Doctoral Thesis

Understanding of Tuberculosis Transmission in Urban Setting
Using Molecular Epidemiology and Geographic Information System

Kiyohiko Izumi

Department of Basic Mycobacteriosis,
Unit of Mycobacteriosis,
Nagasaki University Graduate School of Biomedical Sciences
(Collaborative Course with the Research Institute of Tuberculosis, Japan)

March 2018
This thesis consists of an original study carried out by the author while enrolled in
the doctoral program in the Department of Basic Mycobacteriosis, Unit of
Mycobacteriosis, Nagasaki University Graduate School of Biomedical Sciences,
Japan from April 2014 to March 2018.

Kiyohiko Izumi
CONTENTS

ABBREVIATIONS .............................................................................................................. 5

CHAPTER 1. INTRODUCTION ......................................................................................... 6
  1-1. Active TB case finding among high-risk groups ...................................................... 6
  1.2. Molecular epidemiology for understanding TB transmission ................................. 7
  1-3. Geographic Information System for understanding TB transmission ................... 9
  1-4. Epidemiology of Tuberculosis in Japan .................................................................. 10
  1-5. Objectives of this thesis .......................................................................................... 11

CHAPTER 2. METHODOLOGY ...................................................................................... 12
  2-1. Study design and settings ...................................................................................... 12
    2-1-1. Study design ..................................................................................................... 12
    2-1-2. Settings ............................................................................................................ 12
  2-2. Subjects and materials ........................................................................................... 14
    2-2-1. Case data .......................................................................................................... 14
    2-2-2. Census tract data ............................................................................................. 15
  2-3. Analysis methods .................................................................................................... 18
    2-3-1. Spatial analysis ................................................................................................ 18
    2-3-2. Hotspot analysis .............................................................................................. 19
    2-3-3. Patient characteristics and urban environmental comparison ..................... 20
  2-4. Ethics statement ..................................................................................................... 21

CHAPTER 3. FINDINGS .................................................................................................. 23
  3-1. Characteristics of TB cases .................................................................................... 23
  3-2. Spatial analysis ...................................................................................................... 26
  3-3. Hotspot Analysis .................................................................................................. 31

CHAPTER 4. DISCUSSIONS AND CONCLUSION ....................................................... 40
  4-1. The identified TB hotspots using activity spaces based spatial approach ........... 40
    4-1-1. Identified TB hotspots ..................................................................................... 40
    4-1-2. Possible hotbeds of TB transmission at the facilities used by unspecified
ABBRVDIATIONS

CI  Confidence Interval
CT  Census Tract
DNA  deoxyribonucleic acid
GIS  Geographic Information System
IDW  Inverse Distance Weighted
ISA  Incremental Spatial Autocorrelation
PHC  Public Health Center
RFLP  Restriction Fragment Length Polymorphism
TB  Tuberculosis
CHAPTER 1. INTRODUCTION

1-1. Active TB case finding among high-risk groups

Tuberculosis (TB), a bacterial airborne infectious disease caused by Mycobacterium tuberculosis (M. tuberculosis), is one of the world’s deadliest communicable diseases. TB patients suffer from at least six months of chemotherapy with multiple toxic medications. Multidrug resistant-TB, defined as TB with resistance to both isoniazid and rifampin imposes longer treatment than susceptible-TB leading high individual and financial burden. Estimated 10 million people developed TB and 1.3 million died in 2016 worldwide\(^1\). The majority of those TB cases is reported from high incidence countries. However, TB is still a significant public health issue in low and middle incidence countries, including Japan, especially among certain high-risk groups who are likely to develop TB disease in urban area, i.e., the socio-economically depressed population such as homeless people and the elderly. Therefore, the targeted active screenings for TB to these high-risk groups in addition to the contacts with infectious TB patients are of critical importance of TB control in Japan. Moreover, it is important to understand where and how individuals obtained the TB bacilli to make efficient active TB case finding happen.
Much advance has been made as a result of development of new technologies, such as molecular epidemiological techniques, or molecular fingerprinting, which have enabled revealing of TB transmission dynamics within social networks\textsuperscript{2–5}). Another technique which has recently become available is Geographic Information System (GIS), where geographic data to describe and visualize spatial distributions and to discover spatial association patterns have been applied on the study about TB transmission.

1.2. Molecular epidemiology for understanding TB transmission

Molecular epidemiological methodology for TB appeared in the early 1990s, which is a technique to understand the TB transmission by interpreting genotyping data together with epidemiological information which is obtained through patient interview\textsuperscript{6}). Since then, as a technical aspect, the IS\textit{6110} based restriction fragment length polymorphism (IS\textit{6110}-RFLP) analysis has been widely used as TB genotyping. However, RFLP has some limitations such as necessary of skillful laboratory technician, long time until producing outcomes, and difficulty of comparison. Then, RFLP was recently replaced to Variable Number of Tandem Repeats (VNTR) method, which overcame abovementioned limitations.
A cluster where two or more TB cases shared a same genotype of TB bacilli was interpreted as being linked chains of recent transmission of TB disease, and unique isolates as being cases of reactivation of TB disease resulting from TB infection in the past. Therefore, molecular epidemiological analysis has been used to confirm suspected TB outbreaks and to detect possible TB transmission routes in TB contact investigation.

Although this technique has largely contributed to the TB control, there are several limitations. This method cannot be applied for bacteriologically negative TB cases which equivalent to 14% of all notified pulmonary TB cases in 2016, and latent tuberculosis infection cases which are 7,477 cases reported in 2016, because it requires DNA of TB bacilli. Moreover, the results of genotyping could not be timely feedback to the public health center or hospitals because it takes time to obtain the results. Thus, contact investigation could be supported by the methodology which does not depend on the bacteriological examination, such as GIS technology which relies on the geographical location attributed to the TB cases.
1-3. Geographic Information System for understanding TB transmission

Numerous studies using GIS technology have examined spatial distribution of TB patients\(^8\), socio-economic risk factors\(^9-11\), demographic risk factors\(^12-15\), and TB transmission dynamics in high-incidence areas\(^16\). In order to assess TB epidemiological situation nationwide, some studies used large data-set such as surveillance data\(^17\) and National Health Insurance data related with TB patients\(^18\). GIS has also been used in combination with molecular epidemiological techniques for understand the dynamics of TB transmission\(^3,19,20\). Haase et. al., identified previously unrecognized high-incidence zones in city of Montreal, Canada, and detected ongoing neighborhood transmission by the combination of spatial and genotyping information\(^19\).

Most of the studies which have attempted to reveal where transmission was ongoing using spatial analysis were, however, limited, in the sense that spatial information was based on patients' residential address\(^19\). TB transmissions occurring outside the households has been indicated\(^21\), but no studies have so far attempted to address this issue. In Japan, too, previous non GIS studies have also shown that TB transmissions occur in public facilities used by an unspecified number of people, such as saunas, bars, amusement facilities\(^22\) and areas near
downtown railroad stations.

1-4. Epidemiology of Tuberculosis in Japan

TB was a top leading cause of death until the 1950s, with a mortality rate of 146.4 per 100,000 population in 1950 in Japan. The rate has since dramatically declined to 1.5 in 2016, as a result of introduction of modern TB control measures and improvements in living standards. Japan is on the track towards becoming a low-burden; in 2016, the notification rate was 13.9 cases per 100,000 population. However, the reduction in the TB mortality and notification rate has stagnated in the recent years, primarily due to the existence of TB high-risk groups, such as elderly population developing active TB from long-standing latent infections, homeless people in urban settings, and foreign nationals who were born in TB-prevalent countries. Similar to other low and middle TB burden countries, active TB case finding including contact investigation and understanding of transmission dynamics have become critical in controlling TB in Japan. TB epidemiological studies using GIS, however, have not been fully conducted in Japan and few of those studies have same limitation as residential based approach.
1-5. **Objectives of this thesis**

The objective of our study, therefore, was to identify possible sites of transmission within urban setting using hotspot analysis, in combination with molecular epidemiological and GIS techniques. In order to overcome the limitation of residential based approach, we developed a unique concept of activity space to capture places where TB patients spend most of their waking hours, which may or may not be their residential homes. We further sought to compare spatial distribution patterns among different residential statuses and to describe patient characteristics and urban environmental features of the detected hotspots.
CHAPTER 2. METHODOLOGY

2-1. Study design and settings

2-1-1. Study design

This case-based cross-sectional analysis of notified culture-positive TB patients is part of a population-based DNA fingerprinting surveillance of \textit{M. tuberculosis} in Shinjuku city, Tokyo, which has been conducted since 2002\textsuperscript{27}.

2-1-2. Settings

Shinjuku city (18.3 km\textsuperscript{2}) is one of the most populous of the cities of Tokyo and, is a home to the Tokyo Metropolitan Government office (Fig. 1). A total of 152 census tracts of Shinjuku city consists of diverse urban environments with residential, commercial, and industrial zones. The population of Shinjuku city is 327,712 in 2015, of which 36,016 (11\%) are foreign residents\textsuperscript{28} [28]. Shinjuku city was ranked fourth among all cities in Japan in terms of high proportion of foreign nationals in 2010\textsuperscript{29}. TB incidence in Shinjuku city was 48.2 per 100,000 population in 2012, which was higher than Tokyo and national average in the same year (21.7 and 16.7, respectively).
Fig 1. Shinjuku city. Location of Shinjuku city and daytime population density per square kilometer. Note: segments in the city map are the 152 census tracts. Data source: 2005 Japanese Census, the Statistics Bureau, Ministry of Internal Affairs of Japan.
2-2. Subjects and materials

2-2-1. Case data

We included culture-positive TB patients who resided in Shinjuku city and were reported to the Shinjuku Public Health Center (PHC) between 2003 and 2011 as eligible subjects. Demographic and clinical information, including sex, age, occupation, categories regarding place of birth (born in Japan or foreign born), residential status (domiciled or homeless – in other words, those whose legal address was unknown or unstable in the last two years), and laboratory test results, were collected during the routine patient interview, which is usually conducted within 3 days of diagnosis, by public health nurses at the Shinjuku PHC.

At the same time, additional information regarding “activity space” was collected where a patient was asked to list a maximum of four places where he or she spends most of waking hours in the past two years. The place where the patient spent the longest waking hours was determined as his or her “activity space”, for the purpose of this study. This information was also collected together with the routine interview in order to ensure data validity on activity spaces for estimating possible transmission place to the extent possible.

TB strains isolated from each enrolled patient were subjected to the DNA
fingerprinting using IS6110-based restriction fragment length polymorphism (IS6110-RFLP) analysis\textsuperscript{30}. RFLP and spoligotyping were the standard techniques in DNA fingerprinting method in Shinjuku PHC and were available throughout the study period from 2003 to 2011. A genotype cluster was defined as \( \geq 2 \) isolates, either 1) with \( \geq 6 \) IS6110 bands with identical band patterns or 2) with <6 IS6110 bands with both identical IS6110 band patterns and spoligotyping patterns. We used criteria of at least 2 isolates identical for ensuring maximum sensitivity for detecting genotype clustering, which is a common practice for TB molecule epidemiological study conducted by Japanese PHCs. All eligible 1,021 culture-positive TB patients in a lump was used for genotype cluster analysis and was classified as genotype clustering or non-clustering. Percentages of genotype clustering were calculated as the number of clustered isolated strains divided by number of all isolated strains from subjects who met the inclusion criteria to the current study.

2-2-2. Census tract data

Demographic and urban environmental variables were derived from the 2005 Japanese Census, provided by the Statistics Bureau, Ministry of Internal Affairs of
Japan\textsuperscript{31}. The 2005 data was chosen as it represented approximately mid-point of the study period. It was used to calculate the variables for each census tract (CT) in Shinjuku city, as shown in Table 1.
### Table 1. The demographic and urban environmental variables of census tracts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Daytime population density</td>
<td>Day time population/km² of CT</td>
</tr>
<tr>
<td>Density of foreign population</td>
<td>Foreign population/km² of CT</td>
</tr>
<tr>
<td>Day and night population ratio</td>
<td>Daytime population/resident population</td>
</tr>
<tr>
<td>Percentage of population aged ≥65</td>
<td>Number of persons aged ≥65/resident population ×100</td>
</tr>
<tr>
<td>Household density</td>
<td>Number of households/km² of CT</td>
</tr>
<tr>
<td><strong>Socioeconomic status</strong></td>
<td></td>
</tr>
<tr>
<td>Total floor area per capita</td>
<td>Total floor area (m²)/number of residents in the household</td>
</tr>
<tr>
<td>Percentage of owned house</td>
<td>Number of households with owner-occupants/total number of household×100</td>
</tr>
<tr>
<td>Percentage of non-working households</td>
<td>Number of household with only non-working resident/total number of household×100</td>
</tr>
<tr>
<td>Distance to the nearest railway station</td>
<td>Distance of meter from a CT centroid to the nearest railway station</td>
</tr>
</tbody>
</table>

CT = census tract, *Centroid is the arithmetic mean position of all points in a two-dimensional area
2-3. Analysis methods

All analyses were based on the CT unit. For each CT, we counted the number of cases who nominated the CT as his or her activity space. We attempted to describe the aggregation and dispersion of TB patients’ activity space and identify their tendency by genotype cluster and residential statuses. Thus the annual TB patient density (the number of TB patients per year, per km²) was calculated for each CT.

We conducted three step analyses which were spatial analysis to describe overall spatial aggregation and dispersion, hotspot analysis to detect significant hotspots using spatial statistical method, and assessment of patient characteristics and environmental features of detected hotspots.

2-3-1. Spatial analysis

We assessed the level of spatial aggregation of TB patients in the whole study area by examining a spatial autocorrelation using Global Moran’s I statistics with row-standardized inverse distance weights matrices\(^{32}\). The method measures whether the values of neighboring areas are similar to one another. Thus, significant positive spatial autocorrelation implies that the distribution of TB patients is more spatially aggregation than a random underlying spatial process.
The patient distributions were observed using Inverse Distance Weighted (IDW) interpolation with annual TB patient density. IDW is an interpolation method, and it creates TB patient density surface. In the process, a neighborhood around the interpolated point is identified, and a weighted average is taken of the observation values within this neighborhood. The weight is a function of inverse distance.

2.3.2. Hotspot analysis

We examined whether patients would form statistically significant local aggregation of activity spaces. CT-level hotspots and coldspots were detected by Getis-Ords Gi* statistics, a spatial statistical method available with optimized hotspot analysis function in ArcGIS (Ver.10.2, ESRI Inc., CA, USA). Gi* statistics are the ratios of the local sum of the values in the vicinity of a distance, in other words the scale of the analysis, to the sum of all values. When the local sum is different than the expected local sum, and that difference is too large to be the result of random chance, it results in a statistically significant Z-score result.

Detected significant hotspots/coldspots imply a high/low value and are surrounded by other features with high/low values. Three confidence interval (CI)
levels (90%, 95%, and 99%) are used, with higher confidence levels implying more presumable aggregation of hotspots or coldspots. Detected CTs with more than 90% CI are considered hotspots or coldspots. The scale of analysis was determined through Incremental Spatial Autocorrelation (ISA). This performs the Global Moran's I statistic for a series of increasing distances, measuring intensification of clustering for each distance. The scale of analysis was specified as the distance at which the ISA generated a first Z-score peak, an indicator of pronounced spatial aggregation. The initial distance of ISA was 650 meters, which is a distance at which all features have at least one neighbor. An issue of multiple testing was corrected using False Discovery Rate control, which estimates the expected proportion of incorrectly rejected null hypotheses and adjusts the p-value accordingly.

2.3.3. Patient characteristics and urban environmental comparison

We compared hotspot analysis results for all patients, patient characteristics within hotspots and outside of hotspots, and demographic and urban environmental variables within hotspots and coldspots. Fisher's exact test and Chi-square test were used for frequency comparison. Unpaired t-test and Wilcoxon-
Mann-Whitney test were used for comparing two populations. A p-value of \( \leq 0.05 \) was considered significant.

The spatial analysis was conducted with ArcGIS (Environmental Systems Research Institute, Redlands, CA, USA) and the statistical analysis used STATA (Ver.12.0, StataCorp, TX, USA).

2-4. Ethics statement

Clinical, geographic and tuberculosis strain information are collected as part of the routine work of infectious disease control in Shinjuku PHC in accordance with field epidemiological investigation provided in the Infectious Disease Law Article 15, which, in principle, does not require informed consent from the patients. According to the Ethical Guidelines for Epidemiological Research established by Ministry of Education, Culture, Sports, Science and Technology and Ministry of Health, Labour and Welfare, informed consent is also not necessary when using pre-existing data for research purpose. Oral informed consent, however, was obtained from all of the culture-positive TB patients after the PHC staff gave a thorough explanation of study objectives and confidentiality at the interviews. The approvals given by the participants were recorded on the TB patient card. The study protocol including the
consent procedure was approved by the Institutional Review Board in Research Institute of Tuberculosis (RIT/IRB 25·26).
CHAPTER 3. FINDINGS

3-1. Characteristics of TB cases

During the study period from 2003 to 2011, 1,021 culture-positive TB patients who resided in Shinjuku city and were reported to the Shinjuku PHC were identified eligible subjects. Among all eligible 1,021 patients, 378 were excluded from present study due to the following reasons: no activity space within Shinjuku city (372 records), and registered multiple times (duplication records) due to suspected TB re-activation (6 records). In the latter case, we included the information of their first TB episode and excluded the following episodes. In the end, 643 records of TB patients were analyzed (Table 2). Comparing all eligible 1,021 patients and 643 enrolled patients, there were no significant differences with regard to sex, age, and residential status (data not shown).
Table 2. Demographic and genotyping features of enrolled culture-positive tuberculosis patients registered at the Shinjuku Public Health Center, Tokyo, 2003-2011

<table>
<thead>
<tr>
<th></th>
<th>All enrolled participants (n = 643)</th>
<th>General inhabitants (n = 413)</th>
<th>Foreign-born people (n = 42)</th>
<th>Homeless people (n = 184)</th>
<th>Unknown (n = 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male</td>
<td>491 (76.4)</td>
<td>281 (68.0)</td>
<td>23 (54.8)</td>
<td>184 (100)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>141 (21.9)</td>
<td>122 (29.5)</td>
<td>19 (45.2)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>11 (1.7)</td>
<td>10 (2.4)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Age, years</td>
<td>≤19</td>
<td>15 (2.3)</td>
<td>9 (2.2)</td>
<td>4 (9.5)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td></td>
<td>20-39</td>
<td>119 (18.5)</td>
<td>82 (19.9)</td>
<td>23 (54.8)</td>
<td>14 (7.6)</td>
</tr>
<tr>
<td></td>
<td>40-59</td>
<td>177 (27.5)</td>
<td>83 (20.1)</td>
<td>12 (28.6)</td>
<td>82 (44.6)</td>
</tr>
<tr>
<td></td>
<td>≥60</td>
<td>332 (51.6)</td>
<td>239 (57.9)</td>
<td>3 (7.1)</td>
<td>88 (47.8)</td>
</tr>
<tr>
<td>Genotype cluster&lt;sup&gt;1&lt;/sup&gt;</td>
<td>281 (43.7)</td>
<td>165 (40.0)</td>
<td>7 (16.7)</td>
<td>108 (58.7)</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>1</sup> A genotype cluster was defined as ≥2 isolates, either with 1) with ≥6 IS6110 bands with identical band patterns or with 2) with <6 IS6110 bands with both identical IS6110 band patterns and spoligotyping patterns.
In terms of place of birth and residential statuses, all subjects were classified into the following three categories: Japanese-born residents (general inhabitants), foreign-born residents and Japanese-born homeless people – there were no foreign born homeless people in our study and thus the three classifications were mutually exclusive. Of the enrolled 643 culture positive TB patients, 413 (64.2%) were general inhabitants, 42 (6.5%) were foreign-born, 184 were homeless people (28.6%), and four had an unknown residential status. Over the three quarters of patients were males (76.4%); no female homeless patient participated. While number of patients increased with advancing age among general inhabitants and homeless people, 64.3% of foreign-born people were 39 years old or less. In general, 281 (43.7%) of 643 isolates matched DNA band patterns of at least one other patient isolate. Interestingly, the percentage of genotype clustering among foreign-born people (16.7%) was far lower than among other residential statuses (40.0% of general inhabitants and 58.7% of homeless people). Activity spaces were generally workplaces for businesspersons, households for retired elderly persons, schools for students, and cardboard city or low-cost lodging for homeless people.
3-2. Spatial analysis

The average annual TB patient density was 3.3 per km². Among all 152 CTs, the median value was 2.3. Global Moran’s I analysis for all patients detected spatial autocorrelation with \( p < 0.01 \). In other words, the distribution of TB patients in the study area was significantly more spatially clustered than would be expected in a random process (Table 3). Every residential status was spatially auto-correlated to genotype clustering (p-values ranged from <0.01 to 0.04). However, the non-clustered groups of general inhabitants and foreign-born people were not auto-correlated (Table 3).
### Table 3. Global autocorrelation analysis using Moran’s I statistics by residential and genotype cluster statuses.

<table>
<thead>
<tr>
<th>Residential status</th>
<th>Genotype cluster status</th>
<th>Moran’s I</th>
<th>Z-score</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All patients</td>
<td></td>
<td>0.308</td>
<td>8.040</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>General inhabitants</td>
<td></td>
<td>0.152</td>
<td>3.667</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Clustered</td>
<td>0.083</td>
<td>2.061</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Non-clustered</td>
<td>0.048</td>
<td>1.305</td>
<td>0.19</td>
</tr>
<tr>
<td>Foreign-born people</td>
<td></td>
<td>0.129</td>
<td>3.299</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Clustered</td>
<td>0.191</td>
<td>4.731</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Non-clustered</td>
<td>-0.021</td>
<td>-0.730</td>
<td>0.47</td>
</tr>
<tr>
<td>Homeless people</td>
<td></td>
<td>0.347</td>
<td>9.512</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Clustered</td>
<td>0.287</td>
<td>9.122</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Non-clustered</td>
<td>0.369</td>
<td>9.385</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>
An IDW map identified a patient density peak (41.7/km²/year) near the Shinjuku railway station, which is one of the busiest railway transport hubs in Japan. Lower patient density areas were observed on the east and north sides of the city, both of which are residential zones (Fig. 2). Various trends depended on different residential and genotype-cluster statuses were observed (Fig. 3). Among the general inhabitants, a number of scattering peaks were found. The genotype-clustered group peaked slightly near the Shinjuku station, but few patients defused in the non-clustered group. Foreign-born people concentrated on the west side of the city where expat communities exist. However, as the number of patients was small, the peak (3.2/km²/year) was not as obvious as other residential groups. Homeless patients were aggregated around the station, regardless of genotype statuses.
Fig 2. Inverse Distance Weighted interpolated map with annual number of all tuberculosis patients per square kilometer. Note: the color scale indicates the annual number of tuberculosis patients per square kilometer.
Fig 3. Inverse Distance Weighted interpolated maps with annual number of tuberculosis patients per square kilometer, stratified by residential status and genotype cluster status. Note: (A) General inhabitants. (B) Genotype-clustered general inhabitants. (C) Genotype non-clustered general inhabitants. (D) Foreign-born people. (E) Genotype-clustered foreign-born people. (F) Genotype non-clustered foreign-born people. (G) Homeless people. (H) Genotype-clustered homeless people. (I) Genotype non-clustered homeless people. Note: the color scale indicates the annual number of tuberculosis patients per square kilometer.
### 3-3. Hotspot Analysis

The overall results showed that 13.7% of the study area (2.5 km²) was detected as hotspots and 6.3% as coldspots (1.1 km²) (Fig. 4). Patients from hotspots tended to be male, < 65 years old, homeless, unemployed, smear-positive, and genotype-clustered (Table 4). Compared to patients in outside of hotspots, they were diagnosed through contact investigations, welfare facility visits, or brought to the emergency department by ambulance.
Fig 4. Hotspots and coldspots for all tuberculosis patients detected by Getis-Ords

$G_i^*$, Note: Three confidence levels (90%, 95%, and 99%) are provided; higher confidence levels imply stronger aggregation in either hotspots or coldspots.
Table 4. Comparison of characteristics of tuberculosis patients within hotspots and outside of hotspots of all study population including both genotype clustered and genotype non-clustered.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Patients in hotspots (n = 240)</th>
<th>Patients outside of hotspots (n = 403)</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male (excluded unknown=11)</td>
<td>214 (89)</td>
<td>277 (70)</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Age, years, mean (SD)</td>
<td>54 (16)</td>
<td>60 (20)</td>
<td>&lt;0.01†</td>
</tr>
<tr>
<td>Age ( \geq 65 ) years</td>
<td>63 (26)</td>
<td>194 (48)</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Residential status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General inhabitant</td>
<td>93 (39)</td>
<td>320 (79)</td>
<td>&lt;0.01‡</td>
</tr>
<tr>
<td>Foreign-born people</td>
<td>16 (7)</td>
<td>26 (6)</td>
<td></td>
</tr>
<tr>
<td>Homeless population</td>
<td>127 (53)</td>
<td>57 (14)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>4 (2)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Permanent worker</td>
<td>45 (19)</td>
<td>102 (25)</td>
<td>0.01†</td>
</tr>
<tr>
<td>Student</td>
<td>5 (2)</td>
<td>15 (4)</td>
<td></td>
</tr>
<tr>
<td>Temporary worker</td>
<td>35 (15)</td>
<td>31 (8)</td>
<td></td>
</tr>
<tr>
<td>Housewife</td>
<td>7 (3)</td>
<td>35 (9)</td>
<td></td>
</tr>
<tr>
<td>Infant</td>
<td>0 (0)</td>
<td>1 (0)</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>147 (61)</td>
<td>217 (54)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>1 (0)</td>
<td>2 (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Unemployed who were ≥65 years old</td>
<td>55 (23)</td>
<td>147 (36)</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Smear-positive tuberculosis</td>
<td>162 (68)</td>
<td>217 (54)</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Genotype-clustered</td>
<td>122 (51)</td>
<td>159 (39)</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Diagnosis method</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual health check</td>
<td>3 (1)</td>
<td>15 (4)</td>
<td>&lt;0.01‡</td>
</tr>
<tr>
<td>Periodic health check</td>
<td>13 (5)</td>
<td>42 (10)</td>
<td></td>
</tr>
<tr>
<td>Contact investigation</td>
<td>11 (5)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Other mass health check</td>
<td>3 (1)</td>
<td>9 (2)</td>
<td></td>
</tr>
<tr>
<td>Health care facility visit</td>
<td>118 (49)</td>
<td>242 (60)</td>
<td></td>
</tr>
<tr>
<td>Welfare facility visit, Emergency</td>
<td>73 (30)</td>
<td>35 (9)</td>
<td></td>
</tr>
<tr>
<td>transportation, and others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health services other than TB¹</td>
<td>14 (6)</td>
<td>49 (12)</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>5 (2)</td>
<td>11 (3)</td>
<td></td>
</tr>
</tbody>
</table>

*Fisher's exact test, †Unpaired t test, ‡Chi square test, SD = standard deviation,

¹Diagnosed as tuberculosis though inpatient or outpatient services for disease other than tuberculosis
We attempted to identify the urban environmental features of hotspots and coldspots, using nine variables (Table 5). The median of all population-related variables (except the percentage of 65-year-old and over residents) was higher in hotspots than in coldspots. In particular, hotspots were characterized by significantly higher daytime population density ($p = 0.01$) and density of foreign-born people ($p = 0.02$) than the coldspots. Households in hotspots had median 28.2 m² floor area per capita, which was significantly smaller than household in coldspots (34.4 m²). The distance to the nearest railway station was shorter in hotspots (median = 359.8 m) than in coldspots (median = 549.4 m), but the difference was not statistically significant ($p = 0.10$). Percentage of non-working household (household in which no one held paying employment) was significantly higher in hotspots (56.9%) than in coldspots (35.1%).
Table 5. Comparison of urban environmental factors in hotspot and coldspot census tracts for all tuberculosis patients.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total CTs = 152</th>
<th>Hotspots CTs = 13</th>
<th>Coldspots CTs = 36</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median</td>
<td>IQR</td>
<td>median</td>
</tr>
<tr>
<td><strong>Population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daytime population density (/km²)</td>
<td>27,675.0</td>
<td>(28,794.9)</td>
<td>63,589.5</td>
</tr>
<tr>
<td>Density of foreign population (/km²)</td>
<td>950.0</td>
<td>(1,344.1)</td>
<td>3,406.3</td>
</tr>
<tr>
<td>Day and night population ratio</td>
<td>149.0</td>
<td>(285.0)</td>
<td>445.0</td>
</tr>
<tr>
<td>Percentage of population aged 65 and over (%)</td>
<td>18.9</td>
<td>(5.0)</td>
<td>18.7</td>
</tr>
<tr>
<td><strong>Household Census</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household density (/km²)</td>
<td>10,352.2</td>
<td>(6,899.0)</td>
<td>10,128.6</td>
</tr>
<tr>
<td>Total floor area per capita (m²)</td>
<td>30.5</td>
<td>(6.0)</td>
<td>28.2</td>
</tr>
<tr>
<td>Percentage of owned house (%)</td>
<td>38.7</td>
<td>(15.6)</td>
<td>37.0</td>
</tr>
<tr>
<td>Percentage of non-working households (%)</td>
<td>40.0</td>
<td>(12.4)</td>
<td>56.9</td>
</tr>
<tr>
<td>Distance to the nearest railway station from a CT centroid (m)</td>
<td>669.3</td>
<td>(539.2)</td>
<td>359.8</td>
</tr>
</tbody>
</table>

CTs = Census tracts, IQR = Inter-quartile range, *Hotspots and coldspots were detected by Getis-Ord Gi* statistics, which implies that detected hotspots have high patient density and are surrounded by other features with high patient density.
In the final part, Fig. 5 showed hotspots and coldspots identified among different residential and genotype-cluster statuses. General inhabitants indicated similar patterns of hotspots and coldspots to the results of the overall group. Genotype-clustered general inhabitants formed a hotspot with a low confidence level (CI=90%), whereas the non-clustered general inhabitants indicated almost no hotspots. Foreign-born people had obvious hotspots, but consisted of a few patients. We could not analyze by genotype-cluster status because of small sample size. Homeless people had significant hotspots around the station (CI=99%), regardless of genotype cluster status.
Fig 5. Hotspots and coldspots for tuberculosis patients detected by Getis-Ords Gi*, stratified by residential statuses and genotype-cluster status. (A) General inhabitants. (B) Genotype clustered general inhabitants. (C) Genotype non-clustered general inhabitants. (D) Foreign-born people. (E) Homeless people. (F) Genotype clustered homeless people. (G) Genotype non-clustered homeless people.

Note: Three confidence levels (90%, 95%, and 99%) are provided; higher confidence levels imply stronger aggregation in either hotspots or coldspots.
CHAPTER 4. DISCUSSIONS AND CONCLUSION

4-1. The identified TB hotspots using activity spaces based spatial approach

4-1-1. Identified TB hotspots

We have hypothesized that transmission occurs in “activity space”, which may or may not be their residential homes. Our results showed that area surrounding Shinjuku station is significant hotspots for all patients regardless of genotype cluster and residential statuses. Patients from the hotspots tended to be young and middle age males who were unemployed or homeless. Similar area was identified as hotspots for both genotype clustered general inhabitants and homeless people, which indicate that recent transmission probably took place around Shinjuku station.

The identified hotspots could be characterized by distinct urban features, including large daytime populations, residence of a large proportion of foreign-born people, and proximity to the nearest railway station indicating large number of business and commercial facilities. At the same time, several indicators suggested low socioeconomic status of the hotspots, including smaller residence floor size, a higher proportion of non-working households, and, though not statistically
significant, lower rate of homeownership.

4.1.2. Possible hotbeds of TB transmission at the facilities used by unspecified numbers of people

Past studies in Japan have shown that TB transmission occur in public facilities and near downtown railroad station\(^{22,23}\), and also that homeless and non-homeless people can share the same genotype cluster in an urban setting\(^9\). Patient interviews by the PHC staff have also revealed that many patients had used facilities such as saunas, bars, and pachinko halls, which have previously also been reported as TB outbreak venues\(^{22}\). The Tokyo Metropolitan Government published the Tuberculosis Control and Prevention Plan and pointed out that 89.2% of TB infections from 20's to 50's, while of unclear origin, were found in those who stayed a long time in the business facilities mentioned above\(^{38}\). In other high-incidence countries, epidemiological links between patients have been reported based on facilities used by unspecified numbers of people\(^{17}\). In our study, significant hotspots were found in the commercial area nearby the large station where there are many of such facilities.
4·1·3. Hotspots of non-clustered homeless patients

Area around Shinjuku station was also identified as a hotspot by the non-clustered homeless patients – this can partially be explained by the fact that Shinjuku city is one of the largest homes to homeless people, attracting people living under fragile conditions from around the country. In other words, some homeless people might have been infected prior to coming to Shinjuku, and hence do not belong to the same genotype cluster, but share similar activity spaces with the clustered homeless people, once they arrive in Shinjuku.

4·1·4. Limited impact of TB transmission among elderly residence and foreign-born people

Lastly, the non-clustered general inhabitants spread widely over the city, indicating no hotspots, as shown in Fig 3C. These areas consist largely of residential zone where retired elderly nominated as their activity spaces. In our study, 57% of genotype non-clustered patients were 60 years old and higher. This implies that TB developed from past (latent) infections and ongoing transmission of TB in the area was unlikely. TB transmission among foreign-born population may be limited scale because the percentage of genotype clustering among foreign-born people was still
low at 16.7%.

4-2. Limitations of the study

There are several limitations to our study. Firstly, we could only collect information regarding activity space close to the time of diagnosis, and not around the time of infection. Hotspot may, therefore, not necessarily represent a site of transmission. In order to partially overcome this issue, we conducted the analysis by genotype cluster statuses. Secondly, by utilizing the concept of one activity space per patient, we are assuming that TB transmission is most likely to occur in a place where the patient spends “most of his or her waking hours”. That is to say, our study does not take into consideration the possibility of transmissions occurring through casual contacts with the TB patient with minimal duration of contact, such as in a crowded elevator, which has been indicated as risk of transmission\(^{39}\), and in public transport. Thirdly, the study population only included those who were registered with the Shinjuku PHC – however, considering the city’s large daytime population, we have potentially missed a large number of TB patients whose activity space is located in Shinjuku city but who reside elsewhere. A further research is therefore necessary to incorporate larger area, at least including the Greater Tokyo Area.
Lastly, discussion of risk of TB transmission, such as incidence rate, in the hotspots was out of scope of the study, because appropriate population at risk based on activity spaces was not available.

4-3. Programmatic implication of the findings for the TB control

Despite these limitations, our study used a novel concept of activity space to attempt to identify possible places of TB transmission, which could take into account both inside and outside household.

The programmatic implications of our findings for the TB control are as follows:

- Recent transmission may possibly take place around large train station that was identified as hotspots for both genotype clustered general inhabitants and homeless people.
- TB transmission at the facilities used by unspecified numbers of people is a key subject needed to be investigated to identify unidentified TB transmission routes.
- TB transmission among elderly residence and foreign-born people may be limited scale. However, as increasing number of foreign-born TB cases in recent years, TB transmission between foreign-born and Japanese-born people
should be monitored.

Molecular epidemiological techniques have certainly made much contribution to understanding the dynamics of TB transmission, however, it is also true that epidemiological links are uncertain among some of the cases even identified as same genotype clustered group\(^3\). We have shown that activity space-based spatial analysis provided an alternative approach to residential address-based approaches, and that it can be used in combination with molecular technique in further comprehension of TB transmission dynamics in an urban setting in Japan.

4.4. Conclusion

This study indicated possible area of TB transmission in inner metropolitan setting, Shinjuku city, Tokyo, based on activity spaces using spatial analysis combined with molecular epidemiology, which were characterized by relatively lower some socio-economic factors. The most difficult but crucial question for breaking the chain of transmission is where ongoing TB infection occurs. Despite accompanied by several methodological limitations, activity space-based spatial analysis provided a further insight of TB transmission dynamics in an urban
setting in Japan.
REFERENCE


5) Murray M, Nardell E. Molecular epidemiology of tuberculosis: Achievements and challenges to current knowledge. *Bull. World*


12) Barr RG, Diez-Roux AV, Knirsch CA, Pablos-Méndez A.


17) Feske ML, Teeter LD, Musser JM, Graviss EA. Including the third dimension: a spatial analysis of TB cases in Houston Harris County.


22) Tamaru A. Tuberculosis measurement for public facilities used by an unspecified number of people [Japanese]. *Japan Public Heal. Pract.*, 50


PUBLISHED PAPERS ASSOCIATED WITH THE DOCTORAL THESIS

Kiyohiko Izumi, Akihiro Ohkado, Kazuhiro Uchimura, Yoshiro Murase, Yuriko Tatsumi, Aya Kayebeta, Yu Watanabe, Nobukatsu Ishikawa

Detection of Tuberculosis Infection Hotspots Using Activity Spaces Based Spatial Approach in an Urban Tokyo, from 2003 to 2011., Plos ONE, 10(9): e0138831. 2015
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to the following people for their help to my doctoral work.

I am deeply indebted to my supervisor, Dr. Akihiro Ohkado and Dr. Satoshi Mitarai for their continuous support and incomparable patience. Without their detailed instruction and kindness, I would not have been able to successfully complete my work.

I express my thanks to Mr. Kazuhiro Uchimura, Mr. Yoshiro Murase, Dr. Yuriko Tatsumi, Dr. Aya Kayebeta, Dr. Yu Watanabe, and Dr. Nobukatsu Ishikawa for their kind support to the data collection, analysis and manuscript development of this paper as the co-authors. I am grateful to my colleague, Ms. Lisa Kawatsu in the Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association for careful and detailed proof reading. I also thank the Shinjuku Public Health Center staff for data collection.

I also appreciate the financial support of the Research Program on
Emerging and Re-emerging Infectious Diseases from Japan Agency for Medical Research and development, AMED, Government of Japan, and double-barred cross seal donation fund, Japan Anti-Tuberculosis Association.