Epidemiol. Infect. (2002), **128**, 175–184. © 2002 Cambridge University Press
DOI: 10.1017/S0950268801006690 Printed in the United Kingdom

The molecular epidemiology of tuberculosis in inner London

A. C. HAYWARD[1,*], S. GOSS[2], F. DROBNIEWSKI[2], N. SAUNDERS[3], R. J. SHAW[4], M. GOYAL[5], A. SWAN[6], A. UTTLEY[2], A. POZNIAK[7], J. GRACE-PARKER[8] AND J. M. WATSON[8]

[1] *PHLS Communicable Disease Surveillance Centre – London/University College London, Centre for Infectious Disease Epidemiology, London*
[2] *PHLS Mycobacterium Reference Unit, Dulwich, London*
[3] *PHLS Virus Reference Division, Central Public Health Laboratories, London*
[4] *Director of Medicine and Professor of Respiratory Medicine, Hammersmith Hospitals NHS Trust, London*
[5] *Department of Respiratory Medicine, NHLI, ICSM London*
[6] *PHLS Statistics Unit, London*
[7] *Department of GUM/HIV Medicine, Chelsea and Westminster Hospital, London*
[8] *PHLS Communicable Disease Surveillance Centre, London*

(*Accepted 19 November 2001*)

SUMMARY

The study used DNA fingerprint typing (spoligotyping and Heminested-Inverse-PCR) of *Mycobacterium tuberculosis* from all culture-confirmed inner London patients over a 12-month period to describe transmission. The methodology was evaluated by comparison with standard IS6110 typing and by examining its ability to identify known household clusters of cases. Isolates sharing indistinguishable typing patterns using both techniques were defined as clustered. Clusters were investigated to identify epidemiological links. The methodology showed good discriminatory power and identified known household clusters of cases. Of 694 culture-confirmed cases, 563 (81%) were typed. Eleven (2%) were due to laboratory cross-contamination and were excluded. Of the remaining 552 isolates 148 (27%) were clustered. Multivariate analysis indicated that clustering was more common in those with pulmonary smear positive disease ($P < 0{\cdot}02$); those born in the United Kingdom ($P < 0{\cdot}0003$) and in patients living in south London ($P = 0{\cdot}02$). There was also a trend towards clustering being more common in those not known to have HIV infection ($P = 0{\cdot}051$). The results suggest that in inner London, recent local transmission makes an important contribution to notification rates.

INTRODUCTION

Between 1987 and 1997 tuberculosis notification rates in many parts of inner London more than doubled [1] and rates continue to increase. Although much of the rise is due to immigration from countries with a high prevalence of tuberculosis, local transmission of disease is also likely to play a role [1]. It has been noted that as well as rises in tuberculosis notification rates in non-white ethnic groups in London there has been an increase in notifications in white patients and in Indian Subcontinent ethnic group patients who were born in the United Kingdom [2]. Increases in these latter two groups suggest that transmission of

* Author for correspondence.

tuberculosis may be an important problem. Control of tuberculosis in new entrants relies on measures such as screening and use of chemoprophylaxis [3]. Control of tuberculosis transmission relies on ensuring early diagnosis and effective treatment of cases so that they remain infectious for the minimum time possible [3]. Transmission of tuberculosis in hospital settings has also been a very important problem in some cities [4]. Estimates of the extent of transmission of tuberculosis and risk factors for transmission are therefore needed to identify aspects of tuberculosis control in London that should be strengthened.

Molecular typing (particularly IS6110 Restriction Fragment Length Polymorphism – RFLP) [5] has allowed strain differentiation of *Mycobacterium tuberculosis* enabling transmission patterns to be described. The high degree of heterogeneity means indistinguishable strains are likely to represent chains of recent transmission [5]. This property has been exploited in a number of studies of tuberculosis transmission in different settings [6–10]. However, the time required to culture sufficient *M. tuberculosis* for DNA extraction delays the public health benefits of molecular epidemiological techniques. In this study we report the results of a large scale study using PCR-based typing methods to investigate tuberculosis transmission in inner London.

PCR-based typing techniques have the advantage of requiring less DNA than conventional IS6110 typing. This obviates the need for prolonged culture prior to typing and potentially permits typing from primary specimens. Several PCR-based typing techniques have been described [11–18] but most are not as discriminatory and reproducible as standard IS6110 typing [19]. Spoligotyping is a rapid technique which is less discriminatory than standard IS6110 typing in isolates with high copy numbers of IS6110, but more discriminatory in isolates with few copies [12]. Heminested Inverse PCR (HIPCR) typing [13] is a rapid method based on the IS6110 sequence. HIPCR is discriminating but is technically more demanding than spoligotyping, less amenable to electronic analysis and less reproducible [19]. These factors make it unsuitable for use as a primary typing technique but are of less importance when it is being used as a secondary typing technique (applying the technique to relatively small groups of isolates that have already been shown to be indistinguishable by the primary technique in an attempt to further distinguish between them). Combinations of typing techniques based on different genetic markers produce a high level of discrimination [12]. This study used spoligotyping and HIPCR to determine the role of PCR-based typing systems and to describe recent transmission of tuberculosis in inner London.

METHODS

The target population was all patients with culture-confirmed *M. tuberculosis* living or treated in inner London during 1993. Isolates were obtained from the Public Health Laboratory Service Mycobacterium Reference Unit, or from the Royal Brompton Hospital, London. Since microbiology laboratories from London submit all tuberculosis isolates to one of these reference laboratories it was expected that an almost complete sample of culture confirmed tuberculosis cases would be obtained. Only one isolate was included for each case. All isolates were confirmed as *M. tuberculosis* by biochemical, microscopic and growth characteristics and by molecular DNA hybridization analysis (using Acuprobe, *M. tuberculosis* detection systems) [20].

Spoligotyping (primary method) and HIPCR (secondary method) were used according to previously published methodology [12, 13]. Spoligotyping patterns were entered onto an Excel database and into Gelcompar/Version 4 (Applied Maths, Belgium). Groups of isolates (each group representing a cluster of isolates with indistinguishable spoligotyping patterns) were then subjected to secondary typing using HIPCR in order to distinguish further between them. Amplification products from secondary typing of each spoligotyping cluster were run on the same gel wherever possible. Isolates that were indistinguishable using both typing techniques were defined as clustered.

Cultures from 1993 were selected as epidemiological information on the notified cases available from the 1993 National Survey of Tuberculosis Notifications (one of a series of 5 yearly surveys that collected detailed epidemiological information on notifications) [21]. As the typing of isolates from 1993 was performed in 1996/7 many isolates were non-viable, preventing the acquistion of sufficient DNA subculture for standard IS6110 typing but offering an ideal opportunity to utilize and evaluate PCR-based methodology for population analysis of transmission. A preliminary analysis comparing the discrimination of typing using the PCR-based methods with standard IS6110 typing was conducted using 64 viable isolates to validate the techniques. The techniques were also

validated by determining whether they identified clusters when patients were known to share the same address. The typing and clustering were performed blinded to patient details in order to avoid bias.

Data on the source laboratory and the specimen date were examined for all clustered isolates. Smear negative cultures which were clustered with smear positive cultures with the same specimen date from the same laboratory were defined as being due to laboratory cross contamination if there were no additional positive cultures from that patient and if record-review of the clinical course was consistent with the patient not having tuberculosis.

Baseline epidemiological information on all patients was derived from the 1993 National Survey of Tuberculosis Notifications [21] and from physicians when a minimum data-set was not available from the survey. HIV status was ascertained by matching an anonymised study database to the PHLS HIV and AIDS register using soundex codes and first initials. To establish links, clusters were investigated using information from the survey and where necessary from patient records.

Analysis

The percentage of cases estimated to be due to recent transmission was calculated assuming that each cluster contained one index case (total number of clustered cases-number of clusters/total number of cases) [9]. Univariate analyses were performed using χ^2 tests (SPSS Version 8). Forward and backward logistic regression (SPSS Version 8) were then used to identify factors to include in a model. This was done using 'missing' categories for all variables in order not to lose data. Variables that made a contribution to the model were then forced into a logistic regression model using only those cases where data was complete for all these variables. This allowed identification of risk factors for clustering, calculation of adjusted odds ratios, 95% confidence intervals and associated Likelihood Ratio Test P values.

RESULTS

Validation of the typing techniques

Of the 569 isolates available only 64 (11%) were viable for culture and thus were typebable using standard IS6110 typing. Using the PCR based typing techniques we were able to type 562 (99%) of these 569 isolates. Typing patterns from Standard IS6110 RFLP typing of the 64 viable cultures identified 55 isolates (86%) that were unique. These 55 isolates were also identified as being unique using the combined typing method. Nine isolates fell into four clusters using IS6110 typing. Three of these clusters comprised pairs of isolates with multiple copies of IS6110 which were indistinguishable by all three techniques. One cluster included three patients with a single copy of IS6110. This cluster was split into three types by the combined method No clusters were identified by the combined method which had not been identified by IS6110 RFLP typing.

Amongst the 563 isolates typed there were 10 pairs of isolates from patients who were known to share the same address. Seven of these addresses (A–G) were normal residential houses, two were refugee hostels (H, I) and the other was a large block of council flats (J, the pair did not live in the same flat within this block). In five of the six residential households (A–E) the combined typing technique identified the pairs as sharing indistinguishable typing patterns. The typing patterns in four of these households (A–D) were not seen in any other isolates in the study. In one of these households (E) the typing pattern was seen in another five isolates. In the sixth household (F) the spoligotyping patterns from isolates from two sisters were indistinguishable but the HIPCR pattern in one was indistinct and appeared to differ by two bands from the other. In the seventh household (G, a husband and wife) both spoligotyping typing patterns and HIPCR patterns were markedly different. It was subsequently found that the husband did not have tuberculosis and that his isolate was due to laboratory cross-contamination (it was indistinguishable from a heavily smear positive isolate that was processed in the same laboratory on the same day). In the refugee hostels (H, I) and the large block of flats (J) spoligotyping patterns and HIPCR patterns were markedly different. These results are summarized in Table 1.

Extent of clustering

There were 694 culture-confirmed cases of tuberculosis in inner London during the study period. Of these, 82% (568 isolates) were available for typing. Typing results were obtained for 563 patients (81% of the target population). 11 isolates (2%) met the definition for laboratory cross-contamination and were

Table 1. *Typing results in patients known to be resident at the same address*

Pair	Type of accommodation	Spoligotyping results	HIPCR results
A–E	Normal residential	Indistinguishable	Indistinguishable
F	Normal residential	Indistinguishable	Different
G	Normal residential (1 case due to laboratory cross contamination)	Different	Different
H–I	Refugee hostels	Different	Different
J	Large block of council houses (different appartments)	Different	Different

Table 2. *Distribution of cluster size*

Cluster size	2	3	4	6	7	8	12	15
Number	26	5	3	3	1	2	1	1

excluded, leaving 552 patients. 147 isolates (27%) were clustered, the remainder had unique typing patterns. There were 42 clusters with a size range of 2–15, although most (62%) involved only 2 patients (Table 2). The extent of recent transmission was therefore estimated as 147–42/552 = 19%.

Risk factors for clustering

Table 3 shows the features of the study population, the percentage of clustered isolates in different risk groups and unadjusted odds ratios, 95% confidence intervals and *P* values. In this analysis clustering was associated with having pulmonary disease (especially if smear positive, $P = 0{\cdot}003$), living in South London ($P = 0{\cdot}008$), being white ($P = 0{\cdot}004$), being born in the United Kingdom ($P < 0{\cdot}0005$), being unemployed ($P = 0{\cdot}017$), being classified as alcoholic ($P = 0{\cdot}023$) and being HIV negative ($P = 0{\cdot}023$).

In both forwards and backwards stepwise logistic regression age, birth in the United Kingdom, HIV, unemployment, area of residence in London and site of disease were shown to contribute to the model. Table 4 shows the final logistic regression model using the 354 patients for whom all these data items were complete. All variables in the model are controlled for each other. In this analysis clustering was most strongly related to birth in the United Kingdom ($P = 0{\cdot}0003$), having pulmonary smear positive disease ($P = 0{\cdot}02$), and living in South London ($P = 0{\cdot}02$). Clustering appeared to be less common in those known to be infected with HIV, however this result was not statistically significant ($P = 0{\cdot}051$).

Epidemiological links in clustered cases

Since this was a retrospective study, epidemiological investigation of clusters was restricted to information derived from patient records. For 46% of clustered isolates there were no obvious features suggesting contact. Fifty-four percent (75) of the 147 patients in clusters had features (e.g. living within a two mile radius, same employment or homelessness) which may have made contact with others in their cluster more likely (including ten patients with known household contact); 71% of clusters (30/42) involved multiple ethnic groups.

Suspected incidents of hospital transmission were identified at two hospitals. Each incident involved two patients. In both incidents HIV infected patients appeared to have acquired tuberculosis from patients who were treated at the same hospitals. In one of the incidents the organism was resistant to isoniazid.

The largest cluster of 15 patients included 3 homeless men and a man living near a homeless shelter. In a cluster of 12 patients the apparent source (with the earliest specimen date) was a homeless alcoholic man with heavily smear-positive fully sensitive pulmonary disease who remained smear and culture positive for at least 3 months after treatment was started. The hospital where he was treated had no community tuberculosis services. Members of the cluster included another homeless man, a psychiatrically ill man, and another alcoholic man known to have remained smear and culture positive for at least 2 months. The cluster also included several patients who would not usually be expected to have

Table 3. *Univariate analysis of risk factors for being in a cluster*

Variable/category	Number in category (% of total)	Clustered (% in category)	Unadjusted odds ratio (95% confidence interval)	Unadjusted *P* value
Age				
0–19	35 (6·3)	9 (25·7)	0·77 (0·32–1·82)	0·10
20–39	279 (50·5)	64 (22·9)	0·66 (0·40–1·09)	
40–59	124 (22·5)	41 (27·9)	1·10 (0·63–1·92)	
60+	103 (18·7)	32 (31·1)	1	
Missing	11 (2·0)	1 (9·1)	0·22 (0·28–1·81)	
Gender				
Male	334 (60·5)	90 (26·9)	1	0·93
Female	213 (38·6)	56 (26·3)	0·97 (0·66–1·43)	
Missing	5 (0·9)	1 (20·0)	0·68 (0·07–6·15)	
Area of residence				
North London	378 (68·5)	93 (24·6)	1	0·008
South London	157 (28·4)	53 (33·8)	1·56 (1·04–2·34)	
Treated London – live elsewhere	17 (3·1)	1 (5·9)	0·19 (0·03–1·46)	
Site/smear				
Pulmonary Smear+	164 (29·7)	58 (35·4)	2·65 (1·62–4·34)	0·003
Pulmonary Smear−	121 (21·9)	34 (28·1)	1·89 (1·10–3·27)	
Pulmonary Unknown	28 (5·1)	8 (28·6)	1·94 (0·79–4·78)	
Non Pulmonary	193 (35·0)	33 (17·1)	1	
Missing	46 (8·3)	14 (30·4)	2·12 (1·02–4·41)	
Unemployed				
Yes	135 (24·5)	48 (35·6)	1·61 (1·03–2·53)	0·017
No	255 (46·2)	65 (25·5)	1	
Missing	162 (29·3)	34 (21·0)	0·78 (0·48–1·24)	
Ethnic group				
Indian Subcontinent	120 (21·7)	22 (18·3)	1	0·004
White	176 (31·9)	63 (35·8)	2·48 (1·42–4·33)	
Black African	112 (20·3)	22 (19·6)	1·09 (0·56–2·10)	
Other	45 (8·2)	14 (31·1)	2·01 (0·92–4·40)	
Missing	99 (17·9)	26 (26·3)	1·58 (0·83–3·01)	
Born UK				
Yes	148 (26·8)	62 (41·9)	3·06 (1·98–4·76)	< 0·0005
No	289 (52·4)	55 (19·0)	1	
Missing	115 (20·8)	30 (26·1)	1·50 (0·90–2·50)	
Previous TB				
Yes	42 (7·6)	17 (40·5)	2·04 (1·06–3·94)	0·11
No	372 (67·4)	93 (25·0)	1	
Missing	138 (25·0)	37 (26·8)	1·10 (0·71–1·71)	
Resistant to any drug				
Yes	108 (19·6)	24 (22·2)	1	0·24
No	444 (80·4)	123 (27·7)	1·34 (0·81–2·21)	
HIV				
Yes	43 (7·8)	6 (14·0)	0·42 (0·18–1·03)	0·037
No	509 (92·2)	141 (27·7)	1	
Homeless				
Yes	23 (4·2)	9 (39·1)	1·82 (0·77–4·30)	0·18
No	529 (95·8)	138 (26·1)	1	
Alcoholic				
Yes	34 (6·2)	15 (44·1)	2·31 (1·14–4·67)	0·023
No	518 (93·8)	132 (25·5)	1	

Table 4. *Risk factors for clustering. Final multivariate logistic regression model*

Variable/category	Adjusted OR (95% CI)	Adjusted *P* value*
Age		
0–19	0·81 (0·25–2·67)	0·73
20–39	0·53 (0·22–1·25)	0·14
40–59	1·03 (0·41–2·57)	0·94
60+	1	N/A
Site/smear		
Non-pulmonary	1	N/A
Pulmonary Smear+	2·13 (1·13–4·01)	0·02
Pulmonary Smear−	1·75 (0·87–3·53)	0·12
Pulmonary Smear Unknown	2·16 (0·55–8·50)	0·27
Area of residence		
North London	1	N/A
South London	1·87 (1·08–3·22)	0·02
Treated in London live elsewhere	0·45 (0·04–4·76)	0·51
Unemployed		
Yes	1·50 (0·89–2·55)	0·13
No	1	N/A
Born UK		
Yes	2·64 (1·55–2·49)	0·0003
No	1	N/A
HIV		
Yes	0·21 (0·04–1·01)	0·051
No	1	N/A

* Likelihood Ratio Test *P* value for each category of variable compared to baseline.

prolonged contact with the homeless (e.g. office workers, a public utility worker and students). In a cluster of 7 patients the apparent index case was a homeless man and in a cluster of 6 patients 2 of the early cases were homeless.

DISCUSSION

This study suggests that approximately 19% of tuberculosis in inner London in 1993 was due to recent local transmission rather than reactivation of old infection or importation from elsewhere. This shows that recent local transmission makes an important contribution to levels of tuberculosis in the city. The 19% figure is higher than the traditionally assumed 10% of cases due to recent transmission in developed countries [7] but lower than has been observed in San Francisco [9] and New York [10] (31% and 26–27% respectively).

A number of factors that may have biased the overall estimate of clustering need to be considered [22]. These include the typing techniques used to define clustering, the completeness of the study sample and the duration of the study.

The underlying assumption in tuberculosis typing studies is that isolates sharing the same typing patterns are part of a recent chain of transmission. Conversely, isolates with unique patterns are assumed to be either due to re-activation of infection acquired in the past or due to recent transmission from a patient outside the study sample. In order for these assumptions to be supported the typing techniques must be able to discriminate reliably between very many different isolates but also be sufficiently stable not to change when tuberculosis is transmitted from patient to patient. No typing system is perfect as there is always a trade-off between discriminatory power and the stability of the patterns.

IS6110 combined with a secondary technique such as PGRS or DR typing is currently the accepted 'gold standard' [19]. Although small studies employing PCR-based techniques have been described we wished to determine the role of these techniques in a complete population (inner London) for which cultures were available although non-viable. Our study was retro-

spective and used archived material. We combined spoligotyping which provides a relatively stable but less discriminatory fingerprint [12] with HIPCR typing which is more discriminating but less stable and reproducible than standard IS6110 [13]. By combining these two PCR-based techniques (which are based on different genetic markers) we expected to achieve a high degree of discrimination. Although HIPCR typing has been criticized due to its limited reproducibility (producing different results on the same isolates when used in different laboratories or at different times) [19], we limited the effect of this by using it as secondary typing technique. Relatively small groups of isolates (that were indistinguishable using spoligotyping) were typed using HIPCR at the same time, in the same laboratory and with the resulting amplification products run on the same gel.

The PCR based techniques were able to produce typing results on 99% of isolates even though only 11% of these were viable for subculture. The success of this approach is supported by the results of the comparison with standard IS6110 typing in 64 isolates and by the ability of the techniques to identify household clusters. Comparison with IS6110 demonstrated a similar level of discrimination as might be expected in studies that use IS6110 with secondary typing for isolates with low IS6110 copy numbers. However, this analysis is based on small numbers and from what we know of the discriminatory ability of spoligotyping [13] and HIPCR [19] it is likely that the estimate of clustering produced is somewhat higher than that which would have been produced by ISS110 combined with a discriminatory secondary typing technique. Fourteen of the 553 patients in the study sample were known to have had household contact. The typing (which was performed blind to patient details) successfully 'picked out' 10 of these (five pairs). In one pair sisters had different HIPCR results, suggesting limitations to HIPCR typing or that there had been no transmission between the sisters. The technique also showed that in an apparent household cluster of tuberculosis in a husband and wife pair only the wife really had tuberculosis and the husband's positive isolate was due to previously unsuspected laboratory cross-contamination. The husband's sample had been submitted to the laboratory because he had a cough identified during contact tracing. Although he had no other tuberculosis symptoms, he was treated for tuberculosis. Thus although the results are not directly comparable to results produced by IS6110 we believe that they are sufficiently robust to allow inferences to be made about the transmission of tuberculosis.

We typed 81% of all the *culture-confirmed* cases from inner London in 1993. However, because only 68% of *notified* patients in inner London were culture-confirmed this represented only 56% of all notifications. Although it is known that estimates of clustering are likely to increase as the completeness of the sample increases [23], the effect of excluding a large number of patients who were not culture-confirmed and were thus probably not infectious is difficult to predict. The study was confined to inner London. Notification rates are lower in outer London than in inner London (23·2 *vs.* 35·5 per 100000 per year) [24] but it is not known whether transmission rates are also lower. Without this information it is not possible to predict whether inclusion of outer London cases would have raised or lowered the estimate of clustering. The study took place over 1 year. It is likely that a higher estimate of the level of clustering would have been obtained if the study had continued for longer [6].

The ability to examine risk factors for clustering depends on the availability of accurate and complete epidemiological data. This study used isolates from 1993 so that epidemiological information derived from a previous rigorous survey could be included [21]. Use of this data supplemented by additional enquiries and matching to the PHLS HIV and AIDS registers meant that epidemiological data on subjects was highly complete for most variables and was collected without knowledge of typing results, thus helping to avoid information bias.

The finding that clustering is higher in those with pulmonary smear positive tuberculosis is to be expected since it is known that these patients are the most infectious. Higher levels of clustering in pulmonary smear negative cases than in non-pulmonary cases also supports the view that smear negative patients can sometimes act as sources of transmission. This reinforces the study by Behr et al. which shows transmission from pulmonary smear-negative patients [25].

Lower levels of clustering in the foreign born have been noted in other typing studies [20, 21]. This does not necessarily mean that there is less recent transmission in this group. It could be that they have disease due to recent transmission but that the source case lives abroad and thus is not in the study sample [22]. Also, high incidence rates in the foreign-born mean that even though only a low proportion of cases

are due to recent local transmission the population transmission rate may be high. The fact that 42% of isolates in the UK-born population were clustered suggests that recent local transmission of tuberculosis makes a very substantial contribution to levels of tuberculosis in this group. This may explain the rising tuberculosis rates seen in London in the white ethnic group and in patients of Indian Subcontinent ethnicity who are born in the United Kingdom [2]. Early identification of cases and effective treatment are central to tuberculosis control because they ensure that patients remain infectious for the minimum possible time. High levels of tuberculosis transmission therefore suggest failures in control.

The higher levels of clustering in patients living in south London was not accounted for by controlling for other factors. We were unable to control adequately for social deprivation as information on unemployment or other indicators of social deprivation was incomplete. Inner London areas in south London tend to have higher levels of social deprivation than north London. It could be that this accounts for higher levels of tuberculosis transmission. Social deprivation may be related to transmission through factors such as overcrowding or through poorer access to services. Another possible explanation could be the extent to which services are able to ensure that patients complete treatment. An audit in one area of south London around the time of the study found that at least 19% of patients were lost to follow up before treatment was completed (personal communication, Dr A. Pearson, Nosocomial Infection Surveillance Unit, Central Public Health Laboratory, London). Patients who are lost to follow up could remain infectious for longer periods and thus act as potent sources of transmission. Data on treatment completion are not routinely collected in most areas of London [1], so it is uncertain whether these levels of loss to follow up are higher than in other parts of London.

The importance of ensuring that patients adhere to treatment is also emphasized by the fact that the apparent source case in one of the largest clusters was a homeless man who remained smear and culture positive for several months after starting treatment despite having fully sensitive disease (implying inadequate treatment). Patients need adequate support to ensure that they adhere to treatment. It was noticeable that homeless patients were often involved in large clusters. Homeless patients are likely to find adherence more difficult than those with a more settled lifestyle. Direct observation of treatment is likely to be needed to ensure these patients complete treatment.

The results suggest that tuberculosis transmission in patients with HIV infection was not a major problem in inner London at the time of study. In fact we found a weak association between HIV and clustering which was in the opposite direction to what we had expected. However, this did not reach significance at the $P = 0{\cdot}05$ level and should be treated with caution, especially in view of the use of multiple significance testing in this study. Increased clustering among HIV-infected tuberculosis patients has been found in some studies [7, 8] but not in others [6, 9]. High levels of clustering in HIV infected patients are likely to reflect problems with nosocomial transmission of disease [10]. We identified two HIV infected patients who may have acquired their disease through nosocomial transmission. Two other outbreaks of multidrug resistant disease in London hospitals, both of which occurred outside the study period, have since been reported [25, 26], emphasizing the need for adequate hospital infection procedures.

As in San Francisco no association between drug resistance and clustering was found [9]. In New York where rates of drug resistance were much higher there was a marked association which was largely due to hospital transmission [10]. This study does not suggest that hospital transmission was a major driving force in the spread of drug resistant disease in London in 1993, but the recent nosocomial outbreaks of MDRTB described above illustrate how this finding should not lead to complacency [25, 26].

The fact that definite epidemiological links were only found in 7% of clustered patients needs to be interpreted in light of the fact that epidemiological investigation was limited to a review of what was recorded in patient records. Nevertheless, the fact that other cluster members were very rarely mentioned in contact tracing notes suggests that unrecognized or casual contact may account for the majority of tuberculosis transmission in inner London. This is probably because each infectious case will have had the opportunity to expose large numbers of patients to a low level of risk through very minor contact but will only have had prolonged contact (associated with a higher level of risk) with relatively few people.

The study demonstrates that recent local transmission makes an important contribution to tuberculosis in inner London. Measures aimed at decreasing transmission (such as ensuring that all

patients are diagnosed early and treated effectively) are likely to have an important impact on disease rates. Such measures are likely to be particularly important in groups such as the homeless. Services should routinely monitor treatment completion rates so that action can be taken if they are found to be low.

ACKNOWLEDGEMENTS

We thank J. Van Embden, RIVM Bilthoven Netherlands for the generous provision of materials for spoligotyping; R. George, Laboratory of Respiratory and Enteric Pathogens Central Public Health Laboratories, London for help in design and conduct of the study; H. Maguire, Regional Epidemiologist, CDSC, Thames Unit and D. O'Sullivan, CCDC Lambeth Southwark and Lewisham for their input into the advisory group; D. Kumar and others involved in the National Survey of Tuberculosis Notifications for provision of epidemiological data; M. Kozak for assistance in data collection; S. Wilson, Mycobacterium Reference Unit for helpful advice; A. Middleton and M. Chadwick, Royal Brompton Hospital for technical assistance; physicians, laboratory directors, hospital control of infection officers, tuberculosis nurses and consultants in communicable disease control throughout London for provision of information for the study.

REFERENCES

1. Hayward AC, Coker RJ. Could a tuberculosis epidemic occur in London as it did in New York? Emerg Infect Dis 2000; **6**: 12–6.
2. Ormerod LP, Charlett A, Gilham C, Darbyshire JH, Watson JM. Geographical distribution of tuberculosis in national surveys of England and Wales in 1988 and 1993: report of the Public Health Laboratory Service/British Thoracic Society/Department of Health Collaborative Group. Thorax 1998; **53**: 176–81.
3. The Interdepartmental Working Group on Tuberculosis. The prevention and control of tuberculosis in the United Kingdom: recommendations for the Prevention and Control of Tuberculosis at Local Level. London: HMSO, 1996.
4. Hannan MM, Azadian BS, Gazzard BG, Hawkins DA, Hoffman PN. Hospital infection control in an era of HIV infection and multi-drug resistant tuberculosis. J Hosp Infect 2000; **44**: 5–11.
5. van Embden JDA, Cave MD, Crawford JT, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardised methodology. J Clin Microbiol 1993; **31**: 406–9.
6. van Soolingen D, Martien W, Borgdorff, et al. Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. J Infect Dis 1999; **180**: 726–73.
7. Safi H, Aznar J, Palomares JC. Molecular epidemiology of *Mycobacterium tuberculosis* strains isolated during a 3-year period (1993 to 1995) in Seville, Spain. J Clin Micro 1997; **35**: 2472–6.
8. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City. N Engl J Med 1994; **330**: 1710–6.
9. Small PM, Hopewell PC, Singh SP, et al. The epidemiology of tuberculosis in San Francisco. N Engl J Med 1994; **330**: 1702–9.
10. Frieden TR, Woodley CL, Crawford JT, Lew D, Dooley SL. The molecular epidemiology of tuberculosis in New York City: the importance of nosocomial transmission and laboratory error. Tuber Lung Dis 1996; **77**: 407–13.
11. Kamerbeek J, Schouls L, Kelk A, et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 1997; **35**: 907–14.
12. Goyal M, Saunders NA, van Embden JD, Young DB, Shaw RJ. Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS6110 restriction fragment length polymorphism. J Clin Microbiol 1997; **35**: 647–51.
13. Patel S, Wall S, Saunders NA. Heminested inverse PCR for IS6110 fingerprinting of *Mycobacterium tuberculosis* strains. J Clin Microbiol 1996; **34**: 1686–90.
14. Plikaytis BB, Plikaytis BD, Yakrus MA, et al. Differentiation of slow growing *Mycobacterium* species including *Mycobacterium tuberculosis*, by gene amplification and restriction fragment length polymorphism analysis. J Clin Microbiol 1992; **30**: 1815–22.
15. Plikaytis BB, Crawford JT, Woodley CL, et al. Rapid amplification-based fingerprinting of *Mycobacterium tuberculosis*. J Gen Microbiol 1993; **139**: 1537–42.
16. Friedman CR, Stoeckle MY, Johnson WD, Riley LW. Double repetitive-element PCR method for subtyping *Mycobacterium tuberculosis* clinical isolates. J Clin Microbiol 1995; **33**: 1383–4.
17. Linton CJ, Jalal H, Leeming JP, Miller MR. Rapid discrimination of *Mycobacterium tuberculosis* strains by random amplified polymorphic DNA analysis. J Clin Microbiol 1994; **32**: 2169–74.
18. Neimark H, Ali Baig M, Carleton S. Direct identification and typing of *Mycobacterium tuberculosis* by PCR. J Clin Microbiol 1996; **34**: 2454–9.
19. Kremer K, Van Soolingen D, Frothingham R, et al. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol. 1999; **37**: 2607–18.
20. Heifets LB, Good RB. Current laboratory methods for the diagnosis of tuberculosis. In: Bloom BR, ed. Tuberculosis protection, pathogenesis, prevention and control. Washington DC: American Society for Microbiology, 1994: 85–110.

21. Kumar D, Watson JM, Charlett A, Nicholas S, Darbyshire JH. Tuberculosis in England and Wales in 1993: results of a national survey. Thorax 1997; **52**: 1060–7.
22. Glynn JR, Bauer J, de Boer AS, Borgdorff MW, Fine PE, Godfrey-Faussett P, Vynnycky E. Interpreting DNA fingerprint clusters of *Mycobacterium tuberculosis*. European Concerted Action on Molecular Epidemiology and Control of Tuberculosis. Int J Tuberc Lung Dis 1999; **3**: 1055–60.
23. Glynn JR, Vynnycky 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.
24. Pearson AD, Hamilton G, Healing TD, Maguire H, Feltham RKA. The surveillance, prevention and control of tuberculosis in London. The Report of the Working Party on Tuberculosis of the London Group of Consultants in Communicable Disease Control, 1995.
25. Behr MA, Warren SA, Salamon H, et al. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *Lancet* 1999; **353**: 444–9.
26. Anonymous. Outbreak of hospital acquired multidrug-resistant tuberculosis. C D R 1994; **4**: 1.
27. Breathnach AS, de Ruiter A, Holdsworth GM, et al. An outbreak of multi-drug-resistant tuberculosis in a London teaching hospital. *J Hosp Infect* 1998; **39**: 111–7.