

A STUDY OF SPACE-TIME CLUSTERING IN HODGKIN'S DISEASE IN THE MANCHESTER REGION

M. R. ALDERSON* AND R. NAYAK

Regional Cancer Epidemiology Unit, Manchester

MacMahon (1966), in a comprehensive review of the epidemiology of Hodgkin's disease, provided evidence for the hypothesis that Hodgkin's disease may consist of several separate entities. Cole, MacMahon, and Aisenberg (1968) supported this hypothesis by analysis of regional and age variation in mortality within the United States. Smithers (1970) has reviewed the epidemiological and clinical evidence and he concludes that it is unlikely that Hodgkin's disease is two distinct conditions. Families with more than one affected member have been reported, in which the disease occurred at the same time among relatives of different ages (DeVore and Doan, 1957; Razis, Diamond, and Craver, 1959). On the basis of this, together with the clinical and pathological features of Hodgkin's disease, MacMahon (1966) suggested that one variety may be triggered off by infections. Fraumeni and Li (1969) and Cridland (1961) have shown that an excess of cases presents in the winter months. Clemmesen, Busk, and Nielsen (1952), who used a crude mapping technique, reported microclusters of the disease. Fraumeni and Li (1969) looked for time and space clustering in 314 affected children; they were unable to identify any evidence of clustering of cases using the procedure developed by Ederer, Myers, and Mantel (1964).

Alderson and Nayak (1971) carried out a preliminary analysis of incidence data for 370 patients presenting with Hodgkin's disease over a period of four years in the Manchester area; they showed a trimodal age distribution and variation in the preponderance of males at different ages, as reported by other authors (MacMahon, 1966; Clemmesen, 1965). There was no appreciable variation in the social class or marital status of the patients compared with the population from which they were drawn, but analysis of the date of first attendance at hospital revealed a significant peak of patients first attending in the second quarter of the year. There was an excess of patients close together in space and time, the clustering being signi-

ficant at the 1% level in patients aged 50 years and over. This paper presents an analysis of additional material, and more detailed examination of the earlier reported 'positive' clustering.

METHOD

The current study has been restricted to patients resident in the Manchester Regional Hospital Area who were registered with the Cancer Registration Bureau as new cases of Hodgkin's disease between 1 January 1962 and 31 December 1968. The routine records in the Bureau include the sex, age, address, and occupation of each patient; the diagnosis, and histology when available; and the date of first hospital attendance. The nominal roll of patients registered with the Bureau was checked against the diagnostic index at the Christie Hospital to ensure that all known patients were included in the study. The diagnostic particulars were reviewed before accepting each patient into the study; histological confirmation was available for all 737 patients.

The age and sex specific incidence has been calculated using population estimates for the region contained in the Registrar General's Statistical Review of England and Wales for 1965 (1967).

Evidence of time and space clustering has been examined by the method described by Knox (1963) using a programme which reads in dates and map references for each case and categorizes every possible pair. Output consists of a contingency table together with expected values calculated from the marginal totals. Using this method a test for space-time interaction can be performed that is independent of the spatial distribution of the population; the validity of the test has been confirmed by theoretical and simulation studies (David and Barton, 1966). The cluster analysis has been restricted to those individuals resident within the Manchester conurbation. This had the advantage of using that part of the region where map references could be identified with greater accuracy, and where completeness of registration was likely to be uniform. The programme requires as input the date of onset of the disease and the map reference of the patient's

*Present address: Medical Information Unit, Wessex Regional Hospital Board, Highcroft, Romsey Road, Winchester.

address. For onset we have used the date at which the patient first attended hospital for Hodgkin's disease. A map reference has been obtained for each patient by a combined use of all street guides that were available for the conurbation and 2½ in Ordnance Survey maps (scale 1 : 25,000). On a long road the street maps were provided with house numbers at intervals so that the location of the house could be pinpointed with considerable accuracy; for shorter streets and blocks of flats the mid-point of the street has been used. It is felt that this was usually accurate to 0.1 kilometre, and always within 0.5 kilometre. Five cases had to be excluded from the analysis because it was not possible to specify the map reference with precision. In producing the time and distance table, time in days was divided as follows: 0-, 30-, 60-, 90-, 120-, 150-, 180-, 240-, 300-, 360+; distance in kilometres was divided 0-, 2-, 3-, 4-, 5-, 6-, 8-, 12+.

RESULTS

The figure illustrates the variation in incidence with age and sex; there is a bimodal distribution with a pronounced trough in the incidence curve at age 45-49. The bimodal swing in incidence by age is present for both males and females, but, as Table I shows, the percentage of the combined incidence that is due to males varies with age.

Table II shows the month of first attendance at hospital of the patients; there is an excess of cases referred to hospital in May and June, and a deficiency of cases referred to hospital at the end of the year. The spring peak shows particularly in patients aged 50 and over, and in males. Applying a test

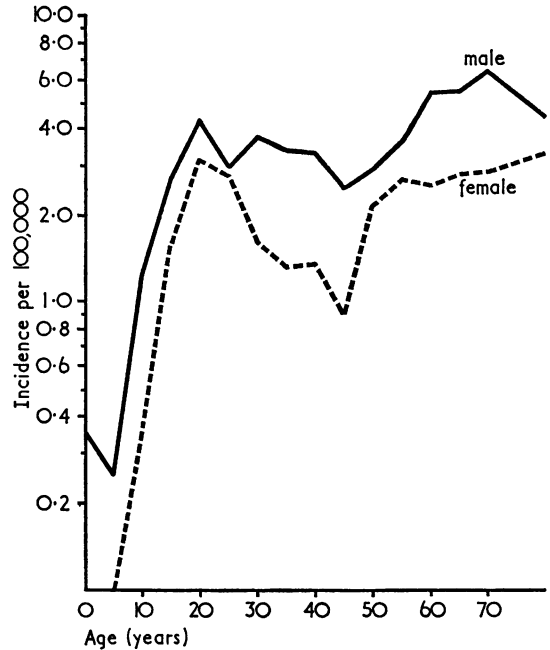


FIGURE. Average annual incidence of Hodgkin's disease per 100,000 by age and sex, in Manchester Regional Hospital area, 1962-8.

for cyclic trend devised by Edwards (1961), this finding is shown to be very unlikely to be due to chance ($0.005 > P > 0.001$). However, seasonal swings in incidence of, and mortality from, other malignant disease have been reported (Lee, 1962, 1967). A comparison was therefore made of the date of first hospital attendance for patients registered with (a) carcinoma of the lung and (b) carcinoma of the breast. These data are also included in Table II, which shows that the seasonal distribution is very similar to that for Hodgkin's disease; there is a slightly earlier peak in lung cancer and the swing in incidence of breast cancer is pronounced.

Table III provides a simple extract of the time and space analysis for the total period for which data are available, 1962-8. This is representative of all the divisions by distance and time for this

TABLE I

HODGKIN'S DISEASE: INCIDENCE IN MALES AS A PERCENTAGE OF MALE PLUS FEMALE INCIDENCE AT VARIOUS AGES

Age (yr)	Incidence in Males	Total No. of Patients
0-	78%	28
15-	61%	227
35-	72%	131
50+	62%	351

TABLE II

MONTH OF FIRST HOSPITAL ATTENDANCE FOR PATIENTS REGISTERED WITH HODGKIN'S DISEASE, CARCINOMA OF THE BREAST, AND CARCINOMA OF THE LUNG

Disease	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Total
Hodgkin's disease	61	53	51	73	80	82	56	61	59	58	59	44	737
Carcinoma of breast	52	61	60	71	71	51	50	52	52	49	36	13	618
Carcinoma of lung	27	23	33	15	32	26	21	17	21	11	9	5	240

TABLE III
SPACE-TIME DISTRIBUTION OF PAIRS OF CASES FOR 1962-8 BY AGE

Time Interval (days)	All Ages				Aged 15-44				Aged 50+			
	< 2 km		< 4 km		< 2 km		< 4 km		< 2 km		< 4 km	
	O	E	O	E	O	E	O	E	O	E	O	E
< 30	56	52.0	186	176.6	10	10.0	42	37.3	16	11.3	38	38.4
< 60	111	100.1	342	339.7	21	18.5	68	68.8	25	22.1	68	75.4
< 120	212	201.1	685	682.6	31	35.9	132	133.3	52	45.8	157	156.3
< 180	310	296.7	972	1007.3	49	53.1	179	197.1	80	68.6	238	234.2

Based on 422 cases of all ages, 198 aged 15-44 and 191 aged 50+ years.
O = observed; E = expected.

material, and does not reveal any appreciable difference between the observed and expected pairs of cases for all ages, those aged 15-44, and those aged 50 and over.

A significant clustering in those aged 50 and over who were registered during the period 1962-5 was reported in the previous paper. An examination has therefore been made of the data for each individual year from 1962 to 1968 for patients aged 15-44, 50 and over, and all ages. Table IV provides a summary of the results obtained for 1963 and 1964. Thirty-six comparisons between observed and expected figures are provided; in only one of them is the observed number of pairs less than expected, and three of the observed figures are significantly greater than the expected figure ($P < 0.05$).

For each of the other five years for which data were available there was no adequate evidence of clustering. Out of a total of 90 comparisons, a significant difference was found in only one subgroup of the data, patients aged 50 and over registered in 1968. This material is presented in Table V.

For simplicity the tables presenting the above results have been restricted to two distance categories and three time categories. These have been selected as representative of the data and suitable for indicating whether or not there is evidence of

clustering. An appendix table is provided (p. 173), which gives the distribution of pairs of cases for all the time and distance intervals for all cases living in the Manchester conurbation registered in 1962-8 (422 cases). Similar tables have been examined for each of the seven calendar years, for all ages and two age subgroups. These additional twenty-three tables are not presented here but are available on request from the authors.

DISCUSSION

Three of the observations from this study warrant discussion. The first is the variation in the age and sex specific incidence of Hodgkin's disease; the data reported here are in line with the observations of other workers. The picture is atypical for malignant diseases; apart from tumours particularly affecting young children, most malignancies show a steadily rising incidence with advancing age. This is compatible with heterogeneity within the general class of Hodgkin's disease, or some peculiar sex-linked variation in susceptibility to the disease.

The seasonal swing is similar in Hodgkin's disease, carcinoma of the breast, and carcinoma of the lung. Further study is required to show whether this is an artefact of the process by which a patient contacts his family doctor and is referred to hos-

TABLE IV
SPACE-TIME DISTRIBUTION OF PAIRS OF CASES FOR 1963 AND 1964 BY AGE

Year and Time Interval (days)	All Ages				Aged 15-44				Aged 50+			
	< 2 km		< 4 km		< 2 km		< 4 km		< 2 km		< 4 km	
	O	E	O	E	O	E	O	E	O	E	O	E
1963												
< 30	9	8.0	38	27.7	2	1.0	9	4.3*	2	1.7	8	6.3
< 60	20	15.1	60	52.5	4	2.1	12	9.0	5	3.1	13	11.8
< 120	35	28.2	111	98.1	5	4.1	22	17.5	7	6.0	24	22.5
1964												
< 30	11	7.8	30	23.7	3	1.5	5	5.2	3	2.5	7	6.2
< 60	20	13.4	54	40.6*	6	2.5*	10	8.4	6	4.2	13	10.2
< 120	31	25.7	91	77.6	8	4.9	20	16.5	9	8.3	24	20.2

Based in 1963 on 61 cases of all ages, 30 aged 15-44, 24 aged 50+.
in 1964 on 63 cases of all ages, 34 aged 15-44, 28 aged 50+.

* $P < 0.05$.

TABLE V
SPACE-TIME DISTRIBUTION OF PAIRS OF CASES BY AGE AND CALENDAR YEAR

Year and Time Interval (days)	All Ages				Aged 15-44				Aged 50+			
	< 2 km		< 4 km		< 2 km		< 4 km		< 2 k		< 4 km	
	O	E	O	E	O	E	O	E	O	E	O	E
1962												
< 30	5	4.8	14	15.4	1	0.9	3	3.0	2	1.1	3	3.0
< 60	9	10.1	32	32.2	1	1.9	5	6.0	2	2.2	5	6.0
< 120	14	16.9	52	53.9	1	3.1	6	9.8	5	4.0	11	10.8
1965												
< 30	11	11.6	33	32.2	1	2.0	5	6.3	4	2.8	10	8.6
< 60	19	21.2	55	59.0	2	3.6	9	11.2	6	5.4	16	16.1
< 120	36	40.2	109	111.6	5	6.4	19	19.7	12	10.3	34	31.0
1966												
< 30	6	8.5	26	25.3	0	1.5	3	4.6	2	1.1	4	5.5
< 60	16	17.4	54	51.7	3	3.2	10	9.7	2	2.4	9	11.9
< 120	27	31.8	94	94.6	5	5.6	18	16.8	4	4.5	22	22.7
1967												
< 30	8	8.7	24	24.4	3	2.1	11	6.6	1	1.7	4	4.5
< 60	13	16.1	45	45.5	3	3.8	14	11.6	2	3.2	8	8.7
< 120	28	29.4	77	83.1	4	5.7	17	17.8	5	6.2	13	17.0
1968												
< 30	4	3.9	16	15.1	0	0.8	4	5.0	2	0.5*	2	1.5
< 60	9	7.2	25	28.3	2	1.6	6	9.8	2	0.9	3	2.7
< 120	13	14.0	53	54.9	2	3.1	15	18.9	2	1.8	4	5.5

*P < 0.05.

pital, or is actually related to a seasonal swing in onset of the disease.

The earlier paper of Alderson and Nayak (1971) reported a statistically significant space-time cluster in Hodgkin's disease; before attempting to interpret the extended results presented here it is important to consider the possible deficiencies of the data:

(a) We have no clear indications as to the completeness of the registered cases. If missing cases occurred at random this should make it more difficult to identify clustering.

(b) With a short time interval between initiation and development of disease the technique used here might be expected to pick up clustering. There are two separate problems with the extremely crude routine data processed in this study. The time of onset of Hodgkin's disease (here date of hospital attendance) may be remote in time from the initiating event, and if there is considerable variation about the mean latent period, clustering will be extremely difficult to identify. If the 'environmental agent' acts away from the home environment, then home address may be a poor index of spatial clustering. These two factors could reduce the likelihood of detecting space-time clustering in Hodgkin's disease.

(c) The variation in age specific incidence has been advanced as one indicator of heterogeneity in Hodgkin's disease. There is evidence from histological studies that there is a variety of histological types, associated with the age of the patient and response to treatment. Though there was histological confirmation of the diagnosis for the patients

included in this study, the routine reports available were insufficient to allot the cases to more specific subgroups of Hodgkin's disease. It is conceivable that clustering occurs in one variety of the condition to a greater extent than in other varieties; analysis of all material together would lessen the chance of detecting clustering.

(d) The validity of the statistical technique used for examination of the data has been supported by Monte Carlo studies carried out by Pike (1966); the above remarks emphasize, however, that the model under consideration may be a gross oversimplification of the actual situation.

One other point to bear in mind whilst attempting to interpret data is the number of subgroups of the data that have been examined. An examination for clustering has been made for seven separate calendar years, and for the total seven-year period 1962-8; for two age groups, 15-44 and 50+, and for all ages; for two distances; and for three lengths of time. In all, 144 comparisons between observed and expected numbers have been presented; though each of these is not independent (some being combinations of smaller subgroups) one could expect by chance a number of tests to be significant at the 5% level, and one at the 1% level.

The above considerations indicate that the interpretation of the data is by no means straightforward. Examination of the total data showed no evidence of clustering; however, there was evidence for low-level clustering in 1963 and 1964. The important question arises as to whether this is an

artefact or whether it represents a transient instability in the transmission pattern or distribution of an external agent. Routine data can only be used as pointers for further study; it is suggested therefore, that an examination of Hodgkin's cases occurring in different geographical areas should be undertaken. Improved techniques of testing for clustering are required for the study of disease with a long latent period between initiation and onset. A more detailed study to investigate fully the possible mechanisms of clustering in Hodgkin's disease is not warranted on the basis of the material presented here.

SUMMARY

Analysis of data for patients presenting with Hodgkin's disease in the Manchester Regional Hospital area from 1962 to 1968 showed a bimodal distribution of incidence by age, with variation in the preponderance of males at different ages.

There was a significant peak of patients first attending hospital in the second quarter of the year. A similar seasonal variation was present in data for patients with lung cancer and breast cancer.

Space-time analysis showed no evidence of clustering in the total data, but analysis year by year showed evidence of clustering among patients aged 15 to 44 in 1963 and 1964. Apart from the fact that one would expect by chance a few differences significant at the 5% level, the long latent period of the disease and errors in the data would tend to lower the likelihood of detecting clustering. It is suggested that further examination of material from other geographical areas should take place before more detailed studies are begun in search of aetiological factors.

We are most grateful to Professor A. Smith, who encouraged us to pursue this matter, to the colleagues who have discussed this problem with us, and to the many clinicians and staff of records departments who co-operated in providing data for the cancer registry. Mrs. V. Hillier arranged for computer analysis of the data using the programme kindly provided by Professor G. Knox.

R.N. was a Commonwealth Medical Fellow from Orissa Medical School, India.

REFERENCES

- ALDERSON, M. R., and NAYAK, R. (1971). Epidemiology of Hodgkin's disease. *J. chron. Dis.*, in press.
- CLEMMESSEN, J. (1965). *Statistical Studies in the Aetiology of Malignant Neoplasms. I. Review and Results*. Munksgaard, Copenhagen.
- , BUSK, T., and NIELSEN, A. (1952). The topographical distribution of leukemia and Hodgkin's disease in Denmark 1942–1946. *Acta radiol. (Stockh.)* **37**, 223.
- COLE, P., MACMAHON, B., and AISENBERG, A. (1968). Mortality from Hodgkin's disease in the United States. *Lancet*, **2**, 1371.
- CRIDLAND, M. D. (1961). Seasonal incidence of clinical onset of Hodgkin's disease. *Brit. med. J.*, **2**, 621.
- DAVID, F. N., and BARTON, D. E. (1966). Two space-time interaction tests for epidemicity. *Brit. J. prev. soc. Med.*, **20**, 44.
- DEVORE, J. W., and DOAN, C. A. (1957). Studies in Hodgkin's syndrome. XII. Hereditary and epidemiologic aspects. *Ann. intern. Med.*, **47**, 300.
- EDERER, F., MYERS, M. H., and MANTEL, N. (1964). A statistical problem in space and time: Do leukemia cases come in clusters? *Biometrics*, **20**, 626.
- EDWARDS, J. H. (1961). The recognition and estimation of cyclic trends. *Ann. hum. Genet.*, **25**, 83.
- FRAUMENI, J. F. JR., and LI, F. P. (1969). Hodgkin's disease in childhood: an epidemiologic study. *J. nat. Cancer Inst.*, **42**, 681.
- KNOX, G. (1963). Detection of low intensity epidemicity. Application to cleft lip and palate. *Brit. J. prev. soc. Med.*, **17**, 121.
- LEE, J. A. H. (1962). Seasonal variation in the clinical onset of leukaemia in young people. *Brit. med. J.*, **1**, 1737.
- (1967). Summer and death from neuroblastoma. *Brit. med. J.*, **2**, 404.
- MACMAHON, B. (1966). Epidemiology of Hodgkin's disease. *Cancer Res.*, **26**, 1189.
- PIKE, M. C. (1966). Unpublished work quoted in David and Barton (1966) q.v.
- RAZIS, D. V., DIAMOND, H. D., and CRAVER, L. F. (1959). Familial Hodgkin's disease: its significance and implications. *Ann. intern. Med.*, **51**, 933.
- REGISTRAR GENERAL'S STATISTICAL REVIEW OF ENGLAND AND WALES FOR 1965 (1967). Part I. Tables, Medical H.M.S.O., London.
- SMITHERS, D. W. (1970). Hodgkin's disease: one entity or two? *Lancet*, **2**, 1285.

APPENDIX TABLE
 HODGKIN'S DISEASE: SPACE-TIME DISTRIBUTION OF PAIRS OF ALL AGES REGISTERED 1962-8*

Days		< 2 km	< 3 km	< 4 km	< 5 km	< 6 km	< 8 km	< 12 km	Total
< 30	O	56	109	186	270	378	614	1163	
	E	52.0	107.1	176.6	263.0	364.0	601.3	1128.6	2176
< 60	O	111	200	342	501	707	1161	2224	
	E	100.1	206.0	339.7	505.8	700.0	1156.5	2170.6	4185
< 90	O	157	288	513	755	1070	1740	3317	
	E	151.9	312.6	515.6	767.7	1062.5	1755.4	3294.5	6352
< 120	O	212	394	685	1007	1399	2295	4363	
	E	201.1	413.9	682.6	1016.4	1406.8	2324.2	4361.9	8410
< 150	O	263	489	831	1236	1710	2821	5374	
	E	248.9	512.4	845.0	1258.3	1741.5	2877.1	5399.8	10411
< 180	O	310	578	972	1460	2016	3358	6390	
	E	296.7	610.8	1007.3	1500.0	2076.0	3429.9	6437.1	12411
< 240	O	419	777	1294	1917	2658	4459	8459	
	E	392.2	807.3	1331.4	1982.4	2743.8	4533.1	8507.6	16403
< 300	O	502	953	1588	2371	3284	5523	10471	
	E	486.0	1000.3	1649.6	2456.3	3399.6	5616.7	10541.2	20324
< 360	O	591	1137	1901	2814	3931	6571	12431	
	E	577.8	1189.4	1961.4	2920.7	4042.3	6678.4	12533.9	24166
Total		2124	4372	7210	10736	14859	24549	46073	88831

*This is a cumulative frequency table based on 422 cases showing the observed and expected frequencies of pairs within particular distance and time intervals. The marginal totals include pairs separated by more than 12 kilometres or 360 days; the observed and expected values for these cells are identical.