BJU Int 2023 doi:10.1111/bju.15959

BJUI
BJU International

Original Article

Machine-learning approach for prediction of pT3a upstaging and outcomes of localized renal cell carcinoma (UroCCR-15)

Astrid Boulenger de Hauteclocque[1], Loïc Ferrer[3], Damien Ambrosetti[4], Solene Ricard[1], Pierre Bigot[6], Karim Bensalah[7], François Henon[8], Nicolas Doumerc[9], Arnaud Méjean[10], Virginie Verkarre[10], Charles Dariane[10], Stéphane Larré[14], Cécile Champy[15], Alexandre de La Taille[15], Franck Bruyère[16], Morgan Rouprêt[11], Philippe Paparel[17], Stéphane Droupy[18], Alexis Fontenil[18], Jean-Jacques Patard[19], Xavier Durand[12], Thibaut Waeckel[20], Herve Lang[21], Cédric Lebâcle[22], Laurent Guy[23], Geraldine Pignot[24], Matthieu Durand[5], Jean-Alexandre Long[26], Thomas Charles[27], Evanguelos Xylinas[13], Romain Boissier[25], Mokrane Yacoub[2], Thierry Colin[3] and Jean-Christophe Bernhard[1]

[1]Department of Urology, [2]Pathology, Bordeaux University Hospital, Bordeaux, [3]SOPHiA GENETICS, Radiomics R&D Department, Pessac, [4]Department of Pathology, [5]Urology, Nice University Hospital, Nice, [6]Department of Urology, Angers University Hospital, Angers, [7]Department of Urology, Rennes University Hospital, Rennes, [8]Department of Urology, Lille University Hospital, Lille, [9]Department of Urology, Toulouse University Hospital, Toulouse, [10]Department of Urology, Georges Pompidou European University Hospital, [11]Department of Urology, La Pitié-Salpêtrière University Hospital, [12]Department of Urology, Saint-Joseph Hospital Foundation, [13]Department of Urology, Bichat University Hospital, Paris, [14]Department of Urology, Reims University Hospital, Reims, [15]Department of Urology, Henri Mondor University Hospital, Créteil, [16]Department of Urology, Tours University Hospital, Tours, [17]Department of Urology, Lyon Sud University Hospital, Lyon, [18]Department of Urology, Nîmes University Hospital, Nîmes, [19]Department of Urology, Mont de Marsan Hospital, Mont de Marsan, [20]Department of Urology, Caen University Hospital, Caen, [21]Department of Urology, Strasbourg University Hospital, Strasbourg, [22]Department of Urology, Bicêtre University Hospital, Le Kremlin-Bicêtre, [23]Department of Urology, Clermont-Ferrand University Hospital, Clermont-Ferrand, [24]Department of Urology, Paoli-Calmettes Institute, [25]Department of Urology, Marseille University Hospital, Marseille, [26]Department of Urology, Grenoble University Hospital, Grenoble, and [27]Department of Urology, Poitiers University Hospital, Poitiers, France

Objectives

To assess the impact of pathological upstaging from clinically localized to locally advanced pT3a on survival in patients with renal cell carcinoma (RCC), as well as the oncological safety of various surgical approaches in this setting, and to develop a machine-learning-based, contemporary, clinically relevant model for individual preoperative prediction of pT3a upstaging.

Materials and Methods

Clinical data from patients treated with either partial nephrectomy (PN) or radical nephrectomy (RN) for cT1/cT2a RCC from 2000 to 2019, included in the French multi-institutional kidney cancer database UroCCR, were retrospectively analysed. Seven machine-learning algorithms were applied to the cohort after a training/testing split to develop a predictive model for upstaging to pT3a. Survival curves for disease-free survival (DFS) and overall survival (OS) rates were compared between PN and RN after G-computation for pT3a tumours.

Results

A total of 4395 patients were included, among whom 667 patients (15%, 337 PN and 330 RN) had a pT3a-upstaged RCC. The UroCCR-15 predictive model presented an area under the receiver-operating characteristic curve of 0.77. Survival analysis after adjustment for confounders showed no difference in DFS or OS for PN vs RN in pT3a tumours (DFS: hazard ratio [HR] 1.08, $P = 0.7$; OS: HR 1.03, $P > 0.9$).

Conclusions

Our study shows that machine-learning technology can play a useful role in the evaluation and prognosis of upstaged RCC. In the context of incidental upstaging, PN does not compromise oncological outcomes, even for large tumour sizes.

Keywords

disease-free survival, partial nephrectomy, machine learning, renal cell carcinoma, TNM staging

Introduction

Locally advanced (pT3a) RCC has a worse prognosis than pT1–T2 [1]. Except for macroscopic renal vein thrombosis, preoperative imaging performs moderately well in accurately diagnosing T3a stage [2,3], resulting in a significant number of clinically localized tumours being upstaged to pT3a after nephrectomy. Although the current international recommendations consider radical nephrectomy (RN) as the 'gold standard' treatment for locally advanced disease [4], the extension of partial nephrectomy (PN) indications to larger renal masses may question the safety of such a procedure in this situation. Moreover, in this era of personalized medicine, the prediction of pT3a upstaging should be interesting to help choose the best treatment strategy among active surveillance, ablation, surgery and peri-operative (including neoadjuvant) treatments with immune-checkpoint inhibitors. Some predictive factors of pT3a upstaging, such as age and tumour size, have been previously reported [5,6] but so far there are no practical tools available to preoperatively assess individual patient risk and guide individual treatment.

Artificial intelligence has opened up a whole new era in clinical decision-making [7]. It has made it possible to process large amounts of medical data within limited time and with greater accuracy. With the capacity to automatically find the best action to achieve a given goal, machine-learning algorithms (MLAs) are powerful tools for statistical analyses [8]. However, MLAs require large, multi-institutional databases to follow the consecutive stages of training, testing and validating before any findings can be extrapolated to a new population. Few studies so far have explored RCC from a computational viewpoint [9-11].

Therefore, using machine-learning processes, we aimed to assess the impact of pT3a upstaging on survival as well as the oncological safety of various surgical approaches and develop a contemporary, clinically relevant model to preoperatively predict pT3a upstaging.

Materials and Methods

Study Population

We conducted a retrospective analysis of patients from the French multi-institutional kidney cancer database UroCCR. All data from patients were prospectively collected in the UroCCR database (NCT03293563). A signed consent was obtained from all the patients after clear information and an ethical board approbation was obtained for this study. The data collection and analysis were authorized by the National Information science and Liberties Commission (CNIL) under number DR-2013-206. The study cohort consisted of patients treated surgically for RCC from 2000 to 2019, either by laparoscopic (pure or robot-assisted) PN, open PN or laparoscopic radical nephrectomy (RN) in all UroCCR centres. All tumours were preoperatively staged as ≤cT2aN0M0 and postoperatively staged as ≤pT3aN0M0. Hereditary and non-primary tumours were excluded. In the causal analysis, only patients with available survival data were included. Data extraction, including follow-up, was performed on 20 May 2019.

Variables

Demographic and clinical variables were collected, including age at diagnosis, sex, American Society of Anesthesiologists (ASA) score, Eastern Cooperative Oncology Group performance status (ECOG PS) and presence of symptoms at diagnosis. Preoperative tumour variables included the RENAL nephrometry score and details of its items: tumour size, rim location, polar location, exophytic/endophytic nature and hilar location. Data from the pathological examination included: pathological tumour size, histological subtype (clear-cell RCC, papillary RCC, chromophobe RCC and other types), tumour stage, Fuhrman grade, presence of sarcomatoid features, necrosis, microvascular invasion, positive surgical margin and type of pT3a invasion.

Staging

Clinical staging was assigned according to the 2009 American Joint Committee on Cancer/Union Internationale Contre le Cancer classification [12]. Clinical staging was based on preoperative imaging and tumour size was defined as the maximum tumour diameter. All clinical stages were cross-referenced with tumour size variables and reclassified accordingly. Pathological staging was assessed by expert uropathologists. Cases involving surgery before 2010 were reclassified.

Outcomes for Analysis

Peri-renal fat invasion, sinus fat invasion and/or renal or segmental vein thrombus were staged as pT3a. Disease-free survival (DFS) was defined as the time from nephrectomy to diagnosis of local recurrence and/or metastatic progression and/or death of any cause. Local recurrence was defined as recurrence either on the ipsilateral kidney after PN or in the ipsilateral renal fossa after RN whether metastatic progression was defined as distant recurrence diagnosed on a CT scan or an MRI at any time during follow-up. Overall survival (OS) was defined as the time from nephrectomy to death of any cause.

Statistical Analysis

Effect of pT3a Upstaging on Survival

Disease-free survival and OS between pT3a-upstaged and non-upstaged tumours were displayed by Kaplan–Meier curves and compared with log-rank tests.

Effect of Surgical Approach on Survival Outcomes after Upstaging

Causal analyses were performed to assess the effect of treatment type (PN vs RN) on DFS and OS in the pT3a-upstaged cohort. Marginal survival curves and marginal causal effects were estimated using G-computation, which involves a two-step estimation based on the counterfactual framework [13]. Cox models were firstly fitted by adjusting on both treatment type and all preoperative variables. Three feature selection strategies were used to specify which preoperative variables should be selected as confounders in the model: (i) no selection (all variables kept); (ii) using a backward stepwise selection procedure to eliminate at each step the variable with higher associated *P* value, until achieving all variables with associated *P* value <0.1; (iii) using a backward stepwise selection procedure based on maximum likelihood ratio test, with *P* value cut-off at 0.1. The average treatment effects and the marginal survival curves were then estimated using the prediction of the counterfactual outcomes of each patient under each treatment type.

To take into account pathological features that might not be designated as potential confounders, we subsequently performed a non-causal analysis with Cox regression multivariable analyses after adjusting for pre- and postoperative covariates.

A sub-analysis was also ruled out to compare surgical approaches between open PN and laparoscopic PN (gathering both pure-laparoscopic PN and robot-assisted PN) in the upstaged cohort in terms of DFS and OS.

Causal analyses were performed with R software environment for statistical computing (version 4.0.3; http://www.r-project.org; main packages: survival v.3.2.7, survminer v.0.4.9, arsenal v.3.5.0, RISCA v.0.9.), $P < 0.05$ denoting statistical significance, unless specifically specified.

Upstaging Predictive Model

The study population was split into train and test cohorts. Sampling was stratified on centres with a 70:30 ratio based on the average pT3a rate. Centres were different from one cohort to the other (17 and eight centres, respectively), in order to perform an external validation.

The Mann–Whitney *U*-test for continuous variables and the chi-squared test for categorical variables were used to measure the association between covariates and pT3a status in the whole dataset and according to the two cohorts.

We tested several strategies of feature selection to specify which variables should be included in the statistical learning process to train the MLAs.

Our decision to exclude each variable was based on several indicators: the estimated level of bivariate association between each feature and outcome, the rate of missing values, and the consistency of the feature distribution between train and test datasets.

Missing values of the selected variables were imputed with the K-nearest neighbours' algorithm using $K = 10$ neighbours.

Seven supervised MLAs [14] were trained: Logistic regression, Binary Decision Tree, Random Forest, eXtreme Gradient Boosting tree (XGBoost), Gradient Boosting Machine (GBM), LightGBM and Support Vector Machine. Hyperparameters of each algorithm were optimized using a 10-fold cross-validation approach on the training dataset. The predictive abilities of each algorithm were evaluated on both train and test datasets using precision/recall curves, which are preferable in imbalanced datasets [15], and receiver-operating characteristic (ROC) curves. The best area under the precision/recall curve (prAUC) was 1 and the baseline was the actual rate of pT3a in the cohort.

The most effective algorithm was chosen to build the predictive model. A calibration curve was applied to the test dataset to validate the model, and to define risk-group cut-offs. Finally, a model based on SHapley Additive exPlanations [16] (SHAP) values was built to explain each patient's probability of being upstaged.

Statistical analyses for developing the prediction model were performed with Python software (version 3.8.15, main libraries: scikit-learn v.0.22.2, pandas v.1.2.4, Shap v.0.39.0, Pillow v.8.0.1, Matplotlib v.3.3.2, Streamlit v.0.82.0).

Results

Effect of pT3a Upstaging on Survival

Among a total of 4395 patients surgically managed for clinically localized RCC, 667 (15%) had pT3a tumours at final pathology. Patients were treated with either PN ($n = 3454$, 79%) or RN ($n = 941$, 21%). Demographics and tumour characteristics are reported in Table 1. Patients with a pathologically upstaged RCC to pT3a had a significantly worse prognosis (Fig. 1). At postoperative Year 2, DFS and OS were 81% and 93%, respectively, for upstaged tumours vs 93% and 97% for non-upstaged tumours.

Effect of Surgical Approach on Survival Outcomes after Upstaging

In the upstaged cohort, survival data were available for 602 patients for DFS analysis and 604 for OS analysis. Patients and tumour characteristics stratified by type of surgery are reported on Tables 2 and 3. After a median follow-up of 21 months, 102 patients had disease recurrence. Local

Table 1 Baseline demographic and clinical characteristics of the study cohort.

	Total (*N* = 4395)	pT1-2 (*N* = 3728)	pT3a (*N* = 667)	*P* value
Age, median (IQR) years	61.4 (52.0–69.3)	61.0 (52.0–68.9)	66.0 (57.6–73.0)	<0.001
Male gender, *n* (%)	2958 (67)	2478 (67)	480 (72)	0.006
ASA score, *n* (%)				
1	1184 (31)	1064 (33)	120 (21)	<0.001
2	1908 (50)	1609 (50)	299 (54)	
3–4	702 (19)	564 (17)	138 (25)	
Missing observations	601	491	110	
ECOG PS, *n* (%)				
0	2612 (72)	2256 (73)	356 (69)	0.14
1	776 (22)	660 (21)	116 (23)	
2	196 (5.2)	157 (5.1)	39 (7.5)	
3–4	30 (0.8)	27 (0.9)	3 (0.5)	
Missing observations	781	628	153	
Symptoms at diagnosis, *n* (%)				
No symptoms	3192 (75)	2777 (77)	415 (65)	<0.001
Local symptoms	884 (21)	705 (20)	179 (28)	
General symptoms	187 (4)	138 (3)	49 (7)	
Missing observations	132	108	24	
Clinical tumour size, median (IQR) cm	3.8 (2.6–5.0)	3.5 (2.5–5.0)	5.5 (4.0–7.0)	<0.001
Clinical T stage, *n* (%)				
cT1a	2597 (59)	2405 (65)	192 (29)	<0.001
cT1b	1442 (33)	1114 (30)	328 (49)	
cT2a	356 (8)	209 (5)	147 (22)	
RENAL score, median (IQR)	7 (5–9)	7 (5–9)	8 (7–10)	<0.001
Missing observations	968	816	152	
Exophytic nature, *n* (%)				
≥50%	1507 (50)	1293 (51)	214 (44)	0.001
<50%	1233 (41)	1000 (39)	233 (48)	
0%	284 (9)	247 (10)	37 (8)	
Missing observations	1371	1188	183	
Posterior location, *n* (%)	1232 (43)	1039 (44)	193 (41)	0.4
Missing observations	1539	1341	198	
Hilar location, *n* (%)	2876 (79)	536 (18)	208 (36)	<0.001
Missing observations	775	688	87	
Polar location, *n* (%)				
Superior	1201 (37)	1025 (36)	176 (37)	0.3
Equatorial	986 (30)	832 (30)	154 (32)	
Inferior	1099 (33)	954 (34)	145 (31)	
Missing observations	1109	917	192	
Partial nephrectomy, *n* (%)	3454 (79)	3117 (84)	337 (51)	–

ASA, American Society of Anesthesiologists; ECOG PS, Eastern Cooperative Oncology Group performance status; IQR, interquartile range.

Fig. 1 Kaplan-Meier curves of disease-free survival (**A**) and overall survival (**B**) according to upstaging status of the tumour (upstaged = pT3a vs non-upstaged = pT1-2).

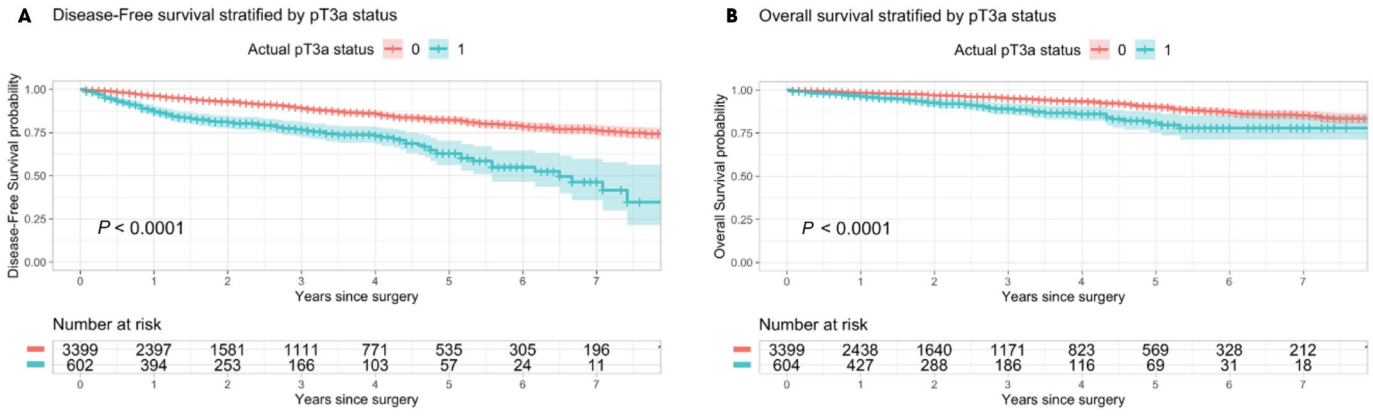


Table 2 Clinical and surgical characteristics of the upstaged cohort.

	PN (*N* = 300)	RN (*N* = 304)	*P* value
Age, median (IQR) years	65.0 (57.8–72.0)	67.0 (58.1–75.9)	0.02
Male gender, *n* (%)	228 (76)	210 (69)	0.07
ASA score, *n* (%)			
1	59 (22)	46 (20)	0.18
2	149 (55)	113 (50)	
3–4	61 (23)	68 (30)	
Missing observations	31	77	
Symptoms at diagnosis, *n* (%)			
No symptoms	214 (74)	159 (55)	<0.001
Local symptoms	62 (22)	103 (35)	
General symptoms	13 (4)	29 (10)	
Missing observations	11	13	
Clinical tumour size, median (IQR) cm	4.5 (3.3–6.0)	6.0 (5.0–8.0)	<0.001
Clinical T stage, *n* (%)			
cT1a	142 (47)	37 (12)	–
cT1b	122 (41)	168 (55)	
cT2a	36 (12)	99 (33)	
RENAL score, median (IQR)	8 (6–9)	9 (8–10)	<0.001
Missing observations	27	121	
Exophytic nature, *n* (%)			
≥50%	137 (59)	51 (27)	<0.001
<50%	90 (38)	121 (63)	
0%	6 (3)	19 (10)	
Missing observations	67	113	
Hilar location, *n* (%)	78 (29)	108 (43)	0.002
Missing observations	32	52	
Polar location, *n* (%)			
Superior	93 (43)	68 (33)	0.002
Equatorial	50 (23)	82 (40)	
Inferior	73 (34)	57 (27)	
Missing observations	84	97	
Surgical approach, *n* (%)			
Robot-assisted PN	180 (60)	–	–
Pure laparoscopic PN	17 (6)	–	
Open PN	103 (34)	–	
Robot-assisted RN	–	32 (11)	
Pure laparoscopic RN	–	270 (89)	
Missing observations	–	2	

ASA, American Society of Anesthesiologists; IQR, interquartile range; PN, partial nephrectomy; RN, radical nephrectomy.

recurrence was found in 36 patients and distant progression (DP) in 83 patients. DP accounted for 81% of recurrences.

Disease recurrence was diagnosed in 64 and 38 cases after RN and PN, respectively.

Descriptive analysis found that treatment type was not significantly associated with survival outcomes (DFS: hazard ratio [HR] 0.74 [95% CI 0.52, 1.05], $P = 0.09$; OS: HR 0.65 [95% CI 0.37, 1.14]; $P = 0.13$).

After G-computation adjustment on confounders, no significant difference in DFS and OS were observed between PN and RN, irrespective of the selection procedure (second procedure, DFS: marginal HR 1.08 [95% CI 0.74, 1.51], $P = 0.7$; OS, marginal HR 1.03 [0.57, 1.72], $P > 0.9$; Fig. 2A, B), for the first and third procedures results are shown in Fig. S1. There was also no difference in DFS and OS between OPN and LPN (DFS: marginal HR 0.82 [0.44, 1.41], $P = 0.4$; OS: marginal HR 1.04 [0.36, 2.56], $P = 0.9$).

Multivariable analyses for pre- and postoperative predictors of disease recurrence and all-cause mortality in upstaged tumours are shown in Table S1. After adjusting for other covariates, type of surgery (PN vs RN) was not significantly associated with disease recurrence (HR 1.13 [95% CI 0.74, 1.70]; $P = 0.6$) nor with all-cause mortality (HR 0.92 [95% CI 0.48, 1.79]; $P = 0.8$); neither was the surgical approach (laparoscopic vs open) in PN (disease recurrence: HR 1.25; $P = 0.5$; all-cause mortality: HR 1.02; $P > 0.9$).

Upstaging Predictive Model

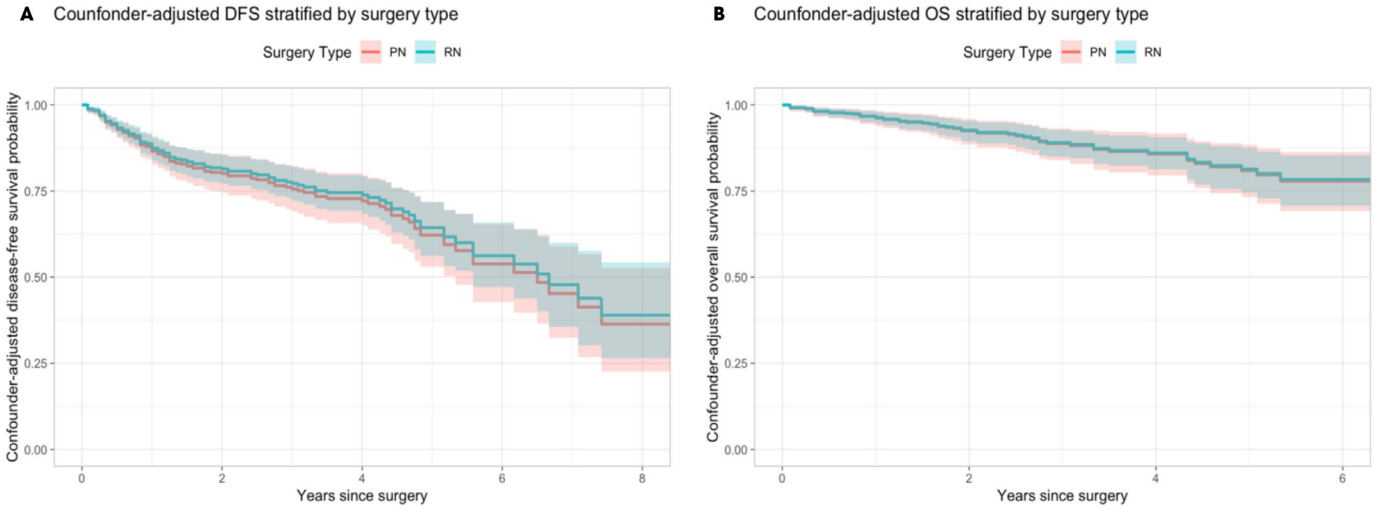
The cohort was split into one training cohort ($n = 2636$ from 17 centres) and one test cohort ($n = 1759$ from eight other centres) with an equivalent pT3a rate (15% each). ECOG PS, polar and rim location were not significantly associated with pT3a status, whereas the association between exophytic nature and pT3a status was not

Table 3 Pathological characteristics of the upstaged cohort.

	PN (*N* = 300)	RN (*N* = 304)	*P* value
Pathological tumour size, median (IQR) cm	4.0 (3.0–5.5)	6.0 (4.9–7.9)	<0.001
Missing observations	2	1	
Histology, *n* (%)			
Clear cell	220 (73)	263 (86)	<0.001
Papillary	41 (14)	17 (6)	
Chromophobe	27 (9)	12 (4)	
Other	12 (4)	12 (4)	
Furhman grade, *n* (%)			
1	5 (2)	3 (1)	<0.001
2	120 (41)	99 (33)	
3	133 (46)	128 (43)	
4	31 (11)	70 (23)	
Missing observations	11	4	
Perinephric fat invasion, *n* (%)	213 (72)	154 (52)	<0.001
Missing observations	4	7	
Sinusal fat invasion, *n* (%)	70 (29)	158 (57)	<0.001
Missing observations	61	28	
Renal vein invasion, *n* (%)	46 (19)	82 (28)	0.031
Missing observations	64	6	
Microvascular invasion, *n* (%)	51 (18)	91 (32)	<0.001
Missing observations	12	16	
Necrosis, *n* (%)	84 (31)	123 (42)	0.006
Missing observations	25	9	
Sarcomatoid feature, *n* (%)	20 (7.2)	31 (11)	0.15
Missing observations	24	20	
Positive surgical margin, *n* (%)	29 (10)	4 (1.6)	<0.001
Missing observations	21	59	

PN, partial nephrectomy; RN, radical nephrectomy.

Fig. 2 Estimated disease-free survival (**A**) and estimated overall survival (**B**) according to surgical type (partial vs radical nephrectomy) after G-computation adjustment on confounders with a backward stepwise selection procedure until reaching only *P*-values <0.1, in the upstaging cohort. PN, partial nephrectomy; RN, radical nephrectomy.



consistent between the training and test cohorts (Table S2). These four features did not improve the predictive performances of the multivariable MLA and were excluded from the prediction model. Preoperative features independently associated with pT3a upstaging and included in the final UroCCR-15 predictive model are shown in Table S3 with their corresponding odds ratios (ORs).

Among the seven MLAs tested, logistic regression was the most effective, with a prAUC of 0.41 on the test dataset and

Fig. 3 Example of an Individual prediction of the risk of upstaging by means of SHAP values. The patient had a predicted 32% probability of upstaging influenced by three pejorative factors (ASA score 2, male gender and a tumour size of 8 cm) counterbalanced by two protective factors (young age and non-hilar location of the tumour).

an area under the ROC of 0.77 (Fig. S2). The calibration curve for the model is shown in the Fig. S3.

Using SHAP values, we obtained an individual prediction for the risk of upstaging for every single patient. Figure 3 shows one example.

Discussion

From this multi-institutional contemporary cohort study, we developed the UroCCR-15 predictive model for predicting the upstaging of individual tumours from clinically localized renal tumours to locally advanced tumours (pT3a) on final pathology. The model could help decision making in the treatment of these tumours. In addition, we found that, even when PN was performed laparoscopically, it did not seem to undermine the oncological outcomes of unexpected pT3a RCC.

In RCC, prognosis worsens with stage, and T3 stage has a cancer-specific survival HR of 5.20 (4.36–6.21) compared to T1 stage [4]; therefore, follow-up after surgery needs to be stratified based on tumour characteristics. It is currently recommended to rely on validated prognostic models such as the University of California Los Angeles Integrated Staging System (UISS) and Leibovich score [4]. In these models, T3a tumours are of intermediate or high risk of recurrence, require closer follow-up than T1–T2 tumours and are potential candidates for postoperative treatment.

Adjuvant therapies for intermediate- and high-risk RCC are currently being assessed with one positive study (S-TRAC [17]) on DFS in high-risk RCC after nephrectomy with tyrosine-kinase inhibitors. More recently, the KEYNOTE-564 phase III study obtained positive results comparing a checkpoint inhibitor (pembrolizumab) vs placebo on DFS (pembrolizumab; HR 0.68; 95% confidence interval [CI] 0.53–0.87; $P = 0.002$) [18]. Whereas, three other trials, two with an adjuvant setting (IMmotion-010 [19] and CheckMate-914 [NCT03138512]) and one with a peri-operative setting (PROSPER [NCT03055013]), failed to show any improvement of DFS with immunotherapy. More evidence about the benefits of such therapies on DFS and OS are still pending, with ongoing phase III trials in a neoadjuvant and adjuvant setting.

Prediction of pT3a could lead to adequate selection of patients with estimated worse prognosis who may then reap greater benefits from preoperative systemic therapy. Neoadjuvant treatment could also allow more technically challenging conservative surgeries or PN for solitary or bilateral kidney tumours.

In accordance with the current trend to perform PN on larger tumours, we wanted to get assurance of its oncological safety although larger tumour size and greater stage have been widely shown to be associated with an increased risk of pathological upstaging in RCCs ≤7 cm. cT2 tumours were then included in our study, filling a gap in the literature since only two previous studies by Hamilton et al. [1] and Patel et al. [20] have reported data on both cT1 and cT2 tumours. A recent review by Chung et al. [21], gathering 12 studies on upstaged cT1/pT3a tumours, reported similar recurrence-free survival (RFS) for both PN and RN and better OS for PN (HR 0.74; 95% CI, 0.57–0.95; $P = 0.02$). This suggests that the benefit of nephron-sparing surgery (NSS) on cardiovascular health might be superior to the potential risk of insufficient cancer control.

Concerns about the risk of local recurrence after PN have tempered its spreading in the past. However, nowadays there is growing evidence that the recurrence rate after PN is indeed low, estimated at 1.4%–6.4% compared to 1.4%–2.9% for RN [22]. Patel et al. [20] reported no difference in recurrence rate for PN vs RN in pT3a upstaged tumours (26.3% vs 29.7%; $P = 0.287$). Additionally, local relapse seems to represent a modest part of recurrences compared to DP. In recent studies, upstaged tumours recurred in approximately 14.4%–29.7% of the cases [1,18,19]. DP accounted for nearly 90% of all recurrences, representing a greater ratio than for localized tumours (72%–82%). Our findings are consistent with these data in that our recurrence rate for upstaged tumours was 16.9% with 81.4% DP. These results suggest that pT3a RCC is more likely to disseminate in a systemic fashion than locally. PN may not be detrimental in this incidental upstaging setting since the battlefield seems to be, not the remaining kidney, but instead the risk of micrometastases. Thus, offering PN for large tumours when technically feasible seems to be justified and reasonable.

Major strengths of our study are the large cohort size, the use of a group of external centres for validation, optimizing the predictive performances of our model and allowing the generalizability of our results at least for the European population, then, the development of a ‘user-friendly’ tool

rather than traditional nomograms. Moreover, the predictive ability of our model was properly evaluated thanks to the use of precision/recall curves. The prAUC is a summary metric that reflects the ability of the predictive signature to predict patients with pT3a status. One also talks about average precision, as it is calculated as the precision averaged across all values of recall between 0 and 1. The values of the prAUC range from 0 to 1, with 1 for a perfect classifier and the baseline (expected value for random guessing) being the actual rate of pT3a in the cohort (0.15). Contrary to the area under the ROC curve, where the baseline value is 0.5, the prAUC thus depends on the observed prevalence and tends to 0 when the latter decreases. In case of unbalanced outcome distribution, such as pT3a rate in our study, the area under the ROC tends to be underestimated, whereas prAUC is not [15].

In this study, the upstaging rate (15.2%), for both cT1 and cT2 RCCs, is consistent with previous literature [1,15], as are the preoperative variables included in our model. Veccia et al. [6] found that age, tumour size and RENAL score were baseline predictors of cT1/pT3a upstaging (OR 1.03, $P < 0.00001$; OR 1.51, $P < 0.00001$; OR 2.80, $P = 0.0004$, respectively). In our model, these three covariates were among the four most important predictors, in addition to hilar location, hitherto reported in other studies [18,20] as well as male gender [5,21].

The present study is the first to date to use multiple MLAs with the aim of predicting pT3a upstaging in localized renal tumours. A similar method was applied for predicting recurrence within 5 years of RCC surgery by Kim et al. [23]. By comparing eight algorithms, they obtained the highest area under the ROC curve of 0.836 from a naive Bayes model.

Three studies have already been published on the development of artificial intelligence models for predicting pT3a upstaging [9-11] in a total of 146 RCC patients. De la Barra et al. evaluated both cT1 and cT2 tumours, as we did, whereas others focused only on cT1. However, none of them developed competing algorithms and only one study generated a training/test split method [11].

While PN and RN groups were not comparable, to ensure the validity of our comparative analysis, we had to account for confounders in the causal analysis. Propensity scores have long been popular for dealing with such pitfalls when randomization is not suitable, although concerns about inherent biases have risen. G-computation, which is also a way of estimating causal effects, is based on the prediction of potential outcomes for each subject under each exposure status. G-computation has been proven to reduce residual variance and increase the accuracy of predictions [13].

Another strength of this study is the contemporary character of our cohort, as it reflects current surgical and pathological practice compared to standard prognostic models (UISS [24] and Leibovich score [25]), which are based on historical cohorts.

In the scope of personalized medicine, individual risk estimation from the UroCCR-15 predictive model seems to be more accurate than traditional risk groups where all the patients in the same group have the same probability of outcome. Precise risk estimation is crucial for clinical decision making as patient-specific rate can be used as a threshold for the choice of disease management. In the setting of pT3a upstaging risk, prediction has a potential role in screening low-risk patients for alternative local treatments such as active surveillance or ablative therapies. In the elderly and comorbid population, oncological safety has been shown to be relatively achievable with surveillance of small renal masses. Beisland et al. [26] reported a 5-year cancer-specific survival rate of 93.3% in monitoring cT1 tumours. A low-risk prediction from the UroCCR-15 model could empower such a decision.

We must acknowledge several limitations of our study. First, the study lacks centralized reviewing of imaging and pathological data, specimens were exclusively examined by expert uropathologists. As a collective database, data completion is left to the discretion of each centre, possibly resulting in evaluation or follow-up biases. It has been shown that tumour staging is subject to inter-observer variability [27]. Second, the retrospective design of our study has possibly led to selection and information biases. Some of these were prevented by the prospective collection of data but we must consider that the cohort gathers patients from a large period with a risk of changes in disease management or data acquisition. Third, the amount of missing data should be mentioned, even though it was managed by the machine-learning imputation strategy [28]. Fourth, the median follow-up in our cohort was relatively short (21 months) and may thus have prevented us from observing most cases of relapse since mean delay of recurrence for RCC has been evaluated at 77.5 months [29]. Longer follow-up is warranted to accurately explore DFS. Moreover, the use of MLAs is new in the scope of renal tumours. More studies are needed to ascertain the interpretability and clinical utility of our prediction model; validation using external datasets in particular is necessary.

Finally, we did not consider biological factors in our predictive model due to lack of data. For example, it has been recognized that systemic inflammatory response, especially neutrophils, have decisive roles in initiating cancer, promoting metastases and helping tumour growth via different cellular and biochemical pathways [30]. Inflammatory biomarkers such as neutrophil-to-lymphocyte ratio have become important prognostic

markers in many types of cancer, including RCC. A recent work showed that inflammatory markers, including neutrophil-to-lymphocyte ratio, could predict pT3a upstaging and other adverse pathological features [11]. Therefore, future predictive models should integrate biological data as part of this emergent precision medicine. Further work with prospective validation cohorts will be conducted using the UroCCR database to identify the best predictive model for future clinical management.

Acknowledgements

Additional statistical support was provided by (alphabetical order): G. Etchepare and C. Lopez of Sophia Genetics, France. The authors wish to thank Teresa Sawyers, British Medical Writer at the B.E.S.P.I.M. (Laboratoire de Biostatistique, Epidémiologie clinique, Santé Publique, Innovation et Méthodologie), Nîmes University Hospital, France, for her help in editing this text.

Funding

No funding or other financial support was received for this study.

Disclosure of Interests

None declared.

References

1 **Hamilton ZA, Capitanio U, Pruthi D et al.** Risk factors for upstaging, recurrence, and mortality in clinical T1-2 renal cell carcinoma patients upstaged to pT3a disease: an international analysis utilizing the 8th edition of the tumor-node-metastasis staging criteria. *Urology* 2020; 138: 60–8

2 **Renard AS, Nedelcu C, Paisant A et al.** Is multidetector CT-scan able to detect T3a renal tumor before surgery? *Scand J Urol* 2019; 53: 350–5

3 **Landman J, Park JY, Zhao C et al.** Preoperative computed tomography assessment for perinephric fat invasion: comparison with pathological staging. *J Comput Assist Tomogr* 2017; 41: 702–7

4 **Ljungberg B, Albiges L, Abu-Ghanem Y et al.** European association of urology guidelines on renal cell carcinoma: the 2019 update. *Eur Urol* 2019; 75: 799–810

5 **Ghanie A, Formica MK, Wang D, Bratslavsky G, Stewart T.** Pathological upstaging of clinical T1 renal cell carcinoma: an analysis of 115,835 patients from National Cancer Data Base, 2004–2013. *Int Urol Nephrol* 2018; 50: 237–45

6 **Veccia A, Falagario U, Martini A et al.** Upstaging to pT3a in patients undergoing partial or radical nephrectomy for cT1 renal tumors: a systematic review and meta-analysis of outcomes and predictive factors. *Eur Urol Focus* 2021; 7: 574–81

7 **Obermeyer Z, Emanuel EJ.** Predicting the future – big data, machine learning, and clinical medicine. *N Engl J Med* 2016; 375: 1216–9

8 **Jayatilake SMDAC, Ganegoda GU.** Involvement of machine learning tools in healthcare decision making. *J Healthc Eng* 2021; 2021: e6679512

9 **de la Barra CC, González PG, Baeza MÁ, Pérez OP, Cruzat JD.** A preoperative model to predict pT3 upstaging in clinically localized renal cell carcinoma. *Cent Eur J Urol* 2020; 73: 173–7

10 **Nocera L, Stolzenbach LF, Ruvolo CC et al.** Predicting the risk of pT3a stage in cT1 clear cell renal cell carcinoma. *Eur J Surg Oncol J Eur Soc Surg Oncol Br Assoc Surg Oncol* 2021; 47: 1187–90

11 **Liu H, Wang Z, Peng E, Chen Z, Tang K, Xia D.** Added value of systemic inflammation markers in predicting clinical stage T1 renal cell carcinoma pathologically upstaged to T3a. *Front Oncol* 2021; 11: 679536

12 **Edge SB, Compton CC.** The American joint committee on cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* 2010; 17: 1471–4

13 **Chatton A, Le Borgne F, Leyrat C et al.** G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study. *Sci Rep* 2020; 10: 9219

14 **Uddin S, Khan A, Hossain ME, Moni MA.** Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak* 2019; 19: 281

15 **Saito T, Rehmsmeier M.** The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 2015; 10: e0118432

16 **Lundberg SM, Lee SI.** A unified approach to interpreting model predictions. In **Curran Associates Inc.** ed, *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., 2017: 4768–77

17 **Ravaud A, Motzer RJ, Pandha HS et al.** Adjuvant sunitinib in high-risk renal-cell carcinoma after nephrectomy. *N Engl J Med* 2016; 375: 2246–54

18 **Choueiri TK, Tomczak P, Park SH et al.** Adjuvant pembrolizumab after nephrectomy in renal-cell carcinoma. *N Engl J Med* 2021; 385: 683–94

19 **Pal SK, Uzzo R, Karam JA et al.** Adjuvant atezolizumab versus placebo for patients with renal cell carcinoma at increased risk of recurrence following resection (IMmotion010): a multicentre, randomised, double-blind, phase 3 trial. *Lancet Lond Engl* 2022; 400: 1103–16

20 **Patel SH, Uzzo RG, Larcher A et al.** Oncologic and functional outcomes of radical and partial nephrectomy in pT3a pathologically upstaged renal cell carcinoma: a multi-institutional analysis. *Clin Genitourin Cancer* 2020; 18: e723–9

21 **Chung DY, Kang DH, Kim JW, Kim DK, Lee JY, Cho KS.** Comparison of oncologic outcomes between partial nephrectomy and radical nephrectomy in patients who were upstaged from cT1 renal tumor to pT3a renal cell carcinoma: an updated systematic review and meta-analysis. *Ther Adv Urol* 2020; 12: 1756287220981508

22 **Wood EL, Adibi M, Qiao W et al.** Local tumor bed recurrence following partial nephrectomy in patients with small renal masses. *J Urol* 2018; 199: 393–400

23 **Kim H, Lee SJ, Park SJ, Choi IY, Hong SH.** Machine learning approach to predict the probability of recurrence of renal cell carcinoma after surgery: prediction model development study. *JMIR Med Inform* 2021; 9: e25635

24 **Patard JJ, Kim HL, Lam JS et al.** Use of the University of California los Angeles integrated staging system to predict survival in renal cell carcinoma: an international multicenter study. *J Clin Oncol* 2004; 22: 3316–22

25 **Leibovich BC, Blute ML, Cheville JC et al.** Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *Cancer* 2003; 97: 1663–71

26 **Beisland C, Hjelle KM, Reisaeter LAR, Bostad L.** Observation should be considered as an alternative in management of renal masses in older and comorbid patients. *Eur Urol* 2009; 55: 1419–27

27 **Spaliviero M, Poon BY, Aras O et al.** Interobserver variability of R.E.N.a.L., PADUA, and centrality index nephrometry score systems. *World J Urol* 2015; 33: 853–8

28 **Troyanskaya O, Cantor M, Sherlock G et al.** Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001; 17: 520–5

29 **Bernhard JC, Pantuck AJ, Wallerand H et al.** Predictive factors for ipsilateral recurrence after nephron-sparing surgery in renal cell carcinoma. *Eur Urol* 2010; 57: 1080–6

30 **Long W, Chen J, Gao C, Lin Z, Xie X, Dai H.** Brief review on the roles of neutrophils in cancer development. *J Leukoc Biol* 2021; 109: 407–13

Correspondence: Astrid Boulenger de Hauteclocque, Department of Urology, Bordeaux University Hospital, Place Amélie Raba-Léon, 33000 Bordeaux, France.

e-mail: astrid.dehauteclocque@gmail.com

Abbreviations: ASA, American Society of Anesthesiologists; DFS, disease-free survival; DP, distant progression; ECOG PS, Eastern Cooperative Oncology Group performance status; HR, hazard ratio; MLA, machine-learning algorithm; OS, overall survival; PN, partial nephrectomy; prAUC, area under the precision/recall curve; RN, radical nephrectomy; ROC, receiver-operating characteristic; SHAP, SHapley Additive exPlanations; UISS, University of California Los Angeles Integrated Staging System.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Estimated disease-free survival (**A**) and estimated overall survival (**B**) according to surgical type (partial vs radical nephrectomy) after adjustment for confounders in the upstaging cohort.

Figure S2. Precision/recall curves (**A**, **B**) and receiver-operating curves (**C**, **D**) of the upstaging prediction algorithms applied to the training cohort (**A**, **C**) and the test cohort (**B**, **D**).

Figure S3. Calibration curve of the logistic regression for the prediction of upstaging.

Table S1. Filtered multivariable Cox regression analyses for predictors of disease recurrence and overall mortality in upstaged tumours.

Table S2. Baseline demographic and clinical characteristics of the train and test cohorts.

Table S3. Multivariable logistic regression for the prediction of upstaging.