Molecular epidemiology of tuberculosis in London 1995–7 showing low rate of active transmission

H Maguire, J W Dale, T D McHugh, P D Butcher, S H Gillespie, A Costetos, H Al-Ghusein, R Halland, A Dickens, L Marston, P Wilson, R Pitman, D Strachan, F A Drobniewski, D K Banerjee

ORIGINAL ARTICLE

Background: Tuberculosis notification rates for London have risen dramatically in recent years. Molecular typing of Mycobacterium tuberculosis has contributed to our understanding of the epidemiology of tuberculosis throughout the world. This study aimed to assess the degree of recent transmission of M. tuberculosis in London and subpopulations of the community with high rates of recent transmission.

Methods: M. tuberculosis isolates from all persons from Greater London diagnosed with culture positive tuberculosis between 1 July 1995 and 31 December 1997 were genetically fingerprinted using IS6110 restriction fragment length polymorphism (RFLP) typing. A structured proforma was used during record review of cases of culture confirmed tuberculosis. Cluster analysis was performed and risk factors for clustering were examined in a univariate analysis followed by a logistic regression analysis with membership of a cluster as the outcome variable.

Results: RFLP patterns were obtained for 2042 isolates with more than four copies of IS6110; 463 (22.7%) belonged to 169 molecular clusters, which ranged in size from two (65% of clusters) to 12 persons. The estimated rate of recent transmission was 14.4%. Young age (0–19 years) (odds ratio (OR) 2.65, 95% confidence interval (CI) 1.59 to 4.44), birth in the UK (OR 1.55, 95% CI 1.04 to 2.03), black Caribbean ethnic group (OR 2.19, 95% CI 1.15 to 4.16), alcohol dependence (OR 2.33, 95% CI 1.46 to 3.72), and streptomycin resistance (OR 1.82, 95% CI 1.15 to 2.88) were independently associated with an increased risk of clustering.

Conclusions: Tuberculosis in London is largely caused by reactivation or importation of infection by recent immigrants. Newly acquired infection is also common among people with recognised risk factors. Preventative interventions and early diagnosis of immigrants from areas with a high incidence of tuberculosis, together with thorough contact tracing and monitoring of treatment outcome among all cases of tuberculosis (especially in groups at higher risk of recent infection), remains most important.

Ethical approval for the study was obtained from all district and trust ethics committees in London. Isolates of M. tuberculosis from persons diagnosed with tuberculosis between 1 July 1995 and 31 December 1997 in any participating laboratory in Greater London (54 NHS laboratories inside the London orbital motorway, M25) were included. Isolates archived at the Public Health Laboratory Service (PHLS) Mycobacterium Reference Unit (MRU) at Dulwich Public Health Laboratory and at the reference facility of the Royal Brompton Hospital were obtained. Isolates of mycobacteria other than tuberculosis were excluded.

Molecular analysis

Isolates of M. tuberculosis were genetically fingerprinted using IS6110 RFLP typing using the international standard protocol at St George’s Hospital Medical School (SGHMS) and the Royal Free & University College Medical School (RF&UCMS). All patterns were entered by one researcher (AD) onto a database using GelCompar software (Version 4.0, Applied Maths, Kortrijk, Belgium) at RF&UCMS and then were analysed independently with the aid of GelComp at the University of Surrey. Cluster analysis was performed using the Dice coefficient. Similarity defined by this coefficient was calculated using the parameter settings at 1.2% band position.
tolerance with optimisation. A molecular cluster was defined as a series of isolates with identical banding patterns (100% identity), subject to visual verification. Strains that differed by one band were regarded as not belonging to the same molecular cluster. Isolates with only 1–4 copies of IS6110 were considered unevaluable and were excluded from the analysis. These isolates were submitted to secondary typing analysis and the data will be presented elsewhere.

Isolates were also excluded from the study if they were the result of a laboratory contamination event, fulfilling all of the following criteria: (1) only one culture was obtained from the patient, (2) there was no evidence of smear positivity on any sample, (3) the date of specimen processing in the source laboratory was within five working days of an indistinguishable isolate from another patient, and (4) the clinical course (in retrospect) was not consistent with tuberculosis infection. The percentage clustering was calculated by subtracting the number of clusters from the number of isolates in the clusters and dividing by the total number of isolates. This procedure avoids double counting of clustered isolates by reducing the size of each cluster by one."

**Epidemiological data collection**

A structured proforma was used during record review to collect information about cases of culture confirmed tuberculosis. Over 150 clinicians and clinic staff at more than 50 hospitals (or clinics) assisted with data collection unaware of cluster status. Two health authorities (in east and south east London) were able to provide data that were stored in local databases. Additional microbiological data from the Mycobacteri al Reference Unit and the Royal Brompton Hospital were obtained from the PHLS UK antimicrobial resistance surveillance network (Mycobnet) database. Ethnic groups were those used in the 1991 census. Potential links between members of the clusters were determined by reference to these proformas after molecular cluster results were known.

**Cluster analysis**

Established links between members of a cluster were considered to exist if members were known to each other, were named contacts or family members, or had close social or institutional contact. Possible links were judged to exist if two or more members lived in the same or an adjacent four digit postcode area and either (a) one member had a risk factor (HIV, alcohol misuse, homelessness) or (b) two members had the same ethnic group or country of birth.

Risk factors for “clustering” were examined initially in a univariate analysis. Two models were then constructed in a multi-...
RESULTS

Epidemiological and risk factor data were collected on 2495 patients (table 1). RFLP patterns were obtained for 2779 isolates of M tuberculosis. After eliminating multiple isolates from the same patient, 2500 individual patients were included in the study comprising 77% of the total number of isolates from individuals in London (n=3260) reported as culture confirmed cases in the relevant time frame to Mycobnet. Ten further isolates were excluded as showing evidence of originating through cross contamination, making a final study population of 2490 individuals. Missed isolates were not systematically different in respect of age, sex, ethnic group, or source hospital.

There was a bimodal distribution of IS6110 copy number (fig 1): 448 isolates (17.9%) had 1–4 copies and 2042 isolates had five or more copies (modal value=11). The low copy number isolates are generally regarded as not fully evaluable with IS6110 due to the low degree of discrimination shown, although there is no consistent definition of the cut off. In this study, analysis of clustering by band number showed that the isolates with five copies were well differentiated (fig 2) and we therefore included these in the multicopy group.

Of the 2042 isolates with more than four copies of IS6110 there were 463 individuals (22.7%) whose isolates fell in 169 clusters which ranged in size from 2 to 12 persons (fig 3). Most of these clusters were small: 110 (65%) contained only two members and only 14 (8.3%) had more than four members.

Assuming that clustering represents recent transmission, the estimated rate of active transmission of tuberculosis in London was 14.4%.

Description of clusters

The largest cluster (designated cluster 3908) involved 12 people; eight were known to have been born in Somalia, all of whom had arrived in the UK since 1990 (five during or since 1995). All 12 members lived in widely separated areas of London (table 2) in seven different districts and they were diagnosed at 11 different hospitals. There were no known connections between them apart from their ethnic group. Three had pulmonary disease and one was known to have been smear positive. Seven had organisms that were sensitive to all the main antimicrobial agents (pyrazinamide, isoniazid, rifampicin, ethambutol, streptomycin); four had streptomycin resistance and the second case with a specimen date in September 1995 who had arrived from Somalia earlier that year had multidrug resistant (MDR) tuberculosis (resistant to isoniazid and rifampicin).

The next largest cluster involved 10 patients (cluster 2066). Members were of mixed ethnic group (four white, one black African, one black Caribbean, and one Chinese). Three were known to have been born outside the UK (one China, one Ireland, one Nigeria), arriving in 1983 from Ireland and in 1992 from Nigeria. Members mainly lived in different parts of London (seven districts, diagnosed at eight different hospitals; table 2), although two were in adjacent four digit postcode areas. Eight had pulmonary disease and six were sputum smear or culture positive. No connections were found between them except that three were alcohol dependent.

There were two clusters involving seven people. The first (cluster 905) involved seven men aged 29–64 years; four of five whose ethnicity was known were white and the other was black Caribbean, and four of these five were UK born including the black Caribbean. The country of birth for the remainder of this cluster was unknown. Four lived in south east London, two of whom were in adjacent four digit postcode areas. Three were alcohol dependent including one with HIV infection and other substance misuse who was non-compliant with treatment, and a second patient who also misused other recreational drugs. An additional 42 year old white patient was homeless. The second cluster of seven (cluster 4350) comprised people of mixed ethnicity of whom four were known to have been born in the UK.

One cluster of six people (cluster 1171) included a husband (with retroviral infection noted to be poorly compliant with treatment) and wife who were both black Caribbean with MDR tuberculosis. Four of the six had adjacent postcodes. The other cluster of six (cluster 4143) involved five white men and one white woman aged 33–66 years (three alcoholic and one with retroviral infection). Three were contacts and poor compliance with treatment was noted.

**Figure 1** Frequency distribution of number of copies of IS6110 for 2490 clinical isolates of Mycobacterium tuberculosis.

**Figure 2** Percentage of clustered isolates against IS6110 copy number demonstrating the poor differentiation of strains with <4 copies.

**Figure 3** Frequency distribution of cluster size for isolates with ≥5 copies of IS6110.
Transmission links

For 14 clusters (8.3%), including clusters 1171 and 4143 (table 2), transmission links were established. In a further 27 clusters (15.9%) links were possible (including clusters 905, 2066, and 3908; table 2), but 128 clusters (79.8%), including cluster 4350, remained unexplained. Transmission routes that were identified included household spread, family links, and close social contact or other close contact within an institution.

Statistical analysis of clustering

Detailed epidemiological and risk factor information was obtained for 1578 (77%) of the 2042 individuals with a high copy number (>4 IS6110 bands) strains; age, sex, and antimicrobial sensitivity data were derived from Mycobnet for all 2042 cases.

In the multivariate statistical analysis (table 3) young age (0–19 years, OR 2.65, 95% CI 1.59 to 4.44), birth in the UK (OR 1.55, 95% CI 1.04 to 2.03), black Caribbean ethnic group (OR 2.19, 95% CI 1.15 to 4.16), alcohol dependence (OR 2.33, 95% CI 1.46 to 3.72), and streptomycin resistance (OR 1.82, 95% CI 1.15 to 2.88) were independently associated with increased clustering. Factors not independently associated with clustering included HIV infection, MDR tuberculosis, black African or Indian subcontinent ethnicity, and drug misuse.

DISCUSSION

The overall rate of apparently recent transmission in London from 1995 to 1997 estimated in this study (14.4%) is substantially lower than that reported in other population based studies (21–35%). This comparison needs to be treated with caution because of considerable differences in study design and analysis. We considered the possibility that a longer time period might be necessary to disclose the full extent of clustering. However, analysis of our data by time periods showed that the rate of clustering had reached 11% after only 1 year and thereafter increased only slightly, suggesting that a longer time period would have made very little difference. On the other hand, the maximum cluster size did continue to increase throughout the study, as would be expected, which may contribute to the difference between this and other studies where larger cluster sizes have been reported. Howse ever, it seems unlikely, even in a much longer study, that recruitment to clusters would achieve the levels reported in the Netherlands and Denmark as these relate to specific and known chains of transmission.

Table 2: Composition and characteristics of the large clusters

In this study we have excluded isolates with few copies of IS6110, although in some comparable studies these isolates were subjected to secondary typing—for example, by PGRS typing. There has been no consensus cut off definition for low copy strains, so that isolates with fewer than six IS copies or fewer than five IS copies and identical by both IS6110 and PGRS typing are also counted as clustered. We have chosen to consider these low copy isolates separately as there is at present no certainty that the two definitions are equivalent, and merging the data may produce unreliable results. However, it seems highly unlikely that this could contribute significantly to the lower level of clustering found in London in this study.

It therefore seems likely that the relatively low rate of clustering and the absence of really large clusters are a genuine reflection of differences between the epidemiology of tuberculosis in London at the time of this study and that in the other populations. Our results are comparable to those from a previous but smaller study (n=555) in inner London in 1993 which found a rate of clustering of 19%. Recent infection, as indicated by clustering, was found to be more common in our study in people with recognised risk factors such as alcohol misuse, as well as in younger patients. Similar associations have been found in some other studies.
not associated with clustering, which is similar to the findings of some previous studies but not others. Previous studies have shown an association of multiple drug resistance with apparent recent transmission. We did not find such an association and these results are similar to those of van Soolingen et al. An association was noted with birth in the UK. We suggest that in those not born in the UK a high proportion will have acquired tuberculosis outside the UK and thus will be part of clusters not seen in this study. A further association with clustering is observed with streptomycin resistance but not with any other antimicrobial agent. We conclude that this observation is confounded by an association between Somali born and both clustering and streptomycin resistance, noted in cluster 3908; country of birth was not a factor in the multivariate analysis.

Our study suggests that recent transmission in London plays a lesser role in the occurrence of tuberculosis than in the other cities studied, and that most tuberculosis in London is due to reactivation of previous infections or importation of infection by recent immigrants. It is notable that the largest infection by recent immigrants. It is notable that the largest

<table>
<thead>
<tr>
<th>Variable</th>
<th>No</th>
<th>% clustering</th>
<th>OR (95% CI)</th>
<th>OR (95% CI)</th>
</tr>
</thead>
</table>
| Age
| 0-19        | 150| 33.3         | 2.49        | 1.59 to 3.89 | 2.65 | 1.59 to 4.44 |
| 20-34        | 831| 24.2         | 1.54        | 1.11 to 2.14 | 1.51 | 1.02 to 2.22 |
| 35-59        | 648| 23.6         | 1.59        | 1.14 to 2.24 | 1.43 | 0.97 to 2.11 |
| 60+          | 335| 16.7         | 1.00        | Reference    | 1.00 | Reference    |
| Sex
| Male        | 1169| 24.5        | 1.00        | Reference    | 1.00 | Reference    |
| Female      | 827 | 21.0         | 0.50        | 0.22 to 1.12 | 0.46 | 0.10 to 2.15 |
| Unknown*    | 50  | 14           | 0.53        | 0.37 to 0.75 | 0.62 | 0.41 to 0.92 |
| Ethnic group
| White       | 274 | 37.2        | 1.00        | Reference    | 1.00 | Reference    |
| ISC         | 340 | 13.2        | 0.26        | 0.17 to 0.38 | 0.34 | 0.22 to 0.52 |
| BC          | 48  | 54.2        | 1.99        | 1.07 to 3.70 | 2.19 | 1.15 to 4.16 |
| BA          | 409 | 33.5        | 0.85        | 0.62 to 1.17 | 1.02 | 0.68 to 1.53 |
| Black other | 24  | 29.2        | 0.69        | 0.28 to 1.73 | 0.76 | 0.30 to 1.96 |
| Other       | 156 | 19.9        | 0.42        | 0.26 to 0.66 | 0.52 | 0.31 to 0.86 |
| Unknown*    | 328 | 23.8        | 0.53        | 0.37 to 0.75 | 0.62 | 0.41 to 0.92 |
| Birth in UK
| UK born     | 221 | 40.3        | 2.12        | 1.56 to 2.88 | 1.55 | 1.04 to 2.03 |
| Not-UK born | 928 | 24.1        | 1.00        | Reference    | 1.00 | Reference    |
| Unknown*    | 431 | 26.4        | 1.13        | 0.87 to 1.47 | 1.17 | 0.87 to 1.57 |
| Other factors
| Alcohol dependence | 94  | 48.9        | 2.78        | 1.82 to 4.23 | 2.33 | 1.46 to 3.72 |
| Streptomycin resistant | 115 | 32.2        | 1.66        | 1.11 to 2.49 | 1.82 | 1.15 to 2.88 |

BA=black African; BC=black Caribbean; ISC=Indian subcontinent.

*Information not known following record review.

ACKNOWLEDGEMENTS
We would like especially to thank the following: H Atkinson, P Atkinson, O J Billington, G Bothamley, J Carless, M Chadwick, R Coker, M Connolly, S Crossland, N Crowcroft, G French, H Gaya, V Gileece, S Handford, J Herbert, P Horby, K Lau, D Lamprecht, H Milburn, L Newport, J O’Sullivan, F Palmer, B Pankhania, M Perkin, A Pozniak, I Roddick, R Shaw, J Watson, M Yates. We also thank consultant microbiologists, chest physicians, TB nurses, consultants in communicable disease control, and other staff right across London. We are indebted to all colleagues who gave time and resources to the study.

Authors’ affiliations
H Maguire, Public Health Laboratory Service (PHLS) Communicable Disease Surveillance Centre, London W2 3QR, UK
J W Dale, Molecular Microbiology Group, School of Biomedical & Life Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK
T D McHugh, S H Gillespie, A Dickens, Department of Medical Microbiology, Royal Free & University College Medical School, Royal Free Campus, London NW3 2PF, UK
P D Butcher, A Costetso, H Al-Ghusein, R Holland, D K Banerjee, Department of Medical Microbiology, St George’s Hospital Medical School, London SW17 0RE, UK
L Marston, D Strachan, Public Health Sciences, St George’s Hospital Medical School, London SW17 0RE, UK
F Wilson, Microbiology Department, The Royal London Hospital, London E1 1BB, UK
R Pitman, Respiratory Division, PHLS Communicable Disease Surveillance Centre, London NW9 5EQ, UK
P A Drobniewski, PHLS Mycobacterial Reference Unit, Public Health Laboratory and Medical Microbiology, Guys, Kings & St Thomas’ School of Medicine, London SE22 8GF, UK

DKB is Chairman and Project Leader of the Molecular Epidemiology of Tuberculosis in London Steering Committee and responsible for overall coordination of the project. The project was conceived of and instigated by DKB, PDB, JWD, SHG, HM, TDMcH, and PW. All authors contributed to the study design and interpretation of results. PDB, JWD, FAD, SHG, and TDMcH were responsible for the design and supervision of molecular...
microbiological methods. HA-G, AD, and TDMcH performed the typing methods. AD and JWD undertook the typing data entry and analysis. Epidemiological methodology was developed by HM. AC and RH were responsible for collation of the epidemiological data, supervised by HM, JWD, and DS. Statistical analysis and interpretation was performed by HM, LM, RP, and DS. Linkage of datasets was performed by RP and JWD. HM produced the first draft of this paper and revised drafts with HM, LM, RP, and DS. Linkage of datasets was performed by RP and JWD with feedback from all authors. All authors have seen and approved the final version for submission. DBK is guarantor.

Funding: This work was undertaken by DBK who received funding from the NHS Executive London, Research & Development Programme. The views expressed in the publication are those of the authors and are not necessarily those of the NHS Executive or the Department of Health. The authors are also grateful for support from the European Union under grants BMH4-CT97-91202 and SMT4-CT96-2097 (provision of GelCompar software).

REFERENCES