Temporal Evaluation of Risk Factors for Acute Myocardial Infarction Readmissions

Gregor Stiglic
Faculty of Health Sciences
University of Maribor
Maribor, Slovenia
gregor.stiglic@um.si

Adam Davey
Department of Public Health
Temple University
Philadelphia, PA, USA
adavey@temple.edu

Zoran Obradovic
Center for Data Analytics and Biomedical Informatics
Temple University
Philadelphia, PA, USA
zoran.obradovic@temple.edu

***Abstract*—Risk-adjusted 30-day readmission rates for specific diagnoses including myocardial infarction are now used to index hospital reimbursement rates, making prediction of this outcome particularly salient. In order to consider how the importance of various predictors may be changing over time, we apply a modified prequential evaluation technique with an extended training set to this classification problem. Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression using cyclical coordinate descent was used for classification. This paper proposes a bootstrapping based approach to evaluation of sparse coefficients in large sparse datasets with binary and numerical features. It was evaluated on an eight-year dataset of hospital discharge records of myocardial infarction patients consisting of 312,309 discharge records. Results indicate diagnoses (clustered around related disease systems) and length of stay are the most important positive predictors, whereas procedures and diagnoses correcting for small groups of patients and total charges are more important among negative predictors. Temporal comparisons tend to suggest that the importance of features themselves is changing, rather than their prevalence.**

Keywords*— *acute myocardial infarction; hospital readmission classification; sparse logistic regression

I. Introduction

Beginning October 1, 2012, under section 3025 of the Affordable Care Act, hospital reimbursements are now based upon performance relative to preventable 30-day Medicare hospital readmission rates compared with hospitals having similar predicted risk profiles. Initially, readmission rates are tracked for three specific diagnoses: acute myocardial infarction (MI), congestive heart failure (HF), and pneumonia (PN). This change in the structure of Medicare reimbursements places increasing importance on the ability of health care providers to identify predictors of 30-day hospital readmissions as well as to identify characteristics of individuals and providers associated with above-average levels of readmission risk. Under-performing hospitals will see reduction of up to 1% in Medicare base reimbursements for services related to all diagnostic-related groups (DRGs). In 2010, these targets would have placed half of all hospitals in the under-performing group. Future discharge diagnoses to be considered include COPD, asthma, elective surgeries, and vascular procedures.

For comparative purposes, in the United Kingdom, approximately one-third of all hospital admissions are unplanned with annual cost of £11B [1]. In the United States, some 18% of Medicare patients were readmitted within 30 days with approximately 13% of these readmissions considered to be avoidable with associated costs of $12 billion [2]. In the United States, the Medicare Payment Advisory Commission [3] identified conditions and procedures that account for approximately 30% of 7 potentially preventable readmissions. These included COPD, pneumonia, AMI, CABG, percutaneous transluminal coronary angioplasty and other vascular procedures.

Despite considerable interest in this topic, the accuracy of predictive models for 30-day hospital readmission is not particularly strong. Horwitz et al. [4], for example, found in-sample prediction by area under the curve of 0.61, 0.63, and 0.61 for MI, HF, and PN, respectively using Medicare claims data. More recently, focusing only on MI readmissions, Krumholz et al. [5] used 2006 Medicare claims data to compare models relying on claims data versus the combination of claims data and medical record data. These authors found high agreement. (r=0.98), but their overall model had an area under the curve of 0.63.

We aim to identify sparse models predicting 30-day hospital readmission risk with an emphasis on how the predictors change over time. In contrast to most previous work in this area, we use multiple years of medical record data and emphasize out-of-sample prediction via cross-validation.

II. Data and Methods

A. Data

Our analyses relied on data from the discharge data from California, State Inpatient Databases (SID), Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality [6]. The SID is a component of the HCUP, a partnership between federal and state governments

978-0-7695-5089-3/13 $26.00 © 2013 IEEE
DOI 10.1109/ICHI.2013.87

and industry, tracking all hospital admissions at the individual level. We included all data from January 2003 through December 2011. Patients were excluded from the analysis if they did not have a diagnosis of MI, HF, or PN, if they died prior to discharge, were discharged on the same day as admission, were transferred to another institution, or were missing data on the unique patient identifier or sex. After pre-processing, we obtained the final dataset containing 312,309 discharge records with 59,962 positive (readmitted within 30 days) and 252,347 negative records.

Two sets of features were used in our experiments. Clinical Classifications Software (CCS) for ICD9-CM coding developed as part of the HCUP was used to construct the first set of features. CCS codes aim to cluster patient diagnoses and procedures into a manageable number of clinically meaningful categories (343 diagnosis and 296 procedure codes). An initial set of features (CCS) consisted of 510 features (extended to 561 after dummy variable coding). The second set included CCS and 3-digit ICD9 features (CCS+ICD9) and consisted of 1630 features (1681 after dummy coding).

Both sets included 21 additional features like sex, age, month of admission, length of stay, total charges in USD, etc. Three numerical features (age, length of stay and total charges in USD) were also log transformed.

B. *Classification*

Our aim was to identify a small set of predictors in each epoch for ease of model interpretation and temporal comparisons (i.e., we optimize for model sparseness and interpretability rather than overall accuracy or stability of prediction). Thus, to reduce the number of features, we use Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression [7] using cyclical coordinate descent, computed along a regularization path as proposed by Friedman et al. [8]. An implementation in R language (glmnet package) was used to handle large datasets with sparse features.

III. EXPERIMENTAL SETUP

A. *Evaluation*

To evaluate the relative importance of features in the readmission classification problem, we used a modified prequential evaluation (PE) technique [9]. Prequential evaluation is based on a very simple one-step-ahead cumulative prediction error described by Wagenmakers et al. [10]. The original PE approach is based on the following steps:

1. Split data in temporally aligned batches (e.g. months, days, hours, …).
2. Use first i-1 batches to build the predictive model.
3. Use batch i to test the model from step 2.
4. Increase i and repeat steps 2 and 3 for i = n…m, where m is a number of available batches and n represents a minimal number of batches to build an efficient model.

We used an extended training set of 12 instead of just 1 month for evaluation to produce more stable evaluation results at the cost of multiple inclusions of the same record in the evaluation phase. In case of readmission problems with large sparse matrices of data it is important to control the stability of results by testing the built models on larger sets of records. Area under ROC Curve (AUC) was averaged over all one-step-ahead (12 months) evaluations using a sliding 12 month window for training and testing. We can therefore describe the modified evaluation approach with the following steps:

1. Use 12 months of data preceding month i to build the predictive model.
2. Use months i to i+11 to evaluate the model from step 1.
3. Increase i and repeat steps 1 and 2 for i = 13…m-12, where m is a number of available months of data.

B. *Variable Importance*

Generally L1-penalized logistic regression (LASSO) allows very effective feature selection and achieves high classification performance at the same time. However, we had to deal with some specific characteristics of our data (large number of samples, large number of very sparse features: diagnoses/ groups of diagnoses). All diagnosis-related features are binary and therefore do not present a problem in the feature selection process. However, we must accommodate features on other scales (e.g., age, sex, length-of-stay). Feature normalization makes little sense for binary variables representing the presence of absence of diagnoses and procedures. Instead, we rank features by the number of their non-zero coefficients in all 84 models built on 12-month training windows, presenting results for positive and negative non-zero coefficients separately. Initially, we replace all positive coefficients with 1 and all negative coefficients with -1. In the next step we sum up the values for all 84 models. This way, we reduce the importance of the features that frequently switch between positive and negative coefficients. Even more robust ranking of the features is obtained by averaging the counts of non-zero coefficients over 100 bootstrap samples used for training in each 12-month window.

IV. RESULTS

To evaluate the prediction performance of the sparse logistic regression model, we observed the Area under ROC curve (AUC) performance metric of four different models (CCS only, CCD + OPT, ICD9 only, ICD9 + OPT). Initially, we compared the CCS and CCS+ICD9 sets of features using the largest value of lambda (the multiplier on the coefficient norm in LASSO) with training data such that error rate was within 1 standard error of the minimum selected using 5-fold cross validation. The comparison was repeated for the optimal lambda value using 5-fold cross validation (CCS OPT and CCS+ICD9 OPT). Figure 1 presents results from all four models demonstrating differences between the use of different lambda and feature sets. Using optimal lambda value the mean AUC was higher than stricter lambda by 0.004 in both CCS and ICD9+CCS cases (paired t-test $p < 0.001$). Mean number of selected features for CCS model was 31.3 and 36.1 when

TABLE I. BOOTSTRAPED ESTIMATIONS OF FEATURE IMPORTANCE OVER 84 EVALUATION WINDOWS, FIRST 42 AND LAST 42 MONTHS

Positive coefficients	**Months**				**Negative coefficients**	**Months**			
Attribute	***All***	***1-42***	***43-84***	***Δ***	***Attribute***	***All***	***1-42***	***43-84***	***Δ***
Congestive heart failure	84.00	42.00	42.00	0.00	Routine discharge	80.87	39.82	41.05	1.23
Atherosclerosis	83.99	42.00	41.99	-0.01	Tracheostomy	79.37	41.16	38.21	-2.95
Diabetes mellitus w complications	83.76	42.00	41.76	-0.24	Race missing	77.60	37.74	39.86	2.12
Hemodialysis	83.66	41.66	42.00	0.34	Other non-OR cardiovascular procedure	75.12	37.07	38.05	0.98
Chronic obstructive pulmonary disease	81.56	39.56	42.00	2.44	Micropolitan adjacent to large metro	66.11	30.88	35.23	4.35
Diabetes mellitus w/o complications	81.02	42.00	39.02	-2.98	Coronary artery bypass grafting	66.02	33.35	32.67	-0.68
Chronic kidney disease	80.51	41.98	38.53	-3.45	Acute cerebrovascular disease	64.33	33.76	30.57	-3.19
Length of stay	79.83	39.47	40.36	0.89	Medical exam	61.45	36.49	24.96	-11.53
Peripheral atherosclerosis	75.74	38.15	37.59	-0.56	Discharge missing	55.65	35.83	19.82	-16.01
Blood transfusion	74.82	35.72	39.10	3.38	Cardiac catheterization	54.06	31.82	22.24	-9.58
Intestinal infection	74.54	36.55	37.99	1.44	Other aftercare	52.78	30.51	22.27	-8.24
Black (race)	71.43	40.04	31.39	-8.65	Nutritional/Endocrine/Metabolic disorders	51.56	31.31	20.25	-11.06
No "Do Not Resuscitate" order	68.40	32.79	35.61	2.82	Discharged to short-term hospital	50.34	13.87	36.47	22.60
Anemia	67.38	25.38	42.00	16.62	Total charges	49.76	30.63	19.13	-11.50
Hypertension w complications	66.86	40.98	25.88	-15.10	Coma	49.45	27.87	21.58	-6.29
Gangrene	64.41	36.43	27.98	-8.45	Intestinal obstruction	49.14	14.26	34.88	20.62
Heart valve disorders	61.47	39.63	21.84	-17.79	Aspiration pneumonitis	47.49	14.84	32.65	17.81
Dysrhythmias	61.46	37.55	23.91	-13.64	Syncope	44.36	18.57	16.94	-1.63
Female	60.05	39.19	20.86	-18.33	Nutritional deficiencies	43.43	34.80	8.63	-26.17
Spondylosis	56.15	21.52	34.63	13.11	Urinary tract infection	43.36	19.20	24.16	4.96

additional 1120 binary ICD9 features were added. For optimal lambda we can observe the mean number of selected features of 104.6 and 138.1 for CCS and ICD9 models.

From Figure 1 it can also be observed that all models improved their performance over time. To identify the reasons for the lift of performance, we checked the distribution of the target class over time. We could not identify any correlations of the ratio between positive and negative class with the classification performance in any time period. Exploring the reasons for a significant lift in the second half of the observed time interval further, we compared the variable importance in both time periods.

We used 100 bootstrapped samples for each 12-month window to estimate the feature importance. Table 1 presents the results of feature importance based on CCS set of features and is divided in two parts – features with positive and negative influence on the readmission. In all feature importance analyses we use only 84 months with first and last year of data excluded to narrow down the performance shift period. Results include top 20 features ordered by feature importance score for the whole observed period for positive and negative coefficients. Initial column represents information on importance over the whole period of 84 evaluation months, while next two columns present the comparison of first versus last 42 months to summarize changes in variable importance over two longer periods of time. The last column contains the information on difference in average inclusion of the feature for both observed periods.

V. DISCUSSION AND CONCLUSIONS

A modified approach to prequential evaluation was useful for identifying a stable set of features associated, positively or negatively, with 30-day risk of hospitalization for acute myocardial infarction, one of the outcomes used to establish hospital Medicare reimbursement rates under the Affordable Care Act. Features associated with higher risk of readmission were mostly comorbid diagnoses in related organ systems, notably diabetes mellitus, chronic kidney disease, and chronic obstructive pulmonary disease, and other cardiovascular diseases. Longer hospital stays also predicted readmission.

In contrast, coefficients associated with lower risk of readmission were more likely to reflect procedures and diagnoses that tend to correct for relatively smaller groups of individuals (e.g., "other" categories, tracheostomy, cardiac catheterization, aspiration pneumonia, intestinal obstructions, nutritional deficiencies, UTIs) or be related to care

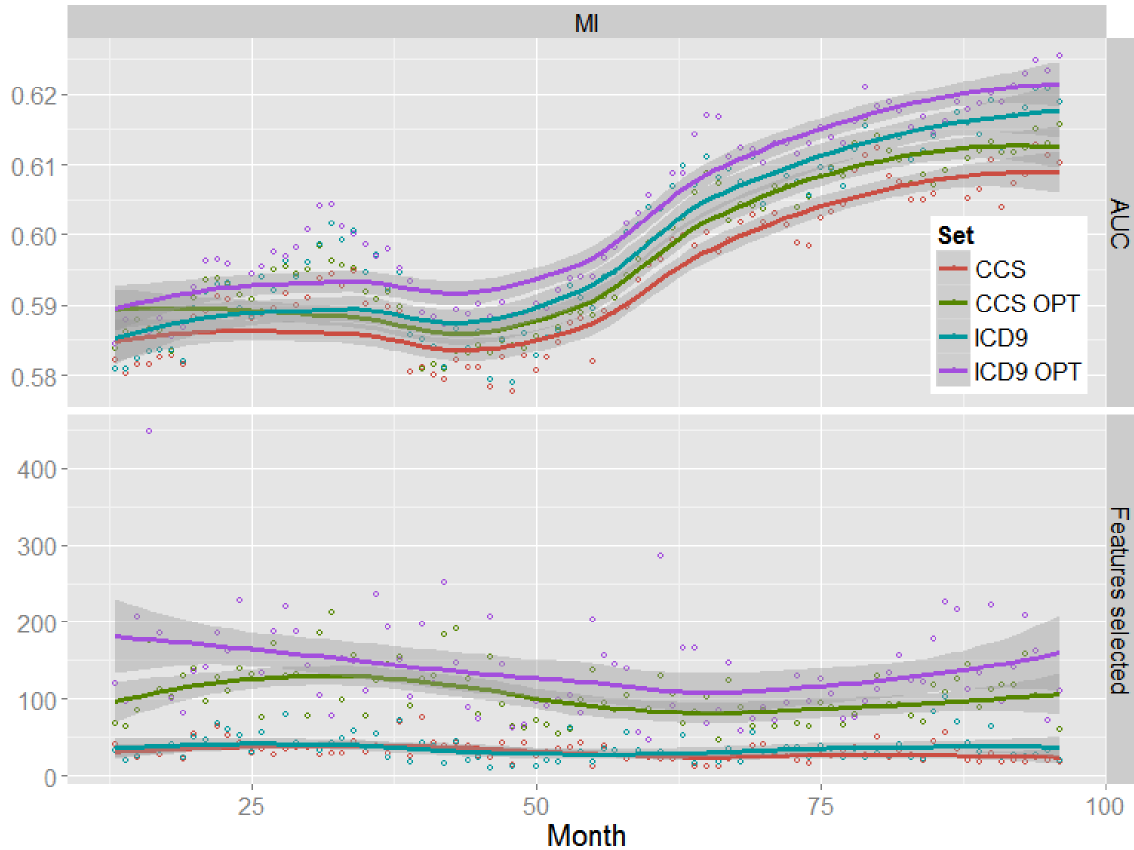

Fig. 1. Comparison of AUC for four different models using only CCS or combination of 3-digit ICD9 and CCS codes. Additionally, we compare the optimal lambda (OPT) to more selective 1-SE lambda value for sparse logistic regression.

characteristics (medical care, aftercare, discharge to a short-stay hospital, location of hospital).

Total cost of care predicted lower risk of readmission overall, but not in individual 42-month care windows. It is also worth noting that the relative importance of features is higher for positive than negative coefficients, on average. This might suggest that current models of hospital care devote more attention to risk factors for readmission than protective factors.

In the first versus last 42 month comparison, we identified some features with significant feature importance changes. Especially in the negative coefficients group, where we identified a few features where importance was much higher in the last 42 months compared to the initial 42 months. Examples of such features include discharges to short-term hospitals, intestinal obstruction or aspiration pneumonitis. On the other side, there are some features with a decrease in importance when comparing the last 42 months with the first 42. Such features include nutritional deficiencies or an indicator of missing value for patient discharge type. Interestingly, "nutritional deficiencies" is one of the features with relatively high prevalence that significantly increases in the last 42 months (Table 2). Additionally there are three features we would like to point out and are not included in the Table 1 due to their lower overall importance score. One of them is CCS code for Diagnostic endocrine procedures that was included in the model as a negative coefficient (average score of 33.84) in the first 42 months and turned up as a mostly positive coefficient in models for the last 42 months (average score of 6.35). This was the only case where we could identify a shift from a negative into positive coefficient or vice versa. Another important feature might be "Other CT scan", where negative importance dropped from 36.03 to 4.05 hinting at a lower predictive power of the CT scans in more recent period.

We also identified a feature representing diagnosis "Open wounds of head, neck and trunk" that was used with a negative coefficient only rarely in the first 42 months (average importance score of 1.96), but was significantly more important in the last 42 months (average negative coefficient score of 33.32). Direct interpretation of this feature importance shift points at high usability of this feature in the last 42 months to identify patients that will not return in the next 30 days.

In the first group of features comparing positive coefficients in two time periods, we were not able to identify many major differences in importance. The most notable

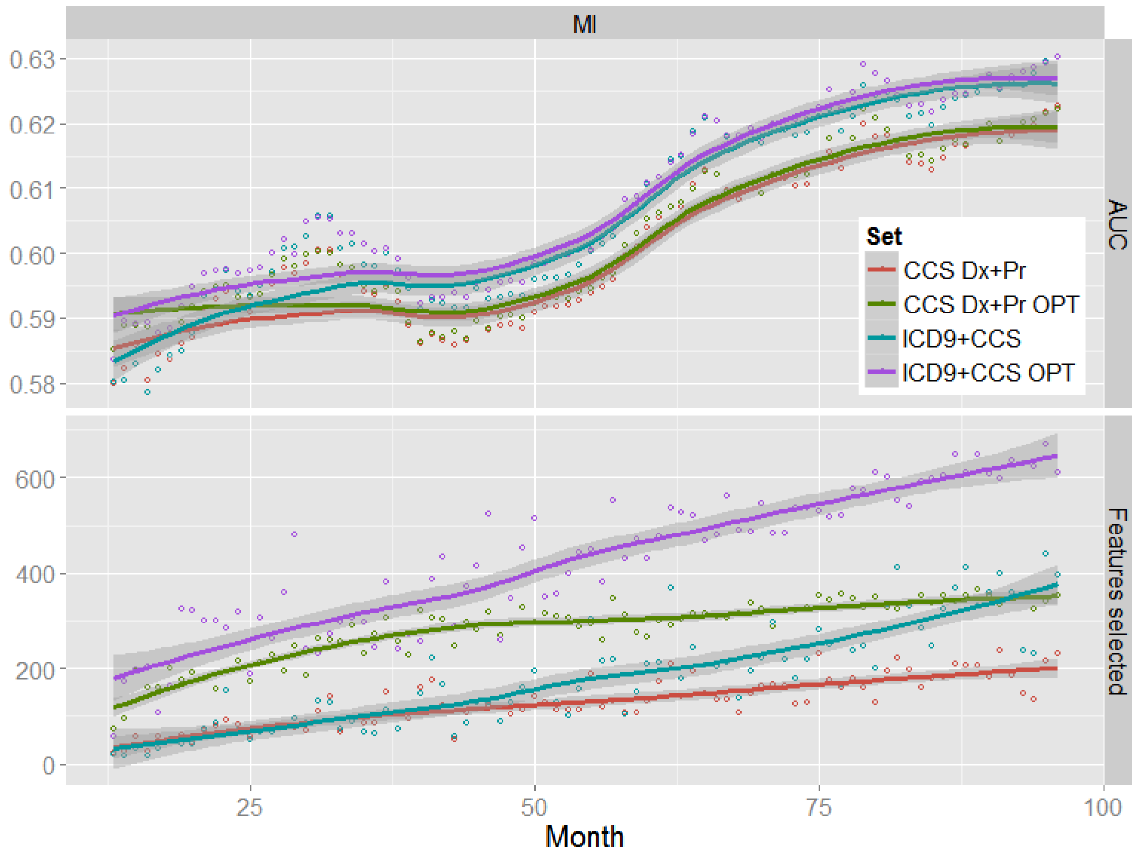

Fig. 2. Comparison of AUC for four different models using only CCS or combination of 3-digit ICD9 and CCS codes. Additionally, we compare the optimal lambda (OPT) to more selective 1-SE lambda value for sparse logistic regression.

change in importance with the increasing trend was identified for anemia, while the negative trend includes features like sex, heart valve disorders and hypertension with complications.

It has to be noted that Table 1 represents only CCS aggregated diagnosis and procedure codes that tend to cluster different number of diagnoses or procedures into groups. Therefore, it would be necessary to analyze more specific ICD9-CM codes to obtain a better insight into the underlying reasons for importance variability.

Finally, we conducted another experiment with cumulative training set, where training windows were not 12-month windows anymore, but ranged from 12 to 96 months of data. This way the last model used almost 300,000 records for training. Figure 2 presents the same results as Figure 1, but with cumulative training set. The most noticeable difference can be observed in the number of selected features when the number of records used is getting higher. Optimal lambda does not play an important role once the number of records gets higher.

Overall, our final models achieved out-of-sample AUCs over 0.62. Additionally, in cumulative setting we were able to achieve out-of-sample AUCs of 0.63. Thus, these models perform on par with other research using gold-standard Medicare claims data [5]. However, our models rely on more features and a longer temporal window to achieve these values.

TABLE II. PREVALENCE OF FEATURES WITH THE HIGHEST CHANGE IN VARIABLE IMPORTANCE

Feature name	First 42 months		Last 42 months	
	≤30 days	*>30 days*	*≤30 days*	*>30 days*
Diagnostic endocrine procedures	0.000%	0.014%	0.027%	0.014%
Other CT scan	0.250%	0.331%	0.261%	0.266%
Open wounds head/neck/trunk	0.358%	0.382%	0.386%	0.500%
Nutritional deficiencies	3.899%	3.627%	9.895%	8.192%
Discharged to short-term hospital	5.942%	5.808%	3.845%	4.660%
Intestinal obstruction	1.761%	1.607%	1.934%	1.748%

At the same time, our models point to several important extensions with potential to achieve improved prediction. For example, the current models do not yet include hospital identifiers or provider characteristics, which are likely to increase the predictive power of models considerably.

Analyses of related outcomes with a different set of observations indicate that a considerable proportion of the systematic variance is associated with hospital identifiers, making the ability to identify better and poorer performing hospitals and important goal.

Acute myocardial infarction is not a homogeneous diagnosis. The set of positive predictors suggests that additional diagnoses related to cardiovascular disease are important in determining the probability of 30-day readmissions. It will be important to extend these analyses to evaluate prediction within subpopulations as a function of the specific comorbidity constellations (for example, diabetes mellitus, chronic kidney disease, congestive heart failure, chronic obstructive pulmonary disease). It will be important to identify how the importance of different subsets of predictors affects readmission risk as a function of these comorbidities, and how these predictors have changed over time.

Related to this, the features associated with lower risk of hospital readmissions are especially interesting. Medical examinations and after care are both associated with lower readmission probability, as is discharge to a short-stay hospital, suggesting that wider implementation of these procedures may be useful in preventing readmissions. Unfortunately, as shown in Table 2, while the prevalence of identifying nutritional deficiencies increases from the first to the second time periods considered, discharges to short-stay hospitals appears is becoming less common. Prevalence of most other features appears quite stable over time periods, suggesting that the importance of these features themselves is changing, rather than their prevalence. Further preventing readmissions may require more intensive services during the original hospital stay as evidenced by the negative coefficients for total charges.

Although they track all California hospital admission for the time period considered, our data source lack some important variables that are available from other sources, particularly those associated with established treatment guidelines for Acute Myocardial Infarction [11], many of which are pharmacological and/or laboratory based. In future work, we plan to extend these models using medical claims data in order to capture information about laboratory tests and prescription medications.

ACKNOWLEDGMENT

This research was partially supported by the Slovenian Research Agency through a bilateral project grant ARRS-BI-US-JR/2011/16 and in part by DARPA Grant FA9550-12-1-0406 negotiated by AFOSR.

REFERENCES

[1] P. E. Cotter, V. K. Bhalla, S. J. Wallis, R. W. S. Biram, “Predicting readmissions: poor performance of the LACE index in an older UK population,” Age and Ageing, vol. 41, no. 6, 784-789, 2012.

[2] Report on Medicare Compliance, “CMS targets readmission trhough payment, audits; ‘Coaching’ model reduces rates,” Report on Medicare Compliance, vol. 17, no. 24, 2008.

[3] Medicare Payment Advisory Commission (MedPAC), Report to Congress: Promoting Greater Efficiency in Medicare, 2007. Available at: http://www.medpac.gov/documents/Jun07_EntireReport.pdf.

[4] L. Horwitz, C. Partovian, Z. Lin, J. Herrin, J. Grady, M. Conover et al., “Hospital-wide (all-condition) 30-day risk-standardized readmission measure,” 2011. Available at: http://www.cms.gov/MMS/Downloads/MMSHospital-WideAllConditionReadmisionRate.pdf

[5] H. M. Krumholz, Z. Lin, E. E. Drye, M. M. Desai, L. F. Han, M. T. Rapp, et al. “An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction,” Circulation: Cardiovascular Quality and Outcomes 4, no. 2, pp. 243-252, 2011.

[6] HCUP State Inpatient Databases (SID), Healthcare Cost and Utilization Project (HCUP). 2003-2011. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp.

[7] R. Tibshirani, “Regression shrinkage and selection via the lasso,” J. Royal. Statist. Soc B., Vol. 58, No. 1, pp. 267-288, 1996.

[8] J. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent." Journal of statistical software, vol. 33, no. 1, 2010.

[9] A. P. Dawid, “iPresent position and potential developments: Some personal views,” Statistical Theory. The prequential approach, Journal of the Royal Statistical Society, Series A, vol. 147, pp. 278-292, 1984.

[10] E. J. Wagenmakers, P. Grünwald, and M. Steyvers, “Accumulative prediction error and the selection of time series models,” Journal of Mathematical Psychology, vol. 50, no. 2, pp. 149-166, 2006.

[11] E. M. Antman, D. T. Anbe, P. W. Armstrong, E. R. Bates, L. A. Green, M. H. Judith, et al. “ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction—executive summaryA report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 1999 Guidelines for the Management of Patients With Acute Myocardial Infarction) ,” Journal of the American College of Cardiology, vol. 44, no. 3, pp. 671-719, 2004.