

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60**1 Title**

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

5 Authors and affiliations

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

22 Corresponding Author

23 Kiyohiko Izumi
24 Department of Epidemiology and Clinical Research, Research Institute of Tuberculosis, Japan Anti-

1
2
3
4
5
6 25 Tuberculosis Association, 3-1-24 Matsuyama, Kiyose, Tokyo 204-8533, JAPAN
7

8 26 [Tel:+81-42-4935711](tel:+81-42-4935711)
9

10 27 Fax: +81-42-492-4600
11

12 28 Email: kizumi95@gmail.com
13
14

15 29
16

17 30 KEY WORDS: RFLP, clustering rate, homeless, foreign-born
18

19 31
20

21 32
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 33 **ABSTRACT**

7
8 34 **Objective:** Molecular epidemiology study is a promising tool to understand tuberculosis transmission
9
10 35 dynamics but has not been sufficiently conducted in Asian countries. The aim of this study was to
11
12 36 estimate the proportion of TB cases attributable to recent transmission and to identify the risk
13
14 37 factors of genotype clustering and the development of large clusters within three years in urban
15
16 38 setting in Japan.

17
18
19 39 **Design and setting:** Long-term cross-sectional observational study combining the characteristics of
20
21 40 culture-positive TB patients notified to Shinjuku City, Tokyo (2002-2013) with the genotype data of
22
23 41 *Mycobacterium tuberculosis*.

24
25
26 42 **Primary outcome measure:** Genotype clustering rate and association between genotype clustering
27
28 43 status and explanatory variables.

29
30
31 44 **Results:** Among 1,025 cases, 515 (50.2%) were in 113 genotype clusters. The overall clustering rate
32
33 45 was 39.2%. The rate was significantly higher in patients aged <40 years (adjusted odds ratio (aOR)
34
35 46 =1.73, 95%CI=1.23-2.44) , native Japanese (aOR=3.90, 95%CI=2.27-6.72), fulltime worker (aOR=1.63,
36
37 47 95%CI=1.17-2.27) and part-time/daily worker (aOR=2.20, 95%CI=1.35-3.58), those receiving public
38
39 48 assistance (aOR=1.81, 95%CI=1.15-2.84), and the homeless (aOR=1.63, 95%CI=1.02-2.62). A
40
41 49 significant predictor of large genotype clusters within three years was a registration interval ≤ 2
42
43 50 months between the first two cases in a cluster.

44
45
46 51 **Conclusion:** Our results indicated that a large proportion of culture-positive TB patients were
47
48 52 involved in the recent TB transmission chain. Foreign-born persons still have a limited impact on
49
50 53 transmission in the Japanese urban setting. Intensified public health interventions including active
51
52 54 case finding need to focus on those with socioeconomic risk factors that are significantly associated
53
54 55 with tuberculosis transmission and clusters with shorter registration intervals between the first two
55
56 56 cases.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

57 **Word Count in Abstract= 255 words**

For peer review only

1
2
3
4
5
6 59 **Article Summary**

7
8 60 **Strengths and limitations of this study**

9
10 61 This study is a one of the longest population-based studies on the molecular epidemiology for
11
12 62 culture-positive notified tuberculosis patients in an Asian large urban setting.

13
14 63 Interview conducted by the experienced public health nurses at the Public Health Center using
15
16 64 standardized questionnaire provided high data quality and less interviewer bias.

17
18 65 We may have underestimated genotype clustering due to the large population flow in and out of the
19
20 66 city.

21
22
23
24 67

25
26 68 **Funding statement**

27
28 69 This research was supported by the Research Program on Emerging and Re-emerging Infectious
29
30 70 Diseases from the Japan Agency for Medical Research and Development, AMED No. JP18fk0108041.

31
32
33 71

34
35 72 **Competing interests:** None declared.

36
37
38 73

39
40 74 **Patient consent:** Not required.

41
42
43 75

44 76 **Data sharing statement:** Due to data restrictions, we are unable to share any aspect of the data.

45
46
47 77

48 78 **Word count:** 2,683 words

49
50
51 79

80 INTRODUCTION

81 Tuberculosis (TB) remains a major public health threat worldwide. An estimated 10 million people
82 developed TB, and one million died from TB in 2016 worldwide¹. Although the majority of cases have
83 been reported in TB high-burden countries, TB remains a persistent health problem in low- and
84 middle- burden countries, because it is concentrated in specific vulnerable and hard-to-reach
85 populations, such as homeless people and foreign-born persons from TB high-burden countries².
86 These specific high-risk populations tend to live in large cities where they are seeking jobs, which
87 potentially poses challenges to TB control in urban areas^{3,4}. Many TB low- and middle- burden
88 countries have recently adopted TB elimination strategies^{2,5}, which emphasize the importance of
89 molecular epidemiology in TB control especially in urban areas^{2,4}.

91 TB molecular genotyping using restriction fragment length polymorphisms (RFLPs) and, recently,
92 variable numbers of tandem repeats (VNTR) combined with epidemiological information identifies
93 TB cases possibly involved in the same chain of transmission⁶. This method differentiates recent
94 transmission or endogenous reactivation from remote infection, and therefore has revealed that a
95 substantial proportion of TB cases are due to recent transmission in TB low-burden countries⁷⁻⁹. This
96 method also identifies the proportion of cases attributable to recent transmission and determines
97 the risk factors for transmission. Moreover, the factors predicting large TB genotype clusters have
98 been investigated by evaluating the characteristics of the first two cases in the same genotype
99 cluster¹⁰⁻¹³. These predictors included socially vulnerable populations and shorter intervals between
100 the registration dates of the first two cases. However, these population-based molecular
101 epidemiological studies were limited in some European countries and U.S. and have not been
102 sufficiently conducted in Asian countries.

103

1
2
3
4
5
6 104 In Japan, a TB middle-burden country, newly notified TB cases have been decreasing from 32,828
7
8 105 (25.8 per 100,000 populations) in 2002 to 17,625 (13.9 per 100,000 populations) in 2016¹⁴, while TB
9
10 106 outbreaks have been constantly occurring at approximately 40 events annually over the last decade.
11
12 107 This suggests that TB transmission may be occurring in some groups such as homeless people, who
13
14 108 constitute a high-risk group for recent TB transmission in urban areas¹⁵. Considering the steady
15
16 109 increase in the proportion of TB cases among foreign-born individuals in Japan (7.9% of all cases in
17
18 110 2016¹⁶), transmission between foreign-born persons and local residents must be monitored. In light
19
20 111 of Japan's transition towards becoming a low-burden country, understanding TB transmission
21
22 112 patterns has become increasingly important. However, few population-based molecular
23
24 113 epidemiological studies have identified the transmission patterns and their risk factors. Additionally,
25
26 114 no study has attempted to evaluate the factors predicting the development of large clusters in Japan.
27
28
29
30
31 115
32
33 116 Therefore, we aimed to estimate the proportion of TB cases attributable to recent transmission, to
34
35 117 identify the risk factors for recent transmission, and to predict the risk factors for the development
36
37 118 of large clusters in the urban setting.
38
39
40
41

42 120 **METHODS**

43 121 **Study population**

44
45
46 122 We included all culture-positive TB patients notified in Shinjuku City from September 2002 to
47
48 123 December 2013 as eligible study population in this cross-sectional observational study. This study
49
50 124 was part of a population-based study on DNA fingerprinting surveillance of *M. tuberculosis* in
51
52 125 Shinjuku City started in 2002. Shinjuku City (18.3 km²) is one of the most populous (342,867 residents
53
54 126 in 2018¹⁷) cities in Tokyo and its TB notification rate was 33.7 per 100,000 people in 2016¹⁸, which
55
56 127 was higher than the rates in Tokyo and the nation (17.2 and 13.9, respectively¹⁴). Experienced public
57
58
59
60

1
2
3
4
5
6 128 health nurses at the Shinjuku Public Health Center (PHC) interviewed and collected information from
7
8 129 all culture-positive TB patients at the time of registration using standardized questionnaire to avoid
9
10 130 possible interviewer bias. Study variables and definitions are described in Table 1.
11
12

13 131

15 132 **Patient and public involvement**

16
17 133 Patients or the public were not involved in the design of this study.
18
19 134

21 135 **DNA fingerprinting and genotype cluster**

22
23 136 Clinical isolates from each of the enrolled TB patients were sent to the Research Institute of
24
25 137 Tuberculosis (RIT), Tokyo, where the TB strains were subjected to DNA fingerprinting using insertion
26
27 138 sequence *6110* by RFLP (*IS6110*-RFLP) analysis¹⁹. *IS6110*-RFLP and spoligotyping were the standard
28
29 139 methods used in the Shinjuku PHC and were available throughout the study period. The Shinjuku
30
31 140 PHC switched from RFLP to VNTR a few years ago, but many TB cases had RFLP profiles. Thus, we
32
33 141 employed RFLP due to the sufficient sample size. A genotype cluster was defined as ≥ 2 isolates with
34
35 142 either (1) ≥ 6 *IS6110* bands with identical band patterns or (2) < 6 *IS6110* bands with both identical
36
37 143 *IS6110* band patterns and spoligotyping patterns. Details on the data collection and genotyping
38
39 144 method were previously described¹⁵.
40
41
42
43
44
45
46
47

145

146 **Data analysis**

147 We calculated the genotype clustering rate by the “n-1 method” according to the formula $\{(n - c)/N\}$,
148 where N is the total number of cases sampled, c is the number of clusters, and n is the total number
149 of cases in the clusters⁹. We also calculated the cumulative clustering rate by calculating the
150 clustering rate in 2002 and then adding the TB patients every year up to 2013. The characteristics of
151 clustered cases, which were the cases belonging to any genotype clusters, were compared with those

1
2
3
4
5
6 152 with unique strain patterns using χ^2 tests. We performed univariate logistic regression to identify risk
7
8 153 factors for genotype clustering using odds ratios (ORs) and multivariate logistic regression including
9
10 154 variables selected by the stepwise maximum-likelihood estimation with a significance level of less
11
12 155 than 0.2 using adjusted ORs (aORs). Potential interactions were assessed using likelihood ratio tests.
13
14
15 156
16
17 157 Additionally, we compared the characteristics of the first two cases in each genotype cluster to
18
19 158 predict the risk factors for the development of a large cluster within three years. For this purpose, a
20
21 159 cluster episode was defined as newly arising genotype clusters in and after 2003 without any TB cases
22
23 160 of that genotype notified prior to that year. We classified cluster episodes into two groups according
24
25 161 to the system in a previous study¹⁰; (1) "large clusters within three years," were cluster episodes with
26
27 162 five or more cases (large clusters) occurring within three years, and (2) "small clusters and large
28
29 163 clusters after three years," were cluster episodes with two to four cases (small clusters) and cluster
30
31 164 episodes that became large clusters after three years. We identified the first two cases in each cluster
32
33 165 episode and compared their characteristics between these two groups. We performed univariate
34
35 166 and multivariate logistic regression analyses to identify predictors of the development of large
36
37 167 clusters within three years.
38
39
40
41
42 168
43
44 169 A *p*-value of 0.05 was set as the statistically significant level. If there are variables with missing values
45
46 170 exceeding 5%, the multiple imputation method is considered. We used Stata version 12 (Stata Corp.,
47
48 171 College Station, TX, USA) for the statistical analysis. Written informed consent was waived because
49
50 172 DNA fingerprinting analysis was part of the routine TB control activities in Shinjuku City. However,
51
52 173 oral informed consent was obtained after the PHC staff gave a thorough explanation of study
53
54 174 objectives and confidentiality. The study protocol was approved by the Institutional Review Board of
55
56 175 the Research Institute of Tuberculosis (RIT/IRB27-9).
57
58
59
60

1
2
3
4
5
6 1767
8 177 **RESULTS**9
10 178 **Study population and clustering rate**

11
12
13 179 In total, 1,885 TB patients were notified in Shinjuku City during the study period, and 1,310 were
14
15 180 culture-positive cases (Figure 1). Of those, 285 patients were excluded from the analysis, mainly due
16
17 181 to the unavailability of culture-positive isolates and lack of implementation of RFLP. Finally, 1,025
18
19 182 (78.2%) patients were included in the analysis. Figure 2 shows the cumulative number of TB patients
20
21 183 and clustering rates from 2002 to 2013. The number gradually increased over a decade, while the
22
23 184 cumulative clustering rates sharply increased in the first four years, from 10% in 2002 to 28% in 2005,
24
25 185 with an average percent change of +43%; the clustering rates continues to increase but at a slower
26
27 186 rate, from 30% in 2006 to 39% in 2013, with an average percent change of +4.2%.

28
29
30
31 187

32
33 188 We identified a total of 113 genotype clusters consisting of 515 (50.2%) patients (Figure 1), and the
34
35 189 genotype clustering rate was 39.2%. The average cluster size was 4.56 cases (range 2-30). There were
36
37 190 57 (50.4%) genotype clusters consisting of only two TB patients, and 36 (31.9%) genotype clusters
38
39 191 had five or more TB patients. We further investigated the status of homelessness and place of birth
40
41 192 among genotype clustered cases. Of the 113 genotype clusters, 45 (39.8%) comprised only
42
43 193 nonhomeless individuals, 7 (6.2%) only homeless individuals, and 61 (54.0%) both homeless and
44
45 194 nonhomeless individuals (mixed cluster). We compared the characteristics of nonhomeless patients
46
47 195 in clusters of only nonhomeless patients and those in mixed clusters. While the finding was not
48
49 196 statistically significant (Pearson χ^2 test, $p=0.17$) proportion of nonhomeless patients receiving public
50
51 197 assistance in the latter group (13.8%) was higher than those in the former group (8.8%). There were
52
53 198 no differences between the two groups regarding sex, age, and place of birth. Of the 113 genotype
54
55 199 clusters, 94 (83.2%) consisted of only individuals born in Japan, 2 (1.8%) consisted of only foreign-

1
2
3
4
5
6 200 born individuals, and 17 (15.0%) consisted of both.
7

8 201

9
10 202 **Factors associated with genotype clustering**

11
12 203 Clustered cases were significantly more likely to be male (OR=1.62), Japan-born (OR=3.74), receiving
13 204 public assistance (OR=2.25), homelessness (OR=2.45), misusing alcohol (OR=1.37), and engaging in
14
15 205 fulltime work (OR=1.53), part-time/daily work (OR=2.29), and Jobless aged 15-59-year-old (OR=2.05)
16
17 206 (Table 2). A significant interaction among the explanatory variables was not identified. The
18
19 207 multivariate analysis demonstrated that the factors associated with genotype clustering were age
20
21 208 <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
22
23 209 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),
24
25 210 receiving public assistance (aOR=1.81, 95%CI=1.15-2.84), and homelessness (aOR=1.63, 95%CI=1.02-
26
27 211 2.62)(Table 3).
28
29
30
31
32

33 212

34
35 213 We further investigated the temporal trends in the characteristics of genotype clustered cases. The
36
37 214 number of all homeless cases and their proportion among all study cases decreased (Figure 3A and
38
39 215 Figure 3B), while the proportion of genotype clustered cases among the homeless increased over
40
41 216 time (Figure 3B). No typical trend was observed among those who were foreign-born, receiving
42
43 217 public assistance, misusing alcohol, and occupation (data not shown).
44
45
46

47 218

48
49 219 **Factors associated with large genotype clustering within three years**

50
51 220 We identified 104 genotype cluster episodes according to the definition. Of these, 14 were “large
52
53 221 clusters within three years”, which was equivalent to 13.5% (14/104) of all genotype clusters and
54
55 222 48.3% (14/29) of large genotype clusters, and 90 clusters were “small clusters and large clusters after
56
57 223 three years”. The univariate analysis indicated that clusters with registration intervals of 0-2 months
58
59
60

1
2
3
4
5
6 224 were 9.51 times more likely to become large genotype clusters within three years compared with
7
8 225 clusters with registration intervals of ≥ 12 months (Table 4). After selecting variables by the stepwise
9
10 226 method, only the variable of registration interval remained for the multivariate model.
11
12
13 227

15 228 **DISCUSSION**

17 229 In this long-term population-based study, we included 1,025 patients and identified a total of 113
18
19 230 genotype clusters, and the genotype clustering rate was 39.2%. Our results indicated that clustered
20
21 231 cases were more likely to have socioeconomic risk factors, namely, being homeless, receiving public
22
23 232 assistance, and having an unstable job at tuberculosis diagnosis. A shorter registration interval
24
25 233 between the first two cases was a statistically significant predictor of the development of a large
26
27 234 genotype cluster within three years.
28
29

30
31 235

33 236 **Clustering rate**

35 237 We identified 515 (50.2%) genotype clustered cases and estimated a clustering rate of 39.2%. The
36
37 238 rate was the same as the pooled clustering rate (40.9%) reported in the meta-analysis of population-
38
39 239 based studies of TB low-incidence countries²⁰. However, it was different from previous estimates in
40
41 240 Japanese studies, which were 27.6% in Shinjuku and 24.6% in Osaka^{15,21}. As the meta-regression
42
43 241 analysis clarified that longer study durations increase the clustering rate²⁰, this difference could be
44
45 242 due to shorter study durations together with smaller sample sizes in the previous studies (388
46
47 243 patients in five years and 195 patients in one year, respectively). In our study, the cumulative
48
49 244 clustering rate rapidly increased in the first four years and increased more slowly thereafter, which
50
51 245 is a similar trend to studies in the U.S. and Malawi^{22,23}.
52
53
54

55
56 246

58 247 **Factors associated with genotype clustering**

1
2
3
4
5
6 248 Our results indicated that clustered cases were more likely to have socioeconomic risk factors,
7
8 249 namely, being homeless, receiving public assistance, and having an unstable job at TB diagnosis.
9
10 250 Similarly, previous studies suggested that being homeless significantly contributed to clustering in
11
12 251 Shinjuku City¹⁵ and other counties²⁰. In our study, more than half of the genotype clusters were
13
14 252 mixtures of nonhomeless and homeless patients. Moreover, the nonhomeless patients in the mixed
15
16 253 clusters tended to be financially unstable, with a higher proportion receiving public assistance than
17
18 254 those among clusters of only nonhomeless cases, which could imply that relatively poor
19
20 255 nonhomeless patients share activity spaces with homeless patients, such as urban areas around the
21
22 256 large train stations that were reported as significant hotspots for homeless patients in Shinjuku City²⁴.
23
24 257 An analysis of the time trend of clusters involving the homeless showed that the proportion of
25
26 258 clustered cases among the total homeless patients increased over the study period. These findings
27
28 259 could suggest that contact investigations of homeless TB patients need to be actively expanded to
29
30 260 possible contact persons who are nonhomeless, especially those who are in financial difficulty.
31
32
33
34
35
36

37 261
38 262 The meta-analysis based on studies in European countries where foreign-born patients substantially
39
40 263 contribute to TB epidemiology reported the range of the proportion of mixed clusters composed of
41
42 264 native and foreign-born patients (0% to 36.5%) and concluded that foreign-born patients did not
43
44 265 have a significant influence on TB in the native population²⁵. In our study, the proportion of mixed
45
46 266 clusters (15.0%) fell into this range. Thus, the impact of TB transmission between native and foreign-
47
48 267 born populations was probably still limited in this urban setting²⁶. However, considering the recent
49
50 268 increase in immigrant TB patients in urban cities, TB transmission between native and foreign-born
51
52 269 populations needs to be closely monitored.
53
54
55
56

57 270

58 271 **Factors associated with large genotype clustering within three years**

1
2
3
4
5
6 272 A shorter registration interval (≤ 2 months) was a significant predictor of the development of a large
7
8 273 genotype cluster within three years, which is compatible with findings of previous studies in the
9
10 274 Netherlands and London^{10,12}. Therefore, when TB patients with identical genotypes have shorter
11
12 275 registration intervals, thorough active case findings need to be considered to investigate potential
13
14 276 infection sources and infected patients to prevent further transmission. On the other hand, it is
15
16 277 difficult to assume that the first patient infected the second patient because the window of two
17
18 278 months seems too short. Thus, we think that there may be a true but unidentified first case that our
19
20 279 study did not identify. A cluster episode was defined as a cluster without any TB patients in 2002 and
21
22 280 at least two patients with identical genotypes in and after 2003. Therefore, a possible true first case
23
24 281 may have been registered before 2002, which was outside of our study period, or registered outside
25
26 282 of Shinjuku City.
27
28
29
30
31
32

33 284 **Limitations**

34
35 285 Our study has some limitations. First, the study population consisted only of TB patients living in
36
37 286 Shinjuku City. Considering the large population flow in and out of the city, as we mentioned above,
38
39 287 we potentially missed patients living outside of the city who shared TB strain types with patients
40
41 288 living in the city. Actually, previous Japanese studies reported that there are clusters with TB patients
42
43 289 living across broad geographic areas²⁷. Consequently, we may have underestimated genotype
44
45 290 clustering. Second, even TB patients with identical genotyping patterns may not suggest recent
46
47 291 transmission if the strain is a nationwide endemic TB strain²⁸. As a result, this could have led to an
48
49 292 overestimated clustering rate. Lastly, IS6110 RFLP has relatively lower discriminatory power
50
51 293 compared with VNTR²⁹, which may have led to overestimation.
52
53
54
55
56

57 295 **Conclusion**

1
2
3
4
5
6 296 This study is a one of the longest-term studies on molecular epidemiology for notified tuberculosis
7
8 297 patients in an Asian large urban setting. Our results indicated that a large proportion of culture-
9
10 298 positive TB patients were involved in the recent TB transmission chain. Homeless persons were
11
12 299 involved in more than half of the genotype clusters. For foreign-born persons, they still have a limited
13
14 300 impact on transmission in the Japanese urban setting, but considering recent increases in foreign-
15
16 301 born TB patients, transmission between native and foreign-born populations should be routinely
17
18 302 evaluated. Intensified public health interventions, such as active case findings, may focus on those
19
20 303 who have socioeconomic risk factors that are significantly associated with TB transmission and
21
22 304 clusters with shorter registration intervals between the first two cases, which are predictors of the
23
24 305 development of large clusters within three years.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 307 **ACKNOWLEDGMENTS**
7

8 308 We would like to thank the Shinjuku Public Health Center staff for collecting the data.
9

10 309

11
12 310 **Author Statement:** Authors roles

13
14
15 311 KIz: concept of study, statistical analysis, interpretation of data, drafting and finalization of the

16
17 312 manuscript

18
19 313 YM: genotyping analysis, interpretation of data and finalization of the manuscript

20
21 314 KU: concept of study, interpretation of data and finalization of the manuscript

22
23 315 AK: concept of study, collection of data, and finalization of the manuscript

24
25 316 KIs: concept of study, collection of data, and finalization of the manuscript

26
27 317 SK: concept of study, collection of data, and finalization of the manuscript

28
29 318 TT: interpretation of data and finalization of the manuscript

30
31 319 AK: concept of study, interpretation of data, drafting and finalization of the manuscript

32
33 320

34
35 321

36
37 322
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

323 Reference

- 324 1. World Health Organization. *Global Tuberculosis Report 2017*.; 2017. doi:WHO/HTM/TB/2017.23.
- 325 2. Lönnroth K, Migliori GB, Abubakar I, et al. Towards tuberculosis elimination: An action
326 framework for low-incidence countries. *Eur Respir J*. 2015;45(4):928-952.
327 doi:10.1183/09031936.00214014.
- 328 3. de Vries G, Aldridge RW, Cayla J a, et al. Epidemiology of tuberculosis in big cities of the
329 European Union and European Economic Area countries. *Euro Surveill Bull Eur sur les Mal*
330 *Transm = Eur Commun Dis Bull*. 2014;19(9):1-8.
- 331 4. Hest NA, Aldridge RW, de Vries G, et al. Tuberculosis control in big cities and urban risk groups in
332 the European Union: A consensus statement. *Euro Surveill*. 2014;19(9):1-13. doi:10.2807/1560-
333 7917.ES2014.19.9.20728.
- 334 5. Jereb JA. Progressing toward tuberculosis elimination in low-incidence areas of the United
335 States. Recommendations of the Advisory Council for the Elimination of Tuberculosis. *MMWR*
336 *Recomm reports*. 2002;51(RR-5):1-14.
- 337 6. Borgdorff MW, van Soolingen D. The re-emergence of tuberculosis: What have we learnt from
338 molecular epidemiology? *Clin Microbiol Infect*. 2013;19(10):889-901. doi:10.1111/1469-
339 0691.12253.
- 340 7. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City. An analysis by
341 DNA fingerprinting and conventional epidemiologic methods. 1994;330(D):1710-1716.
- 342 8. van Soolingen D, Borgdorff MW, de Haas PE, et al. Molecular epidemiology of tuberculosis in the
343 Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis*. 1999;180:726-736.
344 doi:10.1086/314930.
- 345 9. PM S, PC H, SP S, et al. The epidemiology of tuberculosis in San Francisco. A population-based
346 study using conventional and molecular methods. *N Engl J Med*. 1994.

- 1
2
3
4
5
6 347 10. Hamblion EL, Le Menach A, Anderson LF, et al. Recent TB transmission, clustering and predictors
7
8 348 of large clusters in London, 2010-2012: Results from first 3-years of universal MIRU-VNTR strain
9
10 349 typing. *Thorax*. 2016;71(8):749-756. doi:10.1136/thoraxjnl-2014-206608.
11
12
13 350 11. Driver CR, Macaraig M, McElroy PD, et al. Which patients' factors predict the rate of growth of
14
15 351 *Mycobacterium tuberculosis* clusters in an urban community? *Am J Epidemiol*. 2006;164(1):21-
16
17 352 31. doi:10.1093/aje/kwj153.
18
19 353 12. Kik S V., Verver S, Van Soolingen D, et al. Tuberculosis outbreaks predicted by characteristics of
20
21 354 first patients in a DNA fingerprint cluster. *Am J Respir Crit Care Med*. 2008;178(1):96-104.
22
23 355 doi:10.1164/rccm.200708-1256OC.
24
25
26 356 13. Althomsons SP, Kammerer JS, Shang N, Navin TR. Using Routinely Reported Tuberculosis
27
28 357 Genotyping and Surveillance Data to Predict Tuberculosis Outbreaks. *PLoS One*. 2012;7(11):1-8.
29
30 358 doi:10.1371/journal.pone.0048754.
31
32
33 359 14. The Research Institute of Tuberculosis J. Statistics of TB 2016.
34
35 360 <http://www.jata.or.jp/rit/ekigaku/en/statistics-of-tb/>. Accessed June 20, 2018.
36
37
38 361 15. Ohkado A, Nagamine M, Murase Y, et al. Molecular epidemiology of *Mycobacterium*
39
40 362 *tuberculosis* in an urban area in Japan, 2002-2006. *Int J Tuberc Lung Dis*. 2008;12(5):548-554.
41
42 363 16. Tuberculosis Surveillance Center. *Tuberculosis in Japan – Annual Report 2017*. Tokyo, Japan:
43
44 364 Department of Epidemiology and Clinical Research, the Research Institute of Tuberculosis; 2017.
45
46 365 17. Shinjuku city. Description of Shinjuku.
47
48 366 http://www.foreign.city.shinjuku.lg.jp/en/aramashi/aramashi_1/. Accessed January 9, 2018.
49
50
51 367 18. Shinjuku City. *Tuberculosis Statistics in Shinjuku City 2017 [Japanese]*.; 2018.
52
53 368 19. Van Embden JDA, Cave MD, Crawford JT, et al. Strain identification of *Mycobacterium*
54
55 369 *tuberculosis* by DNA fingerprinting: Recommendations for a standardized methodology. *J Clin*
56
57 370 *Microbiol*. 1993;31(2):406-409. doi:10.1128/JCM.39.4.1683.2001.
58
59
60

- 1
2
3
4
5
6 371 20. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: A
7
8 372 systematic review of population-based molecular epidemiology studies. *Int J Tuberc Lung Dis*.
9
10 373 2008;12(5):480-492.
11
12
13 374 21. Wada T, Maeda S, Hase A, Kobayashi K. Evaluation of variable numbers of tandem repeat as
14
15 375 molecular epidemiological markers of *Mycobacterium tuberculosis* in Japan. *J Med Microbiol*.
16
17 376 2007;56(8):1052-1057. doi:10.1099/jmm.0.46990-0.
18
19 377 22. Ellis B a, Crawford JT, Braden CR, McNabb SJN, Moore M, Kammerer S. Molecular epidemiology
20
21 378 of tuberculosis in a sentinel surveillance population. *Emerg Infect Dis*. 2002;8(11):1197-1209.
22
23 379 doi:10.3201/eid0811.020403.
24
25
26 380 23. Glynn JR, Crampin AC, Yates MD, et al. The Importance of Recent Infection with *Mycobacterium*
27
28 381 *tuberculosis* in an Area with High HIV Prevalence: A Long-Term Molecular Epidemiological Study
29
30 382 in Northern Malawi. *J Infect Dis*. 2005;192(3):480-487. doi:10.1086/431517.
31
32
33 383 24. Izumi K, Ohkado A, Uchimura K, et al. Detection of tuberculosis infection hotspots using activity
34
35 384 spaces based spatial approach in an urban Tokyo, from 2003 to 2011. *PLoS One*. 2015;10(9):1-
36
37 385 16. doi:10.1371/journal.pone.0138831.
38
39
40 386 25. Sandgren A, Schepisi MS, Sotgiu G, et al. Tuberculosis transmission between foreign- and native-
41
42 387 born populations in the EU/EEA: A systematic review. *Eur Respir J*. 2014;43(4):1159-1171.
43
44 388 doi:10.1183/09031936.00117213.
45
46 389 26. MURASE Y, OHKADO A, WATANABE Y, et al. Transmission dynamics of mycobacterium
47
48 390 tuberculosis between foreign-nationals and Japanese tuberculosis patients living in Shinjuku-
49
50 391 City, Tokyo, Japan. *Kekkaku*. 2017;92(5):431-439.
51
52
53 392 27. Murase Y, Izumi K, Ohkado A, et al. Prediction of local transmission of mycobacterium
54
55 393 tuberculosis isolates of a predominantly Beijing lineage by use of a variable-number tandem-
56
57 394 repeat typing method incorporating a consensus set of hypervariable loci. *J Clin Microbiol*.

- 1
2
3
4
5
6 395 2018;56(1):1-11. doi:10.1128/JCM.01016-17.
7
8 396 28. Wada T, Iwamoto T, Tamaru A, et al. Clonality and micro-diversity of a nationwide spreading
9
10 397 genotype of Mycobacterium tuberculosis in Japan. *PLoS One*. 2015;10(3):1-13.
11
12 398 doi:10.1371/journal.pone.0118495.
13
14
15 399 29. Murase Y, Mitarai S, Sugawara I, Kato S, Maeda S. Promising loci of variable numbers of tandem
16
17 400 repeats for typing Beijing family Mycobacterium tuberculosis. *J Med Microbiol*. 2008;57(7):873-
18
19 401 880. doi:10.1099/jmm.0.47564-0.
20
21
22 402
23
24 403
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1		
2		
3		
4		
5		
6		A time between the onset of symptoms and the initial doctor visit
7	Status of patient delay	
8		longer than two months
9		
10		
11		A time between the initial doctor visit and diagnosis longer than
12	Status of doctor delay	
13		one month
14		
15		A time between onset of symptoms and TB diagnosis longer than
16	Status of total delay	
17		three months
18		
19		
20	Registration interval	The duration in months between the registration dates of the first
21		
22		two cases in each of the genotype clusters
23		
24	406	
25		
26	407	
27		
28	408	
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		
41		
42		
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		

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		

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

411 RFLP: restriction fragment length polymorphism, OR: odds ratio, CI: confidence interval, TB:

412 tuberculosis, DM: diabetes mellitus, *p<0.05, **p<0.01, ***p<0.001

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

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

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

418

419

420 Table 3. Factors associated with TB clustering; multivariate logistic regression analysis, RFLP, Shinjuku
 421 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

422 RFLP: restriction fragment length polymorphism, aOR: adjusted odds ratio, CI: confidence interval,

423 *p<0.05, **p<0.01, ***p<0.001

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

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

427 § Alcohol misuse means excessive drinking practice judged by the public health nurses conducting
428 the interviews.
429

For peer review only

430 Table 4. Analysis of factors associated with large genotype clusters within three years using the
 431 characteristics of the first two cases in each TB genotype cluster; univariable and multivariable
 432 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					

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

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

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

439 † “Large clusters within three years” were cluster episodes with five or more cases (large clusters)
 440 within three years

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

443

444

1
2
3
4
5
6 445 **Figure legend**
7

8 446
9

10 447 Figure 1. Number of reported cases of TB, including culture-positive cases, strain-typed cases and
11
12 448 genotype clusters, Shinjuku, 2002–2013

13
14
15 449 RFLP: restriction fragment length polymorphism, TB: tuberculosis
16

17 450
18

19 451 Figure 2. Cumulative clustering rate, RFLP, Shinjuku 2002–2013
20

21 452 RFLP: restriction fragment length polymorphism
22

23 453
24

25
26 454 Figure 3. Time series of the number (A) and proportion (B) of clustered homeless cases, RFLP,
27
28 455 Shinjuku 2003-2012
29

30 456 RFLP: restriction fragment length polymorphism
31

32
33 457 Note. The 2002 and 2013 data were omitted because of the small total number of cases (31 and 23
34
35 458 cases, respectively). The percentage of clustered cases among the homeless was calculated as the
36
37 459 number of clustered homeless cases divided by the number of all homeless cases, and the
38
39 460 percentage of the homeless among all study cases was calculated as the number of homeless cases
40
41
42 461 divided by the number of all study cases.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

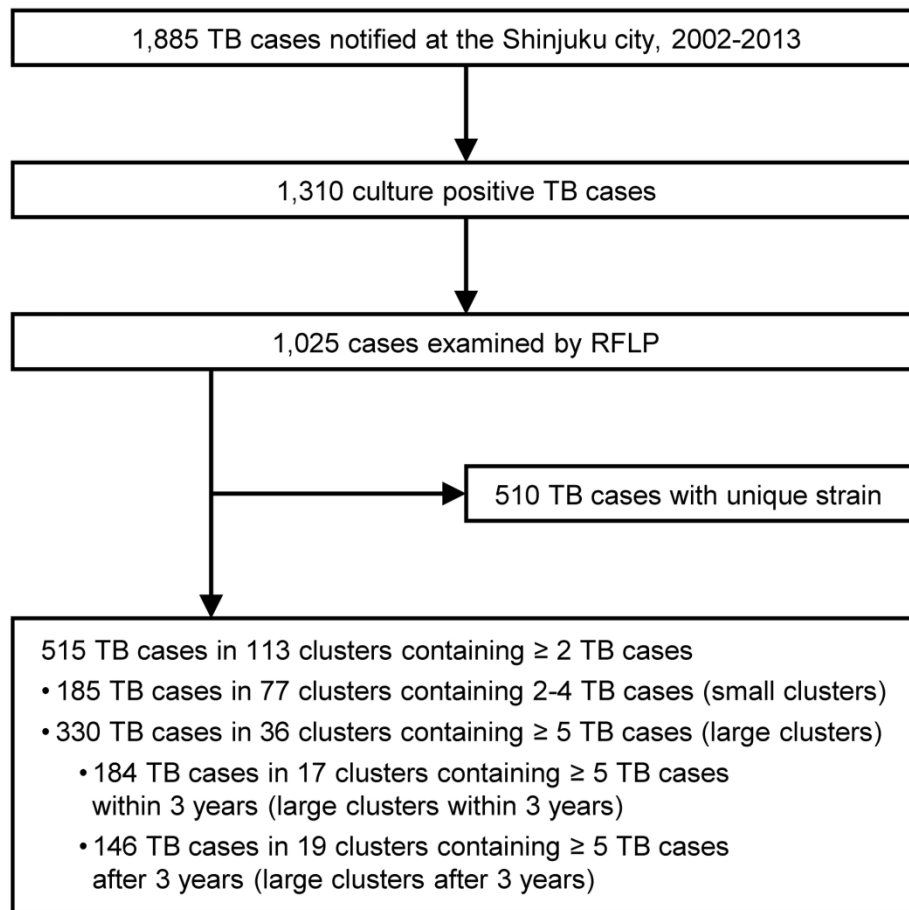


Figure 1. Number of reported cases of TB, including culture-positive cases, strain-typed cases and genotype clusters, Shinjuku, 2002–2013

RFLP: restriction fragment length polymorphism, TB: tuberculosis

90x90mm (600 x 600 DPI)

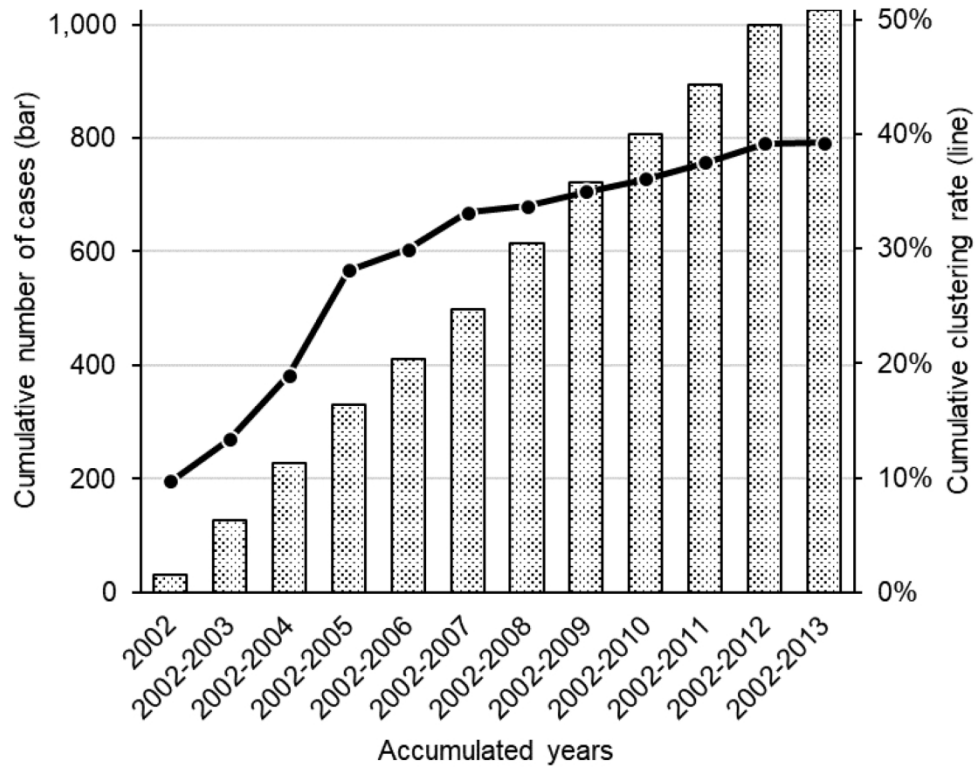


Figure 2. Cumulative clustering rate, RFLP, Shinjuku 2002–2013
RFLP: restriction fragment length polymorphism

90x90mm (600 x 600 DPI)

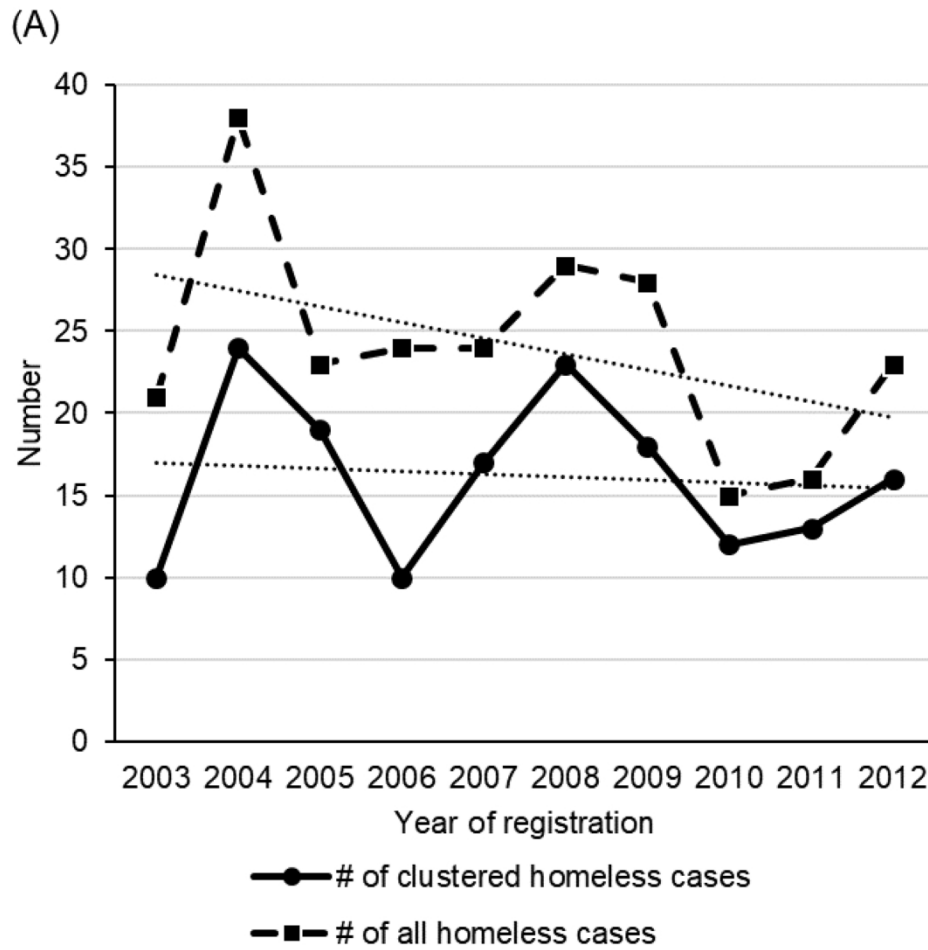


Figure 3. Time series of the number (A) and proportion (B) of clustered homeless cases, RFLP, Shinjuku 2003-2012

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)

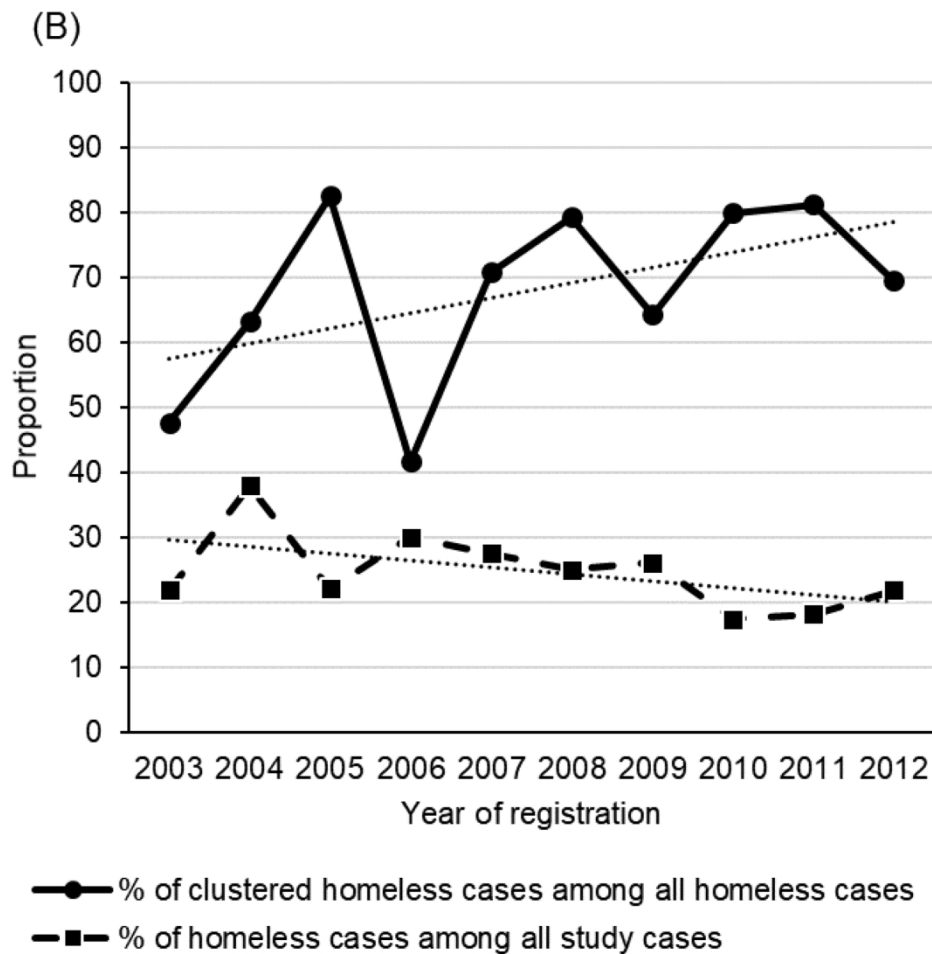


Figure 3. Time series of the number (A) and proportion (B) of clustered homeless cases, RFLP, Shinjuku 2003-2012

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 <i>*We took all eligible study participants in Shinjuku city.</i>
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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3
4
5
6 1 **Title**
7

8 2 Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in
9
10 3 Tokyo, Japan: a population-based molecular epidemiological study
11
12

13 4

14
15 5 **Authors and affiliations**
16

- 17 6 1. Kiyohiko Izumi. Department of Epidemiology and Clinical Research, The Research Institute of
18
19 7 Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: kizumi95@gmail.com
20
21
22 8 2. Yoshiro Murase. Department of Mycobacterium Reference and Research, The Research
23
24 9 Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email:
25
26 10 ymurase@jata.or.jp
27
28 11 3. Kazuhiro Uchimura. Department of Epidemiology and Clinical Research, The Research Institute
29
30 12 of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: uchimura@jata.or.jp
31
32
33 13 4. Aya Kaebeta. Shinjuku Public Health Center. Email: aya.kaebeta@city.shinjuku.lg.jp
34
35 14 5. Keiko Ishihara. Shinjuku Public Health Center. Email: keiko.ishihara@city.shinjuku.lg.jp
36
37 15 6. Sumi Kaguraoka. Shinjuku Public Health Center. Email: sumi.kaguraoka@city.shinjuku.lg.jp
38
39 16 7. Takemasa Takii. The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association,
40
41 17 Tokyo, Japan. Email: t-takii@jata.or.jp
42
43 18 8. Akihiro Ohkado. Department of Epidemiology and Clinical Research, The Research Institute of
44
45 19 Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo, Japan. Graduate School of Biomedical
46
47 20 Sciences, Nagasaki University. Email: ohkadoa@jata.or.jp
48
49
50

51 21

52
53 22 **Corresponding Author**
54

55 23 Kiyohiko Izumi
56

57 24 Department of Epidemiology and Clinical Research, Research Institute of Tuberculosis, Japan Anti-
58
59
60

1
2
3
4
5
6 25 Tuberculosis Association, 3-1-24 Matsuyama, Kiyose, Tokyo 204-8533, JAPAN
7

8 26 [Tel: ±81-42-4935711](tel:+81-42-4935711)
9

10 27 Fax: +81-42-492-4600
11

12 28 Email: kizumi95@gmail.com
13
14

15 29
16

17 30 KEY WORDS: RFLP, clustering rate, homeless, foreign-born
18

19 31
20

21 32
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 33 **ABSTRACT**
7

8 34 **Objective:** Molecular epidemiology is a promising tool for understanding tuberculosis transmission
9
10 35 dynamics but has not been sufficiently utilized in Asian countries including Japan. The aim of this
11
12 36 study was to estimate the proportion of TB cases attributable to recent transmission and to identify
13
14 37 risk factors of genotype clustering and the development of large clusters within three years in an
15
16 38 urban setting in Japan.

17
18
19 39 **Design and setting:** Long-term cross-sectional observational study combining the characteristics of
20
21 40 culture-positive TB patients notified in Shinjuku City, Tokyo (2002-2013), with genotype data of
22
23 41 *Mycobacterium tuberculosis*.

24
25
26 42 **Primary outcome measure:** Genotype clustering rate and association between genotype clustering
27
28 43 status and explanatory variables.

29
30
31 44 **Results:** Among 1,025 cases, 515 (50.2%) were localized within 113 genotype clusters. The overall
32
33 45 clustering rate was 39.2%. Significantly higher rates were found in patients aged <40 years (adjusted
34
35 46 odds ratio (aOR)=1.73, 95% CI=1.23-2.44), native Japanese individuals (aOR=3.90, 95% CI=2.27-6.72),
36
37 47 fulltime workers (aOR=1.63, 95% CI=1.17-2.27), part-time/daily workers (aOR=2.20, 95% CI=1.35-
38
39 48 3.58), individuals receiving public assistance (aOR=1.81, 95% CI=1.15-2.84), and homeless people
40
41 49 (aOR=1.63, 95% CI=1.02-2.62). A significant predictor of large genotype clusters within three years
42
43 50 was a registration interval ≤ 2 months between the first two cases in a cluster.

44
45
46 51 **Conclusion:** Our results indicated that a large proportion of culture-positive TB patients were
47
48 52 involved in the recent TB transmission chain. Foreign-born persons still have a limited impact on
49
50 53 transmission in the Japanese urban setting. Intensified public health interventions, including the
51
52 54 active case finding, need to focus on individuals with socioeconomic risk factors that are significantly
53
54 55 associated with tuberculosis transmission and clusters with shorter registration intervals between
55
56 56 the first two cases.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

57 **Word Count in Abstract= 261 words**

For peer review only

1
2
3
4
5
6 59 **Article Summary**

7
8 60 **Strengths and limitations of this study**

- 9
10 61 • This study is one of the longest population-based studies focusing on the molecular
11 epidemiology of culture-positive notified tuberculosis patients in a large Asian urban setting.
12
13 62
14
15 63 • Interviews conducted by the experienced public health nurses at the Public Health Center using
16 a standardized questionnaire provided high-quality data and less interviewer bias.
17
18 64
19 65 • We may have underestimated genotype clustering due to the large population flow in and out
20 of the city.
21
22 66
23
24 67

25
26 68 **Funding statement**

27
28 69 This research was supported by the Research Program on Emerging and Re-emerging Infectious
29 Diseases from the Japan Agency for Medical Research and Development (AMED No. JP18fk0108041).
30
31 70
32

33 71
34
35 72 **Competing interests:** None declared.
36
37 73

38
39 74 **Patient consent:** Not required.
40
41 75

42
43 76 **Data sharing statement:** Due to data restrictions, we are unable to share any aspect of the data.
44
45 77
46

47
48 78 **Word count:** 2,776 words
49
50 79
51
52
53
54
55
56
57
58
59
60

80 INTRODUCTION

81 Tuberculosis (TB) remains a major public health threat worldwide. In 2017, an estimated 10 million
82 people worldwide developed TB, and 1.27 million died from TB¹. Although the majority of cases have
83 been reported in countries with a high TB burden, TB remains a persistent health problem in low-
84 and medium-burden countries because it is concentrated in specific vulnerable and hard-to-reach
85 populations, such as homeless people and foreign-born persons from TB high-burden countries².
86 These specific high-risk populations tend to live in large cities where they are seeking jobs, which
87 potentially poses challenges to the control of TB in urban areas^{3,4}. Many countries with a low or
88 medium TB burden have recently adopted TB elimination strategies^{2,5}, which emphasizes the
89 importance of molecular epidemiology in TB control, particularly in urban areas^{2,4}.

90
91 TB molecular genotyping using restriction fragment length polymorphisms (RFLPs) and, more
92 recently, variable numbers of tandem repeats (VNTRs) combined with epidemiological information
93 identifies TB cases that are likely involved in the same transmission chain⁶. This method differentiates
94 recent transmission or endogenous reactivation from remote infection and has therefore revealed
95 that a substantial proportion of TB cases are due to recent transmission in low-TB-burden countries⁷⁻
96 ⁹. This method also identifies the proportion of cases attributable to recent transmission and
97 determines the risk factors for transmission. Moreover, various factors predicting large TB genotype
98 clusters, including socially vulnerable populations and shorter intervals between the registration
99 dates of the first two cases, have been investigated by evaluating the characteristics of the first two
100 cases in the same genotype cluster¹⁰⁻¹³. These population-based molecular epidemiological studies
101 were conducted in some European countries^{8,10,12}, the U.S.^{7,9,11} and some Asian countries¹⁴⁻¹⁹.

102
103 In Japan, a country with a medium TB burden, the number of newly notified TB cases decreased from

1
2
3
4
5
6 104 32,828 (25.8 per 100,000 populations) in 2002 to 17,625 (13.9 per 100,000 populations) in 2016²⁰,
7
8 105 but the central government has constantly been reported of TB outbreaks by local governments at a
9
10 106 rate of approximately 40 events annually over the last decade. This information suggests that TB
11
12 107 transmission might be occurring in some groups, such as homeless people, who constitute a high-
13
14 108 risk group for recent TB transmission in urban areas¹⁴. Considering the steady increase in the
15
16 109 proportion of TB cases among foreign-born individuals in Japan (7.9% of all cases in 2016²¹),
17
18 110 transmission between foreign-born persons and local residents must be monitored. In addition, in
19
20 111 light of Japan's transition toward becoming a low-TB-burden country, understanding TB transmission
21
22 112 patterns has become increasingly important. However, few population-based molecular
23
24 113 epidemiological studies have identified the transmission patterns in Japan and their risk factors.
25
26 114 Additionally, no study has attempted to evaluate the factors predicting the development of large
27
28 115 clusters in Japan.
29
30
31
32
33 116
34
35 117 Therefore, we aimed to estimate the proportion of TB cases attributable to recent transmission, to
36
37 118 identify the risk factors for recent transmission, and to predict the risk factors for the development
38
39 119 of large clusters in an urban setting.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

121 **METHODS**

122 **Study population**

123 We included all culture-positive TB patients notified in Shinjuku City from September 2002 to
124 December 2013 as the eligible study population in this cross-sectional observational study. This study
125 forms part of a population-based study on DNA fingerprinting surveillance of *M. tuberculosis* in
126 Shinjuku City that was started in 2002. Shinjuku City (18.3 km²) is one of the most populous (342,867
127 residents in 2018²²) cities in Tokyo, and its TB notification rate in 2016 was 33.7 per 100,000 people²³,

1
2
3
4
5
6 128 which was higher than the rates in Tokyo and the nation (17.2 and 13.9, respectively²⁰). Experienced
7
8 129 public health nurses at the Shinjuku Public Health Center (PHC) interviewed and collected
9
10 130 information from all culture-positive TB patients at the time of registration using a standardized
11
12 131 questionnaire to avoid possible interviewer bias. The study variables and definitions are described in
13
14
15 132 Table 1.

16
17 133

18 19 134 **Patient and public involvement**

20 135 Neither the patients nor the public were involved in the design of this study.

21
22 136

23 24 137 **DNA fingerprinting and genotype cluster**

25
26 138 Clinical isolates from each of the enrolled TB patients were sent to the Research Institute of
27
28 139 Tuberculosis (RIT), Tokyo, where the TB strains were subjected to DNA fingerprinting using insertion
29
30 140 sequence *6110* by RFLP (*IS6110*-RFLP) analysis²⁴. One clinical isolate per person was used for the
31
32 141 clustering analysis. *IS6110*-RFLP and spoligotyping are the standard methods used in the Shinjuku
33
34 142 PHC and were available throughout the study period. The Shinjuku PHC switched from RFLP to VNTR
35
36 143 a few years ago, but the RFLP profiles of many TB cases were available. Thus, we employed RFLP due
37
38 144 to the sufficient sample size. A genotype cluster was defined as a group of TB patients whose isolates
39
40 145 showed either (1) ≥ 6 identical *IS6110* band patterns or (2) < 6 identical *IS6110* band patterns
41
42 146 confirmed by identical spoligotyping patterns. The data collection and genotyping methods were
43
44 147 previously described in detail¹⁴.

45
46 148

47 48 149 **Data analysis**

49
50 150 We calculated the genotype clustering rate by the “n-1 method” according to the formula $\{(n - c)/N\}$,
51
52 151 where N is the total number of cases sampled, c is the number of clusters, and n is the total number
53
54
55
56
57
58
59
60

1
2
3
4
5
6 152 of cases in the clusters⁹. We also calculated the cumulative clustering rate by calculating the
7
8 153 clustering rate in 2002 and then adding the TB patients every year up to 2013. The characteristics of
9
10 154 clustered cases, which were the cases belonging to any genotype clusters, were compared with those
11
12
13 155 with unique strain patterns through χ^2 tests. We performed univariate logistic regression to identify
14
15 156 risk factors for genotype clustering using odds ratios (ORs) and multivariate logistic regression using
16
17 157 adjusted ORs (aORs). Any potential interactions were assessed using likelihood ratio tests.

18
19 158
20
21 159 Additionally, we compared the characteristics of the first two cases in each genotype cluster to
22
23 160 identify risk factors for the development of a large cluster within three years. For this purpose, a
24
25 161 cluster episode was defined as a newly arising genotype cluster in or after 2003 without any TB cases
26
27 162 of that genotype notified prior to that year. We classified cluster episodes into the following two
28
29 163 groups according to a system developed in a previous study¹⁰: (1) "large clusters within three years"
30
31 164 were cluster episodes with five or more cases (large clusters) occurring within three years and (2)
32
33 165 "small clusters and large clusters after three years" were cluster episodes with two to four cases
34
35 166 (small clusters) and cluster episodes that became large clusters after three years. We identified the
36
37 167 first two cases in each cluster episode based on the notification date and compared their
38
39 168 characteristics between these two groups. We performed univariate and multivariate logistic
40
41 169 regression analyses to identify predictors of the development of large clusters within three years.
42
43
44
45

46 170
47
48 171 A *p*-value of 0.05 was set as the level indicating statistical significance. For variables with more than
49
50 172 5% missing values, the multiple imputation method was considered. The variables used for
51
52 173 multivariate logistic regressions were selected by the stepwise maximum-likelihood estimation with
53
54 174 a significance level of less than 0.2. We used Stata version 12 (Stata Corp., College Station, TX, USA)
55
56 175 for the statistical analyses. Written informed consent was waived because DNA fingerprinting
57
58
59
60

1
2
3
4
5
6 176 analysis forms part of the routine TB control activities in Shinjuku City. However, oral informed
7
8 177 consent was obtained after the PHC staff provided a thorough explanation of the study objectives
9
10 178 and confidentiality. The study protocol was approved by the Institutional Review Board of the
11
12 179 Research Institute of Tuberculosis (RIT/IRB27-9).
13
14
15 180

17 181 **RESULTS**

18 182 **Study population and clustering rate**

19
20
21 183 In total, 1,885 TB patients in Shinjuku City were notified during the study period, and 1,310 were
22
23 184 culture-positive cases (Figure 1). Of these, 285 patients were excluded from the analysis, mainly due
24
25 185 to the unavailability of culture-positive isolates and the lack of implementation of RFLP. As a result,
26
27 186 1,025 (78.2%) patients were included in the analysis. Figure 2 shows the cumulative number of TB
28
29 187 patients and the clustering rates from 2002 to 2013. The number of TB cases gradually increased
30
31 188 over the tested decade. In contrast, the cumulative clustering rates sharply increased in the first four
32
33 189 years, from 10% in 2002 to 28% in 2005, with an average percent change of +43%, and then
34
35 190 continued to increase at a slower rate, from 30% in 2006 to 39% in 2013, with an average percent
36
37 191 change of +4.2%.
38
39
40
41
42 192

43
44 193 We identified a total of 113 genotype clusters consisting of 515 (50.2%) patients (Figure 1). The
45
46 194 genotype clustering rate was 39.2%, and the average cluster size was 4.56 cases (range 2-30). Fifty-
47
48 195 seven (50.4%) genotype clusters consisted of only two TB patients, and 36 (31.9%) genotype clusters
49
50 196 had at least five TB patients. We further investigated the homelessness status and place of birth of
51
52 197 the patients in the genotype clusters. Of the 113 genotype clusters, 45 (39.8%) comprised only
53
54 198 nonhomeless individuals, seven (6.2%) included only homeless individuals, and 61 (54.0%) contained
55
56 199 both homeless and nonhomeless individuals (mixed cluster). We compared the characteristics of the
57
58
59
60

1
2
3
4
5
6 200 nonhomeless patients in the clusters of only nonhomeless patients with those in the mixed clusters,
7
8 201 and although the finding was not statistically significant (Pearson χ^2 test, $p=0.17$), the proportion of
9
10 202 nonhomeless patients receiving public assistance in the latter group (13.8%) was higher than that in
11
12 203 the former group (8.8%). No differences in sex, age, and place of birth were found between the two
13
14 204 groups. Of the 113 genotype clusters, 94 (83.2%) consisted of only individuals born in Japan, two
15
16 205 (1.8%) consisted of only foreign-born individuals, and 17 (15.0%) consisted of both individuals born
17
18 206 in Japan and foreign-born individuals.
19
20
21
22
23

207

208 **Factors associated with genotype clustering**

209 The clustered cases were significantly more likely to consist of male individuals (OR=1.62, 95%
210 CI=1.20-2.19), Japan-born individuals (OR=3.74, 95% CI=2.25-6.44), individuals receiving public
211 assistance (OR=2.25, 95% CI=1.69-3.00), homeless individuals (OR=2.45, 95% CI=1.80-3.34),
212 individuals who misuse alcohol (OR=1.37, 95% CI=1.02-1.83), individuals engaging in fulltime work
213 (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
214 individuals aged 15-59 years (OR=2.05, 95% CI=1.43-2.94) (Table 2). A significant interaction among
215 the explanatory variables was not detected. The multivariate analysis demonstrated that the factors
216 associated with genotype clustering were age <40 years (aOR=1.73, 95% CI=1.23-2.44), born in Japan
217 (aOR=3.90, 95% CI=2.27-6.72), working fulltime (aOR=1.63, 95% CI=1.17-2.27), having part-
218 time/daily work (aOR=2.20, 95% CI=1.35-3.58), receiving public assistance (aOR=1.81, 95% CI=1.15-
219 2.84), and homelessness (aOR=1.63, 95% CI=1.02-2.62)(Table 3).

220

221 **Factors associated with large genotype clustering within three years**

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

1
2
3
4
5
6 224 48.3% (14/29) of the large genotype clusters, and 90 clusters were “small clusters and large clusters
7
8 225 after three years”. The univariate analysis indicated that clusters with registration intervals of 0-2
9
10 226 months were 9.51 times more likely to become large genotype clusters within three years compared
11
12 227 with clusters with registration intervals of ≥ 12 months (Table 4). After selecting variables using the
13
14 228 stepwise method, only the “registration interval” variable remained for the multivariate model.
15
16
17 229

19 230 **DISCUSSION**

21 231 In this long-term population-based study, we included 1,025 patients, identified a total of 113
22
23 232 genotype clusters, and obtained a genotype clustering rate of 39.2%. Our results indicated that the
24
25 233 clustered cases were more likely to have certain socioeconomic predictive factors, namely, being
26
27 234 homeless, receiving public assistance, and having an unstable job, at the time of tuberculosis
28
29 235 diagnosis. A shorter registration interval between the first two cases was a statistically significant
30
31 236 predictor of the development of a large genotype cluster within three years.
32
33
34
35 237

37 238 **Clustering rate**

39 239 We identified 515 (50.2%) genotype clustered cases and estimated a clustering rate of 39.2%. The
40
41 240 rate was the same as the pooled clustering rate (40.9%) obtained in a previous meta-analysis of
42
43 241 population-based studies of countries with a low TB incidence¹⁹ but differed from previous estimates
44
45 242 obtained in Japanese studies, which were 27.6% in Shinjuku and 24.6% in Osaka^{14,25}. Because the
46
47 243 meta-regression analysis clarified that longer study durations are associated with an increased
48
49 244 clustering rate¹⁹, this difference could be due to shorter study durations combined with the smaller
50
51 245 sample sizes of the previous studies (388 patients in five years and 195 patients in one year,
52
53 246 respectively). In our study, as expected, the cumulative clustering rate rapidly increased in the first
54
55 247 four years and increased more slowly thereafter, which is similar to the trend observed in the
56
57
58
59
60

1
2
3
4
5
6 248 previous studies^{26,27}.

7
8 249

9
10 250 **Factors associated with genotype clustering**

11
12 251 Our results indicated that the clustered cases were more likely to have socioeconomic predictive
13
14 252 factors, namely, being homeless, receiving public assistance, and having an unstable job, at the time
15
16 253 of TB diagnosis. Similarly, previous studies suggested that being homeless significantly contributed
17
18 254 to clustering in Shinjuku City¹⁴ and other counties¹⁹. In our study, more than half of the genotype
19
20 255 clusters were mixtures of nonhomeless and homeless patients. Moreover, the nonhomeless patients
21
22 256 in the mixed clusters tended to be financially unstable, and a higher proportion of these patients
23
24 257 were receiving public assistance compared with the proportion among clusters of only nonhomeless
25
26 258 cases, which could imply that relatively poor nonhomeless patients share activity spaces with
27
28 259 homeless patients, such as urban areas around the large train stations that were reported to be
29
30 260 significant hotspots for homeless patients in Shinjuku City²⁸. These findings could suggest that
31
32 261 contact investigations of homeless TB patients need to be actively expanded to possible contact
33
34 262 persons who are not homeless, particularly those who are facing financial difficulty.
35
36
37
38
39

40 263

41
42 264 A meta-analysis based on studies conducted in European countries where foreign-born patients
43
44 265 substantially contribute to TB epidemiology found that the proportion of mixed clusters composed
45
46 266 of native and foreign-born patients ranged from 0% to 36.5% and concluded that foreign-born
47
48 267 patients did not have a significant influence on TB in the native population²⁹. In our study, the
49
50 268 proportion of mixed clusters (15.0%) fell into this range. Thus, the impact of TB transmission between
51
52 269 native and foreign-born populations likely remains limited in this urban setting³⁰. However,
53
54 270 considering the recent increase in immigrant TB patients in urban cities, TB transmission between
55
56 271 native and foreign-born populations needs to be closely monitored.
57
58
59
60

1
2
3
4
5
6 2727
8 273 **Factors associated with large genotype clustering within three years**

9
10 274 A shorter registration interval (≤ 2 months) was identified as a significant predictor of the
11
12 275 development of a large genotype cluster within three years, which is compatible with findings of
13
14 276 previous studies conducted in the Netherlands and London^{10,12}. Therefore, when TB patients with
15
16 277 identical genotypes have shorter registration intervals, a thorough active case findings need to be
17
18 278 performed to investigate the potential infection sources and infected patients in order to prevent
19
20 279 further transmission. However, it is difficult to assume that the first patient infected the second
21
22 280 patient because a window of two months appears too short. Thus, we believe that a true but
23
24 281 unidentified first TB case was not identified in our study. A cluster episode was defined as a cluster
25
26 282 without any TB patients in 2002 and at least two patients with identical genotypes in and after 2003.
27
28 283 Therefore, a possible true first TB case might have been registered before 2002, which was outside
29
30 284 of our study period, or registered outside of Shinjuku City.

31
32
33
34
35 28536
37 286 **Limitations**

38
39 287 Our study has some limitations. First, the study population consisted only of TB patients living in
40
41 288 Shinjuku City. Considering the large population flow in and out of the city, as mentioned above, we
42
43 289 potentially missed patients living outside of the city who shared TB strain types with patients living
44
45 290 in the city. In fact, previous Japanese studies reported clusters with TB patients living across broad
46
47 291 geographic areas³¹. Consequently, we may have underestimated the identified genotype clusters.
48
49 292 Second, even the existence of TB patients with identical genotyping patterns may not suggest recent
50
51 293 transmission if the strain is a nationwide endemic TB strain³², which could have led to an
52
53 294 overestimated clustering rate. Third, IS6110 RFLP has relatively lower discriminatory power
54
55 295 compared with VNTR³³ and whole-genome sequencing^{34,35}, which might have led to overestimation.

1
2
3
4
5
6 296 Lastly, information of epidemiological linkage among TB patients was not available in our study.
7
8 297 Therefore, we could not assess and discuss the current practices involving epidemiological
9
10 298 investigations done by the public health center, which could weaken the programmatic implications
11
12
13 299 of our results.
14

15 300

16
17 301 **Conclusion**

18
19 302 This study constitutes a one of the longest-term studies on the molecular epidemiology of notified
20
21 303 TB patients in a large Asian urban setting. Our results indicated that a large proportion of culture-
22
23 304 positive TB patients were involved in the recent TB transmission chain. Homeless persons were found
24
25 305 to be involved in more than half of the genotype clusters. Foreign-born persons continue to have a
26
27 306 limited impact on TB transmission in the Japanese urban setting, but considering recent increases in
28
29 307 foreign-born TB patients, transmission between native and foreign-born populations should be
30
31 308 routinely evaluated. Intensified public health interventions, such as active case findings, should focus
32
33 309 on those with socioeconomic risk factors that are significantly associated with TB transmission and
34
35 310 clusters with shorter registration intervals between the first two cases because these variables could
36
37 311 serve as predictors of the development of large clusters within three years.
38
39
40

41
42 312
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 313 **ACKNOWLEDGMENTS**
7

8 314 We would like to thank the Shinjuku Public Health Center staff for collecting the data.
9

10 315

11
12 316 **Author Statement:** Authors roles

13
14
15 317 KIz: conception of study, statistical analysis, interpretation of data, and drafting and finalization of

16
17 318 the manuscript

18
19 319 YM: genotyping analysis, interpretation of data, and finalization of the manuscript

20
21 320 KU: conception of study, interpretation of data, and finalization of the manuscript

22
23 321 AK: conception of study, collection of data, and finalization of the manuscript

24
25 322 KIs: conception of study, collection of data, and finalization of the manuscript

26
27 323 SK: conception of study, collection of data, and finalization of the manuscript

28
29 324 TT: interpretation of data and finalization of the manuscript

30
31 325 AO: conception of study, interpretation of data, and drafting and finalization of the manuscript

32
33 326

34
35 327

36
37 328
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

329 References

- 330 1. World Health Organization. *Global Tuberculosis Report 2018*. Geneva; 2018. doi:ISBN 978 92 4
331 156539 4.
- 332 2. Lönnroth K, Migliori GB, Abubakar I, et al. Towards tuberculosis elimination: An action
333 framework for low-incidence countries. *Eur Respir J*. 2015;45(4):928-952.
334 doi:10.1183/09031936.00214014.
- 335 3. de Vries G, Aldridge RW, Cayla J a, et al. Epidemiology of tuberculosis in big cities of the
336 European Union and European Economic Area countries. *Euro Surveill Bull Eur sur les Mal*
337 *Transm = Eur Commun Dis Bull*. 2014;19(9):1-8.
- 338 4. Hest NA, Aldridge RW, de Vries G, et al. Tuberculosis control in big cities and urban risk groups in
339 the European Union: A consensus statement. *Euro Surveill*. 2014;19(9):1-13. doi:10.2807/1560-
340 7917.ES2014.19.9.20728.
- 341 5. Jereb JA. Progressing towards tuberculosis elimination in low-incidence areas of the United
342 States. Recommendations of the Advisory Council for the Elimination of Tuberculosis. *MMWR*
343 *Recomm reports*. 2002;51(RR-5):1-14.
- 344 6. Borgdorff MW, van Soolingen D. The re-emergence of tuberculosis: What have we learnt from
345 molecular epidemiology? *Clin Microbiol Infect*. 2013;19(10):889-901. doi:10.1111/1469-
346 0691.12253.
- 347 7. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City. An analysis by
348 DNA fingerprinting and conventional epidemiologic methods. 1994;330(D):1710-1716.
- 349 8. van Soolingen D, Borgdorff MW, de Haas PE, et al. Molecular epidemiology of tuberculosis in the
350 Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis*. 1999;180:726-736.
351 doi:10.1086/314930.
- 352 9. PM S, PC H, SP S, et al. The epidemiology of tuberculosis in San Francisco. A population-based

- 1
2
3
4
5
6 353 study using conventional and molecular methods. *N Engl J Med.* 1994.
7
8 354 10. Hamblion EL, Le Menach A, Anderson LF, et al. Recent TB transmission, clustering and predictors
9
10 355 of large clusters in London, 2010-2012: Results from first 3-years of universal MIRU-VNTR strain
11
12 356 typing. *Thorax.* 2016;71(8):749-756. doi:10.1136/thoraxjnl-2014-206608.
13
14
15 357 11. Driver CR, Macaraig M, McElroy PD, et al. Which patients' factors predict the rate of growth of
16
17 358 *Mycobacterium tuberculosis* clusters in an urban community? *Am J Epidemiol.* 2006;164(1):21-
18
19 359 31. doi:10.1093/aje/kwj153.
20
21 360 12. Kik S V., Verver S, Van Soolingen D, et al. Tuberculosis outbreaks predicted by characteristics of
22
23 361 first patients in a DNA fingerprint cluster. *Am J Respir Crit Care Med.* 2008;178(1):96-104.
24
25 362 doi:10.1164/rccm.200708-1256°C.
26
27
28 363 13. Althomsons SP, Kammerer JS, Shang N, Navin TR. Using Routinely Reported Tuberculosis
29
30 364 Genotyping and Surveillance Data to Predict Tuberculosis Outbreaks. *PLoS One.* 2012;7(11):1-8.
31
32 365 doi:10.1371/journal.pone.0048754.
33
34
35 366 14. Ohkado A, Nagamine M, Murase Y, et al. Molecular epidemiology of *Mycobacterium*
36
37 367 *tuberculosis* in an urban area in Japan, 2002-2006. *Int J Tuberc Lung Dis.* 2008;12(5):548-554.
38
39
40 368 15. Ohkado A, Murase Y, Mori M, et al. Transmission of specific genotype streptomycin resistant
41
42 369 strains of *Mycobacterium tuberculosis* in the Tokyo Metropolitan Area in Japan. *BMC Infect Dis.*
43
44 370 2009;9:138. doi:10.1186/1471-2334-9-138.
45
46
47 371 16. Xu G, Mao X, Wang J, Pan H. Clustering and recent transmission of *Mycobacterium tuberculosis*
48
49 372 in a Chinese population. *Infect Drug Resist.* 2018;11:323-330.
50
51 373 17. Yang C, Shen X, Peng Y, et al. Transmission of *Mycobacterium tuberculosis* in China: a
52
53 374 population-based molecular epidemiologic study. *Clin Infect Dis.* 2015;61(2):219-227.
54
55 375 doi:10.1093/cid/civ255.
56
57
58 376 18. Mears J, Abubakar I, Cohen T, McHugh TD, Sonnenberg P. Effect of study design and setting on
59
60

- 1
2
3
4
5
6 377 tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable
7
8 378 Number Tandem Repeats (MIRU-VNTR): A systematic review. *BMJ Open*. 2015;5(1).
9
10 379 doi:10.1136/bmjopen-2014-005636.
11
12
13 380 19. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: A
14
15 381 systematic review of population-based molecular epidemiology studies. *Int J Tuberc Lung Dis*.
16
17 382 2008;12(5):480-492.
18
19 383 20. The Research Institute of Tuberculosis J. Statistics of TB 2016.
20
21 384 <http://www.jata.or.jp/rit/ekigaku/en/statistics-of-tb/>. Accessed June 20, 2018.
22
23
24 385 21. Tuberculosis Surveillance Center. *Tuberculosis in Japan – Annual Report 2017*. Tokyo, Japan:
25
26 386 Department of Epidemiology and Clinical Research, the Research Institute of Tuberculosis; 2017.
27
28 387 22. Shinjuku city. Description of Shinjuku.
29
30 388 http://www.foreign.city.shinjuku.lg.jp/en/aramashi/aramashi_1/. Accessed January 9, 2018.
31
32
33 389 23. Shinjuku City. *Tuberculosis Statistics in Shinjuku City 2017 [Japanese]*.; 2018.
34
35 390 24. Van Embden JDA, Cave MD, Crawford JT, et al. Strain identification of Mycobacterium
36
37 391 tuberculosis by DNA fingerprinting: Recommendations for a standardized methodology. *J Clin*
38
39 392 *Microbiol*. 1993;31(2):406-409. doi:10.1128/JCM.39.4.1683.2001.
40
41
42 393 25. Wada T, Maeda S, Hase A, Kobayashi K. Evaluation of variable numbers of tandem repeat as
43
44 394 molecular epidemiological markers of Mycobacterium tuberculosis in Japan. *J Med Microbiol*.
45
46 395 2007;56(8):1052-1057. doi:10.1099/jmm.0.46990-0.
47
48
49 396 26. Ellis B a, Crawford JT, Braden CR, McNabb SJN, Moore M, Kammerer S. Molecular epidemiology
50
51 397 of tuberculosis in a sentinel surveillance population. *Emerg Infect Dis*. 2002;8(11):1197-1209.
52
53 398 doi:10.3201/eid0811.020403.
54
55 399 27. Glynn JR, Crampin AC, Yates MD, et al. The Importance of Recent Infection with *Mycobacterium*
56
57 400 *tuberculosis* in an Area with High HIV Prevalence: A Long-Term Molecular Epidemiological Study

- 1
2
3
4
5
6 401 in Northern Malawi. *J Infect Dis.* 2005;192(3):480-487. doi:10.1086/431517.
- 7
8 402 28. Izumi K, Ohkado A, Uchimura K, et al. Detection of tuberculosis infection hotspots using activity
9
10 403 spaces based spatial approach in an urban Tokyo, from 2003 to 2011. *PLoS One.* 2015;10(9):1-
11
12 404 16. doi:10.1371/journal.pone.0138831.
- 13
14
15 405 29. Sandgren A, Schepisi MS, Sotgiu G, et al. Tuberculosis transmission between foreign- and native-
16
17 406 born populations in the EU/EEA: A systematic review. *Eur Respir J.* 2014;43(4):1159-1171.
18
19 407 doi:10.1183/09031936.00117213.
- 20
21
22 408 30. MURASE Y, OHKADO A, WATANABE Y, et al. Transmission dynamics of mycobacterium
23
24 409 tuberculosis between foreign-nationals and Japanese tuberculosis patients living in Shinjuku-
25
26 410 City, Tokyo, Japan. *Kekkaku.* 2017;92(5):431-439.
- 27
28
29 411 31. Murase Y, Izumi K, Ohkado A, et al. Prediction of local transmission of mycobacterium
30
31 412 tuberculosis isolates of a predominantly Beijing lineage by use of a variable-number tandem-
32
33 413 repeat typing method incorporating a consensus set of hypervariable loci. *J Clin Microbiol.*
34
35 414 2018;56(1):1-11. doi:10.1128/JCM.01016-17.
- 36
37
38 415 32. Wada T, Iwamoto T, Tamaru A, et al. Clonality and micro-diversity of a nationwide spreading
39
40 416 genotype of *Mycobacterium tuberculosis* in Japan. *PLoS One.* 2015;10(3):1-13.
41
42 417 doi:10.1371/journal.pone.0118495.
- 43
44
45 418 33. Murase Y, Mitarai S, Sugawara I, Kato S, Maeda S. Promising loci of variable numbers of tandem
46
47 419 repeats for typing Beijing family *Mycobacterium tuberculosis*. *J Med Microbiol.* 2008;57(7):873-
48
49 420 880. doi:10.1099/jmm.0.47564-0.
- 50
51
52 421 34. Nikolayevskyy V, Kranzer K, Niemann S, Drobniewski F. Whole genome sequencing of
53
54 422 *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks : A
55
56 423 systematic review. *Tuberculosis.* 2016;98:77-85. doi:10.1016/j.tube.2016.02.009.
- 57
58 424 35. Jajou R, Neeling A De, Hunen R Van, et al. Epidemiological links between tuberculosis cases
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

425 identified twice as efficiently by whole genome sequencing than conventional molecular typing :

426 A population-based study. *PLoS One*. 2018;13(4):e0195413.

427

428

For peer review only

429 **TABLE**

430 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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

431
432
433

For peer review only

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		

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

436 RFLP: restriction fragment length polymorphism, OR: odds ratio, CI: confidence interval, TB:

437 tuberculosis, DM: diabetes mellitus, *p<0.05, **p<0.01, ***p<0.001

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

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

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

444

445

446 Table 3. Factors associated with TB genotype clustering; multivariate logistic regression analysis,
 447 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

448 RFLP: restriction fragment length polymorphism, aOR: adjusted odds ratio, CI: confidence interval,

449 *p<0.05, **p<0.01, ***p<0.001

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

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

1
2
3
4
5
6 453 §Alcohol misuse refer to excessive drinking, as judged by the public health nurses conducting the
7
8 454 interviews.
9

10 455
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

456 Table 4. Factors associated with large genotype clusters within three years using the characteristics
 457 of the first two cases in each TB genotype cluster; univariable logistic regression, RFLP, Shinjuku,
 458 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					

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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		

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

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

465 +“Large clusters within three years” refers to cluster episodes with five or more cases (large clusters)
 466 within three years.

467 +“Small clusters and large clusters after three years” refers to cluster episodes with two to four cases
 468 (small clusters) and cluster episodes that became large clusters after three years.

469

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

470

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

471 **Figure legends**

472

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

475 RFLP: restriction fragment length polymorphism, TB: tuberculosis

476

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

478 RFLP: restriction fragment length polymorphism

479

For peer review only

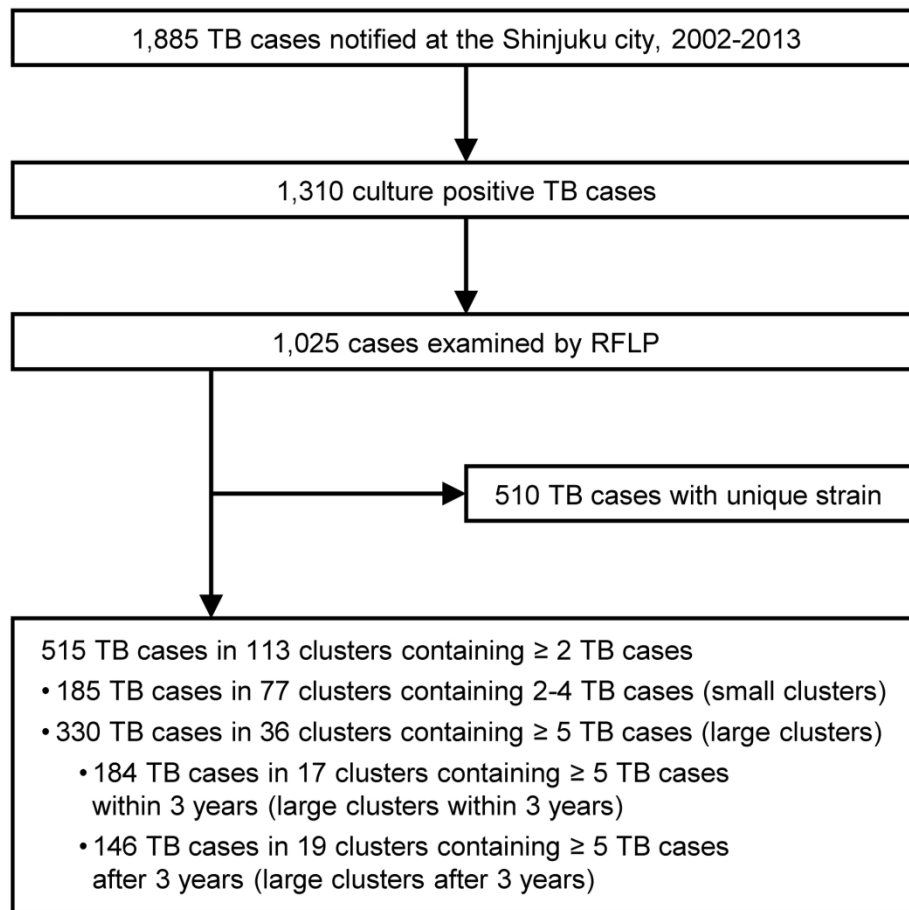


Figure 1. Number of reported cases of TB, including culture-positive cases, strain-typed cases and genotype clusters, Shinjuku, 2002–2013

RFLP: restriction fragment length polymorphism, TB: tuberculosis

90x90mm (600 x 600 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

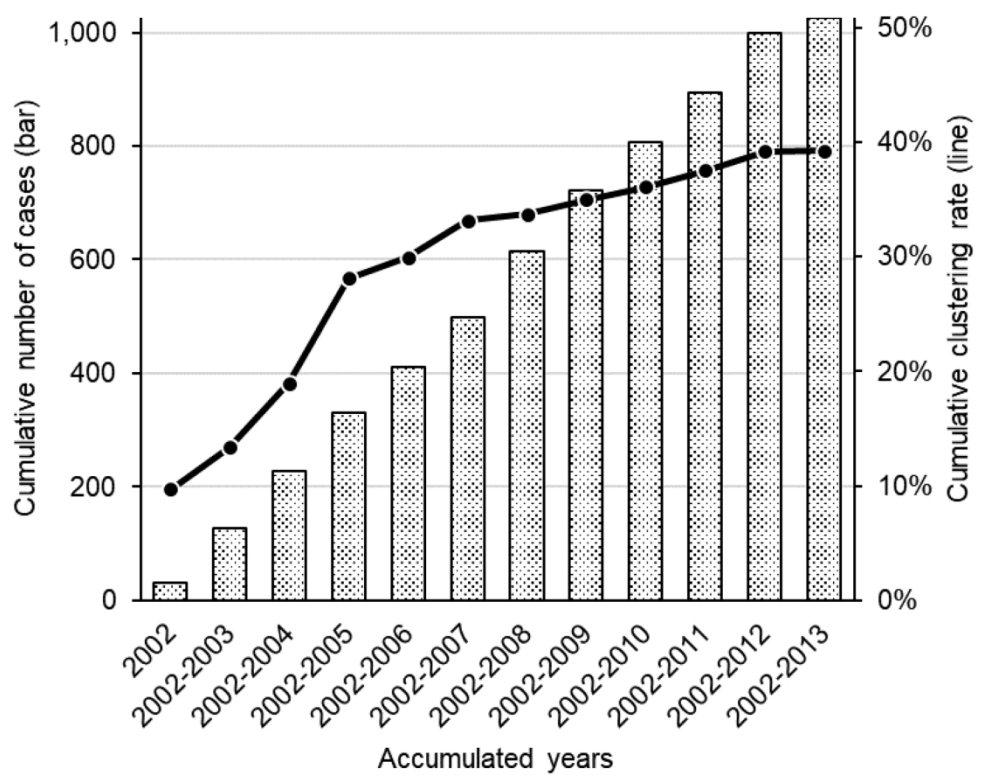


Figure 2. Cumulative clustering rate, RFLP, Shinjuku 2002–2013
RFLP: restriction fragment length polymorphism

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3
4
5
6 1 **Title**
7

8 2 Transmission of tuberculosis and predictors of large clusters within three years in an urban setting in
9
10 3 Tokyo, Japan: a population-based molecular epidemiological study
11
12
13 4

14
15 5 **Authors and affiliations**
16

- 17 6 1. Kiyohiko Izumi. Department of Epidemiology and Clinical Research, The Research Institute of
18
19 7 Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: kizumi95@gmail.com
20
21
22 8 2. Yoshiro Murase. Department of Mycobacterium Reference and Research, The Research
23
24 9 Institute of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email:
25
26 10 ymurase@jata.or.jp
27
28 11 3. Kazuhiro Uchimura. Department of Epidemiology and Clinical Research, The Research Institute
29
30 12 of Tuberculosis, Japan Anti- Tuberculosis Association, Tokyo, Japan. Email: uchimura@jata.or.jp
31
32
33 13 4. Aya Kaebeta. Shinjuku Public Health Center. Email: aya.kaebeta@city.shinjuku.lg.jp
34
35 14 5. Keiko Ishihara. Shinjuku Public Health Center. Email: keiko.ishihara@city.shinjuku.lg.jp
36
37 15 6. Sumi Kaguraoka. Shinjuku Public Health Center. Email: sumi.kaguraoka@city.shinjuku.lg.jp
38
39 16 7. Takemasa Takii. The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association,
40
41 17 Tokyo, Japan. Email: t-takii@jata.or.jp
42
43 18 8. Akihiro Ohkado. Department of Epidemiology and Clinical Research, The Research Institute of
44
45 19 Tuberculosis, Japan Anti-Tuberculosis Association, Tokyo, Japan. Graduate School of Biomedical
46
47 20 Sciences, Nagasaki University. Email: ohkadoa@jata.or.jp
48
49
50

51 21

52
53 22 **Corresponding Author**
54

55 23 Kiyohiko Izumi
56

57 24 Department of Epidemiology and Clinical Research, Research Institute of Tuberculosis, Japan Anti-
58
59
60

1
2
3
4
5
6 25 Tuberculosis Association, 3-1-24 Matsuyama, Kiyose, Tokyo 204-8533, JAPAN
7

8 26 [Tel: ±81-42-4935711](tel:+81-42-4935711)
9

10 27 Fax: +81-42-492-4600
11

12 28 Email: kizumi95@gmail.com
13
14

15 29
16

17 30 KEY WORDS: RFLP, clustering rate, homeless, foreign-born
18

19 31
20

21 32
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 33 **ABSTRACT**

7
8 34 **Objective:** Molecular epidemiology is a promising tool for understanding tuberculosis transmission
9
10 35 dynamics but has not been sufficiently utilized in Asian countries including Japan. The aim of this
11
12 36 study was to estimate the proportion of TB cases attributable to recent transmission and to identify
13
14 37 risk factors of genotype clustering and the development of large clusters within three years in an
15
16 38 urban setting in Japan.

17
18
19 39 **Design and setting:** Long-term cross-sectional observational study combining the characteristics of
20
21 40 culture-positive TB patients notified in Shinjuku City, Tokyo (2002-2013), with genotype data of
22
23 41 *Mycobacterium tuberculosis*.

24
25
26 42 **Primary outcome measure:** Genotype clustering rate and association between genotype clustering
27
28 43 status and explanatory variables.

29
30
31 44 **Results:** Among 1,025 cases, 515 were localized within 113 genotype clusters. The overall clustering
32
33 45 rate was 39.2%. Significantly higher rates were found in patients aged <40 years (adjusted odds ratio
34
35 46 (aOR)=1.73, 95% CI=1.23-2.44), native Japanese individuals (aOR=3.90, 95% CI=2.27-6.72), fulltime
36
37 47 workers (aOR=1.63, 95% CI=1.17-2.27), part-time/daily workers (aOR=2.20, 95% CI=1.35-3.58),
38
39 48 individuals receiving public assistance (aOR=1.81, 95% CI=1.15-2.84), and homeless people
40
41 49 (aOR=1.63, 95% CI=1.02-2.62). A significant predictor of large genotype clusters within three years
42
43 50 was a registration interval ≤ 2 months between the first two cases in a cluster.

44
45
46 51 **Conclusion:** Our results indicated that a large proportion of culture-positive TB patients were
47
48 52 involved in the recent TB transmission chain. Foreign-born persons still have a limited impact on
49
50 53 transmission in the Japanese urban setting. Intensified public health interventions, including the
51
52 54 active case finding, need to focus on individuals with socioeconomic risk factors that are significantly
53
54 55 associated with tuberculosis transmission and clusters with shorter registration intervals between
55
56 56 the first two cases.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

57 **Word Count in Abstract= 260 words**

For peer review only

1
2
3
4
5
6 59 **Article Summary**

7
8 60 **Strengths and limitations of this study**

- 9
10 61 • This study is one of the longest population-based studies focusing on the molecular
11 epidemiology of culture-positive notified tuberculosis patients in a large Asian urban setting.
12
13 62
14
15 63 • Interviews conducted by the experienced public health nurses at the Public Health Center using
16 a standardized questionnaire provided high-quality data and less interviewer bias.
17
18 64
19 65 • We may have underestimated genotype clustering due to the large population flow in and out
20 of the city.
21
22 66
23
24 67

25
26 68 **Funding statement**

27
28 69 This research was supported by the Research Program on Emerging and Re-emerging Infectious
29 Diseases from the Japan Agency for Medical Research and Development (AMED No. JP18fk0108041).
30
31 70
32

33 71
34
35 72 **Competing interests:** None declared.
36
37 73

38
39 74 **Patient consent:** Not required.
40
41 75

42
43 76 **Data sharing statement:** Due to data restrictions, we are unable to share any aspect of the data.
44
45 77
46

47
48 78 **Word count:** 2,774 words
49
50 79
51
52
53
54
55
56
57
58
59
60

80 INTRODUCTION

81 Tuberculosis (TB) remains a major public health threat worldwide. In 2017, an estimated 10 million
82 people worldwide developed TB, and 1.27 million died from TB¹. Although the majority of cases have
83 been reported in countries with a high TB burden, TB remains a persistent health problem in low-
84 and medium-burden countries because it is concentrated in specific vulnerable and hard-to-reach
85 populations, such as homeless people and foreign-born persons from TB high-burden countries².
86 These specific high-risk populations tend to live in large cities where they are seeking jobs, which
87 potentially poses challenges to the control of TB in urban areas^{3,4}. Many countries with a low or
88 medium TB burden have recently adopted TB elimination strategies^{2,5}, which emphasizes the
89 importance of molecular epidemiology in TB control, particularly in urban areas^{2,4}.

90
91 TB molecular genotyping using restriction fragment length polymorphisms (RFLPs) and, more
92 recently, variable numbers of tandem repeats (VNTRs) combined with epidemiological information
93 identifies TB cases that are likely involved in the same transmission chain⁶. This method differentiates
94 recent transmission or endogenous reactivation from remote infection and has therefore revealed
95 that a substantial proportion of TB cases are due to recent transmission in low-TB-burden countries⁷⁻
96 ⁹. This method also identifies the proportion of cases attributable to recent transmission and
97 determines the risk factors for transmission. Moreover, various factors predicting large TB genotype
98 clusters, including socially vulnerable populations and shorter intervals between the registration
99 dates of the first two cases, have been investigated by evaluating the characteristics of the first two
100 cases in the same genotype cluster¹⁰⁻¹³. These population-based molecular epidemiological studies
101 were conducted in some European countries^{8,10,12}, the U.S.^{7,9,11} and some Asian countries¹⁴⁻¹⁹.

102
103 In Japan, a country with a medium TB burden, the number of newly notified TB cases decreased from

1
2
3
4
5
6 104 32,828 (25.8 per 100,000 populations) in 2002 to 17,625 (13.9 per 100,000 populations) in 2016²⁰,
7
8 105 but the central government has constantly been reported of TB outbreaks by local governments at a
9
10 106 rate of approximately 40 events annually over the last decade. This information suggests that TB
11
12 107 transmission might be occurring in some groups, such as homeless people, who constitute a high-
13
14 108 risk group for recent TB transmission in urban areas¹⁴. Considering the steady increase in the
15
16 109 proportion of TB cases among foreign-born individuals in Japan (7.9% of all cases in 2016²¹),
17
18 110 transmission between foreign-born persons and local residents must be monitored. In addition, in
19
20 111 light of Japan's transition toward becoming a low-TB-burden country, understanding TB transmission
21
22 112 patterns has become increasingly important. However, few population-based molecular
23
24 113 epidemiological studies have identified the transmission patterns in Japan and their risk factors.
25
26 114 Additionally, no study has attempted to evaluate the factors predicting the development of large
27
28 115 clusters in Japan.
29
30
31
32
33 116
34
35 117 Therefore, we aimed to estimate the proportion of TB cases attributable to recent transmission, to
36
37 118 identify the risk factors for recent transmission, and to predict the risk factors for the development
38
39 119 of large clusters in an urban setting.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

121 **METHODS**

122 **Study population**

123 We included all culture-positive TB patients notified in Shinjuku City from September 2002 to
124 December 2013 as the eligible study population in this cross-sectional observational study. This study
125 forms part of a population-based study on DNA fingerprinting surveillance of *M. tuberculosis* in
126 Shinjuku City that was started in 2002. Shinjuku City (18.3 km²) is one of the most populous (342,867
127 residents in 2018²²) cities in Tokyo, and its TB notification rate in 2016 was 33.7 per 100,000 people²³,

1
2
3
4
5
6 128 which was higher than the rates in Tokyo and the nation (17.2 and 13.9, respectively²⁰). Experienced
7
8 129 public health nurses at the Shinjuku Public Health Center (PHC) interviewed and collected
9
10 130 information from all culture-positive TB patients at the time of registration using a standardized
11
12 131 questionnaire to avoid possible interviewer bias. The study variables and definitions are described in
13
14
15 132 Table 1.

16
17 133

18 19 134 **Patient and public involvement**

20
21 135 Neither the patients nor the public were involved in the design of this study.

22
23 136

24 25 137 **DNA fingerprinting and genotype cluster**

26
27 138 Clinical isolates from each of the enrolled TB patients were sent to the Research Institute of
28
29 139 Tuberculosis (RIT), Tokyo, where the TB strains were subjected to DNA fingerprinting using insertion
30
31 140 sequence 6110 by RFLP (IS6110-RFLP) analysis²⁴. One clinical isolate per person was used for the
32
33 141 clustering analysis. IS6110-RFLP and spoligotyping are the standard methods used in the Shinjuku
34
35 142 PHC and were available throughout the study period. The Shinjuku PHC switched from RFLP to VNTR
36
37 143 a few years ago, but the RFLP profiles of many TB cases were available. Thus, we employed RFLP due
38
39 144 to the sufficient sample size. A genotype cluster was defined as a group of TB patients whose isolates
40
41 145 showed either (1) ≥ 6 identical IS6110 band patterns or (2) < 6 identical IS6110 band patterns
42
43 146 confirmed by identical spoligotyping patterns. The data collection and genotyping methods were
44
45 147 previously described in detail¹⁴.

46
47 148

48 49 149 **Data analysis**

50
51 150 We calculated the genotype clustering rate by the “n-1 method” according to the formula $\{(n - c)/N\}$,
52
53 151 where N is the total number of cases sampled, c is the number of clusters, and n is the total number
54
55
56
57
58
59
60

1
2
3
4
5
6 152 of cases in the clusters⁹. We also calculated the cumulative clustering rate by calculating the
7
8 153 clustering rate in 2002 and then adding the TB patients every year up to 2013. The characteristics of
9
10 154 clustered cases, which were the cases belonging to any genotype clusters, were compared with those
11
12
13 155 with unique strain patterns through χ^2 tests. We performed univariate logistic regression to identify
14
15 156 risk factors for genotype clustering using odds ratios (ORs) and multivariate logistic regression using
16
17 157 adjusted ORs (aORs). Any potential interactions were assessed using likelihood ratio tests.

18
19 158
20
21 159 Additionally, we compared the characteristics of the first two cases in each genotype cluster to
22
23 160 identify risk factors for the development of a large cluster within three years. For this purpose, a
24
25 161 cluster episode was defined as a newly arising genotype cluster in or after 2003 without any TB cases
26
27 162 of that genotype notified prior to that year. We classified cluster episodes into the following two
28
29 163 groups according to a system developed in a previous study¹⁰: (1) “large clusters within three years”
30
31 164 were cluster episodes with five or more cases (large clusters) occurring within three years and (2)
32
33 165 “small clusters and large clusters after three years” were cluster episodes with two to four cases
34
35 166 (small clusters) and cluster episodes that became large clusters after three years. We identified the
36
37 167 first two cases in each cluster episode based on the notification date and compared their
38
39 168 characteristics between these two groups. We performed univariate and multivariate logistic
40
41 169 regression analyses to identify predictors of the development of large clusters within three years.
42
43
44
45

46 170
47
48 171 A *p*-value of 0.05 was set as the level indicating statistical significance. For variables with more than
49
50 172 5% missing values, the multiple imputation method was considered. The variables used for
51
52 173 multivariate logistic regressions were selected by the stepwise maximum-likelihood estimation with
53
54 174 a significance level of less than 0.2. We used Stata version 12 (Stata Corp., College Station, TX, USA)
55
56 175 for the statistical analyses. Written informed consent was waived because DNA fingerprinting
57
58
59
60

1
2
3
4
5
6 176 analysis forms part of the routine TB control activities in Shinjuku City. However, oral informed
7
8 177 consent was obtained after the PHC staff provided a thorough explanation of the study objectives
9
10 178 and confidentiality. The study protocol was approved by the Institutional Review Board of the
11
12 179 Research Institute of Tuberculosis (RIT/IRB27-9).
13
14
15 180

17 181 **RESULTS**

18 182 **Study population and clustering rate**

19
20
21 183 In total, 1,885 TB patients in Shinjuku City were notified during the study period, and 1,310 were
22
23 184 culture-positive cases (Figure 1). Of these, 285 patients were excluded from the analysis, mainly due
24
25 185 to the unavailability of culture-positive isolates and the lack of implementation of RFLP. As a result,
26
27 186 1,025 (78.2%) patients were included in the analysis. Figure 2 shows the cumulative number of TB
28
29 187 patients and the clustering rates from 2002 to 2013. The number of TB cases gradually increased
30
31 188 over the tested decade. In contrast, the cumulative clustering rates sharply increased in the first four
32
33 189 years, from 10% in 2002 to 28% in 2005, with an average percent change of +43%, and then
34
35 190 continued to increase at a slower rate, from 30% in 2006 to 39% in 2013, with an average percent
36
37 191 change of +4.2%.
38
39
40
41
42 192

43
44 193 We identified a total of 113 genotype clusters consisting of 515 patients (Figure 1). The genotype
45
46 194 clustering rate was 39.2%, and the average cluster size was 4.56 cases (range 2-30). Fifty-seven
47
48 195 (50.4%) genotype clusters consisted of only two TB patients, and 36 (31.9%) genotype clusters had
49
50 196 at least five TB patients. We further investigated the homelessness status and place of birth of the
51
52 197 patients in the genotype clusters. Of the 113 genotype clusters, 45 (39.8%) comprised only
53
54 198 nonhomeless individuals, seven (6.2%) included only homeless individuals, and 61 (54.0%) contained
55
56 199 both homeless and nonhomeless individuals (mixed cluster). We compared the characteristics of the
57
58
59
60

1
2
3
4
5
6 200 nonhomeless patients in the clusters of only nonhomeless patients with those in the mixed clusters,
7
8 201 and although the finding was not statistically significant (Pearson χ^2 test, $p=0.17$), the proportion of
9
10 202 nonhomeless patients receiving public assistance in the latter group (13.8%) was higher than that in
11
12 203 the former group (8.8%). No differences in sex, age, and place of birth were found between the two
13
14 204 groups. Of the 113 genotype clusters, 94 (83.2%) consisted of only individuals born in Japan, two
15
16 205 (1.8%) consisted of only foreign-born individuals, and 17 (15.0%) consisted of both individuals born
17
18 206 in Japan and foreign-born individuals.
19
20
21
22

207

208 **Factors associated with genotype clustering**

209 The clustered cases were significantly more likely to consist of male individuals (OR=1.62, 95%
210 CI=1.20-2.19), Japan-born individuals (OR=3.74, 95% CI=2.25-6.44), individuals receiving public
211 assistance (OR=2.25, 95% CI=1.69-3.00), homeless individuals (OR=2.45, 95% CI=1.80-3.34),
212 individuals who misuse alcohol (OR=1.37, 95% CI=1.02-1.83), individuals engaging in fulltime work
213 (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
214 individuals aged 15-59 years (OR=2.05, 95% CI=1.43-2.94) (Table 2). A significant interaction among
215 the explanatory variables was not detected. The multivariate analysis demonstrated that the factors
216 associated with genotype clustering were age <40 years (aOR=1.73, 95% CI=1.23-2.44), born in Japan
217 (aOR=3.90, 95% CI=2.27-6.72), working fulltime (aOR=1.63, 95% CI=1.17-2.27), having part-
218 time/daily work (aOR=2.20, 95% CI=1.35-3.58), receiving public assistance (aOR=1.81, 95% CI=1.15-
219 2.84), and homelessness (aOR=1.63, 95% CI=1.02-2.62)(Table 3).

220

221 **Factors associated with large genotype clustering within three years**

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

1
2
3
4
5
6 224 48.3% (14/29) of the large genotype clusters, and 90 clusters were “small clusters and large clusters
7
8 225 after three years”. The univariate analysis indicated that clusters with registration intervals of 0-2
9
10 226 months were 9.51 times more likely to become large genotype clusters within three years compared
11
12 227 with clusters with registration intervals of ≥ 12 months (Table 4). After selecting variables using the
13
14 228 stepwise method, only the “registration interval” variable remained for the multivariate model.
15
16
17 229

19 230 **DISCUSSION**

21 231 In this long-term population-based study, we included 1,025 patients, identified a total of 113
22
23 232 genotype clusters, and obtained a genotype clustering rate of 39.2%. Our results indicated that the
24
25 233 clustered cases were more likely to have certain socioeconomic predictive factors, namely, being
26
27 234 homeless, receiving public assistance, and having an unstable job, at the time of tuberculosis
28
29 235 diagnosis. A shorter registration interval between the first two cases was a statistically significant
30
31 236 predictor of the development of a large genotype cluster within three years.
32
33
34
35 237

37 238 **Clustering rate**

39 239 We identified 515 genotype clustered cases and estimated a clustering rate of 39.2%. The rate was
40
41 240 the same as the pooled clustering rate (40.9%) obtained in a previous meta-analysis of population-
42
43 241 based studies of countries with a low TB incidence¹⁹ but differed from previous estimates obtained
44
45 242 in Japanese studies, which were 27.6% in Shinjuku and 24.6% in Osaka^{14,25}. Because the meta-
46
47 243 regression analysis clarified that longer study durations are associated with an increased clustering
48
49 244 rate¹⁹, this difference could be due to shorter study durations combined with the smaller sample
50
51 245 sizes of the previous studies (388 patients in five years and 195 patients in one year, respectively). In
52
53 246 our study, as expected, the cumulative clustering rate rapidly increased in the first four years and
54
55 247 increased more slowly thereafter, which is similar to the trend observed in the previous studies^{26,27}.
56
57
58
59
60

248

249 Factors associated with genotype clustering

250 Our results indicated that the clustered cases were more likely to have socioeconomic predictive
251 factors, namely, being homeless, receiving public assistance, and having an unstable job, at the time
252 of TB diagnosis. Similarly, previous studies suggested that being homeless significantly contributed
253 to clustering in Shinjuku City¹⁴ and other counties¹⁹. In our study, more than half of the genotype
254 clusters were mixtures of nonhomeless and homeless patients. Moreover, the nonhomeless patients
255 in the mixed clusters tended to be financially unstable, and a higher proportion of these patients
256 were receiving public assistance compared with the proportion among clusters of only nonhomeless
257 cases, which could imply that relatively poor nonhomeless patients share activity spaces with
258 homeless patients, such as urban areas around the large train stations that were reported to be
259 significant hotspots for homeless patients in Shinjuku City²⁸. These findings could suggest that
260 contact investigations of homeless TB patients need to be actively expanded to possible contact
261 persons who are not homeless, particularly those who are facing financial difficulty.

262

263 A meta-analysis based on studies conducted in European countries where foreign-born patients
264 substantially contribute to TB epidemiology found that the proportion of mixed clusters composed
265 of native and foreign-born patients ranged from 0% to 36.5% and concluded that foreign-born
266 patients did not have a significant influence on TB in the native population²⁹. In our study, the
267 proportion of mixed clusters (15.0%) fell into this range. Thus, the impact of TB transmission between
268 native and foreign-born populations likely remains limited in this urban setting³⁰. However,
269 considering the recent increase in immigrant TB patients in urban cities, TB transmission between
270 native and foreign-born populations needs to be closely monitored.

271

272 **Factors associated with large genotype clustering within three years**

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

285 **Limitations**

286 Our study has some limitations. First, the study population consisted only of TB patients living in
287 Shinjuku City. Considering the large population flow in and out of the city, as mentioned above, we
288 potentially missed patients living outside of the city who shared TB strain types with patients living
289 in the city. In fact, previous Japanese studies reported clusters with TB patients living across broad
290 geographic areas³¹. Consequently, we may have underestimated the identified genotype clusters.
291 Second, even the existence of TB patients with identical genotyping patterns may not suggest recent
292 transmission if the strain is a nationwide endemic TB strain³², which could have led to an
293 overestimated clustering rate. Third, IS6110 RFLP has relatively lower discriminatory power
294 compared with VNTR³³ and whole-genome sequencing^{34,35}, which might have led to overestimation.
295 Lastly, information of epidemiological linkage among TB patients was not available in our study.

1
2
3
4
5
6 296 Therefore, we could not assess and discuss the current practices involving epidemiological
7
8 297 investigations done by the public health center, which could weaken the programmatic implications
9
10 298 of our results.

11
12
13 299

14
15 300 **Conclusion**

16
17 301 This study constitutes a one of the longest-term studies on the molecular epidemiology of notified
18
19 302 TB patients in a large Asian urban setting. Our results indicated that a large proportion of culture-
20
21 303 positive TB patients were involved in the recent TB transmission chain. Homeless persons were found
22
23 304 to be involved in more than half of the genotype clusters. Foreign-born persons continue to have a
24
25 305 limited impact on TB transmission in the Japanese urban setting, but considering recent increases in
26
27 306 foreign-born TB patients, transmission between native and foreign-born populations should be
28
29 307 routinely evaluated. Intensified public health interventions, such as active case findings, should focus
30
31 308 on those with socioeconomic risk factors that are significantly associated with TB transmission and
32
33 309 clusters with shorter registration intervals between the first two cases because these variables could
34
35 310 serve as predictors of the development of large clusters within three years.

36
37
38
39
40 311
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 312 **ACKNOWLEDGMENTS**
7

8 313 We would like to thank the Shinjuku Public Health Center staff for collecting the data.
9

10 314

11
12 315 **Author Statement:** Authors roles

13
14
15 316 KIz: conception of study, statistical analysis, interpretation of data, and drafting and finalization of

16
17 317 the manuscript

18
19 318 YM: genotyping analysis, interpretation of data, and finalization of the manuscript

20
21 319 KU: conception of study, interpretation of data, and finalization of the manuscript

22
23 320 AK: conception of study, collection of data, and finalization of the manuscript

24
25 321 KIs: conception of study, collection of data, and finalization of the manuscript

26
27 322 SK: conception of study, collection of data, and finalization of the manuscript

28
29 323 TT: interpretation of data and finalization of the manuscript

30
31 324 AO: conception of study, interpretation of data, and drafting and finalization of the manuscript

32
33 325

34
35 326

36
37 327
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

328 References

- 329 1. World Health Organization. *Global Tuberculosis Report 2018*. Geneva; 2018. doi:ISBN 978 92 4
330 156539 4.
- 331 2. Lönnroth K, Migliori GB, Abubakar I, et al. Towards tuberculosis elimination: An action
332 framework for low-incidence countries. *Eur Respir J*. 2015;45(4):928-952.
333 doi:10.1183/09031936.00214014.
- 334 3. de Vries G, Aldridge RW, Cayla J a, et al. Epidemiology of tuberculosis in big cities of the
335 European Union and European Economic Area countries. *Euro Surveill Bull Eur sur les Mal*
336 *Transm = Eur Commun Dis Bull*. 2014;19(9):1-8.
- 337 4. Hest NA, Aldridge RW, de Vries G, et al. Tuberculosis control in big cities and urban risk groups in
338 the European Union: A consensus statement. *Euro Surveill*. 2014;19(9):1-13. doi:10.2807/1560-
339 7917.ES2014.19.9.20728.
- 340 5. Jereb JA. Progressing towards tuberculosis elimination in low-incidence areas of the United
341 States. Recommendations of the Advisory Council for the Elimination of Tuberculosis. *MMWR*
342 *Recomm reports*. 2002;51(RR-5):1-14.
- 343 6. Borgdorff MW, van Soolingen D. The re-emergence of tuberculosis: What have we learnt from
344 molecular epidemiology? *Clin Microbiol Infect*. 2013;19(10):889-901. doi:10.1111/1469-
345 0691.12253.
- 346 7. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City. An analysis by
347 DNA fingerprinting and conventional epidemiologic methods. 1994;330(D):1710-1716.
- 348 8. van Soolingen D, Borgdorff MW, de Haas PE, et al. Molecular epidemiology of tuberculosis in the
349 Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis*. 1999;180:726-736.
350 doi:10.1086/314930.
- 351 9. PM S, PC H, SP S, et al. The epidemiology of tuberculosis in San Francisco. A population-based

- 1
2
3
4
5
6 352 study using conventional and molecular methods. *N Engl J Med.* 1994.
7
8 353 10. Hamblion EL, Le Menach A, Anderson LF, et al. Recent TB transmission, clustering and predictors
9
10 354 of large clusters in London, 2010-2012: Results from first 3-years of universal MIRU-VNTR strain
11
12 355 typing. *Thorax.* 2016;71(8):749-756. doi:10.1136/thoraxjnl-2014-206608.
13
14
15 356 11. Driver CR, Macaraig M, McElroy PD, et al. Which patients' factors predict the rate of growth of
16
17 357 *Mycobacterium tuberculosis* clusters in an urban community? *Am J Epidemiol.* 2006;164(1):21-
18
19 358 31. doi:10.1093/aje/kwj153.
20
21 359 12. Kik S V., Verver S, Van Soolingen D, et al. Tuberculosis outbreaks predicted by characteristics of
22
23 360 first patients in a DNA fingerprint cluster. *Am J Respir Crit Care Med.* 2008;178(1):96-104.
24
25 361 doi:10.1164/rccm.200708-1256°C.
26
27
28 362 13. Althomsons SP, Kammerer JS, Shang N, Navin TR. Using Routinely Reported Tuberculosis
29
30 363 Genotyping and Surveillance Data to Predict Tuberculosis Outbreaks. *PLoS One.* 2012;7(11):1-8.
31
32 364 doi:10.1371/journal.pone.0048754.
33
34
35 365 14. Ohkado A, Nagamine M, Murase Y, et al. Molecular epidemiology of *Mycobacterium*
36
37 366 *tuberculosis* in an urban area in Japan, 2002-2006. *Int J Tuberc Lung Dis.* 2008;12(5):548-554.
38
39 367 15. Ohkado A, Murase Y, Mori M, et al. Transmission of specific genotype streptomycin resistant
40
41 368 strains of *Mycobacterium tuberculosis* in the Tokyo Metropolitan Area in Japan. *BMC Infect Dis.*
42
43 369 2009;9:138. doi:10.1186/1471-2334-9-138.
44
45
46 370 16. Xu G, Mao X, Wang J, Pan H. Clustering and recent transmission of *Mycobacterium tuberculosis*
47
48 371 in a Chinese population. *Infect Drug Resist.* 2018;11:323-330.
49
50 372 17. Yang C, Shen X, Peng Y, et al. Transmission of *Mycobacterium tuberculosis* in China: a
51
52 373 population-based molecular epidemiologic study. *Clin Infect Dis.* 2015;61(2):219-227.
53
54 374 doi:10.1093/cid/civ255.
55
56
57 375 18. Mears J, Abubakar I, Cohen T, McHugh TD, Sonnenberg P. Effect of study design and setting on

- 1
2
3
4
5
6 376 tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable
7
8 377 Number Tandem Repeats (MIRU-VNTR): A systematic review. *BMJ Open*. 2015;5(1).
9
10 378 doi:10.1136/bmjopen-2014-005636.
11
12
13 379 19. Fok A, Numata Y, Schulzer M, FitzGerald MJ. Risk factors for clustering of tuberculosis cases: A
14
15 380 systematic review of population-based molecular epidemiology studies. *Int J Tuberc Lung Dis*.
16
17 381 2008;12(5):480-492.
18
19 382 20. The Research Institute of Tuberculosis J. Statistics of TB 2016.
20
21 383 <http://www.jata.or.jp/rit/ekigaku/en/statistics-of-tb/>. Accessed June 20, 2018.
22
23
24 384 21. Tuberculosis Surveillance Center. *Tuberculosis in Japan – Annual Report 2017*. Tokyo, Japan:
25
26 385 Department of Epidemiology and Clinical Research, the Research Institute of Tuberculosis; 2017.
27
28 386 22. Shinjuku city. Description of Shinjuku.
29
30 387 http://www.foreign.city.shinjuku.lg.jp/en/aramashi/aramashi_1/. Accessed January 9, 2018.
31
32
33 388 23. Shinjuku City. *Tuberculosis Statistics in Shinjuku City 2017 [Japanese]*.; 2018.
34
35 389 24. Van Embden JDA, Cave MD, Crawford JT, et al. Strain identification of Mycobacterium
36
37 390 tuberculosis by DNA fingerprinting: Recommendations for a standardized methodology. *J Clin*
38
39 391 *Microbiol*. 1993;31(2):406-409. doi:10.1128/JCM.39.4.1683.2001.
40
41
42 392 25. Wada T, Maeda S, Hase A, Kobayashi K. Evaluation of variable numbers of tandem repeat as
43
44 393 molecular epidemiological markers of Mycobacterium tuberculosis in Japan. *J Med Microbiol*.
45
46 394 2007;56(8):1052-1057. doi:10.1099/jmm.0.46990-0.
47
48
49 395 26. Ellis B a, Crawford JT, Braden CR, McNabb SJN, Moore M, Kammerer S. Molecular epidemiology
50
51 396 of tuberculosis in a sentinel surveillance population. *Emerg Infect Dis*. 2002;8(11):1197-1209.
52
53 397 doi:10.3201/eid0811.020403.
54
55
56 398 27. Glynn JR, Crampin AC, Yates MD, et al. The Importance of Recent Infection with *Mycobacterium*
57
58 399 *tuberculosis* in an Area with High HIV Prevalence: A Long-Term Molecular Epidemiological Study
59
60

- 1
2
3
4
5
6 400 in Northern Malawi. *J Infect Dis.* 2005;192(3):480-487. doi:10.1086/431517.
- 7
8 401 28. Izumi K, Ohkado A, Uchimura K, et al. Detection of tuberculosis infection hotspots using activity
9
10 402 spaces based spatial approach in an urban Tokyo, from 2003 to 2011. *PLoS One.* 2015;10(9):1-
11
12 403 16. doi:10.1371/journal.pone.0138831.
- 13
14
15 404 29. Sandgren A, Schepisi MS, Sotgiu G, et al. Tuberculosis transmission between foreign- and native-
16
17 405 born populations in the EU/EEA: A systematic review. *Eur Respir J.* 2014;43(4):1159-1171.
18
19 406 doi:10.1183/09031936.00117213.
- 20
21
22 407 30. MURASE Y, OHKADO A, WATANABE Y, et al. Transmission dynamics of mycobacterium
23
24 408 tuberculosis between foreign-nationals and Japanese tuberculosis patients living in Shinjuku-
25
26 409 City, Tokyo, Japan. *Kekkaku.* 2017;92(5):431-439.
- 27
28 410 31. Murase Y, Izumi K, Ohkado A, et al. Prediction of local transmission of mycobacterium
29
30 411 tuberculosis isolates of a predominantly Beijing lineage by use of a variable-number tandem-
31
32 412 repeat typing method incorporating a consensus set of hypervariable loci. *J Clin Microbiol.*
33
34 413 2018;56(1):1-11. doi:10.1128/JCM.01016-17.
- 35
36
37 414 32. Wada T, Iwamoto T, Tamaru A, et al. Clonality and micro-diversity of a nationwide spreading
38
39 415 genotype of *Mycobacterium tuberculosis* in Japan. *PLoS One.* 2015;10(3):1-13.
40
41 416 doi:10.1371/journal.pone.0118495.
- 42
43
44 417 33. Murase Y, Mitarai S, Sugawara I, Kato S, Maeda S. Promising loci of variable numbers of tandem
45
46 418 repeats for typing Beijing family *Mycobacterium tuberculosis*. *J Med Microbiol.* 2008;57(7):873-
47
48 419 880. doi:10.1099/jmm.0.47564-0.
- 49
50
51 420 34. Nikolayevskyy V, Kranzer K, Niemann S, Drobniewski F. Whole genome sequencing of
52
53 421 *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks : A
54
55 422 systematic review. *Tuberculosis.* 2016;98:77-85. doi:10.1016/j.tube.2016.02.009.
- 56
57
58 423 35. Jajou R, Neeling A De, Hunen R Van, et al. Epidemiological links between tuberculosis cases
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

424 identified twice as efficiently by whole genome sequencing than conventional molecular typing :

425 A population-based study. *PLoS One*. 2018;13(4):e0195413.

426

427

For peer review only

428 **TABLE**

429 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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

430
431
432

For peer review only

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		

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

435 RFLP: restriction fragment length polymorphism, OR: odds ratio, CI: confidence interval, TB:

436 tuberculosis, DM: diabetes mellitus, *p<0.05, **p<0.01, ***p<0.001

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

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

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

443

444

445 Table 3. Factors associated with TB genotype clustering; multivariate logistic regression analysis,
 446 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

447 RFLP: restriction fragment length polymorphism, aOR: adjusted odds ratio, CI: confidence interval,

448 *p<0.05, **p<0.01, ***p<0.001

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

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

1
2
3
4
5
6 452 §Alcohol misuse refer to excessive drinking, as judged by the public health nurses conducting the
7
8 453 interviews.
9

10 454
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

455 Table 4. Factors associated with large genotype clusters within three years using the characteristics
 456 of the first two cases in each TB genotype cluster; univariable logistic regression, RFLP, Shinjuku,
 457 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					

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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		

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

461 RFLP: restriction fragment length polymorphism, OR: odds ratio, aOR: adjusted odds ratio, CI:
 462 confidence interval, TB: tuberculosis, DM: diabetes mellitus, Ref: reference, NA: not applicable,

463 * $p < 0.05$, ** $p < 0.01$

464 +“Large clusters within three years” refers to cluster episodes with five or more cases (large clusters)
 465 within three years.

466 +“Small clusters and large clusters after three years” refers to cluster episodes with two to four cases
 467 (small clusters) and cluster episodes that became large clusters after three years.

468

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

469

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

470 **Figure legends**

471

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

474 RFLP: restriction fragment length polymorphism, TB: tuberculosis

475

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

477 RFLP: restriction fragment length polymorphism

478

For peer review only

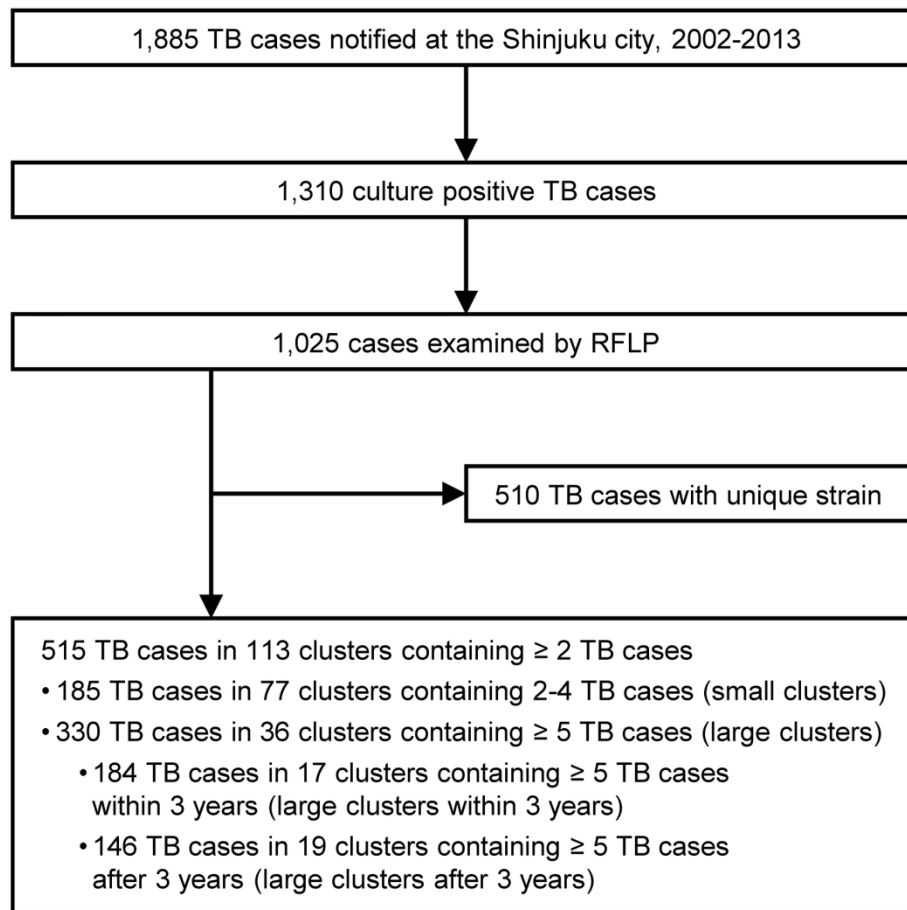


Figure 1. Number of reported cases of TB, including culture-positive cases, strain-typed cases and genotype clusters, Shinjuku, 2002–2013

RFLP: restriction fragment length polymorphism, TB: tuberculosis

90x90mm (600 x 600 DPI)

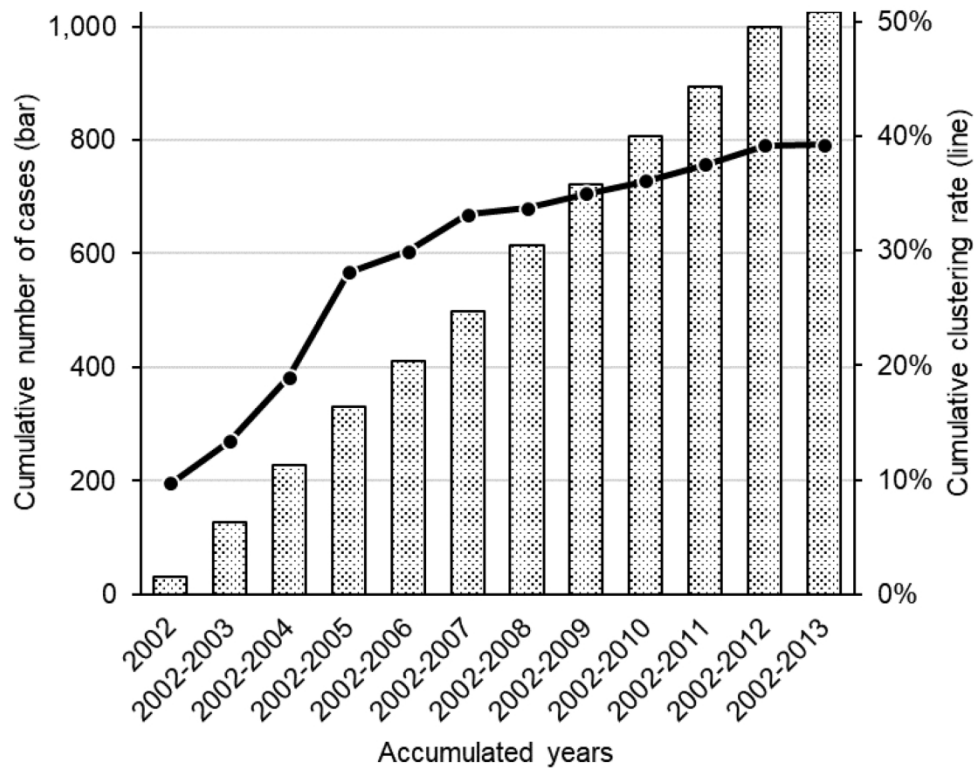


Figure 2. Cumulative clustering rate, RFLP, Shinjuku 2002–2013
 RFLP: restriction fragment length polymorphism

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 <i>*We took all eligible study participants in Shinjuku city.</i>
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.