

Response to the Reviewers' comments on manuscript "Prediction of Appointment No-shows using Electronic Health Records"

Thank you very much for giving us the chance to revise the manuscript. Below we provide detailed responses to the reviewer comments; we have incorporated all of their suggestions into this revision. In the revised manuscript, all corrections, changes and additions (except for minor edits) have been marked in red in order to facilitate your review. We hope that you will now find the manuscript acceptable for publication.

Response to Associate Editor

Your paper was reviewed by two referees. Both provided constructive comments on your manuscript with the recommendation of rejection & resubmission and acceptance, respectively. My own reading of the paper concludes that a major revision is certainly needed at this point. Please make a major revision by addressing all comments from both reviewers. Provide point-to-point responses and indicate changes made by colored text in your revised manuscript.

Response: We appreciate the reviewers' feedback. We have incorporated their suggested changes in the manuscript and have answered their comments point-by-point in this document.

Response to Referee 1

There are several problems with this paper that mean I cannot accept it for publication:

1. *They make no mention of previous landmark studies into missed appointments (e.g. Morbidity, mortality and missed appointments in healthcare: a national retrospective data linkage study by McQueenie et al, Understanding repeated non-attendance in health services: a pilot analysis of administrative data and full study protocol for a national retrospective cohort by Williamson et al, Demographic and practice factors predicting repeated non-attendance in primary care: a national retrospective cohort analysis by Ellis et al.*

Response: Thank you for pointing this out. We have introduced the missing literature in the Introduction of the manuscript by adding the following paragraphs.

Patients that do not show to their scheduled appointment or cancel the appointment late are commonly referred in the literature as appointment *no-shows*. A missed appointment could negatively impact patient's health condition and even increase the risk of premature death in severe cases. McQueenie et al. [2019] have studied the effect of no-shows on all-cause mortality in patients with long-term mental and physical health conditions, using a 3-year period across Scotland. They have found that patients with long-term mental conditions have a higher risk of mortality if missing two or more appointments.

In this work, we are interested in prediction models that can be used to identify specific risk factors and patient subpopulations that have a higher risk of not showing to their appointments. This information can be used to aim intervention strategies to the fraction of incoming patients with highest no-show proneness. Notably, Williamson et al. [2017] and Ellis et al. [2017] have made an effort to link patients' risk of missing appointments with patient-level factors and practice-level factors in the UK. Williamson et al. [2017] used a retrospective

cohort design and Ellis et al. [2017] used a negative binomial model for calculating no-shows risks and a cohort analysis based on patient and practice characteristics. Our paper differs significantly from previous studies both in data structure and methodology. Our data contains summaries of previous appointments for a specific patient such as the proportion of missed appointments in the last three months and the number of times appointments were rescheduled. Using this information and other patient characteristics (demographic features, health conditions, etc) we aim to predict no-show appointments. Our study identifies possible predictors for high-risk appointment no-shows from a large sparse dataset, and we employ Bayesian statistical learning methods for this prediction task. Here, we are able to identify the factors that significantly affect the risk of no-show appointments among hundreds of potential factors to then take specific actions such as targeted reminder phone-calls and optimal overbooking in order to reduce the risk.

2. *If the authors had read the above papers, they would know that a large factor in missed appointments is governed by how many appointments the patients themselves make. Hence, for example - if a patient makes 10 appointments and misses 5, that's a significant problem. However if a patient makes 300 and misses 5, that's significantly less of a problem. Papers highlighted above controlled for number of appointments made. This paper does not. This is a major limitation.*

Response: This is certainly a good point. We do not have access to all the previous appointments a patient has had but our data contains summaries of this information. For example, the proportion of missed appointments in the last three months for each of the patients was used as a predictor variable in our model. This variable certainly plays an important role in predicting future no-show appointments as it is shown in Table 1 in the manuscript where this is listed as one of the ten most frequent relevant predictors across all providers indicating increased risk of no-show in patients.

3. *The paper immediately talks about healthcare cost concerns (lines 44-47) but does not mention the actual health of the patients. This is an alarming and depressing omission.*

Response: We really appreciate the referee's feedback on this aspect of the paper. It was a major oversight in our initial discussion. We have incorporated the recommended literature in the manuscript as shown in point 1.

4. *Smaller issues:*

- *Left hand graph of figure 1 is difficult to read*
- *Figure two x axis labels should not be variable names ("African_American")*
- *Tables "Name of Predictor" should not be variable names, it makes the tables very difficult to read*

Response: We have incorporated these minor changes in the paper.

Response to Referee 2

This paper, applies sparse Bayesian modeling approaches based on Lasso and automatic relevance determination to predict and identify the most relevant risk factors of no-show patients.

Major points:

- *This paper is well written.*
- *The real application is very interesting.*
- *The paper is proper for this journal.*

Given the reasons above, I suggest accepting this paper for publication.

Response: We appreciate the positive feedback about the paper.

References

- M. El-Sharo, B. Zheng, S. W. Yoon, and M. T. Khasawneh. An overbooking scheduling model for outpatient appointments in a multi-provider clinic. *Operation Research for Health Care*, 6:1–10, 2015.
- D. A. Ellis, R. McQueenie, A. McConnachie, P. Wilson, and A. E. Williamson. Demographic and practice factors predicting repeated non-attendance in primary care: a national retrospective cohort analysis. *The Lancet Public Health*, 2(12), 2017.
- R. McQueenie, D. A. Ellis, P. Wilson A. McConnachie, and A. E. Williamson. Morbidity, mortality and missed appointments in healthcare: a national retrospective data linkage study. *BMC medicine*, 17(1), 2019.
- A. E. Williamson, D. A. Ellis, P. Wilson, R. McQueenie, and A. McConnachie. Understanding repeated non-attendance in health services: a pilot analysis of administrative data and full study protocol for a national retrospective cohort. *BMJ Open*, 7(2), 2017.