Modeling Hospital Discharge Counts Across Zip Code Areas

K. P. Singh[1], **J. T. Wulu, Jr.**[2], **S. Bae**[1], **A. A. Bartolucci**[3], **F. Trevino**[1]

[1]School of Public Health, University of North Texas Health Science Center, Ft. Worth, Texas 76107, USA

[2]Analysis & Research, Bureau of Primary Health Care/ HRSA/DHHS, Bethesda, MD 20814, USA

[3]School of Public Health, University of Alabama at Birmingham, AL 35294, USA

Abstract: When a data set is over-dispersed or under-dispersed, the Poisson regression model does not fit the data well. A generalization of Poisson regression (generalized Poisson regression) model is proposed for modeling such data. This paper models hospital discharge counts using Poisson regression (PR), generalized Poisson regression (GPR) and negative Binomial (NB) models. It is shown that the generalized Poisson regression (GPR) model has statistical advantages over Poisson regression and negative binomial regression models. The GPR model outperforms Poisson regression models except where the variation in diagnosis-related groups (DRGs) hospital discharge counts across zip code areas is relatively small. The results of this study show that household size, education, and income are positively related to DRGs hospital discharges. Also, the results support the hospital discharge findings of Wilson and Tedeschyi (*Health Services Research* 1984), Kudur and Demlo (*Iowa Medicine* 1985), Wennberg and Freeman (*The Lancet* 1987), Wolfe et. al. (*Statistics in Medicine* 1991), and Gittlesohn et al. (*American Journal of Public Health* 1991). The method of maximum likelihood is used for estimation. Approximate tests for the dispersion and goodness-of-fit measures for comparing models are discussed. The results of the paper will enhance our understanding variations and related reasons for the utilization of hospital services and health care medical facilities by community residents.

Keywords: *Poisson regression; generalized Poisson regression; negative binomial regression; maximum likelihood method; goodness-of-fit; deviance; diagnosis-related groups*

1. INTRODUCTION

Studies have indicated variations and related reasons for the utilization of hospital services and health care medical facilities by community residents. Gittelsohn et. al. (1991) studied surgical and medical discharge abstracts from 1985 to 1987 for Maryland patients admitted to acute care hospitals. Their study indicated that race, income and education are significant factors in differential hospital utilization rates. They found that admission rates for medical reasons declined with increasing community income levels and were elevated in blacks. Kudur and Demlo(1985) found average income levels in areas of high admission rates were significantly lower than those areas of low admission. Wilson and Tedeschyi (1984) found income levels positively associated with surgical discharges rates and a positive association between the percent of Medicaid population and medical discharge rates. Wennberg and Freeman (1987) found in a population-based study that the relationship between hospitalization rates for avoidable hospital conditions (AHC) cases and median household income, revealed consistent correlation between low income and high rate of hospitalization. Also, Codman Research Group (1991), a group of investigators in California found a negative correlation between income and the rate of hospitalization for these AHC and suggested that a reduction in these AHC's offers a considerable cost savings to the community.

deShazo (1997) admits that the random nature of the discharge count data of his research suggests a fit of the pure Poisson model because the data indicates the average number of discharges per person per time interval. He indicated that the goodness-of-fit of pure Poisson model for the count data was poor and observed that the count data indicated extra-Poisson variation; and thus, decided to fit a log-linear model. deShazo argued that the log-linear model was observed to be appropriate for this purpose since it can be extended to include measured community characteristics in a regression model, while still yielding an estimate of the amount of systematic variability beyond that predicted by the factors included in the model. The results of deShazo's research support the hospital discharge findings of Wilson and Tedeschyi (1984), Kudur and Demlo (1985), Wennberg and Freeman (1987), Wolfe et al. (1991), and Gittlesohn et al. (1991).

This paper models hospital discharge counts using Poisson regression (PR) model, generalized Poisson regression (GPR) model and negative Binomial (NB) model. The GPR model can be used to model data that shows over-dispersion or under-dispersion or equi-dispersion. It has been shown to have statistical advantages over standard Poisson regression (PR), compound Poisson regression (CPR), inverse-Gaussian-Poisson (IGP), and negative binomial regression (NBR) models.

The paper addresses the following questions: (i) What is the relationship between hospital DRGs counts and socio-demographic covariates such as, household income, race/ethnicity, education, family size, financial status, and population size for specific areas? (ii) Which covariates (factors or variables) make significant contributions to explain the particular hospital discharge count distributions?

2. THE GENERALIZED POISSION REGRESSION MODEL

Suppose Y_i is a count response variable that follows a generalized Poisson distribution. To model accident data, we define Y_i (i = 1, 2,…, 595) as the number of automobile accidents involving elderly drivers. The probability function of Y_i is given by

$$f_i(y_i;\mu_i,\alpha)=\left(\frac{\mu_i}{1+\alpha\mu_i}\right)^{y_i}\frac{(1+\alpha y_i)^{y_i-1}}{y!}\exp\left(\frac{-\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}\right),$$

$$y_i=0,\ 1,\ 2,\ \ldots \qquad (2.1)$$

and $\mu_i=\mu_i(x_i)=\exp(x_i\beta)$, where x_i is a (*k*-1) dimensional vector of covariates including demographic factors, driving habits and medication use, and β is a *k*-dimensional vector of regression parameters. For more details, the reader is referred to Famoye and Singh (1995). The mean and variance of Y_i are respectively given by

$$\mathrm{E}(Y_i\,|\,x_i)=\mu_i,$$

and

$$\mathrm{V}(Y_i\,|\,x_i)=\mu_i(1+\alpha\mu_i)^2.$$

The generalized Poisson regression model (2.1) is a generalization of the standard Poisson regression (PR) model. When α = 0 the probability function in (3.1) reduces to the PR model. Within the framework of PR model, the equality constraint is observed between the conditional mean $\mathrm{E}(Y_i\,|\,x_i)$ and the conditional variance $\mathrm{V}(Y_i\,|\,x_i)$ of the dependent variable for each observation. In practical applications and in "real" situations, this assumption is questionable since the variance can either be larger or smaller than the mean. If the variance is not equal to the mean, the estimates in PR model are still consistent but are inefficient, which leads to the invalidation of inference based on the estimated standard errors.

When α > 0, the GPR model represents count data with over-dispersion and when α < 0, the GPR model represents count data with under-dispersion. In (2.1), α is called the dispersion parameter and it can be estimated along with the regression coefficients in the GPR model. Using the method of maximum likelihood the estimates of α and β in the GPR model (2.1) are given by Famoye and Singh (1995).

3. GOODNESS-OF-FIT AND TEST FOR DISPERSION

The goodness-of-fit of GPR model can be based on the deviance statistic that is defined in Famoye and Singh (1995). The deviance statistic can be approximated by a chi-square distribution when

μ_i's are large. We can also use the log-likelihood value to measure the goodness-of-fit of the regression models. The regression model with a larger log-likelihood value is better than the one with a smaller value. To assess the adequacy of the GPR model over the PR model, we test the hypothesis

$$H_0 : \alpha = 0 \text{ against } H_a : \alpha \neq 0. \quad (3.1)$$

The test in (3.1) is for the significance of the dispersion parameter. Whenever H_0 is rejected, one should use the GPR in place of the PR model for the given data. To carry out the test in (3.1), one can use the asymptotically normal Wald type "t" statistic defined as the ratio of the estimate of α to its standard error. An alternative test for the hypothesis in (3.1) is to use the likelihood ratio test statistic, which is approximately chi-square distributed with one degree of freedom when the null hypothesis is true.

4. DESCRIPTION OF HOSPITAL DISCHARGE DATA STRUCTURE

The hospital discharge data set originates from the Finite Information Systems for Hospitals (FISH) Database and is published by the Birmingham Regional Hospital Council in Alabama. deShazo (1997) constructed and analyzed data sets from the FISH database. The data set used in deShazo's research contains inpatient hospital utilization rates, 86,726 inpatient discharge counts, socio-demographic and socio-economic factors (e.g., age, race, education, income, sex, level of poverty). The inpatient hospital discharges during a period of four years (1987-90), in two Alabama counties (Jefferson and Shelby) were classified by deShazo(1997) according to diagnosis related groups (DRGs) and small areas (or zip codes). Using the age-specific hospital utilization rates for the two counties, deShazo applied indirect standardization methods to obtain adjusted or expected discharge counts for all zip codes.

The overall age-adjusted rates per 1,000 for LV, AHC, and ACS discharges over the four- year period was 7.0, 16.1, and 5.7, respectively. The range of the discharge rates in the 60 zip codes of the two-county area was 15.4 to 2.1 per 1,000. The variables considered in this paper are as follows:

Number of observed LV Discharges (OBSLV); Number of observed ACS Discharges (OBSACS); Number of observed AHC Discharges (OBSAHC); Number of adjusted or expected LV Discharges (ADJLV); Number of adjusted or expected ACS Discharges (ADJACS); Number of adjusted or expected AHC Discharges (ADJAHC); Low Variation (LV) discharge rate (LVRATE); Ambulatory Care Sensitive (ACS) discharge rate (ACSRATE); Avoidable Hospital Condition (AHC) (THH90) ; Median 1990 household income (HHINC90); Percent of total black population (PCTBLACK); Percent of household with less than $7,000 annual income(LT7K); Percent of adult with less than High School Education (PCTLTHSG). The results of our analysis are given in the Appendix.

5. DISCUSSION

The issue of hospital discharges of patients is important, because it is economically driven, and while it is economically driven, it places significant implications on the accessibility and quality of care. Studies have shown that early discharge can lead to missed diagnosis (Kiely et al., 1998) and hospital readmission (Camberg et al., 1997), factors that influence the economics of healthcare. Thus, it is universally believed that discharges at anytime, particularly early discharge, without carefully assessing the complexity of the medical and psychosocial factors is unacceptable health care (Kiely et al., 1998). In public health and medical literature, a number of studies have been done on the racial disparity in early hospital discharge. For example, Moy et al. (1996) examined the relationship between race and discharge against medical advice from hospitals.

Investigators of this study found that in 1990, there were an estimated 241,911 discharges against medical advice, accounting for 0.92% of all live discharges. African-American patients were 1.78 times more likely than White patients to be discharged against medical advice (Moy et al., 1996). The result of these early discharges may be due to minorities being less satisfied with the medical care they received. Among Medicare beneficiaries, African-Americans and Hispanic patients report less satisfaction with care than White patients, and African-American patients report less confidence in their physicians (Physician Payment Review Commission, 1994). Blendon et al., (1989) revealed that racial differences in dissatisfaction with care are more visible in the inpatient setting than in the outpatient setting. Dissatisfaction with inpatient care may lead a patient to leave a hospital against medical advice.

Such discharges tend to place the patients at risk for adverse medical outcomes, disrupt the therapeutic relationship, demoralize staff, waste medical resources and expose the hospital to liability (Corley et al., 1981; Endicott et al., 1994). Additional studies that assess the risk of early discharge need to be conducted in order to evaluate the effectiveness of the current healthcare system. While cost is a major component of healthcare, it is important that accessibility and quality of care is not sacrificed in the restructuring of our system.

In this paper, we described nonlinear regression techniques appropriate for the analysis of hospital discharge data. The GPR models outperformed or performed equally as well as the PR and NBR models (Appendix). The average rate for all three DRGs was 28.8 per 1,000 across the four year period for all zip code areas. AHC DRGs showed the highest number or rate of discharges, while ACS DRGs showed the smallest.

6. ACKNOWLEDGEMENTS

The authors thank Dr. W.F deShazo of the University of Alabama Birmingham for providing the data used in this paper.

7. REFERENCES

Blendon, R. J., Aiken, L. H., Freeman, H. E. and Corey, C. R., 1989: Access to medical care for black and white Americans. *Journal of the American Medical Association* 261, 278-281.

Camberg, L. C., Smith, N. E., Beaudet, M., Daley, J., Cagan, M. and Thibault, G., 1997:

Discharge Destination and repeat hospitalizations. *Medical Care* 35, 756-767.

Corley, M. C. and Link, K., 1981: Men patients who leave a general hospital against medical advice: Mortality rates within six months. *Journal of the Studies of Alcohol* 42, 1058-1061.

deShazo, W. F., 1997: DrPH Dissertation (Unpublished), A Population Based Approach to Monitoring Inpatient Utilization and Determining the Influence of Sociodemographic Factors on Three Theoretical Categories of Diagnosis Related Groups in a Two County Area of Alabama. University of Alabama at Birmingham, USA.

Endicott, P. and Watson, B., 1994: Interventions to improve the AMA-discharge rate for Opiate-addicted patients. *Journal of Psychosocial Nursing* 32, 36-40.

Famoye, F. and Singh, K. P., 1995: Regression models for cytogenetic dose-response data. *Journal of Applied Statistical Science* 2(4), 319-333.

Gillum, B. S., Grave, E. J. and Wood, E., 1998: National hospital discharge survey. *Vital and Health Statistics* 13, 1-51.

Gittelsohn, A., Halpern, J., Sanchez, R., 1991: Income, race and surgery in Maryland. *American Journal of Public Health* 81, 1435-1438.

Kudur, J. M., Demlo, L. K., 1985: Small area variation in Iowa hospital utilization. *Iowa Medicine* 75, 213-217.

Wennberg, J. and Freeman, J., 1987: Are hospital services rationed in New Haven or over-utilized in Boston? *The Lancet* 1, 1185-1188.

Wilson, P. and Tedeschyi, P., 1984: Community correlates of hospital use. *Health Services Research* 19, 133-146.

Wolfe, R. A., Petroni, G. R., McLaughlin, C. G., and McMahon, Jr., L.F., 1991: mpirical Evaluation of Statistical Models for Counts or Rates. *Statistics in Medicine* 10, 1405-1416.

APPENDIX: Results

Table: Assessing the Effects of Covariates on the Count of ACS and LV Hospital Discharges Using the Poisson Regression (PR), Generalized Poisson Regression Model (GPR) and Negative Binomial Regression (NBR) Models

	PR model				GPR model				NBR model			
Variable	Estimate	SE	Wald t-test		Estimate	SE	Wald t-test		Estimate	SE	Wald t-test	Sig
CONSTANT	-5.7323	0.1084	-52.8860	*	-6.0248	0.2260	-26.6596	*	-5.9241	0.2316	-25.5790	*
ACSRATE	0.0987	0.0120	8.2286	*	0.1078	0.0211	5.1022	*	0.1016	0.0227	4.4758	*
LVRATE	0.0665	0.0100	6.6436	*	0.0679	0.0166	4.0882	*	0.0681	0.0179	3.8045	*
THH90	0.0353	0.0258	1..3701		0.1265	0.0689	1.8356		0.0936	0.0646	1.4489	
HHINC90	0.0278	0.0150	1.8512		0.0468	0.0262	1.7840		0.0409	0.0283	1.4452	
PCTBLACK	0.0002	0.0006	0.3154		0.0007	0.0011	0.6129		0.0005	0.0012	0.4167	
LT7K	0.0014	0.0024	0.5834		0.0017	0.0037	0.4700		0.0021	0.0040	0.5250	
PCTLTHSG	0.0106	0.0030	3.5572	*	0.0158	0.0059	2.6617	*	0.0140	0.0061	2.2951	*
α-dispersion parameter					0.0047	0.0012	4.0782	*				
Chi-square			170.9721				205.2169				51.9806	
Deviance			174.7723				63.1112				53.8090	
Log-Likelihood			-280.5107				-253.7330				-283.5126	

SE: Standard Error; * Significant at 0.05 level