Reviewer's report

Title: A Pipeline To Extract Drug-Adverse Event Pairs From Multiple Data Sources

Version: 2 Date: 12 October 2013

Reviewer: Jose Luis Oliveira

Reviewer's report:

The manuscript summarizes a text mining approach to leverage the wealth of spontaneous AE reporting and of public sources in the context of PV.

In the state of art section the authors missed many projects and scientific papers quite relevant in this subject. The section starting with "Various studies have looked at mining potential AEs from unstructured text sources.." lacks a deeper overview. The authors need to expand on the findings of other groups in this respect, for example the EU-ADR project, the Mini-Sentinel, the Observational Medical Outcomes Partnership (OMOP) and several others. The discussion section should highlight also in what way authors’ methodology is able to achieve better results than previous solutions.

The unstructured data were obtained through web crawling of three health-related websites. However it is not described how drug-event associations were obtained. I guess that many symptom-drug associations will be gathered through this method. The same may occur in PV-TPX.

Despite the paper is well written and organized, it includes many references to websites, Wikipedia (e.g. REST, SOAP, ..). Even Wikipedia is included. I recommend a more scientific approach for references.

The pipeline is well described and it seems adequate. However, besides the NER part, which is a previous work from authors, it uses some known data and methods (FAERS, BCPNN algorithm). The authors should emphasis the novelty of their approach.

The manuscript could be improved by performed a thoroughly validation of the obtained results. As it is now, the results are presented in a descriptive way showing events from the several methods (table 2).

Finally, despite part of the system is available for registered users, I would suggest to have a set of guess accounts for this kind of test.

Level of interest: An article of limited interest

Quality of written English: Needs some language corrections before being published
**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests