Pergamon

European Journal of Cancer, Vol. 35, No. 3, pp. 439–444, 1999

Printed in Great Britain
0959-8049/99/$ - see front matter

PII: S0959-8049(98)00385-2

Original Paper

Population Density and Childhood Leukaemia: Results of the EUROCLUS Study

F.E. Alexander,[1] P. Boyle,[2] P.-M. Carli,[3] J.W. Coebergh,[4,5] A. Ekbom,[6] F. Levi,[7] P.A. McKinney,[8] W. McWhirter,[9] J. Michaelis,[10] R. Peris-Bonet,[11] E. Petridou,[12] V. Pompe-Kirn,[13] I. Plěsko,[14] E. Pukkala,[15] M. Rahu,[16] C.A. Stiller,[17] H. Storm,[18] B. Terracini,[19] L. Vatten[20] and N. Wray[1] on behalf of the EUROCLUS project

[1]Department of Public Health Sciences, The University of Edinburgh, Medical School, Teviot Place, Edinburgh, EH8 9AG, U.K.; [2]Division of Epidemiology and Biostatistics, Milan, Italy; [3]Registre des Hémopathies Malignes de la Côte d'Or, Laboratoire d'Hématologie, Hospital du Bocage, Dijon, France; [4]Department of Epidemiology and Biostatistics, Erasmus University, Rotterdam; [5]Dutch Childhood Leukaemia Study Group (DCLSG), The Hague, The Netherlands; [6]Uppsala Universitet, Cancer Epidemiology Unit, University Hospital, Uppsala, Sweden; [7]Institut Universitaire de Médecine Sociale et Préventive, Régistres Vaudois et Neuchâtelois des Tumeurs, Lausanne, Switzerland; [8]Information and Statistics Division of the NHS in Scotland, Edinburgh, U.K.; [9]Department of Child Health, The University of Queensland, Royal Children's Hospital, Queensland, Australia; [10]Johannes Gutenberg-Universitat Mainz Klinikum, Institut fur Med Statistik und Dokumentation, Mainz, Germany; [11]Instituto de Estudios Documentales e Historicos sobre la Ciencia (IEDHC) (Universidad de Valencia–CSIC), Valencia, Spain; [12]Department of Hygiene and Epidemiology, University of Athens, Athens (GUDI) Greece; [13]Register Raka za Slovenijo (Cancer Registry of Slovenia), Institute of Oncology, Ljubljana, Slovenia; [14]National Cancer Registry, Bratislava, Slovak Republic; [15]Finnish Cancer Registry, Helsinki, Finland; [16]Department of Epidemiology and Biostatistics, Institute of Experimental and Clinical Medicine, Tallinn, Estonia; [17]Childhood Cancer Research Group, Department of Paediatrics, University of Oxford, Oxford, U.K.; [18]Danish Cancer Registry, Division for Cancer Epidemiology, Kobenhaven Ø, Denmark; [19]Department of Biological Sciences and Human Oncology, University of Turin, Torino, Italy; and [20]Institute of Community Medicine, University Medical Centre, Trondheim, Norway

The EUROCLUS study assembled incidence data for 13 551 cases of childhood leukaemia (CL) diagnosed between 1980 and 1989 in 17 countries (or regions of countries). These were referenced by location at diagnosis to small census areas of which there were 25 723 in the study area. Population counts, surface area and, hence, population density were available for all these small areas. Previous analyses have shown limited extra-Poisson variation (EPV) of case counts within small areas; this is most pronounced in areas of intermediate population density (150–499 persons/km^2). In this study, the data set was examined in more detail for evidence that variations in incidence and EPV of CL are associated with population density. Incidence showed a curvilinear association with population density and was highest in areas which were somewhat more densely populated (500–750 persons/km^2), where the incidence rate ratio relative to areas having $\geq$1000 persons/km^2 was 1.16 (95% confidence interval 1.07–1.26) and the *P* value for quadratic trend across eight strata of population density was 0.02. Incidence in these areas is uniformly elevated and showed no evidence of heterogeneity (i.e. EPV). Statistically significant evidence of EPV was evident amongst some of the areas previously classified as

Correspondence to F.E. Alexander.
Received 15 Jan. 1998; revised 8 Oct. 1998; accepted 30 Oct. 1998.

intermediate density areas (specifically, those with a density of 250–499 persons/km^2, $P < 0.001$ for CL). These results were interpreted in terms of the current aetiological hypotheses for CL which propose that exposure to localised epidemics of one or more common infectious agent may contribute to the development of leukaemia. They suggest that such epidemics arise regularly in moderately densely populated areas and also sporadically in areas which are somewhat less densely populated. Although other interpretations are possible, these results may assist in the identification of characteristics which infectious agents must possess if direct or indirect causes of CL.

Key words: childhood leukaemia, clustering, population density, infections, epidemiology, Poisson regression
Eur J Cancer, Vol. 35, No. 3, pp. 439–444, 1999

INTRODUCTION

THE EPIDEMIOLOGICAL profile of childhood leukaemia (CL), especially the development of the childhood peak of acute lymphoblastic leukaemia (ALL), is consistent with the hypothesis that risk of the disease is increased by certain patterns of exposure to infectious agents and, specifically, limited exposure during infancy [1–3]. In general, population density, including household density, is associated with frequency of infant exposure to infectious agents [4, 5] with those living in high density situations having earlier exposure and thus improved immunity. The hypothesis would, therefore, predict higher CL incidence where the population density is low. Direct investigation of the effects of population density on CL incidence has been limited but, in general, supports this [6, 7], and further support comes from reports of higher incidence where socio-economic status is higher [8, 9]. A series of studies by Kinlen [10] have found excess CL in situations where population mixing has occurred in communities which were previously isolated. The final population density achieved in these communities need not be low. Kinlen has stressed the importance of population density in several of his studies [11]. Isolation, *per se*, has been linked to excess CL incidence in the U.K. [6] but, elsewhere in Europe, reports have noted higher incidence in urban areas [12, 13]. The association appears complex, and one study conducted in the U.S. [14] has found CL incidence to increase with population density, but absolute levels of density are not given.

The EUROCLUS study [15] was set up to investigate and interpret clustering of CL in small areas in 17 countries, mainly in Europe. The possibility was recognised, in advance, that clustering might depend on population density and a consistent classification was adopted: sparsely populated was defined to be < 150 persons/km^2 and densely populated ≥ 500 persons/km^2. The results [16] have shown statistically significant evidence of clustering but of small magnitude; the clustering was strongest in areas with intermediate population density (150–499 persons/km^2).

The work of Kinlen [10] and others [17, 18] suggests that the risk of CL may be highest when micro-epidemics of aetiological agents occur, but gives little guidance regarding characteristics of the unknown agent or agents. The present analysis explores a more extensive range of population densities. Associations with population density of both incidence rates and evidence of extra-Poisson variation (EPV, clustering) have been considered. Variations in incidence rates will indicate overall increased risk within certain strata of population density; EPV within a population density stratum will demonstrate that risk within this stratum is heterogenous while absence of EPV suggests homogeneity within the stratum.

MATERIALS AND METHODS

The EUROCLUS methods have been presented in detail elsewhere [15]. Exact geographical referencing of residence at diagnosis is available for population-based series of cases of CL ($n = 13\,551$) diagnosed between 1980 and 1989 in 17 countries (or regions within countries). Residences at diagnosis are located within the small census areas ($n = 25\,723$, covering the relevant geographical areas). Population counts (hence, densities with the exception of Slovakia and Slovenia for which the geographical areas were not available and which have been excluded from the present report) have been derived from two censuses, around 1980 and 1989. Counts for inter-census years have taken those of the closest census; this process is simpler than linear interpolation and gives results of corresponding accuracy over a 10 year period [9]. Expected numbers of cases were computed by applying combined-country age- and sex-specific rates [19] to the person-years at risk.

Analyses were applied to total CL and to ALL in the childhood peak defined in three ways (0–4 years, 1–7 years, 2–4 years). In conformity with the EUROCLUS protocol, analyses were conducted twice: for all small areas and for all which can reasonably be considered 'small'—defined as those with expected numbers of CL < 5. Most attention was placed on the analyses with the 'larger' areas excluded, since it must be presumed that population density within each of these areas is heterogeneous. Leukaemia subtype was not reliably available for Estonia so these data were excluded from analyses of ALL.

Three strata of population density (0–149, 150–499, 500– persons/km^2) have been recommended for EU use and formed the a priori EUROCLUS strata [15, 16]. In the present study the following more detailed strata were considered: 0–49, 50–99, 100–149, 150–249, 250–499, 500–749, 750–999 and ≥ 1000 persons/km^2. We emphasise that the choice of these ranges was somewhat data dependent in that it was motivated by results for 248 pairs of small areas out of the total number of 25 723; the choice is not dependent on any examination of data for the whole study area. Incidence rate ratios were computed for each stratum which referred to the largest (highest density) group as baseline using Poisson regression modelling with statistical testing for overall

heterogeneity (by strata) and for linear and quadratic trends across strata. The latter provided a powerful test of the prior hypothesis of a unimodal pattern with the highest risk for areas with a density of 500–749 persons/km^2.

An additional analysis took log (density + 25) as the explanatory variable (with density > 50 000 put to 50 000 to avoid numerical instability). This has the advantage of using the actual density values but the disadvantage of placing attention on differences between areas which are already densely populated. Terms for country were included in all models so that all analyses adjusted for between-country differences in incidence. The approximate χ^2 distribution for the difference in deviation between nested models was used for statistical testing.

EPV or 'clustering' within small census areas was examined using the Potthoff–Whittinghill method [20, 21]. This method provides an estimation of the percentage extra-Poisson component, β, of variance,

$$V(O_i) = E_i + \beta E_i/100,$$

where O_i, E_i are observed and expected numbers in the ith area as in the initial EUROCLUS analysis [16]. In the present analyses, the Potthoff–Whittinghill test was conditioned on observed numbers *within* population density stratum *within* country so that the test detected heterogeneity of risk within stratum after allowing for variation between countries. This meant that high risk, relative to the country in which it lies, in *all* areas within a stratum would not appear as clustering. The asymptotic standardised normal distribution for the test statistic was used for statistical testing, but the adequacy of this approximation for EUROCLUS data was tested by Monte-Carlo methods. All tests were one-sided since we are concerned only with EPVs.

The demonstration of EPV in reference [16] and also in the present results, is one reason why the results of the Poisson modelling need not hold; a further reason for caution is that the χ^2 distribution for the difference in deviation may not apply when the data are very sparse [22]. For these reasons, we also conducted simulations to check the robustness of the model and the validity of the P values in a similar way to that previously described [22]. Specifically we generated 499 replicates with the total observed numbers in each country allocated to the available small areas using a multinominal distribution with probabilities (P_i):

(a) proportional to the fitted values derived from the final model (containing country and population density at eight levels);
(b) proportional to the fitted values from the 'null' model (containing country only) but perturbed by up to 5% or 10% to allow for EPV (i.e. $P_i = E_i[1 + k(nd)_i]$ where $(nd)_i$ are random numbers drawn from the standard normal distribution and $k = 0.05$ or 0.10);
(c) a combination of (a) and (b) with the fitted values of (a) perturbed by up to 5% as in (b) to include EPV.

The use of (a) allowed us to estimate the goodness-of-fit of the Poisson model; (b) modelled EPV and permitted estimation of the distribution of the deviance statistics under the null hypothesis but in the presence of EPV; and (c) allowed estimation of the goodness-of-fit of the EPV model. The magnitude of the EPV considered here was larger than the overall 2% of Poisson variation estimated in these data [16]. Since this approach uses a large amount of computer time, it was applied in full for ALL at ages 1–7 years and in part for ALL at ages 2–4 years and total CL.

All Poisson regression, including the simulations described above, were implemented in GLIM. For the Potthoff–Whittinghill test, FORTRAN programs were used as previously described [16].

RESULTS

The numbers of small areas in each stratum and the number of cases of CL in these areas are shown in Table 1. In general, the population at risk (and, hence, the mean number of cases/area) increases across the strata. The incidence rate ratios (Table 2) show that, after controlling for country, incidence was highest in the moderately densely populated areas (500–749 persons/km^2). This was also the case with little change with inclusion of large areas (with expected numbers of CL > 5). Statistical testing confirmed the unimodal pattern, since the quadratic trend across density groups, but not the linear trend, was statistically significant. When the analyses were repeated with log (density + 25) as the explanatory variable, the linear term never improved the model significantly but linear and quadratic terms did.

The simulation approaches (Table 3) showed that, although observed deviances suggested well-fitting models (since they were substantially less than the degrees of freedom), fit of the Poisson model was only moderate. The fit with EPV included was excellent. However, the P values for the *differences* in deviation derived from the asymptotic χ^2 [2] distribution, approximated very well to those obtained from the simulations which included 5% EPV.

Individual countries were examined (Table 4). Since the data were now very sparse many of the rate ratios were unstable and most of the confidence intervals (not shown) were wide. These data showed similar patterns in most individual countries to those in the total data set, with peak rate ratios close to the 500–749 persons/km^2 stratum. Some individual countries, notably England and Wales, also showed high incidence in some low population density strata. These results also showed the large variation between countries with, for example, Norway having no cases (and no areas) in the most densely populated stratum and The Netherlands having no cases (and very few areas) in the four lowest strata of population density.

Table 1. *Population density strata*

	Population density (persons/km^2)							
	0–49	50–99	100–149	150–249	250–499	500–749	750–999	≥1000
Number of areas	2711	4635	4291	4361	3719	1341	994	4150
Number of cases observed (childhood leukaemia)	866	971	1028	1286	1556	783	561	3114
Average number of cases/area	0.32	0.21	0.24	0.29	0.42	0.58	0.56	0.75

The Potthoff–Whittinghill test of EPV clustering (Table 5) provided estimates of β for each population density stratum. There was evidence that β did not exceed 0 (i.e. no EPV) for most strata, including the moderately dense areas. The high incidence noted above for this stratum must, therefore, apply relatively uniformly. However, the upper end of the intermediate stratum (250–499 persons/km^2) showed statistically significant evidence of EPV whose magnitude was considerably in excess of that noted earlier [16] for the entire intermediate strata. The EPV in this stratum persisted (with larger β, for example, = 15.2 for CL) if the large areas were included; inclusion of these areas also led to evidence of clustering in the strata with higher population density, but this was attributable to a very small number of very large areas.

DISCUSSION

These analyses were conducted to clarify and refine the EUROCLUS studies already reported [16, 23]. We found that incidence was higher (relative to the country/region as a whole) in the stratum of areas classified as having a moderately dense population (500–749 persons/km^2) compared with those having a higher or lower density; these results applied to the total EUROCLUS data and were broadly evident in most individual countries/regions, although some countries also had excesses within the sparsely populated areas. The simulation studies conducted demonstrated that our results are unlikely to be attributable to failure of the statistical distributional assumptions.

EPV was effectively limited to the slightly less dense areas (250–499 persons/km^2) where it is quite substantial. In this stratum of population density, the average elevation of risk over total countries was smaller but the EPV provided evidence that risk varied between areas within the same country.

Combining together the two pieces of information suggests that the risk of CL is fairly uniformly increased across areas with a population density in the range of 500–749 persons/km^2,

Table 2. Incidence rate ratios (RR) and 95% confidence intervals (CI) by diagnostic age group and population density strata*

		Diagnostic age group			
Population density (persons/km^2)†		ALL 0–4 years	ALL 1–7 years	ALL 2–4 years	CL
0–49	RR	0.98	1.02	0.99	1.03
	(CI)	(0.85–1.13)	(0.90–1.15)	(0.85–1.16)	(0.96–1.10)
50–99	RR	1.01	1.06	1.02	1.09
	(CI)	(0.89–1.14)	(0.95–1.18)	(0.89–1.17)	(1.00–1.18)
100–149	RR	1.05	1.08	1.04	1.10
	(CI)	(0.93–1.18)	(0.97–1.21)	(0.91–1.11)	(1.02–1.19)
150–249	RR	1.06	1.10	1.06	1.09
	(CI)	(0.94–1.18)	(1.01–1.21)	(0.94–1.21)	(1.01–1.17)
250–499	RR	1.11	1.10	1.11	1.07
	(CI)	(1.01–1.23)	(1.01–1.21)	(1.00–1.25)	(1.00–1.15)
500–749	RR	1.16	1.19	1.20	1.16
	(CI)	(1.02–1.31)	(1.06–1.32)	(1.05–1.38)	(1.07–1.26)
750-999	RR	1.01	1.03	1.05	1.03
	(CI)	(0.88–1.16)	(0.91–1.16)	(0.88–1.22)	(0.94–1.13)
≥ 1000	(RR)	1.00	1.00	1.00	1.00
P linear trend‡		0.96	0.43	0.94	0.07
P quadratic trend§		0.04	0.014	0.054	0.007
P heterogeneity		0.23	0.10	0.23	0.02

ALL, acute lymphoblastic leukaemia; CL, childhood leukaemia. *Incidence rate ratios are expressed relative to the most densely populated areas as reference. †Large areas (expected numbers of CL ≥ 5) excluded. ‡Trend across eight strata fitted -1 df. §Terms for linear and quadratic trend fitted -2 df.

Table 3. Simulation results for the Poisson regression analyses

		Observed		Proportion in simulated data < observed	
Simulation model	Disease group	Deviance*	Difference in deviance†	Deviance	Difference in deviance
(a) Fitted values, final model	ALL 1–7	14 354	8.59	0.42	0.61
(b) Fitted values, null model, 5% EPV	ALL 1–7	14 354	8.59	NA	0.014‡
(b) Fitted values, null model, 10% EPV	ALL 1–7	14 354	8.59	NA	0.052‡
(c) Fitted values, final model, 5% EPV	ALL 1–7	14 354	8.59	0.43	0.73
(b) Fitted values, null model, 5% EPV	ALL 2–4	11 249	5.84	NA	0.058§
(b) Fitted values, null model, 5% EPV	CL 0–14	18 071	9.97	NA	0.008¶

EPV, Extra-Poisson variation; ALL, acute lymphoblastic leukaemia; CL, childhood leukaemia. *The deviance after fitting country and linear, quadratic terms for population density stratum; distributional assumptions applied elsewhere indicated this to be from a χ^2 distribution with 26 174 df. †The difference between above and deviance after fitting country; under similar distributional assumptions this is a χ^2 distribution on 2 df under the null hypothesis and a non-central χ^2 distribution on 2 df if a Poisson distribution applied to the data generated from the final model (i.e. (a)). ‡Monte-Carlo *P* value (*P* value for $\chi^2(2)$ was 0.01). §Monte-Carlo *P* value (*P* value for $\chi^2(2)$ was 0.054). ¶Monte-Carlo *P* value (*P* value for $\chi^2(2)$ was 0.007).

Table 4. Rate ratios by population density strata for individual countries/regions for total childhood leukaemia (excluding large† areas) (numbers of cases given in parentheses)*

Country	Population density (persons/km^2)							
	0–49	50–99	100–149	150–249	250–499	500–749	750–999	≥1000
Australia	1.10	*0.72*	1.59	1.17	1.56	1.69	0.93	1.00
	(81)	(8)	(15)	(13)	(26)	(31)	(11)	(68)
Denmark	1.12	1.22	1.03	1.17	1.44	1.36	2.40‡	1.00
	(54)	(90)	(35)	(39)	(48)	(30)	(12)	(34)
England & Wales	1.06	1.25‡	1.16‡	1.16‡	0.98	1.12	1.05	1.00
	(56)	(249)	(259)	(320)	(495)	(221)	(259)	(1738)
Estonia	*0.39*	0.97	/	/	/	/	/	1.00
	(2)	(35)	(0)	(0)	(0)	(0)	(0)	(4)
Finland	0.91	0.70	1.12	1.15	0.62	*0.68*	0.79	*1.00*
	(213)	(30)	(12)	(27)	(14)	(5)	(13)	(6)
Germany	*1.07*	1.02	1.08	1.06	1.12	1.11	1.15	1.00
	(3)	(373)	(594)	(698)	(623)	(230)	(104)	(44)
Greece	0.93	0.99	0.90	1.17	0.90	1.35	1.52‡	1.00
	(160)	(50)	(28)	(28)	(36)	(32)	(28)	(176)
Italy	1.31	1.19	1.31	1.24	1.15	1.60	*0.87*	1.00
	(14)	(25)	(25)	(30)	(43)	(40)	(4)	(25)
The Netherlands	0	/	/	0	*1.37*	1.19	0.92	1.00
	(0)	(0)	(0)	(0)	(2)	(12)	(23)	(615)
Norway					Not applicable§			
	(165)	(40)	(15)	(18)	(26)	(2)	(0)	(0)
Scotland	1.14	1.10	*1.29*	0.85	1.12	1.09	0.98	1.00
	(91)	(23)	(7)	(24)	(50)	(57)	(30)	(92)
Spain	1.00	*0.74*	*1.36*	1.22	1.54	1.17	1.16	1.00
	(10)	(4)	(8)	(13)	(41)	(10)	(13)	(87)
Sweden	0.67	1.12	1.05	0.94	1.02	1.09	0.73	1.00
	(17)	(36)	(25)	(73)	(147)	(103)	(64)	(220)
Switzerland	/	*1.02*	*0.70*	*0.41*	*0.73*	0.79	/	*1.00*
	(0)	(8)	(5)	(3)	(5)	(10)	(0)	(5)

*Figures in *italics* indicate relative risks based on observed numbers < 10. †Areas with an expected number ≥ 5.0. ‡Rate ratio differs significantly from 1 ($P < 0.05$). §Reference group has no expected cases.

*Table 5. Estimates of extra-Poisson component (%) of variation (β) by diagnostic age group and population density**

Population density/km^2		ALL 0–4 years	ALL 1–7 years	ALL 2–4 years	CL
0–49	β	−0.69	−1.02	0.74	0.17
50–99	β	−0.41	0.24	−2.03	2.59
100–149	β	−1.10	1.73	−2.56	1.73
150–249	β	−1.27	0.62	−1.02	0.89
250–449	β	3.79	3.43	2.12	7.31
	(P value)†	(0.05)	(< 0.01)		(< 0.001)
500–749	β	−3.35	−3.89	−4.48	0.03
750–999	β	−3.32	−1.17	−2.17	−0.54
≥1000	β	1.32	1.35	0.73	−0.20

ALL, acute lymphoblastic leukaemia; CL, childhood leukaemia. *Large areas (with expected numbers of cases of CL ≥ 5) excluded. Monte-Carlo P values (only given when $P \leq 0.05$).

but that clustering, or local elevation of risk, is most common in somewhat less densely populated areas (250–499 persons/km^2).

The Potthoff–Whittinghill test is a simple method of detecting spatial clustering within small census areas. This may not be relevant to the distribution of diseases with infectious aetiologies for two reasons; firstly, the analyses are constrained by the census boundaries and these reflect varying sizes of area both between- and within-countries [15] and, secondly, many authors argue that analyses of space–time clustering are more appropriate when specific infectious agents are involved. The first objection certainly applies here. However, in succeeding in assembling such a large data set for analyses, it was essential to compromise regarding the quantity and precision of the data to be collected. In addition, when preparing our first report, we searched very carefully, but unsuccessfully, to identify any differences, other than population density, in the small area characteristics of the national data sets which could explain presence/absence of EPV in their data. Since the latent period between exposure and leukaemia is likely to be both lengthy (months or years, rather than days) and variable, we do not consider the second objection valid. In addition, critical times of exposure are likely to apply for specific diagnostic/age subgroups. EUROCLUS has applied space–time interaction methods with attention to putative critical times [23] for cases in the 248

areas discussed earlier and found highly significant evidence of predicted space–time aggregations.

It has been noted above that available evidence is consistent with, and supportive of, a hypothesis under which CL arises, in part, as a rare response to one or more common infectious agent and that risk is increased at times when small epidemics have occurred. If this hypothesis is true then the present results suggest that: (a) the agent commonly presents as epidemics in areas with population densities in the region of 500–749 persons/km^2; (b) such epidemics in these areas have relatively constant frequency in time; and (c) the agent less commonly, and possibly also with varying frequency, presents as epidemics in areas with population densities in the region of 250–499 persons/km^2. The lower relative incidence, in general, in areas which are either more or less dense than these could be interpreted as attributable to endemic distribution in the more densely populated areas and absence of infection in the majority of the sparsely populated areas. None of these interpretations is a necessary conclusion from the present results and these results require testing elsewhere, but they are intriguing speculations. The limited progress towards understanding CL underlines the need to follow up such speculations as clues leading to further research; the present results may assist in focusing attention on certain candidate agents by assisting in the identification of epidemicity parameters [24].

1. Doll R. The epidemiology of childhood leukaemia. *J Royal Stat Soc, Series A* 1989, **152**, 341–351.
2. Greaves MF, Alexander FE. An infectious aetiology of common acute lymphoblastic leukaemia in childhood. *Leukaemia* 1993, **7**, 349–360.
3. Greaves MF. Aetiology of acute leukaemia. *Lancet* 1997, **349**, 344–349.
4. Cochi SL, Fleming DW, Hightower AW, *et al.* Primary invasive haemophilus influenza type b disease. A population based assessment of risk factors. *J Pediatrics* 1986, **108**, 887–896.
5. Sumaya CV, Henle W, Henle G, Smith MHD, LeBlanc D. Seroepidemiologic study of Epstein–Barr virus infections in a rural community. *J Infectious Dis* 1975, **3**, 403–408.
6. Alexander FE, Ricketts TJ, McKinney PA, Cartwright RA. Community lifestyle characteristics and risk of acute lymphoblastic leukaemia in children. *Lancet* 1990, **336**, 1461–1464.
7. Staines A. The geographical epidemiology of childhood insulin dependent diabetes and childhood acute lymphoblastic leukaemia in Yorkshire. PhD Thesis, University of Leeds, U.K.
8. McWhirter WR. The relationship of incidence of childhood lymphoblastic leukaemia to social class. *Br J Cancer* 1982, **46**, 640–645.
9. Elliott P, McGale P, Vincent TJ. Description of population data and definitions of areas. In Draper, ed. *The Geographical Epidemiology of Childhood Leukaemia and Non-Hodgkin Lymphomas in Great Britain, 1966–1983*. London, OPCS, HMSO, 1990, 17–23.
10. Kinlen LJ. Epidemiological evidence for an infective basis in childhood leukaemia. *Br J Cancer* 1995, **71**, 1–5.
11. Kinlen L, Clark K, Hudson C. Evidence from population mixing in British New Towns 1946–1985 of an infective basis for childhood leukaemia. *Lancet* 1990, **336**, 577–582.
12. Petridou E, Revinthi K, Alexander FE, *et al.* Space–time clustering of childhood leukaemia in Greece: evidence supporting a viral etiology. *Br J Cancer* 1996, **73**, 1278–1283.
13. Michaelis J, Schüz J, Meinert R, *et al.* Childhood leukaemia and electromagnetic fields: results of a population-based case–control study in Germany. *Cancer Causes Control* 1997, **8**, 167–174.
14. Muirhead CR. Childhood leukaemia in metropolitan regions in the United States: a possible relation to population density? *Cancer Causes Control* 1995, **6**, 383–388.
15. Alexander FE, Wray N, Boyle P, *et al.* Clustering of childhood leukaemia: a European study in progress. *J Epidemiol and Biostatist* 1996, **1**, 13–24.
16. Alexander FE, Boyle P, Carli P-M, *et al.* Spatial clustering of childhood leukaemia: summary results from the EUROCLUS project. *Br J Cancer*, **77**, 818–824.
17. Alexander FE, Chan LC, Lam TH, *et al.* Clustering of childhood leukaemia in Hong Kong: association with the childhood peak and common acute lymphoblastic leukaemia and with population mixing. *Br J Cancer* 1997, **75**, 457–463.
18. Stiller CA, Boyle P. Effects of population mixing and socio-economic status in England and Wales, 1979–85 on lymphoblastic leukaemia in children. *Br Med J* 1996, **313**, 1297–1300.
19. Parkin DM, Stiller CA, Draper GJ, Bieber CA, Terracini B, Young JL. *International Incidence of Childhood Cancer.* IARC Publication No. 87. Lyon, IARC, 1988.
20. Potthoff RF, Whittinghill M. Testing for homogeneity II. The Poisson distribution. *Biometrika* 1966, **53**, 183–190.
21. Muirhead CR, Butland BK. Testing for over-dispersion using an adapted form of the Potthoff–Whitinghill method. In Alexander FE, Boyle P, eds. *IARC Scientific Publication No. 135*. Lyon, IARC, 1996, 40–54.
22. Bithell JF, Dutton SJ, Neary NM, Vincent TJ. Controlling for socio-economic confounding using regression methods. *J Epidemiol Comm Hlth* 1995, **49**, 15–19.
23. Alexander FE, Boyle P, Carli P-M, *et al.* Spatial and temporal patterns in childhood leukaemia: further evidence of an infectious origin. *Br J Cancer* 1998, **77**, 812–817.
24. Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control*. Oxford, Oxford University Press, 1991.

Acknowledgements—The co-ordination of this project is funded by the European Union under its BIOMED programme of Concerted Actions as Project No. PL93-1785*26.02.1993. Dr Freda Alexander is also partially supported by the Leukaemia Research Fund and the Kay Kendall Leukaemia Fund. The collaboration of Mrs M. C. Chassot and C. Pashe for socio-demographic data collection in selected small areas of Vaud and Neuchatel Cantons, and the helpful advice of Dr Martin Schuler, Geographer, Polytechnical Federal School (EPFL) is acknowledged by Dr Levi. The Australian Paediatric Cancer Registry is supported by the Queensland Cancer Fund. The data for Greece were generated with contributions from Professor S. Haida, Professor M. Kalmanti, Dr D. Koliouskas, Dr H. Kosmidi and Dr F. Piperopoulou. The Childhood Cancer Research Group is supported by the U.K. Department of Health. The Childhood Cancer Registry of Piedmont is supported by the Italian National Council, contract 95.00449.PF39 and by the Italian Association for Cancer Research. The collaboration of Dr G. Couillault and Dr Maynadie with Professor Carli is acknowledged. Mrs Katherine Smith and Ms Nadine Russell are thanked for their assistance with Scottish data. The contribution of Professor Boyle was within the framework of support of the Associazione Italiana per la Ricerca sul Cancro (AIRC). Dr Colin Muirhead is thanked for his contributions to meetings of the Statistics Subcommittee for this project. The Childhood Cancer Registry of Valencia is supported by the Conselleria de Sanitat i Consum of the Generalitat Valenciana. Professor Leo Kinlen is thanked for reading an early draft of this paper and for his perceptive comments. Mrs Rosemarie Bland and Miss Patricia Bisset are thanked for preparing the manuscript and Mr Darren Downing for assisting with computer graphics.