보건학 석사학위논문

A Spatial Analysis of Tuberculosis and its Relationship to Social Characteristics in Korea

결핵과 사회적 특성의 지역적 연관성 분석

2014 년 8 월

서울대학교 대학원
보건학과 보건학전공
백 연 수

A Spatial Analysis of Tuberculosis and its Relationship to Social Characteristics in Korea

지도교수 조 성 일

이 논문을 보건학 석사학위논문으로 제출함

2014 년 4 월

서울대학교 대학원

보건학과 보건학전공

백연수

백연수의 석사학위논문을 인준함

2014 년 6 월

위 원 장	성 주 헌	(인)
부 위 원 장	정 해 원	(인)
위 원	조 성 일	(인)

Abstract

A Spatial Analysis of Tuberculosis and its Relationship to Social Characteristics in Korea

Baik Yeon Soo
Department of Public Health
The Graduate School of Public Health
Seoul National University

Tuberculosis (TB) remains unceasing public health burden in Korea. Even the newly notified cases have been increasing steadily, while the total case has not varied much and mortality has decreased. The study focuses on social characteristics pertinent TB incidence based on the concept that TB is a social disease.

This study is an ecological approach to identify whether TB and social characteristics were significantly local correlated in South Korea, using surveillance data collected during 2012. It differentiates from preceding studies by conducting spatial analysis which leads a

conclusion that social characteristics or structures at contiguous areas can affect TB incidence in the neighborhood in specific regions.

Distribution of TB in South Korea was mapped at Si, Gu, and Gun, district level in the aspects of age and sex standardized incidence. Analysis for the local correlation between social characteristics and TB incidence was conducted by using OSL regression, univariate local index of spatial association and bivariate local index of spatial association (BiLISA).

Economic status, Occupational status, Education level, Overcrowded accommodation, housing environment, social participation and TB incidence rate were hardly spatial autocorrelated. What was significantly related to TB distribution in South Korea were proportion of service workers or salesclerk and proportion of residents living in households with insufficient housing facilities. This conclusion should be considered that these factors were also related to poor social determinants in neighbor districts according to bivariate LISA analysis, therefore the correlation coefficient or precision could be exaggerated. However, there were statistically significant high-high and low-low clusters, located in eastern coast and south coast areas and northwestern of the country respectively for one of occupational status and of housing environment variables.

In conclusion, this study suggests that this study also can help decision makers and field staff direct prevention and control interventions for improving living conditions or operational environment in order to control high TB incidence rates in the high risk local areas; eastern and southern part of South Korea.

Keywords: Tuberculosis(TB), social characteristics, bivariate spatial analysis

Student Number: 2012-21890

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1-1. Background

Korea has overcome various infectious diseases and continued to achieve effective control of infectious diseases, however, still faces an unceasing burden of tuberculosis to public health. Though tuberculosis (TB) prevalence has decreased effectively compared to 50 years ago, more than 30000 TB cases have been newly notified since 2001; 36000 new cases in 2010, 39500 new cases in both 2011 and 2012. Even the newly notified cases have been increasing steadily, while total case has not varied much and mortality has decreased.

In order to analyze and expound above unique circumstances, this study started to make an issue from different approach beyond the general definition, TB is a social disease. Neighborhood environments are able to influence disease patterns through several pathways such as social interaction and some conceptual factors associated with resource deprivation. (Oren, E., et al, 2012) Given the infectious nature of TB and the geographical autocorrelation between TB and social characteristics, any analysis of social determinants of TB should consider spatial patterning. (Harling, G., et al, 2014) One study explored which risk factors associated with TB after allowing for spatial autocorrelation. TB in Brazil has continued to be ecologically associated with a lot of low social and economic status indicators. (Harling, G., et al, 2014) Besides, a number of preceding analyses of the geographical distribution of TB have been conducted in the context of socioeconomic determinants. (Harling, G., et al, 2014) Another has found TB cluster around high-deprivation areas of Hermosillo, Mexico. (G.Alvarez-Hernandez. et al., 2010) The other study has found TB clusters surrounding a homeless shelter near urban center in Texas. (Moonan, P.K., et al., 2004) At the city level, spatial studies of TB in several countries have found associations with broad social deprivation indices, low education and asset ownership, and household crowding. (Vendramini, S.H.F., et al., 2006) Lower SES neighbors may have social backgrounds with attributes which encourage behaviors known as risk factors for TB, such as increased interaction in crowded areas, poor nutrition and smoking. (Oren, E., 2012). Most of the proximate risk factors for tuberculosis are associated with social conditions. Risk factors that seem to be of importance at the population level include poor living and working conditions associated with high risk of TB transmission, and factors that impair the host's defense against TB infection and disease, such as HIV infection, malnutrition, smoking, diabetes, alcohol abuse, and indoor air pollution. (Lee, J.H., 2011) People from low socioeconomic status groups typically have more frequent contact with people with active disease, a higher likelihood of crowded living and working conditions, greater food insecurity, lower levels of health awareness or less power to act on existing knowledge concerning healthy behavior, and less access to quality health care than do those from high socioeconomic groups. Malnutrition, crowding, and exposure to indoor air pollution are direct markers of poverty. (Knut Lönnroth, K. G. C., et al., 2010) Among them, disease-related risk factors or individual-level risk factors such as smoking, or malnutrition were not considered for this study.

1-2. Objectives

The aim of this study is to capture how associations between social determinants and TB incidence are spatially clustered in South Korea. It uses reported TB cases in Korea (2012) and spatial geocoded data, and has two specific objectives; first, it characterizes the spatial pattern of TB in Korea, second, it analyses how TB rates are locally correlated with social factors. It differentiates from preceding studies by conducting spatial analysis which leads a conclusion that social characteristics or structures at contiguous areas can affect TB incidence in the neighborhood in specific regions; local correlation. Since there is an issue of that TB has occurred steadily in spite of national control and treatment in Korea, spatial information can help inform efforts to target ongoing strategies, and risk factors epidemiology provides important information on which region remains at risk of infection. (Harling, G., et al, 2014)

Chapter 2. Methods

2-1. Study Design and Geographic information

This study is an ecological approach to identify whether TB and social characteristics were significantly local correlated in South Korea, using surveillance data collected from January 1, 2012 to December 31, 2012. South Korea is comprised of 9 'Do's and 8 'Metropolitan-Si(City)'s as the highest level(Province), Seoul-Capital-Si, Busan-Metropolitan-Si, Daegu-Metropolitan-Si, Incheon-Metropolitan-Si, Guangju-Metropolitan-Si, Daejeon-Metropolitan-Si, Ulsan-Metropolitan-Si, Sejong-Metropolitan-Si, Geoungi-Do, Gangwon-Do, Chungcheongbuk-Do, Chungcheongnam-Do, Jeonlabuk-Do, Jeonlanam-Do, Geoungsangbul-Do, Geoungsannam-Do, and Jeju-Do. The population density varies from 88.2(Busan-Metropolitan-Si) to 16188.9(Seoul-Capital-Si) in 2010. Each Do and Metropolitan-Si is again divided into several lower level(district), 'Si', 'Gun', and 'Gu's, and 249-Si-Gun-Gu in all. TB cases were geo-coded to 'Si-Gun-GU' level.

2-2. Data Collection

Notified Tuberculosis (2012)

TB is categorized into Group III infectious diseases which require continuous surveillance and control measures against the outbreak thereof as they might prevail intermittently. It is mandatory to report the occurrence of TB in that TB is one of 75 national notifiable infectious diseases. The Notified Tuberculosis data collected through the national TB monitoring system according to the Act on the prevention and management of infectious diseases, Article 11(Notification by doctors) and Article 12(Notification by other reporters) and Act on TB prevention, Article 8(Notification obligations of health care providers). People who had been reported to Korea Control and Prevention Disease Center (KCDC) were considered as study population. The case notification region, province and district, of the report is based on the location of the patient but TB patients without residential address are counted as the resident of the reported region.

Data for variables about social characteristics

The best fitted indicators with regard to social characteristics of each Si-Gun-Gu selected from the national surveys and were also collected from the national surveys. Most of them were conducted in

2010 as Population and Housing Census was quinquennial. Gross Regional Domesitc Product (GRDP), as an indicator of economic status, however, had a different version of latest information from provinces to provinces and even Seoul-Capital-Si(City) did not provide any years' GRDP. Data, except Seoul, in 2009 was applied to the GRDP variable unavoidably. The latest data was used for Social Survey (2011). All these national data are available on the Korean Statistical Information Service (KOSIS) website.

Spatial data

KOSIS provides a Si-Gun-Gu unit map as a geospatial vector data format or shape file.

2-3. Measures

Since the number of notified TB cases in this study was based on the annual report from The National Approval Statistic, Statistics Korea, and the notified cases of tuberculosis were corrected to omit missing resident-registration-number data, there may be difference between notified new cases notification rate and exact TB incidence rate. However, this study calculated both the crude incidence rate and the age-sex-standardized incidence rate from the national statistic.

The new tuberculosis case notification rates can be obtained by dividing the total number of notified new TB cases by a mid-year population to show in a way that the number of patient per every 100,000. The mid-year population of the year was used as the standard population for total number of the notified new TB cases, while, the number of events of each province was used as the standard for that of districts under the province. Using this set of population data, TB incidence was indirectly adjusted into the age-sex-standardized incidence rate per 100000 persons. How to calculate this indirect adjusted TB incidence rate was delineated Appendix.

Data for Gross Regional Domestic Product was provided by each province but available on KOSIS website. GRDP was used as it was.

Unemployment data was from Population Census (2010). The rate of unemployment was calculated as the number of people who have not worked for a week (as of the survey date) per national population over 15 years old.

Occupational status was observed by five indicators; proportion of residents working in agriculture and fishery, proportion of service workers or salesclerk, and proportion of outdoor workers. This selection was according to references explained in the introduction part. Data was from Population Census (2010). Outdoor workers were defined as excluding people who works in the office or indoor manual works, relevant to job category I, II, and III (following by KSCO standardization). The standard of organizing follows the Korean Standard Classification of Occupations (KSCO).

Educational level was detected by proportion of people below primary educated. Data was also from Population Census (2010). People below primary educated were defined as people who have not educated in any official way, had quitted primary school, and have been in primary school. The denominator was people who were over 6 years old.

Population density was calculated as the number of national population (2010) per area by province, county, and district level (Si-Gun-GU). Area data was from Statistical Yearbook of Ministry of Land, Transport and Maritime Affairs, 2010.

Overcrowded accommodation was identified from three indicators; proportion of public collective accommodation, and proportion of public institution. Data was available in Population Census (2010).

Housing environment had three indicators; proportion of residents living in households with insufficient housing facilities. Data was from Population Census (2010).

Population movement was determined by total number of people moving in or out. Data was also from Population Census(2010).

Social participation was defined as proportion of residents participating social group activities. Social Survey (2011) data was used for the indicator, however, survey region covered province and cities (the highest level). Each province or city data was representative to Si, Gu, and Gun under each province or city.

2-4. Data Analysis

This study began data analysis via non-spatial descriptive statistics for TB AIR (indirectly Adjusted-Incidence Rate) and 11 indicators. For non-spatial analysis, ordinary least squares regression was conducted with TB AIR and 11 indicators in South Korea, 2012.

Following spatial analysis assessed whether TB AIR and indicators were spatially clustered. TB adjusted incidence rates were estimated for each district level. The rates were then incorporated into geo-demographic maps. The spatial clustering of TB was examined using the Moran I index, which measures similarities among neighboring areas. Local index of spatial association (LISA) identifies a pattern of spatial dependence inside study region limits. (Lee, S.-I., 2001) In other words, LISA for each observation is defined as an indication which extent of significant spatial clustering is likely to have similar values around the observation. (Anselin L, 1995)

A Local spatial autocorrelation statistic provides a measure, for each unit in the region, of the unit's tendency to have attributable values that are correlated with values in nearby areas. This computes a measure of spatial association for each individual location. (Anselin L, 2002)

Lastly, a bivariate spatial association analysis was performed with BiLISA (Bivariate Local Indicators of Spatial Association) index. The bivariate LISA is a straightforward extension of the LISA functionality to two different variables, one for the location and another for the average of its neighbors. (Geoda center, 2004) Local relationships between the social determinants and TB adjusted incidence rates were assessed using a bivariate local index of spatial association (biLISA). A bivariate spatial association index captures the relationship between two variables, taking the topological relationship among observations into account. (Lee, S.-I., 2001) In other words, it depicts the correlation between one variable at a location and a different variable at the contiguous locations. (Anselin L, 2006)

Using a similar rationale as in the original development of local indicators of spatial association (LISA) (Anselin L, 2005), its multivariate generalisation can be defined as follows (Anselin L et al, 2002):

$$I_{ko}^{i} = y_k^i \sum_j w_{ij}\, y_o^j$$

Above each spatial analysis was conducted leading in results of Moran's I. Although each Moran's I in each step implies different meanings, the basic concepts are coming from the same rules; a positive value for I indicates that a feature has similar neighboring features which implies a part of a cluster. A negative value for I indicates that a feature has dissimilar neighboring features, which means an outlier. In either case, the p-value for the feature must be considered statistically significant. (ArcGIS Resource Center, 2014)

Since spatial statistics integrate the relationship between space and spatial directly into mathematic equations, typically these spatial relationships are defined formally through values called spatial weights. This weights matrix is used to make a neighborhood structure reconstructed on the data to assess the extent of similarity between neighbors and interested values. There are two basic categories of neighbor definitions; contiguity (shared borders) and distance. Since weights matrices are necessary when creating spatial lags that define neighboring values, chosen weights matrix will define neighbors. (GeoDa center, 2014) In this study, a distance-based weight, threshold distance was selected to analyze further bivariate analysis. A contiguity weight was also considered, however, it has high probability to have neighbor-less areas, which indicates that using this weights matrix made unspecified results in further bivariate LISA analysis.

2-5. Statistical Tools

Descriptive analyses were conducted using SAS, Version 9.3 and Open GeoDA. Quantum GIS were used for spatial analysis.

Chapter3. Results

3-1. Adjusted TB incidence rate and Variables of interest

TB cases notified from public health centers, hospitals and medical centers for 2012 were 49532 in Korea, and 39545(79.8%) of these were new cases. These confirmed subjects for analysis excluded cases whose resident registration number was omitted, in that they included foreigners who had alien registration number(1510 of total notified cases, 1227(3.1%) of newly notified cases in 2012). Adjusted TB incidence rate (TB AIR) was calculated only for 242 districts among 249 districts, this is because of differences between geocoding of districts mapping and notifying districts. The mean of TB AIR was 80.1 and the range of its value was widely, 36.8 to 175.7. (Table 1) Descriptive statistics for epidemiological characteristics of TB AIR (indirectly Adjusted-Incidence Rate) and 11 indicators were shown in Table1.

3-2. Relationship between TB AIR and social characteristics

As a result of general regression of social characteristics and TB AIR, of the 11 indicators, the number of people moving in and out districts, population density, proportion of residents living in households with insufficient housing facilities, and the proportion of service workers or salesclerk were identified that they were significantly associated with TB AIR. (Table 2) The number of people moving in and out districts showed a negative coefficient correlation and the proportion of residents living in households with insufficient housing facilities was nearly statistical significant. Table 2 summarized the general relationship between different social characteristics and TB AIR.

Table 1. Descriptive statistics for epidemiological characteristics of TB AIR(indirectly Adjusted-Incidence Rate) and 14 indicators

Variable	Indicator	N	Mean	Std Dev	Med	Range
TB incidence rate	TB indirectly adjusted incidence rate	242	80.09	21.35	76.61	36.81-175.73
Economic status	Gross regional domestic product(GRDP) *	223	3610470	3626222	2479041	181723-21949162
Employment	Unemployment rate	249	0.42	0.05	0.43	0.26-0.54
Occupational status	Proportion of residents working in agriculture and fishery	249	0.19	0.21	0.08	0.0005-0.70
	Proportion of outdoor workers	249	2.93	2.00	2.19	0.44-10.43
	Proportion of service workers or salesclerk	249	0.20	0.06	0.20	0.06-0.41
Education level	Proportion of people below primary educated	249	0.32	0.12	0.27	0.14-0.60
Population density	Population density **	249	4014.6	6128	514.0	17.74-26963.47
Overcrowded	Proportion of public collective accommodation	249	0.0015	0.0011	0.0011	0.0002-0.0089
accommodation	Proportion of public institution	249	0.0007	0.0004	0.0007	0.0001-0.0025

	Proportion of residents living in public institution	249	0.0222	0.0170	0.0185	0.0019-0.0916
Housing environment	Proportion of households with insufficient housing facilities	249	0.0097	0.0065	0.0078	0.0005-0.0390
	Proportion of residents living in households with insufficient housing facilities	249	0.0145	0.0077	0.0132	0.0013-0.0727
Population movement	The number of people moving in and out districts ***	249	11384.25	11475	7498.50	0.0-80304
Social participation	Proportion of residents participating social group activities	249	50.15	4.74	48.9	42.60-70.00

* KRW(1019KRW=1USD)

** persons/km^2

*** persons

Table 2. Ordinary Least Squares regression

	Ordinary Least Squares Estimation					
	R^2	Loglikelihood	AIC	SC	Coefficient	*p*value
Model fit	0.205	-1124.55	2279.11	2331.87		
Variable						
GRDP					$9.28*10^{-7}$	0.104
Unemployment rate					28.22	0.582
Proportion of residents working in agriculture and fishery					33.31	0.376
Proportion of service workers or salesclerk***					231.62	*0.00028*
Proportion of outdoor workers					667.32	0.242
Proportion of people below primary educated					33.37	0.55
Population density**					0.00079	*0.040*
Proportion of public collective accommodation					205.62	0.281
Proportion of public institution					-469.69	0.353

Proportion of residents living in households with insufficient housing facilities*	4.374	*0.058*
The number of people moving in and out districts**	-0.00054	*0.014*
Proportion of residents participating social group activities	0.036	0.904
Constant	-7.34	0.798

*This was also included in interpretation in spite of having weak p-value.

**p <.05 ***p <.001

3-3. Spatial autocorrelation of TB AIR and 14 indicators

A distance-based weights matrix was applied to this study. Applying a distance-based weights matrix with threshold distance of 159794.480931 km, which was set in a sense of none-neighborless, there was barely spatial autocorrelation of TB adjusted incidence rates (Moran's I = 0.0838, *p<0.001*). (Table 3) Given the low index value, it could be said that TB incidence rates in contiguity districts were not likely to have similar values. In other words, TB incidence rates were not spatially auto-correlated.

Not only TB AIR but also Gross regional domestic product (GRDP), unemployment rate, proportion of residents working in agriculture and fishery, proportion of service workers or salesclerk, proportion of people below primary educated, population density, proportion of residents living in households with insufficient housing facilities, the number of people moving in and out districts, and proportion of residents participating social activities, 9 indicators had weak spatial auto-correlation, though each Moran's I was statistically significant. Given both low index value and non-significant p-value, proportion of public collective accommodation, and proportion of public institution were observed dispersion. (Table 3) Therefore, it could be said that there was no spatial autocorrelation of TB AIR and social characteristics in general in that all variables were obtained a low Moran's I value.

Table 3. Univariate spatial associations analysis for age-gender adjusted tuberculosis incidence rates in South Korea in a district level.

	LISA	
Indicators	Moran's I	*p-value*
TB indirectly adjusted incidence rate	0.084	*<0.001*
Gross regional domestic product(GRDP)	0.028	*0.01*
Unemployment rate	0.041	*0.004*
Proportion of residents working in agriculture and fishery	0.17	*<0.001*
Proportion of service workers or salesclerk	0.16	*<0.001*
Proportion of people below primary educated	0.20	*<0.001*
Population density	0.12	*<0.001*
Proportion of public collective accommodation	0.0034	0.631
Proportion of public institution	0.020	0.181
Proportion of resident living in households with insufficient housing facilities	0.049	*<0.001*
The number of people moving in and out	0.047	*0.005*

districts		
Proportion of residents participating social group activities	0.050	*<0.001*

However, with regard to analysis of TB AIR, there were 176 significant clusters (72.7%) were detected when using univariate local index of spatial association (LISA) analysis, among which 33 high-high clusters (18.8%) and 85 low-low clusters (48.3%) (Figure 1). Those high-high clusters were all in the southeast, especially aligned along the coastline while low-low clusters were all to the northwest, Gyunggi-Do Province and Chungcheongnam-Do Province area.

3-4. Bivariate LISA analysis (social characteristics - TBAIR)

Though LISA univariate analysis concluded that each variable did not tend to form spatial clusters except few part of study region, bivariate LISA analysis identified 111 significant high-high or low-low conglomerates for proportion of residents working in agriculture and fishery and for proportion of people below primary educated, 115 conglomerates for proportion of service workers or salesclerk, and119 conglomerates for proportion of residents living in households with insufficient housing facilities. It was shown in the southeast, especially aligned along the coastline and the northwest, Gyunggi-Do Province and Chungcheongnam-Do Province area. (Figure 2) This was irrelevant with the Moran index obtained by bivariate analysis which was also very small, which explained the observed dispersion. (Table 4) Those values were statistical significance, however, very close to zero, which explained weak correlation to TB incidence rate from the nationwide as local level analysis.

Like the preceding results, the value was almost same as zero without regard to significant p-values, which means very weak correlation between these indicators and TB incidence rate in the spatial analysis.

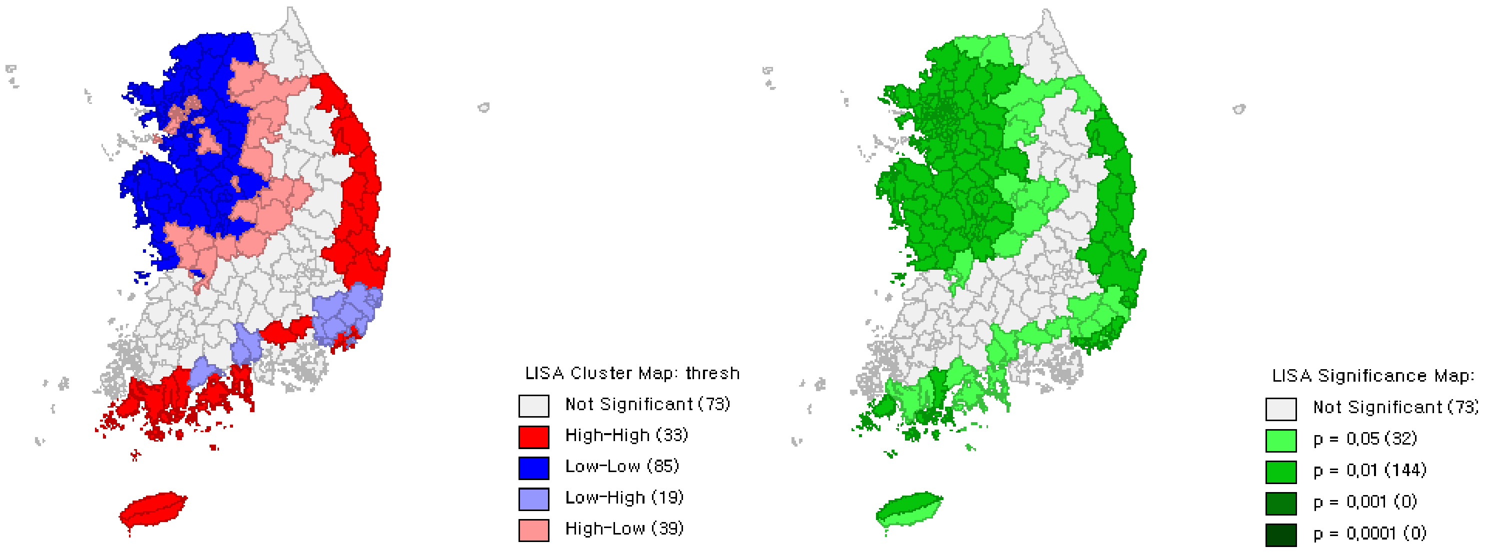


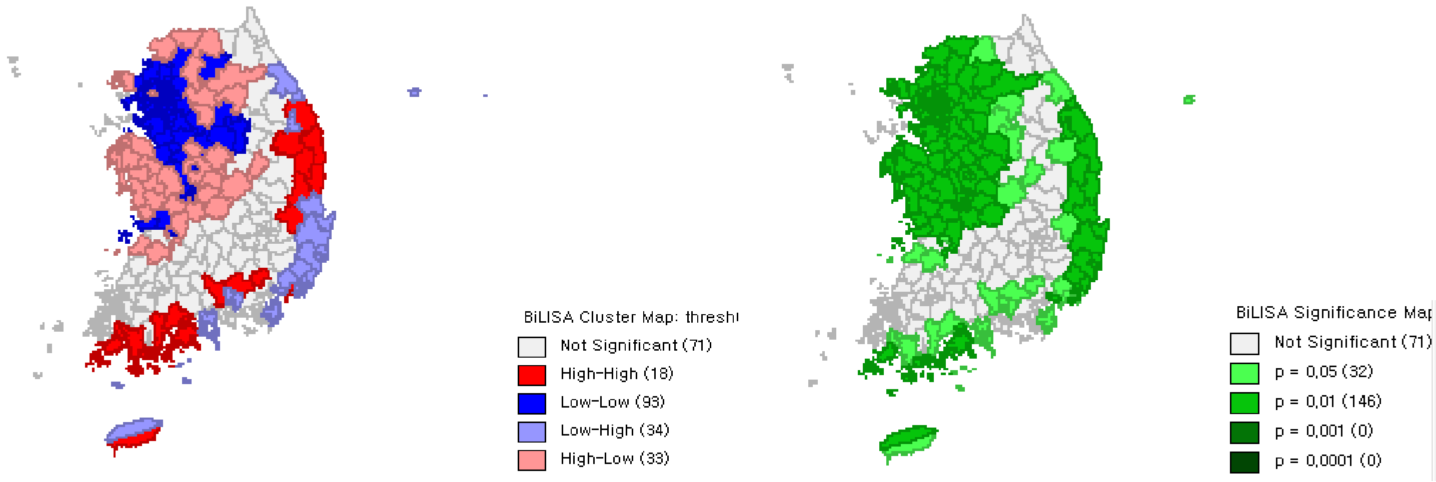
Figure 1 Univariate Local Moran's I of TB AIR (Adjusted TB incidence rate)

Table 4. Bivariate spatial associations analysis for age-gender adjusted tuberculosis incidence rates in South Korea in a district level.

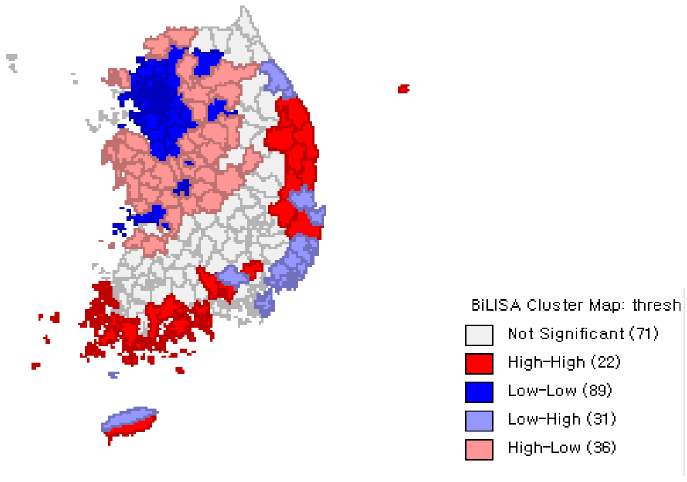
	BiLISA		
Indicators	Moran's I	Std-err	*p*-value
TB indirectly adjusted incidence rate	-	-	-
Gross regional domestic product(GRDP)	0.0181	0.0165	0.275
Unemployment rate	-0.015	0.0165	0.364
Proportion of residents working in agriculture and fishery**	0.0573	0.0161	*<0.001*
Proportion of service workers or salesclerk**	0.0659	0.016	*<0.001*
Proportion of people below primary educated**	0.681	0.016	*<0.001*
Population density*	-0.075	0.0158	*0.022*
Proportion of public collective accommodation	0.0089	0.0165	0.591
Proportion of public institution	0.015	0.0165	0.365
Proportion of resident living in households with insufficient housing facilities**	0.0663	0.016	*<0.001*
The number of people moving in and out districts*	0.0459	0.0163	*0.005*
Proportion of residents participating social	0.0159	0.0165	0.338

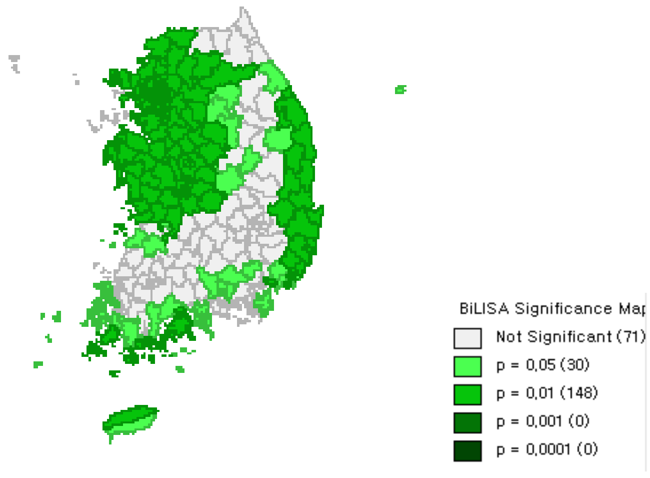
group activities

* $p<.05$ **$p<.001$

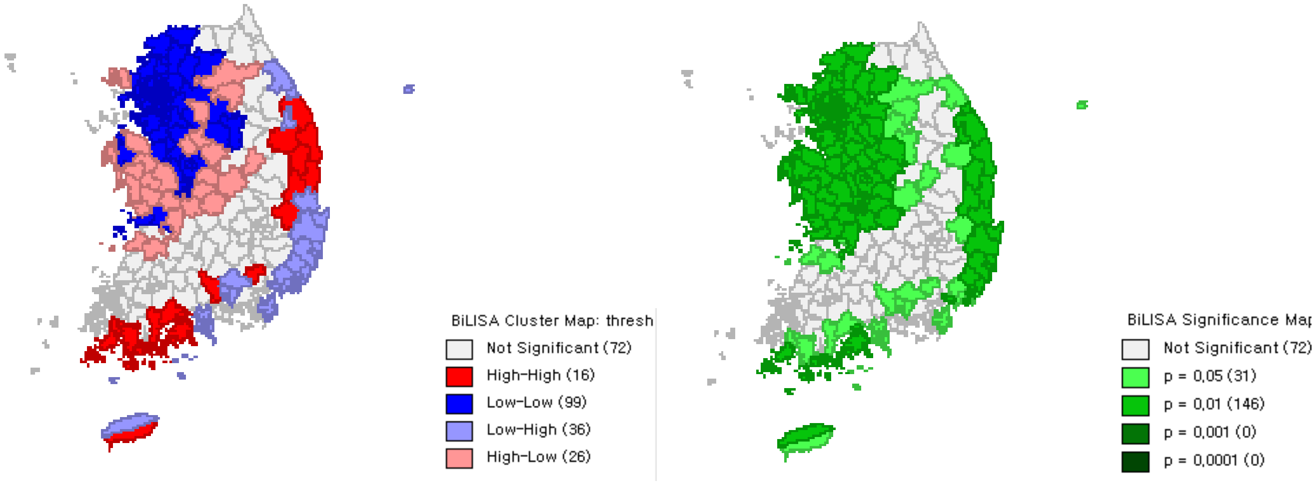


(a) Proportion of residents working in agriculture and fishery – TB AIR

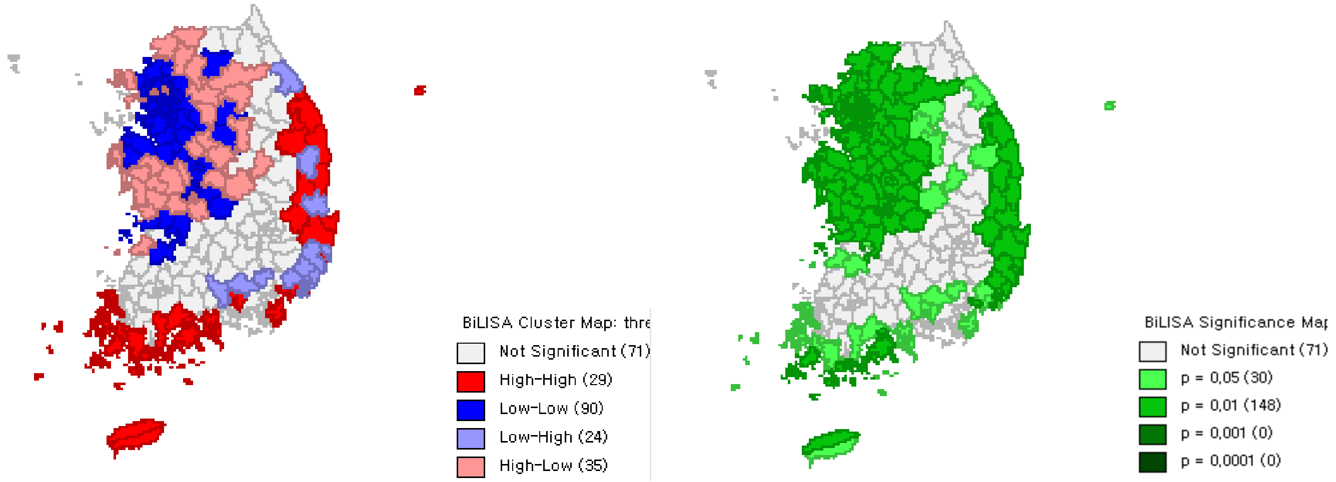




(b) Proportion of people below primary educated – TB AIR



(c) Proportion of service workers or salesclerk - TB AIR



(d) Proportion of residents living in households with insufficient housing facilities - TB AIR

Figure 2 (a), (b), (c), (d) Cluster and significance Map for BiLISA (Bivariate Local Index of spatial association) analysis between significant social indicators and TB AIR (Adjusted Incidence Rate)

Chapter 4. Discussion

4-1. Interpretation of results

In general, this research identified high-high clustering for proportion of service workers or salesclerk and for proportion of residents living in households with insufficient housing facilities in the part of study region, South Korea, where aligning with the eastern coastline to Jeonlanam-Do province areas. Studies that use GIS and spatial analysis as methods have demonstrated their usefulness in identifying high-risk areas in small units (district units in this study). (G.Alvarez-Hernandez., et al., 2010) The results of this study indicated that there were clusters of cases and incidence rates in districts in the eastern coastline and the southern areas of the country, which is useful in understanding disparities in TB distribution, although TB in South Korea did not follow a spatial autocorrelation pattern. Even bivariate LISA between social determinants and TB incidence rates represented large conglomerates, and this have meant inequality of TB distributions according to area.

According to the results of OLS regression analysis, proportion of service workers or salesclerk, population density, proportion of residents living in households with insufficient housing facilities, and the number of people moving in and out districts had statistical significant correlation coefficient value. Unlike several preceding studies, however, correlation coefficient value of population density was very low and that of the number of people moving in and out districts was negative. Both two indicators had negative bivariate local association values in BiLISA analysis. Therefore it would be possible to insist that population density and movement variable were irrelevant variables in this study and these two indicators were excluded.

In the sense of proportion of service workers or salesclerk and proportion of residents living in households with insufficient housing facilities, the results of non-spatial OLS regression analysis were significant at the same time the Moran's I of univariate local autocorrelation analysis and that of bivariate local association analysis were significant in a part of study region. This implied significant meanings that the correlation coefficients might appear bigger than they really were because the variables in nearby areas affect each other. In addition, it may lead to an exaggerated precision error. (Briggs Henan University, 2010)

Besides, while proportion of people below primary educated and proportion of residents working in agriculture and fishery had significance clustering in both spatial autocorrelation analysis and bivariate local association, multivariable regressions gave no correlation among TB AIR and these indicators. This could lead a conclusion that spatial autocorrelation of two variables made bivariate local association seemed to be correlated in each indicator.

Another significant finding of the study is that two distinct high-high and low-low conglomerates were separated distinguishably and each was correlated with social characteristic variables. In other words, TB incidence rate of one centroid districts in the east to south coastline districts was relevant to social status or social determinants of neighboring districts. Since this study set the weighted matrix as threshold distance, neighboring districts meant not 'real' neighbors which share borders but wider areas whose distance was inside the threshold distance.

Like above, though proportion of residents working in agriculture and fishery, proportion of service workers or salesclerk, proportion of people below primary educated, population density, proportion of residents living in households with insufficient housing facilities, and the number of people moving in and out districts showed significant clustering regarding association between TB AIR, it was possible to exclude irrelevant variables and confounding factors. In other words, the spatial pattern of TB in South Korea reflects unequal proportion of service workers or salesclerk and disparate proportion of residents living in households with insufficient housing facilities. This result included affection from neighbor districts so that a cautious conclusion should be needed that correlation coefficients may be larger than their true value.

It is likely to say that service workers or salesclerk who have more outdoor works or have higher contact rates with other people were easily infected TB from other carrier people or TB patients. For clarifying this inference can be obvious if the period of newly notification is considered. Especially for the eastern coastline, some of districts are famous for sightseeing or vacation. If TB statistic is increasing during July to August, the job category of service workers or salesclerk could affect significantly on TB incidence rate in this region.

Likewise, residents living in households with insufficient housing facilities can reflect on TB incidence rate in that insufficient housing facilities is one of representative indicators for poor living environment.

4-2. Strengths

Classical statistical tests rely on the assumption that subjects are independent of one another. (Arc GIS, 2014) If subjects are not independent of one another, degrees of freedom will be overestimated, which increases the probability of rejecting a true null hypothesis and committing a type I error. Since univariate LISA of TB incidence rates was significantly autocorrelated or centered but not independent in southeastern and northwestern part of the country in this study, ordinary correlation analysis might overestimate precision.

Bivariate LISA analysis shows how the nature and strength of the association between two variables varies over the study region. (UT-Dallas, B., 2007) Here, results of bivariate LISA analysis shows eastern to southern coastline part of the Korean Peninsula was in particular correlated with a decreasing social characteristics, detecting those areas with most severe TB rates.

When identifying neighborhood TB incidence was highly correlated with social determinant indicators, a common bias which implies correlation coefficients may be higher than they really are, because the areas with higher concentrations of events will affect a greater impact on the model estimate, was complemented by conducting spatial association analysis.

This study was based on the ecology approach and performed spatial analysis, however, area-based inequality assessments have been shown to generate effect estimates analogous to those yielded by individual-level measures.(Oren, E., et al,, 2012) Individual-level educational and employment opportunities, as well as other demographic factors, can be partially determined by neighborhoods in which one lives.

4-3. Limitations

The geocoding step had problems that district-level units were different from data to data. Sejong Metropolitan-Si (it is province level and district level), Dongnam-Gu (in Cheonahn-Si), Changwon-Si (in Geoyungsangnam-Do), and Mapo Huiwon-GU (in Geoyungsangnam-DO) were typical examples. These examples have one thing in common that they changed their name or incorporated into the other district. Some of insignificance clustering areas on the map corresponded to these districts.

There was a systematic error that missing data or null data were not shown in the map and included in spatial analysis.

Even though weighted matrix was selected in a reasonable way, distance band matrix has limitations as other matrices have. Also results could be varied by setting different threshold distance.

Lastly indicators used in this study were from different source and up to a few years old, while TB incidence information was collected in 2012. The possible latency period of TB was considered, nonetheless, it might be difficult to attribute observed TB incidence rates to current levels of social characteristics.(Hoa, N. B., et al., 2010)

4-4. Further studies

The study conducted analysis using sex and age standardized TB incidence rates from newly notified TB cases rather than using TB prevalence rates from total notified TB cases, because improvements in effective identification and interruption of ongoing TB transmission are important (Oren, E., et al,, 2012) to reach the TB elimination goals. However, it is difficult to explain results from only a year data of TB incidence rates and social characteristics. Even ecological studies have largely found an inverse association between SES level and TB incidence.(Oren, E., et al,, 2012) Alternative methods are suggested in the further studies, such as carrying out a comparative analysis of factors that could be key direct or indirect determinants of TB incidence trends over decades(Couceiro, L., et al., 2011) or considering to use prevalence or mortality.

Lack of district-level data such as occupational environment is leaving much to be studied. It would be possible to find out risk factors which have strongly local correlation to TB incidence rates. Also many diseases were excluded in the range of this study, however, they were well known as significant risk factors related to TB. Further studies are expected to identify a clear correlation in the local level spatial analysis as well.

Chapter 5. Conclusion

This study performed a spatial analysis ranging national data and including 249 district-level units across the country. Of 11 social characteristic indicators, 4 indicators; proportion of service workers or salesclerk, population density, proportion of residents living in households with insufficient housing facilities, and the number of people moving in and out were statistically significantly correlated according to non-spatial regression in this study. After assessing univariate LISA and bivariate LISA analysis, here, the study identified eastern to southern coastline districts in the Korean Peninsula and Jeju island, where Gangwon-Do(province) and Jeonlanam-Do(province) is located, had high TB incidence rates even after sex and age were adjusted in several social indicators.

However, what was significantly related to TB distribution in South Korea were proportion of service workers or salesclerk and proportion of residents living in households with insufficient housing facilities. This conclusion should be considered that these factors were also related to poor social determinants in neighbor districts according to bivariate LISA analysis, therefore the correlation coefficient or precision could be exaggerated. In other words, neighborhood TB incidence was highly correlated with social determinant indicators, detecting those areas with most severe TB. (E.L.N., 2010)

The study demonstrates that spatial analysis is a useful approach for the identification of geographical clusters of variables. The study suggests that more effective strategies should be taken targeting high risk local areas into account, considering the results of biLISA that a few part of significant local regions were significant. Finally, this study also can help decision makers and field staff direct prevention and control interventions for improving living conditions or operational environment in order to control high TB incidence rates in the high risk local areas; eastern and southern part of South Korea.

Chapter 6. Reference

Anseliin, L. (1995). Local Indicators of Spatial Association-LISA Geographical Analysis. 27.

Anseliin, L. (2003). GeoDa 0.9 User's Guide

Center, A. R. Spatial Weights. Retrieved April, 2014, from http://resources.esri.com/help/9.3/arcgisengine/java/gp_toolref/spatial_statistics_toolbox/spatial_weights.htm

Couceiro, L., Santana, P., & Nunes, C. (2011). Pulmonary tuberculosis and risk factors in Portugal: a spatial analysis. Int J Tuberc Lung Dis, 15(11), 1445-1454, i. doi: 10.5588/ijtld.10.0302

Grange, J. M. (2002). The global emergency of tuberculosis: what is the cause? The Journal of the Royal Society for the Promotion of Health, 122(2), 78-81. doi: 10.1177/146642400212200206

Gupta, S., Shenoy, V. P., Mukhopadhyay, C., Bairy, I., & Muralidharan, S. (2011). Role of risk factors and socio-economic status in pulmonary tuberculosis: a search for the root cause in patients in a tertiary care hospital, South India. Trop Med Int Health, 16(1), 74-78. doi: 10.1111/j.1365-3156.2010.02676.x

GeoDa. Glossary of Key Terms. Retrieved April, 2014, from https://geodacenter.asu.edu/node/390#wghts

Haining, R. Bivariate Correlation with Spatial Data.

Hargreaves, J. R., Boccia, D., Evans, C. A., Adato, M., Petticrew, M., & Porter, J. D. (2011). The social determinants of tuberculosis: from evidence to action. Am J Public Health, 101(4), 654-662. doi: 10.2105/AJPH.2010.199505

Harling, G., & Castro, M. C. (2014). A spatial analysis of social and economic determinants of tuberculosis in Brazil. Health Place, 25, 56-67. doi: 10.1016/j.healthplace.2013.10.008

Hino, P., Villa,T.C., daCunha,T.N., dosSantos,C.B., (2011). Spatial patterns of tuberculosis and its association with living conditions in the city of Ribeirao Preto in the state of Sao Paulo. Ciência Saúde Coletiva16, 4795–4802.

Hoa, N. B., Sy, D. N., Nhung, N. V., Tiemersma, E. W., Borgdorff, M. W., & Cobelens, F. G. (2010). National survey of tuberculosis prevalence in Viet Nam. Bull World Health Organ, 88(4), 273-280. doi: 10.2471/BLT.09.067801

Kim, H. J. (2012). Current Status of Tuberculosis in Korea. Korean Journal of Medicine, 82(3), 257. doi: 10.3904/kjm.2012.82.3.257

Knut Lönnroth, K. G. C., Jeremiah Muhwa Chakaya, Lakhbir Singh Chauhan, Katherine Floyd, Philippe Glaziou, Mario C Raviglione. (2010). Tuberculosis control and elimination 2010-50: cure, care, and social development. Lancet, 375. doi: 10.1016/s01406736(10)60483-7

Lai, P. C., Low, C. T., Tse, W. S., Tsui, C. K., Lee, H., & Hui, P. K. (2013). Risk of tuberculosis in high-rise and high density dwellings: an exploratory spatial analysis. Environ Pollut, 183, 40-45. doi: 10.1016/j.envpol.2012.11.025

Lee, S.-I. (2001). Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. Journal of Geographical Systems, 3.

Lonnroth, K., Jaramillo, E., Williams, B. G., Dye, C., & Raviglione, M. (2009). Drivers of tuberculosis epidemics: the role of risk factors and social determinants. Soc Sci Med, 68(12), 2240-2246. doi: 10.1016/j.socscimed.2009.03.041

Moonan, P.K., Bayona,M., Quitugua, T.N., Oppong.J., Dunbar,D., JostJr., K.C., Burgess, G., Singh,K.P., Weis,S.E. (2004). Using GIS technology to identify areas of tuberculosis transmission and incidence. Int.J. Health Geogr.3,23.

Nicole A Olson, A. L. D., Carla A Winston, Michael P Chen, Julie A Gazmararian4 and Dolores J Katz. (2012). A national study of socioeconomic status and tuberculosis rates by country of birth, United States, 1996-2005. BMV Public Health, 12(365).

Olivia Oxlade, M. M. (2012). Tuberculosis and Poverty: Why Are the Poor at Greater Risk in India? PLOS ONE, 7(11), e47533. doi: 10.1371/journal.pone.0047533

Oren, E., Koepsell, T., Leroux, B. G., & Mayer, J. (2012). Area-based socio-economic disadvantage and tuberculosis incidence. Int J Tuberc Lung Dis, 16(7), 880-885. doi: 10.5588/ijtld.11.0700

Protecttion, K. C. f. D. C. a. (2012). Annual Report on the Notified Tuberculosis Patients in Korea 2011. (11-1351159-000012-10). Korea Centers for Disease Control and Protecttion.

Souza, W.V., Ximenes,R., Albuquerque,M.F., Lapa,T.M., Portugal,J.L., Lima,M.L., Martelli, C.M., (2000). The use of socioeconomic factors in mapping tuberculosis risk areas in a city of northeastern Brazil. Rev. Panam. Salud Publica 8, 403–410.

The Commission on Social Determinants of Health Knowledge Networks, J. H. L., Ritu Sadana. (2011). Improving Equity in Health by Addressing Social Determinants.

University, B. H. (2010). Local Measures of Spatial Autocorrelation.

UT-Dallas, B. (2007). Spatial Statistics.

van Leth, F., Guilatco, R. S., Hossain, S., Van't Hoog, A. H., Hoa, N. B., van der Werf, M. J., & Lonnroth, K. (2011). Measuring socio-economic data in tuberculosis prevalence surveys. Int J Tuberc Lung Dis, 15 Suppl 2, S58-63. doi: 10.5588/ijtld.10.0417

Vendramini, S.H.F., Santos,N.S., Santos,M.d.L.S.G., Chiaravalloti-Neto,F., Ponce,M.A., Gazetta,C.E., Villa,T.C.S., Netto,A.R., 2010. Spatial analysis of tuberculosis/ HIV coinfection: its relation with socioeconomic levels in a city in south- eastern Brazil. Rev. Soc. Bras. Med. Trop. 43,536–541.

Zuur, A. Spatial Autocorrelation. Retrieved April, 2014, from http://userwww.sfsu.edu/efc/classes/biol710/spatial/spat-auto.htm

손미아. 직업, 교육수준 그리고 물질적 결핍이 사망률에 미치는 영향. 예방의학회지. 2002,35(1):76-82

조승희. (2011). Tuberculosis screening for vulnerable groups in Korea, 2011. 질병관리본부.

Appendix

Methods: age-gender standardized incidence rate

Step 1. Download crude TB notified cases data.

The data must have crude number of notified cases by province level, district level, age and gender.

Step 2. Calculate the notification rate in the standard population (Province level).

In indirect standardization, the standard population is used as providing the rates. Each Province level data would be applied to the standard population in this study. The rate is calculated as follows,

(Crude) Rate = The number of Events / The number of (Mid-year) Population

Example2. Notification rate in Seoul(Province level)								
Age groups	Events			(Mid-year) Population			Rate	
	Mal	Femal	Tota	Male	Femal	Total	Male	Female

	e	e	l		e			
0_4	3	5	8	215 520	204 435	419 955	0.000013 92	0.000024 46
5_9	5	1	6	211 559	198 918	410477	0.000023 63	0.000005 03
10_14	23	15	38	273 857	252 857	526714	0.000083 99	0.000059 32
:	:	:	:	:	:	:	:	:
80+	228	292	520	49 562	116 238	165800	0.004600 30	0.002512 09
Sum	477 6	3485	8261	497147 6	509083 3	1006230 9		
							0.00082098	

As it is shown in the above example, the rate of Seoul, one of standard populations in this study, is 8261/10062309=0.00082098

Step3. Carry out Indirect Standardized Incidence Ratio (SIR) in the study population (District level).

Expected Events could be calculated as multiplying rate in standard Population by population in study population.

Example3. Notification rate in Gangnam-gu(District level)

Standard Population(Province level)							Study Population(District level)							
Age groups	Events		(Mid-year) Population		Rate		(Mid-year) Population		Observed Events			Expected Events		
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Total	Male	Female	Total
0_4	3	5	215 520	204 435	0.00001392	0.00002446	10 900	10 374	-	-	-	0.151719	0.253724	0.405443
5_9	5	1	211 559	198 918	0.00002363	0.00000503	10 647	9 970	-	-	-	0.2516319	0.050121	0.301753
10_14	23	15	273 857	252 857	0.00008399	0.00005932	17 050	14 954	-	-	-	1.4319097	0.887102	2.319012
:	:	:	:	:	:	:	:	:	:	:		:	:	:
80+	228	292	49 562	116 238	0.00460030	0.00251209	2 331	6 175	-	-	-	10.7209959	15.512139	26.23313
Sum	4776	3485	4971476	5090833			266260	288610			334			440.89940

The SIR is a ratio of observed to expected number of incidence. It is usually expressed as a ratio of two weighted averages of age-stratum-specific notification rates,

$$\text{SIR (Standardized Incidence Ratio)} = \sum \text{Observed Events} / \sum \text{Expected Events}$$

Therefore, as following above example, SIR in Gangnam-gu district level is 334/440.89940 = 0.7575

Step4. Calculate Indirect Adjusted Incidence Rate (AIR) in each district level.

An indirectly adjusted incidence rate can be obtained by multiplying SIR by the (crude) rate of the standard population as below;

$$\text{Indirectly-Adjusted Incidence Rate (AIR)} = \text{SIR} \times (\text{Crude rate in the standard Population})$$

Therefore, AIR in Gangnam-gu district is 0.7575*0.00082098=0.00062193. This value is rearranged as per 100,000 population so that 0.00062193*100,000= 62.193.

In this way, it is possible to get each AIR in 249 districts.

Step 5. Summary

Standard Population	서울특별시	부산광역시
Crude Rate	0.000821	0.000936

Study Population	강남구	강동구	강서구
0_4 Expected Events	0.40544 3	0.406997	0.48382 9
5_9 Expected Events	0.30175 3	0.294899	0.35181 0
10_14 Expected Events	2.31901 2	1.875815	2.08844 4
:	:	:	:
75_79 Expected Events	19.3412 5	18.27813	23.8377 0
80+ Expected Events	26.2331 3	23.03423	28.8109 1
Sum of Expected Events	440.8994 02	388.25028 8	63.49407 2
Crude number of newly notified TB, Observed Events	334	340	506
SIR =Observed Events / Sum of Expected Events	0.75754 24	0.875723 8	1.11723
AIR = SIR * Crude rate	62.1930 61	71.89556 5	91.7228 9

국문초록

결핵과 사회적 특성의 지역적 연관성 분석

백연수

보건학과 역학전공

보건대학원

서울대학교

결핵은 사회경제적 수준이 낮은 나라에서 흔한 질병으로 알려져 있지만 우리나라에서 아직까지 매년 35000 건 이상이 신고되는 감염성 질환이다. 본 연구에서는 결핵과 사회적 특성의 연관성 분석에서 더 나아가 결핵의 분포에 영향을 미치는 사회적 지표와 그 영향을 분석하였다.

2012 년 전국 249 개 시군구의 결핵 신환자율을 간접표준화법을 통해 성·연령이 보정된 발생률로 계산한 후, 이용하여 Ordinary Least Squares 회귀분석과 Bivariate Local Index of Spatial Association 공간 분석을 하였다.

결핵 발생률과 유의한 상관성이 나타난 사회적 지표는 서비스업 종사자 비율, 인구 밀도, 비주거형태 거주자 비율, 시군구간 총 인구이동건수 였다. 각 변수들은 지역적으로 공간 상관성은 보이지 않았고

해당 시군구의 결핵 발생률이 인접한 시군구의 사회적 특성으로부터 받는 영향도 크지 않았다. 그러나 강원도에서 전라남도까지 이어지는 해안선 주변의 지역에서 비거주형태 거주자 비율이 높은 이웃지역이 높은 결핵 유병률을 일으킨 클러스터링, 경기도 및 충청남도에서 그와 반대되는 클러스터링이 각각 유의하게 대조되고 있었다. 이는 우리나라의 결핵 발생률 분포에 서비스업 종사자 비율과 비거주형태 거주자 비율이 영향을 미치며, 그 상관성의 정도는 다중회귀분석 결과보다 낮아질 수 있다는 점을 고려해야 함을 시사한다.

우리나라에서는 지역적 위치가 결핵과 사회적 특성의 연관성에 크게 영향을 미치지는 않지만, 클러스터링이 나타난 고위험 지역군에서는 비거주형태 거주자들의 거주환경을 개선시키기, 서비스업 종사자들의 직업환경을 개선시키기 및 위생수준을 향상시키기 등의 관리가 필요하다고 볼 수 있다.

주요어: 결핵, 사회적 특성, 공간분석

학번: 2012-21890