Author's response to reviews

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

Authors:

Justin Doods (justin.doods@uni-muenster.de)
Martin Dugas (martin.dugas@uni-muenster.de)
Fleur Fritz (fleur.fritz@uni-muenster.de)

Version: 4 Date: 10 December 2013

Author's response to reviews: see over
Author's covering letter for manuscript submission

Dear Sir or Madam,

Hereby I resubmit a second revision of the manuscript "A European inventory of common EHR data elements for clinical trial feasibility" (MS:  2446001331049500) for publication in Trials. Please find the answers to editorial remarks and the reviewer’s comments below.

I am looking forward to receiving your final answer about acceptance.

Sincerely yours,

Justin Doods

Contact:
Justin Doods
Institute of Medical Informatics
Albert-Schweitzer-Campus 1 / A11
48149 Münster, Germany
Justin.Doods@uni-muenster.de
+49 (0)251 – 83 52587

Reviewer 1: Felix Koepcke

- Major Compulsory
  none
- Minor Essential Revisions
  1) hardly used data elements are moved to the wish list. Still some data elements in the data inventory show less than 10% usage (e.g. substance abuse 2%). Should these not therefore be moved to the wish list?
    In the initial version of the manuscript we described in detail which elements were moved to the wish list, but due to removing of all references to two iterations those details were removed as well. From the original manuscript: "For example, one important decision after the first data export was to remove data elements from the inventory that were available in less than half of the source systems. … . After the second data export, elements were moved to the wish list that were not available or not used at any of the sites."
    Therefore the data inventory contains also some very few elements with less than 10% usage.

  2) In the background you introduce fully structured, semi-structured and free text data elements, but semi-structured elements never appear again in the manuscript. The distinction should be referenced or explained.
    We believe that the structured data and semi-structured data are well known and established terms in the field and thus do not have to be explained. Although semi-structured is not further used, it is introduced because the background section usually also talks about general concepts that might not be further discussed.

  3) There is also still no definition for availability. (available yes/no) is not sufficient.
    In our opinion the whole sentence explains well what is meant with availability.
    "Data exports at the eleven EHR4CR sites were conducted to capture the availability of each element (available yes/no) […] at the source systems."
    "Available yes/no" is not the definition but is meant to indicate what the sites were supposed to answer in the data export.
4) I think frequency and usage mean the same thing. One of the terms should be replaced by the other throughout the manuscript. (Personally I would prefer usage.)

They mean the same thing, yes, but depending on what we are describing one fits better than the other. For example when we talk about how often a data element is available in a source system we use frequency. When we talk about documentation habits we use the term usage.

5) I would still like to see which site searched a subsystem and which their EHR. general categories are sufficient, e.g. EHR, intensive care, cardiology, pathology, ... . This information could easily be added as a header line in the heat map.

Sites were asked to report numbers on different kinds of data (laboratory, administrative, diagnostic data etc.). This data could have come from different subsystems and the sites were not required to note from what subsystem (LIS,...) the data came from.

6) English language should be improved.
We thoroughly read the manuscript and improved the language at different locations.

- Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)
  7) by a group of up to 30 people > by a group of 30 people
  8) standard definitions were identified and added > definitions were added
  9) To verify their availability > To measure their availability
 10) Rankings of data elements were created from the results of the data exports. > Data elements were ranked according to usage.
 11) In conclusion: The data elements in the Data Inventory were identified with the knowledge of domain experts from pharmaceutical companies. > This is not a conclusion. It could be better to elaborate on your second sentence: What is available? What not?
 12) Generally, the background could use more references.
 13) Table 1 is unneccessary. It is the same as figure 2. If you want to keep it, data item and data group part should be indicated.