PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	A pragmatic method for electronic medical-record-based observational studies: developing an electronic medical records retrieval system for clinical research
AUTHORS	Yamamoto, Keiichi ; Sumi, Eriko; Yamazaki, Toru; Asai, Keita; Yamori, Masashi; Teramukai, Satoshi; Bessho, Kazuhisa; Yokode, Masayuki; Fukushima, Masanori

VERSION 1 - REVIEW

REVIEWER	Chunhua Weng, PhD Assistant Professor of Biomedical Informatics Columbia University New York, NY, USA 10032 I do not have competing interests with the authors.
REVIEW RETURNED	23-Jul-2012

THE STUDY	The study lacks a clear statement of the research questions or testable hypotheses. The methods lack clarity and novelty. The overall results did not add much value to the literature.
RESULTS & CONCLUSIONS	the results are not compelling that the authors contribution worthy of publication.
GENERAL COMMENTS	This paper presents a method to reuse EHR data to improve the efficiency of clinical research. The topic has high significance. The authors also did a nice work in referencing related work. However, the paper has poor readability and needs substantial revision. The authors presented much low-level engineering technical details but failed to clearly present a generalizable and scalable method to identify patients from EHR and extract corresponding data for research uses. The concepts used by the authors such as converting free-text eligibility criteria to computable formats are not new. This paper in its current form did not add new knowledge or methods to the literature. Major points: • Claimed as one of the major contributions of this study, the process for converting free-text eligibility criteria to computable formats lacks novelty and clarity. Is the process manual or automated? How were the ICD codes identified for the disease diagnoses? How were “no BP medication administration” translated into the category of exclusion criteria? If the process is manual, the approach is not scalable and hence not generalizable. If the process is automated, there are a lot of known challenges for automating this task but the authors did not mention how they overcome these challenges. The methodology was not described in a scientific manner. • It is unclear what the author meant by “entity-level” and “attribute-level” data elements.

	• The evaluation was only one case study showing the example query without measuring the reliability and accuracy of the method. Minor points: • The authors may want to read the “instructions for authors”, which says each paper should contain no more than five tables and figures. Currently the paper contains 5 figures and 3 tables. Table 3 is very long. There is a word limit as well for regular submissions. • It is unclear why some texts are underlined. • Figure 5 contains Japanese characters that may not be understandable by international readers.

REVIEWER	Dr. Jos Aarts Senior Research Scientist Institute of Health Policy and Management Erasmus University Rotterdam Rotterdam, The Netherlands
REVIEW RETURNED	02-Aug-2012

THE STUDY	The authors write that they have applied their method of extracting eligibility data from EMRs to obtain a cohort of patients for a clinical study. It is unclear to me whether the researchers have included patients for this study in a different, i.e. more traditional way to allow comparison of two methods. I believe they didn't and then the authors should be clear about the limitations of their study. On page 22, the statistical analysis is unclear (and then I refer to my remarks above) and the authors write that the researchers were basically satisfied. But the authors do not report how many researchers they recruited for interviewing, nor is it clear how they were interviewed (structured, semi-structured or open). The authors claim that they did qualitative analyses. I understand completely (and I like it actually very much) the development and validation of the ERS model, but the authors do not specify what the qualitative analyses exactly entailed. Qualitative research also requires an accurate description of the methods, including a description how respondents (the researchers of the clinical study) were included.
GENERAL COMMENTS	I like the paper very much. It builds on solid earlier research. It reports a relevant practical tool to recruit patients for a clinical study using data in an EMR. I appreciate if authors reflect on my comments, because I think that it would make the paper stronger and publishable.

REVIEWER	Landen Bain CDISC Liaison to Healthcare US no competing interests.
REVIEW RETURNED	03-Aug-2012

THE STUDY	No statistical methods are used. This is a methodological description, not a statistical study.
RESULTS & CONCLUSIONS	The general approach of using OLAP and Data Marts is not novel. The paper does not acknowledge prior work in this area.
GENERAL COMMENTS	The use of OLAP and data marts as described is not that new and interesting. What is of interest is this statement: "Medical concepts expressed as narrative criteria are mapped onto entities in the data model and converted into entity-level criteria." The paper should delve into this mapping more.

REVIEWER	Prof. Dr. Hans-Ulrich Prokosch Chair of Medical Informatics University of Erlangen Germany no competing interests
REVIEW RETURNED	17-Aug-2012

GENERAL COMMENTS	The manuscript does not directly fit into the structure of typical medical study descriptions. It does **not** describe any new clinical research, but rather describes the development of an IT infrastructure / IT components which allow to reuse data from electronic medical records (EMR) for recruiting patients into observational studies and also to “collect necessary from EMRs for clinical research”. This is actually a very innovative and interesting challenge, which has been tackled in recent years by several research groups around the world. Typically however, publications on such projects appear in medical informatics journals, less in medical journals. Thus the first question is, if the BMJ Open would find such a publication also interesting and informative for its target audit (which I would generally support), even if it does not directly postulate a new clinical hypothesis or research question, but rather describes new approaches, how to pursue clinical research in future. In this case however, one could again aim at two different approaches to publish this new approach. a) One could require from the authors to also postulate a concrete research question, e.g. a. “Is the quality of EMR data good enough to be directly reused for clinical research? Or b. “Can automatic patient retrieval functions within EMRs improve the recruitment of patients for observational studies”? or c. What type of IT components are necessary in order to establish a system for efficiently identifying eligible patients for observational studies from EMRs? Unfortunately the authors have not stated any such research question, therefore it remains unclear for me,

	what they wanted to finally measure in order to provide quantative indications for their results. Also this lead to the fact, that no study design has directly been described in the manuscript and no outcome measures or statistical methods are described. There do exist publications in the medical informatics literature, where such study questions associated with clear outcome measures have been defined, therefore such a publication on the efficiency of new IT measures is possible. Retrospectively however, I am afraid, that such measures are not available for the work described in the manuscript. For such a purpose the manuscript would need a major restructuring and rewriting! b) If the Editors of BMJ Open however believe, that such a new IT-based approach is interesting to be presented for its readers anyway (because the readers should become at least more familiar with such new ideas), the paper could probably be published with only minor revisions. In my opinion, the authors in their currently submitted manuscript just present a nice story on a new approach established and pursued within the last years at their institution. For this, they refer to several earlier publications in which some parts of their work are probably described in more detail. **Methods** On page 13, starting in line 27 the authors write “Medical concepts expressed as narrative criteria are mapped onto entities in the data model and converted into entity-level criteria. For each entity, a criterion is created to extract patients who meet each condition.“ It remained unclear to me, if this mapping process was completely manual (which would on the long term be a heavy workload for every new research project) or any automated computer-based mapping. The examples of exclusion criteria used in the exemplary application of the ERS (page 11 bottom, page 12 top) are very simple and thus probably relatively easy to transfer into computable criteria. In many other research projects and trials inclusion and exclusion criteria are much more vague or complex. Thus, I believe that is not a good example to already underpin the functionality of the approach described for a broader set of studies. **Results**

	As an example for not applying a concret outcome measure for their “results” I would like to cite the last sentence of the RESULTS chapter “The investigators evaluating the system **mentioned that** 1) it enabled them to extract thenecessary data for diagnosis and drug administration without exception; 2) by screening the entire patient population at the hospital using the ERS, they could identify not only eligible patients in the department of oral and maxillofacial surgery but also all eligible patients, which reduced the study bias; and 3) by creating reports for confirmation, it enabled investigators to devote their time to reading images, thus effectively reducing the time required for reviewing medical records.” This means that the measures for this new approach are “impressions” from investigators (which from our research in this field are probably exactly describing advantages of the new approach), but that no real hard facts exist to prove the advantage of the new approach compared to traditional methods. Beginning at page 21 “Among the approximately 800,000 cases at our hospital 8,772 were categorised using the terms ‘Inclusion criteria: Osteoporosis diagnosis’; among this group,. . .” it seems to me, that such numbers should at least provide some hard measures, but as far as I understood, this only describes how many patients the retrieval system did categorize as belonging to the different eligibility categories. Also some numbers presented in this context seem to be inconsistent. E.g. “Among those on the targeted patient list, … 84 [were placed] under ‘Inflammatory jaw condition diagnosis’, …” (page 21, line 31) compared to “Among 72 classified under ‘Inflammatory jaw condition diagnosis’, 35 cases and 37 non-cases were identified.” In the latter sentence I did not understand with cases and non-cases. Should this be a first information on false positively identified patients? In a proper study design one would have measured true positive, false positive, true negative and false negative rates for the patient identification algorithm and thus been able to calculate sensitivity and specificity of the EMR retrieval system (ERS). **References** The citation style of the authors in my opinion is a little strange and to “broad”. Very often they do not cite literature from which they have directly cited a statement, or which would at least in its conclusion underpin a statement of the authors, but rather they cite very general literature which is somewhat related to the last sentence, but does really not have any direct linkage. E.g. page 9, lines40 to 46: “this allows for logical operations and

	creating eligible patient lists for each respective parameter in a study [51]". Reference 51 is "C.J.Date. An Introduction to Database Systems 8th ed. Boston; Addison Wesley. 2003" and thus an absolutely basic introduction into database management systems. page 9, lines 58 to page 10 line 7: "To ensure that the data retrieval process is practical and independent of the EMR system structure, a data warehouse (i.e., data mart [52]) was created on a relational database management system by extracting, transforming, and loading information from the EMR system [16-23]." Reference 52 is "W. H. Inmon, Claudia Imhoff, Ryan Sousa. Corporate Information Factory. Hoboken, New Jersey, John Wiley & Sons 2ed; 2000.", again a very general book. References 16-23 are various different publications on data warehouse technology and applications, but have no direct relationship to the sentence they follow. If there is any citation at the end of this sentence, I would have expected a reference to a publication of the authors, which would describe in more detail how they "created a data warehouse by extracting, transforming, and loading information from the EMR system." This are only three examples, but many more such "strange citations" appear in the manuscript. **Other comments** In many parts of the manuscripts the authors remain at the high level of the IT system/methodology description. From the view point of a medical informatics researcher I do have many more detailed questions on particular implementation details. I know that too much detail would not be understood by a pure medical reader. However, by remaining on such a general high level one risks to "just tell a nice story" but not provide hard measured proofs for research results. Figure 5 would probably be not helpful for a reader because of its Japanese characters on the screenshot.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Chunhua Weng, PhD

Assistant Professor of Biomedical Informatics, Columbia University, New York, NY.

1) Claimed as one of the major contributions of this study, the process for converting free-text eligibility criteria to computable formats lacks novelty and clarity. Is the process manual or automated?

Response:

The process is manual. We propose that the computable criteria should not be a result of the automated conversion of narrative criteria but rather a result of the research preparation involving medical concepts that are not expressed logically or explicitly in the narrative criteria. Therefore, we thought that much of the conversion of eligibility criteria to computable criteria should be executed manually at the protocol development stage. We have updated sentences in the MATERIALS AND METHODS section (page 13, line 12) and DISCUSSION section (page 23, line 14 to page 24, line 14).

2) How were the ICD codes identified for the disease diagnoses?

Response:

At the protocol development stage, investigators defined the list of ICD codes. We have updated sentences in the MATERIALS AND METHODS section (page 14, line 3 to line 4).

3) How were “no BP medication administration” translated into the category of exclusion criteria?

Response:

The “no BP medication administration” criterion was not translated into the category of exclusion criteria. Instead, the “Intravenous BP administrations” criterion was translated into this category (see also Table 1). We have added the schema of the test research as Figure 4 and added the following reference: “47. Yamazaki T, Yamori M, Yamamoto K, et al, Risk of osteomyelitis of the jaw induced by oral bisphosphonates in patients taking medications for osteoporosis: A hospital-based cohort study in Japan, Bone 2012: 51(5) 882–887”.

4) If the process is manual, the approach is not scalable and hence not generalizable. If the process is automated, there are a lot of known challenges for automating this task but the authors did not mention how they overcome these challenges. The methodology was not described in a scientific manner.

Response:

This research aimed to demonstrate an example of a hospital-based cohort study that identified patients and exposure with an ERS and to evaluate the feasibility and usefulness of the method. Though the process is manual, we consider our method to be generalisable. We have updated sentences in the ABSTRACT section (page 3, line 6 to line 9) and OBJECTIVE section (page 8, line 15 to line 18).

5) It is unclear what the author meant by “entity-level” and “attribute-level” data elements.

Response:

To convert computable criteria appropriately, the investigators are needed to make high-level medical decisions concerning the research question of the protocol. Therefore, we thought that the tasks to convert eligibility criteria to computable criteria should be divided into attribute-level tasks and entity-level tasks that require higher medical decisions. To emphasise this claim, we have updated sentences in the MATERIALS AND METHODS section (page 10, line 4 to line 10) and DISCUSSION section (page 23, line 14 to page 24, line 9).

6) The evaluation was only one case study showing the example query without measuring the reliability and accuracy of the method.

Response:

Because this research aimed to demonstrate an example of a hospital-based cohort study with our ERS, we did not compare our method with other methods. We have included this limitation of this research in the DISCUSSION section (page 21, line 18 to page 22, line 1).

7) The authors may want to read the “instructions for authors”, which says each paper should contain no more than five tables and figures. Currently the paper contains 5

figures and 3 tables. Table 3 is very long. There is a word limit as well for regular submissions.

Response:

We agree with your assessment. We reviewed the figures and tables and eliminated the unnecessary items.

8) It is unclear why some texts are underlined.

Response:

We eliminated the unnecessary underlining.

9) Figure 5 contains Japanese characters that may not be understandable by international readers.

Response:

We eliminated the figure to improve clarity and prevent confusion.

Reviewer 2: Dr. Jos Aarts

Senior Research Scientist of Institute of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, Netherlands

1) The authors write that they have applied their method of extracting eligibility data from EMRs to obtain a cohort of patients for a clinical study. It is unclear to me whether the researchers have included patients for this study in a different, i.e. more traditional way to allow comparison of two methods. I believe they didn't and then the authors should be clear about the limitations of their study.

Response:

We agree with your assessment. Because this research aimed to demonstrate an example of a hospital-based cohort study with our ERS, we did not compare our method with other methods. To emphasise this claim, we have included this limitation of the research in the DISCUSSION section (page 21, line 18 to page 22, line 1).

2) On page 22, the statistical analysis is unclear (and then I refer to my remarks above) and the authors write that the researchers were basically satisfied. But the authors do not

report how many researchers they recruited for interviewing, nor is it clear how they were interviewed (structured, semi-structured or open). The authors claim that they did qualitative analyses. I understand completely (and I like it actually very much) the development and validation of the ERS model, but the authors do not specify what the qualitative analyses exactly entailed. Qualitative research also requires an accurate description of the methods, including a description how respondents (the researchers of the clinical study) were included.

Response:

We have provided more information about the interview. Additionally, because the interview did not contain facts to prove the advantage of the new approach compared to traditional methods, we moved these sentences to the DISCUSSION section (page 21, line 4 to line 16).

Reviewer 3: Landen Bain
CDISC, Liaison to Healthcare, US

1) No statistical methods are used. This is a methodological description, not a statistical study.

Response:

We agree with your assessment. We updated the description of the research design in the Abstract to now say “System development and evaluation” (page 3, line 6 to line 9).

2) The general approach of using OLAP and Data Marts is not novel. The paper does not acknowledge prior work in this area. The use of OLAP and data marts as described is not that new and interesting. What is of interest is this statement: “Medical concepts expressed as narrative criteria are mapped onto entities in the data model and converted into entity-level criteria.” The paper should delve into this mapping more.

Response:

We described the mapping process in detail. We have updated sentences in the MATERIALS AND METHODS section (page 10, line 4 to line 10, page 14 line 7 to line 13) and in the DISCUSSION section (page 23, line 14 to page 24, line 14).

Reviewer 4: Prof. Dr. Hans-Ulrich Prokosch PhD.

Chair of Medical Informatics, University of Erlangen, Germany

1)Method - On page 13, starting in line 27 the authors write “Medical concepts expressed as narrative criteria are mapped onto entities in the data model and converted into entity-level criteria. For each entity, a criterion is created to extract patients who meet each condition.” It remained unclear to me, if this mapping process was completely manual (which would on the long term be a heavy workload for every new research project) or any automated computer-based mapping.

Response:

We described the mapping process in detail. To convert computable criteria appropriately, high-level medical decisions are required. Therefore, we thought that large numbers of the conversions of the eligibility criteria to computable criteria should be manually executed at the protocol development stage. To emphasise this claim, we have updated sentences in the MATERIALS AND METHODS section (page 10, line 4 to line 10) and the DISCUSSION section (page 23, line 14 to page 24, line 14).

2) The examples of exclusion criteria used in the exemplary application of the ERS (page 11 bottom, page 12 top) are very simple and thus probably relatively easy to transfer into computable criteria. In many other research projects and trials inclusion and exclusion criteria are much more vague or complex. Thus, I believe that is not a good example to already underpin the functionality of the approach described for a broader set of studies.

Response:

Because this test research was a pharmacoepidemiological study, the eligibility criteria were simpler than those of clinical trials. However, in the test research, we defined the entity-level criteria according to the entered diagnosis and ordered treatments rather than the diagnostic criteria of the disease according to the actual EMR use. The eligibility criteria include many medical concepts that are not stated explicitly, and the computable criteria are not simply converted from the eligibility criteria. Thus, we believe this study provides a good example of computable criteria.

3) Results - As an example for not applying a concrete outcome measure for their “results” I would like to cite the last sentence of the RESULTS chapter “The investigators evaluating the system mentioned that 1) it enabled them to extract the

necessary data for diagnosis and drug administration without exception; 2) by screening the entire patient population at the hospital using the ERS, they could identify not only eligible patients in the department of oral and maxillofacial surgery but also all eligible patients, which reduced the study bias; and 3) by creating reports for confirmation, it enabled investigators to devote their time to reading images, thus effectively reducing the time required for reviewing medical records." This means that the measures for this new approach are "impressions" from investigators (which from our research in this field are probably exactly describing advantages of the new approach), but that no real hard facts exist to prove the advantage of the new approach compared to traditional methods.

Response:

We agree with your assessment. The interviews did not provide hard facts to prove the advantage of the new approach compared to traditional methods. We moved these sentences to the DISCUSSION section (page 21, line 4 to line 16).

4) Beginning at page 21 "Among the approximately 800,000 cases at our hospital 8,772 were categorised using the terms „Inclusion criteria: Osteoporosis diagnosis"; among this group,. . ." it seems to me, that such numbers should at least provide some hard measures, but as far as I understood, this only describes how many patients the retrieval system did categorize as belonging to the different eligibility categories. Also some numbers presented in this context seem to be inconsistent. E.g. "Among those on the targeted patient list, … 84 [were placed] under „Inflammatory jaw condition diagnosis", …" (page 21, line 31) compared to "Among 72 classified under „Inflammatory jaw condition diagnosis", 35 cases and 37 non-cases were identified." In the latter sentence I did not understand with cases and non-cases. Should this be a first information on false positively identified patients? In a proper study design one would have measured true positive, false positive, true negative and false negative rates for the patient identification algorithm and thus been able to calculate sensitivity and specificity of the EMR retrieval system (ERS).

Response:

This test research aimed to estimate some risks related to the oral BP related osteonecrosis of the jaw. In this test research, two oral and maxillofacial surgeons diagnosed cases by chart review. They examined a total of 1,986 patients who were categorised with an "Inflammatory jaw condition diagnosis", "Other suspicious disease

diagnosis” and the "Exclusion criteria (medical records review): radiation therapy to the maxillofacial". Thus, we calculated only the ICD10 code (osteomyelitis of the jaw) sensitivity for which we could determine true values. We have added figure 4 to show the schema of data collection and the confirmation of our research.

5) The citation style of the authors in my opinion is a little strange and to “broad”. Very often they do not cite literature from which they have directly cited a statement, or which would at least in its conclusion underpin a statement of the authors, but rather they cite very general literature which is somewhat related to the last sentence, but does really not have any direct linkage.

If there is any citation at the end of this sentence, I would have expected a reference to a publication of the authors, which would describe in more detail how they ”created a data warehouse by extracting, transforming, and loading information from the EMR system.”

Response:

We appreciate this helpful suggestion. We reviewed and updated the references. Additionally, we have updated sentences about our ERS architecture in the MATERIALS AND METHODS section to provide more detail (page 10, line 17 to page 11, line 5).

6) In many parts of the manuscripts the authors remain at the high level of the IT system/methodology description. From the view point of a medical informatics researcher I do have many more detailed questions on particular implementation details. I know that too much detail would not be understood by a pure medical reader. However, by remaining on such a general high level one risks to “just tell a nice story” but not provide hard measured proofs for research results.

Response:

We have updated sentences about our ERS architecture in more detail. Additionally, we updated the research aim to demonstrate an example of a hospital-based cohort study that identified patients and exposure with an ERS. We propose that, although our manuscript does not provide hard measured proofs for research results, our manuscript is still useful.

7) Figure 5 would probably be not helpful for a reader because of its Japanese

characters on the screenshot.

Response:
After a review, we have eliminated this figure.

VERSION 2 – REVIEW

REVIEWER	Dr. Jos Aarts Senior Research Scientist Institute of Health Policy and Management Erasmus University Rotterdam Rotterdam, The Netherlands
REVIEW RETURNED	03-Oct-2012

GENERAL COMMENTS	I compared the revised version of the paper with the original submission. I had little comments then. As far as I can see the authors have responded appropriately to the comments of the reviewers.