

Data Analytics and Modeling for Appointment No-show in Community Health Centers

Journal of Primary Care & Community Health

Volume 9: 1–11

© The Author(s) 2018

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/2150132718811692

journals.sagepub.com/home/jpc

Iman Mohammadi¹, Huanmei Wu¹, Ayten Turkcan², Tammy Toscos³, and Bradley N. Doebbeling⁴

Abstract

Objectives: Using predictive modeling techniques, we developed and compared appointment no-show prediction models to better understand appointment adherence in underserved populations. **Methods and Materials:** We collected electronic health record (EHR) data and appointment data including patient, provider and clinical visit characteristics over a 3-year period. All patient data came from an urban system of community health centers (CHCs) with 10 facilities. We sought to identify critical variables through logistic regression, artificial neural network, and naïve Bayes classifier models to predict missed appointments. We used 10-fold cross-validation to assess the models' ability to identify patients missing their appointments. **Results:** Following data preprocessing and cleaning, the final dataset included 73811 unique appointments with 12,392 missed appointments. Predictors of missed appointments versus attended appointments included lead time (time between scheduling and the appointment), patient prior missed appointments, cell phone ownership, tobacco use and the number of days since last appointment. Models had a relatively high area under the curve for all 3 models (e.g., 0.86 for naïve Bayes classifier). **Discussion:** Patient appointment adherence varies across clinics within a healthcare system. Data analytics results demonstrate the value of existing clinical and operational data to address important operational and management issues. **Conclusion:** EHR data including patient and scheduling information predicted the missed appointments of underserved populations in urban CHCs. Our application of predictive modeling techniques helped prioritize the design and implementation of interventions that may improve efficiency in community health centers for more timely access to care. CHCs would benefit from investing in the technical resources needed to make these data readily available as a means to inform important operational and policy questions.

Keywords

access to care, community health centers, predictive modeling, appointment non-adherence, electronic health records

Introduction

Community health centers (CHCs) are safety-net clinics providing primary care for underserved and uninsured populations. For individuals at or below the US federal poverty level, CHCs provide a vital safety health care net. CHCs provide primary care services for acute and chronic diseases, injuries, and preventive services. High missed appointment rates have been identified as one of the most significant barriers to access to care for these populations.^{1,2} In semistructured interviews conducted at CHCs, clinic staff and providers agreed that a high missed appointment rate is a major problem.³

Given financial challenges of delivering quality health care in the United States, finding ways to improve performance is critical in the plight to provide greater access to care. Optimizing scheduling systems has been identified as

one system level approach to address access needs. For example, reducing the number of missed appointments is crucial as when appointment slots go unused it effectively reduces access to others in need of an appointment.⁴ In addition to underutilizing providers' time, missed

¹Department of BioHealth Informatics, School of Informatics and Computing, Indiana University, Indianapolis, IN, USA

²Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA, USA

³Parkview Research Center, Parkview Health System, Fort Wayne, IN, USA

⁴College of Health Solutions, Arizona State University, Phoenix, AZ, USA

Corresponding Author:

Iman Mohammadi, Department of BioHealth Informatics, Indiana University, 719 Indiana Avenue, Indianapolis, IN 46202, USA.
Email: imanmoha@iu.edu



appointments impact waits and delays for others, increase health care costs, and increase possibility for adverse health outcomes.^{5,6} Research has shown that lowering missed appointment rates can improve clinical efficiency and utilization, reduce waste, improve provider satisfaction and lead to better health outcomes for patients.^{7,8} Missed appointment rates range from 10% to 50% across healthcare settings in the world with an average rate of 27% in North America.⁶ Patients with higher missed appointment rates are significantly more likely to have incomplete preventive cancer screening, worse chronic disease control and increased rates of acute care utilization.⁹ In previous studies, missed appointments have been due to logistical issues, lack of understanding of the scheduling system, patients not feeling respected by healthcare providers or the health system, affordability, timeliness, patients forgetting appointment and patient severity of illness.^{6,10}

To understand the complexity of appointment adherence in different health care settings, different datasets, variables, and data volumes have been studied. Medium-scale studies (ranging from 6,000 to 8,000 patients) focused on a few patient characteristics or a single (eg, time) component.¹¹⁻¹³ For example, a large-scale no-show modeling of a Veterans Affairs (VA) outpatient clinic included 555,183 patients, which scheduled 25,050,479 appointments; however, the study only considered a few variables such as the patient gender, the date of the appointment, and new versus established patients.¹⁴ Most studies developed regression models to predict appointment nonadherence.^{12,15} Most similar to the present study, one study identified predictors of missed clinic appointments among an underserved population.¹⁶ These results revealed predictors for a missed appointment included percentage of no-shows in patients previous appointments (no-show or cancellation within 24 hours), wait time from scheduling to appointment, season, day of the week, provider type, and patient age, sex, and language proficiency. In other studies of predictive modeling in health care arena using electronic health record (EHR) data, other predictive modeling techniques such as naïve Bayes classifier¹⁷ and neural network¹⁸ were used to predict hospital readmissions. In this study, we apply and build on these techniques to predict appointment no-show in CHCs.

Here, we test missed appointment prediction models by analyzing EHR and scheduling data. We aim to exploit predictive modeling to improve understanding of the complexity of appointment adherence in underserved populations. Information about patients, providers, appointments and time are used to predict patients' adherence to appointments. The main contributions of this study are to (a) build on previous no-show modeling in community health centers by expanding the focus on various outpatient specialties and underserved population specific predictors; (b) compare different predictive modeling methodologies, namely logistic regression, naïve Bayes classifier, and artificial

neural networks (specifically multilayer perceptron); and (c) investigate the impact of clinic characteristics on predictors of the no-show.

Materials and Methods

Participants

Data for this project were collected from a large urban multisite community health center, involving 10 locations in Indianapolis, most of which are considered federally qualified health centers (FQHC). This CHC has provided care for more than 100,000 patients during 2014 to 2016. Health care services provided by this CHC include but not limited to primary care, pediatrics, family practice, internal medicine, obstetrics/gynecology, dental care, vision care, behavioral health services, and preventive care. The goal of the no-show modeling was to focus on primary care, so data on dental and vision care visits was not considered. All study methods were approved by our institutional review board.

Data Collection and Sample Size

We extracted and deidentified semistructured data from over 17 tables in the CHC's database from 2010 to 2016 to address the study aim. EHR data, including clinic (ie, operational and financial data) and patient (ie, patient demographics and clinical characteristics) information, were included and linked at the patient level. The data was stored in a secure Microsoft SQL Server with limited access. For this study, we created a dataset of patients' encounters from January 1, 2014 to April 30, 2016. The dataset included 599,636 appointments by 76,453 unique patients (Table 1).

Data Preprocessing

Appointment compliance field was the dependent variable in this analysis, which included the categories of checkout (ie, complete) appointment, no-show, cancelled, rescheduled, and others. A no-show appointment is defined as a patient who did not keep the prescheduled appointment and did not cancel the appointment at least 24 hours ahead of the appointment time. We focused on appointments scheduled with medical doctor, nurse practitioner, or certified nurse-midwife. All other nurse visit appointments were excluded from analyses. We performed the following data filtering steps:

- Filtering appointment categories: To create the binary outcome variable in this study, we only included no-show and checkout appointments in the final analysis, and observations having other appointment compliance, such as rescheduled, cancelled, and so on, were censored from the dataset.

Table 1. Distribution of Patient Characteristics Versus Appointment Adherence.

Patient Characteristics		Appointment Adherence		P ^a
		Attended (n = 61,419)	Missed (n = 12,392)	
<i>Categorical variables, Percentages</i>				
New patient	Yes	2.1	2.4	.0455
Translator needed	Yes	15.2	8	<.0001
Ethnicity	Hispanic or Latino	19.6	11.9	<.0001
	Not Hispanic or Latino	75	80.2	
	Unspecified	5.4	7.9	
Race	American Indian or Alaska Native	0.1	0.1	<.0001
	Asian	4.2	2	
	Black	30.3	37.7	
	Multiple races	3.9	3.7	
	Native Hawaiian and Other Pacific Islander	1.1	0.7	
	White	60.4	55.7	
Gender	Female	61.4	64.8	<.0001
Marital status	Divorced	3.3	3.1	<.0001
	Legally separated	1.3	1.7	
	Married	12.8	9.5	
	Partner	0.4	0.3	
	Single	80.8	83.4	
	Widowed	1.2	0.8	
Cell phone ownership	No	18.2	26.4	<.0001
Email availability	No	70.6	74.5	<.0001
Using patient portal	No	78.2	83.5	<.0001
Employment status	Employed full-time	13	10.8	<.0001
	Employed part-time	5.1	5.5	
	Not employed	79.6	82.4	
	Retired	1.5	0.4	
	Self-employed	0.5	0.3	
	Insurance	Commercial	14.8	8.4
	Marketplace	0.6	0.3	
	Medicaid	66.8	69	
	Medicare	5.6	3.6	
	Self-pay	12.2	18.7	
Tobacco use	Current every day smoker	22.8	35.5	<.0001
	Current some day smoker	2.8	3.4	
	Former smoker	13	12	
	Never smoker	61.3	49.1	
<i>Continuous variables, Mean (SD)</i>				
Age (years)		21.1 (19.4)	21.4 (16.9)	.1393
Annual income		\$2748 (8421)	\$2046 (7109)	<.0001
Prior no-show Rate		0.11 (0.2)	0.2 (0.3)	<.0001

^aT test for continuous variables and chi-square for categorical variables.

- Ensuring appointment independences: To ensure observations are independent from each other, we only included the last appointment of each patient in the final analysis.
- Handling missing information: unstructured free text fields, such as schedulers' notes, were used to complete any missing values in fields, such as appointment

type, patient age or gender. Simple rules were used to find visit types from scheduler notes. For example, if the note contained "acute" and visit type field was missing, visit type field was filled by "Acute care", and other types can be seen in Table 3. All other observations with missing information were removed from the dataset.

Table 2. Distribution of Provider Characteristics Versus Appointment Adherence.

Variables	Appointment Adherence		P ^a	
	Attended (n = 61,419)	Missed (n = 12,392)		
Categorical variables, Percentages				
Provider specialty	Behavioral Health	1.8	4.3	<.0001
	Certified Nurse-Midwife	9.5	12.7	
	Family Medicine	17.1	14.7	
	Internal Medicine	11.5	11.7	
	Nurse Practitioner	9.9	7.3	
	Obstetrics/Gynecology	4.3	5.9	
	Pediatrics	33.6	30.3	
	Podiatry	0.7	1.4	
Patient's primary care practitioner?	No	83.2	86.6	<.0001

^aT test for continuous variables and chi-square for categorical variables.

Out of 76,453 unique patients, 2,642 patients were removed because they had observations with missing information that could not be found in the data. The final dataset included 73,811 observations of unique individuals, and whether they showed for their last appointment during the study period. Data imputation was not necessary as we had sufficient number of observations for our analyses.

Variable Preparation

Data fields included visit characteristics (facility/clinic type, date of visit, date contacted the clinic for scheduling the visit, time of visit, visit duration, and visit type), patient characteristics (patient pseudo-ID, age, race, ethnicity, gender, marital status, cell phone ownership, email availability, whether using patient portal, employment status, tobacco use, income, needing translator and primary insurance), provider characteristics (whether seeing the patients' primary care practitioner [PCP] or not, specialty and medical license) and appointment compliance ("no-show" or "check out").

In addition to the existing variables in the EHR, we created the following variables to consider in our no-show modeling:

1. Lead time, which is the time difference (in days) between the date of visit and date the patient had contacted the clinic to arrange an appointment.
2. Prior no-show rate, which is the number of no-shows for a given patient prior to the last appointment, divided by the patient's total number of appointments prior to the last appointment. We used this to test the effect of patient no-show behavior on appointment adherence.

3. Days since the last appointment, which is the difference between the date of the last visit and the date of appointment before the last visit.

Statistical Analyses

We hypothesized that patient and provider characteristics and visit features were all predictors of appointment no-show in CHCs. We tested variables individually for relationships with the appointment adherence using a chi-square test for categorical variables and *t* test for continuous variables. We chose variables with a *P* value less than .2 to enter into the model development step. Tables 1-3 list variables that were included in the modeling. The dataset included 73,811 observations, 83% arrived and 17% no-show.

Prediction Model Development

We randomly split the dataset into 2 samples: 70% for the training (or derivation) set and 30% for the test (or validation) set. This training and test set selection was repeated 10 times to overcome selection bias. To decrease potential bias of learning algorithm for training set, we randomly selected training subsets with no-show to checkout ratio of 2 to 1 and repeated this randomization for 10 times. We used the training subsets to develop the no-show prediction model using 3 methodologies:

1. Logistic regression: We used logistic regression in SAS 9.4. to develop the prediction model with a stepwise selection and significance level of $\alpha = .01$. All the variables, shown in Tables 1-3 and their interactions, were included in the model development.

Table 3. Distribution of Visit Characteristics Versus Appointment Adherence.

Variables	Appointment Adherence		P ^a
	Attended (n = 61,419)	Missed (n = 12,392)	
Categorical variables, Percentages			
Appointment duration (minutes)	10	0.8	<.0001
	15	68.3	
	20	14.3	
	30	15.6	
	45	0.5	
	60	0.5	
Lead time	Same day	31.4	<.0001
	Next day	9	
	Within 2 weeks	31.6	
	Between 2 weeks and 1 month	13	
	More than 1 month	15	
Days since last appointment	Within a week	1.4	<.0001
	Between 1 and 2 weeks	1.1	
	Between 2 weeks and 1 month	2.3	
	Between 1 and 3 months	5.6	
	Between 3 and 6 months	6.7	
	Between 6 months and a year	14.5	
	More than a year	53.7	
	No prior appointment since 2014	14.8	
Appointment time	AM	43.8	.1294
Season	Fall	18.1	<.0001
	Spring	29.9	
	Summer	15.1	
	Winter	36.9	
Weekday	Monday	22.3	<.0001
	Tuesday	21.9	
	Wednesday	20.1	
	Thursday	18.8	
	Friday	15.8	
	Saturday	1.1	
	Visit type	Acute care	
Adult routine/Follow-up	17		
Behavioral health	2		
Podiatry	0.7		
Pediatric	37.6		
Pregnant	4.5		
Women	10.5		

^a T test for continuous variables and chi-square for categorical variables.

- Artificial neural network: The large number of features and observations in this study led us to use more complex machine learning algorithms such as multilayer perceptron. Multilayer perceptron consists of multiple linear regression models are advantageous when there is a large number of features (variables) with complex relations among them.¹⁹ Categorical variables were transformed to numeric variables. For example, if a patient is a “New

Patient,” the numeric variable of New Patients would be created with a value of 1. Continuous, binary, and numeric variables were used as inputs for a multilayer perceptron and 1 binary variable (No-show = 1 or 0) was used as output. Matlab software was used to develop the multilayer perceptron in this project having 3 layers of the input layer, hidden layer including 25 nodes and output layer. The training data subsets were used to train the network

by minimizing the mean-square error (MSE) between the desired output and the actual output of the network. The value of the output node determined the classification using a range (between 0 and 1) of cutoff thresholds. Here, we used absolute value of weights for input layer nodes to identify and rank the most important variables contributing to no-show prediction.

3. Naïve Bayes classifier: The majority of predictors in our datasets were categorical; hence, we applied a naïve Bayes classifier that is appropriate to categorical data.²⁰ This classifier computes a conditional probability of each category in each variable given the outcome. Then, Bayes rules are applied to calculate the probability of the outcome given different categories of variables in the data. We applied the naïve Bayes classifier algorithm implemented in “scikit-learn” in Python over the randomly selected train and test datasets. The smoothing value of 0.1 provided the best performance for the classifier.

Model Validation

Models were assessed by calculating the area under the curve for the receiver operating characteristic (AUC-ROC) curve. Test dataset was used to validate models' ability to discriminate between patients who no-showed versus those who attended. Ten-fold cross-validation was used to validate the 3 models, and average AUCs, sensitivities to predicting no-show and overall model accuracy were the key indicators of model validation.

Results

Statistical Analyses

The final dataset included 73,811 observations with 12,392 missed appointments. Comparative analyses of patient characteristics revealed that black, non-Hispanic or non-Latino, female, single, not employed, Medicaid, self-pay, or smoker patients had a higher chance of missed appointments ($P < .0001$; see Table 1). The average annual income is lower, and the average prior missed appointment rate is higher in patients who no-showed in their last appointment ($P < .0001$). Patients without a cell phone, email, or patient portal had a higher chance of a missed appointment ($P < .0001$). The comparative analysis of the provider characteristics showed that patients scheduled with behavioral health or OB-GYN providers or not scheduled with their primary care providers have higher missed appointment rates compared with other appointment types ($P < .0001$), as demonstrated in Table 2.

The appointment duration, the time between appointment days, and the day appointment requested, the time (daytime,

weekday, or season) of an appointment, and the type of an appointment are statistically significantly different between checkout and missed appointment patients ($P < .0001$), as shown in Table 3.

Table 4 shows characteristics of 10 facilities within this CHC system. Clinics are different in terms of missed appointment rates and distributions of patient type, visit type, and provider type.

Predictive Modeling

As shown in Table 4, clinics had different population sizes, characteristics, and no-show rates. Therefore, we developed a separate logistic regression model for each clinic. Supplementary Table S1 (available in the online version of the article) shows the results from regression model development. These separate models corresponding to individual clinics yielded different predictors for missed appointments. Notably, lead time, prior missed appointment rate, age, insurance type, tobacco use, days since the last appointment, and cell phone ownership were consistent significant factors across clinics.

Patient Characteristics. Table 4 demonstrates that clinic 2 patients had lower prior missed appointment rates compared with other clinics. In all clinics except clinic 6, patients between 18 and 64 years old were 1.6 (99% CI 1.5-1.6) and 3.7 (99% CI 2.9-4.6) times more likely to no-show their next appointments compared with patients between 0 to 17 years old and 65 years and older patients, respectively. Notably, clinic 6 is a pediatric clinic and patients are dominantly between 0 and 17 years old. Patients who needed a translator in their appointments, particularly in clinic 7 (with a high proportion of Hispanic or Latinos), were 0.5 times less likely to no-show in their next appointments (99% CI 0.4-0.5). In 2 clinics, the interaction between age and gender also influenced no-shows.

Insurance status was another significant predictor of missed appointments, such that insured patients were less likely to keep their appointments. In most clinics, patients insured by commercial, marketplace, Medicaid, and Medicare plans were 0.4 (99% CI 0.3-0.4), 0.3 (99% CI 0.2-0.5), 0.7 (99% CI 0.6-0.7), and 0.4 (99% CI 0.37-0.50) times as likely to miss appointments, compared with their uninsured counterparts. Smoking daily increased the likelihood of missed appointments by 95%, compared with patients who never smoked (odds ratio OR = 2, 99% CI 1.8-2.1). Patients using for their clinics patient portal (web-enabled) were less likely to no-show in their appointments (OR = 0.7, 99% CI 0.7-0.8). In clinic 5, patients without an email address recorded in the EHR system are 1.2 times more likely to no-show (99% CI 1.21-1.23). Patients without a cell phone number available in the records were 1.6 times more likely to no-show (99% CI 1.52-1.71).

Table 4. Clinic Characteristics.

Facility	Total No. of Patients	No-show		Clinic Characteristics	Percentage/Mean Among All Clinics
		Frequency	Percentage		
Clinic 1	10,633	2,248	21	<ul style="list-style-type: none"> • Large number (23%) of patients needing translator • Large number (20%) of Asian patients • Highest mean lead time (28.6 days) 	<ul style="list-style-type: none"> • 14%, $P < .0001$ • 4%, $P < .0001$ • 17 days, $P < .0001$
Clinic 2	3,680	660	18	<ul style="list-style-type: none"> • Higher percentage of new patients (10.3%) • Dominantly pregnant and woman patients (98%) • Dominantly certified nurse-midwife and obstetrics/gynecology providers (95%) • Dominantly female patients (98%) • Dominantly adult patients (95%) • Patients with lower prior no-show rates (0.08) 	<ul style="list-style-type: none"> • 2.1%, $P < .0001$ • 15.8%, $P < .0001$ • 15%, $P < .0001$ • 62%, $P < .0001$ • 43%, $P < .0001$ • 0.12, $P < .0001$
Clinic 3	3,206	392	12	<ul style="list-style-type: none"> • Mostly scheduled with patients' primary care practitioners (56%) • Patients with lower prior no-show rates (0.08) 	<ul style="list-style-type: none"> • 16.2%, $P < .0001$ • 0.12, $P < .0001$
Clinic 4	6,731	803	12	<ul style="list-style-type: none"> • Majority Black (77%) • Mostly same-day appointments (67%) • Higher number of acute care appointments (46%) 	<ul style="list-style-type: none"> • 32%, $P < .0001$ • 27%, $P < .0001$ • 25%, $P < .0001$
Clinic 5	2,216	480	22	<ul style="list-style-type: none"> • Highest no-show rate 	<ul style="list-style-type: none"> • 17%, $P < .0001$
Clinic 6	7,870	1,543	20	<ul style="list-style-type: none"> • Mostly 20-minute appointments (79%) • Dominantly children (97%) • Majority black (63%) • Dominantly not employed (98%) 	<ul style="list-style-type: none"> • 14%, $P < .0001$ • 55%, $P < .0001$ • 32%, $P < .0001$ • 80%, $P < .0001$
Clinic 7	10,703	1,916	18	<ul style="list-style-type: none"> • Large number (23%) of patients needing translator • Higher number of Hispanic or Latino (34%) 	<ul style="list-style-type: none"> • 14%, $P < .0001$ • 18%, $P < .0001$
Clinic 8	12,016	1,659	14	<ul style="list-style-type: none"> • Large number (22%) of patients needing translator • Higher number of Hispanic or Latino (31%) • Highest income level (\$4553/year) 	<ul style="list-style-type: none"> • 14%, $P < .0001$ • 18%, $P < .0001$ • \$2665, $P < .0001$
Clinic 9	11,521	1,942	17	<ul style="list-style-type: none"> • Dominantly white (85%) 	<ul style="list-style-type: none"> • 60%, $P < .0001$
Clinic 10	5,235	749	14	<ul style="list-style-type: none"> • Patients with lower prior no-show rates (0.08) 	<ul style="list-style-type: none"> • 0.12, $P < .0001$

Scheduling Characteristics. Lead time was the most consistent significant factor across all the clinics. Longer lead time provides greater opportunity for a missed appointment ($P < .0001$). Appointments made more than 1 month in advance are 7.1 (99% CI 6.5-7.5), 2.4 (99% CI 2.2-2.7), 1.7 (99% CI 1.6-1.9), and 1.2 (99% CI 1.1-1.3) times more likely to become a no-show, compared with appointments made on same day, 1 day, 2 weeks, and between 2 weeks and 1 month in advance, respectively. Next day appointments were 2.9 times more likely to become a missed appointment than same day appointments (99% CI 2.6-3.3). Patients with a history of missed appointments were 4.9 times more likely to miss their next appointments (99% CI 4.4-5.8), in all clinics except clinic 2. Patients who had an appointment between 1 and 2 weeks prior to their last appointment were more likely to miss that last appointment compared with patients who had a prior appointment in the last 6 to 12 months (OR = 1.5, 99% CI 1.2-1.8), more than 12 months (OR = 2.2, 99% CI 1.8-2.7), or patients who had no prior appointments (OR = 1.4, 99% CI 1.1-1.7).

Clinic Visit Characteristics. In one-half of the clinics, type of visit predicted appointment adherence. Supplementary Table S1 shows that acute visits had lower missed appointment rates than all other visit types, while behavioral health visits had the highest missed appointment rates. Seasonality of the appointments predicted missed appointments such that appointments occurring during spring or summer had higher missed appointment rates than winter appointments. Notably, patients scheduled with their own PCP were less likely to miss the appointment than the ones scheduled with other providers (OR = 0.8, 99% CI 0.7-0.8). Appointment duration was also a significant factor (particularly in clinics 3 and 5). Longer durations such as 1 hour or 45 minutes were more likely to be no-show than shorter durations such as 15 or 20 minutes.

The ranking of variables contributing to prediction of no-show in the multilayer perceptron are shown in Supplementary Table S2. The ranking is based on the weights nodes in the input layer of multilayer perceptron. The top 10 predictors of the no-show in our multilayer perceptron analyses included: lead-time, provider specialty,

Table 5. Validation and Comparison of Prediction Models.

Modeling Method	Train Set				Test Set			
	AUC	Sensitivity	Positive (No-show) Predictive Value	Overall Accuracy (%)	AUC	Sensitivity	Positive (No-show) Predictive Value	Overall Accuracy (%)
Logistic regression	0.91	0.84	0.58	80	0.81	0.72	0.54	73
Multilayer perceptron	0.77	0.73	0.43	79	0.66	0.63	0.35	71
Naïve Bayes classifier	0.96	0.82	0.67	92	0.86	0.73	0.45	82

Abbreviation: AUC, area under the curve for the receiver operating characteristic curve.

race, employment status, days since last appointment, prior no-show rate, cell phone ownership, tobacco use, marital status, and gender. Similarly, there were multiple variables contributing to no-show (Supplementary Table S3) using the naïve Bayes classifier. prior no-show rate, age group, visit type, lead-time, days since last appointment, duration, insurance, cell phone ownership, tobacco use, and ethnicity are the top 10 factors predicting next appointment no-show. Those variables important in all three types of models included: lead time, patient prior no-show behavior, cell phone ownership, tobacco use, and the number of days since the last appointment of patient. Logistic regression and naïve Bayes classifier have commonly identified visit type, age, and insurance as top 10 predictors.

Model Validation

Table 5 shows the validation results for 3 models. Overall accuracy in Table 5 is the correct classification ratio for the model. The AUC for logistic regression and the naïve Bayes classifier are, respectively, 0.81 and 0.86, which are considered excellent for discriminating between 2 outcomes.⁶ Multilayer perceptron had low AUC of 0.66.

Discussion

We studied missed appointments in 10 separate clinics within one urban community health care system. Our study shows that clinics have different population characteristics, specialties, and patient demographics; thus, it is not surprising that appointment adherence varies across geographic sites. For example, specialty clinics such as pediatric or woman clinics have higher missed appointment rates than the ones providing acute or general primary care. Appointment lead time, past missed appointments, and age group of patients are the common important factors differentiating clinics' overall missed appointment rate. Our study suggests that any attempt to create a missed appointment prediction model or to design interventions for reducing missed appointment rates should be clinic/facility specific and tailored based on clinic, facility, or department characteristics.

Our study has 4 major findings. First, patient, scheduling, and visit characteristics differ across missed and arrived appointments. These characteristics should be of interest to managers and policy makers, in order to better design interventions and policies to reduce missed appointments. Second, the consensus of the logistic regression, multilayer perceptron, and naïve Bayes classification was that lead-time, patient prior missed appointments, cell phone ownership, tobacco use, and the number of days since the last appointment of a patient are the most significant predictors of missed appointments. Other factors were important in certain clinics, even after control for these factors. These findings should help managers in health care systems prioritize the design and implementation of interventions to reduce missed appointments. Third, patient appointment adherence had different determinants in different clinics or facilities within a single health care system. This finding makes sense in a large urban area, where neighborhood, population and clinic characteristics, as well as policies and procedures differ. It also underlines the importance of looking at data at the clinic level, because different clinics, even within the same system may have an important population and organizational differences. Fourth, according to the accuracy of the predictions, logistic regression and Bayes classifiers concluded similarly and perform better in missed appointment modeling than a multilayer perceptron. This might be because of categorical nature of our data. Studies have reported that the discrimination ability of neural networks (such as multilayer perceptron) versus other statistical modeling techniques is data specific.²¹

Poverty, Employment, and Access to Health Information Technology

One key social determinant of health in populations is economic stability; this includes measures such as education, poverty, and employment status.²² We found that lower income and unemployment were associated with more missed medical appointments that would likely impair the health and/or health outcomes of patients. Studies found that socioeconomic characteristics have negative impact on health outcomes.²³

The role of poverty and employment are obviously complex and multifactorial across the United States. Our findings point to the need for social, financial, and educational interventions to help indigent people prosper and communities thrive. Access to emerging technologies such as cell phones, the Internet and social media is another social and financial determinant. We found that patients without access to cell phone, email, and a patient portal were more likely to miss their medical appointments. Therefore, lack of access to these technologies may affect health outcomes. Future research should examine if the provision of these consumer health technologies alone can enhance access to health for individuals in poverty or if our finding is more directly related to financial status alone.

Our results show that patients without insurance for medical services are at risk of not adhering to their appointments and consequently their care plans. This factor is highly correlated with unemployment, which was very high (approximately 80%) in our study population.

Patient Engagement, Tobacco Use, and Promoting Patient Appointment Adherence

In our study, smoking was one of the most significant factors related to missing medical appointments. We hypothesize that this variable as a health behavior, which may be highly related to other health practices, including adherence to scheduled clinic visits. It is beyond the scope of this study to determine whether this variable is a marker for adherence with recommendations or a confounder. Regardless, its importance underscores the importance of engagement of the underserved populations in their care and the role of individual health behaviors, attitudes, and practices.

Common reasons for missed appointments found in prior research include forgetting about the appointment, competing priorities, and demands (such as the need to work or inability to leave work), availability of transportation, or feeling better at the time of the appointment.²⁴ These reasons can be magnified if the lead time (the most important predictor in our study) for appointments is elongated. Interventions such as increasing number of open access (same-day) hours and decreasing number of appointments made more than 1 month in advance should be considered to improve access to care in community health centers. Past missed appointment is an important predictor of future appointment adherence. Our findings are consistent with other research that operationalized passed missed appointments using clinicians' notes containing phrases like "no-show," "did not present," "failed to attend," and "missed appointment." These researchers found that patients that previously missed appointments were more likely to miss future appointments.²⁵ Further investigation of this problem should focus on extracting important information available as free text in patient complaint and reason to seek health.²⁶

Our study found that behavioral health patients were more likely to miss their next appointments than any other type of patients. Differences in adherence with appointments here could either be related to different systems for scheduling and reminding patients of appointments between medical and behavioral health systems, or related to intrinsic differences in practices, attitudes or adherence among behavioral health patients. Further investigation of this problem should focus on differences between the practices and policies for such patients, before efforts to make special accommodations for the population.

Application to Medical Practice

Our study used large patient datasets with multiple potential explanatory variables in order to develop prediction models using various clinics within a health care system. We also used multiple methods to develop and compare the models. Access to health care can affect individuals' health status and quality of life. Missed appointments are one of the most important factors determining access to care. High levels of no-shows are not only an expensive waste of limited provider resources, but they can also lead to unmet health needs and delays in receiving appropriate care. Therefore, predicting and preventing missed appointments can potentially improve access to care.²⁷ The outcomes of this study could help clinicians predict appointment no-shows that can potentially reduce no-show rates in CHCs. Researchers have reported lower no-show rates can improve clinical efficiency and utilization, reduce waste, improve provider satisfaction and lead to better health.²⁸ Redesigning and testing the alternate scheduling processes will help patients get appointments in a timelier manner. These better scheduling systems will improve access for acute patients, increase continuity of care for chronic patients and essentially positively affect health outcomes.

There are 2 possible real-world applications of this study. First, the methodologies and findings of this study can be used to redesign scheduling systems in CHCs to reduce the number of no-show appointments. Second, no-show predictions models can be implemented in EHR systems as decision support systems that would identify patients with a high risk of appointment no-show. Appointments with high risk of no-show may be double booked, or patients with high risk of no-show may be reminded more rigorously.

Limitations

One of the limitations of this study is that it includes only patients from 1 CHC system in Indianapolis. However, this CHC system involves multiple geographic sites and is very diverse from the patient characteristics perspective. Another limitation of this study is that the dataset used in this study did

not have information on the clinical, physical, and functional status of patients (eg, diabetes, depression, congestive heart failure, etc). These attributes can be significant predictors of the no-show. However, visit type variable in our dataset did relate to a patient's clinical characteristics. Findings of this study are drawn from FQHC clinics providing primary care to underserved populations. Whether these results are generalizable to other patient populations will need to be addressed in other studies. Another limitation of this study is that the dataset did not include information about new patients who no-showed in their first appointments; however, sufficient number of observations did not significantly impact the outcomes of this study.

Future Work

These results demonstrate the value of using existing clinical and operational data to address important operational issues. Further resources are needed in CHCs to make these data readily available and to inform important operational and policy questions. Future work might also focus on linking billing information and claims data with EHR to extract important information about patients and appointments. One example could be using evaluation and management codes to adequately identify provider type or provider time spent with patients.

Conclusion

This project developed the statistical model and machine learning models that can be used to predict patients' chance of no-showing to their next medical appointment. Logistic regression, multilayer perceptron, and naïve Bayes classifiers were used to develop and compare the no-show prediction models that resulted in finding lead time, patient prior no-show behavior, cell phone ownership, tobacco use, and the number of days since the last appointment of a patient as significant predictors of appointment adherence. The application of these findings may be used to design new interventions to improve scheduling processes and other policies and practices for better and timelier access to care. We suggest that redesigned operations and policies, from scheduling practices to reminder systems and other technological tools to improve adherence can improve clinic revenues, utilization of resources, and ultimately improve health outcomes.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This

work was supported through a Patient-Centered Outcomes Research Institute (PCORI) Award (IH-12-11-5488).

Supplemental Material

Supplemental Material for this article is available online.

ORCID iD

Iman Mohammadi  <https://orcid.org/0000-0003-2775-1149>

References

1. Mohammadi I, Turkcan A, Toscos T, Miller A, Kunjan K, Doebbeling BN. Assessing and simulating scheduling processes in community health centers. Paper presented at: AMIA 2015 Annual Symposium; November 14-18, 2014; San Francisco, CA.
2. Doroudi R, Mohammadi I, Turkcan A, Toscos T, Wu H, Doebbeling B. Agent-based simulation to test optimal scheduling scenarios and improve access to care for underserved populations. Paper presented at: Institute of Industrial Engineers Annual Conference and Expo; May 21-24, 2016; Anaheim, CA.
3. Toscos T, Carpenter M, Flanagan M, Kunjan K, Doebbeling BN. Identifying successful practices to overcome access to care challenges in community health centers: a "positive deviance" approach [published online March 8, 2018]. *Health Serv Res Manag Epidemiol*. doi:10.1177/2333392817743406
4. Corrigan JM. Crossing the quality chasm. In: Reid PP, Compton WD, Grossman JH, Fanjiang G, eds. *Building a Better Delivery System: A New Engineering/Health Care Partnership*. Washington, DC: National Academic Press; 2005.
5. Deyo RA, Inui TS. Dropouts and broken appointments. A literature review and agenda for future research. *Med Care*. 1980;18:1146-1157.
6. Turkcan A, Nuti L, DeLaurentis PC, et al. No-show modeling for adult ambulatory clinics. In: Denton B, ed. *Handbook of Healthcare Operations Management: Methods and Applications*. New York, NY: Springer; 2013:251-288.
7. Molfenter T. Reducing appointment no-shows: going from theory to practice. *Subst Use Misuse*. 2013;48:743-749.
8. Nguyen DL, Dejesus RS, Wieland ML. Missed appointments in resident continuity clinic: patient characteristics and health care outcomes. *J Grad Med Educ*. 2011;3:350-355.
9. Hwang AS, Atlas SJ, Cronin P, et al. Appointment "no-shows" are an independent predictor of subsequent quality of care and resource utilization outcomes. *J Gen Intern Med*. 2015;30:1426-1433.
10. Daggy J, Lawley M, Willis D, et al. Using no-show modeling to improve clinic performance. *Health Informatics J*. 2010; 16:246-259.
11. Samorani M, LaGanga LR. Outpatient appointment scheduling given individual day-dependent no-show predictions. *Eur J Oper Res*. 2015;240:245-257.
12. Huang Y, Hanauer D. Patient no-show predictive model development using multiple data sources for an effective overbooking approach. *Appl Clin Inform*. 2014;5:836-860.

13. Huang YL, Hanauer DA. Time dependent patient no-show predictive modelling development. *Int J Health Care Qual Assur.* 2016;29:475-488.
14. Davies ML, Goffman RM, May JH, et al. Large-scale no-show patterns and distributions for clinic operational research. *Healthcare (Basel).* 2016;4:E15.
15. Huang Y, Zuniga P. Effective cancellation policy to reduce the negative impact of patient no-show. *J Oper Res Soc.* 2014; 65:605-615.
16. Torres O, Rothberg MB, Garb J, Ogunneye O, Onyema J, Higgins T. Risk factor model to predict a missed clinic appointment in an urban, academic, and underserved setting. *Popul Health Manag.* 2015;18:131-136.
17. Shameer K, Johnson KW, Yahi A, et al. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using Mount Sinai Heart Failure Cohort. *Pac Symp Biocomput.* 2017;22:276-287.
18. Ottenbacher KJ, Smith PM, Illig SB, Linn RT, Fiedler RC, Granger CV. Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke. *J Clin Epidemiol.* 2001;54:1159-1165.
19. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform.* 2002;35:352-359.
20. Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Mach Learn.* 1997;29:131-163.
21. Ayer T, Chhatwal J, Alagoz O, Kahn CE Jr, Woods RW, Burnside ES. Informatics in radiology: comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radiographics.* 2010;30:13-22.
22. Taylor LA, Tan AX, Coyle CE, et al. Leveraging the social determinants of health: what works? *PLoS One.* 2016;11:e0160217.
23. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *J Epidemiol Community Health.* 2001;55:111-122.
24. Kaplan-Lewis E, Percac-Lima S. No-show to primary care appointments: why patients do not come. *J Prim Care Community Health.* 2013;4:251-255.
25. Blumenthal DM, Singal G, Mangla SS, Macklin EA, Chung DC. Predicting non-adherence with outpatient colonoscopy using a novel electronic tool that measures prior non-adherence. *J Gen Intern Med.* 2015;30:724-731.
26. Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: a literature review. *J Biomed Inform.* 2018;77:34-49.
27. Agency for Healthcare Research and Quality. *2015 National Healthcare Quality and Disparities Report and 5th Anniversary Update on the National Quality Strategy.* Rockville, MD: Agency for Healthcare Research and Quality; 2016.
28. Dugdale DC, Epstein R, Pantilat SZ. Time and the patient-physician relationship. *J Gen Intern Med.* 1999;14(suppl 1): S34-S40.

Author Biographies

Iman Mohammadi, PhD, is an expert in health data science. He is experienced in developing and implementing simulation modeling, predictive modeling, and machine learning algorithms their applications in healthcare and engineering domains. He has worked on multiple big data driven projects in healthcare arena using large electronic health records and healthcare claims data.

Huanmei Wu, PhD, is chair of the department of BioHealth Informatics at School of Informatics and Computing at Indiana University. Her background has served her well in multidisciplinary research, leading to projects that explore the roles and applications of data management and analytics to medical and life science research. Her research interests include scientific database design and utilization, big data analytics, machine learning, integrative prediction, bioinformatics, personalized healthcare, and precision medicine through multi-omics data mining.

Ayten Turkcan, PhD, has expertise on application of operations research and statistical methods to improve healthcare systems. She worked on multiple healthcare projects on appointment scheduling, no-show modeling, staffing, and capacity planning in several healthcare settings including primary care, oncology, surgery, mental health and hepatology clinics.

Tammy Toscos, PhD, conducts user-centered design research aimed at developing technologies that enable people of all ages to establish healthy lifestyles aligned with their personal values and goals. She leads an interdisciplinary research team that supports health services research and innovation within a large not-for-profit health system, Parkview Health. She is focused on finding optimal ways to leverage data and technology to enhance health literacy, decision-making, and patient-provider communication as a means of empowering individuals to better manage chronic disease, general health, and wellness.

Bradley N. Doebbeling, MD, MSc, is professor of science of Health Care Delivery and Biomedical Informatics at College of Health Solutions at ASU, Phoenix, AZ. He has expertise in implementation science, system redesign, clinical decision support and data analytics and modeling.