Reviewer's report

Title: The VENUSS prognostic model to predict disease recurrence following surgery for non-metastatic papillary renal cell carcinoma: development and evaluation using the ASSURE prospective clinical trial cohort

Version: 0 Date: 21 May 2019

Reviewer: Kay See Tan

Reviewer's report:

In this study, the authors utilized data from five institutions to develop a scoring system to predict disease recurrence among patients with non-metastatic papillary renal cell carcinoma following curative surgery. The resulting VENUSS score (0 to 11) and VENUSS groups (3 risk categories) were internally validated using bootstrap procedure, and externally validated using a subcohort of patients from a prospective clinical trial. The following comments will focus on the statistical aspects of the paper.

Major comments:

The terminologies used throughout the manuscript implies a formal development and validation of the final VENUSS model - However, the actual reported results deviate from the standard procedure one expects from conventional model development and validation steps. The paper repeatedly stated "validate a prognostic model" - if this were true, then the actual model coefficients should have been utilized in the validation step (as was done in the Leibovich 2018 paper) - not the scores/groups derived from the models.

If the purpose was to validate the VENUSS model and compare to other existing models (not risk groups), then the paper must reflect validation using model coefficients; otherwise, the authors should correct the terminologies and avoid overstating the superior performance of the VENUSS model. Perhaps the intention was to compare the performance of the definitions of the risk groups applied to the cohorts, then the paper should clearly state so, carefully explain the approach in the methods section, and apply appropriate summaries that are better suited to the comparison of categories (e.g., not calibration curves).
Comments:

1) Validation: Please explain how exactly the groups were used in the internal and external validation. Risk groups themselves do not correspond to patient-specific predicted probabilities at a specified timepoint- so how were predicted probabilities derived from the groups to generate C-index or calibration plots? Did the authors generate new models with the groups as the only covariate? Why not compare the models using the reported coefficients from VENUSS, UISS and Leibovich 2018?

2) Calibration plots: These calibration plots are not as expected. E.g., If there were only 3 VENUSS groups, one would expect a maximum of 3 points on the figures corresponding to 3 unique predicted probabilities, yet these plots depicted smoothed curves over continuous probabilities.
   a. Please clearly specify how the calibration plots were generated.
   b. Specify the timepoint (at 5-years etc.) used for the predicted probabilities.
   c. Please provide 95% confidence intervals
   d. Provide relevant references corresponding to the choice of calibration approach: e.g., PMID: 24668611 "Calibration plots for risk prediction models in the presence of competing risks."

3) Decision curve analysis: similar to #2, if there were 3 groups, one would not expect such a smooth net benefit curve - please specify the degree of smoothing used on the figures.

4) The title of the paper should include the outcome of interest (to predict disease recurrence).

5) Throughout the text, please use the correct terminology "cumulative incidence of recurrence" rather than "recurrence probabilities". They are not interchangeable.

6) Abstract: "Specifically, there are no established tools to risk-stratify patients with PRCC..." Please justify this claim given the papRCC risk stratification tool proposed by Leibovich 2018 (Table 3).

7) Please confirm that there were no missing data in the development cohort (i.e., no patient was excluded due to missing data). It would also be helpful to see a consort diagram (supplementary data if necessary) of the exclusions to arrive at the final 556 patients from 5 institutions.

8) The outcomes of interest in UISS (overall survival) and Leibovich (disease-free survival and overall survival) models are different from the VENUSS model (disease recurrence).
a. Please justify the applicability of UISS and Leibovich as comparison models given the tools were originally meant for a different endpoint.

b. Could the poorer performance by UISS and Leibovich in the validation step be attributable to the fact that they were developed for different endpoints?

9) (pg 6, line 10) Please provide the number of recurrences and deaths without recurrence in the ASSURE validation cohort.

10) Was median follow-up duration calculated using reverse Kaplan-Meier approach?

11) (pg 6, line 29) Please specify the date of censoring for those alive without recurrence (last follow-up? Last clinical contact?)

12) More details are required in the statistical methods section (supplemental if necessary):

   a. Please specify the statistical software utilized, along with any specialized packages used for each validation performance component and the multiple imputation procedure

   b. Include standard language about one-/two-sided tests, p<0.05 significance level etc.

   c. As validation in the context of competing risks models is not common, it would benefit the readers to provide references for calibration plots and the decision curve analysis.

   d. Provide details regarding the multiple imputation algorithm used (e.g., software, method of imputation - predictive mean matching).

   e. How exactly were the imputed datasets utilized in external validation? What was the procedure to derive the groups for external validation using the 5 imputed datasets - average of 5 c-indexes and calibration?

13) Table 2:

   a. What is the purpose of "Rank" column?

   b. Why did tumor necrosis and sarcomatoid features remain in the multivariable model if they were not significant in the multivariable model?

   c. Table 3 states that several categories were combined due to similar regression coefficient scores, then present the multivariable model with the combined categories - so that the final multivariable model coefficients reflect the scores derived.

   d. Please include the univariable and multivariable (if applicable) model estimates for fat invasion on Table 2.

14) (Pg 9, line 45) provide 95% confidence intervals around the c-indexes
15) (pg 11, line 37) "...more accurate risk stratification at trial baseline would improve...sample size estimation...". Please clarify how risk stratification can impact sample size estimation.

16) (pg 12, line 25) "...discrimination and calibration were worse than in the validation cohort..." I believe this should have said "development cohort".

17) External cohort: please provide the cumulative incidence of recurrence curves by VENUSS risk groups (similar to Figure 1C).

18) Table 4: Provide 95% confidence intervals

19) Figures of cumulative incidence of recurrence: please include confidence intervals.

20) Figure 1A:
   a. Please specify the competing risks definitions to generate this figure, e.g., for the cumulative incidence of recurrence to abdomen, were all other sites considered as individual competing risks?
   b. If a patient recurred to both abdomen and chest, would the patient be in both curves?
   c. Does "At risk" numbers apply to all 3 curves?

21) Figure 2: Please match the description of the legend to the actual figure (e.g., legend says "black and yellow solid lines" but figure only shows purple and blue lines).

22) Supplementary Figure 2: why would bootstrap resamples be necessary for the external validation?

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Not applicable

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No
Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

**Quality of written English**
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

**Statement on potential review bias**
Please complete a statement on potential review bias, considering the following questions:

1. Did you co-author any publication with an author of this manuscript in the last 5 years?

2. Are you currently or recently affiliated at the same institution as an author of this manuscript?
If you can answer no to all of the above, write 'I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria’. If your reply is yes to any, please give details below.

I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal