Reviewer’s report

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

Version: 2 Date: 19 September 2013

Reviewer: Felix Koepcke

Reviewer’s report:

In the manuscript, the authors create a list of data elements they consider sufficient to evaluate the feasibility of clinical trials in a set of clinical domains and analyse their availability in multiple European EHRs. The list of necessary data elements is created by a survey and abstraction from 17 trial protocols. The method to determine availability is unclear. Predominantly available data elements are added to a “data inventory”, while predominantly unavailable data elements are added to a “wish list”.

I like the concept of the manuscript. The object of research is well defined, unexplored on a European scale and of huge interest to the medical informatics community dealing with routine data based clinical trial support. It should definitely be published in this journal. However, I see some lack of detail in the methods and result section. Furthermore, the discussion is too focused on the methods (I found this actually too much) and would benefit from an examination of the results and their meaning for secondary use in general.

Major Compulsory Revisions

1. If possible remove all references to performing 2 exports and creating 2 data inventories. As nothing appears to have happened between both exports besides adding new data elements to the list (and losing 2 data providers), I cannot see how data inventory 1 was anything but an intermediate version of your results with which you were somehow unsatisfied. Through this, the manuscript would become substantially better to follow.

2. Describe why you decided on defining your own groups and not use existing categories of Luo 2011 (source 19). Luo’s objective was “To semi-automatically induce semantic categories of eligibility criteria from text and to automatically classify eligibility criteria based on their semantic similarity.” They developed a “generic semantic categorization of eligibility criteria” from 4821 criteria sentences and “measured the prevalence of the categories in 27,278 eligibility criteria from 1578 clinical trials”. In my opinion, the method and coverage of Luo is superior to those of this manuscript (no method for derivation of groups given; data elements from survey and 17 trials) and you should have built on that. You should either change your data groups to the semantic classes of Luo or explain in detail, why Luo’s results for clinical trial criteria in general are invalid for feasibility analysis.
3. Diagnosis and procedure are treated very generically (“diagnosis code” and “diagnosis text”), while laboratory findings are considered in detail (41 individual codes / findings). Explain somewhere the rationale behind this decision. Also discuss the meaning of the results for a practical case, which will most likely require specific diagnosis codes.

4. In the discussion, you state that some data providers analysed the availability of data not in the EHR but in some subsystem. Depending on how percentages were calculated and the nature of the subsystem, the results of those data providers might not be comparable with those data providers that used their EHR. Elaborate on this in more detail.

Minor Essential Revisions

Duplicate nature of methods and results:

5. What was the rationale behind performing one export with survey results and data elements from 5 trials, then extending the list by data elements from 12 additional trials and performing the export again? Was this planned from the start? If not, what led to the desire to extend the list?

6. Is it wrong to assume that the differences in ranking between the top data elements from export 1 and 2 are due only to 2 sites delivering no results and adding new items? Otherwise I do not see why any changes to the availability of data elements should have happened at the data provider sites in the meantime.

7. There is only one expert group meeting in your figures 1 and 2. Does this mean the decision to extend by 12 studies was made without consulting the expert group?

8. “After each round of exports the results were analysed and a consensus on the data inventory was agreed. 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. Those data elements were put on a separate list, referred to as 'wish list'. After the second data export, elements were moved to the wish list that were not available or not used at any of the sites.” > how can there be data elements in the second export which are not available in any source system, when all data elements which were available in less than half of all source systems were removed after the first export?

9. Why were percentages grouped for one export and not for the other?

10. In the methods section you say that the definition of each data element’s meaning happened after adding data elements from 12 additional studies and therefore after the first export. If so, how reliable are the results from the first export if data elements were not defined at that time?

Methodology lacks detail:

11. Methodology for creating the “initial list” of data elements is missing: Add at
least (1) how their importance for pharmaceutical companies was measured, (2) the questions of the survey, (3) who (background/qualification) and (4) how many people were questioned, who and how many responded.

12. How was the heat map created?


14. How was the expert group composed?

15. Methodology for selecting the 17 studies is missing.

16. Rationale for requiring included trials to “have run at least at one EHR4CR data provider site” and to represent “each EFPIA company” (the EFPIA homepage says there are 73 member companies) is missing.

17. Chain of events unclear from methods section without figures. I advise to write in chronological order.

18. It remains unclear, what the ranking of data elements is done for. It also appears to make no sense to rank first by percentages and subsequently by availability.

More details on what the sites did needed:

19. Definition for “availability” should be added.

20. Please describe exactly how percentages were calculated.

21. Reason(s) why 2 sites did not perform a second data export are missing.

22. You mention that some sites used a specialized subsystem instead of the EHR. > Which sites used which systems and what was the reason to have them not use the EHR? > I would like a table with data provider, site ID (if this can be disclosed), description of the analysed IT component (specialty, purpose, …).

Additional results I would appreciate:

23. To identify common eligibility criteria you performed 1) a survey and 2) an analysis of trial protocols. Can you describe the match between both methods’ results?

24. Can anything be said on the frequency of the data elements in your inventory in real trial protocols?

25. Results of meetings / telephone conferences are missing.

26. Results section should describe the results in terms of availability and frequency of data elements with numbers.

Others

27. The introduction explains the motivation for the EHR4CR project, but not
specifically those of this manuscript. It remains unclear who will use the inventory to what purpose.

28. “How can a valid and EFPIA accepted inventory be created?” is a research question not being explored by this manuscript. “Creating an inventory” could be an objective.

29. I don’t have the ISO/IRC 11179 to check, but it seems odd to give two definitions for “data element”. Is it possible that the first definition is actually for “data element concept”?

30. I found many undefined concepts in data element definition and subsequent paragraph confusing. Definition requires following attributes for each data element: definition, identification, representation, permissible values, value domain, datatype, representation class, unit of measure. Of these the meaning of “definition”, “identification”, “representation”, “value domain”, “datatype” and “representation class” remain undefined.

31. Furthermore you only define “data group”, “data item” and “data type” for your data elements. In a strict sense, your data elements thus appear to be no data elements according to the given ISO/IRC 11179 definition.

32. You write in the methods section: “capture the availability of each element (available yes/no)” and “availability and frequency of a data element were captured separately.”. But later: “the availability of the elements (available = 0-100%; not available = N/A)”. The second phrase seems to contradict the first one.

33. The analysis seems to be limited to the experience of a given set of pharmaceutical companies and hospitals. Discuss their representativeness for clinical trials in general. How well do the data providers represent the average European hospital and how well do the pharm. companies represent the entirety of clinical trials?


Discretionary Revisions

35. The discussion of Weintraub is too detailed.

36. Paragraph Data Inventory mainly repeats paragraph Data Element.

37. First sentence second paragraph discussion repeats first sentence first paragraph discussion.

38. Non-UK source in addition to [1] would be welcome. Preferably something
with a European focus.

39. Is there a source for “If recruitment could be optimized by a better selection of clinical research centers and if better trial protocols could be created through an improved and more accurate feasibility analysis, clinical studies could be completed faster and more cost-efficient.”?

40. Same paragraph: Elaborate on how the lack of transparency for the sponsor constitutes a problem for the conduct of clinical trials.

41. “but Ross et. al. [3] showed that the majority of criteria in studies are highly complex”. Ross classified 85% of all criteria as complex; A criterion was complex if it included negation (26%), an arithmetic operator (15%) or a Boolean connector (53%). Thus “not pregnant” would already be a complex criterion for Ross, which is quite a low threshold in my opinion. Anyway, Ross uses no category named “highly complex” though one might argument that the 35% of criteria that contained 2 or more semantic patterns could be classified as such.

42. Is there a source for the classification of data as “semi-structured” if it is encoded in a proprietary terminology?

43. “data content and frequency” > maybe something with “element presence” instead of “frequency”

44. “In this process the elements were grouped by their context and the data groups were created as well.” > explain the difference between grouping by context and creating groups

45. Give an example of a typical criteria <> core information – pair.

46. In methods you speak of measuring frequency, relative frequency, percentage, relative percentage, numbers and relative numbers. Please unify or explain the difference.

47. “There is generally no direct incentive for documentation of time intensive scales ….” > From the context it is clear that you mean research related data, but still I fail to understand what time intensive scales are.

48. “The comparison of the data elements of both versions showed that the top 10 elements are stable although 31 elements got added or removed.” Appears out of context.

Spelling errors:

49. “where den validated”

50. “The definitions of each element contains”

51. “ISO/IEC 11179 Standard defines data element in Part 4 [7] as follows "Unit of data for which the defin …” > either introduce “:” or remove “follows”

Tables and figures:
52. Table 4: decimal places are unnecessary.
53. I suggest making figure 1 and 2 a single figure.
54. Sum column in heat map should be replaced by average and median.
55. Discharge date should not be coloured in table 4.
56. Quality of the heat map is too low (at least in the materials I received for review).
57. The heat map seems to be dispensable. It is neither presented nor discussed in the results or the discussion section. I advise to merge it with the supplement table.
58. Similarly, the wish list should be integrated within supplement 1. Following your methods, the wish list should resemble the lower part of the list of all data items, when sorted by availability.
59. Mention of figure 1 should be at the beginning of the text.
60. Table 5 and 6 are not necessary.

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

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

Declaration of competing interests:
I participated in a research project with M Dugas between 2010 and 2012. University hospital Erlangen is one of the data providers in the EHR4CR project, but I am not personally involved.