Reviewer's report

Title: A method for inferring medical diagnoses from patient similarities

Version: 3 Date: 18 May 2013

Reviewer: James J Cimino

Reviewer's report:

I found this an intriguing idea - obvious in theory but you have made a number of interesting methodologic choices to make it practical and reproducible. Nevertheless, there are problems with the paper that make it hard for me to tell if this really worked.

Major Compulsory Revisions:

1) Motivation: The whole premise of the paper is that making an early diagnosis is an important task. This can certainly be true, but it is not always true. Sometimes the diagnosis is straightforward and the task is treatment. Other times the diagnosis is never known and so cannot be confirmed. Early diagnosis would be really useful for the cases in between. I don't know how often those occur nor how you define such situations in your data set. I can't tell if you know when the correct diagnosis is made during an admission (- the discharge ICCD9 codes don't tell this part of the story). I can't tell if some of your training data included admissions of patients that also had admissions in the test set - this might make diagnoses an easy task if the patient was admitted twice for the same diagnosis. I therefore think there needs to be a clearer statement of when early diagnosis is important, how your method can be applied to these situations, how you know when the actual diagnosis was made (early or late) and how you will know whether your prediction was right.

2) Organization: I started to get lost towards the page of Methods and was going to try to go back and dig deeper, but then I found that the Results and Discussion section was not only too confusing for me to follow, but it was full of methods. Frankly, I gave up on making extra effort to understand at this point. I feel the author needs to at least meet me half way, and this dense text is just too much for me to digest. I therefore feel the paper needs major reorganization to separate methods, results, and discussion (interpretation of the results).

3) Examples: This paper would be much more interesting with some real cases of things that were difficult diagnostic cases in which application of your method would suggest the actual diagnosis faster than the clinicians did.

Minor Essential Revisions:

4) The term "electronic health record" (EHR) is generally preferred to "electronic medical record".
5) In the background section, you say "inferring patient diagnostics" - I think you mean "diagnoses".

6) Also in the Background section: Homer Warner's Iliad system was actually based on the use of EHR data to suggest individual patient diagnoses. Your reference [9] seems a bit out-dated (2002) for arguing that something is still unexplored.

7) In your Methods section, in the "sanity check" paragraph, I don't understand what you mean by "enriched".

8) In the Similarity Methods Construction section, you say there are 5 levels and that for the third-level codes (bottom of the ICD-9-CM hierarchy) the number of levels is three. How are there five levels? Are the top two levels number -1 (for chapters) and 0 (of chapter sections)?

9) In the "Blood Test Similarity" section, you talk about similar blood tests but I don't understand if you are comparing subjects that have the test or that have similar results of the test. If the latter, what is "similar" - normal versus high versus low or exact value? What is "difference in distribution? You might comment someplace that the elimination of unusual tests might hurt your method, since those tests might be key distinguishers in unusual cases.

10) In "Gender Similarity" do you consider only two genders?

11) Throughout, you use the term "true diagnosis". How do you determine this?

12) The conclusion section is full of discussion that should be in the Discussion section.

Quality of written English: Acceptable

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

Declaration of competing interests:

I declare that I have no competing interests.