

Effect of clinical and socio-economic factors on the rate of clustering of *Mycobacterium tuberculosis* clinical isolates in Elche (Spain)

M. RUIZ¹*, J. F. NAVARRO², J. C. RODRÍGUEZ¹, J. A. LARROSA³ AND G. ROYO¹

¹ Servicio de Microbiología, Hospital General Universitario de Elche, Universidad Miguel Hernández, Elche (Alicante), Spain

² Servicio de Medicina Preventiva, Hospital General Universitario de Elche, Departamento de Salud Pública, Universidad Miguel Hernández, Elche (Alicante), Spain

³ Departamento Geografía Humana, Universidad de Alicante, San Vicente del Raspeig, Alicante, Spain

(Accepted 4 August 2003)

SUMMARY

We studied the association that exists between the epidemiological type clustering of isolates of *Mycobacterium tuberculosis* and certain clinical, epidemiological and socio-economic characteristics of tuberculosis patients in the Elche health district of Spain. A total of 144 patients diagnosed between 1993 and 1999 and whose isolates had been genotyped by IS6110-RFLP in an earlier study were included. Multivariate analysis showed that the independent variables associated with clustering of strain types were: age (1–25 years, OR 2·92, 95% CI 0·83–10·3), a high percentage of infection in the first circle of contacts (OR 2·89, 95% CI 0·96–8·68), urban dwelling (OR 2·12, 95% CI 0·73–6·2), use of bronchoscopy to obtain samples (OR 16·3, 95% CI 2·3–11·5) and working contact with many people (OR 2·81, 95% CI 0·94–8·3). These data contribute to a better understanding of the epidemiology of tuberculosis and improved systems of control.

At present it is estimated that between 25 and 29% of the population of Spain is infected by *Mycobacterium tuberculosis* [1]. Analysis of age curves of tuberculosis patients shows that the situation in this country is similar to that which exists in developing countries, with a predominance of the disease in young people [2]. Tuberculosis is therefore a serious health problem in Spain and particularly among AIDS patients, approximately 40% of whom are infected with *M. tuberculosis* [3]. Recently we described the molecular epidemiology of *M. tuberculosis* in Elche, Spain by applying the IS6110-RFLP technique [4] to the clinical isolates in this region. We found 25 different genotype clusters (two or more isolates with identical RFLP patterns) among 165 isolates from 147 patients. Three clusters were each represented by more than six strains which together accounted for 32% of all

clustered strains. An overall rate of clustering of 52·4% was obtained [5].

Clustering of cases of tuberculosis in a population is conditioned by many factors such as the thoroughness of the microbiological diagnosis, the duration of the study, contagiousness of the index and secondary cases, susceptibility of the contacts, and the efficacy of prophylaxis for the contacts [6]. Socio-economic factors also play a major role and include overcrowded living conditions, attendance at a community centre and health-care facilities, etc. [7]. It is therefore essential to determine the epidemiological characteristics of a population in order to correctly interpret strain genotyping data [8]. To this end we set out to determine the influence of clinical, epidemiological and socio-economic characteristics of our local tuberculosis patient population on clustering of isolates of *M. tuberculosis*.

Of the original 147 tuberculosis patients studied previously [5], three cases with incomplete data were

* Author for correspondence: Dr Montserrat Ruiz García, S. Microbiología, Hospital General Universitario de Elche, Cami de L'Almazara, 11, 03203 Elche (Alicante), Spain.

eliminated, leaving a total of 144 cases available for analysis. All patients were diagnosed between 1993 and 1999 in the Elche health district, Spain, and accounted for 59.3% of all patients with microbiologically confirmed tuberculosis [5]. The following data were recorded: age, sex, profession, place of residence, date of diagnosis, HIV status, parenteral drug user (PDU), alcoholism, pulmonary cavitation, localization of the disease, type of sample and quantification of direct staining of sputum smears. Professions were classified empirically according to whether they involved contact with many people (e.g. teachers and shop assistants), contact with few people (e.g. bricklayers and farmers), or employment in the footwear industry which is a major local industry attracting a significant number of immigrant workers from other parts of Spain [9]. Criteria and grading of contact were in accordance with the CDC Guidelines [10]: first circle (people living under the same roof, co-workers who spend many hours together in a closed room); second circle (contacts on non-consecutive days in closed places); third circle (occasional contacts). All data were obtained prospectively from files of the microbiology laboratory, medical records or personal interviews with patients and contacts.

Statistical analysis was performed using an SPSS package. Univariate analysis of the relationship between clustering and the different co-variables was done to calculate the odds ratio (OR) with the corresponding 95% confidence intervals (CI). Statistical significance was confirmed using Pearson's χ^2 test or Fisher's exact test. An unconditioned logistic regression procedure based on the likelihood ratio statistic was used to introduce the variables in the multivariate analysis. The categorical variables were factorized (dummy variables), and missing values were considered to be another category. In order to avoid a loss of information in cases due to missing values, we recoded these values as an extra category that was also converted as a dummy variable. Missing values were more common among contact study variables, HIV status and profession. Possible confounding variables (sex and age) or factors traditionally related to clusters in other studies (HIV, PDU) were included in the final model, even though they showed no statistical significance. Since the sample was small, we included in the final model predictive variables with a substantial OR whose statistical significance (*P* value) was close to 0.10.

Table 1 shows the frequency of clustering of strains according to clinical and health-care variables.

Patients with HIV (71.4 vs. 49%, OR 2.6), drug abuse (72 vs. 50%, OR 2.57) had a greater rate of clustering of isolates than negative controls; the effect of alcoholism was less dramatic (63.60 vs. 49.5%, OR 1.79). Clustering was also more evident in patients whose direct sputum smears were positive but the highest OR for clustering was shown for patients whose sample was obtained by bronchoscopy (80 vs. 54.3%, OR 3.37). The presence of cavitating tuberculosis and pulmonary involvement was marginally more associated with clusters than controls.

The rate of clustering was greater when the percentage of those infected in the first circle increased and when infection extended to contacts in the third circle studied or when a greater number of families were infected (Table 2). Subjects who had previous contact with tuberculous patients had similar rates of clustering as those with no contact, but the rate was slightly higher for those with more recent (since 1993) contacts (58.3 vs. 40%, OR = 1.5).

A significantly higher rate of strain clustering was found in young patients (1–25 years) compared with older patients (70.4 vs. 38%, OR 3.87). There was also more clustering in patients in occupations which brought them more into contact with the public (66.7 vs. 43.1%, OR 2.64). Other less evident associations were observed for male sex and residence in an urban area (Table 3).

Independent predictive factors of strain clustering (Table 4) on multivariate analysis were: age 1–25 years (OR 2.92, 95% CI 0.83–10.3) and age 26–50 years (OR 1.27, 95% CI 0.45–3.51) vs. cases aged over 50 years; cases with 25–50% of contacts infected in the first circle (OR 1.72, 95% CI 0.44–6.73) and above 50% (OR 2.89, 95% CI 0.96–8.7) vs. studies with non-infected persons; residence in the town (OR 2.12, 95% CI 0.73–6.2) vs. rural dwelling, microbiological diagnosis made on bronchoscopy (OR 16.3, 95% CI 2.3–11.5) vs. sputum diagnosis, multiple contacts at work and work in the footwear industry (OR 2.41, 95% CI 0.69–8.3) vs. cases whose work did not involve contact with many people.

When independent variables are included, the model correctly predicts the appearance of 73.7% of the cases falling within clusters compared to 67% among individuals with unique strain genotypes.

We believe that the patient population analysed here is representative of all patients with microbiologically confirmed tuberculosis since they were selected over a long time period in a manner analogous to consecutive sampling of cases. Our earlier

Table 1. *Univariate analysis of the association between cases and strain clustering in relation to health-care variables*

Variable/category	Frequencies	% of clusters	OR	Significance (P value)	OR CI (95%)
PDU*					
No	114 (79.2%)	50.0	1		
Yes	25 (17.4)	72.0	2.57	0.05	0.99–6.6
Unknown	5 (3.5%)	20.0	0.25	0.22	0.03–2.3
Alcoholism					
No	95 (66.0%)	49.5	1		
Yes	44 (30.6%)	63.6	1.79	0.12	0.85–3.7
Unknown	5 (3.5%)	20.0	0.25	0.23	0.03–2.4
HIV+					
No	51 (35.4%)	49	1		—
Yes	21 (14.6%)	71.4	2.6	0.09	0.87–7.7
Not known	72 (50.0%)	50	1.04	0.91	0.51–2.13
Quantification of staining					
Negative	94 (65.2%)	51.1	1		
Slight	3 (2.1%)	100	n.a.	0.73	n.a.
Moderate	9 (6.3%)	55.6	1.2	0.80	0.3–4.7
Abundant	38 (26.4%)	52.6	1.06	0.87	0.5–2.3
Cavitation					
No	102 (70.8%)	52.9	1		—
Yes	26 (18.1%)	57.7	1.21	0.66	0.51–2.9
Unknown	16 (11.1%)	43.8	0.69	0.50	0.24–1.99
Site					
Extrapulmonary	35 (24.3%)	42.9	1		—
Pulmonary	109 (75.7%)	56.0	1.69	0.18	0.78–3.7
Sample type					
Sputum	94 (65.3%)	54.3	1		—
Bronchoscope	10 (6.9%)	80	3.37	0.14	0.68–16.7
Others	40 (27.8%)	42.5	0.62	0.21	0.29–1.31

* Parenteral drug user.

genotyping study of the strains from these patients had shown a rate of clustering relatively higher than that reported from a number of other countries [5]. This is probably due to the fact that the present study covered a longer period of time. However, these data may be affected by other factors, such as the characteristics of tuberculosis in Spain where the prevalence is greater than in some other countries [3].

Although recent transmission of tuberculosis has been related to clustering using IS6110-RFLP [11], it is clear that transmission is also associated with socio-economic factors of the study population such as mobility, social mix and concentration [12]. In addition, the existence of endemic strains that are

preferentially transmitted in a population has been suggested [12]. A study carried out in Arkansas [8] on a stable rural population with a low prevalence of HIV infection showed that there was no epidemiological relationship in 60% of the patients in whom clustering was found using RFLP and, furthermore, there was evidence of previous tuberculosis in a third of these. This suggested that some of the clusters were due to simultaneous reactivations of strains that were endemic to the region. However, it should be borne in mind that epidemiological studies using classical methods identify less than half of the patients in whom clustering is found using RFLP [8, 11, 13], since it has been suggested that very brief contacts [14, 15], or contagion arising from contacts of which the patients

Table 2. *Univariate analysis of the association between cases and strain clustering in relation to degree of contact*

Variable/category	Frequencies	% of clusters	OR	Significance (<i>P</i> value)	OR CI (95%)
First circle					
0%	45 (31.2%)	44.4	1	—	—
1–25%	16 (11.1%)	56.3	1.61	0.42	0.51–5.07
25–50%	16 (11.1%)	56.3	1.61	0.42	0.51–5.07
51–100%	41 (28.5%)	65.9	2.41	0.05	1.00–5.77
Not known	26 (18.1%)	42.3	0.92	0.86	0.35–2.43
Contacts					
Up to first circle	40 (27.8%)	50	1	—	—
Up to second circle	58 (40.3%)	55.2	1.23	0.61	0.55–2.76
Up to third circle	20 (13.9%)	65	1.86	0.27	0.61–5.63
Not known	26 (18.1%)	42.3	0.73	0.54	0.27–1.98
No. of families					
1	32 (22.2%)	43.8	1	—	—
2–4	55 (38.2%)	54.5	1.54	0.33	0.64–3.71
≥5	31 (21.5%)	67.7	2.70	0.06	0.97–7.54
Not known	26 (18.1%)	42.3	0.94	0.91	0.33–2.68
Contact with TB					
No	60 (41.7%)	48.3	1	—	—
Yes after 1993	48 (33.3%)	58.3	1.5	0.30	0.70–3.22
Yes before 1993	10 (6.9%)	40	0.71	0.63	0.18–2.78
Not known	26 (18.1%)	57.7	1.46	0.43	0.58–3.69

Table 3. *Univariate analysis of the association between cases and strain clustering in relation to socio-economic variables*

Variable/category	Frequencies	% of clusters	OR	Significance (<i>P</i> value)	OR CI (95%)
Age (years)					
≥51	50 (34.7%)	38	1	—	—
26–50	67 (46.5%)	56.7	2.14	0.05	1.01–4.52
1–25	27 (18.8%)	70.4	3.87	0.01	1.42–10.57
Residence					
Other (non-urban)	25 (17.4%)	40.0	1	—	—
Elche (urban)	119 (82.6%)	55.5	1.87	0.16	0.78–4.5
Profession					
No contact with people	65 (45.1%)	43.1	1	—	—
Footwear industry	23 (16.0%)	65.2	2.48	0.07	0.92–6.6
Contact	33 (22.9%)	66.7	2.64	0.03	1.1–6.3
Unknown	23 (16.0%)	47.8	1.2	0.69	0.47–3.1
Sex					
Female	44 (30.6%)	47.7	1	—	—
Male	100 (69.4%)	55.0	1.34	0.42	0.66–2.73

were unaware, play an important role in the transmission of tuberculosis.

There is a greater proportion of clustering of strains in under-developed countries with a high incidence

of tuberculosis such as central Africa, for example, where there is more recent transmission [16]. On the other hand, in countries with few cases of tuberculosis such as The Netherlands, there is greater strain

Table 4. Multivariate analysis of independent variables associated with cases belonging to a strain cluster

	Adjusted OR	OR CI (95 %)	Significance (P value)
Age (years)			
≥51	1	—	
26–50	1.27	0.45–3.51	0.65
1–25	2.92	0.83–10.3	0.09
First circle			
0 %	1	—	
1–25 %	1.45	0.36–5.82	0.60
25–50 %	1.72	0.44–6.73	0.43
51–100 %	2.89	0.96–8.68	0.06
Residence			
Other (non-urban)	1	—	
Elche (urban)	2.12	0.73–6.2	0.17
HIV			
No	1	—	
Yes	2.0	0.41–9.6	0.38
Unknown	0.95	0.40–2.2	0.90
PDU*			
No	1	—	
Yes	2.72	0.58–12.8	0.20
Unknown	0.17	0.01–2.96	0.22
Sex			
Male	1	—	
Female	1.26	0.53–3.01	0.60
Contact at work with people			
No	1	—	
Footwear industry	2.41	0.69–8.3	0.17
Yes	2.81	0.94–8.3	0.06
Unknown	1.84	0.56–6.1	0.31
Sample type			
Sputum	1	—	
Bronchoscope	16.3	2.3–115	0.005

* Parenteral drug user.

heterogeneity and many cases are reactivated rather than new acquisitions [17]. In Norway, the country with the lowest incidence of tuberculosis in the world, there was clustering in only 19.7% of patients studied [18]. In contrast, in developed countries with a high rate of incidence of the disease such as Japan, with 41.9 cases per 100 000 inhabitants in 1990, the isolates exhibit a high degree of polymorphism indicating that although the incidence is high, it could be mainly due to reactivation of disease [19]. The rate of clustering obtained in our population is a reflection of a region in which tuberculosis is endemic, with a relatively high percentage of HIV + patients and where there is also a significant circulation of strains. This percentage

varies depending on the socio-economic characteristics of each area.

Various studies have highlighted greater clustering of strains in HIV + patients [20–22]; however, some reports contradict this [11, 15]. This discrepancy may be related to nosocomial transmission, the patients' socio-economic situation, the effect of how quickly the diagnosis is made, and the efficacy of treatment on the transmission and progression of the disease in these patients. In our study, we found a greater rate of clustering in patients who were infected with HIV, and this is statistically significant in the univariate analysis, although it does not appear as an independent risk factor in the multivariate analysis. Circulating strains may exist in the area, mainly transmitted between HIV + patients during contacts involving risk practices [23]. We also found a higher rate of clustering in PDU patients but there is considerable overlap between these groups since 80% of HIV-infected patients were also PDU.

In our study, the younger age group showed the greater the risk of clusters, which is in agreement with the literature [11, 20, 22, 24]. This could be a consequence of social peer groups [8, 13, 15] together with the lower risk in these patients of reactivations after a long period of latency. The greater susceptibility of such persons to new infections also plays a part as in Spain people younger than 20–25 years are no longer vaccinated with BCG [25, 26]. There was a clear relation between residence in the most urban area and strain clustering confirming earlier findings of population density with transmission risk [15, 22, 24].

The strong association between samples obtained by bronchoscopy and clusters has not been reported in the literature, except for contaminations associated with the procedure [27]. Our patients who underwent bronchoscopy form part of clusters with patients diagnosed by other procedures and in different years, and so transmission by bronchoscopy may be ruled out. Despite the small number of patients diagnosed using this procedure, we believe that it may be important since bronchoscopy, in addition to its use for the diagnosis of patients co-infected with HIV, is necessary in patients in whom traditional microbiological diagnosis is difficult (non-obvious pulmonary lesions) and in patients in whom suspicion of tuberculosis is based on contact study data. These cases frequently correspond to recent infections with a short clinical history, and might be considered evidence of the improvement made in early diagnosis of this disease in Spain. However, further studies should be

done on larger numbers of patients to confirm this association.

The association between clustering and the greater proportion of patients infected in the first circle of the contact study seems logical and could be explained by the greater capacity and duration of contagion in some cases. We believe that this variable may be the best marker of capacity for contagion of each patient within the family setting (the usual setting investigated in contact studies). This result might also be due to the existence of more contagious patients in our area [5, 19]. The link between work occupation and clustering also appears consistent and complementary to the association found in contact studies in the home setting and probably reflects the capacity for transmission of the organism outside of the domestic area of certain cases of tuberculosis.

In conclusion we recommend that the results obtained from RFLP genotyping of patients' isolates should be interpreted in the light of the clinical and epidemiological data to enable a better understanding of the epidemiology and transmission dynamics of tuberculosis. This analysis permits us to identify factors associated with recent transmission in our setting and the possible existence of endemic strains in our community.

REFERENCES

1. Rey R, Ausina V, Casal M, et al. Situation of tuberculosis in Spain. A precarious health perspective with regard to developed countries [in Spanish]. *Med Clin (Barc)* 1995; **105**: 703–707.
2. Caminero JA, Rodríguez de Castro F, Carrillo Rodríguez JC. Importance of age groups as an epidemiological parameter of tuberculosis [in Spanish]. *Med Clin (Barc)* 1991; **96**: 156–157.
3. Rodríguez E. Present situation of tuberculosis in Spain: incidence and mortality since 1995. Characteristics of cases of tuberculosis and tuberculous meningitis declared in 2000 [in Spanish]. *Boletín epidemiológico. Centro Nacional de Epidemiología. Instituto de Salud Carlos III*. 2001; **9**: 293–304.
4. van Embden JDA, Cave D, Cawford JT, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993; **31**: 406–409.
5. Ruiz M, Rodríguez JC, Navarro JF, Samper S, Martín C, Royo G. Molecular epidemiology of tuberculosis in Elche, Spain: a 7-year study. *J Med Microbiol* 2002; **51**: 273–277.
6. 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.
7. Solsona J, Cayla JA, Verdú E, et al. Molecular and conventional epidemiology of tuberculosis in a inner city district. *Int J Tuberc Lung Dis* 2001; **5**: 724–731.
8. Braden CR, Templeton GL, Cave MD, et al. Interpretation of restriction fragment length polymorphism analysis of *Mycobacterium tuberculosis* isolates from a state with a large rural population. *J Infect Dis* 1997; **175**: 1446–1452.
9. Larrosa JA. Social and demographic atlas of the town of Elche [in Spanish]. *Publicaciones de la Universidad de Alicante*. Alicante, 2000.
10. Centers for Disease Control and Prevention. Targeted tuberculin testing and treatment of latent tuberculosis infection. *MMWR* 2000; **49**: 1–132.
11. 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**: 726–736.
12. Hermans PWM, van Soolingen D, Dale JW, et al. Insertion element IS986 from *Mycobacterium tuberculosis*: a useful tool for diagnosis and epidemiology of tuberculosis. *J Clin Microbiol* 1990; **28**: 2051–2058.
13. Diel R, Schneider S, Meywald-Walter K, Ruf CM, Rüsche-Gerdes S, Nieman S. Epidemiology of tuberculosis in Hamburg, Germany: long term population-based analysis applying classical and molecular epidemiological techniques. *J Clin Microbiol* 2002; **40**: 532–539.
14. Bauer J, Kok-Jensen A, Faurschou P, Thuesen J, Taudorf E, Andersen AB. A prospective evaluation of the clinical value of tuberculosis isolates in Denmark. *Int J Tuberc Lung Dis* 2000; **4**: 295–299.
15. Weis SE, Pogoda JM, Yang Z, et al. Transmission dynamics of tuberculosis in Tarrant County, Texas. *Am J Resp Crit Care Med* 2002; **166**: 36–42.
16. Kallenius G, Koivula T, Ghebremichael S, et al. Evolution and clonal traits of *Mycobacterium tuberculosis* in Guinea-Bissau. *J Clin Microbiol* 1999; **37**: 3872–3878.
17. van Soolingen D, Hermans PWM, de Haas PEW, Soll DR, van Embden JDA. The occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains; evaluation of insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J Clin Microbiol* 1991; **29**: 2578–2586.
18. Dahle UR, Sandven P, Heldal E, Caugant DA. Molecular epidemiology of *Mycobacterium tuberculosis* in Norway. *J Clin Microbiol* 2001; **39**: 1802–1807.
19. Takahashi M, Kazumi Y, Fukasawa Y, et al. Restriction fragment length polymorphism analysis of epidemiologically related *Mycobacterium tuberculosis*. *Microbiol Immunol* 1993; **37**: 289–294.
20. Small P, Hopewell PC, Singh SP, et al. The epidemiology of tuberculosis in San Francisco: a population based study using conventional and molecular methods. *New Engl J Med* 1994; **330**: 1703–1709.

21. Genewein A, Telenti A, Bernasconi C, et al. Molecular approach to identifying route of transmission of tuberculosis in the community. *Lancet* 1993; **342**: 841–844.
22. Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York city. An analysis by DNA fingerprinting and conventional epidemiological methods. *N Engl J Med* 1994; **330**: 1710–1716.
23. Camarena JJ, Artero A, Nogueira JJ, Navarro JC, Olmos A, Blanquer R. Tuberculosis in AIDS patients: the contribution of an analysis of the restriction fragment length polymorphism of *Mycobacterium tuberculosis* isolates [in Spanish]. *Med Clin (Barc)* 1998; **111**: 721–724.
24. van Deutekom H, Gerritsen JJJ, van Soolingen D, van Ameijden EJC, van Embden JDA, Coutinho RA. A molecular epidemiological approach to studying the transmission of tuberculosis in Amsterdam. *Clin Infect Dis* 1997; **25**: 1071–1077.
25. Vynnycky E, Fine PE. The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol Infect* 1997; **119**: 183–201.
26. Vynnycky E, Nagelkerke N, Borgdorff MW, van Soolingen D, van Embden JD, Fine PE. 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.
27. Gubler JG, Salfinger M, Gravenenitz A. Pseudoepidemic of nontuberculous mycobacteria due to contaminated bronchoscope cleaning machine. Report of an outbreak and review of the literature. *Chest* 1992; **101**: 1245–1256.