Author’s response to reviews

Title: Leveraging the EHR4CR Platform to Support Patient Inclusion in Academic Studies: Challenges and Lessons Learned

Authors:

yannick girardeau (yannick.girardeau@aphp.fr)
Justin Doods (justin.doods@ukmuenster.de)
Eric Zapletal (eric.zapletal@aphp.fr)
Gilles Chatellier (gilles.chatellier@aphp.fr)
Christel Daniel (christel.daniel@aphp.fr)
Anita Burgun (anita.burgun@aphp.fr)
Martin Dugas (martin.dugas@ukmuenster.de)
Bastien Rance (bastien.rance@aphp.fr)

Version: 1 Date: 20 Nov 2016

Author’s response to reviews:

Dear Editor,

We would like to thank the reviewers for insightful comments. Please find below our point by point response to the comments.

Best regards,

Dr. Yannick Girardeau

---

Reviewer #1:

The authors presented a well written manuscript on their attempt to find and assess a general process on how to normalize patient data of EHRs for the application of patient recruitment. As a statistician in medical research I often encounter the problem of inadequate datasets for
analysing risk factors and similar. Therefore, the process and problems of normalization is an important concept for data assessment considering the mass of data reported each day. Of course the presented methods can only be a guideline on how patient data can be translated from EHRs. I have some minor remarks, I would like to authors to consider for their manuscript:

1) Describe early on how structured and unstructured data is characterized (first explained on page 10 line 191).

We modified the Material section and added a description of structured and non-structured data earlier in the text (Material section, line 120-125, page 5-6).

2) Describe in short how the institutional studies were selected and why with respect to the presented normalization problem.

We added a short description on how we selected the three different studies at the end of the material section line 179-181, page 8.

3) Give information on how time consuming the process was and how many physicians and other researchers were involved - since you talk about cost-reduction, an estimation of the workload is of interest. In this concept, also discuss the issue of feasibility, since a lot of work requires the physicians' time and expertise in several platform terminologies. The degree of feasibility is relevant, since the issue of usage of EHRs by physicians is a general problem discussed in medical research (time consuming/complexity).

We have not formally evaluated the time needed to perform the different tasks of the process. We recognize that it would have been relevant. The physician who took charge is trained in Clinical Research and Epidemiology and has experience in Medical Informatics, but no formal training in the latter.

If the approach discuss in the paper was to generalize, a bigger team would be involved including data-manager and terminologist who would strongly speed the process-up.

We added a brief paragraph about the feasibility in the discussion (page 20-21, line 440-446)

4) Is a double-check or sensitivity analysis required for certain steps? E.g. Page 10 Line 199 and following - it would be very likely that the childbearing age also has an upper limit; thus, if implemented "incompletely" or on a subjective basis, bias may occur.

We think that the reviewer makes an important point here. Double-check is certainly needed for the translation of criteria into the concepts from terminologies. Several issues can occur, including definition that may be subjective, bias, but also the meaning of the absence of information (in this case, does the absence of indication of pregnancy strictly equivalent to no pregnancy) – known as open-world vs close world assumption in database.

We added a paragraph on this issue in the discussion (page 18, line 378-383).
5) Details in discussion: Please give more detail on how the internationally recognized medical terminologies had an impact on the assessment (page 15, line 303) and whether/how this influenced the workload.

We have not evaluated strictly the impact of the use of standard (against local or terminologies). Among other aspects, there are several benefits for using standards, including, but not limited to, (a) more synonyms available for semi-automated matching, or concept identification, (b) reduced risk of misunderstanding, (c) sharing across recruitment platforms, which would not be possible otherwise. In contrasts, the workload would not be reduced if we had worked with local or dedicated terminologies, has at least one of the two partners would have had to go through all the processes.

We added a paragraph in the discussion section (page 17, line 344-346).

6) Consider rewriting the conclusion of the abstract, the given information and implications are formulated too generally.

We rewrote the conclusion of the abstract (page 3, line 49-51 and 53-54).

Recommendations/Correction:

1) Keyword: Patient Selection - better use "recruitment" instead of selection. I recommend avoiding this term since it can easily be associated with selection bias in patient recruitment.

We changed in the keywords, patient selection for patient recruitment, line 56, page 3.

2) Table 3: The main reason for the reformulation in A3 is the "negative" expression in the criteria; I would add this information in the table as well.

We made the reason clearer by adding “(negation removed)” in the table 3, page 10.

3) Table 5: Please correct the rounding of the percentages in the part "eligibility criteria": Doesn't it have to be 100% for the total of 67 criteria? Similarly, 50/67 to 75%, 36/67 to 54%, and 8/12 to 67%.

Done

Reviewer #2: This is an interesting study with the potential of a long term impact on the recruitment of patients to academic studies. The authors describe the automated operationalisation of inclusion/exclusion criteria using "Electronic Health Records" for clinical research.

Listed below are a number of issues that the authors might want to consider.

- Under "Ethics" the authors report that the study did not require ethical approval, however, they do not mention whether they had the permission from patients involved in the listed studies
to use their data for research other than the original intended studies. I am assuming here that the data presented in Additional file 3 is based on the participants of the three studies. I think that the authors should have discussed the ethical issues surrounding use of patient data for purposes other than the original studies that they have agreed to be involved in. Also, it would have been useful to highlight how one can operationalise the systems so that such approval is sought and to discuss the triggers where additional approval might/should be sought.

The outcomes provide in our study concern only the criteria of the studies themselves. We have not used at any time of our study the data concerning the patients who have been included in those studies. Moreover, Additional file 3, present only the number of patients present in the clinical data warehouse of Georges Pompidou Hospital for whom the information is present and structured for each medical concept. This information does not correspond to the number of patients per criteria or the number of patients eligible for a study.

- In the abstract a brief insight into what are non-structured data, normalisation in what sense, what do medical domains and concepts refer to would be helpful.

We rewrote the abstract to make these concepts clearer to the general public (page 2, line 42-43).

- It would have been useful to show a scenario of how using the suggested platform would have improved the screening of patients for inclusion in a clinical study.

We do agree with the reviewer that this is an interesting topic, but we think this should be the topic of another study as it is a very different work from our proposition. To be clear, evaluation of the improvement of the screening is of the topics we would like to explore next.

- To appeal to wider audience the authors should ensure that abbreviations are spelled on first appearance and try to keep these to a minimum. For example, EFPIA, HEGP, i2b2, UKM, DB, PI,.. ....

We thank the reviewer of his/her comment. We clarified the abbreviations and acronyms.

EHR4CR (Electronic Health Records for Clinical Research); page 4, line 89.
EFPIA (European Federation of Pharmaceutical Industries and Associations); page 5, line 95
HEGP (Georges Pompidou European Hospital); page 5, line 115
UKM (Munster University Hospital); page 6, line 124
ECLECTIC (Eligibility Criteria Language for Clinical Trial Investigation and Construction); page 7, line 150
SQL (Structured Query Language); page 7, line 154

- Tables 1 and 2 should be introduced within the text.
Done, page 6, line 131 and page 8, line 173

- Lines 144-155 further information about the "Selected studies" would have been helpful, for example, time frame of the study and target population, where each was conducted HEGP/UKM.

We have added the information (page 7-8, line 160-173)

  o Highlight whether aXa is a case-control study. Is it possible that this study was registered under another trials register such as www.clinicaltrialsregister.eu?

  aXa is a case control-study (information added page 7, line 159. It has been finally registered in clinical trial.gov, NCT02898051; page 7, line 159

  o Give figures for the enrolment rates

We did not work on the study itself, but on the design part and the potential use of recruitment platforms. Enrolment rates will be needed for the follow-up study, but we are afraid it would be too confusing in the present article.

- A clear indication as to what is considered as structured data versus unstructured data early on in the manuscript.

We modified the Material section and added a description of structured and non-structured data earlier in the text (Material section, line 120-125, page 5-6).

- Would it be possible in the supplementary files to have in addition to the listed criteria in French an English translation?

We added the criteria in English in the supplementary file 2.

- Under "Completeness" highlight the main points from additional file 3.

We added a foot note to remind the reader of the definition of the completeness to clarify.