Two Ways of Modeling Hospital Readmissions: Mixed and Marginal Models

Hui Fen Tan[1 2], Ronald Low[1 3] MD, Shunsuke Ito[1]
Raymond Gregory[1], Leonard Bielory[4] MD, Van Dunn[1] MD

Abstract

The Center for Medicare and Medicaid Services is progressively reducing reimbursement for some hospital readmissions. Understanding factors associated with readmission is increasingly important. Using 13 years of admissions data from New York City's public hospitals, we develop models to predict if a pneumonia patient will be readmitted to a hospital within 30 days of discharge, using covariates such as hospital conditions, patient medical history and demographics, weather, and pollution levels. As a patient could return to the hospital several hundred times over 13 years or never return at all, we use two types of models to account for correlation between observations: a marginal model estimated using generalized estimating equations, and a mixed model with random effects for patient and hospital. The latter model has a higher prediction accuracy of 89.47%. Gender, insurance status, and past medical history were significant predictors. Having a history of serious medical diseases increases readmission risk greatly. However, the two models ask different questions, with the former model being perhaps more relevant for hospital administrators who wish to know the effect of covariates on the population average instead of individual patients. We illustrate the similarities and differences between the empirical results of both models.

Key Words: hospital readmissions, longitudinal data, marginal model, generalized estimating equations, mixed model, random effects

1. Introduction

The Hospital Readmissions Reduction Program was instituted under the Patient Protection and Affordable Care Act[6] to reduce health care costs and improve health care quality. Under this program, starting October 2012, Centers for Medicare and Medicaid Services (CMS), will reduce Medicare[7] payments to hospitals with "higher than average" 30-day readmission rates for certain diseases. In the first two years of this program, three diseases are considered: acute myocardial infarction, heart failure and pneumonia [3]. Since 2009, more than 20,000 papers on hospital readmissions have been published.

---

[1] New York City Health and Hospitals Corporation, 125 Worth Street, New York, NY 10013
[2] Contact information: ht2287@columbia.edu
[3] New York University, 550 First Avenue, New York, NY 10016
[4] Star X Allergy Center, 400 Mountain Avenue, Springfield, NJ 07081
[6] The Patient Protection and Affordable Care Act, informally known as 'ObamaCare', was signed into law on March 23, 2010. The Hospital Readmissions Reduction Program was instituted under Section 3025 of this law. [1]
[7] Medicare is a health insurance program administered by the federal government for people 65 and older, and people under 65 with certain disabilities. Medicare consists of four parts: Medicare Part A covers inpatient care in hospitals. [2]

CMS considers the following to be a "readmission":

> Within 30 days of discharge from an acute care hospital, the patient is admitted to an acute care hospital. The readmission need not be to the original hospital. With certain exceptions, the diagnoses for the repeat admission are not considered.

2. Objective

We develop models to predict if a patient will be readmitted to a hospital within 30 days of discharge. The Hospital Readmissions Reduction Program currently tracks 30-day readmission rates for three diseases. We focus on one of them: pneumonia. We identify factors that are related with 30-day readmission of pneumonia. We hope that our models will be useful in identifying at-risk patients and reducing hospital readmissions.

3. Data

Our data comes from the New York City Health and Hospitals Corporation[8] (HHC). Data for this report was abstracted from an ongoing project to reduce costs and hospital admissions. That clinical effort addresses individual patients; only summary data are reported in this paper.

Our data set of 207,103 admissions of 159,674 pneumonia patients to New York City's 11 public hospitals covers 13 years from December 1st, 1998 to December 31st, 2011. For each pneumonia admission, subsequent admissions within 30 days to any of New York City's 11 public hospitals for any health condition were considered "readmissions". Admissions where patients died while at the hospital are not included; neither are admissions in December 2011, since in both cases, 30-day readmission outcomes cannot be observed.

We consider 37 potential predictor variables from the following sources:

- HHC: Inpatient and emergency department patient volumes on days of admission and discharge, patient co-morbidities, admission and discharge dates, patient's age, gender, race, ethnicity, preferred language, country of birth, insurance status, length of hospital stay, whether patient was treated for other health conditions besides pneumonia. Derived variables include month, day of the week, day of the month, and holiday from admission date.
- Census Bureau: zip code level measures of income, poverty and education.
- National Weather Service: temperature, humidity, precipitation. Derived values include extreme weather conditions.
- Environmental Protection Agency: levels of air pollutants (small particles, SO_2, NO_x, NO_2, CO, O_3, Pb).
- Bureau of Labor Statistics: Six measures of unemployment.
- Laboratory of one of the co-authors (LB): Airborne allergens (trees, grasses, weeds, and ragweed pollen counts and mold spores counts).

---

[8] New York City Health and Hospitals Corporation operates New York City's public healthcare system, which consists of 11 public acute care hospitals, 4 nursing facilities, 6 diagnostic and treatment centers, and more than 70 community clinics. [6]

One posited theory is that busy hospitals tend to discharge patients prematurely to make space for incoming patients [4]. The number of inpatient admissions and emergency department admissions give an indication of how busy the hospital was. The likelihood that a patient will develop a hospital-acquired infection is positively related to hospital length of stay [4].

Many of our predictor variables cannot be controlled by hospitals but may still affect readmissions. Hospital administrators are concerned that providing patients discharge instructions, whether verbally or on paper, in a language that is not the patient's primary language deters the patient from proper care after discharge and hence increases readmission risk. This scenario is especially prevalent in public hospitals where the bulk of patients are minorities. Temperature and pollutant levels, particularly Nitrogen Dioxide and small particles, are related to pneumonia readmission [4].

4. Exploratory Data Analysis

We examined the relationship between numeric variables and 30-day readmissions using box plots (Figures 1 and 2). We cross-tabulated categorical variables with 30-day readmissions.

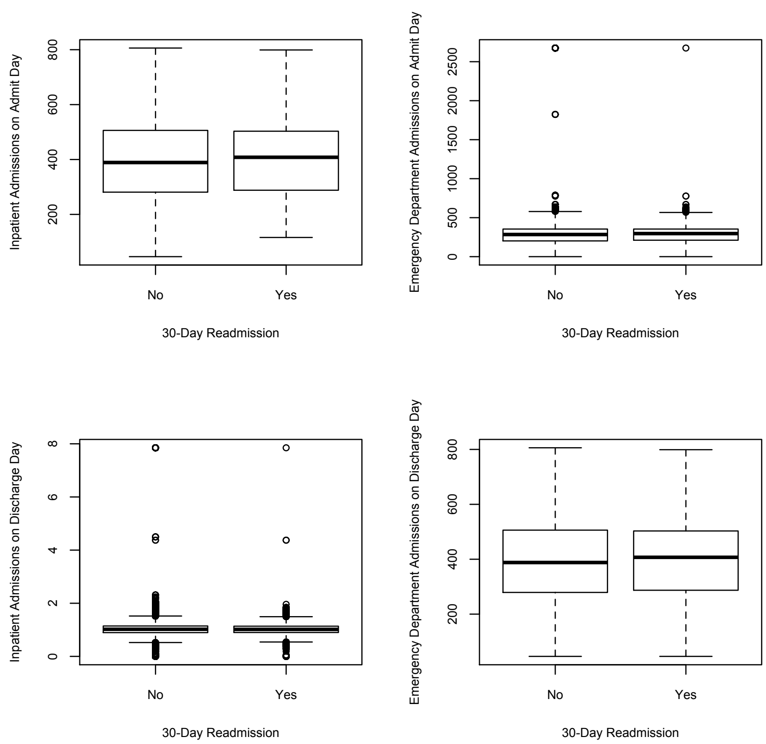

Figure 1: Boxplot of hospital condition variables against 30-day readmission

Surprisingly, hospital condition variables do not appear to have different distributions by 30-day readmission outcomes, as highlighted by previous work on hospital readmissions. Several other numeric variables have similar distributions for 30-day readmission outcomes; it is unlikely that many of these variables will be strong predictors of 30-day readmission. Some examples are shown in Figure 2.

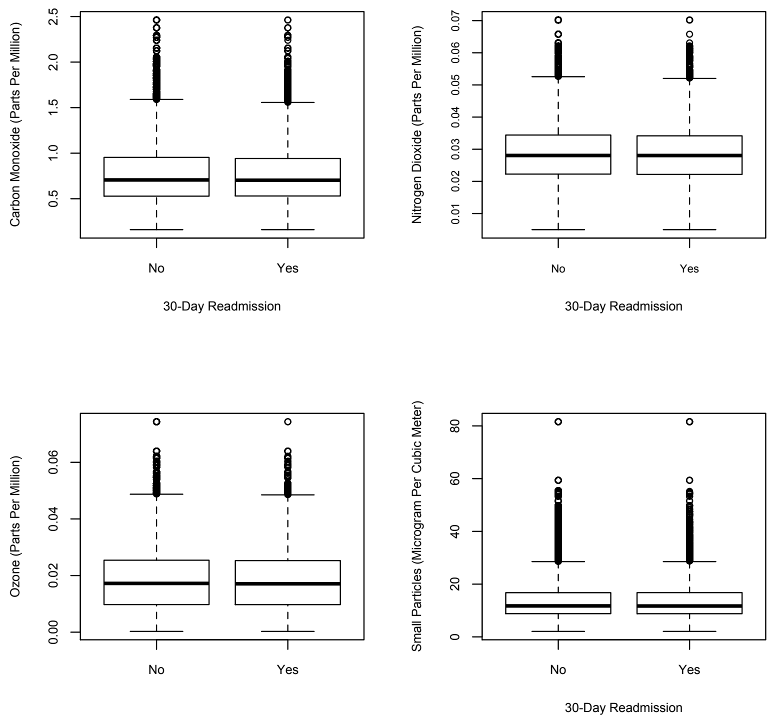

Figure 2: Boxplot of selected numeric variables against 30-day readmission

Overall, the only variables that we found graphically to have different distributions by 30- day readmission outcomes were unemployment, admission type and patients' medical history.

5. Models

Simple logistic regression would seem appropriate for 30-day readmissions, a binary variable. However, hospital admissions are not wholly independent of each other – some patients may only be admitted to the hospital once in 13 years, but there will be others who keep returning to the hospital. To account for correlation between admissions, we use and compare two types of models - marginal models and generalized linear mixed models. Marginal and mixed models answer different questions. The former models the mean response for each admission and "does not incorporate dependence on any random effects or previous admissions" [10], in contrast to the later. The resulting regression coefficients in marginal models contrast the mean responses in subpopulations with common covariate values, allowing inferences to be made about population averages; for this reason, marginal models are frequently called "population-average models". In contrast, mixed models are often called "subject-specific models" because the regression coefficients in mixed models describe the change in mean response for an individual holding other covariates and random effects constant, allowing inferences to be made about individuals. Marginal models can be estimated using generalized estimating equations, whereas mixed models are estimated using maximum likelihood methods.

5.1 Mixed Model

We incorporate two random effects terms to account for within-patient and within-hospital correlation between observations. The model formulation is as follows:

Let p_i be the 30-day readmission probability of admission i:

$$
\begin{aligned}
\text{logit}(p_i) = \alpha^{\text{patient}}_{j[i]} &+ \gamma^{\text{hospital}}_{k[i]} + \beta_0 \\
&+ \beta_1 * \text{length of admission} \\
&+ \beta_2 * \text{treated for other health condition besides pneumonia} \\
&+ \beta_3 * \text{admission type.inpatient} \\
&+ \beta_4 * \text{number of inpatient admissions on admit date} \\
&+ \beta_5 * \text{number of emergency department admissions on admit date} \\
&+ \beta_6 * \text{number of inpatient admissions on discharge date} \\
&+ \beta_7 * \text{number of emergency department admissions on discharge date} \\
&+ \beta_{8-18} * \text{discharge month (11 levels)} + \beta_{19-24} * \text{discharge day of week (6 levels)} \\
&+ \beta_{25} * \text{NYC carbon monoxide level on discharge date} \\
&+ \beta_2 6 * \text{NYC nitrogen dioxide level on discharge date} \\
&+ \beta_{27} * \text{NYC ozone level on discharge date} \\
&+ \beta_{28} * \text{NYC small particles level on discharge date} \\
&+ \beta_{29} * \text{NYC lowest temperature on discharge date} \\
&+ \beta_{30} * \text{NYC mean dew point on discharge date} \\
&+ \beta_{31} * \text{NYC mean sea level pressure on discharge date} \\
&+ \beta_{32} * \text{NYC mean visibility on discharge date} \\
&+ \beta_{33} * \text{NYC maximum wind speed on discharge date} \\
&+ \beta_{34} * \text{NYC total precipitation on discharge date} \\
&+ \beta_{35-38} * \text{NYC extreme weather indicators on discharge date (4 levels)} \\
&+ \beta_{39} * \text{NYC total pollen level on discharge date} \\
&+ \beta_{40} * \text{NYC unemployment on discharge date} \\
a^{\text{patient}}_j \sim N(&\theta_1 * \text{age} + \theta_2 * \text{gender.male} \\
&+ \theta_3 * \text{race.asian} + \theta_4 * \text{race.black} + \theta_5 * \text{race.Hawaiian or Pacific Islander} \\
&+ \theta_6 * \text{race.Hispanic} + \theta_7 * \text{race.other} + \theta_8 * \text{race.South Asian or Middle East} \\
&+ \theta_9 * \text{race.white} \\
&+ \theta_{10} * \text{prefers english} + \theta_{11} * \text{foreign born} + \theta_{12} * \text{no medicare or medicaid} \\
&+ \theta_{13} * \text{has history of asthma} + \theta_{14} * \text{has history of congestive hearth failure} \\
&+ \theta_{15} * \text{has history of myocardial infarction} + \theta_{16} * \text{has history of stroke} \\
&+ \theta_{17} * \text{household income} + \theta_{12} * \text{below poverty level} + \theta_{13} * \text{years of education}, \sigma^2_{\text{patient}}) \\
a^{\text{hospital}}_k \sim N(&0, \sigma^2_{\text{hospital}}) \quad k = 1, \ldots, 11
\end{aligned}
$$

5.2 Marginal Model

Because admissions do not occur at the same time across patients, unlike in a traditional repeated measures data set where some outcome is measured on all subjects at the same time, the data is unbalanced. Predictor variables could be dependent on the particular admission (e.g. hospital conditions) or not dependent on the particular admission (e.g. gender and other demographic variables are always fixed for each patient). So we have a repeated measures model with patient ID being the variable that defines the correlation between admissions.

6. Results

The geeglm function in R's geepack package was used to estimate the marginal model, and the lmer function in R's lme4 package was used to estimate the mixed model. The models were first estimated using all 37 predictor variables. A compound symmetry correlation structure was used in the estimation of the marginal model.

We expected many of the predictor variables to be redundant since from the exploratory data analysis done above, few variables had good discriminatory power for 30-day readmission. Indeed, as many as 20 predictor variables were not statistically significant in the initial mixed and marginal models. A model that can be easily interpreted by hospital administrators needs to be relatively parsimonious; hence for our final models we performed variable selection. Comorbidities, social demographic factors stood out as reliable predictors and met our entry criteria of statistical and medical significance for our final models.

6.1 Marginal model with 28 predictor variables and exchangeable (compound symmetry) correlation structure

Many variables found to be important in other models of hospital readmissions were not statistically significant in this model, including variables such as race and language. Of the 14,498 admissions that resulted in 30-day readmissions, this correctly predicted 2 of them.

6.2 Mixed model with 28 predictor variables and patients and hospitals random effects

This model's predictive power is a substantial improvement over the marginal model; of the 14,498 admissions that resulted in 30-day readmissions, this model correctly predicted 780 of them. While this level of accuracy still leaves much to be desired, it is much higher than the 2 visits predicted correctly by the marginal model. Statistically significant variables include patient demographics such as gender, insurance status, and past medical history. The medical history variables have the largest regression coefficients.

Most of the regression coefficients are as expected. Male patients had exp(0.1038) = 1.109 times the odds of being readmitted in 30 days compared to females. Patients who paid for their hospital visit themselves had exp(-0.2063)=0.8136 times the odds of being readmitted in 30 days compared to patients who had Medicare or Medicaid. The variables with the largest coefficients are medical history variables. For example, patients with a medical history of asthma had exp(0.98) = 2.66 times the odds of being readmitted.

We examined the deviance residuals and plotted them by each of our predictor variables to see if any of the predictor variables have some relationship with the residuals. Most of these plots look like the following plot of gender and residuals, where no trend can be picked out.

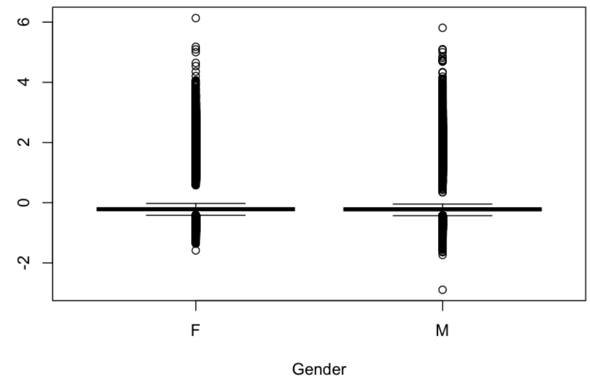

Figure 3: Boxplot of gender against residuals of Model 4

To test for how good the models are, we look at deviance to measure goodness-of-fit but we also look at prediction accuracy. The mixed model has a higher prediction accuracy of 89.47%.

7. Discussion

Our marginal model's bad results are not very surprising; The most appealing part of GEE estimation – that consistent estimates of the regression coefficients are yielded even if the assumed model for covariances among the repeated measures is not correct – may not hold when the data is unbalanced. Our data is certainly extremely unbalanced, with patients returning at all possible times and certainly not at fixed intervals between admissions.

We note that income, poverty level and education information were not collected at the patient level and had to be approximated using zip code level aggregated information. We might have had different findings if we had been able to measure these variables more precisely. It is surprising that three of the four variables that measure how "busy" the hospital is turned out to be statistically insignificant.

CMS hopes to incentivize hospitals to reduce their readmission rates, as high readmission rates are often deemed markers for the quality of care provided during the initial admission and follow-up visit [5]. However, CMS' method for measuring readmission has not been without controversy. For one, CMS-measured readmissions do not distinguish between "planned, scheduled and staged" follow-up treatments and unplanned treatments, and per some medical researchers, this causes readmission rates to be inflated by as much as 25% [4].

Our lack of physical examination or laboratory data on the day of discharge does not allow us to model whether patients were appropriately discharged.

8. Conclusion

Our mixed model performs better than the marginal model for this data set, with the marginal model essentially unable to predict readmission. The marginal model is more restrictive in its assumption of the balanced-ness of the data, which is not fulfilled by our data since admissions do not occur at the same time across patients, unlike in a traditional repeated measures data set where some outcome is measured on all subjects at the same time.

Our results indicate that patient demographics variables such as gender, insurance status, and past medical history are statistically significant. Having a history of serious medical diseases increases readmission risk greatly.

For future work, we are tackling the problem of patient attrition by obtaining a data set that tracks patients outside of the NYC public hospitals system utilizing data from Metroplus Health Plan, an insurance company and subsidiary of HHC. More explanatory variables can be used to augment these models to provide an even clearer picture of the mechanisms behind hospital readmissions.

References

[1] U.S. Department of Health and Human Services. “The Patient Protection and Affordable Care Act”. Web. 29 April 2011. http://www.healthcare.gov/law/full/index.html.
[2] Centers for Medicare and Medicaid Services. “Medicare Benefits”. Web. 29 April 2011. http://www.medicare.gov/navigation/medicare-basics/medicare-benefits/medicare-benefits-overview.aspx
[3] Center for Medicare and Medicaid Services. “Readmissions Reduction Program”. Web. 29 April 2011. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ AcuteInpatientPPS/Readmissions-Reduction-Program.html
[4] Bardi, Jason. “Hospital Readmission Rates Misleading, UCSF Medical Center Study Finds”. Web. 17 April 2011. http://www.ucsf.edu/news/2012/04/11871/ hospital-readmission-rates-misleading-ucsf-medical-center-study-finds
[5] Weissman, J.S., J.Z. Ayanian, S. Chasan-Taber, M. J. Sherwood, C. Roth, A. M. Epstein. “Hospital readmissions and quality of care”. Med Care. May 1999.
[6] New York City Health and Hospitals Corporation. “About Us”. Web. 29 April 2011. http://www.nyc.gov/html/hhc/html/about/about.shtml.
[7] Inter-University Consortium for Political and Social Research. “Census of Population and Housing, 2000 [United States]: Selected Subsets From Summary File 3”. 2000. http: //dx.doi.org/10.3886/ICPSR13402.v2
[8] National Climatic Data Center. “Global Daily Surface Data”. http://www.ncdc.noaa.gov/cdo-web/search#t=secondTabLink.
[9] Environmental Protection Agency. “Air Quality System Data”. http://www.epa.gov/ttn/airs/airsaqs/detaildata/downloadaqsdata.htm.
[10] Fitzmauurice, Garrett M., Nan M. Laird, James H. Ware. “Applied Longitudinal Analysis”. John Wiley & Sons, Inc. 2004.

Appendix

Table 1: Marginal model with 28 predictor variables and exchangable (compound symmetry) correlation structure

Variable	*Estimate*	*Standard Error*	*P-Value*
Intercept	-1.604	0.398	0.000
Hospital 2	-0.416	0.146	0.005
Hospital 3	0.288	0.088	0.001
Hospital 4	-0.633	0.181	0.000
Hospital 5	-0.333	0.118	0.005
Hospital 6	-0.143	0.075	0.058
Hospital 7	-0.468	0.168	0.005
Hospital 8	-0.574	0.153	0.000
Hospital 9	-0.694	0.208	0.000
Hospital 10	-0.600	0.176	0.000
Hospital 11	-0.333	0.115	0.004
Age	-0.000	0.000	0.042
Gender - Male	0.103	0.018	0.000
Race – Asian	0.040	0.136	0.768
Race – Black	0.074	0.131	0.572
Race – Hawaiian / Pacific Islander	0.125	0.502	0.802
Race – Hispanic	0.101	0.131	0.441
Race – Other	-0.015	0.137	0.913
Race – South Asian / Middle East	0.112	0.141	0.428
Race – White	-0.002	0.134	0.985
Language – English	-0.026	0.024	0.264
Foreign Born	-0.207	0.021	0.000
No Medicare or Medicaid	0.263	0.025	0.000
History of asthma	-0.725	0.019	0.000
History of congestive heart failure	1.044	0.027	0.000
History of myocardial infarction	0.322	0.039	0.000
History of stroke	0.360	0.036	0.000
Length of hospital stay	0.002	0.000	0.033
Only treated for pneumonia and nothing else	-0.057	0.025	0.026
Admitted to inpatient	-0.420	0.023	0.000
Median household income	-0.000	0.000	0.251
Percentage below poverty level	-0.237	0.257	0.356
Mean years of education	0.005	0.014	0.710
Inpatient busy-ness on admit day	-0.001	0.000	0.017
Emergency department busy-ness on admit day	-0.000	0.000	0.104
Inpatient busy-ness on discharge day	0.000	0.000	0.665
Emergency department busy-ness on discharge day	-0.000	0.000	0.905
Month – February	0.053	0.041	0.197
Month – March	0.064	0.043	0.134
Month – April	-0.061	0.053	0.245
Month – May	-0.000	0.059	0.995
Month – June	-0.024	0.070	0.732
Month – July	0.1080	0.079	0.171
Month – August	-0.008	0.079	0.918
Month – September	0.099	0.070	0.160
Month – October	0.127	0.055	0.022
Month – November	0.127	0.047	0.007
Month – December	0.047	0.040	0.250

Day – Tuesday	0.016	0.036	0.655
Day – Wednesday	0.020	0.037	0.577
Day – Thursday	0.006	0.037	0.872
Day – Friday	0.107	0.037	0.004
Day – Saturday	0.079	0.044	0.067
Day – Sunday	-0.040	0.046	0.389
Minimum temperature on discharge day	-0.000	0.001	0.914
COppm levels on discharge day	-0.052	0.048	0.389
Extreme weather index of 1/3 on discharge day	0.021	0.032	0.528
Extreme weather index of 1/2 on discharge day	1.054	0.350	0.003
Extreme weather index of 2/3 on discharge day	-0.025	0.027	0.369
Extreme weather index of 1 on discharge day	0.002	0.031	0.942
Pollen levels on discharge day	0.000	0.000	0.331
Unemployment level	0.004	0.003	0.277

Table 2: Mixed model with 28 predictor variables and patients and hospitals random effects

Variable	*Estimate*	*Standard Error*
Intercept	-1.06	0.37
Age	0.00	0.00
Gender - Male	0.11	0.02
Race – Asian	0.17	0.19
Race – Black	0.09	0.18
Race – Hawaiian / Pacific Islander	0.28	0.70
Race – Hispanic	0.16	0.18
Race – Other	-0.04	0.19
Race – South Asian / Middle East	0.26	0.20
Race – White	0.10	0.19
Language – English	-0.08	0.03
Foreign Born	-0.21	0.03
No Medicare or Medicaid	-0.30	0.03
History of asthma	0.66	0.03
History of congestive heart failure	0.98	0.04
History of myocardial infarction	0.27	0.05
History of stroke	0.32	0.05
Length of hospital stay	0.00	0.00
Only treated for pneumonia and nothing else	-0.03	0.03
Admitted to inpatient	-0.42	0.03
Median household income	0.00	0.00
Percentage below poverty level	-1.51	0.30
Mean years of education	-0.05	0.02
Inpatient busy-ness on admit day	0.00	0.00
Emergency department busy-ness on admit day	0.00	0.00
Inpatient busy-ness on discharge day	0.00	0.00
Emergency department busy-ness on discharge day	0.00	0.00
Month – February	0.04	0.05
Month – March	0.07	0.05
Month – April	-0.03	0.06
Month – May	0.03	0.07
Month – June	0.02	0.08
Month – July	0.17	0.09
Month – August	0.04	0.09
Month – September	0.12	0.08

Month – October	0.13	0.06
Month – November	0.13	0.05
Month – December	0.06	0.05
Day – Tuesday	0.02	0.04
Day – Wednesday	0.03	0.04
Day – Thursday	0.03	0.04
Day – Friday	0.15	0.04
Day – Saturday	0.13	0.04
Day – Sunday	0.00	0.05
Minimum temperature on discharge day	0.00	0.00
COppm levels on discharge day	-0.04	0.05
Extreme weather index of 1/3 on discharge day	0.02	0.04
Extreme weather index of 1/2 on discharge day	0.89	0.41
Extreme weather index of 2/3 on discharge day	-0.01	0.03
Extreme weather index of 1 on discharge day	0.02	0.03
Pollen levels on discharge day	0.00	0.00
Unemployment level	0.00	0.00

Error terms:

Groups	Name	Standard Deviation
Subject	Intercept	1.43
Facility	Intercept	0.01
Residual		1.00