Invited Commentary

Invited Commentary: Do Clusters of Leukemia and Lymphoma Provide Evidence for an Infectious Cause?

Graham R. Law

From the Epidemiology and Genetics Unit, Department of Health Sciences, University of York, York, United Kingdom.

Received for publication May 17, 2005; accepted for publication July 18, 2005.

In this issue of the Journal, Heath (1) presents a detailed account of investigations into eight clusters of childhood leukemia and lymphoma carried out by the US Centers for Disease Control and Prevention (CDC) during the 1960s and 1970s. The clusters, chosen for characteristics suggestive of an infectious etiology, were selected from 50 such investigations undertaken during this time period. Heath concludes that indicators of interpersonal contact suggest that infectious disease underlies the etiology of childhood leukemia and lymphoma. Do clusters really provide such evidence?

A cluster is an excess incidence of related health events occurring at the same place, the same time, or (more usually) both. The difficulties involved in identifying clusters of chronic disease have been discussed extensively (e.g., see Rothman (2)). Heath states that the CDC considered post-hoc investigation of a cluster with formal statistical evaluation irrelevant, which is in line with the position of many researchers (e.g., see Alexander (3)). Indeed, one might argue that, were statistical testing required to prove the existence of a cluster, that cluster would be less than useful for causal inference. There are more objective arguments that tests of statistical significance are not appropriate. Principal among these are the definitions of closeness between cases in space, time, and diagnosis employed. These assumptions motivated Grufferman (4) to compare cluster identification to the method of the “Texas sharp-shooter,” who shoots at a wall and then draws a target around the bullet holes: Spatial, temporal, and diagnostic boundaries may be engineered to provide the biggest possible cluster. Heath did not make clear how the CDC investigations of these clusters drew their boundaries, but at least some of the clusters were brought to the CDC’s attention through cases living in very close proximity.

Aldrich and Sinks (5) suggested a hierarchical classification system for clusters, with the top and most frequent level being a “perceived” cluster. This describes the initial report of an aggregation of cases, no matter the source. Historically, most perceived clusters have not been identified through routine surveillance of disease in a community, but rather through public and clinician anxiety, media reports, and, more recently, the promise of litigation. In the United Kingdom, the Small Area Health Statistics Unit was established in 1987 (6) to provide statistical analysis of any perceived cluster to public health departments. Following some screening process, such as that provided by the Small Area Health Statistics Unit, to remove truly spurious aggregations, a subset of clusters may be deemed “observed” and therefore worthy of further investigation. The CDC investigated 50 such clusters during this time period. A subset of observed clusters may be classified as “etiologic” clusters, where the aggregation of cases is suggestive of a causal factor. Heath argues that the eight clusters described in his paper (1) may provide causal evidence.

Rothman, a critic of cluster investigation (2), pointed out that clusters have been useful in identifying previously unknown disease entities and their cause, such as acquired immunodeficiency syndrome and human immunodeficiency virus. However, Rothman also commented that these were rare situations or new diseases, and formal investigation of such striking clusters was unnecessary. One way to understand cluster investigation is to recognize that we are trying to establish whether a causal component clusters, which places a further restriction on the utility of a cluster to provide etiologic insight. McNally and Eden (7) have suggested that clusters and clustering are, in themselves, evidence for the involvement of infection in the leukemia disease process. This is difficult to assess, but clustering due to infectious disease will only be apparent when the induction period is short or the infection is rare or affects only a small number of susceptible persons (7).
The interest in these eight clusters lies not in their existence per se but in the community characteristics that were also aggregated. The inference that infectious disease might have been involved in these eight clusters is based on unusual community characteristics, examination of community patterns of illness, and population change. The clusters that were linked with patterns of membership in churches and attendance at certain schools may have become apparent precisely because of the membership of families in these organizations; a perceived cluster may have been identified where otherwise it would not have been. Such a mechanism would prejudice investigation of these factors because of their role in the original identification of the cluster. With no comparison group available, we are not able to assess the likelihood that these patterns were unusual. Furthermore, the true link between these community characteristics and infectious disease is based on supposition; this would benefit from assessment of the relations of infectious disease load, biologic effective dose, and immunity to community characteristics such as population increase.

There is growing evidence that, for acute lymphoblastic leukemia, at least two events are required to cause the disease (8). The initiating event may occur before or shortly after birth, but there has been little progress in identifying the causal component. The CDC cluster investigations did not make clear distinctions between time windows of exposure, an avenue that should be borne in mind for future investigations. In fact, clusters tend to be identified at the time of diagnosis, whereas in-utero events might be expected to aggregate at the time of birth.

Three hypotheses have proposed that infection is involved in the disease process of all or some subtypes of childhood leukemia and lymphoma (9–11). None of these hypotheses may be directly tested through cluster investigation. At best, we should consider the investigation of clusters an opportunity for far-reaching and intensive exploration of risk factors. A word of caution: the absence of evidence from a cluster investigation should not be used as evidence for an absence of that factor. Indeed, there were 42 clusters investigated by the CDC during this period in which comparisons of apparent disease clusters around a point source: the UK Small Area Health Statistics Unit. J Public Health Med 1999; 21:289–98.


Inevitably, epidemiologists will continue to investigate clusters. As Martin Gardner put it, “We should not hold our hopes too high but in the case of childhood leukaemia, a situation of relative ignorance about causes but of widely and publicly recognised importance as a disease to be understood, any leads will be worthwhile” (12, p. 130).

REFERENCES