ORIGINAL ARTICLE

Using statistical models to assess medical cost of hepatitis C virus

Mohsen Vahedi[1], Asma Pourhoseingholi[2], Sara Ashtari[3], Mohamad Amin Pourhoseingholi[2], Maryam Karkhane[3], Bijan Moghimi-Dehkordi[3], Azadeh Safaee[3,] Zahra Kimia[3], Seyed Moayed Alavian[4]

[1]*Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran*

[2]*Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[3]*Research Center for Gastroenterology and Liver Disease, Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[4]*Baqiyatallah Research Center for Gastroenterology and Liver Disease, Baqiyatallah University of Medical Sciences, Tehran, Iran*

ABSTRACT

Aim: This study compared PR and NB in predicting HCV patient costs. The objective of this study was to predict the direct cost of the HCV patient in Iran.

Background: Hepatitis C virus (HCV) is a common and expensive infectious disease in Iran.
Cost associated with HCV and its complications has not been well characterized. Analysis of cost data is important in providing consistent information to aid budgeting decisions and certain statistical regression models need for prediction mean costs. Poisson regression (PR) and negative binomial regression (NB) are more common in cost prediction study.

Patients and methods: This study designed as a cross-sectional clinic base from 2001 to 2010. First treatment period of each patient bring in study. We evaluated the doctor visiting, drugs, and hospitalization and laboratory tests of patients. Cost per person per one treatment period estimated in purchasing power parity dollars (PPP$). The PR is one of the models from general linear models (GLM) for describing count outcomes. The NB is another model from (GLM) as an alternative to the PR model.

Results: According to Likelihood ratio test NB was found to be more appropriate than PR (P<0.001). Genotype, marriage, medication, and SVR were being significant. Genotype 3 versus 1 decreasing cost while marriage, consuming pegasys and SVR increasing.

Conclusion: choosing best model in cost data is important because of specific feature of this data. After fitting the best model, analyzing and predicting future cost for patient in different situation is possible.

Keywords: Chronic hepatitis C, Costs, Predictive models, Poisson regression, Negative binomial regression.
(Please cite as: **Vahedi M, Pourhoseingholi A, Ashtari S, Pourhoseingholi MA, Karkhane M, Moghimi-Dehkordi B, et al. Using statistical models to assess medical cost of hepatitis C virus. Gastroenterol Hepatol** Bed Bench **2012;5(Suppl. 1):S31-S36).**

Introduction

[1]This days Hepatitis C virus (HCV) infection is a major cause of liver-related morbidity and mortality worldwide and a major public health problem (1-4). According to epidemiologic studies it is estimated that around 170-200 million

individuals are living with HCV infection worldwide (2, 3, 5). It seems the prevalence of HCV is rising in Iran (5, 6). Recent study reported the seroprevalence of HCV in the population studied is 0.5%, which is higher than previous estimates for Iran (5). HCV infection is responsible for 20% of acute hepatitis cases, 70% of all chronic hepatitis cases, 40% of all cases of liver cirrhosis,

60% of hepatocellular carcinomas (HCC), and 30% of liver transplants in Europe (5, 7).

Chronic HCV infection is also a significant health care economic burden. Although serious and costly complication of HCV infection may develop, such as liver failure, the need for liver transplantation, and cancer, patients with chronic HCV may delay treatment until after symptoms emerge because of the significant direct and indirect cost associated with current treatments (7). Thus, it is not surprising that the health costs of people attract the attention of many policy makers and academics in many countries (7,8). It always desirable to measure economic burden and health care effectiveness in order to understand and evaluated various intervention programs in the country. Recently there is a one study for estimating of average diagnosis and treatment costs of hepatitis C in research center of gastrointestinal and liver disease of Shahid Beheshti of Medical University. This study is under the publishing.

After estimating the cost of HCV it seems analyzing and predicting of this cost is important. Specific characterize of cost data is its distributions that are difficult to describe using standard approaches like ordinary least square regression for analyzing (9). Poisson model is one of the approaches that use for analyzing data such as cost data. But due to over-dispersion, a related problem of Poisson regression, that arise in count data frequently, another model like Negative binomial used for this data(10). The application of these models and their comparisons with each others has increased in medical and health fields recently (11-18). In this paper we used Poisson regression (PR), negative binomial (NB), for analyzing the cost HCV.

Patients and Methods

All data for this cross-sectional study were collected from medical records of 200 patients with hepatitis C, who referred to a private gastroenterology clinic between years 2000 through 2009 in Tehran.

We concluded that patients have some common costs during their diagnosis and treatment. These costs are as follows:

- Diagnostic tests includes: Endoscopy, Sonography, liver biopsy, Pathology and Electrophoresis.
- Monthly laboratory tests and Measurement of hepatic markers during the treatment, including CBC-diff, AST, ALT, ALP, total and direct Bill, Genotyping, PCR and Viral load, etc.
- Short term of hospitalization due to liver biopsy.
- The cost of routine visits by a gastroenterologist.
- Medication (drug) fees.

Diagnosis and treatment costs of HCV in this study were calculated per patient during in one course of treatment and patients were followed over six-month period after the stopping of treatment. The cost of short term hospitalization due to liver biopsy was obtained from the medical records. Methodology of cost analysis in this paper is based on Centers for Disease Control and Prevention (19) "cost analysis introduction" and also is similar to another Iranian studies (20-22). Purchasing power parity dollar (PPP$) was used in order to make inter-country comparisons.

Statistical methods

The poisson regression (PR) is one of the models from general linear models (GLM) for describing count outcomes or proportion/rates (10). This model assume response had a poisson distribution. Count data often much more than that we would expect if the response distribution truly were poisson. In this case the variances are much larger than the means, whereas poisson distributions have identical mean and variance. The phenomenon of the data having greater variability than expected for a general linear model is called over-dispersion. A common cause of over-dispersion is heterogeneity among subjects

(10). The negative binomial(NB), is another models from GLM as an alternative to the PR model, is a solution to account for over-dispersion due to unobserved heterogeneity (23). This model helps in adjusting the standard errors of the regression coefficients and provides a more flexible approach for prediction of the count outcome.

Results

A total of 284 patients entered in this study. Mean age (± standard deviation) of patient with HCV infection in this study was 41.69± 11.64. 225 (79.2%) patients were male. Majority of patients 203 (71.5%) were married. The distributions of covariates considered in the analysis are shown in table 1.

Table 1. Distribution of covariates in the population of study

Variables	n	%
Gender		
Male	225	79.2
Female	59	20.8
Age group		
14-35	89	31.3
36-57	171	60.2
>58	24	8.5
Marital status		
Single	81	28.5
Married	203	71.5
Outcome		
SVR	147	51.8
Not SVR	137	48.2
Medication		
Interferon + Ribavirin	126	44.4
Peg-interferon + Ribavirin	158	55.6
Genotype		
1	214	75.4
2	4	1.4
3	66	23.2
Education		
Lower diploma	206	72.5
Upper diploma	78	27.5

According to results of the 284 patients who entered in this analysis, 214 (75.4%) patients of them were infected with1, 4 (1.4%) with 2 and 66 (23.2%) patient with genotype 3. Of the 284 patients who participated in this study, 126 (44.4%) patients had combination therapy of standard Interferon plus ribavirin and the others 158 (55.6%) patients had combination therapy of Peg-interferon plus ribavirin. Since the costs for each patients is different with respect to their treatment regimen. Diagnosis and treatment costs were calculated for each patient who entered in this study. The mean and standard deviation of the costs per patient were 9435.88 and 7249013 PPP\$ respectively. Median of this cost was 5432.5 PPP\$. In (PR) models all covariates were statistically significant. The significant Pearson chi square goodness of fit (gof) test (p < 0.001) along with other characteristics of model fit indicated that the (PR) model produced a poor fit for cost data. So it seems the results of this model were not trustworthy. In the (NB) model, the estimated dispersion statistic (α) was 5.26 (95% CI: 4.34, 6.25). A significant likelihood ratio test (p < 0.001) of dispersion statistic from zero favored the NB model over the PR model. So (NB) model was the best model for analyzing this data. In this model Genotype, marriage, medication, and SVR were being significant. These results showed that SVR (ADJ.OR=1.49; 95% CI 1.34, 1.66; P<0.001), combination therapy of Peg-interferon plus Ribavirin (ADJ.OR=2.88; 95% CI 2.58, 3.21; P<0.001) and marriage (ADJ.OR=1.19; 95% CI 1.05, 1.35; P<0.001) effected to increase the chance of increasing in costs. On the other hand genotype 3 (ADJ.OR=0.64; 95% CI 0.56, 0.73; P<0.001) decrees the chance of increasing in costs. Table 2 showed the result of (NB) model.

Table 2. Results of fitting NB regression model to cost of HCV

Variables	Adj.OR[*](0.95% CI)	p-value
Age	0.996(0.991-1.001)	0.002
Gender		
Female[†]		
Male	0.995(0.871-1.136)	0.942
Outcome		
Not SVR[†]		
SVR	1.493(1.343-1.660)	<0.001
Medication		
Interferon + Ribavirin[$]		
Pegasys + Ribavirin	2.881(2.585-3.210)	<0.001
Marital status		
Single[†]		
Married	1.193(1.051-1.354)	0.006
Education		
Lower diploma[†]		
Upper diploma	1.064(0.943-1.202)	0.310
Genotype		
1[†]		
2	1.167(0.747-1.822)	0.497
3	0.645(0.567-0.734)	<0.001

[*] Adjusted Odds Ratio; [†] Reference Category

Discussion

Cost analyzing and related studies in clinical research, has been must attention in last year's and there are lots of study in this area in the world (19, 24). medical cost data typically show three characteristics that need to be accounted for in modeling (24). First, the data often show a substantial percentage of zeros corresponding to individuals with no expenses over the time of observation. This phenomenon called zero inflated. Second, for those individuals who do have expenses, the distribution of expenses is often highly skewed to the right. Furthermore, when using traditional regression techniques to develop models for those individuals with expenses, the assumption of homoscedasticity (constant variance) is often violated; that is, the expense data exhibit variability that tends to increase as the mean expense increases. Our data have no zero in HCV expense because all of patients have treatment. So we did not use models to account for zero inflation.

The problem of skewness and heteroscedasticity is often dealt with by transforming costs and using traditional linear regression techniques on the transformed data. Under the assumption that the variability in costs is proportional to the square of mean costs, the appropriate variance-stabilizing transformation is the logarithm (25). This transformation provides approximate homoscedasticity while at the same time it often serves to make the distribution of expenses more symmetric. Both of these results permit the use of traditional regression techniques, which assume homoscedasticity, and normality of underlying distributions. Although highly skewed cost data often still do not have a normal distribution when log-transformed, the assumption of normality is not critical (26) . In fact, using ordinary least quares to estimate model parameters make only first and second order moment assumptions on log(y). Where y is expense: the mean of log(y) is linearly related to the covariates, and the variance of log(y), conditional on values of the covariates is constant. But the main problem related to this transformed expense is that all inference must be done on the log-dollar scale, not on the original dollar scale. So instead transformed data, we use a (GLM), which explicitly takes into account heteroscedasticity. Rather than transforming expenses, (GLM) represent a reparameterization of the model. Furthermore, (GLM) can accommodate skewness in the expense distribution. So PR model and NB model that belong to (GLM) have been used for the cost of HCV patient in this study. Blough offered GLM models for medical cost data and was expressed

this model have better fit than ordinary least square regression in cost data (24). Mora in his paper for studying and predicting individual patient costs in adult intensive care units (ICUs) compared GLMs and ordinary least squares regression (OLS)(27). Barber considered (GLM) with either an identity link function and applied to estimate the treatment effects in two randomized trials adjusted for baseline covariates (28). So it seems application of (GLM) for cost will lead to better results. On the other hand if we want to talk about the interpretation of the results, our result showed who achieved SVR had more cost than others. The reason for this result may be was that the patient without SVR, abandoned the treatment before it was complete so they had less cost. The odds ratio of increasing cost in Genotype 3 was 1.61 times of Genotype 1. It seems that the lower cost for Genotype 3 relative to Genotype 1 because of difference in protocol of treatment. So in conclusion after fitting the best model, we can predict future cost for patient in different situation of significant variables.

References

1. Alter MJ. Epidemiology of hepatitis C virus infection. World J Gastroenterol 2007; 13:2436-41.

2. Alavian SM. Hepatitis C virus infection: Epidemiology, risk factors and prevention strategies in public health in I.R. IRAN. Gastroenterol Hepatol Bed Bench 2010; 3:5-14.

3. Alavian SM. New globally faces of hepatitis B and C in the world. Gastroenterol Hepatol Bed Bench 2011; 4:171-74.

4. Alavian SM, Alavian SH, Ashayeri N, Babaei M, Daneshbodi M, Hajibeigi B. Prediction of liver histological lesions with biochemical markers in chronic hepatitis B patients in Iran. Gastroenterol Hepatol Bed Bench 2010; 3:71-76.

5. Shepard CW, Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. Lancet Infect Dis 2005; 5:558-67.

6. Alavian SM, Adibi P, Zali MR. Hepatitis C virus in Iran: Epidemiology of an emerging infection. Arch Iran Med 2005; 8:84-90.

7. Alberti A, Benvegnu L. Management of hepatitis C. J Hepatol 2003; 38:S104-18.

8. Merat S, Rezvan H, Nouraie M, Jafari E, Abolghasemi H, Radmard AR, et al. Seroprevalence of hepatitis C virus: the first population-based study from Iran. Int J Infect Dis 2010; 14:e113-16.

9. Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for analyzing health care utilization and costs. Annu Rev Public Health 1999; 20:125-44.

10. Agresti A. An Introduction to Categorical Data Analysis. 2 ed. New York: John Wiley & Sons, Inc; 2007.

11. Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychol bull 1995; 118:392-404.

12. Hardin J, Hilbe J. Generalized Linear Models and Extensions: Stat-Corp LP, Texas: A Stata Press Publication; 2007.

13. mullahy j. specification and testing of some modified count data models. J Econometrics 1986; 33:341-65.

14. Lambert D. Zero-inflated Poisson regression, with application to defects in manufacturing. Technometrics 1992; 34:1-14.

15. Vuong Q. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 1989; 57:307–33.

16. Picard R, Cook DJ. Cross-Validation of Regression Models. J Am Stat Assoc 1984; 79:575–83.

17. Baughman L. Mixture model framework facilitates understanding of zero-inflated and hurdle models for count data. J Biopharmac Stat 2007; 17:943-46.

18. Gilthorpe MS, Frydenberg M, Cheng Y, Baelum V. Modelling count data with excessive zeros: the need for class prediction in zero-inflated models and the issue of data generation in choosing between zero-inflated and generic mixture models for dental caries data. Stat Med 2009; 28:3539-53.

19. Myers RH, & Montgomery DC. A tutorial on generalized linear models. J Qual Tech 1997; 29:274-91.

20. Poynard T, McHutchison J, Goodman Z, Ling MH, Albrecht J. Is an "a la carte" combination interferon alfa-2b plus ribavirin regimen possible for the first line treatment in patients with chronic hepatitis C? The ALGOVIRC Project Group. Hepatology 2000; 31:211-18.

21. Fried MW, Shiffman ML, Reddy KR, Smith C, Marinos G, Goncales FL, Jr, et al. Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection. N Engl J Med 2002; 347:975-82.

22. Poynard T, Yuen MF, Ratziu V, Lai CL. Viral hepatitis C. Lancet 2003; 362:2095-100.

23. Dwivedi AK, Dwivedi SN, Deo S, Shukla R, Kopras E. Statistical models for predicting number of involved nodes in breast cancer patients. Health 2010; 2:641-51.

24. Blough DK. Using Generalized Linear Models to Assess Medical Care Costs. Health Serv Outcomes Res Methodol 2000; 1:185-202.

25. Duan N, Manning Jr, Morris CN, Newhouse JP. A Comparison of Alternative Models for the Demand for Medical Care. J Business Econ Stat 1993; 1:115-26.

26. Ramsey S, Newton K, Blough D, McCulloch D, Sandhu N, Wagner E. Patient-Level Estimates of the Cost of Complications in Diabetes in a Managed Care Population. Pharmacoeconomics 1999; 196:285-95.

27. Moran JL, Solomon PJ, Peisach AR, Martin J. New models for old questions: generalized linear models for cost prediction. J Eval Clin Pract 2007; 13:381-89.

28. Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. J Health Serv Res Policy 2004; 9:197-204.