Journal of Epidemiology and Community Health 1992; **46**: 566–572

Leukaemia clusters in Great Britain. 1. Space-time interactions

E G Knox, Estelle Gilman

Abstract

***Study objective*—The aim was to test a large set of childhood leukaemia and lymphoma registrations for the presence of clusters in space and in time.**

***Design*—The study was a space-time cluster analysis.**

***Setting*—England, Wales and Scotland.**

***Patients*—All registrations for leukaemia and lymphoma between 1966 and 1983 in children aged 0 to 14 years were examined. The records included date and age of registration, sex, diagnosis, and the map reference of the postcode of residence. Of the 9411 registrations, 8888 were suitable for inclusion.**

***Main results*—There was a statistically significant excess of case pairs occurring jointly within 0·5 km and 60 d of each other: 68 pairs compared with 50·0 expected. The excess was detectable in central England, in the north of England and Scotland, but not in the south west of England. It was concentrated within the age band 4 to 7 years and among the lymphatic leukaemias. Several potential artefacts were considered and excluded, but the possibility remained that clustered detections might be triggered by haematological examinations undertaken for some communicable disease.**

***Conclusions*—There was strong evidence of joint spatial-temporal clustering, with an excess of pairs separated by very short time and distance intervals. The causes are probably biological rather than artefactual, but further work will be necessary in order to exclude the latter.**

J Epidemiol Community Health 1992; **46**: 566–572

We recently reported clustering among childhood leukaemias and lymphomas in Great Britain.[1] The clusters occurred jointly in space and in time, as judged by the dates and the addresses at registration. The full database consisted of all registrations (9411) of leukaemias and of non-Hodgkin's lymphomas in children aged 0–14 years in England, Wales, and Scotland between 1966 and 1983. The analysis itself was based upon a subset with the most accurately recorded dates and times. More pairs of registrations occurred jointly in the same calendar month and within a short distance of each other than should occur by chance. The cluster pattern was present in several different diagnostic subsets. It was also present among pairs spanning different diagnoses and different age groups. Space-time interaction in leukaemia had been described before[2–6] but never on the basis of so large a data set, assembled and validated to a uniform standard, and free of suspicion that the material was tested because clustering had already been recognised within it.

Here, we report a more detailed analysis. We address four specific questions. First, could the clustering be an artefact of the registration process? It is unlikely with these diseases, and with a high grade registration system,[7] that it could represent time limited patches of overdiagnosis or of varied ascertainment level. However, one case might precipitate the registration of another close by, which had previously been overlooked. Second, could the clustering be an artefact arising from temporal-spatial instability of the population itself? If the density pattern of a population should drift, then any events arising within the population will reflect that changing pattern. Third, we must ask whether the interaction is statistically primary, genuinely dependent upon the dates and places of registration, or whether it is statistically secondary to an interaction based upon another date and place (of birth, say), the temporal-spatial coordinates of which are correlated with those of registration. Fourth, we must ask what biological models—including infective or toxic or radiation exposures—are compatible with the temporal-spatial patterns described.

Methods

THE DATA

The origins, the preparation, the validation, and the detailed contents of the data set are fully described elsewhere.[7] It consisted of 9411 registrations of leukaemias and non-Hodgkin's lymphomas in children aged 0–14 years, in England, Wales, and Scotland, between 1966 and 1983. This national registration file was abridged and consolidated before presentation to research groups in order to preserve confidentiality. Each abridged registration record comprised a unique index number, the sex of the child, the ICDO coded type and site of the neoplasm, the "anniversary date" (day, month, year), the age at anniversary-0 in completed months, the census tract and the enumeration district codes, and a map reference indicating the position of the postcode. Anniversary-0 is the date on which the diagnosis was made and treatment begun.

The postcode coordinates recorded in the file had been obtained initially from the Central Postcode Directory. The Directory records its postcode map references to a resolution of 100 metres, or 10 metres in Scotland. They relate to the first address within the postcode. These map references were then checked, corrected, and supplemented—where this was necessary and possible—against a current commercial index of

Department of Public Health and Epidemiology, University of Birmingham, Edgbaston, Birmingham B15 2TJ, United Kingdom
E G Knox
E Gilman

Correspondence to:
Professor Knox

Accepted for publication March 1992

postcode centroids, supplied by Pinpoint Analysis Ltd. The map references finally entered to the consolidated/abridged research file were recorded as seven digit eastings and seven digit northings, formally correct to one metre. However, many of them—those not cross checked against the Pinpoint index—were still in practical terms recorded only to a resolution of 100 metres. For this and for other reasons—to which we return later—the actual resolutions of these values are to a degree uncertain. However, the case records contained indicators of the origins of the addresses, and of the presumed reliability of the allocated postcodes and coordinates, and this enabled us to exclude from analysis those of known doubtful validity.

The relationships between these first address and centroid coordinates, and the actual addresses within the postcodes, are also uncertain; but a published analysis claims[8] that over 50% of the coordinates are accurate to within 100 metres, and 90% to within 400 metres. There are about 17 addresses per postcode, so that most individual addresses would then be represented to an accuracy of about 200 metres. Address separations less than this will generally have identical coordinates; although some will not; and some with wider separations will. We shall pursue the issue of resolution at a later stage.

The date of diagnosis/registration was also sometimes uncertain and substitute dates had sometimes been entered. Again, however, the record contained an indicator of reliability, and doubtful dates were excluded from analysis. Using these reliability indicators we prepared a subfile which excluded all those dates and places the inaccuracy of which might invalidate our conclusions. Diagnoses were then blocked into three main groups comprising (a) lymphocytic and unspecified leukaemias (ICD 980·0–980·4, 982·0–982·5, 985·0), (b) myeloid leukaemias (ICD 984·0–984·9, 985·1–994·0), and (c) the non-Hodgkin's lymphomas together with the unspecified lymphomas (ICD 951·1–975·0). A number of other diagnoses—for example specified leukaemias and lymphomas not falling within these groups—were excluded. This resulted in a reduced file of 8889 registrations.

Dates of registration were converted to "day of 20th century", and approximate dates of birth were calculated by subtracting "age in completed months at registration". Clearly, these estimated dates of birth have an error of about plus or minus 15 days. Nevertheless, this procedure allowed us to identify a pair of like sexed concordant twins with the same type of leukaemia and the same map position, both registered during the first year of life. We excluded one of the pair, treating the other as a single case, thus leaving 8888 events for analysis.

ANALYSIS

The standard methods for the detection of space-time interactions are based upon an examination of all possible pairs. There are n(n-1)/2 possible pairs from n events; for example there are 39 493 828 pairs from among the 8888 events. This permits a tabulation of distance separations against time separations, to see whether there is a disproportionate excess of pairs combining short distances with short time intervals.

For a dataset which extended over 18 years and over 1000 km, the question arises whether a false interaction might be generated through movements in the density pattern of the population at risk. We took precautions against such mis-

Table I Spatial-temporal distribution of pairs of leukaemias registered within 1000 days and 100 km of each other. Registrations of leukaemias and lymphomas (8888 cases)

Time separation (days)	*Distance separation (km)*							
	0–0·5	*0·5–1·0*	*1·0–1·5*	*1·5–2·0*	*2·0–5·0*	*5·0–20·0*	*20·0–100·0*	*0–100·0*
0–15								
Observed	22	23	43	36	385	3867	32 031	36 407
Expected	12·99	24·26	31·44	40·10	380·78	3915·26	32 002·17	—
O/E	1·69	0·95	1·37	0·90	1·01	0·99	1·00	—
p	0·014	—	0·029	—	—	—	—	—
16–30								
Observed	13	25	27	43	332	3735	30 548	34 723
Expected	12·40	23·73	29·98	38·24	363·16	3734·17	30 521·91	—
O/E	1·05	1·05	0·90	1·12	0·91	1·00	1·00	—
p	0·061	—	—	—	—	—	—	—
31–60								
Observed	33	46	53	60	691	7496	60 590	68 969
Expected	24·62	45·94	59·55	75·96	721·34	7417·03	60 624·54	—
O/E	1·34	1·00	0·89	0·79	0·96	1·01	1·00	—
p		—	—	—	—	—	—	—
61–120								
Observed	48	102	125	148	1489	14 854	121 844	138 610
Expected	49·47	92·35	119·69	152·67	1449·71	14 906·03	121 839·78	
O/E	0·97	1·10	1·04	0·97	1·03	1·00	1·00	—
p	—	—	—	—	—	—	—	—
121–365								
Observed	182	342	459	630	5758	59 697	487 043	554 111
Expected	197·78	369·18	478·47	610·32	5795·39	59 589·91	487 069·94	—
O/E	0·92	0·93	0·96	1·03	0·99	1·00	1·00	—
p	—	—	—	—	—	—	—	—
366–1000								
Observed	482	918	1180	1490	14 201	145 363	1 188 861	1 352 495
Expected	482·74	901·12	1167·87	1489·70	14 145·62	145 449·30	1 188 858·65	—
O/E	1·00	1·02	1·01	1·00	1·00	1·00	1·00	
p	—	—	—	—	—	—	—	—
Total	780	1456	1887	2407	22 856	235 012	1 920 917	2 185 315

O/E = ratio of observed to expected registrations.
The 2185315 pairs registered within 1000 d and 100 km of each other represent 5·53% of all possible pairs (39493828) from 8888 registrations.
Only significant p values are shown.

interpretation in our previous paper[1] by examining the material in three separate time periods. However there was still a danger that major movements of population might have taken place within one or more of these intervals; so in the present analysis we have adopted another and more restrictive approach. The events were sorted in ascending date order, and the material was then examined for interactions within a moving time window of arbitrary length—usually 1000 days. Technically, each case was paired with every subsequent case within this upper limit. This approach has the dual advantage of greatly reducing the numbers of non-informative pairs examined, while limiting even further the scope for detecting and misinterpreting major population drifts. Its only disadvantage is that it effectively limits the examination to detecting geographical movements which occur within this time. We also restricted the geographical scope of the interaction search, usually to within 100 km.

We also made use of further novel techniques. One was a modification to the "all possible pairs" technique in which the distances between pairs were standardised for the local density of events. This was carried out using a modification of a method described and used elsewhere.[9] Each record was examined to see how many other cases were located within a nominated distance, and the number (plus 1) stored as a "proximity count" within the registration record. Different nominated distances—typically 4 km—were used in different runs. These individual proximity counts were later divided by the overall mean proximity count, the ratio then providing a crude indicator of the density of other registrations in the immediate vicinity of the index case. In the subsequent analysis of pairs, each of the geographical separations was then stretched or shrunk according to the local density indices associated with the two registrations. Technically, the measured distance was multiplied by the harmonic mean of the two density ratios. The general intention was to provide a transformed index of distance scaled in terms of the numbers of intermediate personal contacts, rather than the number of metres.

The second novel technique was an elaboration of the examination based upon pairs. We sought space-time interaction among triplets of cases. In practice we used a subset of all possible triplets of events—n(n-1)(n-2)/6 from n events: in this case, approximately $1{\cdot}1698 \times 10^{11}$. This required an extension of the notions of "distance apart" and "time apart", beyond the usages appropriate to pairs. A triplet "distance" was defined as the mean of the three distances between the three cases: one third of the perimeter of the triangle. Time intervals were defined as the interval between the first and the last of the three events. An analysis conducted on a date sorted file, within moving limits of 400 days and 100 km, generated 32 400 050 triplets, approximately 0·028% of all possible triplets. This method, including extensions to quadruplets and higher n-tuplets, has also been described and used in other contexts.[9]

Results

BASIC INTERACTIONS

Table I gives the numbers of pairs observed within different distance intervals and time intervals, among all those pairs occurring within a sliding time window of 1000 days; and within 100 km of each other. Each cell of the table records the observed number of pairs (0), the expected number of pairs (E), the observed/expected ratio (O/E), and the p value, where it is significant. The p values were based upon χ^2 testing where the numbers were large, but usually upon directly calculated single tailed Poisson probabilities. Table II provides the same material in cumulative form. Table III provides cumulative data for the same map locations but using estimated date of birth instead of registration.

There was a significant excess of pairs within 15 days and within 0·5 km of each other at the time of registration. Significant cumulative excesses also extended into time intervals beyond those likely to correspond with short term anticipation of diagnosis—that is, resulting from the local stimulus of one case leading to the recognition or registration of another. Also, there was no evidence of the later compensatory depletion of other short range, short time secondary registrations which such anticipatory diagnoses would then have caused. The examination of dates of birth

Table II Cumulative spatial-temporal distribution of pairs registered within 1000 days and 100 km of each other. Registrations of leukaemias and lymphomas (8888 cases)

Time separation (days)		Distance separation (km) 0–0·5	0–1·0	0–1·5	0–2·0	0–5·0	0–20·0	0–100·0
0–15	Observed	22	45	88	124	509	4376	36 407
	Expected	12·99	37·25	68·69	108·79	489·57	4404·83	—
	O/E	1·69	1·21	1·28	1·14	1·04	0·99	—
	p	0·014	0·119	0·014	—	—	—	—
0–30	Observed	35	83	153	232	949	8551	71 130
	Expected	25·39	72·78	134·20	212·55	956·49	8605·91	—
	O/E	1·38	1·14	1·14	1·09	1·99	0·99	—
	p	0·041	—	—	—	—	—	—
0–60	Observed	68	162	285	424	1832	16 930	140 099
	Expected	50·01	143·35	264·32	418·63	1883·92	16 950·37	—
	O/E	1·36	1·13	1·08	1·01	0·97	1·00	—
	p	0·009	—	—	—	—	—	—
0–120	Observed	116	312	560	847	3744	33 696	278 709
	Expected	99·48	285·17	525·84	832·82	3747·81	33 720·59	—
	O/E	1·17	1·09	1·06	1·02	1·00	1·00	—
	p	0·057	—	—	—	—	—	—
0–365	Observed	298	836	1543	2460	11 115	100 764	832 820
	Expected	297·26	852·14	1571·27	2488·57	11 198·96	100 761·65	—
	O/E	1·00	0·98	0·98	0·99	0·99	1·00	—
	p	—	—	—	—	—	—	—
0–1000		780	2236	4123	6530	29 386	264 398	2 185 315

O/E = ratio of observed to expected registrations. Only significant p values are shown.

Table III Cumulative spatial-temporal distribution of pairs born within 1000 days and 100 km of each other. Birth dates of leukaemias and lymphomas

Time separation (days)		*Distance separation (km)*						
		0–0·5	*0–1·0*	*0–1·5*	*0–2·0*	*0–5·0*	*0–20·0*	*0–100·0*
0–15	Observed	11	30	60	90	414	3456	28 102
	Expected	9·94	27·60	51·73	82·62	374·62	3406·31	—
	O/E	1·11	1·09	1·16	1·09	1·11	1·01	—
	p	—	—	—	—	0·024	—	—
0–30	Observed	23	67	121	191	804	6874	55 606
	Expected	19·67	54·61	102·36	163·48	741·26	6740·14	—
	O/E	1·17	1·23	1·18	1·17	1·08	1·02	—
	p	—	—	—	0·019	0·012	—	—
0–60	Observed	47	121	222	346	1479	13 296	109 923
	Expected	38·88	107·95	202·35	323·17	1465·35	13 324·04	—
	O/E	1·21	1·12	1·10	1·07	1·01	1·00	—
	p	—	—	—	—	—	—	—
0–120	Observed	95	238	429	680	2957	26 628	218 343
	Expected	77·23	214·43	401·93	641·93	2910·91	26 465·91	—
	O/E	1·23	1·11	1·07	1·06	1·02	1·01	—
	p	0·028	—	—	—	—	—	—
0–365	Observed	249	657	1188	1921	8730	80 072	661 020
	Expected	233·79	649·14	1216·77	1943·35	8811·57	80 121·48	—
	O/E	1·07	1·01	0·98	0·99	0·99	1·00	—
0–1000		631	1752	3284	5245	23 782	216 244	1 784 007

The 1 784 007 pairs born within 1000 d of each other and registered within 100 km of each other represent 4·52 % of all possible pairs. The numbers selected within the 1000 d band are smaller than in tables I and II because the birth dates are spread over a broader time band than are the registrations.

(table III) did not show a short time, short distance interaction comparable with that in table II but there was an excess of pairs within 30 days and at rather greater distances, ranging from about 1 km up to 5 km. The slightly more diffuse temporal boundary might be due to the relative uncertainty of the estimated birth dates, and the wider geographical limit to local migration between birth and registration. It is unlikely that the less striking birth interaction could be the indirect source of the registration interaction, although the reverse is possible. Alternatively, the two findings could be secondary to an interaction at an intermediate date; or both could be independent phenomena.

DIAGNOSTIC AND AGE HETEROGENEITIES

The short time, short distance interaction shown in table II for date and place of registration was disaggregated (1) into the three different diagnostic groups, (2) between boys and girls, and (3) between different age bands. Interactions were sought within each subgroup, and among pairs spanning the disaggregated groups.

Among the different diagnostic groups (table IV) the effect was concentrated among the lymphatic/unspecified leukaemias. It could not be detected in the other groups, possibly because there were so few short range pairs. Nor could it be demonstrated among cross pairs spanning different diagnostic groups.

Table IV Cumulative spatial-temporal interaction. Registrations disaggregated by diagnostic group

	<0·5 km, <15 d	*<0·5 km, <30 d*	*<0·5 km, <60 d*
Groups 1+2+3[a]			
Observed (O)	22	35	68
Expected (E)	12·99	25·39	50·01
O/E	1·69	1·38	1·36
p	0·014	0·041	0·009
Group 1			
Observed (O)	13	21	36
Expected (E)	6·36	12·30	24·29
O/E	2·04	1·71	1·48
p	0·014	0·015	0·015
All other pairs[b]			
Observed (O)	9	14	32
Expected (E)	6·63	13·09	25·72
O/E	1·36	1·07	1·24
p	—	—	—

[a] Group 1 is lymphatic/un-specified leukaemias. Group 2 is other leukaemias. Group 3 is non-Hodgkin's/unspecified lymphomas.
[b] All "other" pairs include 2:2, 3:3, 1:2, 1:3, 2:3. Group 1 pairs comprise only 1:1.

Space-time clustering was evident both among boys and among girls. Within 0·5 km and 60 days there were 27 MM pairs against an expected 17·68 ($p = 0·023$); and there were 16 FF pairs against an expected 8·56 ($p = 0·005$). Oddly, there was no interaction among the MF cross pairs.

Space-time interaction was initially sought separately in children aged 0–59 months and in children aged 60 months or more. The O/E ratios were greater among the older children than among the younger ones but stronger still, and highly significant, among pairs spanning the chosen age boundary. For example, there were 26 cross group pairs within 0·5 km and 60 days against an expected 11·96 ($p < 0·001$). The data were therefore re-examined in three age bands: 0–3 years, 4–7 years, and 8–14 years. The results are given in table V, confirming that the main strength of the interaction lies in the intermediate age band (4–7 years). Furthermore, diagnostic disaggregation within this age band confirmed that the main interaction related to the lymphatic/unspecified leukaemias, possibly combined with the lymphomas. In this combined grouping (lymphatic/unspecified leukaemia + lymphoma) in children aged 4–7 years, there were 17 pairs within 1·5 km and 30 days of each other against an expected 5·93 ($p < 0·001$).

REGIONAL HETEROGENEITIES AND POPULATION DENSITIES

Close space-time pairs were located on a map. They showed a strong apparent concentration along a corridor stretching from the conurbations of north west England (Liverpool, Manchester, Lancashire), through the industrial Midlands, towards the south east (London, Dover). This is as might be expected for close spatial pairs, irrespective of their time separations, because high population densities increase their frequencies. Geographical heterogeneities should not in themselves give rise to artefactual space-time interactions, but they are capable under some circumstances of masking genuine ones. If the critical distance for identifying interaction pairs

Table V Cumulative temporal-spatial interaction: all diagnoses. Registrations disaggregated by age band.

		<0·5 km, <15 d	*<0·5 km, <30 d*	*<0·5 km, <60 d*	*<0·5 km, <90 d*	*<1·5 km, <30 d*	*<1·5 km, <60 d*	*<1·5 km, <90 d*
All ages	O	22	35	68	92	153	285	422
	E	12·99	25·39	50·01	74·74	134·20	264·32	395·06
	O/E	1·69	1·38	1·36	1·25	1·14	1·08	1·07
	p	0·014	0·041	0·009	0·029	0·59	0·108	0·093
Band 1: 0–47 months	O	4	6	10	12	19	38	52
	E	1·94	3·83	7·57	11·30	17·16	33·94	50·66
	O/E	2·06	1·57	1·32	1·06	1·11	1·12	1·03
	p	0·132	0·189	0·232	0·456	0·360	0·265	0·444
Band 2: 48–83 months	O	2	6	6	8	19	24	31
	E	0·78	1·51	2·95	4·47	7·76	15·20	23·01
	O/E	2·57	3·98	2·03	1·79	2·45	1·58	1·35
	p	0·184	0·005	0·079	0·084	<0·001	0·022	0·064
Band 3: 84–167 months	O	3	5	12	15	12	33	51
	E	2·00	3·93	7·72	11·52	21·20	41·66	62·16
	O/E	1·5	1·27	1·55	1·30	0·57	0·79	0·82
	p	0·323	0·357	0·093	0·186	0·988	0·926	0·934
Age band cross pairs	O	13	18	40	57	103	190	288
	E	8·27	16·12	31·77	47·45	88·08	173·52	259·23
	O/E	1·57	1·12	1·26	1·20	1·17	1·04	1·11
	p	0·078	0·352	0·089	0·097	0·065	0·114	0·041

O = observed, E = expected registrations

were to vary in different density zones—as it might do in the context of an infective model—then the phenomenon might be blurred out within a dataset containing the pooled cases from different density zones.

The interaction search was therefore repeated for three zones separately: the central corridor, and the two zones on either side. The results are given in table VI. Independent space-time interactions were detected in the central corridor and in the north eastern region, but not in the south west. This may have been due to the smaller number of cases and the sparsity of short range pairs in the south western zone.

This set of results did indeed show some evidence of heterogeneities of critical distance. In the dense central corridor, the most significant interactions were at 0·1 km and 35 days (O, E, O/E = 10, 2·18, 4·59; $p < 0·0001$) and at 0·1 km and 10 days (O, E, O/E = 5, 0·64, 7·78; $p < 0·001$). In the north eastern region, the most significant interaction was within 0·3 km and 25 days (O, E, O/E = 9, 3·10, 2·90; $p < 0·01$).

Standardisation for population density was carried out, as described earlier. The effect of density standardisation was to reallocate some of the short distances in the central corridor and in the north east to rather longer distances, and to do the opposite in the south west. Results for the three separate zones are given in table VII. In the event, these manoeuvres made little difference. There was still evidence of an interaction in the central and north eastern zones, but none in the south western zone.

Table VI Cumulative spatial-temporal distribution of pairs: all diagnoses. Registrations by separate geographical zones

		Distance separation (km)					
Time separation (days)		*0–0·1*	*0–0·3*	*0–0·6*	*0–1·2*	*0–2·0*	*0–100·0*
				South western zone			
0–15	O	0	2	3	3	5	1489
	E	0·20	0·60	1·72	3·78	7·91	
0–30	O	0	3	5	9	21	4147
	E	0·55	1·68	4·78	10·54	22·02	
0–50	O	0	4	10	17	36	6947
	E	0·92	2·82	8·02	17·66	36·88	
0–100	O	0	7	18	38	67	13 715
	E	1·82	5·56	15·82	34·86	72·82	
0–1000	O	17	52	148	326	4446	128 269
				Central zone			
0–15	O	5†	6*	11	31	70	15 581
	E	0·95	3·11	9·04	26·48	63·12	
0–30	O	9‡	11*	22	61	135	30 482
	E	1·86	6·09	17·68	51·80	123·49	
0–50	O	11‡	15	35	99	210	50 431
	E	3·07	10·07	29·25	85·70	204·30	
0–100	O	12*	28	64	174	394	99 892
	E	6·08	19·95	57·93	169·74	404·68	
0–1000	O	57	187	543	1591	3793	936 280
				North eastern zone			
0–15	O	0	7‡	11*	24*	45*	6872
	E	0·38	1·95	5·72	15·93	34·42	
0–30	O	2	9*	14	35	73	13 224
	E	0·72	3·75	11·02	30·65	66·23	
0–50	O	2	13*	26*	58	123	21 686
	E	1·19	6·15	18·07	50·26	108·61	
0–100	O	2	22†	39	113	227	42 840
	E	2·34	12·14	35·69	99·29	214·55	
0–1000	O	22	114	335	932	2014	402 136

South western zone is defined as (N + 1·2E) < 730·0; North eastern zone is defined as (N + 1·2E) > 860·0
O = observed, E = expected number of registrations
*p < 0·05, †p < 0·01, ‡p < 0·001

TRIPLETS

The analysis of pairs was supplemented by an analysis of all possible triplets, using the technique described above. The results confirmed the presence of space-time clustering in the central and north eastern regions and this time provided some positive evidence in the south western zone as well. The central zone gave significant findings at several levels, for example at 200 days and 0·5 km (O, E, O/E = 7, 2·78, 2·51). In the north east there was an excess at a range of 10 km and 20 days (O, E, O/E = 137, 111·13, 1·23). In the south west there was a significant although small proportional excess at 20 km and 160 days (O/E = 1·09).

Discussion

The major findings and inferences arising from this study are as follows.

(1) One important finding emerged during the preparation of the file for analysis, namely the near absence of concordant twins. Only one concordant set of twins was detected among 8889 registrations. In a file from which exact dates of birth were excluded, concordant twin pairs who had changed addresses between the two diagnoses might be missed; but a careful search and special enquiry (G J Draper, personal communication) failed to reveal any others. About one in 40 infants in the general population are from multiple pregnancies, leading to an expectation of 223 such infants in this sample, including 75 from monozygotic (MZ) twin pairs. We already knew that MZ twins were often discordant for these dis-

eases,[10] although where twins are concordant, they are almost always MZ. However, it is only in a sample such as this that the highly exceptional nature of concordance is apparent. Since only one concordant set of twins was found, and 75 MZ twins were expected from a file of this size, it must represent about 74 discordant MZ pairs. The single concordant pair (both males and both occurring during the first year of life) is no more than might be expected from one of the very rare genetically determined immunological deficiencies predisposing to malignancies.

The lack of concordant MZ twins in this dataset suggests that the major causes of leukaemia and lymphoma impinge at a time later than that corresponding with the division of the ovum to form an MZ pair.[10] Furthermore, these environmental causes—as they necessarily are—must operate in a highly stochastic manner, such as to strike the twin of an affected subject at a frequency not much greater than that in the population as a whole.

(2) The second major finding was a highly significant pattern of space-time clustering with respect to dates and times of onset, as reflected in the anniversary dates. This confirms findings in the several previous studies quoted. The interaction occurred at very short ranges, both geographically and in time, for example within boundaries of 30 days and 300–500 metres. It is extremely unlikely that this could be the result of localised overdiagnosis—say a mistaken allocation of a leukaemia diagnosis to children with glandular fever or whooping cough. Nor was there any evidence of short term temporal "anticipation" of a second local diagnosis within a period of some weeks; this should have produced a secondary medium term temporal deficiency, at the same short range, to match the short term temporal excess. No such deficiency occurred. Finally, the registration date interaction did not appear to be secondary to a birth date interaction. There may be an independent tendency for dates and places of birth to cluster; or, possibly, both findings may be secondary to a clustering tendency at some intermediate date. We shall examine this further in a later report.

Within the constraints imposed by small numbers, the space-time clusters seemed to be concentrated in children between the ages of 4 and 7 years, and (probably) among those suffering from lymphatic/unspecified leukaemias, or lymphomas. Boys and girls were involved equally. Children under 4 years of age and over 8 years of age and those suffering from myeloid leukaemias were not obviously involved.

(3) The space-time interactions themselves showed geographical heterogeneities. They were clearly detectable in a very highly populated central corridor of Great Britain, stretching from the north west conurbations of Lancashire, Manchester, and Liverpool, through the Midlands towards the south east and London, and in the north eastern region of England, together with Scotland, but not in the more dispersed populations of Wales and the south west of England. The examination was repeated using density standardisation but the conclusion was the same. An examination of triplets introduced weak evidence of an interaction in the south west, but did not fundamentally alter the picture. The critical distance for detecting interactions was less in zones of high population density; and the existence of a sufficient number of close geographical pairs, itself depending upon a high density of population, appeared to be a prerequisite for demonstrating space-time interactions. It is however difficult to distinguish between a biological necessity and a simple statistical necessity.

(4) This demonstration of space-time clustering raises several additional and important questions. The first is whether the space-time clusters are freely located in the population: or whether they are superimposed upon, and interact with, a consistent underlying geographical heterogeneity of risk. To answer this question it will be necessary first to see whether geographical heterogeneities exist, and whether any spatial clusters are characterised by the same short ranges as those observed in the space-time clusters. The observed geographical scale dependency of the space-time clusters provides a firm prior hypothesis against which to test the characteristics of the geographical distributions. However, the question of time independent geographical variation was not within the primary objectives of the present analysis, and its detailed examination is deferred to a companion report.

(5) It is extremely difficult to construct a biological model capable of generating space-time clustering within such short distances and such short times, yet which scarcely ever repeats within the same family and which almost never occurs in both members of a twin pair. This is equally true for radiation, for toxic exposures, and for person to person transmissions of infection. Environmental exposures leading to a primary oncogenic mutation, and simultaneously affecting a local population—such as transient and local obstetric radiology fashions, or group overdoses from faulty equipment, or parental occupational

Table VII Cumulative spatial-temporal distribution of pairs. Registrations by separate geographical zones: standardised distances

		Distance separation (km)					
Time separation (days)		*0–0·1*	*0–0·3*	*0–0·6*	*0–1·2*	*0–2·0*	*0–100·0*
			South western zone				
0–15	O	1	2	2	7	32	2841
	E	0·48	0·98	3·40	11·09	44·31	
0–30	O	1	4	5	15	75	5572
	E	0·55	1·68	4·78	10·54	22·02	
0–50	O	1	4	10	27	129	9282
	E	1·58	3·21	11·11	36·22	144·78	
0–100	O	3	6	18	54	268	18 238
	E	3·10	6·31	21·83	71·17	284·48	
0–1000	O	29	59	204	665	2658	170 407
			Central zone				
0–15	O	5*	8	25	80	354	18 812
	E	1·79	4·57	19·06	70·54	347·50	
0–30	O	9†	15*	46	151	697	36 745
	E	3·49	8·93	37·24	137·79	678·75	
0–50	O	11*	19	70	236	1150	60 849
	E	5·78	14·79	61·66	228·18	1124·00	
0–100	O	18*	30	128	465	2237	120 300
	E	11·42	29·25	121·91	451·12	2222·18	
0–1000	O	107	274	1142	4226	20 817	1 126 948
			North eastern zone				
0–15	O	4*	10†	24†	60*	211	9970
	E	1·25	3·52	13·46	45·40	188·89	
0–30	O	6*	12*	33	98	389	19 308
	E	2·42	6·83	26·06	87·92	365·80	
0–50	O	9	16*	56*	155	619	31 745
	E	3·97	11·22	42·85	144·56	601·42	
0–100	O	15*	27	96	298	1227	62 911
	E	7·87	22·24	84·92	286·48	1191·88	
0–1000	O	74	209	798	2692	11 200	591 170

O = observed, E = expected numbers of registrations
*$p < 0·05$, †$p < 0·01$

exposure to radiation or to toxic materials, or even cosmic ray showers—are unlikely to initiate disease natural histories of such uniformity as to achieve paired registrations on dates so narrowly separated. This suggests that there must be a pair shared disease-promoting event quite close to the date of registration; and therefore probably precipitating the presentation of disease already present.

One possibility would be an infective agent which provoked haematological examinations. The space-time clusters could then reflect the spread of infectious mononucleosis, or asthma microepidemics, or swollen lymph nodes from respiratory infection. This might lead to the clustered discovery of leukaemias which had not yet presented clinically. We shall reconsider these possibilities when we have examined the question of geographical clustering, independent of time.

We are grateful to Dr G J Draper and the Childhood Cancer Research Group for providing the data on which these analyses are based. Thanks are also due to the cancer registries and the UK Children's Cancer Study Group who originally provided the cancer registrations.

1 Gilman EA, Knox EG. Temporal-spatial distribution of childhood leukaemias and non-Hodgkin lymphomas in Great Britain. In: Draper G, ed. *The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966–83.* (Studies on Medical Population Subjects, No 53.) London: OPCS, 1991.
2 Knox EG. Epidemiology of childhood leukaemia in Northumberland and Durham. *Br J Prev Soc Med* 1964; **18**: 17–24.
3 Spence-Meighan MS, Knox EG. Leukaemia in childhood. Epidemiology in Oregon. *Cancer* 1965; **18**: 811–4.
4 Knox EG. Detection of low intensity epidemicity: application to cleft lip and palate. *Br J Prev Soc Med* 1963; **17**: 121–7.
5 Gunz FW and Spears GFS. Distribution of acute leukaemia in time and space. Studies in New Zealand. *BMJ* 1968; iv: 604–8.
6 Smith PG. Spatial and temporal clustering. In: Schotterfeld D, Fraumeni JF, eds. *Cancer epidemiology and prevention.* Philadelphia: WB Saunders, 1982: chapter 21.
7 Stiller CA, O'Connor CM, Vincent TJ, Draper GJ. The National Registry of Childhood Tumours and the leukaemia/lymphoma data for 1966–83. In: Draper G, ed. *The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966–83.* (Studies on Medical and Population Subjects, No 53.) London: OPCS, 1991.
8 Wilson PR, Elliot DJ. An evaluation of the Postcode Address File as a sampling frame and its use within OPCS. *J R Stat Soc A* 1987; **150**: 230–40.
9 Knox EG, Lancashire RJ. *Epidemiology of congenital malformations.* London: HMSO, 1991.
10 Knox EG, Marshall T, Barling R. Leukaemia and childhood cancer in twins. *J Epidemiol Community Health* 1984; **38**: 12–16.