

# An assessment of spatial clustering of leukaemias and lymphomas among young people in New Zealand

John D Dockerty, Katrina J Sharples, Barry Borman

## Abstract

**Study objective**—To assess spatial clustering of childhood leukaemias and lymphomas in New Zealand, using a national dataset from a country with no nuclear installations.

**Design**—New Zealand Map Grid coordinates, derived from the birth addresses of cases and controls were used in clustering analyses that applied Cuzick and Edwards' method.

**Setting**—The whole of New Zealand.

**Participants**—The cases were ascertained from the New Zealand Cancer Registry. They were diagnosed with leukaemia or lymphoma at ages 0-14 years during the period 1976 to 1987. For Hodgkin's disease, the age range was extended to include those aged from 0-24 years. The cancer registrations were linked with national birth records, to obtain the birth addresses of the cases. The controls were selected at random from birth records, with matching to cases (1:1) on age and sex. The analyses included 600 cases and 600 controls.

**Main results**—There was no statistically significant spatial clustering for any tumour group overall, including acute lymphoblastic leukaemia, acute non-lymphoblastic leukaemia, other leukaemias, non-Hodgkin's lymphomas, Hodgkin's disease, and all these combined. Significant clustering was found in a sub-analysis for one of three age specific subgroups of acute lymphoblastic leukaemia (ages 10-14 years,  $p=0.003$ ).

**Conclusion**—The subgroup finding may have been real or a chance association, as several comparisons were made. This study found little evidence for spatial clustering of leukaemias or lymphomas in a population with no nuclear installations.

(J Epidemiol Community Health 1999;53:154-158)

In 1937, Kellett<sup>1</sup> identified what appeared to be a cluster of leukaemia among children and adults near Ashington, a small mining town in northern England. Periodically, there have been other reports of an increased incidence of childhood leukaemia in a particular area and time, for example among the residents of a small town or even one street.<sup>2,3</sup> Much attention has been focused on reports of clusters of childhood leukaemia around nuclear reprocessing plants in the United Kingdom.<sup>4</sup> Post hoc reports such as these are very difficult to interpret, because clusters will occur occasionally in populations as

"random" statistical events.<sup>2</sup> Rothman<sup>5</sup> argued that the post hoc study of an individual cluster of a disease was almost always unhelpful in understanding the aetiology of the disease, although there are a few notable exceptions. So far, the clusters of childhood leukaemia that have been studied have not resulted in the identification of new causal risk factors.

The types of studies that are needed to assess whether childhood leukaemia occurs in clusters are large studies with complete case ascertainment in defined areas where clustering has not already been studied.<sup>2</sup> Since the 1960s, at least 32 datasets from around the world have been studied to assess whether there is spatial or space-time clustering of childhood leukaemias.<sup>2-6-14</sup> The studies have mostly been small, and have used a variety of methods. Tables of the results of the various studies in review articles and the results of new studies<sup>2-6-14</sup> show that of the 32 datasets assessed, 15 showed some statistically significant evidence of clustering of childhood leukaemia, 12 showed no evidence of clustering, and five showed possible clustering limited to one or several subgroups. Bithell and Draper<sup>15</sup> summed up the studies of clustering of childhood leukaemia by saying that "In general, the accumulated evidence can fairly be described as weak, whether it is addressing a generalised tendency to case aggregation or possible proximity to specific risk sources."

Nevertheless, the apparent excesses of childhood leukaemia near nuclear reprocessing plants in Britain have caused great concern and generated much research.<sup>16,17</sup> The causes of these clusters remain unknown.<sup>17</sup> This study was conducted to find out whether childhood leukaemia clusters could be found in a country with no nuclear establishments (New Zealand). Childhood leukaemia clustering in New Zealand would require a non-nuclear explanation; such as another possible risk factor with a localised source; the person to person spread of an infectious agent; or an artefact caused by bias, confounding, or chance variation.<sup>2,18,19</sup>

It has been over 25 years since a study was conducted to assess childhood leukaemia clustering in New Zealand.<sup>20</sup> The present study was conducted using a new dataset, and with a method that avoids problems related to population changes of the type that hindered interpretation of the previous study.<sup>20,21</sup> The previous New Zealand study focused on childhood leukaemias, but lymphomas are also of interest,<sup>22,23</sup> and both tumour groups were included in this study.

Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand  
J D Dockerty  
K J Sharples

Health Funding Authority, P O Box 10097, Wellington, New Zealand  
B Borman

Correspondence to:  
Dr J Dockerty, Childhood Cancer Research Group, University of Oxford, 57 Woodstock Road, Oxford OX2 6HJ.

Accepted for publication 6 October 1998

## Methods

The study involved obtaining the residential addresses at birth of cases and controls, for an analysis using the method of Cuzick and Edwards.<sup>21</sup> Addresses at the time of birth (rather than cancer registration) were used because of the marked early peak in the age distribution of the incidence of childhood leukaemia in New Zealand. The peak for acute lymphoblastic leukaemia occurs at ages 2–3 years.<sup>24</sup> Hence, exposure to relevant aetiological factors for leukaemia would often have been in utero or in early childhood.

Ethical approval for this study was granted by the ethical committee of the Otago Area Health Board. The cases were ascertained from the New Zealand Cancer Registry. The study included nationwide leukaemia and lymphoma registrations, for the period 1976–87. The age range was 0–14 years for acute lymphoblastic leukaemia (ALL), acute non-lymphoblastic leukaemia (ANLL), other and unspecified leukaemias, and non-Hodgkin's lymphomas (NHLs). For Hodgkin's disease, the range was 0–24 years, because of the older age distribution, and the spatial clustering found for that age group elsewhere.<sup>25</sup> For each cancer registration, the dataset included the full name, sex, date of birth, date of diagnosis/admission, International Classification of Diseases (ICD) site code, and ICD-Oncology code.

The Registrar General permitted us to have access to national birth records. Cancer registrations were linked manually with these. Adopted people were excluded from the study because the law did not permit us to access their original birth records.

A single individually matched control was selected for each case, from other births registered at the Registrar General's Office. The records of all children born in New Zealand are stored in quarterly sets for each year (for example, January to March 1980). The quarters relate to the dates of registration, not birth. It is a legal requirement for births to be registered within two months. Within quarters, the birth records are numbered sequentially. The process for control selection was as follows. The date on which the birth of each case child was registered was obtained from his or her birth record. The matched control was selected from a list of possible controls, produced by a computer program. The program randomly generated the (folio) numbers of the birth records of 12 children whose births were registered in the same quarter and year as the case. The birth record of the first of the 12 was then checked, for matching on sex and confirmation of eligibility. The other eligibility criteria were that the record had to be for the initial registration of the birth (not a re-registration), and that it could not be for a stillborn or adopted child. The clerical worker would continue through the list of the 12 possible controls (which were ordered randomly), until an eligible match for the case child was obtained. In this way, the control for each case was matched on quarter (and year) of birth registration, and on sex. Very occasionally, none of the 12 possible controls could be matched to

the case. In such a situation, a second list of 12 was produced, and a match was found.

The geographical data used were the addresses from birth records, normally given as street addresses. Each address was assigned to its meshblock (a small geographical unit) by the Department of Statistics. At the time of this study, there were 35 152 meshblocks in New Zealand. For 93% of the cases and controls, the meshblock was assigned precisely, using the exact address. For 7%, the meshblock was assigned on the basis of less detailed address information (for example, the central point of a given town or location was identified, and assigned to its meshblock). For each case and control, the centroid of the meshblock was assigned to its two seven digit New Zealand Map Grid coordinates. The map grid coordinates were used in the analyses.

Cuzick and Edwards' test for spatial clustering is a nearest neighbour test. It is based on the locations of  $n_0$  cases and  $n_1$  (randomly selected) controls from a specified region.<sup>21</sup> The test counts the number of cases among the  $k$  nearest neighbours of each case as follows.<sup>21</sup> Let  $(z_1, \dots, z_n; n=n_0+n_1)$  be the locations of both the cases and the controls in random order. For  $i=1$  to  $n$ , define:

$$\delta_i = \begin{cases} 1 & \text{if } z_i \text{ is a case} \\ 0 & \text{if } z_i \text{ is a control} \end{cases}$$

and  $d_i^k$  = the number of  $k$  nearest neighbours to  $z_i$  that are cases.<sup>21</sup> The test statistic for the  $k$  nearest neighbours ( $T_k$ ) is<sup>21</sup>

$$T_k = \sum_{i=1}^n \delta_i d_i^k$$

The approach has been extended by Jacquez to allow upper and lower bound test statistics to be calculated when ties exist in the data.<sup>26</sup> In this study, a tie could occur when two different street addresses were located in the same meshblock. Both addresses would be assigned to the same meshblock centroid, even though their exact locations were different. Jacquez' extension of Cuzick and Edwards' method has been implemented in the package Stat!,<sup>27</sup> which was used for these analyses. Details of the distribution of the  $T_k$  statistics are given by Jacquez.<sup>26</sup> The choice of the optimal  $T_k$  statistic is difficult, as it is dependent on the setting. There is no clear information available about which value of  $k$  is best in this New Zealand setting. We assessed  $p$  values for several  $T_k$  statistics ( $k=1-10$ , as produced by Stat!). Stat! also produces Simes  $p$  values, to allow consideration of the effects of multiple testing.

Analyses were conducted for the following groups: ALL (ages 0–4, 5–9, 10–14, and 0–14 years); ANLL (ages 0–14); other leukaemias (ages 0–14); Hodgkin's disease (ages 0–14, 15–24, and 0–24); NHL (ages 0–14); and the whole sample (combined leukaemias and lymphomas in children aged 0–14 plus Hodgkin's disease aged 15–24 years). Age specific analyses were conducted for ALL (which had the largest numbers) because the incidence of that cancer increased significantly during the period

Table 1 Numbers of cases and controls in the analyses for each diagnostic group

Diagnostic group	Age range (y)	Number of cases	Number of controls
ALL	0-4	162	162
	5-9	62	62
	10-14	52	52
	0-14	276	276
ANLL	0-14	71	71
	0-14	17	17
"Other" leukaemias	0-14	34	34
	15-24	124	124
Hodgkin's	0-24	158	158
	0-14	78	78
NHL	0-14 or 0-24	600	600

ALL = acute lymphoblastic leukaemia, ANLL = acute non-lymphoblastic leukaemia, NHL = non-Hodgkin's lymphoma.

from 1973-77 to 1988-90 among children aged 0-4 years in New Zealand, but not among those aged 5-9 or 10-14 years.<sup>24</sup> Hodgkin's disease was separated into children (0-14 years) and young adults (15-24 years). The ages were calculated using the cases' birth dates (from their birth certificates) and dates of diagnosis/admission (from their cancer registrations). For each group defined by diagnosis and age, only the matched controls for the cases were included.

**Results**

There were 748 leukaemia and lymphoma registrations (1976-87) in the study. Birth records

were found for 630 (84%). Seventy six of the 118 unlinked cancer registrations gave a country of birth other than New Zealand.

Twenty seven of the 630 birth records were for adopted children, who were excluded, leaving 603. Three of the cases' birth records did not give the address at registration. Thus, exactly 600 cases and 600 matched controls could be included in the analyses. Table 1 shows the numbers in each diagnostic group.

Table 2 shows the main results of the clustering analyses. There was no statistically significant spatial clustering for any tumour group overall (including ALL ages 0-14, ANLL ages 0-14, other leukaemias ages 0-14, Hodgkin's disease ages 0-24, NHL ages 0-14, and all these combined). There was significant clustering for one of the three age specific subgroups of ALL (ages 10-14 years). Statistical significance was achieved for each of the 10  $T_k$  statistics for this age group, and the Simes p value was 0.003 (table 2). There was no significant clustering for either subgroup of Hodgkin's disease (ages 0-14 and 15-24 years).

Table 3 gives further details of the results for spatial clustering of acute lymphoblastic leukaemia at ages 10-14 years, with observed and expected values of the  $T_k$ .

Table 2 Results of the analyses for spatial clustering of leukaemias and lymphomas among young people in New Zealand

Diagnostic group, age range	p value for $T_k$ statistic										Simes p value	
	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$	$T_7$	$T_8$	$T_9$	$T_{10}$		
ALL												
0-4												
Lower bound*	0.26	0.35	0.49	0.49	0.43	0.35	0.41	0.48	0.45	0.50	0.50	
Upper bound	0.26	0.35	0.49	0.49	0.43	0.35	0.41	0.47	0.45	0.50	0.50	
5-9												
No ties	0.66	0.28	0.19	0.50	0.57	0.52	0.63	0.68	0.64	0.67	0.67	
10-14												
No ties	0.051	0.019	0.003	0.002	0.006	0.061	0.041	0.005	0.001	0.011	0.003	
0-14												
Lower bound	0.41	0.38	0.43	0.42	0.41	0.28	0.32	0.38	0.38	0.36	0.36	
Upper bound	0.38	0.37	0.43	0.41	0.41	0.27	0.32	0.38	0.38	0.35	0.35	
ANLL												
0-14												
No ties	0.20	0.65	0.83	0.92	0.86	0.67	0.49	0.33	0.41	0.40	0.40	
Other leukaemias												
0-14												
No ties	0.36	0.32	0.28	0.41	0.52	0.40	0.55	0.81	0.77	0.58	0.58	
Hodgkin's												
0-14												
No ties	0.97	0.89	0.78	0.56	0.73	0.70	0.47	0.72	0.78	0.61	0.61	
15-24												
Lower bound	0.74	0.84	0.58	0.73	0.78	0.65	0.76	0.87	0.92	0.86	0.86	
Upper bound	0.74	0.81	0.55	0.70	0.76	0.65	0.73	0.86	0.91	0.85	0.85	
0-24												
Lower bound	0.84	0.75	0.74	0.75	0.74	0.68	0.68	0.70	0.74	0.72	0.72	
Upper bound	0.84	0.73	0.73	0.74	0.73	0.68	0.67	0.70	0.73	0.72	0.72	
NHL												
0-14												
Lower bound	0.89	0.86	0.88	0.84	0.75	0.82	0.79	0.80	0.78	0.94	0.94	
Upper bound	0.89	0.86	0.86	0.78	0.75	0.82	0.79	0.80	0.78	0.94	0.94	
Total												
0-14 or 0-24												
Lower bound	0.62	0.69	0.64	0.60	0.53	0.46	0.49	0.51	0.49	0.46	0.46	
Upper bound	0.44	0.61	0.55	0.57	0.49	0.44	0.46	0.49	0.45	0.43	0.43	

\* When ties were present, the software (Stat!) gave p values for the lower and upper bound test statistics.

Table 3 Detailed results from Cuzick and Edwards' test of spatial clustering: acute lymphoblastic leukaemia, ages 10-14 years

Test	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$	$T_6$	$T_7$	$T_8$	$T_9$	$T_{10}$
Observed value	32	63	97	127	153	171	201	240	274	292
Expected value	25.75	51.50	77.24	102.99	128.74	154.49	180.23	205.98	231.73	257.48
Variance	14.57	30.39	49.96	71.65	92.96	113.90	142.06	173.97	199.63	226.78
p Value	0.051	0.019	0.003	0.002	0.006	0.061	0.041	0.005	0.001	0.011

### Discussion

This study of spatial clustering of leukaemias and lymphomas in young people was conducted using data from New Zealand. The completeness of case ascertainment is likely to have been high during the period assessed (1976–87).<sup>24–28</sup> Cancer registrations from the whole of New Zealand were linked satisfactorily with national birth records. A population-based control group was selected, and Cuzick and Edwards' method was used. Participation bias was avoided because this was a records-based study. The method of Cuzick and Edwards is particularly suitable for use in populations that have an uneven geographical distribution,<sup>21</sup> as in New Zealand.

Many studies of spatial or space-time clustering and childhood leukaemia have been criticised for methodological difficulties or problems. These have included low statistical power and a lack of sub-type specificity for the diagnoses studied.<sup>6</sup> Fraumeni and Miller<sup>18</sup> noted that leukaemia clusters may have been found because of bias and error, including "random" variation, the practice of defining geographical boundaries after a cluster had already been found in the region, improved case ascertainment in areas suspected of having clusters, and factors to do with population density, migration or age distribution. Several researchers have conducted multiple comparisons with different combinations of distances and time periods.<sup>2</sup> False positives are a particular concern when the same data are tested for clustering using different methods, or when the same region is tested repeatedly for many different short time periods,<sup>29</sup> as has happened in several of the leukaemia studies. Confounding (by factors such as population density, age, sex, and ethnicity) has been described as the biggest problem in studies of clustering.<sup>19</sup>

In this study, we avoided some, but not all, of the problems discussed above. The controls were matched to the cases on age and sex, to reduce potential confounding. Large geographical variations in the age and sex distributions could affect the distribution of the  $T_k$  statistics. However, most of the *p* values were very large (table 2), so it seems unlikely that the regional differences in the age and sex distributions could explain the findings. Limited information was available from birth records, and it was not practical to match on factors other than age and sex. For this reason, there could be residual confounding because of uncontrolled factors such as social class and ethnic group. The numbers were small for some analyses (table 1), even though this was a national study including data from cancer registrations over a 12 year period. Thus, some clusters might not be detectable with our data and the method of Cuzick and Edwards. On the other hand, several comparisons were made (for groups defined by diagnosis and age), and this would increase the likelihood of a chance association.

We did not find spatial clustering for ALL at ages 0–14, ANLL at ages 0–14, other leukaemias at ages 0–14, Hodgkin's disease at ages 0–24, NHL at ages 0–14, or all these combined. Spatial clustering was, however,

#### KEY POINTS

- In New Zealand, spatial clustering of leukaemias or lymphomas at birth was not found overall for children aged 0–14 years.
- Significant clustering was found in a subgroup (acute lymphoblastic leukaemia at ages 10–14 years).
- This could be attributable to chance variation, a previously suggested infectious aetiology, or an unknown localised risk factor.

demonstrated in one of the three subgroup analyses undertaken for acute lymphoblastic leukaemia (that for children aged from 10 to 14 years (table 2)). This finding could be aetiologicaly relevant, or it could be caused by chance variation.

In a previous New Zealand study, Glass and colleagues (1971) assessed space-time clustering of childhood leukaemia.<sup>20</sup> This included 288 children aged 0–14 who died from leukaemia during 1953–64. The addresses at which the children lived at the time of disease onset were plotted on a map, and inter-point distances were calculated. Space-time clustering was assessed using two different methods.<sup>30–31</sup> The data were divided into 12 subgroups based on age and year of onset, and clustering was assessed for pairs of cases within different distances and time periods. The authors found statistically significant clustering of childhood leukaemia, especially at distances of less than two miles and times of under three months apart. The clustering was statistically significant for children aged under 6 years at diagnosis,<sup>20</sup> in contrast with the findings of the present study. Glass and colleagues' findings were not the same for different six year periods of the study. They concluded that rapid overall population growth, especially in urban areas, could have produced the significant space-time clustering they found.<sup>20</sup> Such an explanation cannot be invoked for the subgroup finding in our study, because Cuzick and Edwards' method should be free from problems related to the population density.<sup>21–32</sup>

A study such as ours would not be expected to be able to detect all types of clustering. Statistical methods will differ in their ability to detect particular types. Cuzick and Edwards' method might not be powerful if space-time clusters were present and existed for very short time periods. When this study was planned in 1990, there was much interest in spatial clustering, because of the childhood leukaemia clusters demonstrated near nuclear reprocessing plants in Britain.<sup>4</sup> Cuzick and Edwards' method has been shown to perform reasonably well in comparison with other methods for detecting spatial clusters.<sup>32–33</sup> Very little is known about the aetiology of childhood leukaemias, and it is not possible to predict what type(s) of clusters (including shape, size, and duration) might occur in New Zealand (if any). Although Cuzick and Edwards' method was (in our view) the method of choice for our situation, there is limited information about the power of the method for detecting different

types of clusters in the New Zealand setting. For the smaller subgroups, power is likely to be limited, particularly as smaller clusters seem to be harder to detect.<sup>32</sup>

If the clustering of acute lymphoblastic leukaemia for ages 10–14 is not because of chance, bias or confounding, it could suggest a localised environmental risk factor or the person to person spread of an infectious agent. Several factors determine whether an infectious disease will maintain itself in a population, including the density (in the population) of susceptibles and the frequency of contacts between individuals.<sup>34</sup> These factors vary greatly in different localities, depending on the environmental and social conditions.<sup>34</sup> Community size and population density are important in determining whether the infectious disease will maintain itself within the population.<sup>34</sup> Compared with European countries, New Zealand is sparsely populated. Differences between New Zealand communities and those in other countries could produce differences in the transmission of a suggested infectious risk factor. Studies of clustering in New Zealand might therefore give findings that would differ from those in other countries. Similarly, other relevant (but unidentified) non-nuclear environmental exposures that might have localised origins could differ in different settings.

In conclusion, we found no evidence of spatial clustering of leukaemias or lymphomas among young people in combined age groups, but some evidence of spatial clustering of ALL in a subgroup aged from 10–14 years.

The cancer registration data were provided by Health Statistical Services. Mr Brian Clarke (the Registrar General) kindly permitted us to have access to records at his office in Lower Hutt. We are grateful to Mr Tim Davey, Mr Nicholas Gorman, Mr David Methven, Ms Michaela Stevens, and Mr Stefan White for their careful work with record linkage and birth certificate finding. The meshblock centroids were assigned to the street addresses by the Department of Statistics. Professor Mark Elwood, Professor David Skegg, and Associate Professor Brian Heenan gave helpful advice during the planning and setting up of this study. Professor Sarah Darby gave encouragement and advice during the planning and early conduct of this work. Dr Rob Edwards kindly shared information, advice, and computer programs.

Funding: Cancer Society of New Zealand.  
Conflicts of interest: none.

- 1 Kellett CE. Acute myeloid leukaemia in one of identical twins. *Arch Dis Child* 1937;12:239–52.
- 2 Smith PG. Spatial and temporal clustering. In: Schottenfeld D, Fraumeni JF Jr, eds. *Cancer epidemiology and prevention*. Philadelphia: Saunders, 1982;391–407.
- 3 Hyams J. Leukaemia and the cluster enigma: The towns united by fear. *Australian Women's Weekly* 1988;Oct:44–6.
- 4 Beral V, Roman E, Bobrow M, eds. *Childhood cancer and nuclear installations*. London: BMJ Publishing Group, 1993.
- 5 Rothman KJ. A sobering start for the cluster busters' conference. *Am J Epidemiol* 1990;132 (suppl 1):S6–13.
- 6 Linet MS. *The leukemias: epidemiologic aspects*. New York: Oxford University Press, 1985.

- 7 Alexander FE. Viruses, clusters and clustering of childhood leukaemia: a new perspective? *Eur J Cancer* 1993;29A:1424–43.
- 8 Gilman EA, Knox EG. Childhood cancers: space-time distribution in Britain. *J Epidemiol Community Health* 1995;49:158–63.
- 9 Hjalmar U, Kulldorff M, Gustafsson G, et al. Childhood leukaemia in Sweden: using GIS and a spatial scan-statistic for cluster detection. *Statist Med* 1996;15:707–15.
- 10 Muirhead CR. Childhood leukemia in metropolitan regions in the United States: a possible relation to population density? *Cancer Causes Control* 1995;6:383–8.
- 11 Knox EG, Gilman EA. Spatial clustering of childhood cancers in Great Britain. *J Epidemiol Community Health* 1996;50:313–19.
- 12 Alexander FE, Chan LC, Lam TH, et al. Clustering of childhood leukaemia in Hong Kong: association with the childhood peak and common acute lymphoblastic leukaemia and with population mixing. *Br J Cancer* 1997;75:457–63.
- 13 Petridou E, Revinthi K, Alexander FE, et al. Space-time clustering of childhood leukaemia in Greece: evidence supporting a viral aetiology. *Br J Cancer* 1996;73:1278–83.
- 14 Alexander FE, Boyle P, Carli P, et al. Spatial clustering of childhood leukaemia: summary results from the EURO-CLUS project. *Br J Cancer* 1998;77:818–24.
- 15 Bithell JF, Draper GJ. Apparent association between benzene and childhood leukaemia: methodological doubts concerning a report by Knox. *J Epidemiol Community Health* 1995;49:437–9.
- 16 Wakeford R, Binks K, Wilkie D. Childhood leukaemia and nuclear installations. *J R Stat Soc A* 1989;152:61–86.
- 17 Doll R, Evans HJ, Darby SC. Paternal exposure not to blame. *Nature* 1994;367:678–80.
- 18 Fraumeni JF Jr, Miller RW. Epidemiology of human leukemia: Recent observations. *J Natl Cancer Inst* 1967;38:593–605.
- 19 Wartenberg D, Greenberg M. Solving the cluster puzzle: clues to follow and pitfalls to avoid. *Statist Med* 1993;12:1763–70.
- 20 Glass AG, Mantel N, Gunz FW, et al. Time-space clustering of childhood leukemia in New Zealand. *J Natl Cancer Inst* 1971;47:329–36.
- 21 Cuzick J, Edwards R. Spatial clustering for inhomogeneous populations. *J R Stat Soc B* 1990;52:73–104.
- 22 Draper GJ, Stiller CA, O'Connor CM, et al. Introduction and objectives. In: Draper G, ed. *The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966–83*. London: HMSO, 1991:1–6.
- 23 Alexander FE, Daniel CP, Armstrong AA, et al. Case clustering, Epstein-Barr virus Reed-Sternberg cell status and herpes virus serology in Hodgkin's disease: results of a case-control study. *Eur J Cancer* 1995;31A:1479–86.
- 24 Dockerty JD, Cox B, Cockburn MG. Childhood leukaemias in New Zealand: time trends and ethnic differences. *Br J Cancer* 1996;73:1141–7.
- 25 Alexander FE, Ricketts TJ, McKinney PA, et al. Community lifestyle characteristics and incidence of Hodgkin's disease in young people. *Int J Cancer* 1991;48:10–14.
- 26 Jacquez GM. Cuzick and Edwards' test when exact locations are unknown. *Am J Epidemiol* 1994;140:58–64.
- 27 Jacquez GM. *Stat! Statistical software for the clustering of health events*. Ann Arbor, MI: BioMedware, 1994; (computer program).
- 28 Dockerty JD, Becroft DMO, Lewis ME, et al. The accuracy and completeness of childhood cancer registration in New Zealand. *Cancer Causes Control* 1997;8:857–64.
- 29 Wartenberg D, Greenberg M. Detecting disease clusters: The importance of statistical power. *Am J Epidemiol* 1990;132 (suppl 1):S156–66.
- 30 Knox G. Epidemiology of childhood leukaemia in Northumberland and Durham. *Br J Prev Soc Med* 1964;18:17–24.
- 31 Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res* 1967;27:209–20.
- 32 Alexander FE, Boyle P, eds. *Methods for investigating localized clustering of disease*. IARC Scientific Publication No 135. Lyon: International Agency for Research on Cancer, 1996: 1–247.
- 33 Oden N, Jacquez G, Grimson R. Realistic power simulations compare point- and area-based disease cluster tests. *Stat Med* 1996;15:783–806.
- 34 Anderson RM, May RM. Directly transmitted infectious diseases: control by vaccination. *Science* 1982;215:1053–60.