Pergamon

European Journal of Cancer Vol. 31A, No. 9, pp. 1479–1486, 1995
Elsevier Science Ltd
Printed in Great Britain
0959–8049/95 $9.50+0.00

0959-8049(95)00117-4

Original Paper

Case Clustering, Epstein–Barr Virus Reed–Sternberg Cell Status and Herpes Virus Serology in Hodgkin's Disease: Results of a Case–Control Study

F.E. Alexander, C.P. Daniel, A.A. Armstrong, D.A. Clark, D.E. Onions, R.A. Cartwright and R.F. Jarrett

The Leukaemia Research Fund Data Collection Study (DCS) is a specialist registry of leukaemias and lymphomas. The present study involves 494 cases of Hodgkin's disease (HD) registered with the DCS between 1985 and 1989. This entire data set has been tested for localised spatial clustering using an established nearest neighbour method with 18% of all cases in young people classified as clustered ($P < 0.05$). No clustering was found in older cases. Subsamples were selected from the registered cases for a pilot study in which case clustering, herpes virus antibody titres and Epstein–Barr virus (EBV) presence within the Reed–Sternberg (RS) cells (EBV-RS status) were investigated together. Firstly, a case–control study of HD in young people or nodular sclerosing (NS) subtype (39 HD cases and 26 healthy controls) found significant elevation of antibody titres to EBV-viral capsid antigen (VCA), EBV-early antigen (EA) and human herpes virus 6 (HHV-6) in HD cases compared with controls. EBV viral genome was present in 5 cases and 4 of these were in clusters of HD in young people. Elevation of antibody titres to the EBV antigens was not associated with case clustering or EBV-RS status. Antibody titres to HHV-6 differed significantly between EBV-RS+ and EBV-RS− cases ($P = 0.04$). Geometric mean titres for HHV-6 for EBV-RS+ and EBV-RS− cases were 11.5 and 73.7, respectively, with the former lower than the control value of 20.5. Secondly, a cluster study included all other cases ($n = 14$) in clusters containing known EBV-RS+ cases. 3 further cases were EBV-RS+ positive but no cluster consisted entirely of positive cases. Overall, 5/16 clustered, 2/12 peripheral and 1/25 random cases in these studies were EBV-RS+ ($P = 0.017$). The interpretation of these results in terms of shared aetiological exposures of cases within clusters and the roles of EBV and HHV-6 is discussed, and hypotheses for testing in future studies proposed.

Key words: Hodgkin's disease, clustering, Epstein–Barr virus, human herpes virus 6, serology, molecular epidemiology

Eur J Cancer, Vol. 31A, No. 9, pp. 1479–1486, 1995

INTRODUCTION

AGE-SPECIFIC INCIDENCE for Hodgkin's disease (HD) in developed societies reveals marked peaks amongst young adults [1, 2]. Consideration of this pattern and other features led MacMahon [3] to propose that the disease in younger (aged ≤34 years) and older (≥50 years) people was aetiologically distinct, with cases in the first group having an infectious aetiology; in addition, MacMahon suggested children and young adults might have distinct aetiologies. Under the "late host response" model [4], HD among young adults arises as an unusual response to delayed first exposure to some unknown but common infectious agent. This model is supported by international comparisons [1] and ecological, case–control and cohort studies linking socio-economic, demographic and lifestyle factors with HD in young adults [3–7].

The leading candidate as an infectious agent involved in the pathogenesis of HD has been the Epstein–Barr virus (EBV).

Correspondence to F.E. Alexander at the Department of Public Health Sciences, The University of Edinburgh, Medical School, Teviot Place, Edinburgh EH8 9AG, U.K.
C.P. Daniel, A.A. Armstrong, D.A. Clark and R.F. Jarrett are at the Leukaemia Research Fund Virus Centre, University of Glasgow, Bearsden Road, Glasgow G61 1QH; F.E. Alexander and R.A. Cartwright are at the Leukaemia Research Fund Centre for Clinical Epidemiology, The University of Leeds, 17 Springfield Mount, Leeds LS2 9NG, U.K.; D.E. Onions is now at the Department of Veterinary Pathology, The University of Glasgow, U.K.
Received 9 June 1994; accepted 15 Dec. 1994.

Serological case–control studies have repeatedly found higher titres of antibody to viral capsid antigen (VCA) and early antigen (EA) in HD cases, and these results have been confirmed by nested case–control studies using stored serum collected prospectively [8–10]. Higher antibody titres to human herpes virus 6 (HHV-6) have also been reported [11], whilst epidemiological associations with other herpes viruses are available [12].

Recent advances in molecular biology and related techniques have made it possible to test for the presence of the EBV genome and/or viral protein expression in the Reed–Sternberg cells (RS), the presumed malignant cells (e.g. [13, 14]). According to these criteria, 17–48% of HD cases have been classified as EBV-associated [15] in several studies, and the consensus is that EBV is causally involved [16, 17]. We refer to EBV-associated cases as EBV-RS+ and other cases as EBV-RS−.

The first report of a cluster of HD cases was associated with one high school in Albany, U.S.A. [18]; this was striking but not suitable for formal statistical evaluation. Since then, several anecdotal reports of HD clusters have been published, but results of formal analyses have been equivocal and the subject has been controversial [19, 20]. Analyses of space–time clustering have low statistical power for putative aetiological models for HD [21]. Recent statistical methodologies have permitted analyses of spatial clustering, and these have shown consistent evidence of spatial clustering of HD in young people [22–25] and possibly the nodular sclerosing subtype (NS) at all ages [22]. One methodology [22] permits identification of small areas as sites where clustering has occurred and classification of certain groups of cases as clusters.

Spatial clustering, elevated antibody titres and EBV genome within RS cells are three sources of evidence for an infectious aetiology for HD. They have, to date, been reported independently in separate studies. The motivation for the present study came from a belief that the investigation of these factors should be integrated. The underlying hypothesis was that EBV-RS status and/or elevated antibody titres might be associated with aggregation of exposures during school-age and adolescence [26] and hence with clustering by residence at diagnosis or at ages 5–18 years.

PATIENTS AND METHODS

The Leukaemia Research Fund Data Collection Study

Cases of HD diagnosed while resident in the Yorkshire Health Region are routinely registered by the Leukaemia Research Fund Data Collection Study (DCS) [27]. This specialist registry uses methods which ensure rapid ascertainment and diagnostic accuracy and avoid geographical bias. 494 cases of HD registered in 1985–1989 form the case series for this study. Analyses have been applied to the total series and to two subsamples (see below).

Spatial clustering

A nearest-neighbour (NN) test [22] has been used throughout. This test must be applied to all cases diagnosed in a geographical area over a period of at least 3 and preferably 5 years. It considers each case in turn and examines the location of its three nearest neighbours for unusual proximity. A simple test based on Poisson distribution and allowing for the heterogeneity of the population at risk determines whether the case is the centre of a small aggregation of cases (Figure 1 and Appendix). If so, the case is called clustered and its NNs are called peripheral.

If the allocation of cases to the population at risk were random, then clustered cases would still arise and, therefore, a global test of clustering is required. For this purpose, we examined the overall numbers or percentages of clustered cases [22, 28]. Before the present study commenced, we had already applied the test to HD cases in young people (≤34 years) diagnosed in the Yorkshire Health Region for 1984–1987 [22]. The percentages of clustered cases was 17% whereas only 8% would have arisen by chance ($P < 0.05$). A second application to NS cases at all ages for the same time period also showed significant clustering [22]. The circles containing each clustered case and its peripheral cases were designated prospectively as cluster sites for case sampling (see below).

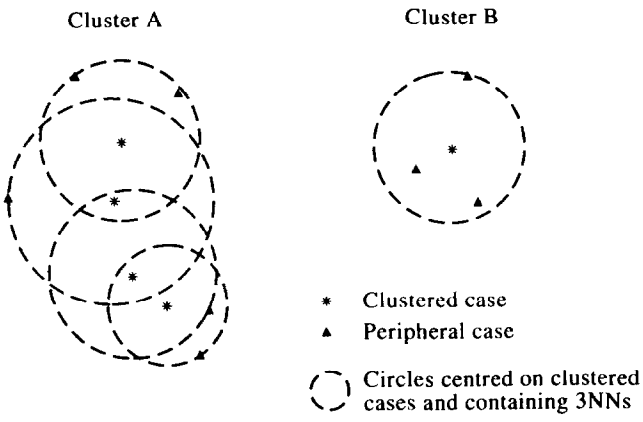

Figure 1. Nearest neighbours and clusters.

The NN test has now been applied to the Yorkshire Health Region for the period 1985–1989 so as to test for clustering and to classify cases as clustered, peripheral or random. Clustered cases have been grouped into clusters by aggregating those which were NNs of one another; see Figure 1 which shows two clusters. Each cluster will normally have a number of peripheral cases around its edge.

The analytical studies: sampling and data collection (Table 1)

The case–control study was restricted to DCS cases aged ≤34 years or of the NS subtype diagnosed in 1988 and 1989. For young people, 66 cases were eligible and 27 cases were selected for inclusion. For older NS cases, 36 cases were eligible and 12 cases were selected. The intention was to favour clustered cases in the sampling process, therefore, the 39 cases selected for the case–control study included all eligible cases ($n = 13$) resident in the areas which had been identified prospectively as cluster sites (see above). 26 other cases were selected randomly from the other eligible cases preserving the age, sex and Rye-type distribution. In addition, 26 healthy controls were randomly selected from representative GP practices—two from each of the 13 practices with cases in the prospective cluster sites—and were frequency matched to the cases by sex and age group (0–18, 19–34, 35–49, 50+ years). All subjects selected for the study agreed to participate and all, apart from one control, provided blood samples for serological analysis.

The original intention was to include matched case–control comparisons of clustered and random cases. However, this was not possible since the prospective sites failed to identify all clustered cases.

For the cluster study, all other young cases ($n = 14$) registered with the DCS between 1985 and 1989 and in the same clusters of young-onset HD as EBV-RS+ cases from the case–control study were identified.

Residential and school histories were collected from all sub-

jects in either study and paraffin-embedded blocks were retrieved for all HD cases so that EBV-RS status could be assigned.

Laboratory methods: serological analyses

Antibodies against EBV-VCA, -EA and HHV-6 were detected using standard indirect immunofluorescence assay (IFA) as described previously [8, 11, 29]. All sera were initially screened at a 1:10 dilution, and those which were scored as positive were diluted two-fold until the end titre was reached. The antibody end titre was defined as the reciprocal of the serum dilution at which specific immunofluorescence was last seen.

Laboratory methods: detection of EBV within RS cells

The presence of EBV within RS cells, referred to as the EBV-RS status, was determined using immunohistochemistry and *in situ* hybridisation as described below.

Immunohistochemical analysis. Sections of paraffin-embedded material from 51 cases (37 in the case–control study and all 14 in the cluster study) were examined for the presence of the EBV-encoded latent membrane protein-1 (LMP-1) using methodology which has been described previously [29]. Satisfactory material was not available for 2 cases. Briefly, sections were incubated with a mixture of monoclonal antibodies (CS1-4) raised against LMP-1 and reactivity was detected using an ABC technique (Dako U.K. Ltd), incorporating fast red as the chromogenic substrate. A known positive sample was included in each assay and sections from all cases were tested using an antibody to rotavirus to control for non-specific background staining.

In situ *hybridisation.* The expression of the EBV-encoded RNA, EBER-1, which is abundantly transcribed in cells latently infected by EBV, was investigated in a subset of cases. These included 2 cases in which the results of the immunohistochemical analyses were unclear, all 14 cases in the cluster study and a random selection of 20 from the other 37 cases. The material from the case–control study had been analysed by immunohistochemistry before the *in situ* hybridisation assay was established in our laboratory. Since concordance of the two assays was excellent—see Results—further application of this assay was not considered necessary. Sections from routinely fixed, paraffin-embedded material were hybridised with a biotinylated oligonucleotide probe complementary to the EBER-1 RNA. Hybridisation was detected using an ABC technique (Dako), utilising nitroblue tetrazolium as the chromogenic substrate. A known positive sample was included in each assay and a biotinylated nonsense oligonucleotide was used as a negative control.

Cases were considered EBV-associated or EBV-RS+ if the RS cells were positive in either the immunohistochemical or *in situ* hybridisation analyses.

Statistical methods

Statistical analysis has used logistic regression [30] and exact methods of analysis for 2 × k tables [31] with implementation in the statistical package EGRET and preference for exact methods in univariate analysis. For tabulation and statistical analysis all serological outcomes have been split into four levels with as near equal frequencies as possible.

Table 1. Subjects included in the analytical studies

Study	Status	0–34 years	35 + years
Case–control	Cases (NS)	22	12
	Cases (AOS)	5	—
	Controls	18	8
Cluster (additional)	Cases (NS)	11	—
	Cases (AOS)	3	—
Cluster (total)	Cases (NS)	18	—
	Cases (AOS)	5	—
Total	Cases (NS)	33	12
	Cases (AOS)	8	—
	Controls	18	8

NS, nodular sclerosing; AOS, all other subtypes.

RESULTS

When DCS registration data for young people in the entire Yorkshire Health Region (1985–1989) were tested for spatial clustering, 18% of all cases were classified as clustered ($P < 0.05$). No clustering was found for older people (9.0% clustered) or for NS cases at all ages (11.5% clustered, $P > 0.05$).

Case–control study

Of the 39 cases in the study, 14 were classified as clustered or peripheral (6 clustered, 8 peripheral) and 25 cases as random. The random cases included all 12 NS cases older than 34 years and the only paediatric case, which was also NS.

Univariate analyses of serological data (Table 2) showed statistically significant elevations of antibody activity for each antigen when HD cases were compared with controls. In multivariate analyses, the differences for HHV-6 persisted after allowing for the EBV antigens, but not conversely.

LMP-1 protein was detected within the RS cells of 4 of the 37 cases examined by immunohistochemistry. Samples from 2 cases which gave rise to equivocal results were subsequently examined using the EBER *in situ* hybridisation assay and 1 was scored as positive. In the 20 additional cases investigated using both the

Table 2. Case–control study: antibody activity for Hodgkin's disease (HD) cases and controls

	HD case	Controls*	OR (CI)
EBV-VCA antibody titres			
Negative–320	6	10	1.00
640	12	7	2.78 (0.6–14.3)
1280–2560	11	6	2.94 (0.6–16.7)
5120–10240	10	2	7.69 (1.1–100.0)†
EBV-EA antibody titres			
Negative	15	11	1.00
10	4	8	0.37 (0.06–1.9)
20–40	11	3	2.71 (0.5–16.7)
80–320	9	3	2.22 (0.4–14.3)†
HHV-6 antibody titres			
Negative	10	11	1.00
10–40	7	8	0.96 (0.2–4.5)
80–160	11	4	3.03 (0.6–16.7)
320–1280	11	2	5.88 (0.9–50.0)†

* For 25 controls (blood sample not available for one control). † Exact two-tailed P-value for trend <0.05. OR, odds ratio; CI, confidence interval; EBV, Epstein–Barr virus; VCA, viral capsid antigen; EA, early antigen; HHV-6, human herpes virus 6.

Table 3. Case–control study: Epstein–Barr virus Reed–Sternberg (EBV-RS) status and antibody titres to EBV and human herpes virus 6 (HHV-6) antigens

	HHV6		EBV-EA		EBV-VCA	
	EBV-RS+	EBV-RS−	EBV-RS+	EBV-RS−	EBV-RS+	EBV-RS−
Level of antibody titres*						
1	3	7	1	13	0	6
2	1	4	1	3	2	9
3	1	10	2	9	0	10
4	0	11	1	7	3	7
P trend†	0.042		0.70		0.26	

EA, early antigen; VCA, viral capsid antigen. * Levels correspond to ranges of titres. For HHV-6: negative, 10–40, 80–160 and 320–1280; for EBV-EA: negative, 10, 20–40 and 80–320; for EBV-VCA: negative–320, 640, 1280–2560 and 5120–10240. † Two-sided test for trend in either direction.

above techniques there was perfect concordance between the results of the two assays. Therefore, 5 of the 39 cases examined were EBV-RS+; these included 2 of the 21 young adult NS cases and 2 of the 5 young cases with HD of other subtypes but none of the older NS cases. The paediatric case was EBV-RS+.

Further analyses investigate associations between pairs of the three available status classifications of HD cases by EBV-RS, clustering and serology.

EBV-RS status and serology. No association was found between antibodies to EBV antigens and EBV-RS status (Table 3). However, there was a statistically significant negative relationship between anti-HHV-6 titres and EBV-RS status. The geometric mean titre (GMT) for anti-HHV-6 activity was 73.7 for the EBV-RS− cases but only 11.5 for the EBV-RS+ cases which was lower than the control level of 20.5. Replication of these analyses restricted to young people yielded almost identical results (data not shown).

Clustering and serology. There was no evidence of association with EBV antibody activity (Table 4), but there was weak evidence of lower antibody titres for HHV-6 in the clustered and peripheral cases.

EBV-RS status and clustering. All of the random cases, apart from the one child, were EBV-RS− and the highest ratio of EBV-RS+ to EBV-RS− was in the clustered cases (Table 5). The trend of the proportion of cases EBV-RS+ across categories of clustering is statistically significant.

The cluster study

Details of DCS cases in young-onset clusters involving the 4 EBV-RS+ cases from the case–control study are given in Table 6. These include 9 other cases from the case–control study and 14 additional cases. Overall, 7 of the cases were EBV-RS+ which is similar to what might be expected for this age range and no single cluster consisted entirely of positive cases. Further comparisons of EBV-RS status and clustering for these new cases and for the two studies combined are provided in Table 5. Overall, 5/16 clustered cases, 2/12 peripheral cases and 1/25 random cases were EBV-RS+ which confirms the tendency of EBV-RS+ cases to be associated with case clusters ($P = 0.017$).

Table 4. Case–control study: antibody activity and case clustering

	Clustered/peripheral cases	Random cases	OR (CI)
EBV-VCA antibody activity			
Negative–320	2	4	1.00
640	5	7	0.71 (0.05–17.7)
1280–2560	3	8	1.31 (0.08–17.4)
5120–10240	4	6	0.76 (0.05–9.1)
EBV-EA antibody activity			
Negative	4	11	1.00
10	2	2	0.39 (0.02–17.0)
20–40	3	8	0.97 (0.12–18.6)
80–320	5	4	0.31 (0.04–2.25)
HHV-6 antibody activity			
Negative	4	6	1.00
10–40	2	5	1.62 (0.15–25.3)
80–160	5	6	0.81 (0.10–6.15)
320–1280	3	8	1.73 (0.20–16.8)

EBV, Epstein–Barr virus; VCA, viral capsid antigen; EA, early antigen; HHV-6, human herpes virus 6; OR, odds ratio; CI, confidence interval.

Table 5. Case–control and cluster studies: Epstein–Barr virus Reed–Sternberg (EBV-RS) status and case clustering

Sample	Clustering classification: Clustered* (EBV-RS+/total)	Peripheral* (EBV-RS+/total)	Random* (EBV-RS+/total)	P†
Case–control study (all)	2/6 (33%)	2/8 (25%)	1/25 (4%)	0.04
Cluster study (all additional cases)	3/10 (30%)	0/4 (0%)	—	—
Total	5/16 (31%)	2/12 (17%)	1/25 (4%)	0.017

* Clustered and peripheral cases are all in young people (ages ≤34 years); the random group includes some older onset nodular sclerosing cases. † Exact one-sided *P* value for trend across the levels of clustering.

Table 6. Cluster study: Epstein–Barr virus Reed–Sternberg (EBV-RS) status within young onset (0–34 years) case clusters involving EBV-RS+ cases in the case–control study

Cluster number	Clustered cases	Peripheral cases	Type	EBV-RS+ status
1*	HD43†	—	NS	+
	HD63†	—	NS	−
	A	—	NS	+
	B	—	NS	−
	C	—	MC	+
	D	—	NS	−
	E	—	NS	−
	—	HD09†	NS	−
	—	F	NS	−
	—	G	NS	−
2	HD11†	—	NS	−
	—	HD61†	NS	−
	—	HD07†	NS	−
	—	HD32†	NS	+
3	H	—	NS	−
	I‡	—	LP	−
	—	HD36†	MC	+
	—	J‡	NS	−
	—	K	NS	−
4	L	−	NS	−
	M	−	NS	+
	N	−	MC	−
	HD08†	−	LP	+

* The three near neighbours of HD43 are A, B and C; of A are HD43, B and C; of C are B, D and F. † Identifiers beginning HD (Hodgkin's disease) indicate cases in the case–control study. ‡ Denotes paediatric case. MC, mixed cellularity; LP, lymphocyte predominant; NS, nodular sclerosing.

The group HD43, A, C of EBV-RS+ cases in cluster one is interesting because of the strong nearest neighbour links. The two EBV-RS+ NNs of HD43 (i.e. A, C) had lived outside the study area until their twenties. Cases HD43 and E had lived close together for a prolonged period and had common (sibship) contact with the same school from 1966 to 1971, but these cases differ for EBV-RS status. Shared school experience cannot then be the explanation for the occurrence of EBV-RS+ cases together in cluster one. HD09 and HD63 (both EBV-RS−) had also lived close together for some time and had attended the same school from 1977 to 1979. No other children in either sample had attended (or had siblings who attended) the same schools at the same time. Three other EBV-RS+ cases (HD32, HD36 and HD08) also moved into Yorkshire after leaving school. It follows that neither residential proximity during school days nor shared contact with particular schools can explain the excess clustering of the EBV-RS+ cases.

DISCUSSION

The interpretation of these data must be cautious because of the small numbers involved, but this study is important because

it is currently unique in integrating clustering and virological classifications for HD.

The analyses of the individual factors are in close agreement with those of other larger studies. Previous applications of the NN test [22, 25] and other methodologies [23, 24] have demonstrated spatial clustering of HD in young people. NS subtypes predominate in young people and it has not been clear whether the clustering extended to NS cases arising at all ages [22, 25] but these new data provide further evidence that this is unlikely. Sero-epidemiological studies have consistently reported elevated antibodies to the EBV antigens VCA and EA and to HHV-6 in HD cases compared with controls. The proportions of HD cases in which direct evidence of EBV within the tumour cells was found (5/39 for the case–control study, 7/23 for the cluster study) is similar to other studies concentrating on young adults and/or the NS subtypes [17, 32].

The aspects of the present results which are new and potentially important are those which involve associations between the three factors.

EBV-RS status and EBV serology

There is, firstly, no evidence that raised antibodies to the EBV antigens are restricted to EBV-RS+ cases. These and other data [33] reveal the complexity of the relationship (if any) between elevated EBV antibody titres and EBV presence in tumour cells. The consistency with which elevated titres to EBV antigens have been reported in HD cases is, however, strong evidence that they indicate (at least as proxies) a causal factor for HD even though that factor need not be EBV itself.

EBV-RS status and clustering

Secondly, EBV-RS+ status, but not EBV-VCA nor EBV-EA antibody titres, are associated with case clustering in these data; the results are statistically significant but based on small numbers. There was little evidence that EBV-RS+ cases were concentrated in the same cluster(s), and no cluster consisted entirely of EBV-RS+ cases. It is mathematically possible (see Appendix) for the spatial clustering to be attributed to these EBV-RS+ cases but the most plausible interpretation of these data is that spatial clustering denotes shared exposure to an unknown factor X which is a co-factor in both EBV-RS+ and (at least some) EBV-RS− disease in young people. EBV association is most frequent in older cases and those which are not of the NS subtype [17, 32], but our results are not applicable to older cases of other subtypes.

Since the latent period between exposure to the putative infectious agent of the late host response model and symptomatic HD is expected to be both long (years rather than weeks) and variable, clustering of location at diagnosis will be, at best, a poor proxy for aggregation of aetiological exposures [20]. It was anticipated that examination of residential and school histories might have provided additional insight. However, the majority of the EBV-RS+ cases had moved into the study area quite recently so that the putative factor X (if associated with both clustering and HD) must act at a late stage.

The significance of the spatial clustering lies in the overall proportion of cases classified as clustered, and statistical methodology cannot distinguish those clustered cases which arise by chance from those which do not and which may be attributable to aggregations of an aetiological agent. Even within the latter clusters, random cases will be found. Thus, both the classification of high proportions of EBV-RS+ cases as clustered or peripheral and the presence of EBV-RS− cases within clusters should be interpreted cautiously. It is appropriate to recall the role of clustering in the known aetiological associations of the feline leukaemia virus (FeLV) with feline leukaemia and EBV with endemic Burkitt's lymphoma (eBL). FeLV is a causal agent for the majority of feline leukaemia cases but case clustering attributable to shared exposure to FeLV is only seen in a minority of cats with unusual living conditions favouring early and high dose exposures [34]. Case clustering of eBL is common although not universal, but almost certainly not attributable to EBV; tumour cells from almost all cases of eBL contain EBV viral genome following, it is believed, early exposure to EBV. Case clustering is due to shared exposure to other co-factor(s), notably malaria [35]. Clustering of EBV-associated HD due to shared exposure to factor X would be similar in many respects to the eBL model, but feline epidemiology indicates that factor X need not be restricted to case clusters.

HHV-6 serology in relation to EBV-RS status and clustering

Antibody titres to HHV-6, although generally elevated in HD cases, were lower than the controls for EBV-RS+ cases. These results are based on small numbers but are statistically significant, and tentatively suggest that high levels of anti-HHV-6 activity indicate a causal exposure which is alternative to and independent of EBV and unlikely to be factor X. Since we have sought (but not found) evidence for HHV-6 viral genomes in HD tumour biopsies [36], it is possible that another agent is directly involved. Our data provide no evidence that HHV-6 antibody levels mark an exposure which is specific to either clustered or random EBV-RS− cases.

In conclusion, this pilot study emphasises the potential strength of epidemiological investigations of HD which integrate clustering, molecular biology and serology. It also raises hypotheses for testing on independent and larger data sets: (i) EBV-RS+ cases in young people are likely to be found within case clusters but clustering is attributable to shared exposure to an unknown co-factor rather than EBV; and (ii) high antibody titres to HHV-6 are markers for causal pathway independent to EBV and are found primarily in EBV-RS− cases. A larger case–control study is now in progress to test these hypotheses and with sufficient geographical and longitudinal scope to investigate aggregation of exposures at earlier time periods as well as clustering at time of diagnosis.

1. Correa P, O'Conor GT. Epidemiologic patterns of Hodgkin's disease. *Int J Cancer* 1971, **8**, 192–201.
2. Muir CS, Waterhouse J, Mack T, Power J, Whelan S. *Cancer Incidence in Five Continents*. Lyon, IARC, Scientific Publications no. 88, 1987.
3. MacMahon B. Epidemiology of Hodgkin's disease. *Cancer Res* 1966, **26**, 1189–1200.
4. Gutensohn N, Cole P. Epidemiology of Hodgkin's disease. *Int J Cancer* 1977, **19**, 595–604.
5. Gutensohn N, Cole P. Childhood social environment and Hodgkin's disease. *N Engl J Med* 1981, **304**, 135–140.
6. Paffenberger RS, Wing AL, Hyde RT. Characteristics in youth indicative of adult-onset Hodgkin's disease. *J Natl Cancer Inst* 1977, **58**, 1489–1491.
7. Munoz N, Davidson RJL, Witthoff B, Ericsson JE, de-The G. Infectious mononucleosis and Hodgkin's disease. *Int J Cancer* 1978, **22**, 10–13.
8. Henle W, Henle G. Epstein–Barr virus-related serology in Hodgkin's disease. *Natl J Cancer (Natl Cancer Inst Monograph)* 1973, **36**, 79–84.
9. Mueller N, Evans A, Harris NL, *et al.* Hodgkin's disease and Epstein–Barr virus. *N Engl J Med* 1989, **320**, 689–695.
10. Mueller N. An epidemiologist's view of the new molecular biology findings in Hodgkin's disease. *Ann Oncol* 1991, **2**, 23–28.

11. Clark DA, Alexander FE, McKinney PA, Onions D, Jarrett R, Cartwright RA. The sero-epidemiology of human herpes virus-6 (HHV-6) from a Yorkshire case–control study of leukaemia and lymphoma. *Int J Cancer* 1990, **45**, 829–833.
12. Serraino D, Franceschi S, Talamini R, *et al.* Socio-economic indicators infectious diseases and Hodgkin's disease. *Int J Cancer* 1991, **47**, 352–357.
13. Weiss LM, Movahed LA, Warnke A, Sklar J. Detection of Epstein–Barr viral genomes in Reed–Sternberg cells of Hodgkin's disease. *N Engl J Med* 1989, **320**, 502–506.
14. Pallesen G, Hamilton-Dutoit SJ, Rowe M, *et al.* Expression of Epstein–Barr virus latent gene products in tumour cells of Hodgkin's disease. *Lancet* 1991, **337**, 320–322.
15. Armstrong AA, Weiss LM, Gallagher A, *et al.* Criteria for the definition of Epstein–Barr virus association in Hodgkin's disease. *Leukaemia* 1992, **6**, 869–874.
16. Anagnostopoulos I, Herbst H, Niedobitek G, Stern H. Demonstration of EBV genomes in HD and RI-1-positive anaplastic large cell lymphoma by combined southern blot and *in situ* hybridisation. *Blood* 1989, **74**, 810–816.
17. Jarrett RF, Gallagher A, Jones DB, *et al.* Detection of Epstein–Barr virus genomes in Hodgkin's disease: relation to age. *J Clin Pathol* 1991, **44**, 844–848.
18. Vianna NJ, Greenwald P, Davies JNP. Extended epidemic of Hodgkin's disease in high-school students. *Lancet* 1971, i, 1209–1211.
19. Smith PG. *Epidemiology of Human Lymphomas.* UICC Technical Report Series 370, 1978.
20. Grufferman S. Clustering and aggregation of exposures in Hodgkin's disease. *Cancer* 1977, **39**, 1829–1833.
21. Chen R, Mantel N, Klingberg MA. A study of three techniques for time-space clustering in Hodgkin's disease. *Stats Med* 1984, **3**, 173–184.
22. Alexander FE, Williams J, McKinney PA, Ricketts TJ, Cartwright RA. A speccialist leukaemia/lymphoma registry in the UK. Part 2: clustering of Hodgkin's disease. *Br J Cancer* 1989, **60**, 948–952.
23. Urquhurt J, Black R, Buist E. *Exploring Small Area Methods. Methodology of Enquiries into Disease Clustering.* London, SAHSU0, 1989.
24. Glaser SL. *Evidence for Spatial Clustering of Hodgkin's Disease in the San Francisco Bay Area.* Abstract of scientific presentations, National Conference on Clustering of Health Events Atlanta, 1989, 28.
25. Alexander FE, McKinney PA, Williams J, *et al.* Epidemiological evidence for the "two-disease hypothesis" in Hodgkin's disease. *Int J Epidemiol* 1991, **20**, 354–361.
26. Alexander FE. Clustering and Hodgkin's disease. *Br J Cancer* 1990, **62**, 708–711.
27. McKinney PA, Alexander FE, Ricketts TJ, Williams J, Cartwright RA. A specialist leukaemia/lymphoma registry in the UK. Part 1: incidence and geographical distribution of Hodgkin's disease. *Br J Cancer* 1989, **60**, 942–947.
28. Alexander FE, Ricketts TJ, Williams J, Cartwright RA. Methods of mapping and identifying small clusters of rare diseases with applications to geographical epidemiology. *Geog Analysis* 1991, **23**, 158–173.
29. Long C, Derge JG, Hampar B. Procedure for activating Epstein–Barr virus early A antigen in nonproducer cells by 5-iododeoxyuridine. *J Natl Cancer Inst* 1974, **52**, 1355–1357.
30. Breslow NE, Day NE. *Statistical Methods in Cancer Research: Volume I—The Analysis of Case–Control Studies.* Lyon, IARC Scientific Publications no. 32, 1980.
31. Mehta CR, Patel NR, Gray R. On computing an exact interval for the common odds ratio in several 2 × 2 contingency tables. *JASA* 1985, **80**, 969–973.
32. Pallesen G, Hamilton-Dutoit SJ, Rowe M, Young LS. Expression of Epstein–Barr virus latent gene products in tumour cells of Hodgkin's disease. *Lancet* 1991, **337**, 320–322.
33. Brousset P, Chittal S, Schlaifer D, *et al.* Detection of Epstein–Barr virus messenger RNA in Reed–Sternberg cells of Hodgkin's disease by *in situ* hybridisation with biotinylated probes on specially processed modified acetone methyl benzoate xylene (ModAMeX) sections. *Blood* 1991, **77**, 1781–1786.
34. Onions D. Animal models: lessons from feline and bovine leukaemia virus infections. *Leukemia Res* 1985, **9**, 709–711.
35. Facer AC, Playfair JHL. Malaria, Epstein–Barr virus and the genesis of lymphomas. *Adv Cancer Res* 1989, **53**, 33–72.
36. Jarrett RF, Gledhill S, Qureshi F, *et al.* Identification of human herpes virus 6-specific DNA sequences in two patients with non-Hodgkin's lymphoma. *Leukemia* 1988, **2**, 496–502.

Acknowledgement—The study was supported by the Leukaemia Research Fund.

APPENDIX

STATISTICAL ASPECTS OF THE NEAREST NEIGHBOUR TEST

1. *The local test*

Each individual case (C) is considered in turn and its three nearest neighbours (NNs) identified. These will be at distances d_1, d_2 and d_3 from C ($d_1 \leq d_2 \leq d_3$), and these distances are the random variables of interest. The locations of all cases are taken from the ordnance survey grid reference of their postcode at diagnosis. Population denominators are not directly available for postcodes, but these have been estimated from those for enumeration districts in the 1981 UK census and 1986 midcensus estimates. Application of Leukaemia Research Fund Data Collection Study (DCS) age- and sex-specific rates (adjusted to ensure that observed and expected counts for each Yorkshire county are equal) to these populations provides expected numbers μ_1, μ_2 and μ_3 in the circles centre C with radii d_1, d_2 and d_3. Under the null hypothesis of a random distribution of cases to the population at risk, case counts in these circles have Poisson distributions with means $\mu_1 - \mu_3$. Thus, for example,

$$P(d_3 > \text{observed value})$$

$$= P(\geq 3 \text{ cases observed} \mid \text{Poisson distribution, mean } \mu_3)$$

$$= \sum_{i=3}^{\infty} \frac{\mu_3^i}{i!} e_3^{-\mu} = P_3,$$

where P is probability.

It would be possible to consider a variety of NNs but standard implementations of the test define:

$$\text{C to be clustered if } P_3 < 0.05 \text{ or } P_2 < 0.05 \text{ (or both)}.$$

A clustered case is thus the centre of a small aggregation of cases. If $P_3 < 0.05$ then all three NNs are said to be peripheral, but if $P_3 > 0.05$ (and $P_2 < 0.05$) only the two nearest neighbours are peripheral. Peripheral cases are associated as a group with local aggregations of cases but the evidence that this applies to each individually is somewhat weak unless, as often occurs, the peripheral cases are themselves clustered (see Figure 1).

2. *The global test*

Two dependent tests have been applied ($P_3 < 0.05$, $P_2 < 0.05$) and local tests for distinct cases (C) are also dependent. It is therefore not possible to provide theoretical estimations of the percentages of cases which would be classified as clustered if the case distribution were random. Simulation studies [28] have shown that this is approximately 8.5%. These studies have also provided 95 and 99% critical values for the percentages of clustered cases; these critical values depend on the total number of cases in the population.

Can the EBV-RS+ cases explain the clustering?

After the completion of the analyses of EBV-RS+ status for the case–control study and the cluster study, the status was established for 15 of the 36 clustered cases of HD in young people diagnosed in Yorkshire, 1985–1989. Of these, only 5 were EBV-RS+. However, the spatial clustering involved an excess of approximately 20 clustered cases over the expected number of 16. It appeared very unlikely that clustering could be explained by the EBV-RS+ cases. Allocating EBV-RS status to the entire case series was not feasible. An alternative, mathematical approach used extreme assumptions and examined their effect on the spatial clustering.

From the observed case distribution the following were removed (as potentially EBV-RS+): all cases known to be EBV-RS+ (5 clustered, 2 peripheral, 1 random), 5 cases selected randomly from the 21 clustered cases whose status was unknown, 27 cases selected randomly from the peripheral and random cases whose status was unknown.

This maintained a presumed overall proportion of 20% of EBV-RS+ cases with the proportion for clustered cases approaching 30%.

When the NN test was applied to these data the proportion of clustered cases was just 8.5% so that the distribution of presumed EBV-RS− cases was essentially random. Since the test conditions on the observed totals within each county, the absence of spatial clustering in the modified data set cannot be attributed to a reduction in total observed numbers. It follows that it *is* mathematically possible for the observed clustering to be attributed to EBV-RS+ cases.