Research Article

Application of Zero-Inflated Poisson Mixed Models in Prognostic Factors of Hepatitis C

Alireza Akbarzadeh Baghban,[1] Asma Pourhoseingholi,[2] Farid Zayeri,[3] Ali Akbar Jafari,[4] and Seyed Moayed Alavian[5]

[1] *Department of Basic Sciences, School of Rehabilitation, Shahid Beheshti University of Medical Science, Tehran, Iran*

[2] *Department of Biostatistics, School of Paramedical Science, Shahid Beheshti University of Medical Science, Tehran, Iran*

[3] *Proteomics Research Center, School of Paramedical Science, Shahid Beheshti University of Medical Science, Tehran, Iran*

[4] *Department of Statistics, Yazd University, Yazd, Iran*

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

Correspondence should be addressed to Asma Pourhoseingholi; asma_phg@yahoo.com

Background and Objectives. In recent years, hepatitis C virus (HCV) infection represents a major public health problem. Evaluation of risk factors is one of the solutions which help protect people from the infection. This study aims to employ zero-inflated Poisson mixed models to evaluate prognostic factors of hepatitis C. *Methods.* The data was collected from a longitudinal study during 2005–2010. First, mixed Poisson regression (PR) model was fitted to the data. Then, a mixed zero-inflated Poisson model was fitted with compound Poisson random effects. For evaluating the performance of the proposed mixed model, standard errors of estimators were compared. *Results.* The results obtained from mixed PR showed that genotype 3 and treatment protocol were statistically significant. Results of zero-inflated Poisson mixed model showed that age, sex, genotypes 2 and 3, the treatment protocol, and having risk factors had significant effects on viral load of HCV patients. Of these two models, the estimators of zero-inflated Poisson mixed model had the minimum standard errors. *Conclusions.* The results showed that a mixed zero-inflated Poisson model was the almost best fit. The proposed model can capture serial dependence, additional overdispersion, and excess zeros in the longitudinal count data.

1. Introduction

In recent years, hepatitis C virus (HCV) infection has been a major cause of liver diseases worldwide and represents a major public health problem [1–5]. Transfusion and contact with infected blood and its products, intravenous drug use, and contamination during medical procedures are among different risk factors of HCV [6–8]. An estimated 130–170 million people worldwide are infected with hepatitis C. The global prevalence of this infection is approximately 0.2%−40% [2, 9]. But there is a difference between developed and undeveloped countries in its prevalence. It is due to difference in health policies and medical care [10]. Apart from few studies that have been done on high-risk groups or in specific locations, no comprehensive and accurate estimate of

HCV infection is available in Iran. According to two available studies which examined Iranian population, the prevalence of HCV infection in the general population is less than 1% [11, 12].

Hepatitis C is a common infection that causes chronic liver disease in the world [13]. The occurrence of end-stage liver disease caused by HCV is estimated to peak around 2020 [10, 14]. According to other studies, 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 [15].

In the coming decades, it is expected that the economic burden and mortality associated with hepatitis C rise [7, 16]. Unfortunately, the majority of infections do not respond to treatment and lead to chronic diseases. So it seems that

controlling HCV infection is an important issue in public health [5, 17]. Risk factor evaluation in order to reduce the problem in the community is one solution to protect people from the infection.

In medical researches statistical modeling is a powerful approach in risk factor evaluation, but selection of good and appropriate model is important. When the response variable is count, there are some models that they use for analyzing such data. Sometimes count data have an overdispersion problem because of having large number of zeros. This phenomenon is called zero-inflation. Using usual count model in zero-inflated data causes misleading results.

Lambert [18] proposed the zero-inflated Poisson (ZIP) regression model for independent count data. For clustered count data, ZIP models have been developed, and different types of such models have been introduced and used in different studies [19, 20]. In this study, the relationship between 3 viral loads of each HCV patient and some risk and demographic factors was investigated using mixed ZIP regression. Details of mixed ZIP modeland its parameter estimation are described in [21, 22].

2. Methodology

2.1. Patient Selection. This is a longitudinal study and all data for this research were drawn from medical records of 186 patients with hepatitis C. All of these patients had been referred to Tehran hepatitis clinic, a clinic of Baqiyatallah Research Center for Gastroenterology and Liver Diseases, from 2005 to 2010. The Information concerning 186 patients includes viral load (HCV-RNA). The viral load had been recorded before the treatment, during the period of treatment, immediately after this period and 3 to 4 months after the end of the treatment. The viral load before treatment has been considered for baseline adjusting. The variables included in the study are as follows: demographic information including sex and age, genotypes including genotypes 1, 2 and 3, treatment protocol including combination therapy of standard Interferon (3 MU three times a week) plus Ribavirin (800–1200 mg per day) for 24 weeks or 48 weeks [23–25] as well as combination therapy of Peg-Interferon (Alfa 2a in a fixed dose of 180 micrograms per week) plus Ribavirin (800–1200 mg per day) for 24 weeks or 48 weeks [24, 26], history of blood transfusion, addiction (IV drug user), and contaminated needle stick. All of these factors were extracted from the patient's medical records. Therefore, five covariates including age, sex, genotype, protocol of treatment, and risk factor were entered in this study. Finally, 558 viral loads of HCV and their related information were extracted; it means that each patient was examined three times (the first time was baseline). On the other hand, negative HCV-RNA is considered as being below 100 and it is taken as zero in the analyses. Generally, HCV-RNA of 100 to 200,000 is considered as being very low; 200,000 to 1,000,000 as low; 1,000,000 to 5,000,000 as medium; 5,000,000 to 25,000,000 as high; and above 25,000,000 as very high.

2.2. Statistical Analysis. Descriptive statistics and frequency distribution such as mean, standard deviation, and

TABLE 1: The distribution of covariates in this study.

Variable	Category	n	%
Sex	Female	55	29.6
	Male	131	70.4
Risk factor	Yes	104	55.9
	No	82	44.1
Genotype	1	142	76.3
	2	4	2.2
	3	40	21.5
Treatment	Interferon plus Ribavirin	100	53.8
	Peg-Interferon plus Ribavirin	86	46.2

TABLE 2: The distribution of six groups of viral load in 186 patients.

Viral load	Male		Female		Total	
	n	%	n	%	n	%
Zero	289	55.2	116	52.7	405	54.5
200 to 200,000	78	14.9	36	16.4	114	15.3
200,000 to 1,000,000	110	21	45	20.5	155	20.8
1,000,000 to 5,000,000	40	7.6	21	9.5	61	8.2
5,000,000 to 25,000,000	7	1.3	1	0.5	8	1.1
Above 25,000,000	0	0	1	0.5	1	0.1
Total	524	100	220	100	744	100

percentage were calculated according to standard methods. The outcome variable of interest is the viral load of HCV patients. For calculating the viral load of HCV patients, where data are clustered on the subjects, a mixed ZIP model was employed. This model is a combination of zero-inflated and random effects models to control both zero-inflated and cluster structure of data [22]. On the other hand, the Poisson random effects model, without considering zero inflated structure of data, was carried out [27]. These two models were compared using standard error of their estimators. Significance was defined as $P < 0.05$. Stata 11 and R 2.13.1 program, were used for the analysis.

3. Results

55 patients of the total 186 patients who were entered into this study were females. The mean and standard deviation of age were 42.88 and 11.17 years, respectively. Their age ranged between 19 and 76 years. Table 1 shows the distribution of covariates in this study. Each patient had four viral loads for evaluating the treatment process. Table 2 shows the distribution of six groups of viral load in 186 patients repeated four times for each. According to these results, 55.2% of patients had negative HCV-RNA, which means that zero inflated models is needed. At the first stage, PR regression with random effects (mixed PR) was fitted. The random effect was entered into this model for adjusting the clustered data structure. According to the results of this model, genotype 3 and treatment protocol were statistically significant. Table 3

TABLE 3: The results of mixed PR model.

Parameter	Category	Estimated	Standard error	P-value
Intercept		0.7320	0.188	<0.001
Age		0.0049	0.0036	0.195
Sex	Female*			
	Male	−0.0487	0.0864	0.580
Riskfactor	No*			
	Yes	0.154	0.0803	0.053
Genotype	1*			
	2	−0.1760	0.2785	0.539
	3	−0.2357	0.0984	0.013
Treatment	Interferon plus Ribavirin*			
	Peg-Interferon plus Ribavirin	−0.2584	0.0728	0.001
Variance component		0.2469		<0.001

*Reference group.

TABLE 4: The results of mixed ZIP model.

Parameter	Category	Estimated	Standard error	P-value
Intercept		3.6517	0.1659	<0.001
Age		0.0046	0.0023	0.037
Sex	Female*			
	Male	−0.978	0.0833	<0.001
Risk factor	No			
	Yes	0.4626	0.074	<0.001
Genotype	1*			
	2	−0.8006	0.2381	<0.001
	3	−1.148	0.0932	<0.001
Treatment	Interferon plus Ribavirin*			
	Peg-Interferon plus Ribavirin	−0.6628	0.0737	<0.001
λ		0.6891		
τ^2		1.6153		
ρ		0.3634		

*Reference group.

shows the results of this model. The significant Pearson Chi square goodness of fit (GOF) test ($P < 0.001$) along with other features of the model fit indicated that the mixed PR model produced a poor fit. On the other hand, a significant likelihood ratio test ($P < 0.001$) of dispersion statistic from zero showed that overdispersion has occurred in this data.

In the next stage, ZIP model with random effects (mixed ZIP) was carried out to account for both clustering and excessive zeros. The covariates of age, sex, and genotype, protocol of treatment, and risk factors had significant effects on developing HCV-RNA at $\alpha = 0.05$. The rate of virological response was higher in younger males. Subjects who had none of the risk factors, including the history of blood transfusion, addiction (IV drug user), and needle stick, were more likely to have virological response than others. Patients with genotype 3 and genotype 2 tended to have more virological response than those with genotype 1. The rate of virological response was also higher in subjects with combination therapy of Peg-Interferon plus Ribavirin. In addition to regression parameter in this model, two parameters of the random effects were estimated. The first estimate of random effect model ($\rho = 0.3543$) indicated the longitudinal correlation between the subjects. Also this random effect shows that the recurrence of the HCV-RNA every time partly depends on its value at the previous count. The second random effects ($\tau^2 = 1.61$) indicated that the variation of data is much greater than the one shown by the first random effect. The results of this model are shown in Table 4. The comparison of these two models is presented in Table 5. The standard errors for covariate effects obtained from ZIP model were generally smaller than those obtained from PR regression with random effects.

4. Discussion

In this paper, a mixed ZIP model was used for clustered count data with excessive zero. Its results were compared with those of mixed PR model.

In mixed ZIP model, all covariates had significant effects on the response variable. In this research, the rate of low viral load in men was more than that in women. The results of the studies done recently on patients with genotype 1 indicate that SVR in men is 2.5 times higher than that in women [28]. In the present study, this rate was 2.7 times as much. It seems that this difference is because of some physiological and psychological differences between men and women in the society. Also, patients with genotypes 3 and 2 had more virological response than patients with genotype 1. The fact that achieving SVR in genotype 1 is more difficult than in other genotypes has also been confirmed by the results of other studies [29]. On the other hand, risk factors decrease the rate of virological response. It seems that such results have been obtained due to the relationship that exists between the risk factors and the genotype. For example, there is a direct relationship between genotype 1 and injecting drug users, blood transfusion, and contact with infected blood as well as its products [30]. Two main protocols of treatment were used in this study based on the genotype of patients. According to the results, combination therapy of Peg-Interferon plus Ribavirin had better results than combination therapy of standard interferon plus Ribavirin. The large number of studies which have been conducted so far showed that Peg-Interferon plus Ribavirin had been most responsive to treatment [31–34]. So it seems that this protocol has been the best choice [35, 36]. Unfortunately in Iran, due to high cost

TABLE 5: The comparison of mixed ZIP and mixed PR.

Parameter	Category	Mixed ZIP model		Mixed PR model	
		Estimated	Standard error	Estimated	Standard error
Intercept		3.6517	0.1659	0.7320	0.188
Age		0.0046	0.0023	0.0049	0.0036
Sex	Female*				
	Male	−0.978	0.0833	−0.0487	0.0864
Risk factor	No				
	Yes	0.4626	0.074	0.154	0.0803
Genotype	1*				
	2	−0.8006	0.2381	−0.1760	0.2785
	3	−1.148	0.0932	−0.2357	0.0984
Treatment	Interferon plus Ribavirin*				
	Peg-Interferon plus Ribavirin	−0.6628	0.0737	−0.2584	0.0728

*Reference group.

of the drug, it is not the first choice for doctors. Usually when patients did not respond to the treatment, doctors decided to prescribe Peg-Interferon plus Ribavirin [36].

Although clustered count data with extra zeros often occur, few methods have been developed for correlated data with extra zeros [37]. There are some studies done on the extension of zero inflated models in order to accommodate random effects [20, 38]. In all of these models, there were two separate random effects in the models; therefore, the interpretation of the results was more difficult and sometimes confusing. A mixed ZIP model that was used in this paper has been introduced by Ma et al. in 2009. This proposed model had a compound Poisson random effect structure. This distribution was very useful for characterizing both the excessive zeros and clustering structure of the data. Another advantage of this model was its computational efficiency which was highly useful for analyzing massive data sets [21]. In this data set, the programs run after two minutes and thirty seconds. A comparison was also made between the results gained by this model and those gained by one of the standard methods for analysis of longitudinal count data (mixed PR model). The standard errors for covariate effects obtained from the mixed ZIP model were generally smaller than those obtained from mixed PR model. Results were compared by standard error, because there is not any goodness of fit criterion for mixed ZIP model yet. Standard errors get larger unless the extra zeros are accounted for. If the excessive number of zeros is not adjusted, the standard deviation gets larger. Therefore, the standard deviation of estimators in ZIP model is smaller than that of the other model. Since the zero inflated structure has not been taken into account in the other model, the standard deviation of estimators gets larger compared to ZIP model.

In conclusion, the mixed zero inflated Poisson models were seen as almost being the best fit. As with this research, clustered zero inflated count data is quite frequent in medical researchers. Since a wrong model would yield unreliable results, therefore, choosing the best and correct model for analyzing the data is highly important.

References

[1] S. M. Alavian, "Are the real HCV infection features in Iranian patients the same as what is expected?" *Hepatitis Monthly*, vol. 5, no. 1, pp. 3–5, 2005.

[2] M. J. Alter, "Epidemiology of hepatitis C virus infection," *World Journal of Gastroenterology*, vol. 13, no. 17, pp. 2436–2441, 2007.

[3] S. M. Alavian, "Hepatitis C virus infection: epidemiology, risk factors and prevention strategies in public health in I.R.IRAN," *Gastroenterology and Hepatology from Bed to Bench*, vol. 3, no. 1, pp. 5–14, 2010.

[4] S.-M. Alavian, "New globally faces of hepatitis B and C in the world," *Gastroenterology and Hepatology from Bed to Bench*, vol. 4, no. 4, pp. 171–174, 2011.

[5] S.-M. Alavian, P. Adibi, and M.-R. Zali, "Hepatitis C virus in Iran: epidemiology of an emerging infection," *Archives of Iranian Medicine*, vol. 8, no. 2, pp. 84–90, 2005.

[6] S. M. Alavian, K. Bagheri-Lankarani, M. Mahdavi-Mazdeh, and S. Nourozi, "Hepatitis B and C in dialysis units in Iran: changing the epidemiology," *Hemodialysis International*, vol. 12, no. 3, pp. 378–382, 2008.

[7] S.-M. Alavian, "We need a new national approach to control hepatitis C: it is becoming too late," *Hepatitis Monthly*, vol. 8, no. 3, pp. 165–169, 2008.

[8] S. M. Alavian, "Optimal therapy for hepatitis C," *Hepatitis Monthly*, vol. 4, no. 2, pp. 41–42, 2004.

[9] M. J. Alter, "Epidemiology of hepatitis C," *Hepatology*, vol. 26, no. 3, supplement 1, pp. 62S–5S, 1997.

[10] S. M. Alavian, K. B. Lankarani, S. H. Aalaei-Andabili et al., "Treatment of chronic hepatitis C infection: update of the recommendations from scientific leader's meeting-28th july

2011-Tehran, IR Iran," *Hepatitis Monthly*, vol. 11, no. 9, pp. 703–713, 2011.

[11] S. M. Alavian, M. Ahmadzad-Asl, K. B. Lankarani, M. A. Shahbabaie, A. B. Ahmadi, and A. Kabir, "Hepatitis C infection in the general population of Iran: a systematic review," *Hepatitis Monthly*, vol. 9, no. 3, pp. 211–223, 2009.

[12] S. Merat, H. Rezvan, M. Nouraie et al., "Seroprevalence of hepatitis C virus: the first population-based study from Iran," *International Journal of Infectious Diseases*, vol. 14, no. 3, pp. e113–e116, 2010.

[13] S. Touzet, L. Kraemer, C. Colin, P. Pradat, D. Lainor, F. Baily et al., "Epidemiology of hepatitis C virus infection in seven European Union countries: a critical analysis of the litreature. HENCORE Group. Hepatitis C European Network for Co-operative Resarch," *European Journal of Gastroenterology & Hepatology*, vol. 12, pp. 667–678, 2000.

[14] D. L. Wyles, "Moving beyond interferon alfa: investigational drugs for hepatitis C virus infection," *Topics in HIV Medicine*, vol. 18, no. 4, pp. 132–136, 2010.

[15] M. H. Ahmadipour, S. M. Alavian, S. Amini, and K. Azadmanesh, "Hepatitis C virus genotypes," *Hepatitis Monthly*, vol. 5, no. 3, pp. 77–82, 2003.

[16] R. S. Brown Jr. and P. J. Caglio, "Scope of worldwide hepatitis C problem," *Liver Transplantation*, vol. 9, no. 11, pp. S10–S13, 2003.

[17] S.-J. Hwang, S.-D. Lee, R.-H. Lu et al., "Hepatitis C viral genotype influences the clinical outcome of patients with acute posttransfusion hepatitis C," *Journal of Medical Virology*, vol. 65, no. 3, pp. 505–509, 2001.

[18] D. Lambert, "Zero-inflated poisson regression, with an application to defects in manufacturing," *Technometrics*, vol. 34, no. 1, pp. 1–14, 1992.

[19] D. Böhning, E. Dietz, P. Schlattmann, L. Mendonça, and U. Kirchner, "The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology," *Journal of the Royal Statistical Society A*, vol. 162, no. 2, pp. 195–209, 1999.

[20] D. B. Hall, "Zero-inflated poisson and binomial regression with random effects: a case study," *Biometrics*, vol. 56, no. 4, pp. 1030–1039, 2000.

[21] R. Ma, M. T. Hasan, and G. Sneddon, "Modelling heterogeneity in clustered count data with extra zeros using compound Poisson random effect," *Statistics in Medicine*, vol. 28, no. 18, pp. 2356–2369, 2009.

[22] M. T. Hasan, G. Sneddon, and R. Ma, "Pattern-mixture zero-inflated mixed models for longitudinal unbalanced count data with excessive zeros," *Biometrical Journal*, vol. 51, no. 6, pp. 946–960, 2009.

[23] R. P. Myers, C. Regimbeau, T. Thevenot et al., "Interferon for interferon naive patients with chronic hepatitis C," *Cochrane Database of Systematic Reviews*, no. 2, Article ID CD000370, 2002.

[24] T. Poynard, M.-F. Yuen, V. Ratziu, and C. Lung Lai, "Viral hepatitis C," *The Lancet*, vol. 362, no. 9401, pp. 2095–2100, 2003.

[25] T. Poynard, J. Mchutchison, Z. Goodman, M.-H. Ling, and J. Albrecht, "Is an "a la carte" combination interferon alfa-2b plus ribavirin regimen possible for the first line treatment in patients with chronic hepatitis C?" *Hepatology*, vol. 31, no. 1, pp. 211–218, 2000.

[26] M. W. Fried, M. L. Shiffman, K. Rajender Reddy et al., "Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection," *The New England Journal of Medicine*, vol. 347, no. 13, pp. 975–982, 2002.

[27] A. Agresti, *An Introduction to Categorical Data Analysis*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2007.

[28] A. Tsubota, N. Shimada, K. Yoshizawa et al., "Contribution of ribavirin transporter gene polymorphism to treatment response in peginterferon plus ribavirin therapy for HCV genotype 1b patients," *Liver International*, vol. 32, no. 5, pp. 826–836, 2012.

[29] F. Ionita-Radu, A. Rascanu, and B. Cheiab, "IL28B polymorphism—predictive factor of HCV infected genotype 1 individuals to treatment response and management of therapy," *Romanian Journal of Internal Medicine*, vol. 49, no. 2, pp. 99–104, 2011.

[30] K. Samimi-Rad, R. Nategh, R. Malekzadeh, H. Norder, and L. Magnius, "Molecular epidemiology of hepatitis C virus in Iran as reflected by phylogenetic analysis of the NS5B region," *Journal of Medical Virology*, vol. 74, no. 2, pp. 246–252, 2004.

[31] P. Ferenci, M. W. Fried, M. L. Shiffman et al., "Predicting sustained virological responses in chronic hepatitis C patients treated with peginterferon alfa-2a (40 KD)/ribavirin," *Journal of Hepatology*, vol. 43, no. 3, pp. 425–433, 2005.

[32] M. P. Manns, J. G. McHutchison, S. C. Gordon et al., "Peginterferon alfa-2b plus ribavirin compared with interferonalfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial," *The Lancet*, vol. 358, no. 9286, pp. 958–965, 2001.

[33] S. J. Hadziyannis, H. Sette Jr., T. R. Morgan et al., "Peginterferon-α2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose," *Annals of Internal Medicine*, vol. 140, no. 5, pp. 346–355, 2004.

[34] D. B. Strader, T. Wright, D. L. Thomas, and L. B. Seeff, "Diagnosis, Management, and Treatment of Hepatitis C," *Hepatology*, vol. 39, no. 4, pp. 1147–1171, 2004.

[35] G. C. Farrell, "New hepatitis C guidelines for the Asia-Pacific region: APASL consensus statements on the diagnosis, management and treatment of hepatitis C virus infection," *Journal of Gastroenterology and Hepatology*, vol. 22, no. 5, pp. 607–610, 2007.

[36] M. L. Shiffman, F. Suter, B. R. Bacon et al., "Peginterferon alfa-2a and ribavirin for 16 or 24 weeks in HCV genotype 2 or 3," *The New England Journal of Medicine*, vol. 357, no. 2, pp. 124–134, 2007.

[37] K. F. Lam, H. Xue, and Y. Bun Cheung, "Semiparametric analysis of zero-inflated count data," *Biometrics*, vol. 62, no. 4, pp. 996–1283, 2006.

[38] Y. Min and A. Agresti, "Random effect models for repeated measures of zero-inflated count data," *Statistical Modelling*, vol. 5, no. 1, pp. 1–19, 2005.

The Scientific
World Journal

Gastroenterology
Research and Practice

MEDIATORS
of
INFLAMMATION

Journal of
Diabetes Research

Disease Markers

Journal of
Immunology Research

International Journal of
Endocrinology

PPAR Research

Hindawi

Submit your manuscripts at
http://www.hindawi.com

BioMed
Research International

Journal of
Ophthalmology

Stem Cells
International

Evidence-Based
Complementary and
Alternative Medicine

Journal of
Obesity

Journal of
Oncology

Computational and
Mathematical Methods
in Medicine

Behavioural
Neurology

Parkinson's
Disease

AIDS
Research and Treatment

Oxidative Medicine and
Cellular Longevity