A PROFILING MODEL FOR READMISSION OF PATIENTS WITH CONGESTIVE HEART FAILURE: A MULTI-HOSPITAL STUDY

Indranil Bardhan
The University of Texas at Dallas
Richardson, TX, U.S.A.
bardhan@utdallas.edu

Jeong-ha (Cath) Oh
The University of Texas at Dallas
Richardson, TX, U.S.A.
jhoh@utdallas.edu

Zhiqiang (Eric) Zheng
The University of Texas at Dallas
Richardson, TX, U.S.A.
ericz@utdallas.edu

Kirk Kirksey
The University of Texas Southwestern Medical Center
Dallas, TX, U.S.A.
kirk.kirksey@utsouthwestern.edu

Abstract

Mitigating preventable readmissions, where patients are readmitted for the same primary diagnosis within thirty days, is a significant challenge in delivery of high quality healthcare. Towards this end, it is imperative to understand the cause, risk propensity and timing associated with patient readmissions. We develop a patient profiling model that can predict the propensity of readmission for a patient as well as the timing of future readmissions. We develop a new model termed as BG/EG Hurdle model that can simultaneously estimate both the propensity and timing of patient readmissions. We test this model using a unique dataset that tracks both patient demographic and clinical data for individual patients across 72 hospitals in North Texas. The results indicate that patient profiles derived from our model can serve as the building block for a clinical decision support system to identify patients with high readmission risk.

Introduction

Readmission of patients with chronic diseases is a significant and growing problem in the USA, 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 recent study by PricewaterhouseCoopers (2010). Experts believe that high readmission rates, when patients return within 30 days of discharge, indicate that the nation's hospitals aren't adequately addressing patients' health issues or are discharging them prematurely. To tackle this problem, starting in 2012, the Department of Health and Human Services (HHS) will publish each hospital's readmission track record. In 2012, Medicare will stop paying hospitals for preventable readmissions tied to health conditions such as heart failure or pneumonia. In 2014, HHS will expand this policy to cover four additional health conditions.

Patients with chronic diseases, such as diabetes, asthma, and heart failure, are especially susceptible to readmission due to several reasons such as lack of access to high quality care, misdiagnosis, and/or lack of understanding of follow-up or post-operative care guidelines. Prior research on readmissions that are associated with congestive heart failure (CHF) has typically been based on small samples, in the range of a few hundred patients. These studies are typically based on samples obtained from one hospital (or a single hospital chain), overlooking the possibility of patient admissions across multiple (disparate) hospitals. Other studies have focused on niche samples of patients (e.g. among older age groups) and have been

limited by the lack of adequate data to track clinical and provider characteristics that may be important determinants of readmission rates.

Considering the increasing impetus to reduce readmission rate in nation's hospitals, it is important to conduct a comprehensive study based on a large, longitudinal panel of patients across multiple hospitals to evaluate the determinants of readmissions and the readmission risk of patients with chronic diseases. We obtained a unique dataset, provided by the Dallas Fort Worth Hospital Council (DFWHC) Foundation, that tracks a large panel of patient readmissions from January 2006 to June 2010 across 72 hospitals in North Texas. The data was first gleaned from hospitals' electronic medical record (EMR) systems and then was syndicated by DFWHC across all 72 hospitals through a unique master patient index. We note that generating such a dataset would not have been possible without health information technology (HIT) integration and only very recently has such a cross-hospital dataset become available for healthcare research. Our research focuses on developing a novel readmission profiling model to provide a better understanding of clinical and patient characteristics that mostly impact patient readmission rates. In particular, we focus on patient demographic and clinical variables for patients diagnosed with CHF since this is one of the first two health conditions that the HHS policy will cover in 2012.

We seek to model both the propensity and timing of patient readmissions. That is, for a given patient, we are interested in knowing how likely she will result in a future readmission and when the readmission will likely to occur if so. Towards this end, we build an integrated model that simultaneously estimates both the propensity and timing of readmissions. This stands in contrast with the large body of readmission literature that mostly focuses on one or the other, but seldom both (Chin and Goldman 1997, Philbin and DiSalvo 1999, Krumholz et al 2000, Silverstein et al 2008). Our proposed model, termed as the *BG/EG Hurdle model*, builds on the extensive customer relationship management (CRM) and business intelligence (BI) literature that models customers' choice, timing decisions and the associated life-time value (Schmittlein et al 1987, Morrison and Schmittlein 1988, Gupta 1991, Seetharaman and Chintagunta 2003, Fader et al 2005, Reinartz and Kumar 2003). Similar to the central role of customer models and customer profiles playing in CRM and BI systems, our patient readmission model can serve as an integral component in any healthcare BI system by providing a better patient profile, especially in terms of patients' readmission patterns.

Literature Review

The information systems research literature has witnessed growing interest related to the impact of health information systems on institutional healthcare performance. Most of the prior studies focus on hospital level performance and the impact of IT on hospitals (Devaraj and Kohli 2000, 2003; Menon, Lee, and Eldenburg, 2000), or on the hospital-level adoption and diffusion of HIT (Angst et al 2010; Agarwal et al 2010). However, there is a growing focus on patient-level analysis as researchers and clinicians have come to recognize that it is important to measure the impact of HIT on patient-level outcomes, i.e. focusing on the patient as the unit of analysis (Wilson and Lankton 2004, Gao et al 2010).

The scientific literature has investigated patient re-admissions along various dimensions, such as understanding the risk factors of readmission, identification of patient characteristics associated with readmission, and the estimation of treatment effects etc. Extant readmission studies typically are based on a single hospital using relatively small samples (on the scale of hundreds), or are restricted to a specific cohort such as elderly patients (e.g. Shelton et al 2000, Silverstein et al 2008), veterans (e.g. Cheng et al 2001, Deswal et al 2004, Muus et al 2010), and a specific racial and income group (e.g. Philbin et al 2001). A few studies have used data from several hospitals (e.g. Philbin et al 2001) but usually these hospitals belong to the same hospital chain (Felker et al 2003, Deswal et al 2004, Silverstein et al 2008). For example, Silverstein et al (2008) obtained data from seven hospitals, all of which belong to one hospital chain, i.e. the Baylor Health Care System. This can lead to serious undercounting of patient readmissions because it is not uncommon for a patient to consult with different hospitals. Hence, it is important to obtain data that truly captures the complete picture of patient admission patterns across multiple hospitals within a geographic region.

Understanding the risk factors associated with readmission has been the focus of the prior literature. However, the large body of literature on CHF yields no consensus on what factors might affect the risk of patients' re-admission. Across various prior studies, a set of socio-demographic and comorbidity

covariates are considered as likely candidates for multivariate models to predict patient readmission risk after a heart failure. These socio-demographic covariates includes age, gender, race, incomes, and types of insurance, while comorbidity covariates associated with heart failure includes diabetes mellitus, hypertension, ischemic heart disease, peripheral vascular disease, anemia, and renal failures. However, the estimation results of these variables vary. As pointed out by Ross et al (2008) there is lack of consensus with respect to the statistical models to accurately profile readmission rates.

Furthermore, we observe several limitations of the models commonly used in prior studies. First, most existing models (as exemplified by the popular proportional hazard model) assume a stationary visit rate, i.e., each patient's re-admission rate is independent of the history of admission. This does not bode well for CHF patients whose visit patterns clearly depend on factors such as the severity of the disease, health status of the patient, treatment received in the last visit etc. This phenomenon is known as 'change of pace' behavior in the literature (Fader and Lattin 1993).

Second, patients can differ significantly with regard to a set of unobserved factors such as the cause of admissions, their underlying health status and reaction to treatments etc. Such cases, which result in unobserved heterogeneity in patient readmission patterns, have seldom been taken into account in prior research. Third, existing models focus only on the propensity of future re-admission, and do not differentiate between the propensity and the frequency of re-admissions. In this paper, we develop a new model that overcomes these limitations.

Model Development

First, we develop a set of baseline estimation models to be consistent with prior studies in the literature.

Baseline Model

We estimate the probability that a readmission is a function of patient demographic variables, patient visit characteristics, insurance type, patient condition at admission, and other patient co-morbidities, as shown in Equation 1. We first estimate this model using a logit regression, following the common practice in the readmission literature.

$$\log\left(\frac{p}{1-p}\right) = x \cdot \beta + \varepsilon \tag{1}$$

where p is the probability of the (readmission) event to occur. All model variables are defined in Table A in the appendix. Per the standard definition in health care, a readmission is considered when a patient is readmitted as an inpatient within a 30-day period from the previous discharge for the same primary diagnosis. The estimation model shown in (1) allows us to estimate the various factors that are associated with the likelihood of patient readmission due to CHF.

Beta Geometric/ Erlang-2 Gamma (BG/EG) Hurdle model

In order to address the deficiencies associated with the baseline models, we now seek to develop a stochastic model, the BG/EG Hurdle model, to provide a better understanding of clinical characteristics and patient characteristics that impact patient readmission patterns. The model consists of two components: a hurdle component which estimates the probability of readmission, and a BG/EG component which models the frequency and timing of the next readmission, accounting for unobserved patient-level heterogeneity and non-stationary readmission rates. .

The hurdle component originates from a Logit regression which models the probability of a zero outcome. In other words, the probability of no readmission at any time in the future,

$$\log\left(\frac{\theta_{0i}}{1-\theta_{0i}}\right) = X_{0i} \cdot \xi_{0i} + \varepsilon_{0i}, \quad i = 1, \ldots, N \tag{2}$$

where θ_{0i} is the probability of zero-outcome (no readmission) for individual patient i, X_{0i} is the set of covariates observed for patient i at their initial admission time with coefficients ξ_{0i}. The propensity of future readmission is then the flip side of hurdle-at-zero. Hurdle regression considers systematically different statistical processes for zero versus nonzero binary outcomes, where the positive counts are conditioned on having a nonzero outcome[1]. The appeal of a hurdle-at-zero formulation is that it can account for a large number of patients with zero readmission (Winkelmann 2010), as with our data where 70% of CHF patients are not readmitted. It also reflects a two-stage process, where the risk factors dominating a patient's readmission frequency are systematically different from the risk factors determining the readmission propensity.

The BG/EG component simultaneously models both the frequency and the timing of the admissions. Suppose we have N patients, where patient i is readmitted J_i times in the period $(0,T_i]$ with the time of admissions occur at $(t_1,t_2,\ldots,t_{J_i})$ where t_0=0 corresponds to the initial admission time (j=0 indexes the initial admission) and T_i represents the censoring point which is the end of the model calibration period for patient i (it is June 2010 in our data). As each patient i has different entering points, T_i varies across patients.

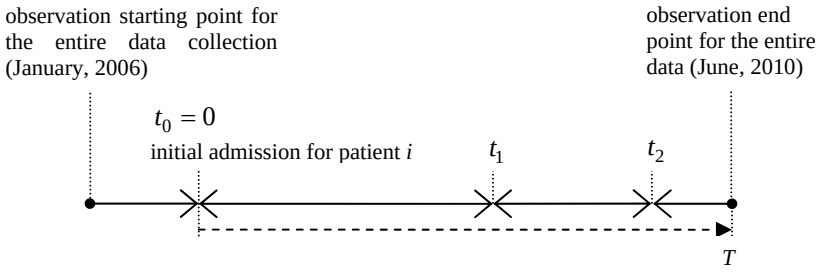

Figure 1 Illustration of Admission Timing Intervals

The total number of admissions of patient i experiencing J_i readmissions is J_i+1 ($J_i \geq 0$). We assume that the time interval between two consecutive admissions follow an Erlang-2 distribution which means that a patient's future readmission timing depends not only on the current visit, but also on the previous admission. This relaxes the restrictive stationary assumption held by many count models such as the popular negative binomial model (NBD) (Winkelmann 2010).

Table 1. Description of Model Parameters

Parameter	Description
i	Subscript for individual patient i
j	Subscript for jth readmission (j=0 refer to the first admission)
t_j	Time at jth readmission
J_i	Total number of readmissions of individual i
T_i	Calibration time of individual i
λ	Base hazard rate, admission rate
γ	Covariate coefficient
r	Shape parameter of the gamma distribution function
α	Scale parameter of the gamma distribution function
p	Dropout probability
a	Dropout parameter of a beta distribution function
b	Non-dropout parameter of a beta distribution function

1 The hurdle model derived its name because it is as if a patient needs to first overcome the first hurdle of being readmitted to be considered or the next level.

We start by specifying individual *i*'s probability, or hazard function, to pay a hospital visit on time duration *t*, given a set of time-varying covariates, X_t, as

$$h(t, X_t) = \lambda_t \cdot e^{X_t \gamma_1} \equiv \lambda_t \cdot \phi(t) \tag{3}$$

where λ_t is the baseline hazard at time *t* (Cox 1972).

We follow Seetharaman and Chintagunta's (2003) formulation of a continuous time proportional hazard model to incorporate the effects of time-varying covariates $(X_{t_1}, X_{t_2}, \ldots, X_{t_{J_i}})$ into the Erlang-2 distribution. The individual level survivor function of the inter admission time distribution between the (j-1)th admission and the jth admission is given by

$$\begin{aligned} S(t_j - t_{j-1}, X_{t_j}) &= (1 + \lambda_j \cdot \int_{t_{j-1}}^{t_j} \phi(u)\, du) \cdot \exp[-\lambda_j \int_{t_{j-1}}^{t_j} \phi(u)\, du] \\ &= (1 + \lambda_j \psi(t_j, t_{j-1})) \exp[-\lambda_j \psi(t_j, t_{j-1})] \end{aligned} \tag{4}$$

where

$$\begin{aligned} \psi(t_j, t_{j-1}) &\equiv \psi(t_j) - \psi(t_{j-1}) \\ \psi(t) &\equiv \int_0^t \phi(u)\, du \\ \phi(t) &\equiv e^{X_t \gamma_1} \end{aligned} \tag{5}$$

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

$$f(t_j - t_{j-1} \mid \lambda_j, \gamma_1, X_{t_j}) = \lambda_j^2 \cdot \phi(t_j) \cdot \psi(t_j, t_{j-1}) \cdot \exp[-\lambda_j \psi(t_j, t_{j-1})]. \tag{6}$$

The likelihood function at the individual patient level is simply the product of equation (9) :

$$L(T_i, \gamma_1 \mid \lambda, X_t) = \prod_{j=1}^{J_i} f(t_j - t_{j-1} \mid \lambda_j, \gamma_1, X_{t_j}) \tag{7}$$

To model the unobserved heterogeneity across patients, we adopt the common mixing distribution for λ (Winkelmann 2010) which is assumed to be gamma distributed with shape parameter *r* and scale parameter α :

$$\begin{aligned} L(T_i, \gamma_1, r, \alpha \mid J_i, X_t) &= \int_0^\infty L(T_i, \gamma_1 \mid \lambda, J_i, X_t) g(\lambda \mid r, \alpha) d\lambda \\ &= \int_0^\infty \lambda^{2J_i} \left(\prod_{j=1}^{J_i} \phi(t_j) \psi(t_j, t_{j-1}) \right) \cdot \exp\left(-\lambda \left(\sum_{j=1}^{J_i} \psi(t_j, t_{j-1}) + \psi(T, t_J) \right) \right) \cdot \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)} d\lambda \\ &= \left(\prod_{j=1}^{J} \phi(t_j) \psi(t_j, t_{j-1}) \right) \cdot \frac{\Gamma(r+2J) \cdot \alpha^r}{\Gamma(r)} \cdot \left(\frac{1}{\alpha + \sum_{j=1}^{J} \psi(t_j, t_{j-1}) + \psi(T, t_J)} \right)^{r+2J} \end{aligned} \tag{8}$$

We further assume that, after every admission, a patient can become inactive with a dropout probability of *p*. This enables us to model the unobserved heterogeneity that causes a patient to drop out of our sample due to reasons such as death, being cured or moving outside the region. It also directly addresses the data censoring issue as the data may only capture a snapshot of certain patients' life-time admissions (though we obtain a relatively long 4.5 years of data) because the probability of not observing the patient in the next period is considered. We specify this 'dropout' pattern to be a common geometric process with a beta mixing function, i.e. the beta-geometric (BG) distribution (Fader et al. 2005) as follows:

$$P(\text{dropout after jth admission}) = p(1-p)^{j-1} \tag{9}$$

The heterogeneity in dropout probabilities follows a beta distribution, the natural mixing distribution for the binary-outcome geometric processes, with parameters *a* and *b*:

$$f(p \mid a,b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} \tag{10}$$

Taking the expectation over λ and *p* yields to the individual likelihood function,

$$\begin{aligned} L_1(r,\alpha,a,b,\gamma \mid X,J,t_J,T) &= A_1 \times \int_0^1 (1-p)^J \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} dp + A_2 \times \int_0^1 p(1-p)^{J-1} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} dp \\ &= \left(\prod_{j=1}^{J} \phi(t_j) \psi(t_j, t_{j-1}) \right) \cdot \frac{\Gamma(r+2J)}{\Gamma(r)} \cdot \alpha^r \cdot \frac{\Gamma(a+b)\Gamma(b+J-1)}{\Gamma(b)\Gamma(a+b+J)} \\ &\cdot \left[(b+J-1) \left(\frac{1}{\alpha + \sum_{j=1}^{J} \psi(t_j, t_{j-1}) + \psi(T, t_J)} \right)^{r+2J} + \delta_{J>0} \cdot a \cdot \left(\frac{1}{\alpha + \sum_{j=1}^{J} \psi(t_j, t_{j-1})} \right)^{r+2J} \right] \end{aligned} \tag{11}$$

Hence, the log-likelihood of *N* individuals is given by

$$LL_1 = \sum_{i=1}^{N} log(L_1(r,\alpha,a,b,\gamma \mid X_i(t), J_i, t_{J_i}, T_i)). \tag{12}$$

The above specification works only for patients who have crossed the hurdle, i.e. those who have at least one readmission. Combining it with the logit hurdle component we described earlier, for a patient with no readmission after the initial admission (J=0), her likelihood simplifies to

$$\begin{aligned} L_0 &= \int \exp\left(-\lambda \psi(T,0)\right) \cdot (1 + \lambda \psi(T,0)) \cdot \frac{\alpha^r \gamma^{r-1} e^{-\alpha\gamma}}{\Gamma(r)} d\lambda \\ &= \left(\frac{\alpha}{\alpha + \psi(T,0)} \right)^r \cdot \left(1 + \frac{r \cdot \psi(T,0)}{\alpha + \psi(T,0)} \right). \end{aligned} \tag{13}$$

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

$$L = \prod_{i=1}^{N} \theta_0^{d_i} \frac{(1-\theta_0)^{1-d_i}}{(1-L_0)^{1-d_i}} \cdot L_1^{1-d_i} \tag{14}$$

where,

$$\theta_{0i} = P(J_i = 0)$$

$$d_i = 1 - min\{J_i, 1\}$$

The log-likelihood of BG/EG Hurdle model is therefore

$$LL = \sum_{i=1}^{N} (d_i \cdot \log \theta_{0i} + (1-d_i) \cdot \log(1-\theta_{0i}) - (1-d_i) \cdot \log(1-L_{0i}) + (1-d_i) \cdot LL_{1i} \quad , \tag{15}$$

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

Data

Our data consists of four and half years (January 2006 to June 2010) of patient admission records from seventy-two hospitals in the North Texas region. Patients' visits across multiple hospitals are tracked by matching the regional master patient index (REMPI), developed by the DFWHC Foundation. This rich dataset includes entire patient admission records across multiple regional hospitals, and is unique because it can identify a single patient's hospitalization history over time and across multiple hospitals within a large metropolitan region.

The total number of observations includes 73,388 admissions for 45,499 distinct patients having CHF as the primary diagnosis. Among all the patients admitted for CHF, 70% of those had a single admission, while around 30% (13,887) of the patients experienced multiple admissions. Table 2 provides a description of the sample.

Table 2. Sample Summary

Length of Observation	4.5 years
Total Number of Observations	73,388
Number of Patients	45,499
Number of Hospitals	72
Number of Attending Physicians (with valid ID)	4,547
Number of Patients with Multiple Hospitalizations	13,887

Our data captures several patient demographic characteristics including gender, racial profile and discharge age. Among all patients, 52% (23,622) were female, 71% (32,503) were white, and 21% (9,625) were African Americans. The average discharge age was 69, with 66% (30,232) patients being 65 or older.

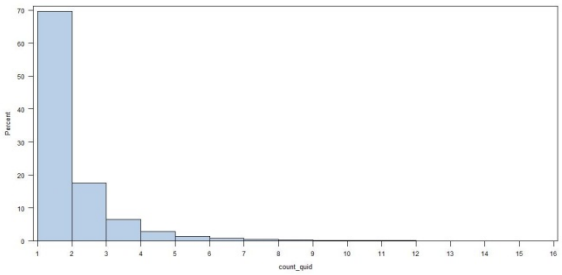

Figure 2. Distribution of the Number of Admissions per Patient with CHF

The data also contains a set of clinical variables which indicate the condition, severity, diagnosis and treatment of patients' admission characteristics. Some key variables that have been commonly used in the readmission literature (Ross et al 2008, Silverstein et al 2008, Mudge et al 2010) include the length of stay (LOS), number of diagnoses, number of procedures, total charges, admission type, and the risk of mortality. The hospital length of stay is defined as the number of days from the date of admission to the date of discharge. In our sample, we observe that the length of stay per admission was 5.45 days on average, with an average of 12.85 diagnoses and 1.09 procedures recorded on each visit. Each admission claimed a total charge of $38,212 on average. On each admission, hospitals record admission types and the risk of mortality. Standard admission type is coded with six classes, from 1 to 5, and 9 (e.g. class 1 denotes the patient was admitted as medical emergency.) In our sample of patients who are admitted with CHF as the principal diagnosis, only admission type classes 1, 2, 3, 5 and 9 are relevant. Risk mortality is coded on a scale from 1 to 4, and represents the death risk of a patient as minor, moderate, major, and extreme, respectively.

For each patient admission, hospital and attending physician identifiers are recorded, that uniquely identify the hospital and the physician. Therefore, on each patient visit, we can track a patient's repeated admission behavior to the same hospital or attending physician. The trivial case is when a patient is hospitalized only once, where the number of hospitals and number of attending physicians are both equal to one. On the other hand, if a patient is hospitalized multiple times, she can either go to the same

hospital or to multiple hospitals over time. In such cases, we track the number of times the patient visits the same hospital prior to a current visit. Sample statistics show that 31% of CHF patients are re-hospitalized, and 37% of those patients with multiple admissions go to different hospitals. Hence, for each admission of a patient, we also count the number of different hospitals visited by the patient prior to the current visit. As a patient's admission pattern may be affected by the quality of care she receives from the hospital and the physicians, we expect these new across-hospital measures to play a significant role in modeling readmission risk. We develop a single measure to take into account both effects: the number of times a patient visited the same hospital as a ratio of the total number of admissions up to the current time.

Congestive heart failure is likely to be accompanied by other disease afflictions, i.e., 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, which are identified to have substantial effects to CHF related admission in the literature (Ross et al. 2008). The comorbidity variables were identified by the Elixhauser index (Elixhauser et al. 1998) based on ICD-9-CM (International Classification of Diseases) diagnosis codes.

Empirical Analysis

Baseline Models and Results

First, we estimate the logistic regression model used to understand patient readmissions. Our intent is to understand whether there will be a resultant readmission at any time in the future after the first admission, and if so, what are the effects of covariates. Our independent variables include demographic variables (patient gender, racial characteristics, discharge age), admission characteristics (number of procedures, and the length of stay), insurance variables (types of insurance), admission condition variables (medical emergency and risk mortality scores), and comorbidity variables (secondary disease afflictions) as described in Table A in the Appendix.

We present our estimates for the patient level baseline model in Table 3. Patient gender and race are significant risk predictors of readmission. Female CHF patients exhibit a 3% lower risk of readmission, while African Americans exhibit a significantly higher risk of readmission up to 57%. These results are consistent with prior studies on female CHF patients (Chin and Goldman 1997, Shelton et al 2000, Silverstein et al 2008), and increased risk of readmission for African American patients (Philbin and DiSalvo 1999, Shelton et al 2000, Silverstein et al 2008). The negative quadratic effects of discharge age indicate that the risk of readmission increases for middle-aged patients, and as the patient gets older, the risk of readmission starts to decrease.

The significant coefficient for count of procedures (*count_proc*) suggests that a thorough treatment helps to reduce the risk of readmission, lowering the risk by 7%. The positive and significant estimates of hospital length of stay (LOS) indicate that longer stay in hospital is an indicator of higher risk of readmission, which is also consistent with previous findings (Mudge et al 2010). An additional day in the hospital increases the risk of readmission by 5%.

Medicare and Medicaid variables are important predictors of higher propensity of future readmissions as well. The risk of readmission of Medicare patients' increases by 18%, and for Medicaid patients, the risk increases by 19% when compared to self-pay patients.

The severity of a patient's condition, such as medical emergency admission and risk mortality scores, are also significant factors. When a patient is admitted due to a medical emergency, his risk of future readmission increases by 10%. Patients with risk mortality level 2 (moderate) have higher risk of readmission compared to risk mortality level 1 (minor) patients, with the risk increasing by 14%.

Among the secondary diagnostic conditions, diabetes mellitus, chronic pulmonary disease, renal failure, drug abuse, and ischemic disease are significant factors associated with increased risk of readmission which is consistent with prior findings (Alexander et al. 1999, Krumholz et al 2000, Philbin and DiSalvo 1999). Drug abuse is the highest risk factor, increasing the risk by 70%. We note that hypertension is associated with lower propensity of future readmissions as our results indicate that CHF patients with

hypertension are likely to exhibit 7.7% fewer readmissions. These results are consistent with earlier findings in the literature (Alexander et al. 1999).

Table 3. Patient Level Model Estimates
(Standard errors in parenthesis; *** p = 0.01; ** p= 0.05; * p=0.10)

		Logistic (Patient Level)		
Parameter		Estimate		Odds Ratio
Intercept		-3.3459	(0.5158)***	0.035
Gender	Female	-0.0263	(0.011)***	0.974
PtRace	American Indian / Eskimo / Aleut	-0.7044	(0.37)**	0.494
PtRace	Asian or Pacific Islander	0.1546	(0.2311)	1.167
PtRace	African American	0.4533	(0.2185)**	1.574
PtRace	Blank	0.0802	(1.0427)	1.083
PtRace	Other	-0.1546	(0.2209)	0.857
log_dischage		1.2482	(0.2404)***	3.484
log_dischage_sqr		-0.2016	(0.0322)***	0.817
count_proc		-0.0754	(0.00639)***	0.927
log_los		0.132	(0.0165)***	1.141
Medicare		0.1634	(0.0261)***	1.177
Medicaid		0.1752	(0.0309)***	1.191
Private Insurance		0.1932	(0.0253)***	1.213
Insurance_other		0.2447	(0.108)**	1.277
admtypcode	Medical Emergency	0.0903	(0.0229)***	1.095
RiskMort	Level 2	0.1305	(0.0328)***	1.139
RiskMort	Level 3	-0.0766	(0.0376)**	0.926
RiskMort	Level 4	-0.195	(0.0534)***	0.823
comorb_diabetes_mellitus		0.1939	(0.0217)***	1.214
comorb_hypertension		-0.0804	(0.0254)***	0.923
comorb_periph_vascular		0.0188	(0.0343)	1.019
comorb_chronic_pulmonary		0.0575	(0.0223)***	1.059
comorb_renal_failure		0.1238	(0.029)***	1.132
comorb_anemia		-0.0631	(0.0244)***	0.939
comorb_alcohol_abuse		-0.0162	(0.0882)	0.984
comorb_drug_abuse		0.534	(0.0696)***	1.706
comorb_ischemic		0.2632	(0.0221)***	1.301
Log-likelihood	-27002.63	AIC		53960.69

BG/EG Hurdle Model and Results[2]

We estimate the BG/EG hurdle model to calculate the probability of having a readmission as well as the timing of the next readmission. The model takes into account unobserved patient-level heterogeneity, non-stationary readmission rate, the timing of readmission patterns as well as data censoring. A unique

2 The BG/EG Hurdle model is comparable to a visit-level baseline model which we don't report in this study. According to the goodness of fit (AIC), the BG/EG Hurdle model outperforms the visit level model.

feature of our model is that it allows estimation of two subtle but distinctive components of patients' readmissions that have been largely overlooked in the literature. We distinguish a patient's propensity of being readmitted (the focus of the extant literature) from the frequency of future readmissions. In our model, we treat the former as a hurdle to be overcome before the latter condition is observed, i.e. a patient first needs to be re-admitted, before we can observe future readmissions.

The two components of our BG/EG hurdle model include estimation of (a) a patient's propensity of being readmitted (logit hurdle), and (b) the frequency of future readmissions (BG/EG). This treatment enables us to discover several interesting patterns. The logit-hurdle part of the model handles the propensity of future readmission. For this component, the same independent variables are used as in the patient-level baseline analysis. For the BG/EG estimates, we include the patient loyalty measure as well.

The logit-hurdle parameter estimates yield results that are consistent with patient-level logit estimates. However, combined with the BG/EG component, interesting patterns emerge for some of the covariates. For example, the number of patient procedures reduces the propensity of readmission by 11.1%. However, once a patient is readmitted, it increases the chance of having more re-admissions by 2.4%. This result may be attributed to the possibility that patients with more procedures are likely to receive greater care, which increases their dropout rate and thereby reduces the propensity of future readmissions. However, if these patients are readmitted again, they are more likely to need future readmissions due to the possibility that they exhibit more complications and number of comorbidities.

Insurance variables are associated with different patterns of the propensity and timing of readmissions after crossing the hurdle-at-zero. Medicare and Medicaid patients have a higher readmission propensity, but once readmitted, they tend to have lower chance of future readmissions by 29.3% and 15.1%, respectively. On the other hand, once patients with private insurance (or other insurance types) are indeed readmitted, their risk of future re-admissions goes up by 8.9% and 22.8%, respectively.

The estimation results show that, compared to non-emergency and mild mortality risk (level 1), medical emergency and moderate mortality risk (level 2) contribute to a higher propensity of re-admission, increasing the risk by 9.3% and 6.8%, respectively. Nevertheless, the admission patterns change once re-admitted. After crossing the re-admission hurdle, medical emergency is associated with a lower propensity of re-admission by 16.4% compared to non-emergency cases. Once re-admitted, patients with moderate mortality risk (level 2) are associated with a lower risk of re-admission by 6.7%, compared to mild mortality risk patients (level 1). These results may indicate that patients admitted with medical emergency and higher mortality risk receive better care, thus reducing the risk of re-admissions in the long run. It also suggest that if a patient with mild mortality risk is re-admitted, special care for follow-up should be in order, as these types of patients are at higher risk for developing more serious CHF episodes requiring frequent admissions to the hospital (Hagland 2011).

Another interesting finding of our study is that repeat care from the same hospital matters. The coefficient estimates for patient loyalty is negative and significant. If this proportional measure is higher (closer to 1), it means that the patient is treated by the same hospital more often, and is less prone to visit multiple hospitals. On the other hand, if the proportional measure is lower (closer to 0), it indicates that the patient is less frequently treated by the same hospital and is more prone to visit multiple hospitals. Our BG/EG analysis shows that, if a patient receives treatments from the same hospital across multiple visits, his risk of future admission decreases by 17.3%. This finding indicates that a patient treated at the same hospital throughout multiple admissions receives better quality of care, which reduces the risk of being readmitted with the same primary diagnosis in the future. This important finding suggests that implementation of EMR at hospitals can reduce readmission risk for CHF patients by providing clinicians with accurate data on patients' medical history and their continuum of care data over time which may help them to provide greater quality of care.

Table 4. BG/EG Model Estimates
(standard error in parenthesis)

Parameters	BG/EG Estimation				Logit Hurdle Estimation			
	Estimates			Hazard Rate	Estimates			Odds Ratio
Gender	-0.1987	(0.0383)	***	0.8042	-0.0926	(0.0221)	***	0.9116
Ptrace: African American	0.2693	(0.0478)	***	1.3379	0.1974	(0.027)	***	1.2182
log_dischage	1.5078	(0.0978)	***	4.8870	1.4596	(0.055)	***	4.3042
log_dischage_sqr	-0.2158	(0.0227)	***	0.7892	-0.4153	(0.0131)	***	0.6601
count_proc	0.0219	(0.0045)	***	1.0241	-0.1169	(0.0064)	***	0.8897
log_los	0.0882	(0.0124)	***	1.1006	0.1818	(0.0163)	***	1.1994
Patient Loyalty	-0.1733	(0.034)	***	0.8271				
Medicare	-0.3155	(0.0314)	***	0.7066	0.0957	(0.0264)	***	1.1004
Medicaid	-0.1491	(0.029)	***	0.8494	0.0816	(0.0308)	***	1.0850
Private Insurance	0.0784	(0.0248)	***	1.0890	0.0021	(0.025)		1.0021
Insurance_other	0.1895	(0.0805)	**	1.2279	0.085	(0.1046)		1.0887
Admtypcode (Emergency)	-0.1635	(0.019)	***	0.8361	0.089	(0.0231)	***	1.0931
RiskMort level 2	-0.0631	(0.027)	**	0.9334	0.066	(0.0321)	**	1.0682
RiskMort level 3	0.0143	(0.0299)		1.0157	-0.2093	(0.0374)	***	0.8112
RiskMort level 4	-0.0511	(0.0394)		0.9458	-0.2168	(0.0519)	***	0.8051
comorb_diabetes_mellitus	-0.1619	(0.0247)	***	0.8375	0.2539	(0.0218)	***	1.2890
comorb_hypertension	0.1708	(0.0214)	***	1.2034	-0.0645	(0.0257)	**	0.9375
comorb_periph_vascular	0.0281	(0.0289)		1.0311	0.047	(0.0349)		1.0481
comorb_chronic_pulmonary	0.0551	(0.0198)	***	1.0618	0.0812	(0.0225)	***	1.0846
comorb_renal_failure	0.0101	(0.0239)		1.0111	0.1395	(0.0293)	***	1.1497
comorb_anemia	-0.0194	(0.0186)		0.9791	-0.0626	(0.0247)	**	0.9393
comorb_alcohol_abuse	0.001	(0.0617)		1.0011	0.0243	(0.0853)		1.0246
comorb_drug_abuse	0.0978	(0.0472)	**	1.1121	0.3384	(0.0708)	***	1.4027
comorb_ischemic	-0.0038	(0.0201)		0.9959	0.2642	(0.0224)	***	1.3024
r	0.4892	Log-likelihood		-204301.09				
alpha	130.0269	AIC		408650.17				
a	2.9105							
b	5.9945							
Expected Daily Admission Rate	0.44%							
Expected Drop-out Rate	32.83%							

Predictive Performance Comparison

To explore the relative performance of the baseline models and the BG/EG Hurdle model, we compare the predictive capabilities of these models. The main interest in the prior readmission literature has been limited to assessing readmission risk of each patient. None of the prior studies attempted to illustrate the predictive power of their models. Our BG/EG model is intended to be a predictive model, the timing and the frequency of future readmissions for any given patient. We now evaluate our model by comparing its predictive performance against that of the baseline models'.

We use two and a half years data (January 2007 to June 2009) for calibrating the training model and we then make prediction on the remaining one-year holdout data. For the logit and BG/EG hurdle models, the last observed record of each patient in the training set is used as the 'snapshot' to build the training model, which then is applied to the testing data to predict readmission occurrences of each patient during the hold out period. The models' predictive accuracies are measured against the actual data in the last one-year period.

Our prediction outcomes are measured using lift curves as depicted in Figure 3. Lift is a commonly used prediction criteria in predictive modeling (Neslin et al. 2006) and is popular in database marketing for customer retention. The lift curve is constructed by first ranking the patients according to their risk of readmission in a descending order, where the highest risk patient appears at the top. The probability of readmission of each patient is derived from the Logit model and our proposed model. Any point (x, y) belongs on the lift curve (see Figure 3) if the top x% of this sorted data captures y% of the actual readmissions. If the model is random, then the top x% of the data would be expected to contain x% of the readmissions, exhibited as a 45-degree line in Figure 3. Hence, the difference (y-x) is the lift obtained as a result of the model. Figure 3 presents an example of using a lift curve to determine the performance of a predictive model. For example, the BG/EG Hurdle model identifies 27% of the true readmissions from the top 20% of the sorted data, whereas the logit model captures 23% of true readmissions. The area between the 45-degree line and the lift curve of the model is a measure of how much "lift" each model provides. The BG/EG model uniformly outperforms the baseline logistic regression across all patient segments.

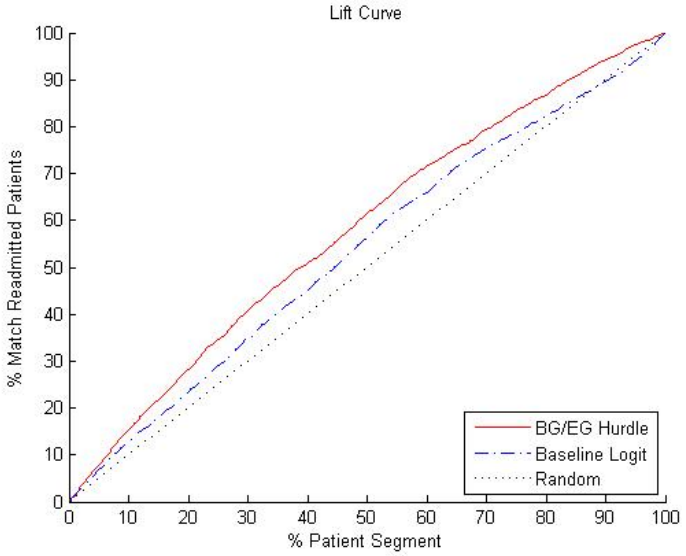

Figure 3. Lift Curves for the Baseline Logit and BG/EG Hurdle models

Information Technology and Readmissions

Once the characteristics of patient readmission patterns are identified, we plan to examine the impact of information technology application on readmission. In addition to the current patient admission data set,

we obtained hospital IT usage archival data from the HIMSS Analytics Database. Clinical IT as well as the electronic medical record (EMR) applications have been considered as essential element to success in reducing avoidable inpatient readmissions (Hagland 2011).

Besides EMR, the use of clinical and administrative IT applications in each hospital also lay a foundation for more effective patient treatment that impacts readmission risks. For example, if a hospital is equipped with a cardiology information system and clinical decision support system, a CHF patient's readmission risk stratification can be assessed more efficiently in real-time. Such risk stratification is critical in identifying high-risk patients and in reducing future readmission by taking preemptive measures. Our ongoing work is focused on studying the specific impact of clinical health IT applications on future readmission and quality of the patient care.

Conclusions

In this study, we examine how patient and clinical characteristics impact the risk propensity of future readmissions for patients with congestive heart failure. Incorporating patient history of readmissions across multiple hospitals, we develop a predictive model, termed as the BG/EG Hurdle, which accounts for the unobserved patient heterogeneity, non-stationary admission rates. Furthermore, we are able to estimate the specific effects related to the propensity of readmission and the timing of future readmissions.

By developing a greater understanding of patient admission data, a hospital can better profile patients who are at higher risk of readmissions and efficiently implement preventive measures to target these patients. Due to the lack of information sharing across multiple hospitals, mostly a result of lacking a master patient index that can track a unique patient's readmissions across different hospitals, the scope of prior readmission studies has been severely limited. In this study, we study the risk factors for patients' readmission patterns across multiple hospitals over four years based on a unique data set obtained through integration of data from EMR systems across 72 hospitals. Our BG/EG Hurdle estimation model provides a novel and unique methodology to overcome the limitations of previous baseline models.

While prior studies have investigated the risk factors associated with the propensity of re-hospitalization, we explicitly consider both the propensity and timing of readmission in our study, thereby providing a more accurate understanding of patient readmission patterns. Allowing the admission rate to change subject to information from the prior admission, we acknowledge that risk factors may have different effects on readmission risk at different stages of patient care. For example, we find that, at the time of initial admission, the number of procedures lowers the propensity of readmission. However, once a patient starts to experience repeated hospitalizations, it increases the chance of having more frequent re-admissions. For hospitals, it means that they should pay more attention to follow-up on patients who are readmitted more than once, because they are more likely to need future hospitalizations due to the possibility that they exhibit more complications and number of comorbidities. From a patient loyalty perspective, we find that repeated care from the same hospital reduces the risk of readmission significantly. This indicates that a patient treated at the same hospital through multiple admissions receives better quality of care, which reduces the risk of being readmitted with the same primary diagnosis in the future.

References

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

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

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

Chin, M.H., and Goldman, L. 1997. "Correlates of Early Hospital Readmission or Death in Patients With Congestive Heart Failure," *Am. J. Cardiol.* (79:12), pp. 1640-1644.

Cox, D.R. 1972. "Regression Models and Life-Tables," *Journal of the Royal Statistical Society. Series B (Methodological)* (34:2), pp. 187–220

Deswal, A., Petersen, N.J., Souchek, J., Ashton, C.M., and Wray, N.P. 2004. "Impact of race on health care utilization and outcomes in veterans with congestive heart failure," *J. Am. Coll. Cardiol.* (43:5), pp. 778-784.

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

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

Elixhauser, A., Steiner, C., Harris, D.R., and Coffey, R.M. 1998. "Comorbidity Measures for Use with Administrative Data," *Medical Care* (36:1), pp. 8-27.

Fader, P., Hardie, B.G.S., and Lee, K. 2005. "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Sci.* (24:2), pp. 275-284.

Fader, P., and Lattin, J. M. 1993. " Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purchase Behavior," *Marketing Sci.* (12:3), pp. 304-317.

Felker, G.M., Leimberger, J.D., Califf, R.M., Cuffe, M.S., Massie, B.M., Adams, K.F., Gheorghiade, M., and O'Connor, C. 2004. "Risk stratification after hospitalization for decompensated heart failure," *J. Card. Fail.* (10:6), pp.460-466.

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

Gupta, S. 1991. "Stochastic Models of Interpurchase Time with Time-Dependent Covariates," *J. Marketing Res.* (28:1), pp. 1-15.

Hagland, M. 2011. "Mastering Readmissions: Laying the Foundation for Change" Healthcare Informatics (28:4), 10-16.

Krumholz, H.M., Chen, Y., Wang, Y., Vaccarino, V., Radford, M.J., and Horwitz, R.I. 2000. "Predictors of readmission among elderly survivors of admission with heart failure," *Am. Heart J.* (139:1), pp.72-77.

Menon, N.M., Lee, B., and Eldenburg, L. 2000. " Productivity of Information Systems in the Healthcare Industry," *Inform. Systems Res.* (11:1), pp.83-92.

Morrison, D.G., and Schmittlein, D.C. 1988. " Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?," *J. Bus. & Econ. Stat.* (6:2), pp.145-159.

Mudge, A.M., Kasper, K., Clair, A., Redfern, H., Bell, J.J., Barras, M.A., Dip, G., and Pachana, N.A. 2010. "Recurrent Readmissions in Medical Patients: a Prospective Study," *J. Hosp. Med.* (6:2), pp. 61-67.

Muus, K.J., Knudson, A., Klug, M.G., Gokun, J., and Sarrazin, M. 2010. "Effect of post-discharge follow-up care on re-admissions among US veterans with congestive heart failure: a rural-urban comparison," *Int. J. Rural and Remote Health Res.* (10:1447), pp. 1-11.

Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., and Mason, C. H. 2006. “Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models,” *J. Marketing Res.* (43:2), pp. 204–211.

Philbin, E.F., and DiSalvo, T.G. 1999. "Prediction of Hospital Readmission for Heart Failure: Development of a Simple Risk Score Based on Administrative Data," *J. Am. Coll. Cardiol.* (33:6), pp. 1560-1566.

Philbin, E.F., Dec, G.W., Jenkins, P.L., and DiSalvo, T.G. 2001. " Socioeconomic status as an independent risk factor for hospital readmission for heart failure," *Am. J. Cardiol* (87:12), pp. 1367-1371.

Reinartz, W.J., and Kumar, V. 2003. "The impact of relationship characteristics on profitable lifetime duration," *J. Marketing* (67:1), pp. 77-99.

Ross, J.S., Mulvey, G.K., Stauffer, B., Patlolla, V., Bernheim, S.M., Keenan, P.S., and Krumholz, H.M. 2008. "Statistical Models and Patient Predictors of Readmission for Heart Failure: A Systematic Review," *Arch. Intern. Med.* (168:13), pp. 1371-1386.

Schmittlein, D.C., Morrison, D.G., and Colombo, R. 1987. "Counting your customers: Who they are and what will they do next?," *Management Sci.* (33:1), pp. 1-24.

Seetharaman, P.B., and Chintagunta, P. 2003. "The Proportional Hazard Model for Purchase Timing: A Comparison of Alternative Specifications," *J. Bus. & Econ. Stat.* (21:3), pp.368-382.

Shelton, P., Sager, M.A., and Schraeder, C. 2000. "The Community Assessment Risk Screen (CARS): Identifying Elderly Persons at Risk for Hospitalization or Emergency Department Visit," *Am. J. Managed Care* (6:8), pp. 925-933.

Silverstein, M.D., Qin, H., Mercer, S.Q., Fong, J., and Haydar, Z. 2008. "Risk Factors for 30-day Hospital Readmission in Patients ≥65 Years of Age," *Baylor U. Medical Center Proc.* (21:4), pp. 363-372.

Wilson, E.V., and Lankton, N.K. 2004. "Modeling patients' acceptance of provider-delivered e-health," *J. Am. Med. Informatics Assoc.* (11:4), pp. 241-248.

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

Appendix

Table A. Variable Descriptions

	Variable Name	Description of Variable	Descriptive Statistics
Demographic Variables	Gender	Patient's gender	Female (52%), Male (48%)
	Race	Patient's race	White (71%), African American (21%)
	log(Discharge age)	Patient's age at the day of discharge log transformed	69 (15.61)[a,b]
	log(Disharge age) square	Quadratic term of log-transformed discharge age	
Visit Characteristics	Count of procedures	Number of procedures on each admission per patient	1.09 (1.99)[a]
	log(LOS)	log-transformed length of stay	5.45 (5.78)[a,c]
Patient Loyalty	Proportion of count of visits to the same hospital	Proportion of the number of times a patient visited the same hospital out of the total number of visits up to the current admission	0.93 (0.19)[a]
Total Charge and Insurance Variables	Medicare	Binary indicator of claim filed to Medicare	45,061 (61.4)[d]
	Medicaid	Binary indicator of claim filed to Medicaid	13,786 (18.79)[d]
	Private	Binary indicator of private insurance	46,306 (63.1)[d]
	Other type of insurance	Binary indicator of other types of insurance excluding Medicare/Medicaid/self pay/private	764 (1.04)[d]
Admission Condition Variales	Admission Typecode (Medical Emergency)	Binary indicator of admission type classified as medical emergency	50,063 (68.22)[d]
	RiskMort	Risk mortality (1: Minor, 2:Moderate, 3:Major, 4:Extreme)	
Comorbidity Variables	comorb_diabetes_mellitus	Comorbidity dummy - Diabetes Mellitus	32,203 (43.88)[d]
	comorb_hypertension	Comorbidity dummy - Hypertension	30,629 (41.74)[d]
	comorb_periph_vascular	Comorbidity dummy - Periphery Vascular	7,882 (10.74)[d]
	comorb_chronic_pulmonary	Comorbidity dummy - Chronic Pulmonary Disease	25,975 (35.39)[d]
	comorb_renal_failure	Comorbidity dummy - Renal Failure	25,840 (35.21)[d]
	comorb_anemia	Comorbidity dummy - Anemia	21,967 (29.93)[d]
	comorb_alcohol_abuse	Comorbidity dummy - Alcohol Abuse	1,143 (1.56)[d]
	comorb_drug_abuse	Comorbidity dummy - Drug Abuse	2,490 (3.39)[d]
	comorb_ischemic	Comorbidity dummy -Ischemic Disease	39,869 (54.33)[d]

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