Reviewer’s report

Title: Evaluation of data completeness in the electronic health record for the purpose of patient recruitment into clinical trials: a retrospective analysis of element presence

Version: 2 Date: 18 December 2012

Reviewer: Monica Horvath

Reviewer’s report:

The purpose of this study is to evaluate whether EHR data contains enough information to determine eligibility for clinical trial protocols.

A major weakness of this paper is that there are many different types of clinical trials and many bodies of eligibility criteria that could be used. Some trials relying upon gender and ICD-9/10 would be easy, while others that need social history could be quite difficult. As a result, the paper is really about how often certain data types are captured in a structured manner in EHRs. It is too broad to state that this study is an assessment of how well EHRs as a whole can provide eligibility criteria. This is an over-generalization.

The authors should propose a common set of data elements in EHRs that seems to be needed by most researchers—thus building upon the Luo data. This would improve the extensibility of the work in creating guidance for EHR deployment and reporting teams.

Major Compulsory Revisions

• A major weakness of the paper is that not all EHRs are created equal. What it means to be an EHR is a function of the care site as well as vendor. The authors should describe how much EHR adoption is present in the systems they are evaluating, perhaps using the common HIMSS EMR adoption levels (http://www.himss.org/content/files/EMR053007.pdf). The authors note that 5 different EHR systems are involved. Much more description of the EHR functionality is required.

• Missing in the study rationale is a discussion of the time it takes to recruit for clinical trials. Why do we need electronic data capture? What gains or savings are to be seen?

• I find the reliance on ‘semantic groups’ troubling. Are there not data standards that should be followed or considered? By defining semantic groups, the authors are designing their own ontology. I think more needs to be explained from the Luo paper.

• In the background you say the EHR doesn’t contain all info necessary to gather eligibility criteria. What is missing? This seems unreferenced.

• The conclusion that EHR data is not acceptable for patient recruitment is too broad and not backed by the data. This interpretation is only relevant for the
specific EHR scenario the authors studied.

- In the discussion, expound more upon what design elements are required of EHRs to better serve researchers.
- Your conclusion refers to the sufficiency of the ‘commonly available EHR’, and yet little description goes into this concept of an ‘average’ EHR. Study conclusions depend heavily on this. At Duke Medicine we are deploying Epic and expect to get many of the data elements you reference easily.

Minor Essential Revisions

- The use of the word ‘decomposed’ to describe creation of data elements is odd. Choose ‘reviewed’ or similar instead.
- In the Background, there are a number of other studies that evaluate the usefulness of EHR data for secondary use, including Meaningful Use reporting in the US. These should be cited and discussed, either here or in the Discussion.
- In the Background, define selection bias clearly as this is an IT journal.
- Cite Table 3 beginning with the ‘health status’ section under Results.
- The use of the Luo data as the focal point of defining eligibility criteria used often in clinical trials should be more prominently noted in the abstract; it is a key aspect of your study.

Discretionary Revisions

- Some of the eligibility criteria seem rather subjective, such as life expectancy. How do study coordinators make this determination and is there an algorithm to follow that could be built in to EHRs to make them serve researchers better?
- Add some thoughts as to why the Luo semantic group distribution differed in places compared to this study.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.