

## The examination of relationship between socioeconomic factors and number of tuberculosis using quantile regression model for count data in Iran 2010-2011

Fatemeh Sarvi<sup>1</sup>, Somayeh Momenian<sup>2</sup>, Mahmoud Khodadost<sup>\*3</sup>  
Bagher Pahlavanzadeh<sup>4</sup>, Mahshid Nasehi<sup>5</sup>, Eghbal Sekhavati<sup>6</sup>

Received: 28 October 2015

Accepted: 16 March 2016

Published: 17 July 2016

### Abstract

**Background:** Poverty and low socioeconomic status are the most important reasons of increasing the global burden of tuberculosis, not only in developing countries but also in developed countries for particular groups. The purpose of this study was to assess the association between socioeconomic factors and the number of tuberculosis patients using quantile regression for count data.

**Methods:** This cross-sectional study was conducted on 11,320 tuberculosis patients from March 2010 to March 2011 in Iran. Data was gathered from the 345 sections of Iran by Ministry of Health and Medical Education and Statistical Center of Iran. The jittering method was applied for smoothing, and then, the quantile regression for count data was fitted. The AIC was used to compare the fitness of quantile regression for count data model and Poisson log-linear model. The R (3.0.1) software and Quantreg and AER packages were used for all analysis and modeling of the data.

**Results:** The results of fitting the quantile regression for count data showed that in all percentiles, the more increase in immigration rate, illiteracy rate, unemployment and urbanization rates, the more tuberculosis morbidity rate was increased. The maximum increase of tuberculosis due to immigration rate, urbanization rate, unemployment rate, and illiteracy rate was in 95th percentile ( $\beta \hat{=} 0.315$ ), 85th percentile ( $\beta \hat{=} 0.162$ ), 75th percentile ( $\beta \hat{=} 0.114$ ), and 95th percentile ( $\beta \hat{=} 0.304$ ), respectively. For 50th percentiles and higher percentiles, with increasing the sum of physicians to the number of population, the tuberculosis morbidity rate was decreased, and the maximum decrease was in 95th percentile ( $\beta \hat{=} -0.1$ ). For all percentiles, the AIC showed that quantile regression for count data had been a better fit to data.

**Conclusion:** With respect to the relationship between socioeconomic factors and TB rate, health care observers should pay close attention to improving these factors in Iran to reduce the TB mortality and morbidity.

**Keywords:** Quantile regression, Poisson regression, Tuberculosis, Iran.

*Cite this article as:* Sarvi F, Momenian S, Khodadost M, Pahlavanzadeh B, Nasehi M, Sekhavati E. The examination of relationship between socioeconomic factors and number of tuberculosis using quantile regression model for count data in Iran 2010-2011. *Med J Islam Repub Iran* 2016 (17 July). Vol. 30:399.

### Introduction

TB that is caused by the bacillus *Mycobacterium tuberculosis* is an infectious disease. It can affect the lungs (pulmonary

TB) and also other sites of the body (extrapulmonary TB). Tuberculosis is a most important global health problem. According to the report of WHO, "In 2012, an

<sup>1</sup>. PhD Student of Biostatistics, Department of Epidemiology and Biostatistics, School of Public Health, Hamadan University of Medical Science, Hamadan, Iran. [fatemehs73@gmail.com](mailto:fatemehs73@gmail.com)

<sup>2</sup>. PhD Student of Biostatistics, Faculty Member, Epidemiology and Biostatistics Department, Qom University of Medical Sciences, Iran. [somy.momenyan@gmail.com](mailto:somy.momenyan@gmail.com)

<sup>3</sup>. (Corresponding author) PhD Student of Epidemiology, Gastroenterology and Liver Diseases Research Center, Baqiyatallah University of Medical Science, & Department of Epidemiology and Biostatistics, Iran University of Medical Sciences, Tehran, Iran. [mahmodkhodadost@yahoo.com](mailto:mahmodkhodadost@yahoo.com)

<sup>4</sup>. PhD Student of Biostatistics, Department of Biostatistics, Shahid Beheshti University of Medical Science, Tehran, Iran. [frema.msh@gmail.com](mailto:frema.msh@gmail.com)

<sup>5</sup>. Associate Professor of Epidemiology, Department of Epidemiology and Biostatistics, Iran University of Medical Sciences, Tehran, Iran. [mnasehi@yahoo.com](mailto:mnasehi@yahoo.com)

<sup>6</sup>. PHD Student of Environmental Engineering, Faculty Member, Larestan School of Medical Sciences, Larestan, Iran. [sekhavati.papers@gmail.com](mailto:sekhavati.papers@gmail.com)

estimated 8.6 million people developed TB and 1.3 million died from the disease” (1). In developing countries that have lower socioeconomic status, TB is one of the most killer diseases and also in developed countries TB has historically been correlated with poverty and low socioeconomic status (2,3). Families with lower socioeconomic status have more children and frequently reside in more crowded quarters and usually live in undesirable environmental conditions (poor environmental sanitation and poor hygienic training). These factors may increase the probability of exposure to infection among poorer and less educated groups too because they have lack information about vaccination and lack access to the medical care (4,5).

With increased urbanization, the rate of TB is increased too, because there are more risk factors for TB in cities like overcrowded housing, high contacts (6). Bhunu et al. studied the effects of poverty on TB transmission. The results showed that transmission rate is higher in poor communities than in the rich ones due to overcrowding, poor nutrition, reduced treatment uptake and lower socioeconomic status (7). Santos M, et al. evaluated the epidemiological status of Tuberculosis concerning to the socioeconomic characteristics of Sao Jose between 1998 and 2004. The socioeconomic factors that were studied were Schooling, Income, and Number of Residents. The results showed that in poorest areas the disease prevalence was higher (8). The higher rate of this disease is among migrant groups. If migrant groups move from areas with low TB rates to one of the high rates, the risk of the face of primary infections is increased. If migration was occurred from areas with high rate of TB to areas with a low rate of TB, the migrant group transmissions this high rate of TB to those areas. Even if the migration occurs between areas with similar incidence, due to the change in the living environment, and exposure to undesirable situations, the probability of contact with TB for the first time increases, so it is likely that TB is prevalent among

migrant groups (9). Jose Leopold et al. investigated the relationship between TB with several population characteristics in Sao Paulo, Brazil, from 1994 to 1998. In their study, the relationship between immigration rate for foreigners and mortality rate of TB and also the relationship between immigration rates for the other Brazilian States with a mortality rate of TB was significant (10).

The counts data are typically modeling with log-linear Poisson model. With this model, the relationship between covariates and the mean of the response can be estimated (11). However, this model cannot be applied to evaluate the effects of covariates on the other aspects of the response distribution, such as the quartiles. Therefore based on the log-linear model the picture of the relationship is incomplete, on the other hand, the assumption of the log-linear model is the equality of mean and variance of the response distribution, whereas in real, the variance of the data may be bigger than the mean. This issue is named overdispersion. The overdispersion may lead to the significant relationship between a response variable and covariates that is not significant in fact and also with overdispersion the variance of the data may be an underestimate (12). To solve this problem of the log-linear model, we can use the quantile regression for count data that investigate how covariates affect the entire distribution of the count responses and also unlike the log-linear model, quantile regression model for count data, has a separate scale parameter, and overdispersion does not occur (13).

The aim of this study was to investigate the relationship between the socioeconomic factors such as immigration rate, unemployment rate, urbanization rate, the sum of physicians to the number of population points and illiteracy rate with the number of tuberculosis patients.

## Methods

This cross-sectional study was conducted on 11,320 TB Patients from March 2010 to

March 2011 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 count data for the skewness nature of response variable and to study the effect of socioeconomic factors on entire distribution,

The estimation of quantile regression that studied by Koenker and Bassett (1978) was offered to continuous data (14). Since the count data have not continuous Quantiles, to fit the model, Machado and Santos Silva (2005) offered the jittering approach for smoothing (15). Using this approach, the response variable becomes continuous, and we can fit the quantile regression for data, so we produced data from uniform distribution form (0,1) and replace the response variable  $Y_k$  with "jittered" responses  $Y_k + U_k$  where  $U_k \sim \text{uniform}(0,1)$ . These jittered responses have continuous real quantiles. The form of model for these data is  $Q_{Y_k+U_k}(\tau|x_k) = \tau + \exp(x_k^T \beta(\tau))$ , where  $x$  is explanatory variable and  $k$  is the number of city,  $\tau$  is the quantile that we fitted the model for them,  $\beta$  is the estimated coefficient for urbanization and  $T$  is the transpose of the explanatory vector. We can implement this model as the equivalent model with this form  $Q_{Y_k^*}(\tau|x) = x_k^T \beta(\tau)$  where  $Y_k^* = \log(Y_k + U_k - \tau)$ , these two models are equivalent because we applied a monotonic transformation (15).

So to apply the quantile regression for count data, we first jittered the response data that was the number of TB patients in 345 cities. The quantile regression model for count data was fitted after jittered the data and applied the transformation. The quantiles that fitted the model was 0.05, 0.1, 0.15, 0.25, 0.5, 0.75, 0.85, 0.9 and 0.95. Since the number of TB patients in every city is dependent on population of the city, we used the population of the city as an offset term. The AIC (Akaike Information Criterion) was used to compare the fitness of quantile regression for count data model and Poisson log-linear model. As the

part of the information may be loosed due to fitting the model, we must apply the model that have a lower loss of information. One of the criterions for considering this issue is AIC. The lower AIC shows, the better model.

We used *R*(3.0.1) software for all computations. We applied the Quantreg and AER packages to fitting the quantile regression for count data and testing the existence of overdispersion, respectively. For all computations, the P value less than 0.05 were considered significant.

## Results

The total number of 11320 TB patients from 345 cities of Iran during March 2010 to March 2011 were enrolled in this study. The mean±SD age of patients was  $47.7 \pm 14.25$  (Years), and the mean±SD for the number of TB patients were  $32.8 \pm 79.11$ . Figure 1 shows the histogram graph of TB patients in every city. This figure displayed the right skewed of the data, and the value of skewness was yield 7.28 with a standard error of 0.13.

Descriptive statistics of variables are shown in Table 1.

For every quantile, the quantile regression model for count data was fitted, and the parameters were estimated. The poisson regression model was fitted to data. The results of estimation using fitting the poisson regression model and quantile regression model for count data was shown in Table 2, 3.

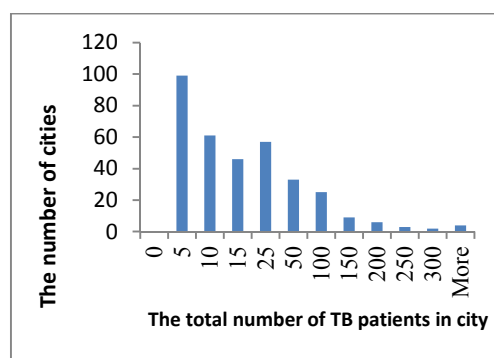


Fig. 1. Histogram graph for number of TB patients in every city

Table 1. Descriptive statistics of study subjects

Variable	Mean (Std. Deviation)	Maximum	Minimum
Immigration rate	-0.03 (6.68)	-17.8	33.70
Urbanization rate	0.52 (.20)	0.99	0.08
Unemployment rate	14.72 (9.84)	59.50	1.05
Illiteracy rate	19.41 (6.17)	47.46	3.96
Sum of physicians to number of population points	4.25(58.5)	1081.200	0.005

Table 2. The results of Poisson regression model for assessing the associations between socio-economic factors and tuberculosis morbidity

Variable	$\beta$	95% CI	Std. Error	p
intercept	-0.914	0.401	1.558	0.713
Immigration rate	0.153	1.165	0.0096	<0.001
Urbanization rate	0.169	1.184	0.013	<0.001
Unemployment rate	0.127	1.135	0.012	<0.001
Illiteracy rate	0.293	1.341	0.0134	<0.001
Sum of physicians to number of population points	0.027	0.973	0.003	<0.001

In Poisson model, there was a significant association between socioeconomic factors and TB morbidity rate. With increasing one unit of these factors, the morbidity rate of TB was increased. For example with increasing one unit in the unemployment rate, the TB morbidity rate was increased by 0.135. The AER package in R(3.0.1) software was used for testing the dispersion and the existence of overdispersion with fitting the Poisson model was confirmed ( $p < 0.001$ ).

The interpret coefficients of the quantile regression for count data model was similar to the Poisson regression model. The results of the quantile regression model for count data showed that in all percentiles, with increasing the immigration rate, the TB morbidity rate was increased. The relationship in higher percentiles was significant. It means that in counties with high TB morbidity rate, immigration and TB have the strongest relationship. For all percen-

tiles of TB morbidity rate, with increasing the urbanization rate, unemployment rate, and illiteracy rate, the TB morbidity rate was increased. The maximum increase for immigration rate was in 95<sup>th</sup> percentile ( $\hat{\beta} = 0.315$ ), for urbanization rate was in 85<sup>th</sup> percentile ( $\hat{\beta} = 0.162$ ), for unemployment rate was in 75<sup>th</sup> percentile ( $\hat{\beta} = 0.114$ ) and for illiteracy rate was in 95<sup>th</sup> percentile ( $\hat{\beta} = 0.304$ ). In 50<sup>th</sup> percentiles and higher percentiles, with increasing the sum of physicians to the number of population, the tuberculosis morbidity rate was decreased, and the maximum decrease was in 95<sup>th</sup> percentile ( $\hat{\beta} = -0.1$ ) (table 3). In higher percentiles of TB morbidity rate, with increasing the number of physicians to population points, TB morbidity rate was decreased and for lower percentiles with increasing average annual income, the TB morbidity rate was decreased, although this relationship was not significant (Fig. 2).

Table 3. The results of quantile regression model for count data for assessing the associations between socio-economic factors and tuberculosis morbidity

Variabl	Results for the following percentiles									
	5%	10%	15%	25%	50%	75%	85%	90%	95%	
Immigration rate	$\beta$	0.077	0.196	0.17	0.141	0.165	0.153	0.098	0.173	0.315
	exp( $\beta$ )	1.08	1.22	1.19	1.15	1.18	1.17	1.103	1.19	1.37
Urbanization rate	$\beta$	0.058	0.109	0.052	0.04	0.08	0.113	0.162	0.142	0.126
	exp( $\beta$ )	1.06	1.115	1.05	1.04	1.08	1.12	1.18	1.165	1.134
Unemployment rate	$\beta$	0.089	0.072	0.083	0.089	0.079	0.114	0.056	0.108	0.05
	exp( $\beta$ )	1.093	1.08	1.09	1.09	1.08	1.13	1.057	1.114	1.05
Illiteracy rate	$\beta$	0.261	0.30	0.12	0.12	0.176	0.235	0.271	0.275	0.304
	exp( $\beta$ )	1.3	1.38	1.13	1.13	1.19	1.26	1.31	1.32	1.26
Sum of physicians to number of population points	$\beta$	0.049	0.044	0.039	0.024	-0.005	-0.027	-0.05	-0.076	-0.1
	exp( $\beta$ )	1.05	1.045	1.039	1.024	0.995	0.973	0.951	0.927	0.9

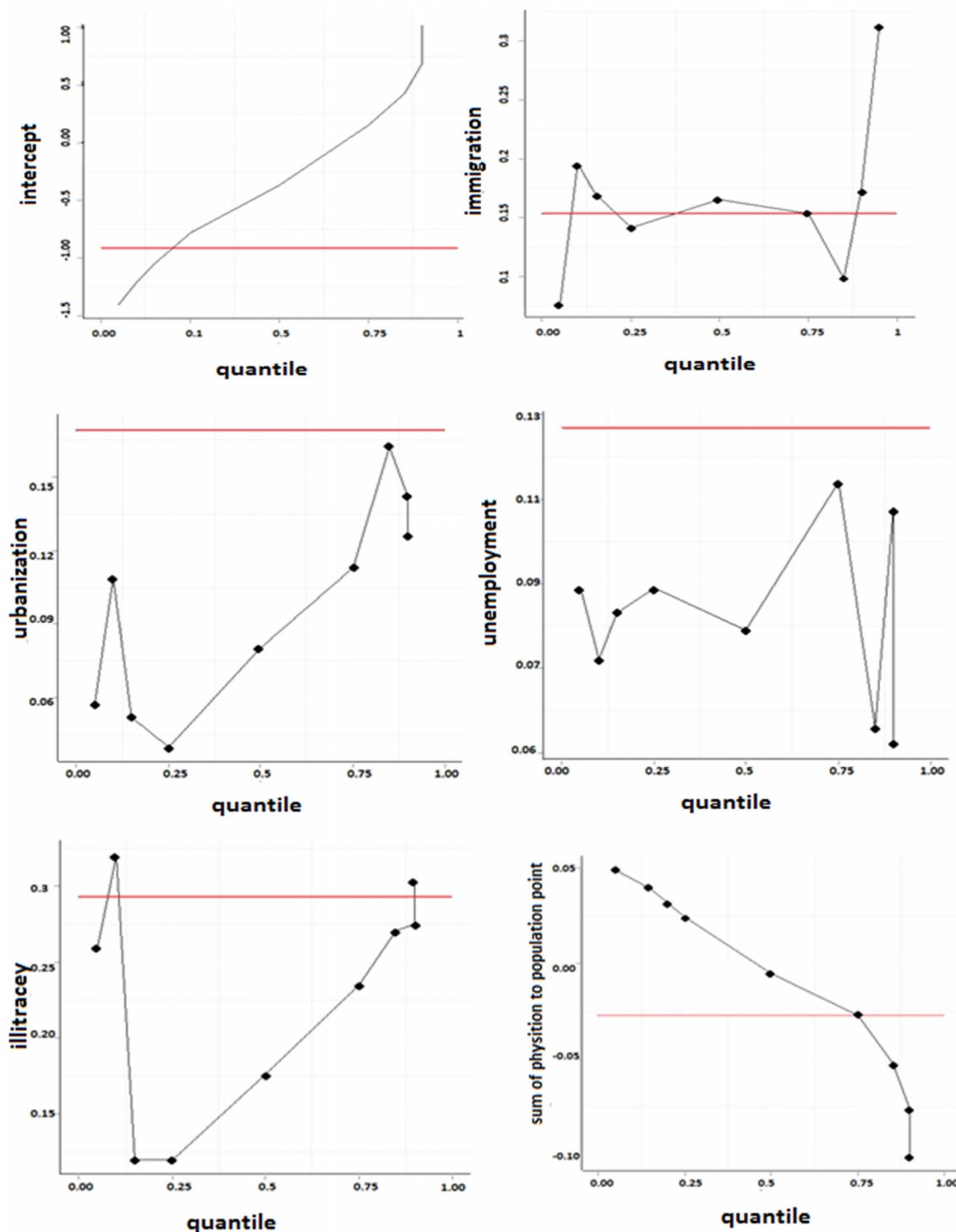


Fig. 2. Coefficients of quantile regression for count data for every quantiles (the red line show the Poisson's estimate)

Also, the AIC was used to compare the fitness of quantile regression for count data model and Poisson log-linear model. The smaller AIC showed the better model. The related AIC was computed for every percentile, and the quantile regression was

the best-fitted model (Table 4).

For all percentiles, the AIC was smaller than one for Poisson regression, and it showed that in all percentiles, the quantile regression for count data was better fitted to data in comparison of Poisson regression model.



Table 4. Results of AIC for assessing the fitness of the models

Percentiles	5%	10%	15%	25%	50%	75%	85%	90%	95%	Poisson model
AIC	964.9	924.6	901.6	860.7	850.7	937.6	1003.8	1070.2	1130.6	6722.9

## Discussion

As the aim of the study is to consider the effect of covariates on entire distribution of the data and because of skewness of the data, we used the quantile regression for count data and also fitted the Poisson regression on data to compared two models. In Poisson regression model, the relationship between socioeconomic factors and TB was significant. This can be due to the existence of overdispersion, but quantile regression has a separate scale parameter and prevent of occurring the overdispersion. So, we used this model for more accurate estimation of parameters in this study. Also, the results of comparing the fitness of two models showed that the AIC was smaller for the quantile regression for count data in all percentiles compared to the Poisson regression model. The results of the quantile regression showed that with increasing the immigration rate, the TB morbidity rate was increased. This result is consistent with the studies of Jose Leopold et al. Talbot et al. and Gilbert et al. (10,16, 17). In their studies, the TB morbidity and mortality rates had an increasing trend with increase the immigration. The foreign immigrants in Iran are almost from Pakistan and Afghanistan countries that are the countries with a high incidence rate of TB and then, immigration from these countries to Iran may increase the morbidity rate of TB. Overcrowded housing and high probability of contacts in cities lead to increasing the TB morbidity. Also in this study, with increasing the urbanization rate, the TB morbidity rate was increased and is consistent with the study of Kwan CK, et al. (6) that reported the urbanization as one of the reasons of increasing the mortality and morbidity of TB. Other socioeconomic factors in this study that affected the TB morbidity rate were unemployment and illiteracy rate. In poorer societies with worse conditions like higher unemployment rate or

lower schooling or high illiteracy rate, the rate of TB morbidity and mortality is high. The results of this study showed that with increasing the unemployment rate, and illiteracy rate, the rate of TB morbidity was increased. Santos and Bhunu et al. reported similar results (7,8). One of the reasons for this relationships is that the illiterate peoples have low information about the vaccination, hygiene, disease prevention and the risk factors. Also, the results of Richard Coke et al. is consistent with results of our study (18). In this study, with increasing the access to medical care, the TB morbidity was decreased. For higher percentiles, with increasing the sum of physicians to the number of population point's factor, the TB morbidity rate was decreased that is consistent with the study of Santos et al. (8). Thus, the results of our study confirmed that the factors related to the low socioeconomic status lead to high morbidity of tuberculosis and with improving these factors we can prevent the TB morbidity and mortality.

## Conclusion

As a result, the quantile regression for count data provided a suitable measurement to examine the relationship between socioeconomic factors and TB morbidity. These findings demonstrated that with increasing the immigration rate, illiteracy rate, unemployment and urbanization rates, the TB morbidity rate was increased. Also, with increasing the sum of physicians to the number of population, the tuberculosis morbidity rate was decreased. So it seems necessary to improve these factors such as increase the proportion of physicians to the number of population to reduce the TB morbidity and mortality and to reach the TB control goals.

## Acknowledgements

Authors wish to express their sincere

thanks to all staffs of the Tuberculosis and Leprosy Control Office of Ministry of Health and Medical Education and also Statistical Center of Iran, especially Dr. Sharifi and Ms. Ahmadi and Mr. Bashiri who sincerely helped the data collection.

### References

1. Organization WH. Global tuberculosis report 2013: World Health Organization 2013.
2. Olson NA, Davidow AL, Winston CA, Chen MP, Gazmararian JA, Katz DJ. A national study of socioeconomic status and tuberculosis rates by country of birth, United States, 1996–2005. *BMC public health* 2012;12(1):365.
3. Ogboi S, Idris S, Olayinka A, Ilyas J. Socio-demographic characteristics of patients presenting pulmonary tuberculosis in a primary health centre, Zaria, Nigeria. *Journal of Medical Laboratory and Diagnosis* 2010;1(2):11-4.
4. Lillebaek T, Andersen AB, Dirksen A, Smith E, Skovgaard LT, Kok-Jensen A. Persistent high incidence of tuberculosis in immigrants in a low-incidence country. *Emerging infectious diseases* 2002;8(7):679-84.
5. Figueroa-Munoz JI, Ramon-Pardo P. Tuberculosis control in vulnerable groups. *Bulletin of the World Health Organization* 2008;86(9):733-5.
6. Kwan CK, Ernst JD. HIV and tuberculosis: a deadly human syndemic. *Clinical microbiology reviews* 2011;24(2):351-76.
7. Bhunu C, Mushayabasa S. Assessing the effects of poverty in tuberculosis transmission dynamics. *Applied Mathematical Modelling* 2012;36(9):4173-85.
8. Santos MdLSG, Vendramini SHF, Gazetta CE, Oliveira SAC, Villa TCS. Poverty: socioeconomic characterization at tuberculosis. *Revista latino-americana de enfermagem* 2007;15(SPE):762-7.
9. Knowelden J. Section of Epidemiology and Preventive Medicine. *Proc R Soc Med* 1956;9:249-52.
10. Antunes JLF, Waldman EA. The impact of AIDS, immigration and housing overcrowding on tuberculosis deaths in São Paulo, Brazil, 1994–1998. *Social science & medicine* 2001;52(7):1071-80.
11. Marzban C. Quantile Regression. Applied Physics Laboratory, Department of Statistics, University of Washington 2003.
12. Agresti A. An introduction to categorical data analysis: John Wiley & Sons 2007.
13. Lee D, Neocleous T. Bayesian quantile regression for count data with application to environmental epidemiology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2010;59(5):905-20.
14. Koenker R, Bassett Jr G. Regression quantiles. *Econometrica: journal of the Econometric Society* 1978:33-50.
15. Machado JAF, Silva JS. Quantiles for counts. *Journal of the American Statistical Association* 2005;100(472):1226-37.
16. Cain KP, Benoit SR, Winston CA, Mac Kenzie WR. Tuberculosis among foreign-born persons in the United States. *Jama* 2008;300(4):405-12.
17. Gilbert R, Antoine D, French C, Abubakar I, Watson J, Jones J. The impact of immigration on tuberculosis rates in the United Kingdom compared with other European countries. *The international journal of tuberculosis and lung disease* 2009;13(5):645-51.
18. Coker R, McKee M, Atun R, Dimitrova B, Dodonova E, Kuznetsov S, et al. Risk factors for pulmonary tuberculosis in Russia: case-control study. *Bmj* 2006;332(7533):85-7.