
Research and Applications

Designing risk prediction models for ambulatory no-shows across different specialties and clinics

Xiruo Ding,¹ Ziad F Gellad,^{2,3} Chad Mather III,² Pamela Barth,⁴ Eric G Poon,⁴ Mark Newman,⁵ and Benjamin A. Goldstein^{1,6}

¹Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, 27710, USA, ²Department of Medicine, Duke University, Durham, North Carolina, 27703, USA, ³Department of Medicine, Durham VA Medical Center, Durham, North Carolina, 27705, USA, ⁴Duke Health Technology Solutions, Duke University, Durham, North Carolina, 27713, USA, ⁵Department of Anesthesiology, University of Kentucky, Lexington, Kentucky, 40536, USA and ⁶Center for Predictive Medicine, Duke Clinical Research Institute, Durham, North Carolina, 27705, USA

Corresponding Author: Benjamin A Goldstein, 2424 Erwin Road, Durham, NC, 27705, USA. E-mail: ben.goldstein@duke.edu

Received 13 July 2017; Revised 25 December 2017; Editorial Decision 31 December 2017; Accepted 7 January 2018

ABSTRACT

Objective: As available data increases, so does the opportunity to develop risk scores on more refined patient populations. In this paper we assessed the ability to derive a risk score for a patient no-showing to a clinic visit.

Methods: Using data from 2 264 235 outpatient appointments we assessed the performance of models built across 14 different specialties and 55 clinics. We used regularized logistic regression models to fit and assess models built on the health system, specialty, and clinic levels. We evaluated fits based on their discrimination and calibration.

Results: Overall, the results suggest that a relatively robust risk score for patient no-shows could be derived with an average C-statistic of 0.83 across clinic level models and strong calibration. Moreover, the clinic specific models, even with lower training set sizes, often performed better than the more general models. Examination of the individual models showed that risk factors had different degrees of predictability across the different specialties. Implementation of optimal modeling strategies would lead to capturing an additional 4819 no-shows per-year.

Conclusion: Overall, this work highlights both the opportunity for and the importance of leveraging the available electronic health record data to develop more refined risk models.

Key words: predictive model, clinical decision making, model comparison, electronic health records

INTRODUCTION

As the adoption of modern electronic health record (EHR) systems has proliferated, one of the key ways that researchers have used EHR data is to build risk prediction models.¹ The majority of risk models have consisted of single center studies creating the question as to whether these risk models are transferable. However, even within a single center, it is questionable as to whether a risk score is generalizable across heterogeneous patient populations. For example, researchers have derived risk scores generally for 30 day readmission^{2,3} – an important driver of reimbursements – as well as models specific to sub-populations such as patients with heart fail-

ure⁴ and chronic kidney disease.⁵ If the underlying risk factors across subpopulations are consistent, then building a model on combined data should be more powerful. However, if there is underlying heterogeneity, then combining sub-populations could result in poorer performing models. Not surprisingly, efforts to compare pooled and separate models have been inconclusive.⁶

In this paper we explore the question of whether a risk model should be pooled or separated across clinical populations. In our use case we sought to derive a risk model for whether a patient would no-show to an outpatient appointment. Like readmissions, there is significant cost

associated with patient No-Shows.^{7,8} Moreover, there is also variability in no show rates across different specialties,⁸ with research exploring risk models within specialty specific domains.^{9,10} Finally, risk factors for “No-Shows” have been well studied and identified,^{11–13} facilitating the ability to create a risk score.

Since various intervention strategies have been identified to reduce “No-Shows”^{14–17} our institution sought to implement a risk model to identify high risk patients. Like most large health care systems, our institution consists of a variety of specialties and clinics, each serving different patient populations. While building separate models is straightforward as a data exercise, implementing, and maintaining multiple risk models can be challenging. Moreover, as clinics are added to a health system, it is valuable to know whether a general model can be directly implemented or whether models need to be trained to clinic specific data to be useful. Therefore, we sought to determine whether the optimal model should be built on the overall system, specialty, or clinic level and understand how the various clinics and specialties may differ in their underlying risk factors for No-Shows.

METHODS

Available Data

Our institution switched to an Epic based system in late 2013. We extracted data on outpatient appointments from 2014 to 2016 from our EHR system. We selected 14 high volume adult specialties that utilize scheduled appointment slots (Cardiology, Dermatology, Endocrinology, Gastroenterology, Neurology, Ophthalmology, Orthopedics, Otolaryngology, Plastic Surgery, Pulmonary and Allergy, Pulmonology, Rheumatology, Urogynecology, and Urology). Additionally, many specialties have multiple clinics, both attached to and detached from the primary hospital, sometimes serving distinct patient populations. In total these specialties encompassed 55 unique clinics.

Outcome definition

Our primary outcome was patient no-show or late cancellation to a scheduled appointment. After consulting with different clinics, we noted that the definition of “late cancellation” differed based primarily on the ease of rescheduling the appointments. For the purpose of this analysis, we used a consistent definition of late cancellation as cancellation on the day of the appointment. This allows any late cancellation to *functionally* be a No-Show, i.e., not intervenable. Therefore, our outcome was a composite of a no-show or same day cancellation, referred to as simply “No-Show.” Any appointments that were canceled prior to 1 day were excluded from analysis. Similarly, any appointments made on the day of the appointment were excluded.

Predictor variables

Defining which variables to integrate into an EHR based risk score can be challenging as one wants to balance parsimony and discovery. Since our ultimate goal was to implement a risk score, our primary consideration in selecting variables was that they had to be easily extractable and calculable from within the EHR system. This excluded some variables such as patient distance from a clinic, which would be more challenging to calculate in a real-time environment.

After consultation with clinical and practice operational colleagues and the literature, we extracted data on 63 predictors of no-show. See [Supplementary Table S1](#) for the full list. Extracted variables could roughly be broken down into demographics (e.g., Age, Sex, Race), comorbidities (e.g., substance abuse, psychiatric diagnoses), service utilization history (e.g., previous appointments in clinic,

previous no-shows), appointment information (e.g., day of week, time of day, appointment length), financial information (e.g., insurance status, copay due), and patient engagement (e.g., active in MyChart [online portal], response to automated pre-phone call). Since many of the variables had multiple levels (e.g., payor type) variables were grouped into meaningful categories that had at least a 1% frequency. Some demographic variables (e.g., race, employment status) had missing values. For these we created a missing category.

For the purposes of risk modeling, predictor values were extracted from the EHR as of 3 days prior to the appointment.

Development of Predictive Models

Predictive models

Given the structure of the available data, we built models on three different levels: System, Specialty or Clinic. A model built on a *higher* level (e.g., System) could include information on a lower level as a covariate (e.g., clinic indicators). Without loss of generality, we express these different models below as a logistic regression model. We define Y to be our indicator for “No-Show,” X the individual level predictors, and S and C the specialty and clinic indicators. The associated parameters are β , γ , and ω .

(a) On the health system level we have:

$$\text{logit}(Y) = X\beta + \varepsilon \quad (1)$$

$$\text{logit}(Y) = X\beta + S\gamma + \varepsilon \quad (2)$$

$$\text{logit}(Y) = X\beta + C\omega + \varepsilon \quad (3)$$

These models use all available observations. The added indicators for specialties and clinics in models (2) and (3) serve to vary the intercept (i.e., No-Show rate) but the effects are assumed constant across each specialty/clinic. We note that (3) does not have an indicator for specialty since clinics are nested in specialties.

(b) Specialty level models:

$$\text{logit}(Y) = X_s\beta_s + \varepsilon \quad (4)$$

$$\text{logit}(Y) = X_s\beta_s + C_s\omega_s + \varepsilon \quad (5)$$

Here we fit a separate model for each specialty (4) adding a clinic specific intercept in (5). Each specialty has its own associated parameter vector, β_s . While in practice we estimate each model separately, this can be thought of as an interaction between specialty and all of the predictors' variables.

(c) Clinic levels model:

$$\text{logit}(Y) = X_c\beta_c + \varepsilon \quad (6)$$

Finally, we have a separate model for each clinic. As above, we fit the model separately, but this can be thought of as an interaction between clinic and the predictor variables.

Model estimation

We divided the data into training and testing sets, using 2014 to 2015 data for training and 2016 for testing. We tested different analytic strategies on a subset of the training data. After deciding on the desired estimation approach, we fitted the above models on the full training data and assessed model performance on the independent test data.

Above, we presented the prediction models as logistic regression models for convenience. In principle we can use any model building strategy. After consideration and testing of different approaches, we decided to use a regularized logistic regression, LASSO (least absolute shrinkage and selection operator).¹⁸ Regularized logistic regression is similar to logistic regression except it “shrinks” β -coefficients

towards zero to generate a more stable risk estimate, that is, avoid over-fitting. This results in biased coefficient estimates that can no longer be interpreted as log odds-ratios, as in a logistic regression. However, another difference between LASSO and logistic regression is the predictors are first standardized to have unit variance. By placing all of the predictor variables on the same scale, this allows one to interpret the coefficients as a *weight* or relative importance of that predictor variable. The amount of shrinkage was chosen separately for each model using 10-fold cross-validation. We also tested the added value of including quadratic and interaction terms but saw minimal added values. Therefore, we only fit a model with main effects. We used the glmnet package¹⁹ in R to fit the models.

Model evaluation

We evaluated discrimination and calibration of the predictions generated on the test data. To assess discrimination, the ability to separate patients into risk groups, we calculated a concordance statistic, i.e., C-statistic. To assess calibration, the alignment of the predicted probabilities with the underlying probabilities, we calculated the calibration-slope.²⁰ We note that a calibration slope of 1 indicates perfect calibration, i.e., the predicted probabilities reflect the underlying probabilities, while deviation from 1 indicates miss-calibration.

To understand underlying differences between the models, for each specialty level model [model (4)], we compared the estimated β -coefficients from the LASSO fit. Finally, for each of the clinic level models [model (6)] we assessed how model fit related to size of the training data.

This study was exempt by our institution's Institutional Review Board.

RESULTS

In total, we had 2 264 235 individual appointments across 14 specialties and 61 clinics. We removed 6 clinics because they did not exist during either training or testing periods. This left us with a total

sample size of 2 232 737 appointments across 55 clinics. [Supplementary Table S1](#) contains descriptive data across the 14 specialties. We note that there are some meaningful differences between the specialties. Of greatest interest is that the specialties saw different types of patients. There was heterogeneity with respect to patient primary payor, age, and gender. Many of the specialties also differed in whether their patient appointments were primarily office or hospital based. [Figure 1](#) shows the No-Show rate for each of 14 specialties. The no-show rates differ across specialties with the lowest in Urogyneecology (13%) and the highest in Pulmonary & Allergy (32%).

Model Performance

System level evaluation

We first considered how adopting a single modeling approach would perform across the full system. [Table 1](#) presents the C-statistics and calibration slopes for each of the 6 models. Each of the models is very well calibrated. The three models that incorporate clinical specific information have the best discrimination, with the clinic specific model performing nominally the best.

Specialty level evaluation

We next evaluated the models within each specialty level. Using models (1–6), we generated fits for each of the 14 specialties for a total of 84 evaluations. [Figure 2](#) shows the C-indices and calibration slopes for the different fits. The C-statistics range from 0.70 to 0.95, with the more granular models generally showing better discrimination than the full system level models. The models are also very well calibrated, with again the more granular models showing better performance. In general the size of the training data does not appear to relate to the model performance.

Clinic level evaluation

Finally, we evaluated models on the clinic level, evaluating a total of 6 models for each clinic (330 total). [Figure 3](#) show the discrimination and calibration results across model fits 1–6. Here we generally observe that

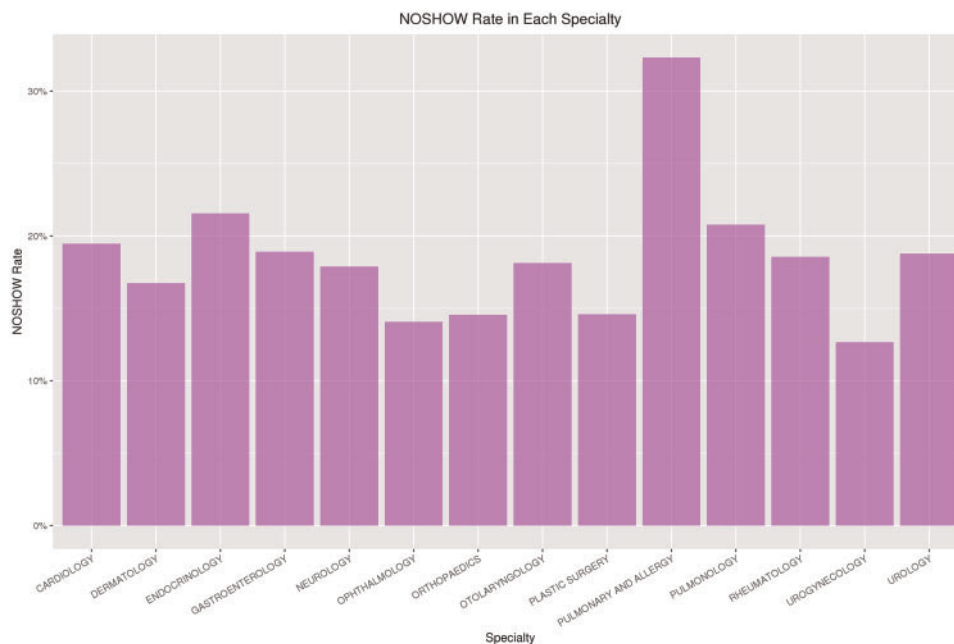


Figure 1. No-show rates across the different specialties. The no-show rates differ across specialties with the lowest in Urogyneecology (13%) and the highest in Pulmonary & Allergy (32%).

the clinic specific models perform better than those at the higher level. This holds even though some of the clinic sizes were quite small in the training data ($n < 200$). Overall clinic size was only moderately correlated with C-statistic, with a Spearman correlation coefficient of 0.32 ($P < 0.05$). We note that for many of the clinics we are able to get quite strong model performance with an average C-statistic of 0.83. Similarly, we find that the clinic level models have the best calibration.

To compare the 6 different models, Table 2 shows the number of clinics where one model performs best in terms of discrimination and calibration. The clinic specific model performs best 67% and 57% of the time for discrimination and calibration, respectively. We assessed also whether clinic size was related to whether the clinical level model performed better than the specialty level model. To do this, we created an indicator for whether the clinical level model performed better and regressed this on clinic size (in the training data). The odd ratio was (per-1000 people) was 1.014 (95% CI, 0.984-1.044), suggesting that size of the clinic training data is not a significant determinant of whether the clinic level model performs better.

Clinical decision making

We assessed the impact of model choice on decision-making. Using a predicted risk threshold of 20% – a value we have implemented in our clinics – we calculated the overall sensitivity for choosing each of the 6 models. The clinic specific model had a sensitivity of 66%

Table 1. Model Performance on the System Level

Models	Discrimination	Calibration
Overall model	0.814	1.02
Overall model with specialty	0.820	1.01
Overall model with clinic	0.814	1.02
Specialty-specific model	0.835	1.00
Specialty-specific model with clinic	0.836	1.00
Clinic specific model	0.841	0.98

while the center general model had a sensitivity of 62%. This corresponds to capturing an additional 4160 no-shows in 2016. Moreover, choosing the optimal model on a clinic-by-clinic basis results in even greater performance. Overall, the optimal model had a sensitivity of 69%, corresponding to capturing an additional 4819 no-shows beyond the clinic specific model.

Important Predictor Variables

Finally, we assessed the drivers of the prediction model. Figure 4 shows the standardized β -coefficients for each of the 14 specialty levels models. These can be thought of as importance weights for the predictive model as opposed to measures of association. For the most part, the direction of the effect was the same across specialties, though the magnitude differed. Table 3 lists the top 5 predictors (in scale of absolute values) for all 14 specialties. For all but 2 of the specialties, appointments rescheduled by the provider were the number one predictor of patient no-shows. While other predictors are shared across specialties (e.g., copays due, number of previous no-shows, number of previous appointments) others are also unique to the specialty (patient sex, appointment length).

DISCUSSION

Our results illustrate the importance of considering the underlying clinical heterogeneity when developing risk prediction models. Many risk prediction models apply a single set of functions across broad clinical populations. However, we have shown that doing so can lead to poorer performing models. Instead, it is likely best to develop models on as local of a level as possible, while assessing whether “higher” levels lead to better performance.

To illustrate this concept, we developed a series of risk models for patient No-Show/Late cancellation across a range of adult specialties. We looked at 14 specialties, constituting 55 clinics. In general we found that clinic level models showed the best discrimination and calibration

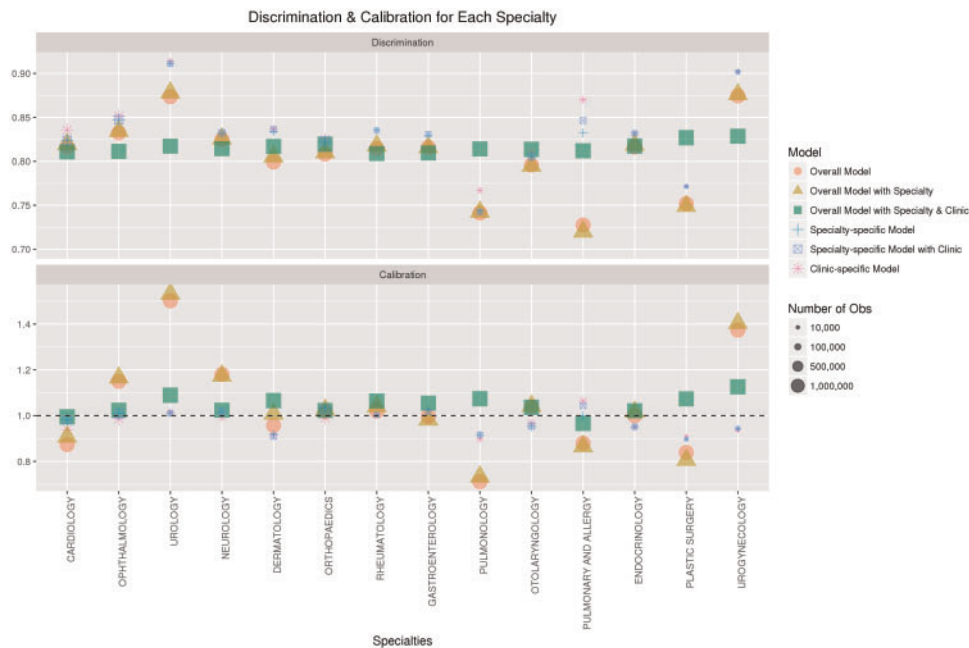


Figure 2. Discrimination and calibration evaluated at the specialty level. The dashed line at ‘1’ indicates perfect calibration. The models that incorporate clinic specific information typically have the best performance. Point size indicates the size of the training data.

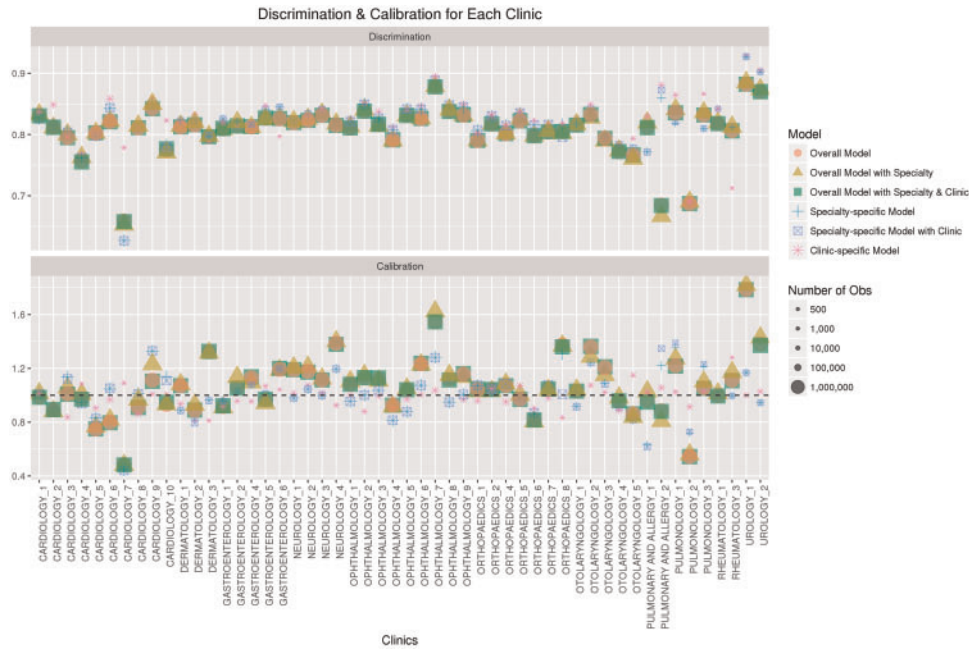


Figure 3. Discrimination and calibration evaluated at the clinic level. Variability is shown in model performance across the different clinics. In general, the clinic specific models (pink star) have the best performance, but this is not uniformly the case.

Table 2. Best Performing Model Level

Models	Discrimination Number of clinics	Calibration Number of clinics
Overall model	0	1
Overall model with specialty	2	6
Overall model with clinic	0	2
Specialty-specific model	7	7
Specialty-specific model with clinic	9	7
Clinic specific model	35	30
Total	53	53

67% and 57% of the time, respectively. One implication of these results is that the risk factors for patient no-shows across the different clinics are different. This was confirmed upon examination of the risk coefficients, with different predictors showing importance for different specialties. This raises the question of when local models perform better than general models. Before this study, we would have hypothesized that clinic size would be significantly related to better performance of a clinic specific model since more training data would be available. While clinic size was moderately correlated with model performance ($r = 0.32$), clinic size was not significantly associated with a better clinic level model. Instead, the difference in performance is likely due to other forms of clinic level heterogeneity. This implies that we cannot apply simple a priori rules as to when local models will outperform general models. Instead, this is something that needs to be learned from the data. As such, one way to consider the impact of model level is as a form of bias-variance trade-off. A model built on a higher level will be less variable but have greater bias. Conversely, a more local model, while producing a better fit, will have more variance.

This work also highlights the potential of modern EHR systems to leverage large amounts of data. Many EHR vendors are offering prediction models out of the box without requiring or recommending clients to do any retraining with local data. The heterogeneity seen within our data set suggests that local validation and recalibration

would lead to better performance and should be part of the implementation of these off-the-shelf models. Since we have a large amount of data (greater than 2 million appointments), we were able to sub-stratify to create more refined models. Obviously all institutions will not have this ability and may be forced to derive higher level models. Even so, there is an importance in deriving models as locally as possible. The importance of accounting for local models has been acknowledged elsewhere,²¹ with some emphasizing the need to understand local clinical practice when developing these models.²² One advantage of developing clinic specific risk models is that it is possible to tailor the outcome to the particular clinic environment. For example, at our institution, different specialties prefer different definitions for late cancellation. For some specialties a late cancellation was the day before the appointment. However, other specialties would ideally define late cancellation as 3 to 5 days before the appointment. By developing clinic specific models, each clinic could define the outcome that makes the most sense for that particular clinic. Even when full clinic specific models are not available due to sample size constraints,²³ random effects models can be a useful for strategy for incorporating group specific variability.^{24,25} This can also be facilitated by using cluster level evaluation metrics.²⁶

While our investigation focused on patient no-shows, these results have implications across health system based risk modeling. Work predicting 30-day readmissions has suggested that condition specific models may do better,⁶ as well as models that target specific causes of readmission²⁷ or patient characteristics.²⁸ Similarly we can consider stratifying other risk models such as surgical complications where condition specific models have been derived (e.g.,²⁹) and it has been suggested that site specific models are better.³⁰ That being said, as this work shows, it is impossible to say categorically whether a local or general model will perform better – in our case, there were environments where each was preferable. Every predictive modeling task ought to consider this within as a tuning parameter in the study design.

Coefficients in Specialty-Specific Models

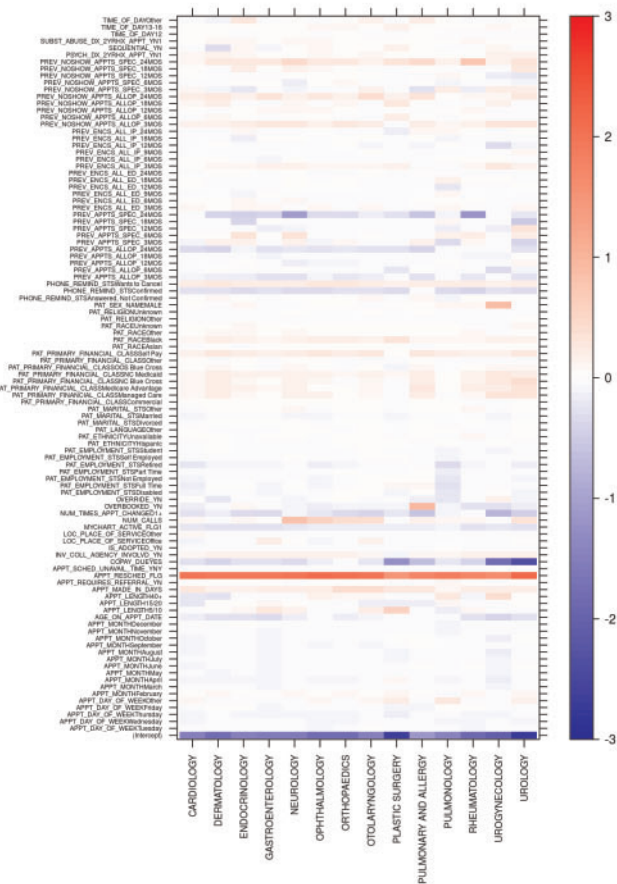


Figure 4. β -coefficients for each of the 14 specialty levels models. Red indicates a risk factor for no-show while blue indicates a protective factor. White indicates no prediction. Coefficients are scaled to have equal scale.

Table 3. Top Predictors for Each Specialty Level Model

Specialty	Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5
Cardiology	APPT RESCHED	Prev Appt All OP 24 months	COPAY DUE	APPT CHANGED1+	PHONE REMIND: confirmed
Ophthalmology	APPT RESCHED	NUM CALLS	Prev Appt Spec 24 months	PHONE REMIND: Confirmed	Prev No Show Spec 24 months
Urology	COPAY DUE	APPT RESCHED	Prev Appt Spec 18 months	APPT CHANGED1+	Prev Appt Spec 3 months
Neurology	APPT RESCHED	Prev Appt Spec 24 months	NUM CALLS	Prev No Show Spec 24 months	PHONE REMIND: confirmed
Dermatology	APPT RESCHED	Prev Appt Spec 24 months	Prev Appt All OP 24 months	COPAY DUE	SEQUENTIAL Appt
Orthopedics	APPT RESCHED	NUM CALLS	Appt made days	APPT CHANGED1+	Prev Appt Spec 24 months
Rheumatology	APPT RESCHED	Prev Appt Spec 24 months	Prev No-Show Spec 24 months	COPAY DUE	PHONE REMIND: confirmed
Gastroenterology	APPT RESCHED	Prev Appt Spec 24 months	Age	APPT CHANGED1+	Prev No Show All OP 24 months
Pulmonology	APPT RESCHED	Employment: Retired	PHONE REMIND: confirmed	Prev Appt Spec 3 months	Employment: Full Time
Otolaryngology	APPT RESCHED	NUM CALLS	APPT CHANGED1+	Prev No Show All OP 24 months	Appt Made Days
Pulmonary and allergy	APPT RESCHED	OVERBOOKED	COPAY DUE	Prev Appt Spec 6 months	APPT CHANGED1+
Endocrinology	APPT RESCHED	Prev Appt Spec 24 months	Prev Appt Spec 18 months	Prev Appt Spec 24 months	APPT CHANGED1+
Plastic surgery	APPT RESCHED	COPAY DUE	APPT LENGTH: 5/10 min	Prev No Show Spec 24 months	PHONE REMIND: Confirmed
Urogyneecology	COPAY DUE	APPT RESCHED	Male	APPT CHANGED1+	APPT LENGTH: 40+ min

There are a few limitations to our approach. Most notably, this study only investigated data from a single center. Currently there is a lack of evidence establishing how well risk models port across institutions.¹ One may extrapolate that the observed disparities in performance would be exasperating had we considered validating across different institutions; however, this is worthy of further examination. We also did not fully leverage the available repeated measurements. While we added covariates based on patients' clinical history, particularly missed appointments, we likely could have created a stronger predictor by modeling the correlation across different visits. For example, if someone missed an appointment yesterday in a cardiology clinic, it is likely that they may miss their appointment tomorrow in a gastrointestinal clinic. In general, more work is needed on how best to model such scenarios, as there are few machine learning methods that properly handle repeated measurements.³¹

There are still some open questions worth considering. Analytically, we only considered discrete models. However, one could better take advantage of the hierarchical nature of the data and derive a Bayesian hierarchical model.³² This would have the advantage of borrowing across the multiple levels of the data and *discover* the optimal level at which to predict. Moreover, this would allow one to naturally handle new, unobserved clinics within in the model, or even reduce the nesting structure lower to the provider level. Similarly, future work could consider how best to borrow information across the different clinical encounters. Finally, this work does not consider the added *cost* in maintaining multiple models. One could apply a cost function to help an institution decide what the optimal trade-off is between model performance and operations.

CONCLUSION

Overall, our results illustrate the value of fitting local level models. Even with small training data sizes, we were able to derive better performing prediction models for patient no-shows. However, this effect was not uniform, suggesting that the optimal model depends on the particular

clinic environment and needs to be learned for the specific context. This highlights the importance of developing “personalized” risk scores.

COMPETING INTERESTS

None.

CONTRIBUTORS

XD analyzed the data and wrote the manuscript. ZFG, ZM, EP, and MN helped design the data extraction protocol. PB extracted all data. BAG designed the analytic study and edited the manuscript.

FUNDING

BAG was supported by National Institute of Diabetes and Digestive and Kidney Diseases grant K25DK097279. ZFG was funded by Veterans Affairs Health Services Research and Development Career Development Award CDA 14-158. This publication was made possible (in part) by Grant Number UL1TR001117 from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCATS or NIH.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

ACKNOWLEDGMENTS

We thank Jenn Gargnon, Mary Schilder, Heidi Banks, Mike Chrestensen, and other members of the PORT team for help with data extraction.

REFERENCES

- Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records: a systematic review. *J Am Med Inform Assoc* 2017;24(1):198–208.
- Hao S, Jin B, Shin AY, et al. Risk prediction of emergency department revisit 30 days post discharge: a prospective study. *PLoS One* 2014;9(11):e112944.
- Low LL, Liu N, Wang S, Thumboo J, Ong MEH, Lee KH. Predicting frequent hospital admission risk in Singapore: a retrospective cohort study to investigate the impact of comorbidities, acute illness burden and social determinants of health. *BMJ Open* 2016;6(10):e012705.
- Eapen ZJ, Liang L, Fonarow GC, et al. Validated, electronic health record deployable prediction models for assessing patient risk of 30-day rehospitalization and mortality in older heart failure patients. *JACC Heart Fail* 2013;1(3):245–251.
- Perkins RM, Rahman A, Bucaloiu ID, et al. Readmission after hospitalization for heart failure among patients with chronic kidney disease: a prediction model. *Clin Nephrol* 2013;80(6):433–440.
- Hebert C, Shivade C, Foraker R, et al. Diagnosis-specific readmission risk prediction using electronic health data: a retrospective cohort study. *BMC Med Inform Decis Mak* 2014;14:65.
- Berg BP, Murr M, Chermak D, et al. Estimating the cost of no-shows and evaluating the effects of mitigation strategies. *Med Decis Mak* 2013;33(8):976–985.
- Kheirkhah P, Feng Q, Travis LM, Tavakoli-Tabasi S, Sharafkhaneh A. Prevalence, predictors and economic consequences of no-shows. *BMC Health Serv Res* 2016;16(1):13.
- Huang Y, Hanauer DA. Patient no-show predictive model development using multiple data sources for an effective overbooking approach. *Appl Clin Inf* 2014;5(3):836–860.
- Cohen AD, Kaplan DM, Kraus M, Rubinshtein E, Vardy DA. Nonattendance of adult otolaryngology patients for scheduled appointments. *J Laryngol Otol* 2007;121(3):258–261.
- Partin MR, Gravelly A, Gellad ZF, et al. Factors associated with missed and cancelled colonoscopy appointments at veterans health administration facilities. *Clin Gastroenterol Hepatol* 2016;14(2):259–267.
- Norris JB, Kumar C, Chand S, Moskowitz H, Shade SA, Willis DR. An empirical investigation into factors affecting patient cancellations and no-shows at outpatient clinics. *Decis Support Syst* 2014;57:428–443.
- Giunta D, Briatore A, Baum A, Luna D, Waisman G, de Quiros FG. Factors associated with nonattendance at clinical medicine scheduled outpatient appointments in a university general hospital. *Patient Prefer Adherence* 2013;7:1163–1170.
- Foley J, O'Neill M. Use of mobile telephone short message service (SMS) as a reminder: the effect on patient attendance. *Eur Arch Paediatr Dent* 2009;10(1):15–18.
- Hogan AM, McCormack O, Traynor O, Winter DC. Potential impact of text message reminders on non-attendance at outpatient clinics. *Ir J Med Sci* 2008;177(4):355–358.
- Woods R. The effectiveness of reminder phone calls on reducing no-show rates in ambulatory care. *Nurs Econ* 2011;29(5):278–282.
- Perron NJ, Dao MD, Kossovsky MP, et al. Reduction of missed appointments at an urban primary care clinic: a randomised controlled study. *BMC Fam Pract* 2010;11:79.
- Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B* 1996;58(1):267–288.
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33(1):1–22.
- Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. *Stat Methods Med Res* 2016;25(4):1692–1706.
- Wynants L, Vergouwe Y, Van Huffel S, Timmerman D, Van Calster B. Does ignoring clustering in multicenter data influence the performance of prediction models? A simulation study. *Stat Methods Med Res* 2016.
- Kappen TH, Vergouwe Y, van Klei WA, van Wolfswinkel L, Kalkman CJ, Moons KGM. Adaptation of clinical prediction models for application in local settings. *Med Decis Mak Int J Soc Med Decis Mak* 2012;32(3):E1–10.
- Wynants L, Bouwmeester W, Moons KGM, et al. A simulation study of sample size demonstrated the importance of the number of events per variable to develop prediction models in clustered data. *J Clin Epidemiol* 2015;68(12):1406–1414.
- Bouwmeester W, Twisk JWR, Kappen TH, van Klei WA, Moons KGM, Vergouwe Y. Prediction models for clustered data: comparison of a random intercept and standard regression model. *BMC Med Res Methodol* 2013;13:19.
- Bardenheier BH, Shefer A, Barker L, Winston CA, Sioneck CK. Public health application comparing multilevel analysis with logistic regression: immunization coverage among long-term care facility residents. *Ann Epidemiol* 2005;15(10):749–755.
- van Klaveren D, Steyerberg EW, Perel P, Vergouwe Y. Assessing discriminative ability of risk models in clustered data. *BMC Med Res Methodol* 2014;14:5.
- Walsh C, Hripicak G. The effects of data sources, cohort selection, and outcome definition on a predictive model of risk of thirty-day hospital readmissions. *J Biomed Inform* 2014;52:418–426.
- Povalej Brzan P, Obradovic Z, Stiglic G. Contribution of temporal data to predictive performance in 30-day readmission of morbidly obese patients. *Peer J* 2017;5:e3230.
- Turner PL, Saager L, Dalton J, et al. A nomogram for predicting surgical complications in bariatric surgery patients. *Obes Surg* 2011;21(5):655–662.
- Shah N, Hamilton M. Clinical review: can we predict which patients are at risk of complications following surgery? *Crit Care* 2013;17(3):226.
- Chen S, Grant E, Wu TT, Bowman FD. Statistical learning methods for longitudinal high-dimensional data. *Wiley Interdiscip Rev Comput Stat* 2014;6(1):10–8.
- Henao R, Lucas JE. Efficient model-based clustering with coalescents: application to multiple outcomes using medical records data. *ArXiv Prepr ArXiv160803191*. 2016.