Appendix (Online Supplement)

Predictive Analytics for Readmission of Patients with Congestive Heart Failure

Indranil Bardhan[a], Jeong-ha (Cath) Oh[b], Zhiqiang (Eric) Zheng[a], Kirk Kirksey[c]

[a] Naveen Jindal School of Management, The Univ. of Texas at Dallas, Richardson, Texas 75080
[b] Robinson College of Business, PO Box 4015, Georgia State University, Atlanta, Georgia 30302-4015
[c] The University of Texas Southwestern Medical Center, 6303 Forest Park Road, Dallas, Texas 75390-8595

Appendix A: Derivation of Equations in Section 3.2

A patient may be inactive either after T (i.e. still active at the end of period but no visit is observed between the last admission and the end of the period) or right after the last admission. We model these cases as follows:

i) A patient is inactive after T:

$$\begin{aligned} &L(\lambda \mid t_1,\ldots,t_J,T, \text{inactive at time } \tau \in (t_J,T]) \\ &\quad = \lambda^2\phi(t_1)\psi(t_1,t_0)\exp[-\lambda\psi(t_1,t_0)]\times\lambda^2\phi(t_2)\psi(t_2,t_1)\exp[-\lambda\psi(t_2,t_1)] \\ &\quad\quad \times\cdots\times\lambda^2\phi(t_J)\psi(t_J,t_{J-1})exp[-\lambda\psi(t_J,t_{J-1})]\times\exp[-\lambda\psi(T,t_J)] \\ &\quad = \lambda^{2J}\left(\prod_{j=1}^{J}\phi(t_j)\psi(t_j,t_{j-1})\right)\exp\left(-\lambda\cdot\left(\sum_{j=1}^{J}\psi(t_j,t_{j-1})+\psi(T,t_J)\right)\right) \end{aligned} \tag{9}$$

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

$$\begin{aligned} &L(\lambda \mid t_1,\ldots,t_J,T, \text{inactive at time } \tau \in (t_J,T]) \\ &\quad = \lambda^2\phi(t_1)\psi(t_1,t_0)\exp[-\lambda\psi(t_1,t_0)]\times\lambda^2\phi(t_2)\psi(t_2,t_1)\exp[-\lambda\psi(t_2,t_1)] \\ &\quad\quad \times\cdots\times\lambda^2\phi(t_J)\psi(t_J,t_{J-1})exp[-\lambda\psi(t_J,t_{J-1})] \\ &\quad = \lambda^{2J}\left(\prod_{j=1}^{J}\phi(t_j)\psi(t_j,t_{j-1})\right)\exp\left(-\lambda\cdot\sum_{j=1}^{J}\psi(t_j,t_{j-1})\right) \end{aligned} \tag{10}$$

This yields the likelihood function:

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

Expectation over the distribution of $\bullet$ yields to

$$L(r,\alpha,p,\gamma \mid X,J,t_J,T) = (1-p)^J\times A_1 + \delta_{J>0}\,p(1-p)^{J-1}\times A_2 \tag{12}$$

where,

$$A_1 = \int_0^{\infty} \lambda^{2J} \left(\prod_{j=1}^{J} \phi(t_j)\psi(t_j, t_{j-1}) \right) exp\left(-\lambda(\sum_{j=1}^{J} \psi(t_j, t_{j-1} + \psi(T, t_J)) \right) \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)} d\lambda$$

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

$$A_2 = \int_0^{\infty} \lambda^{2J} \left(\prod_{j=1}^{J} \phi(t_j)\psi(t_j, t_{j-1}) \right) exp\left(-\lambda \sum_{j=1}^{J} \psi(t_j, t_{j-1}) \right) \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)} d\lambda$$

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

Taking the expectation over $\bullet$ and p yields the individual likelihood function,

$$L_1(r, \alpha, a, b, \gamma \mid X, J, t_J, T) = A_1 \times \int_0^1 (1-p)^J \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} dp + A_2 \times \int_0^1 p(1-p)^{J-1} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} dp$$

$$= \left(\prod_{j=1}^{J} \phi(t_j)\psi(t_j, t_{j-1}) \right) \cdot \frac{\Gamma(r+2J)}{\Gamma(r)} \cdot \alpha^r \cdot \frac{\Gamma(a+b)\Gamma(b+J-1)}{\Gamma(b)\Gamma(a+b+J)} \quad (15)$$

$$\cdot \left[(b+J-1) \left(\alpha + \sum_{j=1}^{J} \psi(t_j, t_{j-1}) + \psi(T, t_J) \right)^{-(r+2J)} + \delta_{J>0} \cdot a \cdot \left(\alpha + \sum_{j=1}^{J} \psi(t_j, t_{j-1}) \right)^{-(r+2J)} \right]$$

Appendix B

Figure B.1. Distribution of the Number of Admissions per Patient with CHF

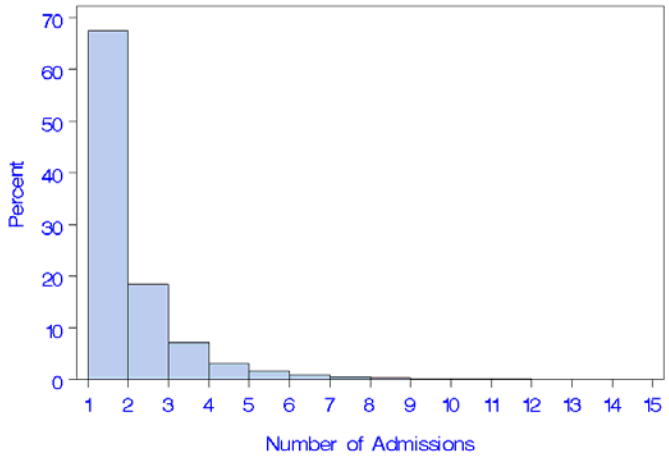

Table B.1. Sample Summary

Length of Admissions	4 years
Total Number of Observations	65,188
Number of Patients	40,983
Number of Hospitals	67
Number of Patients with Multiple Hospitalizations	12,211
Number of patients returning within 30 days from the first visit	2,885

Table B.2. Health Information Technology Factors

Factor	Application	Factor Loading
Administrative IT Systems	Accounts Payable	0.88
	Patient Billing	0.88
	ADT/Registration	0.88
	Chart Deficiency	0.79
	Chart Tracking/Locator	0.79
	General Ledger	0.74
Clinical IT Systems	Operating Room (Surgery) - Post-Operative	0.84
	Operating Room (Surgery) - Peri-Operative	0.83
	Operating Room (Surgery) - Pre-Operative	0.79
	Electronic Medication Administration Record (EMAR)	0.76
	Computerized Practitioner Order Entry (CPOE)	0.70
	Respiratory Care Information System	0.69
	Business Intelligence	0.68
Cardiology IT Systems	Cardiology - Nuclear Cardiology	0.81
	Cardiology - Cath Lab	0.80
	Cardiology - Intravascular Ultrasound	0.80
	Cardiology - CT (Computerized Tomography)	0.79
	Cardiology - Echocardiology	0.77

Appendix C: Lift Table

Our prediction outcomes are measured using a *lift table*, as shown in Table C.1, for the "horserace experiment" where we compare the predictive performance of six models: Random, Baseline Logit, NBD hurdle, EG Hurdle, BG/EG (without hurdle), and our proposed BG/EG Hurdle model. The probability of readmission for each patient is derived from the respective models.[1] If the model is random, then the top *x*% (column 1) of the data is expected to contain *x*% of the readmissions (column 2). If a model outperforms the random model, the top *x*% of sorted customers should capture more than *x*% of actual readmissions, i.e., a lift over the random model.

We observe that the BG/EG Hurdle model exhibits better predictive performance compared to other benchmarking models as a whole, followed by the BG/EG without hurdle, NBD hurdle, logit, EG hurdle and lastly random model, in that order. For example, at the top 5%, the baseline logit model captures 4.43% of the actual readmitted cases, while our model achieves 8.73%. On average, the BG/EG Hurdle model outperforms the logit model by 19.67% in terms of improvement in correctly predicting readmission cases, and outperforms the random and baseline models across all patient segments. Such improvements in prediction performance over the baseline model are substantial, especially if we focus on the top 25% of the readmitted patients where the improvement reaches 27.29%. In terms of prediction performance, the next best comparable model to the BG/EG Hurdle model is the BG/EG model without hurdle. Overall, BG/EG-Hurdle model outperforms BG/EG without hurdle by an average of 3.78%. If we focus on the top 25% of the readmitted patients, the improvement in predictive performance of the BG/EG-Hurdle model over the BG/EG without hurdle model is 4.45%.

[1] We do not present the proportional hazard model because of its low predictive performance. In most cases, the proportional hazard model performs even worse than the random model.

Table C.1. Lift Table: Comparison of Predictive Performance across Multiple Models

Patient segment	Random	Baseline Logit	NBD-Hurdle	EG-Hurdle	BG/EG (no hurdle)	BG/EG-Hurdle
5%	5%	4.43%	6.64%	6.09%	6.86%	8.73%
10%	10%	9.71%	12.65%	11.41%	13.97%	15.20%
15%	15%	15.80%	18.27%	16.95%	20.49%	22.49%
20%	20%	21.51%	25.00%	22.23%	26.02%	28.75%
25%	25%	27.47%	30.88%	28.15%	33.48%	34.97%
30%	30%	33.05%	36.54%	34.07%	40.29%	40.80%
35%	35%	39.14%	42.97%	39.52%	46.34%	47.87%
40%	40%	45.78%	49.70%	44.46%	53.41%	55.07%
45%	45%	51.75%	55.66%	49.28%	59.54%	58.94%
50%	50%	56.43%	60.52%	54.47%	64.22%	63.33%
55%	55%	61.93%	65.46%	58.69%	68.65%	67.29%
60%	60%	66.82%	69.80%	63.03%	72.44%	71.89%
65%	65%	70.83%	74.15%	67.29%	76.06%	76.83%
70%	70%	76.02%	78.24%	71.38%	79.34%	80.71%
75%	75%	80.71%	82.71%	76.36%	82.62%	84.54%
80%	80%	84.41%	86.41%	82.20%	85.90%	88.03%
85%	85%	88.88%	89.95%	87.05%	89.05%	91.40%
90%	90%	92.89%	93.14%	91.23%	93.02%	93.91%
95%	95%	97.02%	96.47%	95.57%	96.47%	97.40%
100%	100%	100.00%	100.00%	100.00%	100.00%	100.00%

Appendix D: Estimation Results for All-period Readmissions Data

Table D.1. Baseline All-period Readmissions

	Logit Model			Proportional Hazard Model		
Variable	**Parameter Estimate**	**Standard Error**	**Odds Ratio**	**Parameter Estimate**	**Standard Error**	**Hazard Ratio**
Intercept	-2.644	(0.504)***	0.071			
Gender: Female	-0.027	(0.011)***	0.973	-0.041	(0.013)***	0.96
Race: American Indian/Eskimo	-0.725	(0.296)***	0.484	-0.368	(0.219)*	0.692
Race: Asian or Pacific Islander	0.213	(0.108)**	1.237	-0.054	(0.06)	0.948
Race: Black	0.464	(0.079)***	1.591	0.192	(0.016)***	1.212
Race: Other	-0.143	(0.086)*	0.867	-0.063	(0.0293)**	0.939
log(disch_age)	1.352	(0.251)***	3.864	1.272	(0.202)***	3.568
log(disch_age)^2	-0.217	(0.034)***	0.805	-0.205	(0.026)***	0.815
Administrative IT	-0.303	(0.056)***	0.739	-0.055	(0.035)*	0.947
Clinical IT	0.074	(0.044)*	1.077	0.050	(0.026)**	1.051
Cardiology IT	-0.207	(0.033)***	0.813	-0.059	(0.019)***	0.943
Number of Procedures	-0.080	(0.007)***	0.924	-0.069	(0.004)***	0.934
log (LOS)	0.129	(0.017)***	1.138	0.065	(0.009)***	1.067
Payer Type: Medicare	0.265	(0.047)***	1.303	0.117	(0.025)***	1.124
Payer Type: Medicaid	0.260	(0.051)***	1.298	0.171	(0.026)***	1.186
Payer Type: Private	0.131	(0.044)***	1.14	0.038	(0.024)*	1.039
Payer Type: Other	0.198	(0.114)*	1.22	0.077	(0.062)	1.08
Admission Type: Emergency	0.107	(0.024)***	1.113	0.089	(0.014)***	1.094
Risk Mortality						
Level 2	0.165	(0.034)***	1.179	0.123	(0.019)***	1.131
Level 3	-0.040	(0.039)	0.961	0.029	(0.023)	1.029
Level 4	-0.337	(0.055)***	0.714	-0.028	(0.033)	0.972
Comorbidities						
DiabetesMellitus	0.205	(0.022)***	1.228	0.086	(0.013)***	1.089
Hypertension	-0.048	(0.026)*	0.953	-0.017	(0.016)	0.984
Periphery Vascular	0.044	(0.036)	1.045	0.019	(0.021)	1.019
Chronic Pulmonary	0.049	(0.023)**	1.051	0.009	(0.013)	1.008
Renal Failure	0.163	(0.029)***	1.177	0.053	(0.017)***	1.055
Anemia	-0.055	(0.025)**	0.946	-0.027	(0.014)**	0.973
Alcohol Abuse	-0.038	(0.092)	0.963	0.042	(0.047)	1.042
Drug Abuse	0.515	(0.073)***	1.673	0.256	(0.030)***	1.291
Ischemic Disease	0.288	(0.023)***	1.334	0.162	(0.013)***	1.176
Number of Beds	0.0002	(0.0001)***	1	0.0000	(0.00003)*	1
Tch_Hosp.	0.153	(0.030)***	1.166	0.081	(0.017)***	1.085
CMI	-0.446	(0.053)***	0.64	-0.259	(0.032)***	0.772
-2 Log L	50400.78			478118.8		
AIC	50466.78			478182.8		

Table D.2. All-period BG/EG Hurdle Estimation Results

	Logit Hurdle			BG/EG Model		
Variable	**Parameter Estimate**	**Standard Error**	**Odds Ratio**	**Parameter Estimate**	**Standard Error**	**Hazard Ratio**
Intercept	-3.399	(0.295)***	0.033			
Gender: Female	-0.081	(0.019)***	0.922	-0.105	(0.011)***	0.894
Black	0.392	(0.036)***	1.480	0.076	(0.023)***	1.087
Asian or Pacific Islander	0.099	(0.021)***	1.105	-0.010	(0.002)***	0.989
Other	0.099	(0.015)***	1.105	0.001	(0.001)	1.001
log(disch_age)	0.517	(0.099)***	1.678	1.414	(0.037)***	4.400
log(disch_age)^2	-0.036	(0.013)***	0.964	-0.293	(0.010)***	0.616
Health IT						
Administrative IT	-0.195	(0.049)***	0.823	-0.298	(0.059)***	0.740
Clinical IT	-0.107	(0.014)***	0.898	-0.098	(0.050)*	0.901
Cardiology IT	-0.284	(0.032)***	0.753	-0.315	(0.023)***	0.705
Number of Procedures	-0.088	(0.007)***	0.916	0.012	(0.002)***	1.012
log (LOS)	0.107	(0.009)***	1.113	-0.006	(0.005)	0.994
Patient Stickiness				-0.151	(0.031)***	0.870
Payer Type: Medicare	0.102	(0.015)***	1.107	-0.304	(0.028)***	0.727
Payer Type: Medicaid	0.142	(0.029)***	1.152	-0.156	(0.011)***	0.839
Payer Type: Private	0.105	(0.022)***	1.110	0.194	(0.029)***	1.236
Payer Type: Other	0.105	(0.027)***	1.111	0.055	(0.008)***	1.063
Admission Type: Emergency	0.116	(0.021)***	1.123	-0.151	(0.016)***	0.870
Risk Mortality						
Level 2	0.306	(0.052)***	1.358	0.020	(0.003)***	1.022
Level 3	0.499	(0.046)***	1.648	-0.073	(0.035)**	0.923
Level 4	0.492	(0.044)***	1.635	-0.082	(0.021)***	0.912
Comorbidities						
DiabetesMellitus	0.024	(0.002)***	1.025	-0.224	(0.017)***	0.786
Hypertension	-0.082	(0.021)***	0.921	0.209	(0.018)***	1.251
Periphery Vascular	0.020	(0.002)***	1.021	0.021	(0.004)***	1.024
Chronic Pulmonary	-0.013	(0.002)***	0.987	0.074	(0.023)***	1.084
Renal Failure	0.112	(0.025)***	1.119	0.097	(0.016)***	1.111
Anemia	-0.022	(0.008)***	0.978	-0.052	(0.025)**	0.945
Alcohol Abuse	-0.312	(0.053)***	0.732	-0.031	(0.009)***	0.966
Drug Abuse	0.706	(0.077)***	2.026	0.115	(0.051)**	1.137
Ischemic Disease	0.185	(0.035)***	1.203	-0.013	(0.007)*	0.987
Number of Beds	0.026	(0.014)*	1.026	-0.065	(0.005)***	0.970
Teaching Hosp.	0.193	(0.044)***	1.213	0.149	(0.019)***	1.172
CMI	-0.297	(0.107)***	0.743	-0.106	(0.030)***	0.906
Previous Discharge to Home/Self				0.051	(0.007)***	1.055
Previous ER reference				0.011	(0.008)	1.011

r	2.867	mu	0.01836
alpha	129.970	sigma	0.11088
a	5.798	log-likelihood	-200082.2
b	6.100	AIC	400304.41