MIS Quarterly

SPECIAL ISSUE

A DATA ANALYTICS FRAMEWORK FOR SMART ASTHMA MANAGEMENT BASED ON REMOTE HEALTH INFORMATION SYSTEMS WITH BLUETOOTH-ENABLED PERSONAL INHALERS[1]

Junbo Son
Alfred Lerner College of Business & Economics, University of Delaware,
Newark, DE 19716 U.S.A. {junboson@udel.edu}

Patricia Flatley Brennan
National Library of Medicine, National Institutes of Health,
Bethesda, MD 20894 U.S.A. {patti.brennan@nih.gov}

Shiyu Zhou
Department of Industrial & Systems Engineering, University of Wisconsin–Madison,
Madison, WI 53706 U.S.A. {shiyuzhou@wisc.edu}

Asthma is a prevalent respiratory chronic disease affecting a large portion of the global population. Patients diagnosed with asthma may experience significantly reduced quality of life if their asthma is not properly controlled. To facilitate better asthma self-management, asthma specialists and engineers have developed the smart asthma management system (SAM). This new health information system provides Bluetooth-enabled inhalers and collects time stamps of every rescue inhaler use. Such detailed inhaler usage logs, which are crucial for investigating patterns of inhaler usage, were not available in traditional clinical trials because clinical trials acquire data only periodically. Due to the low data collection resolution of clinical trials, quantitative asthma studies based on trial data have been focusing mainly on capturing the increasing trend in the number of rescue inhaler uses. Taking advantage of the patient monitoring capability of the SAM system, we developed a data analytics framework for detecting abnormal inhaler use that is out of the patient's normal usage pattern. The new statistical model developed in this paper can address the key features of the data collected from the SAM system such as the heterogeneous impact of environmental factors on inhaler usage behavior and the correlation structure governed by the patient's repetitive routines. We show the satisfactory performance of our data analytics framework through rigorous comparison with various benchmark methods. Furthermore, we give an in-depth discussion on our contribution to the information systems (IS) knowledge base and practical implications of our analytics framework to data-driven asthma management practice.

Keywords: Smart asthma management, design science, healthcare data analytics, health IT, health information system

[1] The accepting senior editors for this paper were Indranil Bardhan, Hsinchun Chen, and Elena Karahanna.

DOI: 10.25300/MISQ/2020/15092

Introduction

Asthma is one of the most prevalent chronic diseases, affecting approximately 24.6 million people in the United States in 2015 (7.8% of the U.S. population) (CDC 2016). Patients diagnosed with asthma may experience symptoms including coughing, wheezing, chest tightness, and shortness of breath. If asthma is not controlled properly, symptoms can significantly reduce patients' quality of life (Carranza et al. 2004). In addition, uncontrolled asthma introduces a substantial amount of economic burden. Just for the United States alone, the asthma-related societal cost is estimated as $81.9 billion (Nurmagambetov et al. 2017). One of the key components for properly controlling asthma is asthma self-management, which requires consistent monitoring of patient's asthma symptoms and medication usage (Ahmed et al. 2016).

In the current practice of asthma self-management, patients are asked to gauge their asthma control level by referring to asthma action plans or guidelines that their doctors provide. Such action plans are mainly based on a threshold specified by the number of times the patient had to use his/her rescue inhaler (NHLBI 2007; Patel et al. 2014). Practitioners have often used threshold-based approaches due to their simplicity, but there are several issues. First, rescue inhaler use does not directly indicate an unbearable asthmatic symptom. Some patients use the inhaler preemptively before they engage in physical activity (Anderson et al. 2006). Also, even if a patient had to use the inhaler because of symptoms, individual patients have different perceptions of their symptoms, with wide variations related especially to gender and age (Zein and Erzurum 2016). Acknowledging the limitations discussed above, asthma self-management practice has been improved by adopting various health information systems (HIS) (Huckvale et al. 2015). At the same time, both care providers and patients have started to notice the potential of model-based and data-driven asthma self-management using advanced health data analytics (Son et al. 2017). Analytics and HIS cannot be separated from each other, because a new form of HIS often calls for a new analytics method, and the knowledge that the analytics method discovers provides useful feedback to the HIS for promoting better asthma self-management.

Currently, there are diverse asthma management HIS, ranging from a simple electronic asthma action plan replacing paper-based asthma action plans (Licskai et al. 2013) to diary-type smartphone applications and web-based summary dashboards (Casper and Brennan 2013; Wu et al. 2015). Asthma management notably benefits from the use of such HIS because they keep records of patients' asthma-related events (Merchant et al. 2016). However, the existing HIS for asthma care rely mainly on patients' self-reported data (Ahmed et al. 2016). The data that patients manually and voluntarily report have several issues, including human input errors and an excessive amount of missing observations (Son et al. 2017). The most critical issue is that patients' personal perceptions often heavily affect the self-reported health data (Janssens et al. 2009). For instance, some patients may report minor coughing as an asthma-related event, whereas other patients may ignore it.

To overcome these issues, a new HIS, the smart asthma management (SAM) system, has been developed. The SAM system is a novel asthma management HIS that uses remote patient monitoring technology. The SAM platform is equipped with a Bluetooth sensor that is attachable to various types of personal inhalers. The sensor attachment records the date and time of every inhaler use of individual patients. The SAM system not only provides a detailed history of the patients' inhaler usage, but also several asthma-related environmental factors (e.g., temperature and air pollution level). The SAM system has been used for monitoring medication adherence, improving asthma control, and providing summary reports to patients and healthcare providers (Chan et al. 2015; Van Sickle et al. 2013). It should be noted that the development of the SAM system provides a new data source for quantitative analysis in asthma management. Traditionally, the major data source for data-driven asthma studies, or quantitative health studies in general, has been clinical trials (Lin et al. 2017). Trial-based health data provide rich information about a patient's asthma condition because the data were obtained through a series of lab tests and diagnosis made by medical experts. On the other hand, data acquisition in the SAM system is remotely done through sensors and wireless network; hence, inevitably, SAM data cannot be as comprehensive as trial-based data. However, the SAM system can monitor asthma patients in almost real time, and the remote patient monitoring technology embedded in the SAM platform provides a significant research opportunity in healthcare data analytics. The data collection resolution of periodically conducted clinical trials is often too low to highlight inhaler usage patterns in detail. In contrast, the high-frequency data acquisition capability of the SAM system now allows us to model the inhaler usage patterns of individual patients. For instance, the SAM data can be used for investigating individual patients' number of rescue inhaler uses in every one-hour window across different days of the week.

Although there are many analytics methods readily available in the literature, it is often necessary to develop a specialized analytics method tailored to the new types of patient-level health data to facilitate knowledge discovery (Kohli and Tan 2016). For that reason, in the recent IS literature, many design science studies have been focusing on developing innovative information technology (IT) artifacts for various health data other than traditional trial-based data (e.g., electronic health records, EHRs) (Lin et al. 2017). In the same spirit, to take full advantage of the data availability provided by the SAM system, we aim to develop an analytics framework that can visualize and model the inhaler usage pattern, so that we can detect an inhaler use that is out of the patient's normal usage pattern. Our research objectives can be summarized as

- To provide a visualization tool highlighting inhaler usage patterns
- To provide a statistical model specifically tailored to the SAM data
- To provide a data-driven method for detecting deviations from normal inhaler usage patterns

Addressing the research questions listed above has a great potential for improving the current practice of asthma self-management because patients diagnosed with asthma can achieve better asthma control by understanding their own inhaler usage patterns. As discussed earlier, so far, asthma patient monitoring has been conducted in a way that emphasizes trends (Son et al. 2016). In other words, the data analytics methods in asthma management have been focusing on identifying increasing trends in the number of inhaler uses. This trend-based approach does not require real-time data acquisition capability; hence, it is often used for traditional trial-based health data. Trend-based methods are useful for assessing the risk of gradual and persistent degradation of asthma control level (Son et al. 2017), but it may not be easy to detect a single unexpected inhaler use that is out of the patient's typical usage pattern. Such out-of-pattern usage provides a different kind of meaningful information, because the abnormal inhaler use might be caused by short-term exposure to an asthma trigger, sudden exacerbation, or exercise-induced symptoms (Anderson et al. 2006).

From the methodological perspective, our research problem boils down to developing a method to investigate the mechanism of a certain type of event occurrence (inhaler use). Therefore, we can find several existing methods that are applicable to our problem. Because the dependent variable in our study can be viewed as count data (the number of rescue inhaler uses within a given time window), we can use the generalized linear mixed-effects models (GLMMs) based on either Poisson or negative binomial distributions (Agresti 2001). In addition, we may treat rescue inhaler usage as a binary variable (i.e., "1" if the patient has used his/her inhaler within a certain time window and "0" otherwise), and use a different type of GLMM such as the logistic regression model (Son et al. 2016). However, the GLMMs share a common drawback for modeling SAM data. SAM data is categorized as observations of daily living (ODLs) and analysis based on ODLs should acknowledge the nuances of the context and environment in which patients live, as well as the patients' daily routine (Backonja et al. 2012). The repetitive routine of the individual patients introduces a unique correlation among data points (e.g., there should be correlated observations if a patient routinely exercises at a certain time across different days of the week). The conventional GLMMs provide a limited flexibility for specifying a correlation structure among event occurrences (Pinheiro and Bates 2000). We may consider a well-established classifier such as the support vector machine (SVM), because the goal of our study is to develop a model for identifying abnormal inhaler use that is out of the patient's regular usage pattern (Bennett and Campbell 2000). We can train the SVM based on the individual patient's SAM data and perform the abnormality detection. However, the SVM has a potential issue in the SAM application because training the SVM is mainly done based on data from a specific individual patient. In other words, individual-level classifiers often ignore the common characteristics of asthma shared by all patients diagnosed with asthma. For instance, asthma patients whose asthma control test scores are low typically have more nocturnal inhaler uses (NHLBI 2007). Such information can be captured better with regression models than the individual-level classifiers. In addition, each individual patient reacts to the same environmental asthma trigger differently, and the environmental factors are assumed to have impacts on the variability of inhaler usage (Su et al. 2017). To properly quantify the effects of environmental asthma triggers, the model should be able to define a sub-model dedicated to characterizing the variability of rescue inhaler use. In general, such dispersion models cannot be established by both the GLMMs and individual-level classifiers.

To fill the research gap discussed above, in this paper we propose a new data analytics method: the generalized linear mixed effects model with a grid-based quasi-Poisson distribution (GLMM-GQP). Our data analytics framework begins with data transformation. The SAM data is longitudinal and, instead of directly modeling the longitudinal data, we transform the data into a grid format. The grid separates the entire week into 168 grids (24 hours a day and 7 days a week) and sums the number of inhaler uses of individual patients within each cell. This data transformation helps us visualize the inhaler usage pattern within a week. Then, we fit the GLMM-GQP to the grid data. The GLMM-GQP has several features that are desirable for the SAM application. First, the GLMM-GQP includes patient-level and grid-level random effects. Therefore, detecting out-of-pattern inhaler usage can be personalized to each patient. Second, the GLMM-GQP has a sub-model for characterizing the variability of inhaler usage behavior. This means that the GLMM-GQP models not only the mean (intensity of inhaler use) but also the dispersion (variability of the inhaler use); hence, we can use it for modeling the impact of environmental asthma triggers. Third, the GLMM-GQP adopts the conditional autoregressive (CAR) model for the grid-level random effects. With the CAR model structure, any kind of correlation structure among rescue inhaler uses can be flexibly specified in the GLMM-GQP reflecting the patient's repetitive daily routine. For instance, if a patient exercises at 4:00 p.m. every day, we can define the correlation structure accordingly so that the inhaler uses observed in the same time windows across different days of the week are correlated.

The rest of paper is organized as follows: In the next section, we provide a review on both the data analytics literature and the IS literature that are relevant to our study. In the model development section, we describe our test bed and the SAM data used in the study followed by a description of the proposed analytics method. Then, we report the performance evaluation results, comparing our approach to the benchmark methods. Finally, we conclude the paper with a summary of our contributions and future research directions motivated by the limitations of the current study. We also discuss the practical implications and provide guidelines to IS scholars and practitioners.

Literature Review

Literature on Statistical Models

In clinical literature, one of the popular statistical models used in trial-based asthma studies is the Markov model (Saint-Pierre et al. 2006). The Markov model in asthma studies can quantify the impacts of various biomarkers on the transition probabilities among different levels of asthma control (Combescure et al. 2003). For instance, the Markov model can tell whether body mass index is positively or negatively associated with the transition probability from optimal asthma control to unacceptable asthma control (Saint-Pierre et al. 2003). With this model, we can assess the risk of transitioning to uncontrolled asthma for individual patients whenever updated measurements from that specific patient become available. Despite the promising capabilities of the Markov model, it is not directly applicable to the SAM platform because one of the crucial inputs for the Markov model is the clinically defined asthma control level that practitioners grade through a series of clinical evaluations and lab tests (Saint-Pierre et al. 2006). Because the SAM system acquires data from the asthma patients remotely through sensors, on-site clinical evaluations of asthma control based on lab tests are not available; hence, fitting the Markov model to the SAM data is practically infeasible. To resolve this issue, an extended Markov model called the correlated gamma-based hidden Markov model (CG-HMM) recently has been developed to detect a persistent and gradual degradation of asthma control (Son et al. 2017). The CG-HMM is based on a hidden Markov model where the true asthma condition is assumed to be hidden but can be inferred by the number of rescue inhaler uses per day. Although the performance of the CG-HMM is satisfactory for identifying gradual degradation in asthma control level, the CG-HMM may not be suitable for our purpose. The objective of our study is to detect a sudden inhaler use that is out of the patient's normal inhaler usage pattern. Because the CG-HMM uses aggregated daily number of inhaler uses, it is challenging to characterize the micro-level rescue inhaler usage patterns.

In addition to the statistical models that frequently have been used in asthma studies, the generalized linear models (GLMs) and GLMMs have been the most natural choice for modeling various event occurrences (Agresti 2001). As the name suggests, the GLMM-GQP proposed in this paper is an extended version of the GLMM, integrating various statistical methods into a unified model structure. The basis of the GLMM-GQP is the GLMM with patient-level and grid-level random effects. To address the common issue of over-dispersion in count data, the GLMM-GQP adopts the quasi-Poisson distribution (Wedderburn 1974). However, the conventional GLMM based on the quasi-Poisson distribution cannot address some of the key characteristics of the SAM data. First, as mentioned earlier, various environmental factors affect the variability rather than the mean of inhaler usage (Su et al. 2017). Therefore, it is important to have a sub-model dedicated to characterizing the dispersion. The quasi-Poisson distribution is defined by two parameters: the mean (intensity) and dispersion, and the GLMM based on the quasi-Poisson distribution focuses on modeling the mean (Clayton and Kaldor 1987). For the GLMs, which are simpler than the GLMMs, there exists a rigorous way to model both the mean and dispersion. Because the model structure is doubled, it is called the double GLM (DGLM) and researchers have often used it for analyzing insurance claims data (Smyth and Jorgensen 2002). The idea of DGLM fits well to the SAM application. However, it is not readily available for the GLMMs. Further, there is another issue with the existing GLMMs: The rescue inhaler usage behavior is inevitably influenced by the individual patient's daily routine. For instance, a patient who regularly exercises at 4:00 p.m. every day should show a positive correlation in the inhaler usage pattern at 4:00 p.m. across different days of the week. In spatial statistics, there are several methods to address this type of structured correlation, such as the conditional autoregressive (CAR) model (Cressie 1993). The CAR model provides a powerful tool for modeling the correlation structure among rescue inhaler uses across different times of the day and dif-

Table 1. Differences in the Relevant Methods

	Characteristics and Capabilities			
Model	Fixed Effects	Random Effects	Dispersion Model	Special Correlation
GLM	✓			
GLMM	✓	✓		Limited
DGLM	✓		✓	
GLMM-GQP	✓	✓	✓	✓

ferent days of the week. However, to the best of our knowledge, the CAR model has been primarily used only in the field of spatial statistics, and it has not been integrated into the GLMMs.

To summarize, the GLMM-GQP integrates three statistical methods to establish a specialized regression model tailored to the SAM data. The GLMM-GQP has the log-normal random effects as in the conventional GLMMs and defines a separate dispersion model as in the DGLM. Furthermore, the GLMM-GQP adopts the CAR model to specify the distribution of the grid-level random effects with a flexible correlation structure among observations. The differences between the GLMM-GQP and other relevant statistical models are highlighted in Table 1.

IS Literature and Positioning of Our Study

In addition to highlighting methodological contributions by contrasting our analytics method with the statistical models that are available in the literature, we would like to clarify the position of our work in the IS literature. Driven by recent advancements and increased penetration of IT in healthcare, it is recognized that the IS community can play an important role in enhancing current healthcare practices (Buntin et al. 2011; Kolodner et al. 2008). Given the diverse domains involved, it is not trivial to realize the potential that IS can offer in terms of improving the quality of healthcare (Kane and Labianca 2011). However, it is possible to further the impact of IS on healthcare by developing healthcare analytics methods (Fichman et al. 2011). As we discussed earlier, analyzing new types of health data provided by the diverse HIS is often challenging with existing analytics methods. Therefore, such research topics have attracted significant interest among IS scholars. We can find examples in recent IS literature: a predictive analytics method on the readmission of patients with congestive heart failure (Bardhan et al. 2015) and a Bayesian multitask learning approach for risk profiling using diabetes as an example (Lin et al 2017), both based on EHRs. This research expansion on healthcare analytics has facilitated knowledge discovery based on medical data that have been either unavailable or underutilized. For instance, even social media data has been investigated in healthcare analytics IS literature (Kallinikos and Tempini 2014), and researchers suggest that social media will be a promising new foundation for medical knowledge creation.

As reflected by the social media example above, the definition of HIS is evolving and IS scholars can make impactful contributions to healthcare research by developing novel IT artifacts that can solve complex problems in modern healthcare practice (Agarwal et al. 2010). The SAM platform introduced in our study is indeed an example of the extended definition of HIS, and it provides a new way of performing data-driven asthma management. By taking advantage of the new data source, we developed an analytics method that may hold promise for reducing asthma-related costs and improving quality of life for asthma patients. Therefore, our study fits to the current IS design science research stream of health analytics. In HIS literature, experts in healthcare, IS, and various engineering disciplines have proposed the vision of smart and connected health, which aims to offer just-in-time and just-for-me clinical interventions through advancing the HIS and analytics methods (Leroy et al. 2014). The analytics framework developed in this study is for detecting an unusual inhaler use that is out of the individual patient's normal usage pattern so that timely intervention can be made prior to serious degradation of the specific patient's asthma status. Thus, the objectives of our solution artifact are well-aligned with the vision of smart and connected health. The data analytics methods developed in this paper significantly improve the existing methods to provide a better solution artifact for the SAM system, a newly introduced asthma management HIS. In other words, our study is a good example of an exaptation study, which a prevalent type of IS design science research (Gregor and Hevner 2013).

Model Development

The proposed data analytics framework for identifying unusual inhaler usage is composed of three distinctive steps as shown in Figure 1. The first step is the data transforma-

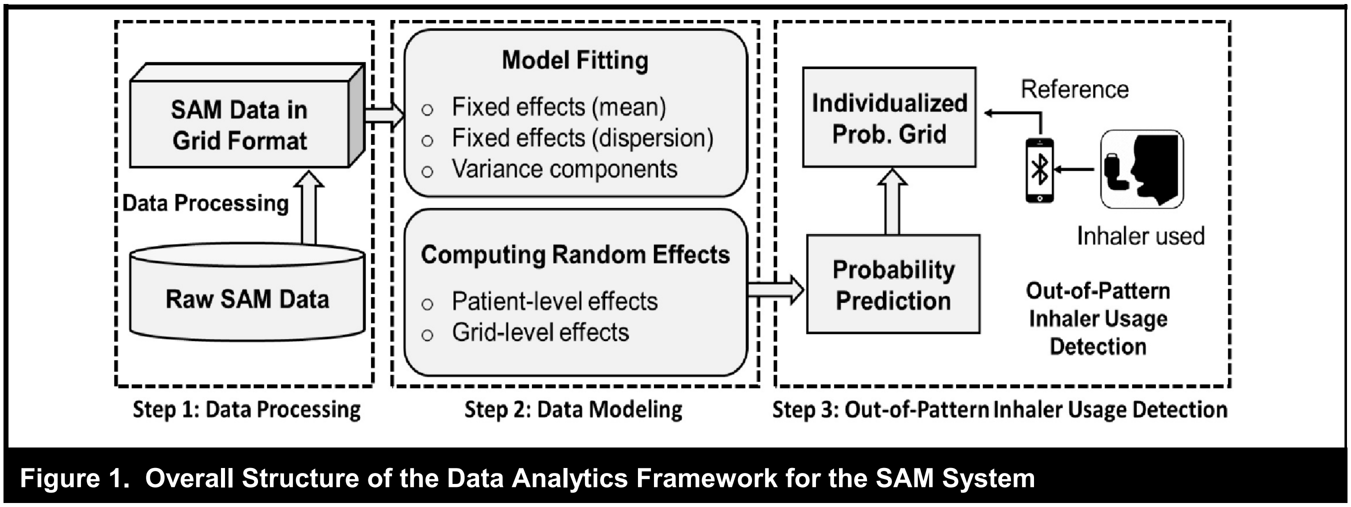


Figure 1. Overall Structure of the Data Analytics Framework for the SAM System

tion. The transformed data provides us not only the basis of model fitting but also a data visualization tool highlighting inhaler usage patterns. The second step is to fit the GLMM-GQP to the SAM data, and the third step is to implement the out-of-pattern inhaler usage detection algorithm.

Description on the Data Collection

To make our discussion easier, we first introduce our test bed and provide a description on the SAM data. We then illustrate the proposed data analytics framework.

Description of the Test Bed

Propeller Health, Madison, WI, implements and manages the SAM platform investigated in this study. Propeller Health has developed a Bluetooth sensor that is attachable to various types of inhaled medications for asthma. The sensor has received 510(k) class II clearance from the U.S. Food and Drug Administration (FDA) (Kim et al. 2016). In addition, according to Federal Communications Commission (FCC) licensing and the FDA's wireless Bluetooth technology testing standards, the Propeller Health sensor demonstrated reliable performance for sensor actuations and data capture (Su et al. 2017). When a patient administers a dose of medication through his/her inhaler, the Bluetooth sensor records the date and time of the inhaler use and transmits the data to the patient's smartphone. For patients who do not own a smartphone, patients can use a wireless hub for data transmission. The collected data travels in an encrypted fashion to the secure Health Insurance Portability and Accountability Act of 1996 (HIPAA)-compliant servers at Propeller Health. If a transmitting device is unavailable at the time of event, the sensor stores the data in its local memory, which can hold approximately 3,900 events, and transmit the information at a later time (Van Sickle et al. 2013).

Our test bed has already demonstrated improved asthma clinical outcomes, such as longer symptom-free days and less frequent use of the inhaler (Barrett et al. 2017; Kim et al. 2016; Merchant et al. 2016). Furthermore, according to a survey study, the participants (asthma patients) were highly satisfied with this new asthma management HIS (Kim et al. 2016). The structure of the SAM platform and its information flows are illustrated in Figure 2.

Data Description

We obtained the data from a real-world study conducted in a mid-size city in the United States from March 2014 to December 2017 using our test bed. The study has an open enrollment procedure, which means that asthma patients can begin and end their participation anytime during the test period. There were 696 participants in total. However, we only used 326 patients in our study, excluding 370 patients because of incomplete demographic data or short participation periods (shorter than two weeks). A list of variables is shown in Table 2 with their summary statistics.

The data contains information about the patients' sex, age, and race, which are the most commonly used demographic variables in asthma research (William et al. 2011). The participants took the asthma control test (ACT) at the time of enrollment. The ACT score reflects the initial asthma control level of each patient, and has been validated for use with subjects at least 12 years of age (Schatz et al. 2006). The ACT score ranges from 5 to 25 for adults. Following the literature,

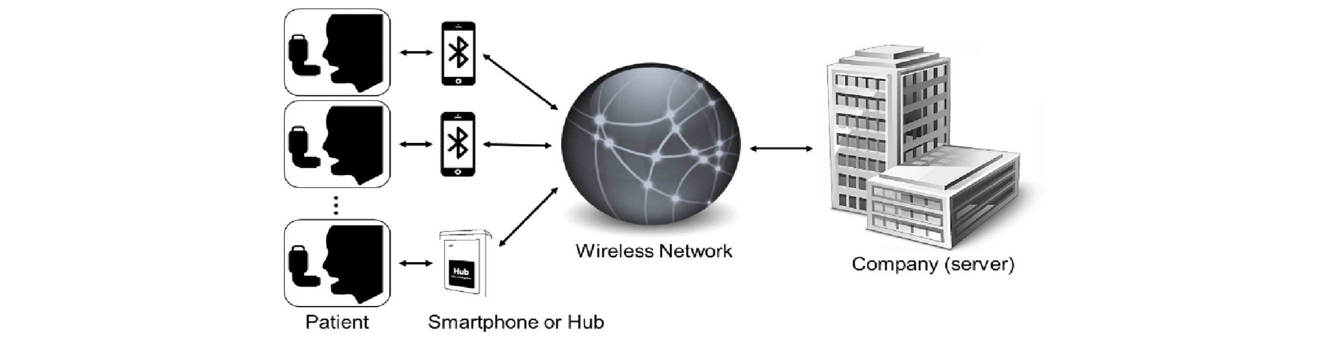


Figure 2. Illustration of the Smart Asthma Management System

Table 2. List of Variables in the SAM Data

Category	Variable	Summary Statistic
Demographic information	Age	Mean: 41.733 (SD: 12.865)
	Sex	Male: 103 patients Female: 223 patients
	Race	White/Caucasian: 224 patients African American: 59 patients Other Race: 43 patients
Patient's asthma-related information	Asthma control test (ACT)-based asthma control level	Mean: 13.917 (SD: 3.945)
Environmental factors	Temperature (°F)	Mean: 61.328 (SD: 17.082)
	Wind speed (meter/s)	Mean: 6.602 (SD: 4.796)
	Humidity (%)	Mean: 65.070 (SD: 18.020)
	Visibility (mile)	Mean: 9.284 (SD: 1.738)
	$PM_{2.5}$ (microgram/m^3)	Mean: 9.510 (SD: 4.616)
Other important variables	Event timestamp	Mean: 0.23 use/day (SD: 1.0615)
	Participation period	Mean: 454.092 days (SD: 354.728)

we categorized participants into two groups based on their ACT scores (controlled and uncontrolled). The patients with ACT scores greater than 19 were categorized as controlled asthma patients (Nathan et al. 2004; Reddel et al. 2009). In terms of environmental factors, we have temperature, wind speed, humidity, visibility, and fine particulate air pollution ($PM_{2.5}$). The $PM_{2.5}$ measures particulate matter with a diameter of ≤ 2.5 microns. Temperature, humidity, and air pollution (and visibility that is closely related to the air pollution) are significant asthma triggers (Vernon et al. 2012). Also, wind speed affects the spread of pollen, which is associated with asthmatic symptoms (Knox 1993). The sensor itself cannot measure the environmental data, so we acquired the data from public data repositories. We obtained temperature, humidity, and wind speed data from the National Oceanic and Atmospheric Administration Quality Controlled Local Climatological Data Repository, and collected the $PM_{2.5}$ data from the Environmental Protection Agency Air Quality System. The SAM system automatically assigned all environmental data to each medication use event from the monitoring station closest to the geographic location of the event.

There are two aspects about the data set that need to be mentioned. First, the sensors periodically send a heartbeat signal to indicate that the Bluetooth sensor is active and alive. Because the battery embedded in the Bluetooth sensor can last more than a year without a charge, we have not noticed any missing data caused by a dead sensor. Second, the SAM data is not a balanced data set. Among 326 participants used in our study, about 68% were female. We had no protocol to balance the number of participants for each gender at the time of enrollment. However, asthma is known to have a higher prevalence in women than in men in adulthood (Postma 2007); hence, it is natural to have more female patients than male. Another imbalance is observed in race of the participants among which 68.7% were White/Caucasian because the city from which the data were collected is not a place with diverse ethnicity (88% White) according to the United States Census Bureau (2016). The initial asthma control level of the participants is also skewed. About 77% of the participants had well-controlled asthma based on ACT scores. This partially explains why we observe few rescue inhaler uses in the SAM data.

Step 1: Data Transformation

The proposed data analytics framework in our study starts from processing the SAM data (i.e., transforming the longitudinal data into a grid format). The SAM data show a unique correlation structure that is defined by the daily routines of patients. The grid transformation performed on the data set enables us to visualize and properly model the correlation structure, which could be challenging to do with data in its original time-series format. During the transformation, we arranged the SAM data into 24 × 7 grids (24 hours a day, 7 days a week). The participation periods for each patient were denoted by m_i (i.e., the number of weeks that patient i participated in the study). We then merged m_i grids for each patient i. We used the one-hour time window in our study because, in the asthma literature, the number of puffs taken within an hour has been widely used as an observational unit in various data analysis (Bender et al. 1998). However, if necessary, the time window can be adjusted. Figure 3 illustrates the data transformation process.

After the transformation, we can see the grid distribution based on the number of inhaler uses as in Figure 4. In Figure 4, we categorized the patients into two groups based on their ACT scores as explained in the data description section. For patients whose asthma is well controlled, it is very rare to see a grid with more than one inhaler use (less than 2%). On the other hand, about 25% of grids have multiple inhaler uses for the poorly controlled asthma patients.

Step 2: Statistical Modeling

The fundamental assumption of the GLMM-GQP is that the number of inhaler uses (y_{ik}) for an individual patient i in a specific day-time grid k follows a quasi-Poisson distribution as

$$y_{ik} \sim \text{QuasiPoisson}(\lambda_{ik}, \theta_{ik}) \text{ where } \mathrm{E}(y_{ik}) = \lambda_{ik} \\ \text{and } \mathrm{Var}(y_{ik}) = \theta_{ik}\lambda_{ik}$$

where mean (intensity) of y_{ik} is denoted as λ_{ik} and the variance of y_{ik} is specified as $\theta_{ik}\lambda_{ik}$. A quasi-Poisson distribution is defined by the first two moments, and it can address the issue of over-dispersion (i.e., $\mathrm{E}(y_{ik}) < \mathrm{Var}(y_{ik})$) by including the dispersion parameter θ_{ik}. In our model, we separately modeled the two parameters of a quasi-Poisson distribution (λ_{ik} and θ_{ik}).

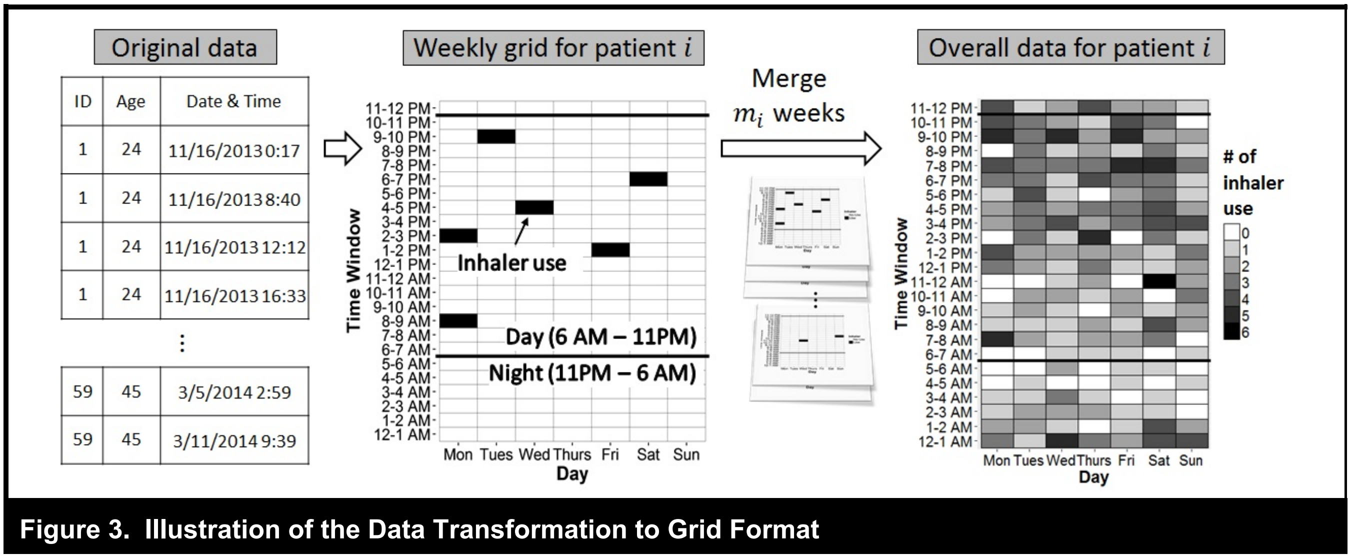
Model Description

Mean Model: The mean model has two types of effects: fixed and random. Therefore, the structure of the proposed model is equivalent to the GLMM. The mean model is defined as

$$\begin{aligned} \log(\lambda_{ik}) = \log(m_i) + \beta_0 + \beta_1 Male_i + \beta_2 Age_i + \\ \beta_3 RaceOther_i + \beta_4 RaceWhite_i + \\ \beta_5 ACTcontrol_i + \beta_6 Weekend_k + u_i + z_k \end{aligned} \tag{1}$$

for $i = 1, \ldots, 326$ (patient ID) and $k = 1, \ldots 168$ (grid index). We considered sex ($Male_i$ = 1 for male patients and 0 for female patients), race ($RaceOther_i$ = 1 for patients who identified themselves as Hispanic, Asian, or Native American and $RaceWhite_i$ = 1 for White/Caucasian where African American is the reference), ACT-based asthma control level ($ACTcontrol_i$ = 1 if the ACT score of patient i was smaller than or equal to 19 points at the time of enrollment), and weekend indicator ($Weekend_k$ = 1 if a specific grid k belongs to weekend). We have coefficients (fixed effects) β's associated with each covariate. Because individual patients have different participation periods (also referred to as exposure), we included an offset variable $\log(m_i)$. For any Poisson-family models, including an offset variable is a standard way for accommodating the difference in lengths of observation period (Agresti 2001).

In addition to the fixed effects, the mean model includes random effects u_i and z_k. The patient-level random effect u_i is assumed to follow an identical normal distribution $N(0, \sigma_u^2)$ where σ_u^2 quantifies the patient-to-patient variability. Due to the severe heterogeneity across the patients, having patient-level random effects is crucial for modeling the SAM data. The grid-level random effects $\mathbf{z} = [z_1 \ldots z_{168}]^T$ follow a multivariate normal distribution as $z \sim N(\mathbf{0}, \tau(\boldsymbol{I} - \rho\boldsymbol{D})^{-1})$. The variance among grids is denoted by τ and the corresponding correlation is quantified by ρ. The matrix $\mathbf{I}$ is an identity matrix and $\mathbf{0}$ is a column vector with appropriate size (in our case, 168). We assume that the correlation exists among some grids, and the 168 × 168 matrix $\mathbf{D}$ determines which grids should be correlated to each other. The matrix $\mathbf{D}$ is defined as $\mathbf{D} = [d_{pq}]$ where $d_{pq} = 1$ if $p \neq q$ and $p, q \in \boldsymbol{\Psi}_j$.



ID	Age	Date & Time
1	24	11/16/2013 0:17
1	24	11/16/2013 8:40
1	24	11/16/2013 12:12
1	24	11/16/2013 16:33
⋮		
59	45	3/5/2014 2:59
59	45	3/11/2014 9:39

Figure 3. Illustration of the Data Transformation to Grid Format

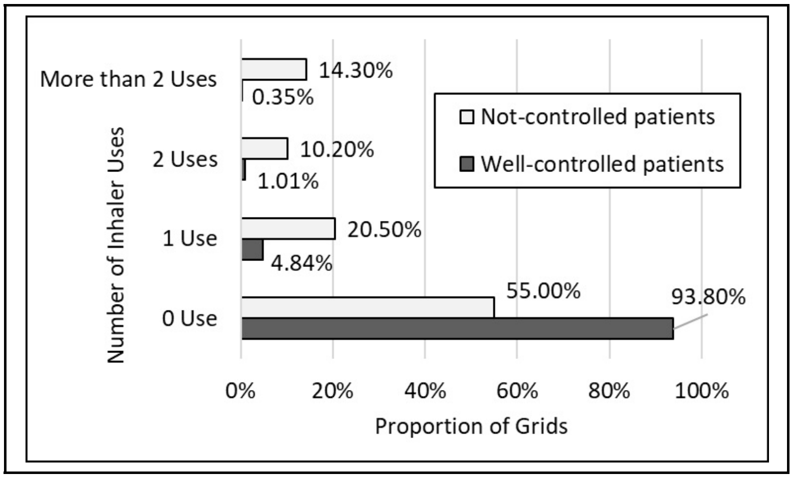


Figure 4. Grid Distribution Based on the Number of Rescue Inhaler Uses

The set of correlated grids is defined as $\Psi_j^T = [j + 24(d-2)]\Big|_{s=1}^{s=7}$ for $j = 1, \ldots, 24$; hence, each column vector Ψ_j contains seven grids. For instance, we have $\Psi_1 = [1\ 25\ 49\ 73\ 92\ 121\ 145]^T$ for $j = 1$ and the grid indices in Ψ_1 respectively represent 12:01 a.m. – 1:00 a.m. time windows from Monday to Sunday. Therefore, we assign $d_{pq} = 1$ to 42 elements ($7 \times 7 - 7 = 42$) in **D**. Similarly, for $j = 2$, we obtain $\Psi_2 = [2\ 26\ 50\ 74\ 98\ 122\ 145]^T$ where its elements represent 1:01 a.m. – 2:00 a.m. time windows from Monday to Sunday. Eventually, among 28,224 ($168^2 = 28{,}224$) elements in **D**, 1,008 elements ($42 \times 24 = 1{,}008$) would be 1. In other words, we assume that the grids in the same one-hour windows are correlated and the conditional correlation between z_k and $z_{k'}$, for $k, k' \in \Psi_j$ is ρ, given the same for all other grids. This formulation produces special dependence in the covariance as a function of the matrix **D** and a fixed unknown correlation parameter ρ (Wall 2004).

We would like to note that our assumption on **D** has been validated in the asthma literature (Milgrom et al. 1996). Furthermore, we found some supporting evidence in the SAM data as well. Figure 5 shows a few example grids of four different patients. The highlighted cells are the one-hour time windows with three or more inhaler uses. As shown in Figure 5, rescue inhaler usage patterns are positively correlated within the same one-hour window across different days of the week.

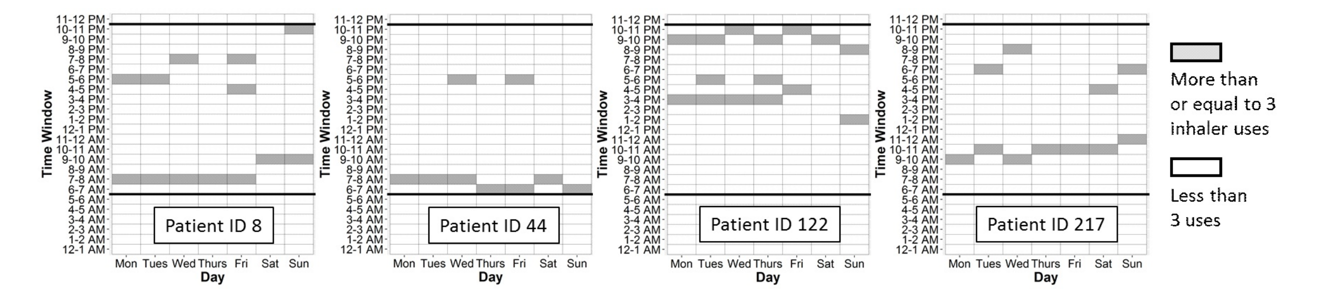


Figure 5. Examples of Correlated Rescue Inhaler Uses Within the Same One-Hour Windows

Dispersion Model: For a conventional quasi-Poisson regression model, a dispersion parameter θ is directly estimated from the data. However, the GLMM-GQP defines a submodel for the dispersion parameter θ_{ik}. We assumed that the dispersion parameter can be modeled by several key environmental factors that are known to be associated with asthma as

$$\log(\theta_{ik}) = \gamma_0 + \gamma_1 Temp_{ik} + \gamma_2 PM25_{ik} \\ \gamma_3 Humidity_{ik} + \gamma_4 Visibility_{ik} + \gamma_5 Windspeed_{ik} \quad (2)$$

The dispersion model includes temperature, humidity, wind speed, visibility, and air pollutant level measured by $PM_{2.5}$. A dedicated model for the dispersion parameter in the GLMM-GQP defines a structure on variability (e.g., the lower temperature increases the variability of the inhaler usage). In asthma literature, the associations between asthma symptoms and the environmental factors in (2) have been already investigated (Knox 1993; Vernon et al. 2012). However, the relationships between the environmental factors and asthma symptoms have been studied mainly in epidemiological settings where data were aggregated to a population level so that the individual patient variability could be suppressed or controlled (Su et al. 2017). When the unit of analysis is at the individual patient level, the impact of the same environmental factor may significantly vary among patients because of the difference in the levels of sensitivity to a specific environmental asthma trigger across patients (Janssens and Ritz 2013). Therefore, for an analysis focusing on personalized asthma management at the individual patient level, it is suggested that the environmental factors are more likely to have a substantial impact on the variability than on the intensity of rescue inhaler usage (Su et al. 2017). Following the literature, we included environmental factors in the dispersion model rather than in the mean model.

The proposed GLMM-GQP uses a quasi-Poisson distribution. In the literature, researchers have often used not only a quasi-Poisson distribution but also a negative binomial distribution for addressing the over-dispersion issue of count data (Ver Hoef and Boveng 2007). The reason we take quasi-Poisson over negative binomial is for the convenience in modeling the dispersion parameter. The idea of separately modeling the dispersion is borrowed from the DGLM. Existing studies have primarily used the DGLM with the quasi-Poisson distribution due to the simple linear relationship between the mean and dispersion parameter of the quasi-Poisson distribution (Smyth and Jorgensen 2002; Smyth and Verbyla 1999). Based on that reason, we chose the quasi-Poisson distribution over negative binomial distribution.

Inference Results: The overall inference follows the guideline of the two-step procedure based on the h-likelihood (Lee and Nelder 1996; Lee et al. 2006). Because a quasi-Poisson distribution does not have a probability function, we used the iteratively weighted least-square algorithm based on the double extended quasi-likelihood (Lee and Nelder 2001) to estimate the parameters in the mean model. Then, the adjusted profile likelihood is used to estimate the parameters in the dispersion model (Ronnegard et al. 2010). To estimate parameters for the grid-level random effects, we used the Eigen decomposition algorithm on the precision matrix (Alam et al. 2015). We did the inference in R (R Core Team 2016) with the *hglm* package (Ronnegard et al. 2010) and *lme4* package (Bates et al. 2015). The estimation results are summarized in Table 3.

In the mean model, sex and ACT-based asthma control level have shown statistical significance ($p = 0.034$ and $p < 0.001$, respectively). The inference results on the dispersion model agree with the clinical intuitions in the asthma literature. The variability of rescue inhaler uses increases as temperature decreases. Humidity, $PM_{2.5}$, and wind speed contribute to increased variability. We observed a noticeable patient-to-patient variability ($\hat{\sigma}_u^2 = 2.284$). The grid-to-grid variance τ is estimated as 0.05 and the conditional correlation coefficient ρ is estimated as 0.15.

Table 3. Inference Results

Mean Model			Dispersion Model		
Parameter	**Estimate**	**SE**	**Parameter**	**Estimate**	**SE**
Intercept (β_0)	-5.8372	0.4112	Intercept (γ_0)	-7.6649	0.6985
Sex=Male (β_1)	0.3952	0.1864	Temperature (γ_1)	-0.0051	0.0011
Age (β_2)	-0.0100	0.0067	$PM_{2.5}$ (γ_2)	0.2811	0.0083
Race=Other (β_3)	-0.1639	0.3124	Humidity (γ_3)	0.0185	0.0029
Race=White (β_4)	0.0632	0.2310	Visibility (γ_4)	0.3158	0.0522
Not Controlled Asthma (β_5)	1.1192	0.2499	Wind Speed (γ_5)	0.1478	0.0117
Weekend (β_6)	0.0038	0.0375			

Step 3: Out-of-Pattern Inhaler Usage Detection

The goal of analyzing the SAM data in our study was to detect out-of-pattern (unusual) inhaler usage for individual asthma patients. Because of the strong association between current asthma control level and future risk of exacerbation (Bateman et al. 2010), our data analytics framework can help with identifying potential exacerbations in the future. The out-of-pattern inhaler usage detection method is based on the individualized probability grid that can be established by the GLMM-GQP. The individualized probability grid is a personal inhaler usage prediction map that shows the likelihood of a patient using his/her rescue inhaler. Each grid k for a patient i has its estimated probability $P(y_{ik}|\hat{\theta}_{ik},\hat{\lambda}_{ik})$ where high probability indicates that, for a specific patient i, there is a high chance of using a rescue inhaler. By referring to this grid, the SAM system can identify if the patient's inhaler usage was out of his/her typical usage pattern. To compute the probability $P(y_{ik}|\hat{\theta}_{ik},\hat{\lambda}_{ik})$, we estimated the patient-level random effects u_i as $\hat{u}_i = \mathrm{E}(u_i|y)$. High and positive values of $\hat{u}_i$ indicate that the patient i tends to use an inhaler more often than expected. Similarly, the grid-level random effects z_k were estimated by $\hat{z}_k = \mathrm{E}(z_k|\mathbf{y})$. The special correlation structure assumed for the distribution of the grid-level random effects should have impact on $\mathrm{E}(z_k|\mathbf{y})$. With all random effects computed for the i^{th} patient and the estimates of the fixed effects $\hat{\beta}$'s, we can obtain $\hat{\lambda}_{ik}$ from equation (1) and the one-week intensity denoted by $\hat{\lambda}^1_{ik}$ is computed as $\log(\hat{\lambda}^1_{ik}) = \log(\hat{\lambda}_{ik}) - \log(m_i)$, subtracting the offset from the original log-intensity estimate (normalizing the difference in the participation periods). A quasi-Poisson distribution does not have a probability distribution function; hence we computed $P(y_{ik}|\hat{\theta}_{ik},\hat{\lambda}^1_{ik})$ approximately from a negative binomial distribution. The approximated $P(y_{ik}|\hat{\theta}_{ik},\hat{\lambda}^1_{ik})$ is denoted by $\tilde{P}(y_{ik}|\hat{\theta}_{ik},\hat{\lambda}^1_{ik})$. A quasi-Poisson random variable has a variance of $\theta\lambda$, while a negative binomial random variable has the variance of $\lambda + \lambda^2/\phi$ where ϕ is the over-dispersion parameter. If we set $\theta = 1 + \lambda/\phi$, then the variance of the negative binomial random variable becomes $\theta\lambda$ because $\lambda + \lambda^2/\phi = \lambda(1 + \lambda/\phi)$. By setting $\phi_{ik} = \hat{\lambda}_{ik}/(\hat{\theta}_{ik}-1)$, we get

$$P(y_{ik}|\hat{\theta}_{ik},\hat{\lambda}_{ik}) \approx \tilde{P}(y_{ik}|\hat{\theta}_{ik},\hat{\lambda}^1_{ik}) = P_{NH}(y_{ik}, |\ \phi_{ik},\ \hat{\lambda}^1_{ik}).$$

We classified the 168 grids into two categories based on the predicted probability of inhaler use: in-pattern and out-of-pattern grids. The in-pattern grids have $\tilde{P}(y_{ik}>0|\hat{\theta}_{ik},\hat{\lambda}^1_{ik}) \geq \delta$ where δ is a decision threshold that can be freely specified by the care provider. We assumed that observing an inhaler use in a grid k for patient i is not alarming if his/her predicted probability of using an inhaler is higher than δ (in-pattern). In contrast, if an inhaler use is observed in one of the out-of-pattern grids, the SAM system should raise a flag to inform both the practitioner and patient so that a closer investigation can be performed. Figure 6 illustrates the overall structure of the data analytics framework for detecting the out-of-pattern rescue inhaler usage.

Performance Evaluation

To demonstrate the satisfactory performance of the proposed analytics framework based on the GLMM-GQP, we conducted a series of performance evaluations. We considered various benchmark methods such as the conventional GLMMs, mixed effects logistic regression model, SVM, and CG-HMM. The CG-HMM is a recently developed statistical model tailored to asthma management (Son et al. 2017). The methods listed above potentially can be used for the out-of-

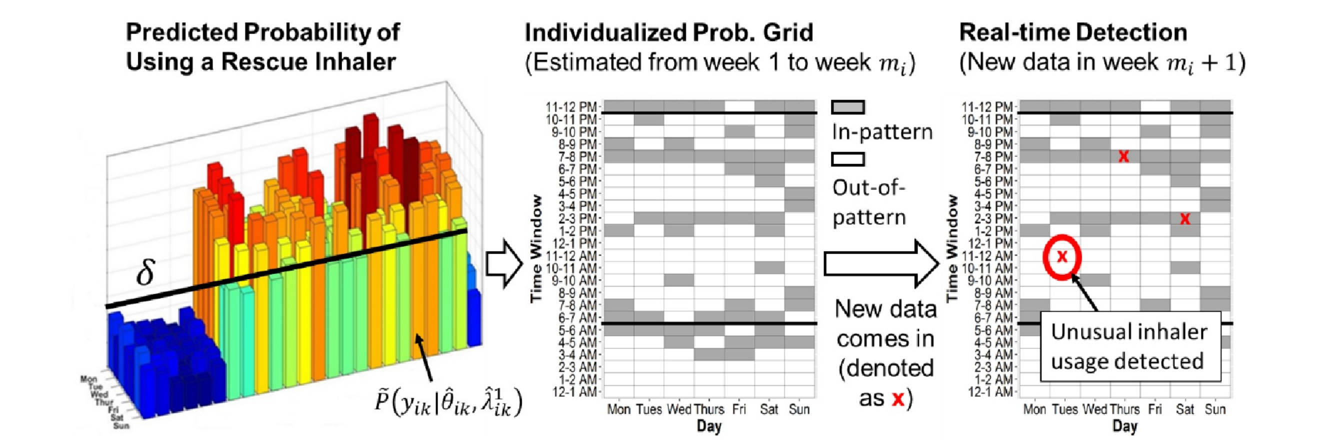

Predicted Probability of Using a Rescue Inhaler
δ
$\tilde{P}(y_{ik}|\hat{\theta}_{ik},\hat{\lambda}^1_{ik})$
Mon
Tue
Wed
Thur
Fri
Sat
Sun
Individualized Prob. Grid
(Estimated from week 1 to week m_i)
Time Window
11-12 PM
10-11 PM
9-10 PM
8-9 PM
7-8 PM
6-7 PM
5-6 PM
4-5 PM
3-4 PM
2-3 PM
1-2 PM
12-1 PM
11-12 AM
10-11 AM
9-10 AM
8-9 AM
7-8 AM
6-7 AM
5-6 AM
4-5 AM
3-4 AM
2-3 AM
1-2 AM
12-1 AM
Mon Tues Wed Thurs Fri Sat Sun
Day
In-pattern
Out-of-pattern
New data comes in (denoted as x)
Real-time Detection
(New data in week $m_i + 1$)
x
Unusual inhaler usage detected

Figure 6. Illustration of the Out-of-Pattern Detection Algorithm

pattern inhaler usage detection instead of the proposed GLMM-GQP. Therefore, in this section, we compare their performance and provide in-depth discussion.

Design of the Evaluation Algorithm and Performance Measure

In design science research on data analytics, one of the popularly used was of contucting a performance evaluation is the hold-out approach (Shmueli and Koppius 2011). This approach excludes a set of observations prior to model fitting and uses the excluded data points to test the performance of the model later. The performance evaluation process is illustrated in Figure 7.

In the performance evaluation algorithm, we considered two scenarios. First, after constructing the individualized probability grid for patient *i*, we replaced the observations in the $(k+1)^{th}$ week of patient *i* with the data from a randomly selected patient *j* ($j \neq i$). Then, we treated the artificially replaced observations as newly collected data of the $(k+1)^{th}$ week for patient *i*. It should be noted that the data used for model fitting only had *k* weeks amount of data from patient *i*. The inhaler usage data is coming from a different patient *j*; hence, the detection method should be able to identify the abnormality in the inhaler usage pattern. This scenario is denoted as true positive (TP) case. If the detection method failed to raise a flag, we record this result as false negative (FN) which is also referred as misdetection.

In the second scenario, we use the data from the i^{th} patient in the $(k+1)^{th}$ week for testing. Because there was no change in the inhaler usage data, the detection method should not generate an alarm. This is a true negative (TN) case. If a wrong alarm was generated, we recorded this result as a false alarm, which is denoted as false positive (FP). We conducted the performance evaluation for various *k* values: $k = 1, 4, 8, 12$, and $m_i - 1$ controlling the amount of data provided to the analytics methods. We excluded the patients who were monitored shorter than 15 weeks from the pool of testing patients to ensure $m_i - 1 \geq 12$. Considering the stochastic nature of the performance evaluation, we repeated the evaluation 100 times for each pair of $\{i, k\}$.

We used the area under the curve (AUC) as our primary performance measure. The curve here represents the receiver operating characteristics (ROC) curve, which shows the relationship between the TP and TN rates. The performance of the out-of-pattern detection methods depends on the decision threshold δ; hence, it is reasonable to use the AUC so that we can see the overall performance of each method considering wide range of possible δ values. The AUC has a range between 0.5 and 1. When the perfect detection is achieved, AUC should have a value of 1, and 0.5 is equivalent to the performance of random guess. The AUC has been widely used for evaluating the performance of analytics methods (Lin et al. 2017) because it considers the trade-off between two types of error: false alarm and misdetection (Bardhan et al. 2014).

Benchmark Methods

In total, we considered six benchmark methods in the performance evaluation. The proposed GLMM-GQP belongs to the GLMM-family. Therefore, we first considered two conventional GLMMs based on Poisson (GLMM-P) and negative

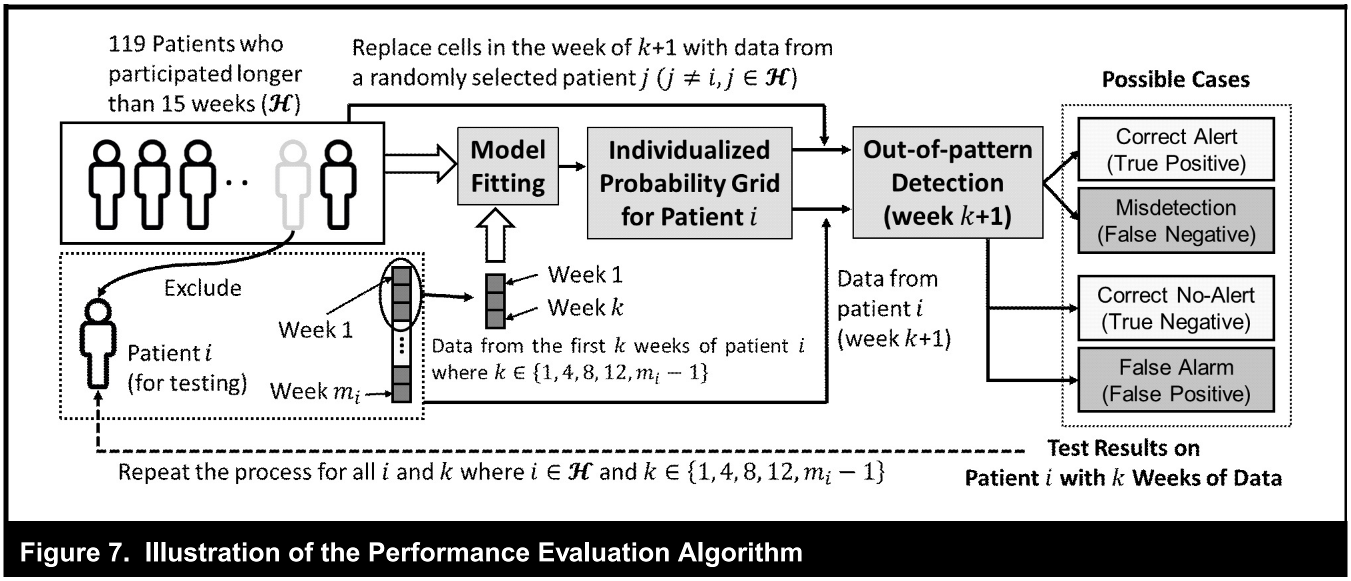

119 Patients who participated longer than 15 weeks ($\mathcal{H}$)
Replace cells in the week of k+1 with data from a randomly selected patient j ($j \neq i, j \in \mathcal{H}$)
Possible Cases
Model Fitting
Individualized Probability Grid for Patient i
Out-of-pattern Detection (week k+1)
Correct Alert (True Positive)
Misdetection (False Negative)
Correct No-Alert (True Negative)
False Alarm (False Positive)
Exclude
Patient i (for testing)
Week 1
Week m_i
Week 1
Week k
Data from the first k weeks of patient i where $k \in \{1, 4, 8, 12, m_i - 1\}$
Data from patient i (week k+1)
Repeat the process for all i and k where $i \in \mathcal{H}$ and $k \in \{1, 4, 8, 12, m_i - 1\}$
Test Results on Patient i with k Weeks of Data

Figure 7. Illustration of the Performance Evaluation Algorithm

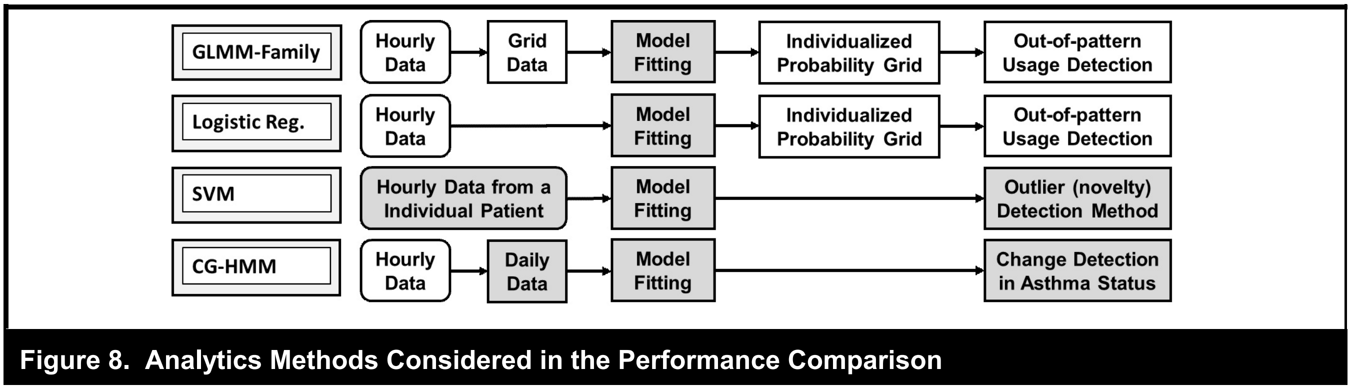

GLMM-Family
Hourly Data
Grid Data
Model Fitting
Individualized Probability Grid
Out-of-pattern Usage Detection
Logistic Reg.
Hourly Data
Model Fitting
Individualized Probability Grid
Out-of-pattern Usage Detection
SVM
Hourly Data from a Individual Patient
Model Fitting
Outlier (novelty) Detection Method
CG-HMM
Hourly Data
Daily Data
Model Fitting
Change Detection in Asthma Status

Figure 8. Analytics Methods Considered in the Performance Comparison

binomial (GLMM-NB) distributions. Also, we defined a simplified GLMM-GQP (GLMM-GQPS) without assuming the CAR model for specifying the special correlation structure among grids. The GLMM-GQPS can be fitted to the data easier than the original GLMM-GQP because the grid-level random effects are simply assumed to follow the typical Gaussian distribution. We denoted those four methods jointly as GLMM-family. Besides the GLMM-family methods, there were other models that needed to be considered. In the out-of-pattern detection analytics framework, the mixed effects logistic regression model can replace the GLMM-family models. We can obtain the individualized probability grid based on the fitted logistic regression model following the same procedure described earlier. One key difference is that we used the hourly usage data, not the transformed data, to ensure a binary response (use/no-use). For the hourly usage data, we can use the support vector machine (SVM) as well. We can train the SVM based on the hourly usage data of a specific individual patient and perform the out-of-pattern usage detection. Because the observations in the SAM data are not labeled, we used the SVM-based outlier (novelty) detection method (Bennett and Campbell 2000). Finally, we included the CG-HMM which is a specifically designed model for monitoring and assessing asthma control (Son et al. 2017). The CG-HMM uses the number of inhaler usage data collected on a daily basis. Therefore, the SAM data need to be converted into daily data (number of inhaler uses per day) prior to model fitting. Then, the fitted CG-HMM can detect the degradation of the asthma control level. Figure 8 summarizes the differences among the seven methods included in our comparative performance evaluation. All methods were implemented R (R Core Team 2016). The *lme4* package (Bates et al. 2015) was used for the GLMM-P, GLMM-NB, and mixed effects logistic regression model. The SVM-based outlier detection algorithm was implemented in *libsvm* package (Chang and Lin 2011).

Table 4. AUC Comparison

	Amount of Data from a Testing Patient Used for Model Fitting (*k*)				
Models	**1 Week**	**4 Weeks**	**8 Weeks**	**12 Weeks**	**Almost All**
GLMM-GQP	0.5754	0.8106	0.8232	0.8517	0.8543
GLMM- GQPS	0.5897	0.7807	0.7949	0.8190	0.8257
GLMM-P	0.5397	0.6387	0.6450	0.6600	0.6734
GLMM-NB	0.5758	0.6930	0.6969	0.7260	0.7409
Logistic Reg.	0.6088	0.6089	0.6099	0.6195	0.6298
SVM	0.5554	0.5664	0.6645	0.6650	0.6736
CG-HMM	0.5000	0.5000	0.5006	0.5039	0.5111

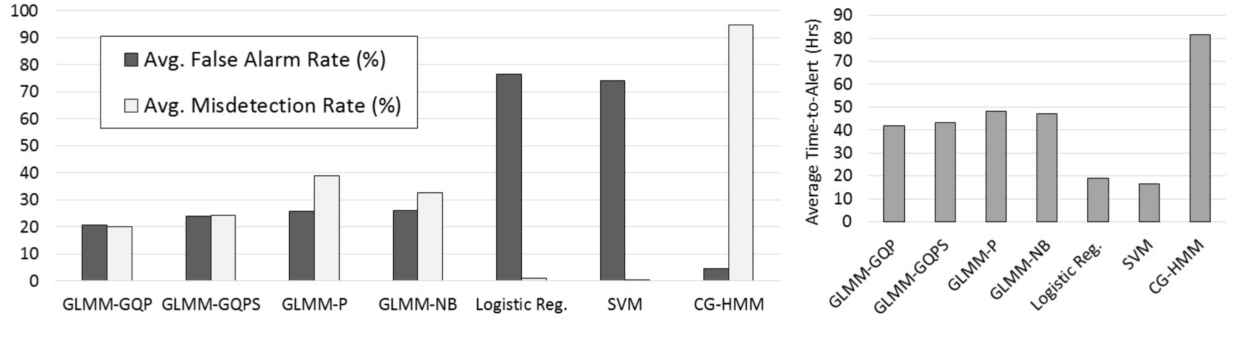
Performance Evaluation Results

Table 4 summarizes the AUCs for the proposed GLMM-GQP and benchmark methods. The performance of the out-of-pattern inhaler usage detection heavily depends on the accuracy of inference (model fitting); hence, we used five different *k* values controlling the amount of data provided to the models. When we provide more data collected from the testing patient, the AUC increases for all seven methods. The right-most column in Table 4 shows the AUCs when we provide every data point of the testing patient *i* except for the last week (m_i) and the GLMM-GQP shows the best performance. We can see that the GLMM-GQP outperforms its simplified version (GLMM-GQPS), and this result suggests that considering the special correlation structure among grids can contribute to the performance improvement.

In general, the GLMM-GQP shows the better performance than the benchmark methods except for the extreme case (k=1). Suppose a new patient is enrolled to the SAM program and he/she has been monitored for only a week (k=1). In that case, the individualized probability grid estimated by the GLMM-GQP for the new patient may not be accurate. However, this issue exists for not only the GLMM-GQP but also for every data-driven method. In Table 4, all methods show unsatisfactory detection performance with a week's worth of data. The GLMM-GQP performs worse than the GLMM-GQPS, GLMM-NB, and logistic regression. This intuitively makes sense because the model structure of the GLMM-GQP is more complex than most of the benchmark methods. Having a flexible model structure often demands more data. However, after acquiring about a month amount of data (k=4), the performance of the GLMM-GQP becomes acceptable. The GLMM-family methods have shown a consistent trend throughout the analysis. The GLMM-GQPS performs slightly worse than the GLMM-GQP but better than the conventional GLMMs (GLMM-P and GLMM-NB). The GLMM-NB outperforms the GLMM-P because, unlike the GLMM-P, the GLMM-NB can address the over-dispersion issue of the inhaler usage data.

The logistic regression model shows AUC around 0.6 with k=1 and the AUC minimally increases even if we provide more data. The SVM achieves AUC comparable to the one for GLMM-P if we have sufficient data of the testing patient. It should be noted that the both logistic regression and SVM were fitted on the hourly usage data without converting them into a grid format. When we do not transform the data into a grid format, the data contain excessive cases of zero inhaler use. Because the inhaler use does not occur frequently, both the logistic regression and SVM become very sensitive to every single inhaler use. In that sense, both logistic regression and SVM are similar to the GLMM-P, which has limited capability of modeling rare events. Thus, observing comparable AUCs of the logistic regression, SVM, and GLMM-P is understandable. The CG-HMM performed the worst in our performance evaluation. The CG-HMM is designed to detect the degradation in asthma control level based on the number of inhaler uses per day. The CG-HMM is a powerful tool for tracing the increasing/decreasing trend in number of rescue inhaler uses (Son et al. 2017). However, it may not be able to detect the unusual inhaler usage which deviates from the normal usage pattern if the number of inhaler uses per day has not shown significant change across time.

The performance of the benchmark methods can be further investigated by their average false alarm and misdetection rates in Figure 9(*a*). As expected, the CG-HMM shows very high misdetection rate because it cannot detect anything unless there was a significant increase in the number of inhaler uses per day. On the other hand, the logistic regression and SVM yield very low misdetection rate but high false alarm rate due to their high sensitivity. The GLMM-family models perform comparably in terms of false alarm rate but the conventional GLMMs (GLMM-NB and GLMM-P) show higher misdetection rate compared to the GLMM-GQP. We



(*a*) Average False Alarm and Misdetection Rates

(*b*) Average Time-To-Alert

Figure 9. Model Comparison Based on False Alarm and Misdetection

can strengthen our findings by comparing the average time-to-alert shown in Figure 9(*b*). The ideal time-to-alert should be 1 hour because it is computed for the cases where the detection method must identify unusual inhaler usage. The logistic regression and SVM have short time-to-alert whereas the CG-HMM rarely raises a flag. In terms of time-to-alert, the SVM and logistic regression seem to be plausible but, as discussed earlier, their excessive sensitivity causes too many false alarms. The GLMM-P and GLMM-NB have slower time-to-alert (about 49 hours) than the GLMM-GQP (about 41 hours). This suggests that the conventional GLMMs are less sensitive and exhibit more misdetections than the GLMM-GQP as shown in Figure 9(*a*).

Discussion and Conclusion

Summary of Contribution

Our study opens a new research direction in data-driven asthma management. The innovations in HIS are fueling a paradigm shift from traditional trial-based healthcare studies to real-time data-driven healthcare analytics. This innovation can potentially transform the current practice of chronic disease management into smart and connected chronic care. The contributions of our study are three-fold. First, for asthma management, our study proposes a new approach that complements conventional guideline-based asthma self-management. Despite the great potential of the SAM system, researchers so far have focused mainly on examining the relationship between the SAM system and improvement in clinical outcomes. The data analytics framework developed in this study shows a way to utilize the advanced HIS for transforming the reactive asthma care practice into data-driven and patient-centered asthma management. From the methodological perspective, we developed the GLMM-GQP, which integrates several statistical models into a unified structure. The GLMM-GQP provides a dedicated sub-model for the dispersion parameter so that the heterogeneous impact of environmental asthma triggers on inhaler usage behavior can be properly modeled. Also, the flexible correlation structure for the random effects allows the model to reflect the repetitive daily routine of individual patients. Furthermore, our analytics framework performs a rigorous model-based abnormality detection and we have shown its promising performance through the comparison study with several benchmark methods. Finally, our study also makes contributions to the IS knowledge base. The SAM platform introduced in this paper further extends the definition of HIS. It provides a good example of how HIS can innovate the current practice of chronic disease management. We also emphasized the importance of developing an IT solution artifact tailored to the new type of HIS considering the disease-specific characteristics. In our study, the data analytics methods that have limited capability for addressing key features of asthma management have shown worse performance than the proposed method specifically designed for the SAM platform. Overall, our study is a good example of exaptation design science research. We proposed a design principle that leads to a solution for a new problem by integrating and improving prior design knowledge. Existing analytics methods have been significantly extended to address the new challenges that were not present in the traditional quantitative asthma studies based on clinical trials.

Our study provides meaningful practical implications as well. For care providers, the IT artifact developed in our study can help with optimizing their resource utilization. Medical experts are a highly trained and expensive resource. Their time should not be wasted by monitoring patients' rescue inhaler usage manually. If data analytics methods could automatically extract the most critical signs from the remote patient monitoring data, healthcare professionals could make a personalized therapeutic decision with less time and effort. For asthma patients, our analytics method serves as an asthma self-management tool. Providing an accessible HIS to patients is crucial but not sufficient. The potential of the HIS in terms of improving the quality of life of patients can be fully realized only if the HIS is coupled with a well-established analytics method. Therefore, we believe that our study takes a critical step closer to realizing how HIS and analytics can collectively contribute to achieve better healthcare practice.

Guidelines for IS Researchers and Practitioners

The proposed analytics framework based on the GLMM-GQP can detect an event that is out of its usual pattern. We developed the method in the context of asthma management HIS, but the application is not limited only to asthma care. The out-of-pattern detection method is applicable to any healthcare applications if a few requirements, listed below, are satisfied:

- A reliable data colleting/transmitting device that provides accurate event times
- Additional biomarkers (covariates) associated with the health-related event of interest
- Sufficient amount of data collected from the specific target patient
- Patient-specific behavioral knowledge that is potentially related to the event patterns

First, the sensors or any type of data collecting device implemented in the monitoring system should be able to provide accurate event times and should guarantee no information loss during data transactions. In other words, the system should have access to the detailed information about every single event rather than summary data such as average blood pressure for each day, because summary data may not be sufficient for characterizing the patterns of event occurrence. Second, the data should contain meaningful biomarkers that are associated with the clinical event of interest. In this study, we considered several demographic and environmental factors because they are closely related to the rescue inhaler usage. Third, in the spirit of personalized medicine, it is recommended to have a sufficient amount of data from the target patient. The progression of a medical condition over time varies across patients. Thus, to provide personalized intervention, practitioners should use the analytics method after accumulating sufficient data collected from the specific target patient. Based on our findings, most of the data-driven methods should perform reasonably well if the patient has been monitored at least a month. Finally, it is important to understand the potential correlation structure among events. Following the asthma literature, we assumed that the same one-hour windows have a positive correlation. However, this assumption may not be valid for other healthcare applications. Thus, the correlation structure needs to be specified adequately for managing the specific medical condition. As shown in this study, considering the patient-specific information has a significant potential to improve the performance of the detection method.

The GLMM-GQP can be directly converted into a working IT artifact. However, we would like to note that there are several aspects that need to be studied before implementing the IT artifact in clinical practice. For instance, providing alerts to patients or practitioners needs to be done cautiously. On one hand, an excessive number of alerts triggers alert fatigue. On the other hand, data-driven healthcare diagnostic tools that show a high misdetection rate may encounter a legal liability issue. Also, it is very challenging to foresee how patients might respond to alerts generated by the HIS because alert compliance depends on various factors. In addition, at the moment, it is not clear how to share the cost of the HIS optimally among the beneficiaries (e.g., hospitals, patients, and insurers). Addressing the issues listed above is crucial for ensuring successful implementation of our solution artifact in clinical practice.

Limitations and Future Research

The major limitation of our study comes from the nature of our test bed. The SAM data contains many useful variables including demographics of the participants and environmental factors. However, because asthma is a complex respiratory chronic condition, the list may not be comprehensive enough. One way to address this issue is to integrate the data collected by the SAM system with EHRs. In this way, we obtain full access to the detailed medical records of individual patients in addition to the continuously monitored rescue inhaler usage logs. Currently, integration between the SAM data and EHRs is under development for future research. Another limitation is that the data set we used in our study was collected from a certain geographical region. As for all data-driven research, our study heavily depends on the specific data set that we have collected from our research test bed. Therefore, we plan to investigate a data set collected from a different geographical region with a different set of participants to further validate the effectiveness of the analytics method proposed in this paper. In addition, we plan to extend our study in the future by including patients diagnosed with chronic obstructive pulmonary disease (COPD). COPD is another prevalent respiratory chronic disease and COPD patients also use a personal inhaler to administer a dose of their rescue medication. Therefore, we believe the solution artifact developed in the current study has great potential in COPD management.

Acknowledgments

We thank the editors and anonymous reviewers for their valuable comments. Also, the comments received at the 2018 *MIS Quarterly* Workshop (Naveen Jindal School of Management, University of Texas-Dallas) were very helpful in improving the quality of the paper. This research would have been very difficult without valuable feedback from Dr. Meredith Barrett at Propeller Health, Dr. Guangquan (Jason) Su at University of California–Berkeley (Division of Environmental Health Sciences, School of Public Health), and Dr. Yeongin Kim at Virginia Commonwealth University (Department of Information Systems, School of Business). Finally, we thank Sunny Rosen, Communications Specialist at Alfred Lerner College of Business & Economics (University of Delaware), for editing the manuscript. This research is partially supported by the University of Delaware General University Research (GUR) grant and National Science Foundation (NSF) grant #1343969.

References

Agarwal, R., Gao, G., DesRoches, C., and Jha, A. K. 2010. "The Digital Transformation of Healthcare: Current Status and the Road Ahead," *Information Systems Research* (21:4), pp. 796-809.

Agresti, A. 2001. *Categorical Data Analysis*, New York: Wiley.

Ahmed, S., Ernst, P., Bartlett, S. J., Valois, M-F., Zaihra, T., Pare, G., Grad, R., Eilayyan, O., Perreault, R., and Tamblym, R. 2016. "The Effectiveness of Web-Based Asthma Self-Management System, My Asthma Portal (MAP): A Pilot Randomized Controlled Trial," *Journal of Medical Internet Research* (18:12), pp. e313.

Alam, M., Rönnegård, L., and Shen, X. 2015. "Fitting Conditional and Simultaneous Autoregressive Spatial Models in hglm," *The R Journal* (7:2), pp. 5-18.

Anderson, S. D., Caillaud, C., and Brennan, J. D. 2006. "Agonists and Exercise-Induced Asthma," *Clinical Reviews in Allergy & Immunology* (31:2-3), pp. 163-180.

Backonja, U., Kim, K., Casper, G. R., Patton, T., Ramly, E., and Brennan, P. F. 2012. "Observations of Daily Living: Putting the 'Personal' in Personal Health Records," in *Proceedings of the 11th International Congress on Nursing Informatics*, June 23-27, 2012, Montreal, Canada.

Bardhan, I., Oh, J. C., Zheng, Z. E., and Kirksey, K. 2015. "Predictive Analytics for Readmission of Patients with Congestive Heart Failure," *Information Systems Research* (26:1), pp. 19-39.

Barrett, M., Combs, V., Su, J. G., Henderson, K., Tuffli, M., and The AIR Louisville Collaborative. 2017. "AIR Louisville: Addressing Asthma with Technology, Crowdsourcing, Cross-Sector, and Policy," *Health Affairs* (37:4), pp. 525-534.

Bateman, E. D., Reddel, H. K., Eriksson, G., Peterson, S., Ostlund, R., Sears, M. R., Jenkins, C., Humbert, M., Buhl, R., Harrison, T. W., Quirce, S., and O'Byrne, P. M. 2010. "Overall Asthma Control: The Relationship between Current Control and Future Risk," *Journal of Allergy and Clinical Immunology* (125:3), pp. 600-608.

Bates, D., Maechler, M., Bolker, B., and Walker, S. 2015. "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software* (67:1), pp. 1-48.

Bender, B., Milgrom, H., Rand, C., and Ackerson, L. 1998. "Psychological Factors Associated with Medication Nonadherence in Asthmatic Children," *Journal of Asthma* (35:4), pp. 347-353.

Bennett, K., and Campbell, C. 2000. "Support Vector Machines: Hype or Hallelujah?," *ACM SIGKDD Explorations* (2:2), pp. 1-13.

Buntin, M. B., Burke, M. F., Hoaglin, M. C., and Blumenthal, D. 2011. "The Benefits of Health Information Technology: A Review of the Recent Literature Shows Predominantly Positive Results," *Health Affairs* (30:3), pp. 464-471.

Carranza, R. J. R., Edwards, L., Lincourt, W., Dorinsky, P., and ZuWallack, R. L. 2004. "The Relationship Between Health-Related Quality of Life, Lung Function and Daily Symptoms in Patients with Persistent Asthma." *Respiratory Medicine* (98:12), pp. 1157-1165.

Casper, G. R., and Brennan, P. F. 2013. "Project HealthDesign: A Preliminary Program-level Report," in *Proceedings of the AMIA Annual Symposium*, American Medical Informatics Association, Washington, DC, pp. 192-199.

CDC. 2016. "Data, Statistics, and Surveillance: Asthma Surveillance Data," National Center for Environmental Health, Center for Disease Control and Prevention (https://www.cdc.gov/asthma/most_recent_data.htm).

Chan, A. H. Y., Harrsion, J., Black, P. N., Mitchell, E. A., and Foster, J. M. 2015. "Using Electronic Monitoring Devices to Measure Inhaler Adherence: A Practical Guide for Clinicians," *Journal of Allergy and Clinical Immunology: In Practice* (3:3), pp. 335-349.

Chang, C-C., and Lin, C-J. 2011. "LIBSVM: A Library for Support Vector Machines," *ACM Transactions on Intelligent Systems and Technology* (2:3), pp. 27:1-27:27.

Clayton, D., and Kaldor, J. 1987. "Empirical Bayes Estimates of Age Standardized Relative Risks for Use in Disease Mapping," *Biometrics* (43), pp. 671-681.

Combescure, C., Chanez, P., Saint-Pierre, P., Daures, J. P., Proudhon, H., and Godard, P. 2003. "Assessment of Variations in Control of Asthma Over Time," *European Respiratory Journal* (22), pp. 298-304.

Cressie, N. A. C. 1993. *Statistics for Spatial Data* (Rev. Ed.), Chichester, UK: Wiley.

Fichman, R. G., Kohli, R., and Krishnan, R. 2011. "The Role of Information Systems in Healthcare: Current Research and Future Trends," *Information Systems Research* (22:3), pp. 419-428.

Gregor, S., and Hevner, A. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355.

Huckvale, K., Morrison, C., Ouyang, J., Ghaghda, A., and Car, J. 2015. "The Evolution of Mobile Apps for Asthma: An Updated Systematic Assessment of Content and Tools," *BMC Medicine* (13:58).

Janssens, T., and Ritz, T. 2013. "Perceived Triggers of Asthma: Key to Symptom Perception and Managemet," *Clinical & Experimental Allergy* (43:9), pp. 1000-1008.

Janssens, T., Verleden, G., Peuter, S. D., Diest, I. V., and Bergh O. V. 2009. "Inaccurate Perception of Asthma Symptoms: A Cognitive–Affective Framework and Implications for Asthma Treatment," *Clinical Psychology Review* (29), pp. 317-327.

Kallinikos, J., and Tempini, N. 2014. "Patient Data as Medical Facts: Social Media Practices as a Foundation for Medical Knowledge Creation," *Information Systems Research* (25:4), pp. 817-833.

Kane, G. C., and Labianca, G. 2011. "IS Avoidance in Health-Care Groups: A Multilevel Investigation," *Information Systems Research* (22:3), pp. 504-522.

Kim, M. S., Henderson, K. A., and Van Sickle, D. 2016. "Using Connected Devices to Monitor Inhaler Use in the Real World," *Respiratory Drug Delivery* (1), pp. 37-44.

Knox, R. B. 1993. "Grass Pollen, Thunderstorms, and Asthma," *Clinical & Experimental Allergy* (23:5), pp. 354-359.

Kohli, R., and Tan, S. 2016. "Electronic Health Records: How Can IS Researchers Contribute to Transforming Healthcare?," *MIS Quarterly* (40:3), pp. 553-573.

Kolodner, R. M., Cohn, S. P., and Friedman, C. P. 2008. "Health Information Technology: Strategic Initiatives, Real Progress," *Health Affairs* (27:Supplement), pp. w391-w395.

Lee, Y., and Nelder, J. A. 1996. "Hierarchical Generalized Linear Models (with Discussions)," *Journal of the Royal Statistical Society, Series B* (58), pp. 619-678.

Lee, Y., and Nelder, J. A. 2001. "Hierarchical Generalized Linear Models: A Synthesis of Generalized Linear Models, Random Effect Models, and Structured Dispersions," *Biometrika* (88), pp. 987-1006.

Lee, Y., Nelder, J., and Pawitan, Y. 2006. *Generalized Linear Models with Random Effects: Unified Analysis via H-Likelihood*. Boca Raton, FL: Chapman & Hall.

Leroy, G., Chen, H., and Rindflesch, T. C. 2014. "Smart and Connected Health," *IEEE Intelligent Systems* (29:3), pp. 2-5.

Licskai, C. J., Sands, T. W., and Ferrone, M. 2013. "Development and Pilot Testing of a Mobile Health Solution for Asthma Self-management: Asthma Action Plan Smartphone Application Pilot Study," C*anadian Respiratory Journal* (20:4), pp. 301-306.

Lin, Y-K., Chen, H., Brown, R. A., Li, S-H., and Yang, H-J. 2017. "Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multitask Learning Approach," *MIS Quarterly* (41:2), pp. 473-495.

Merchant, R. K., Inamdar, R., and Quade, R. C. 2016. "Effectiveness of Population Health Management Using the Propeller Health Asthma Platform: A Randomized Clinical Trial," *Journal of Allergy and Clinical Immunology: In Practice* (4:3), pp. 455-463.

Milgrom, H., Bender, B., Ackerson, L., Bowry, P., Smith, B., and Rand, C. 1996. "Noncompliance and Treatment Failure in Children with Asthma," *Journal of Allergy and Clinical Immunology* (98:6), pp. 1051-1057.

Nathan, R. A., Sorkness, C. A., Kosinski, M., Schatz, M., Li, J. T., Marcus, P., Murray, J. J., and Pendergraft, T. B. 2004. "Development of the Asthma Control Test: A Survey for Assessing Asthma Control," *Journal of Allergy and Clinical Immunology* (113:1), pp. 59-65.

NHLBI. 2007. *National Asthma Education and Prevention Program, Third Expert Panel on the Diagnosis and Management of Asthma. Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma*, National Heart, Lung, and Blood Institute, Bethesda, MD (http://www.ncbi.nlm.nih.gov/books/NBK7232/).

Nurmagambetov, T., Kuwahara, R., and Garbe, P. 2017. "The Economic Burden of Asthma in the United States, 2008–2013," *Annals of the American Thoracic Society* (15:3), pp. 348-356.

Patel, M., Pilcher, J., Reddel, H. K., Qi, V., Mackey, B., Tranquilino, T., Shaw, D., Black, P., Weatherall, M., and Beasley, R. 2014. "Predictors of Severe Exacerbations, Poor Asthma Control, and -Agonist Overuse for Patients with Asthma," *Journal of Allergy and Clinical Immunology: In Practice* (2:6), pp. 751-758.e1.

Pinheiro, J., and Bates, D. 2000. *Mixed-Effects Models in S and S-PLUS*, New York: Springer.

Postma, D. S. 2007. "Gender Differences in Asthma Development and Progression," *Gender Medicine* (4:Supplement 2), pp. S133-S146.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing.

Reddel, H. K., Taylor, D. R., Bateman, E. D., Boulet, L. P., Boushey, H. A., Busse, W. W., Casale, T. B., Chanez, P., Enright, P. L., Gibson, P. G., de Jongste, J. C., Kerstjens, H. A., Lazarus, S. C., Levy, M. L., O'Byrne, P. M., Partridge, M. R., Pavord, I. D., Sears, M. R., Sterk, P. J., Stoloff, S. W., Sullivan, S. D., Szefler, S. J., Thomas, M. D., and Wenzel, S. E. 2009. "An Official American Thoracic Society/European Respiratory Society Statement: Asthma Control and Exacerbations: Standardizing Endpoints for Clinical Asthma Trials and Clinical Practice," *American Journal of Respiratory and Critical Care Medicine* (180:1), pp. 59-99.

Ronnegard, L., Shen, X., and Alam, M. 2010. "hglm: A Package for Fitting Hierarchical Generalized Linear Models," *The R Journal* (2:2), pp. 20-28.

Saint-Pierre, P., Bourdin, A., Chanez, P., Daures, J., and Godard, P. 2006. "Are Overweight Asthmatics More Difficult to Control?," *Allergy* (61:1), pp. 79-84.

Saint-Pierre, P., Combescure, C., Daurs, J., and Godard, P. 2003. "The Analysis of Asthma Control under a Markov Assumption with Use of Covariates," *Statistics in Medicine* (22:24), pp. 3755-3770.

Schatz, M., Sorkness, C. A., Li, J. T., Marcus, P., Murray, J. J., Nathan, R. A., Kosinski, M., Pendergraft, T. B., and Jhingran, P. 2006. "Asthma Control Test: Reliability, Validity, and Responsiveness in Patients Not Previously Followed by Asthma Specialists," *Journal of Allergy and Clinical Immunology* (117:3), pp. 549-556.

Shmueli, G., and Koppius, O. R. 2011. "Predictive Analytics in Information Systems Research," *MIS Quarterly* (35:3), pp. 553-572.

Smyth, G. K., and Jorgensen, B. 2002. "Fitting Tweedie's Compound Poisson Model to Insurance Claims Data: Dispersion Modelling," *ASTIN Bulletin* (32:1), pp. 143-157.

Smyth, G. K., and Verbyla, A. P. 1999. "Adjusted Likelihood Methods for Modeling Dispersion in Generalized Linear Models," *Environmetrics* (10), pp. 695-709.

Son, J., Brennan, P. F., and Zhou, S. 2016. "Rescue Inhaler Usage Prediction in Smart Asthma Management Systems Using Joint Mixed Effects Logistic Regression Model," *IIE Transactions* (48:4), pp. 333-346.

Son, J., Brennan, P. F., and Zhou S. 2017. "Correlated Gamma-Based Hidden Markov Model for the Smart Asthma Management Based on Rescue Inhaler Usage," *Statistics in Medicine* (36), pp. 1619-1637.

Su, J. G., Barrett, M. A., Henderson, K., Humblet, O., Smith, T., Sublett, J. W., Nesbitt, L., Hogg, C., Van Sickle, D., and Sublett, J. L. 2017. "Feasibility of Deploying Inhaler Sensors to Identify the Impacts of Environmental Triggers and Built Environment Factors on Asthma Short-Acting Bronchodilator Use," *Environmental Health Perspective* (125:2), pp. 254-261.

United States Census Bureau. 2016. *Quick Facts Table: Kentucky* (https://www.census.gov/quickfacts/KY).

Van Sickle, D., Meanner, M. J., Barrett, M. A., and Marcus, J. E. 2013. "Monitoring and Improving Compliance and Asthma Control: Mapping Inhaler Use for Feedback to Patients, Physicians, and Payers," *Respiratory Drug Delivery* (1), pp. 119-130.

Ver Hoef, J. M., and Boveng, P. L. 2007. "Quasi-Poisson vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data," *Ecology* (88:11), pp. 2766-2772.

Vernon, M. K., Wiklund, I., Bell, J. A., Dale, P., and Kennneth, R. 2012. "What Do We Know about Asthma Triggers? A Review of the Literature," *Journal of Asthma* (49:10), pp. 991-998.

Wall, M. M. 2004. "A Close Look at the Spatial Structure Implied by the CAR and SAR Models," *Journal of Statistical Planning and Inference* (121), pp. 311-324.

Wedderburn, R. W. M. 1974. "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika* (61), pp. 439-447.

William, L. K., Peterson, E. L., Well, K., Ahmedani, B. K., Kumar, R., Burchard, E. G., Chowdhry, V. K., Favro, D., Lanfear, D. E., and Pladevall, M. 2011. "Quantifying the Proportion of Severe Asthma Exacerbations Attributable to Inhaled Corticosteroid Nonadherence," *Journal of Allergy and Clinical Immunology* (128:6), pp. 1185-1191.

Wu, A. C., Carpenter, J. F., and Himes, B. E. 2015. "Mobile Health Applications for Asthma," *Journal of Allergy and Clinical Immunology: In Practice* (3:3), pp. 446-448.

Zein, J. G., and Erzurum, S. C. 2016. "Asthma Is Different in Women," *Current Allergy and Asthma Reports* (15)

About the Authors

Junbo Son received his B.S. in Industrial Systems and Information Engineering from the Korea University, South Korea, in 2010. He obtained both an M.S. in Statistics and a Ph.D. in Industrial and Systems Engineering from University of Wisconsin–Madison in 2015 and 2016, respectively. He is currently an assistant professor in the Alfred Lerner College of Business & Economics at the University of Delaware. His research interests include medical informatics for advanced healthcare systems, data-driven reliability engineering, and data analytics for solving complex problems in various business sectors.

Patricia Flatley Brennan, RN, Ph.D., is the director of the National Library of Medicine (NLM), a component of the National Institutes of Health (NIH). NLM is the world's largest biomedical library and the producer of digital information services used by scientists, health professionals, and members of the public worldwide. Before joining the NIH, she was the the Lillian L. Moehlman Bascom Professor at the School of Nursing and College of Engineering, University of Wisconsin–Madison. She received a Master of Science in nursing from the University of Pennsylvania and a Ph.D. in industrial engineering from the University of Wisconsin-Madison.

Shiyu Zhou is a professor in the Department of Industrial and Systems Engineering at the University of Wisconsin–Madison. He received his B.S. degree from the University of Science and Technology of China in 1993, and his master's degree and Ph.D. from the University of Michigan in 2000. His research focuses on industrial analytics and system informatics methodologies for quality and productivity improvement and operation optimization. He has received numerous research awards and grants from various federal agencies. He is currently a Fellow of Institute of Industrial and Systems Engineers, American Society of Mechanical Engineers, and Society of Manufacturing Engineers.