

HEALTHCARE PREDICTIVE ANALYTICS FOR RISK PROFILING IN CHRONIC CARE: A BAYESIAN MULTITASK LEARNING APPROACH¹

Yu-Kai Lin

College of Business, Florida State University, Tallahassee, FL 32306 U.S.A. {ylin@business.fsu.edu}

Hsinchun Chen

Eller College of Management, University of Arizona, Tucson, A 895721 U.S.A. {hchen@eller.arizona.edu}

Randall A. Brown

Southern Arizona VA Health Care System, Tucson, AZ 85723 U.S.A. {Randall.Brown4@va.gov}

Shu-Hsing Li

College of Management, National Taiwan University, Taipei City 106, TAIWAN {shli@ntu.edu.tw}

Hung-Jen Yang

The Walter H. Shorenstein Asia-Pacific Research Center, Stanford University,
Stanford, CA 94305 U.S.A. {berniejohan@gmail.com}

Clinical intelligence about a patient's risk of future adverse health events can support clinical decision making in personalized and preventive care. Healthcare predictive analytics using electronic health records offers a promising direction to address the challenging tasks of risk profiling. Patients with chronic diseases often face risks of not just one, but an array of adverse health events. However, existing risk models typically focus on one specific event and do not predict multiple outcomes. To attain enhanced risk profiling, we adopt the design science paradigm and propose a principled approach called Bayesian multitask learning (BMTL). Considering the model development for an event as a single task, our BMTL approach is to coordinate a set of baseline models—one for each event—and communicate training information across the models. The BMTL approach allows healthcare providers to achieve multifaceted risk profiling and model an arbitrary number of events simultaneously. Our experimental evaluations demonstrate that the BMTL approach attains an improved predictive performance when compared with the alternatives that model multiple events separately. We also find that, in most cases, the BMTL approach significantly outperforms existing multitask learning techniques. More importantly, our analysis shows that the BMTL approach can create significant potential impacts on clinical practice in reducing the failures and delays in preventive interventions. We discuss several implications of this study for health IT, big data and predictive analytics, and design science research.

Keywords: Design science, healthcare predictive analytics, Bayesian data analysis, multitask learning, electronic health records, health IT

¹H. Rao was the accepting senior editor for this paper. Huimin Zhao served as the associate editor.

The appendices for this paper are located in the "Online Supplements" section of the *MIS Quarterly's* website (<http://www.misq.org>).

Introduction

The prevalence and the growth rate of patients with chronic diseases are an alarming global phenomenon in many developed and developing countries. According to the World Health Organization (2014a, 2014b), most new cases and mortalities of diabetes and cardiovascular diseases occur in low- and middle-income countries. In the United States, more than 140 million Americans live with one or more chronic conditions, and the population is expected to grow by at least 10 million new cases per decade (Anderson 2010). Meanwhile, the costs of chronic care have also been escalating over the last decades, currently accounting for 86% of health care cost in the United States (Gerteis et al. 2014). The surging cases and costs make patients, clinical experts, and health policymakers around the world believe that effective interventions are needed to prevent, detect, and manage chronic diseases and their sequelae.

With increased adoption of electronic health record (EHR) systems in clinical practices, EHR data analytics for advanced clinical decision support is attracting both scientific and practical interest (Agarwal et al. 2010; Dixon-Woods et al. 2013). Clinical intelligence about a patient's risks of future adverse health events has been a key element for effective decision making in chronic care. This is because patients with chronic diseases often develop complications and comorbidities in their disease course. For instance, patients with diabetes often have higher risks of stroke, heart diseases, eye problems, and renal failure (Centers for Disease Control and Prevention 2014). Similarly, chronic kidney disease can lead to anemia, cardiovascular events, and mortality (Thomas et al. 2008). Accurate predictions of future adverse health events could enable clinicians to take preventive and personalized interventions, which in turn could reduce patients' risks and improve their quality of life.

The value of predictive analytics in healthcare has been repeatedly emphasized in prior information systems (IS) research. As noted by Agarwal and Dhar (2014), healthcare is a domain in which prediction is perhaps more important than explanation, considering the daunting cost of delay in diagnosis and treatment. Chen et al. (2012) discuss the potential of EHR-based healthcare analytics for "smart health and wellbeing" from the perspective of business intelligence. By the same token, after exploring the extant healthcare IS research, Fichman et al. (2011, p. 425) suggest that

Another emerging avenue for knowledge discovery arises from using digital technology to enable new kinds of mathematical healthcare modeling and simulations...use of healthcare analytics tools and how they should be integrated with electronic health records warrants future research attention.

Developing and utilizing information technology (IT) artifacts, such as models, techniques, and systems, to address practical needs has been a focus of IS research since the inception of the discipline (Hevner et al. 2004). This stream of design research is becoming particularly important and relevant given the surging interest in big data and predictive analytics research (Chen et al. 2012; Shmueli and Koppius 2011). The research motivations are often to obtain valuable insights through the development of advanced analytics techniques and the use of large and rich data sources that were previously unavailable or underutilized. For example, Bao and Datta (2014) developed a text analysis method to analyze nearly 15,000 corporate risk disclosures, which has applications in financial accounting. Similarly, Fang et al. (2013) developed a naïve Bayesian method to predict behaviors in large social networks with tens of thousands of users.

Consistent with the design science paradigm and the recent IS research on big data analytics, we develop and evaluate a novel data analytics approach in the context of healthcare. To our knowledge, this is the first IS study to focus on EHR analytics. Our goal is to improve clinical decision making and facilitate preventive and personalized care with data analytics. Specifically, we harvest big EHR data and develop a Bayesian multitask learning (BMTL) approach to predict patients' risks of adverse health events. Delay or failure to provide preventive interventions is one of the common medical errors and results in numerous deaths (Kohn et al. 2000). The EHR-driven BMTL approach could augment healthcare providers' capability in identifying high-risk patients for timely interventions.

The proposed BMTL approach is distinctly different from the existing risk models. Existing healthcare predictive analytics research often focuses on modeling one specific event or outcome.² However, it is rare, especially in chronic care, that patients only face one type of risk. Multifaceted risk profiling with different events or outcomes would provide healthcare professionals greater clinical insights toward a comprehensive and effective care plan. Technically, we can just construct an array of independent risk models, one for each adverse outcome. While this approach is simple and straightforward, it neglects the fact that a patient's risk to one event (say, stroke) is often correlated with his or her risk to other events (say, heart attack). Information contained in one model may be useful to other relevant models, as long as we can coordinate multiple models in a unified modeling framework.

²Healthcare predictive analytics has been phrased in numerous ways in the literature, including prognosis, clinical predictive modeling, and health informatics, among others. We choose to use the phrase "healthcare predictive analytics" throughout this paper as it is more expressive to the IS audience.

This is analogous to the economic effect of knowledge transfers or spillovers, in which the net effect of a group is greater than the sum of that of the individuals (Gupta 2008). We postulate the existence of such spillover effect across individual models in a joint modeling framework and develop a principled approach to exploit this model spillover effect to improve learning performance. Considering risk prediction for a specific adverse health event as a single machine learning task, the key aspect of our approach is to *obtain an improved predictive performance for each individual task by learning multiple related tasks jointly and simultaneously*. Our approach is in sharp contrast with the existing healthcare predictive analytics literature in which the research either considers only one clinical event (e.g., Brownstein et al. 2010), or models multiple clinical events as completely independent tasks, for example the UKPDS Risk Engine (Kothari et al. 2002; Stevens et al. 2001). Taken together, we are interested in studying the following:

- What are a patient's risks to an array of adverse health events?
- How can we model multiple risks simultaneously?
- Does simultaneous learning of multiple event risks improve the overall predictive performance of each event risk?

We chose diabetes as our research case and performed risk profiling on three common, and often fatal, adverse health events: stroke, acute myocardial infarction, and acute renal failure. Our experiments confirmed the postulated model spillover effect. The proposed BMTL approach achieved significantly improved predictive performance for each event compared with independent models that consider the events separately. The BMTL approach also demonstrated competitive and often superior performance in a head-to-head comparison with other multitask learning approaches in the literature. More importantly, our evaluation results provided evidence that the BMTL approach can lead to interventions that reduce risks of the three adverse health events beyond what would occur without the predictions. These findings, along with the artifact, have implications for several areas of IS research, including health IT, big data and predictive analytics, and design science.

The rest of the paper is organized as follows. In the next section, we review related work in healthcare predictive analytics and multitask learning. We then describe the proposed BMTL approach and contrast it with the existing techniques. Following that, we outline a set of experiments and present their results regarding the viability and utility of the BMTL approach. In the final section, we discuss the

contributions of this study to the IS knowledge base, the practical implications of enhanced risk profiling, and directions for future work.

Research Background

Healthcare Predictive Analytics

Healthcare predictive analytics aims to predict future health-related outcomes or events based on clinical and/or non-clinical patterns in the data. The outcomes of interest in healthcare predictive analytics, such as medical complications (Stevens et al. 2001), hospital readmissions (Bardhan et al. 2014), treatment responses (Meyer et al. 2014), and patient mortality (Tabak et al. 2014), are often of great practical importance. While predictive analytics may be used to inform causal inference, the primary goal of prediction models is *not* to unbiasedly explain whether a factor contributes to an outcome, but to predict the outcome in new observations as accurately as possible (Moons et al. 2009; Shmueli and Koppius 2011). This important difference between prediction research and explanatory research drives distinctive principles for model development and evaluation given that *explanatory power does not imply predictive power* (Shmueli and Koppius 2011, p. 553).

There are generally two ways a healthcare predictive model can be developed. One is commonly seen in the medical field in which patient data are purposely collected in clinical trials with a set of predefined protocols. For instance, Tammemägi et al. (2013) developed a risk prediction model for the diagnosis of lung cancer using trial data. There are also several major cohort studies and trials on cardiovascular diseases and diabetes, such as the Framingham Heart Study (D'Agostino et al. 2008) and the UK Prospective Diabetes Study (UKPDS) (Stevens et al. 2001). In addition to clinical trials, the other way to develop a healthcare predictive model is to use existing data that have been routinely collected in clinical practice, such as EHRs, insurance claims, and clinical registries. The surging interest of healthcare predictive analytics in recent years is largely driven by the increasing availability of these data. For example, Tabak et al. (2014) use laboratory test results and diagnoses in EHRs to develop an inpatient mortality predictive model with excellent performance. Similarly, Bardhan et al. (2014) develop their readmission model using the admission data from a regional data exchange registry.

While healthcare predictive analytics can support clinical decisions, actual use of predictive models in clinical practice remains limited (Moons et al. 2009). The barriers for wide-

spread use of predictive models in healthcare include (1) inadequate integration with existing clinical workflow, (2) requiring variables that are expensive to obtain or not immediately accessible, and (3) the need to adapt the models from the study population to the local population. Some of these barriers could be effectively mitigated with the implementation of EHR systems (Moons et al. 2009; Toll et al. 2008). This is because the advent of EHRs not only provides accessible and local cohort data for healthcare predictive analytics, but also offers a platform that seamlessly embeds a predictive model into the clinical workflow.

There is an increased interest in understanding the interface between predictive analytics and decision making. For instance, Meyer et al. (2014) recently proposed a principled machine learning approach for dynamic decision making through the lens of control theory. Their PREDiction of Control Errors in Dynamic Contexts (PROCEDO) approach repeatedly iterates predictions of operation failures (with a C4.5 decision tree) and adjustments of control strategies (based on domain expert's judgment). Our BMTL approach is different because we emphasize improving predictive accuracy rather than modifying the strategies for actions and decision making. This is because once an accurate prediction is made, clinical guidelines usually provide clear strategies for actions and interventions.

As we mentioned, most of the extant healthcare predictive analytics studies are committed to one specific event, and develop models to best capture the characteristics of the event. For example, Bardhan et al. (2014) investigated the readmissions of patients with congestive heart failure, and developed a model to answer whether, when, and how often the patients would have been readmitted. The UKPDS Risk Engine can predict coronary heart disease and stroke in patients with type 2 diabetes (Kothari et al. 2002; Stevens et al. 2001), but these two types of predictions are effectively two independent predictive models. Multiple comorbidities are a common phenomenon among patients with chronic diseases. We hence are motivated to achieve risk profiling with multiple events being considered and modeled simultaneously.

Very few prior studies of healthcare predictive analytics consider multifaceted risk profiling. The closest study to ours is perhaps that by Smith and Mezhir (2014), in which the authors developed a two-part model to predict lymph node ratio and survival in pancreatic cancer patients. Lymph node ratio is a strong predictor on cancer survival, but it is typically unobservable and needs to be estimated. The two-part model first uses a logistic regression to predict lymph node ratio in pancreatic cancer patients and then passes the predicted ratio as an input for a Cox regression. This approach is different

from ours in model design and application. In terms of model design, we shall see later in the model development section that our BMTL approach does not have such a sequential dependency, but instead uses a hierarchical correlation structure to coordinate among multiple baseline models. In terms of applications, BMTL addresses multiple adverse health events in chronic care whereas Smith and Mezhir emphasize only one outcome: cancer survival. Taken together, our BMTL enables a more flexible and holistic approach for multifaceted risk profiling.

Multitask Learning

Multitask learning is a machine learning strategy in which multiple related tasks are trained jointly instead of independently with the goal to improve the overall performance of learning (Caruana 1997).³ Figure 1 provides schematic representations of single-task learning and multitask learning. In multitask learning there is a shared computational structure to tie individual tasks together in a unified training process. This allows training signals to be passed across models as an *inductive bias* (Baxter 2000; Caruana 1997), which in turn improves the generalizability of each trained model. Indeed, research in machine learning (Bishop 2007), statistics (Tibshirani 1996), and artificial intelligence (Mitchell 1982) has suggested the critical and necessary role of biases in the generalizability of learning and prediction. Shmueli and Koppius (2011) also commented on the tradeoff between bias and variance when discussing predictive analytics research in IS.

The literature shows three general approaches to achieve multitask learning. The first approach is through sharing common hidden nodes in artificial neural networks (ANNs) (Bakker and Heskes 2003; Caruana 1997). The architecture of an ANN typically has one input layer, one output layer, and an arbitrary number of hidden layers between the input and the output layers. One can achieve multitask learning by configuring each task as a node in the output layer (Caruana 1997). In doing so, these output nodes receive the same inputs from the nodes in the hidden layer but each with different, task-specific input weights. This approach is most straightforward but the baseline model has to be an ANN. The second approach to implement multitask learning is to minimize (or maximize) an appropriate regularization function over all tasks. This approach can be applied to a wide array of baseline models such as regressions (Huang et al. 2012), support vector machines (SVMs) (Cai and Cherkassky

³We should note that multitask learning is different from ensemble learning. They are similar because both utilize multiple models in the learning process. However, ensemble learning considers only one outcome variable whereas multitask learning considers multiple outcome variables simultaneously.

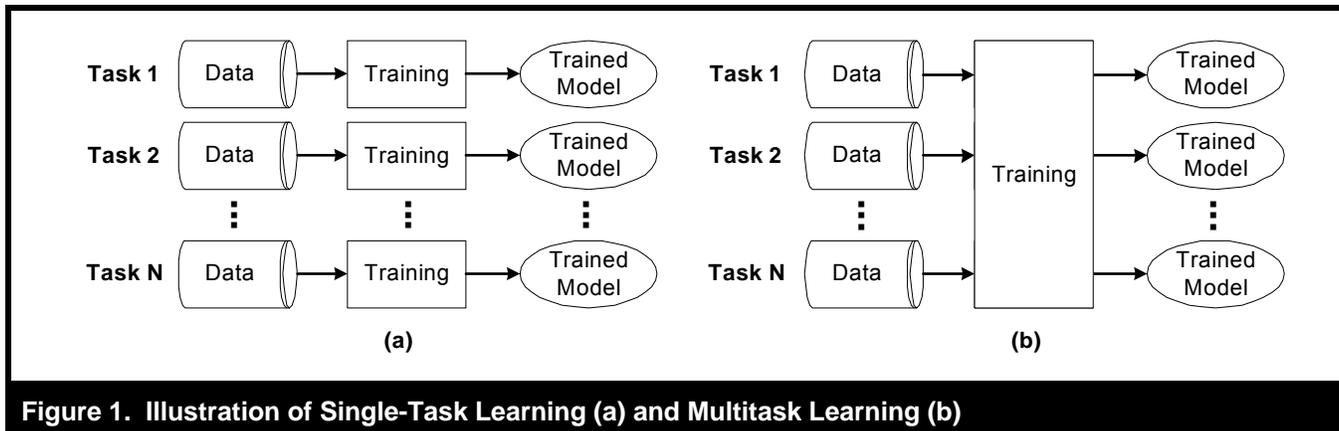


Figure 1. Illustration of Single-Task Learning (a) and Multitask Learning (b)

2012), tree-based models (Simm et al. 2014), and others as long as the learning can be formulated as an optimization problem. However, a complication from having a global regularization function is that tasks often need to be at least moderately positively correlated, otherwise multitask learning could lead to decreased performance compared to single-task learning—a problem known as “negative transfer” (Pan and Yang 2010). The third approach for multitask learning is to impose common prior distributions over tasks in a Bayesian framework (Archambeau et al. 2011; Xue et al. 2007). This approach is perhaps the most flexible one considering that nearly every statistical or machine learning model (including ANNs, SVMs, decision trees, and regressions) can have a Bayesian representation (see Chipman et al. 2002; Neal 1996; Tipping 2001). This approach allows an elaborative structure to transfer information across baseline models and can effectively eliminate the negative transfer issue because unrelated or negatively related tasks are truthfully reflected in the models. The main challenge for the Bayesian approach is that models with even moderate complexity often do not have an analytical solution, making numerical simulation the only route to fit the model.

Multitask learning is an underutilized modeling strategy in the research and practice of healthcare predictive analytics. Most existing multitask learning models were developed for small scale applications, such as text or image classification (Baxter 2000; Pan and Yang 2010). Recently, a small number of healthcare predictive analytics studies have started to leverage the multitask learning strategy. For example, Zhou et al. (2011) formulated cognitive scores of a patient with Alzheimer’s disease at different time points as a multitask learning prediction problem. Singh et al. (2014) presented a similar construction to predict renal function over time. However, the foci of these studies are still just one specific patient outcome. Developing and utilizing multitask learning

strategies to predict multiple patient outcomes remains a research gap in the literature that we aim to fill.

Summary

As the adoption of EHR systems accelerates, EHR-based healthcare predictive analytics is becoming an emerging research area with significant practical values (Chen et al. 2012). While patients with chronic diseases often face higher risks of many adverse health events, the extant research on risk profiling failed to consider the multifaceted nature of risks and focuses on only one specific adverse health event at a time. Modeling risks of multiple adverse health events not only provides better clinical intelligence for comprehensive preventive interventions, but also has the potential to achieve improved predictive performance for each event. However, multifaceted risk profiling is scarce in the extant healthcare predictive analytics research. This is perhaps due to lack of awareness as well as lack of techniques in this area. Research on multitask learning suggests an effective strategy to formulate a unified predictive analytics framework for multiple events and outcomes. We leverage the multitask learning strategy to develop a novel, principled approach to simultaneously model and predict multiple future health events, as discussed in the ensuing section.

Model Development

We first briefly describe single-task learning logistic regression models that have been widely used in prior healthcare predictive analytics research. We then describe the proposed Bayesian multitask learning (BMTL) approach for logistic regression models. Following that, we discuss the generaliz-

ability of our BMTL approach to other baseline models and contrast our approach with existing techniques.

Single-Task Learning with Logistic Regression Models

Given N patients, we are interested in modeling their risks of K different future adverse health events based on available information in EHRs. EHR data contain outcomes of these adverse events as well as the covariates from each patient at any point in time. We can carry out EHR-based risk profiling with the following procedure. We first choose a specific point of time in each patient’s medical history (henceforth denoted by v_{0i}) and then predict whether the patient will experience the K different adverse health events in the next w years. Given a specific w , we let $y_i^{(k)} \in \{0,1\}$ denote patient i ’s observed outcome of event k between v_{0i} and $v_{0i} + w$ years. We use $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iJ}]^T$ to denote a vector containing J predictors, which represent known characteristics of the patient at v_{0i} . Intuitively, $y_i^{(k)}$ follows a Bernoulli distribution; that is, $y_i^{(k)} \sim \text{Bernoulli}(\theta_i^{(k)})$, where $\theta_i^{(k)}$ is the probability of event k given \mathbf{x}_i .

A logistic regression model identifies the relation between $\theta_i^{(k)}$ and \mathbf{x}_i through a logit function. In the single-task learning paradigm, risks of the K events may be modeled as the following:⁴

$$\begin{aligned} \text{logit}(\theta_i^{(1)}) &= \alpha^{(1)} + \sum_{j=1}^J \beta_j^{(1)} x_i \\ \text{logit}(\theta_i^{(2)}) &= \alpha^{(2)} + \sum_{j=1}^J \beta_j^{(2)} x_i \\ &\vdots \\ \text{logit}(\theta_i^{(K)}) &= \alpha^{(K)} + \sum_{j=1}^J \beta_j^{(K)} x_i \end{aligned} \tag{1}$$

or, in a more compact representation,

$$\text{logit}(\theta_i^{(k)}) = \alpha^{(k)} + \sum_{j=1}^J \beta_j^{(k)} x_i, \quad k = 1, \dots, K; i = 1, \dots, N \tag{2}$$

⁴We may consider more sophisticated models than the one presented here. Given that the aim of this paper is to investigate the feasibility, utility, and generalizability of a principled approach (BMTL), we use simpler models for ease of exposition. Extensions such as patient- or physician-level fixed effects or heterogeneity are straightforward and have minimal effects on the implementation of our BMTL approach described here.

In (1) and (2), $\text{logit}(z) = \log(z/(1 - z))$ is a logit function, and $\alpha^{(k)}$ and $\beta_j^{(k)}$ ’s are event-specific intercepts and coefficients. We can then predict whether patient i will experience events 1 through K based on the respective $\theta_i^{(k)}$ in the above system of equations. Notice that these equations do not have explicit relations with other, and each baseline logistic regression is estimated independently in the single-task learning paradigm.

BMTL for Logistic Regression Models

Using the baseline logistic regression models specified as the ones in the previous section, we now describe how to model $\text{logit}(\theta_i^{(k)})$ with the proposed BMTL approach. Our key methodological innovation and contribution to the literature is that in BMTL we formulate a unique hierarchical correlation structure across different tasks. Assuming all tasks have the same set of J predictors, we consider the correlations of the regression coefficients of a predictor across tasks. As shown in Figure 2, we achieve this by explicitly modeling the correlation matrix $(\Omega_j, j = 1, \dots, J)$ for each of the regression coefficients corresponding to a particular predictor.⁵ In doing so, the training of each $\beta_j^{(k)}$ involves not only information within a specific task but also information from other tasks through the respective correlation matrix.

The structure of a BMTL logistic regression model can be represented by a plate diagram as shown in Figure 3. Table 1 provides a description for each of the parameters in Figure 3. Consistent with the conventions of a plate diagram, the symbol at the upper-right corner of each plate (rectangle) indicates the number of nodes in the respective plate, and the single- and double-bordered nodes are used to represent stochastic and deterministic (given their parent nodes) parameters, respectively. To summarize, Figure 3 shows that the probability of event occurrence (θ) is determined by the intercepts (α) and coefficients (β) in a set of K logistic regression models. The regression coefficients for the j^{th} predictor across tasks (β_j) follow a multivariate normal (MVN) distribution with zero means and a scaled covariance matrix $r_j^2 \Sigma_j$. The r_j is shrinkage scalar, and it is used to regulate the original covariance matrix Σ_j . We parameterized Σ_j as follows: $\Sigma_j = \text{diag}(\sigma_j) * \Omega_j * \text{diag}$, where σ_j and Ω_j are, respectively, a $K \times 1$ vector of standard deviations and a $K \times K$ correlation matrix

⁵Our BMTL approach for logistic regressions has similarity with the seemingly unrelated regressions (SURs) in the sense that both approaches consider correlations across different regression equations. However, SURs place correlations on the *disturbance terms* of the equations to avoid bias while BMTL places correlations on the *coefficients* to introduce bias for improved predictive performance.

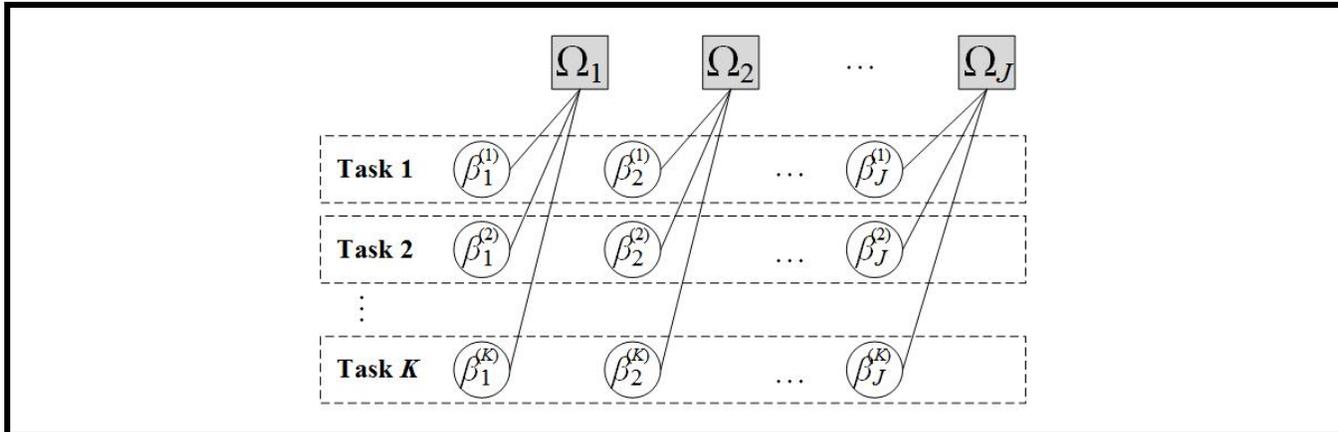


Figure 2. Illustration of Correlation Structure across Tasks in BMTL

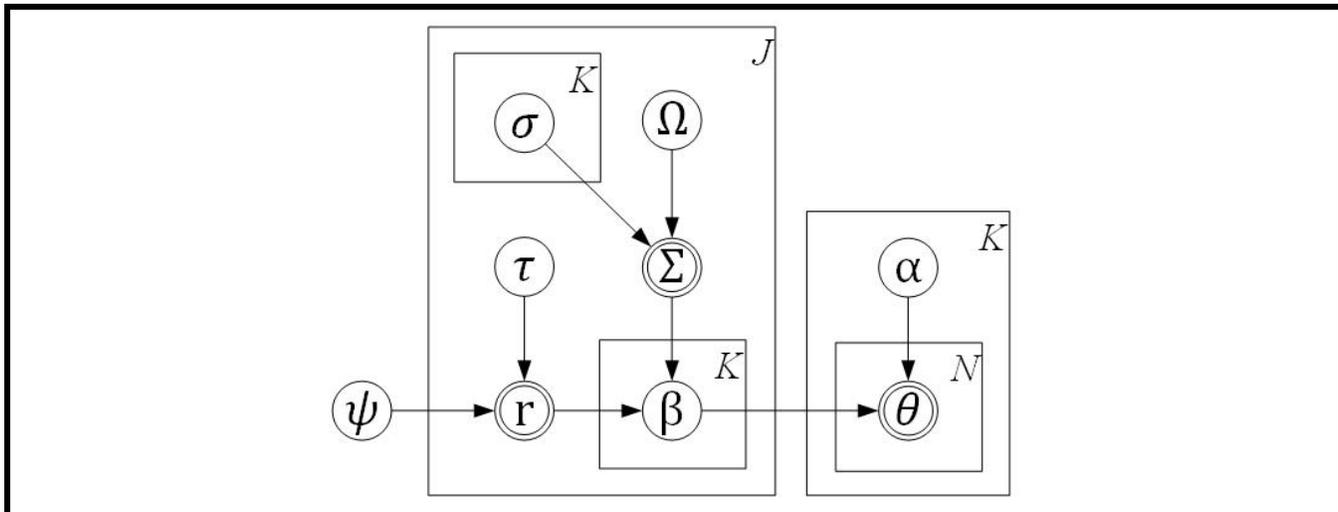


Figure 3. Plate Diagram of BMTL for Logistic Regression Models

for the elements in β_j . Following Gelman et al. (2008), we standardize all nonbinary predictors to have mean 0 and standard deviation 0.5, and then specify weakly informative prior distributions for parameters in the BMTL logistic regression model. In what follows, we provide a detailed description and justification for each of these parameters.

Regression Intercept: $\alpha^{(k)}$

Consistent with Gelman et al. (2008), we set the prior distribution of the regression intercept $\alpha^{(k)}$ to follow a Cauchy distribution with center 0 and scale 10. As shown in Figure 4, a Cauchy distribution has a bell-shape density function like the normal distribution but with thicker tails. Gelman et al. suggest this as the default prior for the intercept term because

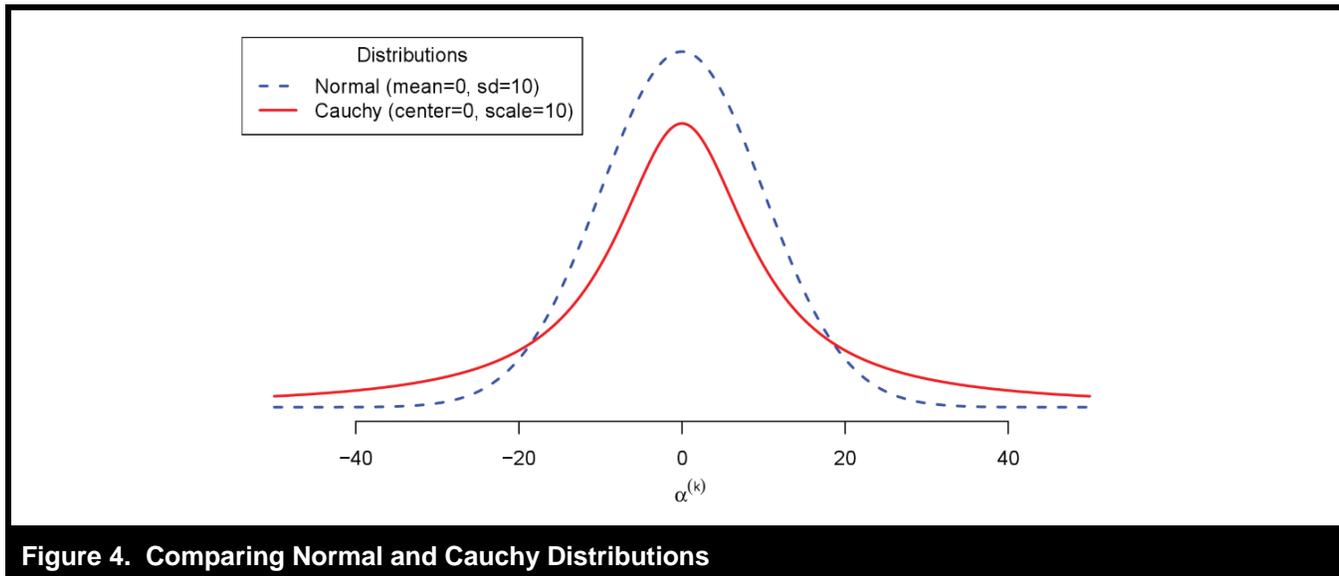
Cauchy allows the occasional possibility of very large values and hence is more robust and conservative than the usual normal distribution. In addition, after standardizing the raw data, the Cauchy density with center 0 and scale 10 is disperse enough to allow the baseline event probability for an average case to range between 10^{-9} (very unlikely to have the adverse health event) and $1 - 10^{-9}$ (very likely to have the adverse health event) in a logistic regression model (Gelman et al. 2008).

Regression Coefficients: β_j

A common approach to model regression coefficients in a Bayesian framework is through MVN distribution (e.g., Ghose et al. 2013; Xu et al. 2014). In doing so, the regression coeffi-

Table 1. Description of Parameters		
Parameter	Distribution/Function Form and Supporting Reference	Description
$\alpha^{(k)}$	Cauchy distribution (Gelman et al. 2008): $\alpha^{(k)} \sim \text{Cauchy}(0, 10)$	Intercept term in a logistic regression. One for each task.
β_j	Multivariate normal (MVN) distribution with the horseshoe prior (Carvalho et al. 2010; Gelman et al. 2008): $\beta_j = [\beta_j^{(1)}, \beta_j^{(2)}, \dots, \beta_j^{(K)}]^T$, $\beta_j \sim \text{MVN}(\mathbf{0}, r_j^2 \Sigma_j)$	Coefficients for the j^{th} predictor across logistic regressions. One for each predictor.
r_j	Horseshoe prior (Carvalho et al. 2010): $r_j = \tau_j \psi$, $\tau_j, \psi \sim \text{Half-Cauchy}(0, 1)$	Shrinkage coefficient for the covariance matrix in the MVN distribution of the j^{th} covariate. One for each predictor.
Σ_j	Covariance matrix (Barnard et al. 2000) $\Sigma_j = \text{diag}(\sigma_j) * \Omega_j * \text{diag}(\sigma_j)$, $\sigma_j = [\sigma_j^{(1)}, \sigma_j^{(2)}, \dots, \sigma_j^{(K)}]^T$	Covariance matrix in the MVN distribution of the j^{th} coefficients across tasks. One for each predictor.
$\sigma_j^{(k)}$	Half-Cauchy distribution (Gelman et al. 2008): $\sigma_j^{(k)} \sim \text{Half-Cauchy}(0, 2.5)$	Standard deviation of the j^{th} coefficient in the k^{th} logistic regression.
Ω_j	Lewandowski, Kurowicka and Joe (LKJ, 2009) distribution: $\Omega_j \sim \text{LKJ}(K, 1)$	Correlation matrix of the j^{th} coefficients across tasks. One for each predictor.

Note: The index j ranges from 1 to J (the total number of predictors), and the index k ranges from 1 to K (the total number of tasks).



coefficients are allowed to be correlated with and influenced by each other. Accordingly, we model $\beta_j = [\beta_j^{(1)}, \beta_j^{(2)}, \dots, \beta_j^{(K)}]^T$ with a MVN distribution:

$$\beta_j \sim \text{MVN}(\mathbf{0}, r_j^2 \Sigma_j) \tag{3}$$

where r_j is a shrinkage scalar and Σ_j is a covariance matrix.

The zero mean in MVN distribution indicates no prior knowledge with regard to the effect of the predictor. On the other hand, the scaled covariance matrix, $r_j^2 \Sigma_j$, is intended to capture the relationships among $\beta_j^{(1)}, \beta_j^{(2)}, \dots$, and $\beta_j^{(K)}$ —the coefficients of the j^{th} predictor in different tasks. We discuss the prior distributions of the shrinkage scalar, r_j , and the original covariance matrix, Σ_j , in the following.

Shrinkage Scalar: r_j

Like many practical big data problems, EHR data have a large number of candidate predictors potentially useful for predictive analytics, including but not limited to patient demographic information and various clinical phenotypes, such as diagnoses, treatments, and laboratory tests. The theoretical and empirical results in Tibshirani (1996) suggest that one can often improve the predictive performance of a linear model by shrinking or setting some of the coefficients toward 0. This reduces, or even eliminates, the effects of the respective predictors.

In a Bayesian linear model, one can achieve shrinkage by setting coefficients' prior distributions to have a zero mean with a smaller variance. In doing so, the posterior distributions of these coefficients will be closer to zero. The r_j in (3) is a shrinkage parameter used to scale the original covariance matrix Σ_j . We set r_j to follow a horseshoe prior distribution (Carvalho et al. 2010). The horseshoe is a robust, adaptive, and effective shrinkage prior because it has a probability density highly concentrated around zero but also with thicker tails than the normal distribution to accommodate occasional extreme values. Consistent with Carvalho et al. (2010), we operationalize the horseshoe prior as follows:

$$r_j = \tau_j \psi \tag{4}$$

$$\tau_j, \psi \sim \text{Half-Cauchy}(0,1) \tag{5}$$

where ψ and τ_j are the global (across all predictors) and local (specific to a predictor) components, respectively, in the horseshoe prior. Both ψ and τ_j follow a half-Cauchy distribution, which is a truncated Cauchy distribution with densities only on positive real numbers. The horseshoe prior is fully specified. That is, the center 0 and scale 1 in the half-Cauchy distribution are fixed values, and we do not need to supply any hyperparameters for the distribution.

Covariance Matrix: Σ_j

The most commonly used prior distribution for covariance matrices is the inverse-Wishart distribution because its conjugacy property with the normal distribution makes it very easy to compute (Xu et al. 2014). However, there is a trade-off between computational convenience and statistical accuracy. Barnard et al. (2000) provide a detailed exposition about the properties and limitations of the usual inverse-Wishart distribution, and suggest a more stable, flexible, and elaborative strategy to model covariance matrices. Specifically, we can write

$$\Sigma_j = \text{diag}(\sigma_j) * \Omega_j * \text{diag}(\sigma_j) \tag{6}$$

where $\sigma_j = [\sigma_j^{(1)}, \sigma_j^{(2)}, \dots, \sigma_j^{(K)}]^T$ is a vector of standard deviations in which $\sigma_j^{(k)}$ is the standard deviation of $\beta_j^{(k)}$, the coefficient for predictor j in task k . The $\text{diag}(\sigma_j)$ is a diagonal matrix with elements of σ_j on the diagonal. The parameter Ω_j is the correlation matrix for the coefficients of predictor j across the K tasks. As motivated earlier and shown in Figure 2, the key element in our BMTL approach is to superimpose a correlation structure over models. By using the decomposition strategy from Barnard et al., we re-parameterize the covariance matrix in the MVN distribution with a correlation matrix, which in turn allows us to explicitly model correlations for the elements in β_j and achieve multitask learning.

Standard Deviation: $\sigma_j^{(k)}$

The standard deviation for the j^{th} coefficient in the k^{th} task is denoted as $\sigma_j^{(k)}$. We set it to have a half-Cauchy prior distribution with center 0 and scale 2.5, as suggested by Gelman et al. (2008). As mentioned earlier, half-Cauchy distribution has probability density only on positive reals. This ensures that we will not generate a negative standard deviation. In addition, since the raw data have been standardized, the variation of the logistic regression coefficients is contained as well. The scale of 2.5 in the half-Cauchy distribution is a conservative choice, and has shown to be effective and robust in many applications (Gelman et al. 2008).

Correlation Matrix: Ω_j

The implementation of our BMTL approach hinges on whether we can capture the correlation among coefficients across tasks. However, modeling a correlation matrix is not easy in practice because of its geometric constraints: symmetric, positive semidefinite, and diagonal elements always equal 1. Prior studies usually generate correlation matrices by modeling the off-diagonal entries in the matrices. However, this approach cannot guarantee the resulting matrix to be positive semidefinite (Rousseeuw and Molenberghs 1994). To address this issue, we follow Lewandowski, Kurowicka and Joe (LKJ, 2009) to generate random samples of Ω_j . Specifically,

$$\Omega_j \sim \text{LKJ}(d, \eta) \tag{7}$$

The first parameter of the LKJ distribution specifies the dimension of the desired correlation matrix. Hence, d equals K , the number of tasks, in this study. The second parameter controls the degree to which the correlation matrix shrinks

toward the identity matrix. When η equals 1, the prior density is uniform over the space of all correlation matrices. As an example, for the two correlation matrices A and B below, they are equally likely to be generated from the LKJ distribution when η equals 1. As η increases, the prior increasingly concentrates around the identity matrix, giving matrix A a higher probability density than matrix B. In BMTL, we set $\eta = 1$, which reflects no prior information on Ω_j .

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.1 \\ 0.1 & 0.1 & 1 \end{bmatrix}$$

Model Fitting

With the above model specification, we now discuss our approach for model fitting. We first note that our model has a rich structure with a set of nonconventional prior distributions. As such, there is no closed form solution for the posterior distribution. We hence fit the BMTL logistic regression model with the No-U-Turn Sampler (NUTS) (Hoffman and Gelman 2014), which is a variant of Hamiltonian Monte Carlo (HMC). As compared to other commonly used Markov chain Monte Carlo (MCMC) algorithms, such as Metropolis algorithm and Gibbs sampler, HMC is more flexible and efficient because it requires no conjugacy, suppresses the local random walk behavior, and explores the marginal variances rather than the conditional variances of the probability space (Neal 2011). These properties make HMC a suitable and sometimes the only feasible option for Bayesian models like ours with high dimensionality, high correlation, and complex hierarchical structure. Interested readers are referred to Neal (2011) and Duane et al. (1987) for detailed exposition of HMC.

Generalizability of the BMTL Approach

Having shown the design of BMTL and its construction for logistic regression models, we now briefly discuss the generalizability of our BMTL approach. As mentioned, we enable multitask learning by imposing a hierarchical correlation structure as a channel to transfer information over tasks. We use logistic regression models to illustrate our BMTL approach because they are the most common technique in predictive analytics research and provide good modeling intuitions compared with other machine learning models. Nevertheless, it is clear that we make no assumptions on the baseline models for BMTL except that the models need to have a Bayesian representation. That is not very restrictive

since many statistical and machine learning models have been implemented in a Bayesian framework, such as ANNs (Neal 1996), SVMs (Tipping 2001) and decision trees (Chipman et al. 2002). To illustrate the generalizability of the BMTL approach, Appendix A provides an example on how to specify ANNs as the baseline models in the BMTL approach.

Contrast with the Literature

We now recap the methodological novelties of this study. We contrast these with respect to the existing healthcare predictive analytics research and the multitask learning research.

Compared with the extant healthcare predictive analytics research, this study is among the first that recognizes the potential benefits of multifaceted risk profiling. Managing multiple comorbidities is particularly relevant for aging populations in developed and developing countries. As such, Parekh and Barton (2010, p. 1304) argue that “transformation from a single chronic condition approach to a multiple chronic conditions approach is needed.” The literature of healthcare predictive analytics has a very limited knowledge base and methodological tools for multifaceted risk profiling. Our novelty is hence providing a principled approach to assess multiple patient outcomes.

Compared with the existing multitask learning literature, we follow the design science paradigm and propose a novel design to achieve multitask learning. The design principle in our approach is utilizing a hierarchical Bayesian structure to establish correlations among the coefficients of the same predictor in a set of baseline models. This design is conceptually intuitive and can potentially be applied to any baseline models as long as they have a Bayesian formulation. We demonstrate the design principle using logistic regressions as our baseline models, and illustrate ANN-based BMTL in Appendix A. This is in contrast with the existing multitask learning approaches, which are designed specifically for a particular type of baseline models, for example, ANNs (Caruana 1997), trees (Simm et al. 2014), or regressions (Huang et al. 2012).

In addition, the BMTL approach can avoid negative transfer in which multitask learning performs worse than single-task learning when the tasks are not positively correlated. Negative transfer is a common problem in many extant multitask learning approaches. Tree-based (Simm et al. 2014) and regression-based (Huang et al. 2012) multitask learning models are often designed to optimize certain regularization functions, and thus are prone to the problem of negative transfer. Our BMTL approach, on the other hand, avoids this problem by modeling the full correlation matrix, which accommodates any pattern of correlation and heteroscedasticity across tasks.

Experimental Study

We choose diabetes as our research case because of its large patient population and its broad societal impact. According to the International Diabetes Federation (2013), the global population of diabetic patients is projected to grow from 382 million in 2013 to 592 million in 2035. Among the new cases, 80% will come from developing countries, including China, India, and Pakistan. In the United States, diabetes is the seventh leading cause of death. Currently, more than 29 million Americans live with diabetes with an estimated medical cost of \$322 billion per year (Centers for Disease Control and Prevention 2014; Dall et al. 2014).

Diabetes is associated with many complications. To demonstrate multifaceted risk profiling, we simultaneously model and predict diabetic patients' risks of three adverse health events: stroke (henceforth denoted by STK), acute myocardial infarction (AMI), and acute renal failure (ARF). These three adverse health events are common among diabetic patients and often lead to premature death. Therefore, accurate predictions of these adverse health events could be used to optimize decisions in care plans and patient education.

We conducted experiments on de-identified EHR data from a major 600-bed hospital in Taiwan. The hospital has over one million registered patients, and provides care to roughly 750,000 outpatients and 20,000 inpatients annually. From our EHR data, we identified a cohort of 14,782 adults with type 2 diabetes using diagnosis codes from the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). For each of these patients, we further extracted their complete medical history in the EHRs. We used standard ICD-9-CM codes to identify whether a patient experienced any of the three adverse health events. Among the patients in the cohort, 2,370 (16%) had an STK event, 722 (5%) an AMI event, and 1,006 (7%) an ARF event.

Our EHR data contain comprehensive information collected from daily clinical practice from 2003 to 2012, including basic patient demographic information, ICD-9-CM diagnosis codes, treatments, laboratory tests, and physiological measures, among others. We note that these data elements are common in most EHR systems across organizations and countries, despite some potential differences in data unit or coding scheme, which have no impact on implementing our BMTL approach.⁶ A total of 179 predictors which have

⁶As an example of the difference in data unit, cholesterol is typically measured by milligrams per deciliter (mg/dL) in Taiwan and the United States, but in the United Kingdom, the standard unit for cholesterol is millimoles per liter (mmol/L). Similarly, some healthcare providers may have moved from the common ICD-9-CM to the new ICD-10-CM coding

values in more than 10% of patients in the cohort are considered in our analysis.⁷ Prior to the analysis, missing values are imputed using the respective column mean. Examples of the final predictors are illustrated in Table 2.⁸

Experiment Design and Performance Measure

We designed experiments to reflect practical uses of a risk prediction model. Figure 5 illustrates the experiment design. For each patient we randomly sampled a visit (v_{0i}) from the first half of the patient's medical history. For example, if a patient visited the hospital eight times during the observation period, v_{0i} would be sampled from the first four visits. Using data from an earlier visit allows us to more realistically assess the predictive performance of the model because we would have less information about the patients. Since the main purpose of risk profiling is to enable preventive interventions, it is also more clinically useful if we are able to make predictions at an earlier stage of disease process, which is another reason why we sample a visit from the first half of the medical history. We then used the "visible" information at v_{0i} for model training. The way we accounted for information prior to v_{0i} depends on the type of variables. For diagnosis and treatment variables, we set their values to 1 if there was evidence, at or before v_{0i} , indicating that the patient had the diagnoses/treatments, and 0 otherwise. For lab and exam variables, we set them to the most recent recorded values available at v_{0i} . We determined the status of $y_i^{(k)}$ by whether patient i experienced an event (STK, AMI, or ARF) in the next w years after v_{0i} . We varied w from 1 to 5 years for two purposes. First, we wanted to examine whether the time window length affects predictive performance. Second, we wanted to understand if the issue of data censoring affects the overall performance of our BMTL approach. With this experiment design, Table 3 summarizes event occurrence with respect to the sampled visit of each patient. Notice that patients could have the events before their respective sampled visit. We excluded patients who experienced all three events (i.e., STK, AMI, and ARF) before their respective v_{0i} . How-

scheme for diagnosis. Regardless of these differences, they do not affect the implementation of our BMTL approach. Moreover, our main objective is not to provide a model, but to provide a principled approach for healthcare institutions, practitioners, and researchers to build their own models using their own data.

⁷There is no gold standard for the cutoff of missing value. As such, we have tried different cutoffs, ranging from 10% to 90%, for the degree of accepted missing values. All of these produced similar results.

⁸A complete list of the variables is available upon request from the first author.

Category	Example Predictors
Patient Information	Age, body weight, male, smoking
Diagnoses	Three-digit ICD-9-CM codes, e.g., 401 for essential hypertension and 427 for cardiac dysrhythmias
Treatments	Aspirin, clopidogrel, insulin, isoket, metformin
Labs and Exams	Computerized tomography, low-density lipoprotein cholesterol, serum creatinine, systolic blood pressure

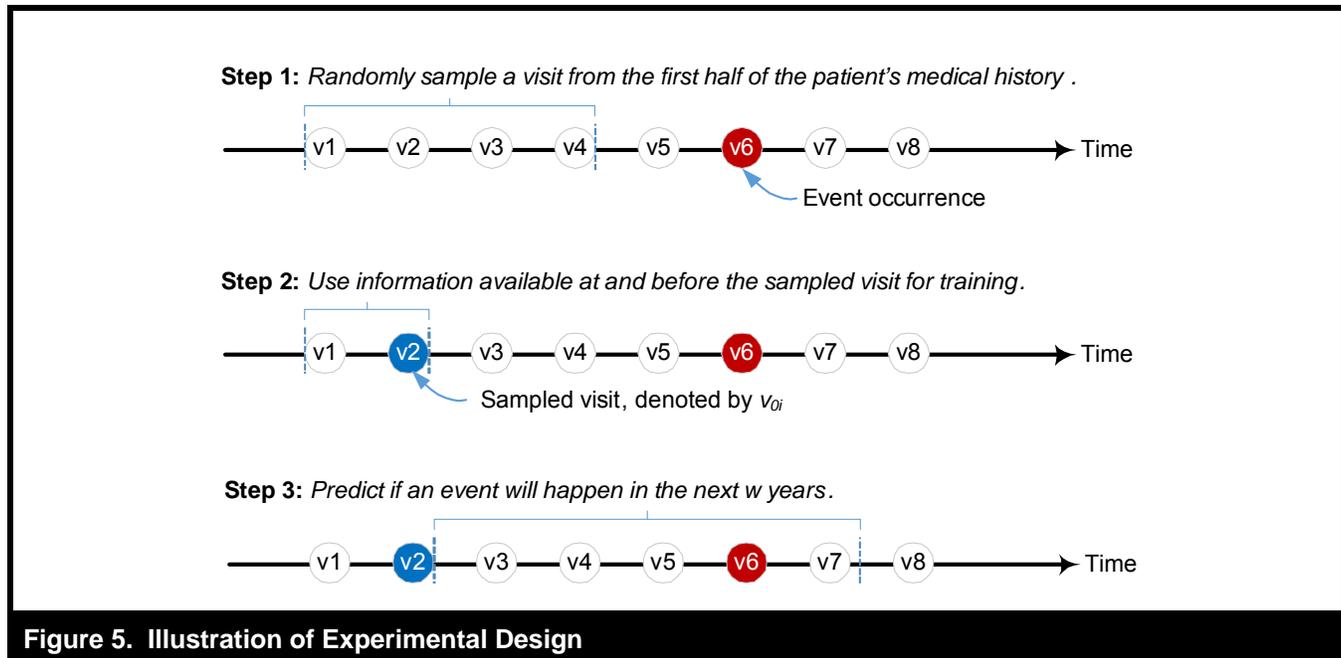


Figure 5. Illustration of Experimental Design

Event	Before v_{0i}	During v_{0i} and $v_{0i} + w$ (cumulative with respect to w)				
		$w = 1$ year	$w = 2$ years	$w = 3$ years	$w = 4$ years	$w = 5$ years
STK	1,507	354	560	685	793	828
AMI	485	75	146	178	210	225
ARF	410	217	399	488	536	571

ever, we retained patients in the cohort if they experienced only one or two of the three events before their v_{0i} , because we needed to predict their risk of the other event. Overall, at v_{0i} the cohort included 12,494 patients with zero events, 2,144 with one event, and 144 with two events. Our evaluations of predictive power use only events that happen during v_{0i} and $v_{0i} + w$ years, and do not include events that happened before v_{0i} because these events are already known to the clinician and the patient at v_{0i} .

The essence of evaluating a predictive model is measuring the model's performance on previously unseen instances in a holdout dataset (Shmueli and Koppius 2011). Cross-validation is the most common approach for evaluating predictive models. In cross-validation, instances are divided into M subsets, and a model is trained on $M - 1$ subsets and tested on the holdout subset. By performing the evaluation M times—each with a different holdout subset—the predictive performance of a model is its average performance across the

M holdout subsets. We used a 10-fold cross-validation design (i.e., $M = 10$) and quantified predictive performance using receiver operating characteristics (ROC) (Fawcett 2006). The ROC space is two-dimensional with true positive rate as the Y axis and false positive rate as the X axis, in which

$$\text{True positive rate} = \frac{\text{Positive correctly predicted}}{\text{Total positives}}$$

$$\text{False positive rate} = \frac{\text{Negatives incorrectly predicted}}{\text{Total negatives}}$$

The area under the ROC curve (AUC; aka, C-statistic or C-index) is a scalar metric ranging between 0.5 (equivalent to a random guess) and 1.0 (perfect performance). The AUC is a standard measure in predictive analytics and quantifies a model's trade-offs between type I and type II errors (Bardhan et al. 2014). Unless otherwise stated, we use AUC as the primary measure to compare different predictive models in our evaluation experiments.

We conducted three sets of evaluations to assess the proposed BMTL approach. In the first set of evaluations, we aimed to understand the utility of multitask learning. We compared the BMTL approach for logistic regression models (denoted by BMTL-Logit) with three single-task learning counterparts: Bayesian logistic regression (denoted by B-Logit), the common maximum likelihood based logistic regression (denoted by Logit), and logistic regression with lasso regularization (denoted by Logit-lasso) (Tibshirani 1996). To fit the BMTL-Logit model for this and the following evaluations, a total of 2,000 samples were drawn from two separate Markov chains after 1,000 warm-up draws from each chain.⁹ For each parameter, convergence was assessed using Gelman and Rubin's (1992) diagnostic test, also known as the \hat{R} statistic, with the value less than 1.2. In the second set of evaluations, we aimed to understand the performance of our BMTL approach against other multitask learning approaches in the literature. We hence compared BMTL-Logit with logistic regression-based multitask learning (MTL-Logit; Huang et al. 2012), tree-based multitask learning (MTL-Tree; Simm et al. 2014) and ANN-based multitask learning (MTL-ANN; Caru-

ana 1997).¹⁰ In the third set of evaluations, we aimed to demonstrate the practical impact of our BMTL approach. We counterfactually analyzed how healthcare predictive models could augment clinicians' capability in identifying high-risk patients and providing guideline-recommended preventive treatments to reduce the risks.

Experiment Results

Evaluation 1: BMTL Versus Single-Task Learning Approaches

In evaluation 1, we compared BMTL-Logit, B-Logit, Logit, and Logit-lasso models to examine the utility of multitask learning over single-task learning. Table 4 summarizes the results in evaluation 1, which are broken down by the length of the window (i.e., w) and by the prediction task (i.e., STK, AMI, or ARF). While each of the single-task learning models is trained independently, the BMTL-Logit model simultaneously learns and predicts the three events in the same window size. The results show that BMTL-Logit consistently outperforms the alternative single-task learning models. The nonparametric DeLong test of AUC (DeLong et al. 1988) shows that all of the performance differences are statistically significant.

The average AUC measures of BMTL-Logit, B-Logit, Logit, and Logit-lasso models across all windows and tasks are 0.774, 0.755, 0.751, and 0.758. The performance difference between BMTL-Logit and the alternative models varies depending on the window length and task. Overall, the mean (maximum and minimum) difference is 0.019 (0.049 and 0.005). We notice a greater performance difference in the AMI task among models. Averaging across the windows, BMTL-Logit attains an AUC of 0.743 in the AMI task whereas B-Logit, Logit, and Logit-lasso are, respectively, 0.713, 0.707, and 0.717. This greater degree of performance improvement from BMTL is likely due to the fact that AMI is a relatively rare event in our cohort compared with STK and ARF. A known challenge in machine learning is "class imbalance," in which negative instances significantly outnumber positive instances (He and Garcia 2009). Learning from imbalanced data is difficult because there is a higher risk for overfitting. By simultaneously learning multiple baseline models, BMTL can mitigate class imbalance because the AMI model can now leverage additional training signals from STK

⁹Readers familiar with the Metropolis algorithm and Gibbs sampler may find the 2,000 MCMC iterations (1,000 burn-in draws and 1,000 sampling draws) extremely small. This is because NUTS is much more efficient than random-walk Metropolis and Gibbs, especially for high dimensional and hierarchical models. See Hoffman and Gelman (2014) for an example of how a 2,000-iteration NUTS outperforms a 1,000,000-iteration Metropolis/Gibbs.

¹⁰For the implementations of MTL-Logit, MTL-Tree, and MTL-ANN, we use open source R packages `gprreg`, `extraTrees`, and `nnet`, respectively.

Table 4. Summary of Results in Evaluation 1

Window (w)	Task	Models			
		BMTL-Logit	B-Logit	Logit	Logit-lasso
1	STK	0.747	0.725***	0.723***	0.735***
1	AMI	0.778	0.744***	0.729***	0.758**
1	ARF	0.863	0.855*	0.847**	0.849***
2	STK	0.744	0.724***	0.722***	0.729***
2	AMI	0.748	0.723**	0.719**	0.721***
2	ARF	0.841	0.831***	0.828***	0.835**
3	STK	0.742	0.724***	0.722***	0.728***
3	AMI	0.736	0.703***	0.699***	0.704***
3	ARF	0.833	0.823***	0.819***	0.823***
4	STK	0.739	0.723**	0.722***	0.725***
4	AMI	0.725	0.694***	0.691***	0.699***
4	ARF	0.824	0.817**	0.814***	0.819**
5	STK	0.739	0.724***	0.723***	0.727***
5	AMI	0.727	0.699***	0.698***	0.704***
5	ARF	0.820	0.812***	0.809***	0.814***

Note: Bolded values highlight the best AUC result in a row.

***The AUC result is statistically significantly different from BMTL-Logit at $\alpha = 0.01$.

**The AUC result is statistically significantly different from BMTL-Logit at $\alpha = 0.05$.

*The AUC result is statistically significantly different from BMTL-Logit at $\alpha = 0.1$.

and ARF models.¹¹ Overall, the results from evaluation 1 confirm our speculations that there exists a spillover effect among individual baseline models and that our BMTL approach can effectively exploit the spillover effect to improve predictive performance.

Evaluation 2: BMTL Versus Other Multitask Learning Approaches

Although it is relatively new to the IS community, multitask learning has been an active area in machine learning and artificial intelligence. A number of multitask learning approaches have been proposed since the seminal paper by Caruana (1997). To determine the standing of our BMTL approach among the existing ones, we conducted a head-to-head comparison of predictive performance with a logistic regression-based MTL-Logit approach (Huang et al. 2012), a recent MTL-Tree approach (Simm et al. 2014), and the classic MTL-ANN approach (Caruana 1997). Unlike BMTL-Logit, all three of these alternative multitask learning approaches require user-specified parameters. For MTL-Logit, users need

to specify the weight for the regulation term; for MTL-Tree, the size of the tree; for MTL-ANN, the number of hidden nodes and the learning rate. We identified the best parameter settings for these approaches through cross-validation before we conducted evaluation 2.

Table 5 reports the results from evaluation 2. The average AUC values of BMTL-Logit, MTL-Logit, MTL-Tree, and MTL-ANN are 0.774, 0.755, 0.736, and 0.690, respectively. The results from evaluation 2 suggest that the BMTL-Logit approach consistently outperformed the alternative approaches—often with a statistically significant margin. Taken together, BMTL-Logit is very competitive among the existing multitask learning approaches for multifaceted risk profiling.

Evaluation 3: Counterfactual Analysis of Practical Use

We have been arguing that healthcare predictive analytics can provide clinical intelligence for preventive care. Grady and Berkowitz (2011) also suggest that clinical predictive modeling should go beyond prediction of risk and provide evidence that “prediction can lead to actions that reduce risk

¹¹We thank an anonymous reviewer for pointing out this technical insight.

Table 5. Summary of Results in Evaluation 2

Window (w)	Task	Models			
		BMTL-Logit	MTL-Logit	MTL-Tree	MTL-ANN
1	STK	0.747	0.746	0.717**	0.660***
1	AMI	0.778	0.767*	0.737**	0.686**
1	ARF	0.863	0.849*	0.831***	0.650***
2	STK	0.744	0.735*	0.708***	0.657***
2	AMI	0.748	0.701***	0.727**	0.734*
2	ARF	0.841	0.817***	0.787***	0.768***
3	STK	0.742	0.730**	0.702***	0.677***
3	AMI	0.736	0.693***	0.727*	0.680***
3	ARF	0.833	0.816***	0.787***	0.763***
4	STK	0.739	0.722**	0.690***	0.675***
4	AMI	0.725	0.701*	0.704*	0.628***
4	ARF	0.824	0.811***	0.773***	0.740***
5	STK	0.739	0.719***	0.686***	0.670***
5	AMI	0.727	0.705**	0.692**	0.653***
5	ARF	0.820	0.809***	0.77***	0.703***

Note: Bolded values highlight the best AUC result in a row.

***The AUC result is statistically significantly different from BMTL-Logit at $\alpha = 0.01$.

**The AUC result is statistically significantly different from BMTL-Logit at $\alpha = 0.05$.

*The AUC result is statistically significantly different from BMTL-Logit at $\alpha = 0.1$.

beyond what would occur without the prediction rule” (p. 1702). Prescribing preventive treatments is perhaps the most critical action in reducing risks. The practical utility of a model is hence its capability in prompting preventive interventions in high risk patients who otherwise would not receive such interventions. The gold standard to determine the impact of a healthcare predictive model is through a randomized clinical trial with two groups of clinicians—one with the predictive model and the other without. However, clinical trials are extremely expensive and time-consuming, which is one of the reasons why very few healthcare predictive models have undergone such evaluation. Accordingly, Reilly and Evans (2006, p. 207) suggest that

The potential impact of a prediction rule can be estimated by assessing its predictive validity and clinical sensibility and by measuring its potential to improve current decision making.

Instead of measuring the *actual* impact with a clinical trial, we proceeded to assess the *potential* impact of our approach through a counterfactual analysis. Specifically, we assume that rational clinicians will always prescribe guideline-recommended preventive interventions if they foresee a high

risk of adverse health events in their patients.¹² We then looked into the patients with STK/AMI/ARF events during v_{0i} and $v_{0i} + 5$ years, and analyzed the proportion of them who had not received any preventive interventions but could have been provided with such interventions at v_{0i} had an indication of “high risk” was provided by a predictive model. In other words, our counterfactual analysis reveals clinicians’ risk assessment capability and preventive treatment behavior, and triangulates that with what could have happened differently with the support from a predictive model.

We identified appropriate preventive treatments for STK, AMI, and ARF by using the *Diabetes Comprehensive Care Plan Guidelines* from the American Association of Clinical Endocrinologists (Table 6). In medical science, it is very common to use 20% risk over 10 years as a cutoff between high- and low-risk patients (e.g., Dhamoon and Elkind 2010; Lackland et al. 2012). Following Dhamoon and Elkind (2010),

¹²While the assumption is practically reasonable and consistent with the economics literature of rational choice (Simon 1955), we nonetheless acknowledge that there are occasions a clinician purposefully chooses *not* to prescribe preventive treatments to high-risk patients. In addition, we also note that getting preventive treatments does not guarantee prevention of adverse health events.

Table 6. Guideline-Recommended Preventive Treatments for STK, AMI and ARF

Adverse Health Event	Preventive Treatment
STK	<ul style="list-style-type: none"> • Antihypertensive agents • Antiplatelet therapy
AMI	<ul style="list-style-type: none"> • Antihypertensive agents • Antiplatelet therapy • Lipid lowering therapy
ARF	<ul style="list-style-type: none"> • Antihypertensive agents • Angiotensin receptor blockers • Angiotensin-converting-enzyme inhibitors

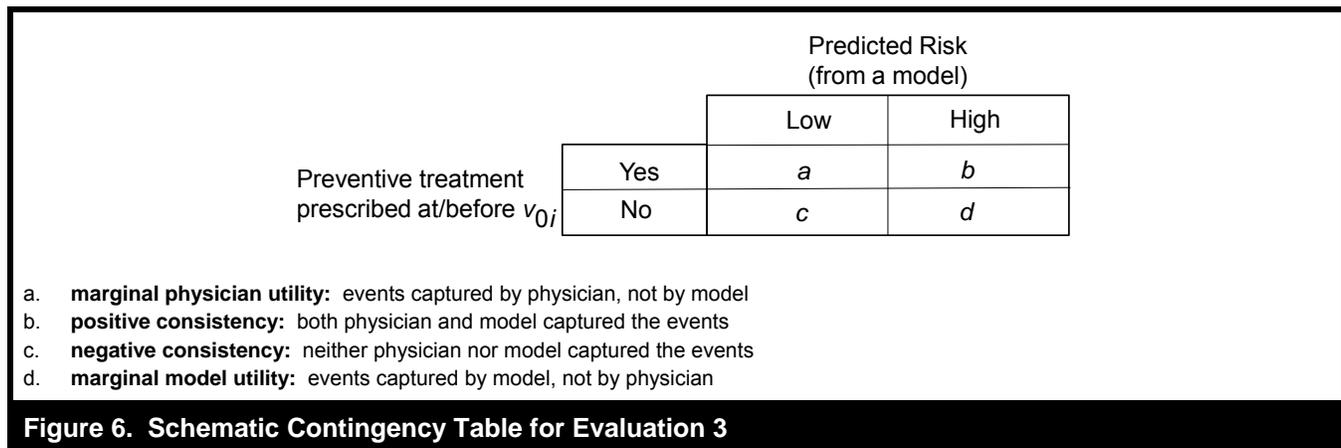


Figure 6. Schematic Contingency Table for Evaluation 3

we chose 10% risk over 5 years as our cutoff level because the median follow-up time among the patients in our data is about 5 years, which makes it impractical to assess 10-year risks. Accordingly, we categorize patients who have 5-year event risk above (below) 10% at v_{0i} as high (low) risk.¹³

With two levels of treatment behavior (with or without preventive treatments) and two levels of predicted risk (high or low), we then created a contingency table like Figure 6. Such a contingency table can provide several useful insights. The *a*, *b*, *c*, and *d* in the contingency table are the number of patients who fit into the respective quadrant. We name the *a* and *d* values in the contingency table as marginal physician utility and marginal model utility, respectively, because they represent the events that are correctly identified only by the physician or only by the predictive model. Values *b* and *c* in the contingency table, on the other hand, show consistency between the physician’s judgment and the model’s prediction—either both correct (*b*; positive consistency) or both

incorrect (*c*; negative consistency). Given that the pool of patients in this analysis are the ones who will have an adverse health event in the next five years, they should be classified as high-risk by a predictive model and provided with at least one preventive treatment by a physician. Therefore, we may consider *a* and *c* as model’s errors, and *c* and *d* as physician’s errors. In light of this analytical exposition, a predictive model is deemed more useful and valuable than another when it has a smaller value in *c* (making fewer mistakes) and a larger value in *d* (augmenting physician’s capability).

Along with our BMTL-Logit model, we also used the Logit-lasso model and the UKPDS Risk Engine (Kothari et al. 2002; Stevens et al. 2001) as our benchmarks in this analysis. The Logit-lasso model is a commonly used technique in data-driven healthcare predictive analytics and performed reasonably well in our evaluation 1. On the other hand, the UKPDS Risk Engine is one of the most authoritative risk models in diabetes care based on a large-scale clinical trial.

Figure 7 reports the results from evaluation 3. Regardless of the event type, we notice that a large portion of these patients did not have the guideline-recommended preventive treat-

¹³As a robustness check (see Appendix B), we examined different cut-off values (5% or 20% risk over 5 years). We found qualitatively similar results as the ones presented here.

STK (# of events = 828)		Predicted Risk (BMTL-Logit)		Predicted Risk (Logit-lasso)		Predicted Risk (UKPDS)	
		Low	High	Low	High	Low	High
Preventive treatment prescribed at/before v_{0i}	Yes	83	369	96	356	398	54
	No	181	195	197	179	329	47

AMI (# of events = 225)		Predicted Risk (BMTL-Logit)		Predicted Risk (Logit-lasso)		Predicted Risk (UKPDS)	
		Low	High	Low	High	Low	High
Preventive treatment prescribed at/before v_{0i}	Yes	107	54	111	50	111	50
	No	61	3	63	1	44	20

ARF (# of events = 571)		Predicted Risk (BMTL-Logit)		Predicted Risk (Logit-lasso)		Predicted Risk (UKPDS)	
		Low	High	Low	High	Low	High
Preventive treatment prescribed at/before v_{0i}	Yes	118	195	121	192	UKPDS does not predict ARF risks.	
	No	154	104	170	88		

Figure 7. Summary of Results in Evaluation 3

ments at or before their respective v_{0i} (STK: 45%; AMI: 28%; ARF: 45%). The nonzero d values suggest that all the models provide some level of practical utility by recognizing high-risk patients who were missed by the clinician. From the c and d values, the UKPDS Risk Engine outperforms the alternative models in the AMI cases. However, it falls short in the STK cases and does not predict ARF risks. With smaller c and larger d values, the BMTL-Logit model consistently outperforms the Logit-lasso model. Overall, we find that risk prediction models can support clinical decision making in a meaningful way. Our BMTL approach can better identify high-risk patients than the common Logit-lasso model. However, UKPDS seems to be the best model in predicting AMI, showing the merits of clinical trials in building certain risk models. In practice, an EHR system should employ both trial-based and data-driven risk models to maximize the opportunity for clinical decision support. A clinician could then choose which model to use based on the predictive accuracy obtained from the historical records of the local patient population (e.g., UKPDS for AMI events and BMTL for STK and ARF events in our focal hospital).

Discussion and Conclusions

The pronounced need to use IT to transform healthcare is widely recognized in the IS community (Agarwal et al. 2010;

Chen et al. 2012; Fichman et al. 2011). Although there is little doubt about the importance of EHR systems in healthcare, the research and practice communities are still exploring ways to fully realize the potential of EHRs. Clearly, the capabilities of EHRs are more than just digitalized patient records per se. Big and longitudinal EHR data can enable various business intelligence and analytics applications for advanced clinical decision support that were previously unavailable.

With the rise of EHR adoption, we ask whether it is possible and advantageous to model risks of different adverse health events simultaneously using EHR data. We further add to the extant research by developing a principled approach, Bayesian multitask learning (BMTL), for multifaceted risk profiling in chronic care. Formulating a comprehensive care plan for people with chronic disease is challenging because there is a need to assess and manage risks of different complications and comorbidities. As an IT artifact for advanced clinical decision support, the BMTL approach can assist healthcare providers in better assessing patients' risks and attaining the goals of preventive and personalized care.

To demonstrate the viability and utility of the BMTL approach, we used diabetes as our research case, and chose stroke, acute myocardial infarction, and acute renal failure as three adverse health events to be modeled simultaneously in diabetic patients. Our experiments showed that the BMTL approach consistently outperformed the respective single-task

Table 7. Linking This Study to Areas of IS Research

IS Research Area	Examples of IS Research	Relevance of This Study
Healthcare IS	Bardhan et al. (2014) Meyer et al. (2014)	<i>Research context:</i> Provide advanced decision support in healthcare
Big data and predictive analytics	Bao and Datta (2014) Fang et al. (2013)	<i>Methodology:</i> Develop an analytics approach for big EHR data
Design science	Abbasi et al. (2012) Chen et al. (2013)	<i>Research paradigm:</i> Address a practical problem with an IT artifact

learning models. In most evaluation settings, BMTL also had significantly better performance compared to the existing multitask learning approaches. Our counterfactual analysis of potential impact further reveals that the BMTL approach can support clinicians by identifying high-risk patients who otherwise would not be prescribed with preventive interventions.

Relevance to IS Research

This study fits into multiple areas of IS research. The ones most relevant are healthcare IS (Agarwal et al. 2010; Fichman et al. 2011), big data and predictive analytics (Goes 2014; Shmueli and Koppius 2011), and design science (Gregor and Hevner 2013; Hevner et al. 2004). These three areas of IS research are not mutually exclusive. In fact, they represent the research context, methodology, and research paradigm of this study. Table 7 summarizes the relevance of this study to these areas, which we discuss in turn.

Healthcare IS. We examine the role of IS in the healthcare context. Healthcare IS research concerns the issues regarding the managerial, organizational, and technical aspects of IS in various healthcare settings. Most of the extant research follows the positivist paradigm and examines the adoption and impacts of health IT (Angst and Agarwal 2009; Venkatesh et al. 2011; Yaraghi et al. 2015). The unique characteristics of healthcare, such as privacy concerns, government regulations, and diverse stakeholders, shed light on new IS theories and empirical findings. Against this backdrop, one specific gap in healthcare IS research is the development of advanced decision support methods or techniques that leverage the large amount of patient-level clinical data in EHRs (Fichman et al. 2011). Recent studies from Bardhan et al. (2014) and Meyer et al. (2014) show promising applications of advanced decision support in healthcare. Following these studies, we explored a risk-profiling application that provides decision support in chronic care. We recognize that our approach can be implemented in various contexts, but we focused on healthcare for reasons of scope. As we have shown, even in this particular context of healthcare, EHR data

analytics is very complex and of significant importance to research and practice (Agarwal et al. 2010; Chen et al. 2012).

Big data and predictive analytics. Developing better algorithms and models to discover useful insights from data has been the focus of big data and predictive analytics research (Chen et al. 2012; Goes 2014; Shmueli and Koppius 2011). As mentioned earlier, Bao and Datta (2014) and Fang et al. (2013) are excellent examples of big data analytics in IS research. In their application contexts, it is beyond the cognitive capability of a human being to harness tens of thousands of financial reports or predict social behaviors in large social networks. Analytics, hence, provides a necessary means to harvest data and facilitate knowledge discovery. There are similar cognitive challenges for clinicians at the point of care. The big EHR data contain longitudinal and detailed information about patients, but it is difficult for clinicians to leverage this rich information. Consistent with big data and predictive analytics research, we developed a big EHR data analytics approach to acquire useful clinical insights for chronic care.

Design science. This study follows the paradigm of design science research. In contrast with the positivist paradigm that emphasizes theory development and testing, the main objective of design science research is to develop IT artifacts to address practical problems (Gregor and Hevner 2013; Hevner et al. 2004). According to Hevner et al. (2004, p. 77),

IT artifacts are broadly defined as constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems).

Indeed, IT artifacts may take on different forms depending on the problem at hand. Chen et al. (2013) provided a novel approach in developing data models. Their goal was to enable efficient information flow in emergency management practice. Similarly, Abbasi et al. (2012) developed a new meta-learning framework to improve the performance of financial fraud detection. Consistent with these studies, we aim to

address a salient practical problem with an IT artifact. Our BMTL approach falls into the methods category of IT artifacts, and provides guidance on “how to search the solution space” (Hevner et al. 2004, p. 79). A perennial need in healthcare is assessing patient risks, and risk profiling is becoming particularly important as it moves toward preventive and personalized care. We develop a new method for multifaceted risk profiling that enables improved performance in risk profiling.

Contributions to the IS Knowledge Base

Our study makes several research contributions. First, we developed an EHR data analytics approach for risk profiling. An essential novelty of our approach is the consideration of multiple adverse health events in a risk-prediction framework. To our knowledge, BMTL is the first approach for multifaceted risk profiling and allows healthcare providers to model an arbitrary number of events and outcomes simultaneously. In contrast with the existing multitask learning techniques, our approach is enabled by a unique hierarchical correlation structure that orchestrates multiple baseline models in a joint modeling framework. Second, we evaluated the proposed approach with real-world EHR data. We obtained empirical evidence that simultaneous learning of multiple event risks improves overall predictive performance of each event risk. That is, a multifaceted risk profiling framework can indeed offer better clinical insights than multiple independent risk models. Finally, we recognize that there are multiple approaches to achieve multitask learning. Our evaluation results further suggest that BMTL outperforms the alternative multitask learning techniques in risk profiling.

Design science research can offer different forms of contributions to the IS knowledge base, including strong theory, partial theory, incomplete theory, or even the instantiation of the solution artifact (Gregor and Hevner 2013). Other than the instantiation of the BMTL approach in healthcare, our theoretical contribution is to motivate, examine, and establish two design principles in data analytics: (1) multitask learning and (2) hierarchical correlation structure for multitask learning. To our knowledge, the two design principles are either new to the IS discipline (design principle 1) or new to the world (design principle 2). These design principles prescribe how to model multiple outcomes simultaneously to attain improved predictive performance. The prescriptive knowledge advanced in this study is generalizable to other predictive analytics contexts as a “nascent design theory” (Gregor and Hevner 2013). Analogous to the effort of theory testing in a positivist manuscript, this study offers proof-of-concept and proof-of-value-added by demonstrating the

viability and utility of these design principles in EHR-based risk profiling.

Practical Implications

Healthcare is in the midst of a paradigm shift—from reactive care to preventive care (Dexter et al. 2001) and from one-size-fits-all medicine to precision medicine (The White House 2015). Prediction of adverse health events in patients with chronic disease plays a significant role in improving healthcare quality and reducing cost of care. According to Hillestad et al. (2005), a fully EHR-enhanced chronic care management system with advanced clinical decision support tools can potentially yield up to \$147 billion in savings per year as a result of preventing medical complications and reducing healthcare acute incidents. Hospitals, physicians, and patients can all benefit from a more comprehensive and accurate risk-profiling application such as BMTL. We discuss key practical implications for these stakeholders in the following.

Hospitals. Hospitals are facing new healthcare delivery models such as accountable care organizations and bundled payments. These are designed to create financial incentives for *better*, rather than *more*, services (Bates et al. 2014). To maximize financial gain, hospitals will need to consider not only the best treatments for a patient’s current condition, but also preventive interventions for possible complications and comorbidities in the future. In other words, there is an increasing need for hospitals to look beyond each specific patient encounter and take a long term prospect for care provision. Multifaceted risk profiling applications like BMTL will facilitate hospitals in identifying patient risks of different adverse health events as well as the most cost-efficient service plans in the long run.

Physicians. The BMTL approach provides physicians with advanced clinical decision support at the point of care. Despite physicians being highly trained professionals, medical errors such as failures and delays in providing preventive interventions are pervasive (Kohn et al. 2000). As Eddy (1990, p. 1272) notes, “The complexity of modern medicine exceeds the inherent limitations of the unaided human mind.” Just like the need for marketers to predict consumer behavior from large marketing databases for effective promotions, physicians, too, have a similar need to predict patient risk from large EHR databases for timely interventions. The results from our evaluations suggest that BMTL can augment physicians in identifying high-risk patients and, hence, reduce medical errors.

Patients. The number of chronic conditions often determines a patient’s quality of life and healthcare spending. Data from

the Medical Expenditure Panel Survey show a strong positive correlation between the number of chronic conditions and healthcare spending (Gerteis et al. 2014). Specifically, having one additional chronic condition can be roughly translated to an increment of \$2,000 in annual healthcare spending. It is hence of significant importance and interest for people with chronic disease to obtain preventive care. With hundreds of millions of patients living with chronic diseases and trillions of dollars spent on chronic care annually, even relatively small improvements in the performance of risk profiling can lead to significant impacts on quality and cost of care.

Beyond these stakeholders, a subtle but important practical implication stemming from our study is the complementarity between trial-based risk models and EHR-based risk models. Randomized controlled trials provide the strongest evidence in quantifying risk factors whereas EHRs permit a holistic and more realistic context of prediction. Our evaluation shows the usefulness of the UKPDS Risk Engine in predicting AMI events, although it cannot predict ARF events and its STK predictions are not as good as EHR-based risk models. Instead of replacing one type of risk models with the other, physicians should attain the best decision support by considering evidence from both clinical trials and EHR data and then triangulating these different sources of information with the unique characteristics of the focal patient. That is, the implementation of clinical decision support systems should include both trial- and EHR-based risk models so as to enable the best care.

Limitations and Future Research

This work has a number of limitations. First and foremost, the “no free lunch” theorem suggests that there will never be a learning method that can guarantee to outperform another method on every possible data set (Wolpert and Macready 1997). Our evaluations are based on one EHR data set from a single hospital. While we have employed cross-validation to train and test models, it is still possible that the better performance of the BMTL approach is limited to the data set under consideration. Future research may experiment the BMTL approach on different data sets and explore the conditions in which it is effective. Second, in our BMTL approach the baseline individual models need to be the same modeling technique (e.g., all logistic regression models). We note that this is a limitation universal to all existing multitask learning approaches in the literature. It is not clear how different techniques can be integrated in a multitask learning framework and share information in the learning process because the parameters from different techniques are not related in any meaningful way. Third, we assume that individual models in BMTL have the same set of predictors.

Despite this assumption, extending the BMTL approach to coordinate models with non-exact predictors is straightforward. If there are K individual models (again, one for each event) and a predictor is used only in H individual models ($1 \leq H \leq K$), we just need to adjust the dimension of the corresponding terms in our BMTL formulation from K to H . Finally, information sharing across models is only through the correlations of the coefficients of the same predictor in different models. Incorporating correlation matrices for each pair of unique predictors will exponentially increase the complexity of model fitting. Therefore, future studies may explore other ways to communicate information among different predictors across models. Despite these limitations, this study is just a first step toward multifaceted risk profiling and EHR data analytics. Most importantly, our principled approach opens a new way to frame and conduct big data and predictive analytics for enhanced performance.

Acknowledgments

We thank the senior editor, the associate editor, and three anonymous reviewers for their suggestions and comments on earlier versions of this manuscript. We are also grateful for the research support from Min-Sheng General Hospital. This study was partially funded by the National Science Foundation (IIP-1417181), E.SUN Bank and E.SUN Foundation, and Ministry of Science and Technology, R.O.C.

References

- Abbasi, A., Albrecht, C., Vance, A., and Hansen, J. 2012. “Metafraud: A Meta-Learning Framework for Detecting Financial Fraud,” *MIS Quarterly* (36:4), pp. 1293-1327.
- Agarwal, R., and Dhar, V. 2014. “Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research,” *Information Systems Research* (25:3), pp. 443-448.
- Agarwal, R., Gao, G., DesRoches, C., and Jha, A. K. 2010. “The Digital Transformation of Healthcare: Current Status and the Road Ahead,” *Information Systems Research* (21:4), pp. 796-809.
- Anderson, G. 2010. “Chronic Care: Making the Case for Ongoing Care,” Robert Wood Johnson Foundation (available at <http://www.rwjf.org/content/dam/farm/reports/reports/2010/rwjf54583>).
- Angst, C. M., and Agarwal, R. 2009. “Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion,” *MIS Quarterly* (33:2), pp. 339-370.
- Archambeau, C., Guo, S., and Zoeter, O. 2011. “Sparse Bayesian Multi-Task Learning,” in *Advances in Neural Information Processing Systems*, pp. 1755-1763.
- Bakker, B., and Heskes, T. 2003. “Task Clustering and Gating for Bayesian Multitask Learning,” *The Journal of Machine Learning Research* (4), pp. 83-99.

- Bao, Y., and Datta, A. 2014. "Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures," *Management Science* (60:6), pp. 1371-1391.
- Bardhan, I., Oh, J., Zheng, Z., and Kirksey, K. 2014. "Predictive Analytics for Readmission of Patients with Congestive Heart Failure," *Information Systems Research* (26:1), pp. 19-39.
- Barnard, J., McCulloch, R., and Meng, X.-L. 2000. "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage," *Statistica Sinica* (10:4), pp. 1281-1312.
- Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., and Escobar, G. 2014. "Big Data In Health Care: Using Analytics to Identify And Manage High-Risk and High-Cost Patients," *Health Affairs* (33:7), pp. 1123-1131.
- Baxter, J. 2000. "A Model of Inductive Bias Learning," *Journal of Artificial Intelligence Research* (12:1), pp. 149-198.
- Bishop, C. M. 2007. *Pattern Recognition and Machine Learning*, New York: Springer.
- Brownstein, J. S., Murphy, S. N., Goldfine, A. B., Grant, R. W., Sordo, M., Gainer, V., Colecchi, J. A., Dubey, A., Nathan, D. M., Glaser, J. P., and Kohane, I. S. 2010. "Rapid Identification of Myocardial Infarction Risk Associated with Diabetes Medications Using Electronic Medical Records," *Diabetes Care* (33:3), pp. 526-531.
- Cai, F., and Cherkassky, V. 2012. "Generalized SMO Algorithm for SVM-Based Multitask Learning," *IEEE Transactions on Neural Networks and Learning Systems* (23:6), pp. 997-1003.
- Caruana, R. 1997. "Multitask Learning," *Machine Learning* (28:1), pp. 41-75.
- Carvalho, C. M., Polson, N. G., and Scott, J. G. 2010. "The Horseshoe Estimator for Sparse Signals," *Biometrika* (97:2), pp. 465-480.
- Centers for Disease Control and Prevention. 2014. "National Diabetes Statistics Report" (available at <http://www.cdc.gov/diabetes/pdfs/data/2014-report-estimates-of-diabetes-and-its-burden-in-the-united-states.pdf>).
- Chen, H., Chiang, R. H. L., and Storey, V. C. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly* (36:4), pp. 1165-1188.
- Chen, R., Sharman, R., Rao, H. R., and Upadhyaya, S. J. 2013. "Data Model Development for Fire Related Extreme Events: An Activity Theory Approach," *MIS Quarterly* (37:1), pp. 125-147.
- Chipman, H. A., George, E. I., and McCulloch, R. E. 2002. "Bayesian Treed Models," *Machine Learning* (48:1-3), pp. 299-320.
- D'Agostino, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., and Kannel, W. B. 2008. "General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study," *Circulation* (117:6), pp. 743-753.
- Dall, T. M., Yang, W., Halder, P., Pang, B., Massoudi, M., Wintfeld, N., Semilla, A. P., Franz, J., and Hogan, P. F. 2014. "The Economic Burden of Elevated Blood Glucose Levels in 2012: Diagnosed and Undiagnosed Diabetes, Gestational Diabetes Mellitus, and Prediabetes," *Diabetes Care* (37:12), pp. 3172-3179.
- DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. 1988. "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics* (44:3), pp. 837-845.
- Dexter, P. R., Perkins, S., Overhage, J. M., Maharry, K., Kohler, R. B., and McDonald, C. J. 2001. "A Computerized Reminder System to Increase the Use of Preventive Care for Hospitalized Patients," *New England Journal of Medicine* (345:13), pp. 965-970.
- Dharmoon, M. S., and Elkind, M. S. V. 2010. "Inclusion of Stroke as an Outcome and Risk Equivalent in Risk Scores for Primary and Secondary Prevention of Vascular Disease," *Circulation* (121:18), pp. 2071-2078.
- Dixon-Woods, M., Redwood, S., Leslie, M., Minion, J., Martin, G. P., and Coleman, J. J. 2013. "Improving Quality and Safety of Care Using 'Technovigilance': An Ethnographic Case Study of Secondary Use of Data from an Electronic Prescribing and Decision Support System," *Milbank Quarterly* (91:3), pp. 424-454.
- Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. 1987. "Hybrid Monte Carlo," *Physics letters B* (195:2), pp. 216-222.
- Eddy, D. M. 1990. "Practice Policies: Where Do They Come From?," *The Journal of the American Medical Association* (263:9), pp. 1265-1275.
- Fang, X., Hu, P. J.-H., Li, Z., and Tsai, W. 2013. "Predicting Adoption Probabilities in Social Networks," *Information Systems Research* (24:1), pp. 128-145.
- Fawcett, T. 2006. "An Introduction to ROC Analysis," *Pattern Recognition Letters* (27:8), pp. 861-874.
- Fichman, R. G., Kohli, R., and Krishnan, R. 2011. "The Role of Information Systems in Healthcare: Current Research and Future Trends," *Information Systems Research* (22:3), pp. 419-428.
- Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. 2008. "A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models," *The Annals of Applied Statistics* (2:4), pp. 1360-1383.
- Gelman, A., and Rubin, D. B. 1992. "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science* (7:4), pp. 457-472.
- Gerteis, J., Izrael, D., Deitz, D., LeRoy, L., Ricciardi, R., Miller, T., and Basu, J. 2014. "Multiple Chronic Conditions Chartbook: 2010 Medical Expenditure Panel Survey Data," AHRQ Publications No. Q14-0038, Rockville, MD: Agency for Healthcare Research and Quality.
- Ghose, A., Goldfarb, A., and Han, S. P. 2013. "How Is the Mobile Internet Different? Search Costs and Local Activities," *Information Systems Research* (24:3), pp. 613-631.
- Goes, P. 2014. "Editor's Comments: Big Data and IS Research," *MIS Quarterly* (38:3), pp. iii-viii.
- Grady, D., and Berkowitz, S. A. 2011. "Why Is a Good Clinical Prediction Rule So Hard to Find?," *Archives of Internal Medicine* (171:19), pp. 1701-1702.
- Gregor, S., and Hevner, A. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355.

- Gupta, S. 2008. "Channel Structure with Knowledge Spillovers," *Marketing Science* (27:2), pp. 247-261.
- He, H., and Garcia, E. A. 2009. "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering* (21:9), pp. 1263-1284.
- Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75-106.
- Hillestad, R., Bigelow, J., Bower, A., Girosi, F., Meili, R., Scoville, R., and Taylor, R. 2005. "Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, and Costs," *Health Affairs* (24:5), pp. 1103-1117.
- Hoffman, M. D., and Gelman, A. 2014. "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo," *Journal of Machine Learning Research* (15), pp. 1593-1623.
- Huang, J., Breheny, P., and Ma, S. 2012. "A Selective Review of Group Selection in High-Dimensional Models," *Statistical Science* (27:4), pp. 481-499.
- International Diabetes Federation. 2013. "IDF Diabetes Atlas, 6th Edition," Brussels, Belgium (available at <http://www.idf.org/diabetesatlas>).
- Kohn, L. T., Corrigan, J., and Donaldson, M. S. (eds.). 2000. *To Err Is Human: Building a Safer Health System*, Washington, DC: National Academies Press.
- Kothari, V., Stevens, R. J., Adler, A. I., Stratton, I. M., Manley, S. E., Neil, H. A., and Holman, R. R. 2002. "UKPDS 60: Risk of Stroke in Type 2 Diabetes Estimated by the UK Prospective Diabetes Study Risk Engine," *Stroke* (33:7), pp. 1776-1781.
- Lackland, D. T., Elkind, M. S. V., D'Agostino, R., Dhamoon, M. S., Goff, D. C., Higashida, R. T., McClure, L. A., Mitchell, P. H., Sacco, R. L., Sila, C. A., Smith, S. C., Tanne, D., Tirschwell, D. L., Touzé, E., and Wechsler, L. R. 2012. "Inclusion of Stroke in Cardiovascular Risk Prediction Instruments a Statement for Healthcare Professionals from the American Heart Association/American Stroke Association," *Stroke* (43:7), pp. 1998-2027.
- Lewandowski, D., Kurowicka, D., and Joe, H. 2009. "Generating Random Correlation Matrices Based on Vines and Extended Onion Method," *Journal of Multivariate Analysis* (100:9), pp. 1989-2001.
- Meyer, G., Adomavicius, G., Johnson, P. E., Elidrissi, M., Rush, W. A., Sperl-Hillen, J. M., and O'Connor, P. J. 2014. "A Machine Learning Approach to Improving Dynamic Decision Making," *Information Systems Research* (25:2), pp. 239-263.
- Mitchell, T. M. 1982. "Generalization as Search," *Artificial Intelligence* (18:2), pp. 203-226.
- Moons, K. G., Royston, P., Vergouwe, Y., Grobbee, D. E., and Altman, D. G. 2009. "Prognosis and Prognostic Research: What, Why, and How?," *BMJ* (338), pp. 1317-1320.
- Neal, R. 2011. "MCMC Using Hamiltonian Dynamics," in *Handbook of Markov Chain Monte Carlo* (Volume 2), S. Brooks, A. Gelman, G. Jones, and X.-L. Meng (eds.), London: Chapman & Hall/CRC Press, pp. 113-162.
- Neal, R. M. 1996. *Bayesian Learning for Neural Networks*, Secaucus, NJ: Springer-Verlag New York, Inc.
- Pan, S. J., and Yang, Q. 2010. "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering* (22:10), pp. 1345-1359.
- Parekh, A. K., and Barton, M. B. 2010. "The Challenge of Multiple Comorbidity for the US Health Care System," *Journal of the American Medical Association* (303:13), pp. 1303-1304.
- Reilly, B. M., and Evans, A. T. 2006. "Translating Clinical Research into Clinical Practice: Impact of Using Prediction Rules to Make Decisions," *Annals of Internal Medicine* (144:3), pp. 201-209.
- Rousseeuw, P. J., and Molenberghs, G. 1994. "The Shape of Correlation Matrices," *The American Statistician* (48:4), pp. 276-279.
- Shmueli, G., and Koppius, O. R. 2011. "Predictive Analytics in Information Systems Research," *MIS Quarterly* (35:3), pp. 553-572.
- Simm, J., de Abril, I. M., and Sugiyama, M. 2014. "Tree-Based Ensemble Multi-Task Learning Method for Classification and Regression," *IEICE Transactions on Information and Systems* (97:6), pp. 1677-1681.
- Simon, H. A. 1955. "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics* (69:1), pp. 99-118.
- Singh, A., Nadkarni, G., Gottesman, O., Ellis, S. B., Bottinger, E. P., and Gutttag, J. V. 2014. "Incorporating Temporal EHR Data in Predictive Models for Risk Stratification of Renal Function Deterioration," *Journal of Biomedical Informatics* (53), pp. 220-228.
- Smith, B. J., and Mezhir, J. J. 2014. "An Interactive Bayesian Model for Prediction of Lymph Node Ratio and Survival in Pancreatic Cancer Patients," *Journal of the American Medical Informatics Association* (21:e2), pp. e203-e211.
- Stevens, R. J., Kothari, V., Adler, A. I., and Stratton, I. M. 2001. "The UKPDS Risk Engine: A Model for the Risk of Coronary Heart Disease in Type II Diabetes (UKPDS 56)," *Clinical Science* (101:6), pp. 671-679.
- Tabak, Y. P., Sun, X., Nunez, C. M., and Johannes, R. S. 2014. "Using Electronic Health Record Data to Develop Inpatient Mortality Predictive Model: Acute Laboratory Risk of Mortality Score (ALaRMS)," *Journal of the American Medical Informatics Association* (21:3), pp. 455-463.
- Tammemägi, M. C., Katki, H. A., Hocking, W. G., Church, T. R., Caporaso, N., Kvale, P. A., Chaturvedi, A. K., Silvestri, G. A., Riley, T. L., Commins, J., and Berg, C. D. 2013. "Selection Criteria for Lung-Cancer Screening," *New England Journal of Medicine* (368:8), pp. 728-736.
- Thomas, R., Kanso, A., and Sedor, J. R. 2008. "Chronic Kidney Disease and its Complications," *Primary Care* (35:2), pp. 329-344.
- Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B (Methodological)* (58:1), pp. 267-288.
- Tipping, M. E. 2001. "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research* (1), pp. 211-244.
- Toll, D. B., Janssen, K. J. M., Vergouwe, Y., and Moons, K. G. M. 2008. "Validation, Updating and Impact of Clinical Prediction Rules: A Review," *Journal of Clinical Epidemiology* (61:11), pp. 1085-1094.
- Venkatesh, V., Zhang, X., and Sykes, T. A. 2011. "'Doctors Do Too Little Technology': A Longitudinal Field Study of an

- Electronic Healthcare System Implementation,” *Information Systems Research* (22:3), pp. 523-546.
- Wolpert, D. H., and Macready, W. G. 1997. “No Free Lunch Theorems for Optimization,” *IEEE Transactions on Evolutionary Computation* (1:1), pp. 67-82.
- White House, The. 2015. “President Obama’s Precision Medicine Initiative” (available at <http://www.whitehouse.gov/node/319876>).
- World Health Organization. 2014a “Cardiovascular Diseases Fact Sheet” (available at <http://www.who.int/mediacentre/factsheets/fs317/en/index.html>).
- World Health Organization. 2014. “Diabetes Fact Sheet” (available at <http://www.who.int/mediacentre/factsheets/s312/en/>).
- Xu, L., Duan, J. A., and Whinston, A. 2014. “Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion,” *Management Science* (60:6), pp. 1392-1412.
- Xue, Y., Liao, X., Carin, L., and Krishnapuram, B. 2007. “Multi-Task Learning for Classification with Dirichlet Process Priors,” *Journal of Machine Learning Research* (8), pp. 35-63.
- Yaraghi, N., Du, A. Y., Sharman, R., Gopal, R. D., and Ramesh, R. 2015. “Health Information Exchange as a Multisided Platform: Adoption, Usage and Practice Involvement in Service Co-Production,” *Information Systems Research* (26:1), pp. 1-18.
- Zhou, J., Yuan, L., Liu, J., and Ye, J. 2011. “A Multi-Task Learning Formulation for Predicting Disease Progression,” in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, CA: ACM Press, pp. 814-822.

About the Authors

Yu-Kai Lin is an Assistant Professor of Business Analytics at Florida State University College of Business. He received a B.B.A. degree from Chung Yuan Christian University (Taiwan), an MBA from National Tsing Hua University (Taiwan), and a Ph.D. in Management Information Systems from the University of Arizona. His current research interests are data analytics for decision support and innovative IT applications, particularly in the healthcare context. His research has appeared in *Journal of Biomedical Informatics* and *IEEE Intelligent Systems*, and he has presented his work at the Workshop on Information Systems Economics, the Workshop on Information Technologies and Systems, the Workshop on Health IT Economics, and INFORMS. He is a member of the AIS and INFORMS.

Hsinchun Chen is Arizona Regents’ Professor, Thomas R. Brown Chair Professor in Management and Technology, and Director of the Artificial Intelligence Lab at the University of Arizona. He graduated with a Bachelor of Science degree from National Chiao-Tong University (Taiwan), an MBA from SUNY Buffalo, and MS

and Ph.D. from New York University. He is a Fellow of ACM, IEEE, and AAAS. Hsinchun recently served as the lead program director of the Smart and Connected (SCH) Program at the NSF (2014-2015), a multi-year multi-agency health IT research program in the United States. His research interests include artificial intelligence, business intelligence and analytics, and design science. He is also a successful IT entrepreneur. His COPLINK/i2 system for security analytics was commercialized in 2000 and acquired by IBM as its leading government analytics product in 2011.

Randall A. Brown received his medical degree from Rush Medical College in Chicago and completed residency training in internal medicine at the University of Michigan, Ann Arbor. He is affiliated with the University of Arizona College of Medicine as Assistant Professor of Medicine, Clinical Scholar. He has over 20 years of clinical medical experience teaching and practicing medicine in academic tertiary medical centers. He holds an MBA degree from the University of Arizona Eller College of Management and has significant healthcare consulting experience. He has worked as a healthcare domain expert and medical consultant to the Artificial Intelligence Laboratory of the Eller College of Management, MIS Department, advising on healthcare data mining, predictive analytics and disease progression research.

Shu-Hsing Li is Executive Vice President for Financial Affairs and a professor of accounting at National Taiwan University (NTU). He was the dean of NTU College of Management from 2010 to 2013 and was the head of the Department of Accounting from 2006 to 2009. He received his bachelor’s degree in business administration from National Chengchi University in Taiwan, and Ph.D. degree in accounting from New York University. His academic publications have appeared in *The Accounting Review*, *Auditing: A Journal of Practice & Theory*, *Journal of Accounting, Auditing and Finance*, *European Journal of Operational Research*, *IEEE Intelligent Systems*, *Decision Support Systems*, and other scholarly journals. He is now the editor of *Asia-Pacific Journal of Accounting & Economics*. He is the leading scholar in Taiwan working on transfer pricing for multinational companies. He was the independent board director for Taiwan Financial Holdings.

Hung-Jen Yang is a physician executive and is currently a visiting fellow at Stanford University. He holds an M.D. degree from National Taiwan University Medical School and a master’s of public health from Harvard. He is the CEO of MissionCare—an international joint commission accredited health care system in Taiwan. Prior to joining the MissionCare Group, he worked at Tenet Healthcare as a financial analyst. He is interested in hospital management, health economics, financial engineering, and strategic management. He received the 2010 Ernst & Young Taiwan Entrepreneur Award for successfully listing his organization on the Taipei OTC, making it the only hospital group listed in Taiwan.

HEALTHCARE PREDICTIVE ANALYTICS FOR RISK PROFILING IN CHRONIC CARE: A BAYESIAN MULTITASK LEARNING APPROACH

Yu-Kai Lin

College of Business, Florida State University, Tallahassee, FL 32306 U.S.A. {ylin@business.fsu.edu}

Hsinchun Chen

Eller College of Management, University of Arizona, Tucson, A 895721 U.S.A. {hchen@eller.arizona.edu}

Randall A. Brown

Southern Arizona VA Health Care System, Tucson, AZ 85723 U.S.A. {Randall.Brown4@va.gov}

Shu-Hsing Li

College of Management, National Taiwan University, Taipei City 106, TAIWAN {shli@ntu.edu.tw}

Hung-Jen Yang

The Walter H. Shorenstein Asia-Pacific Research Center, Stanford University,
Stanford, CA 94305 U.S.A. {berniejohan@gmail.com}

Appendix A

Bayesian Multitask Learning for Artificial Neural Networks

We have shown in the main text how to apply the proposed Bayesian Multitask Learning (BMTL) approach to a set of baseline logistic regression models. The BMTL approach is applicable to other baseline models as well. To demonstrate the generalizability of the BMTL approach, we describe how to apply BMTL to artificial neural networks (ANNs) in this appendix. In the interest of consistency and for ease of exposition, we reuse the notations in equations (1) to (7) in the main text whenever possible.

Consider a feed-forward ANN with one single hidden layer. A typical functional form of the ANN is

$$y_i^{(k)} = \text{logit}(F^{(k)}(x_i)), \quad i = 1 \dots N, \quad k = 1 \dots K \quad (\text{A1})$$

where

$$F^{(k)}(x_i) = \alpha_0^{(k)} + \sum_{h=1}^H \alpha_h^{(k)} \pi \left(\beta_{h0}^{(k)} + \sum_{j=1}^J \beta_{hj}^{(k)} x_{ij} \right) \quad (\text{A2})$$

The π in (A2) is referred to as an activation function in the literature of ANNs and is often nonlinear. Two common choices for π are the logistic and the hyperbolic tangent functions. The $\alpha_0^{(k)}$, $\alpha_h^{(k)}$, $\beta_{h0}^{(k)}$, and $\beta_{hj}^{(k)}$ are task-specific parameters to be fitted. The $\alpha_0^{(k)}$ and $\beta_{h0}^{(k)}$ are the *biases*

for the output and hidden nodes, and the $\alpha_h^{(k)}$ and $\beta_{hj}^{(k)}$ are the *weights* for the respective input units. To achieve BMTL, we set the following prior distributions for these parameters.

$$\alpha_0^{(k)}, \beta_{h0}^{(k)} \sim \text{Cauchy}(0, 10), \quad k = 1, \dots, K; \quad h = 1, \dots, H \tag{A3}$$

$$\alpha_h \sim \text{MVN}(0, u_h^2 A_h), \quad h = 1, \dots, H \tag{A4}$$

$$\beta_{hj} \sim \text{MVN}(0, s_{hj}^2 B_{hj}), \quad h = 1, \dots, H; \quad j = 1, \dots, J \tag{A5}$$

In (A4) and (A5), $\alpha_j = [\alpha_j^{(1)}, \alpha_j^{(2)}, \dots, \alpha_j^{(K)}]^T$ and $\beta_{hj} = [\beta_{hj}^{(1)}, \beta_{hj}^{(2)}, \dots, \beta_{hj}^{(K)}]^T$. At this point, it is straightforward to draw hyper prior distributions for u_h and s_{hj} as we did for r_j in (4) and (5). Similarly, A_h and B_{hj} will follow the same formulation as Σ_j in (6).

Appendix B

Robust Check for Evaluation 3 Using Different Decision Thresholds

STK (# of events = 828)		Predicted Risk (BMTL-Logit)		Predicted Risk (Logit-lasso)		Predicted Risk (UKPDS)			
		Low	High	Low	High	Low	High		
Preventive treatment prescribed at/before v_{0j}	Yes	28	424	Yes	40	412	Yes	344	108
	No	69	307	No	88	288	No	278	98

AMI (# of events = 225)		Predicted Risk (BMTL-Logit)		Predicted Risk (Logit-lasso)		Predicted Risk (UKPDS)			
		Low	High	Low	High	Low	High		
Preventive treatment prescribed at/before v_{0j}	Yes	85	76	Yes	88	75	Yes	65	96
	No	56	8	No	59	5	No	28	36

ARF (# of events = 571)		Predicted Risk (BMTL-Logit)		Predicted Risk (Logit-lasso)		Predicted Risk (UKPDS)			
		Low	High	Low	High	Low	High		
Preventive treatment prescribed at/before v_{0j}	Yes	62	252	Yes	75	238	Yes	UKPDS does not predict ARF risks.	
	No	84	174	No	100	158	No		

Figure B1. Summary of Results in Evaluation 3 Using 5% as the Cut-Off Value for High/Low Risks

STK (# of events = 828)		Predicted Risk (BMTL-Logit)		Predicted Risk (Logit-lasso)		Predicted Risk (UKPDS)			
		Low	High	Low	High	Low	High		
Preventive treatment prescribed at/before v_{0i}	Yes	208	244	Yes	205	247	Yes	436	16
	No	309	67	No	312	64	No	354	22

AMI (# of events = 225)		Predicted Risk (BMTL-Logit)		Predicted Risk (Logit-lasso)		Predicted Risk (UKPDS)			
		Low	High	Low	High	Low	High		
Preventive treatment prescribed at/before v_{0i}	Yes	126	35	Yes	125	36	Yes	140	21
	No	63	1	No	63	1	No	60	4

ARF (# of events = 571)		Predicted Risk (BMTL-Logit)		Predicted Risk (Logit-lasso)		Predicted Risk (UKPDS)			
		Low	High	Low	High	Low	High		
Preventive treatment prescribed at/before v_{0i}	Yes	174	139	Yes	181	132	Yes	UKPDS does not predict ARF risks.	
	No	196	62	No	200	58	No		

Figure B2. Summary of Results in Evaluation 3 Using 20% as the Cut-Off Value for High/Low Risks