Multiple Sclerosis 2005; **11:** 403–408
www.multiplesclerosisjournal.com

Multiple sclerosis in Tayside, Scotland: detection of clusters using a spatial scan statistic

Peter T Donnan[*,1], *John DE Parratt*[2], *Sally V Wilson*[2], *Raeburn B Forbes*[3], *Jonathan I O'Riordan*[2] *and Robert J Swingler*[2]

[1]*Tayside Centre for General Practice, Kirsty Semple Way, University of Dundee, Dundee DD2 4BF, Scotland, UK;* [2]*Neurology Section, Department of Medicine, University of Dundee, Dundee DD1 9SY, Scotland, UK;* [3]*Royal Victoria Hospital, Belfast BT12 6BA, Northern Ireland*

Debate continues over the relative importance of genetic factors over infectious agents in the aetiology of multiple sclerosis (MS). Detection of clusters of MS in space and time in the Tayside region of Scotland, UK would provide valuable evidence for the movement of infectious agents into a genetically susceptible population. A spatial scan statistic was used to detect, locate and provide a robust statistical test of any clusters found, without prior knowledge of their location or size. This was applied to a population-based MS register for the Tayside region of Scotland from 1970 to 1997, allowing for age at symptom onset, gender, population density and social deprivation. There were a total of 772 cases during the study period; an annual incidence of 7.2 per 100 000. The mean age of symptom onset was 35.7 (SD = 10.5) and 73.8% of cases were women. There was a general increase in cases over time probably reflecting gradually better detection and diagnosis. There was a peak around the mid-1990s and some evidence of periodicity. There was a highly significant temporal cluster between 1982 and 1995 (P = 0.002) for the whole region. Additionally, a significant spatial cluster for the time period 1993–1995 was found centred in the rural area south-west of Perth (P = 0.016). Significant temporal and spatial–temporal clusters are consistent with exogenous factors contributing to the distribution of MS in Tayside, Scotland.

Multiple Sclerosis *(2005)* **11**, 403–408

Key words: *cluster analysis; multiple sclerosis; spatial epidemiology; spatial scan statistic*

Introduction

Variation in the global prevalence of multiple sclerosis (MS) has been observed for over a century. The subsequent broad assignment of gross geographical regions to areas of 'high, medium and low' prevalence indicated a preponderance of the disease in latitudinal zones consistent with the settlement and spread of largely Caucasoid populations.[1] Such generic epidemiological observations strongly favour a genetic predisposition in the development of MS. However, this assumes that within any given study population the prevalence of MS is homogeneous.

Intraregional variation in the prevalence of MS has been observed, notably in the Faroe Islands[2] and to a lesser extent in the Orkney and Shetland Islands[3] and Iceland.[4] It has been suggested that this was associated with the introduction of an exogenous agent(s), specifically an infectious agent(s), in a genetically susceptible population by troops garrisoned in these regions during the second world war. This has subsequently been contested in all three areas.[5–7] Furthermore, it has been difficult to confirm these observations in larger populations.

**Correspondence: Peter T Donnan, Tayside Centre for General Practice, Kirsty Semple Way, University of Dundee, Dundee DD2 4BF, Scotland, UK.*
E-mail: p.t.donnan@dundee.ac.uk
Received 17 August 2004; accepted 17 February 2005

© 2005 Edward Arnold (Publishers) Ltd

Clustering of MS has been reported elsewhere in mainland Europe and America.[8–12] However, there is a tendency for many of these apparent 'clusters' to be identified using post hoc analyses. It is more rigorous to analyse a given area for clusters without *a priori* specification of their size and location. Several authors[13–16] have used this approach to identify clusters and have largely based their analysis on the space–time interaction method devised by Knox[17]. This method examines 'closeness' in space and time for pairs of cases compared with pairs of controls. However, dependence is introduced between two variables and consequently, the test statistic and significance value of any given cluster can only be approximated. Moreover, these methods do not take account of the multiple testing inherent in such a procedure and therefore, the possibility of eliciting a significant cluster by chance increases. We wished to look for clustering in a large MS population without *a priori* specification of the size or location (temporal or spatial) of potential clusters recorded in a population-based MS register covering the Tayside region of Scotland for the period 1970–1997.[18] We used the spatial scan statistic, which has been developed to detect and locate aggregations of leukaemia and breast cancer.[19–22] It uses a moving circular window of variable size, which scans the study area to detect clusters without prior knowledge of location or size. It also provides a robust statistical test of any clusters found. In addition, we used capture–recapture methods[23–25] to test the hypothesis that clusters may

10.1191/1352458505ms1191oa

be detected due to improved ascertainment either temporally or geographically.

Materials and methods

Study population

The study was made possible by accumulation of data on every potential case of MS in the Tayside region, Scotland from 1970 to 1997. The first survey was completed in 1996 and the register has been maintained prospectively since that date. The database contains, among other data, information on the date of symptom onset, date of diagnosis, age, gender, clinical information, addresses and postcodes. All patients gave informed consent to be included in the database. A diagnosis of MS was defined according to Poser's criteria.[26] Multiple sources (morbidity records, departmental records and surveys of general practitioners) were used to identify cases and therefore it was possible to estimate the completeness of ascertainment using capture–recapture methods.[24] Forbes *et al*. demonstrated ascertainment of between 93 and 99% using a two-source model in 1996 for the Tayside region.[18] For this analysis, a three-source capture–recapture method was used to test the ascertainment of 'cluster' and 'noncluster' cohorts. The most appropriate model, namely the one that sufficiently accounted for intersource dependencies, was chosen on the basis of goodness-of-fit as represented by the likelihood ratio statistic (G^2).[24]

The geographical area unit used in this study was the postcode sector. Each sector comprises aggregations of smaller area units, called postcodes, which are arbitrary boundaries derived for mail delivery. Postcodes contain approximately 10 houses and postcode sectors vary in population size from 500 to 10 000 individuals. There are 69 postcode sectors in Tayside, Scotland, and the number of cases of MS in each postcode sector was recorded in the database over the study period. The spatial location (postcode sector) of any given individual was recorded at the time the patient was entered on the register. The true geographical location of the postcode sector in terms of longitude and latitude was identified from digitized postcode information and presented as an X,Y co-ordinate for the sector centroid. In addition, the population of each postcode sector is known from census data from 1971, 1981 and 1991, giving accurate population data by age group and gender over the whole study period. Finally, a social deprivation score has been assigned to every postcode in Scotland based on the 1981 and 1991 censuses.[27] In detail, the score is composed of data relating to overcrowding, car ownership, male unemployment and social class of the head of household. This study was approved by the Tayside committee on medical ethics.

Statistical analysis

The algorithm for the spatial scan statistic of Kulldorff[20–22] was used to detect and test for statistical significance of clusters of MS. The geographic unit was the postcode sector and the method assumes a Poisson distribution for the counts of cases in each sector, allowing for population density stratified by age group (<30, 30–40, 40+) and gender. The spatial scan statistic is a generalization of a test first proposed by Turnbull.[28] A series of circles centred at each of the postcode sectors is generated. At each centre the radius increases continuously covering more and more of the adjacent sectors until 50% of the total population is covered, leading to an infinite number of circular zones. This means that before knowing the size of any cluster in the region we set an upper limit of 50% as a sensible maximum. The circles also extend in time giving a cylinder in space and time for each postcode sector. The null hypothesis is that there is complete spatial or temporal randomness, where the probability of falsely detecting a cluster anywhere in the region is equal to the 5% level of significance. The spatial scan statistic results in a likelihood ratio test, conditional on the total number of cases observed. The distribution of the test statistic was obtained from a series of 999 random Monte Carlo replicas of the data set, generated under the null hypothesis. The test statistic was calculated for each replica. The cluster detected is statistically significant at the 5% level if the value of the test statistic from the real data set is among the 5% highest of the 1000 values generated (i.e., in the top 50), including the 999 replicas. The *P*-values obtained for secondary clusters are conservative. A computer programme, SaTScan, is available from the authors.[29] In this study adjustment was made for age, gender, population size and social deprivation. Once significant clusters are detected, the maximum window size was sequentially reduced to assess whether there are subclusters which have enough strength on their own to reject the null hypothesis.[21]

Results

There were a total of 772 cases of clinically definite or probable MS identified in the region of Tayside between 1970 and 1997. The cases of MS were predominantly female (73.8%). The mean age of MS at symptom onset was 35.7 (SD = 10.5) with a range from 13.9 to 69.7 with no significant difference between males and females ($P = 0.66$). The mean age at symptom onset appeared to increase slightly with age 33.2 (SD 9.9) in 1970–1979, 34.9 (SD 10.0) in 1980–1989, and 38.1 (SD 11.0) in 1990–1997. The total number of postcode sectors used in this analysis was 69 covering a total population of 383 405. With 772 cases the disease incidence was 7.2 per 100 000/year (95% CI: 4.5, 9.9 per 100 000/year). The number of cases of MS by postcode sector is shown in Figure 1. The number of cases by year shows a general increasing trend with some indication of a peak around the mid-1990s (Figure 2). There is also some periodicity in the number of cases with peaks occurring every two to three years. A formal time series analysis demonstrated a significantly increasing trend over time and sinusoidal terms ($p < 0.1$) with a frequency of three years. A periodogram showed a peak between 2 and 3 years.

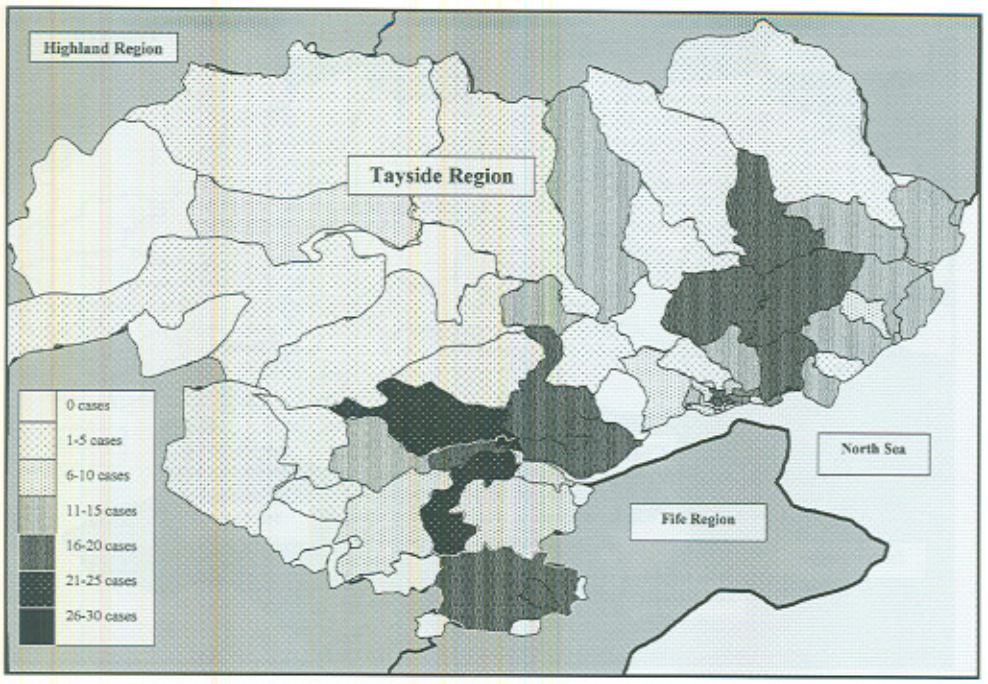

Figure 1 *Counts of cases of MS in the Tayside region of Scotland (1970–1997) by postcode sector.*

The spatial scan analysis identified two clusters. The most likely cluster was exclusively temporal, in that, for the whole of Tayside, there was a statistically significant cluster between 1982 and 1995 ($P=0.002$). The total number of cases was 467 during this time period with an annual incidence of 8.6 per 100 000/year (95% CI: 5.7, 11.5 per 100 000/year). There was also a significant secondary temporal–spatial cluster for the period 1993–1995 for the area to the south-west of Perth, centred in the rural area around Auchterarder (PH5 2), but also including central Perth ($P=0.016$), as shown in Figure 3. There were 42 cases in this area during 1993–1995 giving an annual incidence of 17.1 per 100 000/year. The overall relative risk (RR) was 2.37 (95% CI: 1.65, 3.09) for this area during this time period. When the maximum window was reduced in size there was still a highly significant purely temporal cluster for the whole of Tayside, which was centred around the period 1990/1994 when the peak number of cases occurred (Figure 2). In addition, the original spatial–temporal cluster remained significant for the same time period of 1993–1995, when both spatial and temporal window sizes were reduced to 20%.

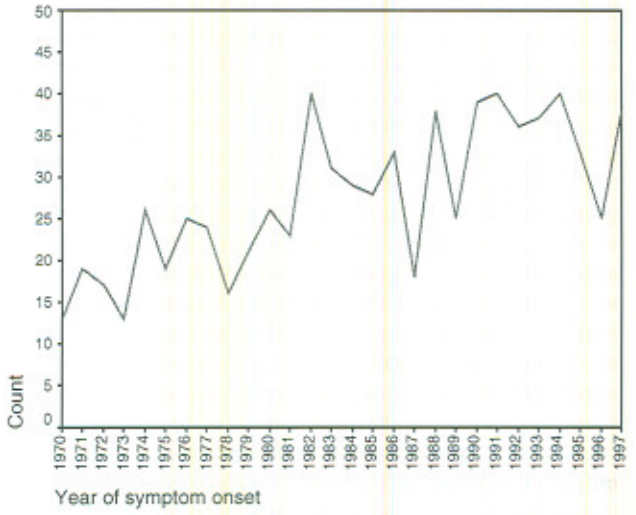

Figure 2 *Counts of cases of MS in the Tayside region of Scotland by year (1970–1997).*

Both the large temporal cluster and the temporal–spatial cluster in Perthshire were examined using capture–recapture methods to determine whether improved ascertainment during these periods might explain their presence. Using a three-source capture–recapture method, the number of missing cases for the whole of Tayside, during the temporal cluster from 1982 to 1995 was estimated to be 29 (95% CI: 20, 41). This model accounted for intersource dependency between morbidity data and general practitioner's (GP) records ($G^2=4.7$, 2 df, $P=0.095$). The number of missing cases for the remainder of the study period, namely 1970–1981 and 1995–1997 combined was estimated to be 28 (95% CI: 19, 40) accounting for interaction between morbidity and GP records ($G^2=8.9$, 2 df, $P=0.012$). For the Perthshire cluster from 1993 to 1995 the estimated number of missing cases was 3 (95% CI: 3, 10) with an independent model ($G^2=3.1$, 3 df, $P=0.38$). For the rest of Tayside during the same period the estimated missing number of cases was one (95% CI: 0, 8) using a model accounting for GP and morbidity data interaction ($G^2=4.5$, 2 df, $P=0.11$).

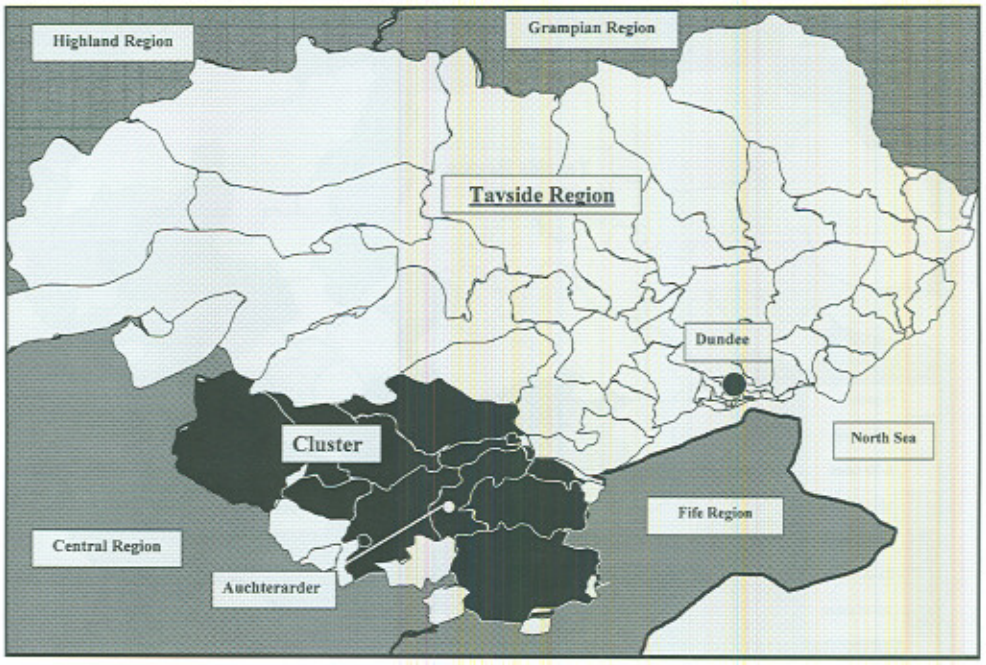

Figure 3 *Geographical location of the cluster detected in Perthshire, Scotland (1993–1995).*

Discussion

We have used the spatial scan statistic,[20–22] a robust method of identifying disease clusters, to demonstrate significant temporal and spatial–temporal aggregations of MS in the Tayside region of Scotland. This approach detects spatial–temporal clusters over and above that expected by chance alone. The clusters identified have a prominent temporal component and it is notable that for the period 1982–1995 the whole region demonstrated a significant increase in MS incidence. Furthermore, the area south-west of Perth city and Perth itself demonstrated a significantly higher number of cases for the period 1993–1995 yielding an annual incidence of 17.1 per 100 000. Increasing the 'resolution' of the analysis by reducing the potential size of the scanning window illustrated that this cluster was stable to a decrease in window size, remaining statistically significant.

There are several factors to consider when scrutinizing the reliability of these clusters. First, whether case ascertainment has been complete and appropriately performed. Secondly, whether confounding variables have been accounted for, and thirdly, whether other explanations of the results can be discounted.

We used capture–recapture methods to determine the ascertainment of 'cluster and noncluster' aggregations of MS. Ascertainment was, in fact, similar for the temporal period of clustering to the period outwith the cluster. Secondly, the estimate of missing cases for the temporal–spatial cluster was similar to that for the rest of the region outwith the geographic aggregation of MS identified. Hence, there was little evidence of differential case ascertainment between clusters and noncluster regions and time periods. It is possible that the upward trend could be partially explained by some deaths in the early part of the study period, which would not appear in the database. However, around 75% of patients remain alive at 20 years postclinical onset and hence this could only account for a few missing cases. For example, we estimate two missing cases in 1979 due to early death. Hence, this potential bias could not explain the peak found in the mid-1990s or the cyclical nature of the temporal pattern. All entries were classified according to Poser's criteria and thus there was uniformity of diagnostic categorization. We used only clinically definite and probable cases of MS to ensure accuracy of ascertainment.

Deprivation has been inversely correlated with MS previously, and furthermore, one would anticipate that this may also hold true for population density according to Alter's hypothesis.[30] Population densities stratified by age group, gender and deprivation status were included as confounding variables in the analysis and so all clusters were identified independently of these factors. The observed aggregations of MS could not be explained by occurrence in populations which were particularly susceptible, namely young female cohorts.

The temporal variable we used for any patient's given geographical position at the proposed onset of the disease was symptom onset. There are two problems with this approach. First, the geographical position for each patient was recorded when that individual was entered on the register. This may have been different from the address at symptom onset. Secondly, the use of symptom onset as a marker for the onset of MS may be inaccurate due to the potentially long, subclinical period of the disease. In this respect, we intend to repeat the analysis, once we have obtained residential histories for our cohort, using adolescence (aged 15) as the pathogenetic onset of the disease. Despite this, we are aware that migration rates for the Tayside region are some of the lowest in the UK (2.3% per

annum estimated for the 1991 census) and thus the potential error introduced by population movements is reduced.

The temporal peak for the whole of Tayside (1990–1995) occurred in recent times and so it is possible that the latency from symptom onset to diagnosis has prevented those cases for 1998 onwards from being registered. We therefore discounted cases from 1998 onwards. The mean lag from symptom onset to diagnosis for the whole cohort is 4.5 years (95% CI: 4.1, 5.0). We are reasonably certain that no cases prior to and including 1997 have been missed due to diagnostic lag. The slight increase in age at onset may be partly due to improved diagnostic techniques and hence detecting MS later in life, even though it was present but not detected when the patients were younger. It may also be due, in part, to the shift upwards of the age distribution of the population.

Magnetic resonance imaging scanning was introduced in Tayside in 1990 and a third neurologist was appointed in 1995. It is therefore possible that a surge in MS diagnoses may have occurred around this period with subsequent stabilization at a higher overall rate. However, this was not borne out by our examination of ascertainment and furthermore, the incidence of MS appears to be continuing to fall despite the availability of improved diagnostic techniques for a decade. Indeed there were only 23 new cases in 1998 and 18 in 1999 overall. In the spatial–temporal cluster, the number of new cases was 7 and 11 in 1998 and 1999 respectively. Furthermore, from Figure 1, it is apparent that the increase in MS incidence is cyclical, occurring approximately every three years, which is more consistent with an exogenous agent rather than service changes.

The cyclical nature and close temporal proximity of the peaks argues against a purely genetic component. Indeed, the population studied can be regarded as ancestrally fairly homogeneous. The Tayside region has one of the lowest migration rates in the UK, 99% of the population is Caucasian, 90% have a Scottish surname and 88.5% were born in Scotland (from the 1991 census). Swingler *et al*. demonstrated that in northern Scotland, the frequency of the HLA-DR2 genotype is equivalent amongst MS cases and healthy controls.[31] It was therefore suggested that in the Scottish population where the prevalence of susceptibility genes is likely to be generally high, other environmental factors become important in modulating the frequency of MS. Taken with the cyclical changes in incidence of MS in Tayside demonstrated in this study, we would suggest such factors are exogenous but the causal agent is unclear.

The spatial cluster detected occurred largely in a rural area which is 60 km from the primary centre in Dundee. It is therefore plausible that within this remoter area, population mixing is lower, resulting in an increased frequency of susceptibility genes. However, this cannot singularly explain why the cluster for this area occurred during a specific two-year period (see Figure 4) and we would again take this to indicate that environmental factors are salient.

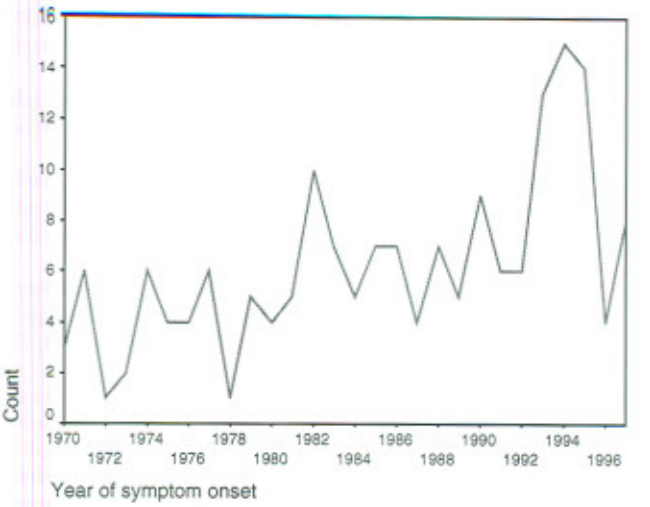

Figure 4 *Counts of cases for the area in which a temporal–spatial cluster was detected (1970–1997).*

Scotland, situated at latitude 56 N, has perhaps the highest prevalence of MS in the world. The point prevalence for clinically probable and definite MS in Tayside region on the 31 January 2002 was 236 per 100 000 (95% CI: 221, 251). Our study indicates that within this area there is clear temporal and geographical variation of prevalence which cannot be explained by overascertainment of clusters or confounding. The cyclical nature of these changes leads us to postulate that infectious agents may be of importance in determining the frequency of MS in this genetically susceptible population but no particular agent has been identified. Recent evidence has suggested a link between antibodies to Epstein–Barr virus and increased risk of development of MS.[32,33] Furthermore, we would suggest use of the spatial scan statistic combined with capture–recapture methods for future examinations of MS clustering.

Acknowledgements

We wish to acknowledge the help and advice of Dr. Simon Ogston, Department of Epidemiology and Public Health, Ninewells Hospital, Dundee with the capture–recapture analysis.

References

1 Kurtzke JF. The epidemiology of multiple sclerosis. In Raine CS, McFarland HF, Tourtellotte WW eds. *Multiple sclerosis. Clinical and pathogenetic basis*, first edition. London: Chapman and Hall, 1997, 105–107.

2 Kurtzke JF, Hyllested K. Multiple sclerosis in the Faroe Islands: I. Clinical and epidemiological features. *Ann Neurol* 1979; **5**: 6–21.

3 Poskanzer DC, Prenney LB, Sheridan JL, Kondy JY. Multiple sclerosis in the Orkney and Shetland Islands. I: Epidemiology, clinical factors, and methodology. *J Epidemiol Community Health* 1980; **34**: 229–39.

4 Kurtzke JF, Gudmundsson KR, Bergmann S. Multiple sclerosis in Iceland: 1. Evidence of a postwar epidemic. *Neurology* 1982; **32**: 143–50.
5 Benedikz J, Magnusson H, Guthmundsson G. Multiple sclerosis in Iceland, with observations on the alleged epidemic in the Faroe Islands. *Ann Neurol* 1994; **36**: S175–79.
6 Poser CM, Hibberd PL, Benedikz J, Gudmundsson G. Analysis of the 'epidemic' of multiple sclerosis in the Faroe Islands. I. Clinical and epidemiological aspects. *Neuroepidemiology* 1988; **7**: 168–80.
7 Rothwell PM, Charlton D. High incidence and prevalence of multiple sclerosis in south east Scotland: evidence of a genetic predisposition. *J Neurol Neurosurg Psychiatry* 1998; **64**: 730–35.
8 Binzer M, Forsgren L, Holmgren G, Drugge U, Fredrikson S. Familial clustering of multiple sclerosis in a northern Swedish rural district. *J Neurol Neurosurg Psychiatry* 1994; **57**: 497–99.
9 Eastman R, Sheridan J, Poskanzer DC. Multiple sclerosis clustering in a small Massachusetts community with possible common exposure 23 years before onset. *N Engl J Med* 1973; **10**: 793–94.
10 Jaworski MD, Slater JD, Severini A, Henning KR, Mansour G, Mehta JG *et al*. Unusual clustering of diseases in a Canadian Old Colony (Chortitza) Mennonite kindred and community. *Can Med Assoc J* 1988; **138**: 1017–25.
11 McDougall EM. Multiple sclerosis in research workers studying swayback lambs: an updated report. *J Neurol Neurosurg Psychiatry* 1985; **48**: 859–65.
12 Stein EC, Schiffer RB, Hall WJ, Young N. Multiple Sclerosis and the workplace: report of an industry based cluster. *Neurology* 1987; **37**: 1672–77.
13 Larsen JP, Riise T, Nyland H Kvale G, Aarli JA. Clustering of multiple sclerosis in the county of Hordaland, Western Norway. *Acta Neurol Scand* 1985; **71**: 390–95.
14 Neutel CI, Walter SD, Mousseau G. Clustering during childhood of multiple sclerosis patients. *J Chronic Dis* 1977; **30**: 217–24.
15 Poskanzer DC, Walker AM, Prenney LB, Sheridan JL. The etiology of multiple sclerosis: temporal–spatial clustering indicating two environmental exposures before onset. *Neurology* 1981; **31**: 708–13.
16 Riise T, Gronning M, Klauber MR, Barrett-Connor E, Nyland H, Albrektsen G. Clustering of residence of multiple sclerosis patients at age 13 to 20 years in Hordaland, Norway. *Am J Epidemiol* 1991; **133**: 932–39.
17 Knox G. The detection of space–time interaction. *Appl Stat* 1964; **13**: 25–29.
18 Forbes RB, Wilson SV, Swingler RJ. The prevalence of multiple sclerosis in Tayside, Scotland: do latitudinal gradients really exist? *J Neurol* 1999; **246**: 1033–40.
19 Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Stat Med* 1995; **14**: 799–810.
20 Hjalmars U, Kulldorff M, Gustafsson G, Nagarwalla N. Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Stat Med* 1996; **15**: 707–15.
21 Kulldorff M, Feuer EJ, Miller BA, Freedman LS. Breast cancer in northeastern United States: A geographical analysis. *Am J Epidemiol* 1997; **146**: 161–70.
22 Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods* 1997; **26**: 1481–96.
23 Forbes RB. *Capture–recapture methods in epidemiological surveys*. University of Dundee, Dundee, 1999.
24 Hook EB, Regal RR. Capture–recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* 1995; **17**: 243–64.
25 Forbes RB, Swingler RJ. Estimating the prevalence of multiple sclerosis in the United Kingdom by using capture–recapture methodology. *Am J Epidemiol* 1999; **149**: 1016–24.
26 Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC *et al*. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983; **13**: 227–31.
27 McLoone P. *Carstairs codes for Scottish postcodes from the 1991 census*. Glasgow: Public Health Research Unit, University of Glasgow, 1991.
28 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**: S136–43.
29 Kulldorff M, Rand K, Gherman G, Williams G, DeFranscesco D. *SaTScan v2.1: software for the spatial and space–time scan statistics*. Bethesda, MD: National Cancer Institute, 1998.
30 Alter M, Zhang ZX, Davanipour Z, Sabel E, Zibulewski J, Schwartz G *et al*. Multiple sclerosis and childhood infections. *Neurology* 1986; **36**: 1386–89.
31 Swingler RJ, Compston D. The distribution of multiple sclerosis in the United Kingdom. *J Neurol Neurosurg Psychiatry* 1986; **49**: 1115–24.
32 Ascherio A, Munger KL, Lennette ET, Spiegelman D, Hernan MA, Olek MJ *et al*. Epstein–Barr virus antibodies and risk of multiple sclerosis: a prospective study. *JAMA* 2001; **286**: 3083–88.
33 Levin LI, Munger KL, Rubertone MV, Peck CA, Lennette ET, Spiegelman D *et al*. Multiple sclerosis and Epstein–Barr virus. *JAMA* 2003; **289**: 1533–36.