CLUSTERING OF CASES OF HODGKIN'S DISEASE AND LEUKEMIA

M. C. PIKE, PHD,* AND P. G. SMITH, BSC,†

Time-space "clusters" of leukemia cases have been reported since the 1930s; since 1960, a great number of such observations has been made. However, it is still an open question whether or not the phenomenon is real.[5] "Clustering" of Hodgkin's disease cases is a relatively new phenomenon.[6,17,18] The reports appear more convincing than those relating to leukemia, but closer analysis reveals possible serious methodological flaws which make their interpretation difficult. In this paper we consider the currently more interesting Hodgkin's disease situation, using the leukemia studies as a guide as to what pitfalls should be avoided. We then reverse the process and suggest cluster studies that should now be conducted in leukemia on the basis of what may be learned from the Hodgkin's disease studies.

Cancer 34:1390–1394, 1974.

PRIOR TO 1971, EPIDEMIOLOGIC EVIDENCE pointing to Hodgkin's disease being infectious was very limited. The literature contained numerous reports of cases occurring in the same family, mostly sib-sib pairs, but these reports are virtually impossible to interpret. The best detailed study was conducted by Razis and his colleagues.[11] They reviewed case notes at Memorial Hospital, NY, over the period 1918–1958, for evidence of Hodgkin's disease in the immediate family members of cases. They took as "controls" cases of leukemia, lymphosarcoma, other malignant tumors, and benign conditions, and also looked for Hodgkin's disease cases among their immediate family members. There was about a three-fold excess of Hodgkin's disease cases in the families of Hodgkin's disease patients compared with the controls. Further analysis of the familial cases showed that, in sib-sib pairs, the dates of onset tended to be at the same time rather than at the same age. The nature of the data is biased towards these effects, but the evidence, such as it was, seemed to favor a small environmental risk rather than a genetic factor.

ALBANY STUDY

Current epidemiologic interest in Hodgkin's disease is founded on the Albany "outbreak."[16,17]

A medical colleague drew the attention of Vianna and his colleagues to a peculiar situation in an Albany high school in which a number of cases of Hodgkin's disease had occurred in school friends. This situation was investigated using classical infectious disease epidemiologic methods, but a long and variable latent period was postulated and a possible carrier state assumed. Working outwards from the initially identified cases they endeavored to "link" further patients either directly or indirectly through a single contact. In this way 31 patients diagnosed in 1950–1970 were linked together: there were 9 instances of case-case "spread," and 25 instances of case-contact-case "spread."

On the face of it a remarkable observation; but could one's intuitive feelings of what is remarkable be very wrong?

These were 31 patients out of the 208 diagnosed in Albany County during the study period. Perhaps if 208 "normal" people were randomly selected from the County it

Presented at The American Cancer Society–National Cancer Institute National Conference on Virology and Immunology in Human Cancer, November 29–December 1, 1973.

Supported by Contract No. PH43-NCI-68-1030 within the Virus Cancer Program of the National Cancer Institute, National Institutes of Health, U.S. Public Health Service.

* Departments of Community Medicine and Pediatrics, University of Southern California School of Medicine, Los Angeles, CA.

† DHSS Cancer Epidemiology and Clinical Trials Unit, Oxford University, Oxford, England.

Address for reprints: M. C. Pike, University of Southern California School of Medicine, 2025 Zonal Ave., Los Angeles, CA 90033.

Received for publication May 28, 1974.

might be possible to link a surprising number in this way.

Take a new Hodgkin's case and consider his contacts during the 10 years before his diagnosis. If he had contact with 500 people in that time, then, assuming an incidence rate of 4/100,000/year, the expected number of Hodgkin's disease cases in that 500 in the 10-year period is 0.2, i.e. he has a 20% probability of having had contact with another Hodgkin's case in the previous 10 years by chance alone. Five hundred contacts may be too high, but this only takes account of direct contact; the percentage will increase considerably if we consider the case-contact-case mode of transmission.

The authors studied two control groups, but as they only chose controls for the linked patients rather than for all 208 cases, it is perhaps not surprising that they were unable to link the controls to the same extent as the cases.

Thus, although this study is truly a landmark in cancer epidemiology, it is in our opinion impossible to assess the significance of the findings.

The relevant lesson to be learned from the leukemia cluster literature is that further reports along these same lines with no possibility of calculating a statistical significance level are of very limited value. Only a first report, such as that of Vianna and his colleagues, is permissible without a statistical test. We must now ask how such a study can be conducted so that a definite answer is obtained. Before going into this further we need to look at certain other studies.

Pseudo-Statistical Studies

Reports, such as that of Klinger and Minton,[6] of apparent small clusters based on investigations prompted by some individual drawing attention to an "anomolous" situation may conveniently be termed pseudo-statistical studies. Klinger and Minton observed in Union County, Ohio (1970 population: 23,786) 12 cases of Hodgkin's disease over the period 1960–71, 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 level (p) of less than 0.05% for this result. This is not correct.

Firstly, Darby Township was not chosen for a priori reasons as likely to have a high Hodgkin's disease rate and the statistical test should take this into account. When this is done we find a p value of approximately 2%.[13] Secondly, even this p value of 2% takes no account of the selection of Union County as the study area because of an initial report of three 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, i.e. we could have a report in the literature every week for the next 3 years and it might still represent no more than a chance phenomenon.

This is, of course, an exaggeration, but it does emphasize the importance of doing controlled studies in which the area to be studied has been chosen "blind," and the great difficulty in interpreting anecdotal reports.

The medical literature contains vast numbers of such reports of "clusters" of leukemia cases; these have contributed heat but no light, they provide confusion rather than information.[5]

Distribution of Onset of Hodgkin's Disease in Space-Time

Two proper studies have been reported of the space-time distribution of place and date of onset of Hodgkin's disease cases; both gave negative results. Alderson and Nayak[1] studied the 737 patients diagnosed in the Manchester (U.K.) conurbation in 1962–1968. They found no evidence of clustering, using Knox's statistical test,[7] with a variety of critical distances up to 12 kilometers and time differences up to a year. Kryscio and colleagues[8] studied the 1985 Hodgkin's disease patients diagnosed in Connecticut in the years 1940–1969; they found no evidence for cases in the same towns to cluster in either 1- or 2-year periods.

In the Manchester study, the Knox test took the date of interest as the date of diag-

nosis. This implies that one is assuming that both the latent period of the disease and the time interval between clinical onset and diagnosis are short and relatively constant. As Alderson and Nayak point out[1] these assumptions are unlikely to be true.

The problem of a long latent period may be overcome by using a generalization of Knox's method.[9] For each patient we postulate a period of susceptibility and a period of infectivity (e.g. susceptible 2–3 years before onset). We can then determine which patients could have infected other patients in time. We may do the same in space, defining the patients effective movements when he was susceptible and infective to determine which persons could have infected others in space. We then count the number of patients who were in the right place at the right time to have caught the disease from another patient: the expected number and variance of the number of such links may be calculated using similar arguments to the Knox test and the statistical significance can thus be found. This overcomes the problem of a long latent period, but it does not help us solve the problem of a variable latent period unless we have a priori methods of assigning a latent period to each patient.

Also, studies using fixed addresses are probably best conducted in rural areas, rather than in conurbations whose residents tend to be very mobile so that they have considerable contact with persons living far from their home address.

The Connecticut study[8] may also be criticized on the grounds that it considered all Hodgkin's disease cases without distinction. The investigators should at least have looked at subdivisions of the cases based on age. There is considerable evidence suggesting that the disease may consist of more than one entity; the Albany findings point to young cases as the "infectious" ones.

The leukemia clustering literature has, unfortunately, many examples of these types of "errors."

HODGKIN'S DISEASE IN SCHOOLS

Recently, Vianna and Polan[18] presented an epidemiologic analysis of young Hodgkin's disease cases in Nassau and Suffolk Counties of New York State. They concluded that the disease clustered in certain schools. Some aspects of their findings are however, open to criticism on the basis of possible severe bias due to nonascertainment of cases.

The authors clearly believe that they have identified nearly all of the cases of Hodgkin's disease in the study area, but there is evidence that this may not be so. They found 465 cases in the 11 years 1960–1970, which gives an average annual crude incidence rate of 1.9/100,000. This compares to an expected rate of between 3 and 4/100,000.[2,3] Thus, the rate in Nassau and Suffolk appears to be about one-third lower than that expected. The consequences of failing to ascertain a proportion of the cases in epidemiologic studies may not be important if the loss is "random," but bias, whether known or unknown, in case ascertainment can invalidate the results. If, for example, cases in certain areas or schools were missed more often than in others, then the authors' "two-time period" approach must be open to question. Such an effect would arise if physicians in certain areas were more likely to refer cases outside of the counties for treatment. We can only speculate on such biases, but the apparent low incidence of Hodgkin's disease in the study period is disturbing.

The discrepancy appears more pronounced if we consider the most relevant very young cases. The authors do not give an age breakdown of the cases diagnosed in the study area, but it is possible to identify at most 26 patients (24 from their Table 1 and 2 from the text) aged 10–19 years diagnosed between 1960 and 1970. The relevant average population in this age group for the two counties is 427,866, giving an annual rate of 0.55/100,000. The annual rate in the same age group in the whole of New York State (excluding N.Y. city) in 1959–1961 was three times this.[3] In contrast to this, the authors quote annual rates of 5.3/100,000 in schools with an enrollment of less than 1500 and 23.8/100,000 in larger schools. Our calculation of a rate of 0.55 in the 10–19 years age group assumes that no cases are diagnosed in persons currently resident in Nassau or Suffolk who schooled outside of the area, which is probably not true. However, to reconcile this rate even with that of 5.3 for small schools, it is necessary to assume that there are about nine times as many persons aged 10–19 years who live in the area, but attended school outside of the area, as persons who both live and attended school in the area. This seems unlikely.

FUTURE STUDIES OF HODGKIN'S DISEASE

To test the Albany findings properly a case-control approach is needed.[10,12]

First, identify in some area all Hodgkin's cases (perhaps with some age restriction) over a period of years, preferably 10 or more (giving a total of n cases). Then for each patient choose a matched control: the factors for which matching is probably necessary are age, sex, socioeconomic status, and possibly, depending on how stringent a test of contagion is envisaged, an area of residence smaller than the total study area.[14] By interviewing all these 2n persons (or, if dead, their relatives or friends—note that the persons interviewed should be the same for both case and control) we may establish whether any two given people had "effective" contact, that is, contact of such a nature at the "right" time, that if they had both been Hodgkin's disease cases then they would have been taken as a "contagion pair" or "link."

The total number of such links between patients can then be compared with the number of links observed between the controls, and between the cases and the controls, and a statistical significance level can be calculated.[10]

A study of this type is nearly complete in Oxford, England.[14] We identified all 97 patients aged under 40 diagnosed in 1962–1971 whose address at clinical onset was in the Oxford Record Linkage area. We restricted ourselves to patients under 40, as the Albany observations suggest that the "transmission" of Hodgkin's disease is most likely to take place between young persons, and we wished specifically to test their findings. For each patient we have chosen a matched control admitted to a hospital in the area with a nonmalignant condition in the same year as the patient was diagnosed. Each person is then being interviewed to build up a brief life history. We record at the initial interview residences, schools, workplaces, outside activities bringing the person into regular contact with others, and names of close friends, all with dates. When all the Hodgkin's disease patients and the controls have been interviewed we will present a list in random order of all patients and controls to each patient and control and ask them if they know anyone on the list.

A problem with the list (and with names of "close" friends) is that Hodgkin's disease patients will tend to meet each other at clinics, and even though we can exclude this contact the problem remains of a patient remembering he knew someone at school, say, because he had talked to him at the Hodgkin's disease clinic. This bias will be difficult to overcome or allow for; positive clustering will have to be intense enough to be detectable without these answers if it is to be convincing.

Using the results of the first interview we may look, for example, for persons who attended the same school at the same time. This will give us both direct and indirect contacts, as even if we cannot link two people directly it is likely that they could be linked with a common third party, such as a teacher.

To date we have followed up 157 persons and have done a preliminary analysis of the data. When we allow for having data on 91 patients and only 66 controls the observed number of links between Hodgkin's patients (of which there are many) are no different from the number "expected" based on the number of links seen between the controls.

Further studies along these lines and along the lines of the Nassau and Suffolk school study are urgently needed. The leukemia clusters were never taken as seriously as the Hodgkin's clusters—a number of Hodgkin's patients have, we believe, even killed themselves as they did not want to "infect" their families and friends. This issue must be settled quickly and, if "infecton" does occur, the infective period in the disease process must be pinpointed.[14] Clearly, if the disease is infectious only before it is clinically apparent, it is a very different matter from its being infectious after diagnosis. It is the latter which will call for the most difficult decisions to be made.

FUTURE STUDIES OF LEUKEMIA

The work of Vianna and his colleagues has shown us that to seriously study the transmission of diseases with long and variable latent periods, detailed questioning of patients must be conducted. Analyses based solely on dates and places of onset must now be considered outmoded.

The subject of leukemia clusters should be reopened with this more elaborate approach in mind. The studies should if possible be done in rural areas and with knowledge of the epidemiologic facts concerning the different

forms of the disease. For example, if one is studying childhood acute lymphatic leukemia, one needs to take account of the evidence pointing towards an in utero induction of the disease,[4,15] and aim some of one's questions to events relating to this time period.

REFERENCES

1. Alderson, M. R., and Nayak, R.: A study of space-time clustering in Hodgkin's disease in the Manchester Region. *Br. J. Prev. Soc. Med.* 25:168–173, 1971.

2. Biometry Branch, National Cancer Institute, National Institutes of Health: U.S. Third National Cancer Survey, Incidence. 1969.

3. Doll, R., Payne, P., and Waterhouse, J.: Cancer incidence in five continents. Berlin, Springer, 1966.

4. Fedrick, J., and Alberman, E.: Reported influenza in pregnancy and subsequent cancer in the child. *Br. Med. J.* 2:485–488, 1972.

5. Glass, A. G., Hill, J. A., and Miller, R. W.: Significance of leukaemia clusters. *J. Pediatr.* 73:101–107, 1968.

6. Klinger, R. J., and Minton, J. P.: Case clustering of Hodgkin's disease in a small rural community, with associations among cases. *Lancet* i:168–171, 1973.

7. Knox, G.: Epidemiology of childhood leukaemia in Northumberland and Durham. *Br. J. Prev. Soc. Med.* 18:17–24, 1964.

8. Kryscio, R. J., Myers, M. H., Prusiner, S. T., Heise, H. W., and Christine, B. W.: The space-time distribution of Hodgkin's disease in Connecticut, 1940–69. *J. Natl. Cancer Inst.* 50:1107–1110, 1973.

9. Pike, M. C., and Smith, P. G.: Disease clustering—A generalisation of Knox's approach to the detection of space-time interactions. *Biometrics* 24:541–556, 1968.

10. Pike, M. C., and Smith, P. G.: Two methods of examining space-time clustering in diseases with long latent periods. *Biometrics* 30:263–279, 1974.

11. Razis, D. V., Diamond, H. D., and Craver, L. F.: Familial Hodgkin's disease—Its significance and implications. *Ann. Intern. Med.* 5:933–971, 1959.

12. Smith, P. G., and Pike, M. C.: Space-time clustering—A case-control method of examining diseases with long latent periods. *In* Host Environment Interactions in the Aetiology of Cancer in Man, R. Doll and I. Vodopija, Eds. I.A.R.C. Scientific Publication no. 7, Lyons, 1973.

13. Smith, P. G., Pike, M. C., and Kinlen, L. J.: Clustering in Hodgkin's disease. *Lancet* i:433–434, 1973.

14. Smith, P. G., and Pike, M. C.: Case clustering in Hodgkin's disease—A brief review of the present position and report of current work in Oxford. *Cancer Res.* 34:1156–1160, 1974.

15. Stewart, A., and Barber, R.: The epidemiological importance of childhood cancers. *Br. Med. Bull.* 27: 64–70, 1971.

16. Vianna, N. J., Greenwald, P., and Davies, J. N. P.: Extended epidemic of Hodgkin's disease in high school students. *Lancet* i:1209–1211, 1971.

17. Vianna, N. J., Greenwald, P., Brady, J., Polan, A. K., Dwork, A., Mauro, J., and Davies, J. N. P.: Hodgkin's disease—Cases with features of a community outbreak. *Ann. Intern. Med.* 77:169–180, 1972.

18. Vianna, N. J., and Polan, A. K.: Epidemiological evidence for transmission of Hodgkin's disease. *N. Engl. J. Med.* 289:499–502, 1973.