Gastroenterology and Hepatology From Bed to Bench.

ORIGINAL ARTICLE

Specific count model for investing the related factors of cost of GERD and functional dyspepsia

Alireza Abadi[1], Asma Pourhoseingholi[2], Samira Chaibakhsh[2], Azadeh Safaee[3], Bijan Moghimi-Dehkordi[3]

[1]Department of Community Medicine & Health, Medical School, Shahid Beheshti University of Medical Sciences, Tehran, Iran

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

[3]Department of Control & Prevention of Diseases, Deputy of Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

ABSTRACT

Aim: The purpose of this study is to analyze the cost of GERD and functional dyspepsia for investing its related factors.

Background: Gastroesophageal reflux disease (GERD) and dyspepsia are the most common symptoms of gastrointestinal disorders. Recent studies showed high prevalence and variety of clinical presentation of these two symptoms imposed enormous economic burden to the society. Cost data related to economic burden have specific characteristics. Therefore, this kind of data needs to be specific models. Poisson regression (PR) and negative binomial regression (NB) are the models that were used for analyzing cost data in this paper.

Patients and methods: This study was designed as a cross-sectional household survey from May 2006 to December 2007 on a random sample of individuals in the Tehran province, Iran to find the prevalence of gastrointestinal symptoms and disorders and its related factors. The cost in each item was counted. PR and NB were carried out to the data respectively. Likelihood ratio test was performed for comparison between models. Also Log likelihood, akaike information criterion (AIC) and bayesian information criterion (BIC) were used to compare performance of the models.

Results: According to likelihood ratio test and all three criterions that we used to compare performance of the models, NB was the best model for analyzing this cost data. Sex, age and insurance statues were being significant.

Conclusion: PR and NB models were carried out for this data and accordingly the results improved to fit the NB model over PR. It clearly indicates that over-dispersion is involved due to unobserved heterogeneity and/or clustering. NB model in cost data is more appropriate fit than PR.

Keywords: GERD, Dyspepsia, Cost, Count models, Zero inflated models.

(Please cite as: **Abadi A, Pourhoseingholi A, Chaibakhsh S, Safaee A, Moghimi-Dehkordi B. Specific count model for investing the related factors of cost of GERD and functional dyspepsia. Gastroenterol Hepatol Bed Bench 2013;6(Suppl.1):S122-S128).**

Introduction

Gastroesophageal reflux disease GERD is one of the most common symptoms of gastrointestinal disorders characterized by heartburn and/or acid regurgitation (1). Recent studies showed high prevalence and variety of clinical presentation of GERD imposed enormous economic burden to the society (2, 3). Another high prevalence gastrointestinal symptom is dyspepsia, which refers to a group of upper gastrointestinal symptoms. An international definition, Rom criterion is used for diagnosis of functional dyspepsia (4). However, this disorder is not life threatening but the impact of this disorder on patient and society is considerable (5). Recently,

Received: 3 June 2013 *Accepted*: 14 August 2013
Reprint or Correspondence: Asma Pourhoseingholi, PhDs. Department of Biostatistics, School of Paramedical Science, Shahid Beheshti University of Medical Science, Iran
E-mail: asma_phg@yahoo.com

there are lots of studies that investigated the economic burden of these symptoms worldwide (2, 3, 6-9). These studies have shown the direct costs of GERD and dyspepsia ranged between PPP $172 (purchasing power parity dollars) and PPP $176 per person per year. Given the importance of affair, there are some studies for estimating of this economics burden in Iran recently. Rezai Lashkajani and Moghimi Dehkordi published their paper in 2007 and 2011 respectively (3, 9). On the other hand, the continuing rise in charge for medical care worldwide has increased interest in precise estimating the cost of disease and the impact of specific treatments on the cost of care (10). So we have seen a heightened interest in studying the cost of health care, and medical cost data in recent years (11). Specific characteristic of cost data is its distributions that are difficult to describe using standard approaches like ordinary least square regression for analyzing (12). Poisson model is one of the approaches that is used for analyzing such data like cost data. However, due to over-dispersion, related problems of poisson regression, that frequently arise in count data, another model like negative binomial use for this data (13). In this paper we use poisson regression (PR) and negative binomial (NB) for analyzing the cost of GERD and dyspepsia.

Patients and Methods

This study was designed as a cross-sectional household survey from May 2006 to December 2007 on a random sample of individuals in the Tehran province, Iran to find the prevalence of gastrointestinal symptoms and disorders and its related factors. Among this random sample, a total of 18180 adults (age > 18) were selected up (more details in refs (9, 23-29)). Subsequently, in the first step, trained health staff from the corresponding local health center contacted all selected random (18180), door to door and face to face, and asked them to participate in the interviews consisting of certain questionnaires. The ethics committee of the research center had approved the research protocol of this study for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Science. The questionnaire was filled; a valid and reliable questionnaire designed specifically for this study. This questionnaire included the symptoms of dyspepsia (23) (based on ROM III criteria), GERD (i.e. heartburn and regurgitation only) and their frequency in the last 6 month (27). In this questionnaire beside the information about listed symptoms, the frequency of the utilization of health services/resources including visits to physicians, drug intake, laboratory tests, hospitalizations and the productivity loss due to GERD/dyspepsia symptoms were reported. GERD was diagnosed as having heartburn or acid regurgitation on a weekly basis for the last 6 month. Dyspepsia was diagnosed, based on ROM III criteria, as a person who experienced one or more of the following symptoms for the past 3 month with onset of symptoms at least 6 month before diagnosis. The symptoms are as follow: bothersome postprandial fullness, early satiety, epigastric pain and epigastric burning sensation. The methodology of cost analysis in this study was similar to that other cost analysis in Iran (3, 30). In the calculation and analysis of cost, direct and indirect sources of expenses including physicians visit, drugs, laboratory test, hospitalization and day off work due to GERD and dyspepsia related symptoms were considered. The basic plans of the economic burden are as following:

Direct cost = physician visit + drug + laboratory test + hospitalization

Indirect cost = number of days with total productivity loss + number of days with at least 30% low functionality

Total cost = direct cost + indirect cost

All estimated costs were converted to PPP$, because of cross-country comparison. PPP$ is an economic technique to determine the relative values of two currencies (9).

Statistical methods

Frequency distribution and descriptive statistics such as mean, standard deviation and percentage were calculated according to standard methods. Outcome variable is the cost of GERD and dyspepsia. Two modeling approach were used for analyzing the relation between some demographic factors such as sex, age, marital status, education and other factors such as insurance status with cost of GERD and dyspepsia .The poisson regression (PR) is one of the models from general linear models (GLM) for describing count outcomes or proportion/rates (13). This model assumes response had a poisson distribution. Count data often vary more than 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 (13). The negative binomial (NB), is another model from GLM as an alternative to the PR model, is a solution to account for over-dispersion due to unobserved heterogeneity (31). This model helps in adjusting the standard errors of the regression coefficients and provides a more flexible approach for prediction of the count outcome. To compare performance of the models, there are various methods such as log likelihood, akaike information criterion (AIC) and bayesian information criterion (BIC). The p-values less than 5% were considered as significant results. All analysis was performed using STATA package.

Results

A total of 1929 patients were eligible and entered this study. Of those in the study, 1186 (61.4%) of patient had costs for gastrointestinal disorders. The mean and standard deviation of the costs per patient were 124.56 and 399.707 PPP$ respectively. Median of this cost was 47PPP$. The mean age was 46.22 (standard deviation, 16.55) years and range 16-98 years.

The results of a univariate analysis of cost of GERD and dyspepsia between categorical variables are shown in table 1. According to these results, there was a statistical significant difference between cost of GERD and dyspepsia in educational level, marital statues and insurance statues. However, there was not a statistically significant difference between male and female.

Table 1. Results of a univariate analysis of categorical variables

	Median	Range	p-value
Sex			0.342
Women	45	7230	
Men	42	7614	
Having insurance			<0.001
Yes	47	7614	
No	38	7230	
Marital statues			0.006
Single	35	1201	
Widow	48	3534	
Married	45	7614	
Education			<0.001
Diploma or Upper diploma	36	7230	
Lower diploma	48	7614	

Finally PR regression and NB regression were performed for analyzing the relation between response and all factors simultaneously. In the PR model all covariates were significant.

The significant Pearson chi-square goodness of fit (GOF) test ($p<0.001$) along with other characteristics of model fit indicated the PR model produced a poor fit for cost data. According to the results of the NB model, age, sex and insurance statues had a significant relation with cost of patient. These results showed that gender of

Table 2. Results of negative binomial model

Variable	Category	Estimation	SE	p-value	Odds ratio
Age		0.01	0.002	<0.001	1.01
Sex					
	Women	0.14	0.06	0.018	1.16
	Men*	-	-		-
Having insurance					
	Yes	0.14	0.06	0.04	1.16
	No*	-	-		-
Marital statues					
	Single	-0.06	0.09	0.53	0.94
	Widow	0.04	0.11	0.66	1.04
	Married*	-	-		
Education					
	Upper diploma	-0.08	0.06	0.6	0.92
	Lower diploma*	-	-		-

* Reference group

female (OR=1.16; 95% CI 1.01, 1.36; P=0.03) and high level of age (OR=1.01; 95% CI 1.00, 1.02; P<0.001) effected to increase the rate of costs. Beside for insured patient the rate of cost of their disease increased in compression of patient without insurance (OR=1.16; 95% CI 1.00, 1.34; P=0.04). On the other hand, in this model, the estimated dispersion statistic (α) was 2.4 (95% CI: 2.29, 2.52). A significant likelihood ratio test ($p < 0.001$) of dispersion statistic from zero favored the NB model over the PR model. The model fit characteristics are shown in table 2.

The entire criterion showed that the NB model was better than PR model. Table 3 showed the results of 3 different criterions for comparison between NB and PR model.

Table 3. Comparison of model fit characteristics

	PR	NB
AIC*	575547.4	21307.6
BIC†	575585.9	21346.1
Logliklihood	-287766.7	-10646

*Akaike Information Criterion; † Bayesian Information Criterion

Discussion

Cost analyzing and related studies in clinical research, has been a must attention in last years (10, 11). In this study we examined and analyzed the imposed cost of two high prevalence gastrointestinal symptoms in Tehran province population. This research was a part of a cross-sectional study that carried out in Tehran province in 2006 to 2007, aimed to determine the prevalence and related costs of some of the most common gastrointestinal disorders in this population (9, 24-28, 34, 35). GERD and dyspepsia are two common gastrointestinal disorders that have different prevalence in the world (1, 4, 5, 7, 8, 34, 36-40). According to this study the prevalence of GERD and dyspepsia was 8.85% and 8.5% respectively (23, 28). On the other hand, the cost of GERD, dyspepsia and overlap of them was 111.4, 120.2 and 135 PPP$ (9). Therefore, it seems the cost of these disorders and the related factors that are affecting these costs may be important. Special feature of cost data makes the data analysis special (41).In another Iranian survey, the direct costs of GERD and dyspepsia have ranged from PPP$172 to PPP$176 per person per year and comprised more than 80% of the total cost of the disease. But this study was not population based, and the subjects were referred to a gastroenterology clinic for upper endoscopy. Their patients might have been suffering from relatively more severe or prolonged GERD and dyspepsia to be included in the cost

analysis, which could cause an overestimation of the economic burden of illness compared with population studies such as ours (3).

The main feature is the cost data are count. Hence, modeling for cost data is needed. The first count model that researchers refer to it first is the PR model. However, because of the related problem of this model (over-dispersion), this model sometimes is not suitable and its results is not fully trusted. According to the results, the NB model and PR model have some differences with each other. In the PR model all the factors were significant. However, in the NB model just 3 of them were significant. Besides, the likelihood ratio test in NB model showed this data were over diapered. In this situation using PR model because the factors that not have significant effect will be significant by mistake.

According to the results of the goodness of fit, improved fit of the NB model over PR clearly indicates that over-dispersion is involved due to unobserved heterogeneity and/or clustering. Also, AIC, BIC and log liklihood criterion showed that NB model was better than the PR model. About related factors that they affected the costs. According to NB (the best model), female and older people had more cost than others. It seems this result was logical. Because according to previous papers all of these factors (sex and age) are the risk factors of GERD and dyspepsia (23, 28). There is some difference for example, in inherent characteristics between men and women caused women to cost more than men.

On the other hand, due to lack of insurance, the rate of costs of patient increased significantly. The reason is may be patients do not have insurance, due to high costs of treatments and lack of financial ability, do not refer for treatment.

References

1. Wang JH, Luo JY, Dong L, Gong J, Tong M. Epidemiology of gastroesophageal reflux disease: a general population-based study in Xi'an of Northwest China. World J Gastroenterol 2004; 10:1647-51.

2. Dean BB, Crawley JA, Schmitt CM, Wong J, Ofman JJ. The burden of illness of gastro-oesophageal reflux disease: impact on work productivity. Aliment Pharmacol Ther 2003; 17:1309-17.

3. Rezailashkajani M, Roshandel D ,Shafaee S, Zali MR. High prevalence of reflux oesophagitis among upper endoscopies of Iranian patients. Eur J Gastroenterol Hepatol 2007; 19:499-506.

4. Drossman DA. Rome III: the new criteria. Chin J Dig Dis 2006; 7:181-85.

5. Mahadeva S, Goh KL. Epidemiology of functional dyspepsia: a global perspective. World J Gastroenterol 2006; 12:2661-66.

6. Brook RA, Wahlqvist P, Kleinman NL, Wallander MA, Campbell SM, Smeeding JE. Cost of gastro-oesophageal reflux disease to the employer: a perspective from the United States. Aliment Pharmacol Ther 2007; 26:889-98.

7. Gisbert JP, Cooper A, Karagiannis D. Impact of gastroesophageal reflux disease on patients' daily lives: a European observational study in the primary care setting. Health Qual Life Outcomes 2009;7: 60.

8. Henke CJ, Levin TR, Henning JM, Potter LP. Work loss costs due to peptic ulcer disease and gastroesophageal reflux disease in a health maintenance organization. Am J Gastroenterol 2000; 95:788-92.

9. Moghimi-Dehkordi B, Vahedi M, Khoshkrood MB, Kasaeian A, Safaee A, Habibi M, et al. Economic burden of gastro-oesophageal reflux disease and dyspepsia: a community-based study. Arab J Gastroenterol 2011; 12:86-89.

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

11. Myers RH, Montgomery DC. A tutorial on generalized linear models. J Qual Technol 1997; 29:274–91.

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

13. Agresti A, Editor. An introduction to categorical data analysis. 2nd ed. New York: John Wiley & Sons; 2007.

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

15. Rose CE, Martin SW, Wannemuehler KA, Plikaytis BD. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. J Biopharm Stat 2006; 16:463-81.

16. 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.

17. Hardin J, Hilbe J, Editors. Generalized linear models and extensions. Texas: A Stata Press Publication; 2007.

18. Mullahy J. specification and testing of some modified count data models. J Econometrics 1986; 33:341-65.

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

20. Picard R, Cook DJ. Cross-validation of regression models. J Amer Statist Assoc 1984; 79:575–83.

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

22. 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.

23. Barzkar M, Pourhoseingholi MA, Habibi M, Moghimi-Dehkordi B, Safaee A, Pourhoseingholi A, et al. Uninvestigated dyspepsia and its related factors in an Iranian community. Saudi Med J 2009; 30:397-402.

24. Khoshkrood-Mansoori B, Pourhoseingholi MA, Safaee A, Moghimi-DehkordiB, Sedigh-Tonekaboni B, Pourhoseingholi A, et al. Irritable bowel syndrome: a population based study. J Gastrointestin Liver Dis 2009; 18(4):413-8.

25. Pourhoseingholi A, Pourhoseingholi MA, Vahedi M, Moghimi-Dehkordi B, Maserat AS, Zali MR. Relation between demographic factors and hospitalization in patients with gastrointestinal disorders, using quantile regression analysis. East Afr J Public Health 2009; 6 Suppl(1):45-7.

26. Pourhoseingholi MA, Kaboli SA, Pourhoseingholi A, Moghimi-Dehkordi B, Safaee A, Mansoori BK, et al. Obesity and functional constipation; a community-based study in Iran. J Gastrointestin Liver Dis 2009; 18:151-55.

27. Solhpour A, Pourhoseingholi MA, Soltani F, Zarghi A, Habibi M, Ghafarnejad F, et al. Gastro-esophageal reflux symptoms and body mass index: no relation among the Iranian population. Indian J Gastroenterol 2008; 27:153-55.

28. Solhpour A, Pourhoseingholi MA, Soltani F, Zarghi A, Solhpour A ,Habibi M, et al. Gastro-oesophageal reflux disease and irritable bowel syndrome: a significant association in an Iranian population. Eur J Gastroenterol Hepatol 2008; 20:719-25.

29. Sorouri M, Pourhoseingholi MA, Vahedi M, Safaee A, Moghimi-Dehkordi B, Pourhoseingholi A, et al. Functional bowel disorders in Iranian population using Rome III criteria. Saudi J Gastroenterol 2010; 16:154–60.

30. Roshandel D, Rezailashkajani M, Shafaee S, Zali MR. A cost analysis of functional bowel disorders in Iran. Int J Colorectal Dis 2007; 22:791-99.

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

32. Cheung, YB. Zero-inflated models for regression analysis of count data: a study of growth and development. Stat Med 2002; 21:1461-69.

33. Yau KK, Wang K, Lee AH. Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros. Biometrical J 2003; 4:437–52.

34. Colin Jones DG, Bloom B, Bodemar G. Management of dyspepsia: report of a working party. Lancet 1988; 1:576-79.

35. Pourhoseingholi A, Pourhoseingholi MA, Vahedi M, Safaee A, Moghimi-Dehkordi B, Ghafarnejad F, et al. Burden of hospitalization for gastrointestinal tract cancer patients - Results from a cross-sectional study in Tehran. Asian Pac J Cancer Prev 2009; 10:107-10.

36. Afifi AA, Kotlerman JB, Ettner SL, Cowan M. Methods for improving regression analysis for skewed continuous or counted responses. Annu REV Publ Health 2007; 28:95-111.

37. Fass R. Epidemiology and pathophysiology of symptomatic gastroesophageal reflux disease. Am J Gastroenterol 2003; 98: S2-7.

38. Poelmans J, Tack J. Extraoesophageal manifestations of gastro-oesophageal reflux. Gut 2005; 54:1492-99.

39. Ruigómez A, García Rodríguez L A, Wallander MA, Johansson S, Graffner H, Dent J. Natural history of gastro-oesophageal reflux disease diagnosed in

general practice. Aliment Pharmacol Ther 2004; 20:751-60.

40. Talley NJ, Stanghellini V, Heading RC, Koch KL, Malagelada JR, Tytgat GN. Functional gastro duodenal disorders. Gut 1999; 45:II37-42.

41. Blough DK. Using generalized linear models to assess medical care costs. Health Serv Outcomes Res Methodol 2000; 1:185-202.

42. Boucher J, Denuit M, Guillen M. Risk classification for claim counts: A comparative analysis of various zero inflated mixed Poisson and hurdle models. N Am Actuar J 2007; 11:110–31.

43. Barondess DA, Meyer EM, Boinapally PM, Fairman B, Anthony JC. Epidemiological evidence on count processes in the formation of tobacco dependence. Nicotine Tob Res 2010; 12:734-41.

44. Bergemann TL, Huang Z. A new method to account for missing data in case-parent triad studies. Hum Hered 2009; 68:268-77.

45. Carrel M, Escamilla V, Messina J, Giebultowicz S, Winston J, Yunus M, et al. Diarrheal disease risk in rural Bangladesh decreases as tube well density increases: a zero-inflated and geographically weighted analysis. Int J Health Geogr 2011; 10:41.

46. Carrel M, Voss P, Streatfield PK, Yunus M, Emch M. Protection from annual flooding is correlated with increased cholera prevalence in Bangladesh: a zero-inflated regression analysis. Environ Health 2010; 9:13.

47. Goetzel RZ, Gibson TB, Short ME, Chu BC, Waddell J, Bowen J, et al. A multi-worksite analysis of the relationships among body mass index, medical utilization, and worker productivity. J Occup Environ Med 2010; 52:S52-58.