Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# Predictive Analytics for Readmission of Patients with Congestive Heart Failure

Indranil Bardhan, Jeong-ha (Cath) Oh, Zhiqiang (Eric) Zheng, Kirk Kirksey

*Information Systems Research*
Vol. 26, No. 1, March 2015, pp. 19–39
ISSN 1047-7047 (print) | ISSN 1526-5536 (online)
http://dx.doi.org/10.1287/isre.2014.0553
© 2015 INFORMS

# Predictive Analytics for Readmission of Patients with Congestive Heart Failure

Indranil Bardhan
Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080, bardhan@utdallas.edu

Jeong-ha (Cath) Oh
J. Mack Robinson College of Business, Georgia State University, Atlanta, Georgia 30302, jhoh@gsu.edu

Zhiqiang (Eric) Zheng
Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080, ericz@utdallas.edu

Kirk Kirksey
University of Texas Southwestern Medical Center, Dallas, Texas 75390, kirk.kirksey@utsouthwestern.edu

Mitigating preventable readmissions, where patients are readmitted for the same primary diagnosis within 30 days, poses a significant challenge to the delivery of high-quality healthcare. Toward this end, we develop a novel, predictive analytics model, termed as the beta geometric Erlang-2 (BG/EG) hurdle model, which predicts the propensity, frequency, and timing of readmissions of patients diagnosed with congestive heart failure (CHF). This unified model enables us to answer three key questions related to the use of predictive analytics methods for patient readmissions: whether a readmission will occur, how often readmissions will occur, and when a readmission will occur. We test our model using a unique data set that tracks patient demographic, clinical, and administrative data across 67 hospitals in North Texas over a four-year period. We show that our model provides superior predictive performance compared to extant models such as the logit, BG/NBD hurdle, and EG hurdle models. Our model also allows us to study the association between hospital usage of health information technologies (IT) and readmission risk. We find that health IT usage, patient demographics, visit characteristics, payer type, and hospital characteristics, are significantly associated with patient readmission risk. We also observe that implementation of cardiology information systems is associated with a reduction in the propensity and frequency of future readmissions, whereas administrative IT systems are correlated with a lower frequency of future readmissions. Our results indicate that patient profiles derived from our model can serve as building blocks for a predictive analytics system to identify CHF patients with high readmission risk.

*Keywords*: patient readmissions; healthcare information technologies; congestive heart failure; predictive healthcare analytics

*History*: Il-Horn Hann, Senior Editor; Sudip Bhattacharjee, Associate Editor. This paper was received on October 14, 2011, and was with the authors 13 months for 4 revisions. Published online in *Articles in Advance* November 24, 2014.

## 1. Introduction

Readmission of patients with chronic diseases is a significant and growing problem in the United States and an increasing burden on the healthcare system. Preventable patient readmissions cost the U.S. healthcare system about $25 billion every year, according to a study by PricewaterhouseCoopers (2010). Experts believe that high readmission rates, when patients are readmitted within 30 days of discharge, indicate that the nation's hospitals are not adequately addressing patient health issues. To tackle this problem, the U.S. Centers for Medicare and Medicaid Services (CMS) has imposed penalties on hospitals for preventable readmissions related to chronic conditions such as heart failure or pneumonia, starting in 2012.

We develop a novel, predictive analytics model to predict patient readmission rates based on clinical, patient, and hospital information technologies (IT) characteristics. We focus on patients diagnosed with CHF, because this represents one of the first two health conditions that the Department of Health and Human Services (HHS) began to monitor, starting in 2012. Specifically, our research seeks to develop a healthcare analytics model that considers the propensity, frequency, and timing of patient readmissions. That is, for a given patient, we are interested in studying: (a) What is the likelihood of a future readmission? (b) How many future readmissions are likely to occur? And (c) when will the next readmission occur? Our model stands in sharp contrast to the existing readmission literature that mostly focuses on one or the other, but seldom addresses all of the above research questions in an integrated manner (Chin and Goldman 1997, Philbin and DiSalvo 1999, Krumholz et al. 2000, Silverstein et al. 2008, Kansagara et al. 2011). Our proposed beta geometric/Erlang-2 gamma (BG/EG) hurdle model addresses these research questions simultaneously.

A key aspect to improving the quality of healthcare and reducing patient readmissions is the implementation of health information technology (HIT). The use of HIT has the potential to improve healthcare quality, reduce readmission rates and costs, and consequently increase productivity (Congressional Budget Office 2008, Miller and Tucker 2011). Past studies have primarily focused on the impact of HIT on the productivity and operational efficiency of hospitals and/or healthcare providers (Bardhan and Thouin 2013, Das et al. 2011, Hillestad et al. 2005). However, to the best of our knowledge, they have not analyzed the impact of HIT on *patient-level* readmissions (e.g., see the extensive review by Kansagara et al. 2011). In their commentary on digital transformation of healthcare, Agarwal et al. (2010, p. 796) identify the measurement and quantification of HIT payoff and its impact on *patient care outcomes* as a significant area for future research. Our research fills this gap in the literature by examining whether adoption of HIT is associated with a reduction in the readmission risk of CHF patients, through our proposed model.

To fully account for patient and hospital heterogeneity, it is important to conduct a comprehensive study based on a large, longitudinal panel of patients across multiple hospitals to evaluate the readmission risk of patients. We obtained such a unique data set, which tracks a large panel of patient admissions across hospitals in North Texas. The data was gleaned from hospitals' administrative claims systems electronically and integrated across all hospitals in the region using a unique master patient index.

Our results indicate that health IT applications, patient demographics, payer type, admission condition, and comorbidities are important determinants of patient readmission risk. Our model offers a more nuanced view of patient readmissions when we differentiate the propensity of (initial) readmission from the frequency of (future) readmissions. We find that usage of cardiology information systems is associated with a reduction in the propensity and frequency of future readmissions, whereas administrative IT system usage is correlated with a lower frequency of future readmissions. Furthermore, we compare the predictive performance of our model with several state-of-the art models that have been proposed in the extant literature. Our "horse race" experiment demonstrates the superiority of our model in predicting both the incidence and timing of future readmissions. Because improved prediction is a foundational step toward mitigating future readmissions, our model can serve as an integral component of a predictive healthcare analytics tool to better profile, predict, and take preventive actions on patients with high readmission risk.

## 2. Background

### 2.1. Readmission Analytics

Extant readmission studies have typically been based on a single hospital using relatively small samples, or are restricted to a specific cohort such as elderly patients (e.g., Joynt et al. 2011, Shelton et al. 2000, Silverstein et al. 2008), veterans (e.g., Muus et al. 2010), or specific racial and income groups (e.g., Philbin et al. 2001). A few studies have used data from several hospitals (e.g., Philbin et al. 2001), but usually they belong to the same hospital system (Amarasingham et al. 2010, Deswal et al. 2004, Silverstein et al. 2008), overlooking the possibility of patient admissions across multiple (disparate) hospitals. This can lead to serious undercounting of patient readmissions because it is not uncommon for patients to be admitted (over time) to different hospitals that are owned by different entities. In fact, our data shows that 37.4% of CHF patients who were readmitted within 30 days of their initial admission visit a different hospital. As Nasir et al. (2010) observe, the same-hospital readmission rate is likely to underreport the actual interhospital readmission rate by as much as 50%, and is of limited value as a benchmark for care quality. Hence, it is important to analyze data that provide a complete picture of patient admissions across hospitals within a geographic region.

Besides modeling the risk propensity of readmission, it is equally important to understand *how frequently* (count) and *when* (timing) readmissions are likely to occur using a healthcare analytics system. Some examples of count and timing models in healthcare include the count of health service utilization, such as physician consultations, emergency room visits, or the amount of home care received (Deb and Trivedi 2002, Winkelmann 2006). Our proposed readmission model draws on count models in statistics (Winkelmann 2010) and consumer repeat-buying models in the marketing literature (e.g., Fader et al. 2005, Gupta 1991).

### 2.2. Health IT and Patient Readmissions

The information systems research literature has witnessed a growing interest in the role of IT in patient diagnosis, healthcare delivery, and treatment. Prior studies have mostly focused on hospital performance and the impact of IT (Devaraj and Kohli 2000, 2003; Menon et al. 2000; Das et al. 2011), or on hospital-level adoption and diffusion of health IT (Angst et al. 2010, Agarwal et al. 2010, Zheng et al. 2005). However, there is a growing emphasis on patient-level analytics as researchers and clinicians have come to recognize that the ultimate impact of HIT has to be measured on patient-level outcomes, and therefore, recent studies have called for greater attention using patient-level data to generate useful and actionable insights (Angst et al. 2010, Gao et al. 2010).

The growth in adoption of electronic medical records (EMR) and HIT systems in recent years has spurred widespread interest in studying the impact of HIT applications on patient care outcomes (Anderson and Agarwal 2011, Bardhan and Thouin 2013). Prior studies report improved quality of care in diabetes treatment with the use of EMR systems (Cebul et al. 2011), and in a recent review, Buntin et al. (2011) report that 92% of recent studies document a positive impact of HIT on hospital outcomes, including healthcare quality and efficiency. Whereas a few studies on the use of computerized provider order entry (CPOE) systems report improvements related to medication errors (Aron et al. 2011, Kaushal et al. 2003), others report unintended adverse consequences (Campbell et al. 2006), such as a sudden increase in mortality rates after implementation of CPOE (Han et al. 2005). Usage of automated notes and record systems, order entry, and clinical decision support systems, has helped lower complications and mortality rates (Amarasingham et al. 2009, McCullough et al. 2010, Miller and Tucker 2011). However, as reported in a recent comprehensive review by Kansagara et al. (2011), there have not been any studies on the relationship between health IT systems and patient readmission rates.

Although recent studies have reported mixed evidence on the impact of HIT on the quality of patient care, they have been limited by data deficiencies and limitations in their econometric estimation methods. For example, Linder et al. (2009) use cross-sectional, pooled data analysis of patient visit data but do not take into account the possibility of serial correlation among multiple visits by the same patient over time, which may lead to biased estimates in ordinary least squares (OLS) regressions. McCullough et al. (2010) and DesRoches et al. (2010) focus specifically on two types of HIT applications—EHR and CPOE systems—and study whether hospitals with these systems exhibit greater levels of process quality compared to hospitals without these systems. They ignore the role of ancillary HIT applications, such as radiology, laboratory, and order communication systems, which support decision making related to patient care, and the impact of nonclinical applications, such as patient scheduling systems, human resource systems, and financial systems, on information workflows. A recent study reports nine potential areas where health IT can be utilized to reduce readmissions directly, including case management, communication, analytics and modeling, postacute follow up, health information exchanges, social media, mobility, robotics, and innovation (HIMSS 2012). In particular, it advocates the use of HIT, such as EMR and risk assessment software, in improving care coordination and transitions from admission to discharge by facilitating patient assessment and discharge planning.

A major difference between our study and the extant literature is our focus on health IT applications and their relationship with the risk, frequency, and timing of patient readmissions. We focus on three classes of hospital IT applications, namely, cardiology specific, general clinical, and hospital administrative systems. *Cardiology information systems* enhance patient safety by serving as a repository of patient cardiac information across the continuum of cardiac care. Such systems support cardiac and peripheral catheterization, hemodynamics monitoring, echocardiography, vascular ultrasound, nuclear cardiology, and ECG management, and integrate information and imaging data from multiple systems, enabling clinicians to make optimal care decisions (Pratt 2010).

*Clinical information systems* improve decision-making capabilities associated with care management. For instance, use of CPOE systems not only speeds up transmission of a patient's prescription to a pharmacy, thereby reducing delays but also (a) reduces the need for nurses or physician assistants to transcribe prescriptions, thereby lowering the potential for medication transcription errors, and (b) provides decision support capabilities to flag possible drug–drug and drug-allergy interactions at the time when a physician enters the prescription. Such IT-enabled capabilities within CPOE applications reduce the incidence of adverse drug events and are expected to yield significant savings in inpatient care as well as outpatient visits (Amarasingham et al. 2009). Clinical systems aid in short-term preventive care as well as disease management of chronic diseases such as CHF. For example, heuristics within EMR systems can identify patients in need of follow-up cardio tests, remind physicians to order needed tests and schedule preventive care visits, and provide consistent records of clinical test results, thereby leading to better clinical outcomes. Case management systems (within EMRs) also help to coordinate workflows, such as communication between multiple specialists and high-risk patients.

Hospital *administrative systems* also play an important role in the delivery and coordination of patient care. Benefits management portals enable cross-functional integration of data across multiple departments, and patient administration systems track patient movement in inpatient settings, allowing clinicians and supporting staff to improve hospital resource utilization by reducing waiting time at the point of admission, discharge, or transfer (Bardhan and Thouin 2013). Other administrative applications, such as personnel management systems, support staff needs related to patient education and discharge transition, which is critical to reducing readmission risk.

## 3. Model Development

We first briefly describe two baseline estimation models that have been widely used in the readmission literature: the logistic regression and the proportional hazard models. We then address their methodological deficiencies and propose a predictive model for readmission analytics to address these challenges.

### 3.1. Baseline Models

The readmission literature has commonly used logistic regression models to estimate the readmission probability of patients (Muus et al. 2010, Philbin and DiSalvo 1999, Shelton et al. 2000, Silverstein et al. 2008). These studies model the incidence of a readmission after a patient's initial visit as a binary outcome, and involve *patient-level analysis* where the unit of analysis is a patient. Hence, the readmission propensity of each patient is defined as a logit function of covariates. Another type of baseline model uses survival analysis (or hazard models) to estimate the time duration between consecutive patient readmissions. It considers each visit as the unit of analysis, and hence, is called *visit-level analysis*. In our case, the hazard rate, $h(t)$, refers to the readmission rate of a patient per unit of time, i.e., the readmission rate of a patient on a given day, which is defined as,

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t \mid T > t)}{\Delta t}.$$

The hazard rate is also often expressed through the survival function, $S(t) = 1 - F(t)$, where $F(t)$ is the cumulative distribution function of the time to failure. The hazard function, $h(t)$, provides the instantaneous readmission rate that a patient, who is not readmitted by time $t$, will be readmitted during the infinitesimally small time interval, $(t,\ t + \Delta t)$. The commonly used survival model is the Cox proportional hazard model (Cox 1972, Krumholz et al. 2000).

Though both approaches are useful in identifying readmission risk factors, they do not provide additional insights to develop a predictive model to estimate the frequency and timing of future readmissions. First, these models are typically limited in tackling the nonstationarity nature of patient readmissions, where a patient's readmission propensity often changes over time depending on her changing condition and treatment during prior admissions, commonly referred to as *state dependency* (Heckman 1991). Prior studies using logistic regression models (Amarasingham et al. 2010, Silverstein et al. 2008) or proportional hazard models (Alexander et al. 1999, Krumholz et al. 2000) typically only account for the first readmission; they do not track multiple readmissions for the same patient over time. Second, such models do not fully account for unobserved patient heterogeneity, wherein some patients may be intrinsically healthier than others to start with.[1] The aforementioned prior studies only account for observed patient heterogeneity such as demographics, comorbidities, utilization patterns (Amarasingham et al. 2010), self-rated health conditions (Mudge et al. 2010), self-reported compliance to prescriptions (Chin and Goldman 1997), or prescreened samples to reduce heterogeneity in patient groups (Krumholz et al. 2000). Though logit or hazard models can be extended to account for unobserved patient heterogeneity (e.g., the mixed proportional hazard model), to the best of our knowledge, such extensions have not been applied to the healthcare literature to model readmissions.

Third, extant readmission models do not capture the timing or frequency of readmissions. For example, logistic regression models record a readmission as a binary outcome based on the occurrence or absence of a readmission (Amarasingham et al. 2010, Silverstein et al. 2008), regardless of the number of occurrences of such readmissions for a patient. Furthermore, although hazard models partially address the issue of data censoring, they typically assume that the censored data follow the same stationary process as the observed data, which is a severe limitation of most readmission studies. In contrast, our proposed model considers the survival after each admission, thus directly addressing the data censoring issue. We summarize the major differences between our model and the extant readmission literature in Table 1.

### 3.2. The BG/EG Hurdle Model

To address the deficiencies of existing baseline models, we now develop a stochastic model, the beta geometric/Erlang-2 gamma hurdle model, to better predict patient readmission patterns. Our model has several distinctive properties compared to the baseline models. It accounts for (a) a patient's readmission propensity, frequency, and timing in an integrated manner; (b) nonstationarity in readmission rates; (c) both observed and unobserved patient heterogeneity; and (d) data truncation due to unobserved

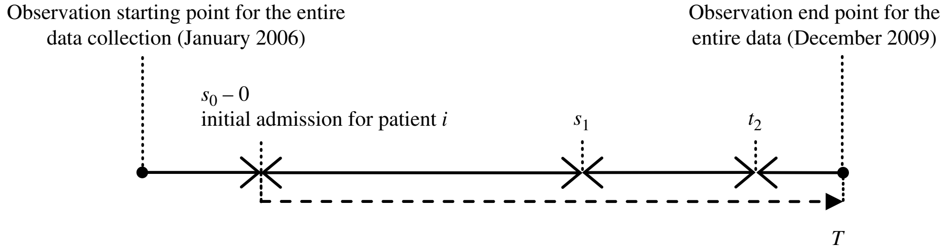
[1] Overlooking unobserved heterogeneity can lead to "weeding out effects" where duration dependence in the observed hazard function becomes more negative as the hazardous population tends to drop out first (Heckman 1991).

**Table 1** **A Contrast of the Literature with Our Study**

| | Domain | Research design | | | | Estimation methods | | Research model | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Research domain | Multi-hospital system analysis | 30-day readmissions | HIT impact on patient-level outcomes | Integrates frequency and timing of events | Trade-offs between Type I and Type II errors | Predictive performance | Individual level unobserved heterogeneity | Data truncation | Account for nonzero hurdle | Non-stationarity | Time-varying covariates | Continuous time |
| Philbin and DiSalvo (1999) | Healthcare | Y | 12-month | N | N | N | Y | N | N | N | N | N | N |
| Silverstein et al. (2008) | Healthcare | N | Y | N | N | N | Y | N | N | N | N | N | N |
| Krumholz et al. (2000) | Healthcare | N[1] | 6-month | N | N | N | N | N | N | N | N | N | N |
| Amarasingham et al. (2010) | Healthcare | N | Y | N | N | N | Y | N | N | N | N | N | N |
| Chin and Goldman (1997) | Healthcare | N | 60-day | N | N | N | N | N | N | N | N | N | N |
| Felker et al. (2004) | Healthcare | N[1] | 60-day | N | N | N | Y | N | N | N | N | N | N |
| Alexander et al. (1999) | Healthcare | Y | 12-month | N | N | N | N | N | Y | N | N | N | N |
| Angst et al. (2010) | Info. Sys. | Y | N.A. | N | N | N.A. | N.A. | N.A. | Y | N.A. | N.A. | Y | N |
| Devaraj and Kohli (2003) | Info. Sys. | N | N.A. | N | N.A. | N.A. | N.A. | N.A. | N.A. | N.A. | N.A. | Y | N |
| Bardhan and Thouin (2013) | Info. Sys. | Y | N.A. | Y | N.A. | N.A. | N.A. | N | N.A. | N | N.A. | Y | N |
| Gupta (1991) | Marketing | N.A. | N.A. | N.A. | Y | N | Y | Y[3] | Y | N | Y[2] | Y | N |
| Fader et al. (2004) | Marketing | N.A. | N.A. | N.A. | Y | N | Y | Y[3] | Y | N | Y | Y | N |
| Fader et al. (2005) | Marketing | N.A. | N.A. | N.A. | Y | N | Y | Y[3] | Y | N | N | N | N |
| Schweidel and Knox (2013) | Marketing | N.A. | N.A. | N.A. | Y | N | Y | Y[3] | Y | N | Y | Y | Y |
| Gönül and Ter Hofstede (2006) | Marketing | N.A. | N.A. | N.A. | Y | N | Y | Y[3] | Y | N | Y | Y | Y |
| Jain and Vilcassim (1991) | Marketing | N.A. | N.A. | N.A. | Y | N | N | Y[3] | Y | N | Y | Y | N |
| Winkelmann (2004) | Economics | N.A. | N.A. | N.A. | N | N | N | Y[3] | Y | Y | N | Y | N |
| This study | Healthcare & Info. Sys. | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |

[1] Patients across multihospital systems are studied but readmission is only to a single hospital.
[2] Nonstationarity through time-varying covariates.
[3] These papers consider unobserved heterogeneity at the population level; we consider both population- and individual-level heterogeneity (thru random effects).

**Figure 1** **Illustration of Patient Readmission Patterns**


Observation starting point for the entire data collection (January 2006)
Observation end point for the entire data (December 2009)
$s_0 - 0$
initial admission for patient $i$
$s_1$
$t_2$
$T$

causes such as patients' death or migration out of the geographic region.[2]

Our model integrates two components: a hurdle component, which estimates the *probability* of readmission, and a BG/EG component, which estimates the *frequency and timing* of future readmissions. The hurdle model is suitable when one believes that patients who are admitted once need to be treated differently from those who are readmitted multiple times. The hurdle component not only estimates the probability of readmission but also addresses the excessive zero count, as observed in our data, where 70% of CHF patients are not readmitted (Winkelmann 2010). Wooldridge (2010, p. 690) refers to this scenario as the *participation decision* because the hurdle model reflects a decision-maker's choice on whether or not to participate in an event (i.e., readmission). Specifically, the hurdle component models the probability of a zero outcome using a logit model as

$$\log\left(\frac{\theta_{0i}}{1-\theta_{0i}}\right) = X_{0i}\cdot\zeta_{0i}, \quad i = 1, \ldots, N, \tag{1}$$

where $\theta_{0i}$ is the probability of no readmission for an individual patient $i$ and $X_{0i}$ is the set of covariates observed for patient $i$ at their initial admission time with coefficients $\zeta_{0i}$. A hurdle regression considers systematically different statistical processes for the zero and nonzero binary outcomes, where the positive counts are conditioned on having nonzero outcomes. It reflects a two-stage process, where the risk factors affecting readmission frequency may be different from those determining the propensity of readmission.

The BG/EG component estimates the frequency and timing of admissions simultaneously. Wooldridge (2010, p. 690) refers to the frequency of events as the *amount decision*. Suppose we have $N$ patients, where patient $i$ is readmitted $J_i$ times at $(t_1, t_2, \ldots, t_{Ji})$ over the period $(0, T_i]$, where $t_0 = 0$ corresponds to the initial admission time and $T_i$ represents the censoring point, which is the end of the model calibration period for patient $i$. As each patient $i$ is admitted and readmitted at different times, $T_i$ varies across patients. Figure 1 provides a schematic representation of a patient's readmission patterns over time.

We assume that the time interval between two consecutive admissions follows an Erlang-2 distribution, which means that the timing of a patient's future readmission depends not only on the current visit but also on the previous admission. This relaxes the restrictive stationary assumption of the exponential distribution that is most common in proportional hazard models (Winkelmann 2010). By treating each admission as an independent random event, it not only overlooks the rich admission history of a patient, but can lead to an erroneous prediction of patients' future readmissions. A patient's readmission rate depends on the medical treatments that she receives, her health status, and prior hospitalization history. Consequently, researchers in the marketing literature have proposed that the Erlang-2 distribution be used to model interpurchase times, as it more closely resembles customer purchase behavior (Chatfield and Goodhardt 1973, Gupta 1991, Jeuland et al. 1980, Morrison and Schmittlein 1981, Fader et al. 2005). Hence, the Erlang-2 distribution is appropriate in our context of patient readmission behavior as it assumes that timing of the next admission is conditionally dependent on the duration of the previous admission.

The Erlang-2 distribution takes the form of $f_i(x, 2, \lambda) = \lambda^2 x e^{-\lambda x}$ for $x$, $\lambda \geq 0$, where $x$ is a continuous random variable and $\lambda$ is the (readmission) rate parameter. We follow Gupta's (1991) general approach, which specifies patient $i$'s probability, or hazard function, to pay a hospital visit in time period $t$, given a set of time-varying covariates, $X_t$, as

$$h(t, X_t) = \lambda_t \cdot e^{X_t \gamma_1} \equiv \lambda_t \cdot \phi(t), \tag{2}$$

where $\lambda_t$ is the baseline hazard at time $t$ (Cox 1972).

We follow Seetharaman and Chintagunta's (2003) formulation of the continuous time hazard model to incorporate time-varying covariates $(X_{t_1}, X_{t_2}, \ldots, X_{t_{Ji}})$

[2] Failure to account for data truncation is especially problematic for CHF, which tends to occur among older patients (who are 69 years old on average at discharge). For example, older patients typically tend to incur more readmissions, which is consistent with our model results.

in the Erlang-2 distribution. The patient-level survivor function of the interadmission time distribution between the $(j-1)$th and the $j$th admission is specified as

$$\begin{aligned} &S(t_j - t_{j-1}, X_{t_j}) \\ &\quad = \left(1 + \lambda_j \cdot \int_{t_{j-1}}^{t_j} \phi(u)\,du\right) \cdot \exp\left[-\lambda_j \int_{t_{j-1}}^{t_j} \phi(u)\,du\right] \\ &\quad = (1 + \lambda_j \psi(t_j, t_{j-1})) \exp[-\lambda_j \psi(t_j, t_{j-1})], \end{aligned} \tag{3}$$

where

$$\begin{gathered} \psi(t_j, t_{j-1}) \equiv \psi(t_j) - \psi(t_{j-1}); \\ \psi(t) \equiv \int_0^t \phi(u)\,du; \ \ \phi(t) \equiv e^{X_t \gamma_1}. \end{gathered} \tag{4}$$

The individual probability density function during the time interval $(t_{j-1}, t_j]$, given a covariate vector $X_{t_j}$, then follows:

$$\begin{aligned} &f(t_j - t_{j-1} \mid \lambda_j, \gamma_1, X_{t_j}) \\ &\quad = \lambda_j^2 \cdot \phi(t_j) \cdot \psi(t_j, t_{j-1}) \cdot \exp[-\lambda_j \psi(t_j, t_{j-1})]. \end{aligned} \tag{5}$$

The overall likelihood at the patient level is simply the product of Equation (5) over $j$ as

$$L(T_i, \gamma_1 \mid \lambda, X_t) = \prod_{j=1}^{J_i} f(t_j - t_{j-1} \mid \lambda_j, \gamma_1, X_{t_j}) + \tau_i, \tag{6}$$

where $\tau_i \sim N(\mu, \sigma^2)$ represents patient-level random effects to capture the unobserved heterogeneity of individual patients (where $\mu$ and $\sigma$ represent the mean and standard deviation of the random effect normal distribution function, respectively). We note that incorporating time-varying covariates into the Erlang-2 Gamma model in this manner is a new methodological contribution to the literature. Although Fader et al. (2004) also incorporate time-varying covariates into the Erlang-2 distribution, they only model the *grouped duration data* where time is grouped into weeks and the covariates are assumed to be constant during a week's interval. Wooldridge (2010) argues that this treatment is unsuitable for multispell data, where the event can occur multiple times during the chosen time interval. Since patient admissions can occur at any time and multiple times during a certain time period (e.g., a month), and considering that the covariates (comorbidities) may change at any time, we need to model readmissions along a *continuous* time frame.

Furthermore, to account for the fact that the rate of patient visits may differ across patients, we adopt the common mixture distribution for $\lambda$ (Winkelmann 2010), which is assumed to be gamma distributed with shape parameter $r$ and scale parameter $\alpha$: $g(\lambda \mid r, \alpha) = \alpha^r \lambda^{r-1} e^{-\lambda\alpha} \Gamma(r)^{-1}$. The flexibility associated with the gamma distribution allows it to fit various shapes of distributions because of its additive and conjugate properties. Gupta (1991) observes that the most appropriate specification for interpurchase time is a model that features an Erlang-2 interpurchase process with gamma-distributed purchase rates (to account for customer heterogeneity). Based on the assumptions that interadmission times are distributed according to (5) and that unobserved heterogeneity in $\lambda$ follows a gamma distribution, we specify the likelihood function as

$$\begin{aligned} L(T_i, \gamma_1, r, \alpha) = {} & \left(\prod_{j=1}^{J} \phi(t_j) \psi(t_j, t_{j-1})\right) \cdot \frac{\Gamma(r + 2J) \cdot \alpha^r}{\Gamma(r)} \\ & \cdot \left(\alpha + \sum_{j=1}^{J} \psi(t_j, t_{j-1}) + \psi(T, t_J)\right)^{-(r+2J)} \\ & + \tau_i. \end{aligned} \tag{7}$$

We further consider the possibility that, after every admission, a patient can become inactive with a dropout probability of $p$ from a geometric process. However, each patient is likely to have a different $p$, the cause of which may not always be observed, such as death or relocation outside the region, which leads to a data truncation problem. For a geometric process, this unobserved heterogeneity is commonly modeled through a beta mixing function for the binary outcome geometric processes (Winkelmann 2010). Taken together, this yields the beta-geometric distribution Fader et al. (2005), which is specified as

$$P(\text{dropout after } j\text{th admission}) = p(1-p)^{j-1}, \tag{8}$$

where the heterogeneity in dropout probabilities follows a beta distribution, with parameters $a$ and $b$ indicating the relative propensity of dropping out or not: $f(p \mid a, b) = p^{a-1}(1-p)^{b-1} B(a, b)^{-1}$.

Thus a patient may be inactive either after $T$ (i.e., no observation is made between the last admission and the end of the period) or right after the last admission. While dropping the subscript for each individual $i$ for brevity, we model these cases as follows:

(i) A patient is inactive after $T$:[3]

$$\begin{aligned} &L(\lambda \mid t_1, \ldots, t_J, T, \text{ inactive at time } \tau > T) \\ &\quad = \lambda^{2J} \left(\prod_{j=1}^{J} \phi(t_j) \psi(t_j, t_{j-1})\right) \\ &\qquad \cdot \exp\left(-\lambda \cdot \left(\sum_{j=1}^{J} \psi(t_j, t_{j-1}) + \psi(T, t_J)\right)\right). \end{aligned} \tag{9}$$

[3] For a full derivation of Equations (9)–(15), see Appendix A in the online supplement (available as supplemental material at http://dx.doi.org/10.1287/isre.2014.0553).

(ii) A patient becomes inactive right after the last admission $J$:

$$\begin{aligned}&L(\lambda \mid t_1, \ldots, t_J, T,\ \text{inactive at time } \tau \in (t_J, T])\\&\quad = \lambda^{2J}\left(\prod_{j=1}^{J} \phi(t_j)\psi(t_j, t_{j-1})\right)\\&\qquad \cdot \exp\left(-\lambda \cdot \sum_{j=1}^{J} \psi(t_j, t_{j-1})\right).\end{aligned} \tag{10}$$

This yields the likelihood function

$$\begin{aligned}&L(\lambda, p, \gamma \mid X, J, t_J, T)\\&\quad = (1-p)^J \lambda^{2J}\left(\prod_{j=1}^{J} \varphi(t_j)\psi(t_j, t_{j-1})\right)\\&\qquad \cdot \exp\left(-\lambda\left(\prod_{j=1}^{J} \varphi(t_j)\psi(t_j, t_{j-1}) + \psi(T, t_J)\right)\right)\\&\qquad + \delta_{J>0} p(1-p)^{J-1}\lambda^{2J}\left(\prod_{j=1}^{J} \varphi(t_j)\psi(t_j, t_{j-1})\right)\\&\qquad \cdot \exp\left(-\lambda\left(\prod_{j=1}^{J} \psi(t_j, t_{j-1})\right)\right) + \tau_i.\end{aligned} \tag{11}$$

Expectation over the distribution of $\lambda$ yields the likelihood function

$$\begin{aligned}&L(r, \alpha, p, \gamma \mid X, J, t_J, T)\\&\quad = (1-p)^J \cdot A_1 + \delta_{J>0} p(1-p)^{J-1} A_2 + \tau_i,\end{aligned} \tag{12}$$

where

$$\begin{aligned}A_1 &= \left(\prod_{j=1}^{J} \phi(t_j)\psi(t_j, t_{j-1})\right) \cdot \frac{\Gamma(r+2J) \cdot \alpha^r}{\Gamma(r)}\\&\quad \cdot \left(\alpha + \sum_{j=1}^{J} \psi(t_j, t_{j-1}) + \psi(T, t_J)\right)^{-(r+2J)},\end{aligned} \tag{13}$$

$$\begin{aligned}A_2 &= \left(\prod_{j=1}^{J} \phi(t_j)\psi(t_j, t_{j-1})\right) \cdot \frac{\Gamma(r+2J) \cdot \alpha^r}{\Gamma(r)}\\&\quad \cdot \left(\alpha + \sum_{j=1}^{J} \psi(t_j, t_{j-1})\right)^{-(r+2J)}.\end{aligned} \tag{14}$$

Taking expectation over $\lambda$ and $p$ yields the individual likelihood function $L_1$, for patients with at least one readmission as follows:

$$\begin{aligned}&L_1(r, \alpha, a, b, \gamma \mid X, J, t_J, T)\\&= A_1 \cdot \int_0^1 (1-p)^J \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\,dp\\&\quad + A_2 \cdot \int_0^1 p \cdot (1-p)^{J-1} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\,dp + \tau_i\\&= \left(\prod_{j=1}^{J} \phi(t_j)\psi(t_j, t_{j-1})\right) \cdot \frac{\Gamma(r+2J)}{\Gamma(r)} \cdot \alpha^r\\&\quad \cdot \frac{\Gamma(a+b)\Gamma(b+J-1)}{\Gamma(b)\Gamma(a+b+J)}\\&\quad \cdot \Bigg[(b+J-1) \cdot \left(\alpha + \sum_{j=1}^{J} \psi(t_j, t_{j-1}) + \psi(T, t_J)\right)^{-(r+2J)}\\&\qquad + \delta_{J>0} \cdot a \cdot \left(\alpha + \sum_{j=1}^{J} \psi(t_j, t_{j-1})\right)^{-(r+2J)}\Bigg] + \tau_i.\end{aligned} \tag{15}$$

Hence, the log-likelihood of $N$ patients, who have at least one readmission, is specified as

$$LL_1 = \sum_{i=1}^{N} \log\big(L_1(r, \alpha, a, b, \gamma \mid X_i(t), J_i, t_{J_i}, T_i)\big). \tag{16}$$

Combining it with the logit hurdle component, the overall likelihood function, $L_0$, for a patient with no readmission after the initial admission ($J = 0$), simplifies to

$$\begin{aligned}L_0 &= \int_0^\infty \exp(-\lambda\psi(T,0)) \cdot (1 + \lambda\psi(T,0)) \cdot \frac{\alpha^r \gamma^{r-1} e^{-\alpha\gamma}}{\Gamma(r)}\,d\lambda\\&= \left(\frac{\alpha}{\alpha + \psi(T,0)}\right)^r \cdot \left(1 + \frac{r\psi(T,0)}{\alpha + \psi(T,0)}\right).\end{aligned} \tag{17}$$

Altogether, this yields a *BG/EG hurdle model*, the likelihood function of which is given by

$$L = \prod_{i=1}^{N} \theta_0^{d_i} \frac{(1-\theta_0)^{1-d_i}}{(1-L_0)^{1-d_i}} \cdot L_1^{1-d_i}, \tag{18}$$

where

$$\theta_{0i} = P(J_i = 0) \quad \text{and} \quad d_i = 1 - \min\{J_i, 1\}.$$

The log-likelihood of the BG/EG hurdle model is therefore

$$\begin{aligned}LL = \sum_{i=1}^{N} \big(&d_i \cdot \log\theta_{0i} + (1-d_i) \cdot \log(1-\theta_{0i})\\&- (1-d_i) \cdot \log(1-L_{0i}) + (1-d_i) \cdot LL_{1i}\big),\end{aligned} \tag{19}$$

where the first two terms of the right-hand side refer to the *hurdle*, and the last two terms are the likelihood of positive count of readmissions.

### 3.3. Contrast with the Literature

We now recap the key methodological contributions that differentiate our study from the prior literature. Table 1 provides a summary of the key differentiators of our study, in contrast to the prior literature, where we draw on multiple disciplines. With respect to research design, there are several distinguishing features of our paper. First, none of the prior studies have focused on the association between health

IT and patient readmission risk. Second, although the marketing literature has numerous studies that integrate the estimation of the propensity and timing of consumer purchases, the prior healthcare research has been limited by its explicit focus on the estimation of readmission propensity, while ignoring the frequency and timing of future readmissions. This is the first study to provide an integrated model to estimate risk propensity *as well as* the frequency and timing of readmissions in a healthcare analytics context.

With respect to model estimation, while the issues of unobserved heterogeneity and data truncation have been previously studied in the marketing literature, this is the first study to explicitly account for the possibility of unobserved heterogeneity and data truncation (e.g., because of death or patient relocation) using patient readmission data. Similarly, extant healthcare studies on readmissions have not accounted for the possibility of nonstationarity in the patient readmission process, as well as the impact of time-varying covariates on readmission risk and frequency (Kansagara et al. 2011). With the exception of Winkelmann (2004), this is the only study to develop a hurdle-based estimation model that estimates the frequency and timing of future readmissions *once the first readmission hurdle has been crossed*. This study is among the first to account for trade-offs between types I and II errors in the development of a predictive model to study patient readmissions. Although a few marketing studies (e.g., Gupta 1991, Fader et al. 2005) also examine predictive performance of their respective models, they focus on prediction accuracy (i.e., type I error) without regard to the potential trade-offs with type II errors.

As shown in the last row of Table 1, this is the only study in the health IT domain to address the limitations of prior studies in terms of their research design, estimation methods, and model assumptions with respect to modeling patient readmissions.

## 4. Data

Our data consist of four years of patient admission records and clinical data from 67 hospitals in the North Texas region starting from January 2006 to December 2009. Patient visits across multiple hospitals are tracked by matching the regional master patient index (REMPI), developed by the Dallas Fort Worth Hospital Council (DFWHC) Research Foundation. A REMPI is a unique ID number assigned to each patient that allows us to track patients over time and across all hospitals in a region. In other words, a REMPI makes it possible to obtain a patient's entire readmission history and enables us to study the patterns of patient care and clinical diagnosis received across multiple hospitals with different ownership. We observe that this is a major improvement in our study compared to previous studies that have been restricted to studying patient readmission data from hospitals that belong to a single hospital (or health) system (Silverstein et al. 2008).

Our data records 65,188 admissions that originate from 40,983 distinct patients with CHF as their primary diagnosis. Among these patients, 70% had a single admission, whereas 30% (12,211) experienced multiple admissions as shown in Figure B.1 in the online appendix. Table B.1 in the online appendix provides a description of our sample data. Our data captures several patient demographic characteristics including gender, racial profile, and discharge age. Among all patients, 52% (21,281) are female, 72% (29,320) are Caucasian, and 21% (8,686) are African Americans. The average discharge age is 69 years, with 66% (27,134) of the patients being 65 years or older.

The hospital health IT usage data are drawn from the HIMSS analytics database for the corresponding four-year period, i.e., 2006 to 2009. After consulting with health IT practitioners, and based on the intensity of health IT usage among our sample of 67 hospitals, we identify 45 applications that are commonly used in the treatment of CHF patients. We omit various IT applications that are only relevant to general hospital management, and focus instead on clinical and administrative functions associated with CHF. Exploratory factor analysis (EFA) on this group of applications results in the selection of 18 health IT applications, where the factor scores are greater than the threshold of 0.6 for significance. Such clustering of HIT applications has been previously employed to group health IT applications according to their primary functionality (Bhattacherjee et al. 2007, Himmelstein et al. 2010, Bardhan and Thouin 2013).

We code each HIT application as a binary variable where one indicates that it has been implemented and operational, and zero indicates otherwise. Using EFA with Varimax rotation, we identify three major classes of health IT applications: administrative IT, clinical IT, and cardiology IT. *Administrative IT* consists of health information management applications, chart tracking, revenue cycle management, and patient billing applications. *Clinical IT* comprises hospital-wide clinical systems such as EMR, operating room IT, and CPOE applications. *Cardiology IT*, which are primarily used to treat CHF patients, include cardiology information systems, cath lab systems, and echocardiology and computerized tomography (CT) systems. The specific applications that comprise our three HIT factors are shown in Table B.2 in the online appendix. We observe that our HIT factors are generally consistent with previous research (e.g., Bhattacherjee et al. 2007, Menachemi et al. 2008).

Based on the EFA results, we calculate a summative index score for each factor that represents a ratio of the number of applications used in each hospital to the total number of applications. We normalize the summative index to a value between zero and one, which represents the percentage of health IT applications being used at a given hospital out of the entire class of applications.[4] As of 2009, hospitals in our sample have (on average) implemented 91% of administrative IT applications, 57% of clinical IT applications, and 23% of cardiology information systems. Compared to hospitals with below-average levels of administrative IT and cardiology IT, hospitals with above-average levels of these two HIT applications enjoy lower 30-day readmission rates by 3% and 4%, respectively. On the other hand, the 30-day readmissions rate of hospitals with an above-average level of clinical IT applications is not significantly different than hospitals in the below-average group.[5]

Our sample statistics show that 37% of patients who are readmitted visit different hospitals. This alarmingly large percentage suggests a severe undercounting for a single-hospital study, the common approach adopted in the extant readmission literature. Hence, for each patient readmission, we count the number of different hospitals visited by the patient prior to their current visit. We develop a new measure, *patient stickiness,* defined as the ratio of the number of times that a patient visits the same hospital to the total number of admissions until the present time. In other words, if a patient is treated within a single hospital (across multiple visits), her stickiness measure is higher compared to patients who are readmitted across multiple hospitals.

We also include other control variables that have been commonly used in the readmission literature (Mudge et al. 2010, Ross et al. 2008, Silverstein et al. 2008) including patient *demographics* (discharge age, gender, race), *length of stay* (LOS), *number of diagnoses, number of procedures, payer type, admission type*, and the *risk of mortality*. LOS is defined as the number of patient days from admission to discharge during an inpatient visit. In our sample, the average LOS is 5.45 days, with a mean of 12.58 diagnoses, 1.08 procedures recorded, and total hospital charges of $37,649. Payer type is classified into five categories according to their claim filing code: Medicare, Medicaid, self-pay, private insurance, and other insurance types. For each admission, hospitals record the admission type and the risk of patient mortality. The standard admission type is coded into six classes (class 1 denotes emergency). Risk mortality is coded on a scale from 1 to 4, which indicates the patient death risk as minor, moderate, major, or extreme, respectively.

CHF is likely to be accompanied by other comorbidities. Frequent comorbidities associated with CHF include diabetes mellitus, hypertension, peripheral vascular disease, chronic pulmonary disease, renal failure, anemia, alcohol abuse, drug abuse, and ischemic heart disease (Ross et al. 2008). We control for these comorbidity variables, which are identified by the Elixhauser index (Elixhauser et al. 1998) based on the ICD-9-CM (International Classification of Diseases) diagnosis codes.

Other hospital characteristics may also affect patient readmission rates. We include hospital-level control variables such as the number of beds, case mix index, and teaching/nonteaching hospital attributes. *Num_beds* is the number of beds available for use in each hospital, which serves as a proxy for hospital capacity. We control for the hospital *case mix index* (CMI), which accounts for the average severity of patients' disease case mix. *Tch-hosp* represents the academic status of the hospitals. Table 2 provides definitions as well as descriptive statistics on our model variables.

## 5. Empirical Analysis

We first discuss identification of the potentially endogenous HIT variables and then present the results of our empirical estimation, starting with the baseline results, followed by the BG/EG hurdle estimation.

### 5.1. Identification of the Health IT Effects

It is likely that the health IT variables that we construct in our study may be subject to endogeneity. For example, having a higher level of readmissions may prompt a hospital to implement HIT (i.e., simultaneity). Identification of the causal effect of health IT can be challenging, because it is hard to isolate it from a hospital's efforts to improve patient outcomes such as readmission reduction. To address endogeneity, we identify two instrumental variables (IV): average level of health IT in (external) peer hospital systems, and the difference in a hospital's level of health IT between the current and previous year. The first IV is defined as the average level of HIT among peer hospitals, after excluding other hospitals within the same health system as the focal hospital. The second IV is derived by taking the difference in the values

[4] We also experimented with factor scores (i.e., using factor loading as the weights) and the results remain consistent. We prefer the normalized summative index because it has a more intuitive interpretation.

[5] Based on a split sample analysis, we observe that the average readmission rate for hospitals with above-average usage of administrative IT is 12%, whereas that for hospitals with below-average administrative IT is 15% (the difference is significant at $p < 0.001$). Similarly, hospitals with above-average usage of cardiology IT have a 10% readmission rate, compared to a 14% readmission rate for below-average hospitals. This difference is also statistically significant ($p < 0.001$). On the other hand, high or low levels of clinical IT are not associated with a significant difference in average readmission rates.

**Table 2 Variable Definitions**

| Variable | Description of variable | Descriptive statistics |
|---|---|---|
| Demographic variables | | |
| *Gender* | Patient's gender | Female (52%), Male (48%) |
| *Patient race* | Patient's race | Caucasian (72%), African American (21%) |
| log(*disch_age*) | Patient age on the day of discharge (log transformed) | 69.17 (15.65)[a,b] |
| (log(*disch_age*))$^2$ | Quadratic term of log-transformed discharge age | |
| Health IT variables | | |
| *Administrative IT* | Normalized summative index of administrative IT applications | 0.91 (0.22)[a] |
| *Clinical IT* | Normalized summative index of clinical IT applications | 0.57 (0.28)[a] |
| *Cardiology IT* | Normalized summative index of cardiology IT applications | 0.23 (0.35)[a] |
| Visit characteristics | | |
| *Number of procedures* | Number of procedures on each admission per patient | 1.08 (1.98)[a] |
| log(*LOS*) | log-transformed length of stay | 5.45 (5.87)[a,c] |
| Patient stickiness | | |
| *Proportion of same hospital visits* | Number of times a patient visited the same hospital divided by the total number of visits up to the current admission | 0.93 (0.18)[a] |
| Payer type variables | | |
| *Medicare* | Binary indicator of claim filed to Medicare | 40,266 (61.77%)[d] |
| *Medicaid* | Binary indicator of claim filed to Medicaid | 4,912 (7.54%)[d] |
| *Private* | Binary indicator of private insurance | 16,650 (25.54%)[d] |
| *Other type of insurance* | Binary indicator of other types of insurance (Veterans Administration or other federal programs) | 686 (1.05%)[d] |
| Admission condition variables | | |
| *Admission type* (medical emergency) | Binary indicator of admission type classified as medical emergency | 44,247 (67.88%)[d] |
| *Risk mortality* | Risk mortality (1: minor (15.14%), 2: moderate (46.88%), 3: major (29.24%), 4: extreme (8.74%)) | |
| Comorbidity variables | | |
| *Diabetes_mellitus* | Binary indicator of diabetes mellitus | 28,436 (43.62)[d] |
| *Hypertension* | Binary indicator of hypertension | 27,399 (42.03)[d] |
| *Periph_vascular* | Binary indicator of periphery vascular | 6,935 (10.64)[d] |
| *Chronic_pulmonary* | Binary indicator of chronic pulmonary disease | 22,949 (35.20)[d] |
| *Renal_failure* | Binary indicator of renal failure | 22,441 (34.43)[d] |
| *Anemia* | Binary indicator of anemia | 19,214 (29.47)[d] |
| *Alcohol_abuse* | Binary indicator of alcohol abuse | 984 (1.51)[d] |
| *Drug_abuse* | Binary indicator of drug abuse | 2,237 (3.43)[d] |
| *Ischemic_disease* | Binary indicator of ischemic disease | 35,301 (54.15)[d] |
| Hospital variables | | |
| *Num_beds* | Number of beds in hospital | 392.70 (298.44)[a] |
| *Tch_hosp* | Binary indicator of teaching/nonteaching hospital (1 = teaching/0 = nonteaching) | 30,341 (46.54%)[d] |
| *CMI* | Case mix index | 1.53 (0.26)[a] |
| Admission and discharge | | |
| *Admission source* | ER reference (1 = ER reference (77%)/0 = non-ER reference) | |
| *Discharge disposition* | Discharged to home/self-care (1 = home (62%)/0 = elsewhere) | |

[a]Mean (standard deviation); [b]statistics on discharge age in years; [c]statistics on length of stay in days; [d]number of occurrences (%, percentage out of the total number of observations).

of the HIT variables across two consecutive years. Using the difference of an endogenous variable as an IV was proposed by Arellano and Bond (1991) and has become a well-accepted method to account for endogeneity. Both IVs are correlated with the current year's level of HIT of the focal hospital since (a) the hospital is likely to monitor and follow the HIT applications that its peer hospitals have implemented (due to peer pressure), resulting in correlated hospital-level IT decisions and (b) the previous year's HIT (and the derived difference) ought to be correlated with the level of HIT in the current year.

At the same time, these two IVs are unlikely to be systematically determined by an individual patient's readmission at the focal hospital. In other words, competition in the local healthcare market may drive health IT implementations such that the IT infrastructure of competing hospital systems may influence the focal hospital's decision to implement HIT. However, it is unlikely that HIT implementation decisions of peer hospitals will be associated with an individual patient's readmission rate at the focal hospital. Likewise, a hospital's current level of HIT may depend on its level in the prior year, but prior year levels (and the derived difference) predates patient outcomes in the current year and thus is unlikely to be systematically codetermined. Hence, both variables fulfill the criteria for IV estimation. We also test the strength and exogeneity of our IV in a nonlinear, two-stage, least-squares model. The $F$-value of the administrative IT, clinical IT, and cardiology IT variables, in the first stage, are 296.85, 860.79, and 578.13, respectively, confirming the strength of these IVs. Furthermore, the Hansen's $J$ test provides a test statistic of $J = 3.305$ ($p = 0.3469$), thus supporting the exogeneity of these IVs.

## 5.2. Baseline Model Results

For patient-level analysis, our objective is to develop a better understanding of the determinants of readmission propensity within 30 days of discharge from the previous admission. The dependent variable in the logit model is measured as a binary variable that takes a value of one in the presence of a 30-day readmission, and zero otherwise. The dependent variable in the proportional hazard model is measured as the time interval between two consecutive admissions that occur within a 30-day window. If a readmission occurs outside of the 30-day window, we treat the subsequent visit as a new admission, a common practice in research and practice (Joynt and Jha 2012). We present our estimation results for the logit and Cox proportional hazard models in Table 3. Based on the logit estimation results, we observe that female patients are 7.6% *less likely* to be readmitted within 30 days compared to male patients (odds ratio = 0.924); their average duration between consecutive readmissions is also 11.7% less than their male counterparts.

The logit results suggest that African American CHF patients are 42% more likely to be readmitted within 30-days of their prior discharge, as compared to their Caucasian counterparts (coeff. = 0.356, $p < 0.10$; odds ratio = 1.42). Our results also suggest that older patients do not necessarily incur higher readmission risks, as demonstrated by the insignificant value of the coefficient of log(*disch_age*). However, the negative and significant coefficient on the quadratic term, $\log(disch\_age)^2$, implies that the risk of readmission within 30 days starts to decrease for patients who are older than 63. We note that this negative effect on the quadratic term can be attributed to the possibility of data truncation (i.e., older patients being closer to the end of their lives), which is not accounted for in the baseline estimation models.[6]

The estimation results of the logit and proportional hazard models indicate that the three classes of health IT applications (i.e., administrative, clinical, and cardiology IT) are not significantly associated with the propensity of 30-day readmission. We also do not observe a significant association between health IT applications and the time duration between consecutive admissions based on the results of the proportional hazard model. We will revisit these factors again during our discussion of the BG/EG hurdle model results.

A positive coefficient (coeff. = 0.112; $p$-value < 0.01) on LOS indicates that longer hospital stays are associated with higher risk of readmission, which is also consistent with previous findings (Mudge et al. 2010). This result may be attributed to the possibility that sicker patients may require longer LOS and are more likely to be readmitted within 30 days.

With respect to the effect of payer type on readmission risk, we observe that patients with Medicaid are at a significant risk of 30-day readmission, compared to self-pay patients. For Medicaid patients, the readmission risk increases by 16.7% compared to self-payers, marginally significant at a $p$-value < 0.10.

The severity of a patient's condition, based on risk mortality scores, also plays a significant role in determining readmission risk. Patients with moderate risk levels (i.e., level 2) have a higher risk of being readmitted within 30 days, as compared to patients with low risk levels (i.e., level 1), with the risk increasing by 36.4%. Higher levels of mortality are also associated with a greater 30-day readmission risk by a degree of 63.9% and 65.2% for levels 3 and 4, respectively. We observe that the average duration between consecutive readmissions is 15.8% less for emergency room patients compared with those that are admitted as inpatients. We also observe that teaching hospitals have a higher risk of 30-day patient readmission compared to their nonteaching counterparts, with the risk of readmission being greater than 14.9%, due to the possibility that such hospitals often treat more complex cases.

## 5.3. BG/EG Hurdle Model Results

Next, we estimate the BG/EG hurdle model, which takes into account the following salient estimation issues: (a) unobserved patient-level heterogeneity, (b) state dependency of patient readmissions that result in a nonstationary readmission rate, and (c) data censoring due to truncation in the data caused by patient dropout from the sample. A unique feature of our model is that it allows estimation of two distinct components of patient readmissions that have been overlooked in the readmission literature: (a) a patient's propensity of being readmitted within 30 days of the prior admission (logit hurdle), and (b) frequency

[6] The BG/EG hurdle model addresses this issue by modeling a "drop out" probability after each visit for a patient. We further conducted a robustness check using the approach described in Gönül and Hofstede (2006) to test the potential nonlinear effect of age and found that the estimation results are consistent with our results reported here.

**Table 3 Baseline Logit and Proportional Hazard Estimation of 30-Day Readmission Model**

| | Logit | | | Proportional hazard | | |
|---|---|---|---|---|---|---|
| Variable | Parameter estimate | Standard error | Odds ratio | Parameter estimate | Standard error | Hazard ratio |
| *Intercept* | −3.585 | (0.877)*** | 0.028 | | | |
| *Gender*: female | −0.078 | (0.022)*** | 0.924 | −0.124 | (0.039)*** | 0.883 |
| *Patient race* | | | | | | |
| Asian or Pacific Islander | 0.108 | (0.261) | 1.115 | −0.133 | (0.179) | 0.876 |
| African American | 0.356 | (0.211)* | 1.428 | −0.013 | (0.052) | 0.987 |
| Other | 0.126 | (0.219) | 1.135 | −0.170 | (0.086)** | 0.843 |
| log(*disch_age*) | 0.502 | (0.426) | 1.653 | 0.181 | (0.262) | 1.198 |
| log(*disch_age*)$^2$ | −0.094 | (0.058)* | 0.911 | −0.055 | (0.039) | 0.947 |
| Health IT | | | | | | |
| *Administrative IT* | 0.038 | (0.115) | 1.039 | −0.087 | (0.10) | 0.916 |
| *Clinical IT* | 0.076 | (0.085) | 1.079 | 0.102 | (0.078) | 1.107 |
| *Cardiology IT* | −0.050 | (0.060) | 0.951 | −0.083 | (0.056) | 0.921 |
| *Number of procedures* | −0.067 | (0.013)*** | 0.935 | −0.063 | (0.012)*** | 0.939 |
| log(*LOS*) | 0.111 | (0.034)*** | 1.117 | 0.120 | (0.031)*** | 1.128 |
| Payer type | | | | | | |
| *Medicare* | 0.112 | (0.087) | 1.119 | 0.106 | (0.081) | 1.111 |
| *Medicaid* | 0.154 | (0.094)* | 1.167 | 0.128 | (0.088) | 1.136 |
| *Private* | 0.070 | (0.082) | 1.073 | 0.061 | (0.077) | 1.063 |
| *Other type of insurance* | 0.057 | (0.236) | 1.058 | 0.070 | (0.199) | 1.073 |
| *Admission type* | | | | | | |
| Medical emergency | 0.06 | (0.057) | 1.062 | 0.147 | (0.043)*** | 1.158 |
| *Risk mortality* | | | | | | |
| Level 2 | 0.310 | (0.073)*** | 1.364 | 0.294 | (0.066)*** | 1.341 |
| Level 3 | 0.494 | (0.080)*** | 1.639 | 0.482 | (0.072)*** | 1.619 |
| Level 4 | 0.502 | (0.102)*** | 1.652 | 0.646 | (0.093)*** | 1.909 |
| Comorbidities | | | | | | |
| *Diabetes_mellitus* | 0.025 | (0.043) | 1.026 | −0.0004 | (0.039) | 1 |
| *Hypertension* | −0.137 | (0.052)*** | 0.872 | −0.103 | (0.046)** | 0.902 |
| *Periph_vascular* | 0.032 | (0.065) | 1.033 | 0.078 | (0.058) | 1.081 |
| *Chronic_pulmonary* | −0.012 | (0.044) | 0.988 | −0.024 | (0.04) | 0.977 |
| *Renal_failure* | 0.089 | (0.056)* | 1.093 | 0.130 | (0.05)*** | 1.139 |
| *Anemia* | −0.005 | (0.047) | 0.995 | 0.033 | (0.042) | 1.033 |
| *Alcohol_abuse* | −0.253 | (0.18) | 0.777 | −0.116 | (0.16) | 0.89 |
| *Drug_abuse* | 0.689 | (0.115)*** | 1.992 | 0.598 | (0.108)*** | 1.819 |
| *Ischemic_disease* | 0.206 | (0.044)*** | 1.229 | 0.175 | (0.04)*** | 1.191 |
| *Num_beds* | −0.0003 | (0.0001)*** | 1 | −0.0001 | (0.0001) | 1 |
| *Tch_hosp* | 0.139 | (0.057)*** | 1.149 | 0.142 | (0.051)*** | 1.153 |
| *CMI* | −0.224 | (0.123)* | 0.8 | −0.503 | (0.098)*** | 0.605 |
| $-2\log L$ | | 17,893.18 | | | 61,254.92 | |
| *AIC* | | 17,959.18 | | | 61,318.92 | |

*Note.* Standard errors in parentheses.
$^{*}p = 0.10$; $^{**}p = 0.05$; $^{***}p = 0.01$.

of future readmissions (BG/EG) after the hurdle has been crossed. We treat the 30-day readmission window as a *hurdle* to be overcome before the latter condition (i.e., future readmission) is observed.

**5.3.1. Logit-Hurdle Analysis.** The logit-hurdle component estimates the propensity of 30-day readmission where we use the same set of independent variables as in patient-level baseline analysis. For the BG/EG estimation of the frequency of future readmissions, we include three new variables in addition to the ones used in the logit hurdle model. These variables include (a) patient stickiness, (b) destination of patient's prior discharge (i.e., if the previous discharge was to home or self-care), and (c) admission source of the patient (e.g., if the patient was admitted to the ER on the prior admission). These variables allow for accurate model identification of our BG/EG estimation model as they are only relevant to the BG/EG component and are not included in the logit hurdle estimation. Our choice of these variables is based on recent anecdotal evidence that suggests that community-level factors, such as challenges faced by patients when they are discharged to their homes, are associated with patient readmissions (e.g., Kansagara et al. 2011).

The logit-hurdle parameter estimates, as shown in the left-hand panel of Table 4, indicate that patient demographics (gender, race, and discharge age) and admission characteristics (number of procedures and LOS) are significant determinants of 30-day readmission risk. We observe that usage of cardiology IT is associated with a significant reduction in 30-day readmission risk (coeff. = −0.086; $p$-value < 0.05). CHF patients who are admitted to hospitals with a high level of cardiology IT applications are about 8.3% less likely to be readmitted within 30 days. On the other hand, higher levels of administrative and clinical IT systems are associated with a slight increase in patient readmission risk by 2.2% and 1.4%, respectively. There are several possible explanations for this result. First, it may be attributed partly to self-selection because sicker patients, who are at higher risk to be readmitted, may seek better-equipped hospitals with greater IT resources. However, since we have controlled for patient risk mortality in our model, this is unlikely to be an issue. As a robustness check, we conduct a Heckit analysis to account for possible sample selection bias arising from sicker patients. The results show that all parameter estimates remain qualitatively unchanged, ruling out possible selection bias. A similar argument may hold that hospitals with higher levels of readmissions self-select into adopting clinical and administrative IT systems. However, we have addressed this potential endogeneity with instrument variables in §5.1.

Because the HIMSS data do not explicitly distinguish between implementation and usage of various health IT applications, another possibility may be that, although hospitals have implemented administrative and clinical IT systems, their actual use of these systems may not have occurred in tandem. As Devaraj and Kohli (2003) observe, the drivers of business value is not implementation of IT per se but its actual usage within business processes. Hospitals with high levels of HIT implementation may not necessarily represent the ones with higher proportion of users of these systems. Our sample period (from 2006–2009) represents the period when hospitals started to implement clinical IT, such as CPOE and EMR systems. Implementation of administrative and clinical IT systems usually requires significant investments in training clinical staff and users, and incurs a time lag of 18 to 24 months before process improvements are realized (Menon et al. 2000). Since we do not observe actual usage of HIT applications, the negative effects of clinical IT on care outcomes may not reflect the time lag between health IT implementation and their actual usage. In this respect, we note that recent meaningful use incentives provided by the federal government to spur *usage* of electronic health records (EHRs) may indeed serve as a catalyst to increase usage of health IT to improve clinical workflows and patient care outcomes (Blumenthal and Tavenner 2010).

**5.3.2. BG/EG Analysis.** Next, we present the estimation results of the BG/EG component in the right-hand panel of Table 4, and observe that interesting patterns begin to emerge for many of the model variables compared to the results of our logit hurdle estimation.[7] We find differential effects of health IT applications on CHF patient readmission rates. For patients who cross the 30-day readmission hurdle, cardiology IT systems are associated with a 23% reduction in frequency of future readmission (coeff. = −0.25, $p < 0.01$, hazard ratio = 0.770). Similarly, our results indicate that administrative IT systems are associated with a 26.3% reduction in the frequency of future readmissions (coeff. = −0.304, $p < 0.01$, hazard ratio = 0.737). This indicates that healthcare IT systems have differential impacts on 30-day patient readmission risk and their frequency of future readmissions, necessitating the separation of the two components in our model. This differential effect reveals the underlying mechanism of the impact of EHRs on clinical information workflow. When a patient is first admitted to a hospital, there is no prior history available on the patient, thereby limiting the benefit of EHRs. However, once the patient is readmitted, administrative and cardiology systems accrue more information on the patient's prior medical history and their benefits start to emerge. Our results also indicate that clinical HIT systems are not significantly associated with the frequency of future readmissions. This may be attributed to the possibility that such systems may improve workflows for clinicians through automated reminders and computerized order entry, but may not have a direct impact on patient readmission risk. Our results suggest that administrative and cardiology IT are particularly important in ensuring that readmitted patients receive high-quality care, which translates into a lower risk and frequency of future readmissions.

An interesting finding is that repeat care from the same hospital reduces the risk of future readmissions significantly, as reflected in the negative coefficient estimate for the *patient stickiness* variable. Our BG/EG results indicate that a 1% increase in patient stickiness

[7] Because our BG/EG model is based on continuous time, where the time unit is one day, we first calculate the values of the density function for each time point and integrate it over a 30-day period to define the probability for 30 days. To calculate the marginal effect for each variable, we compute the point estimate of the density function by plugging in one coefficient at a time holding all other variables at their mean values. We then aggregate the functional estimation over 30 days for each variable.

**Table 4 BG/EG Hurdle Model Estimation Results**

| | 30-day readmission logit hurdle | | | BG/EG | | |
|---|---|---|---|---|---|---|
| Variable | Parameter estimate | Standard error | Odds ratio | Parameter estimate | Standard error | Hazard ratio |
| *Intercept* | −3.379 | (0.364)$^{***}$ | 0.034 | | | |
| *Gender*: female | −0.024 | (0.01)$^{**}$ | 0.976 | −0.094 | (0.057)$^{*}$ | 0.908 |
| *Patient race* | | | | | | |
| African American | 0.343 | (0.039)$^{***}$ | 1.409 | 0.112 | (0.011)$^{***}$ | 1.124 |
| Asian or Pacific Islander | 0.054 | (0.039) | 1.056 | −0.002 | (0.009) | 0.998 |
| Other | 0.075 | (0.018)$^{***}$ | 1.078 | −0.024 | (0.014) | 0.975 |
| log(*disch_age*) | 0.459 | (0.13)$^{***}$ | 1.582 | 1.448 | (0.23)$^{***}$ | 4.295 |
| log(*disch_age*)$^2$ | −0.061 | (0.013)$^{***}$ | 0.941 | −0.272 | (0.044)$^{***}$ | 0.641 |
| Health IT | | | | | | |
| *Administrative IT* | 0.022 | (0.013)$^{*}$ | 1.022 | −0.304 | (0.117)$^{***}$ | 0.737 |
| *Clinical IT* | 0.014 | (0.002)$^{***}$ | 1.014 | 0.006 | (0.008) | 1.006 |
| *Cardiology IT* | −0.086 | (0.034)$^{**}$ | 0.917 | −0.250 | (0.089)$^{***}$ | 0.770 |
| *Number of procedures* | −0.074 | (0.01)$^{***}$ | 0.929 | −0.011 | (0.004)$^{***}$ | 0.989 |
| log(*LOS*) | 0.120 | (0.021)$^{***}$ | 1.128 | −0.016 | (0.017) | 0.984 |
| Payer type | | | | | | |
| *Medicare* | 0.130 | (0.034)$^{***}$ | 1.139 | −0.263 | (0.035)$^{***}$ | 0.764 |
| *Medicaid* | 0.090 | (0.079) | 1.094 | −0.102 | (0.053)$^{*}$ | 0.898 |
| *Private* | 0.103 | (0.019)$^{***}$ | 1.109 | 0.259 | (0.015)$^{***}$ | 1.310 |
| *Other type of insurance* | 0.116 | (0.097) | 1.123 | −0.031 | (0.007)$^{***}$ | 0.967 |
| *Admission type* | | | | | | |
| Medical emergency | 0.081 | (0.019)$^{***}$ | 1.084 | −0.147 | (0.012)$^{***}$ | 0.861 |
| *Risk mortality* | | | | | | |
| Level 2 | 0.227 | (0.039)$^{***}$ | 1.254 | 0.060 | (0.006)$^{***}$ | 1.064 |
| Level 3 | 0.509 | (0.051)$^{***}$ | 1.663 | −0.039 | (0.017)$^{**}$ | 0.960 |
| Level 4 | 0.459 | (0.076)$^{***}$ | 1.583 | 0.005 | (0.002)$^{***}$ | 1.005 |
| Comorbidities | | | | | | |
| *Diabetes_mellitus* | 0.028 | (0.013)$^{**}$ | 1.028 | −0.098 | (0.041)$^{**}$ | 0.903 |
| *Hypertension* | −0.075 | (0.026)$^{***}$ | 0.927 | 0.144 | (0.036)$^{***}$ | 1.160 |
| *Periph_vascular* | 0.020 | (0.015) | 1.020 | −0.019 | (0.004)$^{***}$ | 0.980 |
| *Chronic_pulmonary* | 0.000 | (0.005) | 1.000 | 0.070 | (0.021)$^{***}$ | 1.075 |
| *Renal_failure* | 0.084 | (0.021)$^{***}$ | 1.088 | 0.072 | (0.028)$^{**}$ | 1.078 |
| *Anemia* | 0.039 | (0.01)$^{***}$ | 1.040 | −0.079 | (0.03)$^{***}$ | 0.921 |
| *Alcoho_abuse* | −0.283 | (0.2) | 0.754 | −0.017 | (0.012) | 0.982 |
| *Drug_abuse* | 0.709 | (0.088)$^{***}$ | 2.033 | 0.016 | (0.02) | 1.017 |
| *Ischemic_disease* | 0.179 | (0.035)$^{***}$ | 1.197 | 0.018 | (0.01)$^{*}$ | 1.019 |
| *Num_beds* | 0.009 | (0.006) | 1.009 | −0.004 | (0.003) | 0.997 |
| *Tch_hosp.* | 0.200 | (0.04)$^{***}$ | 1.221 | 0.066 | (0.032)$^{**}$ | 1.070 |
| *CMI* | −0.353 | (0.058)$^{***}$ | 0.703 | −0.154 | (0.048)$^{***}$ | 0.861 |
| *Patient stickiness* | | | | −0.075 | (0.037)$^{**}$ | 0.927 |
| *Previous discharge to home/self* | | | | 0.135 | (0.031)$^{***}$ | 1.148 |
| *Previous ER reference* | | | | −0.002 | (0.008) | 0.997 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $r$ | 2.833 | $a$ | 5.782 | $\mu$ | 0.021 | Expected daily admission rate | 2.18% |
| $\alpha$ | 129.97 | $b$ | 6.096 | $\sigma$ | 0.099 | Expected drop-out rate | 48.68% |
| $-2\log L$ | | 97,239.979 | | AIC | | 97,379.979 | |

*Note.* Standard errors in parentheses.
$^{*}p = 0.10$; $^{**}p = 0.05$; $^{***}p = 0.01$.

reduces the frequency of future readmission by 7.3% (coeff. $= -0.075$; $p$-value $< 0.05$; hazard ratio $= 0.927$). In other words, patients who are treated at the same hospital are less likely to incur future readmissions. This finding can be attributed to the possibility that a patient who is treated at the same hospital may receive a better continuum of care since doctors are likely to have access to her complete medical history and can make more informed decisions related to patient diagnosis and treatment. Another plausible interpretation is that these "sticky" patients are less severe ones and are less likely to incur readmissions in the first place. To account for this possibility, we compare the profiles of highly sticky patients with less stickier ones and do not find a statistical difference across these two samples in terms of risk mortality scores. Hence, our results imply that improving "patient stickiness" allows healthcare providers to reduce the frequency of future readmissions of CHF patients, and provide indirect evidence of the value of information integration across disparate hospitals.

Next, we observe that payer-type variables are associated with different patterns of readmission propensity and frequency of readmissions. Compared to self-pay patients, Medicare patients exhibit a higher propensity of initial 30-day readmission. However, once they cross this readmission hurdle, their frequency of future readmission decreases by 23.6% (coeff. $= -0.263$; $p$-value $< 0.01$; hazard ratio $= 0.764$). Future readmissions of Medicaid patients also decrease by 10.2% (coeff. $= -0.102$; $p$-value $< 0.10$; hazard ratio $= 0.898$). On the other hand, patients with private insurance exhibit a 31.0% higher risk of future readmissions than self-pay patients. In other words, after the first readmission, Medicare/Medicaid patients are likely to incur less frequent hospital admissions relative to self-pay patients, whereas patients with private insurance are likely to be readmitted more frequently. We also observe that ER patients exhibit the same readmission characteristics as Medicare and Medicaid patients.

Our results imply that providers who treat Medicare and Medicaid patients would be well-served to reduce future readmissions, in light of the proposed penalties for not meeting federal readmission requirements (Joynt and Jha 2012). However, for self-pay or private insurance patients, there are no such stipulations and the quality of preventive medical care that they receive is not likely to be on par with that of Medicare patients (Ong et al. 2009). Hospitals are more likely to provide high-quality, preventive care to Medicare and Medicaid patients once they are readmitted by ensuring that they receive extraordinary interventions to reduce future readmissions. For example, a recent partnership between Parkland Hospital in Dallas and the Texas Health Resources system has resulted in the development of a risk stratification model to identify high-risk CHF patients and provide them with a dedicated heart failure team, nurse practitioner, and pharmacist, to reduce the likelihood of future readmissions (Hagland 2011).

We also observe that patients who are discharged to their homes or self-care facilities exhibit a greater frequency of future readmissions (coeff. $= 0.135$; $p < 0.01$; odds ratio $= 1.148$). This result suggests that alternative discharge locations, such as intermediate care or skilled nursing facilities, should be explored for at-risk CHF patients who may otherwise receive inadequate postdischarge care.

### 5.4. Comparison of Predictive Analytics Performance

Although we have, thus far, focused on explanatory results of the BG/EG hurdle model, we now turn our attention to the predictive capabilities of our proposed model and contrast these results against existing models in the literature. In a recent paper, Shmueli and Koppius (2011, p. 553) advocated the importance of predictive analytics and questioned the (almost exclusive) practice of explanatory statistical modeling in the IS research. They argue that

> . . . Despite the importance of predictive analytics, we find that they are rare in the empirical IS literature. Extant IS literature relies nearly exclusively on explanatory statistical modeling . . . . However, explanatory power does not imply predictive power and thus predictive analytics are necessary for assessing predictive power and for building empirical models that predict well . . . .

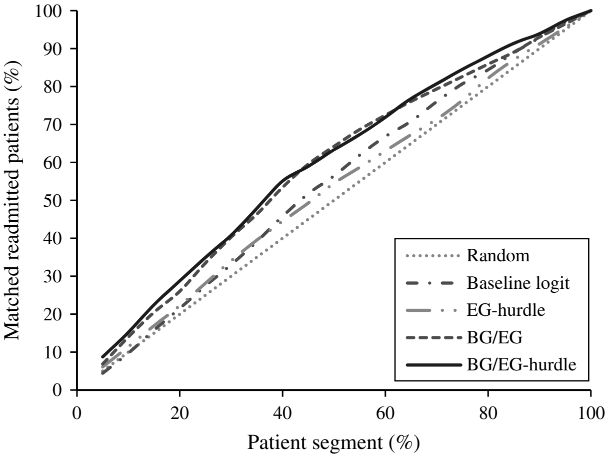
Fader et al. (2005) and Morrice and Bardhan (1995) also call for more attention to employ prediction models as the yardstick for researchers to use when judging model performance.

Our review of the readmission literature reveals the lack of attention on predictive analytics. Whereas the primary focus has been limited to models to estimate patient readmission risk, our BG/EG hurdle model serves as a predictive model that is capable of predicting the propensity and frequency of future readmissions on any given patient. We now evaluate our model using a "horse race" to compare its predictive performance against other benchmark models, which consist of the random estimation, baseline logit model, BG/NBD hurdle model, EG hurdle model, and the BG/EG model without hurdle. The logit model represents the baseline model commonly used in the healthcare literature to model readmissions. The BG/NBD hurdle model (Fader et al. 2005) represents the case when nonstationary readmission patterns are not accounted for, and it only differs from our model by replacing our EG component with the stationary NBD process. The EG hurdle model (Gupta 1991) depicts the case where the drop out component (i.e., the BG process) is not accounted for; and the "BG/EG without hurdle" model is similar to our proposed model without the hurdle component. Collectively, we design this horse race to demonstrate the relative performance of our model compared to the current state-of-the-art, alternative models that do not fully address the complexity of the patient readmission process.

We use two years of data to calibrate the training model and then test its predictive performance by using the next one year as the holdout period. Hence, our training set consists of admission records of CHF patients from January 2007 through December 2008, and the testing set includes one year of admission data from January to December 2009.[8] In the holdout

[8] We drop the first year's (i.e., 2006) data from our training model to alleviate the potential left censoring problem. Initiating the training

**Figure 2 Lift Curve of the Predictive Performance of Different Models**



sample, 2,348 patients were readmitted out of 19,408 patients. This evaluation scheme is consistent with Fader et al. (2005) conditional expectation approach. For the benchmarking models and our BG/EG hurdle model, the last observed visit for each patient in the training set is used as the "snapshot" to build the training model, which is then applied to the test data to predict readmission occurrences of each patient during the holdout period. We measure our model's predictive accuracy against actual readmission data observed during the last year of our sample period.

Figure 2 represents a *lift curve*, which describes the overall lift in predictive performance provided by the six types of estimation methodologies in our horse race experiment. We derive the probability of readmission for each patient for all models considered in our experiment. Figure 3 demonstrates that our BG/EG hurdle model outperforms all other benchmark models as a whole, followed by the BG/EG without hurdle, BG/NBD hurdle, logit, EG hurdle, and the random estimation model, in that order. For example, compared to the logit model, the lift improvement in our BG/EG hurdle model is 27.29% if we focus on the top 25% of readmitted patients. We provide additional details of the lift table for specific patient segments in Appendix C of the online supplement. Based on our sample data of 587 readmitted, high-risk patients, the BG/EG hurdle model accurately profiles 160 more patients than the baseline logit model.

Furthermore, we adopt another criterion to evaluate the predictive power of the BG/EG hurdle model. This criterion focuses on the accuracy of our model's prediction of the frequency of future readmissions. Prediction of readmission frequency has important managerial implications because hospital managers can utilize the results to anticipate future demand and allocate hospital resources accordingly. We evaluate our model by first predicting the timing of readmissions for each patient during the holdout period. Then, by aggregating the occurrence of readmissions for each month across all patients who are identified as candidates for readmission, we obtain an expectation of the total number of monthly readmissions. We then compare this expectation to the number of actual monthly readmissions.

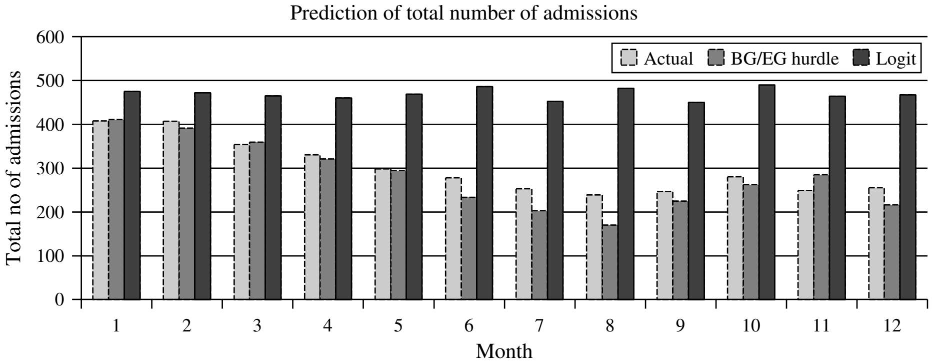
Since the baseline logit model only deals with a binary dependent variable, we use the estimated logit probability along with a simple OLS regression model to predict the frequency of future readmissions. First, for each patient, we calculate the projected number of total admissions based on OLS estimation. At the same time, an estimated logit probability is derived for each patient. Next, we estimate the projected number of admissions for patients with estimated odds greater than 0.5 for each month. Figure 3 reports the actual readmissions versus the predicted readmission frequency based on the BG/EG hurdle and baseline logit estimation for each month of the one-year holdout period. Overall, we observe that the BG/EG hurdle model provides a fairly accurate prediction, where the average difference between the predicted and actual values is 10%. On the other hand, the average difference between the predicted and actual values for the baseline model is significantly worse at 62%. The BG/EG hurdle model is able to outperform the baseline models by accurately identifying and predicting patient readmission patterns. It is worth noting that our hurdle model also achieves a higher accuracy in terms of predicting the 30-day readmission rate, which is within 4.7% of the actual number of readmissions.

### 5.5. Robustness Checks

A common alternative to our approach is to create an out of sample by randomly selecting a subset of patients from our data across all years, and then predicting the readmission patterns for this out of sample given the in sample representing other patients' readmission history (Shmueli and Koppius 2011). We generate such an out of sample by randomly selecting 50% of patients as in sample, and treating the rest as the out of sample. We recalibrate our model based on the in sample, and use this model to predict readmissions for patients in the out of sample. For each out-of-sample patient, we use the BG/EG hurdle model to predict the propensity of a 30-day readmission.

model in 2007 allows us to at least have a one-year lead time to ensure that we measure our dependent variable as accurately as possible. We acknowledge that if the index admission occurred before 2006, it is possible (though with small probability) that we may mislabel such readmissions. However, for 30-day readmissions, having a whole year of data as a buffer ensures that our labels are accurate.

**Figure 3 Predicted versus Actual Number of Patient Admissions**



The BG/EG hurdle model exhibits an overall out-of-sample accuracy of 59%, whereas the overall accuracy of the logit model is 56%. One concern regarding the prediction accuracy of a model among high-risk patient segments is that it may be achieved at the cost of over-predicting readmitted cases. Therefore, we also account for type-II errors, which represent the probabilities of false classification of patients as readmission cases. Our randomly selected out-of-sample prediction shows that the type-II error for the BG/EG hurdle model is lower than the baseline logit models. The *C*-statistic, the standard measure that accounts for types I and II trade-offs, of the BG/EG hurdle model is 0.601, whereas the corresponding *C*-statistic of the logit model is 0.563.

We also check the robustness of our results by analyzing whether our estimations that are based on multihospital data still hold for a single hospital sample. Our analysis of patient data from a large, teaching hospital shows that the traditional, single-hospital readmission rate, when readmissions to other hospitals are not accounted for, is 30.55%. On the other hand, if we also account for patients who are readmitted to other hospitals in the region, the readmission rate climbs to 40.45%, which indicates that a single-hospital view of readmission is erroneous and underestimates the true risk of readmission. For example, although diabetes, renal failure, drug abuse, and ischemic disease are significant factors that increase the risk of CHF readmission in our analysis, a single-hospital view only identifies drug abuse as a readmission risk factor. Although the actual number of CHF admissions to this hospital is 12.3 cases per month, our model predicts an average of 13.2 CHF-related visits per month, whereas the logistic model predicts an average of 27.2 visits per month. This is an average difference across months of 33.3% and 172% for the BG/EG and logistic models, respectively, confirming that the BG/EG model outperforms logit models even based on data obtained from a single hospital.

We further estimate the all-period readmission results, where we study readmissions over the entire four-year period (instead of 30 days), and report these results in Appendix D in the online appendix. These results are qualitatively similar to the main results reported in the paper.

## 6. Conclusions

Understanding the characteristics of patient readmission patterns allows hospitals to develop better predictive capabilities in order to identify and profile patients who pose greater readmission risk. Predicting the propensity of readmission for a CHF patient enables hospitals to identify and deliver appropriate treatment to the right patients and provide more efficient postacute care, which significantly reduces subsequent readmissions. In this study, we examine the association between patient and health IT characteristics and the risk propensity of future readmissions for patients with CHF. By incorporating patient history of readmissions across multiple hospitals, we develop a predictive BG/EG hurdle model that accounts for unobserved patient heterogeneity, nonstationary admission rates, time-varying risk factors, and data censoring. Furthermore, we estimate the specific effects related to the *propensity* as well as the *frequency* of future readmissions. Our proposed model represents a significant methodological improvement over extant models of readmission risk, and delivers superior predictive performance compared to traditional models.

By developing a greater understanding of patient readmission behavior, a hospital can better profile patients who are at higher risk of readmission and implement preventive measures to target these patients effectively. The scope of previous academic research on patient readmissions has been severely limited because of the lack of information sharing across hospitals that can be attributed to the absence of a common master patient index. In this study, we identify the risk factors associated with patient readmissions across multiple hospitals over a four-year period based on a unique data set obtained through electronic integration of patient data across 67 hospitals in a large geographical region.

### 6.1. Discussion

There are several deficiencies in the manner in which the extant healthcare literature treats the readmission problem. Admittedly, monitoring the 30-day hospital readmission rate is one of the key barometers of the current healthcare reform plan. However, overly focusing on short term (e.g., 30-day) readmission targets alone may lead to myopic actions, as evident from the hurdle estimation component of the BG/EG hurdle model. We find that managerial insights obtained from the logit hurdle component and the BG/EG component are drastically different. For instance, we observe that health IT systems are associated with a reduction in the frequency of future readmissions, once the 30-day readmission hurdle has been crossed. Similarly, we find that Medicare patients exhibit a lower frequency of future readmissions, once they cross the 30-day initial readmission hurdle. This paper sends a forward-looking message to policy makers that thinking beyond 30-day readmissions may be necessary.

We observe that health IT applications represent important tools in reducing avoidable inpatient readmissions. For example, Parkland Hospital in Dallas captures clinical, social, and demographic characteristics of patient data in EMR systems, which is used to calculate a risk score for each heart failure patient on admission (Hagland 2011). A high-risk score triggers an alert to a heart failure "SWAT team" for follow-up care. Our results suggest that the use of cardiology and administrative IT applications help to reduce patient readmission risk and lay the foundation for more effective treatment and care delivery. Our empirical results support recent recommendations made by the Hospital Readmission Workgroup, which advocates use of health IT tools, such as case management systems, predictive analytics, and social media, as enablers to offer patients better postdischarge care and reduce the incidence of hospital readmissions (HIMSS 2012).

### 6.2. Implications

Prediction of the timing and frequency of a patient's future readmissions is a unique component of our model, since it enables managers to make better decisions related to hospital capacity planning. When aggregated across thousands of patient readmissions in a given year, even relatively small improvements in predictive modeling of readmission risk and frequency can substantially improve the quality and cost of healthcare delivery. For example, in 2004, the average hospitalization cost for a CHF case was estimated to be $9,400 (Russo et al. 2007). Hence, our model's superior prediction capabilities can potentially provide average annual savings of up to $1,504,000 for these high-risk patients, if our predictions were to result in successful readmission avoidance when we correctly predict their readmission propensity and apply preventive care in advance (e.g., deployment of a dedicated cardiology team and support processes).[9]

Furthermore, disentangling the estimation of the frequency from readmission propensity provides us with a more accurate and nuanced understanding of patient readmission patterns. From a patient perspective, we find that repeat care delivery at the same hospital reduces the risk of future readmissions significantly. This indicates that a patient treated at the same hospital (across multiple visits) tends to receive better quality of care, which reduces their risk of being readmitted for the same diagnosis in the future.

Predictive analytics is central to the Medicare Hospital Readmissions Reduction Program (HRRP), as established by the Affordable Care Act (ACA) to improve the quality of healthcare. To reduce preventable readmissions, the ACA imposes a financial penalty on hospitals with above-average Medicare readmissions, where the penalties are collected from hospitals through a reduction in their base Medicare inpatient claims payments, up to a cap set at 3% each year (James 2013). For instance, one of the hospitals in our sample would have faced a significant penalty if this regulation had been enforced in 2009. Its CHF inpatient-related charges were $9 million, of which $5.7 million was charged to Medicare patients. Under the HRRP, the financial penalty for this hospital would be about $171,000. If the hospital could successfully implement healthcare IT, it could potentially reduce patient readmission rates by 4%, resulting in a reduction of $360,000 in readmission costs as well as avoiding $171,000 in financial penalties due to ACA regulations, for a total cost reduction of $531,000.

### 6.3. Limitations and Future Research

Nevertheless, our study does have a few limitations. Our model is limited to studying CHF patients within one geographic region. Although the North Texas region is fairly diverse in terms of its population, future studies are needed to expand the scope of our models to account for patient demographic characteristics in other regions of the country. Our study is restricted to patients whose primary diagnosis is CHF. Future studies will extend these models to other chronic diseases. Our measures of hospital IT usage

[9] The estimated cost savings are based on the top 25% highest risk patients from the lift curve, which first ranks patients from the highest readmission probability to the lowest, and then counts the rate of true positives. The predictive performance from the lift curve shows that the BG/EG hurdle model accurately profiles 160 more patients than the baseline logit model. Therefore, the potential cost savings of accurately predicting 160 patients is equal to 160× $9,400 per patient, or $1.5 million.

are based on the HIMSS data, which provides information on overall hospital-wide health IT applications instead of their usage for treatment of specific patients. Future studies will be designed to investigate the impact of the usage of different types of health IT for treatment of specific patient and disease clusters. We acknowledge that our observed relationships between health IT and patient readmission risk are associational in nature, although we have accounted for potential endogeneity in terms of readmission risk.

## Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/isre.2014.0553.

## Acknowledgments

The authors gratefully acknowledge the feedback received from the senior editor, associate editor, and the review team. The authors thank seminar participants at the University of Washington, University of Arizona, University of Texas at Austin, University of Texas at Dallas, University at Buffalo, Georgia State University; as well as attendees of the 32nd International Conference on Information Systems, Shanghai, and the 2011 INFORMS Conference on Information Systems and Technology, Charlotte; and Dr. Gary Reed at the University of Texas Southwestern Medical Center for their helpful feedback. The authors also gratefully acknowledge the Dallas Fort Worth Hospital Council Research Foundation and the Healthcare Information and Management Systems Society for their help in providing access to the research data. Financial support from UT Dallas and research grants awarded by the UT Southwestern Medical Center (Contract 34456001 and 34456002) are gratefully acknowledged.

## References

Agarwal R, Gao G, DesRoches C, Jha AK (2010) Research commentary—The digital transformation of healthcare: Current status and the road ahead. *Inform. Systems Res.* 21(4):796–809.

Alexander M, Grumbach K, Remy L, Rowell R, Massie BM (1999) Congestive heart failure hospitalizations and survival in California: Patterns according to race/ethnicity. *Amer. Heart J.* 137(5):919–927.

Amarasingham R, Plantinga L, Diener-West M, Gaskin DJ, Powe NR (2009) Clinical information technologies and inpatient outcomes: A multiple hospital study. *Arch. Intern. Med.* 169(2):108–114.

Amarasingham R, Moore BJ, Tabak YP, Drazner MH, Clark CA, Zhang S, Reed WG, Swanson TS, Ma Y, Halm EA (2010) An automated model to identify heart failure patients at risk for 30-day readmission or death using electronic medical record data. *Medical Care* 48(11):981–988.

Anderson CL, Agarwal R (2011) The digitization of healthcare. *Inform. Systems Res.* 22(3):469–490.

Angst CM, Agarwal R, Sambamurthy V, Kelley K (2010) Social contagion and information technology diffusion: The adoption of electronic medical records in U.S. hospitals. *Management Sci.* 56(8):1219–1241.

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econom. Stud.* 58(2):277–297.

Aron R, Dutta S, Janakiraman R, Pathak PA (2011) Impact of automation of systems on medical errors. *Inform. Systems Res.* 22(3):429–446.

Bardhan I, Thouin M (2013) Health information technology and its impact on the quality and cost of healthcare delivery. *Decision Support Systems* 55(2):438–449.

Bhattacherjee A, Hikmet N, Menachemi N, Kayhan VO, Brooks RG (2007) The differential performance effects of healthcare information technology adoption. *Inform. Systems Management* 24(1):5–14.

Blumenthal D, Tavenner M (2010) The "meaningful use" regulation for electronic healthcare records. *New England J. Medicine* 363(6):501–504.

Buntin MB, Burke MF, Hoaglin MC, Blumenthal D (2011) The benefits of health information technology: A review of the recent literature shows predominantly positive results. *Health Affairs* 30(3):464–471.

Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH (2006) Types of unintended consequences related to computerized provider order entry. *J. Amer. Med. Informatics Assoc.* 13(5):547–556.

Cebul RD, Love TE, Jain AK, Hebert CJ (2011) Electronic health records and quality of diabetes care. *New England J. Medicine* 365(9):825–833.

Chatfield C, Goodhardt GJ (1973) A consumer purchasing model with Erlang interpurchase time. *J. Amer. Statist. Assoc.* 68(344):828–835.

Chin M, Goldman M (1997) Correlates of early hospital readmission or death in patients with congestive heart failure. *Amer. J. Cardiol.* 79(12):1640–1644.

Congressional Budget Office (2008) Evidence on the Costs and Benefits of Health Information Technology. Washington, DC. www.cbo.gov/publication/41690.

Cox DR (1972) Regression models and life-tables. *J. Royal Statist. Soc. Ser. B (Methodological)* 34(2):187–220.

Das S, Yaylacicegi U, Menon NM (2011) The effect of information technology investments in healthcare: A longitudinal study of its lag, duration, and economic value. *IEEE Trans. Engrg. Management* 58(1):124–140.

Deb P, Trivedi PK (2002) The structure of demand for health care: Latent class versus two-part models. *J. Health Econom.* 21(4):601–625.

DesRoches CM, Campbell EG, Vogeli C, Zheng J, Rao SR, Shields AE, Donelan K, Rosenbaum S, Bristol SJ, Jha AK (2010) Electronic health records' limited successes suggest more targeted uses. *Health Affairs* 29(4):639–646.

Deswal A, Petersen NJ, Souchek J, Ashton CM, Wray NP (2004) Impact of race on health care utilization and outcomes in veterans with congestive heart failure. *J. Amer. Coll. Cardiol.* 43(5):778–784.

Devaraj S, Kohli R (2000) Information technology payoff in the health-care industry: A longitudinal study. *J. Management Inform. Systems* 16(4):41–67.

Devaraj S, Kohli R (2003) Performance impacts of information technology: Is actual usage the missing link? *Management Sci.* 49(3):273–289.

Elixhauser A, Steiner C, Harris DR, Coffey RM (1998) Comorbidity measures for use with administrative data. *Medical Care* 36(1):8–27.

Fader PS, Hardie BGS, Huang C-Y (2004) A dynamic changepoint model for new product sales forecasting. *Marketing Sci.* 23(1):50–65.

Fader PS, Hardie BGS, Lee KL (2005) "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Sci.* 24(2):275–284.

Felker GM, Leimberger JD, Califf RM, Cuffe MS, Massie BM, Adams KF, Gheorghiade M, O'Connor CM (2004) Risk stratification after hospitalization for decompensated heart failure. *J. Cardiac Failure* 10(6):460–466.

Gao G, McCullough J, Agarwal R, Jha A (2010) A study of online physician ratings by patients. Working paper, R. H. Smith School of Business, University of Maryland, College Park.

Gönül FF, Hofstede FT (2006) How to compute optimal catalog mailing decisions. *Marketing Sci.* 25(1):65–74.

Gupta S (1991) Stochastic models of interpurchase time with time-dependent covariates. *J. Mktg. Res.* 28(1):1–15.

Hagland M (2011) Mastering readmissions: Laying the foundation for change. *Healthcare Informatics* 28(4):10–16.

Han YY, Carcillo JA, Venkataraman ST, Clark RSB, Watson RS, Nguyen TC, Bayir H, Orr RA (2005) Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 116(6): 1506–1512.

Heckman JJ (1991) Identifying the hand of past: Distinguishing state dependence from heterogeneity. *Amer. Econom. Rev.* 81(2):75–79.

Hillestad R, Bigelow RJ, Bower A, Girosi F, Meili R, Scoville R, Taylor R (2005) Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs* 24(5):1103–1117.

Himmelstein DU, Wright A, Woolhandler S (2010) Hospital computing and the costs and quality of care: A national study. *Amer. J. Medicine* 123(1):40–46.

HIMSS (2012) Reducing readmissions: Top ways information technology can help. The Hospital Readmission Workgroup, Management Engineering-Process Improvement Committee, Chicago. www.himss.org/ResourceLibrary/ResourceDetail.aspx?ItemNumber=10534.

Jain DC, Vilcassim NJ (1991) Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Sci.* 10(1):1–23.

James J (2013) Health policy brief: Medicare hospital readmission reduction program. *Health Affairs* (November 12).

Jeuland AP, Bass FM, Wright GP (1980) A multibrand stochastic model compounding heterogeneous Erlang timing and multinomial choice processes. *Oper. Res.* 28(2):255–277.

Joynt KE, Jha AK (2012) Thirty-day readmissions—truth and consequences. *New England J. Medicine* 366(15):1366–1369.

Joynt KE, Orav EJ, Jha AK (2011) Thirty-day readmission rates for Medicare beneficiaries by race and site of care. *J. Amer. Med. Assoc.* 305(7):675–681.

Kansagara DEH, Englander H, Salanitro A, Kagen D, Theobald C, Freeman M, Kripalani S (2011) Risk prediction models for hospital readmission: A systematic review. *J. Amer. Medical Assoc.* 306(15):1688–1698.

Kaushal RSK, Shojania KG, Bates DW (2003) Effects of computerized physician order entry and clinical decision support systems on medication safety: A systematic review. *Arch. Intern. Med.* 163(12):1409–1416.

Krumholz HM, Chen Y-T, Wang Y, Vaccarino V, Radford MJ, Horwitz RI (2000) Predictors of readmission among elderly survivors of admission with heart failure. *Amer. Heart J.* 139(1): 72–77.

Linder JA, Rigotti NA, Schneider LI, Kelley JHK, Brawarsky PP, Haas JS (2009) An electronic health record–based intervention to improve tobacco treatment in primary care: A cluster-randomized controlled trial. *Arch. Intern. Med.* 169(8):781–787.

McCullough JS, Casey M, Moscovice I, Prasad S (2010) The effect of health information technology on quality in U.S. hospitals. *Health Affairs* 29(4):647–654.

Menachemi N, Chukmaitov A, Saunders C, Brooks RG (2008) Hospital quality of care: Does information technology matter? The relationship between information technology adoption and quality of care. *Health Care Manage. Rev.* 33(1):51–59.

Menon NM, Lee B, Eldenburg L (2000) Productivity of information systems in the healthcare industry. *Inform. Systems. Res.* 11(1):83–92.

Miller AR, Tucker CE (2011) Can health care information technology save babies? *J. Political Econom.* 119(2):289–324.

Morrice DJ, Bardhan IR (1995) A weighted least squares approach to computer simulation factor screening. *Oper. Res.* 43(5):792–806.

Morrison DG, Schmittlein DC (1981) Predicting future random events based on past performance. *Management Sci.* 27(9): 1006–1023.

Mudge AM, Kasper K, Clair A, Redfern H, Bell JJ, Barras MA, Dip G, Pachana NA (2010) Recurrent readmissions in medical patients: A prospective study. *J. Hosp. Med.* 6(2):61–67.

Muus K, Knudson A, Klug M, Gokun J, Sarrazin M (2010) Effect of postdischarge follow-up care on readmissions among U.S. veterans with congestive heart failure: A rural-urban comparison. *Internat. J. Rural Remote Health Res.* 10(2):1447.

Nasir K, Lin Z, Bueno H, Normand S-LT, Drye EE, Keenan PS, Krumholz HM (2010) Is same-hospital readmission rate a good surrogate for all-hospital readmission rate? *Medical Care* 48(5):477–481.

Ong M, Mangione CM, Romano PS, Zhou Q, Auerbach AD, Chun A, Davidson B, et al. (2009) Looking forward, looking back. *Circulation: Cardiovascular Quality Outcomes* 2(6):548–557.

Philbin EF, DiSalvo TG (1999) Prediction of hospital readmission for heart failure: Development of a simple risk score based on administrative data. *J. Amer. Coll. Cardiol.* 33(6):1560–1566.

Philbin EF, Dec GW, Jenkins PL, DiSalvo TG (2001) Socioeconomic status as an independent risk factor for hospital readmission for heart failure. *Amer. J. Cardiol.* 87(12):1367–1371.

Pratt L (2010) N.J. teaching hospital parlays cardiology and IT partnership into new business model. *Health Imaging IT* 8(6):1–4.

PricewaterhouseCoopers (2010) The price of excess: Identifying waste in healthcare spending. PricewaterhouseCoopers Health Research Institute. www.pwc.com/us/en/healthcare/publications/the-price-of-excess.jhtml.

Ross JS, Mulvey GK, Stauffer B, Patlolla V, Bernheim SM, Keenan PS, Krumholz HM (2008) Statistical models and patient predictors of readmission for heart failure. *Arch. Intern. Med.* 168(13): 1371–1386.

Russo CA, Ho K, Elixhauser A (2007) Hospital stays for circulatory diseases, 2004. *HCUP Statistical Brief #26* (Agency for Healthcare Research and Quality, Rockville, MD).

Schweidel DA, Knox G (2013) Incorporating direct marketing activity into latent attrition models. *Marketing Sci.* 32(3):471–487.

Seetharaman PB, Chintagunta PK (2003) The proportional hazard model for purchase timing: A comparison of alternative specifications. *J. Bus. Econom. Statist.* 21(3):368–382.

Shelton P, Sager M, Schraeder C (2000) The community assessment risk screen (CARS): Identifying elderly persons at risk for hospitalization or emergency department visit. *Amer. J. Managed Care* 6(8):925–933.

Shmueli G, Koppius OR (2011) Predictive analytics in information systems research. *MIS Quart.* 35(3):553–572.

Silverstein MD, Qin H, Mercer SQ, Fong J, Haydar Z (2008) Risk factors for 30-day hospital readmission in patients ≥ 65 years of age. *Proc. Baylor U. Medical Center* 21(4):363–372.

Winkelmann R (2004) Health care reform and the number of doctor visits—An econometric analysis. *J. Appl. Econom.* 19(4):455–472.

Winkelmann R (2006) Reforming health care: Evidence from quantile regressions for counts. *J. Health Econom.* 25(1):131–145.

Winkelmann R (2010) *Econometric Analysis of Count Data* (Springer, Berlin).

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. (MIT Press, Cambridge, MA).

Zheng K, Padman R, Johnson MP, Diamond HS (2005) Understanding technology adoption in clinical care: Clinician adoption behavior of a point-of-care reminder system. *Int. J. Medical Informatics* 74(7–8):535–543.