CrossMark
click for updates

Multiple large clusters of tuberculosis in London: a cross-sectional analysis of molecular and spatial data

Catherine M. Smith[1], Helen Maguire[2,3], Charlotte Anderson[2], Neil Macdonald[2] and Andrew C. Hayward[1]

Affiliations: [1]UCL Dept of Infectious Disease Informatics, Farr Institute of Health Informatics Research, University College London, London, UK. [2]Field Epidemiology Service – South East and London, Public Health England, London, UK. [3]Research Dept of Infection and Population Health, Centre for Infectious Disease Epidemiology, University College London, London, UK.

Correspondence: Catherine M. Smith, UCL Dept of Infectious Disease Informatics, Farr Institute of Health Informatics Research, University College London, London NW1 2DA, UK. E-mail: catherine.smith.13@ucl.ac.uk

ABSTRACT Large outbreaks of tuberculosis (TB) represent a particular threat to disease control because they reflect multiple instances of active transmission. The extent to which long chains of transmission contribute to high TB incidence in London is unknown. We aimed to estimate the contribution of large clusters to the burden of TB in London and identify risk factors.

We identified TB patients resident in London notified between 2010 and 2014, and used 24-locus mycobacterial interspersed repetitive units–variable number tandem repeat strain typing data to classify cases according to molecular cluster size. We used spatial scan statistics to test for spatial clustering and analysed risk factors through multinomial logistic regression.

TB isolates from 7458 patients were included in the analysis. There were 20 large molecular clusters (with n>20 cases), comprising 795 (11%) of all cases; 18 (90%) large clusters exhibited significant spatial clustering. Cases in large clusters were more likely to be UK born (adjusted odds ratio 2.93, 95% CI 2.28–3.77), of black-Caribbean ethnicity (adjusted odds ratio 3.64, 95% CI 2.23–5.94) and have multiple social risk factors (adjusted odds ratio 3.75, 95% CI 1.96–7.16).

Large clusters of cases contribute substantially to the burden of TB in London. Targeting interventions such as screening in deprived areas and social risk groups, including those of black ethnicities and born in the UK, should be a priority for reducing transmission.

@ERSpublications
Large clusters contribute substantially to the burden of tuberculosis in London, indicating ongoing transmission http://ow.ly/3xk23068P6w

Introduction

In countries with low incidence of tuberculosis (TB) such as the UK, highest rates are often found in large cities [1]. The rate of TB in London in 2014, for example, was 30 per 100 000 population compared with 12 per 100 000 in the whole of England [2]. This high incidence has led to the city being described as the "TB capital of Western Europe" [3].

Large outbreaks of TB represent a particular threat to control because they reflect multiple instances of active transmission. Such large outbreaks have occurred previously in London and other large cities [4–9]. However, identification of outbreaks of TB is difficult, as it requires cases resulting from active transmission to be distinguished from those resulting from reactivation of latent disease with absent or limited onward transmission. The extent to which they contribute to the overall disease burden is therefore not known.

Molecular strain typing provides one means of linking cases that may be part of outbreaks. Cases that share a molecular strain type may be linked through transmission and therefore form part of large outbreaks, although they may also reflect common endemic strains. In England, prospective molecular strain typing has been conducted since 2010 using the 24-locus mycobacterial interspersed repetitive units–variable number tandem repeat (MIRU-VNTR) method.

Spatial analyses provide another means of investigating potential links between cases of infectious disease [10]. Tests of spatial clustering, for example, can be used to identify cases that occur closer together in space than would be expected by chance. They can therefore be used in combination with molecular data to assess evidence for recent transmission in investigations of TB clusters.

An analysis of the first 3 years of MIRU-VNTR data in London showed that 46% of cases were part of a molecular cluster and identified risk factors for clustering [11]. It also showed that cluster size ranged from two to 55 cases, and that over half of the clusters had only two cases. However, the study did not determine whether risk factors varied by cluster size or assess spatial clustering.

In this study, we investigated the size and distribution of molecular clusters of TB in London between 2010 and 2014 using routine molecular strain typing and epidemiological data. We aimed to quantify the contribution of large molecular clusters to the burden of TB in the city, describe the characteristics of cases by cluster size and identify risk factors. We also aimed to assess evidence for transmission in large molecular clusters by testing for spatial clustering.

Methods

Study population and data sources

This was a cross-sectional analysis of patients notified with TB between January 1, 2010 and December 31, 2014 resident in London. Data were extracted from the Enhanced Tuberculosis Surveillance (ETS) system, a national online register for real-time case reporting that is run by Public Health England (PHE). This system includes demographic (age, sex, ethnic group, country of birth, time since entry to the UK and occupation) and clinical (site of disease, sputum smear status, history of TB disease and treatment, drug sensitivity, and whether the case spent time as a hospital inpatient) characteristics of patients. It also includes patient residential locations and information on social risk factors for TB (whether the patient has a history of homelessness, or problems with illicit drug or alcohol use). Surveillance data from ETS is routinely matched to the National Tuberculosis Strain Typing Service to provide MIRU-VNTR molecular clustering data.

An estimate of level of social deprivation (the index of multiple deprivation (IMD)) is also included in the ETS system. This is obtained through matching of residential postcodes to Lower Layer Super Output Areas (LSOAs), a geographic hierarchy used in England and Wales each encompassing a mean population of 1500. The IMD is a measure of relative deprivation at the LSOA level in England and is based on seven domains of deprivation: 1) income, 2) employment, 3) crime, 4) living environment, 5) barriers to housing and services, 6) health and disability, and 7) education, skills and training [12]. Low ranks indicate higher levels of deprivation. We converted IMD ranks into London-level deprivation quintiles, with the lowest quintile representing the most deprived areas.

Ethical approval was not required for this study because it was based on PHE routine surveillance data. PHE has Health Research Authority approval to hold and analyse national surveillance data for public health purposes.

Molecular clustering analysis

We categorised cases as unique or part of a molecular cluster and by the number of cases in the molecular cluster. We used the PHE convention for assigning cluster status. Molecular clusters were groups of two or

more cases that shared an identical MIRU-VNTR strain type with another case notified in the study region during the study period. Unique cases were individuals whose strain type did not cluster with another case. We excluded cases who did not have an isolate typed by MIRU-VNTR with at least 23 loci and those whose molecular strain type was unique within the study area but shared a molecular strain type with another case in England.

We described the distribution of molecular clusters by size (number of cases in the cluster) and calculated the proportion of cases which were part of clusters. We identified successive cases reported in a molecular cluster using case notification dates and calculated the median and interquartile range (IQR) number of days between successive cases in molecular clusters by cluster size.

In this analysis, we aimed to investigate risk factors for cases belonging to a molecular clusters of different sizes. We therefore categorised cases according to the size of the molecular cluster (not clustered (unique) cases, n=2 cases, n=3–20 cases and n>20 cases). In situations such as this, in which the outcome of interest is a categorical variable, a multinomial logistic regression model can be used [13]. This is an extension of the simple logistic regression model which is used for dichotomous outcomes. Coefficients resulting from the multinomial model are interpreted in a similar way to the odds ratios (ORs) derived from a logistic regression model.

We investigated associations at single-variable analysis and included variables with an association of p<0.2 in the initial multivariable model. A backwards stepwise approach was then used to eliminate variables which did not contribute significantly to produce a final model. LSOA of residence was included as a random effect in models which included the IMD to account for the hierarchical level at which this variable was measured. Social risk factors were considered separately, and as a cumulative count of these risk factors at single-variable analysis and as a count at multivariable analysis.

Spatial clustering analysis

We used spatial scan statistics to assess spatial clustering within molecular clusters, implemented using SaTScan software (www.satscan.org). We tested the hypothesis that cases in large molecular clusters (n>20 cases) were closer together in space than the underlying spatial distribution of TB cases. For each large molecular cluster, we therefore performed a spatial scan under the Bernoulli (case–control) model, using the locations of all other TB cases as controls.

We aimed to identify areas with evidence of local transmission and therefore set the maximum radius of the spatial window at 5 km and identified clusters with a p-value of <0.05 which encompassed at least 10 cases. We plotted the locations of significant spatial clusters for each molecular cluster overlaid on a smoothed incidence map of the relative distribution of the given molecular cluster compared with all other TB cases. These maps were generated through kernel density estimation using a Gaussian kernel of bandwidth 5 km.

Statistical analyses were performed using R version 3.2.3 (www.r-project.org).

Results

Between 2010 and 2014, a total of 15 670 cases of TB were notified in London. Of these, 8148 (52%) cases were successfully typed by MIRU-VNTR with at least 23 loci defined, whilst 6241 (40%) were not culture confirmed and 1281 (8%) were not typed, and therefore excluded from this analysis (figure 1). A further 690 cases were also excluded because they clustered only with cases that were not resident within the study
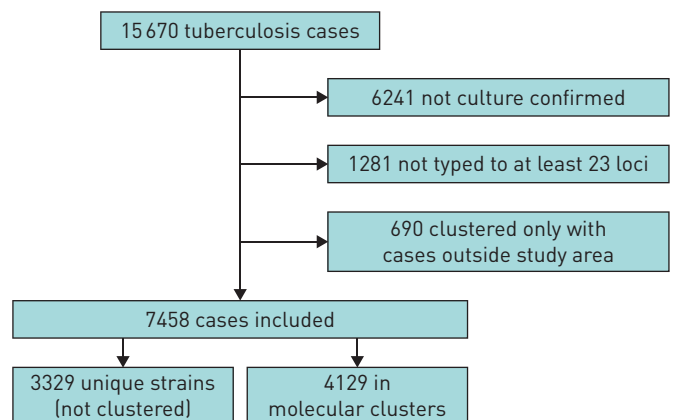

FIGURE 1 Cases included in analysis of molecular clusters of tuberculosis in London (2010–2014).

Cases in cluster	Cases	Clusters
1 (not clustered)	3329 (44.6)	
2	1044 (14.0)	522 (52.4)
3–20	2290 (30.7)	454 (45.6)
>20	795 (10.7)	20 (2.0)
Total	7458 (100)	996 (100)

TABLE 1 Distribution of tuberculosis cases and molecular clusters in London (2010–2014), by cluster size

Data are presented as n (%).

area. This study therefore included 7458 TB cases with a molecular strain type, of which 4129 (55%) were part of 996 molecular clusters and 3329 (45%) had a unique strain.

Cluster size and time between cases
Cluster size ranged from two to 102 cases, with a median of n=2 cases. There were 20 clusters with n>20 cases, including 795 (11%) of all cases (table 1). Over half of the clusters (522 (53%)) comprised pairs of cases, but a larger proportion of cases were in the 454 clusters of n=3–20 cases (2290 (31% of cases)).

Successive cases in clusters were defined using notification dates. Figure 2 displays the distribution of median intervals between successive cases in each cluster by cluster size. Overall, the median (IQR) time between successive cases in a cluster was 114 (32–323) days. For cases in clusters of n>20 cases the median (IQR) time was 23 (8–54) days, for cases in clusters of n=3–20 cases it was 149 (54–335) days) and for clusters of n=2 cases it was 406 (162–752) days.

Factors associated with large clusters
Baseline characteristics of TB cases according to the number of cases in the cluster are shown in table 2 and results of the single-variable multinomial logistic regression analysis are presented in table 3.

For each exposure, an OR was calculated for each of the three cluster size outcomes (n=2, 3–20 and >20 cases), with cases not in a cluster representing the comparison group. For example, the unadjusted ORs for being born in the UK were 4.11 (for cases in clusters of n>20 cases), 2.55 (for cases in clusters of n=3–20 cases) and 1.77 (for cases in clusters of n=2 cases). This means that the odds of cases being in the largest clusters *versus* not being in a cluster for those born in the UK were 4.11 times that of those not born in
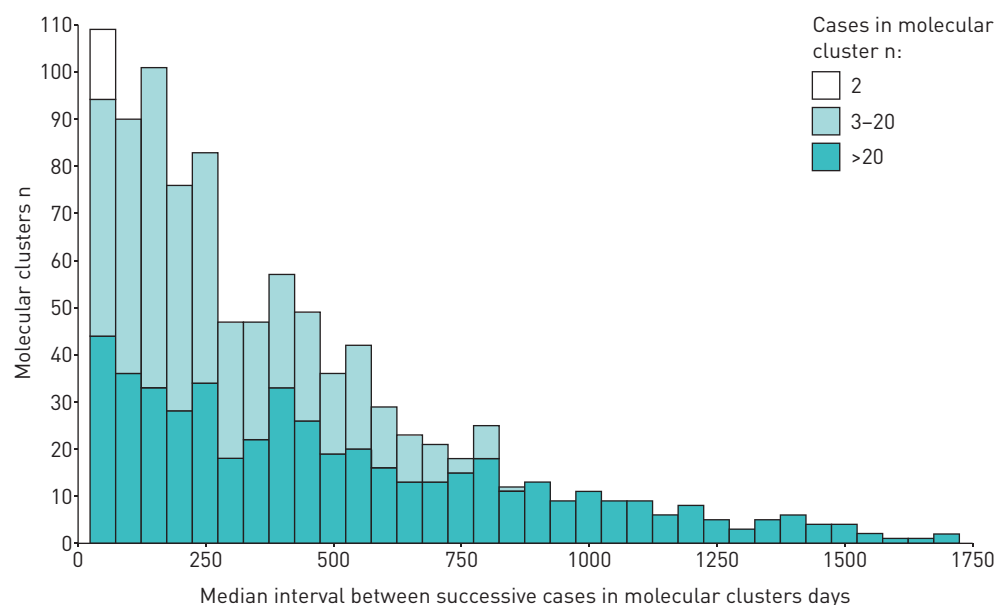

FIGURE 2 Median interval between successive tuberculosis cases in molecular clusters in London (2010–2014), by cluster size.

TABLE 2 Baseline characteristics of tuberculosis cases in molecular clusters of different sizes in London (2010–2014)					
	All cases	Not clustered	Cases in cluster n		
			2	3–20	>20
Sex					
Female	2911	1320 (45.3)	430 (14.8)	874 (30.0)	287 (9.9)
Male	4546	2008 (44.2)	614 (13.5)	1416 (31.1)	508 (11.2)
Age group years					
0–14	158	43 (27.2)	24 (15.2)	60 (38.0)	31 (19.6)
15–44	5284	2346 (44.4)	731 (13.8)	1628 (30.8)	579 (11.0)
45–64	1369	585 (42.7)	207 (15.1)	433 (31.6)	144 (10.5)
⩾65	647	355 (54.9)	82 (12.7)	169 (26.1)	41 (6.3)
Ethnic group					
White	829	308 (37.2)	119 (14.4)	277 (33.4)	125 (15.1)
Black-Caribbean	258	49 (19.0)	38 (14.7)	113 (43.8)	58 (22.5)
Black-African	1722	609 (35.4)	242 (14.1)	614 (35.7)	257 (14.9)
Black-Other	97	28 (28.9)	15 (15.5)	37 (38.1)	17 (17.5)
Indian	2185	1126 (51.5)	303 (13.9)	615 (28.1)	141 (6.5)
Pakistani	673	313 (46.5)	96 (14.3)	181 (26.9)	83 (12.3)
Bangladeshi	365	246 (67.4)	38 (10.4)	69 (18.9)	12 (3.3)
Chinese	86	49 (57.0)	11 (12.8)	21 (24.4)	5 (5.8)
Mixed/other	1177	561 (47.7)	176 (15.0)	347 (29.5)	93 (7.9)
Place of birth					
Non-UK	6223	2990 (48.0)	876 (14.1)	1801 (28.9)	556 (8.9)
UK	1150	301 (26.2)	156 (13.6)	463 (40.3)	230 (20.0)
Time since entry to UK years					
0–1	1095	540 (49.3)	159 (14.5)	298 (27.2)	98 (9.0)
2–4	1391	740 (53.2)	202 (14.5)	347 (24.9)	102 (7.3)
5–9	1156	538 (46.5)	162 (14.0)	339 (29.3)	117 (10.1)
⩾10	1803	785 (43.5)	261 (14.5)	583 (32.3)	174 (9.7)
Occupation					
Other	2465	1146 (46.5)	342 (13.9)	759 (30.8)	218 (8.8)
None	2615	1129 (43.2)	367 (14.0)	788 (30.1)	331 (12.7)
Education	1074	452 (42.1)	148 (13.8)	342 (31.8)	132 (12.3)
Healthcare	275	137 (49.8)	44 (16.0)	81 (29.5)	13 (4.7)
Pulmonary disease					
No	3006	1553 (51.7)	389 (12.9)	815 (27.1)	249 (8.3)
Yes	4452	1776 (39.9)	655 (14.7)	1475 (33.1)	546 (12.3)
Sputum smear					
Negative	2371	1066 (45.0)	338 (14.3)	734 (31.0)	233 (9.8)
Positive	2062	745 (36.1)	314 (15.2)	718 (34.8)	285 (13.8)
Previous diagnosis					
No	6821	3088 (45.3)	941 (13.8)	2078 (30.5)	714 (10.5)
Yes	354	124 (35.0)	57 (16.1)	125 (35.3)	48 (13.6)
Previous treatment					
No	13	6 (46.2)	1 (7.7)	6 (46.2)	
Yes	261	82 (31.4)	41 (15.7)	98 (37.5)	40 (15.3)
Drug resistance#					
No	6738	3033 (45.0)	921 (13.7)	2114 (31.4)	670 (9.9)
Yes	667	276 (41.4)	111 (16.6)	161 (24.1)	119 (17.8)
Inpatient					
No	4649	2100 (45.2)	636 (13.7)	1441 (31.0)	472 (10.2)
Yes	2731	1190 (43.6)	397 (14.5)	831 (30.4)	313 (11.5)
Homeless					
No	6917	3116 (45.0)	969 (14.0)	2135 (30.9)	697 (10.1)
Yes	294	96 (32.7)	44 (15.0)	85 (28.9)	69 (23.5)
Drug use					
No	6822	3110 (45.6)	967 (14.2)	2079 (30.5)	666 (9.8)
Yes	307	75 (24.4)	30 (9.8)	111 (36.2)	91 (29.6)

Continued

TABLE 2 Continued	All cases	Not clustered	Cases in cluster n		
			2	3–20	>20
Alcohol use					
No	6460	2911 (45.1)	908 (14.1)	1979 (30.6)	662 (10.2)
Yes	328	106 (32.3)	42 (12.8)	124 (37.8)	56 (17.1)
Prison					
No	6938	3140 (45.3)	984 (14.2)	2129 (30.7)	685 (9.9)
Yes	225	53 (23.6)	22 (9.8)	78 (34.7)	72 (32.0)
Risk factor count[¶]					
0	6688	3081 (46.1)	950 (14.2)	2020 (30.2)	637 (9.5)
1	508	187 (36.8)	64 (12.6)	180 (35.4)	77 (15.2)
2	159	41 (25.8)	19 (12.0)	58 (36.5)	41 (25.8)
3	84	19 (22.6)	8 (9.5)	26 (31.0)	31 (36.9)
4	19	1 (5.3)	3 (15.8)	6 (31.6)	9 (47.4)
Mean IMD quintile[+]		2.45	2.40	2.41	2.15

Data are presented as n or n (% row). IMD: index of multiple deprivation. [#]: resistance to any first-line antibiotic; [¶]: cumulative number of social risk factors (history of homelessness, illicit drug use, alcohol misuse, imprisonment) reported by each case; [+]: IMD quintile of Lower Layer Super Output Area within London (lowest is most deprived).

TABLE 3 Single-variable multinomial logistic regression analysis for risk factors associated with tuberculosis cases in molecular clusters of different sizes in London (2010–2014)	Cases in cluster n		
	2	3–20	>20
Sex			
Female	1	1	1
Male	0.94 (0.81–1.08)	1.07 (0.95–1.19)	1.16 (0.99–1.37)[#]
Age group years			
0–14	1.79 (1.08–2.97)[#]	2.01 (1.35–2.99)[#]	2.92 (1.82–4.68)[#]
15–44	1	1	1
45–64	1.14 (0.95–1.36)[#]	1.07 (0.93–1.23)	1.00 (0.81–1.22)[#]
≥65	0.74 (0.57–0.96)[#]	0.69 (0.57–0.83)[#]	0.47 (0.33–0.65)[#]
Ethnic group			
White	1	1	1
Black-Caribbean	2.01 (1.25–3.22)[#]	2.56 (1.77–3.72)[#]	2.92 (1.89–4.50)[#]
Black-African	1.03 (0.79–1.33)	1.12 (0.92–1.37)	1.04 (0.81–1.34)[#]
Black-Other	1.39 (0.72–2.69)	1.47 (0.88–2.46)[#]	1.50 (0.79–2.83)[#]
Indian	0.70 (0.54–0.89)[#]	0.61 (0.50–0.73)[#]	0.31 (0.24–0.40)[#]
Pakistani	0.79 (0.58–1.08)[#]	0.64 (0.50–0.82)[#]	0.65 (0.47–0.90)[#]
Bangladeshi	0.40 (0.27–0.60)[#]	0.31 (0.23–0.43)[#]	0.12 (0.06–0.22)[#]
Chinese	0.58 (0.29–1.16)[#]	0.48 (0.28–0.81)[#]	0.25 (0.10–0.65)[#]
Mixed/other	0.81 (0.62–1.06)[#]	0.69 (0.56–0.85)[#]	0.41 (0.30–0.55)[#]
Place of birth			
Non-UK	1	1	1
UK	1.77 (1.44–2.18)[#]	2.55 (2.18–2.99)[#]	4.11 (3.38–4.99)[#]
Time since entry to UK years			
0–1	1	1	1
2–4	0.93 (0.73–1.17)	0.85 (0.70–1.03)[#]	0.76 (0.56–1.02)[#]
5–9	1.02 (0.80–1.31)	1.14 (0.94–1.39)[#]	1.20 (0.89–1.61)
≥10	1.13 (0.90–1.41)	1.35 (1.13–1.61)[#]	1.22 (0.93–1.60)[#]
Occupation			
Other	1	1	1
None	1.09 (0.92–1.29)	1.05 (0.93–1.20)	1.54 (1.27–1.86)[#]
Education	1.10 (0.88–1.37)	1.14 (0.97–1.35)[#]	1.54 (1.21–1.96)[#]
Healthcare	1.08 (0.75–1.54)	0.89 (0.67–1.19)	0.50 (0.28–0.90)[#]

Continued

TABLE 3 Continued

	Cases in cluster n		
	2	**3–20**	**>20**
Pulmonary disease			
No	1	1	1
Yes	1.47 (1.28–1.70)[#]	1.58 (1.42–1.77)[#]	1.92 (1.63–2.26)[#]
Sputum smear			
Negative	1	1	1
Positive	1.33 (1.11–1.59)[#]	1.40 (1.22–1.61)[#]	1.75 (1.44–2.13)[#]
Previous diagnosis			
No	1	1	1
Yes	1.51 (1.09–2.08)[#]	1.50 (1.16–1.93)[#]	1.67 (1.19–2.35)[#]
Previous treatment			
No	1	1	1
Yes	3.03 (0.35–26.12)	1.20 (0.37–3.86)	
Drug resistance[¶]			
No	1	1	1
Yes	1.32 (1.05–1.67)[#]	0.84 (0.68–1.02)[#]	1.95 (1.55–2.46)[#]
Inpatient			
No	1	1	1
Yes	1.10 (0.95–1.27)[#]	1.02 (0.91–1.14)	1.17 (1.00–1.37)[#]
Homeless			
No	1	1	1
Yes	1.47 (1.02–2.12)[#]	1.29 (0.96–1.74)[#]	3.21 (2.33–4.43)[#]
Drug use			
No	1	1	1
Yes	1.29 (0.84–1.98)	2.21 (1.64–2.98)[#]	5.67 (4.13–7.78)[#]
Alcohol use			
No	1	1	1
Yes	1.27 (0.88–1.83)[#]	1.72 (1.32–2.24)[#]	2.32 (1.66–3.25)[#]
Prison			
No	1	1	1
Yes	1.32 (0.80–2.19)	2.17 (1.52–3.09)[#]	6.22 (4.32–8.95)[#]
Risk factor count[+]			
0	1	1	1
1	1.11 (0.83–1.49)	1.47 (1.19–1.82)[#]	1.99 (1.51–2.63)[#]
2	1.50 (0.87–2.60)[#]	2.16 (1.44–3.23)[#]	4.84 (3.11–7.52)[#]
3	1.37 (0.60–3.13)	2.09 (1.15–3.78)[#]	7.89 (4.43–14.06)[#]
4	9.73 (1.01–93.64)[#]	9.15 (1.10–76.07)[#]	43.53 (5.51–344.20)[#]
IMD quintile[§]	0.59 (0.54–0.63)[#]	0.86 (0.81–0.91)[#]	0.62 (0.57–0.68)[#]

Data are presented as adjusted odds ratio (95% CI). IMD: index of multiple deprivation. [#]: p<0.2, included in initial multivariable model. [¶]: resistance to any first-line antibiotic; [+]: cumulative number of social risk factors (history of homelessness, illicit drug use, alcohol misuse, imprisonment) reported by each case; [§]: IMD quintile of Lower Layer Super Output Area (LSOA) within London (lowest is most deprived), included as a continuous variable in multilevel model accounting for random effects of LSOA.

the UK. Similarly, the odds of cases being in a cluster of n=2–20 cases compared with not being in a cluster for those born in the UK were 2.55 times that of those not born in the UK; and the odds of cases being in a cluster of n=2 cases compared with not being in a cluster for those born in the UK were 1.77 times those not born in the UK.

Factors included in the final multivariable model were sex, age, ethnicity, place of birth, occupation, site of disease, drug resistance, number of social risk factors and IMD (table 4 and figure 3). Cases in the oldest age group (⩾65 years) had an adjusted OR (aOR) of 0.52 (95% CI 0.35–0.78); aOR for being born in the UK was 2.93 (95% CI 2.28–3.77). The association between black ethnic groups and larger cluster size was maintained (aOR black-Caribbean ethnicity 3.64, 95% CI 2.23–5.94), whilst the only ethnic group with significantly lower risk than the white population was Bangladeshi (aOR 0.26, 95% CI 0.13–0.50). Students and those working in education had an increased adjusted odds of being in large clusters (aOR 1.31, 95% CI 1.01–1.70), and those working in healthcare had a decreased adjusted odds (aOR 0.47, 95% CI 0.25–0.87).

TABLE 4 Multivariable multinomial logistic regression analysis for risk factors associated with tuberculosis cases in molecular clusters of different sizes in London (2010–2014), adjusted for random effects of Lower Layer Super Output Area (LSOA)

| | Cases in cluster n | | |
	2	3–20	>20
Sex			
Female	1	1	1
Male	0.98 (0.83–1.16)	1.08 (0.95–1.23)	1.14 (0.94–1.38)
Age group years			
0–14	1.15 (0.65–2.02)	1.18 (0.76–1.83)	1.29 (0.75–2.22)
15–44	1	1	1
45–64	1.13 (0.92–1.39)	0.97 (0.82–1.15)	0.82 (0.64–1.04)
⩾65	0.68 (0.50–0.93)	0.71 (0.56–0.90)	0.52 (0.35–0.78)
Ethnic group			
White	1	1	1
Black-Caribbean	2.1 (1.25–3.55)	3.13 (2.08–4.71)	3.64 (2.23–5.94)
Black-African	1.35 (0.99–1.86)	1.86 (1.46–2.38)	2.09 (1.49–2.91)
Black-Other	1.88 (0.92–3.84)	1.92 (1.07–3.45)	2.29 (1.13–4.66)
Indian	0.95 (0.70–1.30)	1.02 (0.80–1.30)	0.78 (0.55–1.11)
Pakistani	1.08 (0.75–1.55)	1.02 (0.76–1.36)	1.51 (1.02–2.24)
Bangladeshi	0.60 (0.38–0.94)	0.53 (0.37–0.76)	0.26 (0.13–0.50)
Chinese	0.78 (0.37–1.64)	0.78 (0.44–1.40)	0.63 (0.24–1.68)
Mixed/other	1.12 (0.81–1.54)	1.11 (0.86–1.43)	0.89 (0.61–1.29)
Place of birth			
Non-UK	1	1	1
UK	1.45 (1.12–1.87)	2.13 (1.75–2.58)	2.93 (2.28–3.77)
Occupation			
Other	1	1	1
None	1.04 (0.85–1.27)	0.96 (0.82–1.12)	1.18 (0.94–1.49)
Education	1.04 (0.82–1.31)	1.04 (0.87–1.24)	1.31 (1.01–1.70)
Healthcare	1.00 (0.69–1.45)	0.82 (0.60–1.11)	0.47 (0.25–0.87)
Pulmonary disease			
No	1	1	1
Yes	1.46 (1.24–1.72)	1.48 (1.30–1.68)	1.47 (1.21–1.79)
Drug resistance[#]			
No	1	1	1
Yes	1.25 (0.96–1.61)	0.82 (0.65–1.03)	1.75 (1.34–2.28)
Risk factor count[¶]			
0	1	1	1
1	0.83 (0.59–1.15)	1.12 (0.88–1.42)	1.36 (0.99–1.87)
2	1.00 (0.53–1.89)	1.59 (1.00–2.52)	2.46 (1.45–4.18)
3	0.86 (0.35–2.12)	1.28 (0.67–2.45)	3.75 (1.96–7.16)
4	4.35 (0.39–48.50)	4.13 (0.45–37.58)	16.64 (1.98–139.88)
IMD quintile[+]	0.98 (0.91–1.04)	1.00 (0.95–1.06)	0.90 (0.83–0.97)

Data are presented as adjusted odds ratio (95% CI). IMD: index of multiple deprivation. [#]: resistance to any first-line antibiotic; [¶]: cumulative number of social risk factors (history of homelessness, illicit drug use, alcohol misuse, imprisonment) reported by each case; [+]: IMD quintile of LSOA within London (lowest is most deprived), included as a continuous variable in multilevel model accounting for random effects of LSOA.

Social risk factors were included in the final model as a count and there was a trend of increased odds with increased number of risk factors, although confidence intervals overlapped (aOR three risk factors 3.75, 95% CI 1.96–7.16; aOR four risk factors 16.64, 95% CI 1.98–139.88). Deprivation was also independently associated with being in a large cluster; the aOR was 0.90 (95% CI 0.83–0.97) for increased IMD quintile and therefore decreased deprivation level.

Black-Caribbean ethnicity, being born in the UK and pulmonary disease were the only factors that also had significantly elevated odds for clusters of n=2 or n=3–20 cases.

Spatial clusters of cases in large molecular clusters
We used SaTScan to test for spatial clustering in the 20 molecular clusters that had n>20 cases. A total of 25 significant spatial clusters (p<0.05) were identified, with at least one significant spatial cluster in 18
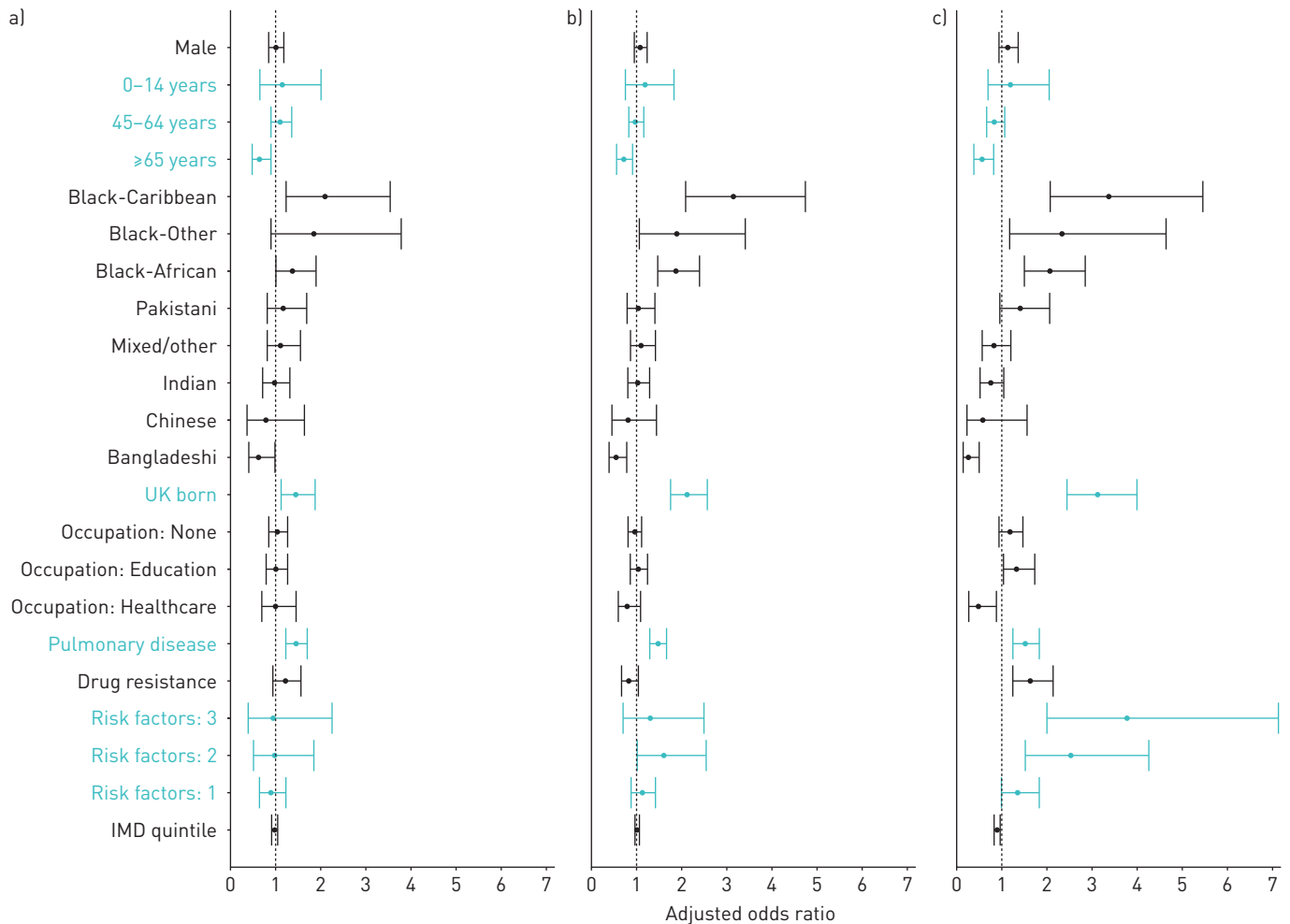
FIGURE 3 Forest plot of adjusted odds ratios (with 95% confidence intervals) from multivariable multinomial logistic regression analysis (table 4), by number of cases in molecular cluster: a) n=2, b) n=3–20 and c) n>20 cases.

(90%) of the molecular clusters, and eight of the spatial clusters included more than 10 cases. These clusters tended to be located in more deprived areas; the median IMD rank of the 4970 LSOAs within London for areas within the clusters was 1110 compared with 2538.5 for areas not in clusters.

The locations of the eight spatial clusters, overlaid on smoothed incidence maps, are shown in figure 4.

Discussion

In this study, we present results from the first 5 years of routine molecular strain typing of TB by MIRU-VNTR in London. There were 20 molecular clusters that had n>20 cases of TB notified between 2010 and 2014. These clusters accounted for 795 (11%) of all typed cases notified during this period, and cases in large clusters also tended to occur closer together in space and time. One of the molecular clusters described in this study is part of a known outbreak of isoniazid-resistant disease that was first identified in 1999 [5], but this is the first analysis to suggest that multiple similar outbreaks may be ongoing.

Cases in large molecular clusters were more likely to have multiple social risk factors, be of black ethnicities, born in the UK, have pulmonary and drug-resistant disease, and live in more deprived areas of London. Small clusters (pairs of cases) and those of intermediate size (n=3–20 cases) were associated with black-Caribbean ethnicity, being born in the UK and pulmonary disease. There was also some association between large clusters and occupation. Large clusters were more likely to include students and those involved in education, which may suggest that outbreaks in schools and universities can spread widely in these settings or through extensive social networks involving students. However, they were less likely to involve healthcare workers. This indicates that there was limited nosocomial transmission and that when transmission involving a healthcare worker did occur it was usually an isolated incident rather than part of a large outbreak.
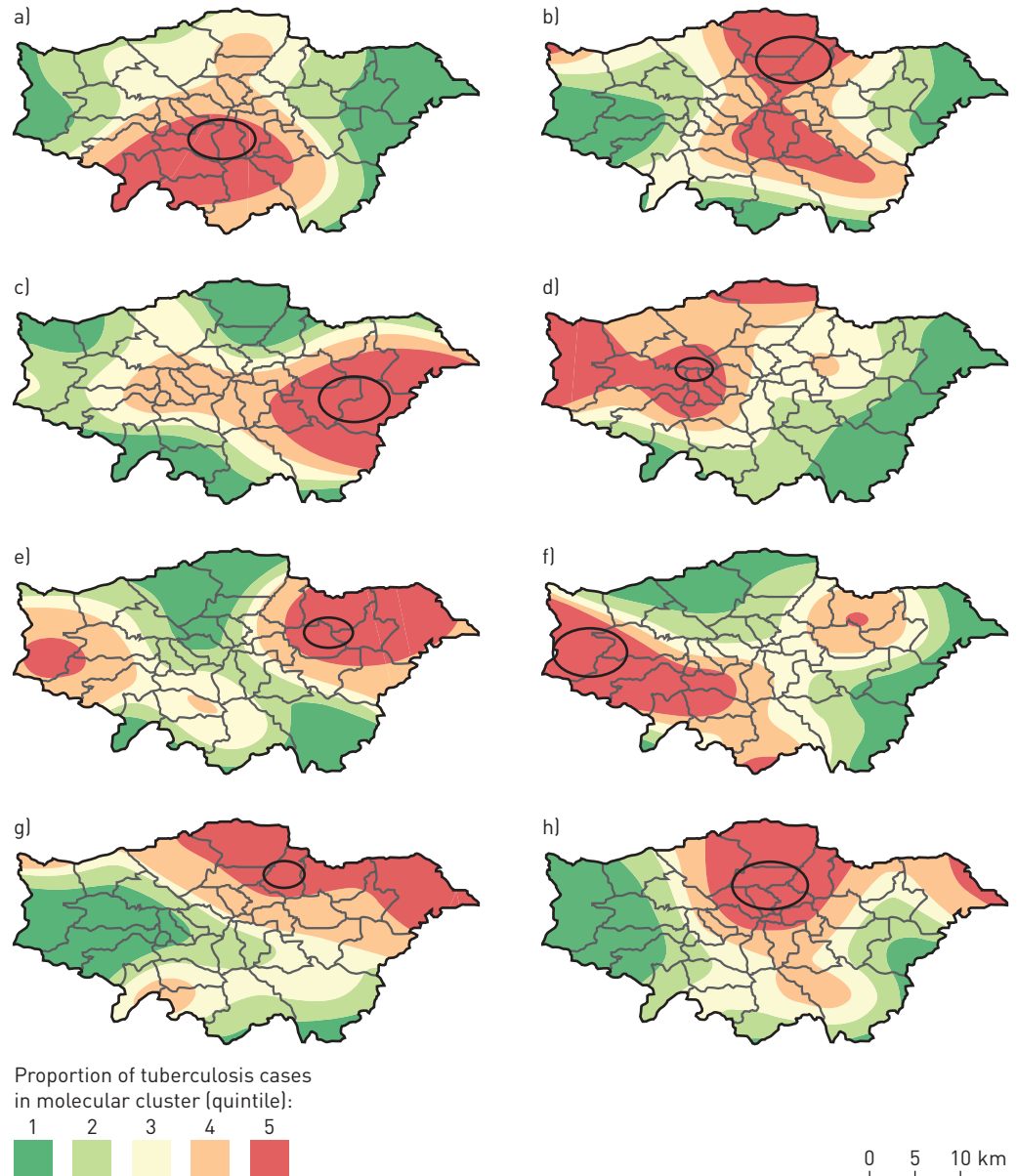
FIGURE 4 a–h) Locations of significant spatial clusters of cases within eight molecular clusters of tuberculosis (TB) in London (2010–2014), overlaid on smoothed incidence maps. Ovals represent areas of significant spatial clustering (p<0.05) with more than 10 cases of the given molecular cluster compared with the general distribution of TB cases. The proportions of cases in molecular clusters compared with all other TB cases are represented through kernel density estimation (bandwidth 5 km).

The majority of large molecular clusters exhibited significant spatial clustering, indicating likely transmission within London. Spatial clusters tended to be in more deprived areas and the IMD was independently associated with being in a large molecular cluster, after accounting for individual risk factors. Studies in other settings, including Lima, Peru [14], northern England [15], Tokyo, Japan [16] and the USA [17], have also investigated TB clustering using molecular and spatial data. Various methods have been used to assess spatial clustering, but all have also identified areas of likely localised TB transmission or "hotspots".

Previous analyses of MIRU-VNTR strain typing and surveillance data in various settings have sought to determine the proportion of cases that were part of a molecular cluster or to identify risk factors for clustering [11, 18–25]. A systematic review of 27 articles found that clustering estimates ranged from 0% to 63% [26], whilst the clustering proportion in the first 3 years of routine molecular strain typing in London was 46% [11]. This indicates that the rate of molecular clustering identified in our analysis (56%) was relatively high and that it has risen with inclusion of more years of data.

A strength of this study was that it was based on routine surveillance data and therefore included all cases of TB in London that were successfully typed by MIRU-VNTR to at least 23 loci over a 5-year period. There was a low level of missing information in the variables used in the risk factor analysis (table 2). As a result it provides a good representation of the population of TB cases in the city. The study adhered to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for reporting of cross-sectional studies [27].

Another strength was the use of multinomial logistic regression to identify risk factors. The advantage of this method was that it allowed associations to be assessed for different sizes of molecular cluster. This is important because larger clusters have more implications for TB control. It therefore extended the previous analysis of risk factors for clustering, which used a binary outcome that was not stratified by cluster size. The combination of molecular with spatial clustering analyses was a further advantage of this study, as it provided further evidence for transmission in some large molecular clusters. It also has practical application, as it could be used to prioritise clusters for further investigation.

A limitation of this study is that we had to restrict our analysis to cases of TB which had been typed by MIRU-VNTR for at least 23 loci. We therefore excluded 7522 cases that were not culture confirmed or typed. This will have resulted in misclassification of some cases as unclustered which did not have a unique strain and therefore underestimated the number of cases in some clusters. A second limitation was that we considered the temporal distribution of cases in molecular clusters by examining the median interval between case notification dates, which are a proxy for dates of onset. This provided some evidence that cases in larger molecular clusters occurred closer together in time than those in smaller clusters. However, a true estimate of serial intervals to assess this robustly would require ascertainment of epidemiological links between cases to establish chains of transmission. Our analysis was also limited because we were unable to assess the importance of other potential factors affecting TB transmission which are not currently collected in surveillance data. HIV status, for example, is not routinely collected, although all TB patients are offered HIV testing and it was taken up by 98% patients in London in 2014 [28]. Modelling estimates show that HIV co-infection in TB patients in England is relatively low, at just 3.4% in 2014 (personal communication, PHE National Infection Service).

The results of this study have implications for the control of TB in London and other high-incidence cities. Targeting interventions to deprived areas should be a priority for reducing transmission, whilst efforts should also be made to raise awareness of the disease amongst at-risk groups, such as those of black ethnicities born in the UK. An example of such an intervention is the "Find and Treat" mobile radiography unit which actively screens for cases in vulnerable populations in London and provides support to help patients complete treatment [29]. Continued support for this service is therefore a key component of TB control in London. Our results also imply that detailed investigations of molecular clusters could be beneficial in preventing large chains of transmission through interventions such as contact tracing and screening. We recommend incorporating routine spatial clustering analysis to assist with prioritising clusters for further investigation, as use of simple thresholds has previously been ineffective in making these decisions [30, 31].

Future work arising from this study could aim to identify which of the components of the IMD may be contributing to TB transmission. This would be useful to inform environmental and housing interventions, such as improving ventilation and reducing overcrowding. More work is also required to determine if the associations observed in clusters of different sizes could be used to predict whether cases in small clusters are likely to form larger clusters. Finally, results from whole genome sequencing of TB isolates when routinely available should add further resolution to networks suggested by molecular clusters. This could provide evidence to support or refute transmission in some instances and direct the focus of intensive investigations [32].

In conclusion, this study shows that large molecular clusters contribute substantially to the burden of TB in London. The results highlight the continued importance of preventing long chains of transmission in order to eliminate TB as a public health problem in large cities.

References
1   de Vries G, Aldridge RW, Cayla JA, *et al.* Epidemiology of tuberculosis in big cities of the European Union and European Economic Area countries. *Euro Surveill* 2014; 19: pii:20726.
2   Public Health England. Tuberculosis in England: 2015 Report Version 1.1. London, Public Health England, 2015.
3   London Assembly Health Committee. Tackling TB in London. London, London Assembly Health Committee, 2015.

4   Maguire H, Brailsford S, Carless J, *et al.* Large outbreak of isoniazid-monoresistant tuberculosis in London, 1995 to 2006: case-control study and recommendations. *Euro Surveill* 2011; 16: pii:19830.

5   Ruddy M, Davies A, Yates M, *et al.* Outbreak of isoniazid resistant tuberculosis in north London. *Thorax* 2004; 59: 279–285.

6   Ghebremichael S, Petersson R, Koivula T, *et al.* Molecular epidemiology of drug-resistant tuberculosis in Sweden. *Microbes Infect* 2008; 10: 699–705.

7   Edlin BR, Tokars JI, Grieco MH, *et al.* An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. *N Engl J Med* 1992; 326: 1514–1521.

8   Frieden TR, Sherman LF, Maw KL, *et al.* A multi-institutional outbreak of highly drug-resistant tuberculosis: epidemiology and clinical outcomes. *JAMA* 1996; 276: 1229–1235.

9   Miller AC, Butler WR, McInnis B, *et al.* Clonal relationships in a shelter-associated outbreak of drug-resistant tuberculosis: 1983–1997. *Int J Tuberc Lung Dis* 2002; 6: 872–878.

10   Smith CM, Le Comber SC, Fry H, *et al.* Spatial methods for infectious disease outbreak investigations: systematic literature review. *Euro Surveill* 2015; 20: pii:30026.

11   Hamblion EL, Le Menach A, Anderson LF, *et al.* Recent TB transmission, clustering and predictors of large clusters in London, 2010–2012: results from first 3 years of universal MIRU-VNTR strain typing. *Thorax* 2016; 71: 749–756.

12   Department of Communities and Local Government. The English Indices of Deprivation 2015. London, Department of Communities and Local Government, 2015.

13   Kwak C, Clayton-Matthews A. Multinomial logistic regression. *Nurs Res* 2002; 51: 404–410.

14   Zelner JL, Murray MB, Becerra MC, *et al.* Identifying hotspots of multidrug-resistant tuberculosis transmission using spatial and molecular genetic data. *J Infect Dis* 2016; 213: 287–294.

15   Saavedra-Campos M, Welfare W, Cleary P, *et al.* Identifying areas and risk groups with localised *Mycobacterium tuberculosis* transmission in northern England from 2010 to 2012: spatiotemporal analysis incorporating highly discriminatory genotyping data. *Thorax* 2016; 71: 742–748.

16   Izumi K, Ohkado A, Uchimura K, *et al.* Detection of tuberculosis infection hotspots using activity spaces based spatial approach in an urban Tokyo, from 2003 to 2011. *PLoS One* 2015; 10: e0138831.

17   Althomsons SP, Kammerer JS, Shang N, *et al.* Using routinely reported tuberculosis genotyping and surveillance data to predict tuberculosis outbreaks. *PLoS One* 2012; 7: e48754.

18   Biadglegne F, Merker M, Sack U, *et al.* Tuberculous lymphadenitis in Ethiopia predominantly caused by strains belonging to the Delhi/CAS lineage and newly identified Ethiopian clades of the *Mycobacterium tuberculosis* complex. *PLoS One* 2015; 10: e0137865.

19   Ojo OO, Sheehan S, Corcoran DG, *et al.* Molecular epidemiology of *Mycobacterium tuberculosis* clinical isolates in Southwest Ireland. *Infect Genet Evol* 2010; 10: 1110–1116.

20   Lim LK-Y, Sng LH, Win W, *et al.* Molecular epidemiology of *Mycobacterium tuberculosis* complex in Singapore, 2006–2012. *PLoS One* 2013; 8: e84487.

21   Tuite AR, Guthrie JL, Alexander DC, *et al.* Epidemiological evaluation of spatiotemporal and genotypic clustering of *Mycobacterium tuberculosis* in Ontario, Canada. *Int J Tuberc Lung Dis* 2013; 17: 1322–1327.

22   Chen KS, Liu T, Lin RR, *et al.* Tuberculosis transmission and risk factors in a Chinese antimony mining community. *Int J Tuberc Lung Dis* 2016; 20: 57–62.

23   Goldblatt D, Rorman E, Chemtob D, *et al.* Molecular epidemiology and mapping of tuberculosis in Israel: do migrants transmit the disease to locals? *Int J Tuberc Lung Dis* 2014; 18: 1085–1091.

24   Toit K, Altraja A, Acosta CD, *et al.* A four-year nationwide molecular epidemiological study in Estonia: risk factors for tuberculosis transmission. *Public Health Action* 2014; 4: Suppl. 2, S34–S40.

25   Yang C, Shen X, Peng Y, *et al.* Transmission of *Mycobacterium tuberculosis* in China: a population-based molecular epidemiologic study. *Clin Infect Dis* 2015; 61: 219–227.

26   Mears J, Abubakar I, Cohen T, *et al.* Effect of study design and setting on tuberculosis clustering estimates using Mycobacterial Interspersed Repetitive Units-Variable Number Tandem Repeats (MIRU-VNTR): a systematic review. *BMJ Open* 2015; 5: e005636.

27   von Elm E, Altman DG, Egger M, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; 370: 1453–1457.

28   Public Health England. Tuberculosis in London: Annual Review (2014 Data). London, Public Health England, 2015.

29   Jit M, Stagg HR, Aldridge RW, *et al.* Dedicated outreach service for hard to reach patients with tuberculosis in London: observational study and economic evaluation. *BMJ* 2011; 343: d5376.

30   Public Health England. TB Strain Typing and Cluster Investigation Handbook. London, Public Health England, 2014.

31   Mears J, Vynnycky E, Lord J, *et al.* The prospective evaluation of the TB strain typing service in England: a mixed methods study. *Thorax* 2016; 71: 734–741.

32   Walker TM, Lalor MK, Broda A, *et al.* Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome sequences: an observational study. *Lancet Respir Med* 2014; 2: 285–292.