outcomes. *BMC Cancer.* 2018;18(1):1224. doi:10.1186/s12885-018-5134-7

29. Chevuru PT, McElree IM, Mott SL, Steinberg RL, O'Donnell MA, Packiam VT. Long-term follow-up of sequential intravesical gemcitabine and docetaxel salvage therapy for non-muscle invasive bladder cancer. *Urol Oncol.* 2023;41(3):148.e1-148.e7. doi:10.1016/j.urolonc.2022.10.030

EDITORIAL COMMENT

High-risk (HR) nonmuscle-invasive bladder cancer (NMIBC) is associated with higher risk of recurrence and progression compared with NMIBC. Bacillus Calmette-Guérin (BCG) is commonly used as first-line therapy and has been shown to reduce recurrence and progression in patients with HR-NMIBC.[1] Predicting response to BCG after transurethral resection of bladder tumor is crucial to avoid delays in surgery decisions and to conserve resources, especially considering current BCG shortages. In this multicenter study, Lotan et al developed and validated artificial intelligence (AI)–based histologic assays to predict recurrence, progression, BCG-unresponsive disease, and cystectomy in patients with HR-NMIBC treated with BCG.[2] Using whole-slide pathology images from 944 cases (303 development, 641 validation) across 12 centers, features were extracted to identify clinical outcome markers. In the validation cohort, "high recurrence risk" patients had significantly worse recurrence-free survival compared with "low recurrence risk" cases (HR, 2.08; $P < .0001$). Similarly, "high progression risk" patients had poorer progression-free survival (HR, 3.87; $P < .001$) and a higher likelihood of cystectomy (HR, 3.35; $P < .001$). Patients with the BCG-unresponsive disease signature experienced a more rapid onset of BCG-unresponsive disease compared with those without the signature (HR, 2.31; $P < .0001$).[2] The novel AI assays provided predictive information beyond traditional clinicopathologic factors, offering valuable insights for improving clinical decision-making in HR-NMIBC management. In addition, from an AI research perspective, the study's large, multicenter population and their ability to identify pathology features associated with adverse oncologic outcomes are critical for translational AI practice. Further validation of this prognostic AI model in larger cohorts will be key to confirming its clinical utility.

Baris Turkbey [1]
[1]*National Cancer Institute, National Institutes of Health Bethesda, Maryland*

REFERENCES

1. Malmström PU, Sylvester RJ, Crawford DE, et al. An individual patient data meta-analysis of the long-term outcome of randomised studies comparing intravesical mitomycin C versus bacillus Calmette-Guérin for non-muscle-invasive bladder cancer. *Eur Urol.* 2009;56(2):247-256. doi:10.1016/j.eururo.2009.04.038

2. Lotan Y, Krishna V, Abuzeid WM, et al. Predicting response to intravesical bacillus Calmette-Guérin in high-risk nonmuscle-invasive bladder cancer using an artificial intelligence–powered pathology assay: development and validation in an international 12-center cohort. *J Urol.* 2025;213(2):192-204. doi:10.1097/ju.0000000000004278