Angiogenesis (2019) 22:383–395
https://doi.org/10.1007/s10456-019-09664-2

ORIGINAL PAPER

A new prognostic model for survival in second line for metastatic renal cell carcinoma: development and external validation

Lisa Derosa[1,2,3,4] · Mohamed Amine Bayar[5,6] · Laurence Albiges[1] · Gwénaël Le Teuff[5,6] · Bernard Escudier[1]

Received: 2 October 2018 / Accepted: 21 January 2019 / Published online: 9 February 2019

Abstract

Background In patients with metastatic renal cell carcinoma (mRCC), the oncologic benefit of second-line treatment for high volume tumors or presence of more than five risk factors remain to be defined. Our aim was to develop and externally validate a new model most likely to correctly predict overall survival (OS) categories in second line.

Method mRCC patients treated within clinical trials at Gustave Roussy Cancer Campus (GRCC) formed the discovery set. Patients from two phase III trials from Pfizer database (PFIZERDB), AXIS (NCT00678392), and INTORSECT (NCT00474786), formed the external validation set. New prognostic factors were analyzed using a multivariable Cox model with a backward selection procedure. Performance of the GRCC model and the prognostic classification scheme derived from it, measuring by R^2, c-index, and calibration, was evaluated on the validation set and compared to MSKCC and IMDC models.

Results Two hundred and twenty-one patients were included in the GRCC cohort and 855 patients in the PFIZERDB. Median OS was similar in the discovery and validation cohorts (16.8 [95% CI 12.9–21.7] and 15.3 [13.6–17.2] months, respectively). Backward selection procedure identified time from first to second-line treatment and tumor burden as new independent prognostic factors significantly associated to OS after adjusting for IMDC prognostic factors (HR 1.68 [1.23–2.31] and 1.43 [1.03–1.99], respectively). Dividing patients into four risk groups, based on the number of factors selected in GRCC model, median OS from the start of second line in the validation cohort was not reached (NE) [95% CI 24.9–NE] in the favorable risk group ($n=20$), 21.8 months [18.6–28.2] in the intermediate-risk group ($n=367$), 12.7 months [11.0–15.8] in the low poor-risk group ($n=347$), and 5.5 months [4.7–6.4] in the high poor-risk group ($n=121$). Finally, this model and its prognostic classification scheme provided the better fit, with higher R^2 and higher c-index compared to other possible classification schemes.

Conclusion A new prognostic model was developed and validated to estimate overall survival of patients with previously treated mRCC. This model is an easy-to-use tool that allows accurate estimation of patient survival to inform decision making and follow-up after first line for mRCC.

Keywords MSKCC · IMDC · Metastatic renal cell carcinoma · Second line · Prognostic model · Tumor burden · Time from first to second line

Introduction

Prognostic models in metastatic renal cell carcinoma (mRCC) can help identifying patients more likely to benefit from standard therapies, as well as interpreting drug effectiveness. Such prognostic models are currently used for treatment selection in treatment naïve patients but also for stratification in clinical trials, for comparison of cohorts in the literature and finally for patient counseling. In second line, prognostic factors are not as commonly used [1, 2]. In mRCC, the current available prognostic models in the second line setting are the Memorial Sloan Kettering Cancer

Gwénaël Le Teuff and Bernard Escudier have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10456-019-09664-2) contains supplementary material, which is available to authorized users.

✉ Lisa Derosa
deros.lisa@gmail.com

Extended author information available on the last page of the article

Center (MSKCC) and the International Metastatic Renal Cell Carcinoma Database Consortium (IMDC) models. The first [3, 4] was derived from a cohort of 251 patients between 1957 and 2002, where 50% of patients received cytokines as first-line treatment. Three factors were considered as prognostic: Karnofsky performance status (KPS), hypercalcemia, and anemia. The IMDC model, built by extension of MSKCC, was initially developed for first-line therapy [5]. Then it has been validated in second-line targeted therapy (TT) from a large cohort of 1021 patients treated with TT between 2005 and 2012 [6]. Only five of six predefined risk factors in first line (KPS, time from diagnosis to treatment, neutrophil, platelet count, hemoglobin, and calcium concentration) were confirmed as independent predictors of overall survival (OS) in second line and were also found prognostic. In this context, the six-factor model showed relatively better predictive ability than the MSKCC model but it remains unclear whether these models can accurately predict the OS of these patients. In fact, the majority of patients (more than 60%) with mRCC after primary care are classified by the IMDC schema in intermediate-risk group [6]. This represents a heterogeneous group of patients that probably has to be better stratified. Besides, no model used so far takes into account some potentially important prognostic determinants for second line, such as tumor burden (TB) [7] or number of metastatic sites prior second line. Finally, the efficacy of second-line treatment seems to be independent of the response to first line [8] but no one has ever investigated whether time from first- to second-line therapy, tumor shrinkage (TS) during first line, TB before second line, and number of metastatic sites prior second line can improve the predictive ability.

Methods

Study design and patients

Discovery set came from the Gustave Roussy Cancer Campus (GRCC) electronic database. We included only patients initially treated within approved clinical trials conducted between Jan 1, 2005, and Dec 30, 2014 and who received a second line TT, VEGF inhibitor, or mTOR inhibitor, after progressive disease on first-line TT (patients who received cytokines as first-line therapy were not eligible). To externally validate the new model, we used the Pfizer database (PFIZERDB), composed by two independent cohorts from two prospective second line phase III studies: AXIS (NCT00678392) and INTORSECT (NCT00474786). In the first trial, axitinib was compared to sorafenib in patients previously treated with any approved therapy for mRCC at that time, mostly sunitinib or cytokines [9]. In the second trial, temsirolimus was compared to sorafenib in patients previously treated exclusively with sunitinib [10]. From PFIZERDB, only patients treated in first line with sunitinib or bevacizumab alone or in combination were selected as validation set, excluding all patients who received cytokines.

Procedures

We collected demographic, baseline characteristics, and outcome before first- and second-line treatment. Target lesions were selected at baseline with tumor measurement according to RECIST version 1.1 [11]. For each baseline and follow-up imaging study, the longest axis of each target lesion was recorded to the nearest millimeter and the sum of the long-axis diameter (SLD) of target lesions was calculated to assess TB [7, 12, 13]. The percentage of change in TB was assessed at every available study time point. For each patient, the time point with the maximum percentage change in the SLD of target lesions was defined as TS [14]. New lesions were not assessed for TS.

Statistical analysis

The main endpoint was OS defined as the time from the start of second-line therapy to the date of death whatever the cause or the date of the last follow-up for patients still alive. OS was estimated by using the Kaplan–Meier method [15]. Median follow-up was estimated by the Schemper's method [16]. We developed a new prognostic model (thereafter called GRCC model) using a Cox proportional hazard model. This model includes the prognostic factors of IMDC [6], i.e., time from diagnosis to treatment (< 1 year), KPS (< 80), hemoglobin [< lower limit normal (LLN)], corrected calcium [> upper limit normal (ULN)], neutrophil (> ULN), and platelet (> ULN). In addition, the potential risk factors, namely, TS during first-line TT, time from first to second line, TB prior to second line, and number of metastatic sites prior second line were tested using a backward selection procedure after controlling for IMDC factors. The proportional hazards assumption was evaluated through the Schoenfeld residual and the functional form of continuous predictors by a spline function. Interactions between IMDC factors and new selected factors were investigated. Some statistical measures were used to evaluate the GRCC model performance. The Kent and O'Quigley R^2 statistic [17] based on the explained variation was used to measure the overall performance (value close to 1 indicates good dependence between predictors and censored outcome). The discriminative ability was evaluated using the Uno's c statistic [18] (value close to 1 indicates good discrimination). The calibration was measured by estimating the slope calibration and by comparing the observed and model-based predicted OS probability at 1, 2, and 3 years. As for MSKCC and IMDC models, we also derived a simple prognostic classification

scheme according to different risk groups defined in counting the number of factors of our final GRCC model. This classification is the result of the comparison of several classifications and the one with the best performance in terms of Akaike information criterion (AIC [19]) and R^2 was retained. In addition to the previous measures used for the assessment of GRCC model, we also evaluated the discriminant ability of the selected prognostic classification scheme by representing the Kaplan–Meier OS curves between the risk groups and by estimating the hazard ratios (HR) between risk groups using a Cox model including an indicator of the risk groups. The performance of GRCC model and the prognostic classification scheme was measured on the validation dataset, and a comparison with MSKCC and IMDC models (and their risk groups) were also performed both in the discovery and validation datasets using likelihood-ratio test for nested models and AIC criterion for non-nested.

All models were derived from the complete cases, i.e., patients with non-missing information for the risk factors and with a significance level of 5% for risk factors selection. Preparation of databases (discovery and validation) was performed using SAS (version 9.4, SAS Institute, Cary, NC) and the statistical analyses with SAS and R. We used the kentoqnr SAS macro for Kent and O'Quigley [20] and the R survival and survAUC packages for survival analyses and Uno's c-index, respectively.

Results

Patients

From 316 patients identified in the GRCC dataset treated with a second line between Jan 1, 2005, and Dec 30, 2014, 222 patients (70.3%) were used to develop a new prognostic model. From 1235 patients enrolled in the validation set (512 and 723 from INTORSECT and AXIS trials, respectively), 855 (69.2%) patients had complete information on prognostic factors at initiation of second-line therapy. To detect a possible selection bias, we compared characteristics of patients included in the analysis to patients excluded for screen failure ($n=18$) or missing information on prognostic factors ($n=92$). No several differences were found except for the presence of mainly female or patients without nephrectomy in the excluded group (Fig. 1, Table S1).

Patients' and disease characteristics of the two datasets are compared in Table 1. The median follow-up was 49.4 months (Q1–Q3: 28.4–92.1) and 16.3 months (Q1–Q3: 10.3–27.8) for the discovery and validation datasets, respectively ($p<0.0001$). This is probably due to the difference between the two databases: a large center database for the discovery cohort versus a clinical trial database with data lock for the validation cohort. In spite of this, no significant

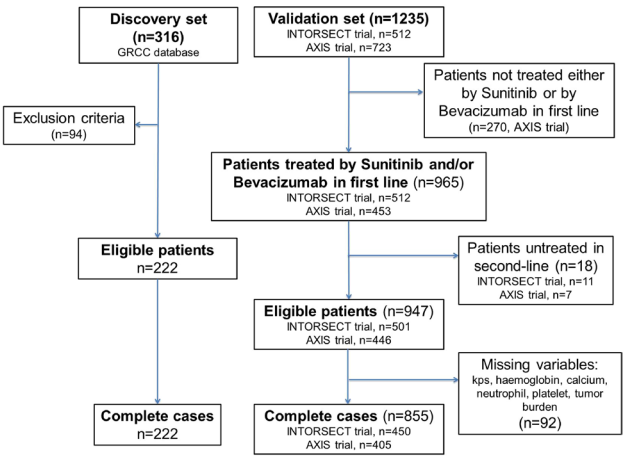


Fig. 1 Patient's flowchart of the discovery and the validation datasets

differences in OS were observed (log-rank test = 0.26) with a median of 16.8 months [95% CI 12.9, 21.7] [176 (79.3%) deaths] versus 15.3 months [13.6–17.2] [455 (53.2%) deaths] for the discovery and validation datasets, respectively (Fig. 2).

Patients' and disease characteristics differ mainly in terms of nephrectomy ($p<0.005$) and type of second line ($p<0.0001$), although mTOR inhibitors strategy appear to be similar. Second-line treatment was mainly everolimus (27.5%), sunitinib (24.8%), and sorafenib (22.1%) in the discovery dataset and sorafenib (50.4%), temsirolimus (25.9%), and axitinib (23.7%) in the validation dataset. For prognostic factors used in MSKCC and IMDC models, significant differences occur for the time from diagnosis to treatment ($p<0.0001$), KPS ($p<0.0001$), calcium ($p=0.002$), and neutrophil ($p=0.03$). The number of metastatic sites before second line and TB was also significantly different between the two cohorts ($p<0.0001$).

Building GRCC model and corresponding prognostic classification scheme

Backward selection procedure identified time from first to second line treatment and TB as new independent prognostic factors significantly associated to OS after adjusting for IMDC prognostic factors. Association between TS during first line and number of metastatic sites prior second line with OS were not significant (data not shown). Based on the non-monotonic relation between the log hazard ratio of death and these two new continuous predictors (Fig. S1), we transformed them into two classes using the cut-offs of 12 months and 100 mm based on statistical and clinical considerations. OS was shorter for patients with time from first to second line < 12 months compared to those with time interval ≥ 12 months (HR 1.68, 95% CI [1.23–2.31],

Table 1 Baseline patients' and disease characteristics in the discovery and validation sets

Characteristics	Discovery set ($n=222$)	Validation set ($n=855$)	p value$^{\perp}$
Sex			0.99
Male	166 (74.8)	639 (74.7)	
Female	56 (25.2)	216 (25.3)	
Clear cell histology			0.42
No	25 (11.3)	80 (9.4)	
Yes	197 (88.7)	767 (90.6)	
Missing	0	8	
Nephrectomy			< 0.005
No	15 (6.8)	128 (15.0)	
Yes	207 (93.2)	727 (85.0)	
Time from diagnosis to first line			< 0.0001
< 1 year	153 (68.9)	733 (85.7)	
≥ 1 year	69 (31.1)	122 (14.3)	
MSKCC at second line			0.37
Good	86 (38.7)	373 (43.6)	
Intermediate	135 (60.8)	476 (55.7)	
Poor	1 (0.5)	6 (0.7)	
IMDC at second line			0.10
Good	20 (9.1)	51 (6.0)	
Intermediate	145 (65.3)	615 (71.9)	
Poor	57 (25.7)	189 (22.1)	
Age at second line			0.59
< 60 years	107 (48.2)	395 (46.2)	
≥ 60 years	115 (51.8)	460 (53.8)	
KPS at second line			< 0.0001
Normal	179 (80.6)	834 (97.5)	
Alteration	43 (19.4)	21 (2.5)	
Number of metastatic sites at second line			< 0.0001
1	43 (19.4)	71 (8.3)	
> 1	179 (80.6)	784 (91.7)	
Type of second line			< 0.0001
Axitinib	19 (8.6)	203 (23.7)	
Sunitinib	55 (24.8)	0 (0.0)	
Sorafenib	49 (22.1)	431 (50.4)	
Temsirolimus	13 (5.9)	221 (25.9)	
Everolimus	61 (27.5)	0 (0.0)	
Nivolumab	12 (5.4)	0 (0.0)	
Other	13 (5.9)	0 (0.0)	
Hemoglobin at second line			0.83
Normal	100 (45.1)	392 (45.9)	
Low	122 (54.9)	463 (54.2)	
Corrected calcium at second line			0.002
Normal	217 (97.8)	784 (91.7)	
High	5 (2.2)	71 (8.3)	
Neutrophil count at second line			0.03
Normal	185 (83.3)	758 (88.7)	
High	37 (16.7)	97 (11.4)	
Platelets count at second line			0.75
Normal	182 (82.0)	693 (81.1)	
High	40 (18.0)	162 (18.9)	

Table 1 (continued)

Characteristics	Discovery set ($n=222$)	Validation set ($n=855$)	p value[⊥]
Time from first to second line (months)			0.07
Mean [min–max]	15.8 [2–77.5]	14.0 [1.1–120.3]	
Tumor burden at second line (mm)			< 0.0001
Mean [min–max]	96.8 [0–517.0]	122.9 [10.0–500.0]	
Median follow-up (months) [Q1–Q3]	49.4 [28.4–92.1]	16.3 [10.3–27.8]	< 0.0001
Deaths	176 (79.3)	455 (53.2)	
Median overall survival (months) 95% CI	16.8 [12.9–21.7]	15.3 [13.6–17.2]	0.26*

[⊥]Chi-square or Fisher's exact test for categorical variables and t test for continuous variables

*Log-rank test comparing the survival curve between the two groups

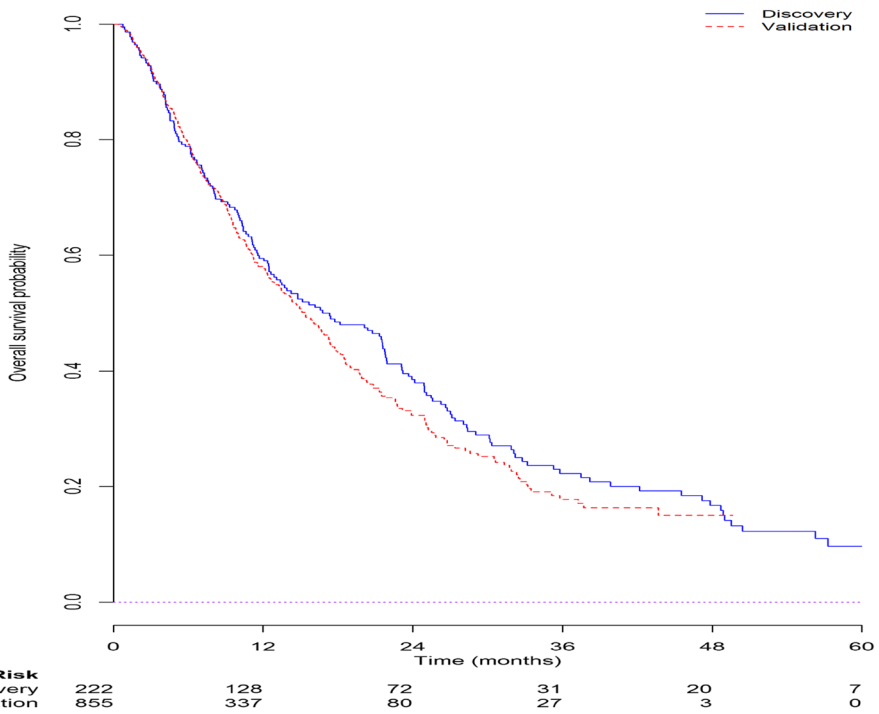


Fig. 2 Kaplan–Meier estimates of overall survival for the discovery and the validation sets

$\chi^2 = 10.4$), as for patients with a TB ≥ 100 mm compared to those with TB < 100 mm (HR 1.43 [1.03–1.99], $\chi^2 = 4.6$) (Table 2). No violation of the proportional hazards assumption was observed (global test, $p = 0.33$) and no significant interaction was observed between these two factors and IMDC prognostic factors (data not shown). Figure S2 shows the distribution of the prognostic index derived from this model in the two datasets.

The overall performance of GRCC model was $R^2 = 0.3496$ and c-index was 0.71 at 1, 2, and 3 years (Table 3). This discriminant ability of our new model can be illustrated by representing the Kaplan–Meier OS curves of groups defined by different quantiles of the prognostic index (Fig. S3, left) (see Table S3 for agreement between observed and model-based predicted OS probabilities at 1, 2, and 3 years). Regarding the number of prognostic factors retained in GRCC model (eight risk factors) compared to the number of deaths (176 deaths) and the empirical shrinkage parameter (0.91, close to 1), the risk of over-fitting seems limited.

From the number of prognostic factors selected in GRCC model (Table 2), a new prognostic classification scheme with four risk groups was defined: good risk 0 risk factor ($n = 13$ (5.9%), 5 deaths), intermediate risk 1–2 risk factors ($n = 98$ (44.1%), 69 deaths), low poor risk 3–4 risk factors ($n = 87$ (39.2%), 78 deaths), and high poor risk ≥ 5 risk factors ($n = 24$ (10.8%), 24 deaths). This model provides the better fit with the lowest AIC and higher R^2 compared to other possible classification schemes (see Table S4). Figure 3 left side reports well-separated Kaplan–Meier OS curves for the four risk groups in the discovery set with a median of 49.5 [26.3–NE], 26.6 [23.1–33.1], 12.5 [10.0–16.6], and 4.3 [3.9–7.2], respectively. Taking "0 risk" group as reference,

Table 2 Hazard ratios of prognostic factors included in the GRCC model for overall survival on the discovery set ($n = 222$, 176 deaths)

Risk factors	# Deaths/*n*	HR [95% CI]
Time from diagnosis to first line		
≥ 1 year	49/69	1
< 1 year	127/153	1.45 [1.02–2.05]
KPS		
≥ 80	137/179	1
< 80	39/43	2.08 [1.41–3.08]
Hemoglobin		
≥ LLN	73/100	1
< LLN	103/122	1.63 [1.15–2.32]
Calcium		
≤ ULN	171/217	1
> ULN	5/5	0.95 [0.38–2.41]
Neutrophil		
≤ ULN	143/185	1
> ULN	33/37	2.22 [1.44–3.42]
Platelet		
≤ ULN	137/182	1
> ULN	39/40	1.74 [1.14–2.66]
Time from first to second line		
≥ 12 months	71/104	1
< 12 months	105/118	1.68 [1.23–2.31]
Tumor burden		
< 100 mm	104/143	1
≥ 100 mm	72/79	1.43 [1.03–1.99]

HR hazard ratio, *CI* confidence interval, *KPS* Karnofsky performance status, *LLN* lower limit normal, *ULN* upper limit normal

the hazard ratios of intermediate, low poor, and high poor risk groups were HR 2.33 [0.94–5.80], 5.12 [2.06–12.73], and 15.03 [5.63–40.12] (Table 4), respectively, with $R^2 = 0.2848$ and c-index = 0.64 at 1, 2, and 3 years (Table 3). There is no substantial difference between observed and model-based predicted OS probabilities at 1, 2, and 3 years for each risk group except for the low risk group at 1 year (Table 5). As expected, the performance of such classification scheme is smaller than that based on the continuous prognostic index.

Validation of GRCC model and corresponding prognostic classification scheme

GRCC model performance, based on the prognostic index, is maintained on the validation dataset with a $R^2 = 0.3049$ and c-index = 0.72, 0.68, and 0.68 at 1, 2, and 3 years (Table 3). The calibration slope is close to 1 [= 0.92 (se = 0.06)] with some degrees of miscalibration especially with an over estimation of the estimated OS probabilities for the intermediate risk (at 2 and 3 years) and low poor risk (at 1 and 2 years) groups (Table S3). When applying the prognostic classification scheme on the patients of the validation dataset, the difference of prognostic between the four risk groups is maintained. The median OS is not reached [24.9–NE], 21.8 [18.6–28.2], 12.7 [11.0–15.8], 5.5 [4.7–6.4] in the good, intermediate, low poor risk, and high poor risk groups, respectively (Fig. 3 right) and the hazard ratios are similar to those estimated in the discovery dataset (Table 4). Some disagreements between observed and predicted probabilities occur for the first three risk groups and more particularly with an over estimation of the predicted probabilities at 2 and 3 years in intermediate-risk group (Table 5).

Comparison between GRCC, MSKCC, and IMDC models

In the discovery, we fitted two Cox models with the prognostic factors defining MSKCC and IMDC models. As these two models are nested in GRCC model, a likelihood-ratio test allows comparing these models. GRCC model greatly improves the fit of data compared to MSKCC model ($\chi^2 = 51.6$, 5 ddl, $p < 0.0001$) and IMDC model ($\chi^2 = 19.0$, 2 ddl, $p < 0.0001$). The HRs of these two models, reported in Table 6, were similar to those previously reported in the literature. GRCC model has best performance compared to these models: $R^2 = 0.3496$ versus 0.1562 and 0.2780 and c-index at 1 year = 0.71 versus 0.64 and 0.69 for MSKCC and IMDC, respectively. On the validation dataset, GRCC model has also a better performance: $R^2 = 0.3049$ versus 0.0983 and 0.2150 and c-index at 1 year = 0.72 versus 0.65 and 0.69 for MSKCC and IMDC, respectively. Similar results were also observed for c-index at 2 and 3 years (Table 3). When comparing the risk groups derived from GRCC, MSKCC, and IMDC models, the new prognostic classification scheme has, in the discovery set, the best goodness-of-fit: $AIC_{GRCC} = 1552$ versus $AIC_{MSKCC} = 1588$ and $AIC_{IMDC} = 1570$ (Tables 4, 7) and the best overall performance: $R^2_{GRCC} = 0.2848$ versus $R^2_{MSKCC} = 0.1345$ and $R^2_{IMDC} = 0.2264$ (Table 3). The discriminant ability of the new classification was also higher (c-index at 1 year = 0.64) compared to MSKCC risk groups (c-index at 1 year = 0.59) and IMDC risk groups (c-index at 1 year = 0.63) (Table 3). The Kaplan–Meier estimates of OS for MSKCC and IMDC are reported in Fig. 4 (left). Similar results were observed in the validation dataset in terms of goodness-of-fit ($AIC_{GRCC} = 4742$ vs. $AIC_{MSKCC} = 4839$ and $AIC_{IMDC} = 4773$) and of explained variation ($R^2 = 0.2760$ vs. 0.1384 and 0.1853 for MSKCC and IMDC, respectively) which validate the performance of the new classification compared to the two existing ones. In terms of discrimination, visual inspection of Kaplan–Meier estimates of OS for GRCC classification (Fig. 3 right) indicates better separation than MSKCC and IMDC classifications (Fig. 4 right). However,

Table 3 R^2 and c-index of MSKCC, IMDC, and GRCC models and corresponding risk groups in the discovery and validation sets

Prognostic models[a]	Discovery set		Validation set	
	R^2	c-index [95% CI][b]	R^2	c-index[b]
MSKCC model (PI)	0.1562	1y: 0.64 [0.57–0.70]	0.0983	1y: 0.65 [0.62–0.67]
		2y: 0.64 [0.58–0.69]		2y: 0.63 [0.61–0.65]
		3y: 0.63 [0.59–0.68]		3y: 0.62 [0.60–0.64]
MSKCC model (risk groups)	0.1345	1y: 0.59 [0.54–0.64]	0.1384	1y: 0.63 [0.60–0.65]
		2y: 0.60 [0.56–0.64]		2y: 0.61 [0.59–0.63]
		3y: 0.60 [0.56–0.64]		3y: 0.60 [0.58–0.62]
IMDC model (PI)	0.2780	1y: 0.69 [0.63–0.75]	0.2150	1y: 0.69 [0.66–0.72]
		2y: 0.69 [0.64–0.74]		2y: 0.66 [0.64–0.68]
		3y: 0.69 [0.64–0.73]		3y: 0.65 [0.63–0.68]
IMDC model (risk groups)	0.2264	1y: 0.63 [0.59–0.68]	0.1853	1y: 0.65 [0.62–0.67]
		2y: 0.63 [0.59–0.67]		2y: 0.62 [0.60–0.64]
		3y: 0.63 [0.59–0.66]		3y: 0.61 [0.60–0.63]
GRCC model (PI)	0.3496	1y: 0.71 [0.66–0.77]	0.3049	1y: 0.72 [0.69–0.75]
		2y: 0.71 [0.66–0.75]		2y: 0.68 [0.66–0.71]
		3y: 0.71 [0.66–0.75]		3y: 0.68 [0.66–0.70]
GRCC model (risk groups)	0.2848	1y: 0.64 [0.59–0.69]	0.2760	1y: 0.65 [0.62–0.67]
		2y: 0.64 [0.60–0.68]		2y: 0.62 [0.60–0.64]
		3y: 0.64 [0.60–0.68]		3y: 0.62 [0.60–0.64]

MSKCC Memorial Sloan Kettering Cancer Center, *IMDC* International Metastatic Renal Cell Carcinoma Database Consortium, *GRCC* Gustave Roussy Cancer Campus

[a]The performance of two models are reported for GRCC, MSKCC, and IMDC: (i) the Cox model with its prognostic index (PI, weighted combination of risk factors comprising the model) and (ii) the classification scheme with corresponding risk groups derived in counting the number of risk factors included in the Cox model. R^2: Kent and O'Quigley estimator of explained variation. R^2 reported for the validation set is estimated from the regression coefficients estimated from the discovery set

[b]c-index at 1, 2, and 3 years according to Uno's method with confidence interval estimated by 1000 bootstrap. c-index reported for the validation set is estimated from the regression coefficients estimated from the discovery set

this does not translate into higher c-index (0.65 at 1 year for GRCC and IMDC classification schemes). Finally, another way to compare GRCC and IMDC models is to count the number of patients who change categories. The number of patients in the first two IMDC risk groups reclassified in the higher contiguous GRCC risk group is 61 (27.5%) and 310 (36.3%) in the discovery and validation datasets, respectively (Table S5).

Discussion

We have developed and externally validated a new prognostic model, called GRCC, to predict the risk of cancer specific mortality over 1 or 2 and 3 years for patients with mRCC treated with second line. The model was well calibrated and showed a better performance compared to MSKCC and IMDC models in both cohorts (discovery and validation). It was designed to provide better information on survival taking into account conventional factors and two new prognostic factors, such as time from first to second line and tumor burden. This height-factor model, based on information likely to be available before commencing a second line, have been designed to help improve the management decision by clinicians in secondary care. This is useful for identifying patients with high risk of death from mRCC due to their high volume tumor or the presence of multiple (even more than five) variables after their initial therapy. At that time, it is important to identify patients that deserve a more aggressive treatment from those in whom palliative care may be more appropriate. Before our data, the oncologic benefit of second-line treatment for these individuals was undefined. Indeed, in comparison with the literature our model takes into account more variables then previous ones. The MSKCC model includes KPS, hemoglobin, and calcium concentration [3]. It can be applied also to patients in second-line setting even if it was built for the primary care. The second model, the IMDC, includes the three factors used in the MSKCC model and time from diagnosis to treatment, neutrophil, and platelet count. It has been validated in second line, although calcium concentration was not confirmed

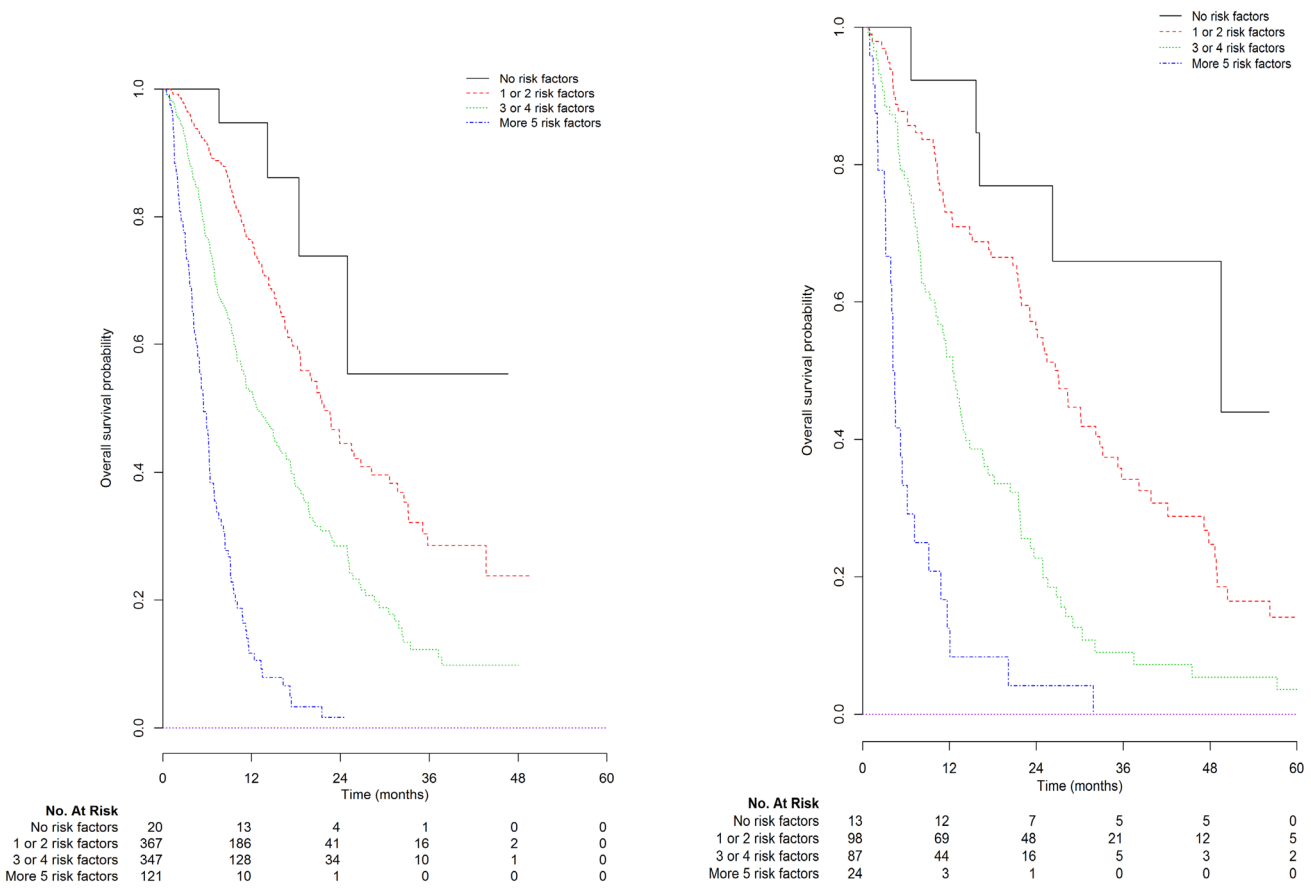


Fig. 3 Kaplan–Meier estimates of overall survival for the four risk groups of the GRCC prognostic classification in the discovery set (left, $n=222$, 176 deaths) and validation set (right, $n=855$, 455 deaths)

Table 4 Hazard ratios of GRCC risk groups for overall survival (OS) in the discovery and validation sets

	Discovery set ($n=222$, 176 deaths)				Validation set ($n=855$, 455 deaths)			
	# Patients	# Death	Median OS	HR (95% CI)	# Patients	# Death	Median OS	HR (95% CI)[a]
GRCC								
Low: 0 RF	13 (5.9)	5 (38.5)	49.5 [26.3–NE]	1.0	20 (2.3)	4 (20)	NE [24.9–NE]	1.0
Intermediate: 1–2 RF	98 (44.1)	69 (70.4)	26.6 [23.1–33.1]	2.33 [0.94–5.77]	367 (42.9)	135 (36.8)	21.8 [18.6–28.2]	2.25 [0.83–6.09]
Low poor: 3–4 RF	87 (39.2)	78 (89.7)	12.5 [10.0–16.6]	5.12 [2.06–12.73]	347 (40.6)	211 (60.8)	12.7 [11.0–15.8]	4.52 [1.68–12.18]
High poor: ≥5 RF	24 (10.8)	24 (100)	4.3 [3.9–7.2]	15.03 [5.63–40.12]	121 (14.2)	105 (86.8)	5.5 [4.7–6.4]	14.85 [5.46–40.67]
AIC				1552				4742

RF risk factor, *NE* not estimable, *HR* hazard ratio, *CI* confidence interval, *GRCC* Gustave Roussy Cancer Campus, *RF* risk factor, *AIC* Akaike criterion. No violation of the proportional hazards assumption is observed (global test: $p=0.722$ and 0.06 for the discovery and validation set)

[a]HRs were estimated from a stratified Cox model. The variable of stratification is the trial's name. Note that similar results are obtained without stratification

as independent predictors of overall survival [6]. The two new variables added in our model increased the predictive ability compared to those models reported before. These factors may also have clinical implications for the decision about duration of first-line treatment and optimal timing of switch from first- to second-line therapy.

Delays in switching may contribute to accumulated drug resistance, advanced immunosuppression, increased

Table 5 Calibration by risk groups comparing the observed (Kaplan–Meier) and model-based predicted overall survival probabilities

Risk group	Time (years)	Discovery set		Validation set	
		Observed Kaplan–Meier survival (SE)	Predicted survival (SE)	Observed Kaplan–Meier survival (SE)	Predicted survival (SE)
Good: 0 RF	1	92.31 (7.39)	87.68 (5.32)	94.74 (5.12)	87.47 (5.88)
	2	76.92 (11.69)	76.55 (9.32)	73.82 (13.97)	73.65 (11.3)
	3	65.93 (14.28)	63.13 (13.15)	55.37 (19.11)	60.74 (15.28)
Intermediate: 1–2 RF	1	73.08 (4.52)	73.57 (3.55)	76.41 (2.44)	73.91 (2.13)
	2	55.99 (5.20)	53.57 (4.64)	44.47 (4.06)	50.14 (3.34)
	3	34.20 (5.38)	34.16 (4.89)	28.56 (4.74)	32.46 (3.98)
Low poor: 3–4 RF	1	52.01 (5.41)	51.04 (4.66)	52.68 (2.91)	54.44 (2.66)
	2	22.73 (4.73)	25.48 (4.41)	28.48 (3.18)	24.94 (2.8)
	3	9.02 (3.51)	9.51 (3.7)	12.27 (3.00)	10.40 (2.3)
High poor: ≥ 5 RF	1	12.5 (6.75)	13.87 (5.86)	11.70 (3.28)	13.48 (2.81)
	2	4.17 (4.08)	1.8 (1.6)	1.65 (1.57)	1.03 (0.55)
	3	NE (NE)	0.1 (0.16)	NE (NE)	0.06 (0.06)

The model-based predictions were estimated from a Cox model with an indicator of the four GRCC risk groups

SE standard error, *NE* not evaluable, *RF* risk factor

Table 6 Multivariable Cox analyses for overall survival in the discovery set ($n = 222$, 176 deaths)

Risk factors	MSKCC model HR [95% CI]	IMDC model HR [95% CI]
Time from diagnosis to first line		
≥ 1 year		1
< 1 year		1.46 [1.04–2.07]
KPS		
≥ 80	1	1
< 80	2.04 [1.40–2.97]	2.12 [1.45–3.09]
Hemoglobin		
≥ LLN	1	1
< LLN	1.86 [1.35–2.56]	1.72 [1.22–2.42]
Calcium		
≤ ULN	1	1
> ULN	0.97 [0.39–2.44]	0.98 [0.38–2.50]
Neutrophil		
≤ ULN		1
> ULN		2.45 [1.61–3.74]
Platelet		
≤ ULN		1
> ULN		1.57 [1.04–2.38]

MSKCC Memorial Sloan Kettering Cancer Center, *IMDC* International Metastatic Renal Cell Carcinoma Database Consortium, *HR* hazard ratio, *CI* confidence interval, *KPS* Karnofsky performance status. No violation of the proportional hazards assumption is observed (global test: $p = 0.557$ and 0.1794 for MSKCC and IMDC models, respectively)

morbidity, and mortality. In our cohort, OS was shorter for patients with time from first to second line < 12 months compared to those with time interval ≥ 12 months (HR 1.68, 95% CI [1.23–2.31], $\chi^2 = 10.4$), as for patients with a TB ≥ 100 mm compared to those with TB < 100 mm (HR 1.43 [1.03–1.99], $\chi^2 = 4.6$). For example, in case of slow-growing disease and low-volume tumor, the clinical practice as well as a phase II trial suggested that a period of drug holiday after a favorable response to TKI can be proposed, until an increase in TB by ≥ 10% is reached [21, 22]. Contrariwise, the tumor growth rate after TT interruption, is negatively associated with prognosis in mRCC [23]. Finally, patients treated with a first-line TT "gained" over time in survival, even if they were before defined as poor-risk patients, for the concept of conditional survival [24]. This is due to the occurrence of good response to TT in poor-risk patients (maybe dependent on TB or systemic inflammatory disease reduction), with further outcome which can be similar to those patients with favorable and intermediate groups. Survival probability in metastatic RCC is dynamic and directly dependent on time elapsed from treatment initiation and duration of therapy, that is why time from first to second line is a new intriguing variable to consider.

Our model has been developed using a large, representative population of patients treated in first line with current standard of care and can be applied to all patients with a mRCC after receiving first-line strategy. However, our model could also be used to risk stratify patients recruited to clinical trials, as has been proposed by Li et al. for similar models [1].

The statistical methods we have used to derive and validate this model are similar to those for other risk prediction

Table 7 MSKCC, IMDC risk groups for overall survival in the discovery set and validation set

	Discovery set ($n=222$, 176 deaths)				Validation set ($n=855$, 455 deaths)			
	# Patients	# Death	Median	HR (95% CI)	# Patients	# Death	Median	HR (95% CI)[a]
MSKCC								
Low: 0 RF	86 (38.7)	60 (69.8)	26.6 [23.1–39.8]	1.0	373 (43.6)	139 (37.3)	22.7 [19.8–26.5]	1.0
Intermediate: 1–2 RF	135 (60.8)	115 (85.2)	12.1 [10.2–15.1]	2.20 [1.59–3.03]	476 (55.7)	310 (65.1)	10.6 [9.4–12.2]	2.33 [1.91–2.85]
Poor: ≥3 RF	1 (0.5)	1 (100)	6.3 [NE–NE]	6.63 [0.90–48.70]	6 (0.7)	6 (100)	4.8 [3.9–NE]	8.51 [3.68–19.53]
AIC				1588				4839
IMDC								
Low: 0 RF	20 (9.0)	9 (45.0)	49.5 [26.3–NE]	1.0	51 (6.0)	18 (35.3)	22.7 [18.4–NE]	1.0
Intermediate:1–2 RF	145 (65.3)	114 (78.6)	21.8 [17.3–25.5]	2.71 [1.37–5.38]	615 (71.9)	286 (46.5)	18.6 [17.2–20.8]	1.47 [0.91–2.37]
Poor: ≥3 RF	57 (25.7)	53 (93.0)	7.1 [5.0–11.7]	7.01 [3.40–14.43]	189 (22.1)	151 (79.9)	6.3 [5.6–7.3]	5.65 [3.45–9.26]
AIC				1570				4773

MSKCC Memorial Sloan Kettering Cancer Center, *IMDC* International Metastatic Renal Cell Carcinoma Database Consortium, *RF* risk factor, *NE* not estimable, *HR* hazard ratio, *CI* confidence interval, *RF* risk factor, *AIC* Akaike criterion. No violation of the proportional hazards assumption is observed (global test: $p=0.699$ and 0.279 for the MSKCC and IMDC for the discovery set and $p=0.337$ for the IMDC for the validation set) except for MSKCC for the validation set ($p=0.005$)

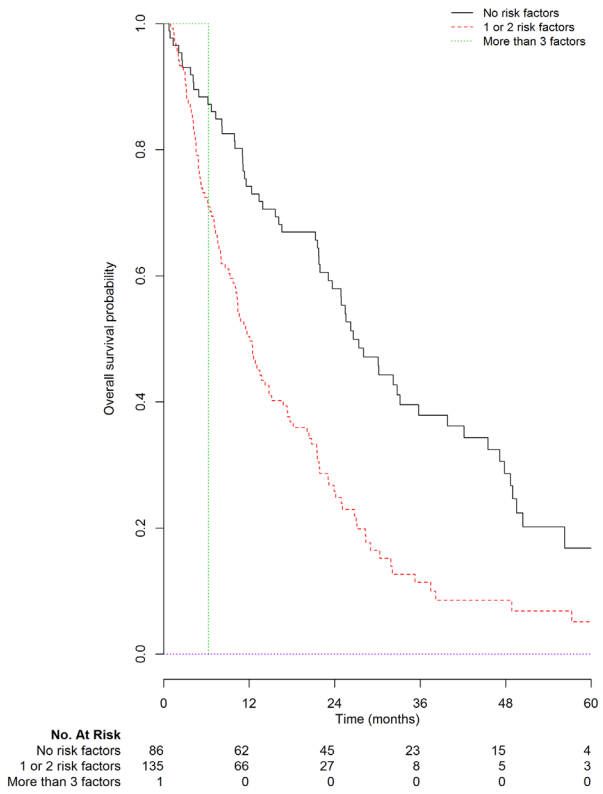
[a]HRs were estimated from a stratified Cox model. The variable of stratification is the trial's name. Note that similar results are obtained without stratification

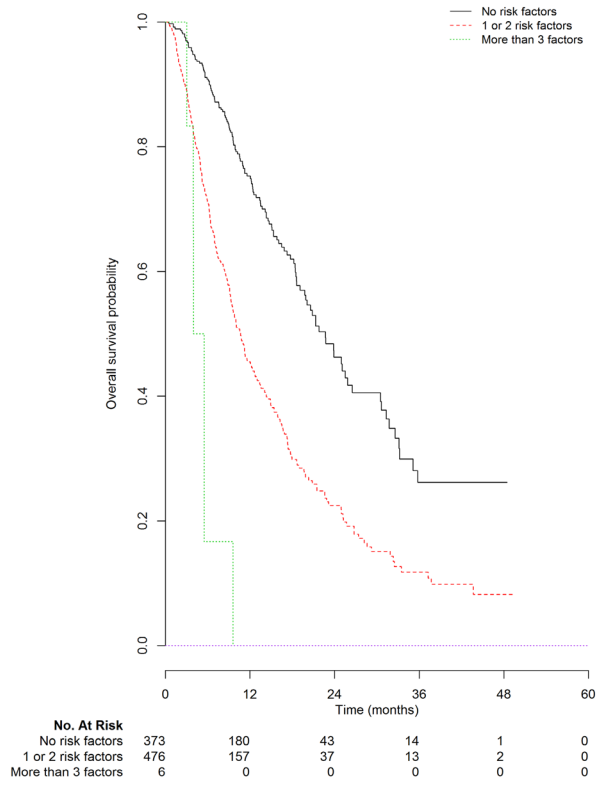
tools [6], although counting the number of risk factors frequently used to build classification scheme has some disadvantages. In summary, key strengths include size, follow-up, and representativeness for the discovery cohort even if the retrospective design could lead a potential bias. However, we have validated the model in a large prospective external multicentric dataset, though other limitation for our study could be the uncommon strategies out of clinical trials used as second-line therapy of some of these patients (such as temsirolimus, used only in clinical trials, and sorafenib, an old standard for second line).

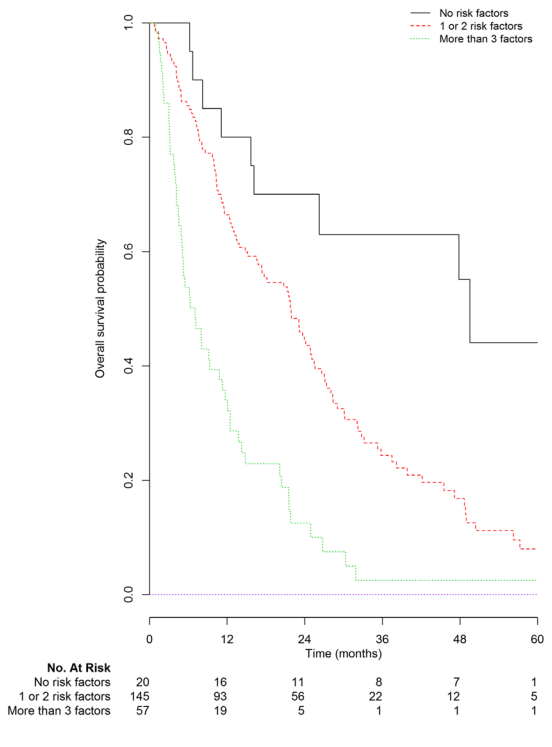
While we have validate the model in the PFIZERDB, it would be interesting to further explore the prognostic value of our model internationally by using additional validation studies in different settings and populations, such as patients receiving new standards cabozantinib including patients treated with nivolumab, in the second-line setting or beyond. This will allow confirmation of our findings and ensure good discrimination and calibration in the whole applicable populations.

Conclusions

The GRCC model is an externally validated prognostic model for second-line therapy in mRCC patients, taking account of variables that characterize post primary care features. It has been validated in a large prospective database and will be useful in clinical practice and for planning future trials with second-line treatment in mRCC.







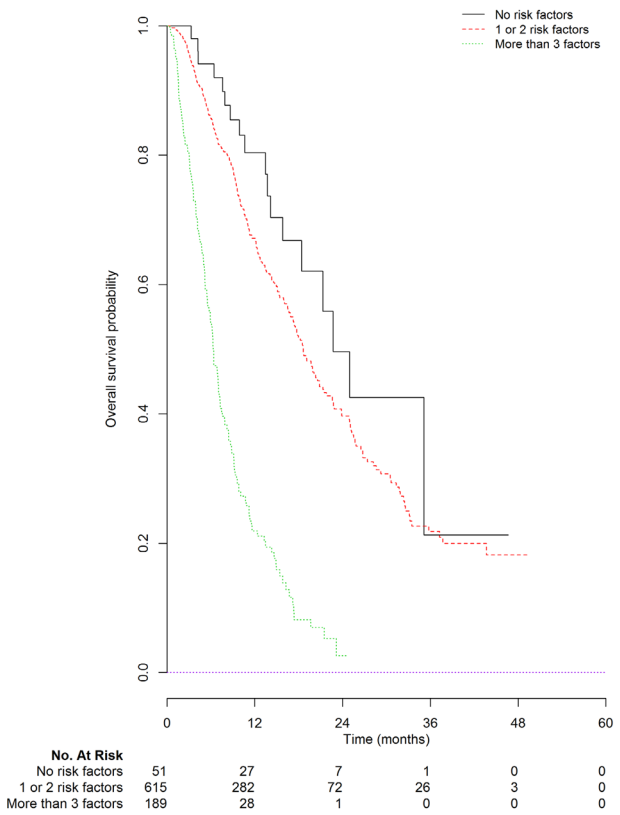


Fig. 4 Kaplan–Meier estimates of overall survival according to the MSKCC, IMDC risk groups for the discovery set ($n = 222$, 176 deaths) and validation set ($n = 855$, 455 deaths)

Acknowledgements LD was supported by Gustave Roussy Fondation Philanthropia, by Gustave Roussy DUERTECC program, and by ESMO translational research fellowship. We thank Pfizer laboratory for provided data used for the external validation.

Author contributions Conception and design: LD, MAB, GLT, BE. Financial support: None. Provision of study materials or patients: LD, BE, LA. Collection and assembly of data: LD. Data analysis and interpretation: MAB, GLT. Manuscript writing: LD, MAB, GLT, BE. Final approval of manuscript: All authors.

Compliance with ethical standards

Conflict of interest The author(s) indicated no potential conflicts of interest. This work won the 2018 ESMO Merit Award.

References

1. Li H, Samawi H, Heng DYC (2015) The use of prognostic factors in metastatic renal cell carcinoma. Urol Oncol 33(12):509–516
2. Shinohara N, Abe T (2015) Prognostic factors and risk classifications for patients with metastatic renal cell carcinoma. Int J Urol 22(10):888–897
3. Motzer RJ, Bacik J, Schwartz LH, Reuter V, Russo P, Marion S et al (2004) Prognostic factors for survival in previously treated patients with metastatic renal cell carcinoma. J Clin Oncol 22(3):454–463
4. Mekhail TM, Abou-Jawde RM, Boumerhi G, Malhi S, Wood L, Elson P et al (2005) Validation and extension of the Memorial Sloan-Kettering prognostic factors model for survival in patients with previously untreated metastatic renal cell carcinoma. J Clin Oncol 23(4):832–841
5. Heng DYC, Xie W, Regan MM, Warren MA, Golshayan AR, Sahi C et al (2009) Prognostic factors for overall survival in patients with metastatic renal cell carcinoma treated with vascular endothelial growth factor-targeted agents: results from a large, multicenter study. J Clin Oncol 27(34):5794–5799
6. Ko JJ, Xie W, Kroeger N, Lee J-L, Rini BI, Knox JJ et al (2015) The International Metastatic Renal Cell Carcinoma Database Consortium model as a prognostic tool in patients with metastatic renal cell carcinoma previously treated with first-line targeted therapy: a population-based study. Lancet Oncol 16(3):293–300
7. Iacovelli R, Lanoy E, Albiges L, Escudier B (2012) Tumour burden is an independent prognostic factor in metastatic renal cell carcinoma. BJU Int 110(11):1747–1753
8. Al-Marrawi MY, Rini BI, Harshman LC, Bjarnason G, Wood L, Vaishampayan U et al (2013) The association of clinical outcome to first-line VEGF-targeted therapy with clinical outcome to second-line VEGF-targeted therapy in metastatic renal cell carcinoma patients. Target Oncol 8(3):203–209
9. Rini BI, Escudier B, Tomczak P, Kaprin A, Szczylik C, Hutson TE et al (2011) Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. Lancet (London) 378(9807):1931–1939
10. Hutson TE, Escudier B, Esteban E, Bjarnason GA, Lim HY, Pittman KB et al (2014) Randomized phase III trial of temsirolimus versus sorafenib as second-line therapy after sunitinib in patients with metastatic renal cell carcinoma. J Clin Oncol 32(8):760–767
11. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer (Oxf England 1990) 45(2):228–247
12. Basappa NS, Elson P, Golshayan A-R, Wood L, Garcia JA, Dreicer R et al (2011) The impact of tumor burden characteristics in patients with metastatic renal cell carcinoma treated with sunitinib. Cancer 117(6):1183–1189
13. Stein WD, Huang H, Menefee M, Edgerly M, Kotz H, Dwyer A et al (2009) Other paradigms: growth rate constants and tumor burden determined using computed tomography data correlate strongly with the overall survival of patients with renal cell carcinoma. Cancer J (Sudbury) 15(5):441–447
14. Miyake H, Miyazaki A, Imai S, Harada K-I, Fujisawa M (2016) Early tumor shrinkage under treatment with first-line tyrosine kinase inhibitors as a predictor of overall survival in patients with metastatic renal cell carcinoma: a retrospective multi-institutional study in Japan. Target Oncol 11(2):175–182
15. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53(282):457–481
16. Schemper M, Smith TL (1996) A note on quantifying follow-up in studies of failure time. Control Clin Trials 17(4):343–346
17. Kent JT, O'quigley J (1988) Measures of dependence for censored survival data. Biometrika. 75(3):525–534
18. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ (2011) On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med 30(10):1105–1117
19. Boisbunon A, Canu S, Fourdrinier D, Strawderman W, Wells MT (2014) Akaike's information criterion, Cp and estimators of loss for elliptically symmetric DISTRIBUTIONS. Int Stat Rev 82(3):422–439
20. Heinzl H (2000) Using SAS to calculate the Kent and O'Quigley measure of dependence for Cox proportional hazards regression model. Comput Methods Programs Biomed 63(1):71–76
21. Mittal K, Derosa L, Albiges L, Wood L, Elson P, Gilligan T et al (2018) Drug holiday in metastatic renal cell carcinoma patients treated with vascular endothelial growth factor receptor inhibitors. Clin Genitourin Cancer. https://doi.org/10.1016/j.clgc.2017.12.014
22. Ornstein MC, Wood LS, Elson P, Allman KD, Beach J, Martin A et al (2017) A phase II study of intermittent sunitinib in previously untreated patients with metastatic renal cell carcinoma. J Clin Oncol 35(16):1764–1769
23. Ferté C, Koscielny S, Albiges L, Rocher L, Soria J-C, Iacovelli R et al (2014) Tumor growth rate provides useful information to evaluate sorafenib and everolimus treatment in metastatic renal cell carcinoma patients: an integrated analysis of the TARGET and RECORD phase 3 trial data. Eur Urol 65(4):713–720
24. Harshman LC, Xie W, Bjarnason GA, Knox JJ, MacKenzie M, Wood L et al (2012) Conditional survival of patients with metastatic renal-cell carcinoma treated with VEGF-targeted therapy: a population-based study. Lancet Oncol. https://doi.org/10.1016/S1470-2045(12)70285-1

Affiliations

Lisa Derosa[1,2,3,4] · Mohamed Amine Bayar[5,6] · Laurence Albiges[1] · Gwénaël Le Teuff[5,6] · Bernard Escudier[1]

1 Departments of Medical Oncology, Gustave Roussy, 114 Rue Edward Vaillant, 94800 Villejuif, France

2 Université Paris-Sud, Université Paris-Saclay, 94085 Villejuif, France

3 Institut National de la Santé Et de la Recherche Médicale (INSERM) U1015, Villejuif, France

4 Equipe Labellisée-Ligue Nationale contre le Cancer, Villejuif, France

5 Department of Biostatistics and Epidemiology and Ligue National Contre le Cancer meta-analysis platform, Gustave Roussy, 94805 Villejuif, France

6 U1018 INSERM, CESP, Université Paris-Sud, Université Paris-Saclay, 94085 Villejuif, France