

# External validation and comparison with other models of the International Metastatic Renal-Cell Carcinoma Database Consortium prognostic model: a population-based study



Daniel Y C Heng\*, Wanling Xie\*, Meredith M Regan, Lauren C Harshman, Georg A Bjarnason, Ulka N Vaishampayan, Mary Mackenzie, Lori Wood, Frede Donskov, Min-Han Tan, Sun-Young Rha, Neeraj Agarwal, Christian Kollmannsberger, Brian I Rini, Toni K Choueiri

## Summary

**Background** The International Metastatic Renal-Cell Carcinoma Database Consortium model offers prognostic information for patients with metastatic renal-cell carcinoma. We tested the accuracy of the model in an external population and compared it with other prognostic models.

**Methods** We included patients with metastatic renal-cell carcinoma who were treated with first-line VEGF-targeted treatment at 13 international cancer centres and who were registered in the Consortium's database but had not contributed to the initial development of the Consortium Database model. The primary endpoint was overall survival. We compared the Database Consortium model with the Cleveland Clinic Foundation (CCF) model, the International Kidney Cancer Working Group (IKCWG) model, the French model, and the Memorial Sloan-Kettering Cancer Center (MSKCC) model by concordance indices and other measures of model fit.

**Findings** Overall, 1028 patients were included in this study, of whom 849 had complete data to assess the Database Consortium model. Median overall survival was 18.8 months (95% CI 17.6–21.4). The predefined Database Consortium risk factors (anaemia, thrombocytosis, neutrophilia, hypercalcaemia, Karnofsky performance status <80%, and <1 year from diagnosis to treatment) were independent predictors of poor overall survival in the external validation set (hazard ratios ranged between 1.27 and 2.08, concordance index 0.71, 95% CI 0.68–0.73). When patients were segregated into three risk categories, median overall survival was 43.2 months (95% CI 31.4–50.1) in the favourable risk group (no risk factors; 157 patients), 22.5 months (18.7–25.1) in the intermediate risk group (one to two risk factors; 440 patients), and 7.8 months (6.5–9.7) in the poor risk group (three or more risk factors; 252 patients;  $p < 0.0001$ ; concordance index 0.664, 95% CI 0.639–0.689). 672 patients had complete data to test all five models. The concordance index of the CCF model was 0.662 (95% CI 0.636–0.687), of the French model 0.640 (0.614–0.665), of the IKCWG model 0.668 (0.645–0.692), and of the MSKCC model 0.657 (0.632–0.682). The reported versus predicted number of deaths at 2 years was most similar in the Database Consortium model compared with the other models.

**Interpretation** The Database Consortium model is now externally validated and can be applied to stratify patients by risk in clinical trials and to counsel patients about prognosis.

**Funding** None.

## Introduction

Treatment of metastatic renal-cell carcinoma (RCC) has been revolutionised by targeted treatments such as those directed against VEGF. This class of agents—which includes sunitinib,<sup>1</sup> sorafenib,<sup>2</sup> bevacizumab,<sup>3,4</sup> pazopanib,<sup>5</sup> and axitinib<sup>6</sup>—has been included in treatment for patients with this advanced disease. The new era of targeted treatment needs new prognostic models and updated survival data for accurate clinical trial design, patient counselling, and risk-specific treatment. Thus, the International Metastatic RCC Database Consortium<sup>7,8</sup> derived the first prognostic model since the development of targeted treatment from a large multicentre cohort. Six independent predictors of poor survival were identified: Karnofsky performance status of less than 80%, less than 1 year from diagnosis to treatment, anaemia (haemoglobin concentration <lower limit of normal), hypercalcaemia (corrected calcium

concentration >upper limit of normal), neutrophilia (neutrophil count >upper limit of normal), and thrombocytosis (platelet count >upper limit of normal). According to the number of poor prognostic factors, patients were segregated into favourable (no factors), intermediate (one or two factors), and poor (more than three factors) risk groups.

Other prognostic models for metastatic RCC exist but are based on outcomes of patients treated with immunotherapy or on single-institution experiences (table 1). The most widely used system is the Memorial Sloan-Kettering Cancer Center (MSKCC) model,<sup>13</sup> which contains many of the same factors as the Database Consortium model. Other models include the Cleveland Clinic Foundation (CCF) model,<sup>9</sup> the updated French model adapted to the AVOREN trial,<sup>10,11</sup> and the International Kidney Cancer Working Group (IKCWG) model.<sup>12</sup>

*Lancet Oncol* 2013; 14: 141–48

Published Online

January 9, 2013

[http://dx.doi.org/10.1016/S1470-2045\(12\)70559-4](http://dx.doi.org/10.1016/S1470-2045(12)70559-4)

See [Comment](#) page 102

\*Joint first authors

Tom Baker Cancer Center, Calgary, AB, Canada (D Y C Heng MD); Dana-Farber Cancer Institute, Harvard University, Boston, MA, USA (W Xie MS, M M Regan ScD, T K Choueiri MD); Stanford University, Stanford Cancer Institute, Stanford, CA, USA (L C Harshman MD); Sunnybrook Odette Cancer Institute, Toronto, ON, Canada (G A Bjarnason MD); Wayne State University, Detroit, MI, USA (U N Vaishampayan MD); London Health Sciences Center, London, ON, Canada (M Mackenzie MD); Queen Elizabeth II Health Sciences Center, Halifax, NS, Canada (L Wood MD); Aarhus University Hospital, Aarhus, Denmark (F Donskov MD); National Cancer Center, Institute of Bioengineering and Nanotechnology, Singapore (M-H Tan MD); Yonsei University Hospital, Seoul, South Korea (S-Y Rha MD); University of Utah, Huntsman Cancer Institute, Salt Lake City, UT, USA (N Agarwal MD); BC Cancer Agency, Vancouver, BC, Canada (C Kollmannsberger MD); and Cleveland Clinic Taussig Cancer Institute, Cleveland, OH, USA (Prof B I Rini MD)

Correspondence to:

Dr Daniel Y C Heng, Tom Baker Cancer Center, University of Calgary, Calgary, AB, T2N 4N2, Canada  
[daniel.heng@albertahealthservices.ca](mailto:daniel.heng@albertahealthservices.ca)

	Database Consortium model <sup>7</sup>	CCF model <sup>9</sup>	French model <sup>10,11</sup>	IKCWG model <sup>12*</sup>	MSKCC model <sup>13</sup>
<b>Risk factors</b>					
Karnofsky performance status or ECOG PS	✓	✓	✓	✓	✓
Time from diagnosis to treatment	✓	✓	..	✓	✓
Time from diagnosis to metastasis	..	..	✓	..	..
Number of metastatic sites	..	..	✓	✓	..
Liver metastasis	..	..	✓	..	..
Previous immunotherapy	..	..	..	✓	..
Haemoglobin concentration	✓	..	..	✓	✓
Corrected or uncorrected calcium concentration	✓	✓	..	✓	✓
Neutrophil count	✓	✓	..	..	..
Platelet count	✓	✓	..	..	..
Lactate dehydrogenase concentration	..	..	..	✓	✓
White blood cell count	..	..	..	✓	..
Alkaline phosphatase concentration	..	..	..	✓	..
<b>Risk groups</b>					
Favourable	0 risk factors	0–1 risk factors	ECOG PS of 0 and one metastasis	Risk score ≤–2.755	0 risk factors
Intermediate	1–2 risk factors	2 risk factors	All others	Risk score >–2.755 to ≤–1.253	1–2 risk factors to ≤–1.253
Poor	≥3 risk factors	≥3 risk factors	Liver and other metastasis and <1 year to metastasis, or ECOG PS >1	Risk score >–1.253	≥3 risk factors
<b>Distribution of risk groups</b>					
In this database (n=672)					
Favourable	117 (17%)	229 (34%)	36 (5%)	91 (15%)	128 (19%)
Intermediate	347 (52%)	246 (37%)	450 (67%)	336 (47%)	398 (59%)
Poor	208 (31%)	197 (29%)	186 (27%)	245 (38%)	146 (22%)
Published previously <sup>7,9,12,13</sup>					
Favourable	133/586 (23%)	63/120(53%)	..	937/3748 (25%)	80/437 (18%)
Intermediate	301/586 (51%)	27/120 (23%)	..	1874/3748 (50%)	269/437 (62%)
Poor	152/586 (26%)	30/120 (25%)	..	937/3748 (25%)	88/437 (20%)
CCF=Cleveland Clinic Foundation. IKCWG=International Kidney Cancer Working Group. MSKCC=Memorial Sloan-Kettering Cancer Center. ECOG PS=Eastern Cooperative Oncology Group performance score. *The cutoff values for the risk groups (–2.78 and –1.25) are the 25th and 75th percentiles of the distribution of risk scores from the IKCWG model.					

**Table 1: Prognostic models and distributions of risk groups**

An ideal prognostic model is easy to use, includes only the most relevant patient and disease characteristics, and is able to accurately distinguish between groups of patients with different outcomes. We tested the validity of the metastatic RCC Database Consortium model in a large international multicentre dataset and compared its accuracy with other prognostic models.

## Methods

### Participants

In this population-based analysis, we included consecutive patients from 13 international cancer centres (five in the USA, five in Canada, one in South Korea, one in Singapore, and one in Denmark). The 645 patients originally used<sup>7</sup> to derive the Database Consortium model were not included in this analysis. We collected data between Aug 15, 2008, and Jan 14, 2011. Included patients

had metastatic RCC treated between 2004 and 2010 with an anti-VEGF targeted treatment (sunitinib, sorafenib, bevacizumab, axitinib, or pazopanib) as their first anti-VEGF agent. Previous immunotherapy was allowed (ie, targeted treatment as second-line treatment). Patients treated with front-line mTOR inhibitors were excluded.

We collected baseline patient characteristics and outcome data with uniform data collection templates as described previously.<sup>7</sup> Laboratory test results were standardised against institutional upper limit of normal and lower limit of normal values when appropriate. The study was approved by the institutional review board at each participating centre.

### Statistical analysis

The primary endpoint was overall survival, defined as the time from start of targeted treatment to death or censored

at date of last follow-up. We assessed the predictive accuracy of the model by the concordance index,<sup>14</sup> which is the area under the receiver operating curve for survival time in the presence of censored data. A concordance index of 0·5 represents no predictive discrimination and an index of 1 represents perfect ability to distinguish patients.

We classified patients into risk groups with four other existing prognostic models (CCF, French, IKCWG, and MSKCC; table 1) and fitted them into Cox regressions for overall survival. Each of these models was compared to the Database Consortium model with: (1) Bayes information criterion, a global measure of model fit in which a low number represents a good fit; (2) generalised R<sup>2</sup>, a statistic between 0 and 1 that is large when the covariates are strongly associated with the dependent variable;<sup>15</sup> and (3) the concordance index.

We also compared performance between models with new measures based on reclassification of risk categories—ie, reclassification calibration<sup>16</sup> and net reclassification improvement.<sup>17</sup> These measures are based on a cross-tabulation comparing the Database Consortium model with the others. The 3×3 table of the risk groups (favourable, intermediate, and poor) for the two models in each comparison provides nine groups. For each group, we compared the reported number of events from Kaplan-Meier estimates with the predicted number of events from Cox regression at 2 years after start of treatment with the following formulas:

Reported number of deaths = number of patients in the group×2-year death rate for that group estimated by Kaplan-Meier method

Predicted number of deaths=number of patients in that group×predicted 2-year risk rate for the risk group according to the model

These calculations generate two reclassification calibration  $\chi^2$  statistics: the model with the smaller  $\chi^2$  has a better fit (K–2 degrees of freedom, where K is the number of cells with at least 20 observations).

For the net reclassification improvement analysis, the 3×3 tables were further stratified by patient survival status at 2 years after the start of treatment. We calculated the proportions of participants who were classified into different risk groups (either a better or worse risk group) in each model separately for dead and alive patients. Patients who had not reached 2 years of follow-up were excluded from this analysis.

Net reclassification improvement=(prob(worse risk group/died)–prob(better risk group/died))+ (prob(better risk group/alive)–prob(worse risk group/alive))

Thus, a higher net reclassification improvement means that a model had a better reclassification compared with the other model—ie, the model is more likely to classify

dead patients to a poor risk group or alive patients to a favourable risk group.

We used multiple imputation to account for missing data.<sup>18</sup> Unlike single imputation methods, multiple imputation yields several plausible imputed datasets to account for the uncertainty caused by missing data. These multiple-imputed datasets are then analysed by using standard procedures for complete data and combining the results obtained from each. For this analysis, five imputation datasets for missing data were created with the ice package of Stata (version 11) to ensure that results were consistent when compared with the complete case analyses. Each imputation dataset was analysed with the same methods as from the original dataset without imputation. Rubin's rules<sup>19</sup> were used to combine results from the five imputation datasets, by computing the mean of the five estimates and a variance estimate that includes components for both within-imputation and across-imputation for each measurement of model fit when appropriate. Sample sizes were determined by the size of consecutive cohorts of patients from each centre. We report results from the complete case analysis in this report and results from the imputation datasets are included in the appendix. Statistical computations were done with SAS (version 9.2) and R (version 2.12). Reclassification measures were done with SAS Macros implemented by Cook and Ridker.<sup>16</sup>

	n/N* (%)
Age ≥60 years	564/1028 (55%)
KPS <80%	261/970 (27%)
Men	765/1028 (74%)
>1 site of metastases	791/1024 (77%)
Brain metastasis	99/1025 (10%)
Liver metastasis	176/895 (20%)
Non-clear cell pathology	122/949 (13%)
Presence of sarcomatoid features	104/857 (12%)
Previous nephrectomy	798/1028 (78%)
Previous immunotherapy	245/1028 (24%)
Treatment	
Sunitinib	844/1028 (82%)
Sorafenib	134/1028 (13%)
Axitinib	1/1028 (<1%)
Bevacizumab	47/1028 (5%)
Pazopanib	2/1028 (<1%)
<1 year from diagnosis to targeted treatment	561/1026 (55%)
Haemoglobin concentration <lower limit of normal	541/968 (56%)
Serum corrected calcium concentration >upper limit of normal	86/894 (10%)
Lactate dehydrogenase concentration >1·5 × upper limit of normal	87/721 (12%)
Neutrophil count >upper limit of normal	175/934 (19%)
Platelet count >upper limit of normal	204/959 (21%)

KPS=Karnofsky performance status. \*Excluding patients with missing values.

**Table 2: Patient and disease baseline characteristics**

	Original model <sup>†</sup> (n=564)*		Validation (n=849)	
	Hazard ratio (95% CI)	p value	Hazard ratio (95% CI)	p value
KPS <80%	2.51 (1.92–3.29)	<0.0001	2.08 (1.71–2.55)	<0.0001
<1 year from diagnosis to treatment	1.42 (1.09–1.84)	0.0098	1.27 (1.05–1.53)	0.0122
Haemoglobin concentration <lower limit of normal	1.72 (1.31–2.26)	0.0001	1.69 (1.38–2.06)	<0.0001
Calcium concentration >upper limit of normal	1.81 (1.29–2.53)	0.0006	1.45 (1.10–1.92)	0.0087
Neutrophil count >upper limit of normal	2.42 (1.72–3.39)	<0.0001	1.64 (1.31–2.05)	<0.0001
Platelet count >upper limit of normal	1.49 (1.09–2.03)	0.0121	1.60 (1.28–2.01)	<0.0001

The concordance index of the original model was 0.73. For the validation group it was 0.71 (95% CI 0.68–0.73). KPS=Karnofsky performance status. \*From multivariable Cox regression.

**Table 3: Associations of overall survival with six prognostic risk factors from the Database Consortium model**

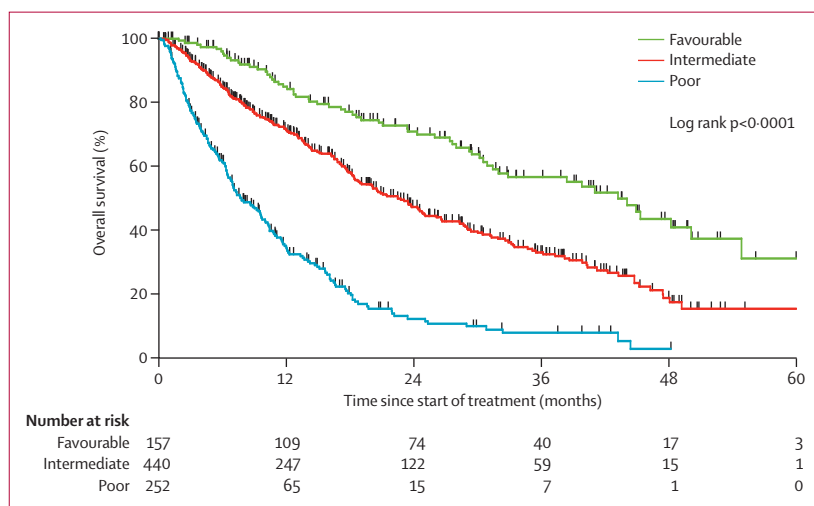


Figure 1: Results of Kaplan-Meier analysis of overall survival for the Database Consortium model

were in the favourable risk group and had a median overall survival of 43.2 months (95% CI 31.4–50.1). 440 patients (52%) were in the intermediate risk group and had a median overall survival of 22.5 months (95% CI 18.7–25.1). 252 patients (30%) were in the poor risk group and had a median overall survival of 7.8 months (95% CI 6.5–9.7). Figure 1 shows clear distinctions between risk groups (log rank  $p < 0.0001$ ). The concordance index of this model was 0.71 (95% CI 0.68–0.73) using the individual risk factors and 0.66 (0.64–0.69) when using the three risk groups.

Only the 672 patients with complete data for all five prognostic models were included in the comparison between the Database Consortium model and other models. The CCF model separates patients into three almost equal groups whereas in the French model, only 36 (5%) of 672 patients are in the favourable risk group. The three other models have about 50% of patients in the intermediate risk group. The Database Consortium model and MSKCC model are highly concordant, with 83% of patients classified into the same risk group by each model. Concordance of the Database Consortium model was 64%, with the CCF model, 61% with the French model, and 69% with the IKCWG model. The x-axes of figure 2 show concordance of risk groups and how many patients would change risk categories depending on the different prognostic criteria of each model.

We calculated concordance indices for each model (table 4). The indices were similar for all models except for the French model, which was slightly lower than the others. Other measures of fit showed that the Database Consortium model was not inferior to other models. Bayes information criterion was lowest for the Database Consortium model, suggesting that this model had the best fit and the generalised  $R^2$  was the highest, suggesting that this model was most strongly associated with outcomes (table 4).

We included analysis with new measures of fit, including the reclassification calibration test. This statistic is formed on the comparison of two models. In this case, nine groups from the  $3 \times 3$  cross-tabulation of the two models were evaluated (x-axes of figure 2). For

See Online for appendix

For R see <http://cran.r-project.org>

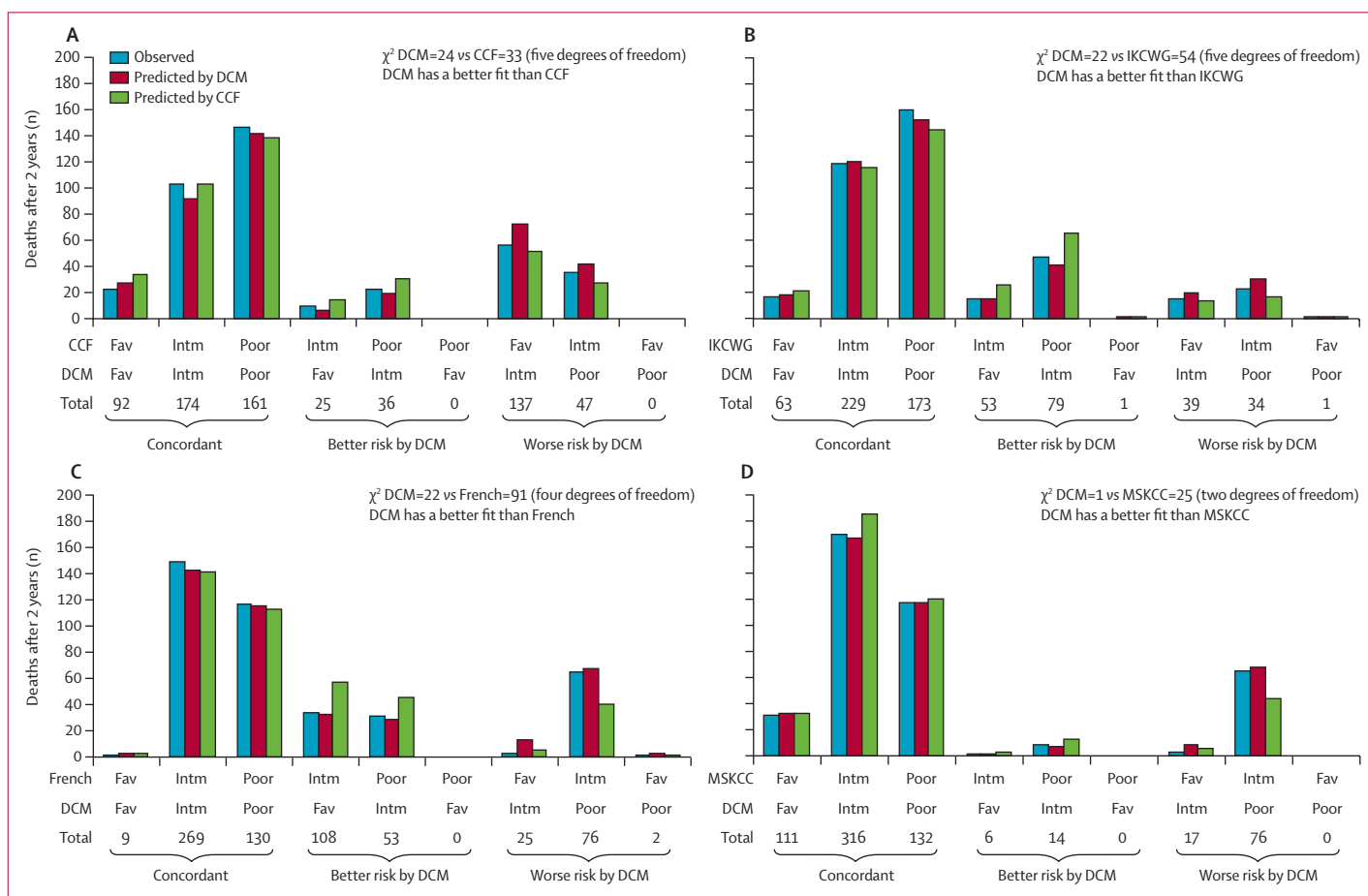
### Role of the funding source

There was no funding source for this study. DYCH, WX, and TKC had full access to all the data. The corresponding author had the final responsibility to submit for publication.

### Results

1028 patients were eligible for this analysis. At the cutoff date, 851 patients (83%) had discontinued targeted treatment and 447 (43%) were alive. Median follow-up in those alive was 16.3 months (IQR 7.4–30.6) and the median overall survival of all patients was 18.8 months (95% CI 17.6–21.4). Table 2 shows baseline characteristics.

Because data were missing for some laboratory measurements, only 849 patients (83%) had complete data for the Database Consortium model and 672 (65%) had complete data for all five models. In multivariable analysis for the Database Consortium model, the six risk factors were independent predictors of poor overall survival (hazard ratios [HRs] ranged from 1.27 to 2.08; table 3). The HRs in the validation dataset were much the same as those in the original model, which suggests excellent external validation. 157 of 849 (18%) patients



**Figure 2: Comparison of the Database Consortium model with other models for prognosis of metastatic renal-cell carcinoma**

Panels show reclassification calibration comparison of the Database Consortium model with the CCF model (A), the IKCWG model (B), the French model (C), and the MSKCC model (D). The x-axis includes nine groups from the 3 × 3 cross-tabulation table of risk groups of the two models being compared. A smaller reclassification calibration  $\chi^2$  suggests a better fit. CCF=Cleveland Clinic Foundation. IKCWG=International Kidney Cancer Working Group. MSKCC=Memorial Sloan-Kettering Cancer Center. DCM=Database Consortium model. Fav=Favourable. Intm=Intermediate.

each group, we compared the observed and predicted number of deaths from each of the two models at 2 years since treatment start. Two  $\chi^2$  values were generated, and a higher  $\chi^2$  shows increasing difference between the model's predicted risk and the actual observed risk—ie, that the model does not fit the data as well as one with a lower  $\chi^2$  result. The Database Consortium model had the lowest  $\chi^2$ , which means that the reported versus predicted number of deaths at 2 years was most similar in the Database Consortium model compared with the other models (figure 2). Particularly, patients who changed risk group (either to better or worse) in the Database Consortium model fitted better with the Database Consortium criteria than with all of the other models (figure 2). For example, 108 patients with intermediate risk according to the French model were classified as favourable risk by the Database Consortium model. The predicted 2 year death rate was 0·52 (56 deaths; table 5) if these patients were assigned to the intermediate risk group by the

	Concordance index (95% CI; rank)*	Bayes information criterion (rank)†	Generalised R <sup>2</sup> (rank)‡
Database Consortium model	0·664 (0·639–0·689; 2)	4341 (1)	0·185 (1)
CCF	0·662 (0·636–0·687; 3)	4361 (3)	0·161 (3)
French	0·640 (0·614–0·665; 5)	4380 (5)	0·136 (5)
IKCWG	0·668 (0·645–0·692; 1)	4370 (4)	0·149 (4)
MSKCC	0·657 (0·632–0·682; 4)	4359 (2)	0·163 (2)

n=672, patients with complete data only. CCF=Cleveland Clinic Foundation. IKCWG=International Kidney Cancer Working Group. MSKCC=Memorial Sloan-Kettering Cancer Center. \*A high concordance index suggests high discriminatory value. †A low value suggests a good global fit. ‡A large R<sup>2</sup> suggests that the model is strongly associated with the outcome.

**Table 4: Measures of model fit from Cox regression in the validation group**

French criteria, compared with 0·30 (32 deaths) if they were assigned to the favourable group by the Database Consortium model. The reported 2 year death rate from the Kaplan-Meier estimate was 0·30 (32 deaths), which was more similar to the prediction using the Database

	Favourable	Intermediate	Poor
DCM	0.30 (0.21-0.37)	0.53 (0.47-0.58)	0.88 (0.83-0.92)
CCF	0.37 (0.31-0.43)	0.60 (0.52-0.66)	0.86 (0.80-0.90)
French	0.19 (0.07-0.30)	0.52 (0.47-0.57)	0.86 (0.80-0.90)
IKCWG	0.35 (0.28-0.43)	0.50 (0.44-0.56)	0.84 (0.78-0.88)
MSKCC	0.30 (0.22-0.37)	0.58 (0.53-0.63)	0.91 (0.84-0.94)

Data are death rate (95% CI). CCF=Cleveland Clinic Foundation. IKCWG=International Kidney Cancer Working Group. MSKCC=Memorial Sloan-Kettering Cancer Center. DCM=Database Consortium model.

**Table 5: Predicted 2 year death rate for each model**

Consortium model (figure 2C). Likewise, 76 patients changed their risk category from the MSKCC intermediate risk group to the poor risk group by the Database Consortium model. The reported 2 year death rate for these patients was closer to the prediction using the Database Consortium model (figure 2D).

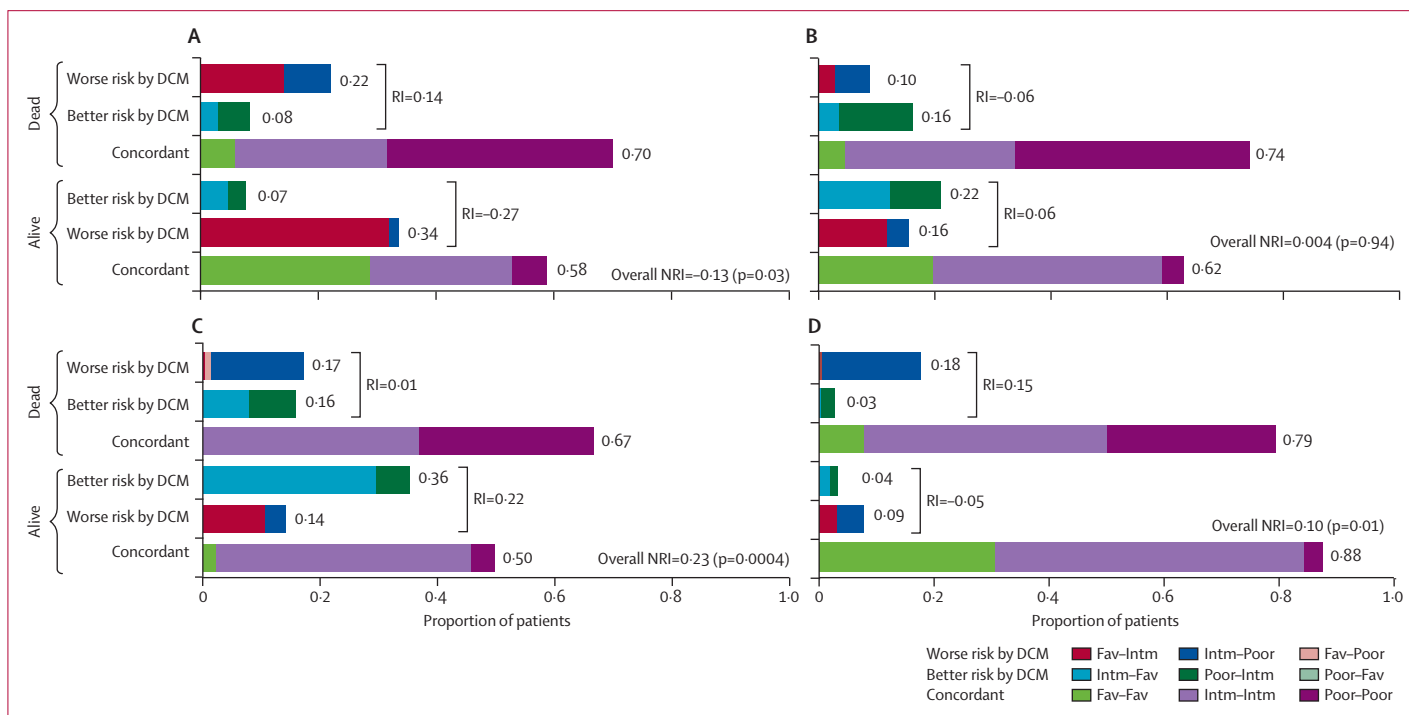
According to net reclassification improvement (figure 3), the Database Consortium model reclassified patients more correctly than did the French model (by 23%) and the MSKCC model (by 10%). We did not detect a significant improvement compared with the IKCWG model. The CCF model reclassified 13% of patients more correctly but had a lower concordance

index than did the Database Consortium model. Although the CCF model improved classification by net reclassification improvement in the first 2 years, no difference existed at 3 and 4 years (net reclassification improvement was 6% at 3 years and -1% at 4 years;  $p>0.05$ ). By contrast, the Database Consortium model classified patients to the correct risk group better than did the French model at 3 and 4 years (data not shown).

We used five imputation datasets to account for missing data in clinical covariates. The aggregated results across five imputation datasets had similar concordance indices, reclassification calibration, and net reclassification improvement compared with complete case analyses (appendix).

### Discussion

To our knowledge, this study is the largest external validation and comparison of prognostic models for metastatic RCC (panel). This study externally validates the International Metastatic RCC Database Consortium model. This study also provides clinicians with long-term overall survival data, which can be used for more accurate prognosis, patient counselling, and clinical trial design. A median overall survival of 43 months after the start of targeted treatment in the favourable risk group has set a new benchmark for this disease and is a testament to the



**Figure 3: Net reclassification improvement comparing the Database Consortium model with other models for prognosis of metastatic renal-cell carcinoma**

Panels show comparison of the DCM with the CCF model (A), the IKCWG model (B), the French model (C), and the MSKCC model (D), calculated according to patient's survival status at 2 years after start of anti-VEGF targeted treatment. 489 patients were included, of whom 161 were alive and 328 were dead. 183 who had not reached 2 years follow-up were excluded. As an example, comparing the French model with the DCM, 14% of alive patients were incorrectly moved to a worse risk group and 36% were correctly moved to a better risk group by the DCM, resulting in a 22% (36% - 14%) improvement. 17% of patients who had died were correctly moved to a worse risk group and 16% were incorrectly moved to a better risk group by DCM, resulting in a 1% (17% - 16%) improvement. The overall net reclassification improvement was 23% (22% + 1%). CCF=Cleveland Clinic Foundation. IKCWG=International Kidney Cancer Working Group. MSKCC=Memorial Sloan-Kettering Cancer Center. DCM=Database Consortium model. Fav=Favourable. Intm=Intermediate. RI=reclassification improvement. NRI=net reclassification improvement.



efficacy of targeted treatment. These survival results might be more generalisable to our clinic populations than are clinical trial data since the Database Consortium criteria are based on an unselected, consecutive series of patients.<sup>20</sup>

The comparisons of fit showed that different models can yield dissimilar prognosis on the basis of inclusion of different clinical factors. Overall, the concordance indices, aside from that of the French model, are within a similar range. Also, more complex models—eg, the IKCWG model, which includes mathematical transformations and more clinical factors—might not add significantly more accuracy or discriminatory power once simplified into three risk groups. The Database Consortium model has a high accordance of risk groups with MSKCC, except that 14% of the population was upgraded to a less favourable risk group by the Database Consortium model. By upgrading these patients to a higher risk (ie, from MSKCC intermediate to Database Consortium model poor), the model showed better fit and reclassification accuracy by both reclassification calibration and net reclassification improvement (figures 2D, 3D).

With at least five clinical prognostic nomograms for metastatic RCC, the use of clinical factors for prognosis has probably reached a limit. We use these models because they are the best available. Other models with similar concordance indices have been published for hepatocellular carcinoma<sup>21</sup> and prostate cancer.<sup>22</sup> The discriminatory ability of the Database Consortium model might be improved by using individual risk factors instead of collapsing them into three risk categories. The addition of tumour-specific or patient-specific biomarkers is the next likely step for improvement of the accuracy of these models. Angiogenesis biomarkers could add prognostic information for overall survival but need external validation.<sup>23</sup> Biomarkers related to germline polymorphisms might be another useful non-clinical factor for prognosis. For example, a study of 136 patients with clear-cell metastatic RCC treated with sunitinib reported a potentially favourable genetic profile. This profile included an A allele in the CYP3A5 6986A/G loci, a missing CAT copy in the NR1/3 haplotype, and a TCG copy in the ABCB1 haplotype. Patients with this profile had improved progression-free and overall survival compared with patients without.<sup>24</sup> In the future, incorporation of these potential biomarkers into the Database Consortium model might improve prognostic accuracy.

Strengths of this study are that it is generalisable and the sample size of previously unanalysed patients was large. We included patients who were treated in clinical trials, off protocol, in academic centres, in community centres, in several countries, and with all RCC histologies. Additionally, modern targeted treatment was used to treat these patients as determined by normal practice at each institution. Thus, these results represent treatment

#### Panel: Research in context

##### Systematic review

A Medline search for “metastatic renal cell carcinoma”, “prognostic factors”, and “external validation” yielded no articles that externally validated or compared the Database Consortium model to other models available for metastatic renal-cell carcinoma. The Database Consortium model was developed as a prognosis model for patients with metastatic renal-cell carcinoma treated in the era of targeted treatment.

##### Interpretation

This study externally validates the Database Consortium model in a new dataset of patients with metastatic renal-cell carcinoma treated with targeted treatment. As a result, this model can be used in clinical practice for patient counselling and risk stratification in clinical trials. A comparison of existing models shows that a ceiling in prognosis has been reached using clinical variables alone. Thus, biomarkers or other prognostic factors that can be added to these criteria should be investigated.

of metastatic RCC in the modern era in which physicians and patients have several treatment options and are not limited to only immunotherapy. This work also expands on early experience, when access to targeted treatments was limited.

Limitations of this retrospective analysis include missing data. We used multiple imputation datasets to account for missing data and they yielded similar results. Additionally, the amount of missing data in all data elements was less than 5% and results were similar between the complete case analyses and the analyses using imputed data for missing values. Finally, comparisons between different first-line drugs are difficult to do because this dataset included patients mainly treated with sunitinib and sorafenib. However, we have shown that this model is not affected by first-line drug choice.<sup>7</sup>

Now externally validated, the International Metastatic RCC Database Consortium model can be applied to stratify patients by risk in clinical trials and to counsel patients about prognosis. This model might be better than others with respect to ease of use and stratification capability. Patient-specific and tumour-specific biomarkers that can predict response and prognosis should be investigated.

##### Contributors

DYCH, WX, and TKC designed the study and wrote the report. DYCH, WX, MM, and TKC analysed data. All authors collected data, interpreted data, and approved the final manuscript.

##### Conflicts of interest

DYCH has an advisory role at Aveo, Pfizer, Novartis, and Bayer. LCH has an advisory role at Novartis and Pfizer and has received research funding from Bristol-Myers Squibb, Novartis, and Genentech. GAB has a consultant and advisory role at Pfizer and has received honoraria and research funding from Pfizer. MM has an advisory role at Novartis and

Pfizer and has received research funding from both. LW has an advisory role at Pfizer and Novartis and has received research funding from Pfizer, Novartis, and GlaxoSmithKline. UNV has received honoraria and research funding from Pfizer, Novartis, and GlaxoSmithKline. S-YR has an advisory role at Novartis, Pfizer, and GlaxoSmithKline, and has received research funding from Novartis and Bayer Korea. FD has received research funding from Novartis. CK has an advisory role at Pfizer, Novartis, and GlaxoSmithKline and has received honoraria and research funding from Pfizer, Novartis, and GlaxoSmithKline. BIR has an advisory role at Pfizer, GlaxoSmithKline, Aveo, Bayer, Onyx, and has received research funding from GlaxoSmithKline and Pfizer. TKC has received research funding from Pfizer and has an advisory role at Aveo, Pfizer, Novartis, GlaxoSmithKline, Genentech, Bayer, and Onyx. WX, MMR, M-HT, and NA declare that they have no conflicts of interest.

#### Acknowledgments

We thank individual private patient donors who helped fund this study.

#### References

- Motzer RJ, Hutson TE, Tomczak P, et al. Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. *N Engl J Med* 2007; **356**: 115–24.
- Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. *N Engl J Med* 2007; **356**: 125–34.
- Escudier B, Pluzanska A, Koralewski P, et al. Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet* 2007; **370**: 2103–11.
- Rini BI, Halabi S, Rosenberg JE, et al. Phase III trial of bevacizumab plus interferon alfa versus interferon alfa monotherapy in patients with metastatic renal cell carcinoma: final results of CALGB 90206. *J Clin Oncol* 2010; **28**: 2137–43.
- Sternberg CN, Szczylik C, Lee E, et al. A randomized, double-blind phase III study of pazopanib in treatment-naive and cytokine-pretreated patients with advanced renal cell carcinoma (RCC). *Proc Am Soc Clin Oncol* 2009; **27** (suppl): abstr 5021.
- Rini BI, Escudier B, Tomczak P, et al. Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *Lancet* 2011; **378**: 1931–39.
- Heng DY, Xie W, Regan MM, et al. 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* 2009; **27**: 5794–99.
- Heng DY, Xie W, Bjarnason GA, et al. A unified prognostic model for first- and second-line targeted therapy in metastatic renal cell carcinoma (mRCC): results from a large international study. *Proc Am Soc Clin Oncol* 2010; **28** (suppl): abstr 4523.
- Choueiri TK, Garcia JA, Elson P, et al. Clinical factors associated with outcome in patients with metastatic clear-cell renal cell carcinoma treated with vascular endothelial growth factor-targeted therapy. *Cancer* 2007; **110**: 543–50.
- Escudier BJ, Ravaud A, Negrier S, et al. Update on AVOREN trial in metastatic renal cell carcinoma (mRCC): efficacy and safety in subgroups of patients (pts) and pharmacokinetic (PK) analysis. *Proc Am Soc Clin Oncol* 2008; **26** (suppl): abstr 5025.
- Negrier S, Escudier B, Gomez F, et al. Prognostic factors of survival and rapid progression in 782 patients with metastatic renal carcinomas treated by cytokines: a report from the Groupe Francais d'Immunotherapie. *Ann Oncol* 2002; **13**: 1460–68.
- Manola J, Royston P, Elson P, et al. Prognostic model for survival in patients with metastatic renal cell carcinoma: results from the international kidney cancer working group. *Clin Cancer Res* 2011; **17**: 5443–50.
- Motzer RJ, Bacik J, Murphy BA, Russo P, Mazumdar M. Interferon-alfa as a comparative treatment for clinical trials of new therapies against advanced renal cell carcinoma. *J Clin Oncol* 2002; **20**: 289–96.
- Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004; **23**: 2109–23.
- Allison PD. Survival analysis using the SAS system: a practical guide. Cary: SAS Institute, 1995.
- Cook NR, Ridker PM. Advances in measuring the effect of individual predictors of cardiovascular risk: the role of reclassification measures. *Ann Intern Med* 2009; **150**: 795–802.
- Pencina MJ, D'Agostino RBS, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; **27**: 157–72.
- Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; **338**: b2393.
- Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley, 1987.
- Heng DY, Choueiri TK, Lee JL, et al. An in-depth multicentered population-based analysis of outcomes of patients with metastatic renal cell carcinoma (mRCC) that do not meet eligibility criteria for clinical trials. *Proc Am Soc Clin Oncol* 2012; **30** (suppl): abstr 4536.
- Op den Winkel M, Nagel D, Sappl J, et al. Prognosis of patients with hepatocellular carcinoma. Validation and ranking of established staging-systems in a large western HCC-cohort. *PLoS One* 2012; **7**: e45066.
- Buyyounouski MK, Pickles T, Kestin LL, Allison R, Williams SG. Validating the interval to biochemical failure for the identification of potentially lethal prostate cancer. *J Clin Oncol* 2012; **30**: 1857–63.
- Tang PA, Vickers MM, Heng DY. Clinical and molecular prognostic factors in renal cell carcinoma: what we know so far. *Hematol Oncol Clin North Am* 2011; **25**: 871–91.
- van der Velde AA, Eechoute K, Gelderblom H, et al. Genetic polymorphisms associated with a prolonged progression-free survival in patients with metastatic renal cell cancer treated with sunitinib. *Clin Cancer Res* 2011; **17**: 620–29.