BMJ Open

Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in Japan: a population-based molecular epidemiological study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029295
Article Type:	Research
Date Submitted by the Author:	24-Jan-2019
Complete List of Authors:	Izumi, Kiyohiko; The Research Institute of Tuberculosis, Murase, Yoshiro; The Research Institute of Tuberculosis Uchimura, Kazuhiro; The Research Institute of Tuberculosis Kaebeta, Aya; Shinjuku Public Health Center Ishihara, Keiko; Shinjuku Public Health Center Kaguraoka, Sumi; Shinjuku Public Health Center Takii, Takemasa; The Research Institute of Tuberculosis Ohkado, Akihiro; The Research Institute of Tuberculosis
Keywords:	RFLP, clustering rate, homeless, foreign-born

Title

Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in Japan: a population-based molecular epidemiological study

Authors and affiliations

1. Kiyohiko Izumi. Department of Epidemiology and Clinical Research, The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: kizumi95@gmail.com
2. Yoshiro Murase. Department of Mycobacterium Reference and Research, The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: ymurase@jata.or.jp
3. Kazuhiro Uchimura. Department of Epidemiology and Clinical Research, The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: uchimura@jata.or.jp
4. Aya Kaebeta. Shinjuku Public Health Center. Email: aya.kaebeta@city.shinjuku.lg.jp
5. Keiko Ishihara. Shinjuku Public Health Center. Email: keiko.ishihara@city.shinjuku.lg.jp
6. Sumi Kaguraoka. Shinjuku Public Health Center. Email: sumi.kaguraoka@city.shinjuku.lg.jp
7. Takemasa Takii. The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: t-takii@jata.or.jp
8. Akihiro Ohkado. Department of Epidemiology and Clinical Research, The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Graduate School of Biomedical Sciences, Nagasaki University. Email: ohkadoa@jata.or.jp

Corresponding Author

Kiyohiko Izumi

Department of Epidemiology and Clinical Research, Research Institute of Tuberculosis, Japan Anti-

Tuberculosis Association, 3-1-24 Matsuyama, Kiyose, Tokyo 204-8533, JAPAN

Tel:+81-42-4935711

Fax: +81-42-492-4600

Email: kizumi95@gmail.com

KEY WORDS: RFLP, clustering rate, homeless, foreign-born

ABSTRACT

Objective: Molecular epidemiology study is a promising tool to understand tuberculosis transmission dynamics but has not been sufficiently conducted in Asian countries. The aim of this study was to estimate the proportion of TB cases attributable to recent transmission and to identify the risk factors of genotype clustering and the development of large clusters within three years in urban setting in Japan.

Design and setting: Long-term cross-sectional observational study combining the characteristics of culture-positive TB patients notified to Shinjuku City, Tokyo (2002-2013) with the genotype data of *Mycobacterium tuberculosis.*

Primary outcome measure: Genotype clustering rate and association between genotype clustering status and explanatory variables.

Results: Among 1,025 cases, 515 (50.2%) were in 113 genotype clusters. The overall clustering rate was 39.2%. The rate was significantly higher in patients aged <40 years (adjusted odds ratio (aOR) =1.73, 95%CI=1.23-2.44) , native Japanese (aOR=3.90, 95%CI=2.27-6.72), fulltime worker (aOR=1.63, 95%CI=1.17-2.27) and part-time/daily worker (aOR=2.20, 95%CI=1.35-3.58), those receiving public assistance (aOR=1.81, 95%CI=1.15-2.84), and the homeless (aOR=1.63, 95%CI=1.02-2.62). A significant predictor of large genotype clusters within three years was a registration interval ≤2 months between the first two cases in a cluster.

Conclusion: Our results indicated that a large proportion of culture-positive TB patients were involved in the recent TB transmission chain. Foreign-born persons still have a limited impact on transmission in the Japanese urban setting. Intensified public health interventions including active case finding need to focus on those with socioeconomic risk factors that are significantly associated with tuberculosis transmission and clusters with shorter registration intervals between the first two cases.

Word Count in Abstract= 255 words

Article Summary

Strengths and limitations of this study

This study is a one of the longest population-based studies on the molecular epidemiology for culture-positive notified tuberculosis patients in an Asian large urban setting.

Interview conducted by the experienced public health nurses at the Public Health Center using standardized questionnaire provided high data quality and less interviewer bias.

We may have underestimated genotype clustering due to the large population flow in and out of the city.

Funding statement

This research was supported by the Research Program on Emerging and Re-emerging Infectious Diseases from the Japan Agency for Medical Research and Development, AMED No. JP18fk0108041.

Competing interests: None declared.

Patient consent: Not required.

Data sharing statement: Due to data restrictions, we are unable to share any aspect of the data.

Word count: 2,683 words

INTRODUCTION

Tuberculosis (TB) remains a major public health threat worldwide. An estimated 10 million people developed TB, and one million died from TB in 2016 worldwide[1]. Although the majority of cases have been reported in TB high-burden countries, TB remains a persistent health problem in low- and middle- burden countries, because it is concentrated in specific vulnerable and hard-to-reach populations, such as homeless people and foreign-born persons from TB high-burden countries[2]. These specific high-risk populations tend to live in large cities where they are seeking jobs, which potentially poses challenges to TB control in urban areas[3,4]. Many TB low- and middle- burden countries have recently adopted TB elimination strategies[2,5], which emphasize the importance of molecular epidemiology in TB control especially in urban areas[2,4].

TB molecular genotyping using restriction fragment length polymorphisms (RFLPs) and, recently, variable numbers of tandem repeats (VNTR) combined with epidemiological information identifies TB cases possibly involved in the same chain of transmission[6]. This method differentiates recent transmission or endogenous reactivation from remote infection, and therefore has revealed that a substantial proportion of TB cases are due to recent transmission in TB low-burden countries[7–9]. This method also identifies the proportion of cases attributable to recent transmission and determines the risk factors for transmission. Moreover, the factors predicting large TB genotype clusters have been investigated by evaluating the characteristics of the first two cases in the same genotype cluster[10–13]. These predictors included socially vulnerable populations and shorter intervals between the registration dates of the first two cases. However, these population-based molecular epidemiological studies were limited in some European countries and U.S. and have not been sufficiently conducted in Asian countries.

In Japan, a TB middle-burden country, newly notified TB cases have been decreasing from 32,828 (25.8 per 100,000 populations) in 2002 to 17,625 (13.9 per 100,000 populations) in 2016[14], while TB outbreaks have been constantly occurring at approximately 40 events annually over the last decade. This suggests that TB transmission may be occurring in some groups such as homeless people, who constitute a high-risk group for recent TB transmission in urban areas[15]. Considering the steady increase in the proportion of TB cases among foreign-born individuals in Japan (7.9% of all cases in 2016[16]), transmission between foreign-born persons and local residents must be monitored. In light of Japan's transition towards becoming a low-burden country, understanding TB transmission patterns has become increasingly important. However, few population-based molecular epidemiological studies have identified the transmission patterns and their risk factors. Additionally, no study has attempted to evaluate the factors predicting the development of large clusters in Japan.

Therefore, we aimed to estimate the proportion of TB cases attributable to recent transmission, to identify the risk factors for recent transmission, and to predict the risk factors for the development of large clusters in the urban setting.

METHODS

Study population

We included all culture-positive TB patients notified in Shinjuku City from September 2002 to December 2013 as eligible study population in this cross-sectional observational study. This study was part of a population-based study on DNA fingerprinting surveillance of *M. tuberculosis* in Shinjuku City started in 2002. Shinjuku City (18.3 km^2) is one of the most populous (342,867 residents in 2018[17]) cities in Tokyo and it's TB notification rate was 33.7 per 100,000 people in 2016[18], which was higher than the rates in Tokyo and the nation (17.2 and 13.9, respectively[14]). Experienced public

health nurses at the Shinjuku Public Health Center (PHC) interviewed and collected information from all culture-positive TB patients at the time of registration using standardized questionnaire to avoid possible interviewer bias. Study variables and definitions are described in Table 1.

Patient and public involvement

Patients or the public were not involved in the design of this study.

DNA fingerprinting and genotype cluster

Clinical isolates from each of the enrolled TB patients were sent to the Research Institute of Tuberculosis (RIT), Tokyo, where the TB strains were subjected to DNA fingerprinting using insertion sequence *6110* by RFLP (IS*6110*-RFLP) analysis[19]. IS*6110*-RFLP and spoligotyping were the standard methods used in the Shinjuku PHC and were available throughout the study period. The Shinjuku PHC switched from RFLP to VNTR a few years ago, but many TB cases had RFLP profiles. Thus, we employed RFLP due to the sufficient sample size. A genotype cluster was defined as ≥2 isolates with either (1) ≥6 IS*6110* bands with identical band patterns or (2) <6 IS*6110* bands with both identical IS*6110* band patterns and spoligotyping patterns. Details on the data collection and genotyping method were previously described[15].

Data analysis

We calculated the genotype clustering rate by the "n-1 method" according to the formula *{(n - c)/N}*, where *N* is the total number of cases sampled, *c* is the number of clusters, and *n* is the total number of cases in the clusters[9]. We also calculated the cumulative clustering rate by calculating the clustering rate in 2002 and then adding the TB patients every year up to 2013. The characteristics of clustered cases, which were the cases belonging to any genotype clusters, were compared with those

with unique strain patterns using χ^2 tests. We performed univariate logistic regression to identify risk factors for genotype clustering using odds ratios (ORs) and multivariate logistic regression including variables selected by the stepwise maximum-likelihood estimation with a significance level of less than 0.2 using adjusted ORs (aORs). Potential interactions were assessed using likelihood ratio tests.

Additionally, we compared the characteristics of the first two cases in each genotype cluster to predict the risk factors for the development of a large cluster within three years. For this purpose, a cluster episode was defined as newly arising genotype clusters in and after 2003 without any TB cases of that genotype notified prior to that year. We classified cluster episodes into two groups according to the system in a previous study[10]; (1) "large clusters within three years," were cluster episodes with five or more cases (large clusters) occurring within three years, and (2) "small clusters and large clusters after three years," were cluster episodes with two to four cases (small clusters) and cluster episodes that became large clusters after three years. We identified the first two cases in each cluster episode and compared their characteristics between these two groups. We performed univariate and multivariate logistic regression analyses to identify predictors of the development of large clusters within three years.

A *p*-value of 0.05 was set as the statistically significant level. If there are variables with missing values exceeding 5%, the multiple imputation method is considered. We used Stata version 12 (Stata Corp., College Station, TX, USA) for the statistical analysis. Written informed consent was waivered because DNA fingerprinting analysis was part of the routine TB control activities in Shinjuku City. However, oral informed consent was obtained after the PHC staff gave a thorough explanation of study objectives and confidentiality. The study protocol was approved by the Institutional Review Board of the Research Institute of Tuberculosis (RIT/IRB27-9).

RESULTS

Study population and clustering rate

In total, 1,885 TB patients were notified in Shinjuku City during the study period, and 1,310 were culture-positive cases (Figure 1). Of those, 285 patients were excluded from the analysis, mainly due to the unavailability of culture-positive isolates and lack of implementation of RFLP. Finally, 1,025 (78.2%) patients were included in the analysis. Figure 2 shows the cumulative number of TB patients and clustering rates from 2002 to 2013. The number gradually increased over a decade, while the cumulative clustering rates sharply increased in the first four years, from 10% in 2002 to 28% in 2005, with an average percent change of +43%; the clustering rates continues to increase but at a slower rate, from 30% in 2006 to 39% in 2013, with an average percent change of +4.2%.

We identified a total of 113 genotype clusters consisting of 515 (50.2%) patients (Figure 1), and the genotype clustering rate was 39.2%. The average cluster size was 4.56 cases (range 2-30). There were 57 (50.4%) genotype clusters consisting of only two TB patients, and 36 (31.9%) genotype clusters had five or more TB patients. We further investigated the status of homelessness and place of birth among genotype clustered cases. Of the 113 genotype clusters, 45 (39.8%) comprised only nonhomeless individuals, 7 (6.2%) only homeless individuals, and 61 (54.0%) both homeless and nonhomeless individuals (mixed cluster). We compared the characteristics of nonhomeless patients in clusters of only nonhomeless patients and those in mixed clusters. While the finding was not statistically significant (Pearson χ^2 test, p=0.17) proportion of nonhomeless patients receiving public assistance in the latter group (13.8%) was higher than those in the former group (8.8%). There were no differences between the two groups regarding sex, age, and place of birth. Of the 113 genotype clusters, 94 (83.2%) consisted of only individuals born in Japan, 2 (1.8%) consisted of only foreign-

born individuals, and 17 (15.0%) consisted of both.

Factors associated with genotype clustering

Clustered cases were significantly more likely to be male (OR=1.62), Japan-born (OR=3.74), receiving public assistance (OR=2.25), homelessness (OR=2.45), misusing alcohol (OR=1.37), and engaging in fulltime work (OR=1.53), part-time/daily work (OR=2.29), and Jobless aged 15-59-year-old (OR=2.05) (Table 2). A significant interaction among the explanatory variables was not identified. The multivariate analysis demonstrated that the factors associated with genotype clustering were age <40 years old (aOR=1.73, 95%CI=1.23-2.44), Japan-born (aOR=3.90, 95%CI=2.27-6.72) , engaging in fulltime work (aOR=1.63, 95%CI=1.17-2.27) and part-time/daily work (aOR=2.20, 95%CI=1.35-3.58), receiving public assistance (aOR=1.81, 95%CI=1.15-2.84), and homelessness (aOR=1.63, 95%CI=1.02-2.62)(Table 3).

We further investigated the temporal trends in the characteristics of genotype clustered cases. The number of all homeless cases and their proportion among all study cases decreased (Figure 3A and Figure 3B), while the proportion of genotype clustered cases among the homeless increased over time (Figure 3B). No typical trend was observed among those who were foreign-born, receiving public assistance, misusing alcohol, and occupation (data not shown).

Factors associated with large genotype clustering within three years

We identified 104 genotype cluster episodes according to the definition. Of these, 14 were “large clusters within three years”, which was equivalent to 13.5% (14/104) of all genotype clusters and 48.3% (14/29) of large genotype clusters, and 90 clusters were “small clusters and large clusters after three years”. The univariate analysis indicated that clusters with registration intervals of 0-2 months

were 9.51 times more likely to become large genotype clusters within three years compared with clusters with registration intervals of ≥12 months (Table 4). After selecting variables by the stepwise method, only the variable of registration interval remained for the multivariate model.

DISCUSSION

In this long-term population-based study, we included 1,025 patients and identified a total of 113 genotype clusters, and the genotype clustering rate was 39.2%. Our results indicated that clustered cases were more likely to have socioeconomic risk factors, namely, being homeless, receiving public assistance, and having an unstable job at tuberculosis diagnosis. A shorter registration interval between the first two cases was a statistically significant predictor of the development of a large genotype cluster within three years.

Clustering rate

We identified 515 (50.2%) genotype clustered cases and estimated a clustering rate of 39.2%. The rate was the same as the pooled clustering rate (40.9%) reported in the meta-analysis of population-based studies of TB low-incidence countries[20]. However, it was different from previous estimates in Japanese studies, which were 27.6% in Shinjuku and 24.6% in Osaka[15,21]. As the meta-regression analysis clarified that longer study durations increase the clustering rate[20], this difference could be due to shorter study durations together with smaller sample sizes in the previous studies (388 patients in five years and 195 patients in one year, respectively). In our study, the cumulative clustering rate rapidly increased in the first four years and increased more slowly thereafter, which is a similar trend to studies in the U.S. and Malawi[22,23].

Factors associated with genotype clustering

Our results indicated that clustered cases were more likely to have socioeconomic risk factors, namely, being homeless, receiving public assistance, and having an unstable job at TB diagnosis. Similarly, previous studies suggested that being homeless significantly contributed to clustering in Shinjuku City[15] and other counties[20]. In our study, more than half of the genotype clusters were mixtures of nonhomeless and homeless patients. Moreover, the nonhomeless patients in the mixed clusters tended to be financially unstable, with a higher proportion receiving public assistance than those among clusters of only nonhomeless cases, which could imply that relatively poor nonhomeless patients share activity spaces with homeless patients, such as urban areas around the large train stations that were reported as significant hotspots for homeless patients in Shinjuku City[24]. An analysis of the time trend of clusters involving the homeless showed that the proportion of clustered cases among the total homeless patients increased over the study period. These findings could suggest that contact investigations of homeless TB patients need to be actively expanded to possible contact persons who are nonhomeless, especially those who are in financial difficulty.

The meta-analysis based on studies in European countries where foreign-born patients substantially contribute to TB epidemiology reported the range of the proportion of mixed clusters composed of native and foreign-born patients (0% to 36.5%) and concluded that foreign-born patients did not have a significant influence on TB in the native population[25]. In our study, the proportion of mixed clusters (15.0%) fell into this range. Thus, the impact of TB transmission between native and foreign-born populations was probably still limited in this urban setting[26]. However, considering the recent increase in immigrant TB patients in urban cities, TB transmission between native and foreign-born populations needs to be closely monitored.

Factors associated with large genotype clustering within three years

A shorter registration interval (≤2 months) was a significant predictor of the development of a large genotype cluster within three years, which is compatible with findings of previous studies in the Netherlands and London[10,12]. Therefore, when TB patients with identical genotypes have shorter registration intervals, thorough active case findings need to be considered to investigate potential infection sources and infected patients to prevent further transmission. On the other hand, it is difficult to assume that the first patient infected the second patient because the window of two months seems too short. Thus, we think that there may be a true but unidentified first case that our study did not identify. A cluster episode was defined as a cluster without any TB patients in 2002 and at least two patients with identical genotypes in and after 2003. Therefore, a possible true first case may have been registered before 2002, which was outside of our study period, or registered outside of Shinjuku City.

Limitations

Our study has some limitations. First, the study population consisted only of TB patients living in Shinjuku City. Considering the large population flow in and out of the city, as we mentioned above, we potentially missed patients living outside of the city who shared TB strain types with patients living in the city. Actually, previous Japanese studies reported that there are clusters with TB patients living across broad geographic areas[27]. Consequently, we may have underestimated genotype clustering. Second, even TB patients with identical genotyping patterns may not suggest recent transmission if the strain is a nationwide endemic TB strain[28]. As a result, this could have led to an overestimated clustering rate. Lastly, IS*6110* RFLP has relatively lower discriminatory power compared with VNTR[29], which may have led to overestimation.

Conclusion

This study is a one of the longest-term studies on molecular epidemiology for notified tuberculosis patients in an Asian large urban setting. Our results indicated that a large proportion of culture-positive TB patients were involved in the recent TB transmission chain. Homeless persons were involved in more than half of the genotype clusters. For foreign-born persons, they still have a limited impact on transmission in the Japanese urban setting, but considering recent increases in foreign-born TB patients, transmission between native and foreign-born populations should be routinely evaluated. Intensified public health interventions, such as active case findings, may focus on those who have socioeconomic risk factors that are significantly associated with TB transmission and clusters with shorter registration intervals between the first two cases, which are predictors of the development of large clusters within three years.

ACKNOWLEDGMENTS

We would like to thank the Shinjuku Public Health Center staff for collecting the data.

Author Statement: Authors roles

KIz: concept of study, statistical analysis, interpretation of data, drafting and finalization of the manuscript

YM: genotyping analysis, interpretation of data and finalization of the manuscript

KU: concept of study, interpretation of data and finalization of the manuscript

AK: concept of study, collection of data, and finalization of the manuscript

KIs: concept of study, collection of data, and finalization of the manuscript

SK: concept of study, collection of data, and finalization of the manuscript

TT: interpretation of data and finalization of the manuscript

AK: concept of study, interpretation of data, drafting and finalization of the manuscript

Reference

1. World Health Organization. *Global Tuberculosis Report 2017*.; 2017. doi:WHO/HTM/TB/2017.23.
2. Lönnroth K, Migliori GB, Abubakar I, et al. Towards tuberculosis elimination: An action framework for low-incidence countries. *Eur Respir J*. 2015;45(4):928-952. doi:10.1183/09031936.00214014.
3. de Vries G, Aldridge RW, Cayla J a, et al. Epidemiology of tuberculosis in big cities of the European Union and European Economic Area countries. *Euro Surveill Bull Eur sur les Mal Transm = Eur Commun Dis Bull*. 2014;19(9):1-8.
4. Hest NA, Aldridge RW, de Vries G, et al. Tuberculosis control in big cities and urban risk groups in the European Union: A consensus statement. *Euro Surveill*. 2014;19(9):1-13. doi:10.2807/1560-7917.ES2014.19.9.20728.
5. Jereb JA. Progressing toward tuberculosis elimination in low-incidence areas of the United States. Recommendations of the Advisory Council for the Elimination of Tuberculosis. *MMWR Recomm reports*. 2002;51(RR-5):1-14.
6. Borgdorff MW, van Soolingen D. The re-emergence of tuberculosis: What have we learnt from molecular epidemiology? *Clin Microbiol Infect*. 2013;19(10):889-901. doi:10.1111/1469-0691.12253.
7. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. 1994;330(D):1710-1716.
8. van Soolingen D, Borgdorff MW, de Haas PE, et al. Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis*. 1999;180:726-736. doi:10.1086/314930.
9. PM S, PC H, SP S, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med*. 1994.

10. Hamblion EL, Le Menach A, Anderson LF, et al. Recent TB transmission, clustering and predictors of large clusters in London, 2010-2012: Results from first 3-years of universal MIRU-VNTR strain typing. *Thorax*. 2016;71(8):749-756. doi:10.1136/thoraxjnl-2014-206608.

11. Driver CR, Macaraig M, McElroy PD, et al. Which patients' factors predict the rate of growth of Mycobacterium tuberculosis clusters in an urban community? *Am J Epidemiol*. 2006;164(1):21-31. doi:10.1093/aje/kwj153.

12. Kik S V., Verver S, Van Soolingen D, et al. Tuberculosis outbreaks predicted by characteristics of first patients in a DNA fingerprint cluster. *Am J Respir Crit Care Med*. 2008;178(1):96-104. doi:10.1164/rccm.200708-1256OC.

13. Althomsons SP, Kammerer JS, Shang N, Navin TR. Using Routinely Reported Tuberculosis Genotyping and Surveillance Data to Predict Tuberculosis Outbreaks. *PLoS One*. 2012;7(11):1-8. doi:10.1371/journal.pone.0048754.

14. The Research Institute of Tuberculosis J. Statistics of TB 2016. http://www.jata.or.jp/rit/ekigaku/en/statistics-of-tb/. Accessed June 20, 2018.

15. Ohkado A, Nagamine M, Murase Y, et al. Molecular epidemiology of Mycobacterium tuberculosis in an urban area in Japan, 2002-2006. *Int J Tuberc Lung Dis*. 2008;12(5):548-554.

16. Tuberculosis Surveillance Center. *Tuberculosis in Japan – Annual Report 2017*. Tokyo, Japan: Department of Epidemiology and Clinical Research, the Research Institute of Tuberculosis; 2017.

17. Shinjuku city. Description of Shinjuku. http://www.foreign.city.shinjuku.lg.jp/en/aramashi/aramashi_1/. Accessed January 9, 2018.

18. Shinjuku City. *Tuberculosis Statistics in Shinjuku City 2017 [Japanese]*.; 2018.

19. Van Embden JDA, Cave MD, Crawford JT, et al. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: Recommendations for a standardized methodology. *J Clin Microbiol*. 1993;31(2):406-409. doi:10.1128/JCM.39.4.1683.2001.

20. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: A systematic review of population-based molecular epidemiology studies. *Int J Tuberc Lung Dis*. 2008;12(5):480-492.

21. Wada T, Maeda S, Hase A, Kobayashi K. Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of Mycobacterium tuberculosis in Japan. *J Med Microbiol*. 2007;56(8):1052-1057. doi:10.1099/jmm.0.46990-0.

22. Ellis B a, Crawford JT, Braden CR, McNabb SJN, Moore M, Kammerer S. Molecular epidemiology of tuberculosis in a sentinel surveillance population. *Emerg Infect Dis*. 2002;8(11):1197-1209. doi:10.3201/eid0811.020403.

23. Glynn JR, Crampin AC, Yates MD, et al. The Importance of Recent Infection with *Mycobacterium tuberculosis* in an Area with High HIV Prevalence: A Long-Term Molecular Epidemiological Study in Northern Malawi. *J Infect Dis*. 2005;192(3):480-487. doi:10.1086/431517.

24. 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):1-16. doi:10.1371/journal.pone.0138831.

25. Sandgren A, Schepisi MS, Sotgiu G, et al. Tuberculosis transmission between foreign- and native-born populations in the EU/EEA: A systematic review. *Eur Respir J*. 2014;43(4):1159-1171. doi:10.1183/09031936.00117213.

26. MURASE Y, OHKADO A, WATANABE Y, et al. Transmission dynamics of mycobacterium tuberculosis between foreign-nationals and Japanese tuberculosis patients living in Shinjuku-City, Tokyo, Japan. *Kekkaku*. 2017;92(5):431-439.

27. Murase Y, Izumi K, Ohkado A, et al. Prediction of local transmission of mycobacterium tuberculosis isolates of a predominantly Beijing lineage by use of a variable-number tandem-repeat typing method incorporating a consensus set of hypervariable loci. *J Clin Microbiol*.

2018;56(1):1-11. doi:10.1128/JCM.01016-17.

28. Wada T, Iwamoto T, Tamaru A, et al. Clonality and micro-diversity of a nationwide spreading genotype of Mycobacterium tuberculosis in Japan. *PLoS One*. 2015;10(3):1-13. doi:10.1371/journal.pone.0118495.

29. Murase Y, Mitarai S, Sugawara I, Kato S, Maeda S. Promising loci of variable numbers of tandem repeats for typing Beijing family Mycobacterium tuberculosis. *J Med Microbiol*. 2008;57(7):873-880. doi:10.1099/jmm.0.47564-0.

404 **TABLE**

405 Table 1. Study variables and definitions

Category	Variables	Definition
Demographic factors	Sex	Men or women
	Age	Age as years old at registration (≥40 or <40)
	Country of birth	Japan-born or foreign-born persons
Social factors	Occupation	Fulltime, Part-time/daily worker, jobless under 60 years old, or others (including infant, student, housewife, retired, unknown)
	Receipt of public assistance	Those who receive government welfare benefits for those whose household income is below the minimum living expense at registration
	Homelessness status	Those whose legal address is unknown or not stable at least in the last two years at registration
	Alcohol misuse	Those who have excessive drinking practice judged by the public health nurses
Clinical factors	Site of disease	Those who have pulmonary or extra pulmonary disease
	Cavity lesions	Those who have cavity lesions in lung field on chest radiography
	Sputum smear microscopy	Those who have positive or negative result in sputum smear microscopy test
	Past TB history	Those who have history of past TB treatment
	Status of diabetes mellitus	Those who have diabetes mellitus status by the patient's self-report
Others	Mode of detection	Those who found through active case finding conducted by public health centers

Status of patient delay	A time between the onset of symptoms and the initial doctor visit longer than two months
Status of doctor delay	A time between the initial doctor visit and diagnosis longer than one month
Status of total delay	A time between onset of symptoms and TB diagnosis longer than three months
Registration interval	The duration in months between the registration dates of the first two cases in each of the genotype clusters

406

407

408

409 Table 2. Factors associated with TB genotype clustering; univariable logistic regression analysis, RFLP,

410 Shinjuku 2002–2013

	Total number of cases (N=1025), n	Clustered cases (N=515), n (%)		OR	(95% CI)	p-value
Age (years old)	1025					
≥40	754	371	(49.2)	Reference		
<40	271	144	(53.1)	1.17	(0.88-1.56)	0.267
Sex	1025					
Female	248	102	(41.1)	Reference		
Male	777	413	(53.2)	1.62	(1.20-2.19)	0.001**
Country of birth	1025					
Foreign	95	22	(23.2)	Reference		
Japan	930	493	(53.0)	3.74	(2.25-6.44)	<0.001***
Occupation	1025					
Fulltime worker	313	165	(52.7)	1.53	(1.15-2.05)	0.004**
Part-time/daily worker	96	60	(62.5)	2.29	(1.45-3.61)	<0.001***
Jobless aged 15-59-year-old	172	103	(59.9)	2.05	(1.43-2.94)	<0.001***
Others †	444	187	(42.1)	Reference		
Public Assistance ‡	1024					
No	720	319	(44.3)	Reference		

Yes	304	195	(64.1)	2.25	(1.69-3.00)	<0.001***
Homelessness	1025					
No	776	349	(45.0)	Reference		
Yes	249	166	(66.7)	2.45	(1.80-3.34)	<0.001***
Alcohol misuse [§]	1025					
No	761	367	(48.2)	Reference		
Yes	264	148	(56.1)	1.37	(1.02-1.83)	0.028*
TB site	1024					
Extra-pulmonary	80	32	(40.0)	Reference		
Pulmonary	944	482	(51.1)	1.56	(0.96-2.58)	0.058
Cavity lesions	1023					
No	565	271	(48.0)	Reference		
Yes	458	243	(53.1)	1.23	(0.95-1.58)	0.105
Smear results	1024					
Negative	406	192	(47.3)	Reference		
Positive	618	322	(52.1)	1.21	(0.94-1.57)	0.132
Past TB history	989					
New	880	441	(50.1)	Reference		
Retreatment	109	59	(54.1)	1.17	(0.77-1.79)	0.429
DM	1005					
No	832	421	(50.6)	Reference		
Yes	173	86	(49.7)	0.97	(0.69-1.36)	0.831
Active case finding	1025					
No	842	412	(48.9)	Reference		

Yes	183	103	(56.3)	1.34	(0.96-1.88)	0.071
Patient delay	1000					
<2m	773	377	(48.8)	Reference		
≥2m	227	127	(55.9)	1.33	(0.98-1.82)	0.057
Doctor delay	1018					
<1m	799	415	(51.9)	Reference		
≥1m	219	97	(44.3)	0.74	(0.54-1.00)	0.045*
Total delay	997					
<3m	777	382	(49.2)	Reference		
≥3m	220	122	(55.5)	1.29	(0.94-1.76)	0.099

RFLP: restriction fragment length polymorphism, OR: odds ratio, CI: confidence interval, TB: tuberculosis, DM: diabetes mellitus, *p<0.05, **p<0.01, ***p<0.001

† Occupation of others includes infant, student, housewife, retired, and unknown.

‡ Public Assistance refers to government welfare benefits for those whose household income is below the minimum living expense.

§ Alcohol misuse means excessive drinking practice judged by the public health nurses conducting the interviews.

Table 3. Factors associated with TB clustering; multivariate logistic regression analysis, RFLP, Shinjuku 2002–2013

Variables		aOR	(95% CI)	p-value
Age (years old)	≥40	Reference		
	<40	1.73	(1.23-2.44)	0.002**
Country of birth	Foreign	Reference		
	Japan	3.90	(2.27-6.72)	<0.001***
Occupation	Fulltime worker	1.63	(1.17-2.27)	0.004**
	Part-time/daily worker	2.20	(1.35-3.58)	0.002**
	Jobless aged 15-59-year-old	1.32	(0.88-1.97)	0.180
	Others †	Reference		
Public Assistance ‡	No	Reference		
	Yes	1.81	(1.15-2.84)	0.011*
Homeless	No	Reference		
	Yes	1.63	(1.02-2.62)	0.042*
Alcohol misuse §	No	Reference		
	Yes	1.29	(0.79-2.11)	0.311
Active case finding	No	Reference		
	Yes	1.39	(0.98-1.99)	0.066

RFLP: restriction fragment length polymorphism, aOR: adjusted odds ratio, CI: confidence interval, *p<0.05, **p<0.01, ***p<0.001

† Occupation of others includes infant, student, housewife, retired, and unknown.

‡ Public Assistance refers to government welfare benefits for those whose household income is below the minimum living expense.

[§] Alcohol misuse means excessive drinking practice judged by the public health nurses conducting the interviews.

Table 4. Analysis of factors associated with large genotype clusters within three years using the characteristics of the first two cases in each TB genotype cluster; univariable and multivariable logistic regression, RFLP, Shinjuku 2003–2013 (N=104 cluster episodes)

Variable	Large clusters within 3 years (N=14), n (%)†		Small clusters and large clusters after 3 years (N=90), n (%)‡		Univariate logistic regression OR	(95% CI)	p Value
Sex							
No male	1	(7.1)	4	(4.4)	Ref		
≥One male	13	(92.9)	86	(95.6)	0.60	(0.06-5.84)	0.664
Age							
No cases <40 years	8	(57.1)	57	(63.3)	Ref		
At least one case <40 years	6	(42.9)	33	(36.7)	1.30	(0.41-4.06)	0.657
Japanese							
Non-Japan born	0	(0.0)	2	(2.2)	Ref		
≥One case Japan born	14	(100.0)	88	(97.8)	NA		
Full- and part-time/daily worker							
No case being full- and part-time/daily worker	6	(42.9)	35	(38.9)	Ref		
≥One case being full- and part-time/daily worker	8	(57.1)	55	(61.1)	0.85	(0.27-2.65)	0.778
Public Assistance							
No case with public assistance	5	(35.7)	41	(45.6)	Ref		
≥One case with public assistance	9	(64.3)	49	(54.4)	1.51	(0.47-4.85)	0.492
Homeless							

No case with currently homeless	6	(42.9)	45	(50.0)	Ref		
≥One case with currently homeless	8	(57.1)	45	(50.0)	1.33	(0.43-4.15)	0.620
Alcohol misuse							
No case with alcohol misuse	5	(35.7)	48	(53.3)	Ref		
≥One case with alcohol misuse	9	(64.3)	42	(46.7)	2.06	(0.64-6.62)	0.227
Cavity lesions							
No case with cavity	2	(14.3)	24	(26.7)	Ref		
≥One case with cavity	12	(85.7)	66	(73.3)	2.18	(0.45-10.47)	0.330
Smear results							
No case with smear positive	1	(7.1)	12	(13.3)	Ref		
≥One case with smear positive	13	(92.9)	78	(86.7)	2.00	(0.24-16.71)	0.522
Past TB history							
No case with past TB history	11	(78.6)	69	(76.7)	Ref		
≥One case with past TB history	3	(21.4)	21	(23.3)	0.90	(0.23-3.52)	0.875
DM							
No case with DM	9	(64.3)	57	(63.3)	Ref		
≥One case with DM	5	(35.7)	33	(36.7)	0.96	(0.30-3.11)	0.945
Active case finding							
No case found through active case finding	8	(57.1)	53	(58.9)	Ref		
≥One case found through active case finding	6	(42.9)	37	(41.1)	1.07	(0.34-3.35)	0.902
Patient delay							
No case with patient delay	9	(64.3)	55	(61.1)	Ref		
≥One case with patient delay	5	(35.7)	35	(38.9)	0.87	(0.27-2.82)	0.820
Doctors delay							

No case with doctor delay	10	(71.4)	57	(63.3)	Ref		
≥One case with doctor delay	4	(28.6)	33	(36.7)	0.69	(0.20-2.38)	0.558
Total delay							
No case with total delay	10	(71.4)	53	(58.9)	Ref		
≥One case with total delay	4	(28.6)	37	(41.1)	0.57	(0.17-1.97)	0.376
Registration interval							
0-2 months between first two cases	7	(50.0)	13	(14.4)	9.51	(2.16-41.89)	0.003**
3-5 months between first two cases	2	(14.3)	5	(5.6)	7.07	(0.95-52.77)	0.057
6-11 months between first two cases	2	(14.3)	19	(21.1)	1.86	(0.29-12.00)	0.514
≥12 months between first two cases	3	(21.4)	53	(58.9)	Ref		

Note. After selecting variables for multivariate logistic regression by the stepwise method, only the variable of registration interval remained in the model. Thus, the table shows only the result of univariate logistic regression.

RFLP: restriction fragment length polymorphism, OR: odds ratio, aOR: adjusted odds ratio, CI: confidence interval, TB: tuberculosis, DM: diabetes mellitus, Ref: reference, NA: not applicable, *p<0.05, **p<0.01

† "Large clusters within three years" were cluster episodes with five or more cases (large clusters) within three years

‡ "Small clusters and large clusters after three years", which were cluster episodes with two to four cases (small clusters) and cluster episodes that became large clusters after three years.

Figure legend

Figure 1. Number of reported cases of TB, including culture-positive cases, strain-typed cases and genotype clusters, Shinjuku, 2002–2013

RFLP: restriction fragment length polymorphism, TB: tuberculosis

Figure 2. Cumulative clustering rate, RFLP, Shinjuku 2002–2013

RFLP: restriction fragment length polymorphism

Figure 3. Time series of the number (A) and proportion (B) of clustered homeless cases, RFLP, Shinjuku 2003-2012

RFLP: restriction fragment length polymorphism

Note. The 2002 and 2013 data were omitted because of the small total number of cases (31 and 23 cases, respectively). The percentage of clustered cases among the homeless was calculated as the number of clustered homeless cases divided by the number of all homeless cases, and the percentage of the homeless among all study cases was calculated as the number of homeless cases divided by the number of all study cases.

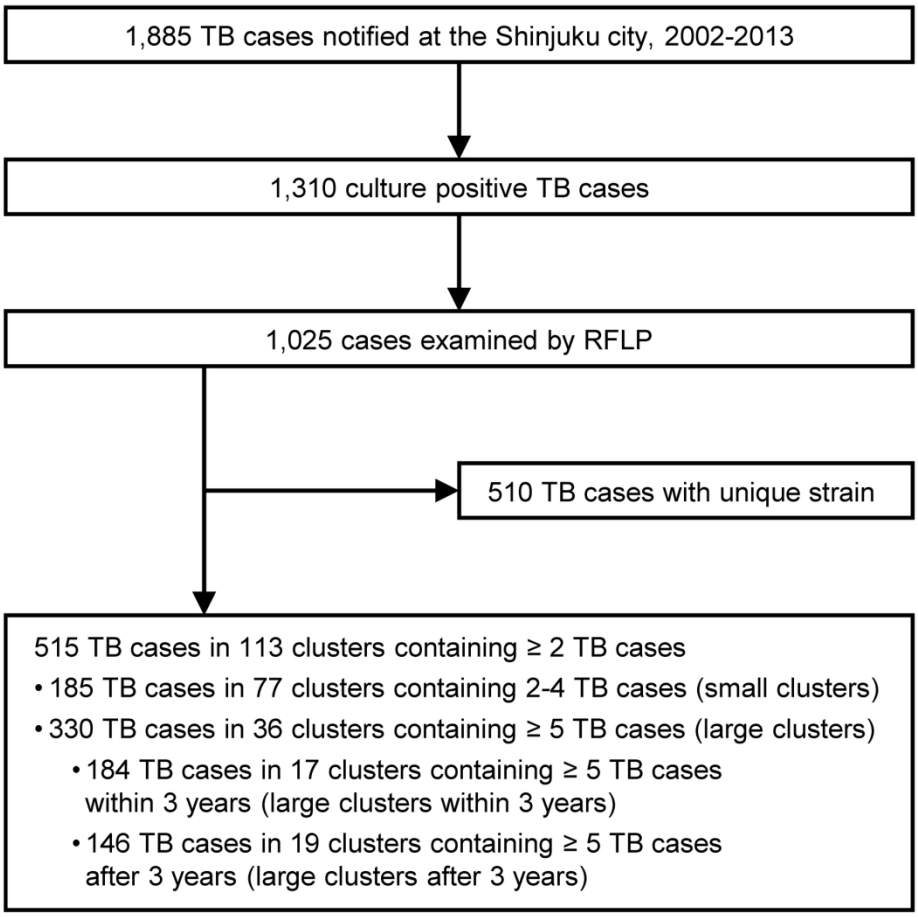

Figure 1. Number of reported cases of TB, including culture-positive cases, strain-typed cases and genotype clusters, Shinjuku, 2002–2013
RFLP: restriction fragment length polymorphism, TB: tuberculosis

90x90mm (600 x 600 DPI)

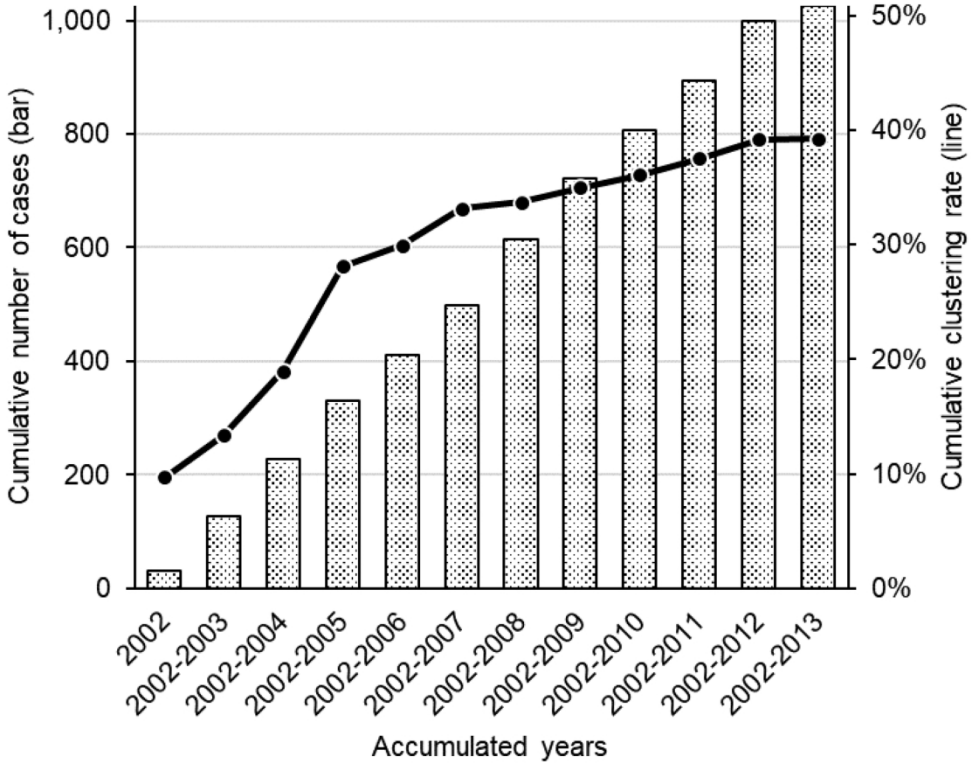

Figure 2. Cumulative clustering rate, RFLP, Shinjuku 2002–2013
RFLP: restriction fragment length polymorphism

90x90mm (600 x 600 DPI)

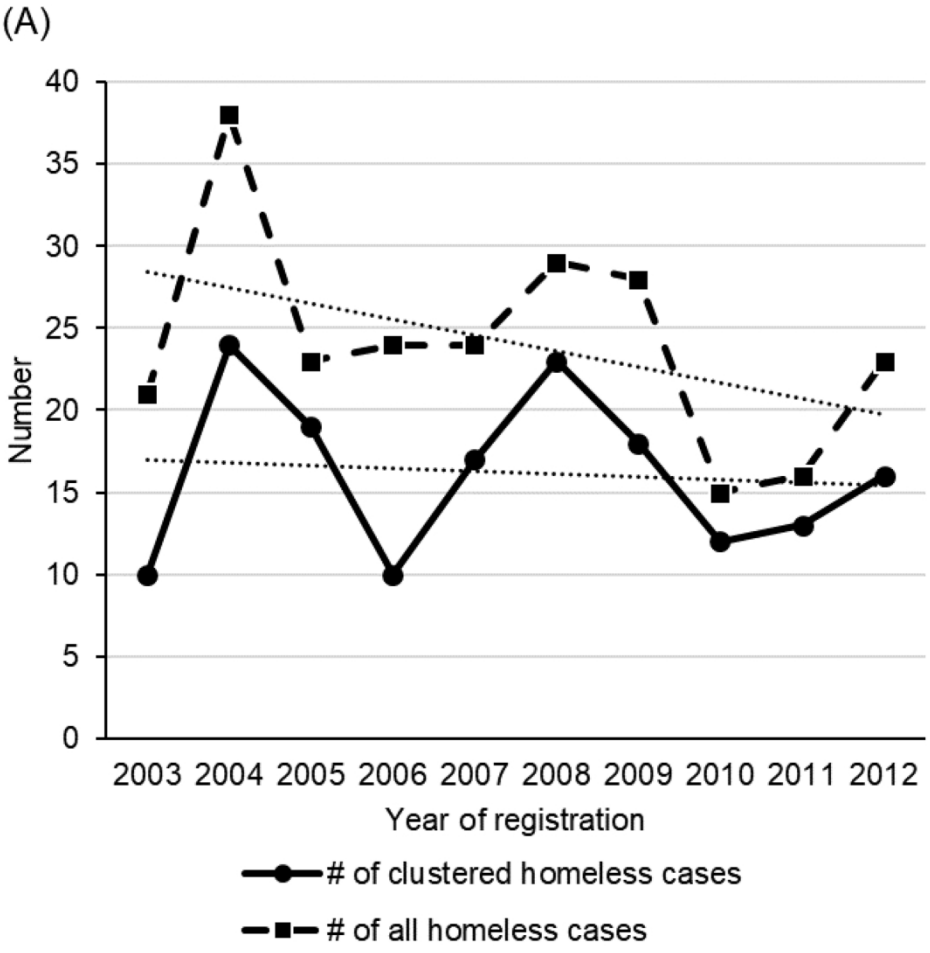

Figure 3. Time series of the number (A) and proportion (B) of clustered homeless cases, RFLP, Shinjuku 2003-2012

RFLP: restriction fragment length polymorphism

Note. The 2002 and 2013 data were omitted because of the small total number of cases (31 and 23 cases, respectively). The percentage of clustered cases among the homeless was calculated as the number of clustered homeless cases divided by the number of all homeless cases, and the percentage of the homeless among all study cases was calculated as the number of homeless cases divided by the number of all study cases.

90x90mm (600 x 600 DPI)

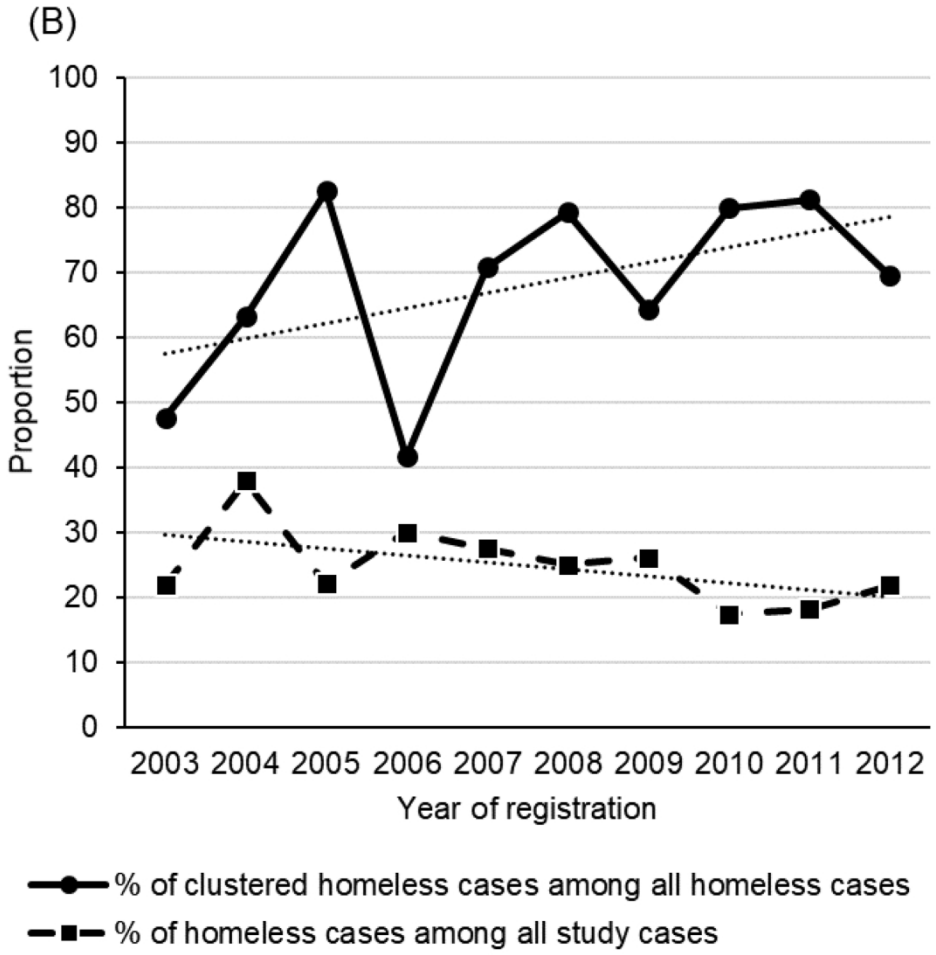

Figure 3. Time series of the number (A) and proportion (B) of clustered homeless cases, RFLP, Shinjuku 2003-2012
RFLP: restriction fragment length polymorphism
Note. The 2002 and 2013 data were omitted because of the small total number of cases (31 and 23 cases, respectively). The percentage of clustered cases among the homeless was calculated as the number of clustered homeless cases divided by the number of all homeless cases, and the percentage of the homeless among all study cases was calculated as the number of homeless cases divided by the number of all study cases.

90x90mm (600 x 600 DPI)

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Table 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	6 **We took all eligible study participants in Shinjuku city.*
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Table 1
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-8
		(*b*) Describe any methods used to examine subgroups and interactions	8
		(*c*) Explain how missing data were addressed	8
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	Not Applicable

		(*e*) Describe any sensitivity analyses	Not Applicable
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8-9 Figure 1
		(b) Give reasons for non-participation at each stage	8-9
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	N in tables
Outcome data	15*	Report numbers of outcome events or summary measures	Tables, 9-10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Tables
		(*b*) Report category boundaries when continuous variables were categorized	Tables
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not Applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Table 4
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-14
Generalisability	21	Discuss the generalisability (external validity) of the study results	11-14
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	4

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in Tokyo, Japan: a population-based molecular epidemiological study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029295.R1
Article Type:	Research
Date Submitted by the Author:	27-Mar-2019
Complete List of Authors:	Izumi, Kiyohiko; The Research Institute of Tuberculosis, Murase, Yoshiro; The Research Institute of Tuberculosis Uchimura, Kazuhiro; The Research Institute of Tuberculosis Kaebeta, Aya; Shinjuku Public Health Center Ishihara, Keiko; Shinjuku Public Health Center Kaguraoka, Sumi; Shinjuku Public Health Center Takii, Takemasa; The Research Institute of Tuberculosis Ohkado, Akihiro; The Research Institute of Tuberculosis
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Epidemiology, Public health
Keywords:	RFLP, clustering rate, homeless, foreign-born

Title

Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in Tokyo, Japan: a population-based molecular epidemiological study

Authors and affiliations

1. Kiyohiko Izumi. Department of Epidemiology and Clinical Research, The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: kizumi95@gmail.com
2. Yoshiro Murase. Department of Mycobacterium Reference and Research, The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: ymurase@jata.or.jp
3. Kazuhiro Uchimura. Department of Epidemiology and Clinical Research, The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: uchimura@jata.or.jp
4. Aya Kaebeta. Shinjuku Public Health Center. Email: aya.kaebeta@city.shinjuku.lg.jp
5. Keiko Ishihara. Shinjuku Public Health Center. Email: keiko.ishihara@city.shinjuku.lg.jp
6. Sumi Kaguraoka. Shinjuku Public Health Center. Email: sumi.kaguraoka@city.shinjuku.lg.jp
7. Takemasa Takii. The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo, Japan. Email: t-takii@jata.or.jp
8. Akihiro Ohkado. Department of Epidemiology and Clinical Research, The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo, Japan. Graduate School of Biomedical Sciences, Nagasaki University. Email: ohkadoa@jata.or.jp

Corresponding Author

Kiyohiko Izumi

Department of Epidemiology and Clinical Research, Research Institute of Tuberculosis, Japan Anti-

Tuberculosis Association, 3-1-24 Matsuyama, Kiyose, Tokyo 204-8533, JAPAN

Tel: ±81-42-4935711

Fax: +81-42-492-4600

Email: kizumi95@gmail.com

KEY WORDS: RFLP, clustering rate, homeless, foreign-born

ABSTRACT

Objective: Molecular epidemiology is a promising tool for understanding tuberculosis transmission dynamics but has not been sufficiently utilized in Asian countries including Japan. The aim of this study was to estimate the proportion of TB cases attributable to recent transmission and to identify risk factors of genotype clustering and the development of large clusters within three years in an urban setting in Japan.

Design and setting: Long-term cross-sectional observational study combining the characteristics of culture-positive TB patients notified in Shinjuku City, Tokyo (2002-2013), with genotype data of *Mycobacterium tuberculosis.*

Primary outcome measure: Genotype clustering rate and association between genotype clustering status and explanatory variables.

Results: Among 1,025 cases, 515 (50.2%) were localized within 113 genotype clusters. The overall clustering rate was 39.2%. Significantly higher rates were found in patients aged <40 years (adjusted odds ratio (aOR)=1.73, 95% CI=1.23-2.44), native Japanese individuals (aOR=3.90, 95% CI=2.27-6.72), fulltime workers (aOR=1.63, 95% CI=1.17-2.27), part-time/daily workers (aOR=2.20, 95% CI=1.35-3.58), individuals receiving public assistance (aOR=1.81, 95% CI=1.15-2.84), and homeless people (aOR=1.63, 95% CI=1.02-2.62). A significant predictor of large genotype clusters within three years was a registration interval ≤2 months between the first two cases in a cluster.

Conclusion: Our results indicated that a large proportion of culture-positive TB patients were involved in the recent TB transmission chain. Foreign-born persons still have a limited impact on transmission in the Japanese urban setting. Intensified public health interventions, including the active case finding, need to focus on individuals with socioeconomic risk factors that are significantly associated with tuberculosis transmission and clusters with shorter registration intervals between the first two cases.

Word Count in Abstract= 261 words

Article Summary

Strengths and limitations of this study

- This study is one of the longest population-based studies focusing on the molecular epidemiology of culture-positive notified tuberculosis patients in a large Asian urban setting.
- Interviews conducted by the experienced public health nurses at the Public Health Center using a standardized questionnaire provided high-quality data and less interviewer bias.
- We may have underestimated genotype clustering due to the large population flow in and out of the city.

Funding statement

This research was supported by the Research Program on Emerging and Re-emerging Infectious Diseases from the Japan Agency for Medical Research and Development (AMED No. JP18fk0108041).

Competing interests: None declared.

Patient consent: Not required.

Data sharing statement: Due to data restrictions, we are unable to share any aspect of the data.

Word count: 2,776 words

INTRODUCTION

Tuberculosis (TB) remains a major public health threat worldwide. In 2017, an estimated 10 million people worldwide developed TB, and 1.27 million died from TB[1]. Although the majority of cases have been reported in countries with a high TB burden, TB remains a persistent health problem in low- and medium-burden countries because it is concentrated in specific vulnerable and hard-to-reach populations, such as homeless people and foreign-born persons from TB high-burden countries[2]. These specific high-risk populations tend to live in large cities where they are seeking jobs, which potentially poses challenges to the control of TB in urban areas[3,4]. Many countries with a low or medium TB burden have recently adopted TB elimination strategies[2,5], which emphasizes the importance of molecular epidemiology in TB control, particularly in urban areas[2,4].

TB molecular genotyping using restriction fragment length polymorphisms (RFLPs) and, more recently, variable numbers of tandem repeats (VNTRs) combined with epidemiological information identifies TB cases that are likely involved in the same transmission chain[6]. This method differentiates recent transmission or endogenous reactivation from remote infection and has therefore revealed that a substantial proportion of TB cases are due to recent transmission in low-TB-burden countries[7–9]. This method also identifies the proportion of cases attributable to recent transmission and determines the risk factors for transmission. Moreover, various factors predicting large TB genotype clusters, including socially vulnerable populations and shorter intervals between the registration dates of the first two cases, have been investigated by evaluating the characteristics of the first two cases in the same genotype cluster[10–13]. These population-based molecular epidemiological studies were conducted in some European countries[8,10,12], the U.S.[7,9,11] and some Asian countries[14–19].

In Japan, a country with a medium TB burden, the number of newly notified TB cases decreased from

32,828 (25.8 per 100,000 populations) in 2002 to 17,625 (13.9 per 100,000 populations) in 2016[20], but the central government has constantly been reported of TB outbreaks by local governments at a rate of approximately 40 events annually over the last decade. This information suggests that TB transmission might be occurring in some groups, such as homeless people, who constitute a high-risk group for recent TB transmission in urban areas[14]. Considering the steady increase in the proportion of TB cases among foreign-born individuals in Japan (7.9% of all cases in 2016[21]), transmission between foreign-born persons and local residents must be monitored. In addition, in light of Japan's transition toward becoming a low-TB-burden country, understanding TB transmission patterns has become increasingly important. However, few population-based molecular epidemiological studies have identified the transmission patterns in Japan and their risk factors. Additionally, no study has attempted to evaluate the factors predicting the development of large clusters in Japan.

Therefore, we aimed to estimate the proportion of TB cases attributable to recent transmission, to identify the risk factors for recent transmission, and to predict the risk factors for the development of large clusters in an urban setting.

METHODS

Study population

We included all culture-positive TB patients notified in Shinjuku City from September 2002 to December 2013 as the eligible study population in this cross-sectional observational study. This study forms part of a population-based study on DNA fingerprinting surveillance of *M. tuberculosis* in Shinjuku City that was started in 2002. Shinjuku City (18.3 km^2) is one of the most populous (342,867 residents in 2018[22]) cities in Tokyo, and its TB notification rate in 2016 was 33.7 per 100,000 people[23],

which was higher than the rates in Tokyo and the nation (17.2 and 13.9, respectively[20]). Experienced public health nurses at the Shinjuku Public Health Center (PHC) interviewed and collected information from all culture-positive TB patients at the time of registration using a standardized questionnaire to avoid possible interviewer bias. The study variables and definitions are described in Table 1.

Patient and public involvement

Neither the patients nor the public were involved in the design of this study.

DNA fingerprinting and genotype cluster

Clinical isolates from each of the enrolled TB patients were sent to the Research Institute of Tuberculosis (RIT), Tokyo, where the TB strains were subjected to DNA fingerprinting using insertion sequence *6110* by RFLP (IS*6110*-RFLP) analysis[24]. One clinical isolate per person was used for the clustering analysis. IS*6110*-RFLP and spoligotyping are the standard methods used in the Shinjuku PHC and were available throughout the study period. The Shinjuku PHC switched from RFLP to VNTR a few years ago, but the RFLP profiles of many TB cases were available. Thus, we employed RFLP due to the sufficient sample size. A genotype cluster was defined as a group of TB patients whose isolates showed either (1) ≥6 identical IS*6110* band patterns or (2) <6 identical IS*6110* band patterns confirmed by identical spoligotyping patterns. The data collection and genotyping methods were previously described in detail[14].

Data analysis

We calculated the genotype clustering rate by the "n-1 method" according to the formula *{(n - c)/N}*, where *N* is the total number of cases sampled, *c* is the number of clusters, and *n* is the total number

of cases in the clusters[9]. We also calculated the cumulative clustering rate by calculating the clustering rate in 2002 and then adding the TB patients every year up to 2013. The characteristics of clustered cases, which were the cases belonging to any genotype clusters, were compared with those with unique strain patterns through χ^2 tests. We performed univariate logistic regression to identify risk factors for genotype clustering using odds ratios (ORs) and multivariate logistic regression using adjusted ORs (aORs). Any potential interactions were assessed using likelihood ratio tests.

Additionally, we compared the characteristics of the first two cases in each genotype cluster to identify risk factors for the development of a large cluster within three years. For this purpose, a cluster episode was defined as a newly arising genotype cluster in or after 2003 without any TB cases of that genotype notified prior to that year. We classified cluster episodes into the following two groups according to a system developed in a previous study[10]: (1) "large clusters within three years" were cluster episodes with five or more cases (large clusters) occurring within three years and (2) "small clusters and large clusters after three years" were cluster episodes with two to four cases (small clusters) and cluster episodes that became large clusters after three years. We identified the first two cases in each cluster episode based on the notification date and compared their characteristics between these two groups. We performed univariate and multivariate logistic regression analyses to identify predictors of the development of large clusters within three years.

A *p*-value of 0.05 was set as the level indicating statistical significance. For variables with more than 5% missing values, the multiple imputation method was considered. The variables used for multivariate logistic regressions were selected by the stepwise maximum-likelihood estimation with a significance level of less than 0.2. We used Stata version 12 (Stata Corp., College Station, TX, USA) for the statistical analyses. Written informed consent was waived because DNA fingerprinting

analysis forms part of the routine TB control activities in Shinjuku City. However, oral informed consent was obtained after the PHC staff provided a thorough explanation of the study objectives and confidentiality. The study protocol was approved by the Institutional Review Board of the Research Institute of Tuberculosis (RIT/IRB27-9).

RESULTS

Study population and clustering rate

In total, 1,885 TB patients in Shinjuku City were notified during the study period, and 1,310 were culture-positive cases (Figure 1). Of these, 285 patients were excluded from the analysis, mainly due to the unavailability of culture-positive isolates and the lack of implementation of RFLP. As a result, 1,025 (78.2%) patients were included in the analysis. Figure 2 shows the cumulative number of TB patients and the clustering rates from 2002 to 2013. The number of TB cases gradually increased over the tested decade. In contrast, the cumulative clustering rates sharply increased in the first four years, from 10% in 2002 to 28% in 2005, with an average percent change of +43%, and then continued to increase at a slower rate, from 30% in 2006 to 39% in 2013, with an average percent change of +4.2%.

We identified a total of 113 genotype clusters consisting of 515 (50.2%) patients (Figure 1). The genotype clustering rate was 39.2%, and the average cluster size was 4.56 cases (range 2-30). Fifty-seven (50.4%) genotype clusters consisted of only two TB patients, and 36 (31.9%) genotype clusters had at least five TB patients. We further investigated the homelessness status and place of birth of the patients in the genotype clusters. Of the 113 genotype clusters, 45 (39.8%) comprised only nonhomeless individuals, seven (6.2%) included only homeless individuals, and 61 (54.0%) contained both homeless and nonhomeless individuals (mixed cluster). We compared the characteristics of the

nonhomeless patients in the clusters of only nonhomeless patients with those in the mixed clusters, and although the finding was not statistically significant (Pearson χ^2 test, p=0.17), the proportion of nonhomeless patients receiving public assistance in the latter group (13.8%) was higher than that in the former group (8.8%). No differences in sex, age, and place of birth were found between the two groups. Of the 113 genotype clusters, 94 (83.2%) consisted of only individuals born in Japan, two (1.8%) consisted of only foreign-born individuals, and 17 (15.0%) consisted of both individuals born in Japan and foreign-born individuals.

Factors associated with genotype clustering

The clustered cases were significantly more likely to consist of male individuals (OR=1.62, 95% CI=1.20-2.19), Japan-born individuals (OR=3.74, 95% CI=2.25-6.44), individuals receiving public assistance (OR=2.25, 95% CI=1.69-3.00), homeless individuals (OR=2.45, 95% CI=1.80-3.34), individuals who misuse alcohol (OR=1.37, 95% CI=1.02-1.83), individuals engaging in fulltime work (OR=1.53, 95% CI=1.15-2.05) and part-time/daily work (OR=2.29, 95% CI=1.45-3.61), and jobless individuals aged 15-59 years (OR=2.05, 95% CI=1.43-2.94) (Table 2). A significant interaction among the explanatory variables was not detected. The multivariate analysis demonstrated that the factors associated with genotype clustering were age <40 years (aOR=1.73, 95% CI=1.23-2.44), born in Japan (aOR=3.90, 95% CI=2.27-6.72), working fulltime (aOR=1.63, 95% CI=1.17-2.27), having part-time/daily work (aOR=2.20, 95% CI=1.35-3.58), receiving public assistance (aOR=1.81, 95% CI=1.15-2.84), and homelessness (aOR=1.63, 95% CI=1.02-2.62)(Table 3).

Factors associated with large genotype clustering within three years

We identified 104 genotype cluster episodes according to the definition. Of these, 14 were “large clusters within three years”, which was equivalent to 13.5% (14/104) of all the genotype clusters and

48.3% (14/29) of the large genotype clusters, and 90 clusters were “small clusters and large clusters after three years”. The univariate analysis indicated that clusters with registration intervals of 0-2 months were 9.51 times more likely to become large genotype clusters within three years compared with clusters with registration intervals of ≥12 months (Table 4). After selecting variables using the stepwise method, only the “registration interval” variable remained for the multivariate model.

DISCUSSION

In this long-term population-based study, we included 1,025 patients, identified a total of 113 genotype clusters, and obtained a genotype clustering rate of 39.2%. Our results indicated that the clustered cases were more likely to have certain socioeconomic predictive factors, namely, being homeless, receiving public assistance, and having an unstable job, at the time of tuberculosis diagnosis. A shorter registration interval between the first two cases was a statistically significant predictor of the development of a large genotype cluster within three years.

Clustering rate

We identified 515 (50.2%) genotype clustered cases and estimated a clustering rate of 39.2%. The rate was the same as the pooled clustering rate (40.9%) obtained in a previous meta-analysis of population-based studies of countries with a low TB incidence[19] but differed from previous estimates obtained in Japanese studies, which were 27.6% in Shinjuku and 24.6% in Osaka[14,25]. Because the meta-regression analysis clarified that longer study durations are associated with an increased clustering rate[19], this difference could be due to shorter study durations combined with the smaller sample sizes of the previous studies (388 patients in five years and 195 patients in one year, respectively). In our study, as expected, the cumulative clustering rate rapidly increased in the first four years and increased more slowly thereafter, which is similar to the trend observed in the

previous studies[26,27].

Factors associated with genotype clustering

Our results indicated that the clustered cases were more likely to have socioeconomic predictive factors, namely, being homeless, receiving public assistance, and having an unstable job, at the time of TB diagnosis. Similarly, previous studies suggested that being homeless significantly contributed to clustering in Shinjuku City[14] and other counties[19]. In our study, more than half of the genotype clusters were mixtures of nonhomeless and homeless patients. Moreover, the nonhomeless patients in the mixed clusters tended to be financially unstable, and a higher proportion of these patients were receiving public assistance compared with the proportion among clusters of only nonhomeless cases, which could imply that relatively poor nonhomeless patients share activity spaces with homeless patients, such as urban areas around the large train stations that were reported to be significant hotspots for homeless patients in Shinjuku City[28]. These findings could suggest that contact investigations of homeless TB patients need to be actively expanded to possible contact persons who are not homeless, particularly those who are facing financial difficulty.

A meta-analysis based on studies conducted in European countries where foreign-born patients substantially contribute to TB epidemiology found that the proportion of mixed clusters composed of native and foreign-born patients ranged from 0% to 36.5% and concluded that foreign-born patients did not have a significant influence on TB in the native population[29]. In our study, the proportion of mixed clusters (15.0%) fell into this range. Thus, the impact of TB transmission between native and foreign-born populations likely remains limited in this urban setting[30]. However, considering the recent increase in immigrant TB patients in urban cities, TB transmission between native and foreign-born populations needs to be closely monitored.

Factors associated with large genotype clustering within three years

A shorter registration interval (≤2 months) was identified as a significant predictor of the development of a large genotype cluster within three years, which is compatible with findings of previous studies conducted in the Netherlands and London[10,12]. Therefore, when TB patients with identical genotypes have shorter registration intervals, a thorough active case findings need to be performed to investigate the potential infection sources and infected patients in order to prevent further transmission. However, it is difficult to assume that the first patient infected the second patient because a window of two months appears too short. Thus, we believe that a true but unidentified first TB case was not identified in our study. A cluster episode was defined as a cluster without any TB patients in 2002 and at least two patients with identical genotypes in and after 2003. Therefore, a possible true first TB case might have been registered before 2002, which was outside of our study period, or registered outside of Shinjuku City.

Limitations

Our study has some limitations. First, the study population consisted only of TB patients living in Shinjuku City. Considering the large population flow in and out of the city, as mentioned above, we potentially missed patients living outside of the city who shared TB strain types with patients living in the city. In fact, previous Japanese studies reported clusters with TB patients living across broad geographic areas[31]. Consequently, we may have underestimated the identified genotype clusters. Second, even the existence of TB patients with identical genotyping patterns may not suggest recent transmission if the strain is a nationwide endemic TB strain[32], which could have led to an overestimated clustering rate. Third, IS*6110* RFLP has relatively lower discriminatory power compared with VNTR[33] and whole-genome sequencing[34,35], which might have led to overestimation.

Lastly, information of epidemiological linkage among TB patients was not available in our study. Therefore, we could not assess and discuss the current practices involving epidemiological investigations done by the public health center, which could weaken the programmatic implications of our results.

Conclusion

This study constitutes a one of the longest-term studies on the molecular epidemiology of notified TB patients in a large Asian urban setting. Our results indicated that a large proportion of culture-positive TB patients were involved in the recent TB transmission chain. Homeless persons were found to be involved in more than half of the genotype clusters. Foreign-born persons continue to have a limited impact on TB transmission in the Japanese urban setting, but considering recent increases in foreign-born TB patients, transmission between native and foreign-born populations should be routinely evaluated. Intensified public health interventions, such as active case findings, should focus on those with socioeconomic risk factors that are significantly associated with TB transmission and clusters with shorter registration intervals between the first two cases because these variables could serve as predictors of the development of large clusters within three years.

ACKNOWLEDGMENTS

We would like to thank the Shinjuku Public Health Center staff for collecting the data.

Author Statement: Authors roles

KIz: conception of study, statistical analysis, interpretation of data, and drafting and finalization of the manuscript

YM: genotyping analysis, interpretation of data, and finalization of the manuscript

KU: conception of study, interpretation of data, and finalization of the manuscript

AK: conception of study, collection of data, and finalization of the manuscript

KIs: conception of study, collection of data, and finalization of the manuscript

SK: conception of study, collection of data, and finalization of the manuscript

TT: interpretation of data and finalization of the manuscript

AO: conception of study, interpretation of data, and drafting and finalization of the manuscript

References

1. World Health Organization. *Global Tuberculosis Report 2018*. Geneva; 2018. doi:ISBN 978 92 4 156539 4.
2. Lönnroth K, Migliori GB, Abubakar I, et al. Towards tuberculosis elimination: An action framework for low-incidence countries. *Eur Respir J*. 2015;45(4):928-952. doi:10.1183/09031936.00214014.
3. de Vries G, Aldridge RW, Cayla J a, et al. Epidemiology of tuberculosis in big cities of the European Union and European Economic Area countries. *Euro Surveill Bull Eur sur les Mal Transm = Eur Commun Dis Bull*. 2014;19(9):1-8.
4. Hest NA, Aldridge RW, de Vries G, et al. Tuberculosis control in big cities and urban risk groups in the European Union: A consensus statement. *Euro Surveill*. 2014;19(9):1-13. doi:10.2807/1560-7917.ES2014.19.9.20728.
5. Jereb JA. Progressing towards tuberculosis elimination in low-incidence areas of the United States. Recommendations of the Advisory Council for the Elimination of Tuberculosis. *MMWR Recomm reports*. 2002;51(RR-5):1-14.
6. Borgdorff MW, van Soolingen D. The re-emergence of tuberculosis: What have we learnt from molecular epidemiology? *Clin Microbiol Infect*. 2013;19(10):889-901. doi:10.1111/1469-0691.12253.
7. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. 1994;330(D):1710-1716.
8. van Soolingen D, Borgdorff MW, de Haas PE, et al. Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis*. 1999;180:726-736. doi:10.1086/314930.
9. PM S, PC H, SP S, et al. The epidemiology of tuberculosis in San Francisco. A population-based

study using conventional and molecular methods. *N Engl J Med*. 1994.

10. Hamblion EL, Le Menach A, Anderson LF, et al. Recent TB transmission, clustering and predictors of large clusters in London, 2010-2012: Results from first 3-years of universal MIRU-VNTR strain typing. *Thorax*. 2016;71(8):749-756. doi:10.1136/thoraxjnl-2014-206608.

11. Driver CR, Macaraig M, McElroy PD, et al. Which patients' factors predict the rate of growth of Mycobacterium tuberculosis clusters in an urban community? *Am J Epidemiol*. 2006;164(1):21-31. doi:10.1093/aje/kwj153.

12. Kik S V., Verver S, Van Soolingen D, et al. Tuberculosis outbreaks predicted by characteristics of first patients in a DNA fingerprint cluster. *Am J Respir Crit Care Med*. 2008;178(1):96-104. doi:10.1164/rccm.200708-1256°C.

13. Althomsons SP, Kammerer JS, Shang N, Navin TR. Using Routinely Reported Tuberculosis Genotyping and Surveillance Data to Predict Tuberculosis Outbreaks. *PLoS One*. 2012;7(11):1-8. doi:10.1371/journal.pone.0048754.

14. Ohkado A, Nagamine M, Murase Y, et al. Molecular epidemiology of Mycobacterium tuberculosis in an urban area in Japan, 2002-2006. *Int J Tuberc Lung Dis*. 2008;12(5):548-554.

15. Ohkado A, Murase Y, Mori M, et al. Transmission of specific genotype streptomycin resistant strains of Mycobacterium tuberculosis in the Tokyo Metropolitan Area in Japan. *BMC Infect Dis*. 2009;9:138. doi:10.1186/1471-2334-9-138.

16. Xu G, Mao X, Wang J, Pan H. Clustering and recent transmission of Mycobacterium tuberculosis in a Chinese population. *Infect Drug Resist*. 2018;11:323-330.

17. Yang C, Shen X, Peng Y, et al. Transmission of Mycobacterium tuberculosis in China: a population-based molecular epidemiologic study. *Clin Infect Dis*. 2015;61(2):219-227. doi:10.1093/cid/civ255.

18. Mears J, Abubakar I, Cohen T, McHugh TD, Sonnenberg P. Effect of study design and setting on

tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR): A systematic review. *BMJ Open*. 2015;5(1). doi:10.1136/bmjopen-2014-005636.

19. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: A systematic review of population-based molecular epidemiology studies. *Int J Tuberc Lung Dis*. 2008;12(5):480-492.

20. The Research Institute of Tuberculosis J. Statistics of TB 2016. http://www.jata.or.jp/rit/ekigaku/en/statistics-of-tb/. Accessed June 20, 2018.

21. Tuberculosis Surveillance Center. *Tuberculosis in Japan – Annual Report 2017*. Tokyo, Japan: Department of Epidemiology and Clinical Research, the Research Institute of Tuberculosis; 2017.

22. Shinjuku city. Description of Shinjuku. http://www.foreign.city.shinjuku.lg.jp/en/aramashi/aramashi_1/. Accessed January 9, 2018.

23. Shinjuku City. *Tuberculosis Statistics in Shinjuku City 2017 [Japanese]*.; 2018.

24. Van Embden JDA, Cave MD, Crawford JT, et al. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: Recommendations for a standardized methodology. *J Clin Microbiol*. 1993;31(2):406-409. doi:10.1128/JCM.39.4.1683.2001.

25. Wada T, Maeda S, Hase A, Kobayashi K. Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of Mycobacterium tuberculosis in Japan. *J Med Microbiol*. 2007;56(8):1052-1057. doi:10.1099/jmm.0.46990-0.

26. Ellis B a, Crawford JT, Braden CR, McNabb SJN, Moore M, Kammerer S. Molecular epidemiology of tuberculosis in a sentinel surveillance population. *Emerg Infect Dis*. 2002;8(11):1197-1209. doi:10.3201/eid0811.020403.

27. Glynn JR, Crampin AC, Yates MD, et al. The Importance of Recent Infection with *Mycobacterium tuberculosis* in an Area with High HIV Prevalence: A Long-Term Molecular Epidemiological Study

in Northern Malawi. *J Infect Dis*. 2005;192(3):480-487. doi:10.1086/431517.

28. 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):1-16. doi:10.1371/journal.pone.0138831.

29. Sandgren A, Schepisi MS, Sotgiu G, et al. Tuberculosis transmission between foreign- and native-born populations in the EU/EEA: A systematic review. *Eur Respir J*. 2014;43(4):1159-1171. doi:10.1183/09031936.00117213.

30. MURASE Y, OHKADO A, WATANABE Y, et al. Transmission dynamics of mycobacterium tuberculosis between foreign-nationals and Japanese tuberculosis patients living in Shinjuku-City, Tokyo, Japan. *Kekkaku*. 2017;92(5):431-439.

31. Murase Y, Izumi K, Ohkado A, et al. Prediction of local transmission of mycobacterium tuberculosis isolates of a predominantly Beijing lineage by use of a variable-number tandem-repeat typing method incorporating a consensus set of hypervariable loci. *J Clin Microbiol*. 2018;56(1):1-11. doi:10.1128/JCM.01016-17.

32. Wada T, Iwamoto T, Tamaru A, et al. Clonality and micro-diversity of a nationwide spreading genotype of Mycobacterium tuberculosis in Japan. *PLoS One*. 2015;10(3):1-13. doi:10.1371/journal.pone.0118495.

33. Murase Y, Mitarai S, Sugawara I, Kato S, Maeda S. Promising loci of variable numbers of tandem repeats for typing Beijing family Mycobacterium tuberculosis. *J Med Microbiol*. 2008;57(7):873-880. doi:10.1099/jmm.0.47564-0.

34. Nikolayevskyy V, Kranzer K, Niemann S, Drobniewski F. Whole genome sequencing of Mycobacterium tuberculosis for detection of recent transmission and tracing outbreaks : A systematic review. *Tuberculosis*. 2016;98:77-85. doi:10.1016/j.tube.2016.02.009.

35. Jajou R, Neeling A De, Hunen R Van, et al. Epidemiological links between tuberculosis cases

identified twice as efficiently by whole genome sequencing than conventional molecular typing : A population-based study. *PLoS One*. 2018;13(4):e0195413.

TABLE

Table 1. Study variables and definitions

Category	Variables	Definition
Demographic factors	Sex	Men or women
	Age	Age at registration (≥40 or <40 years)
	Country of birth	Japan-born or foreign-born persons
Social factors	Occupation	Fulltime, part-time/daily worker, jobless under 60 year of age, or others (including infant, student, housewife, retired, and unknown)
	Receipt of public assistance	Those who were receiving government welfare benefits due to a household income that is below the minimum cost of living at registration
	Homeless status	Those whose legal address was unknown or unstable during the previous two or more years prior to registration
	Alcohol misuse	Those who tend to drink excessively, as judged by the public health nurses
Clinical factors	Site of disease	Those who have pulmonary or extra pulmonary disease
	Cavity lesions	Those who have cavity lesions in lung field on chest radiography
	Sputum smear microscopy	Those who exhibit positive or negative results in the sputum smear microscopy test
	Past TB history	Those with a history of past TB treatment
	Status of diabetes mellitus	Those with diabetes mellitus, as self-reported by the patient
Others	Mode of detection	Those who were identified through active case finding conducted by public health centers

Status of patient delay	A time between the onset of symptoms and the initial doctor visit longer than two months
Status of doctor delay	A time between the initial doctor visit and diagnosis longer than one month
Status of total delay	A time between the onset of symptoms and TB diagnosis longer than three months
Registration interval	The duration in months between the registration dates of the first two cases in each of the genotype clusters

434 Table 2. Factors associated with TB genotype clustering; univariable logistic regression analysis, RFLP,

435 Shinjuku, Tokyo, Japan, 2002–2013

	Total number of cases (N=1025), n	Clustered cases (N=515), n (%)		OR	(95% CI)	p-value
Age (years)	1025					
≥40	754	371	(49.2)	Reference		
<40	271	144	(53.1)	1.17	(0.88-1.56)	0.267
Sex	1025					
Female	248	102	(41.1)	Reference		
Male	777	413	(53.2)	1.62	(1.20-2.19)	0.001**
Country of birth	1025					
Foreign	95	22	(23.2)	Reference		
Japan	930	493	(53.0)	3.74	(2.25-6.44)	<0.001***
Occupation	1025					
Fulltime worker	313	165	(52.7)	1.53	(1.15-2.05)	0.004**
Part-time/daily worker	96	60	(62.5)	2.29	(1.45-3.61)	<0.001***
Jobless (aged 15-59 years)	172	103	(59.9)	2.05	(1.43-2.94)	<0.001***
Others†	444	187	(42.1)	Reference		
Public assistance‡	1024					
No	720	319	(44.3)	Reference		

Yes	304	195	(64.1)	2.25	(1.69-3.00)	<0.001***
Homelessness	1025					
No	776	349	(45.0)	Reference		
Yes	249	166	(66.7)	2.45	(1.80-3.34)	<0.001***
Alcohol misuse§	1025					
No	761	367	(48.2)	Reference		
Yes	264	148	(56.1)	1.37	(1.02-1.83)	0.028*
TB site	1024					
Extrapulmonary	80	32	(40.0)	Reference		
Pulmonary	944	482	(51.1)	1.56	(0.96-2.58)	0.058
Cavity lesions	1023					
No	565	271	(48.0)	Reference		
Yes	458	243	(53.1)	1.23	(0.95-1.58)	0.105
Smear results	1024					
Negative	406	192	(47.3)	Reference		
Positive	618	322	(52.1)	1.21	(0.94-1.57)	0.132
Past TB history	989					
New	880	441	(50.1)	Reference		
Retreatment	109	59	(54.1)	1.17	(0.77-1.79)	0.429
DM	1005					
No	832	421	(50.6)	Reference		
Yes	173	86	(49.7)	0.97	(0.69-1.36)	0.831
Active case finding	1025					
No	842	412	(48.9)	Reference		

Yes	183	103	(56.3)	1.34	(0.96-1.88)	0.071
Patient delay	1000					
<2 m	773	377	(48.8)	Reference		
≥2 m	227	127	(55.9)	1.33	(0.98-1.82)	0.057
Doctor delay	1018					
<1 m	799	415	(51.9)	Reference		
≥1 m	219	97	(44.3)	0.74	(0.54-1.00)	0.045*
Total delay	997					
<3 m	777	382	(49.2)	Reference		
≥3 m	220	122	(55.5)	1.29	(0.94-1.76)	0.099

RFLP: restriction fragment length polymorphism, OR: odds ratio, CI: confidence interval, TB: tuberculosis, DM: diabetes mellitus, *p<0.05, **p<0.01, ***p<0.001

†Others includes infant, student, housewife, retired, and unknown, and this population is considered to be as a low risk of infection.

‡Public assistance refers to government welfare benefits due to household income below the minimum cost of living.

§Alcohol misuse refers to excessive drinking, as judged by the public health nurses conducting the interviews.

Table 3. Factors associated with TB genotype clustering; multivariate logistic regression analysis, RFLP, Shinjuku, Tokyo, Japan, 2002–2013

Variables		aOR	(95% CI)	p-value
Age (years)	≥40	Reference		
	<40	1.73	(1.23-2.44)	0.002**
Country of birth	Foreign	Reference		
	Japan	3.90	(2.27-6.72)	<0.001***
Occupation	Fulltime worker	1.63	(1.17-2.27)	0.004**
	Part-time/daily worker	2.20	(1.35-3.58)	0.002**
	Jobless (aged 15-59 years)	1.32	(0.88-1.97)	0.180
	Others†	Reference		
Public assistance‡	No	Reference		
	Yes	1.81	(1.15-2.84)	0.011*
Homeless	No	Reference		
	Yes	1.63	(1.02-2.62)	0.042*
Alcohol misuse§	No	Reference		
	Yes	1.29	(0.79-2.11)	0.311
Active case finding	No	Reference		
	Yes	1.39	(0.98-1.99)	0.066

RFLP: restriction fragment length polymorphism, aOR: adjusted odds ratio, CI: confidence interval, *p<0.05, **p<0.01, ***p<0.001

†Others includes infant, student, housewife, retired, and unknown.

‡Public assistance refers to government welfare benefits due to a household income below the minimum cost of living.

§Alcohol misuse refer to excessive drinking, as judged by the public health nurses conducting the interviews.

Table 4. Factors associated with large genotype clusters within three years using the characteristics of the first two cases in each TB genotype cluster; univariable logistic regression, RFLP, Shinjuku, Tokyo, Japan, 2003–2013 (N=104 cluster episodes)

Variable	Large clusters within 3 years (N=14), n (%)†		Small clusters and large clusters after 3 years (N=90), n (%)‡		Univariate logistic regression		
					OR	(95% CI)	p Value
Sex							
No male patients	1	(7.1)	4	(4.4)	Ref		
≥One male patient	13	(92.9)	86	(95.6)	0.60	(0.06-5.84)	0.664
Age							
No patients <40 years of age	8	(57.1)	57	(63.3)	Ref		
At least one patient <40 years of age	6	(42.9)	33	(36.7)	1.30	(0.41-4.06)	0.657
Japanese							
No Japan-born patients	0	(0.0)	2	(2.2)	Ref		
≥One Japan-born patient	14	(100.0)	88	(97.8)	NA		
Full- and part-time/daily workers							
No patients with full- and part-time/daily employment	6	(42.9)	35	(38.9)	Ref		
≥One patient with full- and part-time/daily employment	8	(57.1)	55	(61.1)	0.85	(0.27-2.65)	0.778
Public assistance							
No patient receiving public assistance	5	(35.7)	41	(45.6)	Ref		

≥One patient receiving public assistance	9	(64.3)	49	(54.4)	1.51	(0.47-4.85)	0.492
Homeless							
No patient who is currently homeless	6	(42.9)	45	(50.0)	Ref		
≥One patient who is currently homeless	8	(57.1)	45	(50.0)	1.33	(0.43-4.15)	0.620
Alcohol misuse							
No patient who misuses alcohol	5	(35.7)	48	(53.3)	Ref		
≥One patient who misuses alcohol	9	(64.3)	42	(46.7)	2.06	(0.64-6.62)	0.227
Cavity lesions							
No patients with a cavity	2	(14.3)	24	(26.7)	Ref		
≥One patient with a cavity	12	(85.7)	66	(73.3)	2.18	(0.45-10.47)	0.330
Smear results							
No patient with a positive smear	1	(7.1)	12	(13.3)	Ref		
≥One patient with a positive smear	13	(92.9)	78	(86.7)	2.00	(0.24-16.71)	0.522
Past TB history							
No patient with a past history of TB	11	(78.6)	69	(76.7)	Ref		
≥One patient with a past history of TB	3	(21.4)	21	(23.3)	0.90	(0.23-3.52)	0.875
DM							
No patient with DM	9	(64.3)	57	(63.3)	Ref		
≥One patient with DM	5	(35.7)	33	(36.7)	0.96	(0.30-3.11)	0.945
Active case finding							
No patient identified through active case finding	8	(57.1)	53	(58.9)	Ref		
≥One patient identified through active case finding	6	(42.9)	37	(41.1)	1.07	(0.34-3.35)	0.902
Patient delay							

No case with patient delay	9	(64.3)	55	(61.1)	Ref		
≥One case with patient delay	5	(35.7)	35	(38.9)	0.87	(0.27-2.82)	0.820
Doctors delay							
No case with doctor delay	10	(71.4)	57	(63.3)	Ref		
≥One case with doctor delay	4	(28.6)	33	(36.7)	0.69	(0.20-2.38)	0.558
Total delay							
No case with total delay	10	(71.4)	53	(58.9)	Ref		
≥One case with total delay	4	(28.6)	37	(41.1)	0.57	(0.17-1.97)	0.376
Registration interval							
0-2 months between first two cases	7	(50.0)	13	(14.4)	9.51	(2.16-41.89)	0.003**
3-5 months between first two cases	2	(14.3)	5	(5.6)	7.07	(0.95-52.77)	0.057
6-11 months between first two cases	2	(14.3)	19	(21.1)	1.86	(0.29-12.00)	0.514
≥12 months between first two cases	3	(21.4)	53	(58.9)	Ref		

Note. After the variables for multivariate logistic regression were selected using the stepwise method, only the "registration interval" variable remained in the model. Thus, the table shows only the results of the univariate logistic regression.

RFLP: restriction fragment length polymorphism, OR: odds ratio, aOR: adjusted odds ratio, CI: confidence interval, TB: tuberculosis, DM: diabetes mellitus, Ref: reference, NA: not applicable, *p<0.05, **p<0.01

†"Large clusters within three years" refers to cluster episodes with five or more cases (large clusters) within three years.

‡"Small clusters and large clusters after three years" refers to cluster episodes with two to four cases (small clusters) and cluster episodes that became large clusters after three years.

470

Figure legends

Figure 1. Number of reported cases of TB, including culture-positive cases, strain-typed cases and genotype clusters, in Shinjuku during 2002–2013

RFLP: restriction fragment length polymorphism, TB: tuberculosis

Figure 2. Cumulative clustering rate (RFLP, Shinjuku 2002–2013)

RFLP: restriction fragment length polymorphism

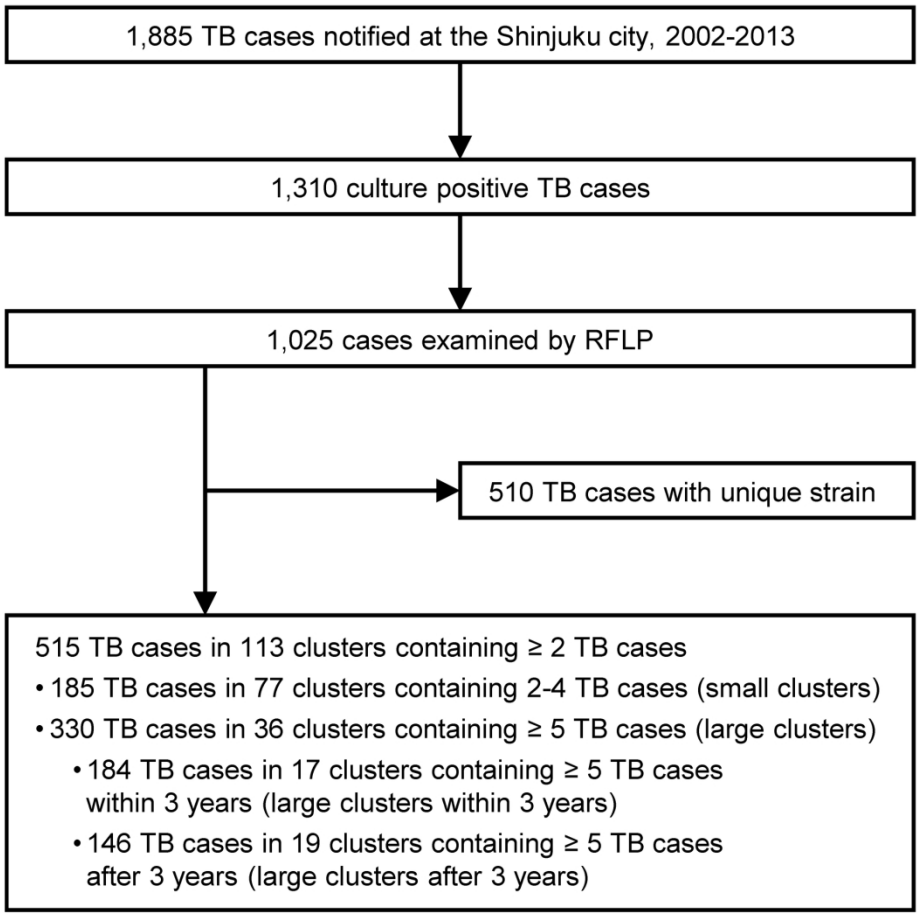

Figure 1. Number of reported cases of TB, including culture-positive cases, strain-typed cases and genotype clusters, Shinjuku, 2002–2013
RFLP: restriction fragment length polymorphism, TB: tuberculosis

90x90mm (600 x 600 DPI)

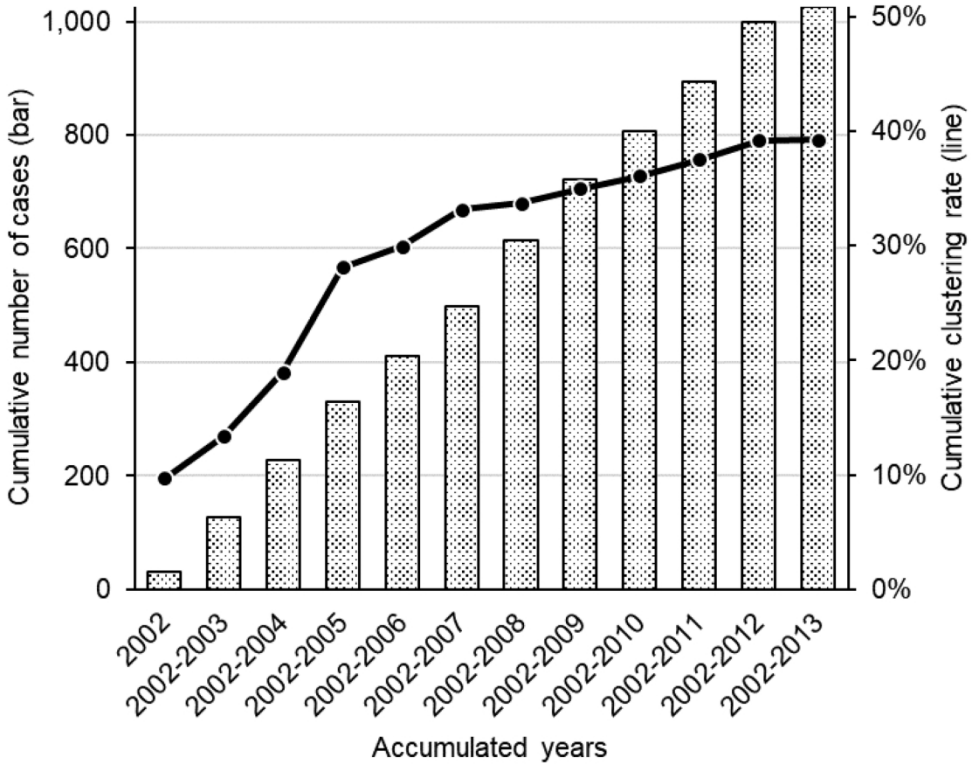

Figure 2. Cumulative clustering rate, RFLP, Shinjuku 2002–2013
RFLP: restriction fragment length polymorphism

90x90mm (600 x 600 DPI)

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Table 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	6 **We took all eligible study participants in Shinjuku city.*
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Table 1
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-8
		(*b*) Describe any methods used to examine subgroups and interactions	8
		(*c*) Explain how missing data were addressed	8
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	Not Applicable

		(*e*) Describe any sensitivity analyses	Not Applicable
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8-9 Figure 1
		(b) Give reasons for non-participation at each stage	8-9
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	N in tables
Outcome data	15*	Report numbers of outcome events or summary measures	Tables, 9-10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Tables
		(*b*) Report category boundaries when continuous variables were categorized	Tables
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not Applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Table 4
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-14
Generalisability	21	Discuss the generalisability (external validity) of the study results	11-14
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	4

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in Tokyo, Japan: a population-based molecular epidemiological study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029295.R2
Article Type:	Research
Date Submitted by the Author:	04-Apr-2019
Complete List of Authors:	Izumi, Kiyohiko; The Research Institute of Tuberculosis, Murase, Yoshiro; The Research Institute of Tuberculosis Uchimura, Kazuhiro; The Research Institute of Tuberculosis Kaebeta, Aya; Shinjuku Public Health Center Ishihara, Keiko; Shinjuku Public Health Center Kaguraoka, Sumi; Shinjuku Public Health Center Takii, Takemasa; The Research Institute of Tuberculosis Ohkado, Akihiro; The Research Institute of Tuberculosis
Primary Subject Heading:	Infectious diseases
Secondary Subject Heading:	Epidemiology, Public health
Keywords:	RFLP, clustering rate, homeless, foreign-born

Title

Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in Tokyo, Japan: a population-based molecular epidemiological study

Authors and affiliations

1. Kiyohiko Izumi. Department of Epidemiology and Clinical Research, The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: kizumi95@gmail.com
2. Yoshiro Murase. Department of Mycobacterium Reference and Research, The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: ymurase@jata.or.jp
3. Kazuhiro Uchimura. Department of Epidemiology and Clinical Research, The Research Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: uchimura@jata.or.jp
4. Aya Kaebeta. Shinjuku Public Health Center. Email: aya.kaebeta@city.shinjuku.lg.jp
5. Keiko Ishihara. Shinjuku Public Health Center. Email: keiko.ishihara@city.shinjuku.lg.jp
6. Sumi Kaguraoka. Shinjuku Public Health Center. Email: sumi.kaguraoka@city.shinjuku.lg.jp
7. Takemasa Takii. The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo, Japan. Email: t-takii@jata.or.jp
8. Akihiro Ohkado. Department of Epidemiology and Clinical Research, The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo, Japan. Graduate School of Biomedical Sciences, Nagasaki University. Email: ohkadoa@jata.or.jp

Corresponding Author

Kiyohiko Izumi

Department of Epidemiology and Clinical Research, Research Institute of Tuberculosis, Japan Anti-

Tuberculosis Association, 3-1-24 Matsuyama, Kiyose, Tokyo 204-8533, JAPAN

Tel: ±81-42-4935711

Fax: +81-42-492-4600

Email: kizumi95@gmail.com

KEY WORDS: RFLP, clustering rate, homeless, foreign-born

ABSTRACT

Objective: Molecular epidemiology is a promising tool for understanding tuberculosis transmission dynamics but has not been sufficiently utilized in Asian countries including Japan. The aim of this study was to estimate the proportion of TB cases attributable to recent transmission and to identify risk factors of genotype clustering and the development of large clusters within three years in an urban setting in Japan.

Design and setting: Long-term cross-sectional observational study combining the characteristics of culture-positive TB patients notified in Shinjuku City, Tokyo (2002-2013), with genotype data of *Mycobacterium tuberculosis.*

Primary outcome measure: Genotype clustering rate and association between genotype clustering status and explanatory variables.

Results: Among 1,025 cases, 515 were localized within 113 genotype clusters. The overall clustering rate was 39.2%. Significantly higher rates were found in patients aged <40 years (adjusted odds ratio (aOR)=1.73, 95% CI=1.23-2.44), native Japanese individuals (aOR=3.90, 95% CI=2.27-6.72), fulltime workers (aOR=1.63, 95% CI=1.17-2.27), part-time/daily workers (aOR=2.20, 95% CI=1.35-3.58), individuals receiving public assistance (aOR=1.81, 95% CI=1.15-2.84), and homeless people (aOR=1.63, 95% CI=1.02-2.62). A significant predictor of large genotype clusters within three years was a registration interval ≤2 months between the first two cases in a cluster.

Conclusion: Our results indicated that a large proportion of culture-positive TB patients were involved in the recent TB transmission chain. Foreign-born persons still have a limited impact on transmission in the Japanese urban setting. Intensified public health interventions, including the active case finding, need to focus on individuals with socioeconomic risk factors that are significantly associated with tuberculosis transmission and clusters with shorter registration intervals between the first two cases.

Word Count in Abstract= 260 words

Article Summary

Strengths and limitations of this study

- This study is one of the longest population-based studies focusing on the molecular epidemiology of culture-positive notified tuberculosis patients in a large Asian urban setting.
- Interviews conducted by the experienced public health nurses at the Public Health Center using a standardized questionnaire provided high-quality data and less interviewer bias.
- We may have underestimated genotype clustering due to the large population flow in and out of the city.

Funding statement

This research was supported by the Research Program on Emerging and Re-emerging Infectious Diseases from the Japan Agency for Medical Research and Development (AMED No. JP18fk0108041).

Competing interests: None declared.

Patient consent: Not required.

Data sharing statement: Due to data restrictions, we are unable to share any aspect of the data.

Word count: 2,774 words

INTRODUCTION

Tuberculosis (TB) remains a major public health threat worldwide. In 2017, an estimated 10 million people worldwide developed TB, and 1.27 million died from TB[1]. Although the majority of cases have been reported in countries with a high TB burden, TB remains a persistent health problem in low- and medium-burden countries because it is concentrated in specific vulnerable and hard-to-reach populations, such as homeless people and foreign-born persons from TB high-burden countries[2]. These specific high-risk populations tend to live in large cities where they are seeking jobs, which potentially poses challenges to the control of TB in urban areas[3,4]. Many countries with a low or medium TB burden have recently adopted TB elimination strategies[2,5], which emphasizes the importance of molecular epidemiology in TB control, particularly in urban areas[2,4].

TB molecular genotyping using restriction fragment length polymorphisms (RFLPs) and, more recently, variable numbers of tandem repeats (VNTRs) combined with epidemiological information identifies TB cases that are likely involved in the same transmission chain[6]. This method differentiates recent transmission or endogenous reactivation from remote infection and has therefore revealed that a substantial proportion of TB cases are due to recent transmission in low-TB-burden countries[7–9]. This method also identifies the proportion of cases attributable to recent transmission and determines the risk factors for transmission. Moreover, various factors predicting large TB genotype clusters, including socially vulnerable populations and shorter intervals between the registration dates of the first two cases, have been investigated by evaluating the characteristics of the first two cases in the same genotype cluster[10–13]. These population-based molecular epidemiological studies were conducted in some European countries[8,10,12], the U.S.[7,9,11] and some Asian countries[14–19].

In Japan, a country with a medium TB burden, the number of newly notified TB cases decreased from

32,828 (25.8 per 100,000 populations) in 2002 to 17,625 (13.9 per 100,000 populations) in 2016[20], but the central government has constantly been reported of TB outbreaks by local governments at a rate of approximately 40 events annually over the last decade. This information suggests that TB transmission might be occurring in some groups, such as homeless people, who constitute a high-risk group for recent TB transmission in urban areas[14]. Considering the steady increase in the proportion of TB cases among foreign-born individuals in Japan (7.9% of all cases in 2016[21]), transmission between foreign-born persons and local residents must be monitored. In addition, in light of Japan's transition toward becoming a low-TB-burden country, understanding TB transmission patterns has become increasingly important. However, few population-based molecular epidemiological studies have identified the transmission patterns in Japan and their risk factors. Additionally, no study has attempted to evaluate the factors predicting the development of large clusters in Japan.

Therefore, we aimed to estimate the proportion of TB cases attributable to recent transmission, to identify the risk factors for recent transmission, and to predict the risk factors for the development of large clusters in an urban setting.

METHODS

Study population

We included all culture-positive TB patients notified in Shinjuku City from September 2002 to December 2013 as the eligible study population in this cross-sectional observational study. This study forms part of a population-based study on DNA fingerprinting surveillance of *M. tuberculosis* in Shinjuku City that was started in 2002. Shinjuku City (18.3 km^2) is one of the most populous (342,867 residents in 2018[22]) cities in Tokyo, and its TB notification rate in 2016 was 33.7 per 100,000 people[23],

which was higher than the rates in Tokyo and the nation (17.2 and 13.9, respectively[20]). Experienced public health nurses at the Shinjuku Public Health Center (PHC) interviewed and collected information from all culture-positive TB patients at the time of registration using a standardized questionnaire to avoid possible interviewer bias. The study variables and definitions are described in Table 1.

Patient and public involvement

Neither the patients nor the public were involved in the design of this study.

DNA fingerprinting and genotype cluster

Clinical isolates from each of the enrolled TB patients were sent to the Research Institute of Tuberculosis (RIT), Tokyo, where the TB strains were subjected to DNA fingerprinting using insertion sequence *6110* by RFLP (IS*6110*-RFLP) analysis[24]. One clinical isolate per person was used for the clustering analysis. IS*6110*-RFLP and spoligotyping are the standard methods used in the Shinjuku PHC and were available throughout the study period. The Shinjuku PHC switched from RFLP to VNTR a few years ago, but the RFLP profiles of many TB cases were available. Thus, we employed RFLP due to the sufficient sample size. A genotype cluster was defined as a group of TB patients whose isolates showed either (1) ≥6 identical IS*6110* band patterns or (2) <6 identical IS*6110* band patterns confirmed by identical spoligotyping patterns. The data collection and genotyping methods were previously described in detail[14].

Data analysis

We calculated the genotype clustering rate by the "n-1 method" according to the formula *{(n - c)/N}*, where *N* is the total number of cases sampled, *c* is the number of clusters, and *n* is the total number

of cases in the clusters[9]. We also calculated the cumulative clustering rate by calculating the clustering rate in 2002 and then adding the TB patients every year up to 2013. The characteristics of clustered cases, which were the cases belonging to any genotype clusters, were compared with those with unique strain patterns through χ^2 tests. We performed univariate logistic regression to identify risk factors for genotype clustering using odds ratios (ORs) and multivariate logistic regression using adjusted ORs (aORs). Any potential interactions were assessed using likelihood ratio tests.

Additionally, we compared the characteristics of the first two cases in each genotype cluster to identify risk factors for the development of a large cluster within three years. For this purpose, a cluster episode was defined as a newly arising genotype cluster in or after 2003 without any TB cases of that genotype notified prior to that year. We classified cluster episodes into the following two groups according to a system developed in a previous study[10]: (1) "large clusters within three years" were cluster episodes with five or more cases (large clusters) occurring within three years and (2) "small clusters and large clusters after three years" were cluster episodes with two to four cases (small clusters) and cluster episodes that became large clusters after three years. We identified the first two cases in each cluster episode based on the notification date and compared their characteristics between these two groups. We performed univariate and multivariate logistic regression analyses to identify predictors of the development of large clusters within three years.

A *p*-value of 0.05 was set as the level indicating statistical significance. For variables with more than 5% missing values, the multiple imputation method was considered. The variables used for multivariate logistic regressions were selected by the stepwise maximum-likelihood estimation with a significance level of less than 0.2. We used Stata version 12 (Stata Corp., College Station, TX, USA) for the statistical analyses. Written informed consent was waived because DNA fingerprinting

analysis forms part of the routine TB control activities in Shinjuku City. However, oral informed consent was obtained after the PHC staff provided a thorough explanation of the study objectives and confidentiality. The study protocol was approved by the Institutional Review Board of the Research Institute of Tuberculosis (RIT/IRB27-9).

RESULTS

Study population and clustering rate

In total, 1,885 TB patients in Shinjuku City were notified during the study period, and 1,310 were culture-positive cases (Figure 1). Of these, 285 patients were excluded from the analysis, mainly due to the unavailability of culture-positive isolates and the lack of implementation of RFLP. As a result, 1,025 (78.2%) patients were included in the analysis. Figure 2 shows the cumulative number of TB patients and the clustering rates from 2002 to 2013. The number of TB cases gradually increased over the tested decade. In contrast, the cumulative clustering rates sharply increased in the first four years, from 10% in 2002 to 28% in 2005, with an average percent change of +43%, and then continued to increase at a slower rate, from 30% in 2006 to 39% in 2013, with an average percent change of +4.2%.

We identified a total of 113 genotype clusters consisting of 515 patients (Figure 1). The genotype clustering rate was 39.2%, and the average cluster size was 4.56 cases (range 2-30). Fifty-seven (50.4%) genotype clusters consisted of only two TB patients, and 36 (31.9%) genotype clusters had at least five TB patients. We further investigated the homelessness status and place of birth of the patients in the genotype clusters. Of the 113 genotype clusters, 45 (39.8%) comprised only nonhomeless individuals, seven (6.2%) included only homeless individuals, and 61 (54.0%) contained both homeless and nonhomeless individuals (mixed cluster). We compared the characteristics of the

nonhomeless patients in the clusters of only nonhomeless patients with those in the mixed clusters, and although the finding was not statistically significant (Pearson χ^2 test, p=0.17), the proportion of nonhomeless patients receiving public assistance in the latter group (13.8%) was higher than that in the former group (8.8%). No differences in sex, age, and place of birth were found between the two groups. Of the 113 genotype clusters, 94 (83.2%) consisted of only individuals born in Japan, two (1.8%) consisted of only foreign-born individuals, and 17 (15.0%) consisted of both individuals born in Japan and foreign-born individuals.

Factors associated with genotype clustering

The clustered cases were significantly more likely to consist of male individuals (OR=1.62, 95% CI=1.20-2.19), Japan-born individuals (OR=3.74, 95% CI=2.25-6.44), individuals receiving public assistance (OR=2.25, 95% CI=1.69-3.00), homeless individuals (OR=2.45, 95% CI=1.80-3.34), individuals who misuse alcohol (OR=1.37, 95% CI=1.02-1.83), individuals engaging in fulltime work (OR=1.53, 95% CI=1.15-2.05) and part-time/daily work (OR=2.29, 95% CI=1.45-3.61), and jobless individuals aged 15-59 years (OR=2.05, 95% CI=1.43-2.94) (Table 2). A significant interaction among the explanatory variables was not detected. The multivariate analysis demonstrated that the factors associated with genotype clustering were age <40 years (aOR=1.73, 95% CI=1.23-2.44), born in Japan (aOR=3.90, 95% CI=2.27-6.72), working fulltime (aOR=1.63, 95% CI=1.17-2.27), having part-time/daily work (aOR=2.20, 95% CI=1.35-3.58), receiving public assistance (aOR=1.81, 95% CI=1.15-2.84), and homelessness (aOR=1.63, 95% CI=1.02-2.62)(Table 3).

Factors associated with large genotype clustering within three years

We identified 104 genotype cluster episodes according to the definition. Of these, 14 were “large clusters within three years”, which was equivalent to 13.5% (14/104) of all the genotype clusters and

48.3% (14/29) of the large genotype clusters, and 90 clusters were “small clusters and large clusters after three years”. The univariate analysis indicated that clusters with registration intervals of 0-2 months were 9.51 times more likely to become large genotype clusters within three years compared with clusters with registration intervals of ≥12 months (Table 4). After selecting variables using the stepwise method, only the “registration interval” variable remained for the multivariate model.

DISCUSSION

In this long-term population-based study, we included 1,025 patients, identified a total of 113 genotype clusters, and obtained a genotype clustering rate of 39.2%. Our results indicated that the clustered cases were more likely to have certain socioeconomic predictive factors, namely, being homeless, receiving public assistance, and having an unstable job, at the time of tuberculosis diagnosis. A shorter registration interval between the first two cases was a statistically significant predictor of the development of a large genotype cluster within three years.

Clustering rate

We identified 515 genotype clustered cases and estimated a clustering rate of 39.2%. The rate was the same as the pooled clustering rate (40.9%) obtained in a previous meta-analysis of population-based studies of countries with a low TB incidence[19] but differed from previous estimates obtained in Japanese studies, which were 27.6% in Shinjuku and 24.6% in Osaka[14,25]. Because the meta-regression analysis clarified that longer study durations are associated with an increased clustering rate[19], this difference could be due to shorter study durations combined with the smaller sample sizes of the previous studies (388 patients in five years and 195 patients in one year, respectively). In our study, as expected, the cumulative clustering rate rapidly increased in the first four years and increased more slowly thereafter, which is similar to the trend observed in the previous studies[26,27].

Factors associated with genotype clustering

Our results indicated that the clustered cases were more likely to have socioeconomic predictive factors, namely, being homeless, receiving public assistance, and having an unstable job, at the time of TB diagnosis. Similarly, previous studies suggested that being homeless significantly contributed to clustering in Shinjuku City[14] and other counties[19]. In our study, more than half of the genotype clusters were mixtures of nonhomeless and homeless patients. Moreover, the nonhomeless patients in the mixed clusters tended to be financially unstable, and a higher proportion of these patients were receiving public assistance compared with the proportion among clusters of only nonhomeless cases, which could imply that relatively poor nonhomeless patients share activity spaces with homeless patients, such as urban areas around the large train stations that were reported to be significant hotspots for homeless patients in Shinjuku City[28]. These findings could suggest that contact investigations of homeless TB patients need to be actively expanded to possible contact persons who are not homeless, particularly those who are facing financial difficulty.

A meta-analysis based on studies conducted in European countries where foreign-born patients substantially contribute to TB epidemiology found that the proportion of mixed clusters composed of native and foreign-born patients ranged from 0% to 36.5% and concluded that foreign-born patients did not have a significant influence on TB in the native population[29]. In our study, the proportion of mixed clusters (15.0%) fell into this range. Thus, the impact of TB transmission between native and foreign-born populations likely remains limited in this urban setting[30]. However, considering the recent increase in immigrant TB patients in urban cities, TB transmission between native and foreign-born populations needs to be closely monitored.

Factors associated with large genotype clustering within three years

A shorter registration interval (≤2 months) was identified as a significant predictor of the development of a large genotype cluster within three years, which is compatible with findings of previous studies conducted in the Netherlands and London[10,12]. Therefore, when TB patients with identical genotypes have shorter registration intervals, a thorough active case findings need to be performed to investigate the potential infection sources and infected patients in order to prevent further transmission. However, it is difficult to assume that the first patient infected the second patient because a window of two months appears too short. Thus, we believe that a true but unidentified first TB case was not identified in our study. A cluster episode was defined as a cluster without any TB patients in 2002 and at least two patients with identical genotypes in and after 2003. Therefore, a possible true first TB case might have been registered before 2002, which was outside of our study period, or registered outside of Shinjuku City.

Limitations

Our study has some limitations. First, the study population consisted only of TB patients living in Shinjuku City. Considering the large population flow in and out of the city, as mentioned above, we potentially missed patients living outside of the city who shared TB strain types with patients living in the city. In fact, previous Japanese studies reported clusters with TB patients living across broad geographic areas[31]. Consequently, we may have underestimated the identified genotype clusters. Second, even the existence of TB patients with identical genotyping patterns may not suggest recent transmission if the strain is a nationwide endemic TB strain[32], which could have led to an overestimated clustering rate. Third, IS*6110* RFLP has relatively lower discriminatory power compared with VNTR[33] and whole-genome sequencing[34,35], which might have led to overestimation. Lastly, information of epidemiological linkage among TB patients was not available in our study.

Therefore, we could not assess and discuss the current practices involving epidemiological investigations done by the public health center, which could weaken the programmatic implications of our results.

Conclusion

This study constitutes a one of the longest-term studies on the molecular epidemiology of notified TB patients in a large Asian urban setting. Our results indicated that a large proportion of culture-positive TB patients were involved in the recent TB transmission chain. Homeless persons were found to be involved in more than half of the genotype clusters. Foreign-born persons continue to have a limited impact on TB transmission in the Japanese urban setting, but considering recent increases in foreign-born TB patients, transmission between native and foreign-born populations should be routinely evaluated. Intensified public health interventions, such as active case findings, should focus on those with socioeconomic risk factors that are significantly associated with TB transmission and clusters with shorter registration intervals between the first two cases because these variables could serve as predictors of the development of large clusters within three years.

ACKNOWLEDGMENTS

We would like to thank the Shinjuku Public Health Center staff for collecting the data.

Author Statement: Authors roles

KIz: conception of study, statistical analysis, interpretation of data, and drafting and finalization of the manuscript

YM: genotyping analysis, interpretation of data, and finalization of the manuscript

KU: conception of study, interpretation of data, and finalization of the manuscript

AK: conception of study, collection of data, and finalization of the manuscript

KIs: conception of study, collection of data, and finalization of the manuscript

SK: conception of study, collection of data, and finalization of the manuscript

TT: interpretation of data and finalization of the manuscript

AO: conception of study, interpretation of data, and drafting and finalization of the manuscript

References

1. World Health Organization. *Global Tuberculosis Report 2018*. Geneva; 2018. doi:ISBN 978 92 4 156539 4.
2. Lönnroth K, Migliori GB, Abubakar I, et al. Towards tuberculosis elimination: An action framework for low-incidence countries. *Eur Respir J*. 2015;45(4):928-952. doi:10.1183/09031936.00214014.
3. de Vries G, Aldridge RW, Cayla J a, et al. Epidemiology of tuberculosis in big cities of the European Union and European Economic Area countries. *Euro Surveill Bull Eur sur les Mal Transm = Eur Commun Dis Bull*. 2014;19(9):1-8.
4. Hest NA, Aldridge RW, de Vries G, et al. Tuberculosis control in big cities and urban risk groups in the European Union: A consensus statement. *Euro Surveill*. 2014;19(9):1-13. doi:10.2807/1560-7917.ES2014.19.9.20728.
5. Jereb JA. Progressing towards tuberculosis elimination in low-incidence areas of the United States. Recommendations of the Advisory Council for the Elimination of Tuberculosis. *MMWR Recomm reports*. 2002;51(RR-5):1-14.
6. Borgdorff MW, van Soolingen D. The re-emergence of tuberculosis: What have we learnt from molecular epidemiology? *Clin Microbiol Infect*. 2013;19(10):889-901. doi:10.1111/1469-0691.12253.
7. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. 1994;330(D):1710-1716.
8. van Soolingen D, Borgdorff MW, de Haas PE, et al. Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis*. 1999;180:726-736. doi:10.1086/314930.
9. PM S, PC H, SP S, et al. The epidemiology of tuberculosis in San Francisco. A population-based

study using conventional and molecular methods. *N Engl J Med*. 1994.

10. Hamblion EL, Le Menach A, Anderson LF, et al. Recent TB transmission, clustering and predictors of large clusters in London, 2010-2012: Results from first 3-years of universal MIRU-VNTR strain typing. *Thorax*. 2016;71(8):749-756. doi:10.1136/thoraxjnl-2014-206608.

11. Driver CR, Macaraig M, McElroy PD, et al. Which patients' factors predict the rate of growth of Mycobacterium tuberculosis clusters in an urban community? *Am J Epidemiol*. 2006;164(1):21-31. doi:10.1093/aje/kwj153.

12. Kik S V., Verver S, Van Soolingen D, et al. Tuberculosis outbreaks predicted by characteristics of first patients in a DNA fingerprint cluster. *Am J Respir Crit Care Med*. 2008;178(1):96-104. doi:10.1164/rccm.200708-1256°C.

13. Althomsons SP, Kammerer JS, Shang N, Navin TR. Using Routinely Reported Tuberculosis Genotyping and Surveillance Data to Predict Tuberculosis Outbreaks. *PLoS One*. 2012;7(11):1-8. doi:10.1371/journal.pone.0048754.

14. Ohkado A, Nagamine M, Murase Y, et al. Molecular epidemiology of Mycobacterium tuberculosis in an urban area in Japan, 2002-2006. *Int J Tuberc Lung Dis*. 2008;12(5):548-554.

15. Ohkado A, Murase Y, Mori M, et al. Transmission of specific genotype streptomycin resistant strains of Mycobacterium tuberculosis in the Tokyo Metropolitan Area in Japan. *BMC Infect Dis*. 2009;9:138. doi:10.1186/1471-2334-9-138.

16. Xu G, Mao X, Wang J, Pan H. Clustering and recent transmission of Mycobacterium tuberculosis in a Chinese population. *Infect Drug Resist*. 2018;11:323-330.

17. Yang C, Shen X, Peng Y, et al. Transmission of Mycobacterium tuberculosis in China: a population-based molecular epidemiologic study. *Clin Infect Dis*. 2015;61(2):219-227. doi:10.1093/cid/civ255.

18. Mears J, Abubakar I, Cohen T, McHugh TD, Sonnenberg P. Effect of study design and setting on

tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR): A systematic review. *BMJ Open*. 2015;5(1). doi:10.1136/bmjopen-2014-005636.

19. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: A systematic review of population-based molecular epidemiology studies. *Int J Tuberc Lung Dis*. 2008;12(5):480-492.

20. The Research Institute of Tuberculosis J. Statistics of TB 2016. http://www.jata.or.jp/rit/ekigaku/en/statistics-of-tb/. Accessed June 20, 2018.

21. Tuberculosis Surveillance Center. *Tuberculosis in Japan – Annual Report 2017*. Tokyo, Japan: Department of Epidemiology and Clinical Research, the Research Institute of Tuberculosis; 2017.

22. Shinjuku city. Description of Shinjuku. http://www.foreign.city.shinjuku.lg.jp/en/aramashi/aramashi_1/. Accessed January 9, 2018.

23. Shinjuku City. *Tuberculosis Statistics in Shinjuku City 2017 [Japanese]*.; 2018.

24. Van Embden JDA, Cave MD, Crawford JT, et al. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: Recommendations for a standardized methodology. *J Clin Microbiol*. 1993;31(2):406-409. doi:10.1128/JCM.39.4.1683.2001.

25. Wada T, Maeda S, Hase A, Kobayashi K. Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of Mycobacterium tuberculosis in Japan. *J Med Microbiol*. 2007;56(8):1052-1057. doi:10.1099/jmm.0.46990-0.

26. Ellis B a, Crawford JT, Braden CR, McNabb SJN, Moore M, Kammerer S. Molecular epidemiology of tuberculosis in a sentinel surveillance population. *Emerg Infect Dis*. 2002;8(11):1197-1209. doi:10.3201/eid0811.020403.

27. Glynn JR, Crampin AC, Yates MD, et al. The Importance of Recent Infection with *Mycobacterium tuberculosis* in an Area with High HIV Prevalence: A Long-Term Molecular Epidemiological Study

in Northern Malawi. *J Infect Dis*. 2005;192(3):480-487. doi:10.1086/431517.

28. 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):1-16. doi:10.1371/journal.pone.0138831.

29. Sandgren A, Schepisi MS, Sotgiu G, et al. Tuberculosis transmission between foreign- and native-born populations in the EU/EEA: A systematic review. *Eur Respir J*. 2014;43(4):1159-1171. doi:10.1183/09031936.00117213.

30. MURASE Y, OHKADO A, WATANABE Y, et al. Transmission dynamics of mycobacterium tuberculosis between foreign-nationals and Japanese tuberculosis patients living in Shinjuku-City, Tokyo, Japan. *Kekkaku*. 2017;92(5):431-439.

31. Murase Y, Izumi K, Ohkado A, et al. Prediction of local transmission of mycobacterium tuberculosis isolates of a predominantly Beijing lineage by use of a variable-number tandem-repeat typing method incorporating a consensus set of hypervariable loci. *J Clin Microbiol*. 2018;56(1):1-11. doi:10.1128/JCM.01016-17.

32. Wada T, Iwamoto T, Tamaru A, et al. Clonality and micro-diversity of a nationwide spreading genotype of Mycobacterium tuberculosis in Japan. *PLoS One*. 2015;10(3):1-13. doi:10.1371/journal.pone.0118495.

33. Murase Y, Mitarai S, Sugawara I, Kato S, Maeda S. Promising loci of variable numbers of tandem repeats for typing Beijing family Mycobacterium tuberculosis. *J Med Microbiol*. 2008;57(7):873-880. doi:10.1099/jmm.0.47564-0.

34. Nikolayevskyy V, Kranzer K, Niemann S, Drobniewski F. Whole genome sequencing of Mycobacterium tuberculosis for detection of recent transmission and tracing outbreaks : A systematic review. *Tuberculosis*. 2016;98:77-85. doi:10.1016/j.tube.2016.02.009.

35. Jajou R, Neeling A De, Hunen R Van, et al. Epidemiological links between tuberculosis cases

identified twice as efficiently by whole genome sequencing than conventional molecular typing : A population-based study. *PLoS One*. 2018;13(4):e0195413.

TABLE

Table 1. Study variables and definitions

Category	Variables	Definition
Demographic factors	Sex	Men or women
	Age	Age at registration (≥40 or <40 years)
	Country of birth	Japan-born or foreign-born persons
Social factors	Occupation	Fulltime, part-time/daily worker, jobless under 60 year of age, or others (including infant, student, housewife, retired, and unknown)
	Receipt of public assistance	Those who were receiving government welfare benefits due to a household income that is below the minimum cost of living at registration
	Homeless status	Those whose legal address was unknown or unstable during the previous two or more years prior to registration
	Alcohol misuse	Those who tend to drink excessively, as judged by the public health nurses
Clinical factors	Site of disease	Those who have pulmonary or extra pulmonary disease
	Cavity lesions	Those who have cavity lesions in lung field on chest radiography
	Sputum smear microscopy	Those who exhibit positive or negative results in the sputum smear microscopy test
	Past TB history	Those with a history of past TB treatment
	Status of diabetes mellitus	Those with diabetes mellitus, as self-reported by the patient
Others	Mode of detection	Those who were identified through active case finding conducted by public health centers

Status of patient delay	A time between the onset of symptoms and the initial doctor visit longer than two months
Status of doctor delay	A time between the initial doctor visit and diagnosis longer than one month
Status of total delay	A time between the onset of symptoms and TB diagnosis longer than three months
Registration interval	The duration in months between the registration dates of the first two cases in each of the genotype clusters

433 Table 2. Factors associated with TB genotype clustering; univariable logistic regression analysis, RFLP,

434 Shinjuku, Tokyo, Japan, 2002–2013

	Total number of cases (N=1025), n	Clustered cases (N=515), n (%)		OR	(95% CI)	p-value
Age (years)	1025					
≥40	754	371	(49.2)	Reference		
<40	271	144	(53.1)	1.17	(0.88-1.56)	0.267
Sex	1025					
Female	248	102	(41.1)	Reference		
Male	777	413	(53.2)	1.62	(1.20-2.19)	0.001**
Country of birth	1025					
Foreign	95	22	(23.2)	Reference		
Japan	930	493	(53.0)	3.74	(2.25-6.44)	<0.001***
Occupation	1025					
Fulltime worker	313	165	(52.7)	1.53	(1.15-2.05)	0.004**
Part-time/daily worker	96	60	(62.5)	2.29	(1.45-3.61)	<0.001***
Jobless (aged 15-59 years)	172	103	(59.9)	2.05	(1.43-2.94)	<0.001***
Others†	444	187	(42.1)	Reference		
Public assistance‡	1024					
No	720	319	(44.3)	Reference		

Yes	304	195	(64.1)	2.25	(1.69-3.00)	<0.001***
Homelessness	1025					
No	776	349	(45.0)	Reference		
Yes	249	166	(66.7)	2.45	(1.80-3.34)	<0.001***
Alcohol misuse§	1025					
No	761	367	(48.2)	Reference		
Yes	264	148	(56.1)	1.37	(1.02-1.83)	0.028*
TB site	1024					
Extrapulmonary	80	32	(40.0)	Reference		
Pulmonary	944	482	(51.1)	1.56	(0.96-2.58)	0.058
Cavity lesions	1023					
No	565	271	(48.0)	Reference		
Yes	458	243	(53.1)	1.23	(0.95-1.58)	0.105
Smear results	1024					
Negative	406	192	(47.3)	Reference		
Positive	618	322	(52.1)	1.21	(0.94-1.57)	0.132
Past TB history	989					
New	880	441	(50.1)	Reference		
Retreatment	109	59	(54.1)	1.17	(0.77-1.79)	0.429
DM	1005					
No	832	421	(50.6)	Reference		
Yes	173	86	(49.7)	0.97	(0.69-1.36)	0.831
Active case finding	1025					
No	842	412	(48.9)	Reference		

Yes	183	103	(56.3)	1.34	(0.96-1.88)	0.071
Patient delay	1000					
<2 m	773	377	(48.8)	Reference		
≥2 m	227	127	(55.9)	1.33	(0.98-1.82)	0.057
Doctor delay	1018					
<1 m	799	415	(51.9)	Reference		
≥1 m	219	97	(44.3)	0.74	(0.54-1.00)	0.045*
Total delay	997					
<3 m	777	382	(49.2)	Reference		
≥3 m	220	122	(55.5)	1.29	(0.94-1.76)	0.099

RFLP: restriction fragment length polymorphism, OR: odds ratio, CI: confidence interval, TB: tuberculosis, DM: diabetes mellitus, *p<0.05, **p<0.01, ***p<0.001

†Others includes infant, student, housewife, retired, and unknown, and this population is considered to be as a low risk of infection.

‡Public assistance refers to government welfare benefits due to household income below the minimum cost of living.

§Alcohol misuse refers to excessive drinking, as judged by the public health nurses conducting the interviews.

Table 3. Factors associated with TB genotype clustering; multivariate logistic regression analysis, RFLP, Shinjuku, Tokyo, Japan, 2002–2013

Variables		aOR	(95% CI)	p-value
Age (years)	≥40	Reference		
	<40	1.73	(1.23-2.44)	0.002**
Country of birth	Foreign	Reference		
	Japan	3.90	(2.27-6.72)	<0.001***
Occupation	Fulltime worker	1.63	(1.17-2.27)	0.004**
	Part-time/daily worker	2.20	(1.35-3.58)	0.002**
	Jobless (aged 15-59 years)	1.32	(0.88-1.97)	0.180
	Others†	Reference		
Public assistance‡	No	Reference		
	Yes	1.81	(1.15-2.84)	0.011*
Homeless	No	Reference		
	Yes	1.63	(1.02-2.62)	0.042*
Alcohol misuse§	No	Reference		
	Yes	1.29	(0.79-2.11)	0.311
Active case finding	No	Reference		
	Yes	1.39	(0.98-1.99)	0.066

RFLP: restriction fragment length polymorphism, aOR: adjusted odds ratio, CI: confidence interval, *p<0.05, **p<0.01, ***p<0.001

†Others includes infant, student, housewife, retired, and unknown.

‡Public assistance refers to government welfare benefits due to a household income below the minimum cost of living.

§Alcohol misuse refer to excessive drinking, as judged by the public health nurses conducting the interviews.

Table 4. Factors associated with large genotype clusters within three years using the characteristics of the first two cases in each TB genotype cluster; univariable logistic regression, RFLP, Shinjuku, Tokyo, Japan, 2003–2013 (N=104 cluster episodes)

Variable	Large clusters within 3 years (N=14), n (%)†		Small clusters and large clusters after 3 years (N=90), n (%)‡		Univariate logistic regression OR	(95% CI)	p Value
Sex							
No male patients	1	(7.1)	4	(4.4)	Ref		
≥One male patient	13	(92.9)	86	(95.6)	0.60	(0.06-5.84)	0.664
Age							
No patients <40 years of age	8	(57.1)	57	(63.3)	Ref		
At least one patient <40 years of age	6	(42.9)	33	(36.7)	1.30	(0.41-4.06)	0.657
Japanese							
No Japan-born patients	0	(0.0)	2	(2.2)	Ref		
≥One Japan-born patient	14	(100.0)	88	(97.8)	NA		
Full- and part-time/daily workers							
No patients with full- and part-time/daily employment	6	(42.9)	35	(38.9)	Ref		
≥One patient with full- and part-time/daily employment	8	(57.1)	55	(61.1)	0.85	(0.27-2.65)	0.778
Public assistance							
No patient receiving public assistance	5	(35.7)	41	(45.6)	Ref		

≥One patient receiving public assistance	9	(64.3)	49	(54.4)	1.51	(0.47-4.85)	0.492
Homeless							
No patient who is currently homeless	6	(42.9)	45	(50.0)	Ref		
≥One patient who is currently homeless	8	(57.1)	45	(50.0)	1.33	(0.43-4.15)	0.620
Alcohol misuse							
No patient who misuses alcohol	5	(35.7)	48	(53.3)	Ref		
≥One patient who misuses alcohol	9	(64.3)	42	(46.7)	2.06	(0.64-6.62)	0.227
Cavity lesions							
No patients with a cavity	2	(14.3)	24	(26.7)	Ref		
≥One patient with a cavity	12	(85.7)	66	(73.3)	2.18	(0.45-10.47)	0.330
Smear results							
No patient with a positive smear	1	(7.1)	12	(13.3)	Ref		
≥One patient with a positive smear	13	(92.9)	78	(86.7)	2.00	(0.24-16.71)	0.522
Past TB history							
No patient with a past history of TB	11	(78.6)	69	(76.7)	Ref		
≥One patient with a past history of TB	3	(21.4)	21	(23.3)	0.90	(0.23-3.52)	0.875
DM							
No patient with DM	9	(64.3)	57	(63.3)	Ref		
≥One patient with DM	5	(35.7)	33	(36.7)	0.96	(0.30-3.11)	0.945
Active case finding							
No patient identified through active case finding	8	(57.1)	53	(58.9)	Ref		
≥One patient identified through active case finding	6	(42.9)	37	(41.1)	1.07	(0.34-3.35)	0.902
Patient delay							

No case with patient delay	9	(64.3)	55	(61.1)	Ref		
≥One case with patient delay	5	(35.7)	35	(38.9)	0.87	(0.27-2.82)	0.820
Doctors delay							
No case with doctor delay	10	(71.4)	57	(63.3)	Ref		
≥One case with doctor delay	4	(28.6)	33	(36.7)	0.69	(0.20-2.38)	0.558
Total delay							
No case with total delay	10	(71.4)	53	(58.9)	Ref		
≥One case with total delay	4	(28.6)	37	(41.1)	0.57	(0.17-1.97)	0.376
Registration interval							
0-2 months between first two cases	7	(50.0)	13	(14.4)	9.51	(2.16-41.89)	0.003**
3-5 months between first two cases	2	(14.3)	5	(5.6)	7.07	(0.95-52.77)	0.057
6-11 months between first two cases	2	(14.3)	19	(21.1)	1.86	(0.29-12.00)	0.514
≥12 months between first two cases	3	(21.4)	53	(58.9)	Ref		

Note. After the variables for multivariate logistic regression were selected using the stepwise method, only the "registration interval" variable remained in the model. Thus, the table shows only the results of the univariate logistic regression.

RFLP: restriction fragment length polymorphism, OR: odds ratio, aOR: adjusted odds ratio, CI: confidence interval, TB: tuberculosis, DM: diabetes mellitus, Ref: reference, NA: not applicable, *p<0.05, **p<0.01

†"Large clusters within three years" refers to cluster episodes with five or more cases (large clusters) within three years.

‡"Small clusters and large clusters after three years" refers to cluster episodes with two to four cases (small clusters) and cluster episodes that became large clusters after three years.

469

Figure legends

Figure 1. Number of reported cases of TB, including culture-positive cases, strain-typed cases and genotype clusters, in Shinjuku during 2002–2013

RFLP: restriction fragment length polymorphism, TB: tuberculosis

Figure 2. Cumulative clustering rate (RFLP, Shinjuku 2002–2013)

RFLP: restriction fragment length polymorphism

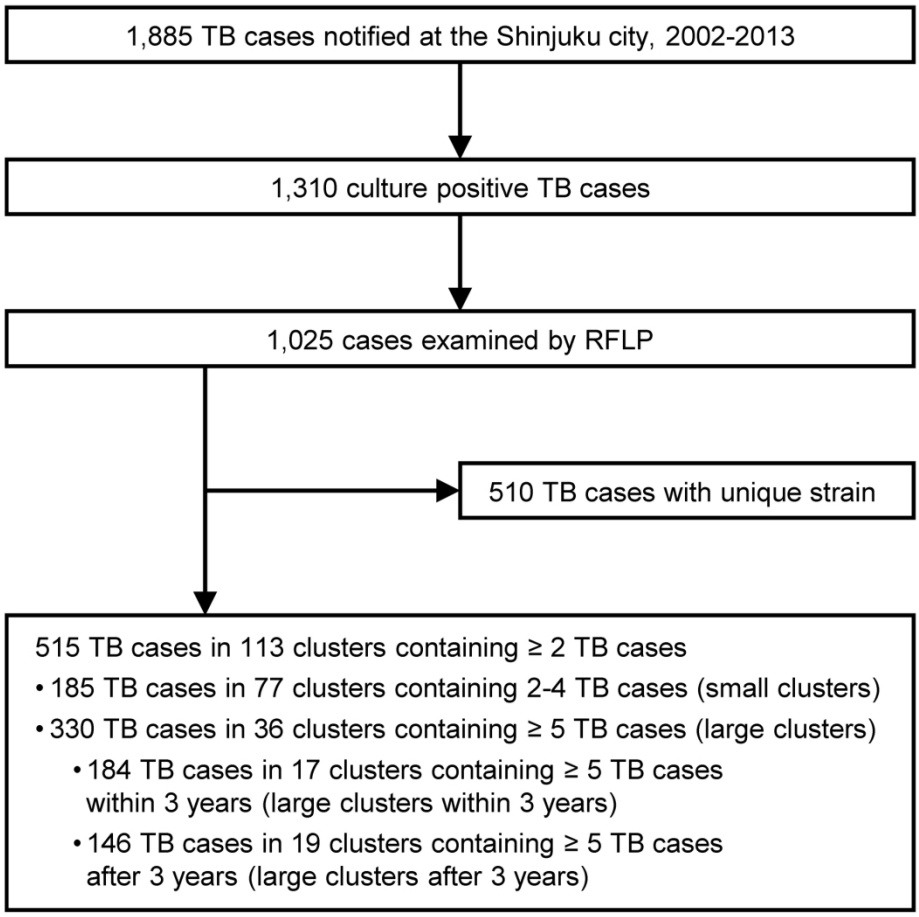

Figure 1. Number of reported cases of TB, including culture-positive cases, strain-typed cases and genotype clusters, Shinjuku, 2002–2013
RFLP: restriction fragment length polymorphism, TB: tuberculosis

90x90mm (600 x 600 DPI)

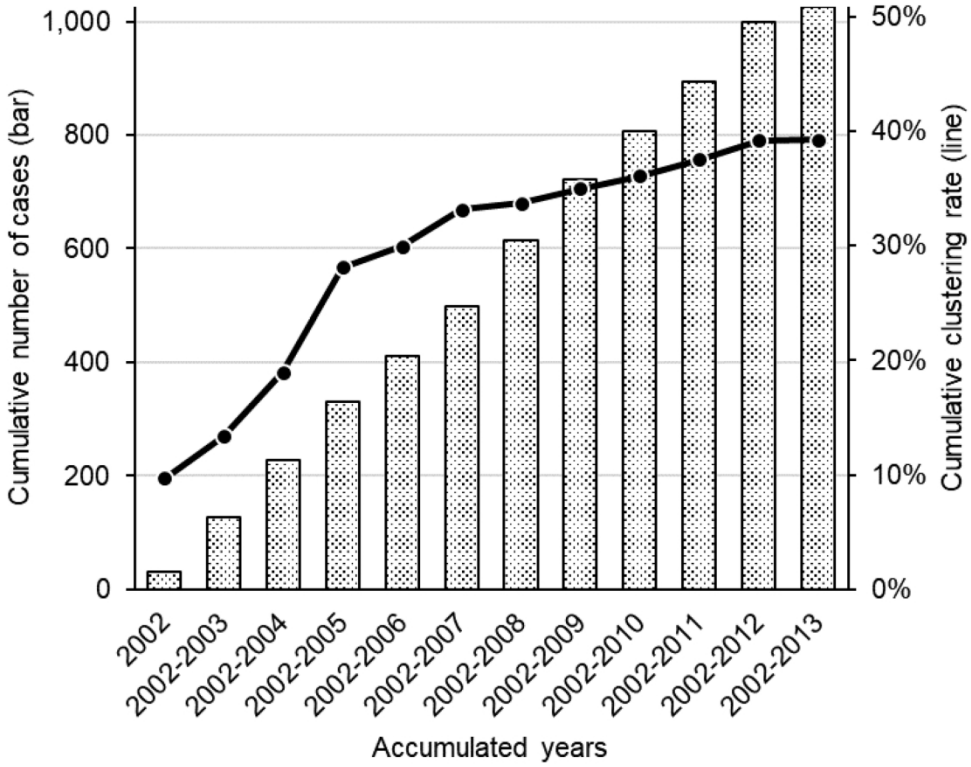

Figure 2. Cumulative clustering rate, RFLP, Shinjuku 2002–2013
RFLP: restriction fragment length polymorphism

90x90mm (600 x 600 DPI)

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Table 1
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	6 **We took all eligible study participants in Shinjuku city.*
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Table 1
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-8
		(*b*) Describe any methods used to examine subgroups and interactions	8
		(*c*) Explain how missing data were addressed	8
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	Not Applicable

		(*e*) Describe any sensitivity analyses	Not Applicable
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	8-9 Figure 1
		(b) Give reasons for non-participation at each stage	8-9
		(c) Consider use of a flow diagram	Figure 1
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	N in tables
Outcome data	15*	Report numbers of outcome events or summary measures	Tables, 9-10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Tables
		(*b*) Report category boundaries when continuous variables were categorized	Tables
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not Applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Table 4
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-14
Generalisability	21	Discuss the generalisability (external validity) of the study results	11-14
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	4

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.