Case Clustering in Hodgkin's Disease: A Brief Review of the Present Position and Report of Current Work in Oxford

P. G. Smith and M. C. Pike

Department of Health and Social Security, Cancer Epidemiology and Clinical Trials Unit, Oxford University, 9 Keble Road, Oxford, United Kingdom

ALBANY STUDY

The most important investigation to date of possible case clustering in HD is that of Vianna et al. (9) in Albany County of New York State. If we accept that their observations were not simply chance effects, then HD was being transmitted in a high proportion of cases either directly from patient to patient or indirectly via a single “carrier.” On this assumption many questions remain to be answered but some tentative conclusions can be drawn.

1. It would seem very likely that there is a transmissible HDA.

2. In this series, the time interval between the diagnosis of the infective patient and the diagnosis of the patient he infected varied from 0 to 16 years, with a median of 3 years. The interval from the period of contact between the infective patient and the susceptible case and the time of diagnosis of the susceptible case varied from 0 to about 17 years with a median of about 6 years (this is all based on the assumption that patients did not meet each other after leaving school). The latent period of HD has therefore a median of about 3 to 6 years; but it can be as brief as 1 to 2 years, and in most instances it is less than 10 years.

3. If the HDA is common then only a few people can infect adults, and these are mainly HD patients and their immediate contacts. Expression of the 1st infection in adults as HD must be low unless either most people are infected in childhood when it must be assumed that expression of the infection as HD is negligible (cf. Epstein-Barr virus and infectious mononucleosis), or infection of adults is uncommon.

4. If the HDA is rare and HD patients are no more effective transmitters of the agent than others who have been infected, then expression of the infection as HD must be high.

Without making assumptions about the relative effectiveness of HD patients and others as HDA transmitters, it is not possible to say whether the HDA is rare or not.

These conclusions will be capable of refinement when all the Albany patients have been investigated (only some 20% have been investigated to date). It will be possible, for example, to estimate the proportion of patients who were probably infected by another patient and this will enable us to consider transmission more quantitatively.

An important question that the Albany group must try to answer from their data is whether HD patients are infectious before and/or after diagnosis. This can be done crudely by constructing a table for each assumed case-to-case infection. If, as shown in Table 1, the period of contact between 2 cases is as in Line i then there is no information about whether Infection of B was before or after the diagnosis of A; if Line ii obtains, then A infected B or B infected A before either A or B were diagnosed; if Line iii obtains, then A infected B after A was diagnosed. Patient 31 and her university roommate in the work of Vianna et al. (9) are evidence for a patient’s being infective before diagnosis, but the published data allow this to be investigated in sufficient detail for very few suggested links.

If hard evidence does become available showing that a HD patient is infective after diagnosis, this will raise very difficult ethical problems in patient care; we need, therefore, to be especially circumspect when discussing this subject.

Table 1 also shows maximum and minimum latent periods and these can be used to estimate the distribution of latent periods (4).

The above discussion is based upon the assumption that the Albany “outbreak” was not simply a chance observation. However, at the present time this is still an open question. We have described elsewhere methods of establishing statistical significance levels of data such as that from the Albany study when a suitable group of matched controls have also been studied (6, 7). It is, however, also possible to derive simple approximate estimates of the number of spurious “links” which may be established in such studies. Consider a new HD patient aged 30 and assume that the latent period of his disease was 10 years or less. Then, if he has had “contact” with 500 persons during those last 10 years, the expected number of new HD cases in his contacts during those 10 years is 0.2 (based on an incidence rate of 4/100,000/year). That is, 20% of new HD patients aged 30 will have known another HD patient in the previous 10 years by chance alone. If 500 seems too high, then 250 gives an estimate of 10% and 125 an estimate of 5%.

1 Presented at the International Symposium on Human Tumors Associated with Herpesviruses, March 26 to 28, 1973, Bethesda, Md.
2 Presented by. Present address: Department of Community Medicine, University of Southern California School of Medicine, 2025 Zonal Avenue, Los Angeles, Calif. 90033.
3 Since this paper was presented, Vianna and Polan (10), in a study of HD in Nassau and Suffolk counties of New York State, have reported considerable clustering of cases of the disease in association with particular schools.
4 The abbreviations used are: HD, Hodgkin’s disease; HDA, Hodgkin’s disease agent; ORLS, Oxford Record Linkage Study.
SPACE-TIME DISTRIBUTION STUDIES

Alderson and Nayak (1) studied the distribution of times and places of onset of the 737 patients diagnosed in the Manchester conurbation in the years 1962 to 1968. They found no evidence of clustering using Knox's statistical test with a variety of critical distances up to 12 km and time differences up to 1 year. Persons in urban areas, however, tend to be very mobile, especially socially mobile, so that they have considerable contact with persons living far from their home address, and such studies are therefore perhaps more likely to be profitable in rural situations.

Kryscio et al. (3) studied the distribution of times and places of onset of all 1985 HD patients diagnosed in Connecticut in the years 1940 to 1969 and found no evidence for cases in the same towns to cluster in either 1- or 2-year periods.

We have studied, by an extension of Knox's method (5), the onset space-time distribution of the 206 HD patients aged under 40 years diagnosed in the Oxford Regional Hospital Board Area in the years 1962 to 1971. We found no evidence of Knox-type clustering but did find a suggestion of significant clustering in that 129 pairs of cases both lived within 5 km of each other and had their onsets between 270 and 539 days apart, whereas the expected value was 108 ($p \approx 0.05$). This is the sort of case clustering one would expect with a reasonably constant latent period of between 9 and 18 months. However, we had no a priori reason for choosing these times; moreover, as the result is only barely significant, it should be treated with extreme scepticism for the moment.

Studies of these types are, however, unlikely to be a useful way to continue to investigate HD in view of its probably long and variable latent period.

Reports, such as that of Klinger and Minton (2), of apparent clusters based on investigations prompted by some individual drawing attention to an anomalous situation are almost impossible to interpret (8). They observed, in Union County, Ohio (1970 population, 23,786), 12 cases of HD over the period 1960 to 1971, an average annual incidence of 4.3/100,000, “close to other published estimates.” However, 5 of these cases occurred in Darby Township (1970 population, 1212), giving a much increased average annual incidence of 34.4/100,000. The authors claim, on the basis (it seems) of a direct comparison of Darby Township to the remainder of the County, a statistical significance ($p$) of less than 0.0005 for this result. This must be considered a much exaggerated $p$ value for 2 reasons.

Firstly, Darby Township was not chosen for a priori reasons as likely to have a high HD rate and the statistical test should take this into account. When this is done we find a $p$ value of approximately 2%. Secondly, even this value of 2% for $p$ takes no account of the selection of Union County as the study area because of an initial report of 3 cases in Darby Township; if these cases were to be excluded then the statistical significance would, of course, completely disappear.

Furthermore, if we divide the population of the United States into units of the same size as Union County we obtain over 8000 units, in 2% of which (160 units) we would expect by chance alone to observe situations more extreme than that in Union County. Thus, further reports of “case-clustering” of this kind must be interpreted with great caution, unless the areas have been chosen “blind.” The medical literature contains many such reports of space-time clusters of leukemia cases, but very little evidence of such clustering has subsequently been found in properly designed studies.

Klinger and Minton linked the 5 Darby Township cases through mutual acquaintances. It may well be, however, that in a small community of 1200 people any 2 persons can be linked together relatively easily in this manner.

Caution is similarly called for in interpreting reports of “lymphoma families.”

CURRENT STUDIES IN OXFORD

At the present time there is an urgent need for studies to test the Albany findings. We are engaged in 2 such studies.

A Case-Control Study of Contacts

**Design.** This study is similar to the Albany one but includes a matched control group. The controls have been selected from patients not suffering from a chronic or malignant disease admitted to hospital in the same year as the HD patient.

The study is being conducted in 2 phases. Initially, a social worker is interviewing all the patients and controls (or the close relatives of patients and the close relatives of controls for those patients who are dead; the matched control for a dead patient will often be alive and in this case we are interviewing the control's relatives before the control himself in order to maintain some comparability). Information is collected on all of the places of residence of the subject (both cases and controls), persons he lived with in those places, and the time periods over which he lived there. All schools attended and all places of employment are recorded as well as any of the subject's social activities that might be linked together relatively easily in this manner.

Kryscio et al. (3) studied the distribution of times and places of onset of all 1985 HD patients diagnosed in Connecticut in the years 1940 to 1969 and found no evidence for cases in the same towns to cluster in either 1- or 2-year periods.
patients and controls have been interviewed. A list of the full names of all patients and controls will be shown to each subject and each will be asked if he has ever been in contact with or known anyone on the list. Any links established between 2 patients subsequent to both of their diagnoses will, of course, be excluded (each matched control will be assumed to have a date of diagnosis of his hypothetical disease the same as that of his patient). Patients with HD are likely to meet at treatment clinics and a possible bias, which may be difficult to overcome in this part of the study, will arise if patients remember having known each other previously (e.g., at school) only because they had met at the treatment clinic.

Selection of the Control Group. We are attempting to identify persons in close contact with the HD patients. The type and number of such contacts will be influenced by the age, sex, and socioeconomic status of the patient and any controls must, at least, be closely matched for these factors. Matching for area of residence is important, as, if this is not done, then mere spacial clustering of cases within the study area might lead to an excess number of “links” between patients because they lived close together. Of course, such spacial clustering might be indicative of the contagious nature of the disease but might also be brought about by a local environmental factor. Nearest “matched” neighbor controls might be the most appropriate but could lead to “overmatching” if neighbors tend to have the same contacts. In any case, existing population registers in England do not contain details of age and social class and such controls would be very difficult to select. However, such details are recorded for most patients admitted to hospitals, and we have therefore selected our controls from admissions of patients to hospitals in the same year as each HD patient was diagnosed. Persons admitted to hospitals may not be a representative sample of the population and may well differ in their social habits and contacts. However, it seems unlikely to us, at this stage, that this will be a serious effect and such persons should constitute a reasonable control group.

The Study Population. The Albany observations suggest that the transmission of HD is most likely to take place between young persons. All cases of the disease, in persons aged less than 40 years, registered with the Oxford Cancer Registry for the 10-year period 1962 to 1971 have therefore been identified. In this way 203 cases were located and only a further 3 cases were found by searching hospital records in the area.

Initially, we have concentrated on the subset of 97 cases who were resident at the time of diagnosis within the area covered by the ORLS. Of these patients, 34 are now known to be dead. The age and sex distribution of the 97 patients is shown in Table 2 together with the 1971 national census estimated populations in the area. The ORLS records hospital admissions of all persons resident in an area of about 50 miles by 30 miles to hospitals within the area. Information on each admission is stored on magnetic tape including each patient’s age, sex, occupation, and area of residence. This computer file provides a very convenient population from which to select a matched control group.

The population of the ORLS area is about 850,000. The area is predominantly rural but includes the towns of Oxford and Reading, each with a population of about 120,000. The area is divided into about 30 administrative subareas, and the matched control for each patient has been chosen to be resident within the same subarea.

Preliminary Findings. To date we have interviewed 48 patients (or their relatives) and 34 controls, including 19 matched pairs. It would be premature to discuss the results of these interviews in detail; however, we have been able to “link” together a number of persons as having attended the same school or having worked in the same place or having lived in the same vicinity. A number of these links have been such that the 2 persons must have been in the same place at the same time (although such time overlap is not strictly necessary if a “carrier” state is postulated), but, as yet, we have not been able to link patients together as was done in Albany. The links we have observed have been distributed as patient→patient, control→control, and patient→control roughly in proportion to the number of links which might be expected having interviewed this number of patients and controls. That is, at the present stage, we have no evidence that HD patients have had more contact with each other prior to diagnosis of the disease than have controls. These are very preliminary observations, and complete results of this initial study should be available shortly.

A Study of the “Transmission” of HD in Schools

Reported “epidemics” have thus far centered around high schools. Apart from the apparent evidence of transmission of the disease between students in Albany High School, Vianna et al. (9) have also recorded 1 instance, in a Downstate New York school, of a teacher with HD who, in a 3-year period, might have infected 4 of his students. The present evidence would thus suggest that the amount and type of contact that occurs between school students, and between school students and their teachers, is favorable for transmission.

If this evidence is sound, then: (a) the students of teachers who develop HD will show a higher incidence of HD than expected; and (b) the classmates of students who develop HD will show a higher incidence of HD than expected. These hypotheses may be tested indirectly by attempting to link teachers and young persons with HD to each other and by doing the same in a control group.

Study Population. Through the courtesy of the Office of Population Censuses and Surveys, we have been able to identify, for all of England and Wales, persons aged less than 65 years, whose occupation was recorded as teacher, and all persons aged less than 25 years, who were certified as dying from HD or another cancer in the period 1959 to 1971 (occupation was not coded for computer storage for the years 1964 to 1969 and thus information on teachers is missing for this period). In addition, all similar persons registered with a cancer registry in the period 1966 to 1969 have been identified.

We are ascertaining, initially through a postal questionnaire, all of the schools in which each teacher had worked, the dates when he moved from one school to another, and
the schools attended by each HD case aged less than 25 years. With this information “links” may be established between HD cases in the school environment.

**Controls.** We are selecting matched controls from patients with a cancer other than HD, other reticuloses or leukemia. This is not an ideal control group, inasmuch as the possibility that other cancers may be contagious cannot be excluded, but we are initially working on the assumption that this is not so and controls will be selected from the information supplied by the Office of Population Censuses and Surveys. Each HD patient is matched with another cancer patient for year of death or registration, age, sex, social class (teachers will be matched with teachers), and area of residence (area matching will not initially be attempted below the level of cancer registry area, as closer matching might lead to a tendency to select a control from the same school as the HD case).

**Analysis.** Evidence for person-to-person transmission of HD will be given by links in Table 3, Column 1, and suitable control links are given in Table 3, Column 2.

The most direct evidence of teacher-to-student transmission of the disease is provided by the number of links of type 1 compared with the number of type 2. The difference between the number of links of type 3 and type 4 would be expected to be zero, significant departures from this would throw doubt on the validity of the comparison of 1 with 2.

Evidence of student-to-student transmission is provided by comparing type 5 links with links of types 6 or 7. The expected number of type 6 links should very nearly equal one-half the number of type 7 links.

If any of these tests are positive, the links will be investigated in detail.

**Expected Numbers of Student-Teacher Links.** In order to calculate the approximate expected number of cases of HD among the students of HD teachers, assuming no transmission of the disease, it is necessary to make a number of sweeping assumptions. If we assume (1) that all the teachers identified have been teaching since 1939 and (2) that an “average” teacher teaches 250 new students each year with the student age distribution of 100 children ages 5 to 9, 100 children ages 10 to 14, and 50 children ages 15 to 19, then we can calculate the man years at risk of the student popu-

---

**Table 2**

Cases of HD aged under 40 years diagnosed in the period 1962 to 1971 in the ORLS area tabulated by age and sex with estimates of population from the 1971 census

<table>
<thead>
<tr>
<th>Age group</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>37,830</td>
<td>35,720</td>
</tr>
<tr>
<td>5-9</td>
<td>39,170</td>
<td>36,757</td>
</tr>
<tr>
<td>10-14</td>
<td>34,310</td>
<td>36,757</td>
</tr>
<tr>
<td>15-19</td>
<td>32,490</td>
<td>36,757</td>
</tr>
<tr>
<td>20-24</td>
<td>38,995</td>
<td>36,757</td>
</tr>
<tr>
<td>25-29</td>
<td>31,525</td>
<td>36,757</td>
</tr>
<tr>
<td>30-34</td>
<td>28,375</td>
<td>29,220</td>
</tr>
<tr>
<td>35-39</td>
<td>26,525</td>
<td>29,220</td>
</tr>
</tbody>
</table>

Cases/100,000/yr:

<table>
<thead>
<tr>
<th>Age group</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>0.26</td>
<td>0.63</td>
</tr>
<tr>
<td>5-9</td>
<td>2.15</td>
<td>3.50</td>
</tr>
<tr>
<td>10-14</td>
<td>1.17</td>
<td>2.03</td>
</tr>
<tr>
<td>15-19</td>
<td>5.90</td>
<td>5.90</td>
</tr>
<tr>
<td>20-24</td>
<td>5.08</td>
<td>5.08</td>
</tr>
<tr>
<td>25-29</td>
<td>2.47</td>
<td>2.47</td>
</tr>
<tr>
<td>30-34</td>
<td>3.77</td>
<td>3.77</td>
</tr>
<tr>
<td>35-39</td>
<td>2.53</td>
<td>1.16</td>
</tr>
</tbody>
</table>

---

**Table 3**

Evidence of person-to-person transmission of HD between teachers and students

<table>
<thead>
<tr>
<th>Evidence for person-to-person transmission</th>
<th>Control links</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. HD teacher-&gt;HD student</td>
<td>2. HD teacher-&gt;C student</td>
</tr>
<tr>
<td>5. HD student-&gt;HD teacher</td>
<td>6. C student-&gt;C student</td>
</tr>
<tr>
<td>9. C teacher-&gt;C teacher</td>
<td>10. HD teacher-&gt;C teacher</td>
</tr>
</tbody>
</table>

*C, control.

---

**References**

6. Pike, M. C., and Smith, P. G. A Case-Control Approach to Examine
P. G. Smith and M. C. Pike


Case Clustering in Hodgkin's Disease: A Brief Review of the Present Position and Report of Current Work in Oxford

P. G. Smith and M. C. Pike

Cancer Res 1974;34:1156-1160.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/34/5/1156.citation

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.