IIE Transactions on Healthcare Systems Engineering (2015) **5**, 14–32

ISSN: 1948-8300 print / 1948-8319 online
DOI: 10.1080/19488300.2014.993006

A hybrid prediction model for no-shows and cancellations of outpatient appointments

ADEL ALAEDDINI[1,*], KAI YANG[2], PAMELA REEVES[3] and CHANDAN K. REDDY[4]

[1]*Department of Mechanical Engineering, University of Texas at San Antonio, San Antonio, TX, USA*
E-mail: adel.alaeddini@utsa.edu
[2]*Department of Industrial and Systems Engineering, Wayne State University, Detroit, MI, USA*
[3]*Dingell VA Medical Center, MI, USA*
[4]*Department of Computer Science, Wayne State University, Detroit, MI, USA*

Received June 2012 and accepted October 2014

A no-show occurs when a scheduled patient neither keeps nor cancels the appointment. A cancellation happens when individuals contact the clinic and cancel their scheduled appointments. Such disruptions not only cause inconvenience to hospital management, they also have a significant impact on the revenue, cost and resource utilization for almost all of the healthcare systems. In this paper, we develop a hybrid probabilistic model based on multinomial logistic regression and Bayesian inference to predict accurately the probability of no-show and cancellation in real-time. First, a multinomial logistic regression model is built based on the entire population's general social and demographic information to provide initial estimates of no-show and cancellation probabilities. Next, the estimated probabilities from the logistic model are transformed into a bivariate Dirichlet distribution, which is used as the prior distribution of a Bayesian updating mechanism to personalize the initial estimates for each patient based on his/her attendance record. In addition, to further improve the estimates, prior to applying the Bayesian updating mechanism, each appointment in the database is weighted based on its recency, weekday of occurrence, and clinic type. The effectiveness of the proposed approach is demonstrated using healthcare data collected at a medical center. We also discuss the advantages of the proposed hybrid model and describe possible real-world applications.

Keywords: Multinomial logistic regression, Dirichlet distribution, Bayesian inference, healthcare operations improvement, no-show and cancellation prediction

1. Introduction

The problem of appointment no-show and cancellation, which is also known as appointment disruption, can cause significant disturbance in the smooth operation of almost all scheduling systems. When scheduled patients do not attend their appointments, resources will be underutilized, while other patients cannot get timely appointments because part of the schedule is filled with patients who will not attend. Also, when scheduled patients cancel their appointments, they often leave the clinic with a very short amount of time to fill the schedule. In such cases, overbooking can help to some extent, but it usually results in clinic congestion and patient dissatisfaction. In fact, appointment no-show and cancellation have far-reaching effects on clinic efficiency, patient outcomes, and healthcare costs, which can reach hundreds of thousands of dollars yearly (Moore *et al.,* 2001; Bech, 2005; Hixon *et al.*, 1999; Rust *et al,*. 1995; Barron, 1980). Hence, accurate prediction of no-show and cancellation probability is a cornerstone for any scheduling system and non-attendance reduction strategy (Daggy *et al.,* 2010; Cayirli and Veral, 2003; Ho and Lau, 1992; Cote, 1999; Hixon *et al.,* 1999; and Moore *et al.,* 2001).

In this article, we develop a hybrid probabilistic model based on multinomial logistic regression and Bayesian inference to predict accurately the probability of no-shows and cancellations in real-time. The result of the proposed method can be used to develop more effective appointment scheduling (Chakraborty *et al.,* 2010; Glowacka *et al.,* 2009; Gupta and Denton, 2008; Hassin and Mendel, 2008; Liu *et al.,* 2009). It can also be used for developing effective strategies, such as selective overbooking for reducing the negative effects of disturbances and filling appointment slots while maintaining short waiting times (Laganga and Lawrence, 2007; Muthuraman and Lawley, 2008; and Zeng *et al.,* 2010).

The rest of the article is organized as follows: Section 2 summarizes some of the related work proposed

*Corresponding author
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uhse.

in the literature. Section 3 describes the general scheme of the proposed model. Section 4 explains our proposed algorithm for predicting no-show and cancellation probabilities. Section 5 presents the results of applying the proposed prediction model to a healthcare dataset collected at a medical center. Finally, Section 6 concludes our work and presents some future extensions of the proposed model.

2. Relevant background and problem formulation

Turkcan *et al.* ([2013)] provide a structured and representative review of no-show literature (see also Rowett *et al.,* 2010; Bowser *et al,.* 2010; Denhaerynck *et al.,* 200; and George *et al.,* 2003). The rate of no-show in the reviewed articles varies significantly for different clinical population of patients, e.g., chronic care, primary care, etc., and locations, e.g., North America, Europe, etc., ranging from close to zero up to 48-64% (see also Mitchell and Selmes, (2007),;Bech, 2005; Cayirli *et al.,* 2006, 2008; and Yehia *et al.,* 2008. Several reasons are reported for no-show including: forgot appointment, conflict with appointment, transportation, physically/mentally unwellness, scheduling system problems, perceived disrespect, bad weather and financial problems (See also Park *et al.,* 2008; Neal *et al.,* 2005; Tuller *et al.,* 2010; Sarnquist *et al.,* 2011; Corfield *et al.,* 2008; Gany *et al.,* 2001). Several factors are studied for predicting non-attendance behavior (Daggy *et al.,* 2010; Zeng *et al.,* 2010; Turkcan *et al.,* 2013; Cashman *et al.,* 2004; Cohen *et al.,* 2008; Savageau *et al.,* 2004; Alafaireet, 2010; Lehmann, 2007). The literature also shows the relationship between no-show and specific health outcomes in some clinical populations including diabetic, dialysis, human immunodeficiency virus (HIV), primary care and psychiatric (Schectman *et al.,* 2008; Obialo *et al.,* 2008; Murphy *et al.,* 2011; Bigby *et al.,* 1984). Ample literature is available discussing interventions to reduce no-show including: appointment reminders, patient education, and follow-up after a no-show appointment, open-access scheduling, and lean process improvement methods (see also Hardy *et al.,* 2001; Guse *et al.,* 2003; Can *et al.,* 2003; Kopach *et al*., 2007; Murray and Tantau, 2000; LaGanga, 2011; Fischman, 2010; Garuda *et al.,* 1998). In fact, the effectiveness of many of the above intervention strategies and consequently clinics' performance significantly rely on the accurate prediction of individual patients' risk of non-attendance (Daggy *et al.,* 2010). Below we have divided the related quantitative methods of predicting appointment disruption into two groups of population-based models and individual-based models.

Population-based techniques mainly use a variety of methods drawn from statistics and machine learning that can be used for predicting no-shows and cancellations (Dove and Schneider, 1981; Kotsiantis, 2007). These methods use the information from the entire population (dataset) in the form of set factors, in order to estimate the (probability of) no-show, cancellation and attendance (Baldi *et al.,* 2000; Kotsiantis, 2007). Logistic regression is one of the most popular statistical methods in this category that is used for binomial regression, which can predict the probability of disturbances by fitting numerical or categorical predictor variables in data to a logit function (Turkcan *et al.,* 2013; Daggy *et al.,* 2010; Hilbe, 2009). There has been some work using tree-based and rule-based models that create if–then constructs to separate the data into increasingly homogeneous subsets, based on which of the desired predictions of disturbances can be made (Glowacka *et al*., 2009). The problem with these population- based methods is that although they provide a reasonable estimate, they do not differentiate between the behaviors of individuals and hence cannot update effectively, especially using small datasets. Another problem with these methods is that once the model has been built, adding new data has an insignificant effect on the result especially when the size of the initial dataset is much larger than the size of the new data. In Section 5, we will compare the performance of above methods with the proposed approach.

Individual based approaches are primarily based on time series and smoothing methods, which are used for predicting the probability of a disruption in an appointment. These methods utilize past behaviors of individuals for the estimation of future no-show and cancellations probability. Potential time series methods for predicting no-shows and cancellations include autoregressive models (Brockwell, 2009), time-frequency analysis models (Chatfield, 1996; and Bloomfield, 1976), nonlinear filtering and Hidden Markov Models (HMM), Stratonovich (1960). Common smoothing algorithms that can be used for no-show and cancellation include moving average, exponential smoothing, and local regression (Simonoff, 1996; Cleveland, 1993; and Winter, 1960). While individual-based methods are fast and effective in modeling the behavioral (no-show) pattern of each individual and work well with a small dataset, they do not provide a reliable initial estimate of no-show and cancellation probabilities, which is especially important in our case. The main reason is that individual based methods usually employ a function of past data (attendance record) to estimate the probability of a future event, e.g. no-show and cancellation. But for the initial state where there is no past history available, such function is inapplicable, and consequently, individual-based methods usually use random guess for the initial estimate, e.g., probability of no-show and cancellation. In Section 5, we will compare the performance of the above- mentioned methods with the proposed method.

As described above, each of the population-based and individual-based approaches have some advantages and disadvantages. However, none of the existing studies for prediction of appointment disruptions have considered using these methods together to overcome their problems and improve their performance, even though related ideas have been successfully employed for fields like universal background model (Reynolds *et al.,* 2000), and recommender systems (Adomavicius and Tuzhilin, 2005). In the next section, we develop a hybrid approach that combines logistic regression as a population-based approach along with Bayesian inference as individual-based approach for prediction of disturbances in appointment scheduling. The proposed approach also contains an efficient linear programming component to optimize its performance.

Indeed, the proposed approach generalizes and extends Alaeddini *et al.* (2010) probabilistic model for no-show estimation to general types of disruptions, e.g. no-show and cancellation. More specifically, this study extends Alaeddini *et al.* (2010) binomial logistic regression model for initial estimation of no-show to a multinomial logistic regression, which can take into account multiple types of disruptions, namely no-show and cancellation. It also generalizes Alaeddini *et al.* (2010) Bayesian updating mechanism for personalization of the no-show estimates for each patient to general types of disruptions using the Dirichlet distribution instead of the Beta distribution. In addition, while Alaeddini *et al.* (2010) used a set of discrete sub-weights to weight the appointments based on their recency, this paper employs a modified versions of the generalized logistic function (Richards, 1959) to design a continuous weighting system. Above generalizations and extensions, enable the proposed approach to develop one general integrated model (instead of multiple standalone models) based on the same set of variables for considering different types of disruptions in appointment scheduling, as they both can have significant but different impacts on the system workload and patients' waiting time; generally the range of possible solutions to a cancellation case is more diverse than a no-show case. The authors recognize that the proposed model implicitly assumes that no-show and cancellation can be predicted by the same set of variables, and while this assumption has been verified for this study, no-show and cancellation may not always have the same predictors.

Some of the major contributions of this research include: (i) Combining the strengths of multinomial logistic regression to provide a reliable initial estimate of no-show and cancellation, and Bayesian updating to personalize the predictions for each patient based on her/his past attendance behavior; (ii) Statistical integration of multinomial logistic regression and Bayesian inference by transforming the output of multinomial logistic regression, which is a multinomial distribution, to its conjugate Dirichlet distribution and applying the Bayesian updating to the Dirichlet distribution, which is a continuous distribution with several good characteristics, such as flexibility and simplicity of updating procedure; (iii) Developing and optimizing a set of appointment weighting factors including: appointment type, recency and weekday of occurrence to further improve the predictions of the proposed model.

Notably, there is a significant difference between the proposed method and Bayesian logistic regression (O'Brien, and Dunson, 2004). Theoretically, in Bayesian logistic regression, the main focus is modeling uncertainties in the parameters of the model, namely the parameter of a multinomial logistic regression. However, in the proposed method, the primary concern is modeling the uncertainty in the outputs of the model, namely no-show, cancellation, and attendance. In addition, unlike the proposed model, in Bayesian logistic regression, modeling a large number of variables can be very challenging; as the number of variables increase, efficient computation of posterior parameters gets very difficult. Finally, while in the proposed method the appropriate type of prior is chosen based on the relationship between logistic regression and the Dirichlet distribution, in Bayesian logistic regression, specifying an appropriate choice of prior is usually trivial.

In Section 5, we compare the performance of the proposed hybrid model with the representative algorithms from each of population-based and individual-based approaches.

3. General scheme of the proposed model

The proposed model for predicting patient no-show and cancellation is composed of three major components shown in Figure 1. First, the data set will be preprocessed, and after coding discrete variables, a multinomial logistic regression model will be built based on the entire populations' general social and demographic information to provide an initial estimate of no-show and cancellation probabilities. Second, each appointment in the database will be weighted based on its recency, weekday of occurrence, and clinic type. Third, the personalized probability of no-show and cancellation will be derived for each individual patient using a comprehensive Bayesian updating mechanism based on the population estimate from Step 1 and weighted records of each individual in Step 2.

In the rest of this section, we will provide a brief description about each of the above components, in the following order: (i) Population-based estimate of appointment disruptions: multinomial logistic regression, (ii) Adaptation: personalization of the predictions, and (iii) Appointment weighting: effect of appointment recency, occurrence on non-working days and clinic type; where we will present the appointment- weighting component after the adaptation component to better demonstrate its impact (in practice, the appointment-weighting component will be applied before the adaptation component). In Section 4, we will show how these components are connected together through the proposed integrated algorithm.

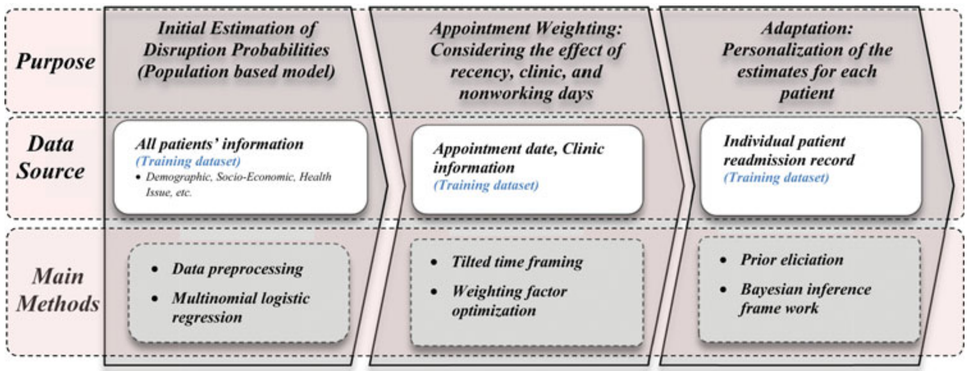


Fig. 1. The general scheme of the proposed method.

3.1. *Population based estimate of appointment disruptions: Multinomial logistic regression*

The first component of the proposed method is a multinomial logistic regression, which is used for initial (prior) estimation of no-show and cancellation probabilities using the (training) dataset of the individuals' general social and demographical information. Let X be the set of factors (explanatory variables) affecting probability of attendance of an individual, and let Y be the attendance type including no-show, cancellation, and attendance. In the preprocessing step of this study, the following factors were identified as significant: (i) Sex, which is considered as a discrete factor with two possibilities of male and female; (ii) Age, which is considered as a continuous factor ranging from 25 to 92; (iii) Marital Status, which is considered as a discrete factor with four categories of never-married, married, divorced and widowed; (iv) Medical Service Coverage, which is considered as a discrete factor with six categories ranging from not service connected, to 50-100% service connected (detail information of each category is available at Department of Veteran Affairs website); (v) Distance to Medical Center (mile), which is considered as a continuous factor and calculated using a free Excel add-on based on the distance between patients' residence ZIP and the medical center; and (vi) Clinic Type (cluster), which is considered as a discrete factor with three categories that are defined by clustering the 271 clinics in the database based on their disruption rates (see the Appendix for details). Likelihood ratio chi-square tests for categorical variable and t-tests for continuous variables have been used to check the significance of the factors.

The multinomial logistic regression model $F(X, B_k)$then takes the form:

$$\begin{cases} P_k = \dfrac{exp\,(XB_k)}{1+\Sigma_{k=0}^{K} exp\,(XB_k)}, k = 1, 2, \ldots, K-1 \\ \quad\text{and } P_k = 1 - \Sigma_{i=0}^{k-1} P_i \\ P_0 = \dfrac{exp\,(XB_k)}{1+\Sigma_{k=0}^{K} exp\,(XB_k)} \end{cases} \tag{1}$$

where P_k is the probability of k^{th} event (In this study, we use $k = 0, 1, 2$, representing no-show, cancellation, and attendance, respectively). The unknown vector of parameters B_k ($B_k = [\beta_{k0}, \beta_{k1}, \ldots, \beta_{kl}]$) can be estimated using iterative procedures, such as Newton-Raphson method or iteratively reweighted least squares (IRLS) (Agresti, 2002). Notably, our preliminary analysis of data also shows negative correlation between the appointments occurred on non-working days and the rates of no-show and cancellation ($P = 0.116$ and 0.218). However, due to its sparsity, addition of such factor makes the data matrix of the regression model severely rank deficient such that the estimation of the model parameter becomes impossible. We address this problem by including this factor to the model in a later stage through another component (appointment weighting), which will be discussed in detail later.

Also, the model proposed in the paper implicitly assumes that no-show and cancellation can be predicted by the same set of variables. In fact, one of the objectives of this research was to develop a unified model (instead of multiple standalone models) to consider different types of disruptions in appointment scheduling, as they can all have significant impact on the system workload and patients' waiting time. However, while our initial statistical analysis has verified above assumption for our dataset, no-show and cancellation may not always have same predictors.

Finally, the proposed multinomial logistic regression approach provides a more compact and convenient model for estimation no-show and cancellation, in comparison to using two separated logistic regression models for no-show/show and cancellation/show.

3.1.1. *Example 1*

For better understanding of how the above model applies to the problem domain, we show a simple case study based on the dataset used in Section 5 experimental results.

The dataset used for this example (the training dataset) is created by dividing the original dataset of this study into two disjoint datasets of approximately equal size, namely training and testing. The original datasets contains the

Table 1. Descriptive statistics of the training dataset used for fitting multinomial logistic regression

	Frequency			Rate		
	No-show	Cancellation	Show-up	No-show	Cancellation	Show-up
Sex						
Female	30	33	89	19.74%	21.71%	58.55%
Male	214	95	617	23.11%	10.26%	66.63%
Marriage Status						
Never Married	63	30	223	19.94%	9.49%	70.57%
Married	80	46	181	26.06%	14.98%	58.96%
Divorced	64	31	121	29.63%	14.35%	56.02%
Widowed	38	19	182	15.90%	7.95%	76.15%
Medical Service Coverage (*Predefined Categories*)						
Service Connected <5	56	16	171	23.05%	6.58%	70.37%
Service Connected, 50% To 100%	70	53	169	23.97%	18.15%	57.88%
Service Connected Less Than 50%	14	4	14	43.75%	12.50%	43.75%
Non-Service Connected	75	51	319	16.85%	11.46%	71.69%
Non-Service Connected, VA Pension	16	5	29	32.00%	10.00%	58.00%
Service Connected <6	6	3	7	37.50%	18.75%	43.75%
Clinic Cluster						
1	10	28	45	12.05%	33.73%	54.22%
2	197	75	618	22.13%	8.43%	69.44%
3	38	23	44	36.19%	21.90%	41.90%
Age (Average)	56.53	52.17	60.12			
Distance To Medical Center (Average)	13.72	15.32	14.87			

information of 1543 attendance records of 99 patients; therefore, 1078 attendance records (from 10/1/2009 to 12/23/2009) are considered for fitting multinomial logistic regression. Table 1 provides some descriptive statistics about the training dataset.

Applying model (1) to the above dataset and using iteratively reweighted least squares (IRLS) method, the parameters of the fitted multinomial logistic regression model along with their standard error are computed as shown in Table 2 (because we are modeling a categorical variable with three mutually exclusive levels, namely no-show, cancellation and attendance, and because the occurrence of any one of them automatically implies the non-occurrence of the remaining two events ($P_2 = 1 - (P_0 + P_1)$), only two sets of regression parameters are estimated (see Model 3)). Meanwhile, the assumption of linear association between the continuous covariates of the model and log odds of no-show and cancellations has been verified to be appropriate.

Now for a sample patient in the dataset with the information shown in Table 3, based on the estimated coefficients (in Table 2), the probability of no-show, cancellation and attendance is estimated as (0.14, 0.09, 0.77).

3.2. *Adaptation: Personalization of the predictions*

The result of above multinomial logistic regression is a multinomial distribution with three variables, namely probabilities of no-show, cancellation and attendance, where the parameters ($P_k, k = 0, 1, 2$) of such distribution is estimated based on all patients' information in the (training) dataset using multinomial logistic regression. Such (initial) estimates can be personalized for each patient based in her/his individual attendance records (in the training dataset) to improve the prediction ability of the model. Bayesian inference is another component of the proposed method, which is used for updating the prior estimate of no-show and cancellation probabilities from multinomial logistic regression using the dataset of individual patient attendance record (training dataset).

To use Bayes' theorem, we need a prior distribution $g\left(\alpha^{pri}\right)$ that gives our belief about the possible values of the parameter vector $\alpha = (\alpha_1, \ldots, \alpha_K)$ representing the probabilities of no-show, cancellation and attendance before incorporating the data (Z). The posterior distribution is proportional to prior distribution times likelihood:

$$g\left(\alpha^{pos}|\mathrm{Z}\right) = \frac{g\left(\alpha^{pri}\right) \times f\left(\mathrm{Z}|\alpha^{pri}\right)}{\int_0^1 g\left(a\right) \times f\left(\mathrm{Z}|a\right) da} \quad (2)$$

In Bayesian statistics, the Dirichlet distribution is a common choice for updating the prior estimate of multinomial distribution parameters (distribution of the dependentvariable in the multinomial logistic regression) because (Bolstad, 2007):

1. The Dirichlet distribution is the conjugate prior of multinomial distribution, giving the same posterior distribution as the prior (Dirichlet).

Table 2. Fitted multinomial logistic regression model, including regression coefficients, standard error and p-values

		Sex	Age	Marriage status Type 1	Type 2	Type 3	Medical service coverage Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Distance to medical center (mile)	Clinic cluster type 1	type 2	Constant
Regression Coefficients	No-show	−1.434	0.332	−0.01	0.432	0.493	−0.363	0.201	0.743	−0.106	0.441	1.667	−0.004	1.307	0.421
	Cancellation	−0.747	0.002	−0.014	0.669	0.471	0.226	0.935	0.866	0.465	0.44	1.526	0.002	−0.094	−1.474
Standard error	No-show	0.619	0.248	0.007	0.217	0.241	0.264	0.229	0.424	0.211	0.364	0.591	0.004	0.428	0.371
	Cancellation	0.758	0.281	0.009	0.289	0.34	0.364	0.334	0.647	0.324	0.57	0.883	0.005	0.381	0.3
p-values	No-show	0.021	0.182	0.116	0.046	0.041	0.169	0.244	0.079	0.217	0.226	0.005	0.232	0.002	0.246
	Cancellation	0.224	0.250	0.144	0.021	0.166	0.235	0.005	0.180	0.152	0.240	0.084	0.248	0.205	0.000

Table 3. Information of a sample patient and a sample appointment in the dataset

Sex	Age	Marriage Status	Medical service coverage	Distance to medical center (mile)	Clinic cluster	Recency	Closeness to non-work days	Probabilities of no-show, cancellation and show-up
Male	78	Widowed	50% -100%	15.90	2	322 days	Not before holiday	(0.14, 0.09, 0.77)

2. The Dirichlet distribution has very efficient updating mechanism, where it only requires adding the number of occurrence of each category to the prior parameters α_k.
3. In the context of our problem, the parameters of the prior Dirichlet distribution is readily available and is equal to the output of the multinomial logistic regression, which is not the case for many other types of prior distributions, such as normal distribution.
4. Unlike multinomial distribution, the Dirichlet distribution is a continuous distribution, which is much easier to work with in terms of inference and updating.
5. The Dirichlet distribution has a few parameters more than multinomial distribution, which allows it to take different shapes and makes it suitable for representing different types of priors.

By definition, the Dirichlet distribution (denoted by $Dir\,(\alpha)$) is a family of continuous multivariate probabilitydistributions parameterized by the vector α of positive reals. The Dirichlet distribution for random variables $Z_1, \ldots, Z_K$with parameters $\alpha_1, ..., \alpha_K > 0$, $K \geq 2$ (our work incorporates $K = 3$, which is based on the number categories: no-show, cancellation, and attendance) and has a probabilitydensityfunction with respect to Lebesgue-measure on the Euclideanspace R^{K-1} given by (Evans *et al.* [2000]):

$$f\,(Z_1, \ldots, Z_K; \alpha_1, \ldots, \alpha_K) = \frac{1}{B\,(\alpha)} \prod_{k=1}^{K} Z_k^{\alpha_k - 1} \tag{3}$$

for all $Z_1, \ldots, Z_K > 0$ where Z_K is an abbreviation for $1 - Z_1 - \ldots - Z_{K-1}$. The density is zero outside this open $(K-1)$-dimensional support of the Dirichlet distribution is the set of K-dimensional vectors Z whose entries are real numbers in the interval (0,1); furthermore, the sum of the coordinates is 1. Another way to express this is that the domain of the Dirichlet distribution is itself a set of probabilitydistributions, specifically the set of K-dimensional discretedistributions. The normalizingconstant is the multinomial betafunction, which can be expressed in terms of the gammafunction:

$$B\,(\alpha) = \frac{\prod_{k=1}^{K} \Gamma\,(\alpha_k)}{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}, \alpha = (\alpha_1, \ldots, \alpha_K) \tag{4}$$

From the Bayesian perspective, the probability density function of the Dirichlet distribution returns the belief that the probabilities of K rival events are Z_j given that j^{th} event has been observed $\alpha_j - 1$ times.

Based on the above discussion, $Dir\,(\alpha)$can be used as prior density of the proposed Baysian update mechanism to update parameters $\alpha = (\alpha_1, \ldots, \alpha_K)$. The result of Bayesian update is a new (posterior) Dirichlet with parameters vector:

$$\alpha_k^{pos} = \alpha_k^{pri} + y_k \tag{5}$$

where $y_k, k = 1, 2, 3$ is the number of occurrence of each category, namely no-show, cancellation and attendance, in the (training) dataset. In other words, the Dirichlet distribution can be updated simply by adding the new occurrence number of each category to the prior parameter α_k (Bolstad, 2007):

$$g\,(\alpha|Z) = \frac{\Gamma\left(\sum_{k=1}^{k} y_k + \alpha_k\right)}{\prod_{k=1}^{K} \Gamma\,(y_k + \alpha_k)} \prod_{k=1}^{K} Z_k^{\alpha_k + y_k - 1} \tag{6}$$

Table 4. Attendance record and Bayesian updated probabilities of no-show cancellation and show-up of the sample patient in Example 1 (No-show, cancellation, and show up are represented by 1, 2, and 3 respectively in the attendance record column)

			Posterior mean of Dirichlet dist.		
Appointment No.	Appointment date	Attendance record	No-Show	Cancellation	Show-up
			0.14	0.09	0.77
1	10/5/2009	1	0.57	0.05	0.38
2	10/29/2009	1	0.71	0.03	0.26
3	11/5/2009	2	0.53	0.27	0.19
4	12/4/2009	1	0.63	0.22	0.15
5	12/7/2009	1	0.69	0.18	0.13

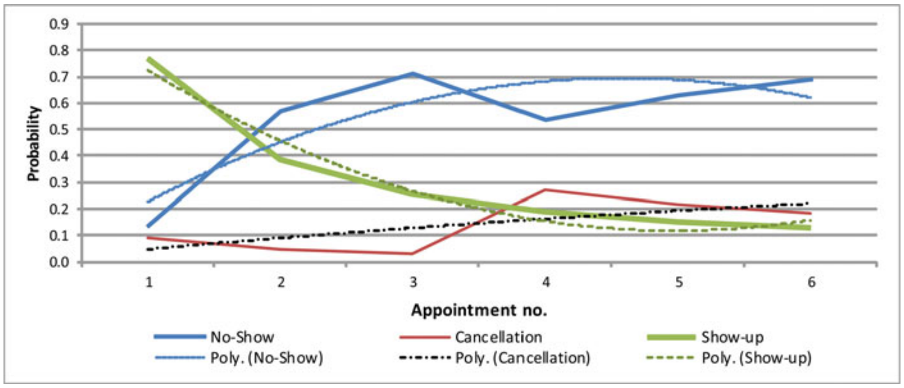


Fig. 2. Changing parameters of Dirichlet distribution for the sample patient during Bayesian update.

The posterior mean, which represents the updated estimate of the multinomial distribution parameters, would then be $E(a_k|y_1,\ldots,y_K) = \frac{y_k+a_k}{\sum_{k=1}^K y_k+\sum_{k=1}^K a_k}$ with variance:

$$Var\,(a_k|y_1,\ldots,y_K) = \frac{(y_k+a_k)\left(\left(\sum_{k=1}^K y_k+\sum_{k=1}^K a_k\right)-(y_k+a_k)\right)}{\left(\left(\sum_{k=1}^K y_k+\sum_{k=1}^K a_k\right)^2\left(\sum_{k=1}^K y_k+\sum_{k=1}^K a_k+1\right)\right)} \tag{7}$$

Bolstad (2007) suggests choosing a prior distribution that matches the belief about the location and scale. This procedure, which is used in this research, can be formulated by letting $\alpha_k^{pri} = P_k$; where P_k is the output of the multinomial logistic regression. As an alternative to the above procedure, several researchers (Leonard, 1973; Aitchison, 1985; Goutis, 1993; and Forster and Skene, 1994) proposed using a multivariate normal prior distribution for multinomial logits.

To conclude this section, when a new patient enters the system (there is no personal history), the initial estimate using multinomial logistic regression will be used as the prediction of no-show, cancellation and attendance. Also, if there is a patient with only one or two appointments, the initial estimate from multinomial logistic regression is updated based on the available appointment/s information.

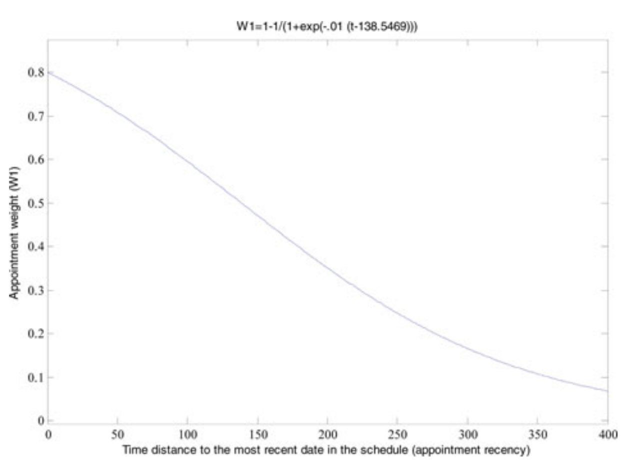


Fig. 3. Tilted time framing using (a modified version of) the generalized logistic function.

3.2.1. *Example 2*

For better understanding of how the Bayesian update can be applied to multinomial regression results, we reconsider the attendance records of the sample patient in Example 1. Table 4 presents the attendance record of the patient during 10/1/2009 to 12/23/2009. Note that no-shows are represented by 1 while cancellation and attendance are represented by 2 and 3, respectively. Using the result of multinomial logistic regression model in Example 1 as the prior parameters of the Dirichlet distribution $\left(\alpha^{pri} = (0.14, 0.09, 0.77)\right)$, the posterior distribution of no-show, cancellation and attendance after each appointment are illustrated in the last three columns of Table 4. In the proposed Bayesian update mechanism, if a patient has multiple appointments on the same day with similar outcomes, e.g., no-show, (there have been no such cases in the database of this study), one of the reviewers suggested to count that outcome (no-show, cancellation, attendance) only once.

Figure 2 also illustrates the changes in the estimated probabilities of no-show, cancellation and attendance (solid lines) after observing each new record of attendance (the estimated trend of each type (dashed lines) is also calculated and shown using polynomials of order three). From Figure 2, that Bayesian update reacts quickly to each new data record can be checked; however, as the number of attendance records increases, the estimates tend to converge to certain probabilities. In other words, in Bayesian updating when the number of updates (records) gets very large, the effect of the new records gets marginal. As a result, the model may not be able to respond effectively to changes in the patients' attendance behavior (it requires many new records). In the next section, we describe another component of the proposed method, which works together with Bayesian

Table 5. Data structure and optimal value of the weighting factors

Appointment Recency ($\tilde{w}_1$) (parameters of the generalized logistic function)		Preceding non-workday ($\tilde{w}_2$)		Clinic Cluster ($\tilde{w}_3$)		
B	M	Not-before holiday	Before holiday	1	2	3
0.01	138.5469	0.6863	0.3648	0.5728	0.3831	0.6921

update mechanism to adjust the weight of appointments based on a number of appointment-related factors including: appointment recency, occurrence on non-working days and clinic type to improve the performance of the proposed method.

3.3. *Appointment weighting: Effect of appointment recency, occurrence on non-working days and clinic type*

The proposed weighting mechanism, which is interlaced with Bayesian update mechanism, will increase the information content of data to improve the prediction ability of no-show and cancellation. For this purpose, a set of weighting factors $W = [\tilde{w}_1, \ldots, \tilde{w}_\omega]$ is designed and weights the appointments in (7) before being applied to Bayesian update. Our weighting scheme includes three weighting factors($\omega = 3$): (i) Recency of the appointment: weights each no-show, cancellation and attendance record based on how recently it occurred. The more recent the appointment the higher the weight; see Figure 3; (ii) occurrence on non-working days: weight the appointments based on whether they occurred on non-working days to adjust the lower rate of no-show and cancellation on those days; (iii) Clinic type: weight each appointment in the dataset with respect to the hosting clinic to adjust different rates of no-show and cancellation in different clinics. Depending on the medical center where the no-show/cancellation prediction model is being applied, one may think of other types of weighting factors as well.

For the first weighting factor ($\tilde{w}_1$), appointment recency, reasonably no-show and cancellation records that occurred a long time ago do not carry the same weight as recent ones. This factor is based on the fact that patients may gradually or abruptly change their behavior and should be reflected in the model. For this purpose, a tilted time framing mechanism is developed and is closely related to exponentially weighted moving average (EWMA) smoothing, Figure 3. A weighting factor of $\tilde{w}_1$ is defined based on a modified version of the generalized logistic function (Richards, 1959):

$$\tilde{w}_1 = 1 - \frac{1}{1 + exp\left(-B\left(t - M\right)\right)} \tag{8}$$

where t is the date of appointment, M is the current date, and B is the growth rate of the logistic function.

For the second weighting factor ($\tilde{w}_2$), occurrence on non-working days, our preliminary study of the data revealed strong negative correlations between no-show and cancellation rates and appointment occurred on non-working days. To adjust the influence of different week days on the probability of disruption, the following two sub-weights

$$\tilde{w}_2 = \begin{cases} w_{21} \ \textit{Monday through Friday} \\ w_{22} \ \textit{Weekend and holidays} \end{cases} \quad \text{are designed.}$$

For the third weighting factor ($\tilde{w}_3$), a strong correlation has also been observed among the rate of no-show and cancellation, and some clinics. Three sub-weights ($\tilde{w}_3 = \{w_{31}, w_{32}, w_{33}\}$) are defined to consider the effect of the hosting clinic on the chance of no-show and cancellation.

The parameters B and M of the first weighting factor ($\tilde{w}_1$), along with the optimal values of the sub-weights of the other two weighting factors, namely $\tilde{w}_2 = \{w_{21}, w_{22}\}$ and $\tilde{w}_3 = \{w_{31}, w_{32}, w_{33}\}$, can be determined by minimizing the mean square difference (MSE) of the estimated probabilities of no-show, cancellation and attendance, and the respective empirical probabilities (in the training dataset). The general formulation for the objective function, which should be minimized with respect to the weights, can be represented as follows:

$$\begin{aligned} &Min\ MSC = \Sigma_{i=1}^{n} \Sigma_{k=1}^{K} \left(\hat{p}_{i(j\in T)k}^{Model} - \hat{p}_{i(j\in T)k}^{Emp}\right)^2 / n \\ &S.T. \\ &\quad 0 \leq W_\omega \leq 1, \omega = 1, 2, 3 \end{aligned} \tag{9}$$

where $i = 1, \ldots, n$, is the index of the patient, $j \in T$ is the index of appointments in training dataset ($T = \{1, \ldots, t\}$), $k = 1, 2, (K = 3)$ is the index of attendance outcome; no-show, cancellation and show- up, and ω is the index of weighting factors ($\omega = 1, \ldots 3$).

In objective function (9), we compare the estimate from the training dataset $\left(\hat{p}_{i(j\in T)k}^{Model}\right)$, which is calculated by applying the Bayesian update mechanism to the weighted appointments in that dataset, with the empirical probability of disruptions $\left(\hat{p}_{i(j\in \mathrm{T})k}^{Emp}\right)$. Since $\hat{p}_{i(j\in \mathrm{T})k}^{Model}$ and $\hat{p}_{i(j\in \mathrm{T})k}^{Emp}$ are providing the estimates of no-show, cancellation, and attendance probabilities of same patients, they are comparable. $\hat{p}_{i(j\in \mathrm{T})k}^{Emp}$, the empirical probability of disruption of type k, for person i after his last appointment in the (training) dataset is calculated as follows:

$$\hat{p}_{i(j\in \mathrm{T})k}^{Emp} = \frac{\sum_{j=1}^{t} Y_{ijk}}{t} \tag{10}$$

Table 6. Attendance record and Bayesian updated probabilities using weighted appointments of the sample patient in Example 1 (No-show, cancellation, and show up are represented by 1, 2, and 3 respectively in the attendance record column)

			Posterior mean of Dirichlet dist.			Weights			Posterior mean of Dirichlet dist.		
Appointment No.	Appointment date	Attendance record	No-Show	Cancellation	Show-up	Recency	Preceding non-work day	Clinic cluster	No-Show	Cancellation	Show-up
			0.14	0.09	0.77				0.14	0.09	0.77
1	10/5/2009	1	0.57	0.05	0.38	0.40	0.36	0.69	0.22	0.08	0.70
2	10/29/2009	1	0.71	0.03	0.26	0.46	0.36	0.38	0.26	0.08	0.66
3	11/5/2009	2	0.53	0.27	0.19	0.48	0.36	0.69	0.24	0.17	0.60
4	12/4/2009	1	0.63	0.22	0.15	0.55	0.69	0.69	0.37	0.14	0.50
5	12/7/2009	1	0.69	0.18	0.13	0.56	0.36	0.38	0.40	0.13	0.47

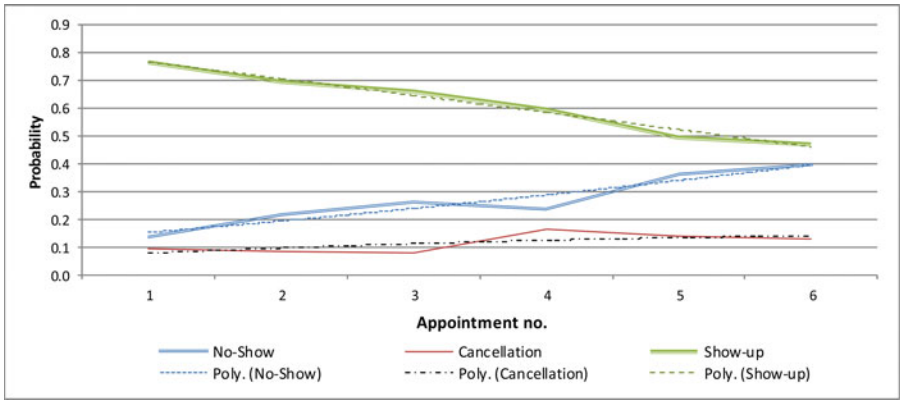


Fig. 4. Changing parameters of Dirichlet distribution based on weighted appointments for the sample patient during Bayesian update.

with Y_{ijk} as a multinomial (random) variable representing attendance record of type k for patient i in the appointment j $\left(Y_{ijk} = (0, 1)\right)$. Also, $\hat{p}^{Model}_{i(j\in T)k}$, the estimated probability of disruption of type $k$for person i is calculated based on the weighted appointments in the training dataset using the proposed model as follows:

$$\hat{p}^{Model}_{ij\in \mathrm{T})k} = \frac{\alpha^{pos}_{i(j\in \mathrm{T})k}}{\sum_{k=1}^{K} \alpha^{pos}_{i(j\in \mathrm{T})k}} \tag{11}$$

where $\alpha^{pos}_{i(j\in \mathrm{T})k}$ is the vector of posterior parameters of the Dirichlet distribution (no-show, cancellation and attendance estimates) calculated based on (5) but using weighted appointments as:

$$\alpha^{pos}_{i(j\in \mathrm{T})k} = \alpha^{pri}_{ik} + \sum_{j=1}^{t} \left(\prod_{\varphi\in\omega} \tilde{w}_{ij\varphi} \right) Y_{ijk} \tag{12}$$

where $\tilde{w}_{ij\varphi}$ is the φ^{th} weight ($\omega = 1, 2, 3$)related to the j^{th} appointment of patient i, and α^{pri}_{ik} is the prior estimate of disruption of type k for patient i resulted from the regression analysis. Substituting $\hat{p}^{Emp}_{i(j\in \mathrm{T})k}$ and $\hat{p}^{Model}_{i(j\in \mathrm{T})k}$ with their equivalent formulations in (10–12), the optimization model in (9) can be rewritten as:

$$\begin{aligned} &Min\ MSE \\ &= \Sigma_{i=1}^{n} \Sigma_{k=1}^{K} \left(\frac{\alpha^{pri}_{ik} + \Sigma_{j=1}^{t} \left(\Pi_{\varphi\in\omega} \tilde{w}_{ij\varphi}\right) Y_{ijk}}{\Sigma_{k=1}^{K} \left(\alpha^{pri}_{ik} + \Sigma_{j=1}^{t} \left(\Pi_{\varphi\in\omega} \tilde{w}_{ij\varphi}\right) Y_{ijk}\right)} - \frac{\Sigma_{j=1}^{t} Y_{ijk}}{t} \right)^2 /n \\ &S.T: \qquad\qquad (13) \\ &0 \le \tilde{w}_{ij\varphi} \le 1 \end{aligned}$$

The above model can be solved by most of the nonlinear programming methods. To ensure having large enough dataset for personalization of the estimated for each patient, we suggest using two-third of data for training and one–third of that for testing. We also suggest using the data in the order they occurred to consider the possible changes in the patient behavior. Table 5 shows the optimal weights from applying optimization model 13 to the training dataset.

3.3.1. *Example 3*

For better understanding of the application of the optimization model, we reconsider Example 2 based on weighted appointments. Incorporating the appointment weights to the Bayesian update mechanism in Example 2 results in the posterior Dirichlet probabilities shown in Table 6. Figure 4 also illustrates the changes in the estimated probabilities of no-show, cancellation and attendance after each new weighted appointment record of attendance (solid lines) as well as the estimated trend (using order three polynomials). Comparing Figure 4 to Figure 2 reveals the changes in the posterior Dirichlet distribution parameters estimates. As shown in Figure 4, weighting the appointments generally decreases the fluctuations of the posterior estimates for the early appointments and increases the responsiveness of the updating mechanism to recent records, which means the quicker response to changes in patient behavior.

In addition, comparison of Tables 4 and 6 reveals that the weighting mechanism moderates the amount of changes (increase/decrease) in the probability of no-show, cancellation, and attendance after each update. For instance in Table 4, the amount of change in the probability of attendance from the first to the second appointment is 0.39. This amount is only 0.02 for appointment 4 to 5. In the meantime, in Table 6 the amount of change from appointment 1 to 2 and appointment 4 to 5 is 0.07.

4. The proposed algorithm

The main idea of the proposed method is to first provide an initial estimate of no-show, cancellation and attendance based on the demographic information of all patients in the (training) dataset, and then improve the initial estimates based on the individuals' no-show, cancellation and attendance behavior using a Bayesian updating mechanism on weighted appointments. *Algorithm 1* describes the proposed algorithm, which comprises the three components, namely, multinomial logistic regression, Bayesian update mechanism, and the optimization procedure that is explained in the previous section.

Algorithm 1: No-show and Cancellation Prediction Algorithm

Input: Training dataset $\left(X_{ij},\, Y_{ij}\right)$, Threshold parameter T
Output: Estimated no-show and cancellation probabilities $\hat{p}^{Model}$, the Dirichlet distribution posterior parameters $\left(\alpha_{ij}^{pos}\right)$, Multinomial logistic regression estimated parameters $\hat{B}_k$
Procedure:

1: /* ***Logistic regression*** */
2: $\hat{B}_k \leftarrow$ *Estimate the parameters of multinomial logistic regression in (3)*
3: $\hat{p}_{ik}^{0}\left(Y_i = 1, 2, 3 | X_{ij}\right) \leftarrow F\left(X_{ij},\, \hat{B}_k\right)$
4: $\alpha_{ik}^{pri} \leftarrow \hat{p}_{ik}^{0}$
5: /* ***Weight optimization*** */
6: $\hat{p}_{i(j\in \mathrm{T})k}^{Emp} = \frac{\sum_{j=1}^{t} Y_{ijk}}{t}$
7: $\tilde{w}_{ij\varphi} \leftarrow Min\, MSE =$
$$\sum_{i=1}^{n}\sum_{k=1}^{K}\left(\frac{\alpha_{ik}^{pri}+\sum_{j=1}^{t}\left(\prod_{\varphi\in\omega}\tilde{w}_{ij\varphi}\right)Y_{ijk}}{\sum_{k=1}^{K}\left(\alpha_{ik}^{pri}+\sum_{j=1}^{t}\left(\prod_{\varphi\in\omega}\tilde{w}_{ij\varphi}\right)Y_{ijk}\right)} - \frac{\sum_{j=v1+1}^{t} Y_{ijk}}{t}\right)^2 /n \; Subjectto 0 \le \tilde{w}_{ij\varphi} \le 1$$
8: /* ***Bayesian update*** */
9: $\alpha_{i(j\in \mathrm{T})k}^{pos} = \alpha_{ik}^{pri} + \sum_{j=1}^{t}\left(\prod_{\varphi\in\omega}\tilde{w}_{ij\varphi}\right)Y_{ijk}$
10: $\hat{p}_{i(j\in \mathrm{T})k}^{Model} = \frac{\alpha_{i(j\in \mathrm{T})k}^{pos}}{\sum_{k=1}^{K}\alpha_{i(j\in \mathrm{T})k}^{pos}}$
11: *Return* $\hat{p}^{Model}$

In the first component *(lines 1 to 3)*, based on the training dataset consisting of individuals' personal information(D_{GI}), (such as gender, marital status, etc.) and their sequence of appointment information (e.g. previous attendance records (D_{NR})), a multinomial logistic regression model $F\left(X_{ij},\, \hat{B}\right)$ is formulated *(line 2)*. Then, using logistic regression, an initial estimate of no-show, cancellation and attendance probabilities are calculated, given by $\hat{p}_{ik}^{0}\left(Y_i = 1, 2, 3 | X_{ij}\right)$ *(line 3)*. This estimate $\left(\hat{p}_{ik}^{0}\right)$is used as the prior of Bayesian update procedure $\left(\alpha_{ik}^{pri}\right)$ in the second and third components (*line 4*). As discussed in Section 2, logistic regression bundles the information of the complete population together and finds a reliable initial estimate of no-show ($\hat{p}_{ik}^{0}$).

In the second component (*lines 5 to 7*), which is interlaced with the third component, the empirical rate of no-show, cancellation and attendance is calculated (*Line 6*). Then, an optimization model is used to find the optimal value of a set of weighting factors (related to appointment recency, occurrence on non-working days and clinic type), which minimizes the sum of squared differences in the predictions form model (using weighted appointments) and empirical estimates of no-show, cancellation and attendance in (*Line 7*). As discussed in Section 3.3, the main purpose of weighting the appointments (in Bayesian update) is to increase the information content of data to improve the prediction ability of no-show and cancellation.

In the third component, using the weighted attendance record of each person $(\prod_{\varphi\in\omega}\tilde{w}_{ij\varphi})Y_{ijk}$, the posterior parameters α_{ijk}^{Pos} and posterior probability of attendance $\hat{p}_{ijk}^{Model}$ is calculated (*lines 9 and 10*). As discussed in Section 2, the reason that Bayesian update procedure is applied to the output of logistic regression is that, typically, regression models cannot consider individual patient's behavior. Also, updating the regression parameters based on new data records is both difficult and only marginally effective (especially when the model is already constructed on a huge dataset) in comparison to the Bayesian update.

In practice, when the database of patient's information is large, once Algorithm 1 is built and executed, only line 9 and 10 of the algorithm is required to be executed again upon receiving new data for a subset of patients. The reason is because the logistic regression and weight optimization parts are already built based on large amount of data and the estimates are confident. Nonetheless, at the individual patient level, the number of records may not be enough, or the patient may change her/his behavior. Therefore, the Bayesian update (*Line 0 to 10*) based on weighted appointments is used to further personalize the estimates. If the size of the database is not large, the clinic may need to run the whole algorithm whenever the size of newly added data is comparable to the size of original dataset, e.g., every month.

5. Experimental results

In this section, we compare the performance of the proposed model with different population-based and individual-based algorithms based on the dataset collected at the Veteran Affairs (VA) Medical Center in Detroit using time-wise analysis. For this purpose, the training and testing data are constructed as follows: appointments that occurred before 2/1/2010 have been used for training and appointments after 2/1/2010 have been considered for testing. The main reason for selecting the above dates is to have approximately two-third of data records for training and the rest of the data for testing.

5.1. *Time-wise analysis*

In this section, we compare the performance of the proposed model with a number of population-based, individual-based and adaptation algorithms using time-wise analysis. The methods used in our comparison along with their information are presented in Table 7.

Figure 5 illustrates the mean squared error (MSE) of the comparing methods. Based on the MSE measure, the proposed model outperforms other methods, while Box

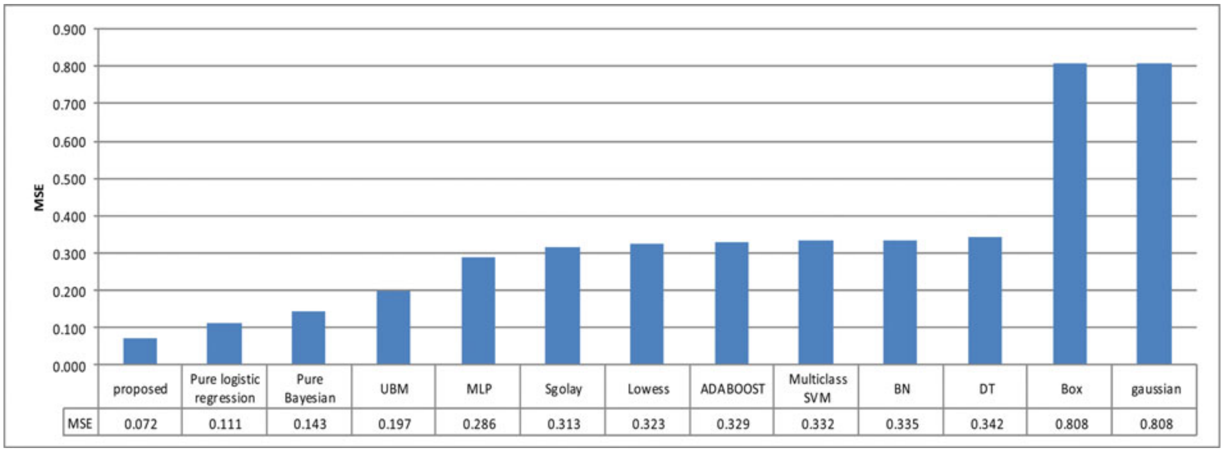


Fig. 5. Mean Squared Error (MSE) of the comparing methods for time-wise analysis.

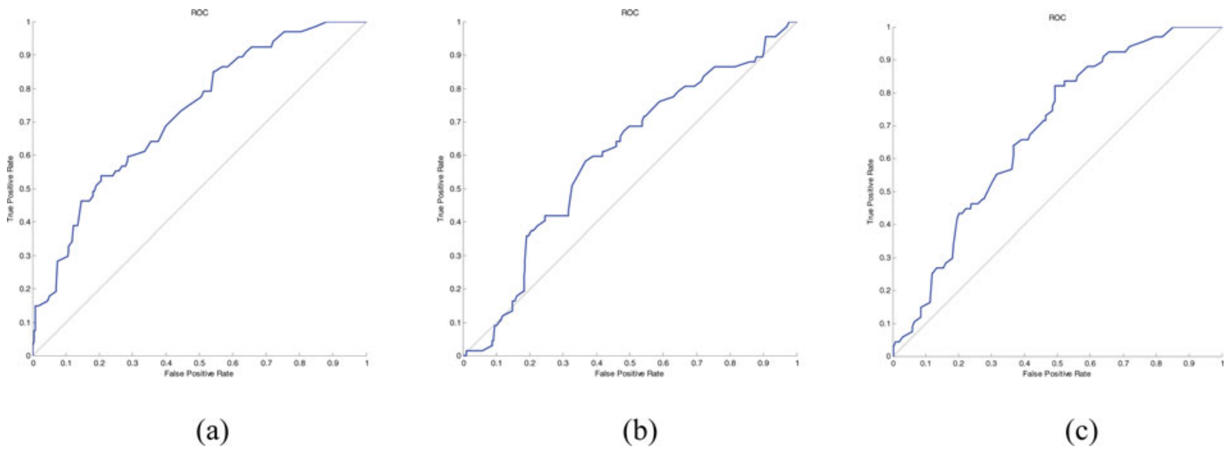


Fig. 6. ROC curves of the comparing methods for no-show prediction: (a) proposed method, (b) pure logistic regression, and (c) pure Bayesian update.

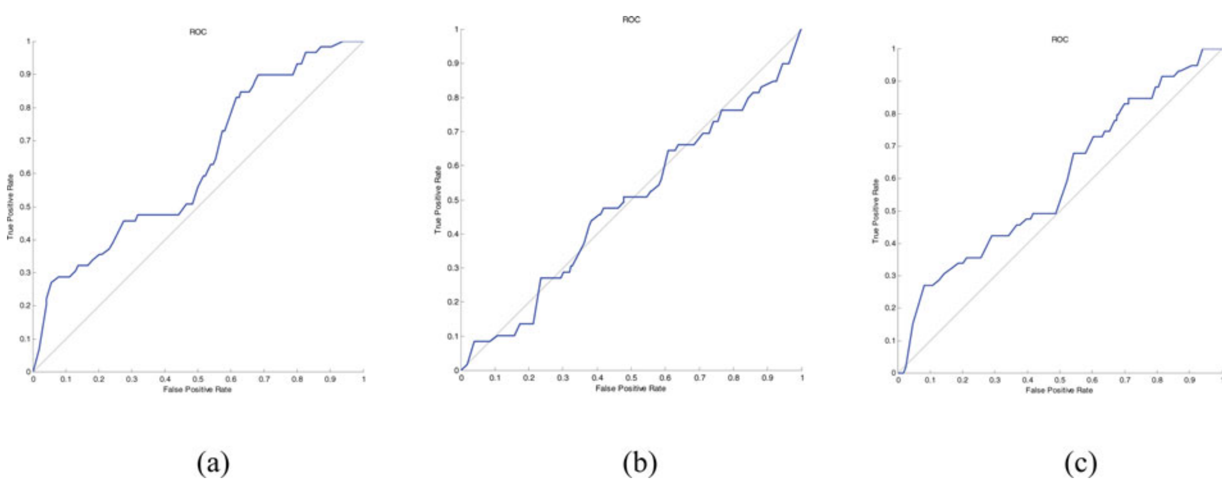
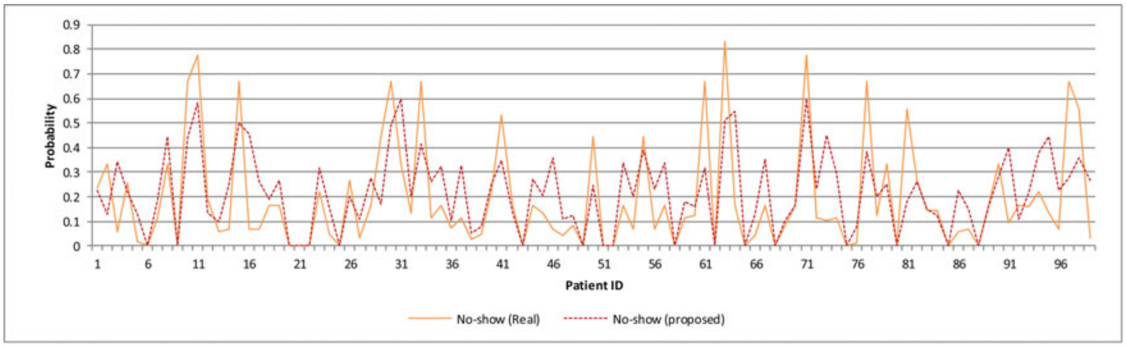
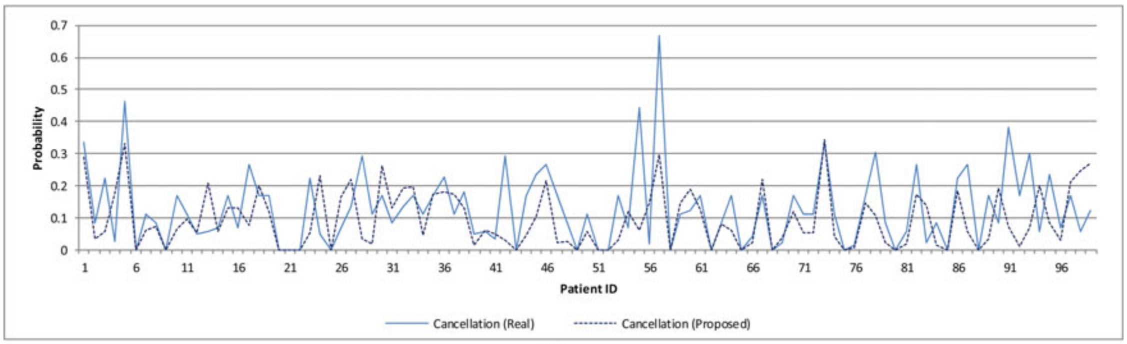


Fig. 7. ROC curves of the comparing methods for cancellation prediction: (a) proposed method, (b) pure logistic regression, and (c) pure Bayesian update.

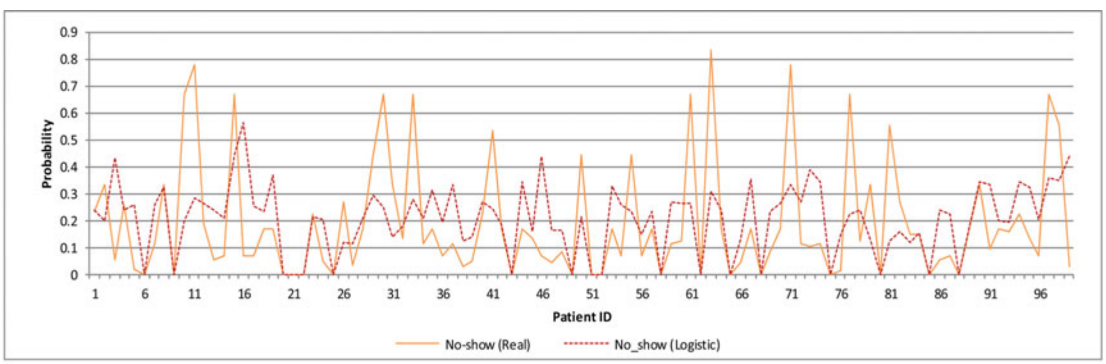


(a)

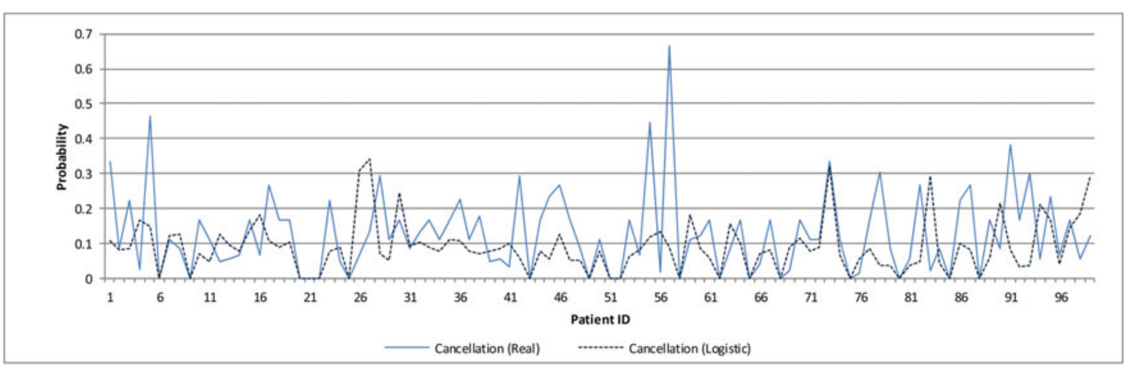


(b)

Fig. 8. Estimated versus empirical probability of appointment disruptions from the proposed approach over different patients: (a) no-show estimation, and (b) cancellation estimation.



(a)



(b)

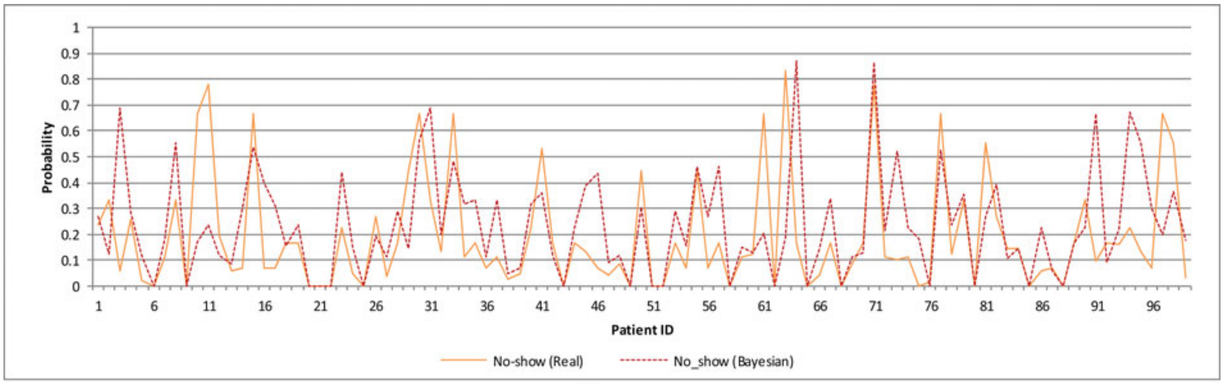
Fig. 9. Estimated versus empirical probability of appointment disruptions from the pure multinomial logistic regression over different patients: (a) no-show estimation, and (b) cancellation estimation.

Table 7. Comparing methods information

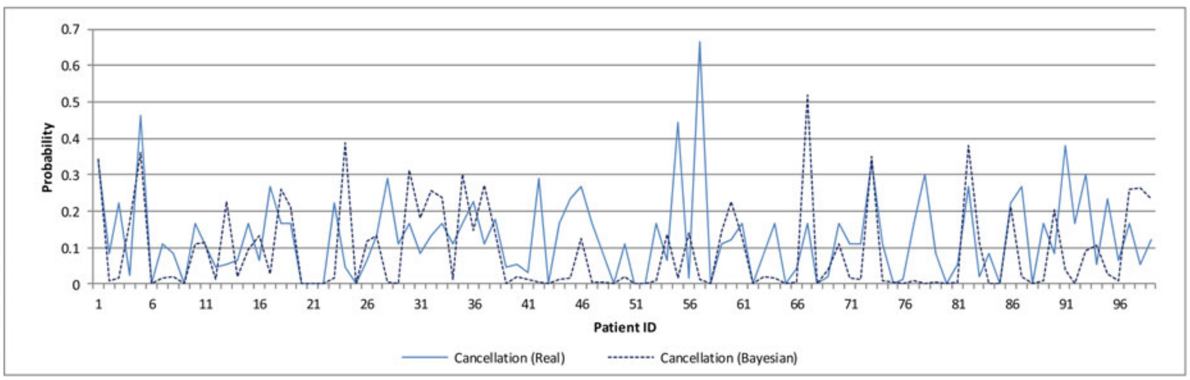
Row	Method	Parameters (estimated)	Reference	Note
1	Proposed method	Sig. level of vars.: ≤ 0.25 Prior par.: logistic regression output		• Data splitting strategy: (2/3 training, 1/3 test) • Model selection: backward and forward
2	locally weighted scatter plot (LOESS)	Moving window size: 5	Simonoff [1996]	• Range of parameter evaluated: (1,7) • Data splitting strategy: (2/3 training, 1/3 test)
3	Box smoothing	Moving window size: 5	Simonoff [1996]	• Range of parameter evaluated: (1,7) • Data splitting strategy: (2/3 training, 1/3 test)
4	Savitzky-Golay smoothing	Polynomial order: 3	Simonoff [1996]	• Data splitting strategy: (2/3 training, 1/3 test)
5	Gaussian smoothing	Std. parameter: 0.65	Simonoff [1996]	• Range of parameter evaluated: (0.2,1) • Data splitting strategy: (2/3 training, 1/3 test)
6	Decision tree (DT)	Confidence factor: 0.25	Quinlan [1986]	• Algorithm: J48 • Data splitting strategy: (1/3 training, 1/3 validation, 1/3 test)
7	Multinomial logistic regression	Sig. level of vars.: ≤ 0.25	Allison [1999]	• Same predictors as used in the proposed model regression part • Data splitting strategy: (2/3 training, 1/3 test) • Model selection: backward and forward
8	Multinomial Bayesian update	Prior par.: (0.33, 0.33, 0.33)	Bolstad [2007]	• Prior choice: Jeffery's prior • Data splitting strategy: (2/3 training, 1/3 test)
9	Bayesian Net	Estimator: Simple ($\alpha = .5$)	Jensen [1996]	• Data splitting strategy: (1/3 training, 1/3 validation, 1/3 test) • Search algorithm: hill climbing
10	Multilayer Perceptron Neural Net (MLP)	Hidden layers: a Learning rate: .3 Momentum: 0.2	Koskela *et al.* [1996]	• Data splitting strategy: (1/3 training, 1/3 validation, 1/3 test)
11	Multi Class Support Vector Machine (SVM)	Kernel fun.: MLP Cache limit: 1000 Method: least square	Weston and Watkins [1998]	• Data splitting strategy: (1/3 training, 1/3 validation, 1/3 test)
12	Boosting	Classifier: Decision Stump Weight threshold: 100	Viola and Jones [2002]	• Algorithm: ADABOOST M1 PART • Data splitting strategy: (1/3 training, 1/3 validation, 1/3 test)
13	Universal Background Model (UBM)	Model selection criterion: AIC	Reynolds *et al.* [2000]	• Data splitting strategy: (1/3 training, 1/3 validation, 1/3 test)

and Gaussian smoothing has the worst performance. As a complement, Figures 6 and 7 show the receiver operating characteristic (ROC) curve of the proposed method along with pure multinomial logistic regression and Bayesian update for no-show and cancellation prediction. Such result can demonstrate that bundling population-based and individual-based methods together (as in our proposed method) can improves the overall performance.

Figures 8 to 10 compare the empirical and estimated probability of no-show and cancellation for the proposed method along with pure multinomial logistic regression and pure multinomial Bayesian update over various patients (the results from other methods along with the source code is available upon request). As shown in Figure 8, the proposed approach often predicts the real pattern correctly with small variance.



(a)



(b)

Fig. 10. Estimated versus empirical probability of appointment disruptions from the pure Bayesian updating mechanism over different patients: (a) no-show estimation, and (b) cancellation estimation.

Figure 9 illustrates the estimates from multinomial logistic regression, a population-based method. The estimates tend to have small fluctuations around an approximately fixed mean however, in general the estimates somehow resembles the true pattern of the real no-show and cancellation. In addition, the difference between the estimated and real estimates significantly increases for patients with the tendency of not cancelling their appointments (those could be either patients with good records of showing-up or those with high rate of no-show).

Finally, Figure 10 shows the results from the pure Bayesian update method, which is a popular individual-based method. The pure Bayesian update can basically detect the fluctuations in the real series correctly; however, the estimates are far from the real ones in considerable number of cases. Further analysis revealed that such cases contain a few numbers of attendance records, which means that the pure Bayesian parameters update could not neutralize the effect of prior especially if that is far from the real case.

Such a result was expected because, as discussed in Section 2, Bayesian update (as an individual based method) can effectively learn the (attendance) behaviors of each patient. However, as discussed in Section 2, Bayesian updating (as an individual based method) typically uses non-informative prior, which assign equal chance (0.3333) to

Table 8. Result of clustering the clinics based on their no-show and cancellation probabilities (show-up probability is the complement of no-show and cancellation probabilities)

Parameter	Attribute	Cluster 1		Cluster 2		Cluster 3	
μ	No-show	0.1300		0.2154		0.3866	
	Cancellation	0.3290		0.0877		0.2563	
	Show-up	0.5410		0.6969		0.3571	
Σ	No-show	0.0036	0.0258	0.0258	0.0039	0.0320	−0.0264
	Cancellation	−0.0001	0.0039	0.0039	0.0074	−0.0264	0.0219

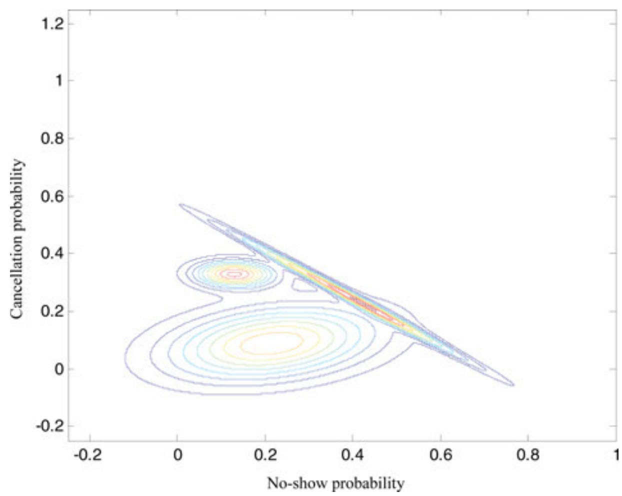


Fig. 11. The contour plot of the mixture of distributions for no-show and cancellation probabilities resulted from applying GMM.

each of no-show, cancellation and attendance,and non-informative prior may not be an appropriate choice in many cases as shown in Figure 10.

In summary, the results from Figures 5 to 10 and their follow up discussions show applying Bayesian update (as an individual-based method) to the initial estimate resulted from multinomial logistic regression (as a population-based method) works superior to Bayesian update and multinomial logistic regression applied individually, and compares favorably with population-based, individual-based and adaptation algorithms.

6. Conclusion and future work

Efficacy of any scheduling system primarily depends on its ability to forecast and manage different types of disruptions and uncertainties. In this paper, we developed a probabilistic model based on multinomial logistic regression and Bayesian inference to estimate individuals' probabilities of no-show, cancellation and attendance in real-time. Based on real patient data collected from a Veterans Affairs medical hospital, we evaluated and showed the effectiveness of the approach. We also modeled the effect of the appointment date and clinic on the proposed method. Our approach is computationally effective and easy to implement. Unlike population-based methods, it takes into account the individual behavior of patients. Also, in contrast to the individual-based methods, it can utilize some valuable information from the complete patient database to provide reliable probabilistic estimates. The result of the proposed method can be used to develop more effective appointment scheduling systems and more precise overbooking strategies to reduce the negative effect of no-shows and fill in appointment slots while maintaining short waiting times.

References

Adomavicius, G., and Tuzhilin, A. (2005) Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, **17**(6), 734–749.

Agresti, A. (2002) *Categorical data analysis* (Vol. 359). John Wiley and Sons.

Aitchison, J. (1985). Practical Bayesian problems in simplex sample spaces. *Bayesian Statistics*, **2**, 15–31.

Alaeddini, A., Yang, K., Reddy, C., and Yu, S. (2011) A probabilistic model for predicting the probability of no-show in hospital appointments. *Health Care Management Science*, **14**(2), 146–157.

Alafaireet, P., Houghton, H., Petroski, G., Gong, Y., and Savage, G. T. (2010) Toward determining the structure of psychiatric visit non-adherence. *The Journal of Ambulatory Care Management*, **33**(2), 108–116.

Allison, P. (1999) *Logistic Regression Using SAS®: Theory and Application*. SAS Publishing.

Alpaydin, E. (2004) *Introduction to Machine Learning*. MIT Press.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A., and Nielsen, H. (2000) Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, **16**(5), 412–424.

Barron, W. M. (1980) Failed appointments. Who misses them, why they are missed, and what can be done. *Primary Care*, **7**(4), 563–574.

Bech, M. (2005) The economics of non-attendance and the expected effect of charging a fine on non-attendees. *Health Policy*, **74**(2), 181–191.

Bigby, J., Pappius, E., Cook, E. F., and Goldman, L. (1984) Medical consequences of missed appointments. *Archives of Internal Medicine*, **144**(6), 1163–1166.

Bloomfield, P. (2004) *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons.

Bolstad, W. M. (2007) *Introduction to Bayesian Statistics*. John Wiley & Sons.

Bowser, D. M., Utz, S., Glick, D., and Harmon, R. (2010) A systematic review of the relationship of diabetes mellitus, depression, and missed appointments in a low-income uninsured population. *Archives of Psychiatric Nursing*, **24**(5), 317–329.

Brockwell, P. J., and Davis, R. A. (2009) *Time Series: Theory and Methods*. Springer.

Can, S., Macfarlane, T., and O'Brien, K. D. (2003) The use of postal reminders to reduce non-attendance at an orthodontic clinic: a randomised controlled trial. *British Dental Journal*, **195**(4), 199–201.

Cayirli, T., and Veral, E. (2003) Outpatient scheduling in health care: a review of literature. *Production and Operations Management*, **12**(4), 519–549.

Cayirli, T., Veral, E., and Rosen, H. (2006) Designing appointment scheduling systems for ambulatory care services. *Health Care Management Science*, **9**(1), 47–58.

Cayirli, T., Veral, E., and Rosen, H. (2008) Assessment of patient classification in appointment system design. *Production and Operations Management*, **17**(3), 338–353.

Chakraborty, S., Muthuraman, K., and Lawley, M. (2010) Sequential clinical scheduling with patient no-shows and general service time distributions. *IIE Transactions*, **42**(5), 354–366.

Chatfield, C. (2013) *The Analysis of Time Series: An Introduction*. CRC Press.

Cleveland, W. S. (1993) *Visualizing Data*. Hobart Press.

Cohen, A. D., Dreiher, J., Vardy, D. A., and Weitzman, D. (2008) Nonattendance in a dermatology clinic–a large sample analysis. *Journal of the European Academy of Dermatology and Venereology*, **22**(10), 1178–1183.

Cote, M. J. (1999) Patient flow and resource utilization in an outpatient clinic. *Socio-Economic Planning Sciences*, **33**(3), 231–245.

Corfield, L., Schizas, A., Williams, A., and Noorani, A. (2008) Non-attendance at the colorectal clinic: a prospective audit. *Annals of the Royal College of Surgeons of England*, **90**(5), 377.

Daggy, J., Lawley, M., Willis, D., Thayer, D., Suelzer, C., DeLaurentis, P. C., . . . and Sands, L. (2010) Using no-show modeling to improve clinic performance. *Health Informatics Journal*, **16**(4), 246–259.

Denhaerynck, K., Manhaeve, D., Dobbels, F., Garzoni, D., Nolte, C., and De Geest, S. (2007) Prevalence and consequences of nonadherence to hemodialysis regimens. *American Journal of Critical Care*, **16**(3), 222–235.

Dove, H. G., and Schneider, K. C. (1981) The usefulness of patients' individual characteristics in predicting no-shows in outpatient clinics. *Medical Care*, **19**(7), 734–740.

Evans, M., Hastings, N., and Peacock, B. (2000) *Statistical Distributions* Wiley.

Fischman, D. (2010) Applying lean six sigma methodologies to improve efficiency, timeliness of care, and quality of care in an internal medicine residency clinic. *Quality Management in Healthcare*, **19**(3), 201–210.

Forster, J. J., and Skene, A. M. (1994) Calculation of marginal densities for parameters of multinomial distributions. *Statistics and Computing*, **4**(4), 279–286.

Gany, F., Ramirez, J., Chen, S., and Leng, J. C. (2011) Targeting social and economic correlates of cancer treatment appointment keeping among immigrant Chinese patients. *Journal of Urban Health*, **88**(1), 98–103.

Garuda, S. R., Javalgi, R. G., and Talluri, V. S. (1998) Tackling no-show behavior: a market-driven approach. *Health Marketing Quarterly*, **15**(4), 25–44.

George, A., and Rubin, G. (2003) Non-attendance in general practice: a systematic review and its implications for access to primary health care. *Family Practice*, **20**(2), 178–184.

Glowacka, K. J., Henry, R. M., and May, J. H. (2009) A hybrid data mining/simulation approach for modeling outpatient no-shows in clinic scheduling. *Journal of the Operational Research Society*, **60**, 1056–1068.

Goutis, C. (1993) Bayesian estimation methods for contingency tables. *Journal of the Italian Statistical Society*, **2**(1), 35–54.

Gupta, D., and Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, **40**(9), 800–819.

Guse, C. E., Richardson, L., Carle, M., and Schmidt, K. (2003) The effect of exit-interview patient education on no-show rates at a family practice residency clinic. *The Journal of the American Board of Family Practice*, **16**(5), 399–404.

Hardy, K. J., O'Brien, S. V., and Furlong, N. J. (2001) Quality improvement report: Information given to patients before appointments and its effect on non-attendance rate. *BMJ: British Medical Journal*, **323**(7324), 1298.

Hassin, R., and Mendel, S. (2008) Scheduling arrivals to queues: A single-server model with no-shows. *Management Science*, **54**(3), 565–572.

Hilbe, J. M. (2009) *Logistic Regression Models*. CRC Press, Boca Raton, FL.

Hixon, A. L., Chapman, R. W., and Nuovo, J. (1999) Failure to keep clinic appointments: implications for residency education and productivity. *Family Medicine-Kansas City*, **31**, 627–630.

Ho, C. J., and Lau, H. S. (1992) Minimizing total cost in scheduling outpatient appointments. *Management Science*, **38**(12), 1750–1764.

Jensen, F. V. (1996) *An Introduction to Bayesian Networks* (Vol. 210). UCL Press, London.

Kopach, R., DeLaurentis, P. C., Lawley, M., Muthuraman, K., Ozsen, L., Rardin, R., . . . and Willis, D. (2007) Effects of clinical characteristics on successful open access scheduling. *Health Care Management Science*, **10**(2), 111–124.

Kotsiantis, S. B. (2007) Supervised machine learning: a review of classification techniques. *Informatica*, **31**(3), 03505596.

Koskela, T., Lehtokangas, M., Saarinen, J., and Kaski, K. (1996, September). Time series prediction with multilayer perceptron, FIR and Elman neural networks. In *Proceedings of the World Congress on Neural Networks* (pp. 491–496).

LaGanga, L. R. (2011) Lean service operations: reflections and new directions for capacity expansion in outpatient clinics. *Journal of Operations Management*, **29**(5), 422–433.

LaGanga, L. R., and Lawrence, S. R. (2007). Clinic overbooking to improve patient access and increase provider productivity. *Decision Sciences*, **38**(2), 251–276.

Lehmann, T. N. O., Aebi, A., Lehmann, D., Balandraux Olivet, M., and Stalder, H. (2007) Missed appointments at a Swiss university outpatient clinic. *Public Health*, **121**(10), 790–799.

Leonard, T. (1973) A Bayesian method for histograms. *Biometrika*, **60**(2), 297–308.

Liu, N., Ziya, S., and Kulkarni, V. G. (2010) Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing and Service Operations Management*, **12**(2), 347–364.

Mitchell, A., and Selmes, T. (2007) A comparative survey of missed initial and follow-up appointments to psychiatric specialties in the United Kingdom. *Psychiatric Services*, **58**(6), 868–871.

Moore, C. G., Wilson-Witherspoon, P., and Probst, J. C. (2001) Time and money: effects of no-shows at a family practice residency clinic. *Family Medicine-Kansas City*, **33**(7), 522–527.

Murphy, K., Edelstein, H., Smith, L., Clanon, K., Schweitzer, B., Reynolds, L., and Wheeler, P. (2011) Treatment of HIV in outpatients with schizophrenia, schizoaffective disorder and bipolar disease at two county clinics. *Community Mental Health Journal*, **47**(6), 668–671.

Murray, M. M., and Tantau, C. (2000) Same-day appointments: exploding the access paradigm. *Family Practice Management*, **7**(8), 45–45.

Muthuraman, K., and Lawley, M. (2008) A stochastic overbooking model for outpatient clinical scheduling with no-shows. *IIE Transactions*, **40**(9), 820–837.

Neal, R. D., Hussain-Gambles, M., Allgar, V. L., Lawlor, D. A., and Dempsey, O. (2005) Reasons for and consequences of missed appointments in general practice in the UK: questionnaire survey and prospective review of medical records. *BMC Family Practice*, **6**(1), 47.

Obialo, C. I., Bashir, K., Goring, S., Robinson, B., Quarshie, A., Al-Mahmoud, A., and Alexander-Squires, J. (2008). Dialysis "no-show" on Saturdays: Implications of the weekly hemodialysis schedules on nonadherence and outcomes. *Journal of the National Medical Association*, **100**(4), 412–419.

O'Brien, S. M., and Dunson, D. B. (2004) Bayesian multivariate logistic regression. *Biometrics*, **60**(3), 739–746.

Park, W. B., Kim, J. Y., Kim, S. H., Kim, H. B., Kim, N. J., Oh, M. D., and Choe, K. W. (2008) Self-reported reasons among HIV-infected patients for missing clinic appointments. *International Journal of STD & AIDS*, **19**(2), 125–126.

Quinlan, J. R. (1986) Induction of decision trees. *Machine Learning*, **1**(1), 81–106.

Reddy, C. K., Chiang, H. D., and Rajaratnam, B. (2008) Trust-tech-based expectation maximization for learning finite mixture models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **30**(7), 1146–1157.

Reynolds, D. A., Quatieri, T. F., and Dunn, R. B. (2000) Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, **10**(1), 19–41.

Richards F. J. (1959) A flexible growth function for empirical use. *Journal of Experimental Botany*, **10**(2), 290–301.

Rowett, M., Reda, S., and Makhoul, S. (2010) Prompts to encourage appointment attendance for people with serious mental illness. *Schizophrenia Bulletin*, **36**(5), 910–911.

Rust, C. T., Gallups, N. H., Clark, W. S., Jones, D. S., and Wilcox, W. D. (1995).Patient appointment failures in pediatric resident continuity clinics. *Archives of Pediatrics & Adolescent Medicine*, **149**(6), 693.

Sarnquist, C. C., Soni, S., Hwang, H., Topol, B. B., Mutima, S., and Maldonado, Y. A. (2011) Rural HIV-infected women's access to medical care: ongoing needs in California. *AIDS Care*, **23**(7), 792–796.

Savageau, J. A., Ferguson, W., Lemay, C. A., and Cashman, S. B. (2004) Patient health status and appointment keeping in an urban community health center. *Journal of Health Care for the Poor and Underserved*, **15**(3), 474–488.

Schectman, J. M., Schorling, J. B., and Voss, J. D. (2008) Appointment adherence and disparities in outcomes among patients with diabetes. *Journal of General Internal Medicine*, **23**(10), 1685–1687.

Simonoff, J. S. (1996) *Smoothing Methods in Statistics*. Springer.

Stratonovich, R. L. (1960) Conditional Markov processes. *Theory of Probability & Its Applications*, **5**(2), 156–178.

Turkcan, A., Nuti, L., DeLaurentis, P. C., Tian, Z., Daggy, J., Zhang, L., . . . and Sands, L. (2013. No-show modeling for adult ambulatory clinics. In *Handbook of Healthcare Operations Management* (pp. 251–288). Springer, New York.

Tuller, D. M., Bangsberg, D. R., Senkungu, J., Ware, N. C., Emenyonu, N., and Weiser, S. D. (2010) Transportation costs impede sustained adherence and access to HAART in a clinic population in southwestern Uganda: a qualitative study. *AIDS and Behavior*, **14**(4), 778–784.

Viola, P., and Jones, M. (2002) Fast and robust classification using asymmetric adaboost and a detector cascade. *Advances in Neural Information Processing Systems*, **2**, 1311–1318.

Weston, J., and Watkins, C. (1998) *Multi-class support vector machines*. Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, May.

Winters, P. R. (1960) Forecasting sales by exponentially weighted moving averages. *Management Science*, **6**(3), 324–342.

Yehia, B. R., Gebo, K. A., Hicks, P. B., Korthuis, P. T., Moore, R. D., Ridore, M., and Mathews, W. C. (2008) Structures of care in the clinics of the HIV Research Network. *AIDS Patient Care and STDs*, **22**(12), 1007–1013.

Zeng, B., Turkcan, A., Lin, J., and Lawley, M. (2010) Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities. *Annals of Operations Research*, **178**(1), 121–144.

Appendix: Clinics Clustering

Due to the variety of clinics (more than 270 in our case), the accuracy of the logistic regression would be severely affected if this explanatory variable is directly used in the model. This problem is solved by clustering similar clinics with respect to their no-show, cancellation and attendance rates. The clinics are originally different in type; hence, grouping them into a set of clusters will result in clusters with different density and dispersion. Such characteristics can be effectively modeled using Gaussian Mixture Models (GMM); when clusters have different sizes and correlation within them, like the clinic data in this paper, GMM can be more appropriate than many of other clustering algorithms as k-means. (Alpaydin [2010], Reddy *et al.* [2008]).

GMM employs Expectation Maximization (EM) algorithm to fit data, which assigns posterior probabilities to each component (clusters of clinics) density with respect to each observation. Choosing a suitable number of components is essential for creating a useful GMM model. Here, we use Bayes Information Criterion (BIC) to determine an appropriate number of components for the model. To address the (potential) problem of clinics with very few appointment records, e.g., 1 or 2 appointments, that may cause formation of clusters with no-show/cancellation/attendance probability of zero or one, clinics have been filtered based on their number of appointments first, and only clinics with more than 6 records have been considered for clustering with GMM (The filter threshold value of 6 has been chosen based on evaluating different (threshold) values and their effect on the size of clusters and robustness of the estimated parameters of the model).

Table 8 shows the result of clustering the clinics based on their probability of no-show and cancellation using GMM (see also Fig. 11). Since the probability of attendance is the complement of other two no-show and cancellation probabilities (dependent variable) and does not contain additional information to the GMM model, it hasn't been considered for the clustering procedure, and its respected value for different clusters has been calculated based on no-show and cancellation. The final result has been verified by a team of experts and clusters represented meaningful groupings.