Author's response to reviews

Title: Identification of Pneumonia and Influenza Deaths using the Death Certificate Pipeline

Authors:

Kailah Davis (kailah.davis@utah.edu)
Catherine Staes (Catherine.staes@hsc.utah.edu)
Jeff Duncan (jduncan@utah.gov)
Sean Igo (sean.igo@utah.edu)
Julio C Facelli (julio.facelli@utah.edu)

Version: 2 Date: 9 December 2011

Author's response to reviews: see over
Dear Editors,

Please find attached our revised manuscript entitled “Identification of Pneumonia and Influenza Deaths using the Death Certificate Pipeline” for your consideration for publication in BMC Medical Informatics and Decision Making. We would like to thank the reviewers for their feedback; we have tried to address their concerns as follows:

**Reviewer I:**

*The description of the methods is somewhat hard to follow which would make it difficult for others to try to replicate this work. Improving the clarity of this section would help a lot.*

- As requested by the reviewer we have updated the method section and provided more details and figures to aid in the clarity of the section.

*I think it would help the readers to show some real life examples of de-identified free text death certificate reports and their associated output codes to help put this into a more concrete context. Also, more descriptions of how the data were transformed or modified at each step along the pipeline process would help.*

-We have modified the Methods section and have provided a snapshot of a paper and electronic death certificate, as well, as a table showing a free text and its output when coded using MetaMap.

[1] Page 5: "However, even these data [are] delayed by approximately 2-3 weeks . . ."  
--> Why is it delayed? Is it due to the need for hand coding?

-We have addressed the questions regarding the time lag from between time of death and reporting and have included an explanation in the Background section.

[2] Page 13: --> The key-word searching approach was very simplistic in terms of the choice of words/phrases. Do those who enter the data into the death certificates ever use abbreviations such as 'PNA' for 'pneumonia', or 'flu' for 'influenza'? In other words, was this key-word search truly a fair comparison to your NLP approach? How were your key-words selected?

-We believe the method for keyword-search is truly a fair comparison to our NLP approach because we followed the process utilized by the Utah Department of Health. This detail has been added to the ‘Comparison Methodologies’ section.

[3] Page 13: "we considered a match if the unique identifier was found by our system/pipeline or keyword searching’s dataset and also the reference standard’s dataset"? --> What if you found extra cases that were missed by the reference standard?

-We considered the reference standard to be the “truth”, therefore, cases that were a potential miss by the reference standard would be counted against the comparator.
[4] Page 14: "SPURIOUS (death found by either method. . .) -->Were both methods lumped together (as suggested by the word 'either') or were the analyses done separately but in parallel for the key-word search and the NLP method?

-The analyses were done separately; the sentence was revised to improve clarity.

[5] Page 15: "Of the 21 false positives, 6 records correctly mentioned pneumonia in the cause of death but their corresponding ICD-10 codes failed to provide any code related to pneumonia, while 2 records were flagged because it included the sub-string 'pneumonia'.

---> If the corresponding ICD-10 codes did not code for pneumonia, does this mean that the state had coded the data wrong? Was your system essentially more correct than the gold standard in this case?

---> Why did the substring 'pneumonia' result in a false positive? What was the full string/context that led to the false positive? In other words, in what context(s) is the term 'pneumonia' used when it really is not pneumonia?

-We have addressed the questions regarding the false positives in the Results section and have provided examples to help readers better understand why these cases were false positives.

[6] Page 15: --> For the false negatives in which a misspelling of pneumonia occurred, was any spelling correction done on that data as well as part of the pipeline process?

- The Utah Department of Health does not perform spelling corrections when searching for pneumonia and influenza deaths; therefore, no spelling correction was done to the data used for keyword search.

[7] Page 15: --> The Keyword search section reports identifying 473 records as pneumonia and influenza. The DCP approach also identified 473 cases. Was this just a coincidence? Were they the same 473 cases using both approaches? If not, how much overlap was there?

-We thank the reviewer for this observation; however, the actual number of cases for DCP was 480. We have corrected this typo and also provided information in the results section on the number of overlap cases between the two methods.

[8] Page 16: --> The analysis of failures is useful to read about. It would be improved with a little more information. For example, when "six cases were caused by the string 'pneumonia' being present in the death literals" but not coded into ICD-10, was this due to a failure of MetaMap? If not, how did this occur? --> It is mentioned that aspiration and pneumonia might have been mentioned in separate text fields. What are these separate text fields? I did not find a clear description of why fields might be separate. Did the different text fields have different meanings?

-The Analysis of failures section was modified and added a paragraph of discussion regarding possible reasons for the missed six cases.
Page 18: ". . . the ability to expand abbreviations . . ." -- This is the first time abbreviations were mentioned in the paper. Was it found to be a major problem? The topic should appear elsewhere for it to logically be discussed in this section. What abbreviations did you come across?

After analyzing the subset for abbreviation we found that there were no abbreviations for pneumonia or influenza. Through communication with the Utah Department of Health it was revealed that abbreviations are no longer a major problem on death certificates, therefore, we have deleted this as a future work.

There are multiple small typographical and grammatical errors scattered throughout the paper. Please have someone read through it very carefully make sure these are all corrected.

The English has been carefully proof read in the revised version.

Page 6: --> The phrase MICAR is used twice towards the bottom of the page. I believe the first should be SuperMICAR.

The abbreviation has been corrected.

Page 12: --> Please expand or explain acronyms such as MRCONSO and MRREL, as their meaning and significance is not clear here.

MRCONSO and MRREL are table names in the UMLS database. We have provided more details about these tables and two tables showing sample rows.

Page 14: "CPU processing time was 0.21 seconds"

For the entire data set? Per record? Based on the rest of the text it seems as though this was for the entire dataset, but it should be made more clear. Additionally, my limited experience with MetaMap is that it can be a bottleneck in terms of processing time. MetaMap was able to process 6450 records in less than 5 seconds? Or was some type of pre-processing done as well?

We have improved the clarity of this section and added information regarding the processing time for MetaMap.

Page 15: "I false negative for the death literal stating 'recurrent aspiration with pneumonia'" --> Was this a false negative or a false positive? It seems like it would have been a false positive if the phrase pneumonia was identified outside the context of "aspiration pneumonia" which was not supposed to count towards the pneumonia cases. Furthermore, was it MetaMap that couldn't handle the phrase "aspiration with pneumonia" properly?

'Recurrent aspiration with pneumonia' was indeed a false negative. The coded provided for this record included J189, which is an appropriate pneumonia or influenza code. MetaMap, however coded it with aspiration. As a result it was missed by our system.
Page 15-16: "thus yielding a recall at 99.8%, being significantly less than keyword searching" --> How was this calculated? Is there p-value? Statistically significant difference or a meaningfully significant difference? The keyword approach doesn't seem too bad! This is related to the statement in the discussion (page 16) which states, "The Death Certificate Pipeline developed here outperformed keyword searching and consistently has a high recall and precision..." It does seem based on the results that it did perform better than keyword searching, but keyword searching did relatively well in this comparison. This is also related to the sentence on page 17 "The sample size was sufficient to show [a] difference between the two methods." Again, statistically significant, but a practically significant difference?

-Fisher’s exact test has been added to the Results section to report the statistical significance for the recall and precision. A McNemar’s test was also conducted to evaluate the significance between the two methods.

Page 15: --> In the discussion of the false positives and negatives, it would be helpful for the reader to provide at least one small example of each so that we can better understand the true context.

-We thank the reviewer for this suggestion and we have revised the explanations of false positive and negatives to help better understand the context.

Page 18: "The generalizability of the findings is limited because the death records were only from one institution" --> Is there generally a lot of variability in death records from one institution to another? Or are they often very similar? How variable are the descriptions in each report? A little more summary data would be helpful here to help the reader put some of this into context: How many words or characters are in the typical text descriptions? How many text fields exist?

-Death certificates are very standard across states; however, institutions differ in their death registration process and reviewing of death certificates. We have revised this sentence and provided further explanation regarding 'generalizability' as a limitation.

Figures 1 and 2 --> These figures are almost too "high level" and a little more detail would be helpful. For example, in figure 1, more detail about the actual data transformations at each step would be informative.

-Figure 1 shows the workflow for the Mortality Medical Death System (MMDS) which is controlled by NCHS, therefore, it would be difficult to show how the data was transformed at each step. However, we did modify the figure to provide more details for each step. Figure 2 was unchanged for we added tables and figures throughout the Methods sections which provide more information for a particular step.
[19] Page 8:
Change "we used the codes produced by this system as the 'true' when comparing. . . "
To "we used the codes produced by this system as the 'gold standard' when comparing. . . "

-The sentence mentioned above was corrected.

[20] Page 11: "ASPELL performed better on three areas of performance which were evaluated. -
-> What were the three areas?

-We addressed the question regarding ASPELL outperforming GPSELL in the Methods section and provided a citation on the evaluation of ASPELL.

[21] Page 13: --> It would be helpful to provide a few examples of the rules so that the reader can get a sense of what these rules looked like, or how complex there were.

-We added a figure to the Methods section showing an excerpt of the rules.

[22] Page 14: "Correct deaths are true positives" --> Why not just call them "true positives" in the text itself rather than calling it "Correct" and then defining it as a true positive?

-We thank the reviewer for this suggestion and have changed the wording.

[23] Page 17: "The simple keyword search method not only decreased recall and precision but also reduced the level of agreement" --> Based on what you found in terms of the errors, could the keyword approach be expanded/modified a bit to achieve almost comparable accuracy with little extra effort?

-The words utilized by UDOH were chosen through a rigorous process to identify frequent words on the death certificates for pneumonia and influenza. Therefore, we can only speculate that modifying the keyword approach can increase accuracy.

[24] Page 17: "Most of the concepts the system did not identify had either the aspiration text in another field or pneumonia was mentioned in the cause of death but not coded." --> For the latter issue, is this a problem with your system or simply a coding error in the official "gold standard". In other words, was your system actually picking up cases that the official approach had missed?

-This problem was due to how the cause of death was reported on the death certificate by the physician. We have added a paragraph in the Discussion section addressing this problem.
Reviewer II

1. In the Background section, the authors should include a short discussion as to why the pneumonia and influenza category is used for surveillance. It is worth noting that only a relatively small fraction (about 10% or so) of the pneumonia deaths in this category are actually due to influenza. See Thompson et al (2003) in JAMA, vol 289, no 2. However, the non-influenza-related pneumonia deaths tend to be stable from year to year and fluctuations in this category are largely driven by the prevalence and severity of seasonal influenza. As a result, the P&I category is an important sentinel indicator.

-We thank the reviewer for directing us towards this reference. We have added the reference mentioned by the reviewer and updated the Background section to provide an explanation as to why the category “pneumonia and influenza” (P & I) is used for influenza surveillance.

2. The authors suggest that their purpose in developing this method is to bypass the lengthy coding process (p. 5). The process for coding cause of death itself is actually not particularly time consuming. The Mortality Medical Data System (MMDS) can actually code P&I deaths in real time. P&I deaths are rarely found among the MMDS rejects (those that need manual coding) and so these will not tend to be delayed by the time needed for manual intervention. The problem is not the coding process, per se, but rather the fact that the vital statistics systems designed to collect and process mortality data (which include coding cause of death) are not configured for surveillance. Thus, MMDS is not brought to bear until later in processing. This is why the coded cause of death data are not typically available for months. In theory...to my knowledge, this has not been tested...coding by MMDS could be done immediately, resulting in codes being assigned the same day the death is reported. All that said, because NCHS controls the application of MMDS to state mortality data, such configuration may not be practical or timely. Thus, the method proposed by the authors may be the best solution...at least in the short term. I suggest that the authors do some rewording to make it clear that it is the configuration of current systems that results in the delay of coded data.

-We thank the reviewer for this explanation regarding the delay in the coding process; we have reworded this paragraph to clarify that the delay is due to the NCHS process and not the actual coding using MMDS.

3. I am also unaware of any publications discussing the overall accuracy of coding using MMDS (p. 7). That said, NCHS does extensive and rigorous QC and has done so for many years. Error rates are very low. I suggest the authors contact NCHS for additional information and to confirm that there are indeed no publications.

-We thank the reviewer for this suggestion and have contacted NCHS for information on the overall accuracy of MMDS, the information provided by NCHS and appropriate references have been added to the revised version.
4. On p. 7 and 8, the authors mention the European Union countries as endorsers of MMDS. This is not the case. The European Union endorsed ACME as the standard for selecting the underlying cause of death, not MMDS or its other components. SuperMICAR cannot be used for processing in non-English languages. I wouldn’t mention South Africa. They have indeed been using MMDS, but not correctly…this is currently being addressed. Canada and the UK are also users of MMDS.

-We thank the reviewer for this clarification and have deleted the countries (European Union and South Africa) from this sentence and have included Canada and the United Kingdom as MMDS users.

5. On p. 6, the authors state that cause of death is “…coded into International Classification of Diseases Tenth Revision Clinical Modification…” The Clinical Modification (CM) of the ICD is not used for mortality coding. The version of the ICD used is the Tenth Revision (ICD 10)...period. The CM version is used (or rather will be, in 2013…ICD-9-CM is currently in use) for clinical data and billing. Reference 29 should be cited here as well.

-We have updated the sentence and included the appropriate reference.

6. It is not clear whether the authors are using underlying cause P&I or any mention of such on the death certificate. I believe the current case definition in use by CDC includes any mention…not just underlying cause. The authors need to make this clear. Also, in comparing codes generated using the authors’ method with those generated by MMDS, the authors need to be clear which set of MMDS codes they are using…entity axis (generated by MICAR) or record axis (generated by TRANSAX). In this case, the entity axis codes are most appropriate as they best reflect what was reported by the certifier. The record axis codes include some combination codes and therefore reporting of pneumonia may be missed. E.g., pneumonia reported with COPD are combined and coded to J44.0 – Chronic obstructive pulmonary disease with acute lower respiratory infection. Thus, while a pneumonia code will be found in the entity axis, it will not be found in the record axis. This, or similar code linkages, may account for the 6 false positives in which pneumonia was mentioned on the death certificate but where pneumonia codes were missing from MMDS.

-We thank the reviewer for this explanation and through communication with UDOH confirmed for our analysis we only had access to the record axis codes. We have updated the Methods section to provide clarification regarding which codes were used for this study.

We hope that with these corrections you will be able to accept the manuscript for publication in BMC Medical Informatics and Decision Making. Many thanks for your consideration

Sincerely,
Julio Facelli, PhD
Professor and Vice-Chair
Department of Biomedical Informatics