This is a peer-review report submitted for the paper “Finding Potential Adverse Events in the Unstructured Text of Electronic Health Care Records: Development of the Shakespeare Method”

Round 1 Review

General Comments

This paper [1] described the “Shakespeare method,” which was designed to discover associations between adverse events (AEs) caused by blood transfusion from unstructured electronic health record (EHR) notes. The authors applied this method on the MIMIC-III data set and seemed to be able to find transfusion AEs (TAEs) and potential TAEs (PTAEs) that were unknown when those EHR notes were developed.

Specific Comments

Major Comments

1. Is there any plan to release all the code/scripts used in this study? The method seems to be complex involving multiple steps; it will be very difficult to reproduce the results if the code is not available.

2. The manuscript should include more details on how the transfusion and comparison groups were created.

3. The author mentioned that the latent Dirichlet allocation (LDA) method they used in topic modeling requires the number of topics to be selected a priori. In this study, they set it to 45. Some questions:
   - How robust is the “Shakespeare method” with respect to this value? If a different value is chosen, will the method find similar topics? Similar notes for manual document review? Similar TAEs/PTAEs?
   - How would you determine this value if the method is applied to detect AEs for other treatments?

Minor Comments

1. In the Abstract section, the authors wrote “We chose the case of transfusion adverse events (TAEs) and potential TAEs (PTAEs) because real dates were obscured in the study data, and new TAE types were becoming recognized during the study data period.” The causal relationship here is a little confusing.

2. On page 3, the authors wrote, “The Shakespeare method has three parts,” but the following bullet-point list has 5 items.

3. On page 8: “The Shakespeare method would likely generalize to other her notes and possibly other types of medical texts.” An additional “her” is inserted.

Round 2 Review

The revision addressed my previous concerns. I have no further comments.

Conflicts of Interest

None declared.

Reference

https://med.jmirx.org/2021/3/e31548