World J Urol
DOI 10.1007/s00345-012-0896-z

ORIGINAL ARTICLE

External validation of the preoperative Karakiewicz nomogram in a large multicentre series of patients with renal cell carcinoma

Paolo Gontero · Maxine Sun · Alessandro Antonelli · Roberto Bertini · Marco Carini · Giorgio Carmignani · Nicola Longo · Giuseppe Martorana · Andrea Minervini · Vincenzo Mirone · Giuseppe Morgia · Giacomo Novara · Marco Oderda · Claudio Simeone · Alchiede Simonato · Salvatore Siracusano · Alessandro Tizzani · Alessandro Volpe · Pierre Karakiewicz · Vincenzo Ficarra · Members of the SATURN Project–LUNA Foundation

Received: 7 February 2012 / Accepted: 12 June 2012

Abstract
Purpose To perform a formal external validation of the preoperative Karakiewicz nomogram (KN) for the prediction of cancer-specific survival (CSS) using a large series of surgically treated patients diagnosed with organ-confined or metastatic renal cell carcinoma (RCC).
Methods Patient population originated from a series of retrospectively gathered cases that underwent radical or partial nephrectomy between years 1995 and 2007 for suspicion of kidney cancer. The original Cox coefficients were used to generate the predicted risk of CSS at 1, 2, 5, and 10 years following surgery and compared to the observed risk of CSS in the current population. External validation was quantified using measures of predictive accuracy, defined as model discrimination and calibration.
Results A total of 3,374 patients were identified. Relative to the original development cohort, the current sample population had a larger proportion of patients with localized (40.0 vs. 26.3 %, $P < 0.001$) and non-metastatic (92.2 vs. 88.1 %, $P = 0.03$) disease at presentation. Model discrimination for the prediction of CSS was 87.8 % (95 % CI, 84.4–91.4) at 1 year, 87.0 % (95 % CI, 84.4–89.5) at 2 years, 84.7 % (95 % CI, 82.3–87.1) at 5 years, and 85.9 % (95 % CI, 83.2–88.6) at 10 years. The relationship between predicted and observed CSS risk was adequate in the calibration plot.
Conclusion The use of the KN for the prediction of CSS in patients diagnosed with renal cell carcinoma was validated in the current study. In consequence, this tool may be recommended for routine clinical counseling in patients with various stages of RCC in the preoperative setting.

Keywords Preoperative nomogram · Renal cell carcinoma · Prognostic Factors · Partial · Nephrectomy · Radical nephrectomy · Metastatic renal cell carcinoma

P. Gontero (✉) · M. Oderda · A. Tizzani
Department of Urology, Molinette Hospital, University of Turin, A.O.U. San Giovanni Battista Molinette, C.so Bramante 88/90, 10126 Turin, Italy
e-mail: paolo.gontero@unito.it

M. Sun · P. Karakiewicz
University of Montreal Health Center, Montreal, QC, Canada

A. Antonelli · C. Simeone
University of Brescia, Brescia, Italy

R. Bertini
Vita-Salute University SanRaffaele, Milan, Italy

M. Carini · A. Minervini
University of Florence, Florence, Italy

G. Carmignani · A. Simonato
University of Genova, Genoa, Italy

N. Longo · V. Mirone
University Federico II, Naples, Italy

G. Martorana
University of Bologna, Bologna, Italy

G. Morgia
University of Catania, Catania, Italy

G. Novara · V. Ficarra
University of Padua, Padua, Italy

S. Siracusano
University of Trieste, Trieste, Italy

A. Volpe
University of Eastern Piedmont, Novara, Italy

Published online: 31 July 2012

Introduction

In the last decade, treatment of renal cell carcinoma (RCC) has significantly improved, and various systems integrating clinical and pathologic prognostic factors have been proposed to improve the prognostication of patients with confined or locally advanced RCC. Available prognostic tools that use pathological variables are considered valuable assets in the post-operative counseling and planning of patients' follow-up schedule following surgical resection of the primary tumor. Moreover, some of these tools may be used in the selection criteria and interpretation of results of ongoing randomized controlled trials evaluating the effect of novel targeted therapies in the adjuvant setting [1].

However, the benefit of these tools is not applicable in the preoperative setting of patients with organ-confined or advanced RCC who are considered suitable surgical candidates. To date, few prognostic models relying on clinical variables exist. Two models that combine clinical tumor size and symptoms showed a model discrimination of less than 70 % in predicting cancer-specific survival (CSS) in patients with non-metastatic RCC [2, 3]. In patients with advanced RCC, patient selection among existing targeted therapies is predominantly based on the risk-group stratification proposed by Motzer et al. [4], which relies on performance status and other laboratory parameters. However, the model's predictive accuracy remains unknown.

Recently, a novel nomogram for the prediction of CSS in the preoperative setting has been proposed by Karakiewicz et al. [5]. The model relied on patient age, gender, mode of presentation, clinical tumor size, clinical stage of the primary tumor, and presence of distant metastases and showed optimal model discrimination. Since its first external validation performed on a cohort of 1,972 patients [6], only one study further tested the predictive accuracy of this nomogram [7], showing its superiority compared with the other tested nomograms and risk groups in predicting survival outcomes in 390 patients with localized RCC.

The aim of the present multicenter study was to perform an external validation of the preoperative Karakiewicz nomogram (KN) for the prediction of CSS using a large series of patients surgically treated for confined or metastatic RCC.

Materials and methods

Sixteen academic institutions participated to the Surveillance And Treatment Update Renal Neoplasms (SATURN) project, promoted by the Leading Urological No-profit foundation Advanced research (LUNA) of the Italian Society of Urology (SIU) in 2008. A computerized databank was generated for data transfer. The initial database comprised 5,463 patients who underwent radical nephrectomy (RN) or partial nephrectomy (PN) between 1995 and 2007 for suspicion of kidney cancer. Excluded patients were those in whom symptom classification ($n = 310$) or clinical tumor size ($n = 1.789$) was missing. These exclusion criteria resulted in 3,364 assessable patients.

The mode of presentation was defined according to the Patard classification [8]. Clinical staging of the primary tumor included a minimum of abdominal computed tomography (CT) scans and chest X-rays and defined according to the 2002 version of the American Joint Committee on Cancer–Union Internationale Contre le Cancer TNM classification [9]. Clinical tumor size, clinical stage of primary tumor, and presence of distant metastasis were assessed according to preoperative CT scans. Bone scans and brain CT scans were obtained only when indicated by signs and symptoms.

Surgery was performed according to the standard criteria for radical nephrectomy, that is, extrafascial dissection of the kidney. The hilar and regional lymph nodes adjacent to the ipsilateral renal pedicle were removed along with enlarged lymph nodes if abnormal on preoperative CT scans or palpable intraoperatively. Extended lymphadenectomy was routinely performed only in few centers. In patients with contralateral normal kidney, elective nephron-sparing surgery had been routinely indicated in the presence of single, peripheral tumors ≤ 4 in size, although some referral centers perform elective nephron-sparing surgery also in case of larger tumors. Imperative nephron-sparing surgery had been performed in patients with bilateral tumors or with neoplasia involving anatomically or functionally solitary kidneys. Nephron-sparing surgery was performed in the form of enucleoresection, simple enucleation, or polar nephrectomy according to the clinical indications and surgeon's preference.

Patients were generally observed every 3–4 months for the first year after surgery, every 6 months from the second through the fifth years, and annually thereafter. Follow-up consisted of a history, physical examination, routine blood work and serum chemistry studies, chest radiography, and radiographic evaluation of the contralateral or remnant kidney. Elective bone scan, chest computed tomography, and magnetic resonance imaging were performed when clinically indicated.

The cause of death was determined by the treating physicians, by chart review corroborated by death certificates, or by death certificates alone. Most patients who were identified as having died of kidney cancer had progressive, widely disseminated metastases at the time of death.

Statistical analyses

Baseline descriptives were reported using median and interquartile range or using mean ± standard deviation for

Table 1 Descriptive characteristics of 3,364 patients from the external validation cohort and 2,474 patients from the previous development cohort

Predictors	External validation cohort ($n = 3,364$)	Development cohort ($n = 2,474$) [5]
Age (year)		
Mean (median)	61.6 (63.0)	60.7 (62.0)
Range	7–92	10–91
Symptoms		
Asymptomatic	2,293 (68.2)	1,142 (46.2)
Local	946 (28.1)	879 (35.5)
Systemic	125 (3.7)	453 (18.3)
Gender		
Male	2,207 (65.6)	1,648 (66.6)
Female	1,157 (34.4)	826 (33.4)
Tumor size (cm)		
Mean (median)	5.6 (5.0)	6.6 (6.0)
Range	1–22	0.5–25
T stage		
T1a	1,354 (40.0)	650 (26.3)
T1b	1,108 (32.9)	565 (22.8)
T2	585 (17.4)	359 (14.5)
T3	326 (9.7)	900 (36.4)
Metastases		
M0	3,103 (92.2)	2,179 (88.1)
M1	261 (7.8)	295 (11.9)

continuously coded variables, and frequencies and proportions for categorically coded variables. Kaplan–Meier plots graphically explored CSS rates. The original Cox regression coefficients were used to generate the predicted risk when relying on the Karakiewicz nomogram and compared with the observed risk of mortality at 1, 2, 5, and 10 years in the current sample population. Model discrimination represents an unbiased measure to discriminate among patients, which was quantified using the area under the receiver operating characteristic curve [10]. A value of 100 % indicates perfect predictions, whereas 50 % is equivalent to a toss of a coin. With censored data, the calculation [11] is slightly modified. However, its interpretation remains the same. It represents the probability that, for a randomly selected pair of patients, the model is capable to discriminate who had a higher risk of the event, hereby death. This methodology was previously used in established publications [12]. The 95 % confidence intervals were computed using 200 bootstrap resampling. Finally, the relationship between predicted and observed rates was assessed using methods of calibration and graphically explored using the *val.surv* function.

All reported P values were two-sided with a significance level set at $P < 0.05$. Statistical analyses were performed with SPSS vers. 16.0 (SPSS Inc, Chicago, IL, USA), S-Plus Professional software (MathSoft, Inc., Washington, USA), and SAS (version 9.0, North Carolina, USA).

Results

The descriptive characteristics of the 3,364 analyzed patients are listed in Table 1. Median age was 63 years (interquartile range, 54–71). Notably, only 261 patients (7.8 %) had synchronous distant metastases at diagnosis. In total, 2,354 (70 %) patients underwent radical nephrectomy, and 1,010 (30 %) patients underwent nephron-sparing surgery. Relative to the original development cohort ($n = 2,474$), the current population had more asymptomatic patients (68.2 vs. 48.2 %, $P < 0.001$). Moreover, most patients were T1a (40.0 vs. 26.3 %, $P < 0.001$) and non-metastatic (92.1 vs. 88.1 %, $P = 0.03$) in the current series. At median follow-up of 48 months (IQR, 26–85), 2,530 patients were alive and disease-free.

The 1-, 2-, 5-, and 10-year CSS probabilities were 96.5 % (95 % confidence interval [CI], 95.8–97.1), 92.1 % (95 % CI, 91.1–93.0), 84.8 % (95 % CI, 83.4–86.2), and 79.4 % (95 % CI, 77.3–81.6), respectively (Fig. 1a). Table 2 shows the results of univariable and multivariable Cox regression analyses for the prediction of CSS. Notably, all variables included in the original preoperative KN emerged as independent predictors of CSS .

Model discrimination for the prediction of CSS was 87.8 % (95 % CI, 84.4–91.4) at 1 year, 87.0 % (95 % CI 84.4–89.5) at 2 years, 84.7 % (95 % CI 82.3–87.1) at 5 years, and 85.9 % (95 % CI 83.2–88.6) at 10 years. Overall, the model was well calibrated, although it demonstrated an overall tendency to overestimate the risk of CSS between 2 and 15 % (Fig. 1b).

Discussion

The current study is a formal external validation of the preoperative KN performed on an independent European multicentre series of patients diagnosed with confined or synchronous metastatic RCC. The model discrimination observed in our cohort is similar to the one originally reported in the external validation cohort of 1,972 patients used by Karakiewicz et al. [5]. Specifically, the 5- and 10-year model discrimination accuracies were 86.8 and 84.2 % in the initial external validation and 84.7 and 85.9 % in the current study, respectively.

Several important considerations should be mentioned in regard to the characteristics of our cohort in comparison with that of the one previously used by Karakiewicz et al. to develop the nomogram ($n = 2,474$), as well as the one

used to perform the external validation ($n = 1{,}972$) in the original study. Specifically, prevalence of symptoms, mean clinical tumor size, and clinical stage were more unfavorable in the original development cohort in comparison with both external validations cohorts (Table 1). The 5- and 10-year CSS rates were 84.8 versus 75.4 % and 79.4 versus 68.3 % in the current cohort relative to the original development cohort, respectively. This may represent a potential explanation for the model's overall tendency to overestimate the predictions in the current external validation.

In order to tailor treatment based on individual patients' needs, it is essential for physicians to have the ability to predict the biological aggressiveness of these neoplasms and to stratify patients into risk groups for risk of recurrence, progression, and death. Defining risk groups is very useful for routine clinical patient counseling, selection of treatment options, scheduling of follow-up, and in selection of patients for participation in clinical trials. The most useful models are probably those able to assist physicians in treatment choices before therapy is determined, particularly in light of the new therapies currently become available or being tested for RCC.

Our results strongly support the use of preoperative KN for the prediction of CSS probabilities both in patients with organ-confined or advanced RCC. In a similar setting, Yaycioglu et al. [2] and Cindolo et al. [3] proposed two models for the prediction of recurrence in patients with non-metastatic RCC. In the former, authors identified patient presentation (symptomatic vs. incidental), as well as clinical size as important prognostic predictors of treatment failure. In the latter, authors proposed the use of clinical presentation, clinical size, TNM, and cellular grade for the prediction of recurrence. Unfortunately, given the lack of information on several variables that were integrated in these models within the current database, we could not externally validate such models. Nonetheless, previous external validation data from a multi-institutional study revealed a 5-year discrimination accuracy of 62 % for the Yaycioglu model and 64 % for the Cindolo score [13].

In the context of patients with synchronous distant metastases, the most commonly used tool to predict overall survival is the Memorial Sloan–Kettering Cancer Center (MSKCC) classification, which includes Karnofsky performance status, lactate dehydrogenase, hemoglobin, corrected calcium, and interval from the initial diagnosis to therapy. According to the Motzer criteria, patients with advanced RCC can be subdivided in low-(0 risk factors), intermediate-(1–2 risk factors), and high-risk (3–5 risk factors) categories [4]. Although this classification has been widely used, it was never subjected to measures of predictive accuracy. Moreover, its applicability in the targeted therapy era is limited [14]. While the Motzer criteria may be appropriate for risk-group stratification, they might not be adequate in the estimation of CSS in individual cases. The preoperative KN could represent a useful tool in this setting.

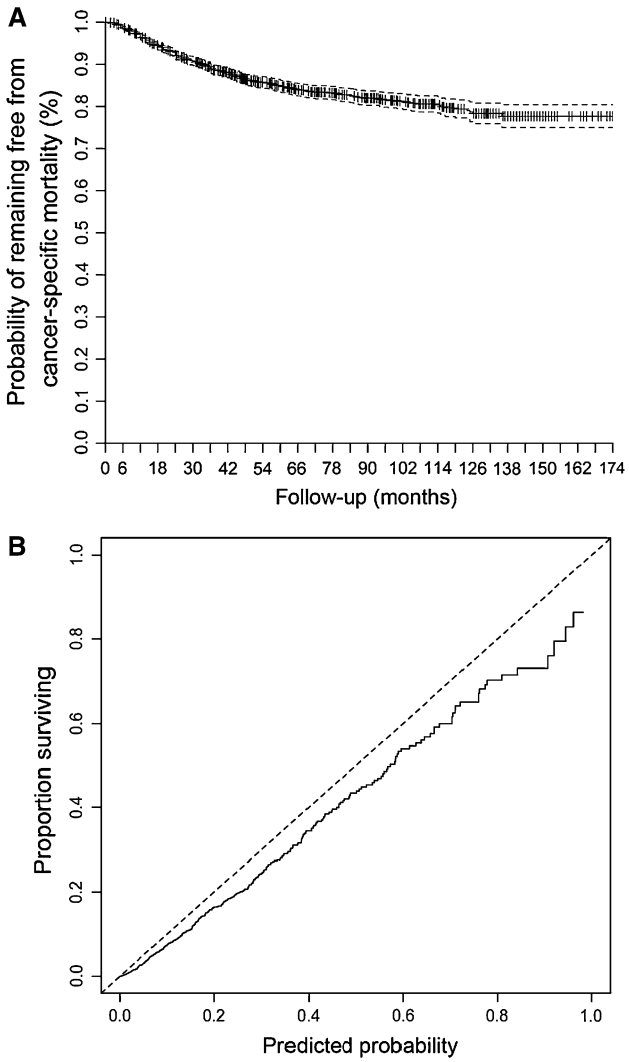

Fig. 1 **a** Cancer-specific survival probabilities in patients included in the analysis. **b** Calibration plot depicting the relationship between the nomogram-predicted CSS probabilities (*x*-axis) and the actual fraction surviving (*y*-axis) within the current external validation cohort

In patients with advanced disease, an attractive future perspective could be the improvement of prognostic accuracy of preoperative KN adding traditional histopathological, molecular, or cytogenetic information coming from the renal tumor biopsies of the primary tumor. Unfortunately, percutaneous renal tumor biopsy has still a limited role in the clinical workup of patients with renal parenchymal tumors, although improvements in the technique significantly reduced the risk of tumor spread and complications.

As KN relies on the old TNM staging 6th edition (2002), the nomogram could be potentially affected by the new TNM classification 7th edition (2010) [15]. From the 6th to

Table 2 Univariable and multivariable Cox regression analyses for the prediction of cancer-specific mortality

Predictors	Univariable		Multivariable	
	Hazard ratio (95 % CI)	*P*	Hazard ratio (95 % CI)	*P*
Age (years)	1.02 (1.01–1.03)	<0.001	1.02 (1.01–1.03)	<0.001
Gender				
Female	1.0 (ref.)	–	1.0 (ref.)	–
Male	1.15 (0.94–1.41)	0.2	1.27 (1.04–1.55)	0.02
Clinical stage				
T1a	1.0 (ref.)	–	1.0 (ref.)	–
T1b	2.12 (1.53–2.94)	<0.001	1.47 (1.04–2.06)	0.03
T2	6.51 (4.80–8.84)	<0.001	2.43 (1.62–3.63)	<0.001
T3	15.61 (11.47–21.25)	<0.001	4.51 (3.04–6.70)	<0.001
Metastases				
M0	1.0 (ref.)	–	1.0 (ref.)	–
M1	12.32 (10.15–14.96)	<0.001	6.39 (5.18–7.89)	<0.001
Tumor size (cm)	1.24 (1.21–1.26)	<0.001	1.08 (1.04–1.12)	<0.001
Symptom classification				
Asymptomatic	1.0 (ref.)	–	1.0 (ref.)	–
Local	2.79 (2.29–3.41)	<0.001	1.70 (1.38–2.09)	<0.001
Systemic	7.48 (5.57–10.05)	<0.001	2.48 (1.81–3.38)	<0.001

the 7th edition, however, the only modifications were to subclassify T2 tumors into T2a and T2b and to redefine the distinction between T3a and T3b, and KN does not consider T2 and T3 subclassifications.

Despite the strengths of the current report, limitations do apply. First, the retrospective nature of the study may have induced bias above all in the correct evaluation of clinical stage of primary tumor. Second, the current validation cohort originated from several institutions in Italy. Third, the exclusion of missing data reduced the number of patients that were assessed. Results may have differed if missing data were handled differently.

Conclusions

The preoperative KN uses the combination of common clinical information such as age, gender, symptoms, clinical tumor size and stage, and the presence of distant metastases to accurately predict the 1-, 2-, 5-, and 10-year CSS of patients with renal cell carcinoma. The current study originating from a multi-institutional series confirmed the benefit of this nomogram using measures of predictive accuracy, defined as model discrimination and calibration.

In consequence, the nomogram is considered a valuable tool for physicians and patients in the preoperative setting. Its predictions may be used to provide a framework for comparisons between nephrectomy and alternative treatment modalities for all stages of RCC.

Conflict of interest None declared.

References

1. Ficarra V, Brunelli M, Cheng L et al (2010) Prognostic and therapeutic impact of the histopathologic definition of parenchymal epithelial renal tumors. Eur Urol 58:655–668
2. Yaycioglu O, Roberts WW, Chan T et al (2001) Prognostic assessment of nonmetastatic renal cell carcinoma: a clinically based model. Urology 58:141–145
3. Cindolo L, De la Taille A, Messina G et al (2003) A preoperative clinical prognostic model for non-metastatic renal cell carcinoma. BJU Int 92:901–905
4. Motzer RJ, Bacik J, Murphy BA et al (2002) Interferon-alfa as a comparative treatment for clinical trials of new therapies against advanced renal cell carcinoma. J Clin Oncol 20:289–296
5. Karakiewicz PI, Suardi N, Capitanio U et al (2009) A preoperative prognostic model for patients treated with nephrectomy for renal cell carcinoma. Eur Urol 55:287–295
6. Karakiewicz PI, Briganti A, Chun FKH et al (2007) Multi-institutional validation of a new renal cancer-specific survival nomogram. J Clin Oncol 25:1316–1322
7. Tan MH, Li H, Choong CV et al (2011) The Karakiewicz nomogram is the most useful clinical predictor for survival outcomes in patients with localized renal cell carcinoma. Cancer 117:5314–5324
8. Patard JJ, Leray E, Cindolo L et al (2004) Multi-institutional validation of a symptom based classification for renal cell carcinoma. J Urol 172:858–862
9. Greene FL (2002) The American Joint Committee on Cancer: updating the strategies in cancer staging. Bull Am Coll Surg 87(7):13–15
10. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36

11. Harrell FE Jr, Califf RM, Pryor DB et al (1982) Evaluating the yield of medical tests. JAMA 247:2543
12. Kattan MW, Eastham JA, Stapleton AM et al (1998) A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. J Natl Canc Inst 90(10):766–771
13. Cindolo L, Patard JJ, Chiodini P et al (2005) Comparison of predictive accuracy of four prognostic models for nonmetastatic renal cell carcinoma after nephrectomy. Cancer 104:1362–1371
14. Karakiewicz PI, Sun M, Bellmunt J et al (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
15. Eggener S (2010) TNM staging for renal cell carcinoma: time for a new method. Eur Urol 58:517–521