

Space-time clustering of childhood cancer in Great Britain: A national study, 1969–1993

Richard J.Q. McNally^{1*}, Freda E. Alexander² and John F. Bithell³

¹School of Clinical Medical Sciences (Child Health) and School of Population and Health Sciences, Sir James Spence Institute, Royal Victoria Infirmary, Newcastle, United Kingdom

²Department of Public Health Sciences, The University of Edinburgh Medical School, Teviot Place, Edinburgh, United Kingdom

³Department of Paediatrics, Childhood Cancer Research Group, University of Oxford, Oxford, United Kingdom

Previous studies have provided evidence that infections may play a part in the aetiology of certain childhood cancers. The finding of space-time clustering indicates the presence of an environmental component to aetiology and is especially supportive of a role for infections. Space-time clustering occurs when excess numbers of cases of a disease are observed within small geographical locations at limited periods of time and this cannot be explained in terms of general excesses in those locations or at those times. To investigate whether infections may be involved in the aetiology of childhood cancer, we have analysed for space-time clustering using a large set of national population-based data from Great Britain for the period 1969–1993. Data were examined by a second-order procedure based on K-functions, with fixed thresholds of closeness in space (0.5–7.5 km) and closeness in time (0.1–1.5 years). Locations were addresses at diagnosis. Tests were repeated, replacing geographical distances with distances to the 19th–33rd nearest neighbours and this provided the primary result for each analysis. There were a total of 32,295 cases of childhood cancer. The analyses showed statistically significant evidence of space-time clustering for acute lymphoblastic leukaemia over the whole age range ($p = 0.04$), but especially for ages 1–4 years ($p = 0.03$). There was less statistically significant evidence for total leukaemia ($p = 0.048$). Significant space-time clustering was also evident for soft tissue sarcomas ($p = 0.03$) and osteosarcomas ($p = 0.02$). Results support other evidence suggesting a role for infections in the aetiology of these particular diagnostic groups.

© 2005 Wiley-Liss, Inc.

Key words: aetiology; cancer; children; infections; space-time clustering

The aetiology of childhood cancer is unclear, but is likely to include both genetic and environmental factors.¹ There has been much discussion concerning the role that infectious agents, including both viruses and bacteria, may have in the aetiology of childhood cancers. A causal link has been established between certain viruses and the development of leukaemia and osteosarcoma in animals.^{2–4} There is substantial evidence indicating that several viruses are causal in the aetiology of certain human malignancies.^{5,6} Acute T-cell leukaemia/lymphoma (ATLL) in adults is caused by exposure to the retrovirus human T-cell lymphotropic retrovirus type 1 (HTLV1). A small percentage of those infected with HTLV1 will subsequently develop ATLL.⁷ Epstein-Barr virus (EBV) has been found in tumour tissue from cases of childhood Burkitt's lymphoma (BL) in certain parts of sub-Saharan Africa, where BL is common. Furthermore, the geographical distribution of BL matches that of endemic malaria.⁸ EBV viral deoxyribose nucleic acid (DNA) has been found in the tumour cells of about 30% of cases of Hodgkin lymphoma (HL) and it is currently agreed that the virus plays a causal role for these cases. However, most cases of HL in children aged 0–9 years are EBV-negative, whilst only a small proportion in older children are EBV-positive.⁹ Kaposi's sarcoma is known to be caused by the human herpes virus type 8 (HHV8), in the presence of human immunodeficiency virus (HIV) infection.¹⁰ The *Helicobacter pylori* bacterium is causally related to the occurrence of gastric carcinoma in adults.^{11,12}

Indirect evidence from case-control, ecological and clustering studies suggests a role for infections in certain childhood cancers, especially leukaemia and central nervous system (CNS) tumours.^{1,13–15} If infections that are neither ubiquitous nor endemic are involved in the aetiology of childhood cancers, then the distribution of cases may be predicted to exhibit space-time clustering. However, this would only happen when the lag time from exposure to diagnosis is short or at least relatively constant. The onset of cancer would occur as a rare response to infection.

Space-time clustering is observed when excess numbers of cases are observed within small geographical locations over limited periods of time and this cannot be explained in terms of general excesses in those locations or at those times. Hence, space-time clustering may be described as 'the irregular grouping of cases of disease simultaneously in space and time'. These irregularities might arise from various situations, for example, a small number of locations with greatly increased incidence over short (but distinct) periods of time, or a large number of limited time periods with moderately increased incidence at limited locations. The analysis controls for constant excesses at specific locations, which occur at all times. It also controls for short temporal peaks in incidence that are geographically widespread (occurring at all locations).

A number of previous studies from parts of Europe, Australia and the USA have demonstrated statistically significant space-time clustering amongst cases of childhood leukaemia and especially among cases of acute lymphoblastic leukaemia (ALL) in the childhood peak.^{16–32}

Statistically significant evidence for space-time clustering amongst cases of certain CNS tumours, especially astrocytoma and ependymoma, soft tissue sarcoma and Wilms tumour, has been found in North West England.^{13,33} Other studies did not find any evidence of space-time clustering amongst cases of childhood cancer.^{1,34} However, many of the earlier studies have used small data sets and lack statistical power.

A number of statistical methods have been developed to test for space-time clustering.^{35–41} Recent analyses have adopted improved methodologies and larger data sets. The earlier methods have been criticised because they rely on an arbitrary choice of the critical values to define 'close in time' and 'close in space'. A more recent method partly overcomes this problem because it uses a large number of different pairs of critical values, rather than just 1 pair.⁴¹

The aim of the study was to test predictions of space-time clustering, which might arise as a result of infectious agents. The study has 3 major strengths: (1) it is the largest study of space-time clustering.

Grant sponsors: Cancer Research UK, Christie Hospital Research Endowment Fund.

*Correspondence to: Sir James Spence Institute, Level 4, Royal Victoria Infirmary, Queen Victoria Road, Newcastle NE1 4LP, UK.
Fax: +44-191-202-3060. E-mail: richard.mcnally@ncl.ac.uk

Received 14 June 2005; Accepted after revision 17 October 2005

DOI 10.1002/ijc.21726

Published online 27 December 2005 in Wiley InterScience (www.interscience.wiley.com).

tering to date, analysing 32,295 cases of childhood cancer; (2) it uses high quality population-based national incidence data on all childhood cancers, diagnosed within a 25-year time period with almost complete ascertainment; and (3) it uses more up-to-date statistical methodology, which overcomes the main criticisms of earlier methods.

Material and methods

Patients

All cases diagnosed with a childhood cancer between the 1st January 1969 and the 31st December 1993 and registered by the National Registry of Childhood Tumours (NRCT) were analysed. The NRCT is a population-based registry covering the whole of England, Wales and Scotland.⁴² It includes records for nearly all children, aged 0–14 years, diagnosed with cancer from 1962 to the present day. The data set analysed here was restricted to the period 1969–1993 because: (i) reliable small area population data required for the analyses by population density were not available prior to 1969 and (ii) at the time of preparation (1999), the data set was only complete up to the end of 1993.

Diagnostic groups

Cases were classified into diagnostic groups according to the International Classification of Diseases for Oncology (ICD-O).⁴³ The following diagnostic groups were specified *a priori* for analysis: (i) leukaemia; (ii) ALL; (iii) acute non-lymphocytic leukaemia (ANLL); (iv) lymphomas; (v) HL; (vi) non-Hodgkin lymphoma; (vii) CNS tumours; (viii) astrocytoma; (ix) primitive neuroectodermal tumours; (x) soft tissue sarcomas; (xi) bone tumours; (xii) osteosarcoma; (xiii) sympathetic nervous system tumours; (xiv) Wilms and other renal tumours; (xv) germ cell and related tumours; (xvi) retinoblastoma; (xvii) hepatic tumours; (xviii) carcinoma; (xix) all cancers except leukaemia and lymphoma and (xx) all cancers.

Grid references

The Ordnance survey (OS) produces grid-based maps for the whole of England, Wales and Scotland. For each case of childhood cancer, OS 4-digit Easting and 4-digit Northing grid references were allocated to the centroid of the post-code of the address at the time of diagnosis. This allowed spatial referencing of both the Easting and the Northing co-ordinates of the address at diagnosis to ~0.1 km. Grid references have had their origin shifted and they have been rotated. This preserved confidentiality of locations without detriment to the analysis.

Prior hypotheses

The following hypotheses were tested. Space-time clustering will be present in and confined to cases in specific diagnostic groups, notably leukaemia, ALL (especially the 'childhood peak'), CNS tumours (especially astrocytoma and ependymoma), soft tissue sarcoma and Wilms tumour. These prior hypotheses were mainly derived from the positive clustering within these groups found from other studies.^{17,30–33}

Statistical analyses

The space-time interactions between times and places of diagnosis were examined. The analyses presented are based on *K*-functions. *K*-function analysis may be regarded as a generalisation of the Knox test.³⁷ For this reason, the Knox test is outlined first of all.

The Knox test

The Knox test regards a pair of cases as being in 'close proximity' if they are both diagnosed at addresses which are close both in space (less than *s* km apart, where, *e.g.*, *s* = 5) and at times which are close (less than *t* months apart, where, *e.g.*, *t* = 12). These limits are arbitrary, but have been used in a number of other space-

time clustering studies in parts of Great Britain.^{30–33,44} The number of pairs of cases that are in close proximity is calculated [*O*]. The number of pairs of cases that are close in space [*D*] and that are close in time [*T*] are calculated. From these, and assuming that spatial and temporal proximity are independent, we have the expectation of *O*, the number of pairs of cases close both in time and space, given by $D \times T / N$, where *N* is the total number of case pairs. In the absence of space-time clustering, an assumption is made that *O* should follow a Poisson distribution with mean equal to *E*. If *O* exceeds *E*, there is space-time clustering and statistical tests can determine whether this excess is statistically significant. The magnitude of the excess is estimated by $S = [(O - E)/E] \times 100$. The variability of *S* depends on *E*. A related quantity whose variability does not depend on *E* is $R = (O - E) / \sqrt{E}$.

The K-function method

There are 2 problems with the Knox approach. First, boundary problems may be important, since it can be difficult for a case near a geographical boundary or the end of a time interval to be close to as many cases as if it occurred near the centre. The second problem is the arbitrariness of the critical values (*s* and *t*). A simplification of a procedure due to Diggle *et al.* is used to overcome the problem of arbitrary critical values.⁴⁰ We use here a set of 225 Knox-type calculations to obtain values of *R*, where the critical values are varied over a pre-specified set of values (for close times, *t* = 0.1, 0.2, . . . , 1.5 years and for close points in space, *s* = 0.5, 1, 1.5, . . . , 7.5 km). The test is based on the observed value of the *K*-function, [*K*₀], which is formed by summing *R* over the pre-specified values of *s* and *t*. The distribution of the *K*-function is unknown and so must be estimated by simulation. For each simulation, the dates of diagnosis were randomly re-allocated to each of the cases in the analysis and a value of the *K*-function [*K*_{*s*}] was then calculated from the simulated data. This was repeated for a total of *n*_{SIM} simulations. The observed value of the *K*-function, *K*₀, was then compared with the simulated values, *K*_{*s*} (where *s* = 1, . . . , *n*_{SIM}). *P*-values were estimated by calculating the proportion of the total number of simulations, *n*_{SIM}, for which *K*_{*s*} > *K*₀. Hence, statistical significance was assessed.

Strength of clustering

The *K*-function analysis yields no measure of the magnitude of the effect. The *P*-values depend both on the magnitude and on the statistical power, which in turn depends on the number of cases. Therefore, in this article, *S*, obtained from the Knox test (with critical values taken as 5 km and 1 year), is provided as some indication of the magnitude. However, a small value of *S* may occur when the true effect is large if the clustering occurs at a scale corresponding to critical values different from 5 km and 1 year.

Nearest neighbour analyses

It is possible that geographical distance may be inappropriate as a measure of closeness, especially when both urban and rural areas are included in the study. Any specified distance between 2 cases may have different meanings in urban and rural areas where, for example, sizes of school catchment areas differ. We therefore feel that it is appropriate to examine the influence of population density on any clustering effect. For a specific 'index' case, the other cases, which were in closest proximity, were termed 'nearest neighbours' (NN). These NN cases were ranked by their distance from the index case: 1, 2, 3, . . . This was done for every case in the data set in turn by treating it as an 'index' case.

On inspection of the whole set of NN distances, it was observed that the average distance to the 26th NN was about 5 km (though these distances varied greatly, between 245.3 km for a very isolated location to 0.7 km in a very densely populated location). The distances to the '*N*th NN (where '*N*' took the values 19, 20, 21, . . . , 33) were used instead of the distances previously considered (namely 0.5–7.5 km). To allow for the effect of varying population density, a second analysis (analysis 'using the NN

thresholds) was undertaken, in which cases were taken to be 'close in space' if either was within the distance to the N^{th} NN of the other, $N = 19, 20, \dots, 33$. All other details of this second analysis correspond to those described for the first analysis. An analogous concept to the NN approach was originally used by Jacques.⁴⁵ The NN threshold method provided the primary results for each analysis.

Analyses by gender

Analyses by gender proceeded by examining 'male:any' and 'female:any' clustering pairs. That is, for 'male:any' clustering pairs, whilst 1 case in a clustering pair was male, the other case could be either male or female.

Analyses by population density

Addresses were classified as being located in a more densely populated area, or being located in a less densely populated area. Fifty percent of cases resided in a ward that had more than 21.26 persons per hectare and these cases were therefore classified as being located in a more densely populated area. Conversely, 50% of cases resided in a ward that had no more than 21.26 persons per hectare and these cases were classified as being located in a less densely populated area. Analysis by population density was undertaken by considering pairs of cases including at least 1 from the 'more densely populated' category ('more densely populated:any' clustering pairs) and pairs of cases including at least 1 from the 'less densely populated' category ('less densely populated:any' clustering pairs). It should be noted that these analyses (especially the analyses of 'less densely populated:any' clustering pairs) are potentially subject to a strong diluting influence from edge effects, since neither the 'more densely populated' areas nor the 'less densely populated' areas form a single spatially contiguous zone.

Choice of statistical significance level

Statistical significance in all analyses was taken as the conventional level $p < 0.05$, chosen because of its wide-spread use and consequent familiarity to readers. Twenty six significance tests were performed. Under an assumption of independence, the proba-

bility that at least 4 tests will be significant at this level is 0.039 and the probability that at least 5 tests will be significant at this level is less than 0.0085.

Results

The study included 32,295 cases of childhood cancer aged 0–14 years. In all analyses, twins have been excluded. Table I gives the total number of cases for each analysis group, the number of males, the number of females, the number of cases located in a more densely populated area and the number of cases located in a less densely populated area.

The primary result from each analysis was taken from the NN threshold method. For statistically significant results, the strength of clustering (S) from the Knox test was also given.

Table II presents the main results of the analyses obtained using the NN threshold method. There was statistically significant space-time clustering for cases of leukaemia, aged 0–14 years as a whole ($p = 0.01$, $S = 0.05\%$ using the geographical distance method and $p = 0.048$, $S = 0.2\%$ using the NN threshold method). This was attributable to statistically significant space-time clustering for cases of ALL ($p = 0.04$, $S = 1.3\%$ using the NN threshold method) but not for cases of ANLL ($p = 0.11$ using the NN threshold method). Furthermore, the space-time clustering was only found for cases of ALL aged 1–4 years ($p = 0.03$, $S = 4.1\%$ using the NN threshold method). The lack of clustering for leukaemias aged 1–4 years can be explained by the diluting effect of non-clustering cases of ANLL. Table II also shows statistically significant space-time clustering for soft tissue sarcomas ($p = 0.03$, $S = 9.8\%$ using the NN threshold method) and osteosarcoma ($p = 0.02$, $S = 25.5\%$ using the NN threshold method). Overall, there were 5 statistically significant ($p < 0.05$) results using the NN threshold method. Assuming independence, the probability of obtaining at least 5 such statistically significant results is 0.0085.

There was statistically significant space-time clustering for leukaemia, aged 1–4 years, bone tumours and the group comprising all cancers except leukaemia and lymphoma. However, these results were based on geographical distance and not on the NN threshold.

TABLE I – NUMBERS OF CASES FOR ANALYSES OF SPACE-TIME CLUSTERING IN CHILDHOOD CANCER IN GREAT BRITAIN, 1969–1993

Diagnostic group and age (years)	Total	Males	Females	In more densely populated area	In less densely populated area
Leukaemia, ages 0–14	10,727	6,059	4,668	5,397	5,330
Leukaemia, ages 1–4	5,092	2,895	2,197	2,546	2,546
Leukaemia, ages 5–14	5,087	2,892	2,195	2,558	2,529
ALL, ages 0–14	8,439	4,841	3,598	4,227	4,212
ALL, ages 1–4	4,343	2,483	1,860	2,166	2,177
ALL, ages 5–14	3,810	2,216	1,594	1,910	1,900
ANLL, ages 0–14	1,736	918	818	893	843
Lymphomas, ages 0–14	3,306	2,326	980	1,679	1,627
HL, ages 0–14	1,364	953	411	692	672
HL, ages 0–9	487	380	107	268	219
NHL, ages 0–14	1,678	1,192	486	848	830
NHL, ages 0–9	1,027	737	290	514	513
CNS tumours, ages 0–14	7,469	4,091	3,378	3,705	3,764
Astrocytoma, ages 0–14	2,824	1,422	1,402	1,392	1,432
PNET, ages 0–14	1,548	976	572	744	804
Soft tissue sarcomas, ages 0–14	2,101	1,185	916	1,008	1,093
Bone tumours, ages 0–14	1,504	775	729	728	776
Osteosarcoma, ages 0–14	811	408	403	394	417
Sympathetic nervous system tumours, ages 0–14	2,110	1,169	941	1,039	1,071
Wilms and other renal tumours, ages 0–14	1,889	945	944	936	953
Germ cell and related tumours, ages 0–14	983	472	511	495	488
Retinoblastoma, ages 0–14	941	488	453	509	432
Hepatic tumours, ages 0–14	257	144	113	138	119
Carcinoma, ages 0–14	877	385	492	445	432
All cancers except leukaemia and lymphoma, ages 0–14	18,262	9,725	8,537	9,071	9,191
All cancers, ages 0–14	32,295	18,110	14,185	16,147	16,148

ALL, acute lymphoblastic leukaemia; ANLL, acute non-lymphocytic leukaemia; HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; CNS, central nervous system; PNET, primitive neuroectodermal tumours.

TABLE II – MAIN RESULTS OF ANALYSES OF SPACE-TIME CLUSTERING IN CHILDHOOD CANCER IN GREAT BRITAIN, 1969–1993, OBTAINED USING THE NN THRESHOLD METHOD

Diagnostic group and age (years)	<i>p</i> -value (strength of clustering (<i>S</i>))
Leukaemia, ages 0–14	0.048 ¹ (0.2%)
Leukaemia, ages 1–4	0.16
Leukaemia, ages 5–14	0.31
ALL, ages 0–14	0.04 ¹ (1.3%)
ALL, ages 1–4	0.03 ¹ (4.1%)
ALL, ages 5–14	0.44
ANLL, ages 0–14	0.11
Lymphomas, ages 0–14	0.14
HL, ages 0–14	0.13
HL, ages 0–9	0.11
NHL, ages 0–14	0.09
NHL, ages 0–9	0.39
CNS tumours, ages 0–14	0.41
Astrocytoma, ages 0–14	0.06
PNET, ages 0–14	0.33
Soft tissue sarcomas, ages 0–14	0.03 ¹ (9.8%)
Bone tumours, ages 0–14	0.08
Osteosarcoma, ages 0–14	0.02 ¹ (25.5%)
Sympathetic nervous system tumours, ages 0–14	0.89
Wilms and other renal tumours, ages 0–14	0.23
Germ cell and related tumours, ages 0–14	0.34
Retinoblastoma, ages 0–14	0.70
Hepatic tumours, ages 0–14	0.31
Carcinoma, ages 0–14	0.07
All cancers except leukaemia and lymphoma, ages 0–14	0.12
All cancers, ages 0–14	0.22

NN, nearest neighbour; ALL, acute lymphoblastic leukaemia; ANLL, non-lymphocytic leukaemia; HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; CNS, central nervous system; PNET, primitive neuroectodermal tumours.

¹Statistically significant.

Table III presents the results of space-time clustering by gender based on 'male:any' and 'female:any' clustering pairs obtained using the NN threshold method. There was statistically significant space-time clustering of 'male:any' pairs for leukaemia ($p = 0.02$, $S = 0.2\%$ using the geographical distance method and $p = 0.04$, $S = 1.1\%$ using the NN threshold method), ALL ($p = 0.03$, $S = 2.5\%$ using the NN threshold method), ANLL ($p = 0.03$, $S = 12.1\%$ using the NN threshold method), astrocytoma ($p = 0.01$, $S = 13.8\%$ using the NN threshold method) and soft tissue sarcoma ($p = 0.02$, $S = 11.2\%$ using the NN threshold method). There was statistically significant space-time clustering of 'female:any' pairs for osteosarcoma ($p = 0.03$, $S = 25.3\%$ using the geographical distance method and $p = 0.03$, $S = 31.4\%$ using the NN threshold method). There were 3 other significant results based on geographical distance but not on the NN threshold method (leukaemia, aged 1–4 years and osteosarcoma for 'male:any' clustering pairs and bone tumours for 'female:any' clustering pairs). Overall, there were 5 statistically significant ($p < 0.05$) results involving males and 1 statistically significant ($P < 0.05$) result involving females using the NN threshold method. Assuming independence, the overall effect for males is unlikely to have occurred by chance (the probability of obtaining at least 5 such significant results is 0.0085), but the single significant result for females may well be spurious (the probability of obtaining at least 1 such significant result is 0.74).

Table IV presents the results of space-time clustering based on 'more densely populated:any' and 'less densely populated:any' clustering pairs obtained using the NN threshold method. Three groups showed statistically significant space-time clustering for 'less densely populated:any' clustering pairs only, and 1 group showed statistically significant space-time clustering for both 'more densely populated:any' and 'less densely populated:any' clustering

pairs (based on the NN threshold method). Specifically, statistically significant space-time clustering was found for ALL ($p = 0.02$, $S = 7.1\%$), soft tissue sarcoma ($p = 0.006$, $S = 19.2\%$) and carcinoma ($p = 0.03$, $S = 26.9\%$), for 'less densely populated:any' clustering pairs only. Statistically significant space-time clustering for osteosarcoma was found for both 'more densely populated:any' ($p = 0.046$, $S = 18.6\%$) and for 'less densely populated:any' ($p = 0.02$, $S = 34.5\%$) clustering pairs. Overall, there were 4 statistically significant ($p < 0.05$) results involving cases from less densely populated areas and 1 statistically significant ($p < 0.05$) result involving cases from more densely populated areas using the NN threshold method. Assuming independence, the overall effect involving cases from less densely populated areas is unlikely to have occurred by chance (the probability of obtaining at least 4 such significant results is 0.039), but the single significant result involving cases from more densely populated areas may well be spurious (the probability of obtaining at least 1 such significant result is 0.74).

Discussion

The analyses presented here have been carried out using predetermined (or well specified) statistical methods on high quality population-based incidence data. It is the largest analysis of space-time clustering of childhood cancer ever done, based on 32,295 cases. Space-time clustering has been particularly identified for cases of leukaemia, ALL (aged 1–4 years), soft tissue sarcoma and osteosarcoma.

The results for leukaemia are in agreement with previous smaller studies from the UK and elsewhere.^{1,27–31} In particular, the finding that the space-time clustering was most pronounced in the 1–4 year olds is consistent with the earlier studies. It should be noted that one of the earlier studies found that clustering of 'childhood peak' cases of ALL was most marked when based on time of diagnosis and place of birth.³¹ However, place of birth was not available in the present study. The current results certainly provide support for an infectious component in the aetiology of childhood leukaemia and are in agreement with prior hypotheses. There are 3 hypotheses relating to the role that infections may play in the aetiology of childhood leukaemia.^{46–49} Kinlen^{47,48} proposed that much higher incidence of leukaemia would be related to situations of very unusual population mixing in previously isolated areas, Greaves⁴⁶ suggested that the second hit precipitating the onset of precursor B-cell ALL is related to delayed exposure to common infections in children whose immune system has been inadequately 'primed' by exposure to infections, whilst Smith⁴⁹ attributed higher incidence rates to an in-utero exposure to infection and DNA change. However, the results from the present study cannot distinguish between these 3 hypotheses, because they are based on time and place of diagnosis. Additional data on time and place of birth would be required, but were not available. The finding of gender differences, especially the excess of males involved in space-time pairs, is consistent with known differences in susceptibility to infection between males and females.^{50–52} The finding, based on ward population density, that there was a statistically significant excess of clustered cases of 'childhood peak' ALL (cases aged 1–4 years) from more 'rural' areas (less densely populated areas) is consistent with Greaves' hypotheses, since residence in such areas is likely to lead to a delay in exposure to common infections. It should be noted that the latter finding relates to space-time clustering and not overall incidence. Most ecological studies have shown higher rates of ALL in more densely populated or urban areas,¹⁵ but the lack of space-time clustering in more 'urban' areas in the current study may be interpreted as indicating a more homogeneous distribution of cases in more 'urban' areas. This would be consistent with earlier exposure to common infections, which are likely to be more prevalent in more 'urban' areas.

There are 3 notable findings for soft tissue sarcomas. First of all, they confirm the results of a much smaller study from North West England,³³ although the earlier study found space-time

TABLE III – RESULTS OF ANALYSES OF SPACE-TIME CLUSTERING IN CHILDHOOD CANCER IN GREAT BRITAIN, 1969–1993, FOR “MALE:ANY” AND “FEMALE:ANY” CLUSTERING PAIRS, OBTAINED USING THE NN THRESHOLD METHOD

Diagnostic group and age (years)	“Male:any”	“Female:any”
	<i>p</i> -value (strength of clustering (<i>S</i>))	<i>p</i> -value (strength of clustering (<i>S</i>))
Leukaemia, ages 0–14	0.04 ¹ (1.1%)	0.14
Leukaemia, ages 1–4	0.14	0.32
Leukaemia, ages 5–14	0.19	0.64
ALL, ages 0–14	0.03 ¹ (2.5%)	0.08
ALL, ages 1–4	0.20	0.20
ALL, ages 5–14	0.48	0.36
ANLL, ages 0–14	0.03 ¹ (12.1%)	0.34
Lymphomas, ages 0–14	0.15	0.17
HL, ages 0–14	0.08	0.33
HL, ages 0–9	0.11	0.42
NHL, ages 0–14	0.14	0.31
NHL, ages 0–9	0.46	0.14
CNS tumours, ages 0–14	0.39	0.48
Astrocytoma, ages 0–14	0.01 ¹ (13.8%)	0.41
PNET, ages 0–14	0.17	0.90
Soft tissue sarcomas, ages 0–14	0.02 ¹ (11.2%)	0.46
Bone tumours, ages 0–14	0.40	0.06
Osteosarcoma, ages 0–14	0.06	0.03 ¹ (31.4%)
Sympathetic NS tumours, ages 0–14	0.90	0.75
Wilms and other renal tumours, ages 0–14	0.16	0.36
Germ cell and related tumours, ages 0–14	0.33	0.46
Retinoblastoma, ages 0–14	0.71	0.83
Hepatic tumours, ages 0–14	0.20	0.48
Carcinoma, ages 0–14	0.06	0.20
All cancers except leukaemia and lymphoma, ages 0–14	0.36	0.08
All cancers, ages 0–14	0.44	0.14

NN, nearest neighbour; ALL, acute lymphoblastic leukaemia; ANLL, acute non-lymphocytic leukaemia; HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; CNS, central nervous system; PNET, primitive neuroectodermal tumours; NS, nervous system.

¹Statistically significant.

TABLE IV – RESULTS OF ANALYSES OF SPACE-TIME CLUSTERING IN CHILDHOOD CANCER IN GREAT BRITAIN, 1969–1993, FOR “MORE DENSELY POPULATED¹:ANY” AND “LESS DENSELY POPULATED²:ANY” CLUSTERING PAIRS, OBTAINED USING THE NN THRESHOLD METHOD

Diagnostic group and age (years)	At least one of clustering pair from more densely populated area	At least one of clustering pair from less densely populated area
	<i>p</i> -value (strength of clustering (<i>S</i>))	<i>p</i> -value (strength of clustering (<i>S</i>))
Leukaemia, ages 0–14	0.07	0.12
Leukaemia, ages 1–4	0.14	0.12
Leukaemia, ages 5–14	0.49	0.45
ALL, ages 0–14	0.06	0.09
ALL, ages 1–4	0.08	0.02 ³ (7.1%)
ALL, ages 5–14	0.48	0.61
ANLL, ages 0–14	0.19	0.11
Lymphomas, ages 0–14	0.25	0.13
HL, ages 0–14	0.15	0.13
HL, ages 0–9	0.38	0.20
NHL, ages 0–14	0.11	0.10
NHL, ages 0–9	0.41	0.64
CNS tumours, ages 0–14	0.81	0.11
Astrocytoma, ages 0–14	0.28	0.10
PNET, ages 0–14	0.52	0.09
Soft tissue sarcomas, ages 0–14	0.42	0.006 ³ (19.2%)
Bone tumours, ages 0–14	0.07	0.10
Osteosarcoma, ages 0–14	0.046 ³ (18.6%)	0.02 ³ (34.5%)
Sympathetic NS tumours, ages 0–14	0.75	0.90
Wilms and other renal tumours, ages 0–14	0.55	0.26
Germ cell and related tumours, ages 0–14	0.52	0.34
Retinoblastoma, ages 0–14	0.47	0.85
Hepatic tumours, ages 0–14	0.22	0.15
Carcinoma, ages 0–14	0.76	0.03 ³ (26.9%)
All cancers except leukaemia and lymphoma, ages 0–14	0.11	0.21
All cancers, ages 0–14	0.24	0.46

NN, nearest neighbour; ALL, acute lymphoblastic leukaemia; ANLL, acute non-lymphocytic leukaemia; HL, Hodgkin lymphoma; NHL, non-Hodgkin lymphoma; CNS, central nervous system; PNET, primitive neuroectodermal tumours; NS, nervous system.

¹More densely populated is >21.26 persons per hectare (50% of cases).–²Less densely populated is ≤21.26 persons per hectare (50% of cases).–³Statistically significant.

clustering particularly for place and time of birth. Second, there was an excess of males in clustering pairs, again in agreement with the earlier study. Finally, there was an excess of clustering cases from less densely populated areas, which was not in agreement with the previous study. However, the present study included 2,101 cases of soft tissue sarcoma, whilst the previous study included only 244 cases.³³ Therefore, the differences may be explained in terms of the power of the studies. It should be noted that an ecological study from North West England found higher incidence of soft tissue sarcomas in areas, which were more rural/less urban.⁵³ The current study is in agreement with the prior hypothesis and is suggestive of an infectious aetiology, although the mechanism, timing and/or infectious agent(s) involved are at present unknown. Further support for an infectious hypothesis for childhood soft tissue sarcoma is provided by the causal link between HHV8 and Kaposi's sarcoma, with HIV also present.¹⁰

There was space-time clustering for cases of osteosarcoma, but not for all bone tumours. The previous study from North West England did not find any space-time clustering amongst cases of bone tumour or osteosarcoma, but case numbers were much smaller.³³ There was a small female excess in clustering pairs and a small excess of cases from 'less densely populated' areas. Whilst these results were not predicted by a prior hypothesis, the current study certainly suggests a role for infectious agents in the aetiology of osteosarcoma, but not other bone tumours. The mechanism is as yet unclear, as is the candidate agent. Simian virus 40 has been shown to induce osteosarcoma in hamsters, but evidence suggesting a link with osteosarcoma in humans is much more tentative.^{54,55}

CNS tumours and astrocytoma, in particular, were only marginally significant, although 'male: any' pairs of astrocytoma were statistically significant. These results contrast with the previous findings of space-time clustering for astrocytoma and ependymoma from North West England, although that study additionally found space-time clustering based on time and place of birth.¹⁴ It is possible that the differences may be explained by the different time periods under consideration or the presence of unknown cofactors, which display large scale geographical variation.

The lack of space-time clustering for Wilms tumour contrasts with the finding of space-time clustering from the earlier and much smaller study from North West England and refutes the prior hypothesis.³³ However, the previous findings from North West England were based on time and place of birth. Furthermore, there are no hypotheses linking infections to the aetiology of Wilms tumour. It is entirely possible that the previous findings from North West England may be linked to an unknown locally occurring environmental factor or they may be due to chance.

The lack of space-time clustering for the lymphomas is consistent with similar findings from North West England.⁴⁴ Furthermore, if lymphomas, especially HL, are linked with the occurrence of a ubiquitous virus, then space-time clustering would not be predicted. The lack of space-time clustering for all the other groups is entirely consistent with the lack of aetiological hypotheses suggesting an infectious origin.

It is conceivable that artefactual space-time clustering may be observed as a result of small-area population shifts over short time intervals. Kulldorff and Hjalmars have proposed a method of

adjusting for such population shifts.⁵⁶ However, it is not possible to adjust for putative population shifts in the present study, because it would require small-area population denominators for short time intervals and such counts are not available in the UK. Furthermore, the statistically significant space-time clustering was limited to very particular groups and for most of these groups, there is good *a priori* evidence that suggests an infectious origin. The specific and limited nature of the space-time clustering and the lack of an overall effect for the "all cancers" group would strongly argue against the proposition that the observed clustering may be artefactual due to population shifts.

The only main limitation of the study was that analyses were based on dates and places of diagnosis. Dates and places of birth were not available but would be more meaningful for some of the early onset childhood tumours where the precipitating environmental exposure is likely to occur in-utero or close to the time of birth. There was some overlap between the current data set and several data sets analysed in previous studies. Specifically, cases from North West England, which were included in the MCTR during the period 1969–1993, were also analysed in a series of studies which covered all or most of the time period from 1954 to 2001.^{14,15,31–33} However, this amounted only to about 1,500 cases. There was also overlap with 2 other studies, but these used more limited statistical methods for analysis.^{27,28}

It is entirely possible that some significant results may have arisen by chance. However, we have regarded the primary results as those obtained using the NN threshold method and given in Table II. There were 26 tests and 5 were significant at the conventional level of $p < 0.05$. Assuming independence, the probability that this will occur is 0.0085. Therefore, the overall pattern is unlikely to be explained by chance resulting from multiple testing. Also, in general, the results are supportive of prior hypotheses (obtained from previous analyses of much smaller data sets). Tables III and IV should be regarded as supplementary analyses that are mainly used to provide further interpretation of the significant results from Table II. Overall, there was evidence that clustering effects involved males ($p = 0.0085$) and cases from less densely populated areas ($p = 0.039$).

In summary, we have used rigorous methods and high-quality data and found evidence of space-time clustering for cases of leukaemia as a whole, ALL, soft tissue sarcoma and osteosarcoma. The results are generally in agreement with the findings of previous much smaller studies, although other previous findings for CNS tumours and Wilms tumour have not been replicated in our study. Space-time clustering is highly suggestive of a role for an infectious component to aetiology. The specific nature of these findings would argue against the results being artefactual. Whilst infectious mechanistic hypotheses have been proposed for leukaemia and ALL, there is a lack of such hypotheses for soft tissue sarcoma and osteosarcoma. Future studies should consider such mechanisms and candidate infectious agents.

Acknowledgements

Much of this work was done when Dr. Richard J.Q. McNally was at the Cancer Research UK Paediatric and Familial Cancer Research Group, University of Manchester.

References

1. Little J. Epidemiology of childhood cancer. Lyon: IARC (IARC Scientific Publications No. 149), 1999.
2. Penrose M. Cat leukaemia. *Br Med J* 1970;i:755.
3. Linenberger ML, Abkowitz JL. Haematological disorders associated with feline retrovirus infections. *Baillieres Clin Haematol* 1995;8: 73–112.
4. Nevins JR. Cell transformation by viruses. In: Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus S, eds. *Fields virology*. Philadelphia: Lippincott Williams and Wilkins, 2005.
5. Evans AS, Mueller NE. Viruses and cancer. Causal associations. *Ann Epidemiol* 1990;1:71–92.
6. Mueller N. Overview: viral agents and cancer. *Environ Health Perspect* 1995;103(Suppl 8):259–61.
7. Kannagi M, Ohashi T, Harashima N, Hanabuchi S, Hasegawa A. Immunological risks of adult T-cell leukemia at primary HTLV-I infection. *Trends Microbiol* 2004;12:346–52.
8. Young LS, Rickinson AB. Epstein-Barr virus: 40 years on. *Nat Rev Cancer* 2004;4:757–68.
9. Jarrett RF. Risk factors for Hodgkin's lymphoma by EBV status and significance of detection of EBV genomes in serum of patients with EBV-associated Hodgkin's lymphoma. *Leuk Lymphoma* 2003; 44(Suppl 3): S27–32.

10. Viejo-Borbolla A, Schulz TF. Kaposi's sarcoma-associated herpesvirus (KSHV/HHV8): key aspects of epidemiology and pathogenesis. *AIDS Rev* 2003;5:222-9.
11. Mladenova I, Pellicano R. Infectious agents and gastric tumours. An increasing role for Epstein-Barr virus. *Panminerva Med* 2003;45:183-8.
12. Pagano JS, Blaser M, Buendia MA, Damania B, Khalili K, Raab-Traub N, Roizman B. Infectious agents and cancer: criteria for a causal relation. *Semin Cancer Biol* 2004;14:453-71.
13. Alexander FE. Viruses, clusters and clustering of childhood leukaemia: a new perspective? *Eur J Cancer* 1993;29A:1424-43.
14. McNally RJQ, Cairns DP, Eden OB, Alexander FE, Taylor GM, Kelsey AM, Birch JM. An infectious aetiology for childhood brain tumours? Evidence from space-time clustering and seasonality analyses. *Br J Cancer* 2002;86:1070-7.
15. McNally RJQ, Eden TOB. An infectious aetiology for childhood leukaemia: a review of the evidence. *Br J Hematol* 2004;127:243-63.
16. Pinkel D, Dowd JE, Bross IDJ. Some epidemiological features of malignant solid tumors of children in the Buffalo, N.Y. area. *Cancer* 1963;16:28-33.
17. Meighan SS, Knox G. Leukemia in childhood. *Epidemiology in Oregon*. *Cancer* 1965;18:811-4.
18. Mainwaring D. Epidemiology of acute leukaemia of childhood in the Liverpool area. *Br J Prev Soc Med* 1966;20:189-94.
19. Gunz FW, Spears GFS. Distribution of acute leukaemia in time and space. *Studies in New Zealand*. *Br Med J* 1968;4:604-8.
20. Klauber MR, Mustacchi P. Space-time clustering of childhood leukemia in San Francisco. *Cancer Res* 1970;30:1969-73.
21. Zahalkova M, Bilek O, Kubikova A, Belusa M. The incidence of leukaemias in children and the clustering in space and time in South Moravian Province, Czechoslovakia. *Blut* 1970;21:180-5.
22. Glass AG, Mantel N, Gunz FW, Spears GFS. Time-space clustering of childhood leukemia in New Zealand. *J Natl Cancer Inst* 1971;47:329-36.
23. Evatt BL, Chase GA, Heath CW. Time-space clustering among cases of acute leukemia in two Georgia counties. *Blood* 1973;41:265-72.
24. Larsen RJ, Holmes CL, Heath CW. A statistical test for measuring unimodal clustering: a description of the test and of its application to cases of acute leukemia in metropolitan Atlanta, Georgia. *Biometrics* 1973;29:301-9.
25. Smith PG, Pike MC, Till MM, Hardisty RM. Epidemiology of childhood leukaemia in Greater London: a search for evidence of transmission assuming a possibly long latent period. *Br J Cancer* 1976;33:1-8.
26. Morris V. Space-time interactions in childhood cancers. *J Epidemiol Commun Health* 1990;44:55-8.
27. Gilman EA, Knox EG. Temporal-spatial distribution of childhood leukaemias and non-Hodgkin's lymphoma in Great Britain. In: Draper GJ, ed. *The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966-1983. Studies on medical and population subjects no. 53*. London: HMSO, 1991;77-99.
28. Gilman EA, Knox EG. Childhood cancers: space-time distribution in Britain. *J Epidemiol Commun Health* 1995;49:158-63.
29. Petridou E, Revinthi K, Alexander FE, Haidas S, Kolioukas D, Kosmidis H, Piperopoulou F, Tzortzatos F, Trichopoulos D. Space-time clustering of childhood leukaemia in Greece: evidence supporting a viral aetiology. *Br J Cancer* 1996;73:1278-83.
30. Alexander FE, Boyle P, Carli PM, Coebergh JW, Draper GJ, Ekblom A, Levi F, McKinney PA, McWhirter W, Magnani C, Michaelis J, Olsen JH, et al. Spatial temporal patterns in childhood leukaemia: further evidence for an infectious origin. EUROCLUS project. *Br J Cancer* 1998;77:812-7.
31. Birch JM, Alexander FE, Blair V, Eden OB, Taylor GM, McNally RJQ. Space-time clustering patterns in childhood leukaemia support a role for infection. *Br J Cancer* 2000;82:1571-6.
32. McNally RJQ, Alexander FE, Birch JM. Space-time clustering analyses of childhood acute lymphoblastic leukaemia by immunophenotype. *Br J Cancer* 2002;87:513-5.
33. McNally RJQ, Kelsey AM, Eden OB, Alexander FE, Cairns DP, Birch JM. Space-time clustering patterns in childhood solid tumours other than central nervous system tumours. *Int J Cancer* 2003;103:253-8.
34. Hjalmar U, Kulldorff M, Wahlqvist Y, Lannering B. Increased incidence rates but no space-time clustering of childhood astrocytoma in Sweden, 1973-1992: a population-based study of pediatric brain tumors. *Cancer* 1999;85:2077-90.
35. Pinkel D, Nefzger D. Some epidemiological features of childhood leukaemia in the Buffalo, NY area. *Cancer* 1959;12:351-8.
36. Ederer F, Myers MH, Mantel N. A statistical problem in space and time: do leukemia cases come in clusters? *Biometrics* 1964;20:626-38.
37. Knox EG. The detection of space-time interactions. *Applied Stats* 1964;13:25-9.
38. Barton DE, David FN, Merrington M. A criteria for testing contagion in time and space. *Ann Hum Genet (Lond)* 1965;29:97-102.
39. Mantel N. The detection of disease clustering and a generalised regression approach. *Cancer Res* 1967;27:209-20.
40. Pike MC, Smith PG. A case-control approach to examine diseases for evidence of contagion, including disease with long latent periods. *Biometrics* 1974;30:263-79.
41. Diggle PJ, Chetwynd AG, Haggkvist R, Morris SE. Second-order analysis of space-time clustering. *Stat Methods Med Res* 1995;4:124-36.
42. Stiller CA, Allen MB, Eatock EM. Childhood cancer in Britain: the National Registry of Childhood Tumours and incidence rates 1978-1987. *Eur J Cancer* 1995;31A:2028-34.
43. Fritz A, Percy C, Jack A, Shanmugaratna K, Sobin L, Parkin DM, Whelan S, eds. *International classification of diseases for oncology (ICD-O)*, 3rd ed. Geneva: World Health Organisation, 2001.
44. McNally RJQ, Alexander FE, Eden OB, Birch JM. Little or no space-time clustering found amongst cases of childhood lymphoma in North West England. *Eur J Cancer* 2004;40:585-9.
45. Jacques GM. A *k* nearest neighbour test for space-time interaction. *Stat Med* 1996;15:1935-49.
46. Greaves MF. Speculations on the cause of childhood acute lymphoblastic leukemia. *Leukemia* 1988;2:120-5.
47. Kinlen L. Evidence for an infective cause of childhood leukaemia: comparison of a Scottish New Town with nuclear reprocessing sites in Britain. *Lancet* 1988;ii:1323-7.
48. Kinlen LJ. Epidemiological evidence for an infective basis in childhood leukaemia. *Br J Cancer* 1995;71:1-5.
49. Smith M. Considerations on a possible viral etiology for B-precursor acute lymphoblastic leukemia of childhood. *J Immunother* 1997;20:89-100.
50. Washburn TC, Medearis DN, Childs B. Sex differences in susceptibility to infections. *Pediatrics* 1965;35:57-64.
51. Purtilo DT, Sullivan JL. Immunological bases for superior survival of females. *Am J Dis Child* 1979;133:1251-3.
52. Taylor GM, Dearden S, Payne N, Ayres M, Gokhale DA, Birch JM, Blair V, Stevens RF, Will AM, Eden OB. Evidence that an HLA-DQA1-DQB1 haplotype influences susceptibility to childhood common acute lymphoblastic leukaemia in boys provides further support for an infection-related aetiology. *Br J Cancer* 1998;78:561-5.
53. McNally RJQ, Alston RD, Cairns DP, Eden OB, Kelsey AM, Birch JM. Geographical and ecological analyses of childhood Wilms' tumours and soft-tissue sarcomas in North West England. *Eur J Cancer* 2003;39:1586-93.
54. Heinsohn S, Scholz RB, Weber B, Wittenstein B, Werner M, Delling G, Kempf-Bielack B, Setlak P, Bielack S, Kabisch H. SV40 sequences in human osteosarcoma of German origin. *Anticancer Res* 2000;20:4539-45.
55. Jasani B, Cristaudo A, Emri SA, Gazdar AF, Gibbs A, Krynska B, Miller C, Mutti L, Radu C, Tognon M, Procopio A. Association of SV40 with human tumours. *Semin Cancer Biol* 2001;11:49-61.
56. Kulldorff M, Hjalmar U. The Knox method and other tests for space-time interaction. *Biometrics* 1999;55:544-52.