**Reviewer’s report**

**Title:** Predicting Opioid Dependence from Electronic Health Records with Machine Learning

**Version:** 0  **Date:** 13 Nov 2018

**Reviewer:** CHELSEA CANAN

**Reviewer's report:**

This paper describes the development of a machine-learning algorithm to identify patients with opioid dependence using EMR data. The analysis is thorough and carefully described, and the objective of the paper is high impact given the current state of the opioid epidemic. My primary concern is that there is a slight disconnect between the predictive aims of the paper and the variables chosen to include in the model. While the goal of this analysis is to help physicians identify patients at risk for opioid dependence, the authors describe an algorithm based primarily on lab values and vital signs +/- 10 days of the substance dependence diagnosis. This may not be the optimal timeframe to predict future dependence, as it may be too close to the diagnosis for any meaningful interventions to occur. Creating an algorithm based on diagnoses and prescriptions in the prior five years might allow for a more meaningful prediction algorithm. Additional comments are provided below:

**Major comments**

1. **What was the reason for excluding people with alcohol/drug related mental health conditions from the control group?** Patients with these conditions are at higher risk for opioid dependence and it seems like we’d want to differentiate between these high-risk patients who don't develop opioid dependence vs. those who do. It might be useful to provide an algorithm to discriminate between the most "high-risk" patients, as this distinction is probably less obvious to providers.

2. **How did you choose to focus on dependence as an outcome rather than something like overdose or addiction?** Did you also look at overdose in this sample? If so, what is the AUROC for overdose? The conclusions state that the predictive model may help to identify overdose and opioid-seeking patients - is this because patients with opioid dependence are more likely to overdose and display opioid-seeking behaviors, or were these outcomes examined specifically?

3. **What is the clinical rationale for hypothesizing that lab values and vital signs may be predictive of opioid dependence?**
4. Is +/- 10 days the most relevant time window for predicting opioid dependence (see comment above)? Also, if this model will be used for predictive purposes, the prediction variables should precede the diagnosis. What was the rationale for including lab values and vital signs taken up to 10 days following the diagnosis?

5. The clinical phenotyping analyses provide a very practical take-away message from this paper. These variables (diagnoses, procedures, and prescriptions from the preceding 5-years) might be worth including in your predictive model, as these behaviors occur early enough that an algorithm created with these measures might allow for actionable interventions. Why did you choose to include only recent vital signs and lab values in the model?

Minor comments

1. Other papers have described similar attempts to use electronic data to identify opioid misuse (e.g. Rice 2012 Pain Med; Cochran 2014 Drug Alcohol Depend; Dufour 2014 Am J Pharm Benefits; Hylan 2015 J Pain). How does this paper fit with the existing literature?

2. Over what time period were data collected?

3. Adding a flow diagram to show how you arrived at the final analytic sample would be useful.

4. How did you define the presence of HIV and hepatitis C?

5. What was the difference in non-methadone opioid prescriptions between cases and controls?

6. The decision to rank predictors by p-value seemed unusual. The value of the p-value doesn't tell us how different the two values are; it only tells us how likely it is that the observed difference is due to chance. Ranking the predictors by difference in standardized effect size is more meaningful.

7. What does this model provide that in-person clinical assessment cannot? Emphasizing the clinical utility of this model would strengthen the discussion.

Level of interest
Please indicate how interesting you found the manuscript:

An article of importance in its field
**Quality of written English**

Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**

Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interest

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

No