Reviewer’s report

Title: Improved de-identification of physician notes through integrative modeling of both public and private medical text

Version: 3 Date: 10 August 2013

Reviewer: Joel Martin

Reviewer’s report:

I believe that the authors have adequately addressed my concerns. This paper will be useful for other researchers because it carefully considers multiple sources of general features.

Discretionary Revisions

I think the authors do a better job of motivating the focus on Recall. It did seem unnecessary to mention F10, because it is for all practical purposes the same as Recall.

I found the new discussion about bigrams and trigrams to be a little dissatisfying. I agree that unigrams are probably more robust and that you are already doing very well. However, what if bigrams and trigrams raise the precision without affecting recall? How much time would it take to try that? My guess is that it would only take a few hours to two days to try it.

Level of interest: An article whose findings are important to those with closely related research interests

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.