Author’s response to reviews

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

Authors:

Felix Koepcke (felix.koepcke@imi.med.uni-erlangen.de)
Benjamin Trinczek (benjamin.trinczek@uni-muenster.de)
Raphael W Majeed (raphael.majeed@chiru.med.uni-giessen.de)
Björn Schreiweis (bjoern.schreiweis@med.uni-heidelberg.de)
Joachim Wenk (joachim.wenk@uni-duesseldorf.de)
Thomas Leusch (thomas.leusch@med.uni-duesseldorf.de)
Thomas Ganslandt (thomas.ganslandt@uk-erlangen.de)
Christian Ohmann (christian.ohmann@uni-duesseldorf.de)
Björn Bergh (bjoern.bergh@med.uni-heidelberg.de)
Rainer Röhrig (rainer.roehrig@chiru.med.uni-giessen.de)
Martin Dugas (dugas@uni-muenster.de)
Hans-Ulrich Prokosch (hans-ulrich.prokosch@uk-erlangen.de)

Version: 3 Date: 12 February 2013

Author’s response to reviews: see over
We thank all referees for taking the time to read the manuscript and suggest possible improvements.

**Referee 1 (John Ainsworth)**

**Major Compulsory Revisions**

**Comment 1**

In Para 5 of the Methods Section, the authors state that "The calculation included all patients admitted in the fourth quarter of 2011 to the clinical department that conducted the trial." We can expect some patients to be not eligible for a trial, because they do not have a specific condition. The absence of a condition is normally not recorded, and so would be missing data. The authors need to discuss how this may affect the results. They may want to consider including a calculation for all patients recruited into a clinical trial in the clinical departments in the fourth quarter of 2011 and presenting these results in Tables 2 & 3.

**Response 1**

We completely rewrote the third paragraph of the discussion to draw attention to the difference between ‘missing data’ and ‘absence of the corresponding patient characteristic’ and the impact thereof.

The fact that absent patient conditions are not recorded is a problem for automatic eligibility screening that can be dealt with in two ways: (1) defining absent data as absent condition or (2) leaving the condition to be evaluated manually by study personnel. (1) is usually preferred if an abundance of potential participants is expected and (2) if a shortage is expected. The first way risks missing eligible patients for whom the data really is missing and the introduction of a selection bias if the distribution of missing data is not completely random. Whether missing data can be used for eligibility calculation or not depends on the patient characteristic, the data element and on how it is used. Our results show missing data as missing data. Whether the data is missing despite the presence of the condition or because the documentation of its absence was not deemed necessary does not influence this figure. In cases where missing data can be interpreted as ‘absent condition’ (i.e. the error introduced is small) a high fraction of missing data for a characteristic will not impede its use for eligibility screening.

**Comment 2**

The method described for matching patient characteristics to EHR data elements is not systematic and the outcome is unlikely to be repeatable. The authors should explain in more detail the steps taken to avoid this problem or discuss the sensitivity of the results to the mapping.

**Response 2**

We made minor changes to the fourth paragraph in the methods section and enhanced the fifth paragraph of the discussion.

The original method for matching eligibility criteria and semantic category developed by Luo et al. is also based on manual assignment. There is however one advantage in our works over the Luo paper
in terms of repeatability. While Luo et al. assigned whole criteria sentences to one semantic category, we added a step to decompose these sentences into single units of information first. We believe the remaining fragments are more easily attributed to semantic categories (and data elements later in the process). All assignments were checked by at least two of the authors. The most important categories (treatment; disease, symptom or sign; lab test; medication; age; gender) were very straightforward and we believe that other groups would arrive at (nearly) the same result here. Some of the rarer categories could be interpreted differently: e.g. special patient characteristic, disease stage, capacity.

Comment 3

Some data elements in any given EHR system will be mandatory for entry of a valid value e.g. gender. The authors should explain the impact this may have on the results.

Response 3

We only included structured data elements. These data elements can only be valued with a given set of options or a number. Thus it is impossible to value them with invalid values.

Referee 2 (Monica Horvath)

Comment 4

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.

Response 4

In the paper, we did not want to judge, whether the available data is sufficient or not for a given trial. This is clearly dependent on the trial and the targets of the recruitment system. In the objectives of the paper we therefore intentionally speak only of assessing the availability of data, not its sufficiency. Everything else you state in this comment is perfectly true. The paper is really about how often (measured as the fraction of all patients that would require to be screened for the corresponding trial) a given list of patient characteristics could be found in the EHR. The only statement we made regarding the suitability of EHR completeness for automatic recruitment was limited to ‘fully computerized recruitment’. We changed the conclusion anyway to prevent misconceptions.

Comment 5

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.

Response 5
We do not think it is possible to give such a set. The intersection of commonly required data seems to encompass only age, gender, pregnancy, alcohol or drug abuse and enrolment in other studies. All other criteria are usually depending on the disease under investigation. Regarding its content, we think it is very difficult to enhance the suitability of an EHR for trials in general. However, it is perfectly possible to enhance its suitability for trials of a specific specialty.

**Major Compulsory Revisions**

**Comment 6**

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.

**Response 6**

We added information on the EHRs to the first paragraph in the methods section.

**Comment 7**

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?

**Response 7**

There is indeed no rationale given for the usage of electronic systems for patient recruitment. After giving the extension of the introduction a lot of thoughts, we feel the target reader of the manuscript should be familiar with the patient recruitment topic. For those who meet this topic here for the first time and are interested to learn more we added reference 3, the most current review on computerized recruitment support by Cuggia et al. The paper offers a much more comprehensive overview of the recruitment process and potential benefits of EHR involvement, as well as the state of the art, then what we could possibly add here in a few sentences.

**Comment 8**

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.

**Response 8**

This is a tough question. We used the results of Luo et al. because they offer the most methodological way to arrive at their categories for eligibility criteria: For a large set of eligibility criteria, each fragment of each sentence was annotated with a corresponding UMLS concept. All annotated sentences are clustered for containing the same concepts. The clusters are manually given names. There is of course a great flexibility on what level to cluster and it is hard to justify the chosen level.
We think what you propose is to use the UMLS concepts directly, without the clustering. This would be a great thing to do, if we had done what we have done for many more trials (thousands), because there are just so many, very detailed concepts in the existing data standards and the eligibility criteria are so diverse. As is, we required the clustering to just 27 categories to arrive at meaningful results.

Before clustering ourselves, we thought best to reuse the clusters of Luo et al. Luo’s methods are too complex to explain in this manuscript without distracting from its actual topic.

Comment 9

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

Response 9

What was meant is that no EHR can contain patient data on all possible eligibility criteria. Every EHR is limited in its content, while the number of patient characteristics that could be required for eligibility criteria is virtually unlimited. Reference 3 (Cuggia et al.) is given (again) for the fact that most recruitment support systems do not evaluate every single eligibility criterion of a trial in all detail. We changed the first sentence of the second paragraph in the background section to make this clearer.

Comment 10

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.

Response 10

See comment 4. We did not intent to make such a conclusion.

Comment 11

In the discussion, expound more upon what design elements are required of EHRs to better serve researchers.

Response 11

While this is certainly interesting to think about, we feel it is out of scope of this manuscript. We do not want to discuss the design of EHRs, but rather its content.

Comment 12

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.

Response 12

We do not intent to conceptualize an ‘average EHR’. The conclusion refers to the ‘commonly available EHR data’. This average content of EHRs has been described in the results section and
quantified in table 3. However, as stated in the discussion, these results are limited to structured
data and German university hospitals. Structured data elements are offered for half of the patient
classifications required for the case trials, which is actually better than what we expected. Judging
from the literature, EHRs of hospitals in the USA are more comprehensive as the clinics have more
manpower for documentation purposes. Furthermore, regarding medical informatics, Duke Medicine
seems to be one of the more advanced hospitals in the USA. Therefore it would not surprise me if
you were able to achieve results above the German average presented in this study.

Minor Essential Revisions

Comment 13

The use of the word ‘decomposed’ to describe creation of data elements is odd. Choose ‘reviewed’
or similar instead.

Response 13

We changed the example given for ‘decomposition’ of criteria sentences in the third paragraph of the
methods section in the hope to illustrate better what is meant. We also replaced the word
decomposed with other words.

Comment 14

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.

Response 14

We are sorry, but we do not understand this comment. The studies which are cited in the
background are also discussed there (third paragraph). Basically, we used the review of Chan to
identify the scope of the existing literature, then reviewed ourselves the papers identified by Chan.
Our results are also compared to similar, but less comprehensive works in the fourth paragraph of
the discussion.

Comment 15

In the Background, define selection bias clearly as this is an IT journal.

Response 15

We replaced selection bias by simpler wording.

Comment 16

Cite Table 3 beginning with the ‘health status’ section under Results.

Response 16

Table 3 is cited right before the details on each topic group. As it is relevant for each topic group we
thought this to be the ideal position.

Comment 17
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.

Response 17

We added a reference to Luo to the abstract.

Discretionary Revisions

Comment 18

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?

Response 18

This question is not in the scope of this manuscript. For each patient characteristic, we screened the EHR for one or more data elements which contain information on this characteristic in a manner that does not necessitate interpretation. For some criteria, the physician or the study coordinator might be able to derive more information on the patient than was actually documented based on his clinical knowledge and experience. Incorporating these abilities into computerized recruitment systems is however a complex matter.

Comment 19

Add some thoughts as to why the Luo semantic group distribution differed in places compared to this study.

Response 19

We actually once had a paragraph on this in the manuscript but removed it, because it was purely speculative. Differences between Luo and us are: International vs German trials; whole sentences vs patient characteristics, random criteria vs random trials. We find it difficult to connect the differences in the distribution of the criteria to the differences in how Luo and we arrived at the numbers.

Referee 3 (Vitaly Herasevich)

Manuscript is addressing important topic described potential of computerized clinical trial recruitment.

Comment 20

The major question that I have is regarding using semantic mapping as eligibility criteria substitute. Semantic mapping is not only one method for enrolment to clinical trials using EMR data. The concept of “digital signatures” allows to do highly successful “mapping” of eligibility criteria based on combination objective and subjective clinical data.

In current manuscript It is difficult understand what exactly was mapped based on names of the trials. Table 1 is counting numbers only. Also tables 2 and 3 have not all eligibility criteria listed.
Method section is lacking details about specific enrollment criteria that probably should be added as appendix.

Based on above title of manuscript not reflects only one method used – semantic mapping. Discussion part and conclusion sections also would have benefits if describe this limitation.

**Response 20**

No substitute for the eligibility criteria was used. First the criterion sentence was broken into single pieces, then for each piece of data needed to evaluate the criterion, we searched for corresponding data elements in the EHR by (1) individual knowledge of the database administrators, (2) searching for keywords in the EHR metadata and (3) involvement of the clinical staff. If a corresponding structured data element was found in the EHR, we also calculated the fraction of patients with available data in this data element. We later mapped the criteria to categories/clusters in order to get result tables with 27 instead of 706 rows (706 if we had presented the results for each patient characteristic individually).

Example:

**Trial 1**

Does the patient suffer from disease A? – category = disease – corresponding data element found? = Yes – patients with data in this data element = 30%

**Trial 2**

Does the patient suffer from disease B? – category = disease – corresponding data element found? = No

**Trial 3**

Has the patient ever suffered from disease C? – category = disease – corresponding data element found? = Yes – patients with data in this data element = 60%

Average result for category ‘disease’ – documentable = 66%, documented = 45%, overall availability = 30%