Ann Ig 2023 Jul-Aug; 34(4): 468-479 doi: 10.7416/ai.2022.2549. Epub 2022 Oct 17.

Tuberculosis incidence and its socioeconomic determinants: developing a parsimonious model

M. Sorokina[1], T. Ukubayev[2], B. Koichubekov[1]

Key words: Tuberculosis, Incidence, Socioeconomic factors, Model
Parole chiave: Tubercolosi, Incidenza, Determinanti socioeconomici, Modello

Abstract

Background. *Tuberculosis is a widespread communicable disease, which is one of the top 10 causes of demise globally. Several regression models have been built, and then utilized for the Tuberculosis incidence projections. However, when fitting a multiple linear regression model, an analysis must account for multicollinearity aspects. The present study aimed to develop a parsimonious model that produces unbiased results based on socioeconomic variables as predictors of Tuberculosis incidence.*
Study design. *Ecological study.*
Methods. *Data were collected from the Karaganda Regional Center of Phthisio-pulmonology and Bureau of National Statistics. By multiple linear regression model, we investigated associations between Tuberculosis incidence rate and socioeconomic determinants in Karaganda region, Kazakhstan, during 2001-2019. A Principal components analysis was performed on the socioeconomic variables with oblique rotation. Furthermore, associations of Tuberculosis incidence with the principal components derived from the Principal components analysis were assessed.*
Results. *The incidence of Tuberculosis in Karaganda region decreased over the period of 2001-2019. Economic development and healthcare capacity were negatively correlated with Tuberculosis incidence. A multiple linear regression equation on Tuberculosis incidence (y) was developed with economic development (x_1) and healthcare capacity (x_2) clustering two components (utilizing Principal components analysis) to eliminate collinearity: $y = 1442 - 454.3x_1 - 211.4x_2$. The incidence of Tuberculosis decreased with the increase of economic development and healthcare capacity.*
Conclusions. *In conclusion, the study indicated that economic development and healthcare capacity are closely associated with the incidence of Tuberculosis. The findings support the implementation of optimal preventive measures for Tuberculosis control, including improving the level of economic status, increasing social protection, health expenditure, and strengthening health sector capacity, which are key determinants of the incidence of Tuberculosis.*

[1] *Department of Informatics and Biostatistics, Karaganda Medical University, Karaganda, Kazakhstan*
[2] *Karaganda Medical University, Karaganda, Kazakhstan*

Authors' contribution: M. Sorokina supervised the study, Ukubayev T. collected the data, performed statistical analysis, and drafted the manuscript; M. Sorokina and B. Koichubekov, reviewed and edited the manuscript. All authors discussed and contributed to the final manuscript.

Annali di Igiene : Medicina Preventiva e di Comunità (Ann Ig)
ISSN 1120-9135 https://www.annali-igiene.it

Introduction

Tuberculosis (TB) is a widespread communicable disease, belonging to the top 10 causes of demise globally. According to World Health Organization (WHO), an approximate 10 million people become ill with TB worldwide in 2019 (1). The number has been decreasing in recent years, even though not fast enough to achieve the milestones of the "End TB strategy" (2). The decline was mostly due to social actions that enhanced overall living standards as well as improvements in healthcare performance (3).

Recent studies of TB usually regard socioeconomic perspectives alongside biomedical and environmental views (4-6). Although the paradigm of TB as a disease of poverty and economic distress has long been recognized, many epidemiological TB studies show great interest in socioeconomic risk factors (7-9). It is well-established that the prevalence of TB fluctuates with social and economic characteristics of the population. Many incident TB cases have been more prevalent among poverty-stricken subgroups (10). In addition, lack of access to healthcare, low capacity of health facilities, and poor healthcare performance are some of the causes of the escalation of TB transmission (11).

In this regard, several regression models have been built, and then utilized for the TB incidence projections (12-14). Multivariate linear regression (MLR) is a statistical method employed to investigate the association a response variable and two or more predictors. MLR is such a suitable statistical tool to approximate complex relationships of the TB incidence with socioeconomic determinants (15). De Castro et al. (12) revealed a significant correlation between the incidence of TB and per-capita income as well as poor living conditions. The study of Wy et al (13) suggested that the deterioration of socioeconomic status of the vulnerable population contributes to reproduction of TB disease. De Castro et al. (12) found a negative association between the TB incidence rate and poverty and unemployment rates. The authors highlighted that health services performance was closely related to the proportion poor population, as well as the unemployment rate.

However, when fitting a multiple linear regression model, an analysis must account for multicollinearity aspects. Multicollinearity is a phenomenon, which occurs when 2 or more predictors are highly correlated (16). Multicollinearity can either inflate or deflate the standard errors of the regression coefficients, which leads to false significance or non-significance. Another consequence of multicollinearity is that the coefficients can change a sign in both directions. Moreover, multicollinearity limits the size of the coefficient of determination (R^2), which makes it difficult to assess the quality of a model (17). One way of recognizing multicollinearity is to scan a correlation matrix of all predictors with high correlation coefficients. Furthermore, to check the effect of multicollinearity, the variance inflation factor (VIF) is to be calculated for each predictor. If the VIF value is greater than 10, multicollinearity exists. Although these statistics aid to identify the presence of collinearity, they cannot detect the predictor causing multicollinearity (18).

Therefore, multicollinearity misrepresents the results from multiple linear regression analysis. The development of a reliable multiple linear regression model requires the careful inspection of multicollinearity and exclusion of multicollinear predictors. Thus, the current study aimed to develop a parsimonious model that produces unbiased results based on socioeconomic variables as predictors of TB incidence rate.

Methods

Settings

The region of Karaganda, located in the central part of Kazakhstan, has a total area of 239,045 sq km, and is made up of 7 districts and 6 cities. The region has a population of 1,372,199.

Study design and Data source

We conducted an ecological study, with the unit of analysis being Karaganda region, Kazakhstan. For the incidence rate of TB, we used data from the National Register of Tuberculosis Patients (NRTP) provided by the Regional Center of Phthisio-pulmonology in Karaganda. The data were available from 2001 through 2019. Surveillance data on socioeconomic indicators were obtained from the Bureau of National Statistics for the period from 2001 through 2019 (19).

Variables

The response variable was the TB incidence rate (new TB cases in a year per 100,000 population), and the predictors were economic factors and healthcare performance characteristics. Previous studies have already identified these variables as TB determinants (12-14).

Socioeconomic determinants were presented in the study by Gross Domestic Product (GDP) per capita, unemployment rate, per-capita income, proportion of the poor population. Additionally, variables describing healthcare system included current health expenditure (CHE) per capita, number of general practitioners (GP), TB specialists, and TB hospital beds capacity.

Data analysis

Statistical analysis was performed in three stages. The first stage included multiple linear regression (MLR) to identify socioeconomic and healthcare related risk factors associated with the incidence rate of TB. We used a backward elimination method in repeated bootstrap samples. The bootstrap procedure is employed to improve estimation and confidence intervals for regression coefficients (20). The bootstrap results were based on 1,000 bootstrap samples. Statistical level of significance was accepted as $p < .05$.

In a second stage, all the variables assessed in the first stage were considered for the Principal Component Analysis (PCA) with oblique rotation (direct oblimin). In PCA, Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett's test of sphericity were conducted to define the suitability of the data for the analysis. An initial analysis was conducted to obtain eigenvalues for each component in the data. We used the Elbow method to determine the number of principal components to retain in the analysis. The criteria for principal components extraction are based on eigenvalues over Kaiser's criterion (greater than 1).

Eventually, we employed MLR for principal components derived from the PCA. Backward elimination method in repeated 1,000 bootstrap samples was applied. Statistical level of significance was accepted as $p < .05$.

Statistical analysis was performed utilizing IBM SPSS Statistics, version 27 (IBM Corporation, 2020).

Results

Figure 1 shows the yearly time series of the incidence rate of TB from 2001 through 2019 in Karaganda region, Kazakhstan. Overall, there was a downward trend in the number of incident cases of TB over the analyzed period. In 2001, there were 2,014 cases per 100,000 population, and then, the incidence rate rose to 2,308/100,000 inhabitants in 2002. The figure then remained stable at around 2,300 till 2005, which was followed by a dramatic decline to 1,308/100,000 inhabitants. In the following years, the number of TB incident cases decreased gradually year on year until they reached 599/100,000 inhabitants in 2019.

Regarding socioeconomic determinants, both per-capita GDP and per-capita income experienced a non-linear growth from 2001 to 2019. By contrast, there was a slow decrease in unemployment rate throughout the period shown; however, the figure for proportion of poor individuals saw a dramatical decline from 2001 to 2009. After that the proportion of the poor remained relatively stable until 2019.

Both the numbers of TB specialists and GPs remained relatively unchanged from 2001 to 2007, at approximately 150 and 600, respectively. After that, the former fluctuated between 100 and 200, whereas the latter witnessed a fluctuation between 600 and 800 in the following years. As regards TB hospital beds, the number experienced an increase from 1,060 in 2001 to 1,243 in 2009. However, this was followed by a steady drop to 454 TB hospital beds in 2019. Conversely, there was a significant growth in per-capita CHE from $51.57 in 2001 to $319.05 in 2017. However, the figure then went down to $281.62 in the final year.

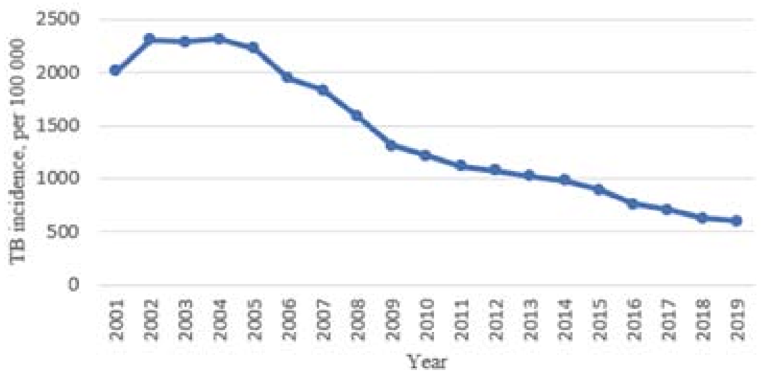

Figure 1 - Tuberculosis incidence rate in Karaganda oblast, Kazakhstan (2001-2019)

We found statistically significant bivariate associations with the TB incidence rate after inspecting the correlation matrix with all predictor variables (Table 1). These include per-capita GDP (r = -.945, p < .001), per-capita income (r = -.949, p < .001), proportion of poor individuals (r = .871, p < .001), unemployment rate (r = .758, p < .001), per-capita CHE (r = -.901, p < .001), number of GPs (r = -.682, p < .001) and TB hospital beds (r = .766, p < .001). However, the correlation coefficient between the TB incidence rate and number of TB specialists was not statistically significant.

The results show that there is a strong relationship among per-capita GDP, per-capita income, proportion of the poor, unemployment rate, and per-capita CHE (p < .001). Furthermore, the number of TB hospital beds is negatively collinear with per-capita GDP (r = -.874, p < .001), per-capita income, (r = -.857, p < .001) and number of GP (r = -.831, p < .001).

Table 2 describes the bootstrap results of the MLR analysis that investigated multivariate associations of TB incidence rate with socioeconomic determinants. The results show that the coefficient of determination (R^2) was 0.985, implying that the model explained 98.5% of the variance of the TB incidence. The bootstrap result for the Durbin-Watson test was 2.540, implying that the model did not show residuals autocorrelation. Regarding multicollinearity aspects,

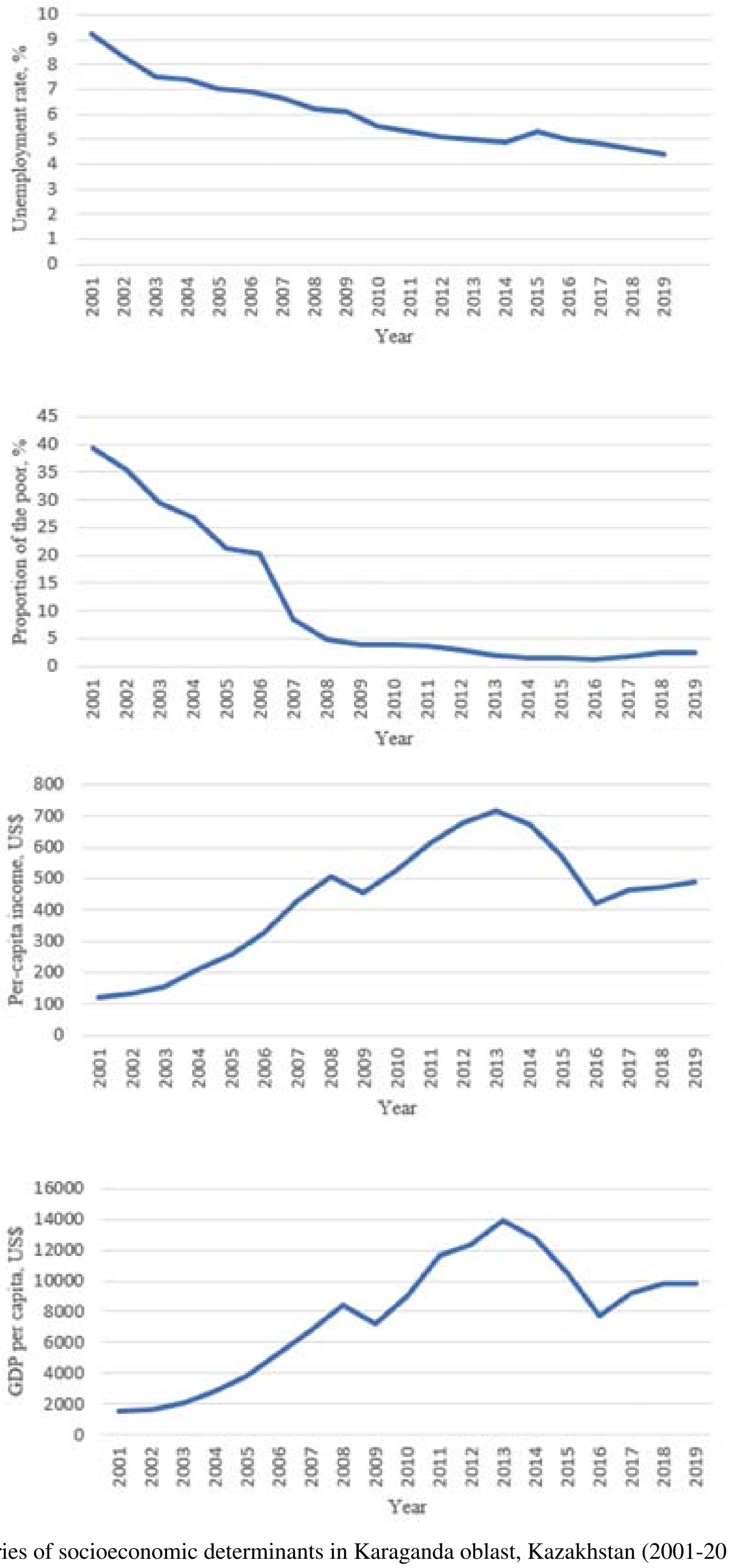

Figure 2 - Time series of socioeconomic determinants in Karaganda oblast, Kazakhstan (2001-2019)

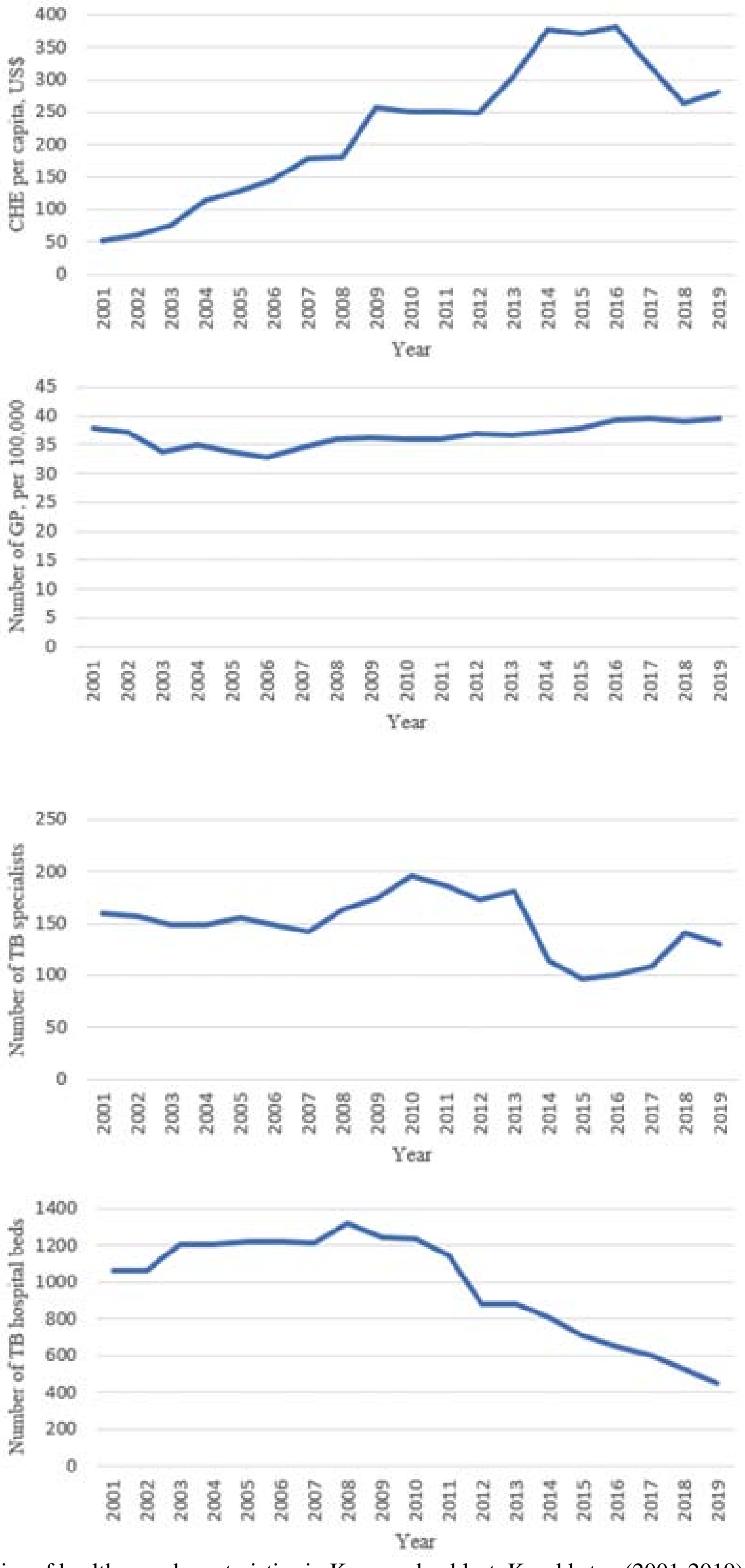

Figure 3 - Time series of healthcare characteristics in Karaganda oblast, Kazakhstan (2001-2019)

Table 1 - Bivariate associations of TB incidence with socioeconomic factors

	1	2	3	4	5	6	7	8	9
1. TB incidence rate	1	-0.945***	-0.949***	0.871***	0.758***	-0.901***	-0.682***	0.359	0.766***
2. Per-capita GDP		1	0.997***	-0.771***	-0.79***	0.836***	0.682***	-0.505*	-0.874***
3. Per-capita income			1	-0.784***	-0.793***	0.835***	0.684***	-0.481*	-0.857***
4. Proportion of the poor				1	0.695***	-0.881***	-0.367	0.204	0.461*
5. Unemployment rate					1	-0.817***	-0.235	0.453*	0.498*
6. Per-capita CHE						1	0.463*	-0.501*	-0.622**
7. Number of GP							1	-0.348	-0.831***
8. Number of TB specialists								1	0.607**
9. Number of TB hospital beds									1

Note: TB – Tuberculosis; GDP – Gross Domestic Product; CHE – Current Health Expenditure; GP – General practitioners. *p < .05, **p < .01; ***p < .001

Table 2 - Multivariate linear regression predicting TB incidence rate

Variable	**B**	**SE**	**95%CI**	***p*-value**
B_0:	5772	3091	311.3;11654.7	0.092
B_1: Per-capita GDP	0.001	0.001	-0.002;0.001	0.476
B_2: Per-capita income	0.002	0.025	-0.052;0.051	0.932
B_3: Proportion of the poor	3.351	20.46	-44.44;42.66	0.816
B_4: Unemployment rate	-0.939	29.11	-79.54;24.84	0.929
B_5: CHE per capita	-2.216	1.925	-6.223;1.163	0.172
B_6: Number of GP	-68.92	60.84	-153.2;78.76	0.262
B_7: Number of TB specialists	-3.198	2.631	-9.25;1.054	0.158
B_8: Number of TB hospital beds	-0.334	1.032	-2.704;0.99	0.657

Note: TB – Tuberculosis; GDP – Gross Domestic Product; CHE – Current health expenditure; GP – general practitioners. Tolerance < 0.1; VIF > 10; $R^2 = .985$; $F(9,10) = 72.15$ ($p < .001$)

the tolerance statistics and VIFs of all variables except for number of GP and TB specialists were less than 0.1 and greater than 10, respectively. This implies that the model showed multicollinearity of the predictors.

The findings indicated non-significant associations with the incidence rate of TB after including all predictor variables in the model. These include per-capita GDP ($B = 0.001$; $p = .476$), per-capita income ($B = 0.002$; $p = .932$), proportion of the poor ($B = 3.351$; $p = .816$), unemployment rate ($B = -0.939$; $p = .929$), CHE per capita ($B = -2.216$; $p = .172$), number of GP ($B = -68.92$; $p = .262$), number of TB specialists ($B = -3.198$; $p = .158$), and number of TB hospital beds ($B = -0.334$; $p = .657$).

Per-capita GDP, per-capita income, proportion of the poor, unemployment rate, CHE per capita, number of GP, and number of TB hospital beds were statistically significant predictors of the TB incidence rate in the bivariate analysis, but not statistically significant in the multivariable model. However, number of TB specialists was not associated with the incidence rate of TB in both analyses.

A PCA was conducted on the socioeconomic factors with oblique rotation (direct oblimin). The KMO measure verified the sampling adequacy for the analysis ($KMO = .709$). Bartlett's test of sphericity indicated that correlations between items were sufficiently large for PCA (X^2 (28) = 238.9, $p < .001$). Two components had eigenvalues over Kaiser's criterion 1 and in combination explained 84.14% of the variance. The correlation between the components was positive and moderate ($r = .533$).

Table 3 describes the factor loadings after rotation. The variables that cluster on the same components suggest that component 1 represents economic development and component 2 healthcare capacity. Economic development is directly associated with CHE per capita ($r = .947$), per-capita GDP ($r = .892$) and income ($r = .887$), and negatively with proportion of the poor ($r = -.919$) and unemployment rate ($r = -.910$). While healthcare capacity is directly correlated with number of GP ($r = .895$) and TB specialists ($r = .668$), and negatively with number of TB hospital beds ($r = -.973$). Since there was no item with a factor loading of less than 0.50, no item was excluded from the analysis.

Table 4 shows multivariate associations between the two components and socioeconomic predictors. The equation on economic development (y_1) is developed with proportion of the poor (x_1), unemployment rate (x_2), CHE per capita (x_3), per-capita income (x_4) and GDP (x_5), number of GP (x_6), TB hospital beds (x_7), and TB specialists (x_8) as follows: $y_1 = -0.98*x_1 - 0.94*x_2 + 0.883*x_3 + 0.639*x_4 + 0.623*x_5 - 0.131*x_6 - 0.103*x_7 - 0.067*x_8$. As for healthcare capacity (y_2), the equation as follows: $y_2 = 0.113*x_1 + 0.057*x_2 + 0.122*x_3 + 0.475*x_4 + 0.496*x_5 + 0.965*x_6 - 0.918*x_7 - 0.632*x_8$.

Table 5 describes the bootstrap results of the MLR analysis that investigated multivariate associations of TB incidence rate with the principal components derived from the PCA. The results show that the coefficient of determination (R^2) was 0.919, implying that the model explained 91.9% of the variance of the TB incidence. With respect to collinearity aspects, the tolerance statistics and VIFs of all variables were greater than 0.1 and less than 10. This implies that the model did not show collinearity of the predictors ($r = .533$, $p < .001$).

The results suggest statistically significant associations with the incidence rate of TB after including the principal components in the MLR model. These include economic development ($B = -454.3$; $p = .002$), healthcare capacity ($B = -211.4$; $p = .029$). Economic development and healthcare capacity were statistically associated with

Table 3 - Bivariate associations of socioeconomic determinants with principal components

Variables	Component 1	Component 2
CHE per capita	.947	.592
Proportion of the poor	-.919	
Unemployment rate	-.910	
Per-capita income	.892	.816
Per-capita GDP	.887	.828
Number of TB hospital beds	-.592	-.973
Number of GP		.895
Number of TB specialists		.668

Note: CHE – Current health expenditure; GDP – Gross Domestic Product; GP – general practitioner; TB – tuberculosis. Factors loading < .50 are suppressed

Table 4 - Multivariate associations of socioeconomic determinants with principal components

Variables	Component 1	Component 2
Proportion of the poor	-.980	.113
Unemployment rate	-.940	.057
CHE per capita	.883	.122
Per-capita income	.639	.475
Per-capita GDP	.623	.496
Number of GP	-.131	.965
Number of TB hospital beds	-.103	-.918
Number of TB specialists	-.067	-.632

Note: CHE – Current health expenditure; GDP – Gross Domestic Product; GP – general practitioner; TB – tuberculosis.

Table 5 - Multivariate linear regression predicting TB incidence with principal components

Variable	**B**	**SE**	**95%CI**	***p*-value**
Intercept	1442	48.24	1364; 1552	< .001
Component 1	-454.3	89.49	-732.7; -351.8	.002
Component 2	-211.4	58.91	-315.6; -71.21	.029

Tolerance > 0.1; VIF < 10; R^2 = .919; F (2,17) = 97.03 (p < .001)

the TB incidence rate in both bivariate (r = -.914, p < .001; r = -.732, p < .001, respectively) and multivariate analyses.

Discussion

In the current study, we examined the associations of socioeconomic determinants with TB incidence in Karaganda region, Kazakhstan. First, we investigated the influence of various socioeconomic indicators on TB incidence. The bivariate analysis showed multicollinearity of the predictors, which led to non-significant coefficients in multivariate analysis. Second, we employed PCA to overcome multicollinearity of the predictors. By performing PCA, we detained two principal components and investigated their impact on TB incidence.

The two principal components represent economic development and healthcare capacity. The economic development includes GDP per capita, proportion of the poor, per-capita income, unemployment rate, CHE per capita, whereas the healthcare capacity number of GPs, TB specialists and TB hospital beds. Our findings show a clear ecological relationship between economic and healthcare development and TB incidence. Economic development and health system capacity have significant negative effects on TB incidence. With the increase of the above principal components, the registered incidence of TB experiences a decline. This finding confirms the results of a recent global study that social protection expenditure is strongly associated with TB incidence rate (21). Furthermore, it also corroborates the report that both economic development and improved accessibility to healthcare diminish TB incidence (22).

Among the limitations of the present study, the ecological fallacy stands out (23), meaning that the data are analyzed at the population level and the findings cannot be interpreted at the individual level. Furthermore, the utilization of secondary data, which may contain typing errors and missing information, which may interfere with the analyses. Another limitation was due to data unavailability and scarcity, implying that other variables believed to be related to TB incidence could not be incorporated, such as alcohol consumption, drug abuse, educational conditions, and nutritional status. Regardless of the limitations, the analysis of relationship in ecological studies is considered vital for generating new hypotheses.

These results support that the incidence of TB is strongly determined by socioeconomic factors. Economic development appears to have a great impact on the health status of the population. In this study, the indicators of the economic development were presented by per-capita GDP, per-capita income, proportion of the poor, unemployment rate and highly correlated with the incidence of TB. The results of the study suggest that healthcare expenditure and capacity is also related to the high incidence of TB. This idea is supported by the associations between the healthcare sector indicators such as per-capita CHE, number of GPs, TB specialists, hospital beds, and TB incidence. The benefits of extended economic investments and strengthened healthcare system go beyond TB control and would be likely to influence the incidence and mortality of other social diseases. The findings suggest that economic investments and efficient healthcare system capacity are associated with the incidence of TB. Furthermore, the results of the study could be used as input parameters to dynamic simulation models that project the effects of socioeconomic interventions on TB incidence. The MLR equation on TB incidence (y) was developed with economic development (x_1) and healthcare capacity (x_2): $y = 1442 – 454.3x_1 – 211.4x_2$. However, further studies are needed to confirm the association between the level of socioeconomic development and TB incidence, as well as to incorporate the equation in the dynamic simulation models and test its validity for projection of TB incidence.

Conclusions

In conclusion, the study indicated that economic development and healthcare capacity are closely associated with the incidence of TB. Furthermore, this study contributes to the process of developing dynamic simulation models that evaluate the effects of socioeconomic interventions on TB incidence. Therefore, the findings support the implementation of optimal preventive measures for TB control, including improving the level of economic status, increasing social protection, health

expenditure, and strengthening healthcare sector capacity, which are key determinants of the incidence of TB.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.
Conflict of interest: The authors declare no competing interests.
Ethical approval: not required
Acknowledgements: We thank the Regional Center of Phthisiopulmonology in Karaganda for access to the NRTP database.

Riassunto

Incidenza della Tubercolosi e suoi determinanti socioeconomici: lo sviluppo di un modello parsimonioso

Premessa. La tubercolosi è una malattia trasmissibile assai diffusa, ed è una delle prime 10 cause di morte a livello globale. Numerosi modelli di regressione sono stati predisposti e quindi utilizzati per le proiezioni dell'incidenza della tubercolosi. Tuttavia, quando si applica un modello di regressione lineare multipla, detta analisi deve tenere conto degli aspetti di multicollinearità. Il presente studio mirava a sviluppare un modello parsimonioso che producesse risultati imparziali basati su variabili socioeconomiche come predittori dell'incidenza della tubercolosi.

Disegno dello studio. Studio ecologico.

Metodi. I dati sono stati raccolti dal Centro regionale di Tisio-pneumologia di Karaganda e dall'Ufficio nazionale di statistica. Con il modello di regressione lineare multipla, abbiamo studiato le associazioni tra il tasso di incidenza della tubercolosi ed i determinanti socioeconomici nella regione di Karaganda, in Kazakistan, nel periodo 2001-2019. È stata eseguita un'analisi delle componenti principali sulle variabili socioeconomiche con rotazione obliqua. Inoltre, sono state valutate le associazioni dell'incidenza della tubercolosi con le principali componenti derivate dall'analisi delle componenti principali.

Risultati. L'incidenza della tubercolosi nella regione di Karaganda è risultata diminuita nel periodo 2001-2019. Lo sviluppo economico e la capacità di assistenza sanitaria risultavano correlati negativamente con l'incidenza della tubercolosi. È stata sviluppata un'equazione basata sul modello di regressione lineare multipla circa l'incidenza della tubercolosi (y) con lo sviluppo economico (x1) e la capacità di assistenza sanitaria (x2) che raggruppano due componenti (utilizzando l'analisi delle componenti principali) per eliminare la collinearità: y = 1442 – 454,3 x 1 – 211,4 x 2. L'incidenza della tubercolosi è diminuita con l'aumento dello sviluppo economico e della capacità di assistenza sanitaria.

Conclusioni. Lo studio ha indicato che lo sviluppo economico e la capacità di assistenza sanitaria sono strettamente e negativamente associati all'andamento dell'incidenza della tubercolosi. I risultati supportano l'attuazione di misure preventive ottimali per il controllo della tubercolosi, compreso il miglioramento del livello di condizioni economiche, l'aumento della protezione sociale e della spesa sanitaria, ed il rafforzamento della capacità del settore dell'assistenza sanitaria, che sono determinanti chiave dell'incidenza della tubercolosi.

References

1. World Health Organization (WHO). Global tuberculosis report 2020 [Internet]. Geneva: World Health Organization; 2020 [cited 2022 Mar 13]. Available on: https://apps.who.int/iris/handle/10665/336069.
2. MacNeil A, Glaziou P, Sismanidis C, Maloney S, Floyd K. Global Epidemiology of Tuberculosis and Progress Toward Achieving Global Targets - 2017. MMWR Morb Mortal Wkly Rep. 2019 Mar 22; **68**(11): 263-6. doi: 10.15585/mmwr.mm6811a3.
3. Bhargava A, Bhargava M, Juneja A. Social determinants of tuberculosis: context, framework, and the way forward to ending TB in India. Expert Rev Respir Med. 2021 Jul; **15**(7): 867-83. doi: 10.1080/17476348.2021.1832469. Epub 2020 Oct 20.
4. Im C, Kim Y. Spatial pattern of tuberculosis (TB) and related socio-environmental factors in South Korea, 2008-2016. PloS One. 2021; **16**(8): e0255727. doi: 10.1371/journal.pone.0255727.
5. Murray M, Oxlade O, Lin H-H. Modeling social, environmental and biological determinants of tuberculosis. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis. 2011 Jun; **15** Suppl 2: 64-70. doi: 10.5588/ijtld.10.0535.
6. Sun W, Gong J, Zhou J, et al. A spatial, social and environmental study of tuberculosis in China using statistical and GIS technology. Int J Environ Res Public Health. 2015 Jan 27; **12**(2): 1425-48. doi: 10.5588/ijtld.10.0535.
7. Duarte R, Lönnroth K, Carvalho C, et al. Tuberculosis, social determinants and co-morbidities (including HIV). Pulmonology. 2018 Mar-Apr;

24(2): 115-9. doi: 10.1016/j.rppnen.2017.11.003. Epub 2017 Dec 21.
8. Acosta LMW, Bassanesi SL. The Porto Alegre paradox: social determinants and tuberculosis incidence. Rev Bras Epidemiol Braz J Epidemiol. 2014; **17** Suppl 2: 88-101. doi: 10.1590/1809-4503201400060008.
9. Zille AI, Werneck GL, Luiz RR, Conde MB. Social determinants of pulmonary tuberculosis in Brazil: an ecological study. BMC Pulm Med. 2019 May 8; **19**(1): 87. doi: 10.1186/s12890-019-0855-1.
10. de Castro DB, de Seixas Maciel EMG, Sadahiro M, Pinto RC, de Albuquerque BC, Braga JU. Tuberculosis incidence inequalities and its social determinants in Manaus from 2007 to 2016. Int J Equity Health. 2018 Dec 29; **17**(1): 187. doi: 10.1186/s12939-018-0900-3.
11. Floyd K, Glaziou P, Zumla A, Raviglione M. The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the End TB era. Lancet Respir Med. 2018 Apr; **6**(4): 299-314. doi: 10.1016/S2213-2600(18)30057-2.
12. Castro DB de, Pinto RC, Albuquerque BC de, Sadahiro M, Braga JU. The Socioeconomic Factors and the Indigenous Component of Tuberculosis in Amazonas. PloS One. 2016; **11**(6): e0158574. doi: 10.1371/journal.pone.0158574.
13. Wu J, Dalal K. Tuberculosis in Asia and the pacific: the role of socioeconomic status and health system development. Int J Prev Med. 2012 Jan; **3**(1): 8-16.
14. Chandra S, Sharma N, Joshi K, Aggarwal N, Kannan AT. Resurrecting social infrastructure as a determinant of urban tuberculosis control in Delhi, India. Health Res Policy Syst. 2014 Jan 17; **12**(3): doi: 10.1186/1478-4505-12-3.
15. Marill KA. Advanced statistics: linear regression, part II: multiple linear regression. Acad Emerg Med Off J Soc Acad Emerg Med. 2004 Jan; **11**(1): 94-102. doi: 10.1197/j.aem.2003.09.006.
16. Beckstead JW. Isolating and Examining Sources of Suppression and Multicollinearity in Multiple Linear Regression. Multivar Behav Res. 2012 Mar 30; **47**(2): 224-46. doi: 10.1080/00273171.2012.658331.
17. P. Vatcheva K, Lee M. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. Epidemiol Open Access [Internet]. 2016 [cited 2022 Mar 13]; 06(02). Available on: https://www.omicsonline.org/open-access/multicollinearity-in-regression-analyses-conducted-in-epidemiologic-studies-2161-1165-1000227.php?aid=69442.
18. Kim JH. Multicollinearity and misleading statistical results. Korean J Anesthesiol. 2019 Dec; **72**(6): 558-69. doi: 10.4097/kja.19087. Epub 2019 Jul 15.
19. Bureau of National Statistics [Internet]. [cited 2022 Feb 18]. Available on: https://stat.gov.kz.
20. Austin PC. Using the bootstrap to improve estimation and confidence intervals for regression coefficients selected using backwards variable elimination. Stat Med. 2008 Jul 30; **27**(17): 3286-300. doi: 10.1002/sim.3104.
21. Zhang Q-Y, Yang D-M, Cao L-Q, et al. Association between economic development level and tuberculosis registered incidence in Shandong, China. BMC Public Health. 2020 Oct 16; **20**(1): 1557. doi: 10.1186/s12889-020-09627-z.
22. Zhang W-H, Xiao C, Ren T, Wang L-P, Wang L, Yuan P. [Ecological Trendofthe Incidence of Tuberculosis in Mianyang City During 2004-2013]. Sichuan Da Xue Xue Bao Yi Xue Ban. 2016 Sep; **47**(5): 759-62. Chinese.
23. Roumeliotis S, Abd ElHafeez S, Jager KJ, et al. Be careful with ecological associations. Nephrology (Carlton) 2021 Jun; **26**(6): 501-5. doi: 10.1111/nep.13861. Epub 2021 Feb 17.

Corresponding author: Temirlan Ukubayev, Karaganda Medical University, 1000008 Gogol Street 40, Karaganda, Kazakhstan
e-mail: ukubaevt@qmu.kz