European Association of Urology

Brief Correspondence

External Validation of the ASSURE Model for Predicting Oncological Outcomes After Resection of High-risk Renal Cell Carcinoma (RESCUE Study: UroCCR 88)

Zine-Eddine Khene [a,], Alessandro Larcher [b], Jean-Christophe Bernhard [c], Nicolas Doumerc [d], Idir Ouzaid [e], Umberto Capitanio [b], François-Xavier Nouhaud [f], Romain Boissier [g], Nathalie Rioux-Leclercq [h], Alexandre De La Taille [i], Philippe Barthelemy [j], Francesco Montorsi [b], Morgan Rouprêt [k], Pierre Bigot [l], Karim Bensalah [a] on behalf of The French Committee of Urologic Oncology*

Article info

Abstract

A prognostic model based on the population of the ASSURE phase 3 trial has recently been described. The ASSURE model stratifies patients into risk groups to predict survival after surgical resection of intermediate- and high-risk localised kidney cancer. We evaluated this model in an independent cohort of 1372 patients using discrimination, calibration, and decision curve analysis. Regarding disease-free survival, the ASSURE model showed modest discrimination (65%), miscalibration, and poor net benefit compared with the UCLA Integrated Staging System (UISS) and Leibovich 2018 models. Similarly, the ability of the ASSURE model to predict overall survival was poor in terms of discrimination (63%), with overestimation on calibration plots and a modest net benefit for the probability threshold of between 10% and 40%. Overall, our results show that the performance of the ASSURE model was less optimistic than expected, and not associated with a clear improvement in patient selection and clinical usefulness in comparison to with available models. We propose an updated version using the recalibration method, which leads to a (slight) improvement in performance but should be validated in another external population.

Patient summary: The recent ASSURE model evaluates survival after surgery for nonmetastatic kidney cancer. We found no clear improvement in patient classification when we compared ASSURE with older models, so use of this model for patients with nonmetastatic kidney cancer still needs to be clarified.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of European Association of Urology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Renal cell carcinoma (RCC) is the most common type of kidney cancer, accounting for more than 150 000 deaths each year [1]. Some 80% of patients newly diagnosed with RCC have nonmetastatic disease and can potentially be cured by surgery. However, 20–30% of these patients will experience recurrence [2,3]. Predictive tools can help to identify patients at risk of progression [4], but there is a lack of consensus regarding risk stratification: almost all of the models were developed using historical cohorts from single institutions without external validation or assessment of calibration or clinical benefit.

Correa and colleagues [5] recently used data from the ASSURE randomised trial to build a prognostic model to predict oncological outcomes (disease-free survival [DFS], overall survival [OS], and early disease progression [EDP]) in patients with nonmetastatic high-risk RCC (defined as pT1b and grade 3–4; pT2/pT3/pT4; N1). The ASSURE model includes seven parameters (vascular invasion, tumour histology, tumour size, tumour grade, presence of tumour necrosis, presence of nodal disease, and age) and stratifies patients according to their risk. There are three prognostic groups (high, intermediate, and low risk) for estimation of DFS and four for prediction of OS (high, intermediate I, intermediate II, and low risk).

Our objective was to validate the ASSURE model in a large, independent, multi-institutional cohort of patients treated in a "real life" setting outside of a clinical trial.

After institutional review board approval, data for patients from ten institutions who underwent radical or partial nephrectomy for nonmetastatic RCC between 2013 and 2019 were analysed. We exclusively focused on patients with pT1b and International Society of Urological Pathology (ISUP) grade 3–4, pT2, pT3, pT4 and N1 disease (ie, those classified as intermediate and high risk according to the ASSURE model). We excluded patients who received adjuvant or neoadjuvant treatment ($n = 19$). All surgical specimens were processed according to standard pathological procedures. Vascular invasion was defined as the presence of any invasion of the vena cava or renal vein and its segmental branches, whether identified grossly or microscopically. Microvascular invasion was defined as invasion by neoplastic cells or tumour emboli in microscopically visible intratumoural vascular vessels or lymphatic vessels.

We plotted survival curves using Kaplan-Meier estimates for a direct visual comparison between groups. Then we used the regression coefficients reported by Correa et al [5] to calculate the individual risk of DFS and OS in our cohort according to the log-normal survival function formula $S(t) = 1 - \Phi\left(\frac{\log(t) - \mu}{\sigma}\right)$.

The performance of the ASSURE model was evaluated in terms of discrimination and calibration. Discrimination (the ability to discern patients who will experience an event from those who will not) was quantified using the concordance index (C-index) of Uno for specific time points and of Harrell for global time assessment. Calibration (the ability to make predictions as close as possible to the occurrence of real events) was assessed using the Brier score and calibration plots. The Brier score is the mean of the square of the difference between predicted and observed probabilities: a score of zero indicates perfect calibration and one is the worst possible calibration. Decision curve analysis (DCA) was used to determine whether the clinical value of the new model increased the net benefit over a realistic range of threshold probabilities. Finally, DCA results for the ASSURE model were compared with those for the older UCLA Integrated Staging System (UISS) [6] and the Leibovich 2018 score [7] (Supplementary Table 1). Missing data were imputed five times, with predictive mean matching for numeric variables and logistic regression for binary variables. Statistical significance was defined as a two-sided p value of <0.05. All statistical analyses were performed with Stata v15.1 (StataCorp, College Station, TX, USA).

A total of 1372 patients were included in the analysis. Supplementary Table 2 shows the characteristics of the population. The median age was 64 yr (interquartile range [IQR] 55–72). The median tumour size was 6.5 cm (IQR 5–9). The majority of tumours were clear cell RCC (77%) of high ISUP grade (72% grade 3 and 4), with 45% exhibiting tumour necrosis and 5% harbouring nodal involvement. The median follow-up was 54 mo (IQR 32–75).

Figure 1 shows Kaplan-Meier curves for DFS and OS by risk group in the ASSURE model. The DFS curves are clearly separated, dividing the cohort into three distinct risk groups. However, the difference is less clear for OS; in particular, the curves for intermediate I and intermediate II risk do not clearly separate during the first 40 mo.

Regarding discrimination, the ASSURE model achieved a C-index of 65% (95% CI 61–72%) for DFS and 63% (95% CI 60–69%) for OS prediction, which is quite modest (Fig. 2A).

The calibration curve presented in the Supplementary material shows that (1) the probability of dying in the validation cohort is constantly overestimated (Fig. 1B) and (2) the probability of relapse in the validation cohort is overestimated in early follow-up (<1 yr) and under-estimated after 18 mo of follow-up (Fig. 1A).

We used the integrated Brier score to summarise both discrimination and calibration of the ASSURE model over an interval of 72 mo. We observed an integrated Brier score of 0.23 and 0.29 for DFS and OS, respectively (Fig. 2B), which is quite modest.

Lastly, DCA demonstrated only a slight improvement with the ASSURE model over UISS and Leibovich scores for DFS and OS predictions: specifically, the ASSURE model showed a small net benefit in the 5–25% range of threshold event probabilities for DFS and in the 10–40% range for OS (Fig. 2C and 2 D).

To mitigate the miscalibration of the ASSURE model and obtain a better fit to our population (patients treated in "real life" outside of a clinical trial), we updated the ASSURE model by making new estimates of both the scale and the mean (Supplementary Table 3). This approach is common and referred to as recalibration in the literature. We found that the updated model had better calibration (Supplementary Fig. 2) and a better net benefit (Supplementary Fig. 3).

The ASSURE model is the most recently reported prognostic model evaluating survival after resection of localised RCC. It was designed by an expert team using
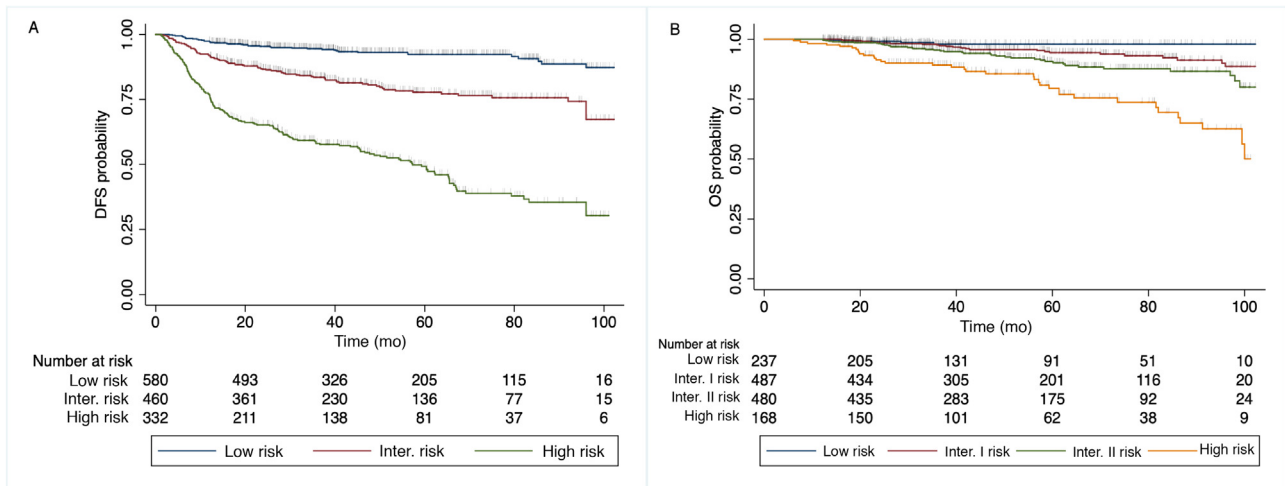
Fig. 1 – Kaplan-Meier curves showing (A) disease-free survival (DFS) and (B) overall survival (OS) for patients after surgical resection of renal cell carcinoma, stratified by prognostic risk group.
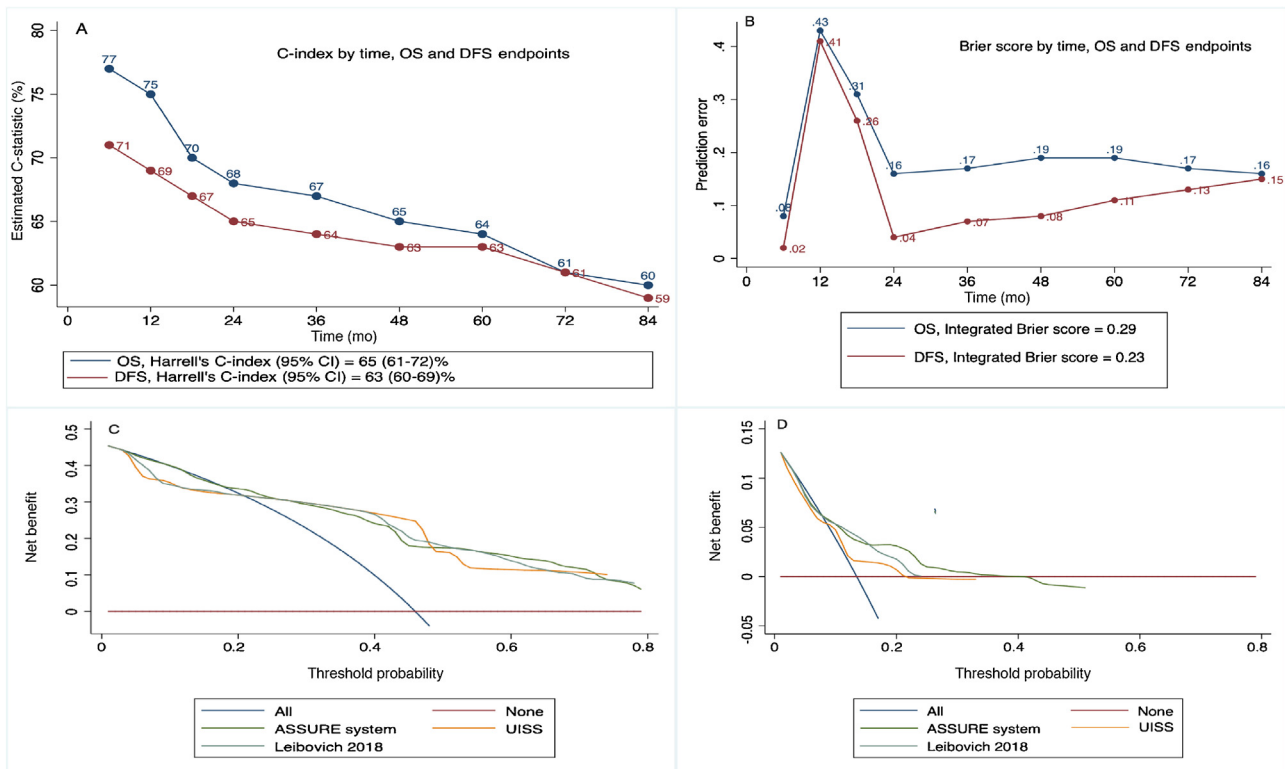

Fig. 2 – (A) Model discrimination and (B) prediction error curves showing the Brier score for each time point. Decision curve analysis showing the net benefit associated with use of the ASSURE model, UISS model, and Leibovitch 2018 model [7] at 60 mo for prediction of (C) disease-free survival (DFS) and (D) overall survival (OS). CI = confidence interval; UISS = UCLA Integrated Staging System.

robust statistics. The authors described their model as superior to classical ones but admitted a lack of discrimination over time.

Our analysis shows that the ASSURE model can indeed classify patients into different prognostic groups. However, we observed poor performance with modest discrimination ability (C-index of 0.65 and 0.63 in our cohort, vs 0.68 and 0.69 in the internal validation of the model), miscalibration,

and, more importantly, a minimal net benefit over the classical UISS and Leibovich models.

Finding lower performance in external validation studies is not surprising. There are some hypotheses that can explain our results. (1) The ASSURE cohort is older than ours (2006–2010 vs 2013–2019) and surgical management of renal tumours has evolved over the last decade, with increased use of partial nephrectomy (vs radical nephrec-

tomy) to treat larger and more complex renal masses [8], which might have had an impact on survival. (2) The larger proportion of pT1b tumours in our cohort (25% vs 11%) may have contributed to the system miscalibration [9]; ISUP grade and microvascular invasion (two parameters included in the ASSURE model) are strong predictors of kidney cancer–specific mortality, but they may not have the same impact when applied to a population with lower-stage tumours. (3) There is well-known variability for stage and grade interpretation among pathologists that might have introduced some bias [10].

The strengths of our analysis are the sample size and the rigorous approach. We included patients who were treated off protocol in academic centres. Patient management was carried out according to modern surgical and medical practices. Regarding the external validation, our study was in accordance with the method described by Royston and Altman [11] for survival prediction models. The retrospective analysis introduced a potential selection bias. Although all centres adhered to the guidelines and terminologies used in current practice, the absence of central review may have led to heterogeneity in imaging reporting and pathological analysis because of the involvement of multiple physicians with various ranges of expertise.

Finally, randomised clinical trials provide a tempting data source for the development of prognostic models. However, we should be aware of some drawbacks, which include selective inclusion of centres, selective eligibility and enrolment, predictor measurement, and the protocol effect [12]. For these reasons we believe that real live data are a valuable addition to the knowledge gained from clinical trial data.

In conclusion, the ability of the ASSURE model to predict oncological outcomes following surgical resection of RCC is comparable to that of classical prognostic systems. To increase the performance of prognostic systems, integration of novel biological and/or imaging biomarkers might be required.

Author contributions: Zine-Eddine Khene had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Khene, Larcher, Montorsi, Bensalah.
Acquisition of data: Khene, Larcher, Bernhard, Doumerc, Ouzaid, Capitanio, Nouhaud, Boissier, Rioux-Leclercq, De La Taille, Barthelemy, Rouprêt, Bigot, Montorsi, Bensalah.
Analysis and interpretation of data: Khene, Larcher, Capitanio, Montorsi, Bensalah.
Drafting of the manuscript: Khene, Larcher.
Critical revision of the manuscript for important intellectual content: Khene, Larcher, Bernhard, Doumerc, Ouzaid, Capitanio, Nouhaud, Boissier, Rioux-Leclercq, De La Taille, Barthelemy, Rouprêt, Bigot, Montorsi, Bensalah.
Statistical analysis: Khene, Bensalah.
Obtaining funding: None.
Administrative, technical, or material support: None.
Supervision: Montorsi, Bensalah.
Other: None.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.euros.2021.09.004.

References

[1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49. http://dx.doi.org/10.3322/caac.21660.

[2] Dabestani S, Thorstenson A, Lindblad P, Harmenberg U, Ljungberg B, Lundstam S. Renal cell carcinoma recurrences and metastases in primary non-metastatic patients: a population-based study. World J Urol 2016;34:1081–6. http://dx.doi.org/10.1007/s00345-016-1773-y.

[3] Dabestani S, Marconi L, Kuusk T, Bex A. Follow-up after curative treatment of localised renal cell carcinoma. World J Urol 2018;36:1953–9. http://dx.doi.org/10.1007/s00345-018-2338-z.

[4] Graham J, Dudani S, Heng DYC. Prognostication in kidney cancer: recent advances and future directions. J Clin Oncol 2018;36:3567–73. http://dx.doi.org/10.1200/JCO.2018.79.0147.

[5] Correa AF, Jegede OA, Haas NB, et al. Predicting disease recurrence, early progression, and overall survival following surgical resection for high-risk localized and locally advanced renal cell carcinoma. Eur Urol 2021;80:20–31. http://dx.doi.org/10.1016/j.eururo.2021.02.025.

[6] Zisman A, Pantuck AJ, Wieder J, et al. Risk group assessment and clinical outcome algorithm to predict the natural history of patients with surgically resected renal cell carcinoma. J Clin Oncol 2002;20:4559–66. http://dx.doi.org/10.1200/JCO.2002.05.111.

[7] Leibovich BC, Lohse CM, Cheville JC, et al. Predicting oncologic outcomes in renal cell carcinoma after surgery. Eur Urol 2018;73:772–80. http://dx.doi.org/10.1016/j.eururo.2018.01.005.

[8] Maurice MJ, Zhu H, Kim SP, Abouassaly R. Increased use of partial nephrectomy to treat high-risk disease. BJU Int 2016;117:E75–86. http://dx.doi.org/10.1111/bju.13262.

[9] Mir MC, Derweesh I, Porpiglia F, Zargar H, Mottrie A, Autorino R. Partial nephrectomy versus radical nephrectomy for clinical T1b and T2 renal tumors: a systematic review and meta-analysis of comparative studies. Eur Urol 2017;71:606–17. http://dx.doi.org/10.1016/j.eururo.2016.08.060.

[10] Williamson SR, Rao P, Hes O, et al. Challenges in pathologic staging of renal cell carcinoma: a study of interobserver variability among

urologic pathologists. Am J Surg Pathol 2018;42:1253–61. http://dx.doi.org/10.1097/PAS.0000000000001087.

[11] Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;13:33. http://dx.doi.org/10.1186/1471-2288-13-33.

[12] Pajouheshnia R, Groenwold RHH, Peelen LM, Reitsma JB, Moons KGM. When and how to use data from randomised trials to develop or validate prognostic models. BMJ 2019;365:l2154. http://dx.doi.org/10.1136/bmj.l2154.

[a]Department of Urology, University of Rennes, Rennes, France
[b]Unit of Urology, Division of Experimental Oncology, Urological Research Institute, IRCCS San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy
[c]Department of Urology, University of Bordeaux, Bordeaux, France

[d]Department of Urology, University of Toulouse, Toulouse, France
[e]Department of Urology, Bichat Claude Bernard Hospital, Paris, France
[f]Department of Urology, University Hospital of Rouen, Rouen, France
[g]Department of Urology, Aix-Marseille University, Marseille, France
[h]Department of Pathology, University of Rennes, Rennes, France
[i]Department of Urology, Mondor Hospital, Créteil, France
[j]Department of Oncology, University of Strasbourg, Strasbourg, France
[k]Department of Urology, La Pitie Salpétrière Hospital, Paris, France
[l]Department of Urology, University of Angers, Angers, France

[*]Corresponding author. Department of Urology, Rennes University Hospital, 2 rue Henri Le Guilloux, 35033 Rennes Cedex, France.
E-mail address: zineddine.khene@gmail.com (Z.-E. Khene).