Hospital Corners and Wrapping Patients in Markov Blankets

Dusan Ramljak Adam Davey Alexey Uversky Shoumik Roychoudhury
Zoran Obradovic*

Abstract

The American health care system is rife with perverse incentives. Providers are reimbursed for providing more care rather than preventive or higher quality care. Starting in 2012, the Affordable Care Act (ACA) sets Medicare reimbursement rates based on hospital performance for 30-day preventable readmissions relative to expectations using 3 specific target diagnoses (AMI, CHF, and PN). This may have introduced an incentive for hospitals to under-diagnose these illnesses by substituting related diagnoses for which they will not be held accountable. We identify Markov blankets, diagnoses that can shield target diagnoses from the rest of the disease network. Each target diagnosis can be accurately identified and inferred from a small subset of related diagnoses. This work suggests several important directions for evaluating implementation of this component of the ACA. Specifically, this method can be used for problems such as identifying true cases with target diagnoses, estimating the extent of gaming via substitute diagnoses, and also to suggest related sets of diagnoses which, in combination, may provide more stable methods for setting reimbursement rates.

1 Introduction

Under §3025 of the Affordable Care Act (ACA), as of October 1, 2012 hospital reimbursements have been based on performance relative to preventable 30-day Medicare hospital readmission rates observed in hospitals with similar predicted risk profiles. Three specific diagnoses are used to track reimbursement rates: acute myocardial infarction (AMI), congestive heart failure (CHF), and pneumonia (PN). As a direct result of this change in the structure of Medicare reimbursements, there is now more focus on problems such as the ability of health care providers to identify changing predictors of 30-day hospital readmissions [4, 5, 13], as well as to identify characteristics of individuals and providers associated with above-average levels of readmission risk. Hospitals that perform below expectations will see a reduction of up to 1% in Medicare-based reimbursements for services related to all diagnostic-related groups (DRGs). Based on performance levels in 2010, these targets would have placed half of all hospitals in the under-performing group. In coming years, additional diagnoses will be added to the list used to determine reimbursement rates.

Because hospitals cannot be penalized for diagnoses they do not make, physicians are incentivized to choose similar, but distinct, diagnoses for criterion diagnoses. For example, a patient who is admitted to a hospital with AMI may initially be diagnosed as having chest pains or coronary atherosclerosis. If this patient was subsequently readmitted within the following 30 days, this diagnosis could not be used to penalize the hospital for poor performance. Similarly, PN may initially be diagnosed as having acute bronchitis or an upper respiratory infection, and CHF may instead be diagnosed at first as chronic obstructive pulmonary disease. In practice, only those assessed as having the lowest risk of 30-day readmission may be likely to receive the target diagnosis.

However, several specific diagnoses are likely to cooccur with the target diagnosis and some procedures (e.g., angioplasty) may be strongly indicative of a specific underlying true diagnoses (e.g., AMI), serving as good proxy indicators of the true diagnosis. Evidence for changes to clinical practice, diagnoses, and associated procedures in response to changes in reimbursement has been well-documented for more than 30 years e.g., [11] and there is reason to suspect that similar changes are already occurring due to the most recent changes enacted under the ACA. One way of estimating the extent of these changes, and identifying cases that represent the true diagnoses of criterion diagnoses, is considering diagnoses as a set of connected nodes in a graph (connected by aspects such as co-occurrence). The Markov blanket (MB) for a node is the set of nodes that shield it from the rest of the network. Previous studies have shown that knowing the Markov blanket of a diagnosis node is all that is required in order to predict the value of the criterion, either by classification or regression [10, 7, 8]. If the MB of a specific diagnosis can be identified prior to a policy change, it may be used to more accurately identify the set of criterion diagnoses

*Center for Data Analytics and Biomedical Informatics, Temple University Philadelphia, PA 19122 USA {dusan.ramljak, adavey, auversky, shoumik.rc, zoran.obradovic}@temple.edu

following the policy change, which can in turn be used to estimate true cases, as well as the extent of gaming of diagnoses which will occur due to the policy change.

In this paper, we investigate these issues in a series of experiments aimed at answering these important questions: 1) Which diagnoses can be used to approximate the MB of AMI, CHF, and PN? 2) Can a similar procedure be used to identify the MB for diagnoses for which similar policy changes have been enacted in the past, such as sepsis?

Our experiments used discharge data from the California State Inpatient Databases (SID), obtained from the Healthcare Cost and Utilization Project (HCUP) provided by the Agency for Healthcare Research and Quality[1]. The SID is a component of the HCUP, a partnership between federal and state governments and industry, which tracks 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 Medicare or Medicaid as the primary payer and if they were younger than 19 years of age. The final dataset included 16,736,927 discharge records, with the primary set of features used in our experiences being the Clinical Classifications Software (CCS) diagnoses for ICD9-CM. CCS codes, developed as part of the HCUP, are designed to cluster patient diagnoses (hereinafter DX) and procedures (hereinafter PX) into a manageable number of clinically meaningful categories (272 diagnoses and 231 procedure codes).

The rest of the paper is organized as follows. The model that we use in the experiments is described in Section 3. Data and experimental setup are described in Section 4, and the results are discussed in Section 5. We conclude the paper by providing interesting areas of future studies.

2 Related Work

Some prior research has examined the role of comorbid conditions with the aim of identifying longer-term effects and mortality risk with a single target diagnosis in mind. Each of our target diagnoses has been considered in this fashion: AMI [12], CHF [1], PN [15], and sepsis [14]. To the best of our knowledge, ours is the first study concerned with identifying co-occurring diagnoses and procedures that can themselves serve as a proxy indicator of the target diagnosis, something necessary to identify potential instances of hospitals gaming the system to reduce risk exposure.

3 Model Description

In our quest to find a minimum subset of the most informative diagnoses associated with the diagnoses we analyse we start with identifying DXs and procedures that most frequently co-occur with our Target DX. The most frequent co-occurrence by itself is necessary but not sufficient to establish appropriate approximation of the Markov Blanket. Therefore we apply two different methodologies: a heuristic we defined based on PageRank, and an established approximation of the Markov Blanket using Hilbert-Schmidt Independence Criterion (HSIC).

3.1 Approximating the Markov Blanket (MB) using PageRank We are not able to build a Bayesian network since it is not clear which diagnoses, procedures are parents, or children and we therefore opt for a Markov Network where the dependencies are defined by co-occurrences. However, every node that co-occurs with our Target DX is an element of a Markov blanket by definition. Therefore, we need to make a distinction between the nodes.

Having the weights defined by co-occurrences and information from the structure of the network of DXs and PXs that co-occur with our Target DX we could use PageRank value as a criterion to identify important nodes. For a subset of highly important nodes, nodes with highest PageRank we could then say that it represents an approximate Markov Blanket for our Target DX.

3.2 Approximating the Markov Blanket (MB) using HSIC As an established machine learning approach we will adopt a feature selection method based on an efficient approximation of the Markov blanket (MB), which is a set of variables that can shield a certain DX from the rest of diagnoses and procedures [9]. MB-based feature selection process has been shown to result in a theoretically optimal set of features [16]. However, it's computational cost is prohibitive for application to high dimensional Electronic Health Records (EHR) data.

Therefore, instead of relying on conditional independence or network structure learning, we will use HSIC as a measure of dependence among variables in a kernel-induced space. This will allow effective approximation of the MB that consists of multiple dependent features rather than being limited to a single feature. Benefits of using HSIC include: 1) It can detect any dependence between two variable sets with a universal ker-

[1] HCUP State Inpatient Databases (SID), Healthcare Cost and Utilization Project (HCUP). 2003-2011. Agency for Healthcare Research and Quality, Rockville, MD. www.hcupus.ahrq.gov/sidoverview.jsp http://www.ahrq.gov/research/data/hcup/index.html

nel in high dimensional kernel space; 2) It can measure the dependence between both discrete and continuous variables; and 3) It is easy to compute from the kernel matrices without density estimation. Given a set of features, we can check whether the set is the MB (MB_i) of feature F_i.

However, evaluating all subsets of F for this property is prohibitively costly. Therefore, to reduce the search cost we will evaluate some MB candidate subsets. Often, there might not be an exact MB for a given feature, but we can still identify an approximating MB, which largely subsumes the information about this feature, so that we can remove this feature with little useful information lost. In this work we use a simple but effective method to find an approximating MB. The proposed method consists of 3 steps: 1) Identifying MB candidates, 2) Screening MB candidates, and 3) Feature selection.

In particular, we compute a Markov blanket candidate MB_i for a feature F_i such that each of its features F_{MB_i} satisfies: $argmax_{F_C} HSIC(K_{F_C}, K_i), where\ F_C \in B_i - MB_i \cup \{F_{MB_i}\}$

In order to determine whether the approximate candidate MB of feature F_i (referred to as MB_i) can be regarded as an actual approximation of the Markov blanket, we use the following screen test:

$$HSIC(MB_i, C) > HSIC(MB_i \cup F_i, C) \tag{3.1}$$

$$\begin{aligned} & HSIC(MB_i, C) > HSIC(F_i, C) \\ \wedge \quad & HSIC(MB_i, F_i) > HSIC(F_i, C) \end{aligned} \tag{3.2}$$

In this test, C is the target variable, and $HSIC(X, Y)$ is defined as the dependence measure between two variable sets X and Y. Condition 3.1 is satisfied by an irrelevant feature, while Condition 3.2 is satisfied by a redundant feature. By applying a Markov blanket to select the minimum subset of the most informative diagnoses and procedures associated with diagnoses that we analyze, we hope to be able to estimate true prevalence of various diseases. Additionally we would like to provide more robust methods to estimate true hospital readmission rates where intentional under-diagnosis of such sentinel diagnoses is likely.

4 Data Description and Experimental Setup

4.1 Data Description The data that we use in our experiments comes from the HCUP family of databases, and the raw data consists of patient hospital visit records from California's SID in the period from January 2003 up to December 2011. Each record consists of a number of attributes, which are explained in detail on the HCUP website[1]. This database contains more than 35 million (35,844,800) inpatient discharge records over the specified 9 years for 19,319,350 distinct patients in 474 hospitals (436 AHAID identified; about 400 per year). The information is not specific to a group of hospitals, but rather represents the data for the entire state. The database also includes demographic information for each patient (like age, birth year, sex, race), diagnosis (primary and up to 24 secondary), procedures (up to 21), information about hospital stays, and other information (including length of stay, total charges, type of payment and payer, discharge month, and survival information).

[1]`http://www.hcup-us.ahrq.gov/db/state/siddist/sid_multivar.jsp`

In addition to addressing the problem that we are considering, using this data could potentially give us insight into many healthcare problems. For instance, we can explore and correct global trends of Target diseases which intrigues many healthcare practitioners [3, 6].

After excluding patients younger than 19 years of age and patients that did not have Medicare or Medicaid as the primary payer, we were left with 16,736,927 discharge records. Table 1 shows the frequency of each Target DX in the SID data set, while Figure 1 displays this frequency over time.

	Target DX			
	Sepsis	AMI	CHF	PN
Freq.	1,027,088	544,228	2,907,625	1,577,822

Table 1: Each Target DX frequency in the data set

4.2 Experimental Setup We first identified subsets of records containing each Target DX (Sepsis, AMI, CHF, PN). Within each subset, diagnoses and procedures were ranked by pagerank values in a network of co-occurrence with Target DX and networks of co-occurrences without Target DX were generated as well.

As proof of concept for this method, exactly 50 DX and PR (roughly 10%, accounting for approximately 95 of records including each Target DX) were identified. In general, the pagerank value of the first selected DX is approximately 10 times higher than the last DX selected.

10 replicate experiments were performed using a randomly selected set of 1000 examples with each Target DX and a set of 1000 examples where the Target DX does not appear. The goal of these analyses was to identify the subset of DXs that can be used to accurately predict the presence of Target DX in the records. To achieve this goal we identified the approximate MB for each Target DX. Performance was evaluated by comparing classification accuracy using only the DXs in the MB against using the entire set of DXs.

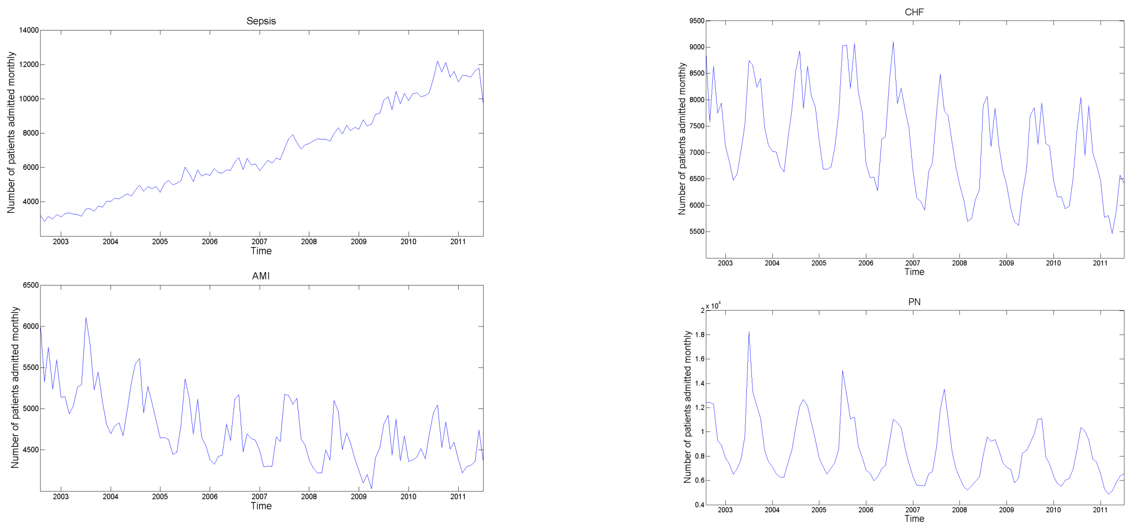

Figure 1: Monthly admission rates in a period 2003 up to 2011 for Target DX (Sepsis, AMI, CHF, PN)

5 Results

Each of the 4 Target DX we examined in this study was fairly prevalent. As shown in Table 1, CHF was most common (17.37%), followed by PN (9.43%), sepsis (6.13%), and AMI (3.25%). There were also considerable seasonal and secular trends in these Target DX. Prevalence of sepsis increased steadily across the study period. The three other Target DX showed gradual decreases in prevalence over time, but also very strong seasonal trends. Table 4 shows the performance of MB and PageRank methods in terms of accuracy, precision, sensitivity, specificity, and F1. Both measures perform well on accuracy and precision for sepsis. For other diagnoses, accuracy is generally higher via MB, but precision is generally higher for PageRank. Excepting sepsis, sensitivity and specificity are higher for PageRank than MB, but F1 values are higher for MB than PageRank.

For consistency with HSIC results (below), based on page ranks, we provide the top 7 DX/PR for sepsis and 5 DX/PR identified for each Target DX (see Table 2). For sepsis, the list included peritonitis, injuries, and shock for diagnoses and tracheostomy, non-cardiac catheterization, ostomy, and intubation. For AMI, no diagnoses were identified, but CABG, angioplasty, coronary thrombolysis, cardiac catheterization, and urinary endoscopy were identified. For CHF, heart valve disorders, carditis, and pulmonary heart disease were identified diagnoses, and heart valve procedures and Swan-Ganz catheterization were the procedures. Finally, for PN, respiratory failure and shock were identified diagnoses and tracheostomy, bronchoscopy, and other respiratory procedures were identified.

As shown in Table 3, the Markov blanket for each Target DX consisted of a small subset of common diagnoses and procedures. For sepsis, the MB included hemorrhagic disorders; intestinal infection; urinary tract infections; skin ulcers; complications of device, implant, or graft; and other injuries as diagnoses and non-cardiac catheterization as a procedure. For AMI, the MB included atherosclerosis and rehabilitation care for diagnoses and endoscopy of the urinary tract, echocardiogram, and other diagnostic procedures. For CHF, the MB included other nervous system disorders, essential hypertension, hypertension with complications, atherosclerosis, and respiratory failure as diagnoses; no procedures were included in the MB. Finally, for PN, the MB included diabetes without complications, hemorrhagic disorders, hypertension with complications, congestive heart failure, and other lower respiratory disease; no procedures were included in the MB.

6 Conclusion

The US healthcare system is rife with opportunities for perverse incentives. Implementation of any new healthcare policy results in changes within healthcare system in order to minimize the adverse consequences of the policy change for healthcare providers. Changes that began in 2012 under the Affordable Care Act can be expected to reduce the number of individuals receiving target diagnoses of AMI, CHF, and PN as healthcare providers move to reduce their exposure to adverse consequences of hospital readmissions.

In this paper, we propose two novel applications to the problem of under-diagnosing, specifically, Markov blankets and page rank. We find that, for each Target DX, and sepsis, a small number of diagnoses and

Sepsis	
Diagnoses	
148	Peritonitis and intestinal abscess
244	Other injuries and cond. due to ext. causes
249	Shock
Procedures	
34	Tracheostomy; temporary and permanent
54	Other vasc. catheteriz.; not heart
73	Ileostomy and other enterostomy
216	Resp. intub. and mech. vent.
AMI	
Diagnoses	
	None
Procedures	
44	Coronary artery bypass graft (CABG)
45	Percutan. translum. coron. angiopl. (PTCA)
46	Coronary thrombolysis
47	Diag. cardiac catheteriz.; coron. arteriogr.
100	Endosc. and endosc. biopsy of the urin. tract
CHF	
Diagnoses	
96	Heart valve disorders
97	Peri-; endo-; and myoc.; cardiomyop. (except TB or STD)
103	Pulmonary heart disease
Procedures	
43	Heart valve procedures
204	Swan-Ganz catheterization for monitoring
PN	
Diagnoses	
131	Resp. failure; insufficiency; arrest (adult)
249	Shock
Procedures	
34	Tracheostomy; temporary and permanent
37	Diag. bronchosc. and biopsy of bronchus
41	Other non-OR therap. proc. on resp. sys.

Table 2: Table showing the diagnoses used to form the PageRank approximated MB for each target DX

Sepsis	
Diagnoses	
62	Coagulation and hemorrhagic disorders
135	Intestinal infection
159	Urinary tract infections
199	Chronic ulcer of skin
237	Complication of device; implant or graft
244	Other injuries and cond. due to ext. causes
Procedures	
54	Other vascular catheterization; not heart
AMI	
Diagnoses	
101	Coron. atheroscl. and other heart dis.
254	Rehab. care; fit. of prosth.; and adj. of devices
Procedures	
100	Endosc. and endosc. biopsy of the urinary tract
193	Diag. ultrasound of heart (echocardiogram)
227	Other diagnostic procedures
CHF	
Diagnoses	
95	Other nervous system disorders
98	Essential hypertension
99	Hypert. with compl. and sec. hypert.
101	Cor. atheroscl. and other heart dis.
131	Resp. fail.; insuffic.; arrest (adult)
Procedures	
	None
PN	
Diagnoses	
49	Diabetes mellitus without complication
62	Coagulation and hemorrhagic disorders
99	Hypert. with compl. and sec. hypertension
108	Congestive heart failure; nonhypertensive
133	Other lower respiratory disease
Procedures	
	None

Table 3: Table showing the diagnoses used to form the HSIC approximation of MB for each target DX

procedures can serve to shield the target diagnosis from the rest of the disease network. Performance using this subset of diagnoses suggests performance that generally quite high for accuracy and precision. Additionally, these diagnoses and procedures often point to clinically meaningful patterns. In general, while nearly all of the selected diagnoses and procedures make sense, with considerable overlap between MB and PageRank methods, it is generally the case that the associations are more directly obvious for selections via PageRank, and somewhat subtler for selections via MB. The results suggest that page rank may be useful for providing an initial screening of potentially useful diagnoses. However, it is unclear which will ultimately prove most useful as the network of diagnoses and procedures surrounding a Target DX change in response to policy. To some extent, this problem is likely to pose a continuously moving target and so future research should more fully develop understanding of the temporal forces as well to determine whether, for example, the indicators of AMI depend on month of admission.

The approach used here is likely to be useful in the analysis of healthcare data in several ways. First, it provides a set of associated diagnoses and procedures that can be used to "impute" missing or unobserved data in an effort to estimate true prevalence of various diseases. Second, it can be used to estimate the extent of "gaming" of diagnoses in response to policy changes.

	DX code			
	Sepsis	AMI	CHF	PN
Acc (All)	91.4%	79.6%	77.5%	73.4%
Acc (PR)	91.4%	71.2%	65%	64.8%
Acc (HSIC)	91.45%	76%	72.6%	74.3%

Table 4: The classification accuracy for each Target DX using the MB alone (generated by PageRank (PR) and generated by HSIC) compared with the entire set of DXs

This suggests that our approach may also prove useful in order to adjust estimates for this kind of gaming and could also provide more robust methods to estimate true hospital readmission rates where intentional under-diagnosis of such sentinel diagnoses is likely. Historically, there are several precedents for this kind of under-reporting. The effects of the Omnibus Budget Reconciliation Act of 1987 (OBRA87) was observed to have considerable impact of medical practice in nursing home settings [2].

7 Acknowledgements

This research was supported by DARPA Grant FA9550-12-1-0406 negotiated by AFOSR and NSF through grant number NSF-14476570. HCUP, Agency for Healthcare Research and Quality, provided data used in this study.

References

[1] J. B. Braunstein, G. F. Anderson, G. Gerstenblith, W. Weller, M. Niefeld, R. Herbert, and A. W. Wu, *Noncardiac comorbidity increases preventable hospitalizations and mortality among medicare beneficiaries with chronic heart failure*, Journal of the American College of Cardiology, 42 (2003), pp. 1226–1233.

[2] R. Elon and L. G. Pawlson, *The impact of obra on medical practice within nursing facilities.*, Journal of the American Geriatrics Society, 40 (1992), pp. 958–963.

[3] K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, and P. Daszak, *Global trends in emerging infectious diseases*, Nature, 451 (2008), pp. 990–993.

[4] P. S. Keenan, S.-L. T. Normand, Z. Lin, E. E. Drye, K. R. Bhat, J. S. Ross, J. D. Schuur, B. D. Stauffer, S. M. Bernheim, A. J. Epstein, et al., *An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure*, Circulation: Cardiovascular Quality and Outcomes, 1 (2008), pp. 29–37.

[5] H. M. Krumholz, Z. Lin, E. E. Drye, M. M. Desai, L. F. Han, M. T. Rapp, J. A. Mattera, and S.-L. T. Normand, *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 (2011), pp. 243–252.

[6] C.-C. Liu, Y.-T. Tseng, W. Li, C.-Y. Wu, I. Mayzus, A. Rzhetsky, F. Sun, M. Waterman, J. J. Chen, P. M. Chaudhary, et al., *Diseaseconnect: a comprehensive web server for mechanism-based disease–disease connections*, Nucleic acids research, 42 (2014), pp. W137–W146.

[7] Q. Lou and Z. Obradovic, *Feature selection by approximating the markov blanket in a kernel-induced space.*, in ECAI, 2010, pp. 797–802.

[8] ———, *Predicting viral infection by selecting informative biomarkers from temporal high-dimensional gene expression data*, in Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on, IEEE, 2012, pp. 1–4.

[9] Q. Lou, H. P. Parkman, M. R. Jacobs, E. Krynetskiy, and Z. Obradovic, *Exploring genetic variability in drug therapy by selecting a minimum subset of the most informative single nucleotide polymorphisms through approximation of a markov blanket in a kernel-induced space*, in Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2012 IEEE Symposium on, IEEE, 2012, pp. 156–163.

[10] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann, 2014.

[11] T. H. Rice, *The impact of changing medicare reimbursement rates on physician-induced demand.*, Medical care, 21 (1983), pp. 803–815.

[12] F. A. Spencer, M. Moscucci, C. B. Granger, J. M. Gore, R. J. Goldberg, P. G. Steg, S. G. Goodman, A. Budaj, G. FitzGerald, K. A. Fox, et al., *Does comorbidity account for the excess mortality in patients with major bleeding in acute myocardial infarction?*, Circulation, 116 (2007), pp. 2793–2801.

[13] G. Stiglic, A. Davey, and Z. Obradovic, *Temporal evaluation of risk factors for acute myocardial infarction readmissions*, in Healthcare Informatics (ICHI), 2013 IEEE International Conference on, IEEE, 2013, pp. 557–562.

[14] D. Weycker, K. S. Akhras, J. Edelsberg, D. C. Angus, and G. Oster, *Long-term mortality and medical care charges in patients with severe sepsis*, Critical care medicine, 31 (2003), pp. 2316–2323.

[15] S. Yende, D. C. Angus, I. S. Ali, G. Somes, A. B. Newman, D. Bauer, M. Garcia, T. B. Harris, and S. B. Kritchevsky, *Influence of comorbid conditions on long-term mortality after pneumonia in older people*, Journal of the American Geriatrics Society, 55 (2007), pp. 518–525.

[16] L. Yu and H. Liu, *Feature selection for high-dimensional data: A fast correlation-based filter solution*, in ICML, vol. 3, 2003, pp. 856–863.