Reviewer’s report

Title: Use of Natural language processing to improve predictive models for imaging utilization in children presenting to the Emergency Department

Version: 1 Date: 14 Jul 2019

Reviewer: Marie-Helene Metzger

Reviewer's report:

Objective: the aim of this study was to compare three predictive models for medical imaging use during the pediatric ED visit, using structured data, unstructured data and both obtained at triage.

The first question is that of the clinical relevance of such work. The hypothesis of the authors would be that predicting the prescription of an imagery at the ED triage would reduce the ED crowding ... Some imaging tests could be prescribed by the triage nurse to allow the doctor to have the elements of decision more quickly, but this is for very specific situations: for example, ankle X-rays for a suspicion of sprain in the context of trauma. There is no need to develop algorithms for this type of symptoms ... Apart from these exceptions, this prescription must be medicalized after a clinical examination (medico-economic risk to carry out imagery examinations that would not be of interest for the clinical management of the patient).

The study based on an outcome as "imaging use" without any precision on the type, does not allow to consider an automation of the corresponding prescription. It would be necessary to detail in a much more precise manner, the exact type of imaging and the corresponding medical context (pulmonary Xray in case of pulmonary symptoms, abdominal ultrasound and / or abdominal CT in front of abdominal pain etc.). The analysis presented in this manuscript does not allow to build algorithms of this level of precision, in particular with the NLP method used (PCA), which does not allow analyzes by medical topic. From my point of view, the study is therefore not clinically relevant as a potential decision-support tool.

Methodologically, the authors have already conducted the same study in adults (see publication ref 43) and there is no methodological contribution to this publication. The following methodological limitations are present:

- absence of information on the temporality of imaging prescription in the patient's pathway (prescription in the emergency ED only or also during the hospitalisation?)
- same list of co-morbidities in pediatrics as in adults, without adaptation of comorbidities most frequently found in pediatric population: cancer, cerebral cardiovascular disease, congestive heart failure, COPD, HIV. Consequently, the number of patients is very low and these variables do not remain in the final predictive models except COPD. Did a pediatrician participate in this study?

-I could not find the dataset using the URL cited by the coauthors. The coauthors should detail the file path (in particular for the unstructured data).

- The AUCs from different models were compared using t-test but the results are not detailed in the results section. The authors wrote (lines 195-196) : "the AUC are significantly different between the models on the unstructured data, structured data and combined data (p<0,001)". The tests should be presented with more precision because according to the AUC 95% CI presented in Table 3, there are no significative differences between "unstructured variables" and "unstructured +structured variables" but there are significant differences between "structured variables" and the two other models. Therefore, the authors should detail the tests results in the results section (tests to do two by two) and adapt their results interpretation accordingly.

Conclusion: this study has no interest in terms of clinical relevance and does not bring any methodological novelty compared to the publication already made by the authors in the adult population.

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.

Unable to assess

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:

Acceptable

**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 organisation 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

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