Reviewer’s report

Title: A European inventory of common EHR data elements for clinical trial feasibility

Version: 2 Date: 2 October 2013

Reviewer: Rachel L Richesson

Reviewer’s report:

MAJOR COMPULSORY REVISIONS:

This is a nicely written paper on an important topic, and this reviewer applauds authors for clearly describing the importance of common data elements to support care delivery and research. Further, authors have done a nice job identifying an approach to this problem in the form of an inventory of data elements that is derived both from elements available in EHRs and important to research. However, I think that the investigation authors have done here is only preliminary and not ready for readers hoping to identify and implementing standards. Authors have not provided enough detail on the methods for readers to assess if their collection of data elements is relevant (either from the perspective of HER developers/implementers or researchers). For example, it is not clear how many EHRs were surveyed or the details about them (e.g., type of system, type of facility, nature of patient populations, size, etc.). In addition, there are no details about who you surveyed in the pharma companies — how many people were surveyed as part of your consensus process and what was their background and training? What types of protocols and studies and forms did you look at for research elements? What disease areas?

The approach I would suggest is to identify clearly important data elements for research (perhaps by surveying a set of CRFs or protocol eligibility criteria as you did but in a bigger way with more details on the sources of your research data elements and supporting data (e.g., type of study, medical domain, type of form, nature of research questions) to demonstrate the representativeness or generalizability of these elements to other research domains. Then, I would attempt to match these elements to data availability in various EHRs and report the coverage of elements (and likely the different variations of value sets and definitions) that are available in different EHRs.

Although, overall the paper is logically organized and nicely written, it would benefit from some editing. The description of multiple iterations of the inventory are confusing and could be condensed to describe the (iterative) development of the final inventory.

The term ‘feasibility analysis (2nd paragraph Background section) needs to be defined more directly and subsequently labeled as ‘trail’ or ‘site’ feasibility. This could be an important and central theme throughout the paper. Overall, paper is fairly abstract and grounded examples would help.
Good point about complexity of eligibility criteria in background. Although it provides nice context, it is distracting to your study and should be removed. More relevant would be text on other CDE elements, including NCI, LOINC, USHIK, CDISC, and FDA efforts – and probably many others. Similarly, the details on data element definitions might be too much for this paper.

Other editorial suggestions include:

Streamlining the Abstract to be a bit more concise. Key points to focus on are study planning (perhaps focus on site identification) and clearly defining your methods – including details regarding terms like “consensus-driven approach.” For example, the abstract background could be trimmed and focused as below.

Clinical studies are a necessity for new medications and therapies, but many fail to meet their recruitment numbers. Study planning could be supported by more reliable site identification in the feasibility analysis. The increasing number of electronic health records (EHRs) in hospitals creates huge databases that could be utilized to support research. The Innovative Medicine Initiative (IMI) funded project “Electronic Health Records for Clinical Research” (EHR4CR) has developed a standardized and homogenous inventory of data elements to support utilization of EHRs for research.

We developed an Inventory of elements for feasibility analysis of new trials.

Further, the examples that authors provide are not well modeled data elements nor should they be promoted as standards of any kind. For example, in the attachment (appx), the first data element DOB does not specify format at MM/DD/YYYY or DD/MM/YYYY and so even in the example (10/10/1967) it is not clear. The description of lab data elements should be more clear and reference LOINC and CDISC and HL7 models that are fairly well developed and standardized.

**Level of interest:** An article of limited interest

**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.