The UROMOL dataset was preprocessed by removing rows with missing values. Clinical features included `Age`, `Sex`, `Concomitant.CIS`, `BCG`, and `UROMOL2021.classification`, which were one-hot encoded. Gene expression data, common between UROMOL (RNA-seq) and Knowles (microarray) cohorts, were reduced using Principal Component Analysis (PCA) with `n\_components=3`, determined via hyperparameter tuning. PCA components were concatenated with clinical features to form the training dataset. A Cox Proportional Hazards (CoxPH) model with elastic net regularization was developed using the `lifelines` library. Hyperparameters were tuned via 5-fold cross-validation (CV) on UROMOL, optimizing the concordance index (C-index), yielding `n\_components=3`, `penalizer=0.1`, and `l1\_ratio=0.9`. Patients were stratified into high- and low-risk groups based on the median risk score. Performance was evaluated using the C-index, Kaplan-Meier curves with log-rank tests, and calibration plots at 20, 40, 60, and 90 months. Internal validation used 5-fold CV on UROMOL, and external validation was conducted on the Knowles cohort. Source code is available at [https://github.com/yiyang2002/520-A4.git].

Internal Validation (UROMOL Cohort): The CoxPH model achieved a 5-fold CV C-index of 0.57 ± 0.04, indicating moderate discriminatory ability. KM curves (Figure 1) showed significant separation between high- and low-risk groups (log-rank p = 0.00046), with the high-risk group exhibiting a steeper RFS decline. Calibration plots (Figure 2) demonstrated reasonable alignment between predicted and observed survival probabilities, improving at later time points (Integrated Calibration Index [ICI] at 90 months = 0.017).

External Validation (Knowles Cohort): The model attained a C-index of 0.67, reflecting robust generalizability despite the Knowles cohort’s lower risk profile and limited clinical data. KM curves (Figure 1) indicated a trend toward significance (log-rank p = 0.087), with the high-risk group showing a sharper RFS decline. Calibration plots (Figure 2) yielded ICI values of 0.152, 0.084, 0.103, and 0.017 at 20, 40, 60, and 90 months, respectively, confirming good calibration, particularly at 90 months. BCG treatment emerged as the only significant predictor (hazard ratio = 1.56, p = 0.05), increasing recurrence hazard by 56%.

Clinicians can utilize this risk score to classify patients into high- and low-risk groups for recurrence-free survival, using the median as a threshold. Patients with scores above the median are identified as “high-risk,” indicating an increased likelihood of cancer recurrence, whereas those scoring below are designated as “low-risk”. This enhanced stratification helps clinicians make more precise decisions regarding follow-up care and treatment options. Specifically, patients categorized as intermediate-risk by EAU guidelines[[1]](#endnote-1) but identified as high-risk by the classifier might benefit from more intensive monitoring, such as cystoscopy every 3–6 months during the initial two years instead of annual checks. Additional treatments, like Bacillus Calmette-Guérin (BCG), could also be considered, though the classifier associates BCG treatment with a higher recurrence risk (hazard ratio = 1.56). Clinicians should interpret this carefully, as the elevated risk likely reflects treatment selection bias—BCG is typically administered to patients already at higher risk rather than causing recurrence directly[[2]](#endnote-2). Therefore, clinicians must thoroughly evaluate factors like the presence of carcinoma in situ (CIS) before making treatment decisions. Conversely, patients classified as intermediate-risk by the EAU but identified as low-risk by the classifier might appropriately undergo less frequent monitoring, such as annual cystoscopy, which can help reduce unnecessary invasive procedures, minimize patient anxiety, and optimize resource allocation. For instance, a 60-year-old male patient with intermediate-risk bladder cancer who previously received BCG treatment might require cystoscopy every six months and potentially reconsider additional BCG treatment if categorized as high-risk by the classifier. Alternatively, a 45-year-old female patient in the same EAU intermediate-risk category but classified as low-risk by the classifier could reasonably follow a less rigorous monitoring schedule, supplemented by lifestyle recommendations such as smoking cessation. Despite the classifier’s clinical usefulness, its moderate predictive performance (C-index: internal 0.57, external 0.67) and marginal external validation results (log-rank p-value = 0.087, not statistically significant at 0.05) highlight its role as a complementary tool rather than a standalone method.

A graph of different colored lines

AI-generated content may be incorrect.A graph of hazard function

AI-generated content may be incorrect.A graph of different colored lines

AI-generated content may be incorrect.A graph of hazard function

AI-generated content may be incorrect.

Figure 1: **Kaplan–Meier Survival Curves**: The first row presents survival analysis plots comparing high- and low-risk patient groups as defined by the median risk score from the Cox model. Panel (1) shows Kaplan–Meier survival curves for the internal UROMOL cohort, where the high-risk group demonstrates significantly reduced recurrence-free survival over time, with clear separation between curves and 95% confidence intervals indicated. Panel (2) displays the corresponding Nelson–Aalen cumulative hazard functions, where the high-risk group accumulates hazard more rapidly, reinforcing the stratification. Panels (3) and (4) replicate these analyses on the external Knowles dataset. Although the separation in Kaplan–Meier curves is less distinct in (3), high-risk patients still show lower survival probabilities. Similarly, the cumulative hazard plot in (4) confirms faster hazard accumulation in the high-risk group, supporting the external validity of the model’s predictive capability.

A close-up of a graph

AI-generated content may be incorrect.

Figure 2: The second row illustrates model calibration across four survival time points using calibration plots. At 20 months (t₀=20), the model shows moderate calibration, with an Integrated Calibration Index (ICI) of 0.152 and a median absolute error (E50) of 0.148, indicating some deviation from the ideal prediction line. At 40 months (t₀=40), calibration improves, reflected by a lower ICI of 0.084 and E50 of 0.091, suggesting better alignment between predicted and observed recurrence probabilities. The model maintains reasonable accuracy at 60 months (t₀=60), with ICI = 0.103 and E50 = 0.110. Finally, at 90 months (t₀=90), calibration is excellent; the smoothed calibration curve closely follows the diagonal, with ICI dropping to 0.017 and E50 to 0.021, indicating high reliability in long-term predictions.

1. Babjuk, M. *et al.* European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (Ta, T1, and Carcinoma in Situ). *Eur. Urol.* **81**, 75–94 (2022). [↑](#endnote-ref-1)
2. Liu, C.-Y. *et al.* Maintenance bacillus Calmette–Guérin therapy prolongs recurrence-free survival in non-muscle-invasive bladder cancer: A real-world experience. *Urol. Sci.* **26**, 96–100 (2015). [↑](#endnote-ref-2)