**Peer Review History**

**MS Title:** **Predicting Mental Health from Followed Accounts on Twitter**

**Author names: Cory Costello, Sanjay Srivastava, Reza Rejaie, Maureen Zalewski**

**Editor decision—Revise & Resubmit**

Nov 12, 2020

Dear Cory,

Thank you for submitting your Stage 2 Registered Report, “Predicting Mental Health from Followed Accounts on Twitter” for consideration to Collabra: Psychology, and for your patience during this challenging time for reviewers. I returned your manuscript to one of the expert reviewers who assessed and approved the original Stage 1 manuscript (the other reviewer was unavailable) and I also read your manuscript carefully myself.

Overall the manuscript is in very good shape: the protocol remains faithful to the original approved methodology, with the reasonable (approved) deviations clearly explained and justified; the results are clearly presented and interpreted reasonably. On a purely subjective note (not affecting the editorial decision) I also found it a very engaging and topical read, and one that approached the workflow of RRs in a different way to most RR submissions I have edited. I’m sure the article will receive significant interest.

The reviewer notes a number of conceptual and analytic issues which it seems to me may benefit from additional consideration in the Discussion section. In revising, please take care not to alter any parts of the manuscript that were approved at Stage 1 , beyond correcting errors of fact or other essential clarifications.

In revising, please include a document with a point-by-point response to the reviewer’s comments, outlining each change made in your manuscript or providing a suitable rebuttal. Provided you are able to address the points raised, I anticipate being able to accept the revised manuscript without further in-depth review.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to resubmission.

If you have any questions or difficulties during this process, please contact the editorial office editorialoffice@collabra.org.

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Chris Chambers

**Reviewer 1**

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

The paper was an enjoyable read and for someone who has conducted large-scale studies - I appreciate the challenges involved. There are several complex issues which I think desereves discussion, and if there is analysis already in paper - reffering to it would be all the better.

1. Unlike many paper in this area, the analysis focus on internalised user metrics such as anxiety, but I am unsure if these are entirely internalised. Classically in NLP, we would deal with targeted responses towards a topic, such as sentiment or stance towards a topic. Then the emotion is both topic specific (and language specific). For example, anger is often easier to detect as it is topic specific. However, psychologic traits are much harder to define and are more latent constructs. Can you comment on your behavior positioning?
2. Related to above, how are the dynamics of the profile managed? For example, in sentiment and stance, we had to normalise the readings every day or normalised it against target baseline topics. What would be your fluctuating baseline?
3. Language: also many of the constructs we have in NLP are English centric. How do you account for differences between racial groups if their first language is different?
4. Have you considered other methods such as GPs? How would that compare against your RF approach?

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Author response**

Nov 16, 2020

Dear Dr. Chambers,

We want to thank you and the reviewer for your helpful feedback and for inviting us to submit a revised manuscript. We want to extend special gratitude to both you and the reviewer for making time for this during this global crisis. Below we provide a response to each of the reviewer’s comments and point to changes in the manuscript (when applicable).

*Unlike many paper in this area, the analysis focus on internalised user metrics such as anxiety, but I am unsure if these are entirely internalised. Classically in NLP, we would deal with targeted responses towards a topic, such as sentiment or stance towards a topic. Then the emotion is both topic specific (and language specific). For example, anger is often easier to detect as it is topic specific. However, psychologic traits are much harder to define and are more latent constructs. Can you comment on your behavior positioning?*

We thank the reviewer for this comment. We agree that these characteristics are manifest in external, observable variables. Indeed, our approach relies on internal characteristics manifesting in externally observable cues. We outline our general theoretical framework for understanding the link between the internal psychological constructs and external observable cues in the introduction from pages 4-6 (under the heading ‘Psychological Traits can be Inferred from Digital Records’). We further detail how the constructs we are studying in particular (anxiety, depression, Post-Traumatic Stress, and Anger) may manifest in the cues we’re studying in particular (followed accounts) from pages 6-9.

*Related to above, how are the dynamics of the profile managed? For example, in sentiment and stance, we had to normalise the readings every day or normalised it against target baseline topics. What would be your fluctuating baseline?*

We thank the reviewer for this comment. Taken with the previous comment, we understood this as a call to clarify how our approach relates to broad and stable individual differences vs. dynamic fluctuations or event-contingent responses. Although we noted throughout the Stage 1 and initial Stage 2 manuscript that our approach was focused on stable, trait-like characteristics, we did not explicitly address how such an approach might relate to dynamic, state-like characteristics or topical responses. To address this, we added a paragraph at page 33 clarifying that our approach is optimized for more stable psychological traits, but we also highlighted more dynamic network ties (e.g., likes and retweets) that could be promising to investigate as a future direction.

*Language: also many of the constructs we have in NLP are English centric. How do you account for differences between racial groups if their first language is different?*

We thank the reviewer for their comment and agree that differences in which language users primarily speak could influence predictive modeling results with social media data generally but believe this is a remote possibility in the present work. To qualify for the study, participants had to currently reside in the US, speak English fluently, and tweet primarily in English. We realize now that the language requirements were accidentally omitted from our initial Stage 1 and Stage 2 submissions; we now include these in the description of the sampling procedure (on p. 9). There were a small number of participants that met these requirements but for whom English was a second language (25 participants or approx. 4% of the total sample). This was also accidentally omitted from the demographic description of participants and we have added to the revised manuscript (on p. 11). Given that all participants were fluent in English and the vast majority spoke English as their first language, we think it is unlikely that language proficiency would have had an impact on our results.

*Have you considered other methods such as GPs? How would that compare against your RF approach?*

We thank the reviewer for their comment. A very large number of supervised machine learning methods are available (and that number is growing all the time), some of which could turn out to have better quantitative performance. Given that the search for the best method could be open-ended, and it was not a research aim of this study to maximize predictive performance at all costs, we had to draw a line somewhere. We thus limited our search to methods that we have seen showing promise in similar studies or applied settings, as reviewed on pp. 14-16. Given that this is a registered report format, we believe that adding such a comparison would have ideally been decided at Stage 1, before seeing the results. An exploratory search through other ML methods could have a very large scope, and in our view, adding one comparison now without a specific reason would not add enough value to justify lengthening the manuscript with additional analyses.

We would like to again extend our sincerest gratitude to you and the reviewer for the thoughtful feedback and opportunity to revise our Stage 2 Registered Report. We look forward to hearing from you.

Sincerely,

Cory K. Costello

Department of Psychology

University of Michigan

530 Church Street, University of Michigan

Ann Arbor, MI 48109

email: costck@umich.edu.

Sanjay Srivastava

Psychology Department

University of Oregon

1227 University of Oregon

Eugene, OR 97403

email: sanjay@uoregon.edu

Reza Rejaie

Department of Computer and Information Science

University of Oregon

1202 University of Oregon

Eugene, OR, 97403

Email: reza@cs.uoregon.edu

Maureen Zalewski

Psychology Department

University of Oregon

1227 University of Oregon

Eugene, OR 97403

email: zalewski@uoregon.edu

**Editor final decision—Accept**

Dec 2, 2020

Dear Cory,

I have now had a chance to read over your revised Stage 2 Registered Report manuscript and I am happy to say that your submission is now officially accepted for publication in Collabra: Psychology. Thank you for the excellent contribution to the journal.

On a personal note, this is one of my last editorial assignments at Collabra: Psychology. I officially left the editorial board last year, but stayed on to handle a small number of remaining Stage 2 submissions to ensure editorial continuity, and it has been a pleasure to go out editing such an interesting and high quality submission.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. Our managing editor will contact you in case there are any pre-prodution file related questions. You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely, Chris Chambers