

# Leukaemia clusters in childhood: geographical analysis in Britain

E G Knox

## Abstract

**Study objective** – To validate previously demonstrated spatial clustering of childhood leukaemias by showing relative proximities of selected map features to cluster locations, compared with control locations. If clusters are real, then they are likely to be close to a determining hazard.

**Design** – Cluster postcode loci and partially matched control postcodes were compared in terms of distances to railways, main roads, churches, surface water, woodland areas, and railside industrial installations. Further supporting comparisons between non-clustered cases and random postcode controls with those map features representable as single grid points were made.

**Setting** – England, Wales, and Scotland 1966–83.

**Subjects** – Grid referenced registrations of 9406 childhood leukaemias and non-Hodgkin's lymphomas, including 264 pairs (or more) separated by < 150 m, and grid references of random postcodes in equal numbers.

**Main results** – The 264 clusters showed relative proximities (or the inverse) to several map features, of which the most powerful was an association with railways. The non-railway associations seemed to be statistically indirect. Some railside industrial installations, identified from a railway atlas, also showed relative proximities to leukaemia clusters, as well as to non-clustered cases, but did not "explain" the railway effect. These installations, with seemingly independent geographical associations, included oil refineries, petrochemical plants, oil storage and oil distribution depots, power stations, and steelworks.

**Conclusions** – The previously shown childhood leukaemia clusters are confirmed to be non-random through their systematic associations with certain map features when compared with the control locations. The common patterns of close association of clustered and non-clustered cases imply a common aetiological component arising from a common environmental hazard – namely the use of fossil fuels, especially petroleum.

Many clusters of childhood leukaemia have been reported, including both simple geographical concentrations and groups confined by joint temporal and geographical limits. They have been extensively discussed.<sup>1,2</sup> Some of the former were related to point-source hazards such as nuclear installations,<sup>3</sup> while others were unrelated to any previously declared geographical feature.<sup>4,5</sup> Some reports of space-time clusters were simple anecdotes, and although formal verification techniques were sometimes used,<sup>6</sup> clustering was confirmed in some sets of data only,<sup>6,7</sup> and not in others.<sup>8–10</sup> Most studies suffered from an arbitrary choice of the bounds within which the clusters were defined; and without a prior scale specific hypothesis it was difficult to separate the defining criteria from those used for statistical testing. The question therefore remained whether clusters of either kind were real, or whether the "positive" reports represented selective publication of extreme examples from a random overall set.

Recent analyses of the full set of leukaemia registrations in England, Wales, and Scotland over a period of 18 years, however,<sup>1</sup> have again shown joint spatial-temporal clusters of registrations spanning approximately 0.5 km and 60 days,<sup>11,12</sup> bounds similar to those identified in other studies. These findings were then used to declare a scale specific prior hypothesis suitable for testing the presence of pure geographical clustering. This was based upon the argument that if space-time clusters and longer term geographical concentrations coexisted within the same set of data, then they could probably be regarded as separate expressions of a common determining mechanism. The approach was successful in that it showed a significant excess of geographically defined pairs and triplets within this same narrow range.<sup>13</sup>

These findings were based upon map references attached by the Post Office to the postcodes of the registration addresses, and stored in the National Childhood Cancer Register. Only the map references were released for this research; not the postcodes themselves. Unfortunately, geographic precision was limited by the frequent allocation of several postcodes to one map reference. Although it was possible to identify the several postcodes attached to a given map reference, and their several address ranges, it was often not possible to identify the registration postcode itself, to check its precise location, or to calculate more exactly the distances between pairs of registration addresses. Important questions therefore remained unanswered. Might the "clusters" represent a variation of postcode sizes only, or of child-resident addresses within certain postcode groups – for example, constellations of tower blocks of flats.

Mill Cottage,  
Great Comberton,  
Worcestershire  
WR10 3DU, UK.  
E G Knox

Correspondence to:  
Professor E G Knox,  
Mill Cottage,  
Great Comberton,  
Worcestershire  
WR10 3DU, UK.

(*J Epidemiol Community Health* 1994;48:369–376)

Clusters defined on so small a scale were too close for comfort to the resolution limits of the map coordinates themselves.

Although the National Childhood Cancer Register was set up to enable research of this kind – and indeed it has no other purpose – the Department of Health (the arbiter in these matters) most unfortunately refused to sanction any research use of the recorded postcodes unless specific permission was first obtained from the ethical committees in every one of some 200 NHS health districts. This requirement was considered impractical. It was unlikely that the total response needed to justify a general release would be obtained; and specific permission relating to a subset of individual children could not be sought because it was not possible to identify them or their physicians to the committees.

If the matter was to be carried further, and the spatial clustering validated, it would be necessary to devise some means of enhancing the positional and relational specificity of the map references within the censored data set. This paper shows how such a technique was devised, and it records the results of its subsequent application.

### Methods

The registration materials have been described in detail elsewhere.<sup>1</sup> They consist of a computer file of all registrations of leukaemias and reticulo-endothelial tumours in children aged 0–14 years in England, Wales, and Scotland in the 18 year period 1966–83. Individual records included the date of registration, age at registration in months, the sex of the child, the ICD code of the tumour, a map reference derived from the postcode of the address at registration, and the census enumeration district code. Certain materials, including exact date of birth, the postcode itself, the address, and identifying data, were not released. This national file provided the data on which the earlier investigations, and the demonstrations of space-time and geographical clustering, had been based.

The present investigation uses the same set of data and it seeks to confirm the reality of the clusters by demonstrating proximities to putative hazards associated with particular cartographic features. It achieves its enhanced specificity through postulating that if the previously demonstrated short range geographical concentrations are genuine, then close pairs and triplets of cases are also likely to be close to a determining hazard; and that these hazards might be recognised, or at least indicated, through inspecting maps or the sites themselves. The present report is based upon map inspections.

This method overcomes the problem that many children may have contacted hazards some distance from the given address in the period before registration. While many day to day movements will be within a few cluster diameters of the home address, others must involve more distant travelling associated with school attendance, shopping, holidays, and house moves. This will necessarily dilute any

true address-proximity effects. However, by concentrating on clusters, with the inference of a nearby hazard common to the pair or triplet of cases, the likelihood of such detection is greatly amplified, and the specificity of any findings much enhanced. The method selects only the most informative events and locations. The lack of positional accuracy resulting from the censorship rules is thus compensated for by a much improved specificity, within this subset, of any detected proximities to apparent external hazards.

Firstly, the file was searched for all pairs of registration coordinates separated by less than 0.15 km. This includes all those separated by  $\leq 0.1$  km in both easting and northing, and the de facto separation limit of such pairs is 0.142 km. This is close to the resolution limits of the postcode coordinates, which are recorded to an accuracy of 0.1 km over the greater part of the country. Accepted pairs were those in which both members suffered from acute lymphatic or acute unspecified leukaemia, or non-Hodgkin's lymphoma: and the more northerly registration was accepted as the position of the cluster.

Some of the pairs were components of larger sets and these groups were condensed to a single coordinate. Experiments with pairs within limits greater than 0.15 km (for example, 0.3 km) resulted in numbers of larger connected groups of limited positional precision, and for this reason the narrower limit was preferred. One concordant twin pair was excluded, as was an artefactual cluster resulting from the assembly of children of armed forces personnel and their families in a favoured treatment centre. This left a working list of 264 locations, each accommodating two or more cases separated by nominal distances of 0.142 km or less.

Control postcodes were chosen as those filed alternately 10 000 before and after the cluster postcode (the first one to carry the cluster map reference), in a serially numbered and easting-sorted file containing 1.28 million residential codes. This was designed to achieve a degree of social and epidemiological pair matching, while leaving sufficient distance for them to differ regarding the presence of short range hazards.

The clusters and the control postcodes were plotted on Ordnance Survey (OS) maps (1:50000 Landranger Series). They were widely scattered throughout the country. For each, distances were carefully measured to the nearest: (a) railway in current use, (b) railway in current or earlier use, (c) surface water, including stream, river, canal, pond, estuary, or coast, (d) wooded area, (e) "A" class road, and (f) church. All of them occur within a few cluster diameters of a large proportion of the general population, unlike sparse features such as nuclear power stations. Rail and road proximities were sought because of pollution associations, and water and woodland proximities because of possible infective (for example, arthropod-borne) associations. Except for churches, which are "points", all these map features have extended forms. Churches were included as a standardising variable, as described later. Measurements were recorded to a precision of 0.01 km, although the dependable accuracy is

probably no better than 0.02 km, on a map of this scale; and with the postcode release problems, it is probably no better than 0.10 km on the ground.

In the event, the first analyses then demanded secondary searches for proximities to other map features. They were definable (like churches) as cartographic points; only the map positions were recorded, and distances calculated rather than measured. Details are given later. The results in turn prompted an examination of distances between these same geographical points and the full set of all registered cases, clustered or not, together with a new set of control locations: 9406 of them, excluding Orkney and Shetland. The new control locations were chosen from the full postcode file through a simple random number process, without matching.

## Results

### CLUSTER-CONTROL RELATIONSHIPS

The mean eastings for the 264 clusters and the 264 cluster-controls were 427.65E and 427.69E, respectively. Mean northings were 304.11N and 334.72N. The difference asymmetry for eastings and northings arises from the prior east-ordering of the file. For the full set of all registrations, mean eastings were 429.46E (cases) and 424.42E (controls); and mean northings were 320.40N and 316.83N. There were no geographical biases here sufficient to perturb any of the subsequent cluster-control or case-control comparisons.

The cluster-control pair matching succeeded to the extent that the mean interpair distance was 124.51 km; while for the cases and their serially corresponding but unmatched random controls, the mean inter-"pair" distance was 239.19 km.

### CLUSTER-CONTROL COMPARISONS OF MEASURED PROXIMITIES

The results of these comparisons are shown in table 1. The most striking finding is a cluster-control asymmetry for mean distance (km) to the nearest railway; a 42% decrease for railways in current use and a 43% decrease if disused and dismantled railways are included. Many of the disused lines would have been in use during the period of the leukaemia registrations. The difference between the group means, and the

deviation of the mean interpair distance from zero, are both highly significant.

The churches and "A" class roads show less striking cluster-control distance contrasts although both are in the same direction as the rail difference, and the church difference is statistically significant. The surface water and woodland correlations, significant for the former, are in the reverse direction.

It was recognised in planning the study that relative proximity to such as a railway might be determined by the characteristics of adjacent neighbourhoods rather than the nominated feature itself. Church distances are measured because they serve as an indicator (in cities) of high population density, older housing, and less affluent life styles: the kind of areas crossed by railways. It was known from another investigation that neural tube defects, with known social correlations, exhibited significant proximities to churches.<sup>14</sup> It was for their controlling function in this respect that church distances were measured. There was indeed a significant correlation between distance to the nearest church and distance to the nearest railway in the combined set of cluster and control data, directly confirming the presence of such an effect. When church distance was used to standardise rail distance, the latter being reduced to an excess/deficit above/below the value expected from the rail/church regression function, the rail association remained significant. The reverse standardisation diminished the church correlation to non-significant levels; the  $\chi^2$  statistic for rail standardised distance asymmetry was 2.3. The inverse correlations of surface water and woodlands probably reflect incompatibilities with dense industrialisation, both in real terms and through the difficulties experienced by cartographers, with competition for map space in complex urban areas.

The possibility of a postcode artefact must again be considered. As noted already, postcodes in areas with dense transport systems may perhaps cover larger numbers of addresses; or postcodes of a given size may accommodate relatively more children. Linear obstructions to postmen's beats (such as railways) must tend to separate individual postcodes rather than traverse them, perhaps creating an ordered "packing" of map references in their immediate vicinities. However, church distance standardisation should have eliminated much of the urban high density effect, and linear packing should affect controls as well as clusters. The absence of a strong association with "A" class

Table 1 Cluster-control comparisons: measured distances to map features

Mean distances	Nearest map feature					
	Active rail	Any rail	Church	Water	Wooded	"A" road
Km from cluster	1.274	0.931	0.562	0.695	1.342	0.614
Km from control	2.198	1.620	0.743	0.546	1.193	0.798
Cluster-control/control	-0.42	-0.43	-0.24	+0.27	+0.12	-0.23
Pairwise <i>t</i> ratio	3.961	4.178	2.960	3.295	1.460	1.871
Group <i>t</i> ratio	3.261	3.450	2.143	2.101	0.940	1.583

roads, which also traverse population dense areas and tend to "organise" the disposition of postcode map references, again supports the intrinsic specificity of the rail connection.

As a further check, mean rail distance comparisons were recalculated separately in areas of very high or lesser population densities and rail network densities – the "very high" comprised map sheets covering London, the West Midlands, Liverpool, Manchester, and Glasgow. The rail distances were greater and the cluster-control differences more extreme in the less dense areas, with mean cluster and control rail distances (current and disused lines) of 1.16 and 2.28 km ( $t=2.989$ ). The mean distances in the higher density areas were 0.73 and 0.85 km respectively, and were not significantly different. These findings do not suggest a simple artefact, but rather that hazard conditions in extremely dense conurbations are sufficiently severe to mask the effects of railways; or perhaps they seldom permit a sufficient distance between railways and control locations, to show real contrasts.

The subsidiary findings in table 1 can probably be regarded as secondary to a primary association between clusters and railways, and the mutual demographic/cartographic associations between railways, roads, churches, and other map features. The evident strength of the rail-cluster correlation, and the failure to dispose of it through a form of density standardisation, suggest that it is statistically real. However, it is not possible to say immediately whether this reflects a direct hazard from the railway itself or some industrial or demographic aspect of the peri-railway environment.

There are many different ways in which railways could offer a hazard. They include atmospheric pollution from diesel fuel exhaust, evaporation of volatile cargo, or spillage of liquid cargo including fuel or toxic chemicals or radioactive waste; sewage and viral contamination of lines from passenger-care lavatories; use of herbicides and pesticides on rail side vegetation; rodent control poisons; viral hazards among railside wildlife ecosystems (for example, rats, shrews, mice, insects); theft of goods or trespass by children or members of their families; or capture of rodents by domestic pets, with resulting toxic or infective contamination of the home.

The clusters were plotted on a national rail map that distinguished high voltage electrified rail lines (25 kV 50 Hz) from lower voltage (750 V dc) "third rail" lines and non-electrified lines, and on regional maps that distinguished industrial from mixed passenger/industrial tracks. The clusters were very widely distributed involving all three main types of locomotive power and both goods-only and mixed goods-passenger lines. That is, the clusters were not specifically associated with either of the electrification formats or specifically with diesel-only tracks; although diesel locomotives can and do travel on any lines, electrified or not. Nor were they specifically associated with sewage leaking from passenger cars.

The need to narrow the field further led to the next stage of the study.

#### CLUSTER-CONTROL COMPARISON OF PROXIMITIES TO POINT HAZARDS

A number of rail associated potential hazards, distinguishable from the railway itself, were identified through a railway atlas.<sup>15</sup> The particular value of the atlas was its recording of rail sidings and industrial branch lines, together with indicators of their uses and users. These include oil refineries, petrochemical complexes, liquid propane gas terminals and tank farms; local oil terminals and depots for transfer from rail to road tankers; cement works and distribution terminals; storage and distribution sidings for agricultural fertilisers (Kemira depots); fossil fuelled and nuclear fuelled power stations; other nuclear installations, sidings and branch lines assigned to the Ministry of Defence (MOD); road stone, quarry, and coal mine terminals; dock sidings, goods traffic marshalling freight, and container terminals; steel works and other (aluminium, tinplate etc) foundries and smelters; gas and coke works; and others. Many of these sidings are also marked on OS maps, furnishing appropriate coordinates, but are shown in formats which are difficult to detect and impossible to identify without the help of the atlas. Some locations are excluded altogether from OS maps under the terms of official falsification policies relating to (some) nuclear and defence installations.

The atlas record was, however, incomplete for present purposes. Some elements recorded on OS maps, for example, several major power stations in the London area and in the West Midlands, are not recorded here, most having been closed down quite recently. Some installations, including docks and liquid propane gas terminals and both nuclear and oil fired power stations, have no rail links and are supplied entirely via sea and estuary, and are legitimately omitted from a rail atlas. Finally, there were railside installations with circular features understood on OS maps as "tanks" but without a siding. They are presumably oil, bitumen, other liquid, or powder rail to road transfer terminals; locomotive refuelling points; or rail-side factories loading directly from tankers halted on the main line or on rail spurs or loops too small to be recorded as sidings. Sewage work tanks are generally well labelled by the OS, although one or two had to be confirmed as such or separated from adjacent oil installations on larger scale maps (1:25000 OS, or "A to Z" city street maps). Cooling towers are recognised from the size of the circles and from their dispositions. Gas holders are more difficult and not always distinguishable from large liquid fuel tanks. With the natural gas revolution, few coal-gas or coke works are now marked and noted.

These various railway associations were identified through a systematic search on OS maps for all features recorded in the atlas; and a supplementary search along all rail lines (approx 3 km either side) on all map sheets with an atlas feature, a case cluster, or a cluster-control) and a search of all major ports and harbours and navigable waters.

Care was taken not to include features such as factories with tanks, identified only because

they were close to identified clusters and which were not close to railways or docks. This would have introduced bias to subsequent cluster-control comparisons. Some locations, for example, power stations with oil tanks or marshalling yards incorporating cement or oil terminals, were recorded under more than one head. Small tank installations on rail lines not covered by the terms of the above search will have been missed; nor was it possible to identify gas and coke works with sufficient consistency to warrant their inclusion. Coal mines, quarries, and stone terminals seemed unlikely candidates both on prior and on intuitive/visual grounds, and were not listed.

Distances from these hazards to the case-clusters, to their matched control locations, to all registrations, and to their unmatched random controls, were subsequently computed (rather than measured) and attached to cluster-control lists and case-control lists for subsequent analysis. Table 2 gives the main results for the cluster-control comparisons. The numbers of hazard points of different classes are given in the final row of this table. Very large installations such as oil refineries were sometimes represented as several points.

The outstanding contrast between the clusters and their matched control locations relates to oil and petroleum depots and terminals, petrochemical factories, oil storage and unloading farms, and refineries. The cluster-control differences are highly significant on several criteria including between group *t* tests, paired *t* tests, and within pair asymmetries tested with  $\chi^2$ . The other putative hazards show much weaker and mainly non-significant differences, some or all of which might be explained as secondary geographical associations.

The mean distances shown in table 2 were greater than those shown in table 1, and there was little overlap between the distributions of

the distances with which these two tables are concerned. Although the search for rail side hazards was undertaken to "explain" the rail cluster associations in indirect terms, the proximity patterns shown in tables 1 and 2 seem to be separate phenomena. There may be a common causal mechanism, related, for example, to diesel and other oil products but if this is so the two proximity distributions are separate manifestations, statistically independent of each other.

#### INDIVIDUAL CASE-CONTROL COMPARISONS OF POINT HAZARD DISTANCES

Confirmation of the findings in table 2 was sought by comparing all 9406 individual cases and their control postcodes, in similar terms. The results are given in table 3. The group means again show highly significant differences for oil installations. The earlier positive findings for power stations, rail yards, and steel works are confirmed, as is the absence of any effect for docks and harbours or for nuclear or MOD installations. The earlier positive finding for fertiliser depots is not confirmed and the earlier negative finding for cement works now appears as a hazard. Significance levels are generally higher than in table 2, due to the larger numbers; although the hazard intensities are reduced compared with the studies using cluster enhancement.

The refinery hazard showed a rather irregular variation of relative risk (RR) with increasing distances up to about 10 km: 422 cases and 360 controls (RR=1.17) within 3.0 km, 256/215 (1.19) at 3 to 5 km, 798/631 (1.26) at 5 to 10 km, and 1600/1491 (1.07) at 10 to 20 km. The risk gradients of the lesser oil hazards were more tightly localised: 552/443 (1.25) within 2 km, 485/409 (1.19) at 2 to 3 km, 978/874 (1.12) at 3 to 5 km, and a reversed ratio thereafter. The

Table 2 Cluster-control comparisons: calculated distances to point map features

Mean distances	Nearest map feature										
	*Oil refinery	Oil depot	Oil either	Kemira	Cement	Power station	Foundry/steelwks	Docks/harbour	Rail yards	Nuclear instaln	Ministry of defence
Km from cluster	45.80	11.36	9.90	54.70	25.37	23.75	41.42	31.42	15.29	57.20	62.03
Km from control	49.49	14.19	12.73	64.06	26.92	29.88	47.19	33.79	18.80	61.70	63.97
Cluster-control/control	-0.07	-0.20	-0.22	-0.15	-0.06	-0.20	-0.12	-0.07	-0.19	-0.07	-0.03
Pairwise <i>t</i> ratio	1.384	2.821	3.159	2.678	0.909	2.860	1.950	1.203	2.518	2.012	0.964
Group <i>t</i> ratio	0.700	1.854	2.102	1.444	0.501	1.900	1.147	0.615	1.612	0.771	0.303
Asymmetry $\chi^2$	2.981	7.670	10.281	2.367	0.015	2.570	2.004	3.186	0.380	0.015	2.207
	(n=51)	(n=183)	(n=234)	(n=16)	(n=44)	(n=67)	(n=44)	(n=82)	(n=107)	(n=19)	(n=25)

\*Oil refinery includes major storage installations.

n = number of hazard points used in analysis.

Asymmetry  $\chi^2$  is based on:  $(cl < co - cl > co)^2 / (cl < co + cl > co)$

Where cl is distance from feature to cluster, and co is distance from feature to control, and  $cl > co$  is number of cases where cl is greater than co.

Table 3 Individual case-control comparisons: calculated distances to point map features

Mean distances	Nearest map feature										
	*Oil refinery	Oil depot	Oil either	Kemira	Cement	Power station	Foundry/steelwks	Docks/harbour	Rail yards	Nuclear instaln	Ministry of defence
Km from cluster	47.94	13.47	11.71	60.59	26.93	26.21	45.25	33.60	17.15	60.82	62.27
Km from control	49.78	14.86	13.14	60.81	30.37	28.68	48.82	33.65	19.20	61.95	64.07
Cluster-control/control	-0.04	-0.09	-0.11	—	-0.11	-0.09	-0.07	—	-0.11	-0.02	-0.02
Group <i>t</i> ratio	2.063	4.976	5.818	0.198	5.989	4.456	4.032	0.080	5.430	1.115	1.669

\*Oil refinery includes major storage installations.

matched cluster-control comparisons had shown a similar range of effectiveness with 24/10 within 3 km of an oil installation of any kind, 23/13 at 3 to 5 km, and no excess beyond 5 km. Refineries and other large installations often spanned several kilometres on the map, sometimes with several refineries in the same area. The difficulties of representing them as single, independent map points may account for the relative irregularity and increased range of their associated gradients, compared with smaller sites.

The relative risks for cement works/terminals, rail yards, and steelworks showed no clear distance gradations at short ranges. The variations were irregular and suggested secondary associations rather than diffusing toxic hazards. Fossil fuelled power stations also showed an irregularity, but one which could be real. There was a modest ratio of 158/144 (1.10) cases/controls within 2 km, a sharp jump to 232/155 (1.50) at 2 to 3 km, then a reversion to 540/516 (1.05) at 3 to 5 km. These results might feasibly reflect a pattern of significant fallout from tall chimneys.

These relationships were explored further through identifying those specific sites whose relative risks, within successive radii, exceeded arbitrary "toxicity" criteria. The most "toxic" sources are listed in table 4, and the criteria are defined in a footnote. This confirms the high risks associated with severe industrial pollution, notably on Merseyside, in north east London, the steel manufacturing area between Sheffield and Rotherham, and parts of the West Midlands; and it confirms the predominance of the oil and petrochemical hazards. It confirms that fertilizer depots, MOD depots, and nuclear installations can be disregarded from this point of view. It shows that the dock and railyard associations are probably indirect: also that the three "toxic" cement works and terminals were close to other more likely sources. An additional examination of the Grimethorpe-Bolsover Coalite plant in Derbyshire, much maligned in the media as a supposed source of dioxin contamination of surrounding farmland, showed no evidence of a leukaemia generating effect.

In contrast with the indirect associations, the high risk power stations were often geographically independent of oil installations, and this apparent hazard cannot be dismissed so readily. The high case-control ratios were generally within 5 km, where intrinsic effects are less likely to be confounded with those of other sources.

The search showed two important heterogeneities. Firstly, certain steelworks showed powerful and probably independent effects while others, notably in Glasgow, the Midlands, and South Wales, failed to do so. Secondly, despite the predominant effects of refinery and petrochemical installations, a group of large refineries seemed to be relatively innocent. This might partly be due to their isolation from large populations, with small numbers of cases and controls, but a grouped assessment of several such refineries (shown in table 4) showed only a short range effect involving relatively few extra cases. Both heterogeneities suggest the likely import-

ance of manufacturing variations and of special processes within these industries or in nearby ancillary plants; or perhaps different crude oil or other raw materials sources.

Within so complex a set of geographical relationships it is difficult to identify cause-effect possibilities among these industrial sources. Caution is clearly necessary. Only the oil and petrochemical installations, the power stations, and possibly some steelworks, show a sufficient strength, consistency, and independence in their leukaemia associations, and sufficient coherence of their risk/distance relationships, to be regarded as genuine potential hazards; and even within these groups there seem to be heterogeneities.

### Discussion

The later steps of this investigation were not envisaged when the study began. The original postulate was that close pairs of leukaemias should themselves be close to a responsible hazard, and appropriate map measurements relating the various cartographic candidates to cluster postcodes and to control postcodes showed a powerful local association with railway lines. A weaker association with churches, and negative associations with open water and wooded areas, were probably the indirect effects of their sharing the high population densities surrounding the rail lines. The question then arose as to whether the high densities or the rail lines themselves should be treated as statistically primary or whether some unconsidered industrial association of the lines might be the true cause; or indeed, whether the findings reflected only some hazard related artefact of the postcode map reference allocation process. The remainder of the study hinged on these questions. In the event, the systematic nature of the findings largely excluded the question of an artefact.

A search for rail associated industrial features was conducted using a railway atlas, and their positions subsequently tested against the clusters and control locations. In a supplementary examination, these features were tested against the full set of leukaemia registrations and a fresh set of controls. The most powerful and consistent findings related to oil and petrochemical installations and to power stations. Other findings were interpreted as secondary to these primary relationships and to the mutual geographical associations of different industrial plants. Although these searches had been designed to see whether the railway effect was secondary to those of nearby installations, the disparate scales of the respective proximity distributions showed them as separate phenomena. Neither the rail proximities nor the oil installation or power station proximities readily "explained" the other.

Although independent statistically, they could each represent a single general class of environmental hazard, namely the spillage or evaporation of petroleum/tar products from diesel locomotives, rail tankers, or static storage tanks together with the partial combustion, fractionation, cracking, and other chemical pro-

Table 4 Locations with high case/control ratios

Place	Grid ref		Case/control 0-5 km	Case/control 0-10 km	Case/control 0-15 km	Notes
	East	North				
<i>Refineries &amp; oil-storage/processing:</i>						
Stanlow Cheshire	343.9	375.8	9/1	56/33	136/98	Refinery centroid
Stanlow Cheshire	343.9	376.8	8/4	59/43	149/109	Shoreline
Stanlow Cheshire	342.5	376.7	13/5	60/35	164/112	Flare stack
Ellesmere Port	336.6	380.2	16/11	101/65	214/152	Oil storage depot
Birkenhead	333.4	387.1	55/35	158/107	247/149	Oil sea terminal
Teesmouth	453.0	523.0	21/11	70/55	94/62	Refinery + chemical works
Canvey/Coryton	575.2	182.4	11/6	51/52	130/109	Refinery E centroid
West Thurrock	560.4	178.6	11/12	69/37	167/123	Tank farm
Purfleet	555.1	178.4	12/6	101/71	260/173	Tank farm
Dagenham	548.6	182.1	48/33	216/137	425/296	Tank farm
Runcorn	350.3	380.2	12/10	53/37	167/129	Chemical works
Partington	373.5	392.3	19/10	107/78	253/216	Chemical works
Saltend/Hull	516.2	427.7	20/9	48/40	66/61	Chemical works
Mexborough	449.2	399.6	20/14	77/61	149/108	Chemical works
Barry, S Glamorgan	314.0	168.6	12/6	41/31	65/49	Chemical works
Oldbury	399.2	288.6	45/51	187/169	463/352	Chemical works
Glazebrook	371.8	392.1	16/7	86/68	231/203	British Tar Products
Rainham, Essex	551.5	182.3	20/18	175/96	342/232	Factory with tanks
Capenhurst	336.7	374.5	19/12	52/40	141/92	Factory with tanks
Doncaster	460.0	406.1	24/13	45/32	70/58	Factory with tanks
Brentwood	556.5	191.4	20/15	68/45	173/125	Factory with tanks
Edmonton	536.8	193.6	61/33	244/166	428/389	Factory with tanks
Romford	550.7	187.9	50/33	141/92	302/190	Factory with tanks
Waltham Abbey	536.7	199.4	38/15	123/87	292/226	Factory with tanks
Newton Aycliffe	428.4	523.3	12/6	44/28	62/49	Factory with tanks
Widnes	352.6	385.6	15/17	83/70	163/132	Oil terminal/factory
Farnborough	488.5	154.2	22/12	59/51	108/91	Oil terminal (Hartwell)
Swindon	416.6	187.2	30/7	40/21	42/27	Oil terminal (Shell)
Bromford Bridge	411.7	290.1	70/40	191/122	291/215	Oil terminal (Cory)
Kingsbury, Warks	422.2	296.8	8/11	59/35	150/80	Terminal (Warks Oil)
Peterborough	518.5	298.8	26/13	36/15	39/24	Oil terminal (Cory)
Staines	503.3	172.2	24/13	100/76	220/182	Oil terminal (Cory)
Brownhills	403.8	304.2	28/13	101/54	215/145	Terminal (Charrington's)
Ferry Rd, Cardiff	317.4	174.3	37/19	60/39	70/51	Oil terminal (BP)
Colnbrook	503.6	175.6	20/9	117/80	215/179	Oil terminal (Elf)
Skellow/Adwick	455.2	409.5	6/9	49/32	88/79	Oil terminal (Elf)
St Helen's	351.3	394.6	23/28	94/73	238/174	Oil terminal (Pilkington)
Sunderland	440.8	557.8	28/23	65/56	150/104	Oil terminal (Petrofina)
Sheffield/Aldwarke	444.5	394.4	33/16	89/58	168/116	Oil-term/steelworks
Swinton	446.3	398.5	27/17	77/51	169/124	Oil-term/chemical works
Rowley Regis	398.1	286.7	33/53	165/162	344/272	Tanks nr rail
Bescott/Walsall	400.7	296.6	49/29	176/136	341/264	Tanks nr rail
Worksop	460.7	179.0	11/6	21/17	49/35	Tanks nr rail
Stanford-le-hope	568.6	181.6	1/3	34/41	80/59	Tanks nr rail
Helsby	347.7	374.6	3/3	53/32	100/72	Tanks nr rail
Grouped refineries	—	—	90/66	304/314	734/698	See footnote
<i>Fertilizer depots</i>						
Helsby	347.4	376.5	4/4	62/41	130/94	Kemira (nr tanks above)
Gloucester	384.5	218.6	24/12	38/22	57/51	Kemira (nr railyard, below)
<i>Cement works/terminals</i>						
Swindon	415.9	185.6	34/14	47/20	52/24	Castle Cement terminal
Widnes	351.6	385.1	21/14	87/67	185/146	Blue Circle terminal
Northfleet	562.4	174.4	30/20	51/29	133/85	Blue Circle Works
<i>Power stations</i>						
Drakelow, Burton	423.3	319.6	18/6	23/16	41/30	
Willington, Burton	430.6	328.9	2/7	58/37	73/60	
Leicester	457.9	302.6	53/30	76/40	89/52	Saffron Lane, disused
Gt Yarmouth	653.0	305.1	15/7	22/13	30/22	Oil tanks, no rail link
Nechels, Birmingham	409.9	289.8	71/43	219/139	313/236	Disused
West Bromwich	413.5	290.6	73/35	185/120	289/207	Disused
Ocker Hill	398.8	294.2	11/11	62/55	184/52	Wednesbury
Fiddlers Ferry	354.4	386.5	21/15	84/71	156/126	Nr Widnes
Ince	347.5	376.2	10/6	65/42	160/111	Nr Runcorn
Doncaster	456.7	403.6	33/18	58/41	103/75	Nr oil source, above
Mexborough	448.8	399.8	23/18	76/65	154/111	Nr chemical works, above
Methil	338.2	700.3	15/7	17/12	31/27	
Tilbury	566.2	175.6	13/15	49/32	116/87	Thames estuary
Gravesend	563.5	174.4	20/19	54/34	127/86	Thames estuary
Dartford	555.9	176.5	14/4	100/70	257/163	Thames estuary
Erith	549.9	180.7	43/21	208/117	398/274	Thames estuary
Dagenham	546.8	182.4	61/28	234/146	465/340	Nr oil terminal, above
<i>Docks</i>						
Runcorn	349.5	381.6	20/10	59/39	185/132	Mersey estuary
Garston	339.8	384.0	35/15	145/75	252/162	Mersey estuary
Seaforth	332.6	397.3	53/34	155/95	220/141	Mersey estuary
Sunderland	440.9	557.5	26/20	63/56	148/100	Nr oil terminal, above
Tyne Dock	434.8	565.6	32/32	127/84	198/143	
Methil	337.5	699.3	14/7	18/13	30/29	Nr power stn, above
Tilbury	563.0	176.0	17/21	55/35	138/88	Nr power stn, above
West Thurrock	558.5	176.0	19/16	76/47	188/132	Nr tank farm, above
<i>Railyards</i>						
Radyr, Llandaf	313.8	180.0	29/17	66/39	85/58	North Cardiff
New Yard, Gloucester	384.7	218.4	24/11	38/22	57/49	
Cocklebury Yard	414.2	185.5	33/15	45/23	51/25	Swindon
Bescott Yards	401.4	296.0	50/34	187/147	351/273	Wednesbury
Washwood Heath Yard	411.0	289.6	74/38	212/124	305/129	Birmingham
West Yard	517.9	300.7	23/11	37/16	38/24	Peterborough
Worksop	457.2	380.6	13/6	21/16	63/44	
Holmes Yard	497.0	371.2	22/14	24/19	24/25	Lincoln
Ditton	348.4	384.7	24/12	96/67	219/156	Nr Widnes
Edge Hill Yard	337.2	389.8	79/41	198/104	263/166	Liverpool
Tinsley Yard	441.5	389.6	35/18	117/66	174/103	Sheffield nr steelworks

Table 4 Locations with high case/control ratios *Contd.*

Place	Grid ref		Case/control 0-5 km	Case/control 0-10 km	Case/control 0-15 km	Notes
	East	North				
Beighton	443.9	384.6	24/13	100/57	180/113	Nr Sheffield steel complex
West Yard	457.9	403.1	27/15	50/40	90/73	Doncaster nr power station
Trafford	380.0	395.6	62/32	184/129	330/242	Freight liner Terminal
Hainault	545.2	192.0	35/27	161/91	372/247	Essex, NE London
Stratford	538.3	184.8	95/52	350/308	663/597	E London
Ilford	544.6	186.7	75/47	247/139	488/377	E London
<i>Steelworks</i>						
Staveley	442.4	375.3	16/10	55/44	122/76	SE of Sheffield
Catcliffe	441.0	389.5	37/23	119/66	169/103	Sheffield/Rotherham
Tinsley Park	440.2	389.6	31/20	115/65	162/99	Sheffield/Rotherham
Templeborough	441.5	391.4	28/19	109/75	168/108	Sheffield/Rotherham
Roundwood	445.0	396.2	30/22	78/50	165/119	Sheffield/Rotherham
Aldwarke	445.2	395.1	37/23	86/56	169/122	Sheffield/Rotherham
Shildon Works	422.6	525.6	15/6	27/19	48/39	Nr Bishop Auckland
Wednesbury	397.8	293.4	44/44	188/170	337/275	Nr Ocker Hill power station

Sites in this list are selected on basis of case/control ratios (RR) and minimum numbers of controls at successive ranges; (RR > = 1.5 and controls > = 5 at 0-5 km) OR (RR > = 1.4 and controls > = 25 at 0-10 km) OR (RR > = 1.3 and controls > = 75 at 0-15 km). Some duplicates removed.

The *Grouped refineries* comprise several which were not selected or listed on these criteria, namely Fawley, Isle of Grain, Milford Haven, Neyland, Angle, Immingham, Grangemouth, Thameshaven/Canvey (W), and Neath.

cessing of fossil fuels. Space-time clustering could presumably have resulted from major discrete escapes. This, if confirmed, would reverse the earlier hypothesis<sup>12</sup> that it might represent a response to an infection and perhaps reinfection, rather than a toxic exposure.

The subject of petroleum exposure in relation to childhood leukaemia has received only sparse attention in published reports. This is surprising, since the benzene component is a well established leukaemogen. The parallel increases in petroleum and other fossil fuel usage and the incidence of acute leukaemia in both children and adults, are also suggestive. The relationships between leukaemia and toxic exposures were studied in a recent case-control study of a cluster of 14 childhood cases in a restricted area in The Netherlands. This showed excessive exposures both to insecticides and to petroleum products.<sup>16</sup> Other locally published Dutch studies incriminating petroleum were mentioned. There are also several studies showing that petroleum and fuel exhaust exposures are leukaemia hazards in industrial workers, and that not all of it is explained by benzene.<sup>17-20</sup> However, no large scale systematic examinations of the question in relation to childhood leukaemia have been found. The present investigation is probably the first clear demonstration of such a relationship within a total population over an extended period, and at the same time provides a confirmation and a general explanation of the geographical clustering associated with this disease.

Thanks are due to Dr G J Draper and the Childhood Cancer Research Group for providing the data on which these analyses were based. The expenses of the work were defrayed through a Leverhulme Emeritus Research Fellowship.

1 Draper G. The geographical epidemiology of childhood

- leukaemia and non-Hodgkin lymphomas in Great Britain, 1966-83. London: OPCS, 1991.
- Cuzick J, Hills M. Clustering and clusters - Summary. In Draper G ed. *The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966-83*. London: OPCS, 1991.
- Black D. *Investigation of the possible increased incidence of cancer in West Cumbria*. London: HMSO, 1984.
- Openshaw S, Craft A. Using geographical analysis machines to search for evidence of clusters and clustering in childhood leukaemia and non-Hodgkin lymphomas in Britain. In Draper G ed. *The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966-83*. London: OPCS, 1991.
- Alexander FE. Investigations of localised spatial clustering, and extra-Poisson variation. In Draper G ed. *The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966-83*. London: OPCS, 1991.
- Knox EG. Epidemiology of childhood leukaemia in North-umberland and Durham. *Br J Prev Soc Med* 1964;18:17-24.
- Glass AG, Mantel N, Gunz FW, Spears GFS. Time-space clustering of childhood leukaemia in New Zealand. *J Nat Cancer Inst* 1971;47:329-36.
- Ederer F, Myers MH, Mantel M. Do leukaemia cases come in clusters? *Biometrics* 1964;20:626-38.
- Lock SP, Merrington M. Leukaemia in Lewisham (1957-63). *BMJ* 1967;3:759-90.
- Glass AG, Mantel N. Lack of space-time clustering of childhood leukaemia in Los Angeles County 1960-64. *Cancer Res* 1969;29:1995-2001.
- Gilman EA, Knox EG. Temporo-spatial distribution of childhood leukaemia 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*. London: OPCS, 1991.
- Knox EG, Gilman EA. Leukaemia clusters in Great Britain. 1. Space-time interactions. *J Epidemiol Community Health* 1992;46:566-72.
- Knox EG, Gilman EA. Leukaemia clusters in Great Britain. 2. Geographical concentrations. *J Epidemiol Community Health* 1992;46:573-7.
- Knox EG, Lancashire RJ. *Epidemiology of congenital malformations*. London: HMSO, 1990.
- Baker SK. *Rail atlas. Great Britain and Ireland*. 7th Ed. Oxford: Oxford Publishing Company, 1992.
- Mulder YM, Drijver M, Kreis IA. Case-control study on the association between childhood haemopoietic malignancies and local, environmental factors in Aalsmeer, The Netherlands. *J Epidemiol Community Health* 1994;160-4.
- Lidquist R, Nilsson B, Eklund G, Gahrton G. Acute leukaemia in professional drivers exposed to gasoline and diesel. *Eur J Haematol* 1991;47:98-103.
- Wongsrichanalai C, Delzell E, Cole P. Mortality from leukaemia and other diseases among workers at a petroleum refinery. *J Occup Med* 1989;31:106-111.
- Austin H, Cole P, McCraw OS. A case-control study of leukaemia at an oil refinery. *J Occup Med* 1986;28:1169-73.
- McCraw OS, Joyner RE, Cole P. Excess leukaemia in a refinery population. *J Occup Med* 1985;27:220-2.