Author's response to reviews

Title: Detecting inpatient falls by using natural language processing of electronic medical records

Authors:

  Shin-ichi Toyabe (toyabe@med.niigata-u.ac.jp)

Version: 4 Date: 26 November 2012

Author's response to reviews: see over
Editor,
BMC Health Services Research

September 26, 2012

Dear Editor,

Thank you very much for giving us an opportunity to revise our manuscript entitled “Detecting inpatient falls by using natural language processing of electronic medical records”, which was previously submitted to BMC Health Services Research.

We have revised the manuscript in accordance with the comments by the referees and have also provided point-by-point replies to questions from the referees in this letter. All of the corrections in the manuscript have been indicated by underlines. In addition, we have had our manuscript checked by a native English speaker according to your advice.

Sincerely yours,

Shin-ichi Toyabe, MD
Professor, Niigata University Crisis Management Office,
Niigata University Medical and Dental Hospital,
Asahimachi-dori 1-754, Chuo-ku, Niigata 951-8520, Japan
E-mail: toyabe@med.niigata-u.ac.jp
Responses to Reviewer 1

*Discretionary Revision: The title of Fig. 1 is not updated. "Sensitivity" -> "Data-containing rate".*

Response: Thank you for your indication. I rewrote the description.
Responses to Reviewer 2

Major Compulsory Revisions 1: The paper reports the need for “quick identification of injurious falls” and “detecting adverse events in a timely and cost-effective way”, and proposes to “promptly detect severe injuries after falls by using the NLP method”. It is not clear what the purpose or benefits of quick identification is? Are you trying to detect falls immediately as it occurs, which makes sense in terms of having appropriate care for the patient. However, this will not be the case with the proposed NLP system as the incident will have to be documented and subsequently scanned by an NLP system to detect the occurrence of falls. As a result, the author needs to clarify why the lag time is such an important factor e.g. why the current lag time exhibited in the study data is not suitable.

Reply: Injurious falls often result in complaints from patients and their families and can sometimes cause litigation. Therefore, the patient safety section or risk management section of a hospital encourage medical staff to report these cases as soon as possible. Therefore, the purpose or benefits of quick detection of the injurious falls is not only for appropriate care for the patients but also for risk management of the hospital. The lag time in reporting is known to be a significant problem in patient safety. It reflects the patient safety climate of the hospital (reference [6]).

Major Compulsory Revisions 2: Thank you for including the overview of the workflow required to enter data into the hospital information system (HIS). It is however not explicit whether the image order entry and discharge summaries are part of the HIS or EMR from reading the description? The first sentence in the “Data collection and ethical consideration” section also aims to address this but did not successfully achieve it. However, from reading the remainder of the paper, it’s inferred that the progress notes and discharge summaries are from the EMR, while incident reports and image order entries are from the HIS. Assuming that this is the case, it seems like that progress notes and discharge notes will never be entered into the HIS, and as a result, how would one derive the lag time and degree of harm for such data sources? The purpose of including data sources from EMRs in the study needs to be justified to make the paper coherent. These system and workflow constraints are important when discussing the implications of lag time and degree of harm in the paper.
Reply: I apologize for the difficulty in understanding the relationships between HIS, EMR, order entries and other systems. I added a description about the organization of a hospital information system.

Large Compulsory Revisions 3: Regarding the NLP rules development, the set of syntactic rules were developed from a development set of incident reports, but was then applied to all other data sources to detect falls. The writing style and contents of an incident report would be vastly different to any of the other data sources in the study. Was there a reason for pursuing such a methodology which would ultimately tailor the NLP algorithm to incident reports?

Reply: In order to perform NLP, a large amount of text corpus specific for inpatient falls is necessary. Incident reports were best suited for a such purpose, and there were no such data sources other than incident reports. In addition, the written style about inpatient falls was not so different among incident reporting, order entries, progress notes and discharge summaries.

Major Compulsory Revisions 4: The evaluation of the system still uses a varied set of measures between the experiments. Noteworthy is that specificity was omitted when analyzing the NLP system against the various data sources (excluding incident reports). Can this measure be included for completeness? Otherwise, justification as to its omission would be beneficial to the reader.

Reply: I added data on specificity when analyzing various data sources to the Results section.

Major Compulsory Revisions 5: Another issue with the results from “Fall events detected from each data source” is that there’s not enough information for the reader to analysis the results. The number of reports for each data source needs to be reported. Furthermore, some of the reported measures do not look right, for example, the image order entries had 10 fall events, however the sensitivity was 0.83 (i.e. you can only get 8 out of 10 recall, but not 8.3 out of 10). Showing the confusion matrix for the 4 data sources would overcome these issues and make it clearer.

Reply: I added two-dimensional tables that showed the number of events from each data source in order to make it possible for readers to recalculate the indexes. As for the
sensitivity of 0.83, I think that you might have misunderstood the results. The total number of fall events in August 2010 was 80, and diagnostic imaging was ordered in 12 of those 80 events. However, only 10 events were judged as positive by the NLP method. The two other events were judged as negative, since information on fall events was not included in the order forms of the two events. Therefore, sensitivity was calculated to be 0.83 (10 out of 12 events), and data-containing rate was calculated to be 0.125 (10 out of 80 events).

Major Compulsory Revisions 6: The degree of injury and lag time between falls and submission of data to hospital information system can be determined regardless of the NLP application. As a result, I would have liked to see a profile of report types against the degree of harm and lag time. However, only incident reports and image order entries were used. How could this be done for other data sources? It just so happens that the selected reports for lag time and degree of harm analysis were both from the HIS. What if the NLP system performed better on other data sources?

Reply: In the analysis of degree of injury and lag time, the NLP method was used in order to detect as many injurious falls as possible. In contrast to the former clause, this analysis was performed using medical records for one year because the number of falls with moderate to severe injury was small. However, it is extremely hard both in time and cost to perform a manual chart review of medical records (a gold standard) for a year. I had no thought to perform it.

Major Compulsory Revisions 7: The reporting of results for the “Comparison of incident reports and image order entries as text data sources” involved no gold standard comparison. We have no idea whether the detected falls were actually correct or the extent of the number of missed falls. The same methodology of analysis would have been better on the data used for “Fall events detected from each data source” where a gold standard did exist.

Reply: Whether the detected falls were actually correct was confirmed by checking the corresponding medical chart. I added the information to the Methods section. As you mentioned, the number of missed falls was unknown because of the lack of a gold standard. However, again, it is extremely hard both in time and cost to perform a manual chart review of medical records for one year.
Major Compulsory Revisions 8: The lag time was found to be significantly shorter for image order entries than incident reports. Since incident reports contain numerous mild incidents, this could cause biases towards the reported results. It would be good to compare the lag time separately for each degree of harm. It could be the case that the lag times for severe cases are much shorter than those that are mild from incident reports.

Reply: I appreciate your advice and comments. According to your advice, I added Table 2 where I showed lag time in the two data sources separately for each degree of harm.

Major Compulsory Revisions 9: It was reported in the Background section (1st paragraph) that the “judgment about adverse events in medical charts depends on the skills of reviewers”. Elaborate on this finding and determine the skill level of the single reviewer that was used in the paper. Justification based on the skill level of the single reviewer is required so that you have evidence to support the use of only one reviewer in the study. Otherwise, the evaluations of the NLP system may not be conclusive.

Reply: I added information on the single reviewer who performed the manual chart review in the Methods section. The reviewer has enough experience and capability in medical record audits and adverse event analysis, because he has had a career in the Department of Medical Information and in the Department of Patient Safety.

Minor Essential Revisions 1: Abstract (paragraph 3): “imager order entries” should be “image order entries”

Reply: Thank you for your indication. I corrected.

Minor Essential Revisions 2: Abstract (paragraph 3): Elaborate and clarify on the purpose of selecting incident reports and image order entries as possible data sources. Was the purpose for reducing under-reporting, delay of submission, and/or for detecting serious falls?

Reply: The purpose of this study was to develop methods to detect serious falls promptly, which I have indicated in the Background section.

Minor Essential Revisions 3: Background (paragraph 2): What were the performances
of the “sophisticated NLP algorithms”? This is worth noting to compare against the proposed “simpler” NLP algorithm.

Reply: I deleted the expression “sophisticated”.

Minor Essential Revisions 4: Methods (NLP of free-text and construction of syntactic category rules): Elaborate on what an “antero-posterior relation” means? This is a medical term used in the context of an NLP relation between morphemes. In the second example, it was rather difficult to comprehend since it has two interpretations. Revise the example and maybe mention that “falling” is an example of “find a situation”.

Reply: I explained what the sign “->” means more precisely. I revised the second example of syntactic category rules.

Minor Essential Revisions 5: Methods (Comparison of lag-time and degree of harm): How many image order entries were used? The degree of injury can be readily determined for the incident report from the HIS (see Methods – Data collection and ethical consideration), however, how was the degree of injury determined for the image order entries?

Reply: The number of image order entries was shown in Table 1, as indicated in the reply to Major Essential Revisions 5. The degree of injury of cases found by image order entries can be easily determined by the corresponding description for EMRs.

Minor Essential Revisions 6: Results (Fall events detected from each data source): The number of patient days was used for number of patients. Detail both or be specific as to what the value refers to.

Reply: I added a description about the number of patients admitted during the period.

Minor Essential Revisions 7: Results (Fall events detected from each data source): The location of Fig 1 reference was inserted when discussing the performance of the NLP system on image order entries. This should either be at the start or end of the results being reported.

Reply: I moved the location of Fig 1.

Minor Essential Revisions 8: Results (Comparison of incident reports and image order
entries as text data sources): The number of mild events found from incident reports, order entries or both does not add up to the total of mild events. Also check the correctness of the number of degrees of injury in fall events found by incident reports and image order entries. This should be reviewed for its correctness.

Reply: I showed the number of fall events based on their degree of harm in Table 2, as indicated in the reply to Major Essential Revision 8.

Minor Essential Revisions 9: Results (falls with moderate to severe injuries): Mention that the statement “image order entries increased the detection rate of falls” was for “moderate to severe injuries”.

Reply: I rewrote the description.

Minor Essential Revisions 10: Results (falls with moderate to severe injuries): Mention that the statement “image order entries increased the detection rate of falls” was for “moderate to severe injuries”. Also the statement “Severe injuries caused by falls included …” should be “Moderate to severe injuries …”.

Reply: I rewrote the descriptions.

Minor Essential Revisions 11: Results (falls with moderate to severe injuries): It would be good to know the 3 injuries that were only from image order entries and the 3 that were from only the incident report. This will help identify the types of moderate to severe injuries that can be found in one but not the other, and consequently provide input towards the best data sources for detecting severe injuries.

Reply: I appreciate your comments and advice, but I could not find any apparent differences between the injuries that were only from image order entries and the injuries that were from only incident reports. Actually, x-ray examination or computed tomography was performed in the three injuries that were from only incident reports. However, no information on patient falls but only information on injuries by the falls was inputted in the order forms in these cases. Consequently, these injuries could not be found from image order entries by using the NLP method. I added the information to the Results section.

Minor Essential Revisions 12: Discussion (paragraph 3): Correct “NPL”; “loss <of>
Minor Essential Revisions 13: Discussion (paragraph 3): Elaborate on what was meant by “positive results were obtained in only 3.0%”?  
Reply: The number was derived from preliminary results that were obtained from data other than the dataset analyzed in this study. If the dataset used in this study was applied, positive results would be obtained in 6% of data that were determined as having information on fall events by using the NLP method. As shown in the Results section (Fall events detected from each data source), PPV to detect fall events was 0.06 when the NLP method was applied to progress notes. I have corrected the description.

Minor Essential Revisions 14: Discussion (paragraph 4): Revise “This report is probably the first…” to “To the best of our knowledge this is the first …”  
Reply: I rewrote the description.