

Comparing linear regression and quantile regression to analyze the associated factors of length of hospitalization in patients with gastrointestinal tract cancers

Mohamad Amin Pourhoseingholi¹, Mohsen Vahedi¹, Asma Pourhoseingholi¹, Bijan Moghimi-Dehkordi¹, Azadeh Safaee¹, Elham Maserat¹, Fatemeh Ghafarnejad¹, Mohammad Reza Zali¹

¹Research Center of Gastroenterology and Liver diseases, Shahid Beheshti University, (M.C), Tehran, Iran
Correspondence to: Bijan Moghimi-Dehkordi, 7th Floor of Taleghani Hospital, Research Center of Gastroenterology and Liver diseases.
E-mail: b_moghimi_de@yahoo.com

Abstract

Objectives: The aim of this study is to compare linear regression and quantile regression to analyze the predictors of duration of staying in hospital for patients with GI cancers.

Methods: This study was designed as a retrospective cross-sectional survey included all consecutive GI cancer patients admitted over one year period in a random selected hospital group located in Tehran metropolitan in 2006. Residence, age, sex and type of cancer were analyzed respectively using linear and quantile regression.

Results: A total of 2,674 GI tract cancer patients were included in the study. There were 1,616 men (60.43%) and 1,058 women (39.57%). Results of the linear regression analyses showed that only type of cancer was significant. The diagnostic criteria showed that only 7% of variation has been predicted by linear regression. In spite of linear regression sex and type of hospital were significance in quantile regression analysis.

Conclusions: The results have demonstrated that if the duration data showed major skewness, using quantile regression leads to better interpretation and richer inference.

Key words: quantile regression, linear regression, GI cancers, length of hospitalization

Introduction

The gastrointestinal cancers are the most frequent cancer among Iranian males and second to breast cancer among females [1].

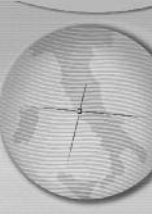
These Cancers has been reported as the most common fatal cancer in some parts of Iran. Overall, GI cancers account for nearly half (44.4%) of all cancer related deaths in Iran [2].

According to the cancer registry program and the Cancer Institute cancer research centre, it is estimated that the majority of the GI cancers occurring in the stomach. The next sites most commonly affected by GI cancers are the colon and rectum (colorectal cancers), esophagus, pancreas and liver [3].

Length of stay is an important measure of hospital activity and health care utilization can recognizing the need to better inform the policy makers, health professionals, and the comprehensive cancer control community about the burden of gastrointestinal cancer, but its

empirical distribution is often positively skewed.

Statistical tools such as linear regression are widely used to assess the associated factors with the duration of staying in hospital. But in most cases of interest, the biometrical measurements like as duration data are not normally distributed, so parametric model such as linear regression that needs normality assumption can not informative enough [4]. In this situation quantiles are more favorable to picture the distribution of response variable. Quantile regression [5,6] 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, that the mean of a continuous response is expressed as a linear function of a set of independent or predictor variables. One of quantile regression's most appealing features is its ability to estimate quantile-specific effects that describe the impact



of covariates not only on the center but 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.

Not only this model frequently used in the econometrics literature [7, 8] but also recently in the field of biostatistics, it is used to find conditional quantile functions to provide a more complete view [9-11].

The aim of this study is comparing quantile regression and linear regression to assess the predictors of length of staying in hospital for patients with GI cancers.

Methods

This study was designed as a retrospective cross-sectional survey included all consecutive gastrointestinal cancer patients admitted over one year period in a random selected hospital group located in Tehran metropolitan in 2006. There are up to 142 medical hospitals in Tehran that covered and organized by 3 medical universities, from that, 19 hospitals selected randomly. The hospitals that have not any gastrointestinal disease or internal medicine wards and also army hospitals excluded from the study and replaced with another one.

Coding of cancer diagnosis samples was based on the international classification of disease for oncology (ICD-O) coding [12]. Study population based on ICD-O coding, including C00-C26. The

cases of interest were all digestive system cancer patients registered between 2005 until 2006 among these esophageal cancer (C15), gastric cancer (C16), colon and rectosigmoid cancer (C18-C21), liver (C22), and pancreas (C25) are investigated here.

The following demographic criteria were analyzed retrospectively in order to drawn up the epidemiology of hospitalization: residence, age, sex, type of cancer and length of hospitalization. Two regression models including quantile regression and multiple linear regression were performed to analyze the data. Linear association between each predictors and length of hospitalization analyzed in specified conditional of first, second and third quartile of the outcome in quantile regression and the overall linear association calculated in linear regression. Stepwise procedure used to find final model in linear regression analysis and adjusted R-square employed to assess the goodness of fit in linear regression. All analysis has been done using SAS, version 9.1.

Results

A total of 2,674 gastrointestinal cancer patients were included in the study. There were 1,616 men (60.43%) and 1,058 women (39.57%) with a male/female ratio of 1.53. The mean±sd of age were 58.87±27.10 years for men and 54.27±14.41 years for women. Table 1 showed the distribution of

Table 1. Distribution of hospitalized patients with sex and type of hospital.

Diagnosis	ICD-OC		Sex		Type of Hospital		Total
			Female	Male	Public	Private	
Colo-Rectal Cancer	18-21	n	491	581	968	104	1,072
		%	45.80	54.20	90.30	9.70	100.00
Esophagus Cancer	15	n	203	256	443	16	459
		%	44.23	55.77	96.51	3.49	100.00
Gastric Cancer	16	n	273	650	886	37	923
		%	29.58	70.42	95.99	4.01	100.00
Liver Cancer	22	n	41	69	109	1	110
		%	37.27	62.73	99.09	0.91	100.00
Pancreatic Cancer	25	n	50	60	109	1	110
		%	45.45	54.55	99.09	0.91	100.00
Total		n	1,058	1,616	2,515	159	2,674
		%	39.57	60.43	94.05	5.95	100.00

diagnosis, sex and type of hospital for inpatients with gastrointestinal cancers. Up to 94 percent of hospitalized cases were reported from public hospitals under the authority of medical universities and only 6 percent were from private hospitals. The vast majority of cancer cases are colorectal cancer (40.0%), followed by gastric cancer (34.5%) and esophagus cancer (17.1%).

The mean of length staying at hospital was 7.4 ± 7.1 days and its skewness equals 5.7 with a large group of patients who stayed at hospital less than 5 days. So the frame of data showed high positive negative skewness.

The results of the linear regression analyses in full and final model showed that only type of cancer (colorectal cancer compare to others) was significant predictor of length of hospitalization (Table 2). The diagnostic criteria showed that only 7% of variation has been predicted by linear regression therefore the fitness of models was really poor. In spite of linear regression, sex and type of hospital were significance in quantile regression analysis (Table 3). Type of cancer was still significance in all quartiles with same coefficient but sex (men compare to women) was significance in first quartile which means that the duration of staying in hospital for men were longer than women for those who allocated in the first quartile of hospitalized duration. Also type of hospital was significant predictors in first quartile indicated that duration of hospitalization was longer for patients in public hospital who allocated in the first quartile. Age was significant just for third quartile but the coefficient was not clinically meaningful.

Discussion

Although quantile regression is frequently used in econometric [5, 6, 8], in the field of medicine where some duration data like as waiting times for receiving medical treatment [13], length of stay in hospital [14] and the score of quality of life [15] that 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 for richer inferences to be drawn a picture of predictors of hospitalized duration. In this study both linear regression and quantile regression demonstrated type of cancer was significant predictor (colorectal cancer compared to others). Although some previous studies on CRC in Iran demonstrated a very low prevalence [16, 17] and the results indicated that time of staying at hospital was significantly shorter than others, a major bulk of inpatients was fund with colorectal and recent study showed that the incidence of colorectal cancer is increased in Iran in recent years [18]. So colorectal cancer will become increasingly important in the next decade.

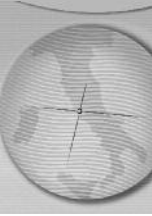
The linear regression provides at best case an incomplete picture of the association between time response and characteristic factors and leads to unreliable results in the case of sever skewness. As a result it can be concluded that researchers should be interested in quantile regression's advantages to evaluate and provide a complete view of changing in time response distribution with patient factors and characteristics when the distribution of data is not normal.

Table 2. Linear regression analysis.

Variable	Full Model			Final Model		
	coefficient	SE	P-value	coefficient	SE	P-value
Constant	6.71	0.74	<0.001			
Age	0.006	0.006	0.32			
Sex	0.17	0.30	0.57			
Type of Cancer	-1.05	0.30	<0.001	-1.14	0.29	<0.001
Type of Hospital	0.68	0.59	0.25			

Table 3. Quantile regression analysis.

Variable	First Quartile			Second Quartile			Third Quartile		
	coefficient	SE	P-value	coefficient	SE	P-value	coefficient	SE	P-value
Constant	3.00	0.00	<0.001	5.00	0.46	<0.001	6.88	1.22	<0.001
Age	0.00	0.00	1.00	0.00	0.002	1.00	0.03	0.01	0.04
Sex	1.00	0.00	<0.001	0.00	0.37	1.00	0.31	0.30	0.29
Type of Cancer	-1.00	0.00	<0.001	-1.00	0.22	<0.001	-1.00	0.30	<0.001
Type of Hospital	1.00	0.00	<0.001	1.00	0.41	0.01	-0.17	0.89	0.84



References

- 1) Mosavi-Jarrahi A, Mohagheghi MA. Epidemiology of esophageal cancer in the high-risk population of Iran. *Asian Pac J Cancer Prev* 2006; 7(3):375-80.
- 2) Naghavi N. Death report from 23 provinces in Iran. 2004 1st edition. Tehran: Ministry of Health.
- 3) Cancer incidence in the Tehran Metropolitan. Second report of the Tehran Population-Based cancer registry. The cancer Institute cancer research centre, 2007.
- 4) Green PJ, Silverman B.W. Nonparametric regression and generalized linear models Chapman & Hall, London 1994.
- 5) Koenker, R and Bassett G. Regression quantiles. *Econometrica* 1978; 46: 33-50.
- 6) Koenker R, Hallock K. Quantile Regression. *Journal of Economic Perspectives* 2001; 51(4): 143-56.
- 7) Levin J. For whom the reductions count: a quantile regression analysis of class size and peer effects on scholastic achievement. *Empirical Economics* 2001; 26:221-46.
- 8) Buchinsky M. Changes in the U.S. wage structure 1963-1987: application of quantile regression. *Econometrica* 1994; 62:405-58.
- 9) Marcotte DE, Wilcox-Gok VJ. Estimating earnings losses due to mental illness: a quantile regression approach. *Ment Health Policy Econ* 2003; 6(3):123-34.
- 10) Austin PC, Schull MJ. Quantile regression: a statistical tool for out-of-hospital research. *Acad Emerg Med* 2003; 10(7):789-97.
- 11) 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. *Statist Med* 2005 ; 24: 791-816.
- 12) Fritz PA, Percy C, Jack A, Shanmugaratnuers K, Solin L, Parkin DM. International classification of diseases for oncology. 3rd edition. Geneva: World Health Organization, 2000.
- 13) The GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *New England Journal of Medicine* 1993; 329:673-82.
- 14) Yoon P, Steiner I, Reinhardt G. Analysis of factors influencing length of stay in the emergency department. *CJEM* 2003; 5(3):155-61.
- 15) Pourhoseingholi MA, Safaee A, Moghimi-Dehkordi B, Zeighami B, Faghihzadeh S, Tabatabaee HR, Pourhoseingholi A. Quality of life in breast cancer patients—a quantile regression analysis. *Asian Pac J Cancer Prev* 2008; 9(3):487-90.
- 16) Haghghi P, Nasr K, Mohallatee EA, Ghassemi H, Sadri S, Nabizadeh I et al. Colorectal polyps and carcinoma in Southern Iran. *Cancer* 1977; 39:274-8.
- 17) Vakili C, Fatourechi V. Age distribution of patients with carcinoma of the colon in a general hospital in Iran. *Surgery* 1976; 79:118-9.
- 18) Moghimi-Dehkordi B, Safaee A, Zali MR. Prognostic factors in 1,138 Iranian colorectal cancer patients. *Int J Colorectal Dis* 2008; 23(7):683-8.