RESEARCH ARTICLE

Relationship between Urbanization and Cancer Incidence in Iran Using Quantile Regression

Somayeh Momenyan [1], Majid Sadeghifar[2], Fatemeh Sarvi[3], Mahmoud Khodadost[4,5]*, Alireza Mosavi-Jarrahi[6], Mohammad Ebrahim Ghaffari[3], Eghbal Sekhavati[7]

Abstract

Quantile regression is an efficient method for predicting and estimating the relationship between explanatory variables and percentile points of the response distribution, particularly for extreme percentiles of the distribution. To study the relationship between urbanization and cancer morbidity, we here applied quantile regression. This cross-sectional study was conducted for 9 cancers in 345 cities in 2007 in Iran. Data were obtained from the Ministry of Health and Medical Education and the relationship between urbanization and cancer morbidity was investigated using quantile regression and least square regression. Fitting models were compared using AIC criteria. R (3.0.1) software and the Quantreg package were used for statistical analysis. With the quantile regression model all percentiles for breast, colorectal, prostate, lung and pancreas cancers demonstrated increasing incidence rate with urbanization. The maximum increase for breast cancer was in the 90th percentile (β=0.13, p-value<0.001), for colorectal cancer was in the 75th percentile (β=0.048 ,p-value<0.001), for prostate cancer the 95th percentile (β=0.55, p-value<0.001), for lung cancer was in 95th percentile (β=0.52, p-value=0.006), for pancreas cancer was in 10th percentile (β=0.011, p-value<0.001). For gastric, esophageal and skin cancers, with increasing urbanization, the incidence rate was decreased. The maximum decrease for gastric cancer was in the 90th percentile(β=0.003, p-value<0.001), for esophageal cancer the 95th (β=0.04, p-value=0.4) and for skin cancer also the 95th (β=0.145, p-value=0.071). The AIC showed that for upper percentiles, the fitting of quantile regression was better than least square regression. According to the results of this study, the significant impact of urbanization on cancer morbidity requires more effort and planning by policymakers and administrators in order to reduce risk factors such as pollution in urban areas and ensure proper nutrition recommendations are made.

Keywords: Urbanization - cancer incidence - quantile regression - Iran

Introduction

Industrialization leads to shift the most important cause of death and disability from nutritional deficiency and infectious diseases to chronic diseases such as cardiovascular disease (CVD), cancer, and diabetes. This issue is named "the epidemiologic transition"(Omran, 1971). One of the major factors that takes part in the transition of global cancer epidemiology is urbanization. The rates of urbanization are increasing globally, from 36.6% of the world population living in urban areas in 1970, to 44.8% in 1994 and by 2025 it is expected to increase to 61.1%. With urbanization, the high-fat diets are common and the energy expenditure (through less physical activity) is decreased. In addition, the use of automobile and bus transportation in rural areas have increased, thus leading to a decrease in physical activities. Certain types of cancer, such as breast, colon, and prostate tumors, are associated with Western lifestyle (sedentary lifestyles) (Stewart et al., 2003).

Risk factors for colorectal cancer that is the fourth most common cancer in men and the third common cancer in women worldwide include obesity, a diet low in fruits and vegetables, insufficient physical inactivity, and smoking (Tomeo et al., 1999; Giovannucci, 2002; Parkin et al., 2005; Botteri et al., 2008). It is a disease that is primarily detected in developed nations with these factors (Popkin, 2004). In South-East England the lung cancer incidence was higher for urban areas than rural (Riaz et al., 2011). In the 1950s for examining the reasons of the

[1]Epidemiology and Biostatistics Department, Qom University of Medical Sciences, Qom, [2]Department of Statistics, Faculty of Basic Science, Bu Ali Sina University, [3]Department of Biostatistics& Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, [4]Gastroenterology and Liver Research Center, Baqiyatallah University of Medical Sciences, [5]Department of Epidemiology, School of Public Health, Iran University of Medical Sciences, [6]Medical School, Shahid Beheshti University of Medical Sciences and Health Services, Tehran , [7]Larestan School of Medical Sciences, Larestan, Iran. *For correspondence: mahmodkhodadost@yahoo.com

Table 1. The Total Number of Cancer Incidence Cases in Iran 2007

Cancer	Skin	Pancreas	Esophageal	Gastric	Prostate	Lung	Breast	Colorectal
Number	9,168.0	360.0	3,164.0	6,235.0	3,164.0	2,066.0	7,192.0	4,887.0

epidemy of lung cancer in the United States and Western Europe, researchers found that the role of cigarette smoking on the lung cancer is sensible (Doll and Hill, 1952; White, 1990). And also Pearce et al. reported that with increasing levels of socioeconomic deprivation, the lung cancer incidence is increased (Pearce and Boyle, 2005). The highest prevalence of smokers are in urban areas and among the socioeconomically groups and this makes the higher incidence of lung cancer in urban areas (Twigg et al.; Robinson and Bugler, 2010). Some studies reported that air pollution in urban areas has a role in the development of lung cancer (Cohen and Pope 3rd, 1995; Nicolich et al., 2001). Only a handful of studies have addressed urban/rural variation in breast cancer incidence and mortality. And also there are studies that reported a statistically significant and excess of breast cancer in urban compared to rural areas (Haenszel et al., 1956; NASCA et al., 1980; Pawlega and Wallace, 1980; Mahoney et al., 1990; Doll, 1991; Howe et al., 1993; Valerianova et al., 1994; Vassallo et al., 1994; Robert et al., 2004).With least square model the relationship between covariates and the mean of the response can be estimated. However,

this model cannot be applied to evaluate the effects of covariates on other aspects of the response distribution, such as the quartiles.

Therefore, based on the least square model, the picture of the relationship is incomplete. To solve this problem of least square model, we can use the quantile regression that investigates how covariates affect the entire distribution of the responses (Lee and Neocleous, 2010). The aim of this study was to investigate the relationship between the urbanization factor and incidence of cancers using quantile regression.

Materials and Methods

This cross-sectional study was conducted on 8 cancers (colorectal, prostate, lung, gastric, pancreas esophageal, skin, and gastric) from March 2007 to March 2008 in 345 cities of Iran. Data was gathered from the Ministry of Health and Medical Education and Statistical Center of Iran for every city. We used the Quantile regression for the skewedness nature of response variable and to study the effect of urbanization factor on entire distribution. The

Table 2. Results of Fitting Quantile Regression for Every Type of Cancer

Cancer		Results for the following percentiles							
		5%	10%	25%	50%	75%	85%	90%	95%
Skin		0.0	-0.07	-0.1	-0.12	-0.11	-0.13	-0.14	-0.145
	Se	0.005	0.019	0.024	0.03	0.029	0.035	0.079	0.08
	P value	0.99	0.001>	0.001>	0.001>	0.001>	0.001>	0.07	0.07
Lung		0.008	0.05	0.09	0.17	0.26	0.3	0.42	0.52
	Se	0.072	0.07	0.08	0.07	0.1	0.098	0.12	0.15
	P value	0.94	0.44	0.45	0.26	0.08	0.03	0.02	0.01
Breast		0.0	0.021	0.056	0.09	0.11	0.114	0.13	0.094
	Se	0.0001	0.002	0.012	0.01	0.013	0.014	0.02	0.042
	P value	0.99	0.001>	0.001>	0.001>	0.001>	0.001>	0.001>	0.001>
Esophageal		0.017	-0.002	-0.023	-0.003	-0.02	-0.03	-0.023	-0.04
	Se	0.005	0.001	0.006	0.002	0.024	0.02	0.03	0.04
	P value	0.001>	0.04	0.001>	0.001>	0.3	0.25	0.4	0.4
Prostate		0.01	0.013	0.02	0.036	0.035	0.038	0.046	0.055
	Se	0.012	0.011	0.01	0.004	0.01	0.012	0.016	0.022
	P value	0.02	0.01	0.01	0.001>	0.001>	0.001>	0.001>	0.001>
Colorectal		0.0	0.012	0.013	0.045	0.048	0.045	0.032	0.039
	Se	0.011	0.007	0.008	0.005	0.01	0.012	0.017	0.021
	P value	0.06	0.04	0.08	0.001>	0.001>	0.001>	0.06	0.001>
Pancreas		0.0013	0.009	0.001	0.011	0.005	0.006	0.009	0.01
	Se	0.0035	0.002	0.001	0.002	0.003	0.004	0.006	0.008
	P value	0.001>	0.001>	0.33	0.001>	0.13	0.14	0.12	0.22
Gastric		0.017	-0.002	-0.023	-0.003	-0.02	-0.03	-0.023	-0.04
	Se	0.005	0.001	0.006	0.002	0.024	0.02	0.03	0.04
	P value	0.001>	0.04	0.001>	0.001>	0.3	0.25	0.4	0.4

Table3. Results of Comparing the Fitness of the Models Using AIC

Cancer	AIC Results for the following percentiles								AIC results for Least square regression
	5%	10%	25%	50%	75%	85%	90%	95%	
Skin	1,050.04	1,022.6	1,015.9	920.35	881.61	729.1	565.7	405.1	952.6
Lung	3,158.8	3,098.02	2,985.5	2,596.1	2,469.12	2,444.52	2,425.6	2,098.9	3,082.18
Breast	1,382.1	1,319.1	1,287.7	1,173.6	1,129.85	987.99	8892.9	856.68	1,254.97
Esophageal	1,597.1	1,456.6	1,338.1	1,166.7	118.3	987.9	852.3	830.1	1,302.9
Prostate	1,981.1	1,834.45	1,757.42	1,633.3	1,588.2	1,443.7	1442.4	1,302.5	1,724.92
Colorectal	1,425.6	1,398.45	1,342.7	1,292.7	1,158.76	1,402.01	1,000.02	952.04	1,342.02
Pancreas	3,237.8	3,076.5	2,797.2	2,406.6	2,320.2	2,089.7	1,845.2	1,005.1	3,255.4
Gastric	1,853.0	1,823.3	1,812.9	1,752.2	1,568.6	1,444.2	1,072.1	925.5	1,806.5

form of the quantile regression model that is studied first time by Koenker and Bassett (1978) is $Y_i = x_i^{'} \beta_\tau + \varepsilon_{\tau i}$ that τ show the quantile. With supposing $Q_Y (\varepsilon_{\tau i}) |X_i) = 0$, the $\tau^{'TH}$ quantile conditional function of Y distribution given X (the explanatory variables) is $\hat{Q}_{Yi} (Y_i | x_i) = x_i^T \beta^{'}(\tau)$ that to estimating the conditional quantiles, the LAD is used. First we fitted the least square regression and then the quantile regression was fitted. The quantiles that we fitted the model for them were 0.05, 0.1, 0.2, 0.5, 0.7, 0.8, 0.9 and 0.9. The AIC (Akaike Information Criterion) was used to compare the fitness of quantile regression for counting data model and Least Square model. As part of information may be lost due to fitting the model, we must apply the model that has a lower loss of information. One of the criteria for considering this issue is AIC. The lower AIC shows the better model. We used software and Quantreg package for all computations. For all computations, the P value less than 0.05 were considered significant.

Results

In this study the relationship between the urbanization and caner (colorectal, prostate, lung, gastric, pancreas esophageal, skin, and gastric) in 345 cities of Iran from March to March 2008 was studied. The incidence cases of these cancers are showed in Table 1.

For investigating the relationship between cancer and urbanization, we fitted least square regression and then quantile regression. The results of fitting these models are
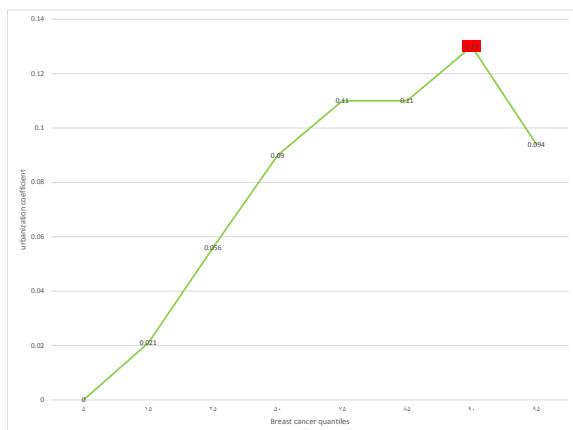
showed in Table 2.

The results of fitting the quantile regression model showed that in all percentiles with increasing the urbanization rate, lung, prostate, pancreas, colorectal and breast cancers incidence increased. For percentiles more than 5, with increasing the urbanization rate, the incidence rates of gastric, esophageal and skin cancers decreased.

The maximum increase for breast cancer was in 90th percentile, for colorectal cancer was in 75th percentile, for prostate cancer was in 90th percentile, for lung cancer was in 90th percentile, for pancreas cancer was in 10'Th percentile For gastric, esophageal and skin cancers, with increasing the urbanization, the incidence rates decreased. The maximum decrease for gastric cancer was in 90th percentile for esophageal cancer was in 90th percentile and for skin cancer was in 90th percentile although this relationship was not significant (Figures 1-7).

Also to compare two models, we used AIC for assessing the goodness of fit of models. The smaller AIC showed the better model. The related AIC was computed for every percentile and the quantile regression was the best fitted model for upper than 50'THpercentiles (Table 2).

Discussion

For skewness of the data and our aim that was to consider the effect of covariates on entire distribution of the data, we used quantile regression and also we
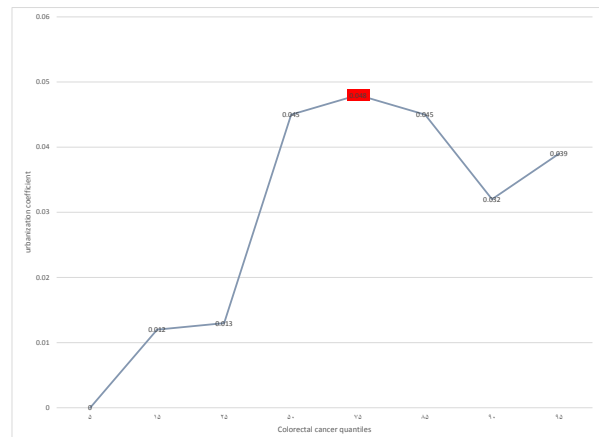

Figure 1. Coefficients of Quantile Regression for, Breast

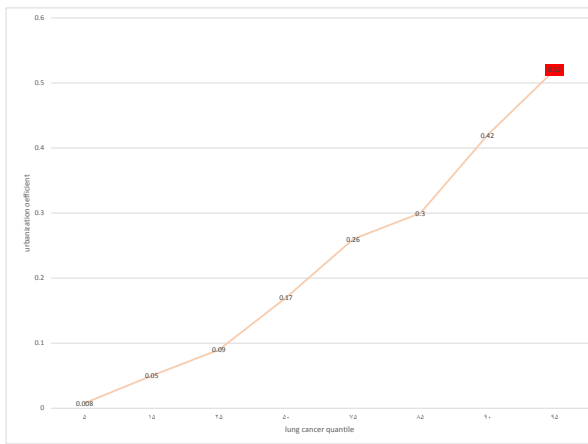Figure 2. Coefficients of Quantile Regression for, Colorectal

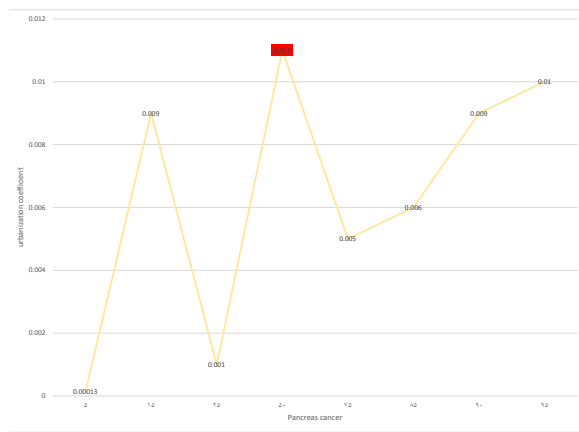Figure 3. Coefficients of Quantile Regression for, Lung

Figure 4.Coefficients of Quantile Regression for, Pancreas
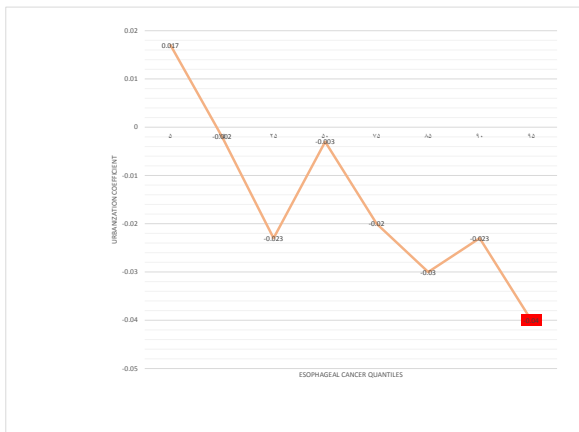

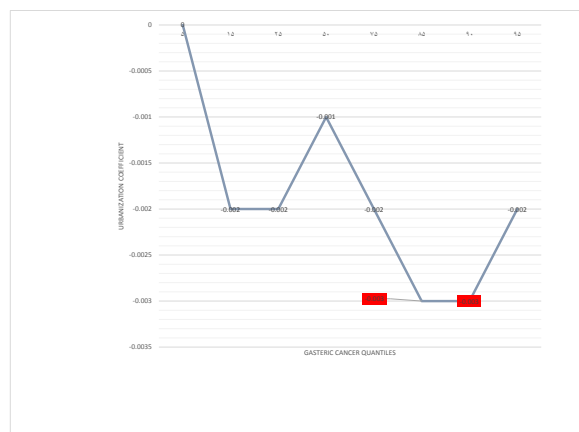Figure 5. Coefficients of Quantile Regression for, Esophageal

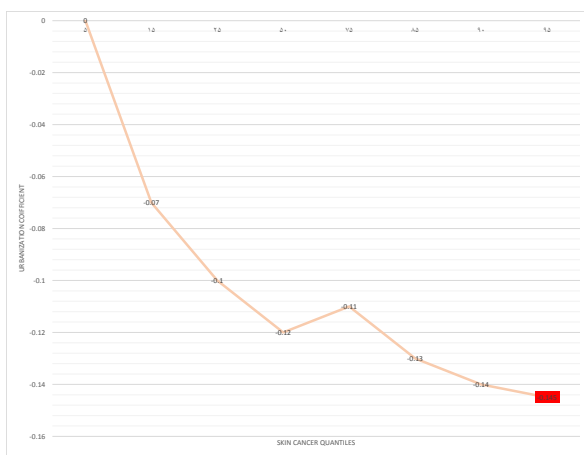Figure 6. Coefficients of Quantile Regression for, Gastric

Figure 7. Coefficients of Quantile Regression for, Skin cancers

fitted the least square regression on data and compared the models. Using least square model, the relationship between urbanization and cancers was significant. Also computing AIC to compare two models showed that quantile regression was better for upper quantiles as the AIC was smaller for the quantile regression in upper percentiles. With respect to coefficients of the quantile regression, we can conclude that with increasing the urbanization, the incidence rate of breast cancer in all percentiles have increased. For instance, for Yazd city that was in 90th percentile of incidence rate of breast cancer,

with increasing one percent in urbanization, the incidence rate increased 0.13. This result is similar to those of the studies of Doll et al. and Nasca et al. and (Haenszel et al., 1956; Nasca et al., 1980; Pawlega and Wallace, 1980; Mahoney et al., 1990; Doll, 1991; Robert et al., 2004), in their studies the cancer incidence had an increasing trend with increasing the urbanization. With urbanization the central factor of breast cancer (e.g. physical inactivity) and other common risk factors increased and increases the incidence rate of breast cancer. Also with increasing the urbanization rate, the colorectal cancer incidence increased and this is similar to the study that is considered by Giovannucci (2002) and Botteri et al (2008). They concluded that urbanization is one of the reasons of increasing the incidence of colorectal cancer because of sedentary lifestyles and prevalence of obesity, a diet low in fruits and vegetables, insufficient physical inactivity and high smoking habit in urban areas. In our study the maximum incidence rate of colorectal cancer was in 90th percentile and for a city like Rasht. Also the incidence rate of prostate and lung and Pancreas increased with increasing the urbanization rate and this is similar to the results that yielded of Stewart et al study, they reported that certain types of cancer, such as breast, colon, and prostate tumors, are associated with Western lifestyle. For a city like Tabriz that is in the 95' TH percentile of incidence of lung cancer, with increasing 1 percent in urbanization, the incidence rate of lung cancer increased 0.5. For skin and gastric and esophageal cancers with increasing the

The Relationship between Urbanization and Cancer Incidence in Iran Using Quantile Regression

urbanization, the incidence rate of these cancers decreased. As urbanization and related factors of industrialization lead to high incidence of cancer, with improving these factors we can prevent the cancer incidence in our society because the cancer incidence and urbanization are closely related.

cancer: a landmark in the history of chronic disease epidemiology. *Yale J Biol Med,* **63**, 29-46.

References

Botteri E, Iodice S, Bagnardi V, et al (2008). Smoking and colorectal cancer: a meta-analysis. *JAMA*, **300**, 2765-78.

Cohen AJ, Pope 3rd C (1995). Lung cancer and air pollution. *Environ Health Perspect*, **103**, 219-24.

Doll R (1991). Urban and rural factors in the aetiology of cancer. *Int J Cancer*, **47**, 803-10.

Doll R, Hill AB (1952). Study of the aetiology of carcinoma of the lung. *BMJ*, **2,** 1271–86.

Giovannucci E (2002). Modifiable risk factors for colon cancer. *Gastroenterol Clin North Am*, **31**, 925-43.

Haenszel W, Marcus SC, Zimmerer EG (1956). Cancer morbidity in urban and rural Iowa. *Public Health Monogr,* **37**, 1-85.

Howe HL, Keller JE, Lehnherr M (1993). Relation between population density and cancer incidence, Illinois, 1986–1990. *Am J Epidemiol*, **138**, 29-36.

Lee D, Neocleous T (2010). Bayesian quantile regression for count data with application to environmental epidemiology. *J R Stat Soc C-Appl,* **59**, 905-20.

Mahoney MC, Labrie DS, Nasca PC, et al (1990). Population density and cancer mortality differentials in New York State, 1978–1982. *Int J Epidemiol,* **19**, 483-90.

Nasca PC, Burnett WS, Greenwald P, et al (1980). Population density as an indicator of urban-rural differences in cancer incidence, upstate New York, 1968–1972. *Am J Epidemiol*, **112,** 362-75.

Nicolich MJ, Gamble JF, Nyberg F, et al (2001). Urban air pollution and lung cancer in Stockholm. *Epidemiology,* **12***,* 590-2.

Omran AR (1971). The epidemiologic transition: a theory of the epidemiology of population change. *Milbank Q,* **83**, 731-57.

Parkin DM, Bray F, Ferlay J, et al (2005). Global cancer statistics, 2002. *CA Cancer J Clin,* **55**, 74-108.

Pawlega J, Wallace R (1980). Nutrition and age at first birth in breast-cancer risk. *Br J Cancer,* **41**, 941-45.

Pearce J, Boyle P (2005). Is the urban excess in lung cancer in Scotland explained by patterns of smoking?. *Soc Sci Med*, **60**, 2833-43.

Popkin BM (2004). The nutrition transition: an overview of world patterns of change. *Nutr Rev,* **62**, 140-3.

Riaz SP, Horton M, Kang J, et al (2011). Lung cancer incidence and survival in England: an analysis by socioeconomic deprivation and urbanization. *J Thorac Oncol*, **6**, 2005-10.

Robert SA, Trentham-Dietz A, Hampton JM, et al (2004). Socioeconomic risk factors for breast cancer: distinguishing individual-and community-level effects. *Epidemiology*, **15**, 442-50.

Stewart BW, Kleihues P, Cancer IAfRo (2003). World cancer report, IARC press Lyon, pp. 198-202.

Tomeo C, Colditz G, Willett W, et al (1999). Harvard Report on Cancer Prevention: volume 3: prevention of colon cancer in the United States. *Cancer Causes Control*, **10**, 167-80.

Valerianova Z, Gill C, Duffy SW, et al (1994). Trends in incidence of various cancers in Bulgaria, 1981–1990. *Int J Epidemiol,* **23**, 1117-26.

Vassallo A, De Stéfani E, Ronco A, et al (1994). Urbanization gradients and cancer mortality in Uruguay, 1988–1992. *Int J Cancer*, **59**, 345-50.

White C (1990). Research on smoking and lung