PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A systematic review of prognostic prediction models for acute kidney injury (AKI) in general hospital populations.
AUTHORS	Hodgson, Luke; Sarnowski, Alexander; Roderick, Paul; Dimitrov, Borislav; Venn, Richard; Forni, Lui

VERSION 1 - REVIEW

REVIEWER	Nick Selby Associate Professor University of Nottingham
REVIEW RETURNED	17-Mar-2017

GENERAL COMMENTS	Thank you for asking me to review the manuscript by Hodgson et al. This is a well written and well designed systematic review, and pulls together studies in an important area that urgently needs more evidence. I agree with the decision not to attempt meta-analysis for the reasons stated and due to heterogeneity. If the PRISMA guidelines were followed then this should be stated in the methods. In the methods, the authors acknowledge that the definitions for AKI may vary between studies. However, the definition for hospital acquired may also differ as noted on page 10. It may help to state in the methods the parameters defined during the search strategy to identify studies that reported hospital acquired AKI and the definitions employed by the different studies in this regard. Page 11. Candidate predictors. How many models included serum creatinine? I think this should be explicitly stated as models that include this are at risk of confusing prediction and detection of AKI events. In the first paragraph of the discussion, the most commonly used predictors are listed but haven't been clearly described in the results. I would add some more detail around this on page 11, and in table 1 if there is room, it would be nice to list the predictors included in each risk model (although that won't be possible for Matheny and Koyner). This would allow the reader to easily see the similarities and differences in these across the studies. On page 15, the authors mention that physiological parameters weren't often included as predictor variables. I think the discussion would benefit from expanding this argument a little more, getting at the point as to whether opportunities exist in future work to improve model performance by including variables that have not been well studied to date. This is important as in some cases, predictors are chosen for pragmatic reasons (i.e. what is possible to collect) rather

	than what may be ideal. Page 17. Discussion of potential for clinical implementation. One of the important results from this review is that despite the heterogeneity in setting and differences in risk score, the different studies generally report AUC values in the same range - reasonable but none >0.8. This moderate performance of risk scores does feed into the discussions as how to think about implementation and may make some future directions more attractive than others. This argument could be brought out more this section of the discussion.

REVIEWER	Azra Bihorac University of Florida Gainesville FL
REVIEW RETURNED	19-Mar-2017

GENERAL COMMENTS	This is an excellent systematic review by Forni et al. This is important topic and the meticulous approach used in this study increases the value of the manuscript. Authors provided excellent review of the limitations of the studies and their effect on the usability of the existing models. Discussion is well written and elucidate remaining important questions to be answered. I congratulate authors on this paper and hard work put in it. I would also suggest editor to consider editorial for this manuscript.

REVIEWER	Dr Kate Birnie University of Bristol, UK
REVIEW RETURNED	24-Apr-2017

GENERAL COMMENTS	1) This study carries out a systematic review of prognostic prediction models for acute kidney injury in general hospital populations. The authors state that they haven't carried out a formal risk of bias, as they say it isn't available in the field. However, the prediction model risk of bias assessment tool (PROBAST) has been presented at Cochrane symposiums in recent years, training workshops have been run and the tool is available for use if you contact the study authors. Details available at http://www.systematic-reviews.com/ or Cochrane Prognostic Methods Group http://methods.cochrane.org/prognosis/scope-our-work. Published systematic reviews have already used this tool, e.g. Ensor J, Riley RD, Moore D, et al. Systematic review of prognostic models for recurrent venous thromboembolism (VTE) post-treatment of first unprovoked VTE. BMJ Open 2016;6:e011190. doi:10.1136/bmjopen-2016- 011190. 2) Methodological and reporting shortcomings in the included studies are summarised in Table 2, but this isn't mentioned until the discussion. It would be better placed citing Table 2 in the results section and then commenting on it in the discussion e.g. "the studies had several methodological limitations…". 3) The conclusion in the abstract states, "Similar candidate predictors reflect an elderly demographic with chronic co-morbidities." But, data are not presented on the age distributions across the included studies.

	4) Please define the CHARMS and TRIPOD abbreviations at first mention in the text. 5) In the results section, for model performance it says: “Median AUROC (or C-Statistic) was 0.74 (range 0.67-0.80) for derivation and 0.75 (range 0.66-0.80) for internal validations reporting discrimination (seven studies).” Please check these results. Should the summary statistics for the derivation AUROC be based on eight studies in Table 1? With a median of 0.745 (range 0.71 to 0.80)? And for internal validation, based on seven studies? With a median of 0.74 (range 0.66 to 0.80)? 6) Table 1: T&O abbreviation is not defined in the footnote. 7) Figure 1: Please check numbers e.g. abstract reading n=1002 articles, from which n=201 were omitted. This leaves 251, but the next box says, full articles reviewed were n=254. 8) eTable 2: Are lines 13-26 used in the Embase search strategy? The final line (no. 27) limits line no. 14 to Human and Publication Type Articles, giving 9102 results. Does this suggests that lines 15-26 were not in used in the final result? 9) eTable 3: The Ovid Medline search term in line no. 7 says “28 OR 29 OR 30 OR 31 OR 32 OR 33”. Should this be “1 OR 2 OR 3 OR 4 OR 5 OR 6”? Similar queries for lines 13-15. 10) eTable 6 includes useful information, but lacks some important demographic information on the study populations e.g. age of participants (see comment no. 3 above) and % male. Also, in the current format it is difficult to compare characteristics across studies. Have other formats been considered? 11) eTable 7 is not referred to in text of the manuscript. Also, there seems to be a lot of abbreviations in eTables 6(i-iv) that are not included in eTable 7. For example, BP, HR, K+, WCC, CRP, RR, AVPU, etc. 12) PRISMA checklist: For item no. 11, the eTable number isn’t given. For item no. 14, Table 1 is given as the reported item. However, these doesn’t seem a relevant response, as PRISMA item 14 is in the methods section and Table 1 in your manuscript is a results table. 13) Searches are until November 2016, please update these prior to publication.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
If the PRISMA guidelines were followed then this should be stated in the methods.
- We have now addressed this in the methods section.

In the methods, the authors acknowledge that the definitions for AKI may vary between studies. However, the definition for hospital acquired may also differ as noted on page 10. It may help to state in the methods the parameters defined during the search strategy to identify studies that reported hospital acquired AKI and the definitions employed by the different studies in this regard.

- Thank for this point, now addressed in the methods section. The definitions employed by the studies are noted in tables 1 and eTable 6. This point has also been addressed by performing a risk of bias assessment (PROBAST).

Page 11. Candidate predictors. How many models included serum creatinine? I think this should be explicitly stated as models that include this are at risk of confusing prediction and detection of AKI events.
- This is a very important point now included in the results and discussion section & is covered by the PROBAST assessment. A significant number of the studies either included admission SCr as a potential predictor or included this in the final model; other studies used admission SCr as the baseline; others did not clarify how or whether they differentiated those with AKI at admission vs hospital acquired. To this end we have added eTable 11 to clarify how each study handled SCr/baseline values and mentioned this in both the results and discussion. We have also now separated CKD from admission SCr in the key predictors Figure 2 & eTables 9-10.

In the first paragraph of the discussion, the most commonly used predictors are listed but haven't been clearly described in the results. I would add some more detail around this on page 11, and in table 1 if there is room, it would be nice to list the predictors included in each risk model (although that won't be possible for Matheny and Koyner). This would allow the reader to easily see the similarities and differences in these across the studies.
- We have now provided further description of the predictors in the results. The most common predictors are in figure 2 and in table form in eTable 9. Furthermore the full list of the predictors are in eTable 10 and we have now highlighted them in eTable 6 in bold.

On page 15, the authors mention that physiological parameters weren't often included as predictor variables. I think the discussion would benefit from expanding this argument a little more, getting at the point as to whether opportunities exist in future work to improve model performance by including variables that have not been well studied to date. This is important as in some cases, predictors are chosen for pragmatic reasons (i.e. what is possible to collect) rather than what may be ideal.
- This is another interesting point that future research will hopefully begin to address, particularly with the rise in track & trigger electronic observations eg in the UK Trusts using standardised EWS such as NEWS. We have mentioned this in the discussion.

Page 17. Discussion of potential for clinical implementation. One of the important results from this review is that despite the heterogeneity in setting and differences in risk score, the different studies generally report AUC values in the same range - reasonable but none >0.8. This moderate performance of risk scores does feed into the discussions as how to think about implementation and may make some future directions more attractive than others. This argument could be brought out more this section of the discussion.
- As with the point above, we have added in the discussion section.

Reviewer: 3
1) This study carries out a systematic review of prognostic prediction models for acute kidney injury in general hospital populations. The authors state that they haven't carried out a formal risk of bias, as they say it isn't available in the field. However, the prediction model risk of bias assessment tool (PROBAST) has been presented at Cochrane symposiums in recent years, training workshops have been run and the tool is available for use if you contact the study authors.
- We have now carried out a risk of bias assessment using the most recent update of the PROBAST tool shared by the author.

2) Methodological and reporting shortcomings in the included studies are summarised in Table 2, but this isn’t mentioned until the discussion. It would be better placed citing Table 2 in the results section

and then commenting on it in the discussion e.g. “the studies had several methodological limitations…”.
- We have now made reference to table 2 in the results and discussion.

3) The conclusion in the abstract states, “Similar candidate predictors reflect an elderly demographic with chronic co-morbidities.” But, data are not presented on the age distributions across the included studies.
- We have now added age in the results text, table 1 and e Table 6.

4) Please define the CHARMS and TRIPOD abbreviations at first mention in the text. - Updated
5) In the results section, for model performance it says: “Median AUROC (or C-Statistic) was 0.74 (range 0.67-0.80) for derivation and 0.75 (range 0.66-0.80) for internal validations reporting discrimination (seven studies).” Please check these results. Should the summary statistics for the derivation AUROC be based on eight studies in Table 1? With a median of 0.745 (range 0.71 to 0.80)? And for internal validation, based on seven studies? With a median of 0.74 (range 0.66 to 0.80)?
– Many thanks this has been updated in the text.

6) Table 1: T&O abbreviation is not defined in the footnote – updated
7) Figure 1: Please check numbers e.g. abstract reading n=1002 articles, from which n=201 were omitted. This leaves 251, but the next box says, full articles reviewed were n=254.
- Records excluded should have read 748 & this has now been corrected. We have reconfigured the PRISMA flow to have articles retrieved from references (n=21) at the top of the flow.

8) eTable 2: Are lines 13-26 used in the Embase search strategy? The final line (no. 27) limits line no. 14 to Human and Publication Type Articles, giving 9102 results. Does this suggests that lines 15-26 were not in used in the final result?
9) eTable 3: The Ovid Medline search term in line no. 7 says “28 OR 29 OR 30 OR 31 OR 32 OR 33”. Should this be “1 OR 2 OR 3 OR 4 OR 5 OR 6”? Similar queries for lines 13-15.
- Many thanks for points 8 & 9 - both the Embase & Medline search strategy summaries have now been corrected – both were incorrectly pasted from the original database searches.

10) eTable 6 includes useful information, but lacks some important demographic information on the study populations e.g. age of participants (see comment no. 3 above) and % male. Also, in the current format it is difficult to compare characteristics across studies. Have other formats been considered?
– We have updated the Table 1 to include age and eTable 6 to include age and sex including statistical comparisons, where reported.

11) eTable 7 is not referred to in text of the manuscript. Also, there seems to be a lot of abbreviations in eTables 6(i-iv) that are not included in eTable 7. For example, BP, HR, K+, WCC, CRP, RR, AVPU, etc.
- We have now added in the missing abbreviations and referred to the table in the text.

12) PRISMA checklist: For item no. 11, the eTable number isn’t given. For item no. 14, Table 1 is given as the reported item. However, these doesn’t seem a relevant response, as PRISMA item 14 is in the methods section and Table 1 in your manuscript is a results table.
- We have updated these omissions.

13) Searches are until November 2016, please update these prior to publication.
- The searches were performed less than 3 months prior to submission to BMJ Open. We believe this is a reasonable timeframe given the extensive searches and critical appraisal of the included articles. For example the recent review, cited by the reviewer in BMJ Open by Ensor et al 2016 completed a

search 2 years prior to publication. Efforts have been made to include any relevant articles identified since the search end dates – no new general population study to date has been described to our knowledge.

VERSION 2 – REVIEW

REVIEWER	Nick Selby University of Nottingham, UK
REVIEW RETURNED	27-Jun-2017

GENERAL COMMENTS	The authors have addressed all previous comments, and are to be congratulated on this important systematic review. Table 1 in particular is now very useful indeed.

REVIEWER	Azra Bihorac University of Florida, Gainesville, Florida, USA
REVIEW RETURNED	28-Jun-2017

GENERAL COMMENTS	Authors have answered all comments raised by reviewers.