**Understanding Cancer Clusters**

*Michael J. Thun, MD, MS; Thomas Sinks, PhD*

**ABSTRACT** Each year, state and local health departments respond to more than 1,000 inquiries about suspected cancer clusters. Three quarters of these reports involve situations that are clearly not clusters and can be resolved by telephone. For the remainder, follow-up is needed, first to confirm the number of persons affected, their age, type of cancer, dates of diagnosis, and other factors, and then to compare cancer incidence in the affected population with background rates in state tumor registries. In approximately 5% to 15% of the reported situations, formal statistical testing confirms that the number of observed cases exceeds the number expected in a specific area, given the age, sex, and size of the affected population. Even in these instances, however, chance remains a plausible explanation for many clusters, and further epidemiologic investigation almost never identifies the underlying cause of disease with confidence. The few exceptions have involved clusters of extremely rare cancers occurring in well-defined occupational or medical settings, generally involving intense and sustained exposure to an unusual chemical, occupation, infection, or drug. This article discusses the resources and scientific tools currently available to investigate cancer clusters. It also provides a framework for understanding cancer clusters and a realistic appraisal of what cluster investigations can and cannot provide in the context of community expectations. (CA Cancer J Clin 2004;54:273–280.) © American Cancer Society, Inc., 2004.

**INTRODUCTION**

The public image of cancer clusters, popularized by Hollywood movies such as *A Civil Action* and *Erin Brockovich*, is that any collection of people diagnosed with cancer may represent a mini-epidemic caused by local environmental contamination.¹ Toxic exposures are presumed to be a major cause of human cancers. Any apparent clustering of cancer cases in a geographic area, time period, and/or defined group of people raises the specter that a localized source of pollution may be causing the problem. Public concern focuses primarily on toxic exposures, even when the perceived cluster involves a school, suburban neighborhood, or office building where the likelihood of such exposures appears no different from that in many other unaffected settings.

Epidemiologists and public health workers who investigate suspected cancer clusters are more skeptical of the scientific value of cluster investigations than is the general public. They recognize the historical examples in which clustering of rare types of cancer among highly exposed occupational and medical populations has led to the recognition of human carcinogens.²,³ However, they distinguish between these situations, where the exposure is high, prolonged, and well defined, and community settings in which exposures are low and poorly defined, where the cases may involve a mix of unrelated, relatively common types of cancer, and the scientific tools available to investigate these situations rarely identify an underlying cause with confidence. More than 1,000 suspected cancer clusters are reported to state health departments each year.⁴–⁶ About three quarters of these are clearly not clusters and can be resolved by telephone if health officials respond promptly and with sensitivity to the requester using clearly defined criteria to evaluate and triage the reports.⁷ In approximately 5% to 15% of the reported situations, formal statistical testing confirms that the number of observed cases exceeds the number expected in the affected population, given the age, sex, number of people at risk, and the time period of observation.⁷ However, even in these settings, epidemiologic studies are rarely definitive, and chance remains a plausible explanation for the clustering.

The goal of this article is to provide a framework for understanding and responding to cancer clusters so that affected communities can realistically anticipate what investigations can and cannot provide. We describe the criteria...
that define a cancer cluster, selected historical examples of clusters that contributed to the discovery of previously unrecognized human carcinogens, the steps involved in investigating a suspected cancer cluster, and considerations that may complicate or impede such investigations in community settings.

**WHAT IS A CANCER CLUSTER?**

The term cancer cluster usually implies that more cases of cancer (usually of the same type) are identified within a certain group of people, geographic area, and time period than are expected, based on the size and age of the population. Usually the term refers to a highly localized situation such as a school, neighborhood, or workplace, although it is sometimes used to refer to a broader geographic area or larger subgroup of the population. Concern about disease clustering is not exclusive to cancer. Similar concerns apply to birth defects, neurologic diseases, and other conditions for which the etiology is obscure. Some suspected cancer clusters involve a combination of cancers with other diseases possibly related to pollution.

Epidemiologists and public health workers who respond to concerns about clusters distinguish between perceived clusters, those that have been noticed and reported but not yet formally evaluated, and confirmed clusters, in which the case diagnoses and their connection to the community have been documented, and statistical testing indicates a very low probability that the observed clustering could occur by chance. The number of perceived cancer clusters reported to public health agencies is much larger than commonly appreciated. Although records are not collected routinely nationwide, 41 state health departments recorded approximately 1,900 inquiries about cancer clusters in 1996.6 Other surveys provide lower estimates, ranging from 1,300 to 1,650 reports in 1989 to 1,100 in 1997.4 Records from the Missouri Department of Health document 101 inquiries about cancer clusters received between 1984 and 1988.9 A similar number of reports were recorded by the Health Departments in Wisconsin and Minnesota during other five-year time periods.7,10 A search of US newspaper articles containing the words “cancer cluster” identified 2,006 reports filed from January 5, 1990 to January 5, 2000.8

In practice, only a small fraction of suspected cancer clusters meet statistical criteria of a confirmed cluster, in which chance is unlikely to explain the excess of observed cases over the expected amount. Of the 101 potential cancer clusters evaluated formally by the Missouri Department of Health between 1984 and 1988, only 17 had a statistically significant excess number of observed compared with expected cases.9 Only 5% of perceived clusters evaluated by the Minnesota Department of Health were statistically significant.7 In many cases, perceived clusters include different types of cancers, benign or metastatic tumors, cases that had little connection with the community, or cases that occurred over a longer time period than appreciated. Even when an investigation documents that a given clustering is “statistically significant” (meaning that there is less than a 5% chance that the observed number of cases could have occurred by chance), this does not rule out chance, given the potential for random aggregation in a country the size of the United States. The interpretation of statistical significance in the context of disease clustering is discussed further below.

**HISTORICALLY INFORMATIVE CANCER CLUSTERS**

There are well-known instances in which the investigation of an unusual cancer cluster has led to the identification of a previously unrecognized human carcinogen. All of the examples listed in Table 1 involved clusters of a rare type of cancer in people with prolonged, high-intensity exposure to industrial or medical carcinogens.2 Each was recognized as extraordinary by an alert clinician and reported to public health and medical officials for evaluation. Although such examples are rare, even in occupational settings, they illustrate how some cancer clusters can provide new scientific information about the causes and prevention of cancers. One of the earliest reports of a cancer...
cluster involved scrotal cancer among London chimney sweeps in the 18th century.11 Young boys employed in this occupation were exposed to soot from coal while crawling through the narrow chimneys and from their unlaundered clothing. Another tragic example, early in the 20th century, was a cluster of women diagnosed with osteosarcoma of the jaw while employed as watch dial painters in New Jersey and Connecticut.12 These women were exposed to ionizing radiation from radium present in the luminous paint when they used their lips to form a sharp tip on the paintbrush. Other clusters involved pleural mesothelioma among asbestos workers in London13 and angiosarcoma of the liver among chemical workers exposed to vinyl chloride monomer.14 In each case, the occupational exposure was high and prolonged. The exposures that result from medical treatment or chronic infections are also considerably higher than those involving exposure to pollutants in the general community. The recognition of a cluster of adenocarcinoma of the vagina in young women whose mothers had been treated with diethylstilbestrol (DES) led to the identification of DES as a transplacental carcinogen.15 A cluster of Kaposi sarcoma and Pneumocystis carinii pneumonia in healthy gay men contributed to the discovery of the acquired immune deficiency syndrome epidemic and the human immunodeficiency virus.16

These examples are much less common than more recent investigations that have not identified any specific cause of the apparent clustering. There have been numerous investigations near high-tension power lines, nuclear facilities, hazardous waste dumps, neighborhoods, schools, and office buildings that have not provided new scientific information about the causes or prevention of cancer, nor have they convincingly identified a reason for apparent clustering.

**HOW ARE SUSPECTED CANCER CLUSTERS INVESTIGATED?**

Public health officials from state and local health departments usually take primary responsibility for responding to perceived clusters. Most states have developed a stepwise approach to triage requests from the public, using established criteria to determine their response.7,17 Some states regularly analyze cancer registry data to identify communities with more cancers than expected. Others do not investigate reported clusters but rather limit their activity to cancer education.

Federal agencies that provide assistance to states in investigating certain clusters include the Centers for Disease Control and Prevention (CDC) (http://www.cdc.gov/nceh/clusters/), the National Cancer Institute (NCI) (http://seer.cancer.gov), and the Environmental Protection Agency (http://www.EPA.gov). The CDC has proposed Guidelines for Investigating Clusters of Health Events (http://www.cdc.gov/mmwr/preview/mmwrhtml/00001797.htm). The National Institute for Occupational Safety (NIOSH) is the lead federal agency within the CDC for investigating occupational cancer clusters. The National Center for Environmental Health and the Agency for Toxic Substances and Disease Registry (http://atsdr1.atsdr.cdc.gov) are other agencies within the CDC that may consult with health departments and are sometimes asked to

### TABLE 1  Examples of Cancer Clusters Leading to Identification of Human Carcinogens

<table>
<thead>
<tr>
<th>Classification</th>
<th>Year</th>
<th>Cancer Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupational</td>
<td>1775</td>
<td>Scrotal cancer in chimney sweeps exposed to soot from coal.</td>
</tr>
<tr>
<td></td>
<td>1929</td>
<td>Osteosarcoma in watch dial painters exposed to radium.</td>
</tr>
<tr>
<td></td>
<td>1965</td>
<td>Mesothelioma and lung cancer in asbestos workers.</td>
</tr>
<tr>
<td></td>
<td>1974</td>
<td>Angiosarcoma of the liver among chemical workers exposed to vinyl chloride monomer.</td>
</tr>
<tr>
<td>Medical</td>
<td>1971</td>
<td>Vaginal clear cell carcinoma in daughters exposed in utero to diethylstilbestrol.</td>
</tr>
<tr>
<td>Other</td>
<td>1981</td>
<td>Kaposi sarcoma in homosexual men with AIDS exposed to human herpes virus B.</td>
</tr>
</tbody>
</table>
conduct field and laboratory studies of community clusters. Both the National Center for Chronic Disease Prevention and Health Promotion of the CDC and the NCI support population-based cancer registries that monitor the background incidence rate of cancers, against which suspected clusters are compared. Other sources of information are the American Cancer Society (ACS) (http://www.cancer.org), the Cancer Information Service (http://www.cis.org),18 and the Council of State and Territorial Epidemiologists (http://www.cste.org).

The initial steps in investigating perceived cancer clusters are straightforward. Health workers inquire about the number of people who have developed cancer, their age, type of cancer, dates of diagnosis, and period of residence in the community. Where appropriate, officials may obtain medical records to confirm the diagnoses and collect supplemental clinical information.19 In many instances, perceived cancer clusters are not confirmed because the cases involve different types of cancers with no known relationship to each other, health conditions other than malignancy, or diagnoses made before moving into the community. Discussions at this point may alleviate public concern by documenting the absence of a cluster. Depending on the circumstances, review of environmental monitoring data may also be indicated.

Formal statistical testing involves comparing the observed number of cases with the number expected, based on the size and age composition of the population. The expected number of cases is estimated by applying background incidence rates at various ages in the general population (from cancer registry data) to the population of interest. For the comparison to be valid, it is essential that identical criteria be used to define cases and persons at risk in the two populations. For example, only people who live in the community at the time of their diagnosis should be counted among the observed cases. Those diagnosed before or after their period of residence should not be included, because state tumor registries only capture cancers diagnosed during the period of residence. The expected number of cases increases with each year of observation. Thus, the number of cases expected in a single year should be multiplied by the number of years over which cases in the perceived cluster occurred.

Complexities of Statistical Testing

Despite the value of statistical testing, chance remains the most plausible explanation for many confirmed cancer clusters, especially those that involve common types of cancer or all cancers combined. Because of the increase in life expectancy and the strong relationship between cancer risk and aging, cancers are more common than recognized. About one of every two men and one in every three women will develop cancer over full life expectancy. Given that an estimated 1,368,000 new diagnoses and 563,700 deaths from cancer are expected in 2004,20 some spatial clustering is inevitable. For instance, a city of 100,000 people with the same age distribution as the United States can, on average, expect 473 new cases and 200 deaths from cancer each year. Even if these cases occur randomly, some clustering will occur by chance. However, the communities affected by clustering may not perceive their experience as part of a larger random pattern, but as the direct consequence of some local underlying cause. This interpretation is analogous to the Texas “sharpshooter” who first fires his shots randomly at a wall and then draws a bull’s-eye around a cluster of bullet holes.21 The fact that the boundaries of a suspected cluster are defined based on when and where the cases actually occurred increases the likelihood that random variation will appear to give rise to clusters.

WHY ARE INVESTIGATIONS OF COMMUNITY CANCER CLUSTERS DIFFICULT?

Many factors limit the information that can be gained by investigating cancer clusters, especially in community settings. The levels of exposure to industrial or agricultural pollutants are much lower and more difficult to assess in nonoccupational settings than in many workplaces; the populations at risk are less clearly
defined. Even when exposure levels exceed environmental standards, the expected increase in risk from community exposures would be detectable only in very large populations rather than in localized clustering. Although certain individuals, such as pregnant mothers and children, may be especially susceptible to toxic and carcinogenic exposures, the size of these groups is usually too small in any single community to support extensive statistical analyses. Epidemiologic methods that can provide strong evidence of association in large studies have limited value in cluster investigations, especially in the absence of documented high-level exposure. Finally, environmental monitoring of current exposures may not satisfy skeptics who contend that past exposures were probably higher and more relevant than current exposures to the development of cancer in the affected individuals.

Another problem that complicates studies in community settings arises from inaccurate data on the population at risk in small geographic areas or demographic subgroups. Census data are less accurate for cities or counties than for states. The uncertainty is greatest for demographic subgroups of the population during the 10-year interval between national census counts. Two recent examples illustrate this problem. The first involves a report of higher cancer incidence and mortality among African Americans in Atlanta than in other areas covered by NCI registries. Com pared with average death rates among African Americans, African American residents of Atlanta appeared to have 40%, 19%, and 16% higher mortality rates from prostate, breast, and colon cancer, respectively, during the 1990s. When updated population data were released from the 2000 census, however, the death rate from these cancers was seen to be similar in African Americans across all of the NCI registries. The higher estimates during the 1990s resulted from an underestimation of migration of African Americans into Atlanta during that period.

A second related example concerns the apparent rapid increase in breast cancer incidence in Marin County, California during the 1990s. Breast cancer incidence was reported to increase by 3.6% per year in Marin County between 1990 and 1999. This increase, which was confined to non-Hispanic White women aged 45 to 64 years, appeared to be six times larger than the increase in other counties in the San Francisco Bay Area. However, a reanalysis based on population data from the 2000 census, rather than projections from the 1990 census, revealed that breast cancer incidence in Marin County had not actually increased more rapidly than in adjoining counties. Rather, projections from the 1990 census underestimated the number of non-Hispanic White women aged 45 to 64 who moved into Marin County in the 1990s. Although breast cancer incidence is high in Marin County, as in other affluent counties, the alarming increase in incidence reported during the 1990s appears to have been an artifact of inaccurate projections of the underlying population.

Regardless of the setting of a suspected cancer cluster, investigations are also complicated by the lack of clinical or molecular tests that can determine the cause of cancer in an individual. Until such tests are developed, researchers must rely on epidemiologic studies that can identify factors associated with risk in groups of people, but not the precise cause of disease in an individual. Because of these difficulties, even extensive investigations of cancer clusters are rarely successful in determining the cause of clusters in community settings. For example, the CDC systematically investigated a series of 108 community cancer clusters reported from 29 states and five foreign countries in the years 1961 to 1982. In none of these did the researchers consider the cause to be well established. NIOSH investigated 61 suspected occupational cancer clusters during the period of 1978 to 1984, most of which included five or fewer cases and had no plausible occupational etiology. In such cases, the apparent cluster is attributed either to chance or to exposures that could not be documented using the investigative tools available at the time.

Despite the many obstacles to investigating cancer clusters in the community, some clusters may nevertheless have common etiologic factors that have not yet been identified. For instance, numerous clusters of childhood
leukemia, and to a lesser extent lymphoma, are reported in the scientific literature. Leukemia clusters have been recorded in Europe since the beginning of the 20th century.27 The first extensive investigations of such clusters were conducted in Northumberland, England28 and Niles, Illinois29 in the early 1960s. Other investigations of childhood leukemia have generated scientific and media interest, such as the cluster near a nuclear power plant in Sellafield, England.30,31 An exceptionally large cluster of childhood leukemia occurred in Churchill County (Fallon), Nevada from 1997 to 2001. Eleven cases of leukemia were identified over a five-year period among children in a community of 26,000 people. Four others who had previously lived in the area but had moved away were also diagnosed with leukemia. Only one case every five years would be expected among the resident population of this age, based on average incidence rates in Nevada.32 Extensive investigation failed to identify an underlying cause for the clustering. Although most statistical analyses suggest that clusters of childhood leukemia occur somewhat more frequently than would be predicted by chance,27,33 such clustering explains only a small fraction of incident cases. Researchers have hypothesized that an as yet unidentified infectious exposure occurring at a particular stage in development may give rise to these clusters.

When is an Extensive Investigation Appropriate?

There are many more reports of suspected cancer clusters than can or should be investigated extensively. The goals of an initial evaluation are to respond to community concerns, to document the facts of what has happened (and thereby minimize the influence of rumor), and to assist the community in determining and implementing the appropriate response. While it is critical to triage reported clusters to determine which should be investigated more thoroughly, it is equally important to hear the community’s concerns and provide information about how reports of cancer clusters are evaluated and what has been learned. Approaches that can improve communication with the community are discussed below.

In some cases, further investigation of a documented cancer cluster is indicated. Increasingly, epidemiologic studies of the community are only conducted when the following conditions are met: (1) the observed number of cases of a specific type of cancer significantly exceeds the number expected; (2) either the type of cancer or age at onset is highly unusual; (3) the population at risk can be defined; and (4) prolonged exposures to known or suspected carcinogens at levels that exceed environmental limits can be documented. The demand for further investigation is greatest when new cases continue to be diagnosed. Further environmental monitoring and/or review of environmental data may be indicated in situations with an identifiable source of contamination. This may be useful to document local contamination and stimulate cleanup. However, the community should be informed in advance that environmental measurements rarely resolve controversy about the cause of the cluster and will not, by themselves, provide scientifically convincing evidence linking the cluster to environmental exposure. The decision of whether or not to conduct further investigation of a cancer cluster is, in most cases, difficult. To some it may appear negligent not to explore every possible explanation for the apparent cluster. However, the desire to “leave no stone unturned” is not in itself a sufficient reason to conduct extensive environmental monitoring or medical testing. Professional judgment about the likelihood of whether further investigation will be informative should help to guide health officials and communities confronting these difficult situations.

Following the completion of an investigation, state health departments may continue to monitor cancer occurrence in the local community and the surrounding county for three to five years.6,9,31 It is presumed that an observed “excess” of cancer cases due to chance will not continue and that the incidence rate will return to the expected range during this
period. If the rate remains elevated, further studies may be performed.  

Talking with the Community

Perhaps the most important challenge for public health agencies that deal with cancer clusters is to communicate effectively with the public. This has been described as the “art of being responsibly responsive.”  
7 State or local health departments usually take primary responsibility for this; physicians in the community can serve an essential role. Communication should begin early, before divergent points of view become highly polarized. It is often helpful to convene a public meeting to hear specific concerns and varying points of view. This provides an opportunity to explain what is known, what steps are being taken to investigate the situation, and to provide background information about suspected cancer clusters. The effectiveness of such a meeting depends on speakers who have considerable experience and credibility in medicine, public health, and cluster investigations and who are able to interact effectively with an alarmed public. Credibility is enhanced by the endorsement of respected leaders of the community with no financial stake in the outcome of an investigation. The goal is to provide a structured process within which individuals can voice their concerns and support informed community decision making.

Potential Roles for Physicians

Physicians are a respected source of information about health and disease. Their extensive interactions with patients and their families provide opportunities to reassure patients in situations that are unlikely to involve a cancer cluster, educate patients about ways to avoid cancers or identify them early, and identify settings that warrant investigation by public health agencies. Physicians may live in communities affected by a suspected cancer cluster. In such cases, an informed doctor can contribute to the public debate by providing background information about cancer and cancer clusters and by realistically describing what can or cannot be learned by exhaustive investigation of environmental exposures. Public concern about cancer clusters provides broader opportunities to educate patients and community leaders about cancer and the value of proven strategies of prevention and early detection.

CONCLUSIONS

In recent decades, considerable public health energy has been invested in the investigation of reported cancer clusters. Responding to such clusters is a legitimate and necessary public health activity,  
7 but many state and local health departments have limited resources to respond to the number of perceived clusters reported each year. Informed clinicians can play an important role by helping to educate patients and their families about cancer and by contributing to public debate and decision making.

REFERENCES