Proceedings of the 6th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2011)
P. Qian, Y. Zhou, C. Rudin, eds.

Understanding the Role of Mental Disorders on HIV Patient Outcomes

Shengfan Zhang
Department of Industrial Engineering
University of Arkansas
Fayetteville, AR
shengfan@uark.edu

Fay C. Payton
College of Management
North Carolina State University
Raleigh, NC

Julie S. Ivy
Edward P. Fitts Department of Industrial and Systems Engineering
North Carolina State University
Raleigh, NC

Abstract

This paper studies the impact of comorbidities, with a particular focus on mental disorders, on HIV patient outcomes as measured by inpatient length of stay (LOS) and total charges. Generalized linear models (gamma models) are developed to characterize the effects of selected comorbidities on HIV patient outcomes in the adult 2006 National Inpatient Sample. Comorbid HIV patients experience different LOS and total charges. In particular, having mental disorders results in a decrease in both LOS (by 19%) and total charges (by 15%) for HIV patients. To characterize the role of individual mental disorder, principal component and cluster analyses on ICD-9 codes are used, and eight conditions are found to be most strongly associated with HIV. These results suggest the complexity and combination of these conditions necessitate an integrated approach for diagnoses, treatment and disease management of comorbid HIV patients, particularly those with substance abuse and depressive mood disorders.

Keywords: Gamma Model, Principal Component Analysis, Cluster Analysis, HIV, Mental Disorder

Introduction

HIV is a life-threatening infectious disease that can lead to AIDS when the human immune system begins to fail. In 2007, 42,655 new cases of HIV/AIDS in adults, adolescents and children were diagnosed in the 33 states with long-term, confidential, name-based HIV reporting [1]. While testing methods have improved, the CDC estimates as many as one in five people are living with undiagnosed HIV. Our focus on HIV is based, in part, on its increasing prevalence and incidence estimate in the United States [10].

Most research has focused on cost-effectiveness analysis or policy planning with respect to HIV in isolation. However, patients with comorbid conditions tend to require additional resources and have different outcome patterns than those with the same primary reason for admission but without additional conditions. A comorbidity refers to the presence of one or more conditions (or diseases) in addition to a primary disease or disorder. Studying the impact of comorbidities provides information regarding disease management, cost structure and resource utilization.

In this study, hypertension, diabetes, mental disorders and obesity are selected as comorbid conditions of interest with respect to HIV, with a particular focus on mental disorders as they have been recognized to have special relationship with HIV in the literature [2, 3, 4]. This study extends earlier research by quantifying the impact of these specific

comorbidities on HIV inpatient length of stay and total charges by using generalize linear models (gamma models) [5]. In addition, variation reduction techniques including principal component analysis and cluster analysis [6] are used to identify specific mental disorder conditions that are most highly associated with HIV patients, and to identify the relationship between these conditions. This, in turn, will help physicians and/or policymakers better identify specific patient groups who may be at greater risk for HIV infection, and the hospital resources can be better planned with estimate on length of stay and total charges given the specific comorbid mental conditions.

Data

The 2006 Nationwide Inpatient Sample (NIS) data by Agency for Healthcare Research and Quality (AHRQ) [7] are used for this study. Only those patients who were 18 and older are considered in this analysis, resulting in 6,712,893 observations. ICD-9-CM diagnosis codes are used to identify HIV patient group (042), as well as comorbid conditions of interest: hypertension (401-405, 430-438, 4258, 4290-4293, 4298-4299, 7962, 36211), diabetes (250), mental disorder (290-311) and obesity (27800, 27801, 27802, V85.30-V85.4). This paper only focuses on primary HIV patients (i.e., HIV is considered the primary condition for hospitalization as defined by diagnosis code DX1), and comorbid conditions are determined if associated ICD-9 codes appear in any of the diagnosis codes DX2 to DX15.

The outcome measures (response variables) include data fields for length of stay (LOS), total charges (TOTCHG). In addition to the ICD-9 codes identifying HIV and comorbid conditions, other exploratory variables are extracted including patient age at admission, gender, race, admission type (ATYPE), total number of diagnoses (NDX), total number of procedures (NPR), vital status (DIED) and indicating variables for transferred (DISPUniform).

Generalized Linear Models (GLM)

Both length of stay and total charges are positive observations, and belong to the exponential family. It is hypothesized that the variance of observations may not be constant [8]. A least squares model is first built with White test to check the heteroscedasticity, and a p-value$<$.0001 strongly suggests rejecting the null hypothesis that the variance of residuals is homogeneous. Thus, generalized linear models, considering mean and variance simultaneously, are developed to study the impact of comorbidity on patient length of stay and total charges. A gamma model is selected since it is commonly used for analyzing positive observations [11, 12], and the model checking using Akaike information criteria has suggested a better fit over log-normal model, another popular model for positive observations.

The expected value of dependent variable Y (LOS and TOTCHG), is related to linear predictor through a monotonic differentiable linear function g, in the form of the following:

$$g(E(Y)) = g(\mu) = X\beta \tag{1}$$

where X is the vector of independent variables, and β is the vector of corresponding coefficients, which are estimated with maximum likelihood.

The variance of the response variable depends on the mean through a variance function V:

$$Var(Y) = V(\mu) = V(g^{-1}(X\beta)) \tag{2}$$

In the gamma model, a log link function $g(\mu) = log(\mu)$ is selected, and the associated variance function for gamma distribution is $V(\mu) = \mu^2$.

The independent variables (X) for the linear predictor include variables of patient demographics and admission information as introduced in the data section, dichotomous variables for individual diseases and interaction terms between HIV and each comorbid condition. Three-way or higher interaction terms are excluded from the model because none have significant coefficient of estimates.

The Akaike information criteria (AIC) for the gamma model is

$$AIC = 2P_g - 2l_g(\theta) \tag{3}$$

where p_g is the number of parameters in the gamma model, and $l_g(\theta)$ is the log likelihood function for the estimated model.

Principal Component and Cluster Analysis

In the GLM analyses above, 30 ICD-9-CM codes (290-319) are used to identify mental disorders. This is a broad category, and the goal of the PCA is to identify the most related of the 30 ICD-9 mental disorder codes for HIV patients. PCA provides a better understanding of the data by interpreting fewer principal components (PC) and removing unnecessary information.

PCA is related to Eigenvalue/Eigenvector theory. Specifically, it can be seen that

$$\lambda_i P_i = \sigma_i P_i \tag{4}$$

where P_i is the ith Eigenvector of the covariance matrix X; σ_i is the variance of ith PC, and λ_i is the corresponding Eigenvalue. The total variance of the system is the sum of all Eigenvalues. The ith PC will explain $\lambda_i/(\lambda_1 + \lambda_2 + ... + \lambda_n)$ proportion of variance.

Because n is large, n=30, the PCs are selected that explain 95% of the variance and thus reduced dimensionality of the ICD-9 codes. After the principal components are selected, for each variable the weighted score is determined by multiplying each PC vector by the corresponding Eigenvalue. Variance structure is used because indicator variable is created for each mental illness and thus the analysis does not have scaling problems. The variables with higher scores are selected because they corresponded to the most prevalent conditions for HIV patients.

CA is closely related to the PCA and also relies on the exploration of the Eigen structures in the dataset. Unlike other cluster analysis on observations (data entries), this study aims to cluster variables (data fields). It explores the correlation among variables and identifies the most significant condition within each cluster. This helps to determine the number of variables necessary for study and to identify what conditions have high correlation with each other.

The CA analysis is performed in SAS according to the following algorithm: cluster components (similar to the principal components) are computed in each iteration, and each variable is assigned to the component with which it has the highest squared correlation. It is then tested to see if the amount of variance explained increases if the variable is assigned to another cluster [9]. The most important variable (ICD-9 code for specific mental disorders) is selected from each cluster, i.e., the variable that has the greatest correlation with its own cluster and least correlated with other clusters. The ratio (1 - R-square) is used to identify these variables. Together with PCA, the dimension of the variables is reduced, and the mental disorders that are most related are identified for the group of primary HIV patients.

Results

GLM results (Table 1) have shown that HIV has a significant impact on both LOS and total charges. The condition increased LOS by approximately 60% (exp(0.4733)-100%) and increased total charges by 76% on average assuming no comorbid condition is present while holding the other independent variables fixed (the same assumption holds for the following discussion). Compared to other races, white patients had significantly shorter LOS but higher charges, while other races had significantly longer LOS and higher charges (except for Native American). For admission type, only type ”New born” had no significant difference in LOS with other admission type.

When looking at the effects of comorbidities on total charges and LOS, the data indicated that hypertension (in LOS only) and obesity (in both LOS and total charges) did not have significant interaction with HIV. Interestingly, the comorbid conditions did not increase length of stay or total charges as the impact of HIV dominated because the absolute coefficient associated with the presence of HIV was much higher than other coefficients. For example, mental disorders decreased length of stay for HIV patients by almost 19%, decrease charges by approximately 15%. The AIC for the LOS model is about 31,205,809 and 143,874,439 for total charges. The high AIC is due to the complexity of dataset which has large number of data fields and observations.

The results for the principal component and cluster analysis are shown in Table 2. Ten components that explained 94.59% of the total variance are used in PCA to calculate score for each conditon. There are eight clusters with the corresponding ICD-9 codes shown in column 2. Based on the (1-R-square) ratio in column 3, the most important variables are highlighted in bold, and the corresponding scores calculated from PCA are summarized in the last column. These results suggest that ”Nondependent abuse of drugs” had a much higher score compared to other conditions.And the most related conditions also include depression, anxiety and mood disorders.

Table 1: Results for Gamma Models on Length of Stay and Total Charges

	LOS		Total Charges			LOS		Total Charges	
Covariate	Estimate	p-value	Estimate	p-value	Covariate	Estimate	p-value	Estimate	p-value
Intercept	0.7337	$< .0001$	9.3302	$< .0001$	NDX	0.0743	$< .0001$	0.0422	$< .0001$
AGE	0.0017	$< .0001$	0.0057	$< .0001$	NPR	0.0772	$< .0001$	0.2274	$< .0001$
FEMALE	-0.0765	$< .0001$	-0.1653	$< .0001$	DIED	0.277	$< .0001$	0.2125	$< .0001$
$RACE_1$	-0.0264	$< .0001$	0.0215	$< .0001$	Transferred	0.4166	$< .0001$	0.2644	$< .0001$
$RACE_2$	0.1628	$< .0001$	0.1144	$< .0001$	HIV	0.4733	$< .0001$	0.5654	$< .0001$
$RACE_3$	0.0268	$< .0001$	0.0976	$< .0001$	H	-0.1131	$< .0001$	-0.012	$< .0001$
$RACE_4$	0.0589	$< .0001$	0.075	$< .0001$	D	-0.0591	$< .0001$	-0.0074	$< .0001$
$RACE_5$	0.0238	$< .0001$	-0.0543	$< .0001$	M	-0.1173	$< .0001$	-0.0202	$< .0001$
$ATYPE_1$	0.0801	$< .0001$	-0.36	$< .0001$	O	-0.0508	$< .0001$	0.0679	$< .0001$
$ATYPE_2$	0.0802	$< .0001$	-0.4985	$< .0001$	HIV×H	-0.0153	0.2884	-0.0663	$< .0001$
$ATYPE_3$	-0.0092	$< .0001$	-0.3233	$< .0001$	HIV×D	-0.1223	$< .0001$	-0.052	0.0214
$ATYPE_4$	0.0125	0.716	-0.5672	$< .0001$	HIV×M	-0.0879	$< .0001$	-0.1369	$< .0001$
$ATYPE_5$	0.2503	$< .0001$	0.2054	$< .0001$	HIV×O	0.0151	0.803	0.0153	0.8052

Note: 1) Race: 1-White 2-African American 3-Hispanic 4- Asian or Pacific Islander 5- Native American 6- Other. 2) Admission type: 1-Emergency 2-Urgent 3-Elective 4-Newborn 5-Trauma Center 6-Other. 3) H = hypertension, D = diabetes, M = mental disorder, O = obesity

Discussion

This study modeled and characterized the role of mental disorders on HIV patient outcomes. It first explored the impact of selected chronic comorbid diseases on HIV patient outcomes. While the results indicated HIV was the dominating condition that resulted in high total charges and longer lengths of stay, comorbid conditions affected HIV patient outcomes.

A particular focus for this paper was on mental disorder impact. Principal component and cluster analyses were used to identify the most correlated mental disorder conditions for HIV patients and reduced the number of ICD-9 codes required to capture mental disorders for HIV patients for future research. The most important conditions were drug-related mental disorders (304, 305), mood disorders (296), depression (311) and anxiety (300). This finding suggests that policymakers must focus on these conditions regarding HIV. Further, clinicians should implement an intregrated health services delivery approach to adherence and treatment to better address chronic diseases and their severity with comorbidity. Druss and Rosenheck [13] reported that, "Successful treatment of mental disorders by primary care physicians has been found to involve a combination of patient education, outcomes assessment and access to psychiatric consultation."

Specifically and in terms of clinical care, the results provide insight into the complexity of treating and managing complex, multi-comorbid patients. Hence, treatment and service delivery to comorbid HIV and mental health partients, particularly those with substance abuse conditions and depressive mood disorders, requires multispecialty service delivery. That is, the infectious disease specialist, psychiatrist and substance abuse counselor and/or social worker can form a multispecialty core team to treat and manage comorbid conditions.

Table 2: Results for Principal Component and Cluster Analysis

Cluster	ICD-9	1- R2 Ratio	ICD-9 Codes Description	Score
1	305	0	**Nondependent abuse of drugs**	0.3892
	313	0.9998	Disturbance of emotions specific to childhood and adolescence	
	315	1.0004	Specific delays in development	
	316	1.0015	Psychic factors associated with diseases classified elsewhere	
	318	1	Other specified mental retardation	
	319	1.0012	Unspecified mental retardation	
2	292	0.9635	Drug-induced mental disorders	
	293	1.0034	Transient mental disorders due to conditions classified elsewhere	
	304	0.0006	**Drug dependence**	0.1670
3	311	0	**Depressive disorder not elsewhere classified**	0.1533
4	296	0.0001	**Episodic mood disorders**	0.1394
	301	0.9968	Personality disorders	
	308	1.0001	Acute reaction to stress	
	309	0.9986	Adjustment reaction	
	310	1	Specific nonpsychotic mental disorders due to brain damage	
	314	0.998	Hyperkinetic syndrome of childhood	
	317	0.9993	Mild mental retardation	
5	290	0.9998	Dementias	
	294	0	**Persistent mental disorders due to conditions classified elsewhere**	0.1085
	307	0.9999	Special symptoms or syndromes not elsewhere classified	
6	297	0.9997	Delusional disorders	
	299	0.9998	Pervasive developmental disorders	
	300	0	**Anxiety, dissociative and somatoform disorders**	0.1121
	302	0.9982	Sexual and gender identity disorders	
	306	0.9983	Physiological malfunction arising from mental factors	
	312	1	Disturbance of conduct not elsewhere classified	
7	291	0.8952	Alcohol-induced mental disorders	
	303	0.0034	**Alcohol dependence syndrome**	0.0762
8	295	0.0002	**Schizophrenic disorders**	0.0706
	298	1.0016	Other nonorganic psychoses	

References

[1] Center for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2007. Vol. 19. Atlanta: US Department of Health and Human Services, CDC: 2009.

[2] Bing, E.G., Burnam, M.A., Longshore, D. et al. Psychiatric disorders and drug use among human immunodeficiency virus-infected adults in the United States. Arch Gen Psychiatry 2001; 58:721-8.

[3] Kilbourne, A.M., Justice, A.C., Rabeneck, L. et al. General medical and psychiatric comorbidity among HIV-infected veterans in the post-HAART era. J Clin Epid 2001; 54(Suppl 1):S22-8.

[4] Stoskopf, C.H., Kim, Y.K., Glover, S.H. Dual diagnosis: HIV and mental illness, a population-based study. Community Ment Hlt J, 2001; 37(6): 469-479.

[5] McCullagh, P., Nelder J.A. Generalized Linear Models. Chapman and Hall, London. 1989.

[6] Tabachnick, B.G., Fidell, L.S. Using Multivariate Statistics. Allyn and Bacon; 2006, 5th Ed.

[7] HCUP, Healthcare Cost and Utilization Project (HCUP) Databases, AHRQ 2008: Rockville, MD.

[8] Myers, R.H., Montgomery, D.C., Vining, G.G. Generalized Linear Models with Applications in Engineering and the Sciences, John Wiley and Sons, New York; 2002.

[9] SAS/STAT 9.1 User's Guide, 2004: 4799-4805.

[10] CDC HIV/AIDS Basic Statistics. Available at: http://www.cdc.gov/HIV/topics/surveillance/basic.htm#hivest

[11] Firth, D. Multiplicative errors: log-normal or gamma? J R Statist Soc B, 1988; 50: 266-68.

[12] Das, R.N., Lee, Y. Log normal versus gamma models for analyzing data from quality-improvement experiments. Quality Engineering 2009; 21(1): 79-87.

[13] Druss, B.G., Rosenheck, R.A. Locus of mental health treatment in an integrated service system. Psychiatr Serv 2000; 51:890-2