EUROPEAN UROLOGY FOCUS xxx (xxxx) xxx–xxx

available at www.sciencedirect.com
journal homepage: www.europeanurology.com/eufocus

Urothelial Cancer

Machine Learning and External Validation of the IDENTIFY Risk Calculator for Patients with Haematuria Referred to Secondary Care for Suspected Urinary Tract Cancer

Sinan Khadhouri[a,b,*], *Artsiom Hramyka*[c], *Kevin Gallagher*[b,d,e], *Alexander Light*[b,f], *Simona Ippoliti*[b,g], *Marie Edison*[b,h], *Cameron Alexander*[b,i], *Meghana Kulkarni*[b,j], *Eleanor Zimmermann*[b,k], *Arjun Nathan*[b,e], *Luca Orecchia*[l], *Ravi Banthia*[m], *Pietro Piazza*[n], *David Mak*[o], *Nikolaos Pyrgidis*[p], *Prabhat Narayan*[q], *Pablo Abad Lopez*[r], *Faisal Nawaz*[s], *Trung-Thanh Tran*[t,u], *Francesco Claps*[v], *Donnacha Hogan*[w], *Juan Gomez Rivas*[x], *Santiago Alonso*[x], *Ijeoma Chibuzo*[y], *Beatriz Gutierrez Hidalgo*[z], *Jessica Whitburn*[aa], *Jeremy Teoh*[bb], *Gautier Marcq*[cc,dd], *Alexandra Szostek*[cc], *Jasper Bondad*[ee], *Petros Sountoulides*[ff], *Tom Kelsey*[c], *Veeru Kasivisvanathan*[b,gg], *IDENTIFY Study Group*[†]

[a] *School of Medicine, University of St Andrews, St Andrews, UK;* [b] *British Urological Researchers in Surgical Training (BURST), London, UK;* [c] *School of Computer Science, University of St Andrews, St Andrews, UK;* [d] *Institute of Cancer and Genetics, University of Edinburgh, Edinburgh, UK;* [e] *Division of Surgery and Interventional Science, University College London, London, UK;* [f] *Department of Surgery and Cancer, Imperial College London, London, UK;* [g] *Department of Paediatric Surgery, Hull Royal Infirmary, Hull University Teaching Hospitals, Hull, UK;* [h] *Department of Urology, Chelsea and Westminster Hospital, London, UK;* [i] *Luton and Dunstable University Hospital, Luton, UK;* [j] *Department of Urology, St George's University Hospitals NHS Foundation Trust, London, UK;* [k] *Department of Urology, Southmead Hospital, Bristol, UK;* [l] *AOU Policlinico Tor Vergata University Hospital of Rome, Rome, Italy;* [m] *University Hospital Coventry Warwickshire, Coventry, UK;* [n] *IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy;* [o] *Royal Wolverhampton Hospitals, Wolverhamp-*

[†] PubMed indexed collaborators: Luca Orecchia, Adan Tijerina, Adrian Simoes, Ahmed Ali, Aisling Nic an Riogh, Albert Wong, Alex Kiciak, Alexander Ridgway, Alexandra Szostek, Ananda Dhanasekaran, Anderson Cheong, Andrew Atayi, Ashna Ashpak, Beatriz Gutierrez Hidalgo, Bernardo Teixeira, Carlo Maria Scornajenghi, Carolina Marramaque, Charmian Reynoldson, Christopher Ho Chee Kong, Claire Crewe, Daniel Griffiths, Daniele Amporore, David Mak, Debashis Sarkar, Desmond Chung Wei Ling, Divya Bheenick, Donnacha Hogan, Dubem Orakwe, Emma Gordon, Enrico Checcucci, Faisal Nawaz, Fernando Ribeiro Gonçalves, Fernando Lozano Palacio, Francesco Prata, Francesco Claps, Francesco Del Giudice, Gaurav Aggarwal, Gautier Marcq, Georgios Hatzichristodoulou, Georgios Karagiannidis, Giacomo Maria Pirola, Giorgio Russo, Hamid Hytham, Hau Chun Khoo, Hesham Abozied, Hiten D Patel, Holly Colvin, Iftikhar Ali, Ijeoma Chibuzo, Ildar Fakhradiyev, Ioannis Sokolakis, Ioannis Tsikopoulos, James Chong, Jasper Bondad, Jayasimha Abbaraju, Jeremy Teoh, Jessica Whitburn, John Hayes, Jose Luis Bauza Quetglas, Jose Antonio Herranz Yague, Juan Gomez Rivas, Juan Colombo Stenstrom, Kathryn Bandeira de Mello, Katie Brodie, Lazaros Tzelves, Lazarou Lazaros, Louise Paramore, Luis Rico, Lydia Lilis, Maha Ullmann, Mahmoud Srour, Matteo Boltri, Mohamed Mustafa, Mohammad Eyad Takahji, Mohammed Almusimie, Mohammud Shakeel Inder, Mostafa Elgamal, Muftah Misurati, Muhammad-Kabir Ali, Muhannud Binnawara, Nabeela Bhaloo, Natalia Vidal Crespo, Nelson Ernesto Morales Palacios, Nicola Santoni, Nikolaos Pyrgidis, Olivia Hamilton, Pablo Abad Lopez, Pattabiraman Maheshkumar, Petros Sountoulides, Pietro Piazza, Pilar Moreno, Piyush Sarmah, Prabhat Narayan, Ravi Banthia, Richard Matulewicz, Roberto Contieri, Rotimi David, Sabir Mohammad, Sadiq Abu, Santiago Alonso, Simona Ippoliti, Stephan Weber, Suleiman Abuhasanein, Taeweon Lee, Tobias Klatte, Tran Trung Thanh, Usman Wazir, Volkan Ulker, WS Yeoh, Zach Feuer, Zain Elahi, and Zara Gall.

[*] Corresponding author. School of Medicine, University of St Andrews, N Haugh, St. Andrews KY16 9TF, UK. Tel. +44 786 274 0049.
E-mail address: sinan.khadhouri@doctors.org.uk (S. Khadhouri), .

https://doi.org/10.1016/j.euf.2024.06.004

Please cite this article as: S. Khadhouri, A. Hramyka, K. Gallagher et al., Machine Learning and External Validation of the IDENTIFY Risk Calculator for Patients with Haematuria Referred to Secondary Care for Suspected Urinary Tract Cancer, Eur Urol Focus (2024), https://doi.org/10.1016/j.euf.2024.06.004

ton, UK; [p] Department of Urology, University Hospital LMU, Munich, Germany; [q] Royal Berkshire Hospital, Reading, UK; [r] Hospital Clinico San Carlos, Madrid, Spain; [s] University Hospitals of Derby and Burton, Derby, UK; [t] Department of Surgery, Hanoi Medical University, Hanoi, Vietnam; [u] Department of Urology, Hanoi Medical University Hospital, Hanoi, Vietnam; [v] University of Trieste, Trieste, Italy; [w] Mercy University Hospital, Cork, Ireland; [x] Urologia-Hospital Clinico San Carlos, Madrid, Spain; [y] Stepping Hill Hospital, Stockport, UK; [z] Clinico San Carlos Hospital, Madrid, Spain; [aa] Oxford University Hospitals, Oxford, UK; [bb] S.H. Ho Urology Centre, Department of Surgery, The Chinese University of Hong Kong, Hong Kong; [cc] Urology Department, Claude Huriez Hospital, CHU Lille, Lille, France; [dd] CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, UMR9020-U1277 – CANTHER – Cancer Heterogeneity Plasticity and Resistance to Therapies, University Lille, Lille, France; [ee] Southend University Hospital, Southend-on-Sea, Essex, UK; [ff] Department of Urology, Aristotle University of Thessaloniki, Thessaloniki, Greece; [gg] University College London, London, UK

Article info

Article history:
Accepted June 8, 2024

Associate Editor: Christian Gratzke

Keywords:
Haematuria
Predictive model
Prediction
Urinary tract cancer
Bladder cancer
Upper tract urothelial cancer
Renal cancer
Validation
Risk calculator
Cancer risk

Abstract

Background: The IDENTIFY study developed a model to predict urinary tract cancer using patient characteristics from a large multicentre, international cohort of patients referred with haematuria. In addition to calculating an individual's cancer risk, it proposes thresholds to stratify them into very-low-risk (<1%), low-risk (1–<5%), intermediate-risk (5–<20%), and high-risk (≥20%) groups.

Objective: To externally validate the IDENTIFY haematuria risk calculator and compare traditional regression with machine learning algorithms.

Design, setting, and participants: Prospective data were collected on patients referred to secondary care with new haematuria. Data were collected for patient variables included in the IDENTIFY risk calculator, cancer outcome, and TNM staging. Machine learning methods were used to evaluate whether better models than those developed with traditional regression methods existed.

Outcome measurements and statistical analysis: The area under the receiver operating characteristic curve (AUC) for the detection of urinary tract cancer, calibration coefficient, calibration in the large (CITL), and Brier score were determined.

Results and limitations: There were 3582 patients in the validation cohort. The development and validation cohorts were well matched. The AUC of the IDENTIFY risk calculator on the validation cohort was 0.78. This improved to 0.80 on a subanalysis of urothelial cancer prevalent countries alone, with a calibration slope of 1.04, CITL of 0.24, and Brier score of 0.14. The best machine learning model was Random Forest, which achieved an AUC of 0.76 on the validation cohort. There were no cancers stratified to the very-low-risk group in the validation cohort. Most cancers were stratified to the intermediate- and high-risk groups, with more aggressive cancers in higher-risk groups.

Conclusions: The IDENTIFY risk calculator performed well at predicting cancer in patients referred with haematuria on external validation. This tool can be used by urologists to better counsel patients on their cancer risks, to prioritise diagnostic resources on appropriate patients, and to avoid unnecessary invasive procedures in those with a very low risk of cancer.

Patient summary: We previously developed a calculator that predicts patients' risk of cancer when they have blood in their urine, based on their personal characteristics. We have validated this risk calculator, by testing it on a separate group of patients to ensure that it works as expected. Most patients found to have cancer tended to be in the higher-risk groups and had more aggressive types of cancer with a higher risk. This tool can be used by clinicians to fast-track high-risk patients based on the calculator and investigate them more thoroughly.

1. Introduction

Haematuria is the most common reason for urgent referral to urology services to investigate suspected urinary tract cancer [1–3]. Early detection is particularly important for patient prognosis in urothelial cancer; therefore, this generates significant pressures on health services to investigate these patients in a timely fashion [4]. The prevalence of urinary tract cancer in patients referred with visible or nonvisible haematuria is 28.2% [5]. Many patients with benign disease or no sinister cause for haematuria are referred with the same level of urgency as those with cancer, which can consume capacity and delay cancer diagnosis. One way to improve this is to risk stratify patients using a cancer predictive model. The IDENTIFY study developed such a model using patient characteristics from a large international cohort of over 10 000 patients [6] to predict urinary tract cancer (bladder, upper tract urothelial, and renal cancers). In addition to calculating the individuals' percentage of cancer risk, it proposes thresholds to stratify them into very-low-risk (<1%), low-risk (1–<5%), intermediate-risk (5–<20%), and high-risk (≥20%) groups. The model showed good internal validity, discrimination, and negligible optimism (overfitting); however, this has not been validated externally.

Please cite this article as: S. Khadhouri, A. Hramyka, K. Gallagher et al., Machine Learning and External Validation of the IDENTIFY Risk Calculator for Patients with Haematuria Referred to Secondary Care for Suspected Urinary Tract Cancer, Eur Urol Focus (2024), https://doi.org/10.1016/j.euf.2024.06.004

Other validated models exist, such as the Haematuria Cancer Risk Score (HCRS) [7], which uses age, sex, type of haematuria, and smoking as predictors. However, this model predicts bladder cancer only and does not take into account upper tract urothelial cancer (UTUC) or renal cancer, which should be considered when investigating patients with haematuria.

We aim to externally validate the IDENTIFY predictive model, compare it with the HCRS model, and explore machine learning modelling techniques to assess whether a better model exists than that developed by traditional regression methods.

2. Patients and methods

2.1. Study design and population

We designed an international multicentre prospective study to evaluate patients with haematuria referred to secondary care for suspected urinary tract cancer with the same eligibility criteria as the original IDENTIFY study [6]. These criteria included adult patients referred with either visible or nonvisible haematuria, and consequently underwent investigation with cystoscopy and upper tract imaging. Nonvisible haematuria was defined as a trace or more on urinalysis, or three or more red blood cells per high power field on microscopy. Microscopy was not required to confirm a urinalysis positive for blood. Patients with previous urological malignancy or on surveillance for cancer recurrence were excluded. Anonymous data were collected prospectively from secondary care urology departments and included variables used in the original predictive model, presence and type of urinary tract cancer, and TNM classification. Data were collected from February 2022 to February 2023. All patients were followed up in this period until a diagnosis of cancer was confirmed, or they were discharged from the clinical care team after completing their investigations. All data were quality assessed for completeness. Cases with missing data from predictors were excluded as the model could not be assessed.

We report this study according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement [8], and Consolidated Reporting Guidelines for Prognostic and Diagnostic Machine Learning Modelling Studies: Development and Validation [9].

2.2. Outcome

The primary outcome was the detection of urinary tract cancer by the IDENTIFY predictive model. The performance of the model was assessed using the area under receiver operating characteristic curve (AUC), calibration coefficient calibration in the large (CITL), and the AUCs of machine learning models alongside their optimism.

The secondary outcomes were the prevalence of urinary tract cancer and stage of cancers stratified by risk categories.

2.3. Risk stratification

We selected a threshold of <1% predicted risk for the very-low-risk group, and 5% and 20% as cut-offs to create low-, intermediate-, and high-risk groups (Supplementary Fig. 1). These thresholds were taken from the original development model and were selected based on clinical reasoning by the study steering committee as this is a more meaningful method than statistical methods.

2.4. Predictors

Data were collected on age, type of haematuria (visible or nonvisible), sex, smoking status, family history of urothelial cancer, previous negative haematuria evaluation, urinary tract infection (single or recurrent), catheter use, previous pelvic radiotherapy, dysuria or suprapubic pain, anticoagulation (any type including antiplatelets), and flank pain.

Age was treated as a continuous variable and centred around its mean due to its interaction terms. The linear coefficients of the multivariable logistic model are shown in Supplementary Figure 2.

2.5. Sample size

The adjusted prevalence of urinary tract cancer from the IDENTIFY study was 28.2% [5]. Using the methods described by Riley et al [10], we calculated a minimum sample size of 2254, and 631 events to target precise performance measures of observed:expected ratio, calibration slope, C-statistic, and standardised net benefit (Supplementary Table 1).

2.6. Statistical analysis

To calculate the predicted probability of urinary tract cancer, we used the mixed-effect multivariable logistic regression formula given in the IDENTIFY predictive model [6]. We calculated the AUC, calibration coefficient and CITL, and Brier score of the predictive model on the validation cohort. Cancer prevalence was calculated as detected cancer cases within the defined population (patients with haematuria referred to secondary care). We also performed a subanalysis on urothelial cancer prevalent countries (over 90% prevalence) to assess whether this improves the model's performance. Conservative recalibration was performed to improve the CITL by adjusting the constant in the formula. We used a decision curve analysis to compare our model with the HCRS [7].

These analyses were performed using Stata version 16.1 (StataCorp, College Station, TX, USA).

2.7. Machine learning

To establish whether other types of models might provide improved predictions, we conducted a comprehensive comparison with the modern machine learning classification algorithms K Nearest Neighbour, Multilayer Perceptron (neural network), Random Forest, and Gradient Boosted trees. Predictive performance evaluation was carried out using the Tripod Cross-Validation method [8]. An analysis was performed using Python 3.8 and the Scikit-learn and XGBoost software libraries.

Initially, the models were trained with default hyperparameter settings and assessed using the IDENTIFY development cohort. Subsequently, the best performing class of model was tuned using Bayesian optimisation [11] to find optimal hyperparameter settings and hence supply a robust estimate for model performance when used to predict unseen instances. The performance of the optimised model was evaluated using a validation cohort.

Missing variable data were imputed using the missForest algorithm [12]. The Random Forest algorithm supplies useful feature importance information as a by-product, and hence we report the strength of each variable as a predictor using this technique.

2.8. Data handling and ethics

Anonymised patient data were securely collected from routinely documented information during the investigation of haematuria, and patient records were accessed only by the direct clinical care team. Participating institutions registered the study locally with their research and development, and approval for study participation was granted locally. In the

Please cite this article as: S. Khadhouri, A. Hramyka, K. Gallagher et al., Machine Learning and External Validation of the IDENTIFY Risk Calculator for Patients with Haematuria Referred to Secondary Care for Suspected Urinary Tract Cancer, Eur Urol Focus (2024), https://doi.org/10.1016/j.euf.2024.06.004

UK, the coordinating country, this study was exempt from ethical approval as it is deemed an audit consistent with the UK Health Research Authority guidelines.

3. Results

Figure 1 describes the flow of patients through the study. There were a total of 3672 patients, of whom 90 (2.5%) were excluded due to missing outcome data. Supplementary Table 2 lists the number of patients from each participating country in the development and validation cohorts. Each cohort had 26 countries, but there were 13 different countries between the validation and development groups, and 57 different centres overall. The unadjusted prevalence of urinary tract cancer was 23.6% (n = 795), bladder cancer 20.7%, renal cancer 1.3%, and UTUC 1.8%.

Table 1 shows the patient characteristics of the validation cohort alongside the development cohort. The proportions and means of these characteristics were similar, and well matched with overlapping confidence intervals (CIs).

Supplementary Figure 3 shows the performance of the IDENTIFY predictive model in the validation cohort. The AUC was 0.78 (95% CI 0.77–0.80), calibration slope coefficient was 0.95 (95% CI 0.90–1.00), CITL was 0.315, and Brier score was 0.15. When tested on urothelial cancer prevalent (>90% prevalence) countries only (ie, excluding Sudan, Egypt, Nigeria, and Libya), the performance improved slightly with an AUC of 0.80 (95% CI 0.78–0.81), a calibration slope of 1.04 (95% CI 0.99–1.09), a CITL of 0.24, and a Brier score of 0.14 (Fig. 2).

Table 2 shows the distribution of cancers, and their stage and grade across the different risk categories in this cohort. There were no cancers in the very-low-risk group (n = 167). The majority (76.6%) of tumours were risk stratified to the high-risk group, with 18.9% and 4.5% in the intermediate- and low-risk groups, respectively. In the low-risk group, there were no metastatic renal cancer or UTUC, nor any invasive UTUC, and there was a lower ratio (0.09) of muscle-invasive to non–muscle-invasive bladder tumours than in the intermediate- and high-risk groups (0.34 and 0.27, respectively). Similarly, there was a lower ratio of high- to low-grade bladder cancer in the low-risk group (1.09) than in the intermediate- and high-risk groups (1.59 and 1.68, respectively).

Figure 3 shows that the IDENTIFY predictive model had a greater clinical net benefit than the HCRS model, especially in the intermediate- and high-risk groups, compared with investigating all or none.

Supplementary Table 3 shows the performance of machine learning models with default hyperparameters when applied to a development cohort. The best performing

Table 1 – Patient characteristics of the development and validation cohorts[a]

	Development cohort (n = 10 282)	Validation cohort (n = 3582)
Age (yr), mean ± SD	64.3 ± 14.6	65.2 ± 16.1
Visible haematuria (%)	69.3 (68.4–70.2)	68.8 (67.3–70.3)
Nonvisible haematuria (%)	30.7 (29.8–31.6)	31.2 (29.7–32.7)
Male (%)	62.5 (61.5–63.4)	64.8 (63.2–66.3)
Female (%)	37.5 (36.5–38.4)	35.2 (33.7–36.8)
Smoker (%)	18.3 (17.6–19.1)	19.0 (17.7–20.2)
Ex-smoker (%)	29.2 (28.4–30.1)	30.9 (29.4–32.4)
Never smoked (%)	52.5 (51.5–53.4)	50.1 (48.5–51.8)
Family history urothelial cancer (%)	2.1 (1.9–2.4)	3.66 (3.1–4.3)
Previous negative haematuria evaluation (%)	10.1 (9.5–10.7)	15.5 (14.4–16.8)
Single UTI (%)	12.3 (11.7–13.0)	11.4 (10.4–12.5)
Recurrent UTI (%)	10.0 (9.5–10.6)	10.5 (9.6–11.6)
Catheter use (%)	3.3 (3.0–3.6)	7.05 (6.2–7.9)
Pelvic radiotherapy (%)	2.0 (1.7–2.7)	2.91 (2.4–3.5)
Dysuria/suprapubic pain (%)	20.9 (20.1–21.6)	26.4 (24.0–35.1)
Anticoagulation (%)	26.3 (25.5–27.2)	26.9 (25.5–28.4)
Flank pain (%)	9.0 (8.4–9.5)	13.0 (11.9–14.2)

SD = standard deviation; UTI = urinary tract infection.
[a] 95% confidence intervals are shown in brackets after each percentage.

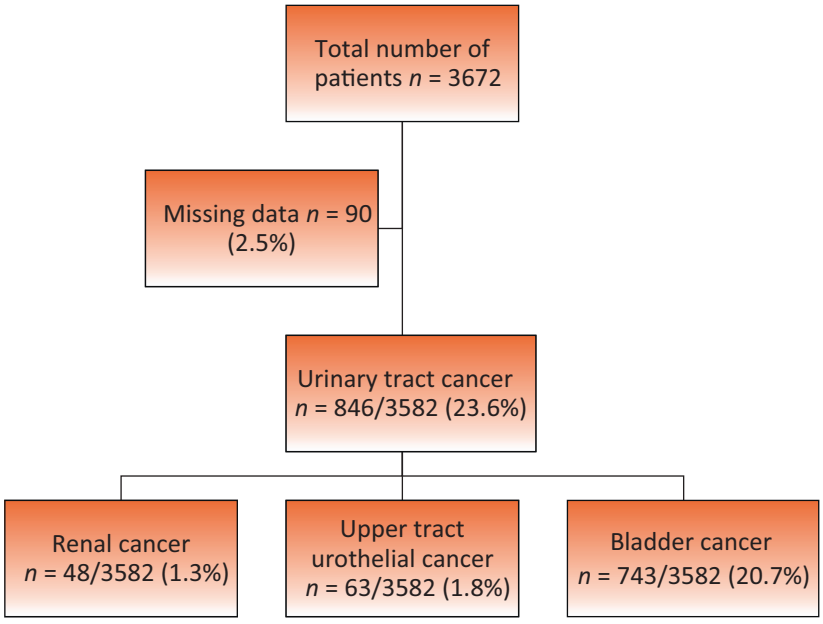

Fig. 1 – Flow chart of the study population.

Please cite this article as: S. Khadhouri, A. Hramyka, K. Gallagher et al., Machine Learning and External Validation of the IDENTIFY Risk Calculator for Patients with Haematuria Referred to Secondary Care for Suspected Urinary Tract Cancer, Eur Urol Focus (2024), https://doi.org/10.1016/j.euf.2024.06.004

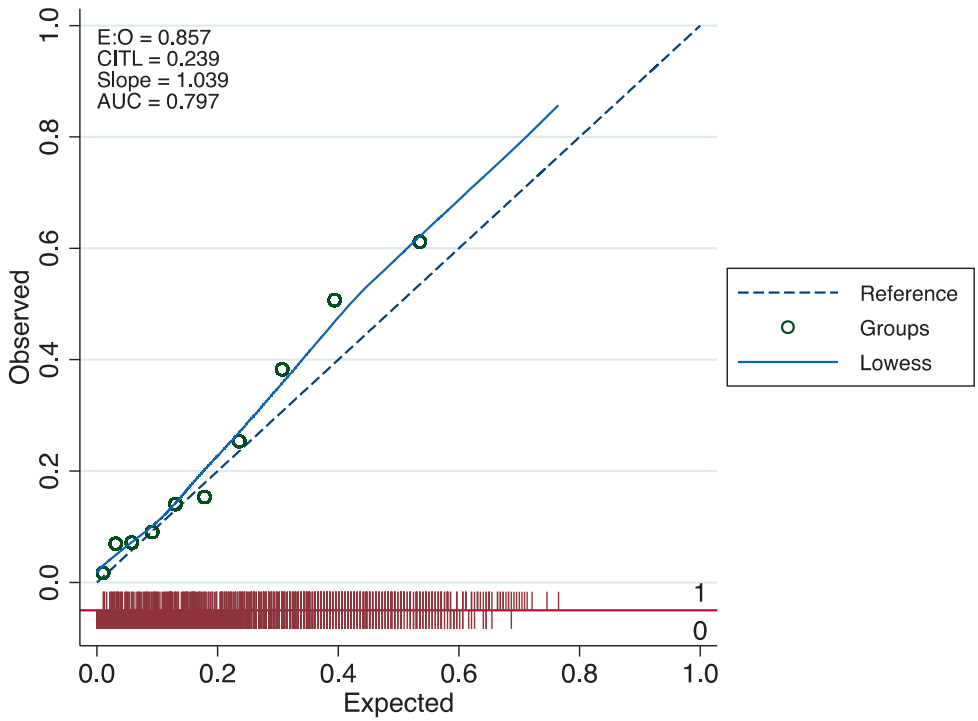

Fig. 2 – Calibration plot of the IDENTIFY predictive model on validation cohort excluding nonurothelial cancer prevalent countries (Sudan, Egypt, Nigeria, and Libya). The plot is of observed versus predicted cancers. Calibration is plotted in groups across the risk spectrum as recommended in the TRIPOD statement. The spike plot denotes the outcome of cancer as 1 and no cancer as 0. A calibration slope of 1 and a CITL of 0 represent perfect calibration. AUC = area under the receiver operating characteristic curve; CITL = calibration in the large; E:O = predicted:observed ratio; Lowess = locally weighted scatterplot smoothing; Slope = calibration slope.

Table 2 – Distribution of cancers and TNM stage across risk categories in the validation cohort excluding nonurothelial cancer prevalent countries (Sudan, Egypt, Nigeria, and Libya)[a]

	Very low risk (n = 167, 4.5%)	Low risk (n = 660, 18.0%)	Intermediate risk (n = 1306, 37.0%)	High risk (n = 1449, 39.5%)
Urinary tract cancer, n (%)		36 (4.5)	150 (18.9)	609 (76.6)
Bladder cancer, n (%)		32 (4.5)	131 (18.6)	542 (76.9)
Low grade		11 (4.6)	44 (18.3)	185 (77.1)
High grade		12 (3.1)	67 (17.1)	312 (79.8)
High:low grade ratio		1.09	1.52	1.68
Missing grade		9 (12.2)	20 (27.0)	45 (60.8)
Tx		1 (11.1)	4 (44.4)	4 (44.4)
Non–muscle invasive (≤T1)		22 (4.3)	85 (16.7)	401 (78.9)
Muscle invasive (≥T2)		2 (1.4)	29 (20.7)	109 (77.9)
Muscle:non–muscle invasive ratio		0.09	0.34	0.27
Tis		5 (31.2)	5 (31.3)	6 (37.5)
Missing T stage		2 (6.2)	8 (25.0)	22 (68.8)
Nx		15 (5.4)	49 (17.8)	212 (76.8)
N0		14 (3.9)	62 (17.1)	286 (79.0)
≥N1		1 (2.9)	12 (35.3)	21 (61.8)
Mx		12 (4.7)	48 (18.8)	196 (76.6)
M0		17 (4.3)	68 (17.0)	315 (78.8)
≥M1	No cancers detected	1 (6.3)	7 (43.8)	8 (50.0)
Renal cancer, n (%)		4 (9.8)	14 (34.2)	23 (56.1)
Tx		1 (33.3)	0 (0)	2 (66.7)
T1		2 (14.3)	5 (35.7)	7 (50.0)
T2		0 (0)	0 (0)	5 (100)
T3		1 (7.7)	4 (30.8)	8 (61.5)
Missing T stage		0 (0)	5 (83.3)	1 (16.7)
Nx		0 (0)	1 (14.3)	6 (85.7
N0		4 (18.2)	7 (31.8)	11 (50)
≥N1		0 (0)	2 (28.6)	5 (71.4)
Mx		1 (14.3)	0 (0)	6 (85.7)
M0		3 (13.0)	8 (34.8)	12 (52.2)
≥M1		0 (0)	2 (33.3)	4 (66.7)
Upper tract urothelial cancer, n (%)		1 (1.8)	5 (8.8)	51 (89.5)
Tx		0 (0)	1 (9.1)	10 (90.9)

(*continued on next page*)

Please cite this article as: S. Khadhouri, A. Hramyka, K. Gallagher et al., Machine Learning and External Validation of the IDENTIFY Risk Calculator for Patients with Haematuria Referred to Secondary Care for Suspected Urinary Tract Cancer, Eur Urol Focus (2024), https://doi.org/10.1016/j.euf.2024.06.004

Table 2 (*continued*)

	Very low risk (*n* = 167, 4.5%)	Low risk (*n* = 660, 18.0%)	Intermediate risk (*n* = 1306, 37.0%)	High risk (*n* = 1449, 39.5%)
≤T1		1 (2.8)	4 (11.1)	31 (86.1)
≥T2		0 (0)	0 (0)	8 (100)
Missing T stage		0 (0)	0 (0)	2 (100)
Nx		0 (0)	0 (0)	11 (100)
N0		1 (2.4)	4 (9.5)	37 (88.1)
≥N1		0 (0)	1 (33.3)	2 (66.7)
Mx		0 (0)	0 (0)	11 (100)
M0		1 (2.5)	4 (10.0)	35 (87.5)
≥M1		0 (0)	1 (20.0)	4 (80.0)

[a] Percentages are row percentages.

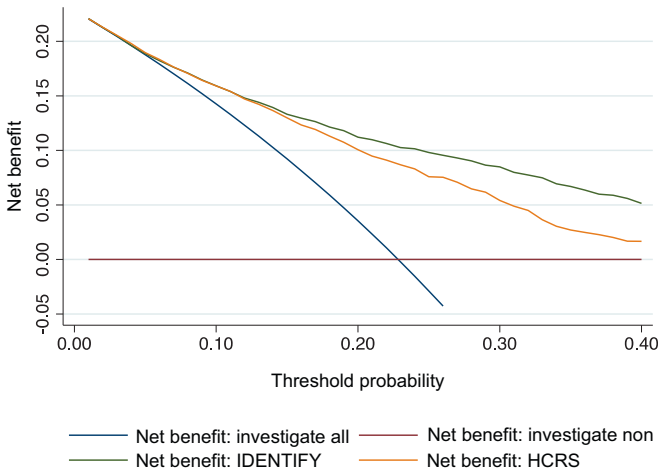

Fig. 3 – Decision curve analysis comparing the net benefit of the IDENTIFY predictive model compared with the Haematuria Cancer Risk Score (HCRS). The decision curve analysis shows that there is a greater net benefit of using the IDENTIFY predictive model over the HCRS model to investigate patients with haematuria, compared with investigating all or none.

model from these, Random Forest, was chosen for Bayesian optimisation, returning an AUC of 0.81 with optimism of 0.02 (compared with the multivariable mixed-effect logistic regression with an AUC of 0.86 and optimism of 0.0049).

When applied to the validation cohort, the optimised Random Forest achieved an AUC of 0.76 (Fig. 4 presents its contingency table [confusion matrix] showing true positive, true negative, false positive, and false negative distribution).

In Figure 5, we present an overview of the relative feature importance retrieved from the Random Forest algorithm. The relative importance of these variables matched the choice of variables used in our logistic regression model.

4. Discussion

We aimed to validate the IDENTIFY predictive model externally [6] using a separate international cohort of patients with haematuria referred to secondary care for the investigation of suspected urinary tract cancer. We showed that in an independent validation cohort, the model performed well with an AUC of 0.78 and good calibration. This improved on a subanalysis of urothelial cancer prevalent countries only, with an AUC of 0.80, a calibration slope of 1.04, a CITL of 0.24, and a Brier score of 0.14. This evaluation gives confidence that the model can be used safely to predict the risk of urinary tract cancer in patients with haematuria, and it can be applied as a risk stratification tool. The majority of cancers were stratified to the high- and intermediate-risk groups. In our patient and public involvement work, development of such a tool was deemed a high priority in urological cancer research.

Conservative recalibration is possible to improve the CITL to 0 by adjusting the constant in the formula. This would reduce the risk of a slight underestimation of cancer by the model.

The model stratified most cancers to the high-risk group (predicted cancer risk ≥20%). Additionally, it stratified more aggressive and advanced cancers to the high-risk group and less so to the low-risk group. For bladder cancer, these were higher muscle to non–muscle invasive and high- to low-grade ratio. Furthermore, there were no metastatic UTUC or renal cell carcinoma in the low-risk group. Finally, there were no cancers in the very-low-risk group (predicted cancer risk <1%), and this could be a group where no investigation is considered.

The IDENTIFY model also showed a greater net benefit than an existing validated predictive model (HCRS) [7],

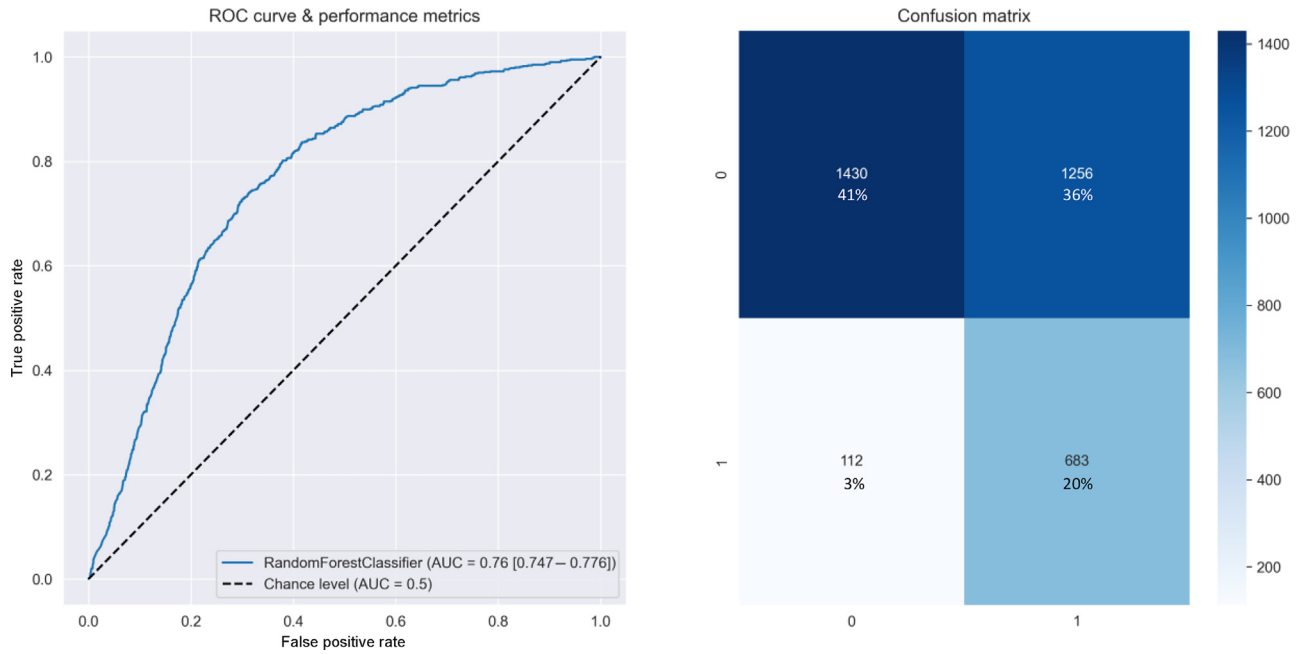

Fig. 4 – (A) ROC curve and (B) confusion matrix illustrating the performance of the optimised Random Forest on the validation cohort. Confusion matrix row: 0 = predicted negative, 1 = predicted positive; column: 0 = actual negative, 1 = actual positive; bottom right = true positive; bottom left = false negative; top left = true negative; and top right = false positive. AUC = area under the receiver operating characteristic curve; ROC = receiver operating characteristic.

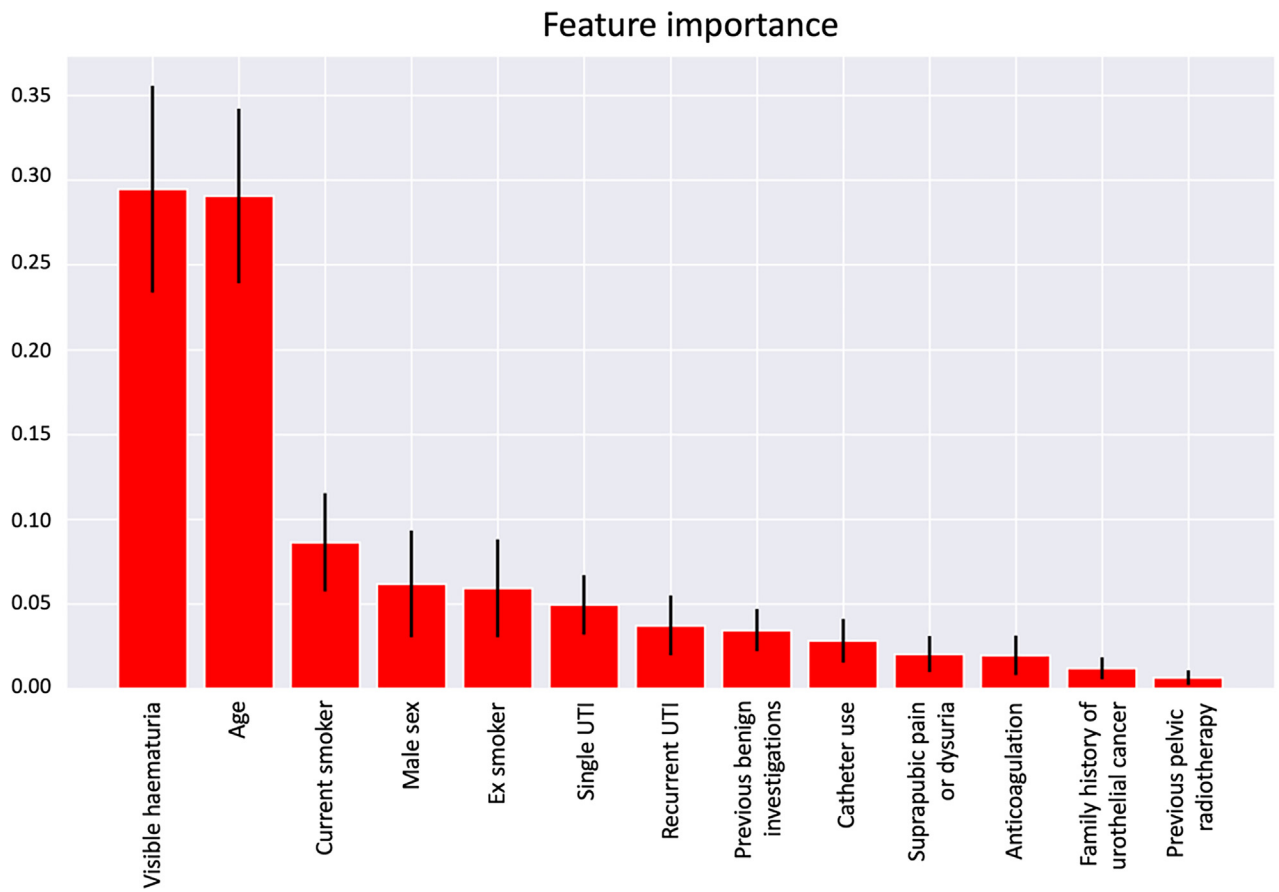

Fig. 5 – Relative feature importance retrieved from the Random Forest algorithm. UTI = urinary tract infection.

especially in the intermediate- and high-risk group where the majority of cancers were stratified, and improved on the HCRS model as it predicts upper tract cancers as well as bladder cancers. The HCRS was developed from a UK cohort recruited for a urinary biomarker trial and validated in a Swiss cohort. The difference in study design and pri-

Please cite this article as: S. Khadhouri, A. Hramyka, K. Gallagher et al., Machine Learning and External Validation of the IDENTIFY Risk Calculator for Patients with Haematuria Referred to Secondary Care for Suspected Urinary Tract Cancer, Eur Urol Focus (2024), https://doi.org/10.1016/j.euf.2024.06.004

mary outcomes of the two studies may provide explanations for differences in model performance. These include selection bias due to trial recruitment, number and choice of predictors (a priori vs. post hoc), methodology of inclusion of predictors in the model, differences in geographical breadth (international vs. single country), and sample size.

The strengths of this validation study lie in its prospective, international, multicentre design and large validation cohort size. As the predictors used in the IDENTIFY model are extensive, it was important to collect the data prospectively, as a retrospective dataset may contain many missing values.

A careful assessment of modern machine learning algorithms shows no improvement to the mixed-effect multivariable logistic regression used in the IDENTIFY predictive model [6], thereby increasing confidence that this model is optimal amongst candidate predictive methods and that the data are not sensitive to the choice of model employed. In addition, the relative importance of variables assessed using Random Forests is a close match to the relative size of coefficients in the original logistic regression model, again validating the increased and reduced risk results for the original study.

One limitation of the study is that the histological type of cancer was not collected in the validation cohort, and so we were unable to perform a subanalysis on urothelial cancers alone. However, we found that this model works better after excluding countries where squamous cell carcinoma is more prevalent. This is likely due to the majority of bladder cancers in the development cohort being urothelial cancer. Therefore, we would caution the use of this calculator in countries where squamous cell cancer of the bladder is more prevalent, as this has different biological causes, and its predictors will be different (such as exposure to schistosomiasis).

Validation of the model allows for the use of this predictive model in clinical use. We envisage its use at the point of contact between the patient and urologist after a history has been taken. The IDENTIFY risk calculator can be used (available as an app "IDENTIFY risk calculator" via Apple and Android) to predict the patient's individual cancer risk, stimulating a discussion with the patient and urologist regarding the need for further investigation, and the urgency and intensity of these investigations rather than a blanket full set of investigations as is commonly seen in practice today. Patients in the very-low-risk group (<1% predicted risk) may avoid investigations altogether, unless there may be a benign cause for their symptoms that requires nonurgent tests. Those in the high-risk and very-high-risk groups can be prioritised for urgent and more intensive investigation with the consideration of streamlining pathways to get histological diagnosis and treatment as quickly as possible.

International guidelines differ in their recommendation for investigation of haematuria and consider different age thresholds for referral, type of haematuria, and various symptoms and patient characteristics [13–15]. All guidelines consider visible haematuria as the most important factor for investigation. The IDENTIFY predictive model builds on these population-targeted guidelines by providing individualised cancer prediction to fine tune a patient's risk and aid decision-making.

5. Conclusions

A future analysis from the IDENTIFY study will consider the diagnostic accuracy of imaging tests after risk stratification using the prediction model, to recommend an improved pathway with better allocation of diagnostic resources according to risk groups, allowing for less invasive tests in low-risk groups and more intense investigations in high-risk groups. This can lead to a clinical trial that assesses the investigations avoided and any cancers missed. The new diagnostic pathway will aim to improve time to cancer diagnosis and treatment, reduce waiting lists and healthcare costs, and free up more radiological services.

***Author contributions*:** Sinan Khadhouri had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Khadhouri, Kasivisvanathan, Hramyka, Kelsey.
Acquisition of data: The IDENTIFY Study Group (see other), all authors.
Analysis and interpretation of data: Khadhouri, Hramyka.
Drafting of the manuscript: Khadhouri.
Critical revision of the manuscript for important intellectual content: Khadhouri, Hramyka, Gallagher, Light, Ippoliti, Edison, Alexander, Kulkarni, Zimmermann, Nathan, Orecchia, Banthia, Piazza, Mak, Pyrgidis, Narayan, Lopez, Nawaz, Tran, Claps, Hogan, Rivas, Alonso, Chibuzo, Hidalgo, Whitburn, Teoh, Marcq, Szostek, Bondad, Sountoulides, Kelsey, Kasivisvanathan.
Statistical analysis: Khadhouri.
Obtaining funding: Khadhouri, Kasivisvanathan.
Administrative, technical, or material support: Khadhouri, Gallagher, Light, Ippoliti, Edison, Alexander, Kulkarni, Zimmermann, Nathan.
Supervision: Kasivisvanathan, Tom Kelsey.
Other: The IDENTIFY Study Group—we request PubMed indexed collaborator authorship for all those named in this section in the manuscript, as they all contributed patient records to the dataset and without them this would not be possible.

Financial disclosures: Sinan Khadhouri certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

***Funding/Support and role of the sponsor*:** None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.euf.2024.06.004.

References

[1] Mathew A, Desai K. An audit of urology two-week wait referrals in a large teaching hospital in England. Ann R Coll Surg Engl 2009;91:310–2.

[2] Hawary AM, Warburton HE, Brough RJ, et al. The "2-week wait" rule for referrals for suspected urological cancers—urgent need for refinement of criteria. Ann R Coll Surg Engl 2008;90:517–22.
[3] Allen D, Popert R, O'Brien T. The two-week-wait cancer initiative in urology: useful modernization? J R Soc Med 2004;97:279–81.
[4] Hollenbeck BK, Dunn RL, Ye Z, et al. Delays in diagnosis and bladder cancer mortality. Cancer 2010;116:5235–42.
[5] Khadhouri S, Gallagher KM, MacKenzie K, et al. The IDENTIFY study: the investigation and detection of urological neoplasia in patients referred with suspected urinary tract cancer; a multicentre observational study. BJU Int 2021;128:440–50.
[6] Khadhouri S, Gallagher KM, Mackenzie KR, et al. Developing a diagnostic multivariable prediction model for urinary tract cancer in patients referred with haematuria: results from the IDENTIFY collaborative study. Eur Urol Focus 2022;8:1673–82.
[7] Tan WS, Ahmad A, Feber A, et al. Development and validation of a haematuria cancer risk score to identify patients at risk of harbouring cancer. J Intern Med 2019;285:436–45.
[8] Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–W73.
[9] Klement W, El Emam K. Consolidated reporting guidelines for prognostic and diagnostic machine learning modeling studies: development and validation. J Med Internet Res 2023;25:e48763.
[10] Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med 2021;40:4230–51.
[11] Victoria AH, Maragatham G. Automatic tuning of hyperparameters using Bayesian optimization. Evol Syst 2021;12:217–23.
[12] Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 2012;28:112–8.
[13] NICE. Overview | Suspected cancer: recognition and referral | Guidance. 2015.
[14] Barocas DA, Boorjian SA, Alvarez RD, et al. Microhematuria: AUA/SUFU guideline. J Urol 2020;204:778–86.
[15] 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;81:75–94.

Please cite this article as: S. Khadhouri, A. Hramyka, K. Gallagher et al., Machine Learning and External Validation of the IDENTIFY Risk Calculator for Patients with Haematuria Referred to Secondary Care for Suspected Urinary Tract Cancer, Eur Urol Focus (2024), https://doi.org/10.1016/j.euf.2024.06.004