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

Kiyohiko Izumi,[1] Yoshiro Murase,[2] Kazuhiro Uchimura,[1] Aya Kaebeta,[3] Keiko Ishihara,[3] Sumi Kaguraoka,[3] Takemasa Takii,[2] Akihiro Ohkado[1]

To cite: Izumi K, Murase Y, Uchimura K, *et al*. Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in Tokyo, Japan: a population-based molecular epidemiological study. *BMJ Open* 2019;**9**:e029295. doi:10.1136/bmjopen-2019-029295

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-029295).

Received 24 January 2019
Revised 4 April 2019
Accepted 5 April 2019

[1]Department of Epidemiology and Clinical Research, The Research Institute of Tuberculosis, Kiyose, Japan
[2]Department of Mycobacterium Reference and Research, The Research Institute of Tuberculosis, Kiyose, Japan
[3]Health Department, Shinjuku Public Health Center, Shinjuku, Japan

Correspondence to
Kiyohiko Izumi;
kizumi95@gmail.com

ABSTRACT

Objective Molecular epidemiology is a promising tool for understanding tuberculosis transmission dynamics but has not been sufficiently utilised 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 3 years in an urban setting in Japan.

Design and setting Long-term cross-sectional observational study combining the characteristics of patients with culture-positive TB 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 1025 cases, 515 were localised 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 to 2.44), native Japanese individuals (aOR=3.90, 95% CI 2.27 to 6.72), full-time workers (aOR=1.63, 95% CI 1.17 to 2.27), part-time/daily workers (aOR=2.20, 95% CI 1.35 to 3.58), individuals receiving public assistance (aOR=1.81, 95% CI 1.15 to 2.84) and homeless people (aOR=1.63, 95% CI 1.02 to 2.62). A significant predictor of large genotype clusters within 3 years was a registration interval ≤2 months between the first two cases in a cluster.

Conclusion Our results indicated that a large proportion of patients with culture-positive TB 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.

Strengths and limitations of this study

- This study is one of the longest population-based studies focusing on the molecular epidemiology of patients with culture-positive tuberculosis in a large Asian urban setting.
- Interviews conducted by the experienced public health nurses at the Public Health Centre using a standardised 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.

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-burden 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 emphasises 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 USA[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 towards 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 patients with culture-positive TB 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 *Mycobacterium 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 Centre (PHC) interviewed and collected information from all patients with culture-positive TB at the time of registration using a standardised questionnaire to avoid possible interviewer bias. The study variables and definitions are described in table 1.

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	Full-time, part-time/daily worker, jobless under 60 years 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 centres
	Status of patient delay	A time between the onset of symptoms and the initial doctor visit longer than 2 months
	Status of doctor delay	A time between the initial doctor visit and diagnosis longer than 1 month
	Status of total delay	A time between the onset of symptoms and TB diagnosis longer than 3 months
	Registration interval	The duration in months between the registration dates of the first two cases in each of the genotype clusters

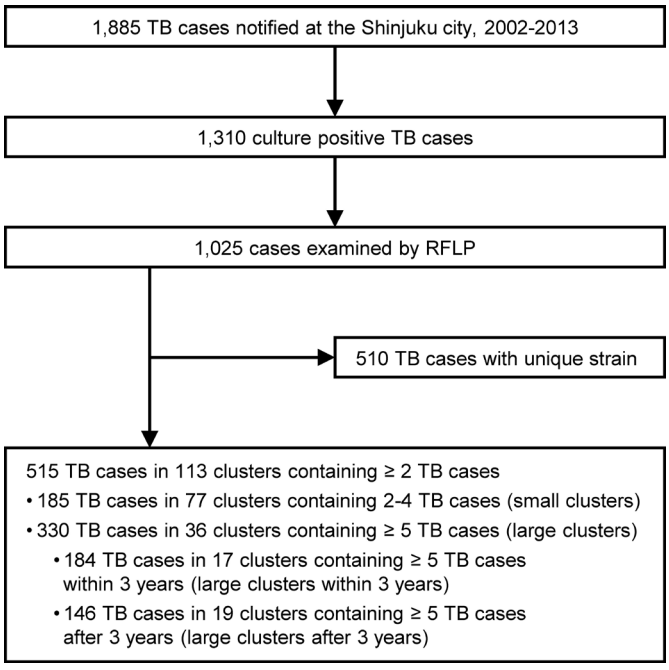

Figure 1 Number of reported cases of TB, including culture-positive cases, strain-typed cases and genotype clusters, in Shinjuku during 2002–2013. RFLP, restriction fragment length polymorphism; TB, tuberculosis.

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 patients with TB 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 patients with TB whose isolates showed either (1) ≥6identical IS*6110* band patterns or (2) <6identical 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 patients with TB 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 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 3 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 3 years' were cluster episodes with five or more cases (large clusters) occurring within 3 years and (2) 'small clusters and large clusters after 3 years' were cluster episodes with two to four cases (small clusters) and cluster episodes that became large clusters after 3 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 3 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 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.

RESULTS

Study population and clustering rate

In total, 1885 patients with TB in Shinjuku City were notified during the study period and 1310 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, 1025 (78.2%) patients were included in the analysis. The figure 2 shows the cumulative number of patients with TB 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 4 years, from 10% in 2002 to 28% in 2005, with an average per cent change of +43%, and then continued to increase at a slower rate, from 30% in 2006 to 39% in 2013, with an average per cent change of +4.2%.

We identified a total of 113 genotype clusters consisting of 515 patients (figure 1). The genotype clustering rate

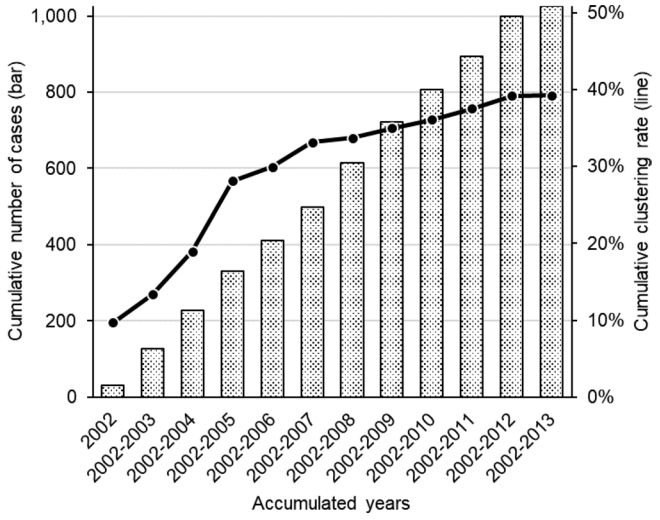

Figure 2 Cumulative clustering rate (restriction fragment length polymorphism, Shinjuku 2002–2013).

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 patients with TB, and 36 (31.9%) genotype clusters had at least five patients with TB. 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 non-homeless individuals, seven (6.2%) included only homeless individuals and 61 (54.0%) contained both homeless and non-homeless individuals (mixed cluster). We compared the characteristics of the non-homeless patients in the clusters of only non-homeless 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 non-homeless 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 to 2.19), Japan-born individuals (OR=3.74, 95% CI 2.25 to 6.44), individuals receiving public assistance (OR=2.25, 95% CI 1.69 to 3.00), homeless individuals (OR=2.45, 95% CI 1.80 to 3.34), individuals who misuse alcohol (OR=1.37, 95% CI 1.02 to 1.83), individuals engaging in full-time work (OR=1.53, 95% CI 1.15 to 2.05) and part-time/daily work (OR=2.29, 95% CI 1.45 to 3.61) and jobless individuals aged 15–59 years (OR=2.05, 95% CI 1.43 to 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 to 2.44), born in Japan (aOR=3.90, 95% CI 2.27 to 6.72), working full-time (aOR=1.63, 95% CI 1.17 to 2.27), having part-time/daily work (aOR=2.20, 95% CI 1.35 to 3.58), receiving public assistance (aOR=1.81, 95% CI 1.15 to 2.84) and homelessness (aOR=1.63, 95% CI 1.02 to 2.62) (table 3).

Factors associated with large genotype clustering within 3 years

We identified 104 genotype cluster episodes according to the definition. Of these, 14 were 'large clusters within 3 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 3 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 3 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 1025 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 3 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 5 years and 195 patients in 1 year, respectively). In our study, as expected, the cumulative clustering rate rapidly increased in the first 4 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,

Table 2 Factors associated with TB genotype clustering; univariable logistic regression analysis, RFLP, 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 to 1.56)	0.267
Sex	1025			
Female	248	102 (41.1)	Reference	
Male	777	413 (53.2)	1.62 (1.20 to 2.19)	0.001**
Country of birth	1025			
Foreign	95	22 (23.2)	Reference	
Japan	930	493 (53.0)	3.74 (2.25 to 6.44)	<0.001***
Occupation	1025			
Full-time worker	313	165 (52.7)	1.53 (1.15 to 2.05)	0.004**
Part-time/daily worker	96	60 (62.5)	2.29 (1.45 to 3.61)	<0.001***
Jobless (aged 15–59 years)	172	103 (59.9)	2.05 (1.43 to 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 to 3.00)	<0.001***
Homelessness	1025			
No	776	349 (45.0)	Reference	
Yes	249	166 (66.7)	2.45 (1.80 to 3.34)	<0.001***
Alcohol misuse§	1025			
No	761	367 (48.2)	Reference	
Yes	264	148 (56.1)	1.37 (1.02 to 1.83)	0.028*
TB site	1024			
Extrapulmonary	80	32 (40.0)	Reference	
Pulmonary	944	482 (51.1)	1.56 (0.96 to 2.58)	0.058
Cavity lesions	1023			
No	565	271 (48.0)	Reference	
Yes	458	243 (53.1)	1.23 (0.95 to 1.58)	0.105
Smear results	1024			
Negative	406	192 (47.3)	Reference	
Positive	618	322 (52.1)	1.21 (0.94 to 1.57)	0.132
Past TB history	989			
New	880	441 (50.1)	Reference	
Retreatment	109	59 (54.1)	1.17 (0.77 to 1.79)	0.429
DM	1005			
No	832	421 (50.6)	Reference	
Yes	173	86 (49.7)	0.97 (0.69 to 1.36)	0.831
Active case finding	1025			
No	842	412 (48.9)	Reference	
Yes	183	103 (56.3)	1.34 (0.96 to 1.88)	0.071
Patient delay	1000			
<2 m	773	377 (48.8)	Reference	
≥2 m	227	127 (55.9)	1.33 (0.98 to 1.82)	0.057

Continued

Table 2 Continued

	Total number of cases (n=1025), n	Clustered cases (n=515), n (%)	OR (95% CI)	P value
Doctor delay	1018			
<1 m	799	415 (51.9)	Reference	
≥1 m	219	97 (44.3)	0.74 (0.54 to 1.00)	0.045*
Total delay	997			
<3 m	777	382 (49.2)	Reference	
≥3 m	220	122 (55.5)	1.29 (0.94 to 1.76)	0.099

*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.
DM, diabetes mellitus; RFLP, restriction fragment length polymorphism; TB, tuberculosis.

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 non-homeless and homeless patients. Moreover, the non-homeless 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 non-homeless cases, which could imply that relatively poor non-homeless 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 patients with TB 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

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

*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.
aOR, adjusted OR; RFLP, restriction fragment length polymorphism.

Table 4 Factors associated with large genotype clusters within 3 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	0.664
≥1 male patient	13 (92.9)	86 (95.6)	0.60 (0.06 to 5.84)	
Age				
No patients <40 years of age	8 (57.1)	57 (63.3)	Ref	0.657
At least one patient <40 years of age	6 (42.9)	33 (36.7)	1.30 (0.41 to 4.06)	
Japanese				
No Japan-born patients	0 (0.0)	2 (2.2)	Ref	
≥1 Japan-born patient	14 (100.0)	88 (97.8)	NA	
Full-time and part-time/daily workers				
No patients with full-time and part-time/daily employment	6 (42.9)	35 (38.9)	Ref	0.778
≥1 patient with full-time and part-time/daily employment	8 (57.1)	55 (61.1)	0.85 (0.27 to 2.65)	
Public assistance				
No patient receiving public assistance	5 (35.7)	41 (45.6)	Ref	0.492
≥1 patient receiving public assistance	9 (64.3)	49 (54.4)	1.51 (0.47 to 4.85)	
Homeless				
No patient who is currently homeless	6 (42.9)	45 (50.0)	Ref	0.620
≥1 patient who is currently homeless	8 (57.1)	45 (50.0)	1.33 (0.43 to 4.15)	
Alcohol misuse				
No patient who misuses alcohol	5 (35.7)	48 (53.3)	Ref	0.227
≥1 patient who misuses alcohol	9 (64.3)	42 (46.7)	2.06 (0.64 to 6.62)	
Cavity lesions				
No patients with a cavity	2 (14.3)	24 (26.7)	Ref	0.330
≥1 patient with a cavity	12 (85.7)	66 (73.3)	2.18 (0.45 to 10.47)	
Smear results				
No patient with a positive smear	1 (7.1)	12 (13.3)	Ref	0.522
≥1 patient with a positive smear	13 (92.9)	78 (86.7)	2.00 (0.24 to 16.71)	
Past TB history				

Continued

Table 4 Continued

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
No patient with a past history of TB	11 (78.6)	69 (76.7)	Ref	0.875
≥1 patient with a past history of TB	3 (21.4)	21 (23.3)	0.90 (0.23 to 3.52)	
DM				
No patient with DM	9 (64.3)	57 (63.3)	Ref	0.945
≥1 patient with DM	5 (35.7)	33 (36.7)	0.96 (0.30 to 3.11)	
Active case finding				
No patient identified through active case finding	8 (57.1)	53 (58.9)	Ref	0.902
≥1 patient identified through active case finding	6 (42.9)	37 (41.1)	1.07 (0.34 to 3.35)	
Patient delay				
No case with patient delay	9 (64.3)	55 (61.1)	Ref	
≥1 case with patient delay	5 (35.7)	35 (38.9)	0.87 (0.27 to 2.82)	0.820
Doctors delay				
No case with doctor delay	10 (71.4)	57 (63.3)	Ref	0.558
≥1 case with doctor delay	4 (28.6)	33 (36.7)	0.69 (0.20 to 2.38)	
Total delay				
No case with total delay	10 (71.4)	53 (58.9)	Ref	0.376
≥1 case with total delay	4 (28.6)	37 (41.1)	0.57 (0.17 to 1.97)	
Registration interval				
0–2 months between first two cases	7 (50.0)	13 (14.4)	9.51 (2.16 to 41.89)	0.003**
3–5 months between first two cases	2 (14.3)	5 (5.6)	7.07 (0.95 to 52.77)	0.057
6–11 months between first two cases	2 (14.3)	19 (21.1)	1.86 (0.29 to 12.00)	0.514
≥12 months between first two cases	3 (21.4)	53 (58.9)	Ref	

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.
*P<0.05, **P<0.01.
†'Large clusters within 3 years' refers to cluster episodes with five or more cases (large clusters) within 3 years.
‡'Small clusters and large clusters after 3 years' refers to cluster episodes with two to four cases (small clusters) and cluster episodes that became large clusters after 3 years.
aOR, adjusted OR; DM, diabetes mellitus; NA, not applicable; Ref, reference; RFLP, restriction fragment length polymorphism; TB, tuberculosis.

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 patients

with TB in urban cities, TB transmission between native and foreign-born populations needs to be closely monitored.

Factors associated with large genotype clustering within 3 years

A shorter registration interval (≤2 months) was identified as a significant predictor of the development of a large genotype cluster within 3 years, which is compatible with findings of previous studies conducted in the Netherlands and London.[10 12] Therefore, when patients with TB 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 2 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 patients with TB 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 patients with TB living across broad geographic areas.[31] Consequently, we may have underestimated the identified genotype clusters. Second, even the existence of patients with TB 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 patients with TB was not available in our study. Therefore, we could not assess and discuss the current practices involving epidemiological investigations done by the public health centre, 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 patients with TB in a large Asian urban setting. Our results indicated that a large proportion of patients with culture-positive TB 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 patients with TB, 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 3 years.

Contributors Klz contributed to the statistical analysis. YM contributed to the genotyping analysis. Klz, YM, KU, TT and AO contributed to the interpretation of data. Klz, KU, AK, Kls, SK and AO contributed to the conception of study. AK, Kls and SK contributed to the collection of data. Klz and AO contributed to the drafting. All authors contributed to the finalisation of the manuscript.

Funding 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 for publication Not required.

Ethics approval The study protocol was approved by the Institutional Review Board of the Research Institute of Tuberculosis (RIT/IRB27-9).

Provenance and peer review Not commissioned; externally peer reviewed.

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

REFERENCES

1. World Health Organization. *Global tuberculosis report 2018*. Geneva, 2018.
2. Lönnroth K, Migliori GB, Abubakar I, *et al*. Towards tuberculosis elimination: an action framework for low-incidence countries. *Eur Respir J* 2015;45:928–52.
3. de Vries G, Aldridge RW, Cayla JA, *et al*. Epidemiology of tuberculosis in big cities of the European Union and European Economic Area countries. *Euro Surveill* 2014;19:1–8.
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;19:1–13.
5. Jereb JA. rogressing 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;19:889–901.
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. *N Engl J Med* 1994;330:1710–6.
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;180:726–36.
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;71:749–56.

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;164:21–31.
12. Kik SV, 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;178:96–104.
13. Althomsons SP, Kammerer JS, Shang N, *et al*. Using routinely reported tuberculosis genotyping and surveillance data to predict tuberculosis outbreaks. *PLoS One* 2012;7:e48754–8.
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;12:548–54.
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.
16. Xu G, Mao X, Wang J, *et al*. Clustering and recent transmission of *Mycobacterium tuberculosis* in a Chinese population. *Infect Drug Resist* 2018;11:323–30.
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;61:219–27.
18. Mears J, Abubakar I, Cohen T, *et al*. 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:e005636.
19. Fok A, Numata Y, Schulzer M, *et al*. Risk factors for clustering of tuberculosis cases: a systematic review of population-based molecular epidemiology studies. *Int J Tuberc Lung Dis* 2008;12:480–92.
20. Statistics of TB. The research institute of tuberculosis J. 2016 http://www.jata.or.jp/rit/ekigaku/en/statistics-of-tb/ (Accessed 20 Jun 2018).
21. Tuberculosis Surveillance Center. Tuberculosis in Japan – Annual Report 2017. Department of Epidemiology and Clinical Research, the Research Institute of TuberculosisTokyo, Japan:; 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 JD, Cave MD, Crawford JT, *et al*. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;31:406–9.
25. Wada T, Maeda S, Hase A, *et al*. Evaluation of variable numbers of tandem repeat as molecular epidemiological markers of Mycobacterium tuberculosis in Japan. *J Med Microbiol* 2007;56:1052–7.
26. Ellis BA, Crawford JT, Braden CR, *et al*. Molecular epidemiology of tuberculosis in a sentinel surveillance population. *Emerg Infect Dis* 2002;8:1197–209.
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;192:480–7.
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:e0138831–16.
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;43:1159–71.
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;92:431–9.
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;56:1–11.
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:e0118495–13.
33. Murase Y, Mitarai S, Sugawara I, *et al*. Promising loci of variable numbers of tandem repeats for typing Beijing family Mycobacterium tuberculosis. *J Med Microbiol* 2008;57:873–80.
34. Nikolayevskyy V, Kranzer K, Niemann S, *et al*. Whole genome sequencing of Mycobacterium tuberculosis for detection of recent transmission and tracing outbreaks: A systematic review. *Tuberculosis* 2016;98:77–85.
35. Jajou R, de Neeling A, van Hunen R, *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:e0195413.