

Keywords: renal cancer; prognostic model; targeted therapy; sunitinib

Development and validation of a prognostic model in patients with metastatic renal cell carcinoma treated with sunitinib: a European collaboration

A Bamias^{*,1}, K Tzannis¹, B Beuselinck^{2,3}, S Oudard⁴, B Escudier⁵, D Diosynopoulos⁶, K Papazisis⁷, H Lang⁸, P Wolter², E de Guillebon⁴, K Stravodimos⁹, M Chrisofos¹⁰, G Fountzilias⁶, R-T Elaidi⁴, M A Dimopoulos¹ and C Bamia¹¹

¹Department of Clinical Therapeutics, University of Athens, Athens, Greece; ²Department of General Medical Oncology and Laboratory for Experimental Oncology, University Hospitals Leuven, Leuven Cancer Institute, Leuven, Belgium; ³Inserm U674 Génomique fonctionnelle des tumeurs solides, Université Paris-5 René Descartes, Paris, France; ⁴Hopital Europeen Georges Pompidou, Paris, France; ⁵Institute Gustav Roussy, Villejuif, France; ⁶Department of Medical Oncology, Papageorgiou Hospital, Aristotle University of Thessaloniki School of Medicine, Thessaloniki, Greece; ⁷Theagenion Hospital, Thessaloniki, Greece; ⁸Centre Hospitalier Régional Universitaire, Strasbourg, France; ⁹First Department of Urology, University of Athens, Athens, Greece; ¹⁰Second Department of Urology, University of Athens, Athens, Greece and ¹¹Department of Hygiene, Epidemiology and Medical Statistics, University of Athens, Athens, Greece

Background: Accurate prediction of outcome for metastatic renal cell carcinoma (mRCC) patients receiving targeted therapy is essential. Most of the available models have been developed in patients treated with cytokines, while most of them are fairly complex, including at least five factors. We developed and externally validated a simple model for overall survival (OS) in mRCC. We also studied the recently validated International Database Consortium (IDC) model in our data sets.

Methods: The development cohort included 170 mRCC patients treated with sunitinib. The final prognostic model was selected by uni- and multivariate Cox regression analyses. Risk groups were defined by the number of risk factors and by the 25th and 75th percentiles of the model's prognostic index distribution. The model was validated using an independent data set of 266 mRCC patients (validation cohort) treated with the same agent.

Results: Eastern Co-operative Oncology Group (ECOG) performance status (PS), time from diagnosis of RCC and number of metastatic sites were included in the final model. Median OS of patients with 1, 2 and 3 risk factors were: 24.7, 12.8 and 5.9 months, respectively, whereas median OS was not reached for patients with 0 risk factors. Concordance (C) index for internal validation was 0.712, whereas C-index for external validation was 0.634, due to differences in survival especially in poor-risk populations between the two cohorts. Predictive performance of the model was improved after recalibration. Application of the mRCC International Database Consortium (IDC) model resulted in a C-index of 0.574 in the development and 0.576 in the validation cohorts (lower than those recently reported for this model). Predictive ability was also improved after recalibration in this analysis. Risk stratification according to IDC model showed more similar outcomes across the development and validation cohorts compared with our model.

*Correspondence: Dr A Bamias; E-mail: abamias@med.uoa.gr

Received 28 February 2013; revised 7 June 2013; accepted 9 June 2013; published online 27 June 2013

© 2013 Cancer Research UK. All rights reserved 0007–0920/13

Conclusion: Our model provides a simple prognostic tool in mRCC patients treated with a targeted agent. It had similar performance with the IDC model, which, however, produced more consistent survival results across the development and validation cohorts. The predictive ability of both models was lower than that suggested by internal validation (our model) or recent published data (IDC model), due to differences between observed and predicted survival among intermediate and poor-risk patients. Our results highlight the importance of external validation and the need for further refinement of existing prognostic models.

Renal cancer is the third most frequent malignancy of the urinary tract and accounts for 3% of all adult malignancies (Cohen and McGovern, 2005). Localised disease can be cured with surgery in most cases. Nevertheless, ~50% of patients with renal cell carcinoma will present with or develop metastatic disease (Flanigan *et al*, 2004; Cohen and McGovern, 2005). In this case, prognosis remains poor and 5-year life expectancy is <20% (Kavolius *et al*, 1998; Flanigan *et al*, 2004).

Recent advances in our understanding of the biology of RCC and especially the role of angiogenesis in the development and expansion of this tumour led to the development of novel vascular endothelial growth factor (VEGF)-targeting therapies (Motzer *et al*, 2007; Escudier *et al*, 2007a, b), which proved to be superior to the previous standard, interferon (IFN). Sunitinib is an inhibitor of the split-kinase-domain family of receptor tyrosine kinases (including -VEGF) (Chow and Eckhardt, 2007). It has been established as first-line treatment for advanced RCC, following the results of a randomised phase III trial, which showed a significant advantage IFNa in progression-free survival (PFS) (Motzer *et al*, 2007). In spite of this undisputed benefit, the prognosis of advanced RCC remains poor, while the toxicity of sunitinib (as well as that of other novel agents) is considerable (Bhojani *et al*, 2008). There is, therefore, a need to select patients likely to benefit from these therapies.

In the era of targeted therapies, specific prognostic algorithms are necessary for clinical trial design, patients counselling and treatment decisions. Until recently, the most widely used prognostic model is that of the MSKCC, which uses five factors: LDH, Karnofsky performance status (KPS), time from nephrectomy, calcium levels and haemoglobin levels, which have all been associated with independent prognostic significance (Motzer *et al*, 1999, 2001; Negrier *et al*, 2005; Escudier *et al*, 2007c). The combination of these factors led to the development of a prognostic model including three patient groups with statistically significant and clinically relevant differences in survival (Motzer *et al*, 2001). The MSKCC model has been used for the design of all phase III trials using modern therapies. Nevertheless, there may be limitations associated with its use in this context. It was developed with patients undergoing treatment with cytokines, while the prognosis of patients with metastatic renal cell carcinoma (mRCC) with targeted therapies has been considerably improved. All randomized studies mainly included patients of low or intermediate risk, that is, populations with different composition than that of those used to develop the MSKCC model. Analyses in contemporary series have not confirmed all factors in the model as significant, while other factors with independent prognostic significance have been suggested (Motzer *et al*, 2009; Beuselinck *et al*, 2011; Karakiewicz *et al*, 2011). Furthermore, certain studies suggested that novel models may perform better than the MSKCC in the targeted therapy era (Karakiewicz *et al*, 2011). Finally, this model is fairly complex, requiring two clinical and three biochemical factors making its application for retrospective analyses somewhat problematic. Information regarding prognostic factors in the targeted therapy era is limited and heterogenous (Choueiri *et al*, 2007; Motzer *et al*, 2008; Karakiewicz *et al*, 2011). One (Motzer *et al*, 2008; Karakiewicz *et al*, 2011) or more (Choueiri *et al*, 2007) agents were used, while PFS (and not overall

survival (OS)) was the endpoint (Choueiri *et al*, 2007; Motzer *et al*, 2008; Karakiewicz *et al*, 2011). Finally, in two of these studies (Motzer *et al*, 2008; Karakiewicz *et al*, 2011) only patients participating in a clinical study were included, thus not being representative of the population treated in everyday practice. None of these models has been externally validated. For these reasons, their accuracy, performance characteristics and impact on clinical decisions remain unknown.

The most accepted model developed in the targeted therapy era is the mRCC IDC model (Heng *et al*, 2009). The data have been derived by a large (645 patients) multinational database including patients treated with first-line anti-VEGF therapy. Six factors (KPS, time from nephrectomy, calcium, haemoglobin, neutrophils and platelets) were used to identify three prognostic groups. This model has been recently externally validated and compared favourably with four other models, which, however, had all been developed in the cytokine era (Heng *et al*, 2013). Nevertheless, the median follow-up was relatively short (16 months) and the model is more complex than that of MSKCC. In addition, treatment was heterogenous with three agents used. Importantly, one of them, sorafenib, can be considered as suboptimal first-line therapy, as, unlikely the other two agents, sunitinib and bevacizumab, it has not shown superiority to IFN in this setting (Escudier *et al*, 2009).

We have recently used the advanced RCC database of the Hellenic Co-operative Oncology Group (HECOG) to study prognostic clinicopathological factors in patients treated with sunitinib. In an initial analysis of 109 patients, we identified PS, time from diagnosis and number of metastatic sites as independent predictors of survival (Bamias *et al*, 2010). The combination of these factors led to the development of a prognostic model with similar performance with that of the more complex MSKCC model. We are now reporting an updated analysis of 170 patients and external validation of this model. We also studied the performance of the IDC model in our development and validation cohorts.

MATERIALS AND METHODS

Patient population. The development cohort included 170 consecutive patients with mRCC from nine Greek centres treated between October 2005 and December 2010. The validation cohort included 266 consecutive patients treated at three French and one Belgian centre between November 2005 and January 2012 (Table 1). The larger part of this database has been used in a previous analysis of prognostic factors for PFS and OS (Beuselinck *et al*, 2011). The analysis was approved by the Institutional Review Board of the participating institutions and informed consent for the use of medical data for research purposes was obtained.

Criteria for inclusion in this analysis included diagnosis of mRCC and treatment with sunitinib. Previous IFNa but not anti-VEGF therapy was allowed. Baseline demographic, clinical and laboratory data with prognostic significance according to published reports and the authors' experience were retrospectively collected from medical charts using uniform database templates to ensure consistent data collection. Overall survival data was available for all

Table 1. Participating centres (number of patients, %) in the development and validation cohorts

Centre	N (%)
Development cohort	
Alexandra	66 (39)
Papageorgiou	40 (24)
Other ^a	64 (37)
Total	170 (100)
Validation cohort	
HEGP	90 (34)
KUL	96 (36)
IGR	63 (24)
Strasbourg	17 (6)
Total	266 (100)

^aSeven centres with fewer than 15 patients each.

patients. The databases were updated in January 2012 before the final analysis.

Statistical analysis. All analyses were carried out in STATA/SE 11.2.

Description of data and model construction. Patients' characteristics were presented through means, medians and proportions. Overall survival was the primary variable and was defined as the time interval between the date of first cycle of sunitinib and the date of death from any cause; patients not dead were censored at the date of last contact. Survival curves were estimated using the Kaplan–Meier method. Factors that were considered for their prognostic ability included: age (≤ 60 vs > 60 years), sex (female vs male), Eastern Co-operative Oncology Group (ECOG) PS (≥ 1 vs 0), time from diagnosis to treatment with sunitinib in months (≤ 12 vs > 12), number of metastatic sites (> 2 vs 0–2), tumour grade (III + IV vs I + II), nephrectomy (no vs yes), previous IFNa (no vs yes), histology (clear cell vs other), alkaline phosphatase (abnormal vs normal), LDH (abnormal vs normal), calcium (> 10 mg dl⁻¹ vs ≤ 10 mg dl⁻¹), platelets ($> 400 \times 10^3$ per mm³ vs $\leq 400 \times 10^3$ per mm³), haemoglobin (≤ 13 g dl⁻¹ for males or ≤ 11.5 g dl⁻¹ for females vs > 13 g dl⁻¹ for males or > 11.5 g dl⁻¹ for females), neutrophils (> 5000 per mm³ vs ≤ 5000 per mm³), WBC ($> 10\,000$ per mm³ vs $\leq 10\,000$ per mm³), liver metastases (yes vs no), brain metastases (yes vs no), bone metastases (yes vs no) and lung metastases (yes vs no).

For some patients in the development cohort, laboratory data (platelets, neutrophils, WBC, haemoglobin, calcium, LDH, ALP) and data on other variables (tumour grade, age, histology, liver metastases, lung metastases, bone metastases and brain metastases) were missing. To account for the missing values, we employed multiple imputations using the Markov Chain Monte Carlo method for arbitrary missing data. The variables used to generate imputed data were number of metastatic sites, PS, time from tumour diagnosis, sex, previous IFNa, previous nephrectomy and survival status.

The associations of each of the above indicated factors (after multiple imputations) with OS were assessed through hazard ratios estimated from univariate Cox proportional hazards models. Factors for which the hazard ratios were statistically significant at the level of significance 0.2 after multiple imputations were then included in a multivariate Cox proportional hazards model. The final predictive model included only those variables for which the corresponding estimated hazard ratios were statistically significant at the level of 5% ($P < 0.05$).

After the final model was defined, patients were classified into risk groups in two ways: four groups on the basis of the actual number of prognostic factors that remained in the final model; and three groups defined as good, intermediate and poor risk on the basis of the 25th and 75th percentiles of the model's prognostic index risk score distribution. The former classification is familiar in the clinical setting, whereas the latter methodology has been suggested in recent studies (Royston *et al*, 2010).

Internal and external validation and calibration of the model. ROC curves and bootstrap-corrected Harrell's C-index were used to assess the model's discriminatory ability (Pencina and D'Agostino, 2004) in the development cohort (internal validation). The C-index was estimated by bootstrapping with 200 resamples to estimate an unbiased measure of the ability of our predictive model to discriminate among patients in the development cohort with respect to their death/survival.

External validation was performed by calculating a risk score for each patient in the validation cohort using the prognostic factors and the respective Cox regression coefficients of the model as estimated in the development data set. Patients were stratified according to their risk of death in the same way as in the development cohort, but using the distribution of risk scores in the validation data set. The model's discriminatory ability in the validation cohort was checked with the C-index.

The model was recalibrated using the method described by Miller and Hui (1991), as this applies in the Cox PH model. According to this method, the need to include in the model a slope for the prognostic index is checked, and if so, the recalibrated model is used to estimate survival probabilities for subjects in the validation cohort.

The predictive ability of the model (with or without calibration) was checked by plotting the observed and the predicted survival curves for the indicated risk groups in the development and in the validation data sets.

RESULTS

Baseline characteristics of the development cohort. The baseline characteristics of the patients in the development and the validation cohorts who were included in the analyses are detailed in Table 2. Median follow-up for the development cohort was 35.51 months and for the validation cohort 37.55 months. During follow-up, 103 patients of the development cohort (61%) and 151 (57%) of the validation cohort died. The median OS in the development and validation data sets was 19.4 months (95% confidence interval (CI) 15.1–24.7) and 26.1 months (95% CI 20.4–31.9), respectively. Significantly more patients in the validation cohort had undergone nephrectomy, had previously received IFNa, had clear cell histology, were more than 12 months from diagnosis of RCC, had normal LDH and had more than two metastatic sites. In addition, more patients of the validation cohort were categorised in the favourable and intermediate risk groups according to the MSKCC and IDC models.

Construction of the predictive model. Model selection on the basis of the development cohort is shown in Table 3. Male sex, PS 0, > 12 months from diagnosis to sunitinib initiation, 0–2 metastatic sites, previous nephrectomy, normal LDH, alkaline phosphatase and calcium and platelet count $\leq 400,000$, lack of anaemia and absence of bone, lung or brain metastases were associated with improved survival in univariate analysis. Our final model (shown in the last two columns of Table 3) included the three factors, which were found significant in multivariate analysis (in order of significance): number of metastatic sites, ECOG PS

Table 2. Baseline characteristics of patients in the development and validation cohorts

	Development data set (n = 170)			Validation data set (n = 266)			P-value
	Median OS	n	(%)	Median OS	n	(%)	
Sex							0.891
Female	11.6	45	(27)	19.1	72	(27)	
Male	22.4	125	(73)	29.3	194	(73)	
Nephrectomy							<0.001
No	7.9	37	(22)	10.6	5	(2)	
Yes	22.3	133	(78)	26.5	261	(98)	
Previous IFNa							<0.001
No	19.4	154	(91)	31.7	148	(56)	
Yes	22.3	16	(9)	22.2	118	(44)	
Time from initial diagnosis to sunitinib therapy							<0.001
≤ 12 months	14.2	90	(53)	21	90	(34)	
> 12 months	33.3	80	(47)	28.8	176	(66)	
Histology							<0.001
Clear Cell	19.4	149	(87)	26.9	256	(96)	
Other	19.8	20	(12)	7.9	10	(4)	
Missing	-	1	(1)		0	(0)	
Tumour grade							0.199
I	NR	3	(2)	29.3	2	(1)	
II	28.8	44	(26)	40.2	53	(20)	
III	16.4	62	(36)	29	113	(43)	
IV	17.2	27	(16)	17.3	55	(21)	
Missing		34	(20)		43	(16)	
Neutrophils status							0.615
≤ 5000	24.7	66	(39)	34.4	130	(49)	
> 5000	15.1	64	(38)	18.8	113	(43)	
Missing		40	(23)		23	(9)	
Platelets status							0.595
≤ 400	22.3	118	(69)	29	217	(82)	
> 400	11.2	27	(16)	13.7	43	(16)	
Missing		25	(15)		6	(2)	
Karnofsky performance status							0.086
0	36.7	96	(57)	37.2	160	(60)	
1	16.2	53	(31)	18.6	91	(34)	
2	6.4	16	(9)	6	10	(4)	
3	3.4	5	(3)	13.6	5	(2)	
Total number of metastatic sites							<0.001
0-2	29	123	(72)	33	131	(49)	
> 2	7.9	47	(28)	18.8	135	(51)	
LDH							<0.001
Normal	29.2	74	(44)	26.1	214	(81)	
Abnormal	13.9	50	(29)	23.6	35	(13)	
Missing		46	(27)		17	(6)	
Calcium							0.537
≤ 10	20.8	102	(60)	24.9	214	(81)	
> 10	9.8	18	(10)	26.9	31	(12)	
Missing		50	(30)		21	(8)	
Haemoglobin							0.309
≤ 13 for males, ≤ 11.5 for females	12	69	(41)	17.9	108	(41)	
> 13 for males, > 11.5 for females	29	77	(45)	30	149	(56)	
Missing		24	(14)		9	(3)	
Heng's risk classification							<0.001
Favourable	37.4	16	(9)	40.2	53	(20)	
Intermediate	28.8	57	(34)	21	126	(47)	
Poor	11.2	32	(19)	13.6	48	(18)	
Missing		65	(38)		39	(15)	
Motzer's risk classification							<0.001
Favourable	NR	16	(9)	38.1	59	(22)	
Intermediate	29.2	61	(36)	21	134	(50)	
Poor	11.6	37	(22)	13.6	44	(17)	
Missing		56	(33)		29	(11)	

Abbreviations: IFN = interferon; LDH = lactate dehydrogenase; NR = not reached; OS = overall survival.

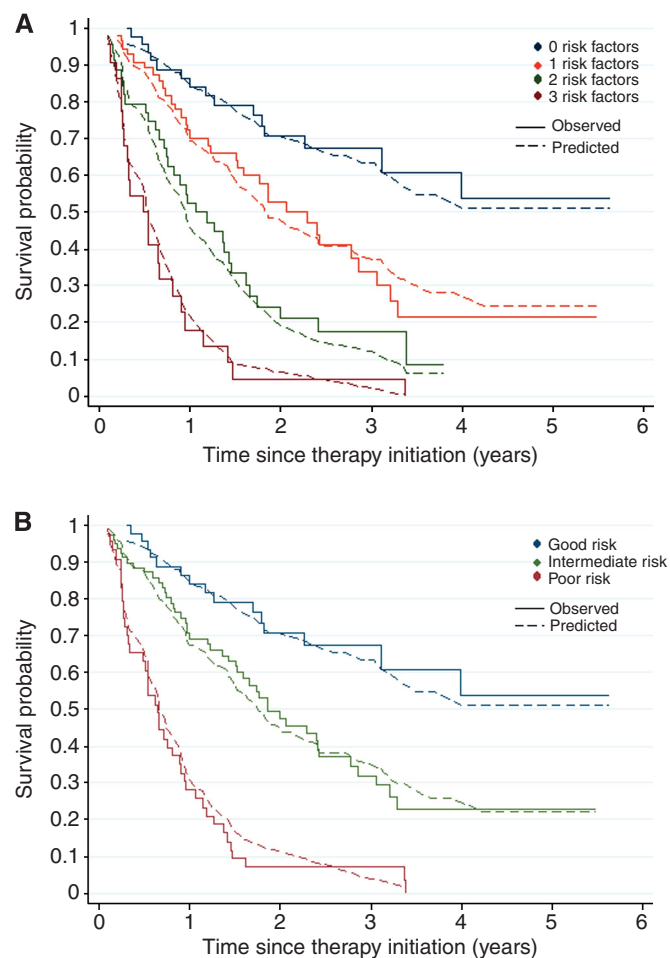


Figure 1. Observed (solid lines) and predicted (dashed lines) overall survival for the development data set by risk classification according to: (A) number of risk factors and (B) according to percentiles of the prognostic index.

and time from diagnosis to sunitinib. Check of the proportionality assumption revealed no violation (P -value = 0.392).

Risk stratification in the development cohort. The prognostic index from the model was estimated for each patient as the sum of the variables included in the final model multiplied by the log of the respective HRs (Table 3). Low values of the index indicate lower probability of death. Patients were classified according to their risk of death in four groups identified by the number of risk factors of the final predictive model that were present in a patient: no factors, any one factor, any two factors and all three factors; and in three groups (good, intermediate and poor risk) by splitting the index values at 0 (25th percentile) and 1.544 (75th percentile). Figure 1A and B (solid lines) show the observed survival curves according to either classification scheme.

Internal validation. Internal validation of the model regarding its discriminatory ability resulted in a C -index of 0.709. After bootstrapping with 200 resamples, the corrected C -index was 0.712 indicating good discriminatory performance of the model, in that subjects with longer predicted survival times also had longer actual survival. The discriminatory ability of the model is also demonstrated in Figure 1A and B (dashed lines): predicted survival curves are very close to observed survival curves.

External validation. External validation was accomplished by applying the log of the HRs shown in Table 3 (7th column) to each

patient in the validation data set to calculate the prognostic index. Risk groups were formed as in the development data set but the cutoffs for the 25th (0) and 75th (1.543) percentile were derived from the distribution of the prognostic index in the validation data set. The C -index from this model in the validation data set was 0.634, indicating that the model did not have as high discriminatory ability in the validation as in the development cohort.

In Figure 2A and B, the observed and predicted survival curves according to the two classification schemes are shown for the validation data set. Predicted survival was similar to the observed in the best prognosis groups of both classifications but deviated in the other risk groups with longer observed survival being longer than the predicted. This is also evident in Figure 3A and B where the observed survival is plotted for the development (solid line) and validation data sets (dashes lines).

When the model was recalibrated, the inclusion of a slope for the prognostic index was deemed statistically significant ($P < 0.001$) and the magnitude of the slope was 0.553 (s.e. = 0.117). Therefore, the calibrated prognostic index was further used to predict survival in the validation data set. The improvement in the prediction of the model is depicted in Figure 2C and D, where the deviation between observed and predicted survival in the validation data set has been decreased compared with the prior-to-calibration analysis. Table 4 provides a summary of survival according to risk classification in the development and validation cohorts (after recalibration).

As our model showed lower discriminatory ability in the validation compared with the development cohort, we also evaluated the performance of the model proposed by IDC, in the development and validation cohorts. In this way, the IDC model was indirectly compared with our predictive model. That is, we estimated a prognostic index on the basis of parameters and respective estimates given by Heng *et al* (2009), and we estimated predicted survival on the basis of this prognostic index in both data sets. Eastern Co-operative Oncology Group (ECOG) PS was converted to KPS by considering KPS of 100 equal to ECOG PS of 0, KPS of 80–90 equal to ECOG PS of 1, and KPS ≤ 70 equal to ECOG PS ≥ 2 . Risk groups were formed on the basis of the following six factors: PS, time from nephrectomy, calcium level, haemoglobin level, neutrophil count and platelet count (categorised as shown in Table 3) according to the published model (Heng *et al*, 2009): favourable risk 0 factors, intermediate risk 1–2, poor risk 3–6. C -index was 0.574 in the development and 0.576 in the validation data sets for the IDC model. This modest discrimination ability was attributed to the worse predicted compared with observed survival in both the development (Figure 4A) and validation (Figure 4B) data sets. When IDC model was recalibrated, a slope of 0.555 (P -value < 0.001) for the development and a slope of 0.580 (P -value < 0.001) were estimated for the validation cohorts. Predicted survival was much closer to observed after recalibration of the IDC model especially in the validation cohort (Figure 4C and D). An overall evaluation of IDC model with respect to survival in the development and validation cohorts is shown in Table 4 after calibration. It should be noted that, in contrast, to our model, survival of each risk group according to IDC classification was quite similar between the development and the validation cohorts. We also compared the predictive performance, in the development and the validation data sets, of our model with that proposed by IDC (Heng *et al*, 2009), as well as, with that proposed by the MSKCC (Motzer *et al*, 2001), using ROC curves (data not shown). No statistically significant differences were seen in the validation data set, whereas our model performed better in the development data set—this was somehow expected, as our model was derived on the basis of the highest predictive ability with respect to survival of subjects in the development data set.

Table 3. Model selection through univariate and multivariate Cox models in the development cohort

Variables	N	Univariate ^a		Multivariate ^a		Multivariate ^a	
		HR (95% CI)	P-value	HR (95% CI)	P-value	HR (95% CI)	P-value
Sex	170						
Male		1	0.013	1	0.490		
Female		1.70 (1.12-2.58)		1.20 (0.71-2.03)			
Performance status	170						
0		1	<0.001	1	0.016	1	<0.001
≥1		2.55 (1.73-3.77)		1.83 (1.12-3.01)		2.06 (1.38-3.08)	
Time diagnosis to sunitinib therapy	170						
> 12 months		1	<0.001	1	0.066	1	0.013
≤12 months		2.19 (1.45-3.29)		1.56 (0.97-2.51)		1.71 (1.12-2.59)	
No. of metastatic sites	170						
0-2		1	<0.001	1	0.011	1	<0.001
>2		3.46 (2.32-5.16)		2.53 (1.24-5.17)		2.75 (1.82-4.15)	
Tumour grade	136						
I+II		1	0.171	1	0.219		
III+IV		1.38 (0.87-2.18)		1.44 (0.80-2.57)			
Nephrectomy	170						
Yes		1	<0.001	1	0.960		
No		2.19 (1.42-3.37)		1.02 (0.56-1.86)			
LDH	124						
Normal		1	0.036	1	0.963		
Abnormal		1.58 (1.03-2.43)		1.01 (0.57-1.80)			
ALP	125						
Normal		1	0.001	1	0.594		
Abnormal		2.12 (1.35-3.33)		1.18 (0.64-2.20)			
Ca	120						
Normal		1	0.137	1	0.815		
Abnormal		1.60 (0.86-2.99)		0.91 (0.42-1.99)			
Neutrophils	130						
≤5000		1	0.381				
>5000		1.22 (0.79-1.88)					
WBC	146						
≤10 000		1	0.487				
>10 000		1.22 (0.70-2.11)					
Platelets	145						
≤400		1	0.015	1	0.851		
>400		1.81 (1.12-2.92)		1.06 (0.59-1.88)			
Histology	169						
Other		1	0.772				
Clear cell		1.10 (0.59-2.05)					
Age	147						
>60		1	0.386				
≤60		1.21 (0.79-1.84)					
Previous IFNa	170						
Yes		1	0.765				
No		1.10 (0.60-2.01)					
Hb	146						
>13 for males		1	0.002	1	0.134		
>11.5 for females		1.92 (1.27-2.91)		1.53 (0.88-2.68)			
≤13 for males							
≤11.5 for females							
Brain metastasis	167						
No		1	0.003	1	0.057		
Yes		2.78 (1.41-5.50)		2.16 (0.98-4.76)			
Liver metastasis	168						
No		1	0.111	1	0.580		
Yes		1.59 (0.90-2.80)		1.20 (0.63-2.31)			
Bone metastasis	169						
No		1	0.009	1	0.666		
Yes		1.73 (1.15-2.61)		0.88 (0.50-1.57)			
Lung	169						
No		1	0.019	1	0.831		
Yes		1.67 (1.09-2.55)		1.06 (0.63-1.78)			

Abbreviations: ALP = alkaline phosphatase; CI = confidence interval; HR = hazard ratio; IFN = interferon; LDH = lactate dehydrogenase; NR = not reached; OS = overall survival; WBC = white blood cells.

^aMultiple imputation was applied for missing values in variables platelets, neutrophils, age, haemoglobin, CA, LDH, ALP, tumour grade, WBC, histology, liver metastasis, lung metastasis, bone metastasis, brain metastasis, which were selected for the multivariate model using backward selection.

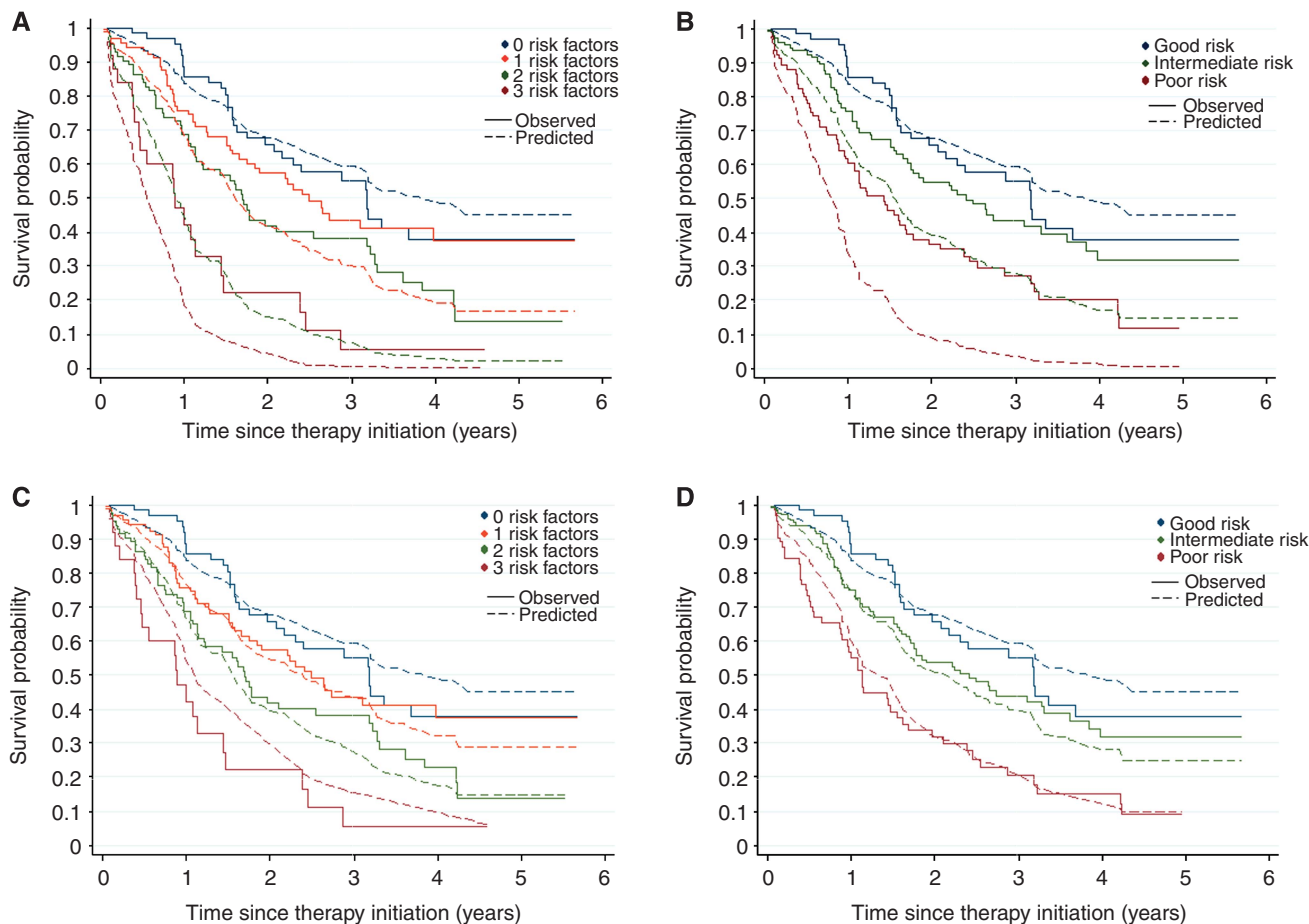


Figure 2. Observed (solid lines) and predicted (dashed lines) overall survival by risk classification according to the number of risk factors (A, C) or the percentiles of the prognostic index (B, D) for the validation data set before (A, B) and after (C, D) recalibration.

Table 4. Survival according to risk stratification in the development and validation cohorts

Prognostic groups	Development data set				Validation data set			
	Deaths/No. of patients	Median observed OS	Median predicted OS	Hazard ratios	Deaths/No. of patients	Median observed OS	Median predicted OS	Hazard ratios
Our model (after calibration in the validation cohort)								
0 risk factors	15/48	NR	NR	1	31/67	38.1	50.2	1
1 risk factors	33/55	24.7	22.3	2.27	45/92	30	29	1.26
2 risk factors	33/45	12.8	11.5	4.32	54/82	20.4	19.2	1.88
3 risk factors	22/22	5.9	6.4	10.48	21/25	10.6	13.6	4.09
Good risk	15/48	NR	NR	1	31/67	38.1	50.2	1
Intermediate risk	46/78	22.4	21.7	2.41	70/135	29	26	1.36
Poor risk	42/44	7.7	8.6	8.18	50/64	13.5	17.2	2.78
Heng's model (after calibration in both cohorts)								
Favourable	6/16	37.4	37.4	1	24/53	40.2	43.2	1
Intermediate	28/57	28.8	24.8	1.62	77/126	21	21.4	1.73
Poor	26/32	11.2	11.2	3.64	37/48	13.6	13.1	2.79

Abbreviations: NR = not reached; OS = overall survival.

DISCUSSION

An ideal prognostic model should be easy to use, include only the most relevant patient and disease characteristics and accurately

distinguish patient groups with different prognosis. Our model fully meets the first two criteria and has satisfactory discriminatory ability, although there is room for improvement.

Other prognostic models for mRCC have been previously proposed by the MSKCC (Motzer *et al*, 2001), the Cleveland Clinic

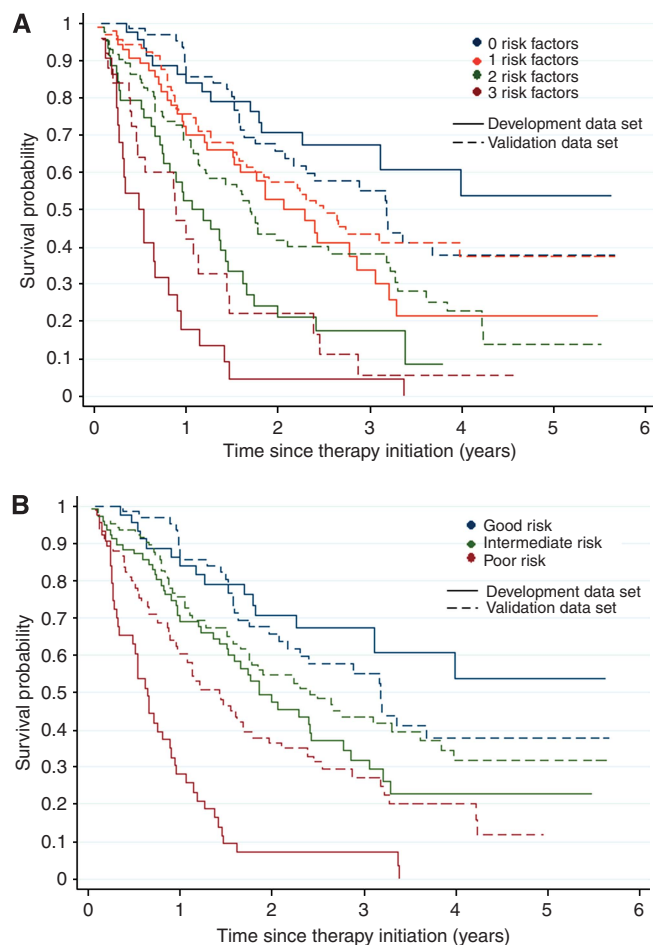


Figure 3. Observed overall survival for the development (solid lines) and validation data sets (dashed lines) by risk classification according to number of risk factors (A) or percentiles of the prognostic index (B).

Foundation (Choueiri *et al*, 2007), French investigators (Negrier *et al*, 2005), the International Kidney Cancer Working Group (IKCWG) (Manola *et al*, 2011) and IDC (Heng *et al*, 2009). All these models, except from the IDC, are on the basis of outcomes of patients treated with immunotherapy or on single-institution experiences and have not always been externally validated. Our model is simpler, including only three clinical factors, usually readily available for every patient with mRCC. This is confirmed by the fact that this information was available for almost all patients in the validation cohort, although patients were not selected on the basis of the availability of such information. Among the three prognostic factors included in our model, PS and time from diagnosis have been consistently found significant in all relevant studies both in cytokine and targeted therapy era (Motzer *et al*, 1999, 2001; Negrier *et al*, 2005; Choueiri *et al*, 2007; Escudier *et al*, 2007c; Motzer *et al*, 2008; Heng *et al*, 2009; Motzer *et al*, 2009; Beuselinck *et al*, 2011; Karakiewicz *et al*, 2011), while number of metastatic sites has been shown to be an independent prognostic factor in several mRCC series (Negrier *et al*, 2005; Escudier *et al*, 2007c; Manola *et al*, 2011; Poprach *et al*, 2012). Our patients were homogeneously treated with sunitinib, which is one of the most active agents in mRCC (Patard *et al*, 2011), the follow-up is among the longest reported in studies with targeted therapies (Heng *et al*, 2009; Beuselinck *et al*, 2011; Karakiewicz *et al*, 2011; Heng *et al*, 2013) and most patients were not included in clinical trials, thus making it applicable in everyday practice. Internal validation showed good discriminatory ability with a C-index of 0.712, similar to that reported for the IDC model (Heng *et al*, 2009).

Two methods for risk stratification were used. No superiority of one over the other was found. We believe that stratification according to the number of risk factors maintains the simplicity of the model and is more easily applicable in a clinical setting. This classification clearly identifies a group of poor prognosis (three risk factors), which does not seem to benefit from sunitinib therapy (median OS, 5.9 months). Such poor-outcome group has not been identified by previous studies and represents an advantage of the proposed model.

External validation yielded a C-index of 0.634, which is lower than that yielded by internal validation. Nevertheless, it is comparable to that reported for other published models, when studied in independent data sets (Heng *et al*, 2013). The less optimal performance was mainly found in the groups with the inferior prognosis, where the observed survival was better than the predicted by our model, although these differences were ameliorated with calibration. The reasons for this discrepancy are obscure. There were imbalances between the two cohorts in certain baseline characteristics as well as in the distribution across MSKCC and IDC risk groups in favour of the validation cohort. This is not infrequent, and has also been reported in other similar studies (Kang *et al*, 2012; Poprach *et al*, 2012; Yi *et al*, 2012). Considering the imbalanced factors, time from diagnosis and number of metastatic sites have been included into the final model and, therefore, their imbalance has been accounted for. As most patients in the validation cohort had undergone nephrectomy, separate validation studies, including only nephrectomised patients, were performed. This did not significantly improve our results (data not shown). The other three imbalanced factors, that is, previous IFN, non-clear histology and LDH were not further investigated. Previous IFNa and histology were not found to be significant in univariate analysis, while LDH was not available in 27% of our patients, limiting the power of further analyses. Other factors may have also affected our results. It has recently been suggested that eligible-for-studies patients may have different outcomes than non-eligible patients (Heng *et al*, 2012), while survival in expanded access programs (EAPs) for sunitinib has been lower than that of the randomised study (Gore *et al*, 2009; Motzer *et al*, 2009). Most patients from the French centres had been included in clinical trials in contrast to Greek and Belgian patients. Although not all our patients would be ineligible for trials, median survival of our cohort resembled that of the EAP, while median OS of the validation cohort approximated that of the randomized study. Inclusion in clinical studies may affect outcome through more thorough tumour evaluation and follow-up. This may be particularly true for the detection of metastatic sites. For these reasons, we performed additional analyses using only Belgian patients as the validation cohort and also using a model with only PS and time from diagnosis. These analyses did not result in better performance of our model (data not shown) but the relatively small numbers included in these subgroups may limit these analyses.

Among the previously developed models, that proposed by IDC (Heng *et al*, 2009) is rapidly gaining acceptance, as it is the only one developed with patients treated with targeted therapies, has been externally validated and seems to have higher stratification capability than the others (Heng *et al*, 2013). We, therefore, attempted to validate this model in our two independent, homogeneously treated, non-selected populations. Median OS of the IDC risk groups was fairly similar between development and validation cohorts, which is an improvement over our model and supports its applicability in mRCC patients. Nevertheless, in both cohorts, C-index was below 0.6, lower than the 0.634 of our model and that of 0.664, yielded by the external validation procedure for IDC (Heng *et al*, 2013). Again, the most notable deviation of the predicted from the observed survival was found in the poor-risk groups, where median OS was higher (13.6 and 11.2 months) than the reported 7.8 months (Heng *et al*, 2013). The latter could be, at

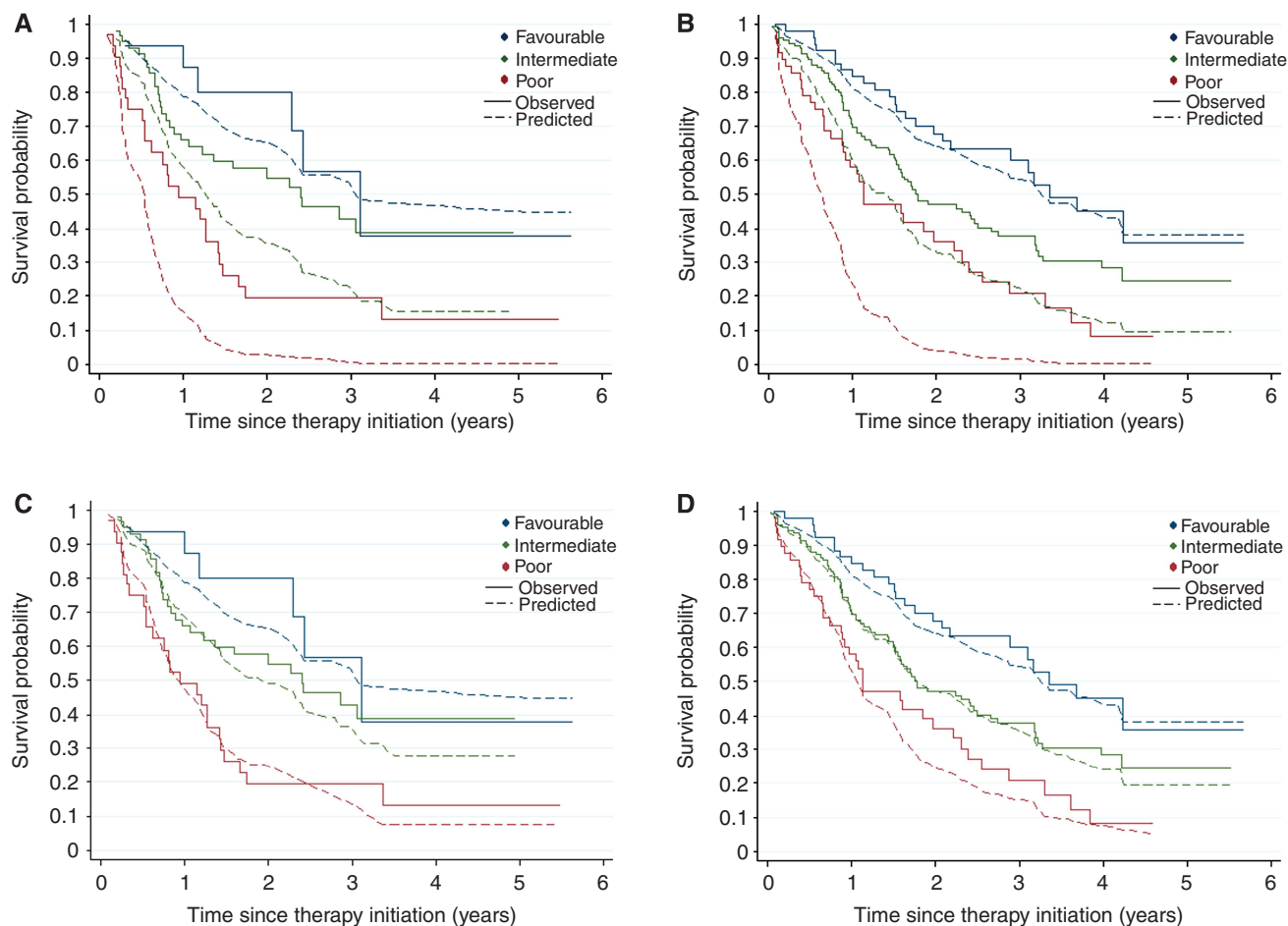


Figure 4. Observed (solid lines) and predicted (dashed lines) overall survival by risk classification according to the number of risk factors using the model proposed by mRCC IDC in the development (A, C) or the validation (B, D) data sets before (A, B) and after (C, D) recalibration.

least partially, attributed to the fact that 1/3rd of those patients received sorafenib as first-line treatment, which is considered inferior to sunitinib. Nevertheless, sorafenib-treated patients had similar OS, while there was no available data regarding the treatment of the poor-risk group. The less satisfactory performance of both models in poor-risk patients, a group under-represented in clinical trials with targeted therapies, underlines the necessity for better characterisation of this group through more focused clinical research. In addition, the lower C-indices yielded by external validation for both models, compared with those by internal validation, underline the importance of external validation and the need for confirmation in multiple data sets before the wide acceptance of a proposed prognostic model. There may exist certain, yet unidentified, factors, which might affect outcome in mRCC patients treated with anti-VEGF therapies and might account for the limitations of the existing models. Recent data (Peña *et al*, 2010; Sun *et al*, 2011) suggest that the introduction of molecular factors may improve the performance of models relying purely on clinical factors.

In conclusion, we externally validated a simple model, which could be used to stratify patients with mRCC offered sunitinib. Although we believe that it could be used for any type of anti-VEGF therapy, this remains to be confirmed. The predictive accuracy of this model appears comparable to that of the more complex IDC model and could, therefore, represent a valid alternative. Both models did not perform equally well in poor-risk populations, which suggests that further refinement in additional independent data sets may be appropriate.

CONFLICT OF INTEREST

Bamias—advisory (compensated): Pfizer, Bayer, Novartis, Roche, GSK, Astellas; honoraria: Novartis, Bayer, GSK. Escudier—honoraria: Bayer, Roche, Pfizer, Novartis, GSK, Aveo, Astellas. Oudard—honoraria: Pfizer, Bayer, Sanofi Aventis, Roche, Keocyt, Janssens, Takeda, Amgen, Novartis. The remaining authors declare no conflict of interest.

REFERENCES

- Bamias A, Karadimou A, Lampaki S, Lainakis G, Malettou L, Timotheadou E, Papazisis K, Andreadis C, Kontovinis L, Anastasiou I, Stravodimos K, Xanthakis I, Skolarikos A, Christodoulou C, Syrigos K, Papandreou C, Razi E, Dafni U, Fountzilas G, Dimopoulos MA (2010) Prognostic stratification of patients with advanced renal cell carcinoma treated with sunitinib: comparison with the Memorial Sloan-Kettering prognostic factors model. *BMC Cancer* **10**: 45.
- Beuselinc B, Oudard S, Rixe O, Wolter P, Blesius A, Ayllon J, Elaidi R, Schöffski P, Barrascout E, Morel A, Escudier B, Lang H, Zucman-Rossi J, Medioni J (2011) Negative impact of bone metastasis on outcome in clear-cell renal cell carcinoma treated with sunitinib. *Ann Oncol* **22**: 794–800.
- Bhojani N, Jeldres C, Patard JJ, Perrotte P, Suardi N, Hutterer G, Patenaude F, Oudard S, Karakiewicz PI (2008) Toxicities associated with the administration of Sorafenib, Sunitinib, and Tamsirolimus and their management in patients with metastatic renal cell carcinoma. *Eur Urol* **53**: 917–930.

- Choueiri T, Garcia J, Elson P, Khasawneh M, Usman S, Golshayan AR, Baz RC, Wood L, Rini BI, Bukowski RM (2007) Clinical factors associated with outcome in patients with metastatic clear-cell renal cell carcinoma treated with vascular endothelial growth factor-targeted therapy. *Cancer* **110**: 543–550.
- Chow LQM, Eckhardt SG (2007) Sunitinib: From rational design to clinical efficacy. *J Clin Oncol* **25**: 884–896.
- Cohen H, McGovern F (2005) Renal Cell Carcinoma. *N Engl J Med* **353**: 3477–3490.
- Escudier B, Eisen T, Stadler WM, Szczylik C, Oudard S, Siebels M, Negrier S, Chevreau C, Solska E, Desai AA, Rolland F, Demkow T, Hutson TE, Gore M, Freeman S, Schwartz B, Shan MH, Simantov R, Bukowski RM (2007a) Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med* **356**: 125–134.
- Escudier B, Pluzanska A, Koralewski P, Ravaud A, Bracarda S, Szczylik C, Chevreau C, Filipek Melichar B, Bajetta E, Gorbunova V, Bay JO, Bodrogi I, Jagiello-Gruszfeld A, Moore N (2007b) Bevacizumab plus interferon alfa-2a for the treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet* **370**: 22–29.
- Escudier B, Choueiri TK, Oudard S, Szczylik C, Négrier S, Ravaud A, Chevreau C, Venner P, Champagne P, Croteau D, Dupont E, Hariton C, Bukowski RM (2007c) Prognostic factors of metastatic renal cell carcinoma after failure of immunotherapy: new paradigm from a large phase III trial with shark cartilage extract AE 941. *J Urol* **178**: 1901–1905.
- Escudier B, Szczylik C, Hutson TE, Demkow T, Staehler M, Rolland F, Negrier S, Laferrriere N, Scheuring UJ, Cella D, Shah S, Bukowski RM (2009) Randomized phase II trial of first-line treatment with sorafenib versus interferon Alfa-2a in patients with metastatic renal cell carcinoma. *J Clin Oncol* **27**: 1280–1289.
- Flanigan RC, Mickisch G, Sylvester R, Tangen C, Van Poppel H, Crawford ED (2004) Cytoreductive nephrectomy in patients with metastatic renal cancer: a combined analysis. *J Urol* **171**: 1071–1076.
- Gore ME, Szczylik C, Porta C, Bracarda S, Bjarnason GA, Oudard S, Hariharan S, Lee SH, Haanen J, Castellano D, Vrdoljak E, Schöffski P, Mainwaring P, Nieto A, Yuan J, Bukowski R (2009) Safety and efficacy of sunitinib for metastatic renal-cell carcinoma: an expanded-access trial. *Lancet Oncol* **10**: 757–763.
- Heng DYC, Choueiri TK, Lee J, Harshman LC, Bjarnason GA, Knox JJ, MacKenzie M, Vaisha mpayan UN, Tan MH, Rha SY, Donskov F, Agarwal N, Kollmannsberger C, North S, Rini BI, Wood L (2012) Outcomes of patients with metastatic renal cell carcinoma that do not meet eligibility criteria for clinical trials. *J Clin Oncol* **30**: (suppl 5: abstract 353).
- Heng D, Wanling X, Meredith M, Warren MA, Golshayan AR, Sahi C, Eigel BJ, Ruether JD, Cheng T, North S, Venner P, Knox JJ, Chi KN, Kollmannsberger C, McDermott DF, Oh WK, Atkins MB, Bukowski RM, Rini BI, Choueiri TK (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**: 5794–5799.
- Heng DYC, Xie W, Regan MM, Harshman LC, Bjarnason GA, Vaishampayan UN, Mackenzie M, Wood L, Donskov F, Tan MH, Rha SY, Agarwal N, Kollmannsberger C, Rini BI, Choueiri TK (2013) External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study. *Lancet Oncol* **14**: 141–148.
- Kang S, Nam B-H, Park J-Y, Seo S-S, Ryu S-Y, Kim JW, Kim S-C, Park S-Y, Nam J-H (2012) Risk assessment tool for distant recurrence after platinum-based chemoradiation in patients with locally advanced cervical cancer: a Korean Gynecologic Oncology Group Study. *J Clin Oncol* **30**: 2369–2374.
- Karakiewicz PI, Sun M, Bellmunt J, Sneller V, Escudier B (2011) Prediction of progression-free survival rates after bevacizumab plus interferon versus interferon alone in patients with metastatic renal cell carcinoma: comparison of a nomogram to the Motzer criteria. *Eur Urol* **60**: 48–56.
- Kavolius JP, Mastorakos DP, Pavlovich C, Russo P, Burt ME, Brady MS (1998) Resection of metastatic renal cell carcinoma. *J Clin Oncol* **16**: 2261–2266.
- Manola J, Royston P, Elson P, McCormack JB, Mazumdar M, Negrier S, Escudier B, Eisen T, Dutcher J, Atkins M, Heng DYC, Choueiri TK, Motzer R, Bukowski R (2011) Prognostic model for survival in patients with metastatic renal cell carcinoma: results from the International Kidney Cancer Working Group. *Clin Cancer Res* **17**: 5443–5450.
- Miller ME, Hui SL (1991) Validation techniques for logistic regression models. *Stat Med* **10**: 1213–1226.
- Motzer R, Bukowski R, Figlin R, Hutson TE, Michaelson MD, Kim ST, Baum CM, Kattan MW (2008) Prognostic nomogram for sunitinib in patients with metastatic renal cell carcinoma. *Cancer* **113**: 1552–1558.
- Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O, Oudard S, Negrier S, Szczylik C, Kim ST, Chen I, Bycott PW, Charles Baum CM, Figlin RA (2007) Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med* **356**: 115–124.
- Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Oudard S, Negrier S, Szczylik C, Pili R, Bjarnason GA, Garcia-del-Muro X, Sosman JA, Solska E, Wilding G, Thompson JA, Kim ST, Chen I, Huang X, Figlin RA (2009) Overall survival and updated results for sunitinib compared with interferon alfa in patients with metastatic renal cell carcinoma. *J Clin Oncol* **27**: 3584–3590.
- Motzer RJ, Mazumdar M, Bacik J, Berg W, Amsterdam A, Ferrara J (1999) Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. *J Clin Oncol* **17**: 2530–2540.
- Motzer RJ, Murphy BA, Bacik J, Russa P, Mazumdar M (2001) Interferon-alfa as a comparative treatment for clinical trials of new therapies against advanced renal cell carcinoma. *J Clin Oncol* **20**: 289–296.
- Negrier S, Gomez F, Douillard J-Y, Ravaud A, Chevreau C, Buclon M, Perol D, Lasset C, Escudier B (2005) Prognostic factors of response or failure of treatment in patients with metastatic renal carcinomas treated by cytokines: a report from the Groupe Français d'Immunothérapie. *World J Urol* **23**: 161–165.
- Patard JJ, Pignot G, Escudier B, Eisen T, Bex A, Sternberg C, Rini B, Roigas J, Choueiri T, Bukowski R, Motzer R, Kirkali Z, Mulders P, Bellmunt J (2011) ICUD-EAU International Consultation on Kidney Cancer 2010: Treatment of Metastatic Disease. *Eur Urol* **60**: 684–690.
- Peña C, Lathia C, Shan M, Escudier B, Bukowski RM (2010) Biomarkers predicting outcome in patients with advanced renal cell carcinoma: Results from sorafenib phase III Treatment Approaches in Renal Cancer Global Evaluation Trial. *Clin Cancer Res* **16**: 4853–4863.
- Pencina MJ, D'Agostino RB (2004) Overall c as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* **23**: 2109–2123.
- Poprach A, Pavlik T, Melichar B, Puzanov I, Dusek L, Bortilicek Z, Vyzula R, Abrahamova J, Buchler T (2012) Skin toxicity and efficacy of sunitinib and sorafenib in metastatic renal cell carcinoma: a national registry-based study. *Ann Oncol* **12**: 3137–3140.
- Royston P, MKB Parmar, Altman G (2010) External validation and updating of a prognostic survival model. Research report No. 307, Department of Statistical Science, University College London <http://www.ucl.ac.uk/statistics/research/pdfs/rr307.pdf>.
- Sun M, Shariat SF, Cheng C, Ficara V, Murai M, Oudard S, Pantuck AJ, Zigeuner R, Karakiewicz PI (2011) Prognostic factors and predictive models in renal cell carcinoma: a contemporary review. *Eur Urol* **60**: 644–661.
- Yi M, Meric-Bernstam F, Kuerer HM, Mittendorf EA, Bedrosian I, Lucci A, Hwang RF, Crow JR, Luo S, Hunt KK (2012) Evaluation of a breast cancer nomogram for predicting risk of ipsilateral breast tumor recurrence in patients with ductal carcinoma in situ after local excision. *J Clin Oncol* **30**: 6000–6007.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.