Epidemiological and spatial factors for tuberculosis: a matched case-control study in Nagata, Japan

Murakami, Risa

(Degree)
博士（保健学）
(Date of Degree)
2019-03-25
(Date of Publication)
2020-03-01
(Resource Type)
doctoral thesis
(Report Number)
甲第7497号
(URL)
https://hdl.handle.net/20.500.14094/D1007497

博 士 論 文

Epidemiological and spatial factors for tuberculosis: a matched case-control study in Nagata, Japan

（神戸市長田区における空間要因を考慮した結核の疫学調査）

平成３１年１月２１日

神戸大学大学院保健学研究科保健学専攻

Risa Murakami

村上　理紗

Epidemiological and spatial factors for tuberculosis: a matched case-control study in Nagata, Japan

Risa Murakami*, Naoto Matsuo†, Kouzou Ueda†, Minato Nakazawa*

*Department of Public Health, Graduate school of Health Sciences, Kobe University, Hyogo, Japan

†Department of Internal Medicine, Kobe Kyodo Hospital, Hyogo, Japan

Running head: Epidemiological and spatial factors for TB

Number of references, tables and figures: 25 references, 3tables and 4 figures

Keywords: Tuberculosis; risk factors; spatial factors; Japan

Correspondence author: Risa Murakami, Kobe University Graduate School of Heath Sciences, 7-10-2, Tomogaoka, Suma-ku, Kobe 654-0142 Hyogo, JAPAN.

E-mail: 115k903k@stu.kobe-u.ac.jp

SUMMARY

SETTING AND OBJECTIVE:

Several studies have found significant association between tuberculosis (TB) and spatial factors. The present study aimed to determine the effect of the host related factors and spatial factors associated with increased risk of TB, and to assess spatial clustering.

DESIGN:

A hospital-based case-control study using medical records was conducted. In total, 103 TB patients as a case were collected and 299 patients without TB as a control in which age and sex were matched with each case sampled from January 2000 to December 2016 at the targeted hospital in Nagata, Kobe, Japan. Logistic regression, kernel density estimation, Cross-L function and Poisson regression model were applied.

RESULTS:

Epidemiological factors associated with TB were healthcare workers (OR: 10.1) and lower serum albumin levels (OR: 0.5). Spatial analyses found that TB was positively associated with population density (RR: 32.1), the proportion of single households (RR: -1.85) and person 65 years of age and higher (RR: 2.65), and one spatial clustering.

CONCLUSION:

These findings could identify high-risk individuals and districts for TB. TB control combined with spatial analysis provided useful information for the assessment of the important factors.

Tuberculosis (TB) is an airborne infectious disease caused by Mycobacterium tuberculosis. In Japan which is one of the middle-burden countries, 17625 new TB cases and 1889 TB deaths were reported in 2016[1]. Although the number of TB patients has been reduced because of recent advances in TB control measures, the incidence of TB is higher compared to other developed countries and thus TB is still public health threat. Therefore, more certainly TB control measures to prevent TB infection and/or onset are critical.

In the past decades, risk factors for TB have been suggested by statistical methods, which were socio-demographic factors (e.g. age, sex and occupation), environmental factors (e.g. indoor air pollution) and clinical characteristics (e.g. smoking, alcohol, HIV infection and diabetes).[2,3,4,5]

Recently, several studies have found significant association between TB and spatial factors.[6,7,8,9] However, most studies investigating risk factors for TB have separately analyzed epidemiological and spatial factors using different designs, which made it impossible to assess their respective effects by controlling the effects of the other factors. By using geographic information system (GIS), it may be helpful to identify the geographical patterns and potential risk factors. Identification of area where TB is concentrated would allow the intensive intervention to such area, which may be highly efficient and feasible for sustainable management. In addition, this is particularly important in low and middle burden countries committed to TB elimination.

We aimed to simultaneously evaluate the epidemiological factors and spatial factors associated with increased risk of TB using the combination of statistical and spatial analysis. We also aimed to identify possible areas of concentrations of TB using hotspot analysis with GIS.

METHODS

Study design and participants

We carried out a retrospective hospital-based case-control study of TB patients and matched controls using medical records in Nagata, Kobe, Japan (Figure 1). The study was designed for a 1:3 case-control ratio. Kobe city where is one of the most populous cities in Japan. It has nine administrative wards, including Nagata (11.36km^2), which has approximately 97,912 inhabitants, 31,825 (32.9%) of whom are aged 65 years of age and higher residents[10]. TB incidence in Nagata was 31.9 per 100,000 population in 2016, which was higher than national average (13.9)[1].

Participants were recruited from January 2000 to December 2016 at Kobe Kyodo Hospital in Nagata (hereafter referred as the target hospital). A total of 103 cases were diagnosed with any type of TB (detailed criteria are described below) during the study period. A total of 299 controls were sampled from patients without TB who visited the hospital. We matched the patients without TB by age within 10 years and sex with the case, then randomly sampled 3 controls for each case.

TB diagnosis was based on clinical findings ascertained by physicians using the same criteria from medical records with either a positive Mycobacterium tuberculosis culture, characteristic chest X-ray abnormalities, a positive Mantoux or QuantiFERON®-TB Gold Test or positive history.

We attempted to minimize selection bias by using random sampling after matching the controls with the cases. We considered potential confounders and addressed confounding by age and sex through matching. We conducted initial descriptive analysis of the cases and the controls to identify potential associated factors.

Statistical analysis

Individual records with missing data were included and excluded pairwise in each analysis: In the worst case, slightly less than 50% variables were missing. R version 3.4.3[11] was used for all statistical analyses. A multivariate logistic regression analysis was conducted to explore the effects of explanatory variables on TB. Medical history was not included because the frequency of distribution of medical history was very similar between the cases and controls: Polychoric correlation coefficient between the distributions was greater than 0.9. The selection of explanatory variables in the final model was based on results of the following analysis: Univariate correlation coefficients with p values less than 0.1, collinearity matrix index with less than 0.7, variance inflation factor with less than 10 and biological relevance, then determined by stepwise selection. Model fit to the data was assessed by Nagelkerke's R squared and Akaike's Information Criterion (AIC).

Spatial analysis

2017 Japanese population census data in a polygon shapefile for each census tract for Kobe was obtained from e-Stat: the portal site of Japanese Government Statistics (Statistics Bureau, Ministry of Internal Affairs and Communications, Japan)[10]. The dependent variable was the number of TB patients in each polygon and variables were also polygon level data including population density, proportion of males, proportion of employee, proportion of single households (for households with single person), proportion of 65 years of age or higher which were created by edging the Population census[10] and the distance from the target hospital in meters as a linear distance matrix between each polygon. An exploratory analysis of spatial patterns in the data was used to visualize TB distribution. The population density of each polygon was calculated by dividing the total population in each polygon by the corresponding area in square kilometers. The proportion of males, employees, private households and 65 years of age and higher were calculated by dividing the total number of each variable by the total population of each polygon. These calculations were performed using the QGIS version 2.18.9[12].

We conducted kernel density estimation (KDE)[13] to identify the geographic hotspots of TB patients in the hospital and generated the KDE maps with a coverage radius of 1 km using QGIS.

All study participant postcodes were geocoded by tree-maps[14]. For spatial independency between the locations of the cases and the controls, we conducted the Cross-L function which can evaluate the dependency of two-point patterns. We also used the Diggle-Cressie-Loosmore-Ford (DCLF) test[15] for the statistical significance of the Cross-L function each with 99 Monte Carlo simulations.

Integrated analysis

We applied Poisson regression model[16,17] to explore the effects of spatial patterns and to measure spatial heterogeneity in the effects of predictor variables of sociogeographic characteristics because the dependent variable was the number of TB cases in each polygon, which was a count variable and met the rare disease condition. All significant variables were tested for multicollinearity, which were the variables with VIF with less than 10. In addition, biological relevance was considered in the selection of the explanatory variables in the final model, followed by stepwise selection. Model fit to the data was evaluated using AIC, and the ratio of residual deviance in degrees of freedom.

Ethical clearance

Due to the study design and pseudonymous data, the rights and privacy of the subjects are protected. The study protocol was approved by the ethics review boards of Graduate School of Health Sciences of Kobe University (574) and Kobe Kyodo Hospital (2016-001).

RESULTS

The 103 cases comprised 64 males and 39 females with an age range of 13-101 years. The proportion of the cases of total number of TB patients in Nagata during this study was less than 20% during the whole period (2000-2016), but about 30% since 2008 (Figure 2). The 299 controls were sampled, allowing matched by age and sex with the case (Figure 3). Patients characteristics of all cases and controls are shown in Table 1. As for the occupation, health workers (physicians, nurses and other healthcare workers) comprised 16.8% of the cases and 2.7% of the control. The proportions of persons with gastrectomy, and HBV/HCV infection among the cases were higher than the controls. Mean BMI, serum albumin level and hemoglobin level were lower in the cases than in the controls.

Table 2 shows the results of the multivariate logistic regression. In the final model (not including body mass index, dietary habit, drinking, family type, hemoglobin level, physical activity, sleeping, smoking), being healthcare workers (OR: 10.10; 95%CI: 1.01, 101.0) and serum albumin levels (OR: 0.50; 95%CI: 0.293, 0.841) were significantly associated with TB.

Figure 1 shows the distribution of clustered TB patients calculated using KDE. One geographical hotspot of the cases (black areas) surrounded by several moderately intensive clusters were observed around the target hospital.

The Cross-L function indicated that the locations of cases and controls exhibited spatial attraction as shown in Figure 4 and this was statistically significant (p=0.01), as confirmed by the DCLF test of the Cross-L function. This result showed significant difference in the spatial pattern between the locations of the cases and the controls.

The results of fitting Poisson regression model are given in Table 3. The incidence of TB per population of each district was positively associated with population density (RR: 32.1;

95%CI: 12.79,49.01), the proportion of single households (RR: -1.85; 95%CI: -3.21, -0.53), the proportion of person 65 years of age and higher (RR: 2.65; 95%CI: 0.1,4.2) and the distance (RR: -0.00029; 95%CI: -0.00037, -0.00022). The reliability and the validity of the regression model were adequate based on post-estimation statistics (Residual deviance: 606.81 on 2606 degrees of freedom; AIC: 755.48).

DISCUSSION

We found one geographical hotspot of the cases near the target hospital and several significant risk factors for TB patients. Our results showed a lower serum albumin level, healthcare worker, 65 years of age and higher, single household, high population density, short-distance from the target hospital were associated with TB.

Spatial hotspots were found only in the cases as the results of KDE and the Cross L-function. Nevertheless, these districts are higher population density and proportion of 65 years of age and higher than the other districts. The detection of TB clusters may mean that TB in the elderly has become a problem. This is biologically plausible because many elderly people who have been infected with TB when Japan was a high burden country, with comorbid illness such as diabetes, immunosuppression due to age and malnutrition[18] may have resulted in a higher probability of developing TB. In Japan, approximately two-third (71.6%) of TB cases occur in the elderly over 60 years of age and higher[1], therefore, the elderly is the largest influential risk groups, whereas in the United Kingdom and the United States, the majority of TB cases are migrants and foreign-born people[19].

Previous studies showed that serum albumin level, high population density and healthcare worker were associated with TB.[20,21,22] Those results may be due to malnutrition, higher exposure to TB patients and overcrowding. However, those results could also be related to any potential bias, such as selection bias. Healthcare workers who worked at the targeted hospital are more likely to attend the targeted hospital for TB screening and treatment. Furthermore, high population density could facilitate TB transmission. These factors could may have led to overestimate the impact of the risk factors with increased risk of TB. However, patients not suffering from TB as the controls would also be subject to the same bias. Moreover, patients can voluntary select the hospital according to Japanese universal health insurance coverage system. It was therefore noteworthy that high population density and healthcare worker were associated with having TB. Two other factors, single households and short-distance, have not been reported previously. Although the short-distance from the targeted hospital seems to be due to the research design (TB patients living in the far place may visit other hospitals), it is impressive that the factor with single household was associated with reduced risk of TB. Few scholars have focused on the household status of TB patients in Japan, but most studies have showed single households to be associated with treatment failure[23] and death due to TB;[24] however, those studies did not mention the risk of TB itself. Our finding on single households may have been due to the absence of TB transmission from other household members. The non-single households may have a higher transmission risk within the household due to poor ventilation and indoor pollution. These findings suggest that current living and crowded condition are associated with an increased risk of TB.

Multivariate logistic regression revealed that being healthcare workers and the serum albumin level were significantly associated with TB. However, confidence intervals (CIs) on healthcare workers were wide, and TB had the lowest prevalence in this study population, which could lead to wide CIs due to the small number of studies. Besides, marital status as an explanatory variable was included in the final model even though p value was not less than 0.1. This was a result of stepwise selection. Although, studies with larger study populations are needed to evaluate the impact of marital status and being healthcare workers in this population.

Our study has implications for future TB control measures. To prevent future TB cases, we recommend targeted screening of the high-risk individuals who are 65 years of age and higher, healthcare workers and those with a low serum albumin level in high-risk districts with high population density and a low proportion of single household; we also suggest that patients and healthcare workers are remained of appropriate cough etiquette and respiratory hygiene[25]. Furthermore, we recommend that TB control measures be linked with spatial distribution of TB, and the risk assessment including sociodemographic factors should be standardized for TB patients and household members. This strategy will enable better understanding of the important factors associated with TB.

Our study had four main limitations. First, the study design was retrospective, so the effect of missing data was unknown. Second, the serum albumin level may have been affected by the inflammation of TB, and thus the relationship between serum albumin and TB may be contrary. Third, our results reflect the experience of the target hospital and may not be generalisable to other settings. Fourth, the postcode area may have limited our capacity to assess the impact of spatial analysis on maternal outcomes. Despite these limitations, our study, based on the patient's socio-demographic and geographic information on associations between TB and risk factors, can provide further insight into TB dynamics in the target population and potentially contribute to planning effective TB control measures in other area of Japan.

CONCLUSION

The retrospective case-control study showed that high TB prevalence was related to individuals who are 65 years of age and higher, being healthcare worker, having a low serum albumin level, and living in the high-risk districts with high population density and a low proportion of single household. Identification of high-risk individuals and districts may improve planning for TB control measure.

Acknowledgment

The authors would like to thank all staffs and study participants involved in the study.

Conflicts of interests: none declared.

References

1 Ministry of Health, Labour and Welfare. Statistics of TB 2016 (Japanese). URL http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000175095.html. Accessed February 2018.

2 Padmanesan Narasimhan, James Wood, Chandini Raina MacIntyre, Dilip Mathai. Risk Factors for Tuberculosis. Pulmonary Medicine. Article ID 828939. 11 pages. 2013; doi:10.1155/2013/828939.

3 C Lienhardt, K Fielding, JS Sillah, et al. Investigation of the risk factors for tuberculosis: a case–control study in three countries in West Africa. International Journal of Epidemiology. 1 August 2005; 34(4): 914–923.

4 P de Colombani, A. Hovhannesyan. Social determinants and risk factors for tuberculosis in national surveillance systems in Europe. Public Health Action. 21 September 2015; 5(3): 194-201(8).

5 Andersson N, Harris E, Mitchell S, Hamel C, Shea B. Risk factors associated with recent transmission of tuberculosis: systematic review and meta-analysis. The International Journal of Tuberculosis and Lung Disease. January 2009; 13(1): 17-26.

6 Yih-Yuan Chen, Jia-Ru Chang, Chih-Da Wu, et.al. Combining molecular typing and spatial pattern analysis to identify areas of high tuberculosis transmission in a moderate-incidence county in Taiwan. Scientific Reports. 2017; 7. Article number: 5394. doi:10.1038/s41598-017-05674-6.

7 Smith CM, Trienekens SCM, Anderson C, et al. Twenty years and counting: epidemiology of an outbreak of isoniazid-resistant tuberculosis in England and Wales, 1995 to 2014. Eurosurveillance. 2017;22(8):30467. doi:10.2807/1560-7917.ES.2017.22.8.30467.

8 Guy Harling, Marcia Castro. A spatial analysis of social and economic determinants of tuberculosis in Brazil. Health & Place. 2014; 25: 56–67.

9 Izumi K, Ohkado A, Uchimura K, et al. Detection of Tuberculosis Infection Hotspots Using Activity Spaces Based Spatial Approach in an Urban Tokyo, from 2003 to 2011. PLOS ONE. 2015; 10(9): e013883.

10 Statistics Bureau. Ministry of Internal Affairs and Communications website (e-Stat). Tokyo, Japan; 2017. URL https://www.e-stat.go.jp/en/. Accessed November 2017.

11 R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. Accessed November 2017.

12 QGIS Development team. QGIS Geographic Information System. Open Source Geospatial Foundation Project. URL http://www.qgis.org/. Accessed November 2017.

13 Diggle PJ. Statistical analysis of spatial point patterns. 2nd ed. London, UK: Arnold, 2003.

14 Tree-maps. (Japanese) URL https://www.tree-maps.com. Accessed November 2017.

15 Chris Brunsdon, Lex Comber. An Introduction to R for Spatial Analysis and Mapping. In: Looking at Marked Point Patterns. London, UK: Sage Publications Ltd, 2015: pp 195-198.

16 Furutani Tomoyuki. R niyoru Kuukandeita no toukeibunseki (Japanese). 3rd ed. Tokyo, Japan: Asakura Publishing Co., Ltd, 2013: pp 155-158.

17 Kin Meitetsu. Series Useful R5 Teiseitekideitabunseki (Japanese). Tokyo, Japan: KYORITSU SHUPPAN CO., LTD, 2016: pp 157-183.

18 M Chan-Yeung, K Noertjojo, J Tan, S L Chan, C M Tam. Tuberculosis in the elderly in Hong Kong. The International Journal of Tuberculosis and Lung Disease. 2002; 6 (9).

19 Uchimura K, Ngamvithayapong-Yanai J, Kawatsu L, et al. Characteristics and treatment outcomes of tuberculosis cases by risk groups, Japan, 2007–2010. Western Pacific Surveillance and Response Journal. 2013; 4 (1). doi: 10.5365/wpsar.2012.3.4.016.

20 Cegielski J P, Arab L, Cornoni-Huntley J. Nutritional risk factors for tuberculosis among adults in the United States, 1971-1992. American Journal of Epidemiology 176, 2012: 409–422. doi:10.1093/aje/kws007.

21 De Abreu E Silva M, Di Lorenzo Oliveira C, Teixeira Neto R G, Camargos P A. Spatial distribution of tuberculosis from 2002 to 2012 in a midsize city in Brazil. BMC Public Health 2016; 16 (1). doi:10.1186/s12889-016-3575-y.

22 Seidler A, Nienhaus A, Diel R. Review of Epidemiological Studies on the Occupational Risk of Tuberculosis in Low-Incidence Areas. Respiration 2005; 72: 431-446.

23 Oki N, Nakamura H. Risk factors for treatment outcomes of tuberculosis (Japanese). Hyogo Prefectural Institute of Public Health Sciences. Annual report 2004; 2: 156-161.

24 Sasaki Y, Yamagishi F, Suzuki K, Saitoh M, Izumizaki M, Koizumi K. A STUDY ON THE BACKGROUND FACTORS OF CASES DIED OF PULMONARY TUBERCULOSIS DETECTED BY SYMPTOMATIC VISIT (Japanese). Kekkaku 1996; 71: 427-430.

25 World Health Organization. WHO policy on TB infection control in health-care facilities, congregate settings and households. Geneva, Switzerland: WHO; 2009.

Table 1 Background characteristics of 103 TB cases and 299 controls*

Variable	Cases n=103	Controls n=299
Age [mean]	61.4 [21.2]	62.1 [20.9]
Sex		
Men	64 (62.1)	184 (61.5)
Female	39 (37.9)	115 (38.5)
Marital status		
Married	66 (64.1)	114 (38.1)
Unmarried/Divorced/Widow	32 (31.1)	91 (30.4)
Missing data	5 (4.9)	94 (31.4)
Family type		
Nuclear	27 (26.2)	44 (14.7)
Extended	70 (68.0)	166 (55.5)
Missing data	6 (5.8)	89 (29.8)
Occupation		
Unemployed	52 (50.5)	165 (55.2)
Healthcare worker	17 (16.5)	8 (2.7)
Other worker	32 (31.1)	82 (27.4)
Missing data	2 (1.9)	44 (14.7)
Drinking		
No	52 (50.5)	101 (33.8)
Sometimes	15 (14.6)	27 (9.0)
Everyday	27 (26.2)	72 (24.0)
Missing data	9 (8.7)	99 (33.1)
Smoking		
No	50 (48.5)	114 (38.1)
Past/Current	41 (39.8)	85 (28.4)
Missing data	12 (11.7)	100 (33.4)
Cancer		
Yes	17 (16.5)	33 (11.0)
Missing data	0	42 (14.0)
Cardiac disease		
Yes	7 (6.8)	27 (9.0)
Missing data	0	42 (14.0)
Cardiovascular disease		
Yes	23 (22.3)	51 (17.1)
Missing data	0	42 (14.0)
Chronic kidney disease		
Yes	5 (4.9)	6 (2.0)
Missing data	0	42 (14.0)
Chronic respiratory disease		
Yes	16 (15.5)	34 (11.4)
Missing data	0	42 (14.0)
Collagen vascular disease		
Yes	1 (1.0)	1 (0.3)
Missing data	0	42 (14.0)
Diabetes		
Yes	23 (22.3)	45 (15.1)
Missing data	0	42 (14.0)
Gastrectomy		
Yes	12 (11.7)	3 (1.2)
Missing data	0	42 (14.0)
HBV/HCV infection		
Yes	16 (15.5)	15 (5.8)
Missing data	0	41 (14.0)
HIV infection		
Yes	0	0
Missing data	0	42 (14.0)
Body mass index [mean]†	20.7 [3.5]	22.1 [4.3]
Missing data	27 (26.2)	145 (48.5)
Serum albumin(G/DL) [mean]	3.3 [0.9]	3.9 [0.7]
Missing data	21 (20.4)	138 (46.2)
Hemoglobin(G/DL) [mean]	12.7 [2.1]	13.81 [5.6]
Missing data	0	92(30.8)

TB = tuberculosis; SD = standard deviation; HBV = hepatitis B virus; HCV = hepatitis C virus; HIV = human immunodeficiency virus.

Table 2 Multivariate logistic regression model for the association of clinical characteristics and socio-demographic variables with TB

Variable	OR	95%CI	P value
Alb (G/DL)	0.50	0.293, 0.841	0.009
Unmarried/Divorce/Widow	0.52	0.229, 1.01	0.125
Healthcare worker	10.10	1.01, 101.0	0.049
Other worker	1.50	0.619, 3.64	0.369

TB = tuberculosis; OR = odds ratio; CI = confidence interval.

Table 3 Poisson regression model for the association of spatial patterns and socio-geographic characteristics with TB

Variable	Coefficient	CI	P value
65 years of age or higher	2.65	0.1, 4.2	0.001
Single households*	-1.85	-3.21, -0.53	0.007
Distance from the target hospital	-0.00029	-0.00037, -0.00022	<0.001
Population density	32.19	12.79, 49.01	<0.001

*households with single person

TB = tuberculosis; CI = confidence interval.

Figure legends

Figure 1. A) Map of study area, Nagata of Kobe city. B) Spatial distribution of clustered TB patients in the targeted hospital. Kernel Density Estimation map was constructed based on Geographically Information System data which were collected by medical records. QGIS was used to create this figure. TB = tuberculosis.

Figure 2. Number of TB patients in Nagata and the targeted hospital by year, 2000-2016. TB = tuberculosis.

Figure 3. Flow diagram for the enrollment of study participants in the matched case-control study. TB = tuberculosis.

Figure 4. Cross-L function between the locations of cases and controls using R.

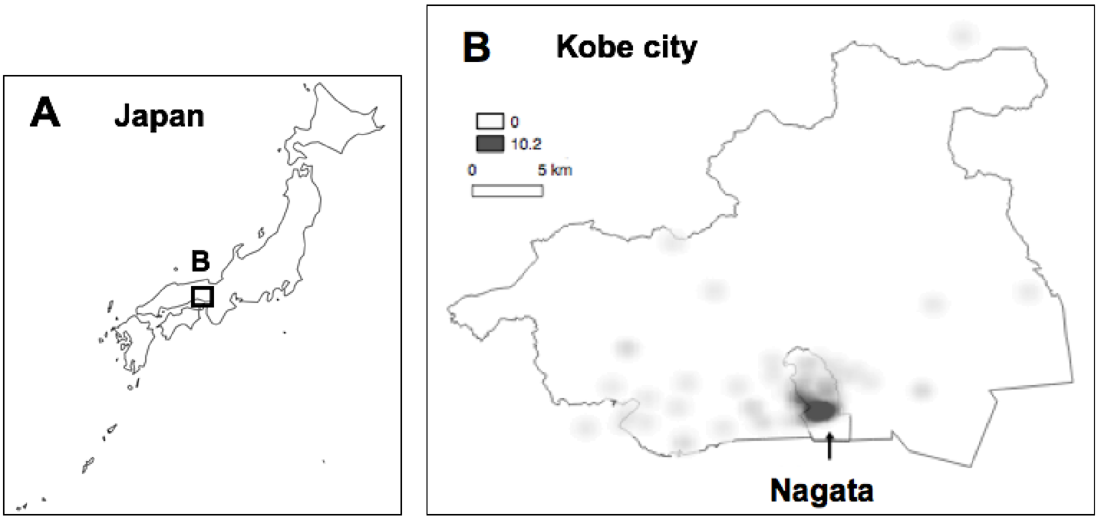

Figure 1. A) Map of study area, Nagata of Kobe city. B) Spatial distribution of clustered TB patients in the targeted hospital. Kernel Density Estimation map was constructed based on Geographically Information System data which were collected by medical records. QGIS was used to create this figure. TB = tuberculosis.

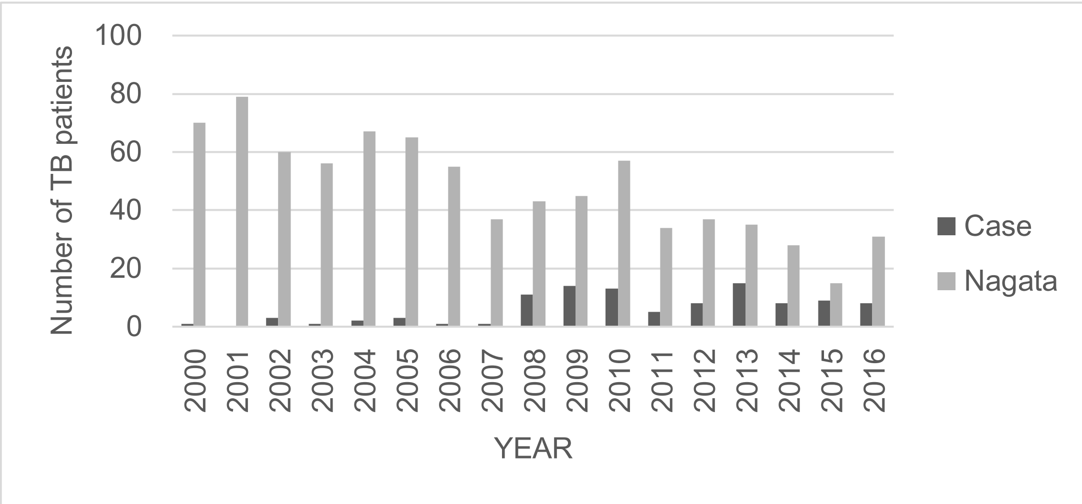

Figure 2. Number of TB patients in Nagata and the targeted hospital by year, 2000-2016.

TB = tuberculosis.

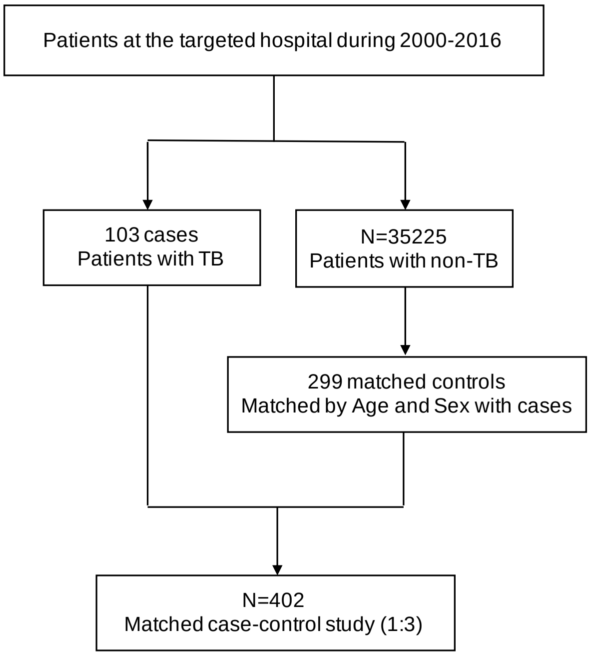

Figure 3. Flow diagram for the enrollment of study participants in the matched case-control study. TB = tuberculosis.

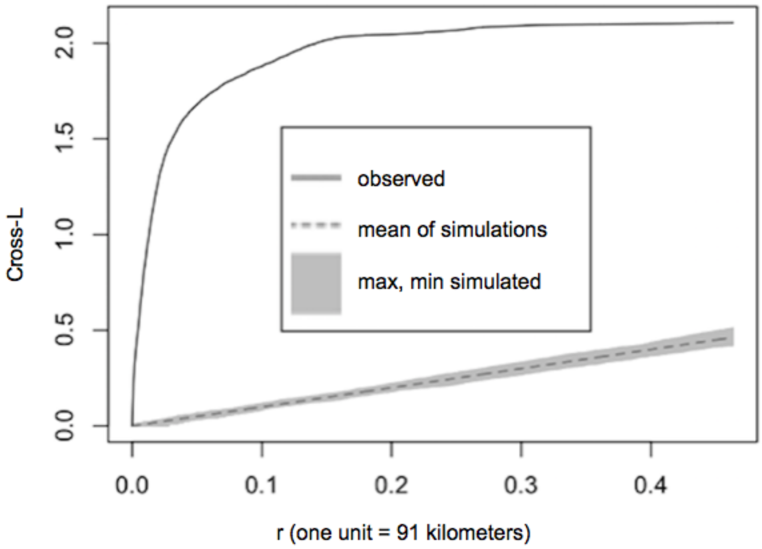

Figure 4. Cross-L function between the locations of cases and controls using R.