# The existence of geographical clusters of cases of inflammatory polyarthritis in a primary care based register

Alan Silman, Beverley Harrison, Elizabeth Barrett, Deborah Symmons

**Abstract**

***Objectives*—To determine whether there is any evidence that there are spatial clusters of rheumatoid arthritis in particular, and inflammatory arthritis in general.**

***Methods*—Setting was a population based incidence register of inflammatory arthritis: the Norfolk Arthritis Register (NOAR). All cases identified between 1990–1995 were mapped to place of residence. Statistical evidence of clustering was determined by calculating Poisson probabilities in putative areas.**

***Results*—Three clusters were identified including one small area (population 85) where five unrelated cases developed during this time period. There was no obvious greater disease homogeneity within clusters and no common environmental factors were identified.**

***Conclusion*—Rare clusters of inflammatory polyarthritis do occur. Their significance and cause remain to be elucidated.**

(*Ann Rheum Dis* 2000;**59**:152–154)

Geographical clustering, or the non-random distribution of cases in space, is considered to be suggestive of a point environmental risk factor for disease susceptibility. Such risk factors include exposure to toxins as well as focal sources of infection. The latter, specifically a viral or other microbial agent, is considered to be one of the most probable environmental triggers for the development of rheumatoid arthritis (RA).[1] It is probable that a number of different organisms may trigger RA but demonstration of local clusters would be of value in highlighting specific groups of patients for further investigation. Indeed the observation of geographical clustering of cases of juvenile arthritis was the first stage in the demonstration of *Borrelia burgdorferri* as the causal organism leading to Lyme disease.[2 3] By contrast most reports of geographical clusters, in the rheumatic diseases are anecdotal and based on chance ascertainment.[4 5] One problem, in interpreting such reports is that some clustering might occur by chance, and in the absence of whole population data are difficult to interpret.

The Norfolk Arthritis Register (NOAR) is a population based register of inflammatory polyarthritis (IP) covering a population of over 500 000. It is a unique population register aiming prospectively to ascertain incident cases of IP as they develop, within an entire community.[6] Data from NOAR were recently analysed for the existence of clustering in time, in space, and in time and space. Using simulation methods, comparing observed with expected distribution, there was no evidence of a non-random distribution of the occurrence of cases in these two dimensions in the NOAR area as a whole.[7] Thus there is no evidence of substantial geographical clustering for either IP in general, or RA in particular. Interestingly, in that original analysis,[7] the NOAR area was then divided into seven sub-areas and the analysis repeated. In one of these areas, the "North West", there was evidence of non-random distribution. These results suggested that although most cases occurred sporadically, clustering in space may occur. The next stage therefore is to examine the exact geographical location of individual cases in an attempt to identify any such clusters that would have explained the result above. If any clusters were identified, we also wished to test the hypothesis that there would be similarity in disease expression within a cluster because of a shared aetiological factor. The ultimate goal would be to investigate possible common aetiological factors, such as environmental exposures.

## Methods

In brief NOAR aims to capture all new cases of IP resident within the Norwich Health Authority (NHA) area, based on first attendance to primary care.[6] Cases are included if they have IP of at least two peripheral joints for at least four weeks. Subjects are excluded if they have an alternative specific diagnosis (apart from RA, psoriatic arthritis or post-viral arthritis) that accounts for their symptoms. All cases identified are interviewed and examined following a structured protocol and blood is taken for rheumatoid factor analysis. Subjects are followed up annually and radiographs taken in those with three or more ACR criteria[8] in all patients at year 1 and in all patients at year 5. All patients on the Register are classified as having either RA or undifferentiated IP based on the 1987 ACR criteria applied at the baseline visit.[8]

Evidence of geographical clustering was sought among all cases appropriately referred to NOAR between 1 January 1990 and 31 December 1995, a total of nearly 1500 cases. Area of residence at the time of onset was mapped using postcodes and the data visually inspected seeking small areas with more than three cases. These possible areas of high incidence were then investigated further. Population data for each of these areas are based on

**ARC Epidemiology Unit, University of Manchester, Stopford Building, Oxford Road, Manchester M13 9PT**
A Silman
B Harrison
D Symmons

**Norfolk Arthritis Register, St Michaels Hospital, Aylsham, Norfolk**
E Barrett

Correspondence to:
Professor Silman

Accepted for publication
12 October 1999

*Table 1 Observed and expected incidence of cases in NOAR and in three sub-populations*

| | | *Inflammatory polyarthritis* | | | | *Rheumatoid arthritis* | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Area* | *Population aged >15* | *Number of cases 1990–95* | *Incidence /100 000 pyr* | *Number of expected cases** | *p Value†* | *Number of cases 1990–95* | *Incidence /100 000 pyr* | *Number of expected cases** | *p Value†* |
| NOAR | 413 421 | 1476 | 51 | | | 962 | 32 | | |
| A | 532 | 7 | 219 | 1.6 | 0.001 | 6 | 188 | 1.0 | 0.0007 |
| B | 694 | 6 | 144 | 2.1 | 0.03 | 5 | 120 | 1.3 | 0.0093 |
| C | 85 | 5 | 980 | 0.28 | 0.00092 | 3 | 588 | 0.16 | 0.0001 |

*Based on NOAR incidence. †Probability based on Poisson distribution of observed or greater number of cases occurring by chance. pyr = person years.

postcodes and were obtained from the 1991 National Census. The analytical approach aimed to determine the likelihood beyond chance of the observed number of cases in such areas occurring, based on the overall incidence of RA and of all IP in the entire population covered by NOAR. The probability that the number of cases (or greater) observed attributable to chance in each area was calculated based on the Poisson distribution. Analyses were undertaken for the entire group of cases and separately for the subgroup that satisfied criteria for RA. When clusters did emerge, information from the NOAR database, and any clinical records available for the cases identified, were examined for similarity in exposure history, clinical presentation and subsequent course.

## Results

In the adult population covered by NOAR, the overall incidence of RA was 32 per 100 000 per year and 51 per 100 000 per year for IP. Five areas, on visual inspection, appeared to include a possible cluster. Three areas (table 1) showed significantly higher incidence rates. The total adult population of area C was only 85, who lived in two adjacent roads. During the period of observation, five of these 85 people developed IP, including three that satisfied criteria for RA. The Poisson probabilities of these number of cases occurring by chance in each of these three areas are, as shown, very small. The proportion of men and women in each of these three areas was identical to that observed in the original NOAR catchment population.[6] Furthermore, the proportion aged over 65 in the three areas were 18%, 20% and 21% respectively, the proportion of the NOAR area as a whole being 23%.[6]

We then compared the incidence of IP and RA in the areas immediately surrounding the clusters with that of the cluster areas. In the postcode areas surrounding areas A, B and C, the overall incidence of IP per 100 000 person years was 94.6, 61 and 55 respectively, in each case higher than the overall incidence of 51/per 100 000 person years.

Limited analysis of the clinical data available on the subjects in each of these three clusters (table 2) did not suggest a high degree of disease homogeneity, though the number of cases was clearly too small to make any definitive statement. In addition to the information presented, we examined the pattern of joint involvement, presence of nodules and the timing of onset, and no obvious similarity was noted within clusters. Perhaps the main point of interest is that, within area C, which had the highest probability of being a cluster, four of the five cases were rheumatoid factor positive. We examined data collected on all people at registration by NOAR for common triggers of disease. However, we could find no similarity between the cases in terms of symptoms of infection before disease onset, prior immunisation, occupation, pregnancy or smoking. Finally, none of the cases were related to one another.

*Table 2 Clinical details of patients within clusters*

| *Area* | *Case number* | *Sex* | *Year onset* | *Age onset* | *ACR criteria for RA* | *RF* | *Erosions* |
|---|---|---|---|---|---|---|---|
| A | 1 | F | 1990 | 48 | + | + | − |
| | 2 | F | 1990 | 54 | + | + | + |
| | 3 | F | 1991 | 20 | − | − | − |
| | 4 | F | 1994 | 48 | + | + | − |
| | 5 | F | 1994 | 40 | + | − | − |
| | 6 | F | 1994 | 45 | + | + | − |
| | 7 | M | 1995 | 33 | + | − | − |
| B | 1 | M | 1990 | 66 | + | + | + |
| | 2 | F | 1990 | 47 | + | − | − |
| | 3 | M | 1992 | 79 | − | − | + |
| | 4 | M | 1994 | 56 | + | + | + |
| | 5 | F | 1994 | 54 | + | − | − |
| | 6 | F | 1995 | 37 | + | − | − |
| C | 1 | M | 1991 | 61 | + | + | + |
| | 2 | F | 1989 | 62 | + | + | + |
| | 3 | F | 1991 | 51 | − | − | − |
| | 4 | M | 1989 | 44 | + | + | + |
| | 5 | F | 1995 | 31 | − | + | − |

## Discussion

The interpretation of reports of clustering is fraught with difficulty, although their presence provides some interesting insights into the role of environmental factors in disease onset. Care has to be taken in the statistical evaluation of potential clustering. In a disease as infrequent in the population as RA, in any small area, the existence of a single case, by definition, would create a greater than expected frequency. Across the NOAR populationas a whole, there was an incidence density of 51/100 000 person years. Consequently, in a population as small a size as area C, only 0.3 cases would have been expected during the six year period of observation. A single case is therefore in excess of this.

The Poisson probabilities calculated allow for this random occurrence and estimate how likely it is that the observed number of cases would have occurred in a denominator population of that size by chance. The difficulty is then determining whether such clusters are "real". It is perhaps useful to distinguish investigations that investigate a large population for the existence of clusters, from those, such as the present investigation, which report on what

are essentially anecdotal collections of cases that appear to occur unusually closer together. The statistical approach used in this study confirms the anecdotal impression that these clusters are unlikely to have occurred by chance. The results obtained should be treated with caution in so far as only a very small proportion of the potential geographical areas that could have been investigated, were actually analysed. As an illustration, in the total NOAR catchment population of 413 000 adults, there are 4860 theoretical sub-populations, the same size as area C and 780 the same as area A and, theoretically, the p values obtained should be adjusted for this multiple significance testing. We have already shown elsewhere that in the entire NHA area there is modest evidence of spatial clustering.[7] The current investigation adds to that observation and suggests that more formal techniques such as the use of a geographical analysis machine[9] and other methods to detect what may be referred to as "smaller area database anomalies"[10] would be appropriate.

One important concern is that any observed clustering represents localised selective over-reporting in specific neighbourhoods. Thus a person developing mild IP might be more likely to be ascertained if there is an awareness of other cases in the immediate vicinity. Similarly, reporting completeness is likely to vary by general practitioner, which might contribute to the likelihood of an apparent cluster. Notification to NOAR is based on attendance at primary care and we cannot exclude the possibility that an important proportion of mild cases is not reported, except in the areas of the apparent clusters observed in this report. Against this theory, however, is the fact that the cases observed in these clusters had similar spectrum of disease severity to NOAR cases as a whole.

The advantage of using a large population register to search prospectively for clusters, over anecdotal reporting of clusters, is that the possibility of clustering occurring by chance can be examined within the target population. As we have previously demonstrated, most cases of IP, including the specific subgroup with RA, are probably sporadic.[7] This more detailed evaluation now suggests that there may be occasional georgraphical clusters that are unlikely to be attributable to chance. We also postulated that, if the same environmental trigger was responsible, then this would lead to a similar disease phenotype. This was not the case in the clusters identified.

The individual cases were 'mapped' to their address at the time of reported disease onset. This particular population is comparatively less mobile than the rest of the United Kingdom. However, given the unknown latency of IP, the relevant address at the time of any putative environmental exposure is unknown.

There are few clues as to the nature of any common environmental agent. Infection with parvovirus does not explain the clusters described in this report.[11] Serum is, however, available for investigation of any future microbial hypothesis. Other putative environmental sources include the domestic water supply and, in this mixed rural-urban area, contamination from agricultural spraying.

We are grateful for the active support and collaboration of Professor David Scott and colleagues from the Norfolk and Norwich Hospital, the local primary care physicians, Sue Whiting for administrative and clerical support and the NOAR metrology team for the clinical assessments.

Funding: this study was funded by a programme grant from the UK Arthritis Research Campaign.

1 Silman AJ. Rheumatoid arthritis and infection: a population approach. Ann Rheum Dis 1989;48:707–10.
2 Maupin GO, Fish D, Zultowsky J, Campos EG, Piesman J. Landscape ecology of Lyme disease in a residential area of Westchester County, York. Am J Epidemiol 1991;133: 1105–13.
3 Kitron U, Kazmierczak JJ. Spatial analysis of the distribution of Lyme disease in Wisconsin. Am J Epidemiol 1997;145: 558–66.
4 Pritchard MH, Matthews N, Munro J. Antibodies to Influenza in a cluster of children with Juvenile Chronic Arthritis. Br J Rheumatol 1988;27:176–80.
5 Valesini G, Litta A, Bonavita MS, Lucan FL, Purpura M, Mariani M, *et al.* Geographical clustering of scleroderma in a rural area of the province of Rome. Clin Exp Rheum 1993;11:41–7.
6 Symmons DPM, Barrett EM, Bankhead CR, Scott DGI, Silman AJ. The incidence of rheumatoid arthritis in the United Kingdom: Results from the Norfolk Arthritis Register. Br J Rheumatol 1994;33:735–9.
7 Silman AJ, Bankhead C, Rowlingson B, Brennan P, Symmons D, Gatrell A. Do new cases of rheumatoid arthritis cluster in time or in space? Int J Epidemiol 1997; 26:628–34.
8 Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, *et al.* The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.
9 Openshaw S. Using a geographical analysis machine to detect the presence of spatial clustering and the location of clusters in synthetic data. In: Alexander FE, Boyle P, eds. *Methods for investigation localised clusters of disease*. IARC Publications No 135. Lyon: International Agency for Research on Cancer, 1996.
10 Newell JN, Besag JE. The detection of small area database anomalies. In: Alexander FE, Boyle P, eds. *Methods for investigation localised clusters of disease*. IARC Publications No 135. Lyon: International Agency for Research on Cancer, 1996.
11 Harrison B, Silman A, Barrett E, Symmons D. Low frequency of recent parvovirus infection in a population-based cohort of patients with early inflammatory polyarthritis. Ann Rheum Dis 1998;57:375–7.