IJE vol.33 no.2 © International Epidemiological Association 2004; all rights reserved. *International Journal of Epidemiology* 2004;**33**:351–357
DOI: 10.1093/ije/dyh021

Transmission of tuberculosis in a high incidence urban community in South Africa

Suzanne Verver,[1,3] Robin M Warren,[2] Zahn Munch,[1] Emilia Vynnycky,[4] Paul D van Helden,[2] Madalene Richardson,[2] Gian D van der Spuy,[2] Donald A Enarson,[5] Martien W Borgdorff,[3] Marcel A Behr[6] and Nulda Beyers[1]

Background The objective of this study was to identify risk factors for ongoing community transmission of tuberculosis (TB) in two densely populated urban communities with a high incidence rate of TB in Cape Town, South Africa.

Methods Between 1993 and 1998 DNA fingerprints of mycobacterial isolates from TB patients were determined by restriction fragment length polymorphism (RFLP). Cases whose isolates shared identical fingerprint patterns were considered to belong to the same cluster and to be attributable to ongoing community transmission.

Results The average annual notification rate of new smear positive TB was 238/100 000. In all, 1023/1526 reported patients were culture positive, and RFLP was available for 768 (75%) of the isolates from these patients. Since some patients experienced more than one infection during the study period, 797 cases were included in the analysis. Of the cases, 575/797 (72%) were clustered. Smear-positive cases and those who were retreated after default were more likely to be clustered than smear-negative and new cases, respectively. Patients from Uitsig were more often part of large clusters than were patients from Ravensmead. Age, sex, year of diagnosis, and outcome of disease were not risk factors for clustering, nor for being the first case in a cluster, although various analytical approaches were used.

Conclusions The incidence and proportion of cases that are clustered in this area are higher than reported elsewhere. An overwhelming majority of TB cases in this area is attributed to ongoing community transmission, and only very few to reactivation. This may explain the lack of demographic risk factors for clustering.

Keywords Tuberculosis, transmission, urban

The molecular epidemiology of *Mycobacterium tuberculosis* is most frequently based on restriction fragment length polymorphism (RFLP) banding patterns generated by Southern hybridization with the probe IS*6110*.[1,2] In low tuberculosis (TB) incidence countries, isolates from cases who are identified during outbreaks of TB usually share identical IS*6110* banding patterns. These so-called 'DNA fingerprint clusters' are therefore assumed to represent recent or ongoing transmission. Conversely, isolates from elderly patients and immigrant patients usually have unique (non-identical) banding patterns.[2,3] Such cases are therefore assumed to be attributed to reactivation or transmission outside the study period or area. However, theoretically, cases may also be part of a cluster if they are experiencing disease as a result of reactivation with an endemic strain. Conversely, cases attributable to recent transmission may not be in a cluster

[1] Department of Paediatrics and Child Health, Stellenbosch University, South Africa.

[2] MRC Centre for Molecular and Cellular Biology, Department of Medical Biochemistry, Stellenbosch University, South Africa.

[3] KNCV Tuberculosis Foundation The Hague, The Netherlands.

[4] Infectious Disease Epidemiology Unit, Department of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, UK.

[5] International Union against Tuberculosis and Lung Diseases (IUATLD), Paris, France.

[6] Division of Infectious Diseases, Department of Medicine, McGill University Health Centre, Montreal, Quebec, Canada.

Correspondence: Suzanne Verver, KNCV Tuberculosis Foundation, PO Box 146, 2501 CC The Hague, The Netherlands. E-mail: ververs@kncvtbc.nl

if the fingerprint pattern of the infecting strain is endemic but is undergoing evolution, or if the source case is not in the study database.

RFLP patterns have been used to identify risk factors for transmission of TB in low incidence areas. Such risk factors include smear positivity, young age, male gender, homelessness, urban residence, duration of residency, human immunodeficiency virus (HIV) positivity, and belonging to certain nationalities and ethnic groups.[2–7]

Few studies in high TB incidence areas have linked epidemiological and RFLP data.[8,9] In Africa, RFLP studies have been small (limited to less than 400 patients per study) and only a few have assessed risk factors for clustering.[10–13] Very few risk factors have typically been found (e.g. older age,[10] treatment failure,[11] and prior imprisonment[12]) perhaps as a result of small numbers of samples. We have therefore assessed risk factors for transmission in a high incidence urban community, using data from 760 patients.

We analysed data from Ravensmead and Uitsig, which are two adjacent high incidence urban communities in Cape Town, South Africa. Several studies have been published on this area, but none has reported the relationship between RFLP clustering and clinical or demographic parameters.[14–20] It was assumed that clustering of RFLP patterns reflects ongoing community transmission. We will discuss the validity of this assumption for a high incidence area. Epidemiological confirmation of clustered samples will be of limited value in such a high incidence area, since many contacts leading to infection may be casual.[15,21]

Methods

Setting

Ravensmead and Uitsig (R/U) have a total population of 38 656 (Statistics South Africa, census 1996), and span an area of 3.4 km^2. Residents of the area are of low socioeconomic status (average household income Rand 6700 (= US$ 700); unemployment rate 70%)(Statistics South Africa, census 1996). Migrant workers make up a small proportion of the population, although there is a high level of mobility both within the community and to and from neighbouring areas. A relatively low proportion of the population is infected with HIV. HIV seroprevalence in women attending antenatal clinics in the Western Cape Province rose from 1.2% to 5.2% from 1994 to 1998 (Directorate of Health Systems Research and Epidemiology, Western Cape Department of Health, 1999). The study area has an extremely high notified disease rate of TB (761/100 000 per year for all forms of TB; 238/100 000 per year for new smear positive disease).

Study population

All bacteriologically confirmed TB patients were eligible for this study if they reported to the clinics in R/U between 1993 and 1998 and were resident at any time during this period in R/U. Informed consent for use of clinical data was obtained from patients or their parents or guardians. We included patients with Ziehl-Neelsen smear positive and smear negative pulmonary as well as extra-pulmonary TB, and those with and without a history of previous treatment (designated 'retreatment' and 'new' TB cases, respectively).[22] Retreatment cases were further stratified according to their previous episode of TB, namely into (1) relapses (those who were found to have TB after having been successfully treated), (2) treatment failures (those who, while on treatment, remained or became smear positive $\geqslant$5 months after starting treatment) or (3) return to treatment after default (those who interrupted treatment for a minimum of 2 consecutive months).[23] In this study we did not distinguish between drug-resistant and drug-sensitive cases since drug susceptibility testing was not carried out systematically. In 1993 in the Western Cape Province, approximately 3.2% of new patients and 8.6% of retreatment patients were resistant to any drug. The corresponding proportions for multi-drug-resistant TB were 1.1% and 4.0% respectively.[24]

In South Africa case detection is passive (patients present themselves with complaints). Before 1996, patients were diagnosed by culture of *Mycobacterium tuberculosis* and/or chest radiography. WHO/International Union Against Tuberculosis and Lung Disease (IUATLD) directly observed treatment, short course (DOTS) guidelines have been followed since 1996 with case finding being based mainly on sputum smear. However, as part of a research project, an attempt is made to obtain sputum for culture from all suspected TB patients in the research area. The Ethics Committee of Stellenbosch University has approved the project.

M. tuberculosis isolates

DNA fingerprints of mycobacterial isolates of bacteriologically confirmed TB patients were determined by RFLP, in a standardized manner.[1,2] The probe IS*6110* was used, and strains harbouring $<$5 IS*6110* copies were subjected to subtyping with MTB*484(1)*.[18] Results were analysed with Gelcompar software. Isolates with identical fingerprint patterns based on IS*6110* and MTB*484(1)* were assigned to the same cluster and assumed to represent ongoing community transmission. Cases were considered to be experiencing disease attributable to re-infection or dual infection if serial isolates from the same disease episode had different IS*6110* patterns and if laboratory error and strain evolution were unlikely. Strain evolution was arbitrarily defined as the occurrence of $<$5 band changes (band shifts, addition or loss of a band) in the banding pattern of isolates with $>$5 bands.[25,26] When evolution within a patient was likely (i.e. the evolved pattern had not been previously identified in the community), only the first pattern was used in the analysis. Both isolates from patients with a re-infection or dual infection were used to determine clustering. For simplicity, these two strains from the same patient will be referred to as two cases from this point. Isolates from retreatment patients were included only when there was no isolate from the first episode, or when the case was experiencing re-infection or dual infection.

Clustering

Clustering was calculated using all strains identified between 1993 and 1998. The proportion of cases due to ongoing transmission was estimated allowing one source case per cluster (i.e. n-1 method).[7] In a sub-analysis we presume that the first case detected in a cluster is the source case of the cluster.[3,4] For that sub-analysis we excluded all cases that occurred during the first study year and clusters that had at least one case observed during the first study year (1993), in order to reduce misclassification of source cases. In the sub-analysis clustered

cases were included only if they were diagnosed within 2 years of their source case, in order to ensure a uniform follow-up period after each source case.

We assessed the influence of study period, sampling in space, and random sampling on clustering percentages.[27] The influence of study period was assessed by calculating the above statistics using time windows of 2 consecutive years, i.e. 1993/94, 94/95, 95/96, etc. The effect of sampling in space was assessed by comparing the proportion of clustered cases within the Ravensmead area with that from within the Uitsig area. Random sampling was carried out by taking 20 random samples each of a varying proportion (i.e. 25%, 30%, 35%, etc … 95%) of the cases.

Risk factors

The following risk factors were assessed for clustering: year of diagnosis, age, sex, smear positivity, (re)treatment status, and outcome of treatment. The risk factor 'retreatment patient' represents the outcome of the previous episode. To assess whether treatment outcome was a risk factor for clustering, defaulters (cases whose treatment was interrupted for a minimum of 2 consecutive months), failures, deaths, and transfers were compared with cases who had been successfully treated (i.e. WHO categories cured or treatment completed). All these risk factors were determined by considering isolates from the first disease episode which had a (new) RFLP pattern. Risk factors were analysed (1) for clustering and (2) for being in a large cluster, arbitrarily chosen as ⩾4 cases. In the sub-analysis, risk factors were analysed (3) for clustering, and (4) for being the first diagnosed case in a cluster as compared with being unique. In the latter method the secondary cases (i.e. those who were in a cluster but not the first diagnosed) were excluded from the analysis. All variables that were significant in univariate analysis were entered in a stepwise forward logistic regression model in SPSS, to adjust for confounding.

Results

Study population

During the period 1993–1998, 1526 tuberculosis patients were reported, of whom 1103 were classified as being new cases. In all, 1094/1526 (72%) patients had bacteriologically confirmed (i.e. smear and/or culture positive) disease, of whom 1023 were culture positive. Of these 1023 cases 68% were smear positive. The average annual notification rate of new bacteriologically confirmed TB (culture and/or smear-positive) was 313/100 000, and of new smear positive TB was 238/100 000 (1993–1998). RFLP patterns were available for 768/1023 (75%) culture positive patients, of whom 328 (43%) had at least 2 isolates. As reported elsewhere, laboratory error rate was 3%.[16] After exclusion of samples which were probably involved in laboratory error, 797 cases were identified among 760 patients (Figure 1).[16] The majority of patients had pulmonary TB, while 19/760 patients had only extra-pulmonary TB.

Representativeness of included patients

Cases for whom the RFLP pattern of their isolates were known did not differ with respect to sex, smear positivity, patient category, and outcome from those for whom the RFLP patterns of their isolates was not known. In multivariate analysis, RFLP results were significantly less likely to be available from culture-positive patients aged <15 years (OR = 0.33, 95% CI: 0.18,

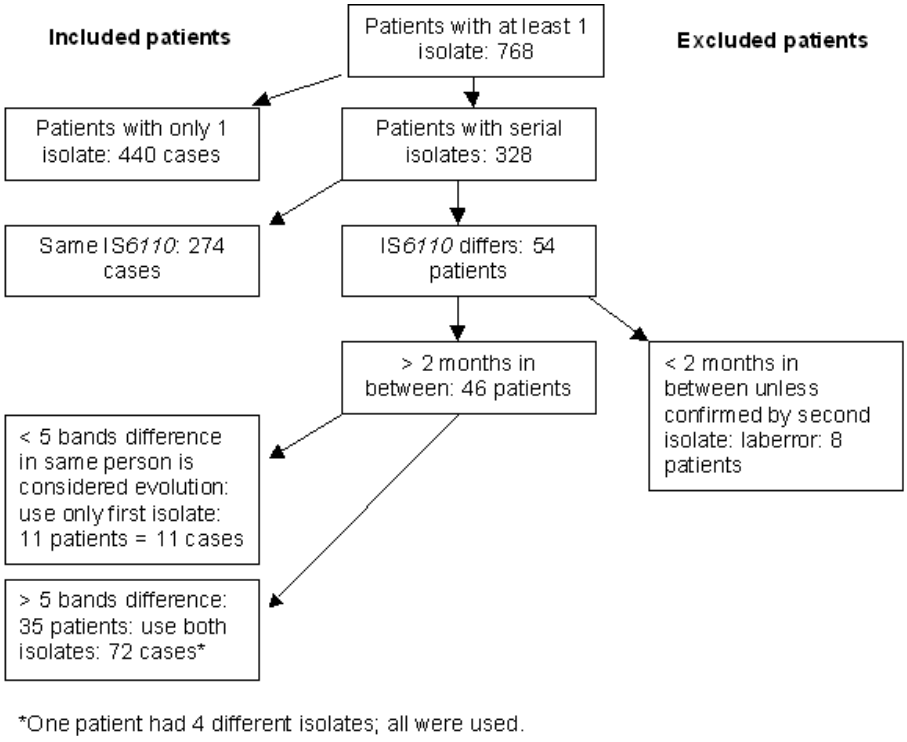

Figure 1 Selection of 797 cases of 760 patients from single and serial isolates from 768 patients

0.56) and from Uitsig (OR = 0.64, 95% CI: 0.48, 0.86), than from culture positive cases aged >15 years and from Ravensmead respectively. Patients with serial isolates were more likely to be smear-positive (OR = 1.99, 95% CI: 1.40, 2.83), to have defaulted from treatment (OR = 2.60, 95% CI: 1.40, 4.85) or to have been retreated after default (OR = 2.48, 95% CI: 1.64, 3.76), than patients with only one isolate.

Clustering

Overall 575/797 cases (72%) were clustered into 115 clusters with cluster sizes ranging from 2 to 40 cases. In total, 392 RFLP patterns were identified. Overall, 396/797 (50%) cases were in clusters comprising ⩾4 cases. The proportion of cases attributable to ongoing transmission (allowing one source case per cluster, i.e. n-1 method) is estimated to be (575–115)/797 = 58%. The clustering proportion decreased with decreasing sample size. When extrapolating the increase in clustering with the increase in sampling fraction to all 1023 culture-positive patients, the proportion of cases clustered would be expected to increase to 76%. When we took random samples of 75% of the cases (i.e. 598), this resulted in 68% of cases being clustered (range: 66–69%). The proportion of cases clustered was reduced both by sampling 2-year study periods (range: 55–64%) and by sampling only one of the suburbs (67% in either one of the two).

Risk factors

Clustering was not associated with sex, year of diagnosis, or suburb of residence (Table 1). Smear-positive cases were more likely to be in clusters than smear-negative cases. Cases who

Table 1 Risk factors for clustering (n = 797)

Risk factor (95% CI)	No. in clusters/ Total no. (%)	Crude odds ratio (95% CI)	Adjusted ratio (95% CI)
Age group			
0–14	25/33 (76)	1.24 (0.54, 2.87)	
15–24	89/124 (79)	1.50 (0.91, 2.48)	
25–34	204/285 (72)	1	
35–44	131/180 (73)	1.06 (0.70, 1.61)	
45–54	73/106 (69)	0.88 (0.54, 1.43)	
⩾55	44/67 (66)	0.76 (0.43, 1.34)	
Unknown	0/2	χ^2 for linear trend = 3.68 (P = 0.055)	
Sex			
Male	331/457 (72)	1	
Female	95/244 (72)	0.98 (0.71, 1.34)	
Unknown	0/1		
Year			
1993	112/150 (75)	1	
1994	128/173 (74)	0.97 (0.59, 1.59)	
1995	82/115 (71)	0.84 (0.49, 1.46)	
1996	87/117 (74)	0.98 (0.57, 1.71)	
1997	69/102 68)	0.71 (0.41, 1.24)	
1998	97/140 (69)	0.77 (0.46, 1.28)	
		χ^2 for linear trend = 1.74 (P = 0.19)	
Suburb			
Ravensmead	357/500 (71)	1	
Uitsig	202/274 (74)	1.12 (0.81, 1.57)	
Other (having another episode in R/U)	16/23 (70)		
Smear			
Negative	132/200 (66)	1	1
Positive	420/562 (75)	1.52 (1.07, 2.16)[a]	1.48 (1.03, 2.13)[a]
Unknown	23/35 (66)		
Patient category			
New	354/500 (71)	1	1
Retreatment after successful treatment	151/207 (73)	1.12 (0.78, 1.61)	1.12 (0.77, 1.63)
Retreatment after default	47/55 (85)	2.44 (1.13, 5.30)[a]	2.36 (1.08, 5.13)[a]
Retreatment after failure	2/4 (50)	0.42 (0.06, 2.98)	0.50 (0.07, 3.64)
Unknown	21/31 (68)	0.87 (0.40, 1.90)	
Outcome			
Successful treatment	384/522 (74)	1	
Default (incl unknown & not treated)	114/159 (72)	0.91 (0.61, 1.35)	
Failure	9/13 (69)	0.81 (0.25, 2.67)	
Transfer	40/65 (62)	0.58 (0.34, 0.98)[a]	
Dead	28/38 (74)	1.01 (0.48, 2.13)	
Total	575/797 (72)		

[a] $P < 0.05$.

had been retreated after default were more likely to be in a cluster than new cases. Cases whose treatment outcome was unknown because they were transferred were less likely to be in a cluster than cases who were successfully treated. The proportion of cases who were in a cluster seemed to decrease with increasing age, but this decrease was not significant (Table 1). In the multivariate analysis, only smear positivity (OR = 1.48, 95% CI: 1.03, 2.13) and retreatment after default (OR = 2.36, 95% CI: 1.08, 5.13) were significant risk factors for being in a cluster. The only risk factors for being part of a larger cluster (size $\geqslant$4) was living in Uitsig (as compared with living in Ravensmead, OR = 1.41, 95% CI: 1.05, 1.89).

Sub-analysis

In the sub-analysis where we excluded all cases in the first study year and clusters in which at least one case had onset during the first study year and excluded clustered cases that were diagnosed >2 years after their source case, 259 cases remained for analysis. Of these, 101 (45%) were clustered in 40 clusters of size 2–5. The only risk factor for being in a cluster that was found with this method was being a retreatment patient after default (OR = 4.10, 95% CI: 1.07, 15.8). No risk factors were found for being the first case in a cluster.

Discussion

We found that 72% of cases were clustered and 58% attributable to ongoing community transmission. The only risk factors for clustering which were identified in this population were smear positivity and being a retreatment patient after default. Patients from Uitsig were more likely to be in large clusters than patients from Ravensmead. Age, sex, year of diagnosis, and outcome of disease were not risk factors for clustering, irrespective of the analytical approach which was used. No risk factors were found for being the first case in a cluster.

The extent of clustering was similar to that observed in South African gold mines, and is higher than that reported from anywhere else to date.[17] Even higher levels of clustering might have been expected, based on a model for an area with a constant annual rate of tuberculosis infection (ARTI) of 3% and European disease risk estimates.[28] The ARI in Ravensmead and Uitsig (3.5% in 1998, B Fourie, N Beyers, unpublished data) is comparable to that used in the model. In such an area almost all cases are predicted to be due to recent infection or recent re-infection and 85% of cases may be clustered. This difference between the expected proportion of clustering and that observed in our study population might be due to a variety of factors including (1) inappropriate assumptions about the disease risks in the model, which might be higher in R/U than those estimated for European settings and (2) sampling and selection of cases, including importation of strains by population movement and infection outside the community, which generally leads to underestimation of clustering,[27,29–31] (3) a faster rate of change of DNA fingerprint patterns than is currently assumed in the model.[25,26,32–34]

In the sub-analysis (after excluding all clusters observed during the first study year and limiting duration of follow-up to 2 years), 45% of cases were clustered. This percentage is lower than that calculated in the main analyses, mainly as a result of smaller numbers, but possibly also because the excluded strains were more likely to be endemic. When we consider the unique cases as being potential source cases, it is estimated that at least one of every five source cases (40/198) leads to at least one secondary case within 2 years within the area.

We investigated sampling bias and found that clustering was affected only to a limited extent by selection in time, space, and random selection. However, sub-sampling within the city (only two suburbs were included) may have been important in view of the mobility of its population. However, this cannot be proved with the available data.

Although we studied risk factors using various analytical approaches, no significant demographic risk factors for clustering could be identified, except that cases from Uitsig were more likely to be in large clusters than cases from Ravensmead. This is possibly due to a slightly higher incidence rate in Uitsig where the population was slightly poorer and more crowded than that in Ravensmead. This result confirms the lack of demographic risk factors found in other studies in high incidence areas where results may have been explained by a lower number of isolates. The lack of demographic risk factors may be due to the fact that most non-clustered cases in the study are also due to ongoing community transmission.

Our finding that in a high-incidence urban community smear-positive cases were more likely to be clustered than were smear-negative cases is consistent with findings in low-incidence areas and one high-incidence mining community.[2,11,35] The extensive clustering found among retreatment cases who had previously defaulted, as compared with new cases might be due to a correspondingly longer duration of infectiousness. This confirms the observation that such cases may be important sources for continuing transmission in high-prevalence communities, such as the Inuit of North America.[36] The low proportion of clustering found among cases that had transferred might be explained by the fact that they are fairly mobile. However, this finding was not significant in the multivariate analysis.

We found that the proportion of cases who were in a cluster seemed to decrease with age, although this decrease was not significant. A decrease in clustering with increasing age was also found in a rural area in Malawi[10] and in low-incidence areas,[2] but this decrease was not seen in other high-incidence areas, where the estimates were based on smaller numbers of cases.[11,12,37] The latter is consistent with model predictions for high-incidence areas where the annual risk of TB infection has not declined over time.[28]

It should be noted that the risk factors for clustering in the main analysis were different from risk factors for being clustered or being the first case in a cluster in the sub-analysis. We expected smear-positivity to be a risk factor for being a source case. The absence of this finding may be due to the presence of source cases outside the study area. When a new cluster within the study area has been caused by a smear-positive common source outside the area, then the first case within the study area may be smear negative.

This study has several limitations. First, bacteriological confirmation was only present for 72% of patients, and of these RFLP was available for the isolates of only 75% of culture-positive patients. Random sampling suggests that this reduced the proportion of clustered cases only to a limited extent. Undersampling has a relatively small effect when cluster sizes are large, as they appear to be in this area.[29]

Second, HIV seropositivity and drug resistance data were not systematically collected in the present study. The number of HIV infected people in the area is relatively low (1.2–5.2% of women attending antenatal clinics), although the proportion of HIV infected people among TB patients may be higher. Other smaller studies have shown no relationship between clustering and HIV.[10–13] Drug-resistant cases would be expected to be more likely to be clustered than drug-sensitive cases due to the fact that they have experienced a longer duration of treatment. This was confirmed in some studies, but the opposite was found in others.[11,38]

This study confirms the importance of detection and treatment of smear-positive cases and completion of treatment. Furthermore, it shows that an overwhelming majority of cases in this high-incidence setting is probably due to recent infection or recent re-infection. This may explain the absence of demographic risk factors for clustering. In high-incidence areas, it may be difficult to identify transmission chains.

Acknowledgements

This study was supported by the GlaxoSmithKline Action TB programme. S Verver, MW Borgdorff and M Behr were supported by the Sequella Global TB Foundation. E Vynnycky was supported by the UK Medical Research Council and the EC Concerted Action Programme on the Molecular Epidemiology of TB. We thank the nursing sisters and patients of the Ravensmead and Uitsig clinics and the Cape Town Unicity, Administration Tygerberg. We thank the TB research team for their valuable assistance.

KEY MESSAGES

- In a high-incidence suburb of Cape Town, South Africa, 72% of cases were clustered and 58% were attributed to ongoing community transmission as opposed to reactivation. These are likely to be underestimates of the true proportions clustered cases and extent of disease attributed to recent transmission.
- Smear positivity and being a retreatment patient after default are risk factors for being involved in ongoing transmission.
- Age, sex, year of diagnosis, and outcome of disease are not risk factors for clustering in this high-incidence area.
- No risk factors were found for being the first case in a cluster.
- The absence of demographic risk factors for clustering may be due to almost all patients having been recently infected or re-infected.

References

[1] Van Embden JDA, Cave MD, Crawford JT *et al*. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;**31:**406–09.

[2] Van Soolingen D, Borgdorff MW, de Haas PEW *et al*. Molecular epidemiology of tuberculosis in The Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis* 1999;**180:**726–36.

[3] Borgdorff MW, Behr MA, Nagelkerke NJ, Hopewell PC, Small PM. Transmission of tuberculosis in San Francisco and its association with immigration and ethnicity. *Int J Tuberc Lung Dis* 2000;**4:**287–94.

[4] Borgdorff MW, Nagelkerke NJD, de Haas PEW, van Soolingen D. Transmission of *Mycobacterium tuberculosis* depending on the age and sex of source cases. *Am J Epidemiol* 2001;**154:**934–43.

[5] Cohn DL, O'Brien RJ. The use of restriction fragment length polymorphism (RFLP) analysis for epidemiological studies of tuberculosis in developing countries. *Int J Tuberc Lung Dis* 1998;**2:**16–26.

[6] Lillebaek T, Andersen AB, Bauer J *et al*. Risk of *Mycobacterium tuberculosis* transmission in a low incidence country due to immigration from high-incidence areas. *J Clin Microbiol* 2001;**39:**855–61.

[7] 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:**1703–09.

[8] Ahn DD, Borgdorff MW, Van LN, van Gorkom T, Kremer K, van Soolingen D. *Mycobacterium tuberculosis* Beijing genotype emerging in Vietnam. *Emerg Infect Dis* 2000;**6:**302–05.

[9] Garcia-Garcia M, Palacios-Martinez M, Ponce-de-Leon A *et al*. The role of core groups in transmitting *Mycobacterium tuberculosis* in a high prevalence community in Southern Mexico. *Int J Tuberc Lung Dis* 2000;**4:**12–17.

[10] Glynn JR, Yates MD, Crampin AC *et al*. Molecular epidemiology of TB in Karonga district, Malawi: preliminary findings. The Hague: Koninklijke Nederlandse Chemische Vereniging (KNCV). *Tuberculosis Surveillance and Research Unit Progress Report*, 2001, pp. 47–60.

[11] Godfrey-Fausset P, Sonnenberg P, Shearer SC *et al*. Tuberculosis control and molecular epidemiology in a South African gold-mining community. *Lancet* 2000;**356:**1066–71.

[12] Lockman S, Sheppard JD, Braden CR *et al*. Molecular and conventional epidemiology of *Mycobacterium tuberculosis* in Botswana: a population-based prospective study of 301 pulmonary tuberculosis patients. *J Clin Microbiol* 2001;**39:**1042–47.

[13] Yang ZH, Mtoni I, Chonde M *et al*. DNA fingerprinting and phenotyping of *Mycobacterium tuberculosis* isolates from human immunodeficiency virus (HIV)-seropositive and HIV-seronegative patients in Tanzania. *J Clin Microbiol* 1995;**33:**1064–69.

[14] Beyers N, Gie RP, Zietsman HL *et al*. The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high incidence community. *S Afr Med J* 1996;**86:**40–44.

[15] Classen CN, Warren R, Richardson M *et al*. Impact of social interactions in the community on the transmission of tuberculosis in a high incidence area. *Thorax* 1999;**54:**136–40.

[16] Van Rie A, Warren R, Richardson M *et al*. Exogenous re-infection as a cause of recurrent tuberculosis after curative treatment. *N Engl J Med* 1999;**341:**1174–79.

[17] Warren RM, Richardson M, Sampson SL *et al*. Molecular evolution of *Mycobacterium tuberculosis*: phylogenetic reconstruction of clonal expansion. *Tuberculosis* 2001;**81:**291–302.

[18] Warren RM, Richardson M, van der Spuy GD *et al*. Genotyping of *Mycobacterium tuberculosis* with additional markers enhances accuracy in epidemiological studies. *J Clin Microbiol* 1996;**34:**2219–24.

[19] Warren RM, Richardson M, van der Spuy GD *et al*. DNA fingerprinting and molecular epidemiology of tuberculosis: use and interpretation in an epidemic setting. *Electrophoresis* 1999;**20:**1807–12.

[20] Munch Z, Van Lill SWP, Booysen CN, Zietsman HL, Enarson DA, Beyers N. Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. *Int J Tuberc Lung Dis* 2003;**7:**271–77.

[21] Behr MA, Hopewell PC, Paz AE, Kawamura ML, Schecter GF, Small PM. Predictive value of contact investigation for identifying recent transmission of *Mycobacterium tuberculosis*. *Am J Respir Crit Care Med* 1998;**158:**465–69.

[22] Enarson DA, Rieder HL, Arnadottir T, Trebucq A. *Management of Tuberculosis. A Guide for Low Income Countries, 5th Edn*. International Union Against Tuberculosis and Lung Disease. St-Just-La-Pendue: Compogravure Impression, Broachage Imprimeriie, 2000, p. 36.

[23] World Health Organization (WHO), International Union Against Tuberculosis and Lung Disease (IUATLD), Koninklijke Nederlandse Chemische Vereniging (KNCV). Revised international definitions in tuberculosis control. *Int J Tuberc Lung Dis* 2001;**5:**213–15.

[24] Weyer K, Groenewald P, Zwarenstein M, Lombard CJ. Tuberculosis drug resistance in the Western Cape. *S Afr Med J* 1995;**85:**499–504.

[25] Warren RM, van der Spuy GD, Richardson M *et al*. Evolution of the IS6110 based RFLP pattern during the transmission of *Mycobacterium tuberculosis. J Clin Microbiol* 2002;**40:**1277–82.

[26] Warren RM, Van der Spuy GD, Richardson M *et al*. Calculation of the stability of the IS*6110* banding pattern in patients with persistent *M. tuberculosis* disease. *J Clin Microbiol* 2002;**40:**1705–08.

[27] Glynn JR, Vynnycky E, Fine PE. Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprint techniques. *Am J Epidemiol* 1999;**149:**366–71.

[28] Vynnycky E, Borgdorff MW, van Soolingen D, Fine PEM. The effect of the magnitude and trend in the annual risk of tuberculous infection on the interpretation of 'clustering' of isolates from tuberculosis cases. *Emerg Infect Dis* 2003;**9:**176–83.

[29] Glynn JR, Bauer J, de Boer AS *et al*. 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.

[30] Murray M, Alland D. Methodological problems in the molecular epidemiology of tuberculosis. *Am J Epidemiol* 2002;**155:**565–71.

[31] Vynnycky E, Nagelkerke N, Borgdorff MW, van Soolingen D, van Embden JDA, Fine PEM. The effect of age and study duration on the relationship between clustering of DNA fingerprint patterns and the proportion of tuberculosis disease attributable to recent transmission. *Epidemiol Infect* 2001;**126:**43–62.

[32] De Boer AS, Borgdorff MW, de Haas PE, Nagelkerke NJ, van Embden JD, van Soolingen D. Analysis of rate of change of IS6110 RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. *J Infect Dis* 1999;**180:**1238–44.

[33] Niemann S, Richter E, Rusch-Gerdes S. Stability of *Mycobacterium tuberculosis* IS6110 restriction fragment length polymorphism patterns and spoligotypes determined by analyzing serial isolates from patients with drug-resistant tuberculosis. *J Clin Microbiol* 1999;**37:**409–12.

[34] Yeh RW, Ponce de Leon A, Agasino CB *et al*. Stability of *Mycobacterium tuberculosis* DNA genotypes. *J Infect Dis* 1998;**177:**1107–11.

[35] Behr MA, Warren SA, Salamon H *et al*. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli. *Lancet* 1999;**353:**444–49.

[36] Enarson DA. Tuberculosis in aboriginals in Canada. *Int J Tuberc Lung Dis* 1998;**2:**S16–22.

[37] Dye C. Williams BG, Espinal MA, Raviglione MC. Erasing the world's slow stain: strategies to beat drug-resistant tuberculosis. *Science* 2002;**295:**2042–46.

[38] Wilkinson D, Pillay M, Crump J, Lombard C, Davies GR, Sturm AW. Molecular epidemiology and transmission dynamics of *Mycobacterium tuberculosis* in rural Africa. *Trop Med Int Health* 1997;**2:**747–53.