



Modeling the Number of Attacks in Multiple Sclerosis Patients Using Zero-Inflated Negative Binomial Model

Marjan Jamalian¹, Vahid Shaygannejad², Morteza Sedehi³, Soleiman Kheiri⁴

¹MSc in Biostatistics, Department of Epidemiology and Biostatistics, School of Public Health, Shahrekord University of Medical Sciences, Shahrekord, Iran

²Professor, Isfahan Neurosciences Research Center, Alzahra Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

³Associate Professor, Department of Epidemiology and Biostatistics, School of Public Health, Shahrekord University of Medical Sciences, Shahrekord, Iran

⁴Professor, Social Determinants of Health Research Center, Shahrekord University of Medical Sciences, Shahrekord, Iran

Abstract

Background and aims: Multiple sclerosis (MS) is an inflammatory disease of the central nervous system. The impact of the number of attacks on the disease is undeniable. The aim of this study was to analyze the number of attacks in these patients.

Methods: In this descriptive-analytical study, the registered data of 1840 MS patients referred to the MS clinic of Ayatollah Kashani hospital in Isfahan were used. The number of attacks during the treatment period was defined as the response variable, age at diagnosis, sex, employment, level of education, marital status, family history, course of disease, and expanded disability as the explanatory variables. The analysis was performed using zero-inflated negative binomial model via Bayesian framework in OpenBUGS software.

Results: Age at diagnosis (CI: -0.04, -0.20), marital status (CI: -0.56, 0.002), level of education (CI: -0.81, -0.26), Job (CI_{Housewives vs Employee}=[0.04, 0.64], CI_{Unemployee vs Employee}=[-1.10,0.008]), and course of disease (CI: -0.51, -0.08) had a significant effect on the number of attacks. In relapsing-remitting patients, the number of attacks was partial significantly affected by expanded disability status scale (EDSS) (CI: -0.019, 0.16).

Conclusion: Aging, being single (never married), high education, and not having a job decrease the number of attacks; therefore, lower age, being married, primary education, and being a housewife increase the number of attacks. An interventional or educational program is suggested in order to prevent the occurrence of further attacks in high-risk groups of patients and to increase their chances of recovery.

Keywords: Multiple sclerosis, Attack, Negative binomial, Zero-inflated, Markov chain Monte Carlo

*Corresponding Author:

Soleiman Kheiri,
Professor, Social
Determinants of Health
Research Center, Shahrekord
University of Medical
Sciences, Shahrekord, Iran
Soleiman.Kheiri@gmail.com

Received: 28 Aug. 2019
Accepted: 1 Jan. 2020
ePublished: 26 Mar. 2020



Introduction

Multiple sclerosis (MS) is a chronic inflammatory disease relevant to the central nervous system. In this illness, neural messages are led more slowly than usual.¹ In recent years, the number of MS patients has been increasing so much in Iran that the MS Association of Iran has announced that there are 70 000 cases of this disease among Iranians. The association has also declared that Isfahan province has the highest number of MS patients among all the provinces in Iran. This disease is more common among 20 to 45-year-old women and is subject to change with the passage of time. The relapsing-remitting is the most common course of MS. Approximately, 0.1% to 0.2% of the population in Europe and North America have MS². In the United States, there are 250 000-350 000

MS patients.³ It has been reported that in western countries, there is one MS patient per 1000 people.⁴ The patients experience attacks during the illness. The attack occurs either when the previous symptoms of the illness worsen or when new symptoms appear suddenly and/or gradually. The attack lasts for least 24 hours, but it could sometimes continue for months. The number of attacks and the time interval between them vary from patient to patient.^{5,6} The overall aim of treating this illness is to prevent attacks as much as possible, reduce the frequency of the attacks, increase the time interval between two attacks, and eventually prevent the progress of the disease and the disability resulting from the disease.⁷

This study aimed to analyze the factors affecting the frequency of attacks in MS patients. Since the

frequency of attacks is count data, a count model must be applied to analyze the data. There are two standard count models: the Poisson regression model and the negative binomial regression model. If the variance is higher than the mean of response variable, there is the problem of “overdispersion”, where the negative binomial model is more applicable than the Poisson model.⁸

Most of the count data in medical studies have a high zero frequency, i.e. the number of zeros is more than expected in a standard count model, which is called zero-inflated response. One of the suggested models to resolve this problem is the zero-inflated model.⁹ Considering that there are both problems of high zero frequency and overdispersion in this study, the zero-inflated negative binomial is used.

The zero-inflated model was introduced by Lambert through Poisson count distribution in 1992. The zero-inflated model is a two-part mixed model in which the first part is the zero inflation ratio and the second part is the complete Poisson count distribution, or complete negative binomial distribution is defined as follows:

$$f_{ZID}(Y) = \begin{cases} p_0 + (1-p_0)f_D(0|\theta) & y = 0 \\ (1-p_0)f_D(y|\theta) & y = 1, 2, \dots \end{cases}$$

$f_D(0|\theta)$ is the mass function of the probability of the distribution of D and p_0 parameter of the zero-inflated ratio. The mean and the variance of the zero-inflated model are respectively obtained as follows⁹:

$$E(Y) = (1-p_0)E(Y_D)$$

$$Var(Y) = (1-p_0) \left[Var(Y_D) + p_0 \left((E(Y_D))^2 \right) \right]$$

In the current study, the probability mass function D is considered based on the negative binomial distribution. This model shows what factors influence the increase or decrease in the number of patients' attacks. Furthermore, the impact of each independent variable on the number of attacks can be assessed. To this end, a logarithm of response variables is linked to explanatory variables. The statistical model is as follows:

$$\log(\mu_i) = \alpha + \beta_i x_i; i = 1, \dots, n$$

The aim of this study was to analyze the number of attacks in MS patients with the zero-inflated and hurdle models.

Materials and Methods

This descriptive-analytical study was conducted on the data collected from the files of 1840 patients who referred to the MS clinic of Ayatollah Kashani hospital of Isfahan. All patients who had MS were examined by a doctor and were included in the study. Their information was recorded. Patients suspected of MS were excluded from this study.

The data needed for this study were extracted by IMED software in which the longitudinal information of the patients had been recorded. The number of attacks during the treatment period was regarded as the response variable. The age at diagnosis, gender, marital status, job, educational level, family history, the course of disease, and the expanded disability status scale (EDSS) were regarded as explanatory variables in the zero-inflated negative binomial count model.

To estimate the parameters and obtain results, the Bayesian analysis was applied. In this method, by constructing the likelihood and the prior distribution that the researcher has considered, the posterior distribution of zero-inflated negative binomial model is formed. Since the researcher had little prior knowledge of the model's parameters, non-informative priors were taken into account. A normal distribution with mean zero and variance 1000 was thus used for the regression coefficient parameters of the model. Moreover, for the prior distribution of the proportion parameter, the Uniform distribution in (0, 1) interval was considered, with the Gamma distribution with parameters of (1, 1) considered for the dispersion parameter. The posterior distribution of zero-inflated negative binomial count model has a complicated and unknown form. As a result, for the inference of parameters of the posterior distribution, MCMC simulation methods were used. In this method, by successive sampling from full conditional posterior distributions, Markov chains were produced whose convergence indicates that they can be taken as samples of posterior distribution, and inference about parameters could be done based on them. To make sure of such convergence, different methods of recognizing the convergence of chains like autocorrelation graphs, density function graphs, and the Brook-Gelman-Rubin method were employed.¹⁰

Posterior summaries for the model parameters were computed based on 40 000 samples after a 5000 burn-in period. To enhance accuracy and precision of parameter estimation and to reduce autocorrelation, 1 out of 10 produced samples was selected for the posterior estimation. The relevant programming was carried out in OpenBUGS software; for recognizing the convergence, BOA package was implied in R software. Since some independent variables such as EDSS and

educational level contained missing data and omitting them would cause loss of data and a reduction in the model's power, MICE replacement method was applied to estimate the missing data, whose information was used in modeling. To estimate the missing values, VIM (Visualization and Imputation of Missing Values) and MICE (Multivariate Imputation by Chained Equations) packages were applied in R software.

Results

In summary, out of the 1840 patients, 1467 (79.7%) were female and 1336 (72.6%) had relapsing-remitting MS. Their age at diagnosis ranged from 9 to 73 with a mean of 34.13 ± 9.86 years. EDSS of patients ranged from 0 to 9 with a mean of 2.74 ± 2.17 . Their follow-up period ranged from 34 to 86 months with a mean of 61.6 ± 14.11 months. Additionally, 1288 (70%) of patients were married and 722 (39.2%) of them had university education. Other characteristics of patients and mean number of attacks of patients are presented in Table 1.

The number of the patients' attacks, as the response

variable, had a lower bound of 0 and an upper bound of 6 with a mean of 0.27 ± 0.67 attacks, while the median and mode of the number of attacks were both 0, showing a high frequency of 0 and a strong skewness in the data (Figure 1). Skewness coefficient of the frequency of attacks was equal to 3.37 with a variance of 0.43; therefore, the variance is 1.59 times as high as the mean, which suggests overdispersion in the number of attacks. The overdispersion was 19.5% which in turn made the overdispersion test of the response variable significant ($P < 0.001$).

The posterior summaries of the zero-inflated negative binomial count model including point estimation, with 90% credible interval, for modeling the number of attacks in MS patients are presented in Table 2. During the follow-up period, patients with relapsing-remitting MS have different variations; therefore, in consecutive periods, the disability can be increased or decreased or remain unchanged. Considering this, a separate count model was employed to study the effect of EDSS on the number of attacks in relapsing-remitting patients. The posterior summaries of EDSS parameter in zero-

Table 1. Mean Number of Attacks in MS Patients (Response Variable) Based on Independent Variables

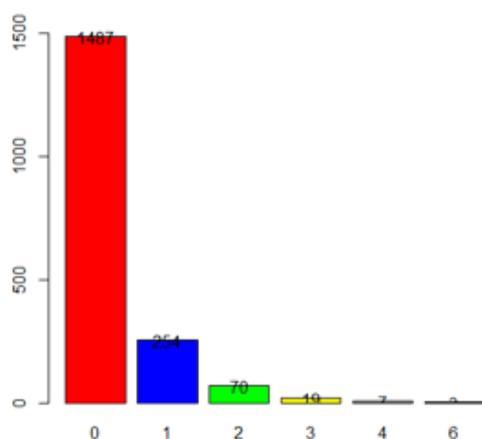
Variable	Level	Number of Patients		Number of Attacks
		Number	%	Mean \pm SD
Sex	Male	373	20.3	0.26 \pm 0.70
	Female	1467	79.7	0.27 \pm 0.64
Age (y)	<20	77	4.2	0.22 \pm 0.45
	20-40	1287	69.9	0.30 \pm 0.70
	40-60	452	24.6	0.21 \pm 0.56
	>60	24	1.3	0.12 \pm 0.45
Marital status	Single	472	25.7	0.23 \pm 0.61
	Married	1288	70	0.28 \pm 0.68
	Other (divorced, widow)	80	4.3	0.26 \pm 0.57
Level of education	Elementary	467	25.4	0.3 ⁴ \pm 0.76
	High school	651	35.4	0.31 \pm 0.67
	University	722	39.2	0.19 \pm 0.55
Job	Employed	500	27.2	0.22 \pm 0.67
	Housewife	1019	50.4	0.31 \pm 0.67
	Unemployed (jobless, retired)	92	5	0.16 \pm 0.50
	Student	229	12.4	0.24 \pm 0.60
Family history	Yes	339	18.4	0.28 \pm 0.67
	No	1501	81.6	0.27 \pm 0.65
Course of disease	RR	1336	72.6	0.25 \pm 0.63
	SP	504	27.4	0.32 \pm 0.70
EDSS	0-3.5	1233	92.3	0.27 \pm 0.65
	3.5-6.5	102	7.6	0.05 \pm 0.29
Follow-up (month)	30-50	539	29.3	0.27 \pm 0.57
	50-70	800	43.5	0.27 \pm 0.68
	>70	501	27.2	0.27 \pm 0.71

Table 2. Posterior Summaries of the ZNB Model for Modeling the Number of Attacks in MS Patients

	Parameters	Mean	Median	Standard Deviation	5%	95%
	Intercept*	-4.981	-4.999	0.446	-5.706	-4.221
	Age*	-0.0328	-0.0328	0.0078	-0.0458	-0.202
	Sex (female)	-0.3117	-0.3062	0.2071	-0.6602	0.0241
	Family history	0.0851	0.08403	0.1453	-0.151	0.3238
Marital status	Single*	-0.2792	-0.2795	0.1702	-0.5599	0.0022
	Divorced	0.1808	0.1833	0.2989	-0.3109	0.6694
Job	Housewives*	0.3197	0.3165	0.1926	0.0038	0.6392
	Unemployed*	-0.5445	-0.5399	0.3273	-1.105	0.0076
	Student	0.1263	0.1260	0.2403	-0.2680	0.5186
Education	High school	-0.1222	-0.1218	0.1456	-0.3607	0.1162
	University*	-0.5331	-0.5314	0.1658	-0.8077	-0.2605
	Course of disease*	-0.2983	-0.2995	0.1316	-0.5133	-0.0819
	Inflated ratio*	0.2357	0.2345	0.1321	0.02868	0.4515
	Dispersion*	0.9252	0.7946	0.4576	0.4792	1.814
	Deviance *	2383	2383	5.271	2376	2393

*Significant at 0.1

Baseline marital status: married; Baseline job: employee; Baseline education: < high school

**Figure 1.** Frequency of Attacks in MS Patients in Follow-up Time.

inflated negative binomial count model including point estimation, with 90% credible interval, for modeling the number of attacks in relapsing-remitting MS patients are illustrated in Table 3.

According to confidence intervals of zero-inflated negative binomial model parameters, independent variables of age, being single vs. married, having an academic degree or not, employment status (including housewives), being unemployed (retired or jobless), and course of the disease had a significant impact on the increase or decrease in the number of attacks during the treatment period. There was no significant relationship between family history and the number of attacks. According to the confidence interval of

90%, EDSS has a significant effect on the frequency of attacks in relapsing-remitting patients.

Discussion

In this study, some factors influencing the number of attacks in MS patients were identified. The patients who were surveyed had referred to the MS clinic of Kashani hospital in Isfahan for follow-up treatment. Most of these patients were female (79.7%) aged 20-40 (69.9%), married (70%), with university education (39.2%), housewives (50.4%), belonging to relapsing-remitting course of the disease (72.6%), and with mild EDSS (92.3%).

The results showed that age, marital status, educational level, employment, and disease course affected the increase or decrease in the frequency of attacks. Older patients, the single, and those with university education experienced fewer attacks. Single patients had 0.75 times fewer attacks than married ones do, which was probably due to marital life difficulties. Patients with university education who underwent treatment experienced 0.59 times fewer attacks compared to those with elementary school education. It can thus be said that education plays a big part in detecting the illness, its process, and controlling it, with well-educated people experiencing fewer attacks indicating their partial recovery. The frequency of attacks in housewives was 38% as high as the employed patients; household chores are regarded as physically demanding activities which may have undesirable impacts on the disease and eventually cause more

Table 3. Posterior Summaries of EDSS Parameters in ZNB Model for Modeling the Number of Attacks in RR MS Patients

Parameter	Mean	Median	Standard Deviation	5%	95%
Intercept*	-5.120	-5.132	0.2524	-5.504	-4.69
EDSS*	0.07	0.0708	0.0544	-0.019	0.16
Inflated ratio*	0.3204	0.3227	0.1611	0.0467	0.566
Dispersion*	0.8274	0.6749	0.5009	0.3558	1.815
Deviance*	1723	1723	2.477	1721	1728

*Significant at 0.1

attacks. Unemployed patients (jobless or retired) had 0.58 times fewer attacks than the employed patients; therefore, the kind of job and activities impact the frequency of attacks. Since this disease is accompanied by physical disability, choosing the appropriate job and having the incentive to work influence the disease. As jobless or retired patients do not have work stress or do taxing activities, they experience fewer attacks. Patients with progressive conditions experience 0.74 times fewer attacks than those with relapsing-remitting conditions, suggesting that the disease usually starts with an attack in patients with progressive conditions and as time passes, the expanded disability deteriorates in turn. However, in relapsing-remitting patients, there are sudden attacks after which the patients experience a quiet period and are partially fine before the occurrence of another attack whose symptoms and intensity can be different from the previous ones. In summary, the frequency of attacks is higher in relapsing-remitting patients. In these patients, the number of attacks has a direct relation with the extent of disability; therefore, those with higher EDSS experience more attacks.

Most of the patients surveyed in this study were women. Similarly, MS is more common among women in other studies conducted in different countries. For instance, in a survey carried out on 1463 patients in Italy, 67.2% of patients were women.¹¹ In another study in Canada, 2837 patients were studied, 70.4% of whom were women.¹² Accordingly, it can be said that MS is 2 or 3 times more common in women than in men. Additionally, 72.6% of the patients in this study belonged to the relapsing-remitting category. A similar study in the United States and Canada showed that 85-90% of the patients belonged to this category.¹³

In count data, especially those with a high frequency of zeros, it is essential to apply zero-inflated count model to obtain efficient and reliable estimates. In this study, the response variable of count data had a high frequency of zeros and significant overdispersion. Hence, the zero-inflated negative binomial count model was applied to analyze the data. One of the limitations of this study was the inequality of the

patients' follow-up period, so to correct the results, each patient's follow-up period was considered as an offset variable in the zero-inflated negative binomial count model.¹⁴

Searching the data banks indicated that no studies have ever been conducted regarding the factors affecting the number of attacks, making this study the first in this area.

Conclusion

Given the fact that patients recognized at an earlier age, married vs. single patients, those with elementary education vs. the ones with an academic degree, employed and housewife patients vs. unemployed ones, and those with relapsing-remitting conditions experienced more attacks in their follow-up treatment, they should be instructed and certain treatment plans be administered to reduce the frequency of attacks and consequently increase their chances of recovery.

References

1. Confavreux C, Vukusic S. Natural history of multiple sclerosis: a unifying concept. *Brain*. 2006;129(Pt 3):606-16. doi: 10.1093/brain/awl007.
2. Wynn DR, Rodriguez M, O'Fallon WM, Kurland LT. A reappraisal of the epidemiology of multiple sclerosis in Olmsted County, Minnesota. *Neurology*. 1990;40(5):780-6. doi: 10.1212/wnl.40.5.780.
3. Wingerchuk DM, Lucchinetti CF, Noseworthy JH. Multiple sclerosis: current pathophysiological concepts. *Lab Invest*. 2001;81(3):263-81. doi: 10.1038/labinvest.3780235.
4. Sadovnick AD, Ebers GC. Epidemiology of multiple sclerosis: a critical overview. *Can J Neurol Sci*. 1993;20(1):17-29. doi: 10.1017/s0317167100047351.
5. Burgess M. Improving relapse management in multiple sclerosis: a nurse-led model. *British Journal of Neuroscience Nursing*. 2011;7(1):441-4. doi: 10.12968/bjnn.2011.7.1.441.
6. Burgess M. *Multiple Sclerosis: Theory and Practice for Nurses*. Wiley; 2002.
7. Rudick RA, Cohen JA, Weinstock-Guttman B, Kinkel RP, Ransohoff RM. Management of multiple sclerosis. *N Engl J Med*. 1997;337(22):1604-11. doi: 10.1056/nejm199711273372207.
8. Dalrymple ML, Hudson IL, Ford RPK. Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. *Comput Stat Data Anal*. 2003;41(3-4):491-504. doi: 10.1016/S0167-9473(02)00187-1.
9. Neelon B, O'Malley AJ, Smith VA. Modeling zero-modified

- count and semicontinuous data in health services research Part 1: background and overview. *Stat Med.* 2016;35(27):5070-93. doi: 10.1002/sim.7050.
10. Kheiri S, Meshkani MR, Faghihzadeh S. A correlated frailty model for analysing risk factors in bilateral corneal graft rejection for Keratoconus: a Bayesian approach. *Stat Med.* 2005;24(17):2681-93. doi: 10.1002/sim.2113.
 11. Liguori M, Marrosu MG, Pugliatti M, Giuliani F, De Robertis F, Cocco E, et al. Age at onset in multiple sclerosis. *Neurol Sci.* 2000;21(4 Suppl 2):S825-9. doi: 10.1007/s100720070020.
 12. Tremlett H, Paty D, Devonshire V. Disability progression in multiple sclerosis is slower than previously reported. *Neurology.* 2006;66(2):172-7. doi: 10.1212/01.wnl.0000194259.90286.fe.
 13. Kremenchutzky M, Rice GP, Baskerville J, Wingerchuk DM, Ebers GC. The natural history of multiple sclerosis: a geographically based study 9: observations on the progressive phase of the disease. *Brain.* 2006;129(Pt 3):584-94. doi: 10.1093/brain/awh721.
 14. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models.* Boston, MA: Springer; 2012.