IJC
International Journal of Cancer

Cross-space-time clustering of childhood cancer in Great Britain: Evidence for a common aetiology

Richard J.Q. McNally [1], Charles Stiller[2], Tim J. Vincent[2] and Michael F.G. Murphy[2]

[1] Institute of Health and Society, Newcastle University, England, United Kingdom
[2] Childhood Cancer Research Group, Department of Paediatrics, University of Oxford, England, United Kingdom

Previously, we identified space-time clustering in certain childhood cancers. This study aimed to determine whether there was cross-space-time clustering between different diagnostic groups. A total of 32,295 cases were diagnosed during 1969–1993. Cross-space-time clustering was analyzed by a second-order procedure based on Diggle's method. Locations were birth and diagnosis addresses. The following space-time combinations were examined: address and date of birth; address at birth and date of diagnosis; address and date of diagnosis. Cross-space-time clustering analyses considered clustering pairs of cases from two different diagnostic groups. Formal statistical significance was taken as $p < 0.00067$ and marginal significance $0.01 > p \geq 0.00067$. Based on address at birth and date of diagnosis, there was statistically significant cross-clustering between cases of HL and intracranial and intraspinal embryonal tumors (IIET), both aged 0–14 years ($p < 0.0001$). Based on address and date of birth, there was marginally significant cross-clustering between cases of lymphoid leukemia (LL) aged 5-14 years and Hodgkin lymphoma (HL) aged 0–14 years ($p = 0.0019$). Based on address and date of diagnosis there was marginally significant cross-clustering between cases of LL aged 1–4 years and soft tissue sarcoma (STS) aged 0–14 years ($p = 0.0041$). Findings from this study are consistent with possible common aetiological factors between different diagnostic groups. They suggest a common aetiology for the following pairs of diagnostic groups: HL and IIET; older cases of LL and HL; younger cases of LL and STS. The possibility of common infectious mechanisms should be explored.

Space-time clustering occurs when excess numbers of cases of cancer are observed within small geographical locations at limited periods and this cannot be attributed to general excesses in those locations or at those times. We have previously demonstrated statistically significant space-time clustering for certain childhood cancers diagnosed in Great Britain during the period 1st January 1969 to 31st December 1993. Analyses based on address and date of birth found space-time clustering for cases of Hodgkin lymphoma (HL, aged 0–14 years) and central nervous system (CNS) tumor (aged 0–14 years). Analyses based on address at birth and date of diagnosis found space-time clustering for cases of leukemia (ages 1–4 years), lymphoma (ages 0–14 years), non-Hodgkin lymphoma (NHL, ages 0–14 years), NHL (ages 0–9 years) and Wilms tumor (ages 0–14 years).[1] Analyses based on address and date of diagnosis found space-time clustering for cases of leukemia (ages 0–14 years), lymphoid leukemia (LL, ages 0–14 years), LL (ages 1–4 years), soft tissue sarcoma (STS, ages 0–14 years) and osteosarcoma (ages 0–14 years).[2] Additionally, other regional studies from the UK support findings of space-time clustering amongst cases of leukemia, CNS tumor, STS and Wilms tumor.[3–7] Together, these findings support the involvement of transient environmental exposures in aetiology.

In general, the aetiology of childhood cancer is not clear. Both genetic predisposition and environmental exposure are likely to be involved, but the former only directly accounts for ~5% of total cases.[8] Preconception, *in-utero* or postnatal environmental exposures could all be implicated. It has been suggested that at least two events are required to trigger the onset of a tumor.[9,10] A role for infections as the agent responsible for triggering the final event has been demonstrated for certain lymphomas[11] and postulated for childhood leukemia, CNS tumors and sympathetic nervous system tumors.[12–15] Support for a possible infectious aetiology for STS is suggested by the causal link between HHV8 and

Epidemiology

Key words: aetiology, cancer, childhood, epidemiology, Great Britain, space-time clustering
Grant sponsor: North of England Children's Cancer Research (NECCR), Newcastle University; **Grant sponsor:** Department of Health/National Cancer Intelligence Network, University of Oxford; **Grant sponsor:** Scottish Government, University of Oxford; **Grant sponsor:** Children with Cancer UK, University of Oxford
DOI: 10.1002/ijc.28332
History: Received 22 Nov 2012; Accepted 28 May 2013; Online 18 Jun 2013

Correspondence to: Dr Richard J.Q. McNally, Institute of Health and Society, Newcastle University, Sir James Spence Institute, Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne NE1 4LP, United Kingdom, Tel.: +44-0-191-282-1356; Fax: +44-0-191-282-4724, E-mail: Richard.McNally@ncl.ac.uk

What's new?

The clustering of childhood cancers within specific geographical areas of Great Britain between 1969 and 1993 has raised questions about the possible etiological involvement of environmental exposures. Here, cross-space-time clustering analysis was used to explore associations between different childhood cancers diagnosed in the region during the 1ST January 1969 to 31ST December 1993 is a 25-year period. Significant clustering was identified between cases of Hodgkin lymphoma and intracranial and intraspinal embryonal tumors, while marginal clustering was observed between lymphoid leukemia and Hodgkin lymphoma and between lymphoid leukemia and soft tissue sarcoma. The findings support the idea that common etiological factors could explain the clustering of these different cancers.

Kaposi's sarcoma, in the presence of HIV infection.[16] Associations with parental farming and residence on a farm suggest that infections may play a part in the development of bone cancer.[17] By contrast, there is little evidence supporting a role for infections in the aetiology of Wilms tumor.[18]

Given that a number of childhood cancers exhibit space-time clustering and have an environmental (possibly infectious) component to aetiology, it is plausible that some (or all) of these different diagnostic groups may share a common aetiology. Only one other smaller study has considered this possibility. A previous, more limited, regional study from northwest England found evidence of cross-space-time clustering between cases of childhood leukemia (especially LL) and CNS tumors (especially astrocytoma). This was interpreted as suggesting a common (possibly infectious) aetiological mechanism.[19]

The aim of this study was to test predictions of cross-space-time clustering, which might arise as a result of a common, possible infectious environmental exposure. The study has major strengths. First, it is the largest and most comprehensive study of cross-space-time clustering to date. Second, it uses high quality population-based national incidence data on all childhood cancers, diagnosed with almost complete ascertainment.

Material and Methods

Cases

All cases diagnosed with childhood cancer during the period 1st January 1969 to 31st December 1993 and registered by the National Registry of Childhood Tumours (NRCT) were included in the study. Anonymous case details were obtained from the NRCT, which is population-based and covers the whole of the UK (England, Wales, Scotland and Northern Ireland).[20] Birth addresses were obtainable for ~92% of registered cases. This analysis was restricted to cases resident in Great Britain (England, Wales and Scotland) at time of diagnosis. There were 10 twin pairs. For each twin pair, the earlier diagnosed case was included and the later diagnosed case was excluded from the analysed data sets.

Diagnostic classification

Cases were divided into diagnostic groups using the International Classification of Childhood Cancer, Third Edition (ICCC-3).[21] Diagnostic groups specified *a priori* for analysis were the following: (*i*) LL [ICCC-3 code I (a)], ages 1–4 years; (*ii*) LL [ICCC-3 code I (a)], ages 5–14 years; (*iii*) HL [ICCC-3 code II (a)]; (*iv*) NHL [ICCC-3 code II (b)]; (*v*) astrocytoma [ICCC-3 code III (b)]; (*vi*) intracranial and intraspinal embryonal tumors [IIET, ICCC-3 code III (c)]; (*vii*) soft tissue and other extraosseous sarcomas (STS, ICCC-3 code IX); (*viii*) osteosarcoma [ICCC-3 code VIII (a)]; and (*ix*) renal tumors (ICCC-3 code VI). Cases of LL were divided into two age-groups because cases of the precursor B-cell sub-type dominate the child peak that occurs at ages 1–4 years, whilst cases aged 5–14 years comprise a more mixed set of LL subtypes.[22]

Grid references

Ordnance Survey (OS) provides grid-based maps for the whole of Great Britain (England, Wales and Scotland). For each childhood cancer case, 4-digit Easting and 4-digit Northing OS grid references were assigned to the centroids of the address post-codes at time of birth and at time of diagnosis. This enabled spatial referencing of the Easting and Northing coordinates of both the address at birth and the address at diagnosis to the nearest 0.1 km. To preserve confidentiality of locations, without compromising the analyses, grid references had their origin shifted and have been rotated.

Prior hypotheses

The following aetiological hypotheses were tested:

- (i) Some of the following groups will share a common environmental aetiology: LL (ages 1–4 years) LL (ages 5–14 years), HL, NHL, astrocytoma, IIET, STS and osteosarcoma;
- (ii) Renal tumors will not share an environmental aetiology with other tumor groups.

Space-time combinations

The following space-time combinations were examined: address and date of birth, address at birth and date of diagnosis and address and date of diagnosis. The interpretation of these interactions has been given in detail previously.[3] A space-time interaction between addresses and dates of birth suggests the involvement of an exposure close to the location of birth, occurring *in utero* or soon after birth. It would

indicate a variable latent period between initial exposure and onset of overt disease. A space-time interaction between addresses at birth and dates of diagnosis would suggest an exposure at heterogeneous times after birth with a short or at least constant latent period. A space-time interaction between addresses and dates of diagnosis would suggest the involvement of an exposure close to the location of diagnosis, occurring close to time of diagnosis.

Statistical methods

Cross-space-time clustering analyses were applied to test for associations between pairs of cases from different diagnostic groups. For these analyses, the test is concerned with clustering pairs "*x*, *y*," where "*x*" represents a case from one diagnostic group and "*y*" represents a case from a different diagnostic group.

The Knox test has been used for many analyses of space-time clustering.[23] A generalized version of the Knox test, based on the method of Diggle *et al.*,[24] was used to analyse cross-space-time clustering, consistent with our previously published analyses.[1,2] For a pair of cases, in the Knox test, if dates of an event (birth or diagnosis) are close and residential addresses (at time of birth or diagnosis) are close, then that pair is said to be in "close proximity." The numbers of cross-pairs of cases observed to be in close proximity is counted ($O_{x,y}$) and the expected number of cross-pairs is calculated ($E_{x,y}$). If $O_{x,y}$ is larger than $E_{x,y}$, a significance test is used to determine if there is space-time clustering. "Strength of clustering" is estimated by calculating $S_{x,y} = ([O_{x,y} - E_{x,y}]/E_{x,y}) \times 100$.[3] It should be noted that the overall observed number of pairs of cases that are in the close proximity in the combined group is $O = O_{x,x} + O_{y,y} + O_{x,y}$, where $O_{x,x}$ and $O_{y,y}$ are the observed numbers of pairs of cases that are in close proximity within diagnostic groups "*x*" and "*y*," respectively. Similarly, the overall expected number in the combined group is $E = E_{x,x} + E_{y,y} + E_{x,y}$, where $E_{x,x}$ and $E_{y,y}$ are the numbers of pairs of cases that are expected to be in close proximity within diagnostic groups "*x*" and "*y*," respectively.

Arbitrary choice of critical values for defining close proximity presents a problem with the Knox test. Also, repeated testing using a number of different critical values would lead to multiple testing. A simplification of the method of Diggle *et al.* was used, thereby partially overcoming the arbitrary choice of critical values and avoiding multiple testing.[24] This approach involved a set of 225 calculations, similar to the Knox test. Other analyses have used critical values changed over a prespecified set (for time: $t = 0.1, 0.2, \ldots, 1.5$ years and for space: $s = 0.5, 1, \ldots, 7.5$ km).[2–7] Analyses based on a nearest neighbor (NN) are more likely to be appropriate when there is heterogeneity in population density, such as when both urban and rural localities are included. To allow for variation in population density, we replaced the fixed geographical distances by variable distances to the $(N - 7)$th, $\ldots$, $(N + 7)$th NNs. This method is similar to one originally proposed by Jacquez.[25] In this study, we have only used the NN method. Thus, the main analyses are based on the method of Diggle *et al.*, with the space-time *K*-statistic computed at equally spaced time, but nonequally spaced distances. For birth locations, the mean distance between the 25th NNs was ~5 km, so the set of fixed distance critical values (0.5, 1, 1.5, ..., 7.5 km) were replaced by variable NN critical values (distances between the 18th, ..., 32nd NNs). For diagnosis locations, the mean distance between the 26th NNs was ~5 km, so fixed critical distances were replaced by variable NN values (distances between the 19th, ..., 33rd NNs). For close times, we used the set of critical values: 0.1, 0.2, ..., 1.5 years.

A quantity $R_{x,y}(s, t) = (O_{x,y} - E_{x,y})/\sqrt{E_{x,y}}$ is defined. Then the observed value of the *K*-statistic is calculated as $K_O = \sum R_{x,y}(s, t)$, summed over all 225 critical pairs of values for *s* and *t*. Since the distribution of the *K*-statistic is not known, it is estimated by simulation using 999 random temporal permutations. For those analyses where $p < 0.001$ we used 9,999 simulations. A realization of the *K*-statistic was obtained at each simulation by randomly reallocating the dates of the event (birth or diagnosis) to each of the cases in the analysis. Comparison of the observed value with the simulated distribution allowed statistical significance to be assessed, using a one-sided test. To allow for multiple testing, the level of formal statistical significance in all analyses was taken as $p < 0.00067$ (determined as 0.05/75).[26] Marginal significance was defined as $0.01 > p \geq 0.00067$.

Since the *K*-statistic does not provide a measure of the magnitude of an effect, $S_{x,y}$, determined from the Knox test is given as an indication of magnitude (with critical values for closeness in space taken as distances between 25th NNs for births, 26th NNs for diagnoses and for closeness in time as 1 year). It should be noted that it is possible for a small value of $S_{x,y}$ to result when the real effect is large if space-time clustering occurs at a different scale (note that to simplify notation "$S_{x,y}$" is denoted "*S*" in subsequent parts of this report).

The distributions of distances between both birth and diagnosis locations were highly skewed. To test whether population density was associated with cross-space-time clustering, separately for birth and diagnosis locations, cases were split into two groups: 50% were classified as belonging to a "more densely populated" group and 50% were classified as belonging to a "less densely populated" group on the basis of whether the 25th (for births) or 26th (for diagnoses) NN was nearer or further away than the median distance of the 25th or 26th NN. Analysis by population density was then done by considering cross-clustering pairs that included at least 1 case from the "more densely populated" category and cross-clustering pairs that included at least 1 case from the "less densely populated" category. It must be noted that analyses of population density (especially analyses of "less densely populated: any" cross-clustering pairs) may be diluted due to edge effects since "less densely populated" areas are sometimes not contiguous.

Table 1. Numbers of cases for analyses of cross-space-time clustering of childhood cancer in Great Britain, 1969–1993

	Total		Males		Females	
Diagnostic group	Birth	Diagnosis	Birth	Diagnosis	Birth	Diagnosis
LL(ages 1–4)	4,140	4,343	2,361	2,483	1,779	1,860
LL(ages 5–14)	3,335	3,810	1,939	2,216	1,396	1,594
HL	1,226	1,364	851	953	375	411
NHL	1,485	1,678	1,050	1,192	435	486
Astrocytoma	2,614	2,824	1,306	1,422	1,308	1,402
IIET	1,410	1,548	889	976	521	572
STS	1,921	2,101	1,083	1,185	838	916
Osteosarcoma	669	811	334	408	335	403
Renal tumors	1,820	1,889	920	945	900	944

HL: Hodgkin lymphoma; IIET: intracranial and intraspinal embryonal tumors; LL: lymphoid leukemia; NHL: non-Hodgkin lymphoma; STS: soft tissue sarcomas.

Table 2. Cross-space-time clustering of cases of HL and NHL with other diagnostic groups

Diagnostic groups	Place of birth and date of birth	Place of birth and date of diagnosis	Place of diagnosis and date of diagnosis
HL × NHL	$p = 0.91$	$p = 0.58$	$p = 0.73$
HL × Astrocytoma	$p = 0.42$	$p = 0.30$	$p = 0.10$
HL × IIET	$p = 0.31$	$p < 0.0001$[1] ($S = 35.20\%$)	$p = 0.69$
HL × STS	$p = 0.05$	$p = 0.55$	$p = 0.68$
HL × Osteosarcoma	$p = 0.60$	$p = 0.72$	$p = 0.34$
HL × Renal tumors	$p = 0.99$	$p = 0.87$	$p = 0.71$
NHL × Astrocytoma	$p = 0.34$	$p = 0.03$	$p = 0.24$
NHL × IIET	$p = 0.27$	$p = 0.50$	$p = 0.73$
NHL × STS	$p = 0.22$	$p = 0.04$	$p = 0.18$
NHL × Osteosarcoma	$p = 0.41$	$p = 0.12$	$p = 0.87$
NHL × Renal tumors	$p = 0.63$	$p = 0.69$	$p = 0.95$

[1]Statistically significant, defined as $p < 0.00067$.
HL: Hodgkin lymphoma; IIET: intracranial and intraspinal embryonal tumors; NHL: non-Hodgkin lymphoma; *S*: strength of clustering ([{observed − expected}/expected] × 100%, counts of pairs that are close in space and time); STS: soft tissue sarcomas.

K-statistic and Knox analyses were run using programs written in FORTRAN (note that *K*-statistic analyses can also be run using the Splancs package in R).[27–29]

Results

The dataset included 32,295 cases of childhood cancer, with complete diagnosis data (address and date) and 29,553 cases with complete birth data (address and date). Only groups for which there were clear prior hypotheses were analysed (Table 1).

Using place of birth and date of diagnosis (Table 2) there was statistically significant cross-space-time clustering between cases of HL and IIET ($p < 0.0001$; $S = 35.20\%$). Using place and date of birth (Table 3) there was marginally significant cross-space-time clustering between cases of LL (ages 5–14 years) and HL ($p = 0.0019$; $S = 19.27\%$). Using place and date of diagnosis (Table 3) there was marginally significant cross-space-time clustering between cases of LL (ages 1–4 years) and STS ($p = 0.0041$; $S = 11.09\%$). Results of analyses of space-time clustering based on "more densely populated: any" and "less densely populated: any" clustering pairs are presented (Tables 4 and 5). There was significant cross-space-time clustering between cases of HL and IIET, which was significant for "more densely populated: any" clustering pairs ($p = 0.0004$, $S = 82.08\%$) and marginally significant for "less densely populated: any" clustering pairs ($p = 0.0039$, $S = 82.44\%$; Table 5). There was marginally significant cross-space-time clustering between cases of LL (5–14 years) and HL, which was confined to "more densely populated: any" clustering pairs ($p = 0.0028$, $S = 61.02\%$ Table 4a).

Discussion

The analyses have been performed using well specified statistical methods on very good quality population-based incidence data. It is the largest study of cross-space-time clustering that has been done, analyzing 32,295 cases. Highly novel statistically significant cross-space-time clustering has been identified between cases of HL and IIET; and marginally significant cross-space-time clustering between cases of LL (ages 5–14 years) and HL; and LL (ages 1–4 years) and STS. Thus there was support for the first prior hypothesis that some of the following groups will share a common environmental aetiology: LL (ages 1–4 years), LL (ages 5–14 years), HL, NHL, astrocytoma, IIET, STS and osteosarcoma, since all of these apart from NHL, astrocytoma and osteosarcoma demonstrated some evidence of cross-space-time clustering. There was also support for the second prior hypothesis that renal tumors will not share an environmental aetiology with

Table 3. Cross-space-time clustering of cases of LL with other diagnostic groups

Diagnostic groups	Place of birth and date of birth	Place of birth and date of diagnosis	Place of diagnosis and date of diagnosis
LL (ages 1–4) × HL	$p > 0.9999$	$p = 0.67$	$p = 0.79$
LL (ages 1–4) × NHL	$p = 0.9998$	$p = 0.47$	$p = 0.52$
LL (ages 1–4) × Astrocytoma	$p = 0.93$	$p = 0.67$	$p = 0.93$
LL (ages 1–4) × IIET	$p = 0.94$	$p = 0.46$	$p = 0.90$
LL (ages 1–4) × STS	$p = 0.46$	$p = 0.51$	$p = 0.0041$[1] ($S = 11.09\%$)
LL (ages 1–4) × Osteosarcoma	$p = 0.99$	$p = 0.43$	$p = 0.29$
LL (ages 1–4) × Renal tumors	$p = 0.89$	$p = 0.53$	$p = 0.80$
LL (ages 5–14) × HL	$p = 0.0019$[1] ($S = 19.27\%$)	$p = 0.59$	$p = 0.43$
LL (ages 5–14) × NHL	$p = 0.21$	$p = 0.17$	$p = 0.54$
LL (ages 5–14) × Astrocytoma	$p = 0.77$	$p = 0.25$	$p = 0.78$
LL (ages 5–14) × IIET	$p = 0.46$	$p = 0.62$	$p = 0.46$
LL (ages 5–14) × STS	$p = 0.75$	$p = 0.80$	$p = 0.92$
LL (ages 5–14) × Osteosarcoma	$p = 0.39$	$p = 0.81$	$p = 0.28$
LL (ages 5–14) × Renal tumors	$p = 0.77$	$p = 0.65$	$p = 0.93$

[1]Marginally significant, defined as $0.01 > p \geq 0.00067$.
HL: Hodgkin lymphoma; IIET: intracranial and intraspinal embryonal tumors; LL: lymphoid leukemia; NHL: non-Hodgkin lymphoma; *S*: strength of clustering ([{observed − expected}/expected] × 100%, counts of pairs that are close in space and time).
STS, soft tissue sarcomas

other tumor groups, since renal tumors did not cross-cluster with any other group.

The finding of significant cross-clustering between cases of HL and IIET indicates that cases of these two distinct diagnostic groups occur together at similar places of birth and similar times of diagnosis. This suggests the aetiological involvement of a common transient exposure at heterogeneous times after birth with a short or at least constant latent period. The analyses indicated that both urban and rural localities were involved. A previous study from Yorkshire identified space-time clustering amongst cases of IIET.[7] A number of transient environmental agents may play a role, including pesticides, insecticides, pollutants and infections. However, infections are a plausible candidate as they have been implicated in both diseases.[13,30–33]

The finding of marginally significant cross-clustering between cases of LL (ages 5–14 years) and HL indicates that cases of these two distinct diagnostic groups occur together at similar places and times of birth. This suggests that these diagnoses may share a common transient aetiological factor, occurring *in utero* or around the time of birth. Furthermore, cross-clustering was confined to clustering pairs that included at least one case from a "more densely populated" area, suggesting an association with more urban locations. In marked contrast, there was no evidence of cross-clustering between cases of LL (ages 1–4 years) and HL. This indicates that younger cases of LL may not share a common aetiological factor with cases of HL. We interpret these findings in the context of other epidemiological evidence regarding the aetiology of LL and HL. Younger cases of LL, forming the childhood peak, mainly comprise the precursor B-cell subtype, whilst older cases have a greater mixture of subtypes.[22] Greaves suggested that the precursor B-cell subtype has a distinctive aetiology related to delayed exposure to common infections.[10] Smith also proposed that *in utero* exposure to infection is responsible for the childhood peak in LL (which is mostly precursor B-cell).[34] There is a lack of similar distinctive hypotheses concerning nonprecursor B-cell LL. Kinlen proposed that childhood leukemia excesses are linked with very unusual population mixing, but did not specify the subtype or age range.[35,36] For LL no single agent has been conclusively linked to aetiology.[12] In contrast, HL has been linked with specific direct transforming infectious agents, including Epstein-Barr virus.[31,32] EBV is especially associated with paediatric cases and the mixed cellularity subtype.[37,38] Our finding of cross-clustering, involving urban settings, suggests that some older paediatric cases of LL (possibly nonprecursor B-cell) may arise due to the same directly transforming agent as some cases of HL. Furthermore, the time of exposure (*in utero* or around the time of birth) suggests that there is a long latency until occurrence of overt disease. Other events are also likely to be involved in the process, as postulated by Knudson.[9] It is also possible that other transient environmental exposures may be implicated including pesticides, fungicides, benzene, consumption of seasonal fruit and vegetables.[11]

The finding of marginally significant cross-clustering between cases of LL (ages 1–4 years) and STS indicates that cases of these two distinct diagnostic groups occur together at similar places and times of diagnosis. This suggests the aetiological involvement of a common transient exposure around the time of diagnosis. Conversely, there was no evidence of cross-clustering between cases of LL (ages 5–14

Table 4. Cross-space-time clustering of cases of LL with other diagnostic groups, by level of population density

Diagnostic groups	Place of birth and date of birth	Place of birth and date of diagnosis	Place of diagnosis and date of diagnosis
(a) "More densely populated: any" cross-clustering pairs			
LL (ages 1–4) × HL	$p = 0.9998$	$p = 0.31$	$p = 0.43$
LL (ages 1–4) × NHL	$p = 0.9992$	$p = 0.31$	$p = 0.71$
LL (ages 1–4) × Astrocytoma	$p = 0.72$	$p = 0.71$	$p = 0.84$
LL (ages 1–4) × IIET	$p = 0.96$	$p = 0.68$	$p = 0.72$
LL (ages 1–4) × STS	$p = 0.77$	$p = 0.46$	$p = 0.02$
LL (ages 1–4) × Osteosarcoma	$p = 0.97$	$p = 0.77$	$p = 0.06$
LL (ages 1–4) × Renal tumors	$p = 0.76$	$p = 0.36$	$p = 0.71$
LL (ages 5–14) × HL	$p = 0.0028$[1] ($S = 61.02\%$)	$p = 0.18$	$p = 0.80$
LL (ages 5–14) × NHL	$p = 0.34$	$p = 0.18$	$p = 0.64$
LL (ages 5–14) × Astrocytoma	$p = 0.28$	$p = 0.32$	$p = 0.73$
LL (ages 5–14) × IIET	$p = 0.12$	$p = 0.91$	$p = 0.87$
LL (ages 5–14) × STS	$p = 0.68$	$p = 0.54$	$p = 0.77$
LL(ages 5–14) × Osteosarcoma	$p = 0.16$	$p = 0.59$	$p = 0.62$
LL (ages 5–14) × Renal tumors	$p = 0.86$	$p = 0.55$	$p = 0.95$
(b) "Less densely populated: any" cross clustering pairs			
LL (ages 1–4) × HL	$p = 0.98$	$p = 0.88$	$p = 0.89$
LL (ages 1–4) × NHL	$p = 0.998$	$p = 0.69$	$p = 0.23$
LL (ages 1–4) × Astrocytoma	$p = 0.90$	$p = 0.68$	$p = 0.89$
LL (ages 1–4) × IIET	$p = 0.56$	$p = 0.25$	$p = 0.73$
LL (ages 1–4) × STS	$p = 0.48$	$p = 0.69$	$p = 0.02$
LL (ages 1–4) × Osteosarcoma	$p = 0.84$	$p = 0.20$	$p = 0.44$
LL (ages 1–4) × Renal tumors	$p = 0.87$	$p = 0.51$	$p = 0.85$
LL (ages 5–14) × HL	$p = 0.05$	$p = 0.77$	$p = 0.14$
LL (ages 5–14) × NHL	$p = 0.28$	$p = 0.48$	$p = 0.60$
LL (ages 5–14) × Astrocytoma	$p = 0.98$	$p = 0.17$	$p = 0.84$
LL (ages 5–14) × IIET	$p = 0.56$	$p = 0.10$	$p = 0.18$
LL (ages 5–14) × STS	$p = 0.37$	$p = 0.95$	$p = 0.97$
LL (ages 5–14) × Osteosarcoma	$p = 0.93$	$p = 0.80$	$p = 0.07$
LL (ages 5–14) × Renal tumors	$p = 0.72$	$p = 0.39$	$p = 0.85$

[1]Marginally significant, defined as $0.01 > p \geq 0.00067$.
HL: Hodgkin lymphoma; IIET: intracranial and intraspinal embryonal tumors; LL: lymphoid leukemia; NHL: non-Hodgkin lymphoma; *S*: strength of clustering ([{observed − expected} / expected] × 100%, counts of pairs that are close in space and time); STS: soft tissue sarcomas

years) and STS. This suggests that older cases of LL may not arise from the same common aetiology as cases of STS. It has been postulated that younger cases of LL, which comprise the childhood peak and are mainly of the precursor B-cell subtype, arise from delayed exposure to common infections.[10] There are no similar mechanisms suggested for STS. However, infectious links for STS are plausible, since HHV8 (in the presence of HIV) is causally associated with Kaposi's sarcoma.[16] Other transient environmental exposures that have been implicated in the aetiology of STS include occupational chemicals, phenoxyacetic acid herbicides, chlorophenols and dioxin.[39] The present finding suggests that, at least for some cases, the final event precipitating a LL (in those aged 1–4 years) or STS (ages 0–14 years) may arise from the same environmental agent (possibly an infection).

A previous study from northwest England found cross-clustering between LL and astrocytoma.[19] In this national study, we did not find such an association. If localized transient agents (such as infections) are involved in aetiology, then it may be predicted that some links are only found in certain geographical regions. Thus, findings from this national study do not necessarily refute the earlier region-specific results, although some could have arisen by chance.

Table 5. Cross-space-time clustering of cases of HL and NHL with other diagnostic groups, by level of population density

Diagnostic groups	Place of birth and date of birth	Place of birth and date of diagnosis	Place of diagnosis and date of diagnosis
(a) "More densely populated: any" cross-clustering pairs			
HL × NHL	$p = 0.77$	$p = 0.37$	$p = 0.55$
HL × Astrocytoma	$p = 0.25$	$p = 0.74$	$p = 0.43$
HL × IIET	$p = 0.40$	$p = 0.0004$[1] ($S = 82.08\%$)	$p = 0.81$
HL × STS	$p = 0.34$	$p = 0.34$	$p = 0.64$
HL × Osteosarcoma	$p = 0.31$	$p = 0.95$	$p = 0.45$
HL × Renal tumors	$p = 0.69$	$p = 0.46$	$p = 0.29$
NHL × Astrocytoma	$p = 0.03$	$p = 0.02$	$p = 0.15$
NHL × IIET	$p = 0.76$	$p = 0.53$	$p = 0.86$
NHL × STS	$p = 0.49$	$p = 0.31$	$p = 0.57$
NHL × Osteosarcoma	$p = 0.47$	$p = 0.14$	$p = 0.71$
NHL × Renal tumors	$p = 0.74$	$p = 0.84$	$p = 0.90$
(b) "Less densely populated: any" cross-clustering pairs			
HL × NHL	$p = 0.92$	$p = 0.86$	$p = 0.38$
HL × Astrocytoma	$p = 0.75$	$p = 0.11$	$p = 0.07$
HL × IIET	$p = 0.53$	$p = 0.0039$[2] ($S = 82.44\%$)	$p = 0.35$
HL × STS	$p = 0.02$	$p = 0.71$	$p = 0.67$
HL × Osteosarcoma	$p = 0.69$	$p = 0.25$	$p = 0.58$
HL × Renal tumors	$p = 0.996$	$p = 0.97$	$p = 0.98$
NHL × Astrocytoma	$p = 0.87$	$p = 0.15$	$p = 0.53$
NHL × IIET	$p = 0.10$	$p = 0.60$	$p = 0.67$
NHL × STS	$p = 0.20$	$p = 0.0107$	$p = 0.14$
NHL × Osteosarcoma	$p = 0.18$	$p = 0.15$	$p = 0.52$
NHL × Renal tumors	$p = 0.42$	$p = 0.35$	$p = 0.84$

[1]Statistically significant, defined as $p < 0.00067$.
[2]Marginally significant, defined as $0.01 > p \geq 0.00067$.
HL: Hodgkin lymphoma; IIET: intracranial and intraspinal embryonal tumors; NHL: non-Hodgkin lymphoma; *S*: strength of clustering ([{observed − expected}/expected] × 100%, counts of pairs that are close in space and time); STS: soft tissue sarcomas.

However, it should be noted that there will be some overlap as a number of the cases in the earlier regional studies from northwest England will also be present in this national study.

The study has some limitations. It must be acknowledged that temporal trends in the incidence of certain childhood cancers may be at least partly influenced by improvements in diagnostic techniques.[40] However, such changes will occur over widespread geographical regions are so are highly unlikely to induce artefactual space-time clustering. The cases were diagnosed *via* the National Health Service in the UK. This is a socialized heath care system with uniform systems and universal coverage. All data were obtained from the NRCT, which has almost complete ascertainment. Thus it is not plausible that diagnostic changes at a local scale would have led to the patterns observed. The choice of the formal significance level of $p < 0.00067$ has guarded against the possible effect of multiple testing leading to spurious nominally significant results. However, strictly adjusting for multiple testing may be too conservative. Therefore, we also defined $0.01 > p \geq 0.00067$ as marginally significant, acknowledging that chance may have played a role in some of these findings. It is possible that apparent space-time clustering may be seen due to shifts in small-area populations over short time periods. A method for adjustment for this type of population shift has been suggested by Kulldorff and Hjalmars.[41] Unfortunately, data on small-area populations for short time periods are not available in GB. Thus, it was not possible to make any adjustments for such putative population shifts. However, statistically significant cross-space-time clustering was specific to particular pairs of diagnostic groups, for which there is some evidence for an environmental (especially infectious) origin. The distinctive findings of cross-space-time clustering provide a strong argument against the possibility that population shifts have led to these observations.

In conclusion, the analyses have been performed on high quality population-based incidence data using appropriate statistical methods. The highly novel findings of cross-space-time clustering from this study are consistent with possible common aetiological factors between different diagnostic groups. Although these findings should be treated tentatively, specifically they suggest a common aetiology for the following pairs of diagnostic groups: HL and IIET; older cases of LL and HL; and younger cases of LL and STS. For cross-clustering groups, the possibility of common infectious mechanisms should be explored.

References

1. McNally RJ, Bithell JF, Vincent TJ, et al. Space-time clustering of childhood cancer around the residence at birth. *Int J Cancer* 2009;124:449–55.
2. McNally RJ, Alexander FE, Bithell JF. Space-time clustering of childhood cancer in Great Britain: a national study, 1969-1993. *Int J Cancer* 2006;118: 2840–6.
3. Birch JM, Alexander FE, Blair V, et al. Space-time clustering patterns in childhood leukaemia support a role for infection. *Br J Cancer* 2000;82:1571–6.
4. McNally RJ, Cairns DP, Eden OB, et al. An infectious aetiology for childhood brain tumours? Evidence from space-time clustering and seasonality analyses. *Br J Cancer* 2002;86:1070–7.
5. McNally RJ, Alexander FE, Birch JM. Space-time clustering analyses of childhood acute lymphoblastic leukaemia by immunophenotype. *Br J Cancer* 2002;87:513–5.
6. McNally RJ, Kelsey AM, Eden OB, et al. Space-time clustering patterns in childhood solid tumours other than central nervous system tumours. *Int J Cancer* 2003; 103:253–8.
7. McNally RJ, James PW, Picton SV, et al. Space-time clustering of childhood central nervous system tumours in Yorkshire, UK. *BMC Cancer* 2012;12:13.
8. Birch JM. Genes and cancer. *Arch Dis Child* 1999;80:1–3.
9. Knudson AG Jr. A two-mutation model for human cancer. *Adv Viral Oncol* 1987;7:1-17.
10. Greaves MF. Speculations on the cause of childhood acute lymphoblastic leukemia. *Leukemia* 1988;2:120–5.
11. McNally RJ, Parker L. Environmental factors and childhood acute leukemias and lymphomas. *Leuk Lymphoma* 2006;47:583–98.
12. McNally RJ, Eden TO. An infectious aetiology for childhood acute leukaemia: a review of the evidence. *Br J Haematol* 2004;127:243–63.
13. Harding NJ, Birch JM, Hepworth SJ, et al. Infectious exposure in the first year of life and risk of central nervous system tumors in children: analysis of day care, social contact, and overcrowding. *Cancer Causes Control* 2009;20:129–36.
14. Barbanti-Brodano G, Martini F, De Mattei M, et al. BK and JC human polyomaviruses and simian virus 40: natural history of infection in humans, experimental oncogenicity and association with human tumors. *Adv Virus Res* 1998;50: 69–99.
15. Menegaux F, Olshan AF, Neglia JP, et al. Day care, childhood infections, and risk of neuroblastoma. *Am J Epidemiol* 2004;159:843–51.
16. Viejo-Borbolla A, Schulz TF. Kaposi's sarcoma-associated herpesvirus (KSHV/HHV8): key aspects of epidemiology and pathogenesis. *AIDS Rev* 2003;5:222–9.
17. Eyre R, Feltbower RG, Mubwandarikwa E, et al. Epidemiology of bone tumours in children and young adults. *Pediatr Blood Cancer* 2009;53: 941–52.
18. Chu A, Heck JE, Ribeiro KB, et al. Wilms' tumour: a systematic review of risk factors and meta-analysis. *Paediatr Perinat Epidemiol* 2010; 24:449–69.
19. McNally RJ, Eden TO, Alexander FE, et al. Is there a common aetiology for certain childhood malignancies? Results of cross-space-time clustering analyses. *Eur J Cancer* 2005;41: 2911–6.
20. 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.
21. Steliarova-Foucher E, Stiller C, Lacour B, et al. International classification of childhood cancer, third edition. *Cancer* 2005;103:1457–67.
22. McNally RJ, Birch JM, Taylor GM, et al. Incidence of childhood precursor B-cell acute lymphoblastic leukaemia in north-west England. *Lancet* 2000;356:485–6.
23. Knox EG. The detection of space-time interactions. *Appl Stats* 1964;13:25–9.
24. Diggle PJ, Chetwynd AG, Haggkvist R, et al. Second-order analysis of space-time clustering. *Stat Methods Med Res* 1995;4:124–36.
25. Jacquez GM. A k nearest neighbour test for space-time interaction. *Stat Med* 1996;15: 1935–49.
26. Holm S. A simple sequentially rejective multiple testing procedure. *Scand J Stat* 1979;6:65–70.
27. Rowlingson BS, Diggle PJ. Splancs: spatial point pattern analysis code in S-Plus. *Comput Geosci* 1993;19:627–55.
28. Bivand R, Gebhardt A. Implementing functions for spatial statistical analysis using the R language. *J Geograph Syst* 2000;2:307–17.
29. Pike MC, Bull D. Algorithm AS69: Knox test for space-time clustering in epidemiology. *J R Stat Soc Ser C (Appl Stat)* 1974;23:92–5.
30. Little J. Epidemiology of childhood cancer. Lyon, France: IARC Scientific Publications No. 149, 1999.
31. 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:S27–S32.
32. Hjalgrim H, Engels EA. Infectious aetiology of Hodgkin and non-Hodgkin lymphomas: a review of the epidemiological evidence. *J Int Med* 2008; 264:537–48.
33. Yeni-Komshian H, Holly EA. Childhood brain tumours and exposure to animals and farm life: a review. *Paediatr Perinat Epidemiol* 2000;14: 248–56.
34. Smith M. Considerations on a possible viral etiology for B-precursor acute lymphoblastic leukemia of childhood. *J Immunother* 1997;20: 89–100.
35. Kinlen L. Evidence for an infective cause of childhood leukaemia: comparison of a Scottish New Town with nuclear reprocessing sites in Britain. *Lancet* 1988;2:1323–7.
36. Kinlen LJ. Epidemiological evidence for an infective basis in childhood leukaemia. *Br J Cancer* 1995;71:1–5.
37. Glaser SL, Lin RJ, Stewart SL, et al. Epstein-Barr virus-associated Hodgkin's disease: epidemiologic characteristics in international data. *Int J Cancer* 1997;70:375–82.
38. Preciado MV, Diez B, Grinstein S. Epstein Barr virus in Argentine pediatric Hodgkin's disease. *Leuk Lymphoma* 1997;24:283–90.
39. Froehner M, Wirth MP. Etiologic factors in soft tissue sarcomas. *Onkologie* 2001;24:139–42.
40. Kroll ME, Carpenter LM, Murphy MF, et al. Effects of changes in diagnosis and registration on time trends in recorded childhood cancer incidence in Great Britain. *Br J Cancer* 2012;107: 1159-62.
41. Kulldorff M, Hjalmars U. The Knox method and other tests for space-time interaction. *Biometrics* 1999;55:544–52.