

Identifying areas and risk groups with localised *Mycobacterium tuberculosis* transmission in northern England from 2010 to 2012: spatiotemporal analysis incorporating highly discriminatory genotyping data

María Saavedra-Campos,^{1,2,3} William Welfare,^{4,5} Paul Cleary,² Andrew Sails,⁶ Andy Burkitt,⁷ Daniel Hungerford,² Ebere Okereke,⁸ Peter Acheson,⁹ Marko Petrovic^{4,5}

For numbered affiliations see end of article.

Correspondence to

María Saavedra-Campos, Field Epidemiology Services, Public Health England, 5th Floor, Rail House, Lord Nelson Street, Liverpool, L1 1JF, USA; maria.saavedra-campos@phe.gov.uk

Received 9 October 2014

Revised 1 March 2015

Accepted 30 March 2015

Published Online First

28 April 2015

ABSTRACT

Background Information on geographical variation in localised transmission of TB can inform targeting of disease control activities. The aim of this study was to estimate the proportion of TB attributable to localised transmission for the period 2010–2012 in northern England and to identify case characteristics associated with spatiotemporal-genotypical clusters.

Methods We combined genotyping data with spatiotemporal scan statistics to define an indicator of localised TB transmission and identified factors associated with localised TB transmission thus defined in a multivariable logistis regression model.

Results The estimated proportion of TB cases in northern England attributable to localised transmission was 10% (95% CI 9% to 12%). Clustered cases (cases which were spatiotemporally clustered with others of identical genotype) were on average younger than non-clustered cases (mean age 34 years vs 43 years; *p* value <0.05). Being UK born (adjusted OR (aOR) 3.6, 95% CI 2.9 to 6.0), presenting with pulmonary disease (aOR 2.2, 95% CI 1.3 to 3.6) and history of homelessness (aOR 2.8, 95% CI 1.2 to 6.8) or incarceration (aOR 2.6, 95% CI 1.2 to 5.9) were independently associated with being part of a spatiotemporal-genotypical cluster in a multivariable model. Belonging to an ethnic group other than white or mixed/other was also significantly associated with localised transmission. We identified localised transmission in 103/1958 middle super output areas mostly in urban areas.

Conclusions Incorporating highly discriminatory genotyping data into spatiotemporal analysis of TB incidence is feasible as part of routine surveillance and can provide valuable information on groups at greater risk and areas with localised transmission of TB, which could be used to inform control measures, such as intensified contact tracing.

INTRODUCTION

The prevention of transmission of TB is the key to reducing incidence. Most TB control programmes rely on contact tracing for the identification of clusters of TB transmission. This method may miss limited or casual contact and thus may underestimate recent transmission.^{1–3} Several countries have used population based genotyping to assess recent transmission, as represented by the number of

Key messages

What is the key question?

- Can we identify risk groups and geographical areas where new TB cases are likely to represent person-to-person transmission, in order to inform targeted control measures?

What is the bottom line?

- Using a novel method applying spatiotemporal statistical methods to a large UK TB data set with highly discriminatory 24-locus mycobacterial interspersed repetitive unit variable number tandem repeat (MIRU-VNTR) profile for culture-positive cases, we estimated that one in every 10 cases of TB in the North of England was highly likely to be the result of localised transmission and we identified the following case characteristics: birth in the UK, pulmonary disease, a history of homelessness or incarceration and belonging to an ethnic group other than white or mixed/other as being significantly associated with localised TB transmission.

Why read on?

- We describe a method using freely available statistical software to add a genetic dimension to spatiotemporal analyses, the application of which has provided valuable information on transmission of TB in the North of England, and which has potential applicability to the study of the distribution of other diseases.

genotypically clustered cases.⁴ Molecular clustering may approximate recent transmission in low incidence populations, but it is not always associated with recent transmission.⁵

In England, universal typing of TB isolates was introduced in 2010. The national strain typing service uses the 24-locus mycobacterial interspersed repetitive unit variable number tandem repeat (MIRU-VNTR) typing method.⁶ One of its objectives is to identify clustered cases that are potentially part of a recent chain of transmission as well as the characteristics associated with recent



► <http://dx.doi.org/10.1136/thoraxjnl-2014-206608>

► <http://dx.doi.org/10.1136/thoraxjnl-2014-206480>



CrossMark

To cite: Saavedra-Campos M, Welfare W, Cleary P, et al. *Thorax* 2016;**71**:742–748.



transmission. The programme defines clustered cases as two or more cases with indistinguishable MIRU-VNTR profiles, with at least one case with a complete 24-locus MIRU-VNTR profile. Additional cases in the cluster may each have one missing locus.⁷ The identification of clustered cases was initially followed by an epidemiological investigation if there were five cases or more resident in the same geographical area under the jurisdiction of a Health Protection Unit, or ten cases within a regional or national boundary in the last 2 years, with two in the previous 6 months.⁷ However a review of the UK strain typing service in March 2013 found no evidence to suggest that cluster investigations undertaken on the basis of specific thresholds were effective or cost-effective.⁸ As a result, cluster investigation undertaken this way was abandoned in favour of a more locally targeted approach.

The Centers for Disease Control and Prevention in the USA recently described a method that combines genotyping data with spatial scan statistics to estimate recent TB transmission.⁹ Spatial scan statistics have been widely used for the identification of clusters of infectious disease.^{10–13} The availability of genotyping data and data on individual cases from the national Enhanced TB Surveillance (ETS) system has provided an opportunity to combine these data for the North of England. The objectives of this study were to employ this method to (1) estimate the proportion of TB in the North of England attributable to localised transmission for the period 2010–2012, (2) identify the areas where localised transmission is occurring and (3) identify the demographic, geographical, clinical and social risk factors associated with localised TB transmission.

METHODS

Study population

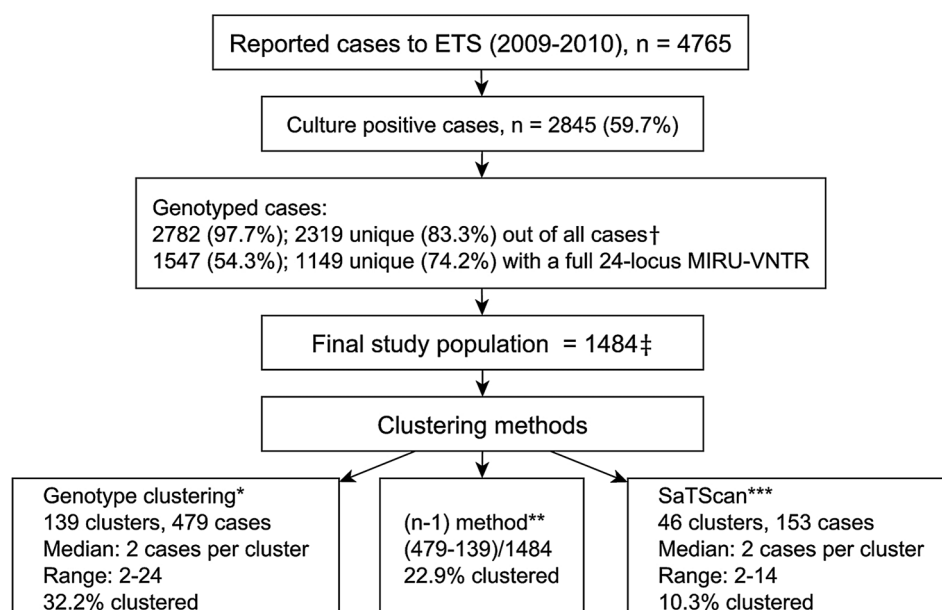
We conducted a cross-sectional study in which cases were defined as patients resident in the North of England (defined as Public Health England region, which includes North-East of England, Cumbria and Lancashire, Yorkshire and the Humber, Greater Manchester and Cheshire and Merseyside) at the time of diagnosis with a positive culture of *Mycobacterium tuberculosis* from a clinical specimen and with complete 24-locus

MIRU-VNTR profile between 2010 and 2012. We obtained the cases from the ETS system. TB nurses and physicians routinely submit to the ETS system information on case demographics, clinical characteristics and risk factors for all TB cases notified in England, which is then linked to microbiological and molecular data. For cases with more than one isolate and discordant genotyping we used the first isolate for the analysis. For cases that did not have a recorded date of onset of symptoms, this was approximated using other dates in the data set and for cases with no recorded residence, the postcode of the hospital reporting the case was used as a proxy measure of residence in order to reduce potential selection bias. Case locations were geocoded with ArcGIS V10.0,¹⁴ using the postcode of residence to assign cases to middle super output areas (MSOA; MSOAs are well-defined census neighbourhoods with a population between 5000 and 15 000 people).¹⁵ The final study population included all cases with confirmed culture-positive TB with a complete 24-locus MIRU-VNTR profile, a MSA within the North of England and onset of symptoms between January 2010 and December 2012 inclusive.

Genotype and geospatial clustering

As an indicator of localised TB transmission, we defined clustered cases as cases with an exact match on all 24 MIRU-VNTR loci within geospatial zones with statistically significant spatio-temporal scan statistics obtained from SaTScan V9.2.¹⁶ SaTScan detects clusters in space and time using a moving cylindrical window method, where the base represents space and the height represents time.¹⁷ For each window, a log-likelihood ratio is generated, comparing observed with expected cases inside and outside the moving window. A p value representing how likely the observed data for a given area and period would be if cases were occurring randomly in accordance with a Poisson distribution and independently of each other is assigned to each window using Monte Carlo simulation.¹⁷ Clusters with a p value of <0.05 were regarded as statistically significant. In our study the temporal window or longest duration of a cluster was set to 3 years, the full study period and the spatial window was

Figure 1 Number of cases of TB reported to Enhanced TB Surveillance (ETS), including culture-positive cases, genotyped cases and clustered cases by the different methods, northern England, 2010 to 2012. †This number includes incomplete 24-locus mycobacterial interspersed repetitive unit variable number tandem repeat (MIRU-VNTR) profiles too. ‡Cases with an onset date that was not between January 2010 and December 2012 were removed. *Genotype clustering defined as two or more cases with identical 24-locus MIRU-VNTR profile identified in northern England. **Calculated using the number of cases that were part of a cluster as per genotype clustering definition minus the total number of clusters identified and divided by the total number of cases with a full 24-locus MIRU-VNTR. ***Two or more cases with the same 24-locus MIRU-VNTR profile found to be spatiotemporally clustered by SaTScan.



set to 50 km. We selected the option for no geographical overlap for reporting secondary clusters.

As the number of cases by genotype was small, we ran SaTScan with all the cases together to first define the spatio-temporal clusters. We then identified cases that had a genotype that appeared more than once within each significant spatio-temporal cluster. For the purpose of this study these cases were considered to be those more likely to represent localised TB transmission. The cases that did not fulfil these criteria were considered more likely to be cases of remotely acquired TB or reactivation of old TB.

We used MSOA population estimates for 2010 from the Office for National Statistics,¹⁸ for all 3 years (the only estimates available at the time of the analysis).

We used the 'n-1' method¹⁹ and genotype clustering (two or more cases with complete identical genotype) as comparator methods for estimating recent TB transmission.

Statistical analysis

We used logistic regression to examine demographic (ie, age, sex, ethnicity, country of birth and year of entry in the UK) and clinical (ie, clinical presentation, sputum smear status, previous diagnosis and patient receiving directly observed therapy (DOT)) characteristics and risk factors (ie, having a social risk factor, homelessness, imprisonment and drug or alcohol abuse)

potentially associated with recent TB transmission. We performed single variable analysis using Pearson's χ^2 test and Fisher's exact test. All variables with a significance value of <0.2 in the single variable analysis were included in the regression model. We used a backwards stepwise approach to identify a final model, eliminating variables with the highest χ^2 values first and examining at each step for possible confounders. The analysis was performed using R V.3.0.3 (The R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

The overall annual incidence of TB for the North of England was 10.7 cases per 100 000 population in 2010, 10.8 in 2011 and 10.7 in 2012. Between January 2010 and December 2012, 4765 cases of TB were reported to ETS. Of these, 2845 (60%) were culture-positive. The proportion of culture-positive cases with a complete genotype result has increased over the 3 years; from 41% (387/942) in 2010; to 57% (562/982) in 2011 and to 58% (535/921) in 2012. A total of 2782 (98%) isolates were characterised using MIRU-VNTR typing. A total of 2090 (75%) had a result with at least 23 loci and 1547 (54%) had a full 24-locus MIRU-VNTR profile. Six cases had multiple isolates and discordant genotypes. We included 1484 TB cases in the analysis, representing 52% of culture-positive cases. Of those 1146 (78%) had a unique 24-locus MIRU-VNTR. A total of 13/94 cases reporting

Proportion of TB clustered cases
by MSOA (per 1,000 population)

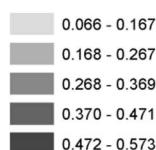


Figure 2 Proportion of TB clustered cases by middle super output areas (MSOAs) and local authority representing the areas where we identified localised TB transmission in northern England, 2010–2012. The map shows the local authorities in which localised transmission was identified and the proportion of clustered cases per 1000 by MSOA within each of the local authorities. The number of MSOAs within each local authority in which we identified localised TB transmission are: Blackburn with Darwen (7), Bolton (3), Bradford (20), Bury (2), Calderdale (3), Doncaster (1), Kirklees (13), Lancashire (7), Leeds (6), Manchester (14), Middlesbrough (3), North Yorkshire (1), Oldham (7), Rochdale (4), Salford (3), Sheffield (3) and Tameside (6).

history of being homeless did not have a postcode. For these cases, we used the postcode of the reporting hospital or laboratory. Only one of these cases was classified as being part of a cluster. Therefore, this procedure was unlikely to cause artificial clustering around reporting hospitals or laboratory.

Estimation of the proportion of TB in the North of England attributable to localised transmission

Of the 1484 cases included in the analysis, 153 (10%; 95% CI 9 to 12) were clustered in space and time and by genotype using the method described in this paper. Eighty-five per cent of clusters had four members or less and 58% of the clusters had only two members (median cluster size 2; range 2–14) (figure 1). The estimate of the proportion due to recent transmission varied depending on the method used, from 32% of all cases in our study population being attributed to recent TB transmission if only genotyping clustering was used to a more conservative estimate of 10% with the method described in this paper (figure 1). Using the 'n-1' method the proportion of clustered cases in our study population was 23%.

Areas where localised transmission was found to be occurring

Of all clustered cases, 3% (5/153) were residents in the North-East, 14% (21/153) in Cumbria and Lancashire, 37% (57/153) in Greater Manchester and 46% (70/153) in Yorkshire and Humberside. No clustering was identified in Cheshire and Merseyside. The proportion of all cases that were clustered (ie, consistent with localised transmission) varied by geographical area. In the North-East, 3% (5/164) of all cases were clustered

compared with 13% (21/183) in Cumbria and Lancashire, 13% (70/591) in Yorkshire and Humberside and 15% (57/429) in Greater Manchester. We identified clustered cases in 103 MSOAs out of a total of 1958 that constitute the North of England. Bradford, Manchester and Kirklees local authorities accounted for the majority of these MSOAs (46%; 47/103) (figure 2).

Identification of characteristics associated with localised TB transmission

Single variable analysis

Clustered cases were on average younger than non-clustered cases (mean 34 years old compared with 43 years; p value <0.05), and almost three times more likely to be UK born (OR 2.7, 95% CI 1.4 to 33.0) (table 1). Having entered the UK in the previous 2–4 years increased the risk of being part of a cluster by OR 2.2 (95% CI 1.0 to 4.7) compared with entering in the year previous to diagnosis (table 1).

Presenting with pulmonary disease was associated with being part of a cluster (OR 2.3, 95% CI 1.5 to 3.7) (table 2). Cases having a social risk factor were two times more likely to be part of a cluster (OR 2.1, 95% CI 1.2 to 3.4) (table 2). Having a history of homelessness (OR 3.0, 95% CI 1.3 to 6.5), being homeless at the time of diagnosis (OR 4.1, 95% CI 1.1 to 12.8) and having been to prison (OR 2.4, 95% CI 1.1 to 5.0) were all associated with being part of a cluster (table 2). Among the cluster cases we identified nine children. Seven of them were born in the UK. Of the other three only one entered the UK within 1 year, the rest had entered the UK more than 5 years

Table 1 Demographic characteristics of clustered cases or cases involved in recent TB transmission events; North of England, 2010–2012

Variable	No. (%) clustered	Total no. available	Crude OR (95% CI)	Adjusted OR (95% CI)†
Sex, n=1478				
Male	94 (61.8)	866	1.16 (0.81 to 1.67)	
Female	58 (38.2)	612	Referent	
Age group (years), n=1484				
0–4	2 (1.3)	15	1.31 (0.29 to 5.95)	
5–14	7 (4.6)	22	3.99 (1.57 to 10.13)	
15–24	29 (19.0)	235	1.20 (0.76 to 1.91)	
25–44	66 (43.1)	630	Referent	
45–64	29 (19.0)	313	0.87 (0.55 to 1.38)	
65+	20 (13.1)	269	0.69 (0.41 to 1.16)	
Ethnicity, n=1484‡				
Black African	25 (10.9)	207	1.14 (0.66 to 1.98)	4.68 (2.24 to 9.75)
Black Caribbean	3 (1.3)	13	3.26 (0.85 to 12.48)	5.56 (1.34 to 23.0)
Indian	27 (11.8)	201	1.05 (0.59 to 1.85)	2.83 (1.31 to 6.08)
Mixed/other	14 (6.1)	107	0.65 (0.3 to 1.42)	2.05 (0.81 to 5.17)
Pakistani	88 (38.4)	418	1.48 (0.97 to 2.27)	4.87 (2.76 to 8.57)
White	68 (29.7)	477	Referent	
Country of birth, n=1381				
UK	75 (51.7)	497	2.7 (1.44 to 2.97)	3.62 (2.19 to 5.99)
Not UK born	70 (45.8)	884	Referent	
Year of entry, n=773				
Entry <1 year	11 (16.7)	178	Referent	
Entry UK 2–4 years	22 (33.3)	173	2.21 (1.04 to 4.71)	§
Entry UK 5 to 9 years	14 (21.2)	177	1.30 (0.58 to 2.96)	
Entry UK > 10 years	19 (28.8)	245	1.28 (0.59 to 2.75)	

n represents the number of persons for whom information for each category was available.

†Final model, n=1356.

‡Bangladeshi, Chinese and black/other were grouped under mixed/other as no clustered cases were identified.

§Not included in the model, to include UK born and non-UK born.

Table 2 Clinical and social risk factor characteristics of clustered cases or cases involved in recent TB transmission events; North of England, 2010–2012

Variable	No. (%) clustered	Total no. available	Crude OR (95% CI)	Adjusted OR (95% CI)*
Clinical presentation of TB, n=1482				
Pulmonary	125 (81.7)	999	2.32 (1.5 to 3.69)	2.17 (1.32 to 3.57)
Extrapulmonary only	28 (18.3)	483	Referent	
Sputum smear positivity n=506				
Yes	56 (76.7)	369	1.26 (0.69 to 2.41)	Referent
No	17 (23.3)	137	Referent	
Previous TB diagnosis, n=1362				
Yes	13 (9.3)	86	1.61 (0.8 to 3.03)	Referent
No	127 (90.7)	1276	Referent	
Patient on DOT, n=1149				
Yes	15 (13.2)	96	1.78 (0.92 to 3.27)	Referent
No	99 (86.8)	1053	Referent	
Having a social risk factor, n=1484				
Yes	25 (16.3)	140	2.07 (1.24 to 3.35)	Referent
No	128 (83.7)	1344	Referent	
History of being homeless, n=1484				
Yes	10 (6.5)	40	3.03 (1.29 to 6.53)	2.84 (1.18 to 6.80)
No	143 (93.5)	1444	Referent	
Currently homeless, n=1484				
Yes	3 (2)	18	4.05 (1.09 to 12.8)	Referent
No	150 (98)	1466	Referent	
Homeless in the last 5 years n=1484				
Yes	5 (3.3)	16	3.04 (0.91 to 9.07)	Referent
No	148 (96.7)	1468	Referent	
Homeless >5 years ago, n=1484				
Yes	1 (0.4)	4	2.91 (0.06 to 36.5)	Referent
No	152 (99.6)	1480	Referent	
History current alcohol use, n=1484				
Yes	10 (6.5)	61	1.76 (0.78 to 3.59)	Referent
No	143 (93.5)	1423	Referent	
History or current drug use, n=1484				
Yes	9 (5.9)	53	2.11 (0.61 to 5.85)	Referent
No	144 (96.4)	1458	Referent	
Current drug use, n=1484				
Yes	5 (3.3)	26	2.3 (0.8 to 5.9)	Referent
No	148 (96.7)	1506	Referent	
Drug use in the last 5 years, n=1484				
Yes	4 (2.5)	21	2.08 (0.5 to 6.47)	Referent
No	149 (97.5)	1463	Referent	
Drug use >5 years ago, n=1484†				
History or currently in prison, n=1484				
Yes	11 (7.2)	52	2.44 (1.1 to 4.96)	2.63 (1.18 to 5.87)
No	142 (92.8)	1432	Referent	
Currently in prison, n=1484				
Yes	2 (1.3)	11	1.95 (0.2 to 9.52)	Referent
No	151 (98.7)	1473	Referent	
Prison UK last 5 years, n=1484				
Yes	6 (2.5)	18	2.53 (0.6 to 8.17)	Referent
No	230 (97.5)	1466	Referent	
Prison UK >5 years, n=1484				
Yes	5 (3.3)	21	2.78 (0.78 to 8.07)	Referent
No	148 (96.7)	1463	Referent	
Prison abroad last 5 years, n=1484				
Yes	1 (0.7)	4	2.91 (0.06 to 36.5)	Referent
No	152 (99.3)	1480	Referent	
Prison abroad >5 years ago, n=1484†				

n represents the number of persons for whom information for each category was available.

*Final model, n=1356.

†No cases with the outcome of interest.

DOT, directly observed therapy.

ago. Out of the nine, four were of Pakistani origin, three were of black African origin and two were mixed/other.

Multivariable analysis

In the final multivariable model, being UK born, being homeless, past incarceration and pulmonary disease remained statistically significantly associated with being part of a cluster (table 2). Cases belonging to an ethnic group other than white or mixed/other were more likely to be clustered, especially cases of black Caribbean origin (aOR 5.6, 95% CI 1.7 to 23.0) or Pakistani origin (aOR 4.9, 95% CI 2.8 to 8.6) (table 1).

DISCUSSION

We estimate that 10% of TB cases in the North of England in 2010–2012 were the result of localised transmission. Being UK born, homelessness, incarceration, presenting with pulmonary disease and being of an ethnic background other than white or mixed/other were associated with localised TB transmission. Our estimate of localised TB transmission showed geographical variation, although due to artificial geographical boundaries local estimates may be an underestimate. Our estimate differs from the estimate using the same method in the Centers for Disease Control and Prevention study.⁹ This might be due to differences in TB epidemiology but the smaller sample size and the use of 24 as opposed to 12-locus MIRU-VNTR may also contribute.^{5 20} In the absence of detailed epidemiological data for all cases of TB, the use of routine epidemiological data together with genotyping data offers an alternative way to assess localised TB transmission at subnational and local levels. This method obtained a more conservative estimate compared with methods previously used, allowing more targeted investigation of clusters. In addition, the method was easy to perform, could be run regularly and requires minimum training.

Studies estimating recent TB transmission have mostly relied on the identification of cases with indistinguishable MIRU-VNTR profiles, assuming that they represent recent transmission. This is not necessarily the case and inferences based on crude estimates need cautious interpretation. Traditional contact-tracing only uncovers a small proportion of epidemiological links between patients in the same molecular cluster.^{1 3 21} Intensive contact-tracing can identify a much higher proportion of links,²² but its application is limited by the available resources especially TB nurse staffing levels.

A precise estimate of recent transmission requires information on parameters including the annual rate of TB infection, existence of effective control programmes, intensity of contact tracing activities, population movements, the number of different circulating strains in the past and new introductions and the number of people infected per case, which may not be known.²³ Interpretation of clustering results based only on genotypical data is challenging. A combination of molecular data with spatial scan statistics provides an innovative way of defining TB clusters. Although there is uncertainty about which definition describes localised TB transmission most accurately, adding a spatiotemporal component has the advantage of identifying areas where the number of cases is higher than expected and where transmission may be ongoing. If performed routinely the method could be used to inform and evaluate TB control activities.

The characteristics identified as associated with being part of a cluster are compatible with reports from London, Denmark and the USA.^{24–27} A previous US study that used similar methods to predict TB outbreaks also reported that the presence of characteristics such as homelessness and incarceration in at least one of the

three first cases in a cluster indicated a likely outbreak.²⁷ Transmission in these groups could reflect chaotic lifestyles or differences in health-seeking behaviour or access to health services that result in delayed diagnosis. They could also reflect complex transmission patterns not yet well understood. Routine contact tracing may miss transient contact in these groups.

The study has a number of limitations. First, our sample population did not include all cases reported to ETS as only 54% (1547/2845) of the cases were culture-positive and had a complete genotype. This is likely to have resulted in an underestimate of recent transmission, if we assume that non-cultured cases were also infectious. Second, our study period was only 3 years. Increasing the length of the study could also increase the proportion of clustered cases. However, using a longer study period can make interpretation of results challenging as some of the cases are likely to be due to reactivation of TB acquired earlier rather than to localised transmission.²⁸ Increasing the length of the study may increase the number of cases due to localised transmission only to reach a plateau determined by the mutation rate of TB.¹⁹ The mutation rate influences the degree of clustering, but uncertainty remains regarding the mutation rate of TB, with several studies suggesting a low frequency of MIRU-VNTR genotype changes.^{29 30} This uncertainty complicates the interpretation of methods using MIRU-VNTR data and highlights the need to estimate the underlying mutation rate. Novel approaches such as whole genome sequencing are encouraging and are likely to improve our understanding of TB microevolution, heterogeneity and transmission.^{31–34} Third, although the definition used in our study is more complex than relying only on genotyping data, it is still only a surrogate for localised TB transmission and does not take into account other factors that affect the interpretation of the results such as migration and diversity of underlying strains circulating or new introductions. For instance, similar ethnic groups might settle in the same areas and introduce genotypes that may reflect common strains in the country of origin and may not reflect local transmission.³⁵ Fourth, by limiting the study to the North of England we were likely to miss cases whose exposure or transmission to others might have occurred outside this geographical area.

Our study suggests that efforts to control TB could target recent or ongoing transmission by prioritising those with identified risk factors. Improving access to healthcare and more intense contact-tracing efforts might be needed but are likely to require additional resources. Routine use of the method described in this paper could assist TB control strategies at a local level by identifying and monitoring where TB transmission events are occurring and the risk factors associated with these events.

Author affiliations

¹Field Epidemiology Training Programme, Public Health England, Liverpool, UK

²Field Epidemiology Services, Public Health England, Liverpool, UK

³European Programme for Intervention Epidemiology Training, European Centre for Diseases Prevention and Control, Stockholm, Sweden

⁴Greater Manchester Public Health England Centre, Public Health England, Manchester, UK

⁵Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

⁶Microbiology Services, Public Health England, Newcastle upon Tyne, UK

⁷Field Epidemiology North East, Public Health England, Newcastle upon Tyne, UK

⁸Yorkshire & Humber Public Health England Centre, Public Health England, Leeds, UK

⁹North East Public Health England Centre, Public Health England, Newcastle upon Tyne, UK

Authors note Dr Marko Petrovic died on 13 March 2015, between submission and publication of this paper. He worked extensively and tirelessly on TB control in North West England. We dedicate this paper to his memory.

Acknowledgements The authors thank Dr Maeve Lalor (Public Health England) for providing data from the national TB data set (matched ETS and laboratory data), Sanch Kanagarajah (Public Health England) for technical assistance, Sara Sarginson and Deborah Osborne (Public Health England, Newcastle Laboratory) for performing the MIRU-VNTR typing, the TB nurses and physicians that input data onto ETS, Dr Sam Bracebridge (Public Health England) and Dr Yvan Hutin (ECDC) for their helpful comments. The authors also thank the PHE National Strain Typing Programme Board for their support.

Contributors WW and MP conceived the study. MS-C and PC performed the analysis. MS-C wrote the first draft. All authors supported interpretation and revised the manuscript.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- Bifani PJ, Mathema B, Liu Z, *et al.* Identification of a W variant outbreak of mycobacterium tuberculosis via population-based molecular epidemiology. *JAMA* 1999;282:2321–7.
- Bishai WR, Graham NH, Harrington S, *et al.* MOlecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. *JAMA* 1998;280:1679–84.
- Small PM, Hopewell PC, Singh SP, *et al.* The epidemiology of tuberculosis in San Francisco—a population-based study using conventional and molecular methods. *N Engl J Med* 1994;330:1703–9.
- Barnes PF, Cave MD. Molecular epidemiology of tuberculosis. *N Engl J Med* 2003;349:1149–56.
- Mathema B, Kurepina NE, Bifani PJ, *et al.* Molecular epidemiology of tuberculosis: current insights. *Clin Microbiol Rev* 2006;19:658–85.
- Supply P, Allix C, Lesjean S, *et al.* Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of mycobacterium tuberculosis. *J Clin Microbiol* 2006;44:4498–510.
- Health Protection Agency. *TB strain typing cluster investigation handbook for health protection units*. 2nd edn. 2011. http://web.archive.nationalarchives.gov.uk/20131102021940/http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317131018354 (accessed 20 Jun 2014).
- Sonnenberg P. Evaluation of TB strain typing service—summary and key recommendations. 2013. http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317140772345 (accessed 20 Jun 2014).
- Moonan PK, Ghosh S, Oeltmann JE, *et al.* Using genotyping and geospatial scanning to estimate recent mycobacterium tuberculosis transmission, United States. *Emerg Infect Dis* 2012;18:458–65.
- Kulldorff M, Heffernan R, Hartman J, *et al.* A space-time permutation scan statistic for disease outbreak detection. *PLoS Med* 2005;2:e59.
- Stelling J, Yih WK, Galas M, *et al.* Automated use of WHONET and SaTScan to detect outbreaks of Shigella spp. using antimicrobial resistance phenotypes. *Epidemiol Infect* 2010;138:873–83.
- Rushton SP, Goodfellow M, O'Donnell AG, *et al.* The epidemiology of atypical mycobacterial diseases in northern England: a space-time clustering and Generalized Linear Modelling approach. *Epidemiol Infect* 2007;135:765–74.
- Briggs ADM, Boxall NS, VAN Santen D, *et al.* Approaches to the detection of very small, common, and easily missed outbreaks that together contribute substantially to human Cryptosporidium infection. *Epidemiol Infect* 2014;142:1869–76.
- ArcGIS Desktop: version 10.0 Redlands. Environmental System Research Institute, CA, USA. 2010.
- Office for National Statistics. Super output areas. <http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas/index.html> (accessed 15 Jun 2013).
- Kulldorff M. *SaTScan version 9.2. Software for the spatial, temporal and space-time scan statistics*. 2013.
- Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods* 1997;26:1481–96.
- Office for National Statistics. Super Output Area mid-year population estimates for England and Wales (experimental), Mid-2002 to Mid-2010 revised. <http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcn%3A77-285154> (accessed 14 Jun 2013).
- Glynn JR, Vyonycky E, Fine PEM. Influence of sampling on estimates of clustering and recent transmission of mycobacterium tuberculosis derived from DNA fingerprinting techniques. *Am J Epidemiol* 1999;149:366–71.
- Allix-Beguec C, Fauville-Dufaux M, Supply P. Three-year population-based evaluation of standardized mycobacterial interspersed repetitive-unit-variable-number tandem-repeat typing of Mycobacterium tuberculosis. *J Clin Microbiol* 2008;46:1398–406.
- Valway SE, Sanchez MP, Shinnick TF, *et al.* An outbreak involving extensive transmission of a virulent strain of Mycobacterium tuberculosis. *N Engl J Med* 1998;338:633–9.
- Van Deutekom H, Hoijng SP, de Haas PEW, *et al.* Clustered tuberculosis cases: do they represent recent transmission and can they be detected earlier? *Am J Respir Crit Care Med* 2004;169:806–10.
- Glynn JR, Bauer J, de Boer AS, *et al.* Interpreting DNA fingerprint clusters of Mycobacterium tuberculosis Position Paper. *Int J Tuberc Lung Dis* 1999;3:1055–60.
- Tuberculosis in the UK: Annual report on tuberculosis surveillance in the UK. 2013. http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317139689583 (accessed 10 Dec 2013).
- Maguire H, Dale JW, McHugh TD, *et al.* Molecular epidemiology of tuberculosis in London 1995–7 showing low rate of active transmission. *Thorax* 2002;57:617–22.
- Kemper-Jorgensen Z, Andersen AB, Kok-Jensen A, *et al.* Clustered tuberculosis in a low-burden country: nationwide genotyping through 15 years. *J Clin Microbiol* 2012;50:2660–7.
- 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.
- Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol Infect* 1997;119:183–201.
- Supply P, Niemann S, Wirth T. On the mutation rates of spoligotypes and variable numbers of tandem repeat loci of Mycobacterium tuberculosis. *Infect Genet Evol* 2011;11:251–2.
- Savine E, Warren RM, van der Spuy GD, *et al.* Stability of variable-number tandem repeats of mycobacterial interspersed repetitive units from 12 loci in serial isolates of Mycobacterium tuberculosis. *J Clin Microbiol* 2002;40:4561–6.
- Van Soolingen D. Whole-genome sequencing of Mycobacterium tuberculosis as an epidemiological marker. *Lancet Respir Med* 2014;2:251–2.
- Walker TM, Lalor MK, Broda A, *et al.* Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK. *Lancet Respir Med* 2014;2:285–92.
- Niemann S, Koser CU, Gagneux S, *et al.* Genomic diversity among drug sensitive and multidrug resistant isolates of Mycobacterium tuberculosis with identical DNA fingerprints. *PLoS ONE* 2009;4:e7407.
- Perez-Lago L, Comas I, Navarro Y, *et al.* Whole genome sequencing analysis of intrapatient microevolution in Mycobacterium tuberculosis: potential impact on the inference of tuberculosis transmission. *J Infect Dis* 2014;209:98–108.
- Borgdorff MW, van Soolingen D. The re-emergence of tuberculosis: what have we learnt from molecular epidemiology? *Clin Microbiol Infect* 2013;19:889–901.