Author’s response to reviews

Title: Building a Tobacco User Registry by Extracting Multiple Smoking Behaviors from Clinical Notes

Authors:

Ellen Palmer (ellelpalmer2018@gmail.com; elp76@case.edu)

Saeed Hassanpour (saeed.hassanpour@dartmouth.edu)

John Higgins (John.H.Higgins@Dartmouth.edu)

Jennifer Doherty (Jen.Doherty@hci.utah.edu)

Tracy Onega (Tracy.L.Onega@Dartmouth.edu)

Version: 1 Date: 15 Mar 2019

Author’s response to reviews:

To the editor and reviewers,

We thank you for your thorough review of our paper, “Building a Tobacco User Registry by Extracting Multiple Smoking Behaviors from Clinical Notes”. After reviewing your comments, we have made the recommended changes as documented below.

We have added some details to the methods, results, and discussion on our hotspotting method and the rationale for choosing this method. These changes can be found in the paragraphs starting at line 114 and 118 (split the paragraph to explain Cohen’s method alone in more detail and to delineate what we did that was novel), line 187 (has minor changes to indicate why we did not test adding words), and a new paragraph in the discussion starting at line 270 (rationale for using this method over others).

We have included additional details about the rationale for both Cohen and our choices for character/word windows around the hotspots starting at line 118. We chose to describe it from a linguistics approach to appeal to a wider audience, whereas Cohen previously described it from an algorithmic approach (bag of words to capture context instead of positional encoding).

We apologize for not including enough details about our SVM approach to reproduce our results. We have added a new paragraph starting at line 126 detailing both the method presented here as well as the other methods we attempted and why we moved forward with the SVM classifier presented here.
We look forward to answering any additional comments or questions you have about our approach,

Ellen “Elle” Palmer, PhD