First, Do No Harm: Predictive Analytics to Reduce In-Hospital Adverse Events

Yu-Kai Lin (corresponding author)

yklin@gsu.edu

Center for Digital Innovation & Department of Computer Information Systems, J. Mack Robinson College of Business, Georgia State University, Atlanta, GA 30303, USA

Xiao Fang

xfang@udel.edu

Department of Accounting and Management Information Systems, Lerner College of Business and Economics, Newark DE 19716

ABSTRACT: Inadequate patient safety is a serious issue in current medical practice. Medical errors cause adverse events (AEs) for patients and lead to premature deaths, unintended complications, prolonged hospital stays, and higher medical costs. Although the importance of AE prediction and prevention is well recognized in the information systems literature, there is a dearth of research on modeling and predicting AEs caused by medical errors. Following the design science research paradigm, this study describes the search, design, and evaluation of a novel in-hospital AE prediction model, called Stochastic Autoregressions for Latent Trajectories (SALT). The proposed model uniquely integrates generalized linear mixed model with multitask learning and stochastic time-series processes. Results from our empirical evaluation show that SALT outperforms prior state-of-the-art techniques in predicting AEs during patients' hospital stays. Through a simulation, we further demonstrate significant cost savings potential when hospitals implement and integrate SALT in their inpatient care. This study contributes to the

design science literature by formalizing the in-hospital AE prediction problem, on the one hand, and developing a novel graphical model to address the prediction problem, on the other. For healthcare practitioners and administrators, our predictive analytics approach unveils important insights to minimize AEs.

Keywords: design science; healthcare predictive analytics; patient safety; medical errors; adverse events

> *The physician must be able to tell the antecedents, know the present, and foretell the future—must mediate these things, and have two special objects in view with regard to disease, namely, to do good or to* ***do no harm***. – Hippocratic Oath

Introduction

Patient safety is the foundation of high-quality healthcare. However, many studies have pointed out that current medical practice is not as safe as it should be in the United States. The landmark "To Err Is Human" report from the Institute of Medicine (IOM) suggested that as many as 98,000 patients per year die in the hospital as a result of medical errors [34]. Using medical claims data, Bos et al. [11] estimated that the annual cost of measurable medical errors in the United States in 2008 was $17.1 billion. More recently, Makary and Daniel [41] argued that medical error could be the third leading cause of death in the US. Although the reliability of these estimates has been challenged, most experts agree that inadequate patient safety is a serious problem in the US healthcare system.

This study aims to develop a novel in-hospital adverse event (AE) prediction model to improve patient safety. Although AEs have been discussed or studied in the information systems (IS) literature, prior studies conceptualize AEs in different ways (see Table A1 in Appendix A). These various notions of AEs can be organized in a layered typology as shown in Table 1. At the

broadest level, some IS studies use AEs to describe undesirable business or societal outcomes. For example, Benaroch and Chernobai [9] regard IT failures and security breaches as AEs. A narrower and healthcare-related notion of AEs is to consider them as deteriorating patient outcomes. Based on this view, Lin et al. [40] and Meyer et al. [44] consider the complications in diabetic patients' nature disease progression as AEs and propose methods to predict the risks of AEs (e.g., given the patient's medical history, latest lab results, and treatment regimes, what is his risk of having a stroke in five year?). Finally, the most specific conceptualization of AEs, which is the focus of our study, follows the IOM's definition: an AE is "an injury caused by medical management rather than the underlying condition of the patient" [34:28]. In other words, AEs are deteriorating patient outcomes caused specifically by medical errors. Consistent with this particular conceptualization of AEs, prior IS research has empirically examined how automation or electronic medical records affect medical errors and patient safety events [4, 32].

[INSERT TABLE 1 ABOUT HERE]

Although in the past decade many IS scholars have acknowledged the importance of AE perdition and prevention and the roles of IS research on this topic [6, 20], there is a surprising dearth of work in the IS literature on *predicting* AEs resulted from harmful medical errors. The closest study to ours is by Abbasi et al. [1], who propose a method for early adverse drug event (ADE) warnings by analyzing co-occurrences of drug and effect mentions on social media. Their work concerns about postmarketing drug surveillance and aims to discover previously unknown drug-related AEs occurring after the drug is introduced to the market but before the AEs become widespread. As such, although Abbasi et al. [1] also consider AEs as harms from medical errors as we do, their study addresses AE *detection* rather than AE *prediction*. AE detection happens after an AE occurs and it explains what drug related factors caused the AE; AE prediction happens before an AE occurs and seeks to prevent it from happening.

To design an in-hospital AE prediction model, this study follows closely the guidelines of design science research (DSR) [26, 30]. Specifically, we first outline the research and practical backgrounds for in-hospital AEs in order to establish the problem relevance. Our review of related work sheds light on the critical role of *latent* AE contributory factors, the need to account for them in AE predictive modeling, and the limitations of extant AE modeling approaches. Consistent with Gregor and Hevner [26], our literature review provides "justificatory knowledge (kernel theory) that was used to inform the construction of the new artifact." We then describe the proposed artifact—a graphical model, called Stochastic Autoregressions for Latent Trajectories (SALT), which integrates generalized linear mixed model (GLMM) with multitask learning and stochastic time-series processes. We assess our model using a large real-world inpatient discharge dataset, and through rigorous evaluations, we demonstrate SALT's superior predictive performance compared to the state-of-the-art techniques. Through an in-depth simulation, we further unveil that SALT can prevent more AEs with fewer false alarms than alternative methods. We estimate that implementing SALT would lead to cost savings of $5.85 million per year in our sample alone.

This study contributes to the IS literature, particularly in DSR. Specifically, we respond to the call for IS research on AE prevention by formalizing the prediction problem and proposing an effective modeling framework for AE prediction [6, 20]. Consistent with the problem-solving paradigm of DSR, our contributions can be characterized through Gregor and Hevner's [26] DSR knowledge contribution framework in which we seek to address a relatively new, yet highly important, application domain where the extant solutions have salient limitations. In designing our innovative solution, we draw from recent empirical findings in the patient safety literature. We emphasize the importance of considering latent organizational factors and multiple categories of AEs in AE predictive modeling. These design principles and our modeling

framework offer a "nascent design theory" [26] that is generalizable to other problem domains that involve multiple outcomes, multi-level observations, and latent factors.

In-Hospital Adverse Events

In DSR projects, it is critical to demonstrate the context and the relevance of the problem [26, 30]. To this end, we begin by motivating the problem of in-hospital AEs. This is followed by an overview of key empirical findings regarding the categories and contributory factors of AEs. Overall, our goal in this section is to surface related empirical background and knowledge so as to motivate the problem and justify our design.

Problem Motivation

Consider the following scenario. *A 70-year-old male was experiencing shortness of breath at home and was immediately transferred to a nearby hospital. At the point of admission, the attending physician gathered basic demographic and clinical information from him. She found that the chief complaint of the patient was heart failure, although he was also showing or indicating chronic health conditions such as hypertension. After being admitted into the hospital, the patient was given hypertensive medication. The particular medication caused him severe headache and dizziness; in response, the attending physician revised the prescription with a different class of medication. Upon reviewing the test results with the attending physician, the patient decided to follow her recommendation and have a coronary artery bypass surgery. The surgery itself was successful, but he had a mild infection afterward. He was given additional antibiotics to control the infection. The patient was discharged 3 days after the surgery and was asked to come back to an outpatient clinic for a follow-up visit next month.*

The scenario illustrates two key points. The first point is that the patient in the scenario had two in-hospital AEs: ADE from the hypertensive medication and the postoperative infection. These are harms caused by medical management during his hospital stay rather than the patient's

nature disease progression. Clearly, AEs like these should be prevented. The question is how to prevent such in-hospital AEs from occurring. This is related to the second key point of the scenario. That is, hospital care generally involves four stages of episodes: (1) before admission, (2) at the point of admission, (3) during the hospital stay, and (4) after discharge. From the perspective of clinical decision support, it would be useful to inform the care team about the patient's risk of AEs during his hospital stay (i.e., stage 3) and the best time to do this is perhaps at the point of admission (i.e., stage 2). This is when basic information about the patient becomes available, but the medical management has not yet started. As such, the primary goal of our study is to reduce preventable AEs by utilizing information available at the point of admission to inform the care team so as to mitigate AE risks in the patient's hospital stay. To that end, it is useful to further characterize in-hospital AEs, as discussed below.

Characterizing In-Hospital Adverse Events

Studies have collectively suggested that the prevalence of AEs may range from 2.9% to 16.6% of all hospital admissions [12, 45]. Meanwhile, researchers have also found that about half of all AEs are preventable [55]. In order to better understand AEs, prior research has examined their categories and their contributory factors.

The patient safety literature has categorized AEs in different ways. Earlier studies tended to consider a dichotomous classification, categorizing medical errors as either operative or non-operative. More recently, some researchers have been interested in defining a more holistic and finer-grained AE classification scheme beyond the operative/non-operative dichotomy. For instance, the Utah–Missouri Adverse Event Classification (UMAEC) categorizes AEs into six groups: ADEs, surgical events, misadventures, device events, infections, and other AEs [31]. These fine-grained AE categories are useful because they offer better insights regarding the nature of the patient safety incidents, which in turn can help care teams identify more promising

ways to intercept and prevent potential AEs in specific circumstances.

Beyond the categories of AEs, recent studies have also analyzed factors that contribute to in-hospital AEs. Lawton et al. [38], for example, identify 20 contributory factor domains and organize these into five categories: *active failures* (e.g., violations), *situational factors* (e.g., task characteristics), *local working conditions* (e.g., leadership), *organizational factors* (e.g., physical environment), and *external factors* (e.g., safety regulations). Similarly, Smits et al. [53] document 5 categories of AE causal factors, including *human causes* (e.g., coordination), *patient-related factors* (e.g., patient's age), *technical factors* (e.g., design and construction), *organizational factors* (e.g., culture and management), and *other factors*.

Accounting for these AE contributory factors is an important step toward better AE prevention and prediction as they inform what to include in the predictive model. However, most of these factors are latent organizational traits that cannot be reliably quantified, e.g., leadership and culture. This is corroborated by a recent case study from the Johns Hopkins Health System, which recognizes that "patient safety encompasses broad areas of risk that are not easily captured by conventional quality and safety reporting" [42:38]. This motivates us to critically analyze the extant AE prediction studies and the latent variable modeling methodologies, as we discuss next.

Adverse Event Prediction and Latent Variable Modeling

In this section, we review the methodological literatures on AE prediction and latent variable modeling. The latter is relevant because it comprises a broad set of methods to tease out latent factors in statistical modeling, and hence can inform our study in understanding how to capture latent AE contributory factors in each hospital. In reviewing these two streams of work, we uncover their potential limitations in yielding effective AE predictions. These methodological limitations hence justify the need to develop a new method for AE predictive modeling.

Adverse Event Prediction

To explore the landscape of AE prediction research, we conducted a systematic literature search by querying the published literature. The process yielded a total of 15 relevant articles (see Appendix B for details about our literature search and synthesis). Overall, our review of this literature reveals three key limitations in the extant work which we aim to fill.

First, despite the fact that there are many categories of AEs [31], the extant AE prediction research mainly focus on ADEs and surgical AEs. One of the reasons for this skewed focus is that other categories of AEs occur less frequently and are harder to recognize. As such, they are more difficult to model and predict because there is a higher noise-to-signal ratio for these incidents in the data. Nevertheless, being able to consider and predict all types of AEs would be useful for helping practitioners prevent any potential safety incident.

Second, extant AE prediction research emphasizes patient's observable characteristics, such as age, gender, and comorbidity, while omitting latent, contextual contributory factors of in-hospital AEs. This is mainly because patient features are easy to measure and obtain. In addition, the techniques used in current AE prediction models often do not have adequate support for including latent factors in the risk estimation. For example, organization culture, staff training, and hospital management have been identified as salient contributory factors for in-hospital AEs [38, 53], but no previous AE prediction models were designed to account for them.

Third, existing AE prediction models focus on patient's characteristics and do not recognize that each hospital is a unique entity and that its latent organizational capabilities and AE contributory factors evolve over time [2]. In other words, there exist cross-hospital heterogeneity and within-hospital temporal dependency in these AE contributory factors. The extant AE predictive models provide no support for these contextual and evolving latent factors.

In sum, the principal differences between our proposed model and the extant AE

prediction models are three. First, we jointly model multiple categories of AEs. Second, we look beyond patient's observable characteristics and consider latent, contextual contributory factors of in-hospital AEs. Third and finally, we recognize that each hospital is a unique entity and that its latent organizational capabilities and AE contributory factors evolve over time. Figure 1 illustrates the high-level AE predictive framework emerged from our literature synthesis.

[INSERT FIGURE 1 ABOUT HERE]

Latent Variable Modeling

Latent variable modeling comprises a broad set of methods and provides a promising methodological framework to address unobservable factors. As such, they are useful candidates to be considered in AE predictive modeling. Here we describe three families of methods in latent variable modeling and discuss issues when applying them in our specific problem domain.

Latent Factor Models: Latent factor models are a large family of methods that discover latent variables in high-dimensional data by means of low-rank approximations. Examples of this include latent semantic indexing, principal component analysis, and other matrix factorization algorithms. They have proven to be very useful in the context of recommendation systems, which tend to have large numbers of users and items and extremely sparse user–item matrices [36]. The low-rank approximations of the original user–item matrix not only resolve the sparsity issue; they also shed light on individual users' latent taste. However, there are two salient limitations for their use in AE prediction problems like ours. First, the low-rank approximations exploit information that is shared by individuals (e.g., users buying the same item) rather than information that is unique to an individual (e.g., a user's order history). As such, latent factor models cannot be used to extract hospital-level latent traits from longitudinal data. Second, these models do not handle multi-level data, and it is difficult to arrange nested AE-admission-hospital observations in a rectangular matrix representation. This means that latent factor models are not

suitable for AE prediction which, we will show, requires the ability to model multi-level data.

Latent chain models: Latent chain models, including state space models and hidden Markov models, are frequently used to model sequences of dependent observations. These models assume that the sequential observations of an outcome variable are generated from some latent states, and typically each state is dependent on the previous one (e.g., a first-order Markov chain). As such, they are powerful models for system dynamics, allowing researchers to infer the transition patterns for latent states and predict the next state given a set of sequential observations. However, latent chain models tend to consider only a small number of latent states (typically, 2 to 6) and highlight the heterogeneity driven by hidden states rather than individual subjects [16, 35]. These make them a less effective tool for AE predive modeling because they are not sensitive to each hospital's unique propensity and trend in making medical errors. Furthermore, latent chain models are designed to analyze the outcome sequences (e.g., the number of AEs in a hospital over time) and are not applicable to binary classification tasks (e.g., whether an admission in a hospital has an AE).

Mixed effects models: Mixed effects models incorporate random effects (REs) to account for latent individual- or group-specific effects. Conceptually, an RE is just an intercept term in a model that varies across subjects, but in practice there are multiple ways to implement it. One common approach is to extend the generalized linear model (GLM), i.e., logistic regression, to include REs. This type of model is known as a GLMM [54]. A recent study by Genovese et al. [24] applied a GLMM to a large clinical registry to predict postoperative AEs. Beyond GLMs, researchers have also incorporated REs in other models. For example, Hajjem et al. [27] developed a method called the *mixed effects regression tree* (MERT), which integrates REs into the classification and regression trees (CART) algorithm. In a follow-up study, the same authors proposed the *mixed effects random forest* (MERF) to combine REs with random

forests [28]. Mixed effects models, when compared to latent factor models and latent chain models, are more relevant and promising for AE predictions because they can easily distinguish multi-level observations and capture hospital-specific unobservables. However, extant mixed effects models assume that the REs are time invariant. This means that the RE component in conventional mixed effects models cannot effectively capture the evolving latent AE contributory factors in each hospital overtime.

Summary of Research Gaps

There has been a growing number of AE prediction studies outside the IS discipline. However, the extant research tends to ignore the fact that there exist different types of AEs and the salient role of latent, hospital-level AE contributory factors. Meanwhile, although latent variable modeling is a useful for capturing unobservable factors, existing approaches are not able to fully accommodate multi-level data format (i.e., AE-patient-hospital) nor can they capture the evolving, idiosyncratic latent factors in each hospital.

Against this background of related modeling work, the methodological development in our study is twofold. First and informed by our literature review, we formulate a novel AE predictive modeling problem in which there are multiple outcomes of interest associated with both observable cross-sectional features and unobservable sequential factors (see the Problem Formulation section below for details). We note that this problem is different from conventional predictive tasks where the focus is primarily on observable cross-sectional features. It is also different from traditional time series analysis because the time-varying factors in our context are latent and unobservable. Compared to other latent factor predictive problems, our predictive problem is distinct in its sequentially dependent latent factors and multiple outcomes of interest. Second, to address this predictive modeling problem we develop a new framework that integrates GLMM, a stochastic time-series process, and multitask learning (see the Model Development

section below for details). As we will show, the integration of the three techniques is non-trivial and yields a new and useful graphical model. Collectively, these methodological components allow us to model latent AE contributory factors in each hospital (through GLMM) that are evolving over time (through stochastic time-series processes) and impacting multiple categories of AEs that are potentially interrelated (through multitask learning).

Problem Formulation

We are now ready to define the AE prediction problem that is studied in this paper. A schematic representation of key elements in our problem formulation is shown in Figure 2. Formally, suppose that there is a set of inpatient admissions in H hospitals over T periods of time, with a total of $N_{h,t}$ admissions in hospital h at time t. We further assume that the outcomes of interest comprise K types/categories of AEs. We denote whether an admission involves a type-k AE by $y_{k,h,t,i}$, where $k = 1, \dots, K$ indexes the types of AEs, $h = 1, \dots, H$ indexes the hospitals, $t = 1, \dots, T$ indexes the time periods, and $i = 1, \dots, N_{h,t}$ indexes the admissions in hospital h at time t. We set $y_{k,h,t,i} = 1$ to denote the presence of a type-k AE during the hospital stay and 0 otherwise. We use $\theta_{k,h,t,i}$ to denote the probability that $y_{k,h,t,i} = 1$. Each admission is further characterized by $\boldsymbol{x}_{h,t,i}$, which is a vector of J features that are observable at the time of admission prior to any in-hospital AEs. These features may describe information about the admission, the patient, the attending physician, and the hospital. More importantly, and consistent with the literature [38, 42], we assume that $\boldsymbol{x}_{h,t,i}$ does *not* include all relevant features and that there exist salient unobservable AE contributory factors. In the presence of latent variables, we aim to learn a model to map $\boldsymbol{x}_{h,t,i}$ to $\theta_{k,h,t,i}$ based on a collection of observations $\{y_{k,h,t,i}, \boldsymbol{x}_{h,t,i}\}$ from H hospitals in T time periods for K different types of AEs, and use the model to predict the risk that a future admission i' characterized by $\boldsymbol{x}_{h,t',i'}$ at hospital h at t' involves type-k AEs.

[INSERT FIGURE 2 ABOUT HERE]

We assume the availability of a small set of easily obtainable covariates in $\boldsymbol{x}$ from multiple hospitals. This stands in sharp contrast to the alternative in which rich and extensive covariates in $\boldsymbol{x}$ are available from one hospital [40]. We also assume that the features in $\boldsymbol{x}$ are all present on admission (POA), rather than real-time recordings of all activities during the hospital stay. Both assumptions are realistic. While it is certainly useful to incorporate rich, real-time data for AE predictions, it is much easier and more realistic for practitioners to obtain simple POA features from multiple hospitals. Indeed, almost every state has a statewide hospital discharge dataset that can be used for such AE predictive modeling [3]. In contrast, very few hospitals have a sophisticated system to supply rich, real-time features for predictive analytics.

Model Development

In this section, we describe the proposed modeling framework for AE predictions: SALT. We will focus on three aspects of our model development and each corresponds to a subsection below: preliminaries, model specification, and model fitting. The goal of the preliminaries subsection is to provide readers with a brief overview of related methodological components used in SALT. After establishing these, we then formally specify the proposed model, including its parameterization, structure, and generative process. In particular, we define and justify the prior distribution or functional form for each parameter in our model. Finally, we discuss our model fitting through variational inference [10], which is an efficient technique to approximate the posterior probability of the unobserved variables in complex Bayesian models, allowing us to learn model parameters from data.

Preliminaries

As mentioned, the proposed model comprises three main methodological components: GLMM, stochastic time-series processes, and multitask learning. They are, respectively, intended to address three unique characteristics in AE predictions: (1) the AE contributory factors are latent,

(2) the AE contributory factors are hospital-specific and evolving, and (3) there are different categories of AEs, which are interrelated to each other. Rather than providing a comprehensive review of these methods (interested readers may consult Stroup [54], Prado and West [47], and Zhang and Yang [57]), our goal here is to highlight their essential methodological foundations and applicability in the context of AE predictions.

GLMM. GLMM is a popular method for latent variable modeling [54]. To predict AEs of a particular type (*k*), e.g., ADEs, a GLMM with hospital-level REs can be expressed as

$$\mathbb{E}\left(\text{logit}\left(\theta_{k,h,t,i}\right)\right) = \alpha_k + \boldsymbol{\beta}_k' \boldsymbol{x}_{h,t,i} + \delta_{k,h}. \quad (1)$$

In Eq. (1), $\boldsymbol{x}_{h,t,i}$ is a vector of J features that are observable at the time of admission, and α_k and $\boldsymbol{\beta}_k = [\beta_{k,1}, \dots, \beta_{k,J}]'$ are an outcome-specific intercept and coefficients. Intuitively, α_k represents the baseline risk of type-*k* AEs in the population, while $\boldsymbol{\beta}_k$ contains the weights for the corresponding features in predicting the outcomes of type-*k* AEs. The last term in Eq. (1), $\delta_{k,h}$, is a hospital-specific, time-invariant, normally distributed random variable. This is how the GLMM captures the latent heterogeneity in the propensity of type-*k* AEs in hospital *h* that cannot be explained by the observable features in $\boldsymbol{x}_{h,t,i}$. It is important to note that more than one RE variable can be incorporated in GLMMs. However, for ease of exposition and comparison, we will only consider hospital REs in this study since we are informed by the patient safety literature to account for hospital-level AE contributory factors.

Stochastic Time-Series Processes. Stochastic time-series processes can be used to describe how a random variable changes over time [47]. Suppose that $\boldsymbol{f}$ is a sequence of observed time-series values over T time periods (i.e., $\boldsymbol{f} = f_1, f_2, \dots, f_T$); time-series processes can be used to model the mean and the variance of $\boldsymbol{f}$ at each time period, i.e., $\mathbb{E}(f_t)$ and $\mathbb{V}(f_t)$. In practice, the mean of a time series is commonly modeled using the autoregressive moving-average (ARMA) process. An ARMA process consists of two parts, one for the autoregression and the other for the

moving average. Given a time series $\boldsymbol{f}$, an autoregressive (AR) process with order p, often written as AR(p), has the following effect on the series: $\mathbb{E}(f_t) = a + \sum_{i=1}^{p}(\phi_i f_{t-i})$, where a is a constant and ϕ_i is the coefficient for the ith lagged value (i.e., f_{t-i}). That is, an AR(p) model predicts the mean value in the current time period based on the observed values in the past p time periods. On the other hand, a moving-average (MA) process with order q, typically written as MA(q), uses previous errors as predictors for the current value in a series: $\mathbb{E}(f_t) = \mu + \sum_{i=1}^{q}(\psi_i e_{t-i})$, where μ is the mean of the series and ψ_i is the coefficient for the ith lagged error (i.e., $e_{t-i} = f_{t-1} - \mathbb{E}(f_{t-1})$). Hence, an ARMA($p$, q) model incorporates AR(p) and MA(q) to model the mean value in time series: $\mathbb{E}(f_t) = \mu + \sum_{i=1}^{p}(\phi_i f_{t-i}) + \sum_{j=1}^{q}(\psi_j e_{t-j})$.

The variance in time series, on the other hand, is often modeled with the autoregressive conditional heteroskedasticity (ARCH) process. An ARCH process models the changes of variance in a time series over time. An ARCH process with order r, or ARCH(r) in short, considers the variance in time t to be $\mathbb{V}(f_t) = \sigma_t^2 = g + \sum_{i=1}^{r}(\omega_i e_{t-i}^2)$, where g is a constant, ω_i is the coefficient for the ith lagged error (e_{t-i}). In other words, the variance of a time series at time t is modeled as linear combination of errors in the past r periods. It is important to note that ARMA and ARCH processes can be integrated together (that is, an ARMA-ARCH process) to simultaneously forecast the movements of the mean and the variance in a time series.

Having introduced the fundamentals of the ARMA and ARCH processes, we now turn to their roles in in-hospital AE predictions. A key limitation of GLMMs is that the REs are time-constant. To account for the evolving latent AE contributory factors in each hospital, a key novelty of the proposed SALT model is that we conceptualize and operationalize hospital-level REs as latent autoregressive trajectories. They are *latent* because most AE contributory factors are unobservable, *autoregressive* because the AE contributory factors are time-varying and sequentially dependent, and *trajectories* because each hospital is assumed to have a unique path

for the evolution of its AE contributory factors over time. For ease of exposition, we choose to use the ARMA(1, 1)-ARCH(1) process in our model specification below, but extending SALT to use a higher-order ARMA/ARCH is straightforward.

Multitask Learning. When there are multiple outcomes of interest, the coefficients of a particular predictor in different models—one for each outcome—may be correlated. For example, if patient's age is a significant predictor in a logistic regression for ADEs, it is possible that the same predictor is significant in another logistic regression for infection and that the coefficients for patient's age in the two regressions are correlated. Therefore, to improve predictive performance it can be useful to simultaneously consider multiple outcomes and allow the coefficients or parameters in different models to influence each other if and when they are related [57]. In our context, it is likely that a patient's risks to different categories of AEs are correlated since they are driven by the same set of patient, physician, and hospital factors. Therefore, it is worthwhile to consider multitask learning in our modeling framework. There are many ways to achieve multitask learning. Recent studies have shown that an effective approach is by superimposing a correlation structure over individual single-task models as a bridge to transfer information among them [40]. This approach is intuitive and can be applied to GLMM. We have therefore incorporated this approach to multitasking into our model.

Model Specification

Having introduced the preliminaries, we now turn to specify the proposed graphical model.[1] Following the best practice in Bayesian data analysis [22], we impose weakly informative priors on all of the non-deterministic parameters. This means that whenever possible, we choose to use

[1] We assume that all input variables have been centered to have a mean of 0 and all non-binary covariates have been scaled to have a standard deviation of 0.5, as suggested by Gelman et al. [23]. This allows us to avoid issues associated with predictors having heterogeneous centers and scales.

a reasonably shaped probability distribution for our parameter, either following related work or through our own theoretical justification. Although we will elaborate and justify our parameterization, for readers' convenience and easy reference we have summarized the description, distribution, functional form, and a relevant reference for each parameter in Table 2.

[INSERT TABLE 2 ABOUT HERE]

Overall, the generative process of AE risks comprises three components: the AE-specific regression intercept α_k, the regression coefficients $\boldsymbol{\beta}_k$, and the REs ($\delta_{k,h,t}$). Formally, we have

$$\mathbb{E}\left(\text{logit}\left(\theta_{k,h,t,i}\right)\right) = \alpha_k + \boldsymbol{\beta}_k' \boldsymbol{x}_{h,t,i} + \delta_{k,h,t}. \quad (2)$$

We discuss the definition and parameterization of each component below.

Regression Intercept Specific to Type-*k* AEs ($\boldsymbol{\alpha_k}$). The first term on the right-hand side of Eq. (2) is α_k. For this parameter, we use the Cauchy prior distribution with location -3 and scale 10, that is, $\alpha_k \sim \text{Cauchy}(-3,\ 10^2)$. The Cauchy distribution is bell-shaped, like the normal distribution, but it has thicker tails. Consistent with Gelman et al. [23], we use the Cauchy distribution for α_k because its thick tails can better accommodate extreme values, making it more robust than the conventional normal distribution. Similar to the mean and standard deviation in the normal distribution, the location and scale are two parameters of the Cauchy distribution, and they define, respectively, the center and the dispersion of the probability density. We chose the location parameter for the Cauchy prior (i.e., -3) because the prevalence of AEs is estimated to be between 2.9% and 16% [45]. On the logistic scale, this prevalence ranges from -3.51 (=logit(0.029)) to -1.66 (=logit(0.16)). Therefore, it is reasonable to set the location parameter for the Cauchy prior to -3 instead of 0. Meanwhile, the scale parameter for the Cauchy prior (i.e., 10) allows a wide variation for α_k, allowing the risk of type-*k* AEs for an average admission to be able to approach 0 (very unlikely) or 1 (very likely) or any values in between [23].

Regression Coefficients Specific to Type-*k* AEs ($\boldsymbol{\beta_k}$). The second term on the right-

hand side of Eq. (2) is $\boldsymbol{\beta}_k' \boldsymbol{x}_{h,t,i}$. As already noted, $\boldsymbol{\beta}_k$ is a vector of *J* coefficients that are used to adjust the risk of type-*k* AEs based on the observable features in $\boldsymbol{x}_{h,t,i}$. To accomplish multitask learning, we jointly model $\boldsymbol{\beta}_k$ for all *K* independent single-task learning models. We first set $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K]'$ and then derive $\boldsymbol{\beta}$ as follows:

$$\boldsymbol{\beta} = \left(\text{diag}(\boldsymbol{\tau}) * \Omega * \text{diag}(\boldsymbol{\tau})\right) * \widehat{\boldsymbol{\beta}}, \tag{3}$$

where $\boldsymbol{\tau}$ is a vector of *K* scalars for standard deviations (that is, $\boldsymbol{\tau} = [\tau_1, \tau_2, \ldots, \tau_K]'$), Ω is a $K \times K$ correlation matrix, and $\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_K]'$ is the matrix of estimates obtained from single-task learning (i.e., considering only one outcome at a time such as Eq. (1) did). The product of $\text{diag}(\boldsymbol{\tau}) * \Omega * \text{diag}(\boldsymbol{\tau})$ gives us the covariance matrix [7]. Through the covariance matrix, $\widehat{\boldsymbol{\beta}}_1$ (coefficients for the first category of AEs from single-task learning) can influence $\widehat{\boldsymbol{\beta}}_2, \ldots, \widehat{\boldsymbol{\beta}}_K$, and vice versa, leading to a multitask learned $\boldsymbol{\beta}$.

In Eq. (3), $\boldsymbol{\tau}$ (standard deviations), Ω (correlation matrix), and $\widehat{\boldsymbol{\beta}}$ (regression coefficients from single-task learning) are themselves random variables. As such, we need to provide a prior distribution for each of them. Following Feit and Bradlow [19], Lewandowski, Kurowicka, and Joe [39], and Ghosh et al. [25], we specify the following priors:

$$\tau_k \sim \text{Cauchy}^+(0, 2.5^2), \tag{4}$$
$$\Omega \sim \text{LKJ}(K, 1), \tag{5}$$
$$\hat{\beta}_{k,j} \sim \text{Student-t}(7, 0, 2.5^2). \tag{6}$$

The variable τ_k is assumed to follow a positive Cauchy distribution, or Cauchy$^+$ in short, with location 0 and scale 2.5. The Cauchy$^+$ distribution looks like a regular Cauchy distribution cut in half, and it has probability density only on positive values. This choice of prior is consistent with the prior choice in Feit and Bradlow [19]. Since standard deviations are always positive, the Cauchy$^+$ distribution ensures that τ_k has a probability mass only on positive values. The correlation matrix Ω is sampled using Lewandowski, Kurowicka, and Joe's [39] approach (hereafter LKJ), which is the state-of-the-art method for generating random correlation matrices.

The LKJ approach has two parameters. The first parameter specifies the dimension of the correlation matrix, which is *K* in our case for a *K* by *K* correlation matrix, to accommodate the correlation among *K* different categories of AEs. The second parameter of the LKJ distribution specifies the *shape* of the random correlation matrix. If the value of the shape parameter is greater than 1, the density increasingly concentrates towards the unit matrix, i.e., it favors less correlation. If it is less than 1, the density increasingly disperses away from the unit matrix, i.e., it favors more correlation. Since we have no prior knowledge about the shape of the correlation matrix, we set the shape parameter for the LKJ distribution to 1 so that the density is uniform over all correlation matrices (i.e., not favoring any shapes).

Regarding $\widehat{\boldsymbol{\beta}}$ (in which $\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\beta}}_1, \dots, \widehat{\boldsymbol{\beta}}_K]'$ and $\widehat{\boldsymbol{\beta}}_k = [\hat{\beta}_{k,1}, \dots, \hat{\beta}_{k,J}]'$), we assign a Student-*t* prior for each individual coefficient $\hat{\beta}_{k,j}$. There are three parameters for the Student-*t* prior: degrees of freedom, mean, and variance. Ghosh et al. [25] find that a Student-*t* distribution with 7 degrees of freedom is robust for coefficients in Bayesian logistic regression. Gelman et al. [23], on the other hand, suggest that the mean and standard deviations of the Student-*t* distribution in Bayesian logistic regression to be 0 and 2.5, respectively.

Random Effect Specific to Type-*k* AEs ($\boldsymbol{\delta_{k,h,t}}$). The last term on the right-hand side of Eq. (2) is $\delta_{k,h,t}$, which is an RE for an unobserved time-varying, hospital-specific risk of type-*k* AEs. As discussed earlier, it is theoretically and methodologically useful in our AE prediction context to model the hospital-level RE as a stochastic time-series process. To this end, we parameterize $\delta_{k,h,t}$ as follows:

$$\delta_{k,h,t} \sim \text{Normal}\left(\lambda_{k,h} z_{h,t},\ \nu^2_{k,h,t}\right). \tag{7}$$

We propose modeling the mean of $\delta_{k,h,t}$ as $\lambda_{k,h} z_{h,t}$, where $z_{h,t}$ represents a hospital's generic AE tendency that encapsulates all latent AE contributory factors, and $\lambda_{k,h}$ is a multiplier that rescales the overall AE tendency $z_{h,t}$ to the risk of type-*k* AEs. The reason for this

parameterization is to reflect the fact that a hospital's AE contributory factors are typically indifferent to the types of AEs. For example, a lack of patient safety culture in a hospital is a generic AE contributory factor that can affect all types of AEs. The same is true for most of the latent AE contributory factors. It is hence useful to model the evolution trajectory of the generic AE tendency $z_{h,t}$ that is indifferent to any specific AE types (i.e., $z_{h,t}$ is not indexed by *k*). Nevertheless, there is still a need to translate this generic AE tendency to each type of AE in AE predictive modeling. Accordingly, we use $\lambda_{k,h}$ as a multiplier for this purpose, which maps the generic AE tendency $z_{h,t}$ to the type-*k* AEs.

Both $z_{h,t}$ and $\lambda_{k,h}$ are random variables and have to be estimated from data. The $z_{h,t}$ variable is a key parameter in this study and follows an ARMA(1, 1)-ARCH(1) process:

$$z_{h,t} \sim \text{Normal}\left(\hat{z}_{h,t},\ \rho_{h,t}^2\right). \tag{8}$$

The mean of $z_{h,t}$, denoted as $\hat{z}_{h,t}$, is determined by an ARMA(1, 1) process. Consistent with the standard definition of an ARMA process as given in the previous subsection:

$$\hat{z}_{h,t} = \mu_h + \phi_h \hat{z}_{h,t-1} + \psi_h e_{h,t-1}, \tag{9}$$

$$e_{h,t-1} = z_{h,t-1} - \hat{z}_{h,t-1}. \tag{10}$$

The priors for the intercept (μ_h) and coefficients (ϕ_h and ψ_h) of the ARMA process follow a Cauchy distribution with location 0 and scale 10. This choice is similar in spirit to the prior distribution of α_k, which is the regression intercept in Eq. (2). As mentioned, REs are individual-specific intercept terms. As such, $\hat{z}_{h,t}$ and its parameters could have a similar Cauchy distribution as α_k that we specified earlier. Here, we set the location parameter for the Cauchy distribution to 0 (instead of -3 for α_k) because we do not have any *a prior* knowledge about the location parameter here.

Similarly, the standard deviation of $z_{h,t}$, denoted as $\rho_{h,t}$ in Eq (8), is defined based on an ARCH(1) process. Consistent with the standard definition of an ARCH(1) that we discussed in

the previous subsection, we specify $\rho_{h,t}$ as

$$\rho_{h,t} = \text{sqrt}\left(g_h + \omega_h\left(z_{h,t-1} - \mu_h\right)^2\right). \quad (11)$$

In Eq. (11), g_h and ω_h are the parameters of ARCH(1) and $z_{h,t-1} - \mu_h$ represents the difference from the mean at $t-1$. To determine the priors for g_h and ω_h, we draw from Gelman [21], who examines different prior choices for variance parameters in Bayesian hierarchical models. Gelman [21] recommends using $\text{Cauchy}^+(0, 5^2)$ as a default prior choice for standard definition. In our setting, the variance is determined by the linear combination of g_h and ω_h. Accordingly, we assign both of their priors to be $\text{Cauchy}^+(0, 5^2)$.

The two remaining parameters in Eq. (7) that we have not discussed are $\lambda_{k,h}$ and $\nu_{k,h,t}$. The parameter $\lambda_{k,h}$ is a scalar that is used to map $z_{h,t}$ to type-*k* AEs, whereas $\nu_{k,h,t}$ is a standard deviation for the RE parameter $\delta_{k,h,t}$. Recall that $\hat{z}_{h,t}$, the mean of $z_{h,t}$, is comprised of a set of three parameters, all of which have a $\text{Cauchy}(0, 10^2)$ prior distribution. This weakly informative prior should cover all plausible values for an RE variable. As such, in theory the scalar adjustor $\lambda_{k,h}$ should not further amplify the scale of $\hat{z}_{h,t}$. In other words, the value of $\lambda_{k,h}$ should be bounded between -1 and 1. Without any other *a priori* information, we therefore assign the prior of $\lambda_{k,h}$ to be a uniform distribution between these theoretical bounds, i.e., $\lambda_{k,h} \sim \text{Uniform}(-1, 1)$. For the standard deviation parameter $\nu_{k,h,t}$, we again follow Gelman's [21] recommendation and use the positive Cauchy prior, i.e., $\nu_{k,h,t} \sim \text{Cauchy}^+(0, 5^2)$.

Model Fitting

SALT has a hierarchical structure with non-conjugate prior distributions and boundary constraints on latent parameters. Fitting a complex model like SALT and obtaining the posterior density of the parameters can be computationally and mathematically challenging. We choose to fit this graphical model using variational inference [10], which provides an efficient means to

approximate posterior density through optimization. The main idea behind variational inference is to replace the joint posterior density with a simpler density that is often referred to as the variational distribution. The variational distribution has its own variational parameters. The task of variational inference is to optimize these variational parameters such that the variational distribution and the true posterior distribution are as close as possible (typically measured with the Kullback-Leibler divergence). We adopt the recent automatic differentiation variational inference (ADVI) technique from Kucukelbir et al. [37], which automates the variance inference procedure and returns the posteriors for the parameters in our model.

Experimental Study

As in any DSR projects, it is important to demonstrate that the proposed artifact is useful [26]. Therefore, the objective of our experimental study here is to evaluate the usefulness and value of the proposed method for in-hospital AE predictions. For the "computational genre" of DSR [48], evaluations are typically conducted through an set of experiments with the following steps: (1) select a suitable test bed, (2) explain the setup of data preparation and evaluation protocols, (3) explain the purpose and rationale of the experiments individually and as a whole, and (4) present and discuss the experimental results (see, e.g., [40, 50, 56, 58]). In what follows, we explicate each of these aspects.

Test Bed

To construct the test bed, we obtained a panel of inpatient discharge records of heart failure (HF) patients for the years 2010 to 2014 from the Florida Agency for Health Care Administration (AHCA). We decided to focus on HF patients because HF is a relatively common, yet severe, condition. There are currently over 5.7 million adults with HF in the US. Depending on the severity of the condition, HF can be treated with medications, surgical procedures, or medical devices, which make HF patients an appropriate sample for predicting all types of AEs (ADEs,

surgical AEs, device AEs, and so on). Meanwhile, the Florida AHCA dataset has two useful characteristics relevant to AE predictions. First, research has shown that the rate of AEs increases strongly with increasing age [12], and Florida has a large older population compared to other states. Second, the AHCA dataset is longitudinal with detailed POA indicators, which allows us to model hospital-level AEs over time and distinguish medical conditions that are POA versus non-POA in predictive modeling and evaluation. In sum, we believe HF admissions in the Florida AHCA dataset is a suitable test bed for evaluating in-hospital AE predictions.

We construct our test bed from the Florida AHCA dataset by identifying all admissions associated with an HF-related admission diagnosis code (that is, 428.*), which yields 152,878 unique admissions for this study. To protect patient privacy, the AHCA dataset does not include any personal identifiable information, nor does it include the exact dates of hospitalization. Instead, each admission is timestamped with a particular year and quarter (e.g., 2010Q1). We therefore number each quarter relative to the first quarter of 2010 and use this number as the time index for model training and testing (i.e., $t = 1$ for 2010Q1, $t = 5$ for 2011Q1, etc.).

We adapt the UMAEC list to determine whether a patient experienced an in-hospital AE (i.e., to label y). The UMAEC list consists of 862 unique AE-related diagnosis codes, which were compiled by a group of clinical experts. We further refine the list in two important ways for this study. First, we consider whether these AE codes are POA conditions. For example, an infection should *not* be considered a medical error if the patient came to the hospital with the condition. Failure to draw this distinction would result in a significant amount of false positive cases in the test bed. Second, to be more conservative in constructing our outcome variables, we consider only AE codes that have a high positive predictive value (PPV). This is to address the problem that even with the POA indicator, some of the codes on the UMAEC list may still not reliably identify AEs. For instance, the UMAEC list includes acute kidney failure and

cardiogenic shock as AEs. While they could be caused by medical errors, it is also possible that these conditions stem from a patient's underlying condition, even if not POA. In contrast, AE codes with a high PPV strongly indicate patient safety incidents when combined with the POA indicator. Examples of these high PPV AE codes include 998.4 (foreign body accidentally left during procedure), E856 (accidental poisoning by antibiotics), E873.2 (overdose of radiation in therapy), and E876.0 (mismatched blood in transfusion). A complete list of our high-PPV AE codes is shown in Table C1 in Appendix C. In sum, admissions are deemed to have an AE when they have a high-PPV AE code and the condition is *not* POA.

We consider five categories of AEs in our experiments: ADE, surgery, infection, device, and other. Table 3 summarizes the prevalence of different categories of AEs over time in the HF test bed. As shown in Table 3, infections are the most common type of AEs, while device and other AEs are much rarer. Overall, we find that 7.64 percent of the HF admissions in our dataset involved at least one AE, but the real AE prevalence among these admissions should be even higher since we only consider high-PPV AE codes when labeling AE incidents in our sample.

[INSERT TABLE 3 ABOUT HERE]

Settings for Predictive Evaluations

There are three key settings in implementing any predictive evaluations: variable construction, data splitting, and performance metrics. We discuss how we approach each of them in turn.

Variable construction. We construct predictors for each admission using only POA information. A total of 158 predictors (i.e., $\boldsymbol{x}$) are constructed but most of them (135) are hospital dummies (see Table C2 in Appendix C for a detailed data summary). Our predictors cover key characteristics related to the admission (e.g., weekday), the patient (e.g., age), the attending physician (e.g., years of practice experience), and the hospital (e.g., ownership). To construct these predictors, we link the AHCA dataset to several secondary data sources, including the

Hospital Compare and Physician Compare datasets from the Centers for Medicare and Medicaid Services and the Healthcare IT database from the American Hospital Association. Almost all our variables come directly from our data sources without any data pre-processing or transformation. The only exceptions are variables involving diagnosis codes. It is practically infeasible to include tens of thousands of diagnosis codes directly into a predictive model. Therefore, to characterize patient's health condition, it is necessary to aggregate diagnosis codes in some ways. Following prior research, we calculate and include five variables that are derived from POA diagnosis codes: number of chronic conditions, the Charlson and Elixhauser comorbidity indexes, and the risk stratification indexes for in-hospital mortality and length of stay.

Data splitting. Predictive accuracy requires an out-of-sample assessment, which involves splitting data into training and test sets [51]. Although cross-validation has been widely used in predictive analytics research, an issue with cross-validation on a longitudinal dataset like ours is that the training and test sets would contain instances from all time periods. This is problematic because in practice, especially in our AE prediction context, model training and model use are separated in time. One should build the model on historical data and use the model for future observations. As such, we consider a before/after data split in our evaluation, using admissions from the earlier period for model training and admissions from the later period for model testing. Given that our test bed contains data from 2010 to 2014, we vary the training period from one year (i.e., 4 quarters in 2010) to four years (i.e., 16 quarters in 2010–2013) and in each case use the rest of the data for testing. This permits an evaluation with temporally separated holdout samples and can shed light on how the length of the training period affects predictive accuracy.

Performance metrics. Consistent with prior predictive analytics studies [50, 51], we calculate the area under the receiver operating characteristics curve (AUC) to represent the performance of each predictive model. A larger AUC means that a model can better distinguish

between admissions with AEs and admissions without AEs. Additionally, we also report other common performance measures in predictive modeling, including precision, recall, and F-score. These additional metrics require a cut-point to turn probability values into binary outcome labels. We determine the cut-point using the Youden index, which maximizes the sum of sensitivity and specificity and has been used in prior research for similar purposes [18].

Three Experiments and Their Purposes

We consider three sets of experiments to evaluate the proposed method. Each of them serves a distinct purpose and as a whole, allow us to better understand the performance and value of the proposed method. In the first experiment (Evaluation 1), we examine whether SALT provides any lift to predictive performance compared with other mixed effects models. Since SALT integrates GLMM with multitask learning and stochastic time-series processes, it is important to tease out the usefulness of these novel design elements. We do so by comparing SALT with GLMM and other modern mixed effects models.

In the second experiment (Evaluation 2), we turn to assessing SALT's performance against a variety of predictive analytics techniques, which are based on trees, neural networks, or other frameworks. Here, the purpose is to demonstrate that the proposed method is competitive when compared to alternative techniques that have been used in prior AE prediction studies as well as the ones that have been commonly used in related predictive modeling tasks. This evaluation is important because these alternative techniques represent the existing state of knowledge for addressing the AE prediction problem. Our design will contribute to the DSR knowledge base only if it offers better predictive performance than the existing solution artifacts.

In our third and final experiment (Evaluation 3), we aim to look beyond the standard model evaluation metrics in the first two experiments and assess SALT's practical usefulness. It is challenging and rare to evaluate the impact of a new methodology like ours in its intended

practical setting. Therefore, simulations have been used to assess the practical usefulness of new methodologies. To simulate the use of AE prediction model in inpatient care, our rationale is that if a model can predict AE risks at the point of admission, alerts can be issued to the care team to take proper precautions for high-risk admissions.[2] This simulation allows us to assess the number of prevented AEs as well as the number of false alerts from each AE prediction model. Figure D1 in Appendix D provides details about our simulation setup.

Evaluation Results

Evaluation 1: Comparison with Existing Mixed Effects Models. In Evaluation 1, we consider three existing mixed effects models in this evaluation: GLMM, MERT, and MERF. The results are shown in Figure 3(a), which comprises a matrix of panels defined by different AE types (columns) and predictive measures (rows). In each panel, the x axis represents the data splitting points for the training sample (the splitting points are included in the training sample; the remainder is used for testing) and the y axis represents the score on the respective predictive measure. Overall, we find that SALT outperforms the alternative models in AUC across all AE categories and training periods. This suggests that SALT has a higher discriminatory power in separating admissions with and without AEs. For the other predictive measures, we also find that SALT tends to perform better than the benchmarks, especially in recall. However, it should be noted that the precision values are low across all models, suggesting that false alarms are common and inevitable in AE predictive modeling.

[INSERT FIGURE 3 ABOUT HERE]

Evaluation 2: Comparison with Alternative Techniques. In Evaluation 2, we consider

[2] Numerous clinical studies have found that alerts can reduce patient safety incidents [33], but the literature has also shown that many factors could also influence alert acceptance and effectiveness. For example, Seidling et al. [49] find that the alert display, frequency, and level (low, moderate, high risk) are significant modulators of providers' alert acceptance.

seven common techniques that have been used in previous AE prediction research or other predictive analytics tasks: CART, deep neural network (DNN), generalized boosted model (GBM), lasso regression (LR), naïve Bayes (NB), random forest (RF), and support vector machine (SVM). Many of these techniques require tuning. We grid-search for the optimal parameters and report the results for the best parameter settings. The results are shown in Figure 3(b).[3] We find that SALT's AUC, F-score, precision, and recall are higher than those of the alternative models in most scenarios.[4]

Evaluation 3: Estimation of SALT's Practical Impact. While Evaluations 1 and 2 show SALT's superior predictive performance compared to an array of competing techniques, Evaluation 3 aims to explore SALT's potential utility in practice. For simplicity, we compare SALT with only GLMM and LR, which are the two best alternative techniques in Evaluations 1 and 2, respectively, in terms of the AUC values. A high AUC value means that a model is better at balancing the trade-off between the hit rate and the false alarm rate, which is the focus of our simulation experiment in Evaluation 3. Indeed, for an AE predictive model to be useful in practice, it should maximize the hit rate while minimizing the false alarm rate.

The results from our simulation are summarized in Table 4. Overall, we find that SALT tends to have fewer false alarms and more prevented AEs than the benchmarks. It is noteworthy that in the surgery and infection categories of AEs, SALT often have more false alarms than the benchmarks. We posit that for these categories of AEs, SALT needed to have a lower cut-point so as to maximize the Youden index, which in turn led to more false alarms as well as more AEs being prevented. Nevertheless, the results in Table 4 suggest that there is greater utility when

[3] For NB and LR, they may be adversely impacted by multicollinearity. We hance conducted a robustness check by using principal component analysis to remove multicollinearity in our input. Our results are robust with this additional data pre-processing procedure.

[4] Detailed results and significance tests from Evaluations 1 and 2 are reported in Tables D1 and D2, respectively, in Appendix D.

applying SALT, rather than the alternative techniques, in clinical practice.

To further estimate SALT's business value, we calculate the potential cost savings if the method is implemented in practice. Previous studies have estimated the costs of ADE, surgery AE, and infection AE incidents.[5] Since our test bed involves admissions between 2010 and 2014, we inflate these cost estimates to 2010 dollars based on the Producer Price Index for primary services at general medical and surgical hospitals. This allows us to calculate the amount the practitioners saved from the prevented ADEs, surgery AEs, and infection AEs. We do not consider the savings for other categories of AEs because, to our knowledge, no such cost estimates are available in the literature. Finally, these savings from prevented AEs should be considered against the costs of false alarms, which reduce the care team's time. The cost savings in the last column of Table 4 are based on the difference between the savings from all prevented AEs and the costs of all false alarms. Based on the 2010–2013 training period (that is, with only test admissions in 2014), the cost savings per year across the three AE categories amount to $5.85 million for the Florida HF admissions in our sample alone.[6]

[INSERT TABLE 4 ABOUT HERE]

Robustness Checks

In the preceding analyses, we did not implement feature engineering or any remedy for the data imbalance issue. To ascertain the robustness of our findings, we further conducted a broad set of

[5] Bates et al. [8] find that each preventable ADE cost $4,685 in 1993. Similarly, Carey and Stefos [14] estimate that each surgery AE and infection AE cost $33,111 and $42,309, respectively, in 2006.

[6] Suppose that a care team consists of a doctor, a registered nurse, a pharmacist, and a technician, and that their hourly rates are $100, $34, $60, and $29, respectively, as per the U.S. Bureau of Labor Statistics. That amounts to an hourly rate of $223 for the entire team. With a median alert dwell time of 8 seconds [43], the cost of each false alarm is roughly 50 cents (223/3600*8). Through an admittedly oversimplified, back-of-the-envelope calculation, we find that extrapolating our cost savings estimate ($5.85 million) to all admissions in the US (over 30 million admissions per year), we could expect the business value of implementing SCARLET to be as high as $6.5 billion per year (i.e., 30,000 (nationwide admissions, in thousands) / 27 (HF admissions in Florida in 2014, in thousands) * 5.85 (cost savings in 2014 for our sample, in millions) = $6.5 billion per year).

analyses with different feature engineering strategies as well as data imbalance solutions. Results from these additional analyses are reported in Appendix E, which we summarize below.[7]

In terms of feature engineering, we consider a variety of ways to transform, discretize, and select features. Overall, we find no evidence that these feature engineering techniques lead to consistent and significant performance improvement in our setting. The results with feature engineering are qualitatively similar to those without. Theoretically, we do not expect feature engineering to significantly improve or alter model performance in our setting. This is because our dataset is very specific to inpatient care with a predefined domain of attributes that are highly relevant to the problem of AEs in hospitals. As mentioned earlier, we do not have a very large number of features. Aside from hospital dummies, there are only 23 features being used to characterize each visit. All our features have a strong face validity in terms of their relevance to the AE prediction problem. This is different from other predictive modeling exercises in which the datasets may come from multiple domains and the attributes may be noisy or with varying level of relevance/importance. As such, there should be no ex-ante reason to believe that feature engineering would significantly improve predictive performance or change the relative standing of the models in our specific setting. More broadly, this is likely why the majority of prior AE prediction studies did not consider feature engineering (see Table B1 in Appendix B).

Moreover, we also consider three common strategies to address data imbalance: Synthetic Minority Over-Sampling Technique (SMOTE), random over-sampling, and random under-sampling. Since AEs are rare, these data imbalance solutions help construct a training dataset in which the number of AE cases are roughly the same as the number of non-AE cases. Overall, we find that SALT remains to be the most accurate model when in the presence of these data

[7] We appreciate insightful comments from the anonymous Associate Editor for improving the discussion of these additional results from our robustness checks.

balancing methods. Nevertheless, it is important to note that these data imbalance solutions do not guarantee to yield better predictive performance. As He and Garcia explained [29], we may miss out important patterns pertaining to the majority class in under-sampling methods, on one hand, and run into overfitting as we exaggerate patterns exhibited in minority class in over-sampling methods, on the other hand. These potential hazards can also prevent us from isolating and identifying the performance of the proposed model, which is why we did not consider data imbalance solutions in our main analyses.

In all, these robustness checks show that SALT still outperformed alternative models when we implemented any of these feature engineering and data balancing strategies. This suggests that the superior predictive power of SALT is reliable and robust to the presence (or absence) of these common data preprocessing strategies in the machine learning practice.

Discussion and Conclusion

Lapses in patient safety are a major problem in healthcare. Empirical evidence has shown that AEs resulting from medical errors are common in current medical practice and often incur significant costs, cause serious injuries, and lead to premature deaths. The importance of preventing in-hospital AEs cannot be overemphasized given that there are more than 30 million hospital stays per year in the U.S. and 2.9% to 16.6% of them experienced at least an AE [12, 45]. With the emergence of healthcare big data and analytics, many IS scholars have suggested that IS research has great potential in transforming healthcare in general and preventing AEs in particular [6, 15, 20]. Consistent with recent DSR [17, 52, 56, 58], we formulate a new AE prediction problem in which latent variables are significant AE drivers, and propose a novel modeling framework—SALT—to proactively predict AEs before the occurrence of medical harms. The evaluation results from our before/after holdout sample and a simulated experiment show that SALT has better predictive power than the alternative techniques and can provide

greater practical utility in terms of more AEs being prevented with fewer false alarms.

Research Contributions

Our study makes two main contributions to IS research. First, we formalize the in-hospital AE prediction problem in accordance with IOM's definition of AEs and recognize that AEs are harms caused by medical management. Unlike prior AE prediction research in the IS literature which emphasizes predicting patients' physiologic deterioration in their nature course of disease progression [40, 44], our problem formulation highlights the fact that latent and evolving organizational factors (such as hospital leadership, patient safety culture, and staffing levels) are salient contributory factors to patient safety incidents in hospital stays. This new perspective of AEs calls for a variety of design considerations, including the needs to consider (1) evolving latent factors in each hospital, (2) multiple potently interrelated categories of AEs, and (3) multi-level observations (AEs in a patient and patients in a hospital), and (4) utilizing only available information at the point of each admission.

Second, we develop a novel graphical model to address this in-hospital AE prediction problem. Our design is justified based on the findings in the empirical literature and the limitations in the extant methods. Prior DSR studies in AE-related topics typically did not consider AE as harms from medical errors [40, 44], and when they do, was focused on post hoc AE detection instead of ex ante AE prediction [1]. Prior AE-related studies also did not address latent, evolving hospital factors in their risk modeling. As such, we believe our model is the first comprehensive solution for predicting harms from medical errors in the IS literature. We demonstrate the superior prediction performance and practical utility of our model compared to alternative techniques in machine learning and latent variable modeling. Two design principles in our predictive modeling are to consider (1) multiple categories of AEs/outcomes and (2) the evolving latent hospital/organizational factors. These principles serve as *nascent design theory*,

according to Gregor and Hevner [26], that can have theoretical implications for DSR projects beyond the current AE prediction context. That is, for predictive modeling it is useful to consider multiple related outcomes together and explicitly account for evolving unobservable factors.

Implications for Praxis

The implications for praxis from this study are twofold. First, hospitals should consider implementing AE prediction models like ours in their clinical workflow. This paper shows the feasibility of implementing an AE prediction technique without "rich" or real-time data. While the availability of such data would certainly enhance predictive performance, it is challenging, if not unrealistic, to require hospitals to muster such data for AE predictions. Therefore, from a practitioner standpoint, being able to use fewer and readily available POA information for AE predictions is not an insignificant strength of the proposed method. As mentioned, the data we used for model training is a state-wide hospital discharge dataset from Florida. Almost every state in the U.S. has been collecting and releasing similar data [3]. With such data, it is straightforward to follow the model specification and model fitting procedure described in this paper to train the proposed SALT model locally in the hospital without huge upfront investments in the data infrastructure.

Second, once the model is trained, it can be integrated with hospital's check-in system so as to enable the automatic usage of POA information to generate risk predictions for different categories of AEs for each admission. Naturally, predictions alone are not useful unless they prompt actions at a right time. Result of this AE risk profiling should be shared with all the members in the care team, including the attending physician, surgent, nurses, etc., so that they can takes proper precautions when interacting with the patient. We demonstrate that preventing harms from medical errors not only is beneficial to public health but also makes sense from the economics perspective. Indeed, our simulation suggested that investment in such AE analytics

capability will likely lead to significant cost savings from the prevented patient safety incidents.

Limitations and Future Directions

This study has several limitations, which creates opportunities for future research. First, we evaluated our method using HF admissions, which we chose for their importance and prevalence in our society. Nevertheless, it is unclear whether the proposed method will still have good predictive performance among patients with other conditions. Future work is needed to identify the generalizability of SALT in different AE contexts.

Second, there have been criticisms on the reliability issue of ICD-9-CM coding in practice [13]. Since our AE labels rely on the ICD-9-CM diagnostic codes in administrative data, there are concerns about the reliability of our outcome variables. To address this reliability issue in our evaluations, we chose to be more conservative and consider only a narrow set of high PPV AE codes from the UMAE list as AEs. We further used the POA indicator to avoid making false positive labels. A promising direction of future research is to integrate AE detection and AE prediction in which the former ensures reliable AE labels for the latter.

Third, concept drift is a common problem in practice. While we do not find evidence of concept drifting in our AE setting (see Table E22 in Appendix E), our data comprises observations from only one population across five years. The threat of concept drifting may emerge with a different population or with a longer observation period. Future research could extend our model in two ways to cope with concept drift. First, one can develop an online learning algorithm such that a model keeps updating its parameters whenever a new batch of data arrives. In a Bayesian framework, Opper [46] showed that we can use the Bayes rule to update the posterior distribution of each parameter for new data points, and there is no need to consider the entire old dataset in the posterior update. Second, researchers have integrated change-point detection in classification in order to address concept drift [5]. In principle, we can train multiple

models—one for each incoming data batch (e.g., each year), and the change-point detection component allows us to identify and discard models before the change point. The benefit of this approach is that we do not need to update those models trained from earlier data batches. However, this approach assumes that the change point is salient and abrupt, making it less applicable to slow-evolving concept drift.

Conclusion

Patient safety is one of the most important topics in modern healthcare. While there have been concerted efforts in advancing the measurement and reporting of patient safety incidents, we argue that hospitals can (and should) also proactively screen and intercept AEs before their occurrence. Our study represents an important step toward reducing in-hospital AEs with predictive analytics. Following the DSR paradigm, we formalized the in-hospital AE prediction problem and proposed a novel approach, SALT, to predict in-hospital AEs caused by medical errors. Our design was informed by recent empirical findings regarding the characteristics of in-hospital AEs and their latent contributing factors. It also addressed several methodological limitations in the prior AE prediction models. Empirical evaluations based on a HF discharge dataset demonstrated that effectiveness of the proposed method in AE prediction and prevention. We hope that our results provide impetus for further work in this high-impact problem domain on improving patient safety with analytics and IT artifacts.

References

1. Abbasi, A.; Li, J.; Adjeroh, D.; Abate, M.; and Zheng, W. Don't Mention It? Analyzing User-Generated Content Signals for Early Adverse Event Warnings. *Information Systems Research*, (2019).
2. Aldrich, H.E. and Ruef, M. *Organizations Evolving*. SAGE Publications Ltd, London, 2006.
3. Andrews, R.M. Statewide hospital discharge data: collection, use, limitations, and improvements. *Health Services Research*, *50*, Suppl 1 (2015); 1273–1299.
4. Aron, R.; Dutta, S.; Janakiraman, R.; and Pathak, P.A. The impact of automation of systems on medical errors: evidence from field research. *Information Systems Research*, *22*, 3 (2011); 429–446.

5. Bach, S. and Maloof, M. A Bayesian approach to concept drift. *Advances in Neural Information Processing Systems*, *23*, (2010); 127–135.
6. Bardhan, I.; Chen, H.; and Karahanna, E. Connecting systems, data, and people: a multidisciplinary research roadmap for chronic disease management. *MIS Quarterly*, *44*, 1 (2020); 185–200.
7. Barnard, J.; McCulloch, R.; and Meng, X.-L. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, *10*, 4 (2000); 1281–1311.
8. Bates, D.W.; Spell, N.; Cullen, D.J.; Burdick, E.; Laird, N.; et al. The costs of adverse drug events in hospitalized patients. *JAMA*, *277*, 4 (1997); 307–311.
9. Benaroch, M. and Chernobai, A. Operational IT failures, IT value destruction, and board-level IT governance changes. *MIS Quarterly*, *41*, 3 (2017); 729–762.
10. Blei, D.M.; Kucukelbir, A.; and McAuliffe, J.D. Variational inference: a review for statisticians. *Journal of the American Statistical Association*, *112*, 518 (2017); 859–877.
11. Bos, J.V.D.; Rustagi, K.; Gray, T.; Halford, M.; Ziemkiewicz, E.; and Shreve, J. The $17.1 billion problem: the annual cost of measurable medical errors. *Health Affairs*, *30*, 4 (2011); 596–603.
12. Brennan, T.A.; Leape, L.L.; Laird, N.M.; Hebert, L.; Localio, A.R.; et al. Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I. *New England Journal of Medicine*, *324*, 6 (1991); 370–376.
13. Burns, E.M.; Rigby, E.; Mamidanna, R.; Bottle, A.; Aylin, P.; et al. Systematic review of discharge coding accuracy. *Journal of Public Health*, *34*, 1 (2012); 138–148.
14. Carey, K. and Stefos, T. Measuring the cost of hospital adverse patient safety events. *Health Economics*, *20*, 12 (2011); 1417–1430.
15. Chen, H.; Chiang, R.H.L.; and Storey, V.C. Business intelligence and analytics: from big data to big impact. *MIS Quarterly*, *36*, 4 (2012); 1165–1188.
16. Chen, W.; Wei, X.; and Zhu, K. Engaging voluntary contributions in online communities: a hidden Markov model. *MIS Quarterly*, *42*, 1 (2018); 83–100.
17. Ebrahimi, M.; Jr, J.F.N.; and Chen, H. Semi-supervised cyber threat identification in dark net markets: a transductive and deep learning approach. *Journal of Management Information Systems*, *37*, 3 (2020); 694–722.
18. Eschmann, E.; Beeler, P.E.; Schneemann, M.; and Blaser, J. Developing strategies for predicting hyperkalemia in potassium-increasing drug-drug interactions. *Journal of the American Medical Informatics Association*, *24*, 1 (2017); 60–66.
19. Feit, E.M. and Bradlow, E.T. Fusion modeling. In C. Homburg, M. Klarmann and A. Vomberg, eds., *Handbook of Market Research*. Springer International Publishing, Cham, 2019; pp. 1–34.
20. Fichman, R.G.; Kohli, R.; and Krishnan, R. The role of information systems in healthcare: current research and future trends. *Information Systems Research*, *22*, 3 (2011); 419–428.
21. Gelman, A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, *1*, 3 (2006); 515–534.
22. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; and Rubin, D.B. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2013.
23. Gelman, A.; Jakulin, A.; Pittau, M.G.; and Su, Y.-S. A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, *2*, 4 (2008); 1360–1383.
24. Genovese, E.A.; Fish, L.; Chaer, R.A.; Makaroun, M.S.; and Baril, D.T. Risk stratification

for the development of respiratory adverse events following vascular surgery using the Society of Vascular Surgery's Vascular Quality Initiative. *Journal of Vascular Surgery*, *65*, 2 (2017); 459–470.
25. Ghosh, J.; Li, Y.; and Mitra, R. On the use of Cauchy prior distributions for Bayesian logistic regression. *Bayesian Analysis*, *13*, 2 (2018); 359–383.
26. Gregor, S. and Hevner, A. Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*, 2 (2013); 337–355.
27. Hajjem, A.; Bellavance, F.; and Larocque, D. Mixed effects regression trees for clustered data. *Statistics & Probability Letters*, *81*, 4 (2011); 451–459.
28. Hajjem, A.; Bellavance, F.; and Larocque, D. Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, *84*, 6 (2014); 1313–1328.
29. He, H. and Garcia, E.A. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*, 9 (2009); 1263–1284.
30. Hevner, A.; March, S.; Park, J.; and Ram, S. Design science in information systems research. *MIS Quarterly*, *28*, 1 (2004); 75–106.
31. Hougland, P.; Nebeker, J.; Pickard, S.; Van Tuinen, M.; Masheter, C.; et al. Using ICD-9-CM codes in hospital claims data to detect adverse events in patient safety surveillance. In K. Henriksen, J.B. Battles, M.A. Keyes and M.L. Grady, eds., *Advances in Patient Safety: New Directions and Alternative Approaches*. AHRQ, Rockville (MD), 2008.
32. Hydari, M.Z.; Telang, R.; and Marella, W.M. Saving patient Ryan—can advanced electronic medical records make patient care safer? *Management Science*, *65*, 5 (2018); 2041–2059.
33. Kaushal, R.; Shojania, K.G.; and Bates, D.W. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. *Archives of Internal Medicine*, *163*, 12 (2003); 1409–1416.
34. Kohn, L.T.; Corrigan, J.; and Donaldson, M.S., eds. *To Err Is Human: Building a Safer Health System*. National Academies Press, 1999.
35. Kokkodis, M.; Lappas, T.; and Ransbotham, S. From lurkers to workers: predicting voluntary contribution and community welfare. *Information Systems Research*, *31*, 2 (2020); 607–626.
36. Koren, Y.; Bell, R.; and Volinsky, C. Matrix factorization techniques for recommender systems. *Computer*, *42*, 8 (2009); 30–37.
37. Kucukelbir, A.; Tran, D.; Ranganath, R.; Gelman, A.; and Blei, D.M. Automatic differentiation variational inference. *Journal of Machine Learning Research*, *18*, 14 (2017); 1–45.
38. Lawton, R.; McEachan, R.R.C.; Giles, S.J.; Sirriyeh, R.; Watt, I.S.; and Wright, J. Development of an evidence-based framework of factors contributing to patient safety incidents in hospital settings: a systematic review. *BMJ Quality & Safety*, *21*, 5 (2012); 369–380.
39. Lewandowski, D.; Kurowicka, D.; and Joe, H. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*, 9 (2009); 1989–2001.
40. Lin, Y.-K.; Chen, H.; Brown, R.; and Li, S.-H. Healthcare predictive analytics for risk profiling in chronic care: a Bayesian multitask learning approach. *MIS Quarterly*, *41*, 2 (2017); 473–495.
41. Makary, M.A. and Daniel, M. Medical error—the third leading cause of death in the US. *BMJ*, *353*, (2016); i2139.
42. Mathews, S.C.; Sutcliffe, K.M.; Garrett, M.R.; Pronovost, P.J.; and Paine, L. A framework

for operationalizing risk: a practical approach to patient safety. *Journal of Healthcare Risk Management*, *38*, 1 (2018); 38–46.
43. McDaniel, R.B.; Burlison, J.D.; Baker, D.K.; Hasan, M.; Robertson, J.; et al. Alert dwell time: introduction of a measure to evaluate interruptive clinical decision support alerts. *Journal of the American Medical Informatics Association*, *23*, e1 (2016); e138–e141.
44. Meyer, G.; Adomavicius, G.; Johnson, P.E.; Elidrisi, M.; Rush, W.A.; et al. A machine learning approach to improving dynamic decision making. *Information Systems Research*, *25*, 2 (2014); 239–263.
45. Murff, H.J.; Patel, V.L.; Hripcsak, G.; and Bates, D.W. Detecting adverse events for patient safety research: a review of current methodologies. *Journal of Biomedical Informatics*, *36*, 1–2 (2003); 131–143.
46. Opper, M. A Bayesian approach to on-line learning. In *On-line learning in neural networks*. Cambridge University Press, USA, 1999; pp. 363–378.
47. Prado, R. and West, M. *Time Series: Modeling, Computation, and Inference*. Chapman and Hall/CRC, 2010.
48. Rai, A. Diversity of design science research. *MIS Quarterly*, *41*, 1 (2017); iii–xviii.
49. Seidling, H.M.; Phansalkar, S.; Seger, D.L.; Paterno, M.D.; Shaykevich, S.; et al. Factors influencing alert acceptance: a novel approach for predicting the success of clinical decision support. *Journal of the American Medical Informatics Association*, *18*, 4 (2011); 479–484.
50. Shi, D.; Guan, J.; Zurada, J.; and Manikas, A. A data-mining approach to identification of risk factors in safety management systems. *Journal of Management Information Systems*, *34*, 4 (2017); 1054–1081.
51. Shmueli, G. and Koppius, O.R. Predictive analytics in information systems research. *MIS Quarterly*, *35*, 3 (2011); 553–572.
52. Silic, M. and Lowry, P.B. Using design-science based gamification to improve organizational security training and compliance. *Journal of Management Information Systems*, *37*, 1 (2020); 129–161.
53. Smits, M.; Zegers, M.; Groenewegen, P.P.; Timmermans, D.R.M.; Zwaan, L.; et al. Exploring the causes of adverse events in hospitals and potential prevention strategies. *Quality and Safety in Health Care*, *19*, 5 (2010).
54. Stroup, W.W. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press, Boca Raton, 2012.
55. Vries, E.N. de; Ramrattan, M.A.; Smorenburg, S.M.; Gouma, D.J.; and Boermeester, M.A. The incidence and nature of in-hospital adverse events: a systematic review. *Quality and Safety in Health Care*, *17*, 3 (2008); 216–223.
56. Xie, J.; Zhang, Z.; Liu, X.; and Zeng, D. Unveiling the hidden truth of drug addiction: a social media approach using similarity network-based deep learning. *Journal of Management Information Systems*, *38*, 1 (2021); 166–195.
57. Zhang, Y. and Yang, Q. A survey on multi-task learning. *arXiv preprint arXiv:1707.08114*, (2017).
58. Zhu, H.; Samtani, S.; Chen, H.; and Jr, J.F.N. Human identification for activities of daily living: a deep transfer learning approach. *Journal of Management Information Systems*, *37*, 2 (2020); 457–483.

Tables

Table 1. Conceptualizations of Adverse Events in the IS Literature

	Conceptualization of AEs	Example AE
More abstract and context-independent conceptualization ↕ ↕ ↕ ↕ More specific and contextualized conceptualization	Level 1: AEs as undesirable business or societal outcomes	A financial service company experiences IT failure [9]
	Level 2: AEs as deteriorating patient outcomes	A diabetic patient develops stroke, heart attack, or renal failure in five years [40]
	Level 3: AEs as harms from medical errors	A postmarketing drug shows previously unknown harms to patients [1]

Table 2. Summary of Parameters in SALT

Parameter	Distribution/Functional Form and Reference	Description
α_k	Cauchy distribution [23]: $\alpha^{(k)} \sim \text{Cauchy}(0, 10^2)$	Regression intercept for type *k* AEs
$\boldsymbol{\beta}_k$	Deterministic function form [7]: $\boldsymbol{\beta} = \left(\text{diag}(\boldsymbol{\tau}) * \Omega * \text{diag}(\boldsymbol{\tau})\right) * \widehat{\boldsymbol{\beta}}$	Multitask learning regression coefficients for type *k* AEs
$\boldsymbol{\tau}$	Positive Cauchy distribution [19]: $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_K], \quad \tau_k \sim \text{Cauchy}^+(0, 2.5^2)$	Standard deviations in *K* categories of AEs
Ω	Lewandowski, Kurowicka and Joe [39] distribution: $\Omega \sim \text{LKJ}(K, 1)$	Correlation matrix across *K* categories of AEs
$\widehat{\boldsymbol{\beta}}$	Student's *t* distribution [25]: $\widehat{\boldsymbol{\beta}} = [\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_K]', \ \widehat{\boldsymbol{\beta}}_k = [\hat{\beta}_{k,j}, \ldots, \hat{\beta}_{k,J}]', \ \hat{\beta}_{k,j} \sim \text{Student-t}(7, 0, 2.5^2)$	Single-task learning regression coefficients for type *k* AEs
$\delta_{k,h,t}$	Normal distribution (proposed in this study): $\pi_{k,h,t} \sim \text{Normal}\left(\lambda_{k,h} z_{h,t}, \ \nu_{k,h,t}^2\right)$	Random effect for type *k* AEs in hospital *h* at time *t*
$\lambda_{k,h}$	Cauchy distribution (proposed in this study): $\lambda_{k,h} \sim \text{Uniform}(0, 1)$	Scalar to map overall hospital AE tendency to type *k* AEs
$z_{h,t}$	Normal distribution (proposed in this study): $z_{h,t} \sim \text{Normal}(\hat{z}_{h,t}, \ \rho_{h,t}^2)$	Overall AE tendency in hospital *h* at time *t*
$\nu_{k,h,t}$	Positive Cauchy distribution [21]: $\nu_{k,h,t} \sim \text{Cauchy}^+(0, 5^2)$	Standard deviation of $\pi_{k,h,t}$
$\hat{z}_{h,t}$	Deterministic function form (definition of ARMA(1, 1)): $\hat{z}_{h,t} = \mu_h + \phi_h \hat{z}_{h,t-1} + \psi_h e_{h,t-1}$	ARMA(1, 1) process
μ_h, ϕ_h, ψ_h	Normal distribution [23]: $\mu_h, \phi_h, \psi_h \sim \text{Cauchy}(0, 10^2)$	Intercept and coefficients of an ARMA(1, 1) process
$e_{h,t-1}$	Deterministic function form (definition of ARMA(1, 1)): $e_{h,t-1} = z_{h,t-1} - \hat{z}_{h,t-1}$	Lagged error in an ARMA(1, 1) process
$\rho_{h,t}$	Deterministic function form (definition of ARCH(1)): $\rho_{h,t} = \text{sqrt}\left(g_h + \omega_h \left(z_{h,t-1} - c_h\right)^2\right)$	ARCH(1) process
g_h, ω_h	Positive Cauchy distribution [21]: $g_h, \omega_h \sim \text{Cauchy}^+(0, 5^2)$	Intercept and coefficient of an ARCH(1) process

Table 3. Prevalence of AEs in the HF Test Bed

AE Category	2010	2011	2012	2013	2014	All
ADE	1,009 (2.90%)	800 (2.49%)	745 (2.52%)	723 (2.53%)	718 (2.57%)	3,995 (2.61%)
Surgery	480 (1.38%)	388 (1.21%)	384 (1.30%)	401 (1.40%)	381 (1.37%)	2,034 (1.33%)
Infection	1,692 (4.87%)	1,517 (4.72%)	1,306 (4.42%)	1,204 (4.21%)	1,095 (3.93%)	6,814 (4.46%)
Device	179 (0.52%)	147 (0.46%)	116 (0.39%)	107 (0.37%)	108 (0.39%)	657 (0.43%)
Others	89 (0.26%)	73 (0.23%)	62 (0.21%)	74 (0.26%)	47 (0.17%)	345 (0.23%)
All	2,916 (8.39%)	2,489 (7.75%)	2,191 (7.41%)	2,111 (7.39%)	1,971 (7.07%)	11,678 (7.64%)

Notes. The values represent the number (and percent) of admissions for each AE category per year. The values in the last row are less than the column sum because some patients experienced multiple AEs during their admissions.

Table 4. Simulation Results from Evaluation 3

AE category	Training	Testing		N. False Alarms			N. AEs Prevented			Cost Saving ($ millions)
	Period	N.	N. AEs	GLMM	LR	SALT	GLMM	LR	SALT	SALT
ADE	2010	118,142	2,986	4,875	4,898	**1,459**	102.1 (8.0)	97.3 (7.9)	**138.3 (9.5)**	0.36 [0.28, 0.44]
	2010-2011	86,035	2,186	3,689	3,514	**730**	81.3 (7.3)	83.7 (7.3)	**101.5 (8.1)**	0.22 [0.16, 0.29]
	2010-2012	56,460	1,441	2,273	2,223	**894**	55.5 (6.1)	52.4 (5.9)	**61.9 (6.5)**	0.20 [0.13, 0.26]
	2010-2013	27,887	718	1,149	1,160	**596**	**32.4 (4.5)**	30.1 (4.4)	29.2 (4.3)	0.13 [0.08, 0.18]
Surgery	2010	118,142	1,554	2,887	**2,301**	2,631	186.5 (10.6)	163.6 (10.1)	**188.5 (10.6)**	1.72 [1.32, 2.11]
	2010-2011	86,035	1,166	**1,373**	1,551	1,427	119.8 (8.6)	121.1 (8.6)	**137.9 (9.1)**	1.99 [1.55, 2.44]
	2010-2012	56,460	782	1,038	1,074	**567**	82.2 (7.1)	82.3 (7.2)	**82.8 (7.1)**	0.90 [0.62, 1.19]
	2010-2013	27,887	381	551	**531**	540	44.4 (5.1)	40.5 (4.9)	**47.9 (5.4)**	0.46 [0.25, 0.66]
Infection	2010	118,142	5,122	**2,654**	2,947	2,781	292.7 (13.5)	305.2 (13.9)	**360.4 (15.0)**	5.75 [4.94, 6.55]
	2010-2011	86,035	3,605	**1,939**	2,074	2,264	206.6 (11.1)	215.4 (11.6)	**289.9 (13.7)**	3.94 [3.27, 4.61]
	2010-2012	56,460	2,299	1,284	1,308	**928**	130.3 (8.8)	131.7 (8.9)	**148.7 (9.9)**	2.17 [1.68, 2.67]
	2010-2013	27,887	1,095	684	**643**	696	65.1 (6.5)	65.2 (6.4)	**81.7 (7.2)**	1.07 [0.74, 1.41]
Device	2010	118,142	478	2,992	**2,070**	2,626	77.0 (6.8)	63.3 (6.1)	**87.1 (7.4)**	Not available
	2010-2011	86,035	331	2,136	2,208	**356**	53.3 (5.6)	**53.7 (5.8)**	45.2 (5.3)	Not available
	2010-2012	56,460	215	1,497	1,572	**348**	35.1 (4.7)	35.4 (4.9)	**38.5 (4.9)**	Not available
	2010-2013	27,887	108	648	588	**181**	18.0 (3.3)	**16.7 (3.2)**	14.9 (3.1)	Not available
Other	2010	118,142	256	1,346	**1,286**	1,301	**18.3 (3.3)**	15.2 (2.9)	16.3 (3.2)	Not available
	2010-2011	86,035	183	**516**	748	578	9.5 (2.4)	10.5 (2.6)	**11.5 (2.7)**	Not available
	2010-2012	56,460	121	494	526	**187**	8.4 (2.3)	9.2 (2.4)	**9.7 (2.5)**	Not available
	2010-2013	27,887	47	227	247	**174**	3.9 (1.6)	3.4 (1.5)	**5.9 (1.8)**	Not available

Notes. Bold font represents the best model in the row with the lowest number of false alarms or the highest number of AEs prevented. For the number of AEs prevented, standard deviations are shown in the parentheses. For cost saving, the 95% confidence interval is shown in the brackets. We only show the cost saving from SALT because compared with the alternative techniques, it tends to have a lower number of false alarms and a higher number of AEs prevented. The cost savings for "Device" and "Other" AEs are not available because there is no study, to our knowledge, estimating the costs of these AEs in the literature.

Figures

Figure 1. Framework of AE Prediction Modeling

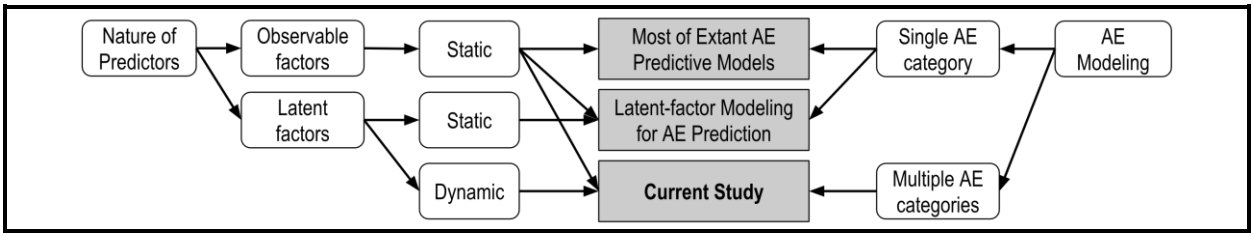


Figure 2. Schematic Illustration of the AE Prediction Problem

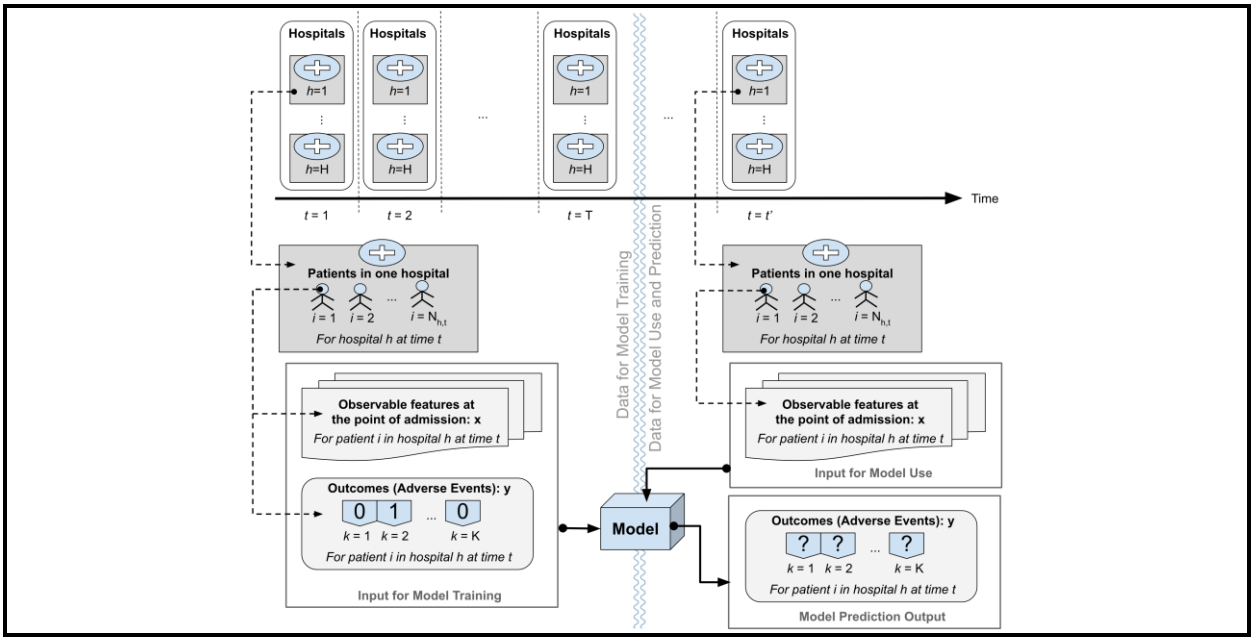
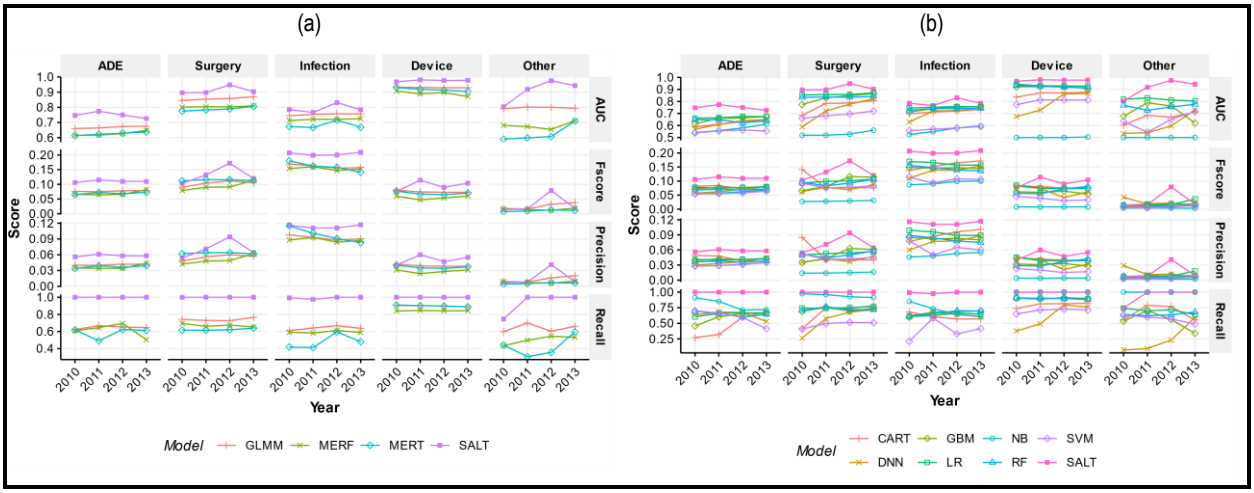


Figure 3. Results from Evaluation 1 (Panel a) and Evaluation 2 (Panel b)



Notes. The year on the x axis indicates that the models are trained using quarterly data between 2010 and the respective year. As an example, year 2012 means that that models are trained using quarterly data between 2010 and 2012, and the models are then tested using holdout observations after 2012.

Online Appendices for

First, Do No Harm: Predictive Analytics to Reduce In-Hospital Adverse Events

Appendix A: "Adverse Events" in Information Systems Literature

Table A1. Mentions of "Adverse Events" (AEs) in Leading Information Systems Journals

Authors	Outlet	Title of the Paper	Conceptualization of AEs
Abbasi et al. [1]	ISR	Don't Mention It? Analyzing User-Generated Content Signals for Early Adverse Event Warnings	The authors developed an algorithm to detect adverse drug events among drugs on the market.
Aron et al. [2]	ISR	The Impact of Automation of Systems on Medical Errors: Evidence from Field Research	The authors consider AEs as medical errors and examine how automation impacts AEs.
Bardhan et al. [4]	MISQ	Connecting Systems, Data, and People: A Multidisciplinary Research Roadmap for Chronic Disease Management	This editorial article describes the roles of IS in chronic disease management, among which AE prediction and prevention is an emergent theme.
Benaroch and Chernobai [6]	MISQ	Operational IT Failures, IT Value Destruction, and Board-Level IT Governance Changes	The authors use AEs as a generic phrase in describing IT failures in financial firms.
Clemons and Hitt [11]	JMIS	Poaching and the Misappropriation of Information: Transaction Risks of Information Exchange	The authors use AEs as a generic phrase in describing risks in transaction cost theory.
Fichman [18]	ISR	Real Options and IT Platform Adoption: Implications for Theory and Practice	The author uses AEs as a generic phrase in real options.
Fichman et al. [19]	ISR	The Role of Information Systems in Healthcare: Current Research and Future Trends	This article describes the roles of IS in healthcare, among which decision support to prevent AEs is a key area.
Hydari et al. [24]	MS	Saving Patient Ryan—Can Advanced Electronic Medical Records Make Patient Care Safer?	The authors consider AEs as patient safety events and show that the adoption of advanced electronic medical records decreases AEs.
Kohli and Tan [25]	MISQ	Electronic Health Records: How Can IS Researchers Contribute to Transforming Healthcare	In this commentary, the authors suggest that incomplete patient information can lead to AEs.
Lin et al. [28]	MISQ	Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multitask Learning Approach	The authors consider the complications in disease progression as AEs and develop an approach to predict them.
Menon and Kohli [29]	ISR	Blunting Damocles' Sword: A Longitudinal Model of Healthcare IT Impact on Malpractice Insurance Premium and Quality of Patient Care	The authors use AEs as a generic phrase in medical malpractice.
Meyer et al. [30]	ISR	A Machine Learning Approach to Improving Dynamic Decision Making	The authors consider the complications in disease progression as AEs and develop an approach to predict them.
Ozdemir et al. [34]	ISR	An Analysis of the Adoption of Digital Health Records Under Switching Costs	The authors mention the prevention of AEs in passing when describing the benefits of health IT.
Park et al. [35]	MISQ	Disaster Experience and Hospital Information Systems: An Examination of Perceived Information Assurance, Risk, Resilience, and HIS Usefulness	The authors mention the roles of AEs in passing when theorizing the perceived usefulness of health IT.
Pinsonneault et al. [37]	JMIS	Integrated Health Information Technology and the Quality of Patient Care: A Natural Experiment	The authors mention AEs in passing when theorizing the role of health IT to quality of care.
Velu et al. [44]	JMIS	Centralizing Data Management with Considerations of Uncertainty and Information-Based Flexibility	The authors mention AEs in passing when theorizing data management decision-making.
Wang et al. [47]	ISR	The Association Between the Disclosure and the Realization of Information Security Risk Factors	The authors mention AEs in passing when describing cybersecurity.

Notes. Here we summarize key published articles that mentioned "adverse events" in *Information Systems Research* (ISR), *Journal of Management Information Systems* (JMIS), *MIS Quarterly* (MISQ), and *Management Science* (MS).

Appendix B: Systematic Review of the Literature on Adverse Event Prediction

We conducted a systematic literature search by querying published literature in the PubMed and the EBSCO (with all available content providers) databases. Our literature search procedure is illustrated in Figure B1. Studies were considered relevant if they are focused on developing predictive models for adverse events (AEs).

We began our search by formulating queries that cover three sets of keywords:

1. **Context keyword set**: health(care), hospital(s), hospitalization, inpatient(s), emergency department(s), surgery, surgical, drug, prescription, device, infect(ion)
2. **AE keyword set**: adverse event(s), medical error(s), patient safety
3. **Prediction keyword set**: prediction(s), predict(s), predictive

The inclusion criterion in our database search is that the articles must contain at least one keyword from each of the above three keyword sets in their title or abstract. After we retrieve the records from both PubMed and the EBSCO databases, we then screen them for duplicates (n=606), exclude articles irrelevant to AE predictive modeling by reviewing their abstracts (n=2,493), and obtain the full texts of the remaining articles (n=15) for our literature review and synthesis. The selected articles and results of our literature synthesis are summarized in Table B1 and discussed in our main text.

Figure B1. Literature Search Procedure

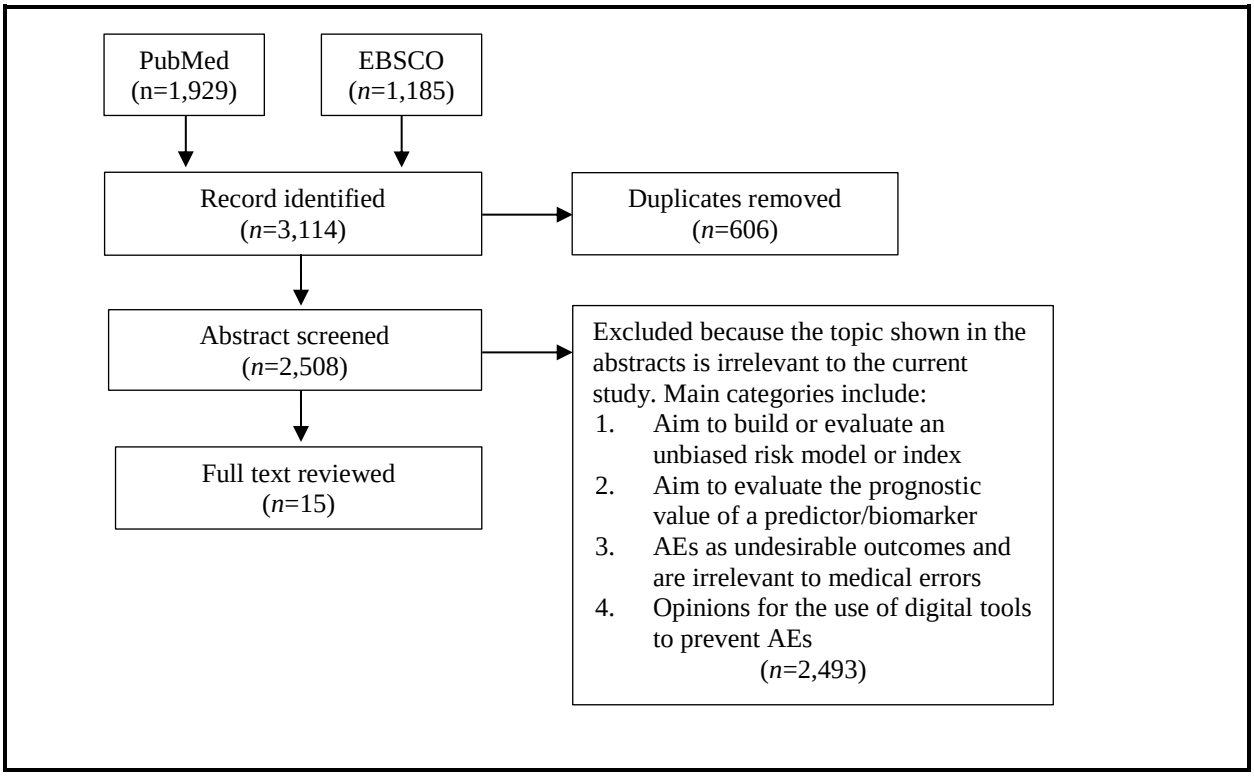


Table B1. Summary of Extant AE Prediction Literature

Authors	Types of AEs					Main Data Source	Data Unit	Methods	Feature Count	Feature Engineering	Predictive Evaluation	AUC
	ADE	Surgery	Infection	Device	Others							
Mufti et al. [33]		V				Clinical registry	5,584 patients	ANN, BN, NB, RF, DT	22	N	10-fold CV	0.77
Buchlak et al. [8]		V				Patient's clinical records	136 patients	GLM	7	N	Not applied	0.71
Ehlers et al. [15]		V				Claims database	410,521 patients	NB	300+	N	10-fold CV	0.79
Ferdousi et al. [17]	V					Drug's biological functions	45,530 drug pairs	SR	12	N	12 holdouts	NA
Genovese et al. [21]		V				Clinical registry	52,562 patients	GLMM	10	N	15% holdout	0.82
Mortazavi et al. [32]		V	V			Patient's clinical records	5,214 patients	GBM, GLM, RF	9828	Y (wrapper and filter methods)	5-fold CV	0.81–0.83
Ratliff et al. [38]		V				Claims database	279,135 patients	BT, GLM, LR, CART	NA	N	20% holdout	0.70
Wang et al. [46]	V					Drugs' genomic expressions	293 drug-reactions	SR	1	N	Not applied	NA
Dodson et al. [13]		V				Clinical registry	240,632 procedures	GLM	21	Y (expert opinions)	30% holdout	0.72
Steeg et al. [43]	NA					Clinical documents	6,096 patients	GLM	9	N	Before/after	0.59
Kusy et al. [27]		V				Patients' genomic records	107 patients	ANN, GEP	10	N	30% holdout	0.82
Krone et al. [26]		V				Clinical registry	41,071 patients	GLM	7	N	Not applied	0.69
Schwilk et al. [40]		V				Patients' clinical records	26,907 procedures	GLM	17	Y (expert opinions)	50% holdout	NA
Geraci et al. [22]		V				Patients' clinical records	2,213 patients	GLM	11	N	50% holdout	0.64
Rosen et al. [39]		V				Patients' clinical records	8,126 patients	GLM	5-10	N	50% holdout	0.64–0.75

Notes. ANN = artificial neural network, AUC = area under the curve, BT = boosting, CART = classification and regression trees, CV = cross-validation, DT = decision tree, GBM = generalized boosted model, GEP = gene expression programming, GLM = generalized linear model, GLMM = generalized linear mixed model, LR = lasso regression, NA = not available, NB = naïve Bayes, RF = random forest, SR = similarity ranking

Appendix C: Detailed Data Description and Statistics

This appendix reports two data-related details related to our study. In Table C1, we list diagnosis codes from the International Classification of Diseases, Ninth Revision (ICD-9) which we use to detect in-hospital adverse events (AEs). As noted in our main text, these diagnosis codes were adapted from the Utah–Missouri Adverse Event Classification (UMAEC), which were originally compiled by a group of clinical experts. We refine the UMAEC list by retaining only AE codes that have a high positive predictive value. In Table C2, we report the list of variables used in our AE predictive modeling and their descriptive statistics. There are 5 outcome variables (i.e., y), one for each AE category. We have a total of 158 predictors (i.e., $\boldsymbol{x}$) but most of them (135) are hospital dummies.

Table C1. Diagnosis Codes related to In-Hospital Adverse Events

AE Category	AE Code	Code Name
ADE	E850.0	Accidental poisoning by heroin
ADE	E850.1	Accidental poisoning by methadone
ADE	E850.2	Accidental poisoning by other opiates and related narcotics
ADE	E850.3	Accidental poisoning by salicylates
ADE	E850.4	Accidental poisoning by aromatic analgesics, not elsewhere classified
ADE	E850.5	Accidental poisoning by pyrazole derivatives
ADE	E850.6	Accidental poisoning by antirheumatics (antiphlogistics)
ADE	E850.7	Accidental poisoning by other non-narcotic analgesics
ADE	E850.8	Accidental poisoning by other specified analgesics and antipyretics
ADE	E850.9	Accidental poisoning by unspecified analgesic or antipyretic
ADE	E851	Accidental poisoning by barbiturates
ADE	E852.0	Accidental poisoning by chloral hydrate group
ADE	E852.1	Accidental poisoning by paraldehyde
ADE	E852.2	Accidental poisoning by bromine compounds
ADE	E852.3	Accidental poisoning by methaqualone compounds
ADE	E852.4	Accidental poisoning by glutethimide group
ADE	E852.5	Accidental poisoning by mixed sedatives, not elsewhere classified
ADE	E852.8	Accidental poisoning by other specified sedative and hypnotic
ADE	E852.9	Accidental poisoning by unspecified sedative or hypnotic
ADE	E853.0	Accidental poisoning by phenothiazine-based tranquilizers

ADE	E853.1	Accidental poisoning by butyrophenone-based tranquilizers
ADE	E853.2	Accidental poisoning by benzodiazepine-based tranquilizers
ADE	E853.8	Accidental poisoning by other specified tranquilizers
ADE	E853.9	Accidental poisoning by unspecified tranquilizer
ADE	E854.0	Accidental poisoning by antidepressants
ADE	E854.1	Accidental poisoning by psychodysleptics (hallucinogens)
ADE	E854.2	Accidental poisoning by psychostimulants
ADE	E854.3	Accidental poisoning by central nervous system stimulants
ADE	E854.8	Accidental poisoning by other psychotropic agents
ADE	E855.0	Accidental poisoning by anticonvulsant and anti-Parkinsonism drugs
ADE	E855.1	Accidental poisoning by other central nervous system depressants
ADE	E855.2	Accidental poisoning by local anesthetics
ADE	E855.3	Accidental poisoning by parasympathomimetics (cholinergics)
ADE	E855.4	Accidental poisoning by parasympatholytics (anticholinergics and antimuscarinics) and spasmolytics
ADE	E855.5	Accidental poisoning by sympathomimetics (adrenergics)
ADE	E855.6	Accidental poisoning by sympatholytics (antiadrenergics)
ADE	E855.8	Accidental poisoning by other specified drugs acting on central and autonomic nervous systems
ADE	E855.9	Accidental poisoning by unspecified drug acting on central and autonomic nervous systems
ADE	E856	Accidental poisoning by antibiotics
ADE	E857	Accidental poisoning by other anti-infectives
ADE	E858.0	Accidental poisoning by hormones and synthetic substitutes
ADE	E858.1	Accidental poisoning by primarily systemic agents
ADE	E858.2	Accidental poisoning by agents primarily affecting blood constituents
ADE	E858.3	Accidental poisoning by agents primarily affecting cardiovascular system
ADE	E858.4	Accidental poisoning by agents primarily affecting gastrointestinal system
ADE	E858.5	Accidental poisoning by water, mineral, and uric acid metabolism drugs
ADE	E858.6	Accidental poisoning by agents primarily acting on the smooth and skeletal muscles and respiratory system
ADE	E858.7	Accidental poisoning by agents primarily affecting skin and mucous membrane, ophthalmological, otorhinolaryngological, and dental drugs
ADE	E858.8	Accidental poisoning by other specified drugs
ADE	E858.9	Accidental poisoning by unspecified drug
ADE	E929.2	Late effects of accidental poisoning
ADE	E930.0	Penicillins causing adverse effect in therapeutic use
ADE	E930.1	Antifungal antibiotics causing adverse effect in therapeutic use
ADE	E930.2	Chloramphenicol group causing adverse effect in therapeutic use
ADE	E930.3	Erythromycin and other macrolides causing adverse effect in therapeutic use
ADE	E930.4	Tetracycline group causing adverse effect in therapeutic use
ADE	E930.5	Cephalosporin group causing adverse effect in therapeutic use
ADE	E930.6	Antimycobacterial antibiotics causing adverse effect in therapeutic use
ADE	E930.7	Antineoplastic antibiotics causing adverse effect in therapeutic use
ADE	E930.8	Other specified antibiotics causing adverse effect in therapeutic use
ADE	E930.9	Unspecified antibiotic causing adverse effect in therapeutic use

ADE	E931.0	Sulfonamides causing adverse effect in therapeutic use
ADE	E931.1	Arsenical anti-infectives causing adverse effect in therapeutic use
ADE	E931.2	Heavy metal anti-infectives causing adverse effect in therapeutic use
ADE	E931.3	Quinoline and hydroxyquinoline derivatives causing adverse effect in therapeutic use
ADE	E931.4	Antimalarials and drugs acting on other blood protozoa causing adverse effect in therapeutic use
ADE	E931.5	Other antiprotozoal drugs causing adverse effect in therapeutic use
ADE	E931.6	Anthelmintics causing adverse effect in therapeutic use
ADE	E931.7	Antiviral drugs causing adverse effect in therapeutic use
ADE	E931.8	Other antimycobacterial drugs causing adverse effect in therapeutic use
ADE	E931.9	Other and unspecified anti-infectives causing adverse effect in therapeutic use
ADE	E932.0	Adrenal cortical steroids causing adverse effect in therapeutic use
ADE	E932.1	Androgens and anabolic congeners causing adverse effect in therapeutic use
ADE	E932.2	Ovarian hormones and synthetic substitutes causing adverse effect in therapeutic use
ADE	E932.3	Insulins and antidiabetic agents causing adverse effect in therapeutic use
ADE	E932.4	Anterior pituitary hormones causing adverse effect in therapeutic use
ADE	E932.5	Posterior pituitary hormones causing adverse effect in therapeutic use
ADE	E932.6	Parathyroid and parathyroid derivatives causing adverse effect in therapeutic use
ADE	E932.7	Thyroid and thyroid derivatives causing adverse effect in therapeutic use
ADE	E932.8	Antithyroid agents causing adverse effect in therapeutic use
ADE	E932.9	Other and unspecified hormones and synthetic substitutes causing adverse effect in therapeutic use
ADE	E933.0	Antiallergic and antiemetic drugs causing adverse effect in therapeutic use
ADE	E933.1	Antineoplastic and immunosuppressive drugs causing adverse effect in therapeutic use
ADE	E933.2	Acidifying agents causing adverse effect in therapeutic use
ADE	E933.3	Alkalizing agents causing adverse effect in therapeutic use
ADE	E933.4	Enzymes, not elsewhere classified, causing adverse effect in therapeutic use
ADE	E933.5	Vitamins, not elsewhere classified, causing adverse effect in therapeutic use
ADE	E933.8	Other systemic agents, not elsewhere classified, causing adverse effect in therapeutic use
ADE	E933.9	Unspecified systemic agent causing adverse effect in therapeutic use
ADE	E934.0	Iron and its compounds causing adverse effect in therapeutic use
ADE	E934.1	Liver preparations and other antianemic agents causing adverse effect in therapeutic use
ADE	E934.2	Anticoagulants causing adverse effect in therapeutic use
ADE	E934.3	Vitamin K (phytonadione) causing adverse effect in therapeutic use
ADE	E934.4	Fibrinolysis-affecting drugs causing adverse effect in therapeutic use
ADE	E934.5	Anticoagulant antagonists and other coagulants causing adverse effect in therapeutic use
ADE	E934.6	Gamma globulin causing adverse effect in therapeutic use
ADE	E934.7	Natural blood and blood products causing adverse effect in therapeutic use
ADE	E934.8	Other agents affecting blood constituents causing adverse effect in therapeutic use
ADE	E934.9	Unspecified agent affecting blood constituents causing adverse effect in therapeutic use
ADE	E935.0	Heroin causing adverse effect in therapeutic use
ADE	E935.1	Methadone causing adverse effect in therapeutic use
ADE	E935.2	Other opiates and related narcotics causing adverse effect in therapeutic use
ADE	E935.3	Salicylates causing adverse effect in therapeutic use

ADE	E935.4	Aromatic analgesics, not elsewhere classified, causing adverse effect in therapeutic use
ADE	E935.5	Pyrazole derivatives causing adverse effect in therapeutic use
ADE	E935.6	Antirheumatics (antiphlogistics) causing adverse effect in therapeutic use
ADE	E935.7	Other non-narcotic analgesics causing adverse effect in therapeutic use
ADE	E935.8	Other specified analgesics and antipyretics causing adverse effect in therapeutic use
ADE	E935.9	Unspecified analgesic and antipyretic causing adverse effect in therapeutic use
ADE	E936.0	Oxazolidine derivatives causing adverse effect in therapeutic use
ADE	E936.1	Hydantoin derivatives causing adverse effect in therapeutic use
ADE	E936.2	Succinimides causing adverse effect in therapeutic use
ADE	E936.3	Other and unspecified anticonvulsants causing adverse effect in therapeutic use
ADE	E936.4	Anti-parkinsonism drugs causing adverse effect in therapeutic use
ADE	E937.0	Barbiturates causing adverse effect in therapeutic use
ADE	E937.1	Chloral hydrate group causing adverse effect in therapeutic use
ADE	E937.2	Paraldehyde causing adverse effect in therapeutic use
ADE	E937.3	Bromine compounds causing adverse effect in therapeutic use
ADE	E937.4	Methaqualone compounds causing adverse effect in therapeutic use
ADE	E937.5	Glutethimide group causing adverse effect in therapeutic use
ADE	E937.6	Mixed sedatives, not elsewhere classified, causing adverse effect in therapeutic use
ADE	E937.8	Other sedatives and hypnotics causing adverse effect in therapeutic use
ADE	E937.9	Unspecified sedatives and hypnotics causing adverse effect in therapeutic use
ADE	E938.0	Central nervous system muscle-tone depressants causing adverse effect in therapeutic use
ADE	E938.1	Halothane causing adverse effect in therapeutic use
ADE	E938.2	Other gaseous anesthetics causing adverse effect in therapeutic use
ADE	E938.3	Intravenous anesthetics causing adverse effect in therapeutic use
ADE	E938.4	Other and unspecified general anesthetics causing adverse effect in therapeutic use
ADE	E938.5	Surface and infiltration anesthetics causing adverse effect in therapeutic use
ADE	E938.6	Peripheral nerve- and plexus-blocking anesthetics causing adverse effect in therapeutic use
ADE	E938.7	Spinal anesthetics causing adverse effect in therapeutic use
ADE	E938.9	Other and unspecified local anesthetics causing adverse effect in therapeutic use
ADE	E939.0	Antidepressants causing adverse effect in therapeutic use
ADE	E939.1	Phenothiazine-based tranquilizers causing adverse effect in therapeutic use
ADE	E939.2	Butyrophenone-based tranquilizers causing adverse effect in therapeutic use
ADE	E939.3	Other antipsychotics, neuroleptics, and major tranquilizers causing adverse effect in therapeutic use
ADE	E939.4	Benzodiazepine-based tranquilizers causing adverse effect in therapeutic use
ADE	E939.5	Other tranquilizers causing adverse effect in therapeutic use
ADE	E939.6	Psychodysleptics (hallucinogens) causing adverse effect in therapeutic use
ADE	E939.7	Psychostimulants causing adverse effect in therapeutic use
ADE	E939.8	Other psychotropic agents causing adverse effect in therapeutic use
ADE	E939.9	Unspecified psychotropic agent causing adverse effect in therapeutic use
ADE	E940.0	Analeptics causing adverse effect in therapeutic use
ADE	E940.1	Opiate antagonists causing adverse effect in therapeutic use
ADE	E940.8	Other specified central nervous system stimulants causing adverse effect in therapeutic use

ADE	E940.9	Unspecified central nervous system stimulant causing adverse effect in therapeutic use
ADE	E941.0	Parasympathomimetics (cholinergics) causing adverse effect in therapeutic use
ADE	E941.1	Parasympatholytics (anticholinergics and antimuscarinics) and spasmolytics causing adverse effect in therapeutic use
ADE	E941.2	Sympathomimetics (adrenergics) causing adverse effect in therapeutic use
ADE	E941.3	Sympatholytics (antiadrenergics) causing adverse effect in therapeutic use
ADE	E941.9	Unspecified drug primarily affecting the autonomic nervous system causing adverse effect in therapeutic use
ADE	E942.0	Cardiac rhythm regulators causing adverse effect in therapeutic use
ADE	E942.1	Cardiotonic glycosides and drugs of similar action causing adverse effect in therapeutic use
ADE	E942.2	Antilipemic and antiarteriosclerotic drugs causing adverse effect in therapeutic use
ADE	E942.3	Ganglion-blocking agents causing adverse effect in therapeutic use
ADE	E942.4	Coronary vasodilators causing adverse effect in therapeutic use
ADE	E942.5	Other vasodilators causing adverse effect in therapeutic use
ADE	E942.6	Other antihypertensive agents causing adverse effect in therapeutic use
ADE	E942.7	Antivaricose drugs, including sclerosing agents, causing adverse effect in therapeutic use
ADE	E942.8	Capillary-active drugs causing adverse effect in therapeutic use
ADE	E942.9	Other and unspecified agents primarily affecting the cardiovascular system causing adverse effect in therapeutic use
ADE	E943.0	Antacids and antigastric secretion drugs causing adverse effect in therapeutic use
ADE	E943.1	Irritant cathartics causing adverse effect in therapeutic use
ADE	E943.2	Emollient cathartics causing adverse effect in therapeutic use
ADE	E943.3	Other cathartics, including intestinal atonia drugs, causing adverse effect in therapeutic use
ADE	E943.4	Digestants causing adverse effect in therapeutic use
ADE	E943.5	Antidiarrheal drugs causing adverse effect in therapeutic use
ADE	E943.6	Emetics causing adverse effect in therapeutic use
ADE	E943.8	Other specified agents primarily affecting the gastrointestinal system causing adverse effect in therapeutic use
ADE	E943.9	Unspecified agent primarily affecting the gastrointestinal system causing adverse effect in therapeutic use
ADE	E944.0	Mercurial diuretics causing adverse effect in therapeutic use
ADE	E944.1	Purine derivative diuretics causing adverse effect in therapeutic use
ADE	E944.2	Carbonic acid anhydrase inhibitors causing adverse effect in therapeutic use
ADE	E944.3	Saluretics causing adverse effect in therapeutic use
ADE	E944.4	Other diuretics causing adverse effect in therapeutic use
ADE	E944.5	Electolytic, caloric, and water-balance agents causing adverse effect in therapeutic use
ADE	E944.6	Other mineral salts, not elsewhere classified, causing adverse effect in therapeutic use
ADE	E944.7	Uric acid metabolism drugs causing adverse effect in therapeutic use
ADE	E945.0	Oxytocic agents causing adverse effect in therapeutic use
ADE	E945.1	Smooth muscle relaxants causing adverse effect in therapeutic use
ADE	E945.2	Skeletal muscle relaxants causing adverse effect in therapeutic use
ADE	E945.3	Other and unspecified drugs acting on muscles causing adverse effect in therapeutic use
ADE	E945.4	Antitussives causing adverse effect in therapeutic use
ADE	E945.5	Expectorants causing adverse effect in therapeutic use

ADE	E945.6	Anti-common cold drugs causing adverse effect in therapeutic use
ADE	E945.8	Other and unspecified respiratory drugs causing adverse effect in therapeutic use
ADE	E946.0	Local anti-infectives and anti-inflammatory drugs causing adverse effect in therapeutic use
ADE	E946.1	Antipruritics causing adverse effect in therapeutic use
ADE	E946.2	Local astringents and local detergents causing adverse effect in therapeutic use
ADE	E946.3	Emollients, demulcents, and protectants causing adverse effect in therapeutic use
ADE	E946.4	Keratolytics, keratoplastics, other hair treatment drugs and preparations causing adverse effect in therapeutic use
ADE	E946.5	Eye anti-infectives and other eye drugs causing adverse effect in therapeutic use
ADE	E946.6	Anti-infectives and other drugs and preparations for ear, nose, and throat causing adverse effect in therapeutic use
ADE	E946.7	Dental drugs topically applied causing adverse effect in therapeutic use
ADE	E946.8	Other agents primarily affecting skin and mucous membrane causing adverse effect in therapeutic use
ADE	E946.9	Unspecified agent primarily affecting skin and mucous membrane causing adverse effect in therapeutic use
ADE	E947.0	Dietetics causing adverse effect in therapeutic use
ADE	E947.1	Lipotropic drugs causing adverse effect in therapeutic use
ADE	E947.2	Antidotes and chelating agents, not elsewhere classified, causing adverse effect in therapeutic use
ADE	E947.3	Alcohol deterrents causing adverse effect in therapeutic use
ADE	E947.4	Pharmaceutical excipients causing adverse effect in therapeutic use
ADE	E947.8	Other drugs and medicinal substances causing adverse effect in therapeutic use
ADE	E947.9	Unspecified drug or medicinal substance causing adverse effect in therapeutic use
ADE	E948.0	Bcg vaccine causing adverse effect in therapeutic use
ADE	E948.1	Typhoid and paratyphoid vaccines causing adverse effect in therapeutic use
ADE	E948.2	Cholera vaccine causing adverse effect in therapeutic use
ADE	E948.3	Plague vaccine causing adverse effect in therapeutic use
ADE	E948.4	Tetanus vaccine causing adverse effect in therapeutic use
ADE	E948.5	Diphtheria vaccine causing adverse effect in therapeutic use
ADE	E948.6	Pertussis vaccine, including combinations with pertussis component, causing adverse effect in therapeutic use
ADE	E948.8	Other and unspecified bacterial vaccines causing adverse effect in therapeutic use
ADE	E948.9	Mixed bacterial vaccines, except combinations with pertussis component, causing adverse effect in therapeutic use
ADE	E949.0	Smallpox vaccine causing adverse effect in therapeutic use
ADE	E949.1	Rabies vaccine causing adverse effect in therapeutic use
ADE	E949.2	Typhus vaccine causing adverse effect in therapeutic use
ADE	E949.3	Yellow fever vaccine causing adverse effect in therapeutic use
ADE	E949.4	Measles vaccine causing adverse effect in therapeutic use
ADE	E949.5	Poliomyelitis vaccine causing adverse effect in therapeutic use
ADE	E949.6	Other and unspecified viral and rickettsial vaccines causing adverse effect in therapeutic use
ADE	E949.7	Mixed viral-rickettsial and bacterial vaccines, except combinations with pertussis component, causing adverse effect in therapeutic use

ADE	E949.9	Other and unspecified vaccines and biological substances causing adverse effect in therapeutic use
Surgery	998	Postoperative shock, not elsewhere classified
Surgery	998.2	Accidental puncture or laceration during procedure
Surgery	998.3	Disruption of operation wound
Surgery	998.4	Foreign body accidentally left during procedure, not elsewhere classified
Surgery	998.51	Infected postoperative seroma
Surgery	998.59	Other postoperative infection
Surgery	998.6	Persistent postoperative fistula, not elsewhere classified
Surgery	998.7	Acute reaction to foreign substance accidentally left during procedure, not elsewhere classified
Surgery	998.81	Emphysema (subcutaneous) (surgical) resulting from a procedure
Surgery	998.83	Non-healing surgical wound
Surgery	998.9	Unspecified complication of procedure, not elsewhere classified
Surgery	E870.0	Accidental cut, puncture, perforation, or hemorrhage during surgical operation
Surgery	E870.1	Accidental cut, puncture, perforation, or hemorrhage during infusion or transfusion
Surgery	E870.2	Accidental cut, puncture, perforation, or hemorrhage during kidney dialysis or other perfusion
Surgery	E870.3	Accidental cut, puncture, perforation, or hemorrhage during injection or vaccination
Surgery	E870.4	Accidental cut, puncture, perforation, or hemorrhage during endoscopic examination
Surgery	E870.5	Accidental cut, puncture, perforation, or hemorrhage during aspiration of fluid or tissue, puncture, and catheterization
Surgery	E870.6	Accidental cut, puncture, perforation, or hemorrhage during heart catheterization
Surgery	E870.7	Accidental cut, puncture, perforation, or hemorrhage during administration of enema
Surgery	E870.8	Accidental cut, puncture, perforation, or hemorrhage during other specified medical care
Surgery	E870.9	Accidental cut, puncture, perforation, or hemorrhage during unspecified medical care
Surgery	E871.0	Foreign object left in body during surgical operation
Surgery	E871.1	Foreign object left in body during infusion or transfusion
Surgery	E871.2	Foreign object left in body during kidney dialysis or other perfusion
Surgery	E871.3	Foreign object left in body during injection or vaccination
Surgery	E871.4	Foreign object left in body during endoscopic examination
Surgery	E871.5	Foreign object left in body during aspiration of fluid or tissue, puncture, and catheterization
Surgery	E871.6	Foreign object left in body during heart catheterization
Surgery	E871.7	Foreign object left in body during removal of catheter or packing
Surgery	E871.8	Foreign object left in body during other specified procedure
Surgery	E871.9	Foreign object left in body during unspecified procedure
Surgery	E872.0	Failure of sterile precautions during surgical operation
Surgery	E872.1	Failure of sterile precautions during infusion or transfusion
Surgery	E872.2	Failure of sterile precautions during kidney dialysis and other perfusion
Surgery	E872.3	Failure of sterile precautions during injection or vaccination
Surgery	E872.4	Failure of sterile precautions during endoscopic examination
Surgery	E872.5	Failure of sterile precautions during aspiration of fluid or tissue, puncture, and catheterization
Surgery	E872.6	Failure of sterile precautions during heart catheterization
Surgery	E872.8	Failure of sterile precautions during other specified procedure
Surgery	E872.9	Failure of sterile precautions during unspecified procedure

Surgery	E873.0	Excessive amount of blood or other fluid during transfusion or infusion
Surgery	E873.1	Incorrect dilution of fluid during infusion
Surgery	E876.0	Mismatched blood in transfusion
Surgery	E876.2	Failure in suture and ligature during surgical operation
Surgery	E876.3	Endotracheal tube wrongly placed during anesthetic procedure
Surgery	E876.4	Failure to introduce or to remove other tube or instrument
Surgery	E876.5	Performance of inappropriate operation
Surgery	E878.0	Surgical operation with transplant of whole organ causing abnormal patient reaction, or later complication, without mention of misadventure at time of operation
Surgery	E878.3	Surgical operation with formation of external stoma causing abnormal patient reaction, or later complication, without mention of misadventure at time of operation
Surgery	E878.4	Other restorative surgery causing abnormal patient reaction, or later complication, without mention of misadventure at time of operation
Surgery	E878.5	Amputation of limb(s) causing abnormal patient reaction, or later complication, without mention of misadventure at time of operation
Surgery	E878.6	Removal of other organ (partial) (total) causing abnormal patient reaction, or later complication, without mention of misadventure at time of operation
Surgery	E878.8	Other specified surgical operation and procedure causing abnormal patient reaction, or later complication, without mention of misadventure at time of operation
Surgery	E878.9	Unspecified surgical operation and procedure causing abnormal patient reaction, or later complication, without mention of misadventure at time of operation
Surgery	E879.0	Cardiac catheterization as the cause of abnormal reaction of patient, or of later complication, without mention of misadventure at time of procedure
Surgery	E879.1	Kidney dialysis as the cause of abnormal reaction of patient, or of later complication, without mention of misadventure at time of procedure
Surgery	E879.2	Radiological procedure and radiotherapy as the cause of abnormal reaction of patient, or of later complication, without mention of misadventure at time of procedure
Surgery	E879.3	Shock therapy as the cause of abnormal reaction of patient, or of later complication, without mention of misadventure at time of procedure
Surgery	E879.4	Aspiration of fluid as the cause of abnormal reaction of patient, or of later complication, without mention of misadventure at time of procedure
Surgery	E879.5	Insertion of gastric or duodenal sound as the cause of abnormal reaction of patient, or of later complication, without mention of misadventure of time of procedure
Surgery	E879.6	Urinary catheterization as the cause of abnormal reaction of patient, or of later complication, without mention of misadventure at time of procedure
Surgery	E879.7	Blood sampling as the cause of abnormal reaction of patient, or of later complication, without mention of misadventure at time of procedure
Surgery	E879.8	Other specified procedure as the cause of abnormal reaction of patient, or of later complication, without mention of misadventure at time of procedure
Surgery	E879.9	Unspecified procedure as the cause of abnormal reaction of patient, or of later complication, without mention of misadventure at time of procedure
Infections	038	Streptococcal septicemia
Infections	038.1	Unspecified staphylococcal septicemia
Infections	038.11	Staphylococcus aureus septicemia
Infections	038.19	Other staphylococcal septicemia
Infections	038.3	Septicemia due to anaerobes
Infections	038.4	Septicemia due to unspecified gram-negative organism
Infections	038.41	Septicemia due to hemophilus influenzae (H. influenzae)

Infections	038.42	Septicemia due to Escherichia coli (E. coli)
Infections	038.43	Septicemia due to pseudomonas
Infections	038.44	Septicemia due to serratia
Infections	038.49	Other septicemia due to gram-negative organism
Infections	038.8	Other specified septicemia
Infections	038.9	Unspecified septicemia
Infections	421	Acute and subacute bacterial endocarditis
Infections	421.1	Acute and subacute infective endocarditis in diseases classified elsewhere
Infections	421.9	Unspecified acute endocarditis
Infections	424.99	Other endocarditis, valve unspecified
Infections	481	Pneumococcal pneumonia (streptococcus pneumoniae pneumonia)
Infections	482	Pneumonia due to Klebsiella pneumoniae
Infections	482.1	Pneumonia due to Pseudomonas
Infections	482.2	Pneumonia due to Hemophilus influenzae (H. influenzae)
Infections	482.3	Pneumonia due to unspecified Streptococcus
Infections	482.31	Pneumonia due to Streptococcus, group A
Infections	482.32	Pneumonia due to Streptococcus, group B
Infections	482.39	Pneumonia due to other Streptococcus
Infections	482.4	Pneumonia due to Staphylococcus, unspecified
Infections	482.41	Pneumonia due to Staphylococcus aureus
Infections	482.49	Other Staphylococcus pneumonia
Infections	482.81	Pneumonia due to anaerobes
Infections	482.82	Pneumonia due to escherichia coli (E. coli)
Infections	482.83	Pneumonia due to other gram-negative bacteria
Infections	482.84	Legionnaires' disease
Infections	482.89	Pneumonia due to other specified bacteria
Infections	482.9	Unspecified bacterial pneumonia
Infections	483.8	Pneumonia due to other specified organism
Infections	485	Bronchopneumonia, organism unspecified
Infections	486	Pneumonia, organism unspecified
Infections	519.01	Infection of tracheostomy
Infections	536.41	Infection of gastrostomy
Infections	569.61	Infection of colostomy or enterostomy
Infections	590.8	Unspecified pyelonephritis
Infections	590.9	Unspecified infection of kidney
Infections	595	Acute cystitis
Infections	595.9	Unspecified cystitis
Infections	599	Urinary tract infection, site not specified
Infections	670	Major puerperal infection, unspecified as to episode of care
Infections	670.02	Major puerperal infection, delivered, with mention of postpartum complication
Infections	670.04	Major puerperal infection, postpartum
Infections	682.3	Cellulitis and abscess of upper arm and forearm
Infections	682.4	Cellulitis and abscess of hand, except fingers and thumb

Infections	790.7	Bacteremia
Infections	958.3	Posttraumatic wound infection not elsewhere classified
Device	349.1	Nervous system complications from surgically implanted device
Device	519.02	Mechanical complication of tracheostomy
Device	536.42	Mechanical complication of gastrostomy
Device	569.62	Mechanical complication of colostomy and enterostomy
Device	E874.0	Mechanical failure of instrument or apparatus during surgical operation
Device	E874.1	Mechanical failure of instrument or apparatus during infusion and transfusion
Device	E874.2	Mechanical failure of instrument or apparatus during kidney dialysis and other perfusion
Device	E874.3	Mechanical failure of instrument or apparatus during endoscopic examination
Device	E874.4	Mechanical failure of instrument or apparatus during aspiration of fluid or tissue, puncture, and catheterization
Device	E874.5	Mechanical failure of instrument or apparatus during heart catheterization
Device	E874.8	Mechanical failure of instrument or apparatus during other specified procedure
Device	E874.9	Mechanical failure of instrument or apparatus during unspecified procedure
Device	E878.1	Surgical operation with implant of artificial internal device causing abnormal patient reaction or later complication without mention of misadventure at time of operation
Device	E878.2	Surgical operation with anastomosis, bypass or graph with natural or artificial tissues used as implant causing abnormal patient reaction or later complication without mention of misadventure at time of operation
Other	668.01	Pulmonary complications of anesthesia or other sedation in labor and delivery, delivered, with or without mention of antepartum condition
Other	668.04	Pulmonary complications of anesthesia or other sedation in labor and delivery, postpartum condition or complication
Other	668.21	Central nervous system complications of anesthesia or other sedation in labor and delivery, delivered, with or without mention of antepartum condition
Other	668.22	Central nervous system complications of anesthesia or other sedation in labor and delivery, delivered, with mention of postpartum complication
Other	668.81	Other complications of anesthesia or other sedation in labor and delivery, delivered, with or without mention of antepartum condition
Other	668.82	Other complications of anesthesia or other sedation in labor and delivery, delivered, with mention of postpartum complication
Other	669.41	Other complications of obstetrical surgery and procedures, delivered, with or without mention of antepartum condition
Other	669.42	Other complications of obstetrical surgery and procedures, delivered, with mention of postpartum complication
Other	673.22	Obstetrical blood-clot embolism, delivered, with mention of postpartum complication
Other	673.24	Obstetrical blood-clot embolism, postpartum condition or complication
Other	674.12	Disruption of cesarean wound, delivered, with mention of postpartum complication
Other	674.14	Disruption of cesarean wound, postpartum condition or complication
Other	674.32	Other complications of obstetrical surgical wounds, delivered, with mention of postpartum complication
Other	674.34	Other complications of obstetrical surgical wounds, postpartum condition or complication
Other	997.02	Iatrogenic cerebrovascular infarction or hemorrhage
Other	999.2	Other vascular complications of medical care, not elsewhere classified
Other	999.3	Other infection due to medical care not elsewhere classified

Other	999.5	Other serum reaction not elsewhere classified
Other	999.8	Other and unspecified transfusion reaction not elsewhere classified
Other	999.9	Other and unspecified complications of medical care, not elsewhere classified
Other	E873.2	Overdose of radiation in therapy
Other	E873.3	Inadvertent exposure of patient to radiation during medical care
Other	E873.4	Failure in dosage in electroshock or insulin-shock therapy
Other	E873.5	Inappropriate (too hot or too cold) temperature in local application and packing
Other	E873.6	Nonadministration of necessary drug or medicinal substance
Other	E873.8	Other specified failure in dosage
Other	E873.9	Unspecified failure in dosage
Other	E875.0	Contaminated substance transfused or infused
Other	E875.1	Contaminated substance injected or used for vaccination
Other	E875.2	Contaminated drug or biological substance administered by other means
Other	E875.8	Other contamination of patient during medical care
Other	E875.9	Unspecified contamination of patient during medical care
Other	E876.1	Wrong fluid in infusion
Other	E876.8	Other specified misadventure during medical care
Other	E876.9	Unspecified misadventure during medical care

Table C2. Data Statistics

Variable Category	Variables and Descriptive Statistics	HF Testbed (n=152,878)
Outcome dummies (i.e., the dependent variables)	**Y1: AEs related to ADEs (binary)**	
	Mean (SD)	0.00992 (0.0991)
	Median [Min, Max]	0.00 [0.00, 1.00]
	Y2: AEs related to Surgeries (binary)	
	Mean (SD)	0.00911 (0.0950)
	Median [Min, Max]	0.00 [0.00, 1.00]
	Y3: AEs related to Infections (binary)	
	Mean (SD)	0.0221 (0.147)
	Median [Min, Max]	0.00 [0.00, 1.00]
	Y4: AEs Related to Devices (binary)	
	Mean (SD)	0.00114 (0.0338)
	Median [Min, Max]	0.00 [0.00, 1.00]
	Y5: Other AEs (binary)	
	Mean (SD)	0.00191 (0.0436)
	Median [Min, Max]	0.00 [0.00, 1.00]
Admission	**X1. Year (numerical)**	
	2010	16094 (19.8%)
	2011	16434 (20.2%)
	2012	16278 (20.0%)
	2013	16244 (20.0%)
	2014	16263 (20.0%)
	X2. Quarter (numerical)	
	1	20768 (25.5%)
	2	20195 (24.8%)
	3	20300 (25.0%)
	4	20050 (24.7%)
	X3. Quarter Relative to 2010Q1 (numerical)	
	Mean (SD)	9.96 (5.82)
	Median [Min, Max]	10 [1, 20]
	X4. Priority of Admission (categorical)	
	Elective	3460 (4.3%)
	Emergency	71818 (88.3%)
	Urgent	6035 (7.4%)
	X5. Source of Origin for Admission (categorical)	
	Clinic	2818 (3.5%)
	Emergency Department	8964 (11.0%)
	Non-Health Care Facility	67331 (82.8%)
	Others	296 (0.4%)
	Transfer	1904 (2.3%)

	X6. Admission during Weekday (binary)	
	FALSE	19297 (23.7%)
	TRUE	62016 (76.3%)
	X7. Admission for Elective Surgery (binary)	
	FALSE	69131 (85.0%)
	TRUE	12182 (15.0%)
Patient	**X8. Gender (binary)**	
	Female	38583 (47.4%)
	Male	42730 (52.6%)
	X9. Race (categorical)	
	Asian	317 (0.4%)
	Black	25224 (31.0%)
	Hispanic	12094 (14.9%)
	Other	1565 (1.9%)
	White	42113 (51.8%)
	X10. Age (numerical)	
	Mean (SD)	51.9 (19.3)
	Median [Min, Max]	53.0 [0.00, 111]
	X11. Number of Chronic Conditions (numerical)	
	Mean (SD)	3.21 (1.69)
	Median [Min, Max]	3.00 [1.00, 12.0]
	X12. Charlson Comorbidity Index (numerical)	
	Mean (SD)	3.08 (2.46)
	Median [Min, Max]	3.00 [1.00, 20.0]
	X13. Elixhauser Comorbidity Index (numerical)	
	Mean (SD)	3.94 (5.96)
	Median [Min, Max]	3.00 [-16.0, 44.0]
	X14. RSI for 30-day Length of Stay (numerical)	
	Mean (SD)	0.404 (0.772)
	Median [Min, Max]	0.592 [-6.27, 1.83]
	X15. RSI for In-hospital Mortality (numerical)	
	Mean (SD)	0.0780 (0.947)
	Median [Min, Max]	0.00 [-3.78, 13.6]
Attending Physician	**X16. Gender (binary)**	
	Female	17085 (21.0%)
	Male	64228 (79.0%)
	X17. Year of Experience (numerical)	
	Mean (SD)	18.7 (9.44)

	Median [Min, Max]	18.0 [0.00, 60.0]
	X18. Number of Hospital Affiliations	
	Mean (SD)	2.55 (1.41)
	Median [Min, Max]	2.00 [0.00, 5.00]
Hospital	**X19. Located in Rural Area (binary)**	
	FALSE	78712 (96.8%)
	TRUE	2601 (3.2%)
	X20. Ownership (categorical)	
	Government	11667 (14.3%)
	Proprietary	29186 (35.9%)
	Voluntary	40460 (49.8%)
	X21. Size (Number of Beds) (numerical)	
	Mean (SD)	551 (557)
	Median [Min, Max]	349 [15.0, 2400]
	X22. Year of Experience with EHRs (numerical)	
	Mean (SD)	5.17 (4.76)
	Median [Min, Max]	4.00 [-3.00, 18.0]
	X23. "Meaningful Use" of EHRs (binary)	
	FALSE	31590 (38.8%)
	TRUE	49723 (61.2%)
	X24 to X158. Hospital Dummies (binary)	Omitted to conserve space

Notes. ADEs = adverse drug events; AEs = adverse events; EHRs = electronic health records; RSI = risk stratification index

Appendix D: Details about the Evaluations

In this appendix, we report details about the results and simulation setup involved in evaluations 1, 2, and 3. In Tables D1 and D2, we tabulate detailed results from evaluations 1 and 2, respectively, including the area under the curve (AUC), precision, recall, and F-score of each model under different adverse event (AE) categories and training/test periods. We also report the DeLong [12] test of AUC among models in the same AE category and training/test period to verify whether the difference in AUC values between the proposed model (SALT) and the benchmark model is statistically significant. In Figure D1, we provide a flow diagram to illustrate the simulation procedure used in evaluation 3 to determine the potential practical value of the proposed mode.

Table D1. Detailed Results from Evaluation 1

AE Category	Training Period	Test Period	Model	AUC	DeLong Test H_a: AUC(Model) ≠ AUC(SALT)	Precision	Recall	F-score
ADE	2010	2011-2014	GLMM	0.660	$p < 0.01$	0.04	0.623	0.076
ADE	2010	2011-2014	MERF	0.615	$p < 0.01$	0.035	0.608	0.066
ADE	2010	2011-2014	MERT	0.614	$p < 0.01$	0.034	0.618	0.065
ADE	2010	2011-2014	SALT	0.747	—	0.056	1.000	0.106
ADE	2010-2011	2012-2014	GLMM	0.664	$p < 0.01$	0.04	0.667	0.075
ADE	2010-2011	2012-2014	MERF	0.617	$p < 0.01$	0.034	0.645	0.064
ADE	2010-2011	2012-2014	MERT	0.622	$p < 0.01$	0.039	0.492	0.072
ADE	2010-2011	2012-2014	SALT	0.774	—	0.061	1.000	0.115
ADE	2010-2012	2013-2014	GLMM	0.674	$p < 0.01$	0.042	0.652	0.078
ADE	2010-2012	2013-2014	MERF	0.627	$p < 0.01$	0.034	0.694	0.066
ADE	2010-2012	2013-2014	MERT	0.629	$p < 0.01$	0.036	0.624	0.068
ADE	2010-2012	2013-2014	SALT	0.750	—	0.058	1.000	0.110
ADE	2010-2013	2014	GLMM	0.675	$p < 0.01$	0.042	0.645	0.079
ADE	2010-2013	2014	MERF	0.648	$p < 0.01$	0.044	0.504	0.082
ADE	2010-2013	2014	MERT	0.638	$p < 0.01$	0.039	0.611	0.073
ADE	2010-2013	2014	SALT	0.726	—	0.058	1.000	0.110
Surgery	2010	2011-2014	GLMM	0.845	$p < 0.01$	0.048	0.741	0.090
Surgery	2010	2011-2014	MERF	0.801	$p < 0.01$	0.042	0.693	0.079
Surgery	2010	2011-2014	MERT	0.775	$p < 0.01$	0.062	0.615	0.112
Surgery	2010	2011-2014	SALT	0.896	—	0.054	1.000	0.103
Surgery	2010-2011	2012-2014	GLMM	0.854	$p < 0.01$	0.056	0.729	0.104

Surgery	2010-2011	2012-2014	MERF	0.804	$p < 0.01$	0.048	0.661	0.090
Surgery	2010-2011	2012-2014	MERT	0.783	$p < 0.01$	0.064	0.614	0.116
Surgery	2010-2011	2012-2014	SALT	0.897	—	0.071	1.000	0.132
Surgery	2010-2012	2013-2014	GLMM	0.858	$p < 0.01$	0.06	0.726	0.112
Surgery	2010-2012	2013-2014	MERF	0.803	$p < 0.01$	0.049	0.674	0.091
Surgery	2010-2012	2013-2014	MERT	0.789	$p < 0.01$	0.064	0.620	0.116
Surgery	2010-2012	2013-2014	SALT	0.948	—	0.094	1.000	0.172
Surgery	2010-2013	2014	GLMM	0.870	$p < 0.01$	0.057	0.766	0.106
Surgery	2010-2013	2014	MERF	0.809	$p < 0.01$	0.062	0.654	0.114
Surgery	2010-2013	2014	MERT	0.806	$p < 0.01$	0.062	0.638	0.113
Surgery	2010-2013	2014	SALT	0.903	—	0.064	1.000	0.120
Infection	2010	2011-2014	GLMM	0.745	$p < 0.01$	0.098	0.609	0.169
Infection	2010	2011-2014	MERF	0.713	$p < 0.01$	0.088	0.591	0.154
Infection	2010	2011-2014	MERT	0.673	$p < 0.01$	0.115	0.418	0.180
Infection	2010	2011-2014	SALT	0.785		0.116	0.993	0.207
Infection	2010-2011	2012-2014	GLMM	0.750	$p < 0.01$	0.093	0.641	0.163
Infection	2010-2011	2012-2014	MERF	0.721	$p < 0.01$	0.093	0.581	0.160
Infection	2010-2011	2012-2014	MERT	0.667	$p < 0.01$	0.101	0.411	0.162
Infection	2010-2011	2012-2014	SALT	0.769	—	0.111	0.976	0.199
Infection	2010-2012	2013-2014	GLMM	0.757	$p < 0.01$	0.088	0.668	0.155
Infection	2010-2012	2013-2014	MERF	0.721	$p < 0.01$	0.084	0.612	0.147
Infection	2010-2012	2013-2014	MERT	0.715	$p < 0.01$	0.091	0.596	0.158
Infection	2010-2012	2013-2014	SALT	0.831	—	0.111	1.000	0.200
Infection	2010-2013	2014	GLMM	0.757	$p < 0.01$	0.090	0.637	0.157
Infection	2010-2013	2014	MERF	0.725	$p < 0.01$	0.087	0.588	0.152
Infection	2010-2013	2014	MERT	0.669	$p < 0.01$	0.083	0.480	0.141
Infection	2010-2013	2014	SALT	0.784	—	0.117	1.000	0.209
Device	2010	2011-2014	GLMM	0.933	$p < 0.01$	0.042	0.904	0.080
Device	2010	2011-2014	MERF	0.907	$p < 0.01$	0.031	0.841	0.060
Device	2010	2011-2014	MERT	0.930	$p < 0.01$	0.040	0.910	0.077
Device	2010	2011-2014	SALT	0.968	—	0.040	1.000	0.076
Device	2010-2011	2012-2014	GLMM	0.931	$p < 0.01$	0.039	0.903	0.074
Device	2010-2011	2012-2014	MERF	0.889	$p < 0.01$	0.024	0.846	0.047
Device	2010-2011	2012-2014	MERT	0.919	$p < 0.01$	0.035	0.903	0.067
Device	2010-2011	2012-2014	SALT	0.981	—	0.060	1.000	0.114
Device	2010-2012	2013-2014	GLMM	0.929	$p < 0.01$	0.038	0.898	0.073
Device	2010-2012	2013-2014	MERF	0.897	$p < 0.01$	0.028	0.842	0.054
Device	2010-2012	2013-2014	MERT	0.913	$p < 0.01$	0.034	0.898	0.065
Device	2010-2012	2013-2014	SALT	0.977	—	0.047	1.000	0.090
Device	2010-2013	2014	GLMM	0.929	$p < 0.01$	0.038	0.889	0.073
Device	2010-2013	2014	MERF	0.871	$p < 0.01$	0.031	0.843	0.060
Device	2010-2013	2014	MERT	0.905	$p < 0.01$	0.037	0.889	0.070
Device	2010-2013	2014	SALT	0.978	—	0.055	1.000	0.104
Other	2010	2011-2014	GLMM	0.790	$p < 0.01$	0.009	0.598	0.018
Other	2010	2011-2014	MERF	0.681	$p < 0.01$	0.009	0.434	0.017

Other	2010	2011-2014	MERT	0.590	$p < 0.01$	0.004	0.441	0.007
Other	2010	2011-2014	SALT	0.804	—	0.007	0.746	0.014
Other	2010-2011	2012-2014	GLMM	0.802	$p < 0.01$	0.008	0.699	0.016
Other	2010-2011	2012-2014	MERF	0.673	$p < 0.01$	0.007	0.497	0.014
Other	2010-2011	2012-2014	MERT	0.599	$p < 0.01$	0.005	0.306	0.009
Other	2010-2011	2012-2014	SALT	0.918	—	0.008	1.000	0.016
Other	2010-2012	2013-2014	GLMM	0.800	$p < 0.01$	0.016	0.603	0.032
Other	2010-2012	2013-2014	MERF	0.654	$p < 0.01$	0.006	0.545	0.012
Other	2010-2012	2013-2014	MERT	0.607	$p < 0.01$	0.007	0.355	0.013
Other	2010-2012	2013-2014	SALT	0.975	—	0.041	1.000	0.079
Other	2010-2013	2014	GLMM	0.794	$p < 0.01$	0.02	0.660	0.038
Other	2010-2013	2014	MERF	0.712	$p < 0.01$	0.01	0.532	0.019
Other	2010-2013	2014	MERT	0.709	$p < 0.01$	0.006	0.596	0.011
Other	2010-2013	2014	SALT	0.943	—	0.009	1.000	0.018

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; MERF = mixed effects random forest; MERT = mixed effects regression tree. Test of difference in AUC is based on the DeLong et al [12].

Table D2. Detailed Results from Evaluation 2

AE Category	Training Period	Test Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)	Precision	Recall	F-score
ADE	2010	2011-2014	CART	0.589	$p < 0.01$	0.049	0.271	0.082
ADE	2010	2011-2014	DNN	0.569	$p < 0.01$	0.030	0.638	0.057
ADE	2010	2011-2014	GBM	0.614	$p < 0.01$	0.039	0.458	0.072
ADE	2010	2011-2014	LR	0.660	$p < 0.01$	0.041	0.604	0.078
ADE	2010	2011-2014	NB	0.541	$p < 0.01$	0.028	0.911	0.054
ADE	2010	2011-2014	RF	0.647	$p < 0.01$	0.036	0.691	0.068
ADE	2010	2011-2014	SVM	0.537	$p < 0.01$	0.028	0.701	0.054
ADE	2010	2011-2014	SALT	0.747	—	0.056	1.000	0.106
ADE	2010-2011	2012-2014	CART	0.610	$p < 0.01$	0.048	0.320	0.084
ADE	2010-2011	2012-2014	DNN	0.609	$p < 0.01$	0.033	0.680	0.062
ADE	2010-2011	2012-2014	GBM	0.660	$p < 0.01$	0.042	0.605	0.078
ADE	2010-2011	2012-2014	LR	0.664	$p < 0.01$	0.040	0.660	0.075
ADE	2010-2011	2012-2014	NB	0.557	$p < 0.01$	0.029	0.851	0.057
ADE	2010-2011	2012-2014	RF	0.648	$p < 0.01$	0.038	0.663	0.071
ADE	2010-2011	2012-2014	SVM	0.555	$p < 0.01$	0.029	0.664	0.055
ADE	2010-2011	2012-2014	SALT	0.774	—	0.061	1.000	0.115
ADE	2010-2012	2013-2014	CART	0.631	$p < 0.01$	0.038	0.598	0.072
ADE	2010-2012	2013-2014	DNN	0.640	$p < 0.01$	0.037	0.673	0.071
ADE	2010-2012	2013-2014	GBM	0.663	$p < 0.01$	0.041	0.656	0.077
ADE	2010-2012	2013-2014	LR	0.675	$p < 0.01$	0.041	0.669	0.077
ADE	2010-2012	2013-2014	NB	0.580	$p < 0.01$	0.032	0.713	0.062
ADE	2010-2012	2013-2014	RF	0.622	$p < 0.01$	0.036	0.603	0.068
ADE	2010-2012	2013-2014	SVM	0.562	$p < 0.01$	0.031	0.595	0.058
ADE	2010-2012	2013-2014	SALT	0.750	—	0.058	1.000	0.110
ADE	2010-2013	2014-2014	CART	0.638	$p < 0.01$	0.037	0.688	0.071
ADE	2010-2013	2014-2014	DNN	0.649	$p < 0.01$	0.044	0.532	0.081
ADE	2010-2013	2014-2014	GBM	0.674	$p < 0.01$	0.042	0.675	0.079
ADE	2010-2013	2014-2014	LR	0.674	$p < 0.01$	0.043	0.659	0.080
ADE	2010-2013	2014-2014	NB	0.610	$p < 0.01$	0.035	0.719	0.068
ADE	2010-2013	2014-2014	RF	0.639	$p < 0.01$	0.038	0.643	0.072
ADE	2010-2013	2014-2014	SVM	0.554	$p < 0.01$	0.035	0.414	0.064
ADE	2010-2013	2014-2014	SALT	0.726	—	0.058	1.000	0.110
Surgery	2010	2011-2014	CART	0.685	$p < 0.01$	0.085	0.412	0.141
Surgery	2010	2011-2014	DNN	0.588	$p < 0.01$	0.036	0.264	0.063
Surgery	2010	2011-2014	GBM	0.774	$p < 0.01$	0.035	0.691	0.066
Surgery	2010	2011-2014	LR	0.853	$p < 0.01$	0.050	0.743	0.093
Surgery	2010	2011-2014	NB	0.518	$p < 0.01$	0.014	0.976	0.027
Surgery	2010	2011-2014	RF	0.834	$p < 0.01$	0.051	0.705	0.096
Surgery	2010	2011-2014	SVM	0.659	$p < 0.01$	0.052	0.411	0.092
Surgery	2010	2011-2014	SALT	0.896	—	0.054	1.000	0.103
Surgery	2010-2011	2012-2014	CART	0.784	$p < 0.01$	0.040	0.750	0.076

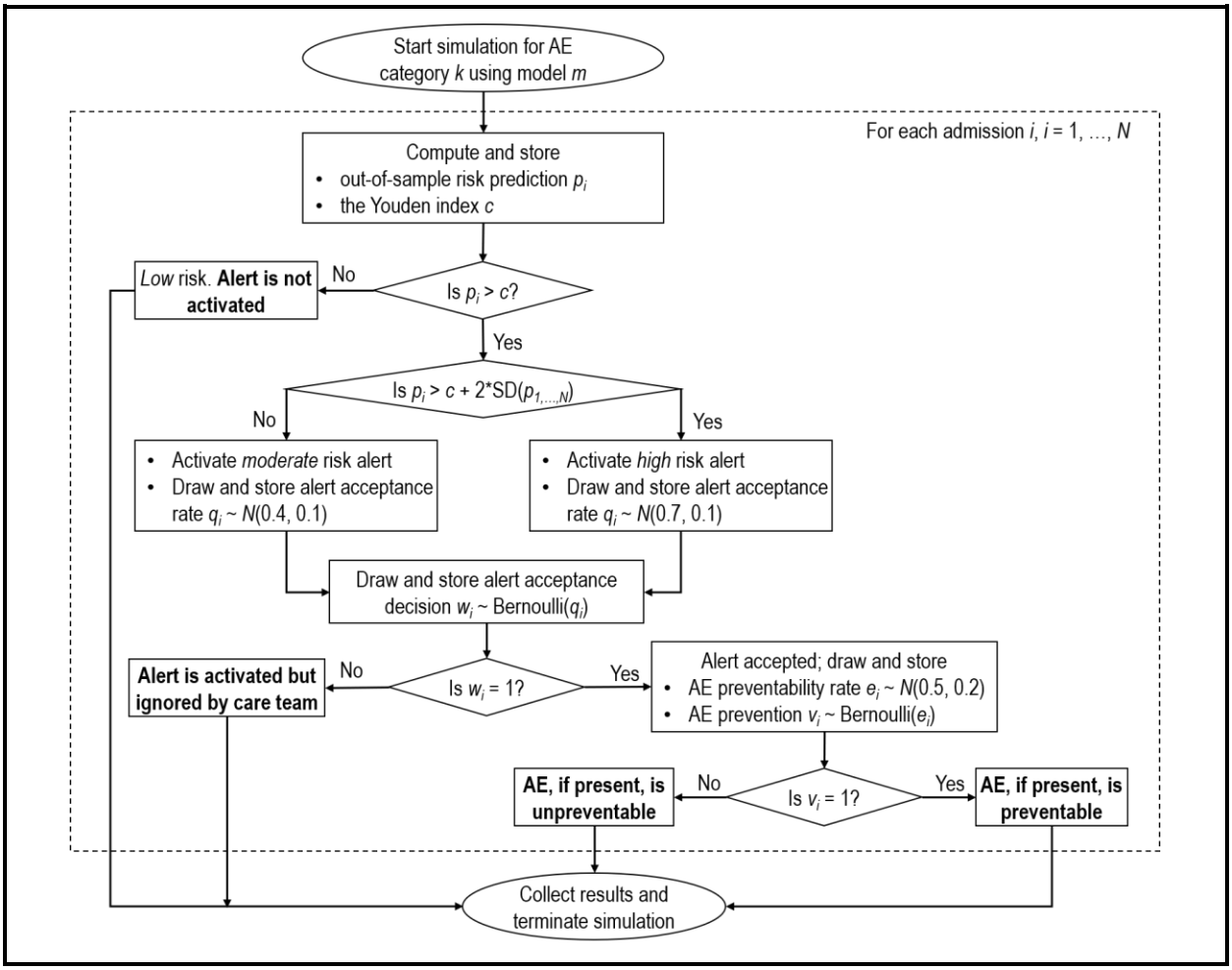
Surgery	2010-2011	2012-2014	DNN	0.721	$p < 0.01$	0.042	0.579	0.078
Surgery	2010-2011	2012-2014	GBM	0.830	$p < 0.01$	0.043	0.756	0.081
Surgery	2010-2011	2012-2014	LR	0.857	$p < 0.01$	0.053	0.741	0.099
Surgery	2010-2011	2012-2014	NB	0.519	$p < 0.01$	0.014	0.960	0.028
Surgery	2010-2011	2012-2014	RF	0.837	$p < 0.01$	0.044	0.765	0.083
Surgery	2010-2011	2012-2014	SVM	0.681	$p < 0.01$	0.041	0.497	0.076
Surgery	2010-2011	2012-2014	SALT	0.897	—	0.071	1.000	0.132
Surgery	2010-2012	2013-2014	CART	0.787	$p < 0.01$	0.041	0.742	0.077
Surgery	2010-2012	2013-2014	DNN	0.777	$p < 0.01$	0.037	0.679	0.069
Surgery	2010-2012	2013-2014	GBM	0.848	$p < 0.01$	0.063	0.696	0.116
Surgery	2010-2012	2013-2014	LR	0.858	$p < 0.01$	0.054	0.752	0.101
Surgery	2010-2012	2013-2014	NB	0.528	$p < 0.01$	0.015	0.927	0.029
Surgery	2010-2012	2013-2014	RF	0.835	$p < 0.01$	0.050	0.721	0.093
Surgery	2010-2012	2013-2014	SVM	0.696	$p < 0.01$	0.042	0.515	0.078
Surgery	2010-2012	2013-2014	SALT	0.948	—	0.094	1.000	0.172
Surgery	2010-2013	2014-2014	CART	0.806	$p < 0.01$	0.040	0.780	0.076
Surgery	2010-2013	2014-2014	DNN	0.822	$p < 0.01$	0.047	0.717	0.088
Surgery	2010-2013	2014-2014	GBM	0.864	$p < 0.01$	0.061	0.738	0.114
Surgery	2010-2013	2014-2014	LR	0.871	$p < 0.01$	0.057	0.769	0.107
Surgery	2010-2013	2014-2014	NB	0.561	$p < 0.01$	0.016	0.913	0.031
Surgery	2010-2013	2014-2014	RF	0.841	$p < 0.01$	0.058	0.719	0.107
Surgery	2010-2013	2014-2014	SVM	0.720	$p < 0.01$	0.045	0.509	0.083
Surgery	2010-2013	2014-2014	SALT	0.903	—	0.064	1.000	0.120
Infection	2010	2011-2014	CART	0.697	$p < 0.01$	0.077	0.683	0.138
Infection	2010	2011-2014	DNN	0.630	$p < 0.01$	0.060	0.623	0.110
Infection	2010	2011-2014	GBM	0.712	$p < 0.01$	0.085	0.599	0.149
Infection	2010	2011-2014	LR	0.745	$p < 0.01$	0.099	0.608	0.170
Infection	2010	2011-2014	NB	0.524	$p < 0.01$	0.046	0.853	0.087
Infection	2010	2011-2014	RF	0.727	$p < 0.01$	0.089	0.616	0.156
Infection	2010	2011-2014	SVM	0.557	$p < 0.01$	0.078	0.213	0.114
Infection	2010	2011-2014	SALT	0.785	—	0.116	0.993	0.207
Infection	2010-2011	2012-2014	CART	0.712	$p < 0.01$	0.087	0.607	0.152
Infection	2010-2011	2012-2014	DNN	0.720	$p < 0.01$	0.077	0.677	0.139
Infection	2010-2011	2012-2014	GBM	0.740	$p < 0.01$	0.082	0.678	0.146
Infection	2010-2011	2012-2014	LR	0.752	$p < 0.01$	0.096	0.626	0.166
Infection	2010-2011	2012-2014	NB	0.549	$p < 0.01$	0.048	0.733	0.090
Infection	2010-2011	2012-2014	RF	0.740	$p < 0.01$	0.084	0.654	0.148
Infection	2010-2011	2012-2014	SVM	0.569	$p < 0.01$	0.050	0.585	0.093
Infection	2010-2011	2012-2014	SALT	0.769	—	0.111	0.976	0.199
Infection	2010-2012	2013-2014	CART	0.720	$p < 0.01$	0.096	0.571	0.165
Infection	2010-2012	2013-2014	DNN	0.725	$p < 0.01$	0.078	0.642	0.139
Infection	2010-2012	2013-2014	GBM	0.754	$p < 0.01$	0.083	0.688	0.148
Infection	2010-2012	2013-2014	LR	0.757	$p < 0.01$	0.089	0.666	0.157
Infection	2010-2012	2013-2014	NB	0.579	$p < 0.01$	0.053	0.646	0.099
Infection	2010-2012	2013-2014	RF	0.742	$p < 0.01$	0.078	0.702	0.140

Infection	2010-2012	2013-2014	SVM	0.579	$p < 0.01$	0.065	0.331	0.108
Infection	2010-2012	2013-2014	SALT	0.831	—	0.111	1.000	0.200
Infection	2010-2013	2014-2014	CART	0.732	$p < 0.01$	0.101	0.570	0.172
Infection	2010-2013	2014-2014	DNN	0.746	$p < 0.01$	0.088	0.605	0.154
Infection	2010-2013	2014-2014	GBM	0.758	$p < 0.01$	0.080	0.699	0.144
Infection	2010-2013	2014-2014	LR	0.756	$p < 0.01$	0.089	0.647	0.157
Infection	2010-2013	2014-2014	NB	0.597	$p < 0.01$	0.055	0.608	0.100
Infection	2010-2013	2014-2014	RF	0.742	$p < 0.01$	0.075	0.697	0.136
Infection	2010-2013	2014-2014	SVM	0.591	$p < 0.01$	0.060	0.416	0.106
Infection	2010-2013	2014-2014	SALT	0.784	—	0.117	1.000	0.209
Device	2010	2011-2014	CART	0.836	$p < 0.01$	0.044	0.738	0.084
Device	2010	2011-2014	DNN	0.674	$p < 0.01$	0.046	0.377	0.082
Device	2010	2011-2014	GBM	0.922	$p < 0.01$	0.032	0.904	0.062
Device	2010	2011-2014	LR	0.943	$p < 0.01$	0.045	0.904	0.085
Device	2010	2011-2014	NB	0.500	$p < 0.01$	0.004	1.000	0.008
Device	2010	2011-2014	RF	0.932	$p < 0.01$	0.029	0.910	0.057
Device	2010	2011-2014	SVM	0.773	$p < 0.01$	0.023	0.651	0.045
Device	2010	2011-2014	SALT	0.968	—	0.040	1.000	0.076
Device	2010-2011	2012-2014	CART	0.871	$p < 0.01$	0.042	0.813	0.080
Device	2010-2011	2012-2014	DNN	0.728	$p < 0.01$	0.040	0.492	0.074
Device	2010-2011	2012-2014	GBM	0.923	$p < 0.01$	0.031	0.888	0.060
Device	2010-2011	2012-2014	LR	0.931	$p < 0.01$	0.039	0.903	0.075
Device	2010-2011	2012-2014	NB	0.500	$p < 0.01$	0.004	1.000	0.008
Device	2010-2011	2012-2014	RF	0.927	$p < 0.01$	0.028	0.897	0.055
Device	2010-2011	2012-2014	SVM	0.813	$p < 0.01$	0.020	0.713	0.039
Device	2010-2011	2012-2014	SALT	0.981	—	0.060	1.000	0.114
Device	2010-2012	2013-2014	CART	0.870	$p < 0.01$	0.040	0.814	0.076
Device	2010-2012	2013-2014	DNN	0.862	$p < 0.01$	0.021	0.795	0.041
Device	2010-2012	2013-2014	GBM	0.926	$p < 0.01$	0.033	0.912	0.064
Device	2010-2012	2013-2014	LR	0.929	$p < 0.01$	0.038	0.898	0.073
Device	2010-2012	2013-2014	NB	0.500	$p < 0.01$	0.004	1.000	0.008
Device	2010-2012	2013-2014	RF	0.918	$p < 0.01$	0.037	0.898	0.072
Device	2010-2012	2013-2014	SVM	0.812	$p < 0.01$	0.015	0.726	0.030
Device	2010-2012	2013-2014	SALT	0.977	—	0.047	1.000	0.090
Device	2010-2013	2014-2014	CART	0.877	$p < 0.01$	0.039	0.824	0.075
Device	2010-2013	2014-2014	DNN	0.867	$p < 0.01$	0.032	0.759	0.062
Device	2010-2013	2014-2014	GBM	0.906	$p < 0.01$	0.028	0.898	0.054
Device	2010-2013	2014-2014	LR	0.928	$p < 0.01$	0.039	0.889	0.074
Device	2010-2013	2014-2014	NB	0.504	$p < 0.01$	0.004	1.000	0.008
Device	2010-2013	2014-2014	RF	0.914	$p < 0.01$	0.042	0.880	0.081
Device	2010-2013	2014-2014	SVM	0.812	$p < 0.01$	0.017	0.713	0.033
Device	2010-2013	2014-2014	SALT	0.978	—	0.055	1.000	0.104
Other	2010	2011-2014	CART	0.603	$p < 0.01$	0.003	0.559	0.007
Other	2010	2011-2014	DNN	0.534	$p < 0.01$	0.029	0.074	0.042
Other	2010	2011-2014	GBM	0.677	$p < 0.01$	0.005	0.535	0.010

Other	2010	2011-2014	LR	0.814	$p < 0.01$	0.007	0.742	0.014
Other	2010	2011-2014	NB	0.500	$p < 0.01$	0.002	1.000	0.004
Other	2010	2011-2014	RF	0.768	$p < 0.01$	0.007	0.641	0.014
Other	2010	2011-2014	SVM	0.635	$p < 0.01$	0.003	0.633	0.007
Other	2010	2011-2014	SALT	0.804	—	0.007	0.746	0.014
Other	2010-2011	2012-2014	CART	0.685	$p < 0.01$	0.004	0.787	0.008
Other	2010-2011	2012-2014	DNN	0.540	$p < 0.01$	0.011	0.098	0.019
Other	2010-2011	2012-2014	GBM	0.789	$p < 0.01$	0.007	0.705	0.013
Other	2010-2011	2012-2014	LR	0.829	$p < 0.01$	0.009	0.699	0.018
Other	2010-2011	2012-2014	NB	0.500	$p < 0.01$	0.002	1.000	0.004
Other	2010-2011	2012-2014	RF	0.726	$p < 0.01$	0.005	0.623	0.011
Other	2010-2011	2012-2014	SVM	0.544	$p < 0.01$	0.002	0.601	0.003
Other	2010-2011	2012-2014	SALT	0.918	—	0.008	1.000	0.016
Other	2010-2012	2013-2014	CART	0.669	$p < 0.01$	0.004	0.769	0.008
Other	2010-2012	2013-2014	DNN	0.596	$p < 0.01$	0.012	0.231	0.022
Other	2010-2012	2013-2014	GBM	0.762	$p < 0.01$	0.009	0.562	0.018
Other	2010-2012	2013-2014	LR	0.811	$p < 0.01$	0.008	0.719	0.016
Other	2010-2012	2013-2014	NB	0.500	$p < 0.01$	0.002	1.000	0.004
Other	2010-2012	2013-2014	RF	0.755	$p < 0.01$	0.006	0.636	0.012
Other	2010-2012	2013-2014	SVM	0.650	$p < 0.01$	0.004	0.570	0.009
Other	2010-2012	2013-2014	SALT	0.975	—	0.041	1.000	0.079
Other	2010-2013	2014-2014	CART	0.712	$p < 0.01$	0.007	0.553	0.014
Other	2010-2013	2014-2014	DNN	0.729	$p < 0.01$	0.006	0.596	0.011
Other	2010-2013	2014-2014	GBM	0.621	$p < 0.01$	0.009	0.340	0.018
Other	2010-2013	2014-2014	LR	0.804	$p < 0.01$	0.018	0.638	0.036
Other	2010-2013	2014-2014	NB	0.500	$p < 0.01$	0.002	1.000	0.003
Other	2010-2013	2014-2014	RF	0.776	$p < 0.01$	0.007	0.681	0.014
Other	2010-2013	2014-2014	SVM	0.716	$p < 0.01$	0.006	0.489	0.012
Other	2010-2013	2014-2014	SALT	0.943	—	0.009	1.000	0.018

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; CART = classification and regression trees; DNN = deep neural network; GBM = generalized boosted model; LR = lasso regression; NB = naïve Bayes; RF = random forest; SVM = support vector machine. Test of difference in AUC is based on the DeLong et al [12].

Figure D1. Flow Diagram of the Simulation Experiment



Notes: In this simulation, we consider three AE risk levels: low, moderate, and high. If the predicted risk is lower than the Youden index, the admission is deemed to be a low AE risk, and an alert will not be activated. Otherwise, a moderate- or high-risk alert will be activated, depending on whether the predicted AE risk is 2 standard deviations above the Youden index. When an alert is activated, the care team can either accept or ignore it. The reported alert acceptance rate ranges from 38.1% [16] to 67% [42], but according to Seidling et al. [41], high-risk alerts are more likely to be accepted (odds ratio = 1.74; $p < 0.001$). We hence consider different alert acceptance rates for different alert levels. The alert acceptance rates for moderate- and high-risk alerts are drawn from $N(0.4, 0.1)$ and $N(0.7, 0.1)$, respectively, so that they are consistent with the range and odds ratios documented in the literature. If an alert is accepted, we assume the care team will take proper precautions. However, not all AEs can be prevented. Studies suggest that the portion of AEs that are deemed preventable is about 50% [45]. As such, we draw the rate of AE preventability from $N(0.5, 0.2)$. Taken together, in our simulation each admission falls into one of the following four scenarios: (1) alert not activated, (2) alert activated but ignored by care team, (3) alert activated and accepted but AE is unpreventable, and (4) alert activated and accepted and AE is preventable. Our primary interest is to estimate the numbers of prevented AEs and false alarms. We replicate the simulation 5,000 times to obtain the empirical distributions for these statistics.

Appendix E: Additional Analyses

In this appendix, we report results from additional analyses that we use to verify the robustness of our results shown in the main text. We have conducted a total of 21 additional analyses related to feature engineering, feature selection, data imbalance, and concept drift. In the results reported in the main text, we found that the proposed model (SALT) achieved a better adverse event (AE) predictive performance than alternative models, but we did not consider feature engineering, feature selection, and remedy for data imbalance in our experiments. Therefore, the objectives of the additional analyses here are to understand the following:

1. Does SALT still outperform alternative models when a feature engineering approach is used in the AE predictive modeling process?
2. Does SALT still outperform alternative models when a feature selection approach is used in the AE predictive modeling process?
3. Does SALT still outperform alternative models when a remedy for data imbalance is used in the AE predictive modeling process?
4. Is there evidence of concept drift in the AE predictive modeling setting?

Accordingly, we devise 6 categories of additional analyses. In category 1, we consider four different variable transformation strategies for feature engineering: log transformation, Yeo-Johnson transformation [49], 95% winsorization [5], and 99% winsorization [5]. In category 2, we consider five different variable discretization strategies for feature engineering: 5 equidistant bins [14], 10 equidistant bins [14], 5-bins discretization with k-means clustering [23], 10-bins discretization with k-means clustering [23], and discretization with decision trees, which detect the optimal number of bins [14]. In category 3, we consider seven feature selection approaches to retain 50% of most important features: chi-squared values [20], information gain [20], mutual information (MI) [20], the joint MI maximization method [7], the double input symmetrical relevance method [31], the joint MI method [48], and the minimum redundancy maximal

relevance method [36]. These are commonly considered as filter-based methods for feature selection in which each of them uses a specific metric to score each individual feature and the features with higher scores are retained for model training and evaluation. The filter-based methods are very popular in machine learning practice because they are usually much faster and less computationally intensive than wrapper or embedded methods for feature selection [9]. In category 4, we consider three strategies that are commonly used to address practical data imbalance issues: Synthetic Minority Over-Sampling Technique (SMOTE) [10], random over-sampling [3], and random under-sampling [3]. In category 5, we combine the best approach we found from each of the prior categories. We determine the best approach in a category based on the mean AUC value across AE categories and training periods. Finally, since concept drift can lead to deterioration of predictive performance, in category 6, we examine whether concept drift is present in our specific AE predictive modeling setting by looking into whether the predictive performance deteriorate over time.

For analyses in categories 1 through 5, we only report results from 2 different training periods (2010 and 2010-2013, which are the shortest and longest possible training periods in our setting) and 2 alternative models (generalized linear mixed model (GLMM and lasso regression (LR), which are the best alternative model from evaluations 1 and 2, respectively). We do this to conserve space and fit each table within one page in order for improved readability. We note that the unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here.

The results of these additional analyses are summarized in Table E1, and details are reported in Tables E2-E22. Overall, we find that each of these feature engineering, feature selection, and data imbalance remedy can have an impact on the performance of AE predictive

modeling. However, the impact can be either negative and positive, depending on the specific AE category, training period, and model being considered. More importantly, we find that SALT still consistently outperforms alternative models in each and every analysis in categories 1 through 5. From the category 6 analysis on concept drift, we do not observe a systematic downward trend in predictive performance when the test periods are further away from the training period. This suggests that concept drift is not present in our setting.

Table E1. Summary of Additional Analyses

Category	Analysis	Results	Finding
1. Apply variable transformation strategies for feature engineering	1.1. Log transformation	Table E2	SALT outperforms alternative techniques in every analysis in this category
	1.2. Yeo-Johnson transformation [49]	Table E3	
	1.3. Winsorization (95%) [5]	Table E4	
	1.4. Winsorization (99%) [5]	Table E5	
2. Apply variable discretization strategies for feature engineering	2.1. Use 5 equidistant bins [14]	Table E6	SALT outperforms alternative techniques in every analysis in this category
	2.2. Use 10 equidistant bins [14]	Table E7	
	2.3. Discretization with k-means clustering (5 clusters/bins) [23]	Table E8	
	2.4. Discretization with k-means clustering (10 clusters/bins) [23]	Table E9	
	2.5. Discretization with decision trees (detecting the optimal number of bins) [14]	Table E10	
3. Apply strategies for feature selection	3.1. Based on chi-squared values [20]	Table E11	SALT outperforms alternative techniques in every analysis in this category
	3.2. Based on information gain [20]	Table E12	
	3.3. Based on mutual information (MI) [20]	Table E13	
	3.4. Based on the joint MI maximization method [7]	Table E14	
	3.5. Based on the double input symmetrical relevance method [31]	Table E15	
	3.6. Based on the joint MI method [48]	Table E16	
	3.7. Based on the minimum redundancy maximal relevance method [36]	Table E17	
4. Apply strategies to address data imbalance	4.1. Synthetic Minority Over-Sampling Technique (SMOTE) [10]	Table E18	SALT outperforms alternative techniques in every analysis in this category
	4.2. Random over-sampling [3]	Table E19	
	4.3. Random under-sampling [3]	Table E20	
5. Combine the best strategy from each of the categories above		Table E21	SALT outperforms alternative techniques
6. Assess the presence of concept drifting		Table E22	No evidence of concept drifting in our setting

Table E2. Feature Engineering thorough Log Variable Transformation (Analysis 1.1)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.598	$p < 0.01$
			LR	0.606	$p < 0.01$
			SALT	0.700	—
	2010-2013	2014	GLMM	0.639	$p < 0.01$
			LR	0.610	$p < 0.01$
			SALT	0.671	—
Surgery	2010	2011-2014	GLMM	0.791	$p < 0.01$
			LR	0.786	$p < 0.01$
			SALT	0.851	—
	2010-2013	2014	GLMM	0.827	$p < 0.01$
			LR	0.797	$p < 0.01$
			SALT	0.853	—
Infection	2010	2011-2014	GLMM	0.694	$p < 0.01$
			LR	0.700	$p < 0.01$
			SALT	0.715	—
	2010-2013	2014	GLMM	0.713	$p < 0.01$
			LR	0.685	$p < 0.01$
			SALT	0.772	—
Device	2010	2011-2014	GLMM	0.885	$p < 0.01$
			LR	0.875	$p < 0.01$
			SALT	0.909	—
	2010-2013	2014	GLMM	0.862	$p < 0.01$
			LR	0.904	$p < 0.01$
			SALT	0.959	—
Other	2010	2011-2014	GLMM	0.727	$p < 0.01$
			LR	0.741	$p < 0.01$
			SALT	0.767	—
	2010-2013	2014	GLMM	0.770	$p < 0.01$
			LR	0.755	$p < 0.01$
			SALT	0.891	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E3. Feature Engineering thorough Yeo-Johnson Variable Transformation (Analysis 1.2)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.639	$p < 0.01$
			LR	0.624	$p < 0.01$
			SALT	0.721	—
	2010-2013	2014	GLMM	0.638	$p < 0.01$
			LR	0.694	$p < 0.01$
			SALT	0.715	—
Surgery	2010	2011-2014	GLMM	0.816	$p < 0.01$
			LR	0.850	$p < 0.01$
			SALT	0.878	—
	2010-2013	2014	GLMM	0.844	$p < 0.01$
			LR	0.850	$p < 0.01$
			SALT	0.914	—
Infection	2010	2011-2014	GLMM	0.734	$p < 0.01$
			LR	0.690	$p < 0.01$
			SALT	0.806	—
	2010-2013	2014	GLMM	0.765	$p < 0.01$
			LR	0.750	$p < 0.01$
			SALT	0.791	—
Device	2010	2011-2014	GLMM	0.936	$p < 0.01$
			LR	0.926	$p < 0.01$
			SALT	0.984	—
	2010-2013	2014	GLMM	0.903	$p < 0.01$
			LR	0.929	$p < 0.01$
			SALT	0.954	—
Other	2010	2011-2014	GLMM	0.769	$p < 0.01$
			LR	0.803	$p < 0.01$
			SALT	0.817	—
	2010-2013	2014	GLMM	0.814	$p < 0.01$
			LR	0.815	$p < 0.01$
			SALT	0.949	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E4. Feature Engineering thorough 95% Winsorization (Analysis 1.3)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.583	$p < 0.01$
			LR	0.588	$p < 0.01$
			SALT	0.685	—
	2010-2013	2014	GLMM	0.593	$p < 0.01$
			LR	0.594	$p < 0.01$
			SALT	0.678	—
Surgery	2010	2011-2014	GLMM	0.758	$p < 0.01$
			LR	0.769	$p < 0.01$
			SALT	0.812	—
	2010-2013	2014	GLMM	0.812	$p < 0.01$
			LR	0.768	$p < 0.01$
			SALT	0.828	—
Infection	2010	2011-2014	GLMM	0.665	$p < 0.01$
			LR	0.634	$p < 0.01$
			SALT	0.710	—
	2010-2013	2014	GLMM	0.665	$p < 0.01$
			LR	0.731	$p < 0.01$
			SALT	0.700	—
Device	2010	2011-2014	GLMM	0.881	$p < 0.01$
			LR	0.874	$p < 0.01$
			SALT	0.897	—
	2010-2013	2014	GLMM	0.888	$p < 0.01$
			LR	0.841	$p < 0.01$
			SALT	0.939	—
Other	2010	2011-2014	GLMM	0.717	$p < 0.01$
			LR	0.727	$p < 0.01$
			SALT	0.744	—
	2010-2013	2014	GLMM	0.702	$p < 0.01$
			LR	0.736	$p < 0.01$
			SALT	0.885	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E5. Feature Engineering thorough 99% Winsorization (Analysis 1.4)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.596	$p < 0.01$
			LR	0.588	$p < 0.01$
			SALT	0.664	—
	2010-2013	2014	GLMM	0.577	$p < 0.01$
			LR	0.618	$p < 0.01$
			SALT	0.684	—
Surgery	2010	2011-2014	GLMM	0.779	$p < 0.01$
			LR	0.800	$p < 0.01$
			SALT	0.831	—
	2010-2013	2014	GLMM	0.822	$p < 0.01$
			LR	0.829	$p < 0.01$
			SALT	0.841	—
Infection	2010	2011-2014	GLMM	0.666	$p < 0.01$
			LR	0.695	$p < 0.01$
			SALT	0.759	—
	2010-2013	2014	GLMM	0.662	$p < 0.01$
			LR	0.696	$p < 0.01$
			SALT	0.745	—
Device	2010	2011-2014	GLMM	0.882	$p < 0.01$
			LR	0.871	$p < 0.01$
			SALT	0.901	—
	2010-2013	2014	GLMM	0.877	$p < 0.01$
			LR	0.882	$p < 0.01$
			SALT	0.909	—
Other	2010	2011-2014	GLMM	0.687	$p < 0.01$
			LR	0.750	$p < 0.01$
			SALT	0.763	—
	2010-2013	2014	GLMM	0.746	$p < 0.01$
			LR	0.743	$p < 0.01$
			SALT	0.896	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E6. Feature Engineering thorough Discretization Using 5 Equidistant Bins (Analysis 2.1)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.541	$p < 0.01$
			LR	0.541	$p < 0.01$
			SALT	0.627	—
	2010-2013	2014	GLMM	0.574	$p < 0.01$
			LR	0.575	$p < 0.01$
			SALT	0.671	—
Surgery	2010	2011-2014	GLMM	0.697	$p < 0.01$
			LR	0.755	$p < 0.01$
			SALT	0.797	—
	2010-2013	2014	GLMM	0.777	$p < 0.01$
			LR	0.774	$p < 0.01$
			SALT	0.832	—
Infection	2010	2011-2014	GLMM	0.611	$p < 0.01$
			LR	0.615	$p < 0.01$
			SALT	0.694	—
	2010-2013	2014	GLMM	0.656	$p < 0.01$
			LR	0.660	$p < 0.01$
			SALT	0.696	—
Device	2010	2011-2014	GLMM	0.841	$p < 0.01$
			LR	0.841	$p < 0.01$
			SALT	0.874	—
	2010-2013	2014	GLMM	0.800	$p < 0.01$
			LR	0.826	$p < 0.01$
			SALT	0.883	—
Other	2010	2011-2014	GLMM	0.663	$p < 0.01$
			LR	0.708	$p < 0.01$
			SALT	0.735	—
	2010-2013	2014	GLMM	0.697	$p < 0.01$
			LR	0.711	$p < 0.01$
			SALT	0.839	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E7. Feature Engineering thorough Discretization Using 10 Equidistant Bins (Analysis 2.2)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.571	$p < 0.01$
			LR	0.543	$p < 0.01$
			SALT	0.654	—
	2010-2013	2014	GLMM	0.600	$p < 0.01$
			LR	0.574	$p < 0.01$
			SALT	0.664	—
Surgery	2010	2011-2014	GLMM	0.801	$p < 0.01$
			LR	0.765	$p < 0.01$
			SALT	0.812	—
	2010-2013	2014	GLMM	0.767	$p < 0.01$
			LR	0.797	$p < 0.01$
			SALT	0.850	—
Infection	2010	2011-2014	GLMM	0.656	$p < 0.01$
			LR	0.638	$p < 0.01$
			SALT	0.682	—
	2010-2013	2014	GLMM	0.672	$p < 0.01$
			LR	0.671	$p < 0.01$
			SALT	0.687	—
Device	2010	2011-2014	GLMM	0.836	$p < 0.01$
			LR	0.871	$p < 0.01$
			SALT	0.848	—
	2010-2013	2014	GLMM	0.871	$p < 0.01$
			LR	0.803	$p < 0.01$
			SALT	0.881	—
Other	2010	2011-2014	GLMM	0.709	$p < 0.01$
			LR	0.732	$p < 0.01$
			SALT	0.745	—
	2010-2013	2014	GLMM	0.664	$p < 0.01$
			LR	0.720	$p < 0.01$
			SALT	0.858	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E8. Feature Engineering thorough Discretization Using K-Means Clustering with 5 Clusters/Bins (Analysis 2.3)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.545	$p < 0.01$
			LR	0.558	$p < 0.01$
			SALT	0.624	—
	2010-2013	2014	GLMM	0.595	$p < 0.01$
			LR	0.582	$p < 0.01$
			SALT	0.642	—
Surgery	2010	2011-2014	GLMM	0.753	$p < 0.01$
			LR	0.724	$p < 0.01$
			SALT	0.805	—
	2010-2013	2014	GLMM	0.762	$p < 0.01$
			LR	0.786	$p < 0.01$
			SALT	0.798	—
Infection	2010	2011-2014	GLMM	0.652	$p < 0.01$
			LR	0.667	$p < 0.01$
			SALT	0.692	—
	2010-2013	2014	GLMM	0.662	$p < 0.01$
			LR	0.680	$p < 0.01$
			SALT	0.699	—
Device	2010	2011-2014	GLMM	0.870	$p < 0.01$
			LR	0.814	$p < 0.01$
			SALT	0.868	—
	2010-2013	2014	GLMM	0.830	$p < 0.01$
			LR	0.838	$p < 0.01$
			SALT	0.852	—
Other	2010	2011-2014	GLMM	0.694	$p < 0.01$
			LR	0.705	$p < 0.01$
			SALT	0.736	—
	2010-2013	2014	GLMM	0.662	$p < 0.01$
			LR	0.674	$p < 0.01$
			SALT	0.868	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E9. Feature Engineering thorough Discretization Using K-Means Clustering with 10 Clusters/Bins (Analysis 2.4)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.590	$p < 0.01$
			LR	0.637	$p < 0.01$
			SALT	0.730	—
	2010-2013	2014	GLMM	0.628	$p < 0.01$
			LR	0.636	$p < 0.01$
			SALT	0.679	—
Surgery	2010	2011-2014	GLMM	0.798	$p < 0.01$
			LR	0.817	$p < 0.01$
			SALT	0.866	—
	2010-2013	2014	GLMM	0.847	$p < 0.01$
			LR	0.835	$p < 0.01$
			SALT	0.889	—
Infection	2010	2011-2014	GLMM	0.747	$p < 0.01$
			LR	0.727	$p < 0.01$
			SALT	0.764	—
	2010-2013	2014	GLMM	0.749	$p < 0.01$
			LR	0.705	$p < 0.01$
			SALT	0.762	—
Device	2010	2011-2014	GLMM	0.921	$p < 0.01$
			LR	0.943	$p < 0.01$
			SALT	0.902	—
	2010-2013	2014	GLMM	0.951	$p < 0.01$
			LR	0.880	$p < 0.01$
			SALT	0.983	—
Other	2010	2011-2014	GLMM	0.780	$p < 0.01$
			LR	0.761	$p < 0.01$
			SALT	0.793	—
	2010-2013	2014	GLMM	0.732	$p < 0.01$
			LR	0.770	$p < 0.01$
			SALT	0.915	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E10. Feature Engineering thorough Discretization Using Decision Trees (Analysis 2.5)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.631	$p < 0.01$
			LR	0.607	$p < 0.01$
			SALT	0.688	—
	2010-2013	2014	GLMM	0.646	$p < 0.01$
			LR	0.634	$p < 0.01$
			SALT	0.680	—
Surgery	2010	2011-2014	GLMM	0.832	$p < 0.01$
			LR	0.836	$p < 0.01$
			SALT	0.826	—
	2010-2013	2014	GLMM	0.816	$p < 0.01$
			LR	0.845	$p < 0.01$
			SALT	0.884	—
Infection	2010	2011-2014	GLMM	0.654	$p < 0.01$
			LR	0.712	$p < 0.01$
			SALT	0.747	—
	2010-2013	2014	GLMM	0.740	$p < 0.01$
			LR	0.674	$p < 0.01$
			SALT	0.759	—
Device	2010	2011-2014	GLMM	0.900	$p < 0.01$
			LR	0.900	$p < 0.01$
			SALT	0.915	—
	2010-2013	2014	GLMM	0.883	$p < 0.01$
			LR	0.888	$p < 0.01$
			SALT	0.923	—
Other	2010	2011-2014	GLMM	0.735	$p < 0.01$
			LR	0.795	$p < 0.01$
			SALT	0.750	—
	2010-2013	2014	GLMM	0.755	$p < 0.01$
			LR	0.782	$p < 0.01$
			SALT	0.894	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E11. Feature Selection based on Chi-squared Values (Analysis 3.1)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.533	$p < 0.01$
			LR	0.602	$p < 0.01$
			SALT	0.665	—
	2010-2013	2014	GLMM	0.551	$p < 0.01$
			LR	0.544	$p < 0.01$
			SALT	0.642	—
Surgery	2010	2011-2014	GLMM	0.740	$p < 0.01$
			LR	0.737	$p < 0.01$
			SALT	0.809	—
	2010-2013	2014	GLMM	0.773	$p < 0.01$
			LR	0.774	$p < 0.01$
			SALT	0.819	—
Infection	2010	2011-2014	GLMM	0.656	$p < 0.01$
			LR	0.671	$p < 0.01$
			SALT	0.718	—
	2010-2013	2014	GLMM	0.642	$p < 0.01$
			LR	0.641	$p < 0.01$
			SALT	0.691	—
Device	2010	2011-2014	GLMM	0.819	$p < 0.01$
			LR	0.823	$p < 0.01$
			SALT	0.876	—
	2010-2013	2014	GLMM	0.829	$p < 0.01$
			LR	0.820	$p < 0.01$
			SALT	0.908	—
Other	2010	2011-2014	GLMM	0.704	$p < 0.01$
			LR	0.745	$p < 0.01$
			SALT	0.762	—
	2010-2013	2014	GLMM	0.676	$p < 0.01$
			LR	0.707	$p < 0.01$
			SALT	0.845	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E12. Feature Selection based on Information Gain (Analysis 3.2)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.537	$p < 0.01$
			LR	0.546	$p < 0.01$
			SALT	0.608	—
	2010-2013	2014	GLMM	0.546	$p < 0.01$
			LR	0.543	$p < 0.01$
			SALT	0.581	—
Surgery	2010	2011-2014	GLMM	0.700	$p < 0.01$
			LR	0.755	$p < 0.01$
			SALT	0.792	—
	2010-2013	2014	GLMM	0.763	$p < 0.01$
			LR	0.772	$p < 0.01$
			SALT	0.787	—
Infection	2010	2011-2014	GLMM	0.625	$p < 0.01$
			LR	0.629	$p < 0.01$
			SALT	0.667	—
	2010-2013	2014	GLMM	0.672	$p < 0.01$
			LR	0.627	$p < 0.01$
			SALT	0.699	—
Device	2010	2011-2014	GLMM	0.807	$p < 0.01$
			LR	0.848	$p < 0.01$
			SALT	0.882	—
	2010-2013	2014	GLMM	0.787	$p < 0.01$
			LR	0.805	$p < 0.01$
			SALT	0.866	—
Other	2010	2011-2014	GLMM	0.698	$p < 0.01$
			LR	0.706	$p < 0.01$
			SALT	0.724	—
	2010-2013	2014	GLMM	0.659	$p < 0.01$
			LR	0.678	$p < 0.01$
			SALT	0.824	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E13. Feature Selection based on Mutual Information (Analysis 3.3)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.565	$p < 0.01$
			LR	0.545	$p < 0.01$
			SALT	0.622	—
	2010-2013	2014	GLMM	0.556	$p < 0.01$
			LR	0.555	$p < 0.01$
			SALT	0.587	—
Surgery	2010	2011-2014	GLMM	0.727	$p < 0.01$
			LR	0.757	$p < 0.01$
			SALT	0.792	—
	2010-2013	2014	GLMM	0.728	$p < 0.01$
			LR	0.746	$p < 0.01$
			SALT	0.810	—
Infection	2010	2011-2014	GLMM	0.604	$p < 0.01$
			LR	0.617	$p < 0.01$
			SALT	0.658	—
	2010-2013	2014	GLMM	0.649	$p < 0.01$
			LR	0.676	$p < 0.01$
			SALT	0.699	—
Device	2010	2011-2014	GLMM	0.822	$p < 0.01$
			LR	0.830	$p < 0.01$
			SALT	0.853	—
	2010-2013	2014	GLMM	0.803	$p < 0.01$
			LR	0.822	$p < 0.01$
			SALT	0.851	—
Other	2010	2011-2014	GLMM	0.692	$p < 0.01$
			LR	0.701	$p < 0.01$
			SALT	0.736	—
	2010-2013	2014	GLMM	0.664	$p < 0.01$
			LR	0.650	$p < 0.01$
			SALT	0.812	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E14. Feature Selection based on the Joint Mutual Information Maximization Method (Analysis 3.4)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.564	$p < 0.01$
			LR	0.550	$p < 0.01$
			SALT	0.638	—
	2010-2013	2014	GLMM	0.594	$p < 0.01$
			LR	0.590	$p < 0.01$
			SALT	0.616	—
Surgery	2010	2011-2014	GLMM	0.780	$p < 0.01$
			LR	0.779	$p < 0.01$
			SALT	0.790	—
	2010-2013	2014	GLMM	0.793	$p < 0.01$
			LR	0.752	$p < 0.01$
			SALT	0.811	—
Infection	2010	2011-2014	GLMM	0.644	$p < 0.01$
			LR	0.640	$p < 0.01$
			SALT	0.658	—
	2010-2013	2014	GLMM	0.658	$p < 0.01$
			LR	0.646	$p < 0.01$
			SALT	0.672	—
Device	2010	2011-2014	GLMM	0.828	$p < 0.01$
			LR	0.883	$p < 0.01$
			SALT	0.897	—
	2010-2013	2014	GLMM	0.882	$p < 0.01$
			LR	0.812	$p < 0.01$
			SALT	0.902	—
Other	2010	2011-2014	GLMM	0.679	$p < 0.01$
			LR	0.711	$p < 0.01$
			SALT	0.716	—
	2010-2013	2014	GLMM	0.702	$p < 0.01$
			LR	0.717	$p < 0.01$
			SALT	0.851	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E15. Feature Selection based on the Double Input Symmetrical Relevance Method (Analysis 3.5)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.606	$p < 0.01$
			LR	0.622	$p < 0.01$
			SALT	0.698	—
	2010-2013	2014	GLMM	0.616	$p < 0.01$
			LR	0.598	$p < 0.01$
			SALT	0.647	—
Surgery	2010	2011-2014	GLMM	0.784	$p < 0.01$
			LR	0.808	$p < 0.01$
			SALT	0.836	—
	2010-2013	2014	GLMM	0.812	$p < 0.01$
			LR	0.768	$p < 0.01$
			SALT	0.837	—
Infection	2010	2011-2014	GLMM	0.650	$p < 0.01$
			LR	0.637	$p < 0.01$
			SALT	0.709	—
	2010-2013	2014	GLMM	0.696	$p < 0.01$
			LR	0.688	$p < 0.01$
			SALT	0.730	—
Device	2010	2011-2014	GLMM	0.904	$p < 0.01$
			LR	0.890	$p < 0.01$
			SALT	0.922	—
	2010-2013	2014	GLMM	0.837	$p < 0.01$
			LR	0.835	$p < 0.01$
			SALT	0.937	—
Other	2010	2011-2014	GLMM	0.715	$p < 0.01$
			LR	0.752	$p < 0.01$
			SALT	0.762	—
	2010-2013	2014	GLMM	0.698	$p < 0.01$
			LR	0.738	$p < 0.01$
			SALT	0.856	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E16. Feature Selection based on the Joint Mutual Information Method (Analysis 3.6)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.556	$p < 0.01$
			LR	0.511	$p < 0.01$
			SALT	0.616	—
	2010-2013	2014	GLMM	0.505	$p < 0.01$
			LR	0.574	$p < 0.01$
			SALT	0.594	—
Surgery	2010	2011-2014	GLMM	0.732	$p < 0.01$
			LR	0.708	$p < 0.01$
			SALT	0.771	—
	2010-2013	2014	GLMM	0.741	$p < 0.01$
			LR	0.709	$p < 0.01$
			SALT	0.769	—
Infection	2010	2011-2014	GLMM	0.611	$p < 0.01$
			LR	0.572	$p < 0.01$
			SALT	0.620	—
	2010-2013	2014	GLMM	0.597	$p < 0.01$
			LR	0.608	$p < 0.01$
			SALT	0.636	—
Device	2010	2011-2014	GLMM	0.808	$p < 0.01$
			LR	0.827	$p < 0.01$
			SALT	0.847	—
	2010-2013	2014	GLMM	0.814	$p < 0.01$
			LR	0.802	$p < 0.01$
			SALT	0.847	—
Other	2010	2011-2014	GLMM	0.657	$p < 0.01$
			LR	0.662	$p < 0.01$
			SALT	0.706	—
	2010-2013	2014	GLMM	0.685	$p < 0.01$
			LR	0.694	$p < 0.01$
			SALT	0.846	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E17. Feature Selection based on the Minimum Redundancy Maximal Relevance Method (Analysis 3.7)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.640	$p < 0.01$
			LR	0.618	$p < 0.01$
			SALT	0.672	—
	2010-2013	2014	GLMM	0.659	$p < 0.01$
			LR	0.606	$p < 0.01$
			SALT	0.712	—
Surgery	2010	2011-2014	GLMM	0.822	$p < 0.01$
			LR	0.770	$p < 0.01$
			SALT	0.872	—
	2010-2013	2014	GLMM	0.823	$p < 0.01$
			LR	0.843	$p < 0.01$
			SALT	0.856	—
Infection	2010	2011-2014	GLMM	0.682	$p < 0.01$
			LR	0.714	$p < 0.01$
			SALT	0.781	—
	2010-2013	2014	GLMM	0.701	$p < 0.01$
			LR	0.712	$p < 0.01$
			SALT	0.779	—
Device	2010	2011-2014	GLMM	0.890	$p < 0.01$
			LR	0.860	$p < 0.01$
			SALT	0.902	—
	2010-2013	2014	GLMM	0.906	$p < 0.01$
			LR	0.875	$p < 0.01$
			SALT	0.920	—
Other	2010	2011-2014	GLMM	0.764	$p < 0.01$
			LR	0.781	$p < 0.01$
			SALT	0.797	—
	2010-2013	2014	GLMM	0.733	$p < 0.01$
			LR	0.739	$p < 0.01$
			SALT	0.915	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E18. Addressing Data Imbalance through the Synthetic Minority Over-Sampling Technique (Analysis 4.1)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.656	$p < 0.01$
			LR	0.691	$p < 0.01$
			SALT	0.766	—
	2010-2013	2014	GLMM	0.694	$p < 0.01$
			LR	0.697	$p < 0.01$
			SALT	0.725	—
Surgery	2010	2011-2014	GLMM	0.883	$p < 0.01$
			LR	0.871	$p < 0.01$
			SALT	0.932	—
	2010-2013	2014	GLMM	0.907	$p < 0.01$
			LR	0.879	$p < 0.01$
			SALT	0.934	—
Infection	2010	2011-2014	GLMM	0.753	$p < 0.01$
			LR	0.742	$p < 0.01$
			SALT	0.813	—
	2010-2013	2014	GLMM	0.745	$p < 0.01$
			LR	0.785	$p < 0.01$
			SALT	0.808	—
Device	2010	2011-2014	GLMM	0.911	$p < 0.01$
			LR	0.976	$p < 0.01$
			SALT	0.991	—
	2010-2013	2014	GLMM	0.931	$p < 0.01$
			LR	0.960	$p < 0.01$
			SALT	0.994	—
Other	2010	2011-2014	GLMM	0.781	$p < 0.01$
			LR	0.837	$p < 0.01$
			SALT	0.850	—
	2010-2013	2014	GLMM	0.780	$p < 0.01$
			LR	0.819	$p < 0.01$
			SALT	0.950	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E19. Addressing Data Imbalance through Random Over-Sampling (Analysis 4.2)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.716	$p < 0.01$
			LR	0.699	$p < 0.01$
			SALT	0.754	—
	2010-2013	2014	GLMM	0.712	$p < 0.01$
			LR	0.661	$p < 0.01$
			SALT	0.734	—
Surgery	2010	2011-2014	GLMM	0.908	$p < 0.01$
			LR	0.880	$p < 0.01$
			SALT	0.919	—
	2010-2013	2014	GLMM	0.928	$p < 0.01$
			LR	0.924	$p < 0.01$
			SALT	0.945	—
Infection	2010	2011-2014	GLMM	0.812	$p < 0.01$
			LR	0.801	$p < 0.01$
			SALT	0.876	—
	2010-2013	2014	GLMM	0.792	$p < 0.01$
			LR	0.784	$p < 0.01$
			SALT	0.839	—
Device	2010	2011-2014	GLMM	0.954	$p < 0.01$
			LR	0.974	$p < 0.01$
			SALT	0.988	—
	2010-2013	2014	GLMM	0.951	$p < 0.01$
			LR	0.968	$p < 0.01$
			SALT	0.980	—
Other	2010	2011-2014	GLMM	0.796	$p < 0.01$
			LR	0.879	$p < 0.01$
			SALT	0.891	—
	2010-2013	2014	GLMM	0.854	$p < 0.01$
			LR	0.821	$p < 0.01$
			SALT	0.973	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E20. Addressing Data Imbalance through Random Under-Sampling (Analysis 4.3)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.659	$p < 0.01$
			LR	0.690	$p < 0.01$
			SALT	0.778	—
	2010-2013	2014	GLMM	0.732	$p < 0.01$
			LR	0.677	$p < 0.01$
			SALT	0.761	—
Surgery	2010	2011-2014	GLMM	0.880	$p < 0.01$
			LR	0.906	$p < 0.01$
			SALT	0.931	—
	2010-2013	2014	GLMM	0.915	$p < 0.01$
			LR	0.918	$p < 0.01$
			SALT	0.935	—
Infection	2010	2011-2014	GLMM	0.784	$p < 0.01$
			LR	0.792	$p < 0.01$
			SALT	0.844	—
	2010-2013	2014	GLMM	0.799	$p < 0.01$
			LR	0.780	$p < 0.01$
			SALT	0.823	—
Device	2010	2011-2014	GLMM	0.955	$p < 0.01$
			LR	0.978	$p < 0.01$
			SALT	0.992	—
	2010-2013	2014	GLMM	0.954	$p < 0.01$
			LR	0.927	$p < 0.01$
			SALT	0.973	—
Other	2010	2011-2014	GLMM	0.836	$p < 0.01$
			LR	0.838	$p < 0.01$
			SALT	0.866	—
	2010-2013	2014	GLMM	0.802	$p < 0.01$
			LR	0.807	$p < 0.01$
			SALT	0.946	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E21. Combining Best Strategies (Analysis 5)

AE Category	Training Period	Testing Period	Model	AUC	Delong Test H_a: AUC(Model) ≠ AUC(SALT)
ADE	2010	2011-2014	GLMM	0.666	$p < 0.01$
			LR	0.626	$p < 0.01$
			SALT	0.734	—
	2010-2013	2014	GLMM	0.645	$p < 0.01$
			LR	0.657	$p < 0.01$
			SALT	0.747	—
Surgery	2010	2011-2014	GLMM	0.820	$p < 0.01$
			LR	0.788	$p < 0.01$
			SALT	0.855	—
	2010-2013	2014	GLMM	0.839	$p < 0.01$
			LR	0.839	$p < 0.01$
			SALT	0.890	—
Infection	2010	2011-2014	GLMM	0.700	$p < 0.01$
			LR	0.741	$p < 0.01$
			SALT	0.739	—
	2010-2013	2014	GLMM	0.731	$p < 0.01$
			LR	0.721	$p < 0.01$
			SALT	0.765	—
Device	2010	2011-2014	GLMM	0.902	$p < 0.01$
			LR	0.923	$p < 0.01$
			SALT	0.943	—
	2010-2013	2014	GLMM	0.913	$p < 0.01$
			LR	0.879	$p < 0.01$
			SALT	0.931	—
Other	2010	2011-2014	GLMM	0.757	$p < 0.01$
			LR	0.809	$p < 0.01$
			SALT	0.831	—
	2010-2013	2014	GLMM	0.757	$p < 0.01$
			LR	0.804	$p < 0.01$
			SALT	0.913	—

Notes. ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from selected training periods—the shortest available period (2010 only) and the longest available (2010-2013) period—and results from two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with different training periods and other predictive models show consistent patterns and qualitatively similar conclusions as reported here. Test of difference in AUC is based on the DeLong et al [12].

Table E22. Assessing the Presence of Concept Drifting (Analysis 6)

AE Category	Training Period	Model	Test Data Period			
			2011	2012	2013	2014
ADE	2010	GLMM	0.544	0.594	0.615	0.631
		LR	0.589	0.650	0.645	0.664
		SALT	0.665	0.714	0.712	0.753
		Average	0.599	0.653	0.658	0.683
	2010-2011	GLMM	—	0.615	0.581	0.617
		LR	—	0.654	0.654	0.630
		SALT	—	0.705	0.737	0.751
		Average	—	0.658	0.657	0.666
	2010-2012	GLMM	—	—	0.647	0.633
		LR	—	—	0.654	0.682
		SALT	—	—	0.754	0.720
		Average	—	—	0.685	0.678
Surgery	2010	GLMM	0.899	0.840	0.802	0.849
		LR	0.940	0.853	0.857	0.903
		SALT	0.928	0.895	0.877	0.918
		Average	0.922	0.863	0.845	0.890
	2010-2011	GLMM	—	0.842	0.840	0.868
		LR	—	0.824	0.887	0.901
		SALT	—	0.921	0.941	0.926
		Average	—	0.862	0.889	0.898
	2010-2012	GLMM	—	—	0.836	0.847
		LR	—	—	0.870	0.851
		SALT	—	—	0.918	0.904
		Average	—	—	0.874	0.867
Infection	2010	GLMM	0.747	0.661	0.649	0.711
		LR	0.735	0.703	0.712	0.755
		SALT	0.772	0.738	0.760	0.781
		Average	0.751	0.700	0.707	0.749
	2010-2011	GLMM	—	0.696	0.669	0.654
		LR	—	0.731	0.684	0.757
		SALT	—	0.690	0.719	0.732
		Average	—	0.706	0.691	0.714
	2010-2012	GLMM	—	—	0.668	0.692
		LR	—	—	0.726	0.725
		SALT	—	—	0.783	0.791
		Average	—	—	0.726	0.736
Device	2010	GLMM	0.692	0.778	0.666	0.704
		LR	0.773	0.787	0.699	0.765
		SALT	0.717	0.767	0.688	0.745
		Average	0.728	0.777	0.684	0.738
	2010-2011	GLMM	—	0.774	0.690	0.728

		LR	—	0.834	0.732	0.753
		SALT	—	0.856	0.849	0.862
		Average	—	0.821	0.757	0.781
	2010-2012	GLMM	—	—	0.652	0.728
		LR	—	—	0.743	0.779
		SALT	—	—	0.891	0.916
		Average	—	—	0.762	0.808
Other	2010	GLMM	0.788	0.798	0.735	0.744
		LR	0.858	0.835	0.830	0.854
		SALT	0.878	0.886	0.811	0.822
		Average	0.841	0.839	0.792	0.807
	2010-2011	GLMM	—	0.829	0.771	0.807
		LR	—	0.884	0.770	0.860
		SALT	—	0.871	0.831	0.907
		Average	—	0.861	0.791	0.858
	2010-2012	GLMM	—	—	0.745	0.812
		LR	—	—	0.787	0.826
		SALT	—	—	0.912	0.859
		Average	—	—	0.815	0.832

Note 1: By definition, concept drift is an issue when a model trained using historical observations cannot reliably predict future observations because the patterns learned from historical observations can no longer be adequately applied to future observations. Following this logic, one way to examine whether concept drifting is present in our setting is to compare a model's accuracy when predicting AEs among testing cases that are 1, 2, 3, and 4 years away from the training periods. For example, we can train a model using inpatient observations from 2010 and 2011 and test the accuracy of the model using inpatient observations from 2012, 2013, and 2014. If we observe a decrease in predictive accuracy going from the 2012 test sample, to the 2013 test sample, and to the 2014 test sample, this would suggest the concept drifting is present in our setting. On the other hand, if the predictive accuracy does not decrease, this would suggest that concept drifting is not present in our setting. Overall, we do not find a significant sign of AUC deterioration when the test sample is further away from the training period, suggesting no evidence of concept drifting in our AE prediction setting.

Note 2: ADEs = adverse drug events; AE = adverse event; AUC = area under the curve; GLMM = generalized linear mixed model; LR = lasso regression. To improve the readability of the results, we only report results from SALT and two alternative models: GLMM, which is the best alternative model from evaluation 1, and LR, which is the best alternative model from evaluation 2. The unreported results with other predictive models show consistent patterns and qualitatively similar conclusions as reported here.

References

1. Abbasi, A.; Li, J.; Adjeroh, D.; Abate, M.; and Zheng, W. Don't mention it? Analyzing user-generated content signals for early adverse event warnings. *Information Systems Research, 30*, 3 (2019); 1007–1028.
2. Aron, R.; Dutta, S.; Janakiraman, R.; and Pathak, P.A. The impact of automation of systems on medical errors: evidence from field research. *Information Systems Research, 22*, 3 (2011); 429–446.
3. Barandela, R.; Valdovinos, R.M.; Sánchez, J.S.; and Ferri, F.J. The imbalanced training sample problem: under or over sampling? In A. Fred, T.M. Caelli, R.P.W. Duin, A.C. Campilho and D. de Ridder, eds., *Structural, Syntactic, and Statistical Pattern Recognition*. Springer, Berlin, Heidelberg, 2004; pp. 806–814.
4. Bardhan, I.; Chen, H.; and Karahanna, E. Connecting systems, data, and people: a multidisciplinary research roadmap for chronic disease management. *MIS Quarterly, 44*, 1 (2020); 185–200.
5. Barnett, V. and Lewis, T. *Outliers in Statistical Data*. Wiley, Chichester, NY, 1994.
6. Benaroch, M. and Chernobai, A. Operational IT failures, IT value destruction, and board-level IT governance changes. *MIS Quarterly, 41*, 3 (2017); 729–762.
7. Bennasar, M.; Hicks, Y.; and Setchi, R. Feature selection using joint mutual information maximisation. *Expert Systems with Applications, 42*, 22 (2015); 8520–8532.
8. Buchlak, Q.D.; Yanamadala, V.; Leveque, J.-C.; Edwards, A.; Nold, K.; and Sethi, R. The Seattle spine score: Predicting 30-day complication risk in adult spinal deformity surgery. *Journal of Clinical Neuroscience, 43*, (2017); 247–255.
9. Chandrashekar, G. and Sahin, F. A survey on feature selection methods. *Computers & Electrical Engineering, 40*, 1 (2014); 16–28.
10. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; and Kegelmeyer, W.P. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, (2002); 321–357.
11. Clemons, E.K. and Hitt, L.M. Poaching and the misappropriation of information: transaction risks of information exchange. *Journal of Management Information Systems, 21*, 2 (2004); 87–107.
12. DeLong, E.R.; DeLong, D.M.; and Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics, 44*, 3 (1988); 837–845.
13. Dodson, J.A.; Reynolds, M.R.; Bao, H.; Al-Khatib, S.M.; Peterson, E.D.; et al. Developing a risk model for in-hospital adverse events following implantable cardioverter-defibrillator implantation: a report from the NCDR (National Cardiovascular Data Registry). *Journal of the American College of Cardiology, 63*, 8 (2014); 788–796.
14. Dougherty, J.; Kohavi, R.; and Sahami, M. Supervised and unsupervised discretization of continuous features. In A. Prieditis and S. Russell, eds., *Machine Learning Proceedings 1995*. Morgan Kaufmann, San Francisco (CA), 1995; pp. 194–202.
15. Ehlers, A.P.; Roy, S.B.; Khor, S.; Mandagani, P.; Maria, M.; et al. Improved risk prediction following surgery using machine learning algorithms. *EGEMS, 5*, 2 (2017); 1–13.
16. Feblowitz, J.; Henkin, S.; Pang, J.; Ramelson, H.; Schneider, L.; et al. Provider Use of and Attitudes Towards an Active Clinical Alert. *Applied Clinical Informatics, 4*, 1 (2013); 144–152.
17. Ferdousi, R.; Safdari, R.; and Omidi, Y. Computational prediction of drug-drug interactions

based on drugs functional similarities. *Journal of Biomedical Informatics*, *70*, (2017);54–64.
18. Fichman, R.G. Real options and IT platform adoption: implications for theory and practice. *Information Systems Research*, *15*, 2 (2004); 132–154.
19. Fichman, R.G.; Kohli, R.; and Krishnan, R. The role of information systems in healthcare: current research and future trends. *Information Systems Research*, *22*, 3 (2011); 419–428.
20. Forman, G. An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, *3*, Mar (2003); 1289–1305.
21. Genovese, E.A.; Fish, L.; Chaer, R.A.; Makaroun, M.S.; and Baril, D.T. Risk stratification for the development of respiratory adverse events following vascular surgery using the Society of Vascular Surgery's Vascular Quality Initiative. *Journal of Vascular Surgery*, *65*, 2 (2017); 459–470.
22. Geraci, J.M.; Rosen, A.K.; Ash, A.S.; McNiff, K.J.; and Moskowitz, M.A. Predicting the occurrence of adverse events after coronary artery bypass surgery. *Annals of Internal Medicine*, *118*, 1 (1993); 18–24.
23. Gupta, A.; Mehrotra, K.G.; and Mohan, C. A clustering-based discretization for supervised learning. *Statistics & Probability Letters*, *80*, 9 (2010); 816–824.
24. Hydari, M.Z.; Telang, R.; and Marella, W.M. Saving patient Ryan—can advanced electronic medical records make patient care safer? *Management Science*, *65*, 5 (2018); 2041–2059.
25. Kohli, R. and Tan, S. Electronic health records: how can IS researchers contribute to transforming healthcare? *MIS Quarterly*, *40*, 3 (2016); 553–572.
26. Krone, R.J.; Laskey, W.K.; Johnson, C.; Kimmel, S.E.; Klein, L.W.; et al. A simplified lesion classification for predicting success and complications of coronary angioplasty. *The American Journal of Cardiology*, *85*, 10 (2000); 1179–1184.
27. Kusy, M.; Obrzut, B.; and Kluska, J. Application of gene expression programming and neural networks to predict adverse events of radical hysterectomy in cervical cancer patients. *Medical & Biological Engineering & Computing*, *51*, 12 (2013); 1357–1365.
28. Lin, Y.-K.; Chen, H.; Brown, R.; and Li, S.-H. Healthcare predictive analytics for risk profiling in chronic care: a Bayesian multitask learning approach. *MIS Quarterly*, *41*, 2 (2017); 473–495.
29. Menon, N.M. and Kohli, R. Blunting Damocles' sword: A longitudinal model of healthcare IT impact on malpractice insurance premium and quality of patient care. *Information Systems Research*, *24*, 4 (2013); 918–932.
30. Meyer, G.; Adomavicius, G.; Johnson, P.E.; Elidrisi, M.; Rush, W.A.; et al. A machine learning approach to improving dynamic decision making. *Information Systems Research*, *25*, 2 (2014); 239–263.
31. Meyer, P.E. and Bontempi, G. On the use of variable complementarity for feature selection in cancer classification. In F. Rothlauf, J. Branke, S. Cagnoni, E. Costa, C. Cotta, et al., eds., *Applications of Evolutionary Computing*. Springer, Berlin, Heidelberg, 2006; pp. 91–102.
32. Mortazavi, B.J.; Desai, N.; Zhang, J.; Coppi, A.; Warner, F.; et al. Prediction of adverse events in patients undergoing major cardiovascular procedures. *IEEE Journal of Biomedical and Health Informatics*, *21*, 6 (2017); 1719–1729.
33. Mufti, H.N.; Hirsch, G.M.; Abidi, S.R.; and Abidi, S.S.R. Exploiting machine learning algorithms and methods for the prediction of agitated delirium after cardiac surgery: models development and validation study. *JMIR medical informatics*, *7*, 4 (2019); e14993.
34. Ozdemir, Z.; Barron, J.; and Bandyopadhyay, S. An analysis of the adoption of digital health records under switching costs. *Information Systems Research*, *22*, 3 (2011); 491–503.

35. Park, I.; Sharman, R.; and Rao, H. Disaster experience and hospital information systems: an examination of perceived information assurance, risk, resilience, and HIS usefulness. *MIS Quarterly*, *39*, 2 (2015); 317–344.
36. Peng, H.; Long, F.; and Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*, 8 (2005); 1226–1238.
37. Pinsonneault, A.; Addas, S.; Qian, C.; Dakshinamoorthy, V.; and Tamblyn, R. Integrated health information technology and the quality of patient care: a natural experiment. *Journal of Management Information Systems*, *34*, 2 (2017); 457–486.
38. Ratliff, J.K.; Balise, R.; Veeravagu, A.; Cole, T.S.; Cheng, I.; et al. Predicting occurrence of spine surgery complications using "big data" modeling of an administrative claims database. *JBJS*, *98*, 10 (2016); 824–834.
39. Rosen, A.K.; Geraci, J.M.; Ash, A.S.; McNiff, K.J.; and Moskowitz, M.A. Postoperative adverse events of common surgical procedures in the Medicare population. *Medical Care*, *30*, 9 (1992); 753–765.
40. Schwilk, B.; Muche, R.; Treiber, H.; Brinkmann, A.; Georgieff, M.; and Bothner, U. A cross-validated multifactorial index of perioperative risks in adults undergoing anaesthesia for non-cardiac surgery. *Journal of Clinical Monitoring and Computing*, *14*, 4 (1998); 283–294.
41. Seidling, H.M.; Phansalkar, S.; Seger, D.L.; Paterno, M.D.; Shaykevich, S.; et al. Factors influencing alert acceptance: a novel approach for predicting the success of clinical decision support. *Journal of the American Medical Informatics Association*, *18*, 4 (2011); 479–484.
42. Shah, N.R.; Seger, A.C.; Seger, D.L.; Fiskio, J.M.; Kuperman, G.J.; et al. Improving acceptance of computerized prescribing alerts in ambulatory care. *Journal of the American Medical Informatics Association*, *13*, 1 (2006); 5–11.
43. Steeg, L. van de; Langelaan, M.; and Wagner, C. Can preventable adverse events be predicted among hospitalized older patients? The development and validation of a predictive model. *International Journal for Quality in Health Care*, *26*, 5 (2014); 547–552.
44. Velu, C.K.; Madnick, S.E.; and Alstyne, M.W.V. Centralizing data management with considerations of uncertainty and information-based flexibility. *Journal of Management Information Systems*, *30*, 3 (2013); 179–212.
45. Vries, E.N. de; Ramrattan, M.A.; Smorenburg, S.M.; Gouma, D.J.; and Boermeester, M.A. The incidence and nature of in-hospital adverse events: a systematic review. *Quality and Safety in Health Care*, *17*, 3 (2008); 216–223.
46. Wang, K.; Weng, Z.; Sun, L.; Sun, J.; Zhou, S.-F.; and He, L. Systematic drug safety evaluation based on public genomic expression (Connectivity Map) data: myocardial and infectious adverse reactions as application cases. *Biochemical and Biophysical Research Communications*, *457*, 3 (2015); 249–255.
47. Wang, T.; Kannan, K.N.; and Ulmer, J.R. The association between the disclosure and the realization of information security risk factors. *Information Systems Research*, *24*, 2 (2013); 201–218.
48. Yang, H. and Moody, J. Data visualization and feature selection: new algorithms for nongaussian data. *Advances in Neural Information Processing Systems*, *12*, (1999); 687–693.
49. Yeo, I. and Johnson, R.A. A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*, 4 (2000); 954–959.