Author’s response to reviews

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

Authors:

Randall Ellis (randy.ellis@icahn.mssm.edu)

Zichen Wang (zichen.wang@mssm.edu)

Nicholas Genes (Nicholas.Genes@mountsinai.org)

Avi Ma'ayan (avi.maayan@mssm.edu)

Version: 1 Date: 24 Dec 2018

Author’s response to reviews:

Replies to Reviewer #1

Reviewer’s comment: The strategy of excluding patients, vital signs, and lab results with sparse data may limit the usability of this work in real practice. Maybe using a data imputation strategy to handle the missing data is more appropriate to resolve this problem.

Response: Thank you for this comment. For the revise version of the paper we added analysis of imputing variables to fill in missing data and retain more samples. Indeed, the results improved the overall quality of the classifiers to ~92%, so we conclude that the imputation did improve performance. However, such improvement is seen only when dropping those patients with sparse data.

Reviewer’s comment: The area under the curve for the dummy classifier is surprisingly high (0.626) in comparison to a truly random classifier (AUC=0.5). The mechanism of this dummy classifier should be described to justify this performance.

Response: Thank you for this comment. After examining the dummy classifier, we realized that we made a mistake. This is now corrected in the revised manuscript where the AUC for the dummy classifier is always around 0.5.

Reviewer’s comment: It is worth comparing the top 10 features from Gini importance and p-values and discussing these feature lists' similarities and differences.
Response: Thank you for this comment. We now added an analysis that compares the ranked features from the two methods, and visualized the results as a scatter plot. This analysis produced a composite list of features that are highly ranked among both methods. Overall we see significant correlation between the two approaches.

Replies to Reviewer #2

Reviewer’s comment: 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.

Response: Thank you for this comment. Indeed, reviewing prior diagnoses and prescriptions before initial dependence diagnosis is another useful way to construct a predictive model. For the revised version of the manuscript we have added such a model and compared its performance to the original models. Both types of models can be useful in alerting physicians to a potential opioid dependence problem since we believe that currently such diagnoses are under detected. Feature importance analysis was done by looking at the model variables that are most predictive. We also added a classifier that uses labs and vitals from -20 days before initial diagnosis, and recomputed the scores which are comparable to our original results.

Reviewer’s comment: 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.

Response: Thank you for this suggestion. The reason we excluded patients with alcohol/drug-related mental disorders from the control population was that we made the assumption that these patients are more likely drug dependent already. Hence, these patients will probably better fit to be in the case population rather than considered as controls. Since psychiatric disorders in general are strongly represented in the case population, it is probable that a lot of patients with
psychiatric disorders, and more specifically those with alcohol/drug-related mental disorders, are drug dependent even though there is no formal record for it. Hence, these “high-risk” patients will be predicted as positive by our model. For the revised manuscript we tested how the model classifies these “high-risk” patients as a test case. Indeed, we see that the model classifies much higher percent of these patients as positive cases. We added text that describes this new analysis and the motivation for it.

Reviewer’s comment: 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?

Response: Thank you for this suggestion. First, we would like to note that we combined dependence and addiction as the same thing. Using the ICD-9 code for overdose, we also now tested whether we can predict overdose using the same approach. While more variable, we see that we can predict overdose well. This analysis is now added to the manuscript.

Reviewer’s comment: What is the clinical rationale for hypothesizing that lab values and vital signs may be predictive of opioid dependence?

Response: The clinical rationale for hypothesizing that lab values and vital signs may be predictive of opioid dependence is counterintuitive. Most clinicians will not guess that it is possible, although agree that it would be helpful if that is the case. Our rationale is that while a single biomarker may not be predictive of opioid dependence, a composite non-linear biomarker can possibly be predictive. So far, lab tests and vital signs were not considered to be predictive of opioid dependence. Hence, the rationale for the project is that these variables are easy to measure, are commonly measured, and can inform us about the molecular and phenotypic changes that occur due to drug dependence, as well as flag patients that are at risk.

Reviewer’s comment: 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?

Response: Thank you for this suggestion. We initially chose this time period since it provides a good representation of the dependence diagnosis signature state. The reasons for this assumption
are: A) the physician had to notice signs of dependence and inquired the patient about it; or B) the patient felt compelled to directly report to their doctor that they are dependent. Including +/- 10 days around the diagnosis provides a time window that is relevant to the diagnosis. However, we agree that for practical implementation, past tests might be more practical. For the revised manuscript, we also attempted a -20 days prior to diagnosis time window. Comparing the results, we see that the performance quality remained approximately the same.

Reviewer’s comment: 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?

Response: Thank you for this suggestion. For the revised version of the manuscript we tried to use diagnoses, procedures, and prescriptions from 5 years prior to diagnosis as features. This analysis is now added to the paper. We run this analysis to obtain the specific performance statistics such as AUROC and AP as well as performed feature importance analysis. Overall we see that including these variables did not improve the results, but these features are also very predictive on their own. The feature importance analysis confirmed the same diagnoses, procedures, and prescriptions we reported in the initial submission of the manuscript under clinical phenotyping.

Reviewer’s comment: 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?

Response: Thank you very much for this useful comment. Indeed, these publications are very relevant to our study. In the Introduction of the revised version of the manuscript we added citations and a new paragraph that discusses these articles where we explain the commonalities and differences between our study and those methods. In principle, our study utilizes a non-linear machine learning approach and can produce predictions that are only based on lab tests and vital signs. The other models use simpler statistics and are insurance claims databases that mostly do not contain molecular measurement data. However, the results of most of those studies are very similar to the conclusions and observations we obtained.

Reviewer’s comment: Over what time period were data collected?
Response: Thank you for this useful comment. We now added the period to the methods section.

Reviewer’s comment: Adding a flow diagram to show how you arrived at the final analytic sample would be useful.

Response: Thank you for this useful comment. We now added a flow diagram to the paper to describe the filtering steps applied to arrive at the final groups.

Reviewer’s comment: How did you define the presence of HIV and hepatitis C?

Response: We define the presence of HIV and hepatitis C by their diagnosis ICD-9 codes.

Reviewer’s comment: What was the difference in non-methadone opioid prescriptions between cases and controls?

Response: Thank you for this useful suggestion. In the revised manuscript we now calculate the difference in non-methadone opioid prescriptions between cases and controls and report this result in the discussion section.

Reviewer’s comment: 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.

Response: Thank you for this useful suggestion. In the revised manuscript we now rank predictors by effect size, rather than p-value. We also re-run the classification using the top 10 features ranked by effect size and calculate AUROC and AP scores.

Reviewer’s comment: What does this model provide that in-person clinical assessment cannot? Emphasizing the clinical utility of this model would strengthen the discussion.

Response: The multi-variate characteristic of the model, or in other words, how unique combinations of the values of many measured values together, can produce predictions not possible by looking at a single biomarker. The complex relationships between measured variables are un-intuitive and would be very difficult to detect via an in-person clinical assessment. We agree that this was not emphasized enough in the original version of the
manuscript. It is now added to the Introduction and Conclusions section of the revised manuscript.