PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Predicting Risk of Hospitalization: A retrospective population-based analysis in a Pediatric Population in Emilia-Romagna, Italy
AUTHORS	Louis, Daniel; Callahan, Clara; Robeson, Mary; Liu, Mengdan; McRae, Jacquelyn; Gonnella, Joseph; Lombardi, Marco; maio, vittorio

VERSION 1 – REVIEW

REVIEWER	Satvinder Dhaliwal Curtin University
REVIEW RETURNED	13-Oct-2017

GENERAL COMMENTS	• Unclear definition on potentially preventable hospitalization The authors identified a total of 24 condition/body system based on References (14, 15, & 16), but References 14 & 15 were published in year 2000 and 2003; Management for certain health conditions have been moved from hospital to primary care settings during the period 2006 to 2015. The identified 24 condition/body system are too broad. For example, "Cancer" – is all 'Cancer' could be managed at primary care settings The authors pointed out, under the Background, that most existing literature (References 4-9) on predicting hospitalisations were based on condition specific, especially chronic health conditions – Is it feasible to examine all-cause hospitalisations? • The authors have stated the C-statistics of the model, however, there was no description of the model. The authors have not stated the effect size of the predictors in the model using odds-ratio or risk-ratio, with associated 95% confidence intervals. Where are the tables for the six multivariable logistic regression models? Which variables were significant? • Lack of explanation on the identified categories of risk - The authors stated that "At higher risk" using a threshold of predicted risk >2.5%, however, there was no explanation on how 2.5% was decided upon • Similar text of this manuscript compared with an earlier

	publication by Louis et al. (2014) entitled "Predicting risk of hospitalisation or death: a retrospective population-based analysis", based on Turnitin software.

REVIEWER	Rhema Vaithianathan Auckland University of Technology, New Zealand
REVIEW RETURNED	29-Oct-2017

GENERAL COMMENTS	1. Concerned about the very high rate of ENT amongst high risk. Can you check if this is right? If correct, are there some readmissions that are ENT driving your results? e.g. If you drop ENT related readmissions, will it drastically change c-statistic? 2. Cite: Billings, J.; Dixon et al. BMJ 2006 3. Would like to see further work using techniques like Random Forest as sensitivity of 0.43 is pretty low. 4. Please let us know how many were dropped in the exclusion criteria, especially those with no valid history (see pg. 6+7).

VERSION 1 – AUTHOR RESPONSE

We have made revisions to our article as recommended, and look forward to moving ahead in the publication process. Changes to our manuscript within the document are in red text. Figure 1 has been replaced because there was an error.

Editor Comments to Author:
- Please include the study design and setting in the title. This is the preferred format of the journal.

We have revised the title to the preferred title format of the journal.

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Satvinder Dhaliwal
Institution and Country: Curtin University

1) Unclear definition on potentially preventable hospitalization

We apologize to the reviewer that we omitted the Appendix material from our original submission. We believe that the inclusion of the Appendix adds the necessary specificity to the definition in the body of our manuscript.

2) The authors identified a total of 24 condition/body system based on References (14, 15, & 16), but References 14 & 15 were published in year 2000 and 2003; Management for certain health conditions have been moved from hospital to primary care settings during the period 2006 to 2015.

The reviewer correctly points out that the work Reference 15 (Shi et al) was published in 2000 and health care management has changed in some cases over that time. However, the AHRQ pediatric quality measures (Reference 16) are quite recent. In addition, the 2 physician authors of the manuscript reviewed all of the Shi criteria and updated the definitions as appropriate. Again, we have to apologize for omitting the Appendix material from our original submission since that material shows

the precise changes we made to update our definition of "hospitalization that could have potentially been prevented or delayed."

3) The identified 24 condition/body system are too broad. For example, "Cancer" – is all 'Cancer' could be managed at primary care settings

We would prefer as much specificity as possible in the definition of the patient characteristics used in the predictor variables. However, even with a relatively large data set available for our research tradeoffs are necessary between specificity and have a manageable number of groups for analytical purposes. We believe that our approach of mapping diseases defined primarily by the affected body system with the exceptions of cancer, genetic conditions, and trauma which were based on etiology achieves a reasonable tradeoff as evidenced by the performance of our models.
We do not understand the second part of the reviewer's comment: "Cancer" – is all 'Cancer' could be managed at primary care settings. Obviously, not all cancer can be "managed at primary care settings." But, we expect that a child with a history of cancer will be at higher future risk of hospitalization.

4) The authors pointed out, under the Background, that most existing literature (References 4-9) on predicting hospitalisations were based on condition specific, especially chronic health conditions – Is it feasible to examine all-cause hospitalisations?

While it may be feasible to examine "all-cause hospitalization" we believe that some hospitalization is not currently predictable. For example, we cannot currently predict who will develop appendicitis. Therefore, including appendicitis in the definition of the dependent variable would not be helpful. On the other hand, complications of appendicitis, such as intestinal obstruction, can potentially be avoided through prompt diagnosis and treatment. Our goals was the development of predictive risk models for the pediatric population to help identify children who are at risk of hospitalization for conditions that may be affected through improved patient care. We believe that the results will be more useful to the pediatricians and other providers than all-cause hospitalization.

5) The authors have stated the C-statistics of the model, however, there was no description of the model.

We agree with the reviewer's comment and have now added supplemental material that shows all six models – 3 age groups times 2 genders = 6 models.

6) The authors have not stated the effect size of the predictors in the model using odds-ratio or risk-ratio, with associated 95% confidence intervals.

We are now providing the log-odds model coefficients and p-values, but have chosen not to report odds ratios and 95% confidence intervals. As you can see, the models include many predictors, many of which are likely to be correlated, and we do not expect that their values to actually represent unbiased independent risk contributions to the predicted probability of hospitalization. Moreover, it is not our goal to evaluate independent risk contributions of any particular risk factors. Our goal was to construct models for generating a predicted probability of hospitalization for each pediatric patient to be used in risk stratifying them and selecting those patients predicted to be at higher risk of hospitalization and compiling, for their pediatrician, a printed descriptive profile for each of their patients at higher risk. We believe that our modeling approach delivered reliable predicted probabilities, as evident from our C-statistic of 0.78, which compare favorably to others in the literature.

7) Where are the tables for the six multivariable logistic regression models? Which variables were significant?

We agree with the reviewer's comment and have now added supplemental material that shows all six models – 3 age groups times 2 genders = 6 models.

8) Lack of explanation on the identified categories of risk - The authors stated that "At higher risk" using a threshold of predicted risk >2.5%, however, there was no explanation on how 2.5% was decided upon

These definitions were chosen arbitrarily to result in a manageable set of profiles for the pediatricians. We have added a sentence to the last paragraph of the methods section to clarify how we selected the threshold of 2.5%.

9) Similar text of this manuscript compared with an earlier publication by Louis et al. (2014) entitled "Predicting risk of hospitalisation or death: a retrospective population-based analysis", based on Turnitin software.

The similarities are ~1% of the overall text according to a software package we tried– which might not be set up like the reviewer's. Basically, there are about 5 blocks of text with matches between the papers… and two of them appear to be common references. We reviewed the matches and believe that they are reasonable given the fact that this manuscript clearly builds on our previous work for an adult population to develop similar models for a pediatric population.

Reviewer: 2
Reviewer Name: Rhema Vaithianathan
Institution and Country: Auckland University of Technology, New Zealand

1. Concerned about the very high rate of ENT amongst high risk. Can you check if this is right? If correct, are there some readmissions that are ENT driving your results? e.g. If you drop ENT related readmissions, will it drastically change c-statistic?

We checked the calculation of the ENT variable in our model and believe it is correct. ENT problems are very prevalent in children. The definition of ENT problems in the independent variable included a history of outpatient visits to otolaryngologists, a very common occurrence in the pediatric population. We believe that the high rate of ENT problems among the high-risk children reflects their increase risk for hospitalization for one of the conditions included in our list of hospitalization that could have potentially been prevented or delayed. We also looked at ENT related readmissions and did not find a high rate that would justify re-running all the models to recalculate the C-statistic.

2. Cite: Billings, J.; Dixon et al. BMJ 2006

We gave now cited both Billings 2006 (BMJ) and also a more recent paper Billings 2013 (BMJ Open) in the first paragraph of the discussion section of our manuscript.

3. Would like to see further work using techniques like Random Forest as sensitivity of 0.43 is pretty low.

We do not agree that "sensitivity of 0.43 is pretty low." When we arbitrarily placed the threshold for defining higher risk at > 2.5% estimated risk of hospitalization (see our rationale explained above in response to Reviewer 1), we are not actually implying that nobody outside that group is predicted to

experience death or hospitalization – which is precisely what would result in a perfect sensitivity (assuming we make at least one true positive). On the contrary, by the definition of the threshold and our prevalence estimate-based risk scores, we expect that many such events will indeed occur among those with an estimated risk below 2.5%. The tradeoff between positive predictive values and sensitivity is addressed in: Billings J, Georghiou T, Blunt I, et al Choosing a model to predict hospital admission: an observational study of new variants of predictive models for case finding. BMJ Open 2013;3:e003352. doi: 10.1136/bmjopen-2013-003352

We thank the reviewer for suggesting Random Forest approach. It is plausible that that approach would arrive at predicted probabilities that would perform comparable to ours, possibly better. However, we feel satisfied with the predictive performance of the logistic regression models reported in our manuscript.

4. Please let us know how many were dropped in the exclusion criteria, especially those with no valid history (see pg. 6+7).

We agree with this suggestion and have made the necessary changes to the methods section and the first paragraph of the results section.