Clinical and Experimental Rheumatology 2003; 21: 635-638.

Identification of possible areas of high prevalence of Paget's disease of bone in Spain

G. López-Abente[1],
A. Morales-Piga[2,3],
F.J. Bachiller-Corral[2],
O. Illera-Martín[2],
R. Martín-Domenech[2],
V. Abraira[4]

[1]Evironmental Epidemiology and Cancer Department, National Center for Epidemiology, Carlos III Institute of Health, Madrid; [2]Department of Rheumatology, Ramón y Cajal Hospital, Madrid; [3]Alcalá University Medical School, Madrid; [4]Unit of Biostatistics, Ramón y Cajal Hospital, Biomathematics Department, Complutensis University of Madrid, Spain.

Gonzalo López-Abente, MD PhD; Antonio Morales-Piga, MD, PhD, Associate Professor of Rheumatology; Francisco J. Bachiller-Corral, MD; Oscar Illera-Martín, MD; Raquel Martín-Domenech, MD; Víctor Abraira, PhD, Associate Professor of Biostatistics.

Please address correspondence to: Gonzalo López-Abente, MD, Area de Epidemiología Ambiental y Cáncer, Centro Nacional de Epidemiología, c/Sinesio Delgado 6, 28029 Madrid, Spain. E-mail: glabente@isciii.es

Received on November 28, 2002; accepted in revised form on July 1, 2003.

Key words: Osteitis deformans, space-time clustering, epidemiology, case-control studies.

ABSTRACT

Objetive: *In view of the fact that Paget's disease of bone (PD) tends to appear in so-called 'foci', a case-control study was undertaken with the dual aim of: 1) identifying areas having a higher likelihood of constituting PD 'foci'; and 2) detecting the geographic origin of 'PD-carrier' families*

Methods: *Two data sets were analysed, one covering the place of birth of 231 cases and 436 controls, and the other covering the place of birth of cases, controls and their parents. Analysis was restricted to six Autonomous Regions accounting for 60% of Spain's towns and cities. To identify geographical areas of high prevalence we used the scan statistic*

Results: *In the first analysis, 6 possible clusters were detected, corresponding to the districts of Avila (Avila), Lozoya-Somosierra (Madrid), Tierra de Campos(Palencia), the Guadalajara Range, South-west Madrid and Cuenca Hills. The second analysis confirmed the 6 groupings identified by the above procedure and, in addition, detected a further 8 possible clusters. Geographical proximity suggests that in some cases, rather than individual groupings, these may instead constitute larger foci.*

Conclusion: *The results point to the possible existence of different PD foci, some coinciding with clusters that have already been reported, and others indicating familial origin in areas that had never previously received PD-specific attention.*

Introduction

As described in the literature, Paget's disease of bone (PD) tends to plot an irregular geographical distribution, with reports of clusters in some enclaves showing a prevalence exceeding 7% in persons over the age of 55 years (1, 2). The best known of these 'foci' is that detected in Lancashire (3). To date, two foci of a similar nature have been reported in Spain, one in the Madrid Range (*Sierra de Madrid*) (4) and another in the Province of Salamanca (5). Both were initially brought to light by clinical observation of an excess frequency of cases coming from these areas and a trend towards familial aggregation. As a result, specific population-based studies were conducted to confirm these findings (4, 5). The trend towards several cases of PD affecting a single family (6,7) would seem to point to factors of a genetic origin and is generating the search for specific markers of susceptibility (8,9). There is no reason to rule out the possibility of there being other enclaves with high prevalence of this disease, nor are the physical limits of previously identified foci known. Detection of such possible clusters may prove very useful for a better understanding of the environmental and genetic mechanisms (and the interaction between the two) underlying PD.

The different methods for detecting clusters can be divided into two categories: 1) tests for overall clustering; and 2) tests for the detection of clusters. Perhaps the best known in the latter group are Openshaw's Geographical Analysis Machine (10) and Besag and Newell's test (11). A test based on that of Turnbull (12) has recently become popular; though similar to the above-mentioned tests, it not only corrects for the problem of multiple comparisons (Kulldorff 1995) but also offers other advantages (13) and therefore has been chosen for this study.

In order to further explore the fact that PD tends to appear in these so-called 'foci', we designed a case-control study with the dual aim of: 1) identifying areas having a higher likelihood of constituting PD 'foci', and 2) detecting the geographic origin of 'PD-carrier' families. This design enables case-clustering to be studied while controlling for spatial variation in the origin of the population attending the out-patient clinics from which the cases come, with the latter information being provided by the geographic origin of controls and their parents.

Materials and methods

Case definition

"Cases" comprised all 231 PD patients registered at our Unit (a facility specialised in PD) over the period January 1992 to June 2001. Approximately half were referred to our Unit by their general practitioners, after detection of an

otherwise unexplained rise in alkaline phosphatase and, less frequently, on the grounds of clinical suspicion. Roughly 25% came from other specialists in our hospital, usually detected by chance for reasons unconnected with PD problems. The remaining 25% were incorporated into our PD stock through sampling performed to conduct previous studies (4,14). Throughout the study period, diagnosis was based on the same radiological criteria (15). All patients were followed up on a regular basis, using the same methods of clinical evaluation. Nine patients (3.9%) diagnosed for the purpose of familial studies were excluded from the analysis when index cases alone had to be taken into account.

Selection of controls

The controls were 436 consecutive ambulatory patients over 45 years of age, attending our clinic for routine periodic evaluation of diverse musculoskeletal complaints, excluding PD.

Variables analysed

The place of birth of cases and controls was registered and coded to 5 digits using the standard codes supplied by the National Statistics Institute (*Instituto Nacional de Estadística*). The father's and mother's place of birth were likewise coded for both cases and controls. Analysis was restricted to the 4800 Spanish towns and cities that make up the Autonomous Regions of Castile-Leon, Castile-La Mancha, Extremadura, Galicia, Andalusia and Madrid, using the Universal Transverse Mercator (UTM) coordinates of their respective population centroids.

Analysis was performed using two data sets, one corresponding to the place of birth of cases and controls, and the other to the place of birth of cases, controls and their parents, thus multiplying the subjects by three in each group. The Regions included in the analysis account for 91% birthplaces of cases and controls and parents.

Statistical analysis

To identify geographical areas of high prevalence, we used the scan statistic proposed by Kulldorff and Nagarwalla (14), which allows for possible clusters in a population having a non-homogeneous spatial density to be located and the statistical significance of the same ascertained. Since this was a case-control study, a binomial distribution was used. To assess the size and location of clusters, this test defines circles centred on every municipal population centroid across the entire surface area of Spain. In each centroid, the size of the respective circle varies from zero to a maximum radius, specified *a priori* without the size of the possible clusters in the data sets being known. For study purposes, this radius was defined as being equal to 1% of the total risk population (represented in our design by the controls). For each circle, the procedure calculates the number of cases situated in its interior and the number of expected cases based on the distribution of the controls. Taking these values as the point of departure, the likelihood ratio associated with each circle is then computed. The likelihood function is maximised over all windows and the window constituting the most likely cluster duly identified. The likelihood ratio for this window is noted and constitutes the maximum likelihood ratio test statistic. To ascertain the distribution of the statistical test, the program generates 999 replications from the data set under the null hypothesis, with the test then being calculated for each replication. The result is significant at the 0.05 level if the value of the test statistic from the real data set ranks among the highest 5% of all 1000 values, including those calculated from the 999 replications. The SaTScan 2.1 computer software program developed and distributed by the National Cancer Institute (USA) was used for the purpose.

Table I. Statistically significant clusters detected. (A) Analysis of the place of birth of cases and controls. (B) Analysis of the place of birth of cases, controls and their parents.

Number[1]	Location	District[2]	Likelihood ratio	Cases	Expected cases	p-value
A)						
1	Avila	Avila	31.47	15	1.90	0.001
2	North Madrid	Lozoya-Somosierra	19.610	12	1.78	0.001
3	Palencia	Tierra de Campos	16.091	11	1.78	0.001
4	Soria-Guadalajara	Guadalajara Range	12.401	9	1.52	0.001
5	South-west Madrid	South-west Madrid	11.966	10	1.90	0.001
6	Cuenca	Cuenca Hills	8.844	7	1.27	0.050
B)						
1	South Avila	Gredos-Valle Bajo Alberche	18.845	17	5.67	0.001
2	South Leon	El Páramo-Esla Campos	18.845	17	5.67	0.001
3	North Madrid	Lozoya-Somosierra	17.727	16	5.33	0.001
4	Salamanca	Ciudad Rodrigo	13.267	12	4.00	0.002
5	Soria-Guadalajara	Guadalajara Range	13.267	12	4.00	0.002
6	South Avila -Madrid	South-west Madrid	13.267	15	5.33	0.003
7	Avila	Avila	13.267	15	5.33	0.003
8	Segovia	Segovia	12.218	14	5.00	0.016
9	Cuenca	Cuenca Hills	11.044	10	3.33	0.016
10	West Madrid	Metropolitan area	11.044	10	3.33	0.041
11	Huelva	Sierra-Andévalo	9.935	9	3.00	0.041
12	North Toledo	Torrijos	9.935	9	3.00	0.041
13	North Madrid	Lozoya-Somosierra	9.287	13	5.00	0.046

[1]Code corresponding to cluster site on map; [2]Name of historical country districts.

Results

Statistically significant clusters detected in the two analyses are shown in Table I and marked on the map of Spain depicted in Figures 1 and 2. In the analysis based exclusively on the place of birth of cases and controls (610 towns), 6 possible clusters were detected, corresponding – in descending order of importance – to the districts of Avila (Avila), Lozoya-Somosierra (Madrid), the Tierra de Campos

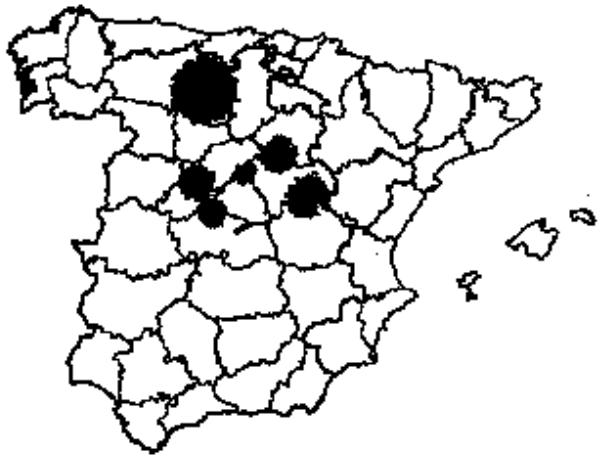

Fig. 1. Detection of possible clusters of birthplaces of Paget's disease of bone patients. Provinces included on the study are those are in the areas marked by a thick line.

Fig. 2. Detection of possible clusters of birthplaces of Paget's disease of bone cases and their parents.

(Palencia), the Guadalajara Range, South-west Madrid and the Cuenca Hills (*Serranía de Cuenca*). The second analysis, which took into account the birthplace of cases, controls and their parents (1811 towns), detected 13 clusters corresponding to the districts of Avila, Leon, Madrid-Segovia, Salamanca, Soria-Guadalajara, Cuenca and Huelva. The size of the clusters detected was smaller, the reason being that, as the number of birthplaces was higher, the imposed constraint of a maximum radius confined the 1% to a smaller surface area. Owing to their proximity, the 4 clusters detected in the Avila and Toledo districts could well constitute a single enclave, as could those corresponding to the Segovian and Madrid sides of the range of mountains that lies just outside Madrid. In contrast, the foci corresponding to Salamanca (Ciudad-Rodrigo district), Leon (El Páramo and Esla-Campos districts), Cuenca (Hill districts) and Huelva (Andévalo districts) were well-differentiated. Detection of this type of cluster points to the families' place of origin. Furthermore, use of the parents' place of birth allowed for the identification of new foci not detected in the first analysis performed on the basis of the cases and controls' place of birth.

Discussion

This study explores: 1) the existence of possible areas of high PD prevalence in Spain using a case-control design, and 2) the geographic origin of the patients' families. Notwithstanding the fact that the trend towards case clustering (existence of foci) is a phenomenon reported in PD and that the study is exploratory in nature, the validity of the method used to detect the same nevertheless calls for in-depth discussion. The justification for undertaking an exploratory study such as this, limited to a single Public Health district, lies in its being a first approach to the problem.

Over half of all PD cases are asymptomatic and the series of cases that was used explicitly excluded all family members identified in specific studies: had this not been done, any possible clusters might merely have indicated areas in which cases had been actively 'searched' for. In our study, the control group was designed to show the spatial heterogeneity of the geographic origin of patients attending the out-patient clinics from which the cases were selected. Selection of controls took place over a shorter time period than, and subsequent to, that of the cases. While the existence of hospital 'catchment areas'and possible modifications to the same could arguably have had an influence on the patterns encountered, no administrative and health care-related changes have taken place in the last ten years which would lead one to suspect this type of effect. Thus we have not any reason to think that the results merely reflect a pattern of patient referral.

The reasons for having included parents'place of birth in one of the analyses were: 1) to enhance the power of the study, since it amounted to multiplying the number of observations by three; and 2) to detect the 'geographic origin' of the disease. PD is a process with a strong hereditary factor (8, 9, 16), with at least 40% of familial cases (7) and a pattern of dominant autosomic transmission (16), implying a more than 50% likelihood of presenting with PD across all generations and both sexes. It is therefore reasonable to conclude that, in many cases, the disease may have been transmitted by one of the two family branches, with some member of the parents' generation or even the parents themselves being affected. Given that both parents' towns/ cities of origin were the same in 56% of cases in our series (or were situated relatively near to each other, since 80% had the same province of origin), such parent-based information can thus be assumed to be a fairly reliable pointer to the ultimate geographic origin of the disease. Indeed, it is this etiopathogenic inference and not the patient's cur-

rent place of residence that was this study's stated goal.

The detection procedure employed yields a principal and secondary clusters, with all statistically significant clusters then being plotted on the maps. In the second analysis more possible clusters were detected in the environs of Madrid. This is probably due to the design and to the fact that Madrid was the hospital catchment area. Similarly, the exclusion of many Autonomous Regions in the analysis was imposed by the geographically limited nature of the study. Nevertheless, when the analysis was repeated without excluding any Autonomous Region, the results proved very similar. All province boundaries have been retained in the maps to facilitate the location of possible foci.

Available data shows a great variation in the occurrence of PD within and among different racial and ethnic groups (17, 18). In our country, mainly in familial cases observed in the regions of central and western Spain, there are some previous indications of association between certain "northern-type" phenotypic traits and PD (7). An inherited factor could be in agreement with a greater susceptibility to develop the disease in populations having some distinctive ethnic origins.

The geographical regions identified by spatial cluster analysis as possible areas of origin of PD patients/families, are located in some delineated territories of central and western Spain (Fig. 1 and 2). In other analyses of the same individuals recruited for the present study, consanguinity appeared to be strongly linked to the disease (OR 4.1, 95% confidence interval 2.1-8.1), and certain phenotypic characteristics such as clear eye colour also showed a statistically significant relationship (OR 1.5, 95% confidence interval 1.1-2.1) that requires further explanation. In Spain there is generally a high degree of intermixing among populations of different racial origins. However, in certain rural parts of "Castilla", areas of Spain where, historically speaking, the influence of invasions by Germanic tribes proved most intense (19), natural and sociologic barriers served to keep the inhabitants fairly isolated until relatively recent times. These conditions may give rise to the speculation that such populations could still be strongly influenced by the genetic background of these ancient peoples and that some ethnic predisposition to PD could hence exist. In this regard, the difference in prevalence observed between native American Indians and Caucasians further supports the influence of racial factors (17). Similar conclusions have been reached in studies conducted in other geographical areas, such as Ireland (1) and New Zealand (20). However, inferences regarding racial susceptibility based on prevalence studies conducted in different geographical areas must be cautiously considered. First, it must be outlined that environmental factors could explain part of the differences in the prevalence rates among different geographical areas.

The dual-analysis procedure conducted points to the possible existence of different PD foci, some coinciding with those that have already been reported (4, 5), and others indicating familial origins in areas (Huelva, Palencia and Cuenca) that had never previously received PD-specific attention. Strictly speaking, the procedure detects the place of birth/family origin of PD patients and we do not know whether these in fact correspond to areas of greater prevalence. This is something that could be confirmed either by upgrading the scope of the study to a multicentre level or by other procedures.

References

1. DETHERIDGE FM, GUYER PB, BARKER DJP: European distribution of Paget's disease of bone. *Br Med J* 1982; 285: 1005-8.
2. BARKER DJP: The epidemiology of Paget's disease of bone. *Br Med Bull* 1984; 40: 396-400.
3. BARKER DJP, CHAMBERLAIN AT, GUYER PB, GARDNER MJ: Paget's disease of bone: The Lancashire focus. *Br Med J* 1980; 280: 1105-7.
4. MORALES-PIGA A, LÓPEZ-ABENTE G, ELENA IBÁÑEZ A, GARCÍA-VADILLO A, GONZALEZ LANZA M, MONGE JODRÁ V: Risk factors for Paget's disease: a new hypothesis. *Int J Epidemiol* 1988; 17: 198-201.
5. MIRÓN CANELO JA, DEL PINO MONTES J, VICENTE ARROYO M, SÁENZ GONZÁLEZ MC: Epidemiological study of Paget's disease of bone in a zone of the Province of Salamanca (Spain). *Eur J Epidemiol* 1997; 13: 801-5.
6. SIRIS E, OTTMAN R, FLASTER E, KELSEY JL: Familial aggregation of Paget's disease of bone. *J Bone Min Res* 1991; 6: 495-500.
7. MORALES-PIGA A, REY REY J, CORRES J, GARCÍA SAGREDO JM, LÓPEZ-ABENTE G: Frequency and characteristics of the familial aggregation in Paget's disease of bone. *J Bone Miner Res* 1995; 10: 1-8.
8. HASLAM S, VAN HUL W, MORALES-PIGA A *et al.*: Paget's disease of bone:Evidence for a susceptibility locus on Chromosome 18q and for genetic heterogeneity. *J Bone Miner Res* 1998; 13: 911-7.
9. LAURIN N, BROWN J, LEMAINQUE A *et al.*: Familial Paget's disease of bone:Mapping of two loci at 5q35-qter and 5q31. *Am J Hum Genet* 2001; 69: 528-43.
10. OPENSHAW S, CRAFT AW, CHARLTON M, BIRCH JM: Investigation of leukaemia clusters by use of a Geographical Analysis Machine. *Lancet* 1988;1: 272-3.
11. BESAG J, NEWELL J: The detection of clusters in rare diseases. *Journal of the Royal Statistical Society* 1991; 154: 143-55.
12. TURNBULL BW, IWANO EJ, BURNETT WS, HOWE HL,CLARK LC: Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *Am J Epidemiol* 1990; 132: S136-S143.
13. KULLDORFF M, NAGARWALLA N: Spatial disease clusters:detection and inference. *Stat Med* 1995; 14: 799-810.
14. MORALES-PIGA A, GONZÁLEZ-LANZA M, ARNAIZ VILLENA A, MARTÍNEZ ESCRIBANO B, ALONSO RUIZ A,ZEA MENDOZA A: Agregación familiar en la enfermedad de Paget: implicaciones etiopatogénicas. *Med Clín (Barc)* 1983; 81: 43-6.
15. MURRAY RO, JACOBSON HG: Miscellaneous disorders of unknown origin. *In:* MURRAY RO (Ed.): *The Radiology of Skeletal Disorders*, Vol II, Edinburgh, Churchill Livingstone Incorporated, 1979.
16. MCKUSICK VA: Paget's disease of the bone. *In* MCKUSICK VA: *Heritable Disorders and Connective Tissue*. St. Louis, Mosby Co., 1972.
17. LAWRENCE JS: Paget's disease in population samples. *Ann Rheum Dis* 1970; 29: 562.
18. MAUTALEN C, PUMARINO H, BLANCO MC, GONZÁLEZ D. GHIRINGHELLI G, FROMM G: Paget's disease: The South American experience. *Semin Arthritis Rheum* 1994; 23: 226-7.
19. BERTRANPETIT J, CAVALLI-SFORZA LL: A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann Hum Genet* 1991; 55: 51-67.
20. REASBECK JC, GOULDING A, CAMPBELL DR, BEALE LR, STEWART RD: Radiological prevalence of Paget's disease in Dunedin, New Zealand. *Br Med J (Clin Res Ed)* 1983; 286: 1937.