

## Association between duration of heartburn and patient characteristics: a quantile regression analysis

Mohamad Amin Pourhoseingholi<sup>1</sup>, Soghrat Faghihzadeh<sup>2</sup>, Manijeh Habibi<sup>1</sup>, Azadeh Safaee<sup>1</sup>, Fatemeh Qafarnejad<sup>1</sup>, Mohammad Rostami Nejad<sup>1</sup>, Bijan Moghimi-Dehkordi<sup>1</sup>, Mohammad Reza Zali<sup>1</sup>

<sup>1</sup> Research Institute for Gastroenterology and Liver Disease, Shahid Beheshti University, M.C., Tehran, Iran

<sup>2</sup> Department of Biostatistics, Tarbiat Modares University, Tehran, Iran.

### ABSTRACT

**Aim:** To determine the associated factors of reflux duration, using quantile regression model.

**Background:** Reflux is one of the most prevalent gastrointestinal disorders. Statistical tools are widely used to assess the associated factors on reflux severity and frequency.

**Patients and methods:** A door to door questionnaire has been used to evaluate gastrointestinal symptoms including reflux and heartburn between May and December 2006 in Tehran. A total of 790 adult patients with gastrointestinal problems were selected using random sampling method. Among them 208 persons were found to have heartburn. A quantile regression was employed to assess the associated factors and the results were compared to linear regression and Cox semi parametric model.

**Results:** Heartburn duration had dramatically high negative skewness. Quantile regression analysis indicated significant increase in the given decile of heartburn duration associated with a unit increase in given patient characteristic containing sex, pain severity and pain after having some specific food. The results from linear regression and Cox model showed poor reliability in comparison to Quantile regression.

**Conclusion:** The results have demonstrated that using quantile regression leads to better interpretation and richer inference about association between heartburn duration and patient characteristics.

**Keywords:** *Quantile regression, Linear regression, Heartburn.*

*(Gastroenterology and Hepatology From Bed to Bench 2008;1(2):71-77).*

### INTRODUCTION

Gastroesophageal reflux disease (GERD) is currently one of the most prevalent gastrointestinal disorders (1). Symptoms associated with gastroesophageal reflux disease are common, with an incidence of approximately 20% in the general population (2, 3). It often interferes with patients' quality of life and can develop into esophagitis or

cause other complications (4, 5). Statistical tools such as linear regression are widely used to assess the factors associated with reflux severity and frequency. Multiple linear regression allows researchers to assess how the mean of a conditional distribution changes with respect to patient characteristics and Cox regression allows one to assess the impact of patient characteristics on the relative instantaneous hazard of an event. But in most cases of interest, the biometrical measurements like duration data are not distributed

*Received:* 21 October 2007 *Accepted:* 11 March 2007

**Reprint or Correspondence:** Mohamad Amin Pourhoseingholi, MD. Research Institute for Gastroenterology and Liver Diseases, Shahid Beheshti University, M.C., Iran.

**E-mail:** Amin\_phg@yahoo.com

normally. This kind of data depends on time so parametric models such as linear regression that need the assumption of normality cannot be informative enough (6). In this situation quantiles are more favorable to depict the distribution of response variable. Quantile regression (7, 8) is an econometric regression model in which a specified conditional quantile (or percentile) of the outcome variable is expressed as a linear function of subject characteristics. This is in contrast to linear regression in which the mean of a continuous response is expressed as a linear function of a set of independent or predictor variables. This model is an important tool for estimating conditional quantile models that has been used in many empirical studies and has been studied extensively in theoretical statistics. One of most appealing features of quantile regression is its ability to estimate quantile-specific effects that describe the impact of covariates not only on the center but also on the tails of the outcome distribution. Therefore, this technique is capable of providing a more complete statistical analysis of the stochastic relationships among random variables. This model is not frequently used only in the econometrics literature (9, 10), but also it has been recently used in the field of biostatistics to find conditional quantile functions to provide a more complete view (11-13).

Duration data are commonly used in applied econometrics and biometrics and although there is a variety of readily available estimators for popular models such as the accelerated failure time model and the proportional hazard model (14), quantile regression is recently emerging as an alternative to these conventional models (15, 16).

Our goal of this study was to determine the associations between heartburn duration and demographic characteristics by quantile regression and compare this model to linear regression and Cox semi parametric model.

## PATIENTS and METHODS

### *Quantile regression analysis*

By modeling the distribution of the duration in a flexible semi-parametric way, quantile regression does not impose modeling assumptions that may not hold. This technique is intended to estimate and conduct inference about conditional quantile functions. In linear regression with least square technique, the coefficients are estimated so as to minimize the sum of squared deviations from the regression lines but in quantile regression, the  $p$ th quantile of the conditional distribution can be modeled as a linear function of subject characteristics. For example if someone wants to estimate median regression (17), then  $p$ th quantile is equal to 0.5 for all observations.

But there is no closed-form expression to estimation and an iterative algorithm must be used to obtain the coefficient estimates (7, 18). Similarly, the standard errors of the regression coefficients can be estimated using methods described by Koenker and Bassett (19) and elaborated by Rogers (20) and another technique that recommended estimating the standard errors of regression coefficients for quantile regression models is bootstrap (21). In practical works, quantile regression is similar to linear regression. Some macros have been developed in software such as SAS (PROC QUANTREG) and R (package quantreg) to fit the models of interest.

### *Cox Proportional Hazards Model:*

Cox model (14) has become the most used procedure for modeling the relationship of covariates to a survival or other time-to-event data (22).

Cox proposed a semi-parametric model for the hazard function that allows the addition of explanatory variables, or covariates, but keeps the baseline hazard as an arbitrary, unspecified, nonnegative functional of time. Because the baseline hazard is not assumed to be of a

parametric form, Cox's model is referred to as a semi-parametric model for the hazard function.

#### *Heartburn duration risk factors*

A door to door questionnaire was used to evaluate gastrointestinal symptoms including heartburn and its history (23). The questionnaire was completed through a cross-sectional study which was conducted between May and December 2006 in Damavand and Firoozkough cities, located in Tehran province. Study subjects were interviewed by using questionnaire. The questionnaire included socio-demographic, health relevant life style and clinical factors and GI symptoms. Those who had any GI symptoms or signs were referred to a general practitioner for further evaluation and functional gastrointestinal questionnaires were completed. We selected 790 adult patients with gastrointestinal problems using random sampling method of which 208 numbers were found with heartburn. Heartburn was defined as retrosternal burning discomfort at least once a week. Criteria fulfilled for the last 3 months with symptom onset at least 6 months prior to diagnosis. Demographic characteristics, including age, sex, body mass index (BMI), feeling of depression and stress (self-report), smoking and the duration of heartburn during the previous six months (in weeks), were reported by means of validated questionnaires and a quantile regression was performed to find the results. Linear regression and Cox model were used to compare the results of models with quantile regression. We conducted all analyses with the use of SAS software (version 9.1).

## RESULTS

A total of 208 patients with heartburn symptoms entered to this study. The mean of patients' age was  $52.5 \pm 17.8$  and the median was 51 years. 72.2% were female, 46.5% reported the pain to be very severe (based on the patient's self report), 35.6%

reported pain after eating, 53.8% experienced pain after ingestion of some certain food, 38.9% felt depressed, 65.9% had some kind of stress, and 7.7% were smokers.

The mean of heartburn duration was  $16.4 \pm 9.3$  based on number of the weeks reported by patients, and the median was 24 weeks. Figure 1 shows the distribution of heartburn duration and it clearly indicates a dramatic high negative skewness that equals -0.58 with a large group of patients who tend to experience symptom more than 15 weeks. This negative skewness was also high for females (-0.65), severe pain (-0.58), pain after eating (-1.03), pain after having some specific food (-0.59), feeling of depression (-1.4), stress (-0.69), and smoking (-1.03).

#### *Linear regression*

An ordinary linear regression was conducted with both untransformed and log-transformed heartburn duration. The final model was based on backward selection in regression with untransformed response and produced significant coefficients for pain after eating, smoking, feeling of depression, and age and in regression with log-transformed response produced these significant coefficients for these variables again except smoking (Table 1). These results indicated that the mean of heartburn duration for patients who had a kind of pain after eating was 2.81 weeks longer than others in linear regression with untransformed and 28.4 percent longer in regression with log-transformed ( $\exp(0.25)=1.28$ ). Based on results, feeling of depression was another risk factor in both regression models. The mean of heartburn duration in regression with untransformed response was 4.79 weeks longer for someone who felt depression and it was 60 percent longer in regression with log-transformed response.

Age had a little effect on the duration of heartburn. According to the linear model with untransformed response, the mean of heartburn duration increased 0.08 week when the age

increased one year and in log-transformed regression the increasing was just 0.7 percent.

Smoking was only significant in regression with

selection. In final model only pain after eating remained with insignificant P-value of 0.2.

**Table 1.** Estimated coefficients and p-value in linear regression and proportional hazard regression

Variable	Linear Regression		Linear Regression on log transformed		Proportional Hazards Regression	
	Full Model	Final Model	Full Model	Final Model	Full Model	Final Model
Sex	-0.45 (0.819)		-0.07 (0.713)		0.06 (0.805)	
Pain Severity	-0.45 (0.762)		-0.05 (0.722)		0.009 (0.960)	
Pain after eating	2.74 (0.084)	2.81 (0.066)	0.25 (0.116)	0.25 (0.099)	0.19 (0.299)	0.22 (0.200)
Pain after some food	-0.28 (0.851)		-0.03 (0.839)		-0.01 (0.932)	
Feeling of Depression	4.98 (0.006)	4.79 (0.001)	0.54 (0.003)	0.47 (0.001)	0.21 (0.306)	
Feeling of Stress	-0.32 (0.867)		-0.05 (0.774)		0.8 (0.743)	
Smoking	4.63 (0.119)	4.40 (0.094)	0.39 (0.193)		0.28 (0.429)	
BMI	-0.07 (0.587)		-0.003 (0.835)		0.006 (0.669)	
Age	0.08 (0.038)	0.08 (0.031)	0.008 (0.043)	0.007 (0.058)	-0.005 (0.344)	

Each cell contains the estimated coefficient and P-value (in parenthesis) associated with the given variable and given mean response

untransformed response. From the linear model with untransformed response it was implied that the mean of heartburn duration increased 0.08 weeks. The mean of heartburn duration was 4.4 weeks more than the patients who did not smoke. Although the results were clinically favorable, the goodness of fit was low for both transformed and untransformed ones. The R square for both models was lower than 0.13 and this case obviously was the result of sever skewness in heartburn duration, which might lead to misleading inference about the parameters of interest.

#### Cox Proportional Hazards Regression

The proportional hazard regression model produced no significant coefficient neither in full model nor in final model based on backward

#### Quantile Regression

The results from quantile regression analysis are contained in Table 2. Estimated coefficients for sex, pain severity and pain after having some specific food described the increase in the given decile of heartburn duration associated with a unit increase in given patient characteristics. The length of heartburn duration for each decile (except sex where the first and second given decile were not significant) were longer in males than females and it increased significantly in expected decile for patients who experienced sever pain and pain after eating some specific food in compare to the others considerably ( $P < 0.001$  for all deciles). The results for feeling of stress and BMI seem skeptical. Although their estimated coefficients were negative, the amount values were not clinically

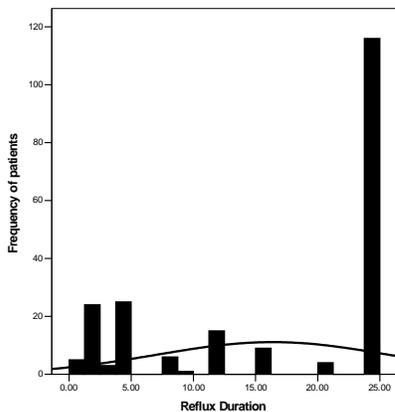
important. Therefore we could ignore these doubtful results. In addition all significant variables in linear and log-linear regression had no significant effect on any decile in quantile regression.

Researchers have traditionally used either OLS regression or proportional hazard model to evaluate the impact of patient characteristics on data duration like waiting times for receiving medical treatment (24) or length of stay in hospital (25).

**Table 2.** Quantile regression coefficients and p-value for the nine Decile of reflux duration

Variable	Decile								
	1	2	3	4	5	6	7	8	9
Sex	0.01 (0.382)	0.02 (0.499)	0.094 ( $<0.001$ )	0.13 ( $<0.001$ )	0.14 ( $<0.001$ )	0.14 ( $<0.001$ )	0.15 ( $<0.001$ )	0.14 ( $<0.001$ )	0.15 ( $<0.001$ )
Pain Severity	0.32 ( $<0.001$ )	0.34 ( $<0.001$ )	0.395 ( $<0.001$ )	0.40 ( $<0.001$ )	0.41 ( $<0.001$ )	0.42 ( $<0.001$ )	0.45 ( $<0.001$ )	0.46 ( $<0.001$ )	0.51 ( $<0.001$ )
Pain after eating	0.07 (0.204)	0.07 (0.200)	0.071 (0.115)	0.03 (0.339)	0.02 (0.471)	0.03 (0.414)	0.02 (0.637)	0.02 (0.644)	0.01 (0.799)
Pain after some food	0.16 ( $<0.001$ )	0.22 ( $<0.001$ )	0.259 ( $<0.001$ )	0.27 ( $<0.001$ )	0.29 ( $<0.001$ )	0.28 ( $<0.001$ )	0.34 ( $<0.001$ )	0.40 ( $<0.001$ )	0.43 ( $<0.001$ )
Feeling of Depression	-0.01 (0.715)	-0.01 (0.570)	0.004 (0.819)	0.004 (0.715)	0.001 (0.933)	-0.0007 (0.945)	-0.003 (0.862)	-0.01 (0.382)	-0.01 (0.462)
Feeling of Stress	-0.09 (0.004)	-0.09 (0.008)	-0.088 (0.002)	-0.08 ( $<0.001$ )	-0.08 ( $<0.001$ )	-0.09 ( $<0.001$ )	-0.09 ( $<0.001$ )	-0.09 ( $<0.001$ )	-0.09 ( $<0.001$ )
Smoking	0.003 (0.923)	0.009 (0.743)	0.022 (0.474)	0.02 (0.341)	0.02 (0.207)	0.02 (0.203)	0.02 (0.318)	0.003 (0.851)	0.000 (0.999)
BMI	-0.04 (0.056)	-0.05 (0.007)	-0.047 (0.001)	-0.05 ( $<0.001$ )	-0.06 ( $<0.001$ )	-0.06 ( $<0.001$ )	-0.07 ( $<0.001$ )	-0.07 ( $<0.001$ )	-0.07 ( $<0.001$ )
Age	0.005 (0.775)	0.005 (0.910)	0.045 (0.318)	0.02 (0.464)	0.01 (0.609)	0.005 (0.809)	-0.001 (0.945)	0.003 (0.897)	-0.005 (0.886)

Each cell contains the estimated coefficient P-value (in parenthesis) associated with the given variable and given quantile regression model.



**Figure 1.** Distribution of reflux duration in patient with GERD symptom

## DISCUSSION

OLS seeks modeling the mean of a random response as a function of observed explanatory factors; quantile regression on the other hand, tries to model this function for each given quantile separately. By modeling the distribution of the duration in a flexible semi-parametric way, quantile regression does not impose modeling assumptions that may not hold.

Although quantile regression is frequently used in econometric (9, 10), in the field of medicine where some duration data tends to concentrate on first or last quantiles researchers recently are interested in using this model. Our study has demonstrated that the use of quantile regression allows the richer inferences to be drawn as a picture of impact of patient characteristics on heartburn duration. In linear regression model with transformed and untransformed responses, study

found pain after eating, depression, smoking and age to be significant. However, depression (26) and certain habits like smoking (27) are reported as the risk factors of reflux, but the diagnostics of regression models indicate poor reliability in modeling because of severe skewness in response. Therefore these results seem unreliable. In contrast to linear regression models, quantile regression indicated a significant relation between sex, pain severity and having pain after some specific foods with duration of heartburn, but age and BMI were found to be insignificant; However, the relationship between GERD and age is controversial, and while some studies have shown a direct relationship (28), others have depicted an inverse relationship (29) and some have reported no association (30). Same controversy exists in the matter of BMI, and some study results for BMI address the association between BMI and gastroesophageal reflux disease (31, 32), while some others have found no significant relationship (33).

Cox proportional hazard showed no significant association for any factors entered to the model. This misleading might be due to the fact that the Cox model imposes a fully parametric relationship between the relative hazard of duration time and the event, while quantile regression does not impose any such relationship. Censored quantile regression discusses this relationship (34) but in this study we had no assumption about censoring. In conclusion, we have demonstrated that using quantile regression allows researchers to find the impact of patient characteristic on the time duration response. The linear regression provides at best case, an incomplete picture of the association between time response and characteristic factors and leads to unreliable results in the case of severe skewness. Cox proportional hazard does not allow one to characterize the distribution of time to factor of interest and can be modeled in unexpected results due to its assumption. As a result, researchers should be interested in advantages of quantile regression for evaluating and providing a

complete view of changing in time response distribution with patient factors and characteristics.

## Acknowledgments

We would like to thank Health Center of Firoozkouch and Damavand cities for data gathering, and Research Institute of Gastroenterology and Liver Diseases for their valuable collaboration in this study.

## REFERENCES

1. Liang H, Chen JD. Assessment of the esophageal pressure in gastroesophageal reflux disease by the local regression *Ann Biomed Eng* 2005;33:847-53.
2. Rosaida MS, Goh KL. Gastroesophageal reflux disease, reflux oesophagitis and non-erosive reflux disease in a multiracial Asian population: a prospective, endoscopy based study. *Eur J Gastroenterol Hepatol* 2004;16:495-501.
3. El-Serag HB, Petersen NJ, Carter J, Graham DY, Richardson P, Genta RM, et al. Gastroesophageal reflux among different racial groups in the United States. *Gastroenterology* 2004;126:1692-99.
4. Dimenas E. Methodological aspects of evaluation of quality of life in upper gastrointestinal disease. *Scand J Gastroenterol* 1993;8(Suppl.199):18-21.
5. Lagergren J, Bergstrom R, Lindgren A, Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *N Engl J Med* 1999;340:825-31.
6. Green PJ, Silverman BW. *Nonparametric regression and generalized linear models*. London: Chapman & Hall; 1994.
7. Koenker R. and Bassett G. Regression quantiles. *Econometrica* 1987;46:33-50.
8. Koenker R, Hallock, K. Quantile Regression. *J Econ Perspect* 2001;51:143-56.
9. Levin J. For whom the reductions count: a quantile regression analysis of class size and peer effects on scholastic achievement. *Empc Econ* 2001;26:221-46.
10. Buchinsky M. Changes in the U.S. wage structure 1963-1987: application of quantile regression. *Econometrica* 1994;62:405-58.
11. Marcotte DE, Wilcox-Gok VJ. Estimating earnings losses due to mental illness: a quantile regression approach. *Ment Health Policy Econ* 2003;6:123-34.

12. Austin PC, Schull MJ. Quantile regression: a statistical tool for out-of-hospital research. *Acad Emerg Med* 2003;10:789-97.
13. Austin PC, Tu JV, Daly PA, Alter DA. The use of quantile regression in health care research: a case study examining gender differences in the timeliness of thrombolytic therapy. *Stat Med* 2005;24: 791-816.
14. Cox DR. Regression models and life-table. *J Royal Stat Soci B* 1972;34:187-220.
15. Koenker R, Biliyas Y. Quantile regression for duration data: a reappraisal of the pennsylvania reemployment bonus experiments. *Emp Econ* 2001;26:199-220.
16. Koenker R, Geling O. Reappraising medfly longevity: a quantile regression survival analysis. *J Am Stat Assoc* 2001;96:458-68.
17. Birkes D, Dodge Y. *Alternative Methods of Regression*. New York, NY: Wiley, Inc; 1993.
18. Gilchrist WG. *Statistical Modeling with Quantile Functions*. Boca Raton, FL: Chapman & Hall/CRC; 2000.
19. Koenker R, Bassett Jr. G. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 1982;50:43-61.
20. Rogers WH. Sg11: quantile regression standard errors. *Stata Tech Bull* 1992;9:16-19.
21. Gould WW. Sg11.1: quantile regression with bootstrapped standard errors. *Stata Tech Bull* 1992;9:19-21.
22. Therneau T, Grambsch P. *Modeling Survival Data: Extending the Cox Model*. New York: Springer-Verlag; 2000.
23. Zarghi A, Pourhoseingholi MA, Habibi M, Rostami Nejad M, Ramezankhani A, Zali MR. Prevalence of gastrointestinal symptoms in population of Tehran, Iran. *Tropic Med Internat Health* 2007;12:181-82.
24. The GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med* 1993;329:673-82.
25. Yoon P, Steiner I, Reinhardt G. Analysis of factors influencing length of stay in the emergency department. *CJEM* 2003;5:155-61.
26. Siupsinskiene N, Adamonis K, Toohill RJ. Quality of life in laryngopharyngeal reflux patients. *Laryngoscope* 2007;117:480-84.
27. Lu CL, Lang HC, Chang FY, Chen TJ, Chen CY, Luo JC, et al. Social and medical impact, sleep quality and the pharmaceutical costs of heartburn in Taiwan. *Aliment Pharmacol Ther* 2005;22:739-47.
28. Isolauri J, Laippala P. Prevalence of symptoms suggestive of gastro-oesophageal reflux disease in an adult population. *Ann Med* 1995;27:67-70.
29. Drossman DA, Li Z, Andruzzi E, Temple RD, Talley NJ, Thompson WG, et al. U.S. householder survey of functional gastrointestinal disorders. Prevalence, sociodemography, and health impact. *Dig Dis Sci* 1993;38:1569-80.
30. Ho KY, Kang JY, Seow A. Prevalence of gastrointestinal symptoms in a multiracial Asian population, with particular reference to reflux-type symptoms. *Am J Gastroenterol* 1998;93:1816-22.
31. Nilsson M, Johnsen R, Ye W, Hveem K, Lagergren J. Obesity and estrogen as risk factors for gastroesophageal reflux symptoms. *JAMA* 2003;290:66-72.
32. Kulig M, Nocon M, Vieth M, Leodolter A, Jaspersen D, Labenz J, et al. Risk factors of gastroesophageal reflux disease: methodology and first epidemiological results of the ProGERD study. *J Clin Epidemiol* 2004;57:580-89.
33. Andersen LI, Jensen G. Risk factors for benign oesophageal disease in a random population sample. *J Intern Med* 1991;230:5-10.
34. Portnoy S. Censored regression quantiles. *J Am Stat Assoc* 2003;98:1001-12.