

Persons using assistive technology might not be able to fully access information in this file. For assistance, please send e-mail to: mmwrq@cdc.gov. Type 508 Accommodation and the title of the report in the subject line of e-mail.

Guidelines for Investigating Clusters of Health Events

Summary

Clusters of health events, such as chronic diseases, injuries, and birth defects, are often reported to health agencies. In many instances, the health agency will not be able to demonstrate an excess of the condition in question or establish an etiologic linkage to an exposure. Nevertheless, a systematic, integrated approach is needed for responding to reports of clusters. In addition to having epidemiologic and statistical expertise, health agencies should recognize the social dimensions of a cluster and should develop an approach for investigating clusters that best maintains critical community relationships and that does not excessively deplete resources.

Health agencies should understand the potential legal ramifications of reported clusters, how risks are perceived by the community, and the influence of the media on that perception. Organizationally, each agency should have an internal management system to assure prompt attention to reports of clusters. Such a system requires the establishment of a locus of responsibility and control within the agency and of a process for involving concerned groups and citizens, such as an officially constituted advisory committee. Written operating procedures and dedicated resources may be of particular value.

Although a systematic approach is vital, health agencies should be flexible in their method of analysis and tests of statistical significance. The recommended approach is a four-stage process: initial response, assessment, major feasibility study, and etiologic investigation. Each step provides opportunities for collecting data and making decisions. Although this approach may not always be followed sequentially, it provides a systematic plan with points at which the decision may be made to terminate or continue the investigation.

INTRODUCTION

Clusters of health events may be identified by an ongoing surveillance system, but more often they are reported by concerned citizens or groups. Although health agencies must respond to these reports, little guidance has been available to them.

These guidelines focus on noninfectious health events such as chronic diseases, injuries, and birth defects. Numerous related issues--such as the epidemiologic workup of infectious disease outbreaks, the assessment of the health effects of environmental exposures, the prospective detection of clusters, and the investigation of interpersonal networks--are not addressed.

Purposes and Assumptions

The purpose of these guidelines is:

- To provide epidemiologic and statistical source material to

state and local health agencies to aid in their development of a systematic approach to the evaluation of clusters of health events.

- To provide generic guidelines for assessing clusters of health

events (e.g., noninfectious diseases, injuries, birth defects, and previously unrecognized syndromes or illnesses).

- To supplement, rather than supplant, existing state and local

plans for evaluating clusters.

Largely on the basis of conference deliberations, the working group (see Preface for list of participants) has used the following operating assumptions:

--In many reports of cluster investigations, a geographic or temporal excess in the number of cases cannot be demonstrated. --When an excess is confirmed, the likelihood of establishing a definitive cause-and-effect relationship between the health event and an exposure is slight. --A cluster may be useful for generating hypotheses but is not likely to be useful for testing hypotheses. Frequently, the issues raised by a cluster cannot be definitively answered by the investigation per se; they require an alternative epidemiologic approach. --From a public health perspective, the perception of a cluster in a community may be as important as, or more important than, an actual cluster. In dealing with cluster reports, the general public is not likely to be satisfied with complex epidemiologic or statistical arguments that deny the existence or importance of a cluster. Achieving rapport with a concerned community is critical to a satisfactory outcome, and this rapport often depends on a mutual understanding of the limitations and strengths of available methods.

Definition, Background, and Characteristics of Clusters

As used in these guidelines, the term "cluster" is an unusual aggregation, real or perceived, of health events that are grouped together in time and space and that are reported to a health agency.

Several breakthroughs and triumphs in infectious disease control have resulted from the epidemiologic evaluation of clusters of cases. Well-known examples include the epidemic of cholera in London in the 1850s (1), the investigation of cases of pneumonia at the Bellevue-Stratford Hotel in Philadelphia in 1976 (2), and the report in 1981 that seven cases of *Pneumocystis carinii* pneumonia had occurred among young, homosexual men in Los Angeles (3).

Investigations of noninfectious disease clusters have also resulted in notable examples of breakthroughs, such as angiosarcoma among vinyl chloride workers (4), neurotoxicity and infertility in kepone workers (5), dermatitis and skin cancer in persons wearing contaminated gold rings (6), adenocarcinoma of the vagina and maternal consumption of diethylstilbestrol (7), and phocomelia and thalidomide (8).

A review of these landmark events and other material on clusters enables public health personnel to identify characteristics of a cluster from which an investigation might lead to important results. Usually, such a cluster has a definable health outcome, either new or rare; a potential exposure or agent is suspected, along with a connection between the exposure and the health event; the situation is highly unusual, and statistical testing confirms the investigator's impression; and the short-term public health impact is immediate and self-evident.

The reported experience of health agencies confirms, however, that major associations between exposures and outcomes are rare. Minnesota, for example, has reported results from over 500 investigations of

clusters (9), six of which were full-scale investigations. In one instance, in an occupational setting, an important public health outcome concerning cancer was documented (10). Missouri (11) and Wisconsin (12) have reported similar experiences: large numbers of requests for investigations have been received, but only an occasional in-depth evaluation is warranted. CDC has been consulted in over 100 such investigations, and again, major associations between exposures and outcomes have been rare (13).

Investigations of clusters reported reflect only a fraction of the activity of health agencies. The unofficial consensus among workers in public health is that most reports of clusters do not lead to a meaningful outcome. Often, a "case" is not clearly defined, and the "cluster" is, in fact, a mixture of different syndromes. Frequently, no exposure or potential cause is obvious, and--to make the investigation even more difficult--there are many possible causes. For example, an inactive toxic waste site may contain hundreds of chemicals. An investigation at the site may indicate no immediate or obvious connection between exposure or disease, and considerable manipulation may be required to demonstrate a statistically significant excess. Finally, the biologic consequences and public health impact often are not clear.

Despite these impediments, reports of clusters cannot be ignored. The health agency must develop an approach that maintains community relations and that manages clusters without excessively depleting resources (14). At times, the health agency must assume a leadership role in recognizing the underlying issues, in understanding the limits of epidemiologic investigation, and in guiding concerns to the appropriate arena.

These guidelines are divided into four sections and an Appendix. --Section 1 addresses the epidemiologic and management skills required for a systematic public health approach to clusters. --Section 2 suggests organizational requirements that can facilitate the management of clusters within a health agency. --Section 3 outlines a systematic, four-stage approach to evaluating clusters. --Section 4 describes statistical and epidemiologic techniques. --The Appendix provides a review and critique of available statistical methods.

SECTION 1. MANAGING INVESTIGATIONS OF CLUSTERS: SKILLS AND KNOWLEDGE REQUIRED

The investigation of a perceived cluster of adverse health effects is not simply an isolated epidemiologic or statistical exercise. Appropriate response by public health agencies to requests for such investigations demands that the complexity of this area be recognized and, in addition, requires the possession and application of skills and knowledge that extend well beyond statistical and epidemiologic tools. These additional skills include a sensitivity to the psychology of the situation, an understanding of the principles of risk perception, a recognition of the functions of public media, and an awareness of potential legal ramifications of the investigation.

Scientific Tools

The investigation of clusters may be best viewed as a form of public health surveillance (i.e., the ongoing collection, analysis, and dissemination of information important to public health practice) that responds to community needs. It is not necessarily a primary mechanism for investigating etiologic relationships. Thus, the investigator may be looking more at patterns (spatial, temporal, or both) in data than searching for specific associations between agent and disease (15). As discussed in Sections 3, 4, and the Appendix, various statistical techniques may be used to detect and characterize such patterns--none of which is consistently the technique of choice or the most appropriate. The investigator should select the epidemiologic or statistical approach to be used according to the circumstances under study (e.g., the nature of the condition, the type of data available on the cases, and the availability of comparison divided by denominator data). In addition to knowing how to apply the selected method, the investigator will need to know its limitations, assumptions, and tendency to give false-positive or false-negative results (and under what conditions it is prone to do so). Finally, the investigator should be familiar with the concept of statistical power and be able to determine the power of any planned study to detect an increased number of cases.

Once the presence of a cluster or an excess of disease or injury is confirmed, a comprehensive investigation may require the capacity to conduct environmental sampling, including the knowledge and equipment necessary to design an appropriate sampling scheme and to collect the specimens; access to

laboratories with adequate facilities and experienced staff to analyze these specimens quickly and accurately, with appropriate attention to quality control/assurance procedures; and the ability to interpret the results. Similarly, the capacity must exist to collect, analyze, and interpret biological monitoring specimens (whether used as measures of exposure or adverse health effects). What is initially thought to be a cluster of cases may, in fact, represent a cluster of incorrectly performed laboratory tests (16).

Psychological Factors

Investigators of clusters should understand the various ways in which individuals respond to stressful situations and react to uncertainties (17). Investigators also should be able to recognize the source of inevitable community suspicions (e.g., of deliberate delay and cover-ups) and demands (e.g., for the unrealistic allocation of resources and schedules). Investigators should respond to these suspicions and demands without hostility and should be able to diffuse them. Finally, investigators must be aware of and responsive to the fact that a perceived problem must be resolved responsibly and sympathetically, even if no underlying community health problem or cluster of disease truly exists. Risk Communication

Once the investigator has estimated the degree of risk inherent in the situation under study, this information should be given to the community. Simply presenting the numbers usually will not suffice. The risk must be put in perspective--in a sensitive, noncondescending manner--through comparison with involuntary risks associated with more familiar activities (18).

In addition, the risk perceived by community members does not necessarily parallel the estimates of risk that are produced by mathematical or scientific assessments (19). This divergence is more than a failure to communicate the true risk or a failure of the community to understand. Rather, it represents a factoring of other aspects of the situation into the reactions of community members (e.g., the extent to which the acceptance of the risk is voluntary or imposed, the degree of control the individual or community has over the source of the risk, the degree to which the source of the risk is familiar and easily comprehended, and the potential adverse social and economic ramifications) (19). Public Media

Public health agencies should be aware of media "imperatives." Investigators must understand the factors that influence the various media in their selection and presentation of stories (e.g., the desire for a pictorial/visual component, the presence of conflict or controversy, the presence of strong emotive content, and the availability of target for blame) (20). Similarly, investigators must recognize that the media tend to simplify complex, technical explanations, thereby losing subtle distinctions or qualifications. Thus, investigators should distill the messages they wish to convey and present them in the way they are most likely to be transmitted without confusion or distortion. Investigators must be prepared to stress key points; provide background necessary for understanding; and be straightforward regarding what is fact, what is speculation, and what is not known. Most of all, investigators must remain cooperative and responsive and must be prepared to provide needed information rapidly, before distortion and discord have been introduced into the public exchanges. Legal Ramifications

Many situations that prompt requests for investigations ultimately involve litigation (ongoing or contemplated litigation may, in fact, stimulate the request for the investigation) or government intervention. Since the investigation report is likely to be used in that litigation or to justify that intervention, members of public health agencies need a basic understanding of the principles of tort law that relate to legal proof of causality and responsibility--and must understand how these differ from the sometimes stricter requirements of scientific proof. Such principles include the concept of negligence, which entails the breach of duty that caused or substantially contributed to harm or damage; the concept of breach of warranty, the understanding that an action or situation is safe; the concept of strict liability, which focuses on the product rather than on the conduct; and the concept of failure to warn (21). Legally establishing a cause-and-effect relationship requires only that a preponderance of the evidence (i.e., the probability is greater than 0.5) indicates the association (21). SECTION 2. ORGANIZATIONAL REQUIREMENTS

A citizen who reports an apparent cluster wants assurance that the appropriate persons will be notified and that immediate action will be taken. The health agency should be organized to receive and respond to

reports of potential clusters so as to systematize the following: --A reporting process that is quick and traceable--and one in which the appropriate person is reached regardless of the first contact made by the concerned citizen. --A response process that is triage-oriented, that can proceed smoothly from one level of action to the next, and that can terminate effectively when resolution is reached. --A feedback and notification process that educates and enlightens with efficiency and courtesy. --A referral process that assures timely and competent field investigation and public health response.

The following organizational components are recommended to assure smooth and timely public health responses:

- A locus of responsibility and control. The health agency should designate an individual with stature in the agency to serve as the identifiable point of responsibility or Program Director (e.g., Director of Environmental Health, State Epidemiologist, Director of Cancer Control Activities, Chronic Disease Director, or County Health Officer). The designation will depend on local circumstances and priorities.
- A process for involving concerned groups and individuals. The health agency should consider the establishment of an advisory committee (or similar group) to oversee the decision-making process for evaluating clusters. Such a committee might include representatives from the health agency, other government agencies, private and voluntary sectors, concerned citizens' groups, and the media, as well as selected individuals. This committee should provide oversight, guidance, and advice to the Program Director. The duties should be carefully specified a priori and agreed to by the committee members. Since the committee is likely to consist of persons with diverse backgrounds, the main focus should be on overseeing the process rather than on making technical decisions.
- A set of operating procedures. The health agency should establish a written protocol for evaluating clusters. The protocol should include mechanisms for reaching responsible persons and a detailed enumeration of the triage process. (Guidelines for such a protocol are provided in Section 3.) This protocol should be disseminated to appropriate points in the health agency to assure proper handling of the report, whatever the point of entry.
- Dedicated resources. Health agencies that have a recurrent need for evaluating clusters probably will require staff and support resources. The Program Director may not be in a position to perform the required day-to-day tasks and may not be the appropriate primary contact. Frequently, responsibilities may be shared by a number of staff members, but the duties and responsibilities should be clearly designated.

Just as no statistical test is best for evaluating all clusters, no organizational structure is best in all situations. The specifics of organization will depend on local circumstances. **SECTION 3. GUIDELINES FOR A SYSTEMATIC APPROACH**

This section outlines a four-stage approach for managing a reported cluster, from original contact to final disposition. The section does not speak directly to the particular outcome of concern (e.g., cancer or birth defects), to the types of data available (mortality, hospital discharge, or disease registries), or to the specific analytic techniques (see Section 4 and the Appendix). Usually, these particulars will be determined by local resources and circumstances. The four stages may be viewed as a series of filters that provide appropriate responses to the reported problem. An assessment of feasibility should be made before the actual study is begun, and the issue of increased frequency of occurrence should be separated from the issue of potential etiologies (Figure 1).

These guidelines should be viewed with the following caveats in mind: --The boundaries between the stages are not fixed. Often, the health agency will choose to follow a different order, to combine steps, or to pursue a problem on several fronts. Considerable local judgment and discretion are required. --The investigation can be resolved at a number of points along the path by a report to "the caller" (the

individual who initiated the contact) and to other interested parties (Figure 1). This step implies that an internal report will be generated for the health agency and its advisory groups. Such reports are useful communication tools, particularly if they are regularly scheduled and available to an established, but flexible, list of recipients. --Although health agencies may have a number of organizational similarities (e.g., the presence of a public affairs office and a cluster advisory committee), their internal structure and function may vary considerably. The guidelines are meant to be tailored to local circumstances. If health agencies choose to establish an advisory committee, the assumption is that the committee will be consulted at critical decision points. Stage 1. Initial Contact and Response

Purpose: to collect information from the person(s) or group(s) first reporting a perceived cluster.

The initial contact is critical. The caller should be referred quickly to the responsible unit in the health agency, and the problem should never be dismissed summarily. Most reports of potential clusters can be successfully closed at the time of initial contact, and the first encounter is often the health agency's best opportunity for communication with the caller about the nature of clusters. Procedures

- A. Gather identifying information on the caller, unless anonymity is requested: name, address, telephone number, and organization affiliation, if any. If anonymity is requested, advise the caller that the inability to follow up may hinder further investigation.
- B. Gather initial data on the potential cluster: suspected health event(s), suspected exposure(s), number of cases, geographic area of concern, time period of concern, and how the caller learned about the cluster.
- C. Obtain identifying information on persons affected: name, sex, age (or birthdate, age at diagnosis, age at death), occupation, race, diagnosis, date of diagnosis, date of death, address (or approximate geographic location), telephone number, length of time in residence at site of interest, contact person (family, friend) and method for contact, and physician contact. In some instances, the health official may choose not to collect identifying information during the first contact but instead to gather it during several contacts.
- D. Discuss initial impressions
D. Discuss initial impressions with a caller. The following frequently arise: --A variety of diagnoses speaks against a common origin. --Cancer is a common illness (with a one in three lifetime probability). The risk increases with age, and cases among older persons are less likely to be true clusters. --Major birth defects are less common than cancer but still occur in 1%-2% of live births. --Length of time in residence must be substantial to implicate a plausible environmental carcinogen because of the long period of latency required for most known carcinogens. --Cases that occurred among persons now deceased may not be helpful in linking exposure to disease because of the lack of information on exposure and because of possible confounding factors. --Rare diseases may occasionally "cluster" in a way that is statistically significant, but such an occurrence may be a statistical phenomenon not related to exposure.
- E. Request further information on cases, obtain more complete enumeration, and plan a follow-up telephone contact, as needed.
- F. Assure the caller that he or she will receive a written response. (Often, the written response simply confirms what has already been communicated by telephone.)
- G. Maintain a log of initial contacts, whether they are made in writing, by telephone, or in person. The log should include the date, time, caller identification, health event, exposure, and geographic area. Follow-up contacts should be logged in as well, with a brief note as to purpose and result. If possible, the log should be cross referenced and computerized so that all personnel concerned will have the same information.

H. Notify the health agency's public affairs office (or equivalent) about the contact. In many agencies, this action is analogous to notifying the commissioner's office of a press contact.

Early in the investigation of a cluster, the health agency may be asked to collect new environmental data or to use previous measurements, although the latter may not exist. Premature environmental measurements should be avoided, since they may be unfocused and uninterpretable. Outcome --If the initial contact suggests that further evaluation is needed (e.g., single and rare disease entity, plausible exposure, or plausible clustering), proceed to Stage 2, Assessment. --If the initial contact permits satisfactory closure, prepare a summary report for the caller and for the advisory committee (or other supervisory group). Stage 2. Assessment

Once the decision has been made to proceed with an assessment, an important step is to separate two concurrent issues: whether an excess has actually occurred and whether the excess can be linked etiologically to some exposure. The first issue usually has precedence, and it may or may not lead to the second. This stage initiates a mechanism for evaluating whether an excess has occurred. Three separate elements are identified: a preliminary evaluation (Stage 2a) to assess quickly from the available data whether an excess may have occurred; case evaluation (Stage 2b) to assure that a biological basis exists for further work; and an occurrence investigation (Stage 2c) for the purpose of obtaining a more detailed description of the cluster through case finding, interaction with the community, and descriptive epidemiology. In addition, the investigators may wish to review the scientific literature and seek consultation with other investigators. These activities are often interrelated and may occur in parallel. The health agency is encouraged to be flexible in conducting this portion of the investigation and to recognize that a linear approach is often not possible. Stage 2a. Preliminary evaluation

Data from the initial contact, possibly with augmentation from other sources, are used to perform an in-house calculation of observed versus expected occurrence.

Purpose: to provide a quick, rough estimate of the likelihood that an important excess has occurred.

Procedures A. Determine the appropriate geographic area and the period in which to study the cluster. B. Determine which cases will be included in the analysis. Because this stage does not involve case verification, all cases will be assumed to be real. However, some cases may need to be excluded from the analysis because they occurred outside the geographic area or the period decided on, or because the health event for the case differs from that of other cases. A helpful step may be to tabulate frequencies of health events and to look at related descriptive statistics. C. Determine an appropriate reference population. Occurrence rates (or other statistics) calculated for the cluster should be compared with those for a reference population in order to identify an excess number of cases. D. If the number of cases is sufficient, and if a denominator is available (e.g., population of a community, number of children in school, or number of employees in a workplace), calculate occurrence rates, standardized morbidity/ mortality ratios, or proportional mortality ratios (see Section 4). Compare the calculated statistic with that for the reference population to assess significance. Chi-square tests and Poisson regression are also commonly used techniques for comparing proportions. E. If the number of cases is not large enough to obtain meaningful rates, or if denominator data are unavailable, use one of the statistical tests developed to assess space, time, or space-time clustering (Appendix).

Outcome --If the preliminary evaluation suggests an excess occurrence, proceed to case evaluation. --If the preliminary evaluation suggests no excess, respond to the caller, indicating findings and advising that no further investigation is needed. --If the preliminary evaluation shows no excess but the data suggest an occurrence of biologic and public health importance, decide if further assessment is warranted. A decision to proceed further at this point should not be based solely on an arbitrary criterion for statistical significance. Stage 2b. Case evaluation

Purpose: to verify the diagnosis. Some health agencies may choose to verify diagnoses before calculating preliminary rates (Stage 2a). Because verification may be costly, however, agencies usually calculate rates first. Procedures

- A. Verify the diagnosis by contacting the responsible physicians or by referring to the appropriate health-event registry. Verification is often a multi-step process, involving initial contact with the patient, family, or friends and subsequent referral to the responsible physicians to obtain permission to examine the records.
- B. If possible, obtain copies of relevant pathology reports or medical examiner's report.
- C. Obtain histologic reevaluation if needed. (Often, however, confirmation and reevaluation are difficult to obtain.)

Outcome --If cases are verified and an excess is confirmed, proceed to Stage 2c, the occurrence evaluation (which already may be under way). --If some (or all) of the cases are not verified and an excess is not substantiated, respond to the caller, outlining findings and advising that further evaluation is not warranted. --If some of the cases are not verified but biologic plausibility persists and the data are suggestive, consider initiating or continuing the occurrence evaluation. Stage 2c. Occurrence evaluation

Purpose: to design and perform a thorough investigation to determine if an excess has occurred and to describe the epidemiologic characteristics.

The occurrence evaluation is meant to define the characteristics of the cluster, often requiring a field investigation. This evaluation begins with a written protocol that outlines the costs and provides information on data collection, the methods to be used, and the plan of analysis. The main product should be a detailed description of the cluster. Up to and including this stage, the allocation of resources is relatively small. Procedures

- A. Determine the most appropriate geographic (community) and temporal boundaries.
- B. Ascertain all potential cases within the defined space-time boundaries.
- C. Identify the appropriate data bases for both numerator and denominator and their availability. D. Identify statistical and epidemiologic procedures to be used in describing and analyzing the data.
- D. Perform an in-depth review of the literature, and consider the epidemiologic and biologic plausibility of the purported association.
- E. Assess the likelihood that an event-exposure relationship may be established.
- F. Assess community perceptions, reactions, and needs.
- G. Complete the proposed descriptive investigation.

Although an advisory committee can be helpful at any point in the process, it may be of particular importance at this point. The occurrence evaluation may vary considerably in size and content; consensus on the appropriate level of effort will facilitate acceptance of the results. Outcome --If an excess is confirmed and the epidemiologic and biologic plausibility is compelling, proceed to Stage 3, the major feasibility study. --If an excess is confirmed but no relationship to an exposure is apparent, terminate the investigation and inform the persons concerned of the possible risks/no risks involved. --If an excess is not confirmed, terminate the investigation and report findings to the caller. Stage 3. Major Feasibility Study

Purpose: to determine the feasibility of performing an epidemiologic study linking the health event and a putative exposure.

The major feasibility study examines the potential for relating the cluster to some exposure. All of the options for geographic and temporal analysis should be considered, including the use of cases that were not part of the original cluster and are of a different geographic locale or time period. In some instances, the feasibility study may provide answers to the basic issue (14). Procedures

- A. Review the detailed literature search with particular attention to known and putative causes of the outcome(s) of concern.
- B. Consider the appropriate study design, with attendant costs and expected outcomes of alternatives (e.g., a consideration of sample size, the appropriateness of using previously identified cases, the geographic area and time period concerned, and the selection of controls).
- C. Determine what data should be collected on cases and controls, including physical and laboratory measurements.D. Determine the nature, extent, and frequency of and the methods used for environmental measurements.
- D. Delineate the logistics of data collection and processing.
- E. Determine the appropriate plan of analysis, including hypotheses to be tested and power to detect differences; assess the epidemiologic and policy implications of alternative results.
- F. Assess the current social and political ambiance, giving consideration to the impact of decisions and outcomes.
- G. Assess the resource implications and requirements of both the study and alternative findings.

Outcome --If the feasibility study suggests that an etiologic investigation is warranted, proceed to Stage 4. The investigation may require extensive resources, however, and the decision to proceed will be related to the allocation of resources. --If the feasibility study suggests that little will be gained from an etiologic investigation, summarize the results of this process (by now rather extensive) in a report to the caller and all other concerned parties. In some circumstances the public or media may continue to demand further investigation regardless of cost or biologic merit. The effort devoted to community relationships, media contacts, and advisory committee interaction will be critical for an appropriate public health outcome.

Stage 4. Etiologic Investigation

Purpose: to perform an etiologic investigation of a potential disease- (or injury-) exposure relationship.

The primary purpose of the study is to pursue the epidemiologic and public health issues that the cluster generated--not necessarily to investigate a specific cluster. In that context, this step is a standard epidemiologic study, for which all the preceding effort has been preparatory. Procedures

Using the major feasibility study as a guide, develop a protocol, and implement the study. The circumstances of most epidemiologic studies tend to be unique; therefore, more specific guidance is not appropriate for inclusion in this publication. Outcome

The results of an etiologic investigation are expected to contribute to epidemiologic and public health knowledge. This contribution may take a number of forms, including the demonstration that an association does or does not exist between exposure and disease, or the confirmation of previous findings.

SECTION 4. STATISTICAL AND EPIDEMIOLOGIC TECHNIQUES

The approach taken to investigate a suspected cluster of health events depends on the nature of the cluster, the data available, and the questions being asked, including the following: --Do the health events cluster in space or time alone, in space and time simultaneously, or in neither? --What are the spatial and temporal boundaries of the cluster? --What are the characteristics of the health events--e.g., acute or chronic disease, long or short latency period, and known or unknown etiology? --What data are available for the health event--e.g., case counts, disease rates, or data on each event, such as place of residence and time of onset of disease or death? --What data are available to describe the population at risk?

A number of problems are encountered in the study of clusters. The health events being investigated (often morbidity or mortality) are usually rare, and increases of these events tend to be small and may occur over

a long period. Another issue that complicates the investigation is that some clusters occur by chance. Information on the population at risk or on the expected rates often is not available. A further complicating factor for methods using aggregated data is that health events occur in space and time continua, thus yielding optional and suboptimal units for displaying a pattern. The choice of a geographic area that is too small or too large, or of a time period that is too short or too long, may result in insufficient statistical power to indicate a cluster. Many of the articles referenced in the Appendix contain informative discussions about issues that can compromise application of statistical methods in investigations of clusters.

Standard statistical and epidemiologic techniques for assessing excess risk can often be used to evaluate reported clusters. Tabulating frequencies of the health event and examining related descriptive statistics is a useful first step in the evaluation. Mapping the data is also helpful. If the number of cases is sufficient and population data are available, examination of rates (possibly age-, race-, and sex-adjusted), standardized mortality/morbidity ratios, or proportional mortality ratios may determine whether there is an excess number of events. If the number of health events is too small to show meaningful rates, pooling across geographic areas or time may be possible. Combinatorial methods are often used for small amounts of data. Other commonly used statistical approaches include chi-square tests of observed versus expected frequencies (based on the Poisson distribution for low-frequency data) and Poisson regression (used for comparison of rates). Confidence intervals may be calculated for point estimates.

Whether the rate for a geographic area or time period is excessive may be determined by comparing it with rates of other areas or times. If a spatial cluster is being assessed, the occurrence in the geographic area can be compared with that in adjacent areas (e.g., a census tract with surrounding census tracts) or with other areas of similar size (e.g., a county with other counties in a state). Alternatively, the rate for an area can be compared with that of a larger area (e.g., the rate for a city with that of the surrounding county). If a temporal cluster is being assessed, the occurrence in that time period can be evaluated in the context of previous or subsequent periods. When such comparisons are made, the referent population must be chosen carefully to ensure its appropriateness. Mortality and morbidity data for referent populations are available from state and national vital statistics systems or registries such as cancer and birth defect registries. Population data are available from the Bureau of the Census. A county-level file with both mortality and population data for 1968-1985 (the Compressed Mortality File) is available for public use from the National Technical Information Service.

If the above standard approaches cannot be used in an investigation of clusters because the number of health events is too small, data on the population at risk are unavailable, or space-time clustering is suspected, numerous statistical tests are available for use in detecting spatial, temporal, and space-time clusters. Although some of these tests may not be familiar to investigators and may require the preparation of more data than required by standard techniques, many of the tests are simple to understand and use. Numerous methods for studying clusters have been reviewed (22,23). Brief descriptions and critiques of some of these techniques are presented in the Appendix.

Most of the tests reviewed in the Appendix use data on individual cases of health events, although a few employ aggregated data such as frequency counts or rates. Information generally required for each case is location of the case (often the geographic coordinates of place of residence) and date of onset of the disease (or injury) or of death. Most of the tests based on aggregated data assume that the number of health events that occur in an area and/or time period follows a Poisson distribution. The tests do not usually require knowledge of the distribution of the population at risk. Instead, they may assume that the population at risk remains constant over time, and they offer special considerations for differing population sizes. The reporting rate for the health event is also assumed to be constant.

The assumption of minimal population shifts over time is frequently violated. More subtly, subgroups of the population with different levels of risk may not remain constant over the time period of interest. Violations of these assumptions can lead to spurious results. An additional problem is encountered when investigators study the occurrence of health events over a long period, i.e., the problem posed by migration. Migration tends to decrease the chance of detecting clustering; however, certain tests account

for non-uniformity of or changes in the population (24-26). As an alternative, adjustments for the size of the population at risk (to account for population changes during the study period) can be made before testing.

In addition to the techniques described in the Appendix, other approaches in use or under investigation for the analysis of clusters include the quality control measure known as the cumulative sum, or cusum, technique (27), the sets technique (28), nearest-neighbor procedures (29,30), and nonlinear and Bayesian time series methods. Normal-theory confidence intervals and bootstrap-prediction intervals for detecting frequencies of disease occurrence above those expected have been explored (31).

Because of the diverse and complicated nature of clusters, there is no omnibus test for assessing them. Investigators are advised to perform several related tests and to report the results that are most consistent with validated assumptions. This process will be aided by the use of CLUSTER, an IBM PC-compatible software program that will soon be commercially available and will offer investigators a choice of statistical procedures to use when investigating clusters.

References

1. Snow J. Snow on cholera. Hafner: New York, 1965.
2. Fraser DW, Tsai TR, Orenstein W, et al. Legionnaires' disease: description of an epidemic of pneumonia. *N Engl J Med* 1977;297:1189-97.
3. CDC. Pneumocystic pneumonia--Los Angeles. *MMWR* 1981;30:250-2.
4. Waxweiler RJ, Stringer W, Wagoner JK, et al. Neoplastic risk among workers exposed to vinyl chloride. *Ann N Y Acad Sci* 1976;271:40-8.
5. Cannon SB, Veazey JM Jr, Jackson RS, et al. Epidemic kepone poisoning in chemical workers. *Am J Epidemiol* 1978;107:529-37.
6. Baptiste MS, Rothenberg R, Nasca PC, et al. Health effects associated with exposure to radioactively contaminated gold rings. *J Am Acad Dermatol* 1984;10:1019-23.
7. Herbst AL, Ulfelder H, Poskanzer DC. Adenocarcinoma of the vagina: association of maternal stilbesterol therapy with tumor appearance in young females. *N Engl J Med* 1971;284:878-81.
8. McBride WG. Thalidomide and congenital abnormalities. *Lancet* 1961;2:1388.
9. Bender AP. Appropriate public health response to clusters: the art of always being wrong. Presented at National Conference on Clustering of Health Events, February 16-17, 1989, Atlanta, Ga.
10. Bender AP, Parker DL, Johnson RA, et al. Minnesota highway maintenance worker study: cancer mortality. *Am J Indus Med* 1989;15:545-56.
11. Devier JR. Development of a public health response to cancer clusters in Missouri. Presented at National Conference on Clustering of Health Events, February 16-17, 1989, Atlanta, Ga.
12. Fiore BJ. State Health Department response to disease cluster reports: a protocol for investigation. Presented at National Conference on Clustering of Health Events, February 16-17, 1989, Atlanta, Ga.
13. Caldwell GC. Twenty-two years of cancer cluster investigations at CDC. Presented at National Conference on Clustering of Health Events, February 16-17, 1989, Atlanta, Ga.
14. Bender AP, Williams AN, Spratka JM, et al. Usefulness of comprehensive feasibility studies in environmental epidemiologic investigations: a case study in Minnesota. *Am J Public Health* 1988;78:287-90.
15. Clapp RW, Wartenberg D, Cupples LA. Statistical methods for analyzing cancer clusters. *Am J Epidemiol* (in press).
16. CDC. Thallium poisoning: an epidemic of false positives--Georgetown, Guyana. *MMWR* 1987;36:481-2, 487-8.
17. Rothenberg RB, Steinberg KK, Thacker SB. The public health importance of clusters: a note from the Centers for Disease Control. *Am J Epidemiol* 1990; 132(suppl 1):S3-5.
18. Smith SJ. On understanding clusters--using the risk comparability chart. *Am J Epidemiol* (in press).
19. Slovik P.

Perception of risk. *Science* 1987; 236:280-5. 20. Greenberg MR, Wartenberg D. Understanding mass media coverage of disease clusters. *Am J Epidemiol* 1990;132(suppl 1):S192-5. 21. Black B. Clustering and tort law: matching scientific evidence with legal requirements. *Am J Epidemiol* (in press). 22. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res* 1967;27:209-20. 23. Klauber MR. Space-time clustering analysis: a prospectus. Proceedings of a SIMS Conference on Epidemiology, Alta, Utah, July 8-12, 1974. 24. Weinstock MA. A generalized scan statistic test for the detection of clusters. *Int J Epidemiol* 1981;10:289-93. 25. Whittemore AS, Friend N, Brown BW, Holly EA. A test to detect clusters of disease. *Biometrika* 1987;74:631-5. 26. Cuzick JC, Edwards R. Spatial clustering for inhomogeneous populations. Proceedings of the Joint Statistical Meetings, Washington, D.C., August 7, 1989. 27. Woodward RH, Goldsmith PL. Cumulative sum techniques. ICI Monograph No. 3, Oliver and Boyd. 28. Chen R. A surveillance system for congenital malformations. *Journal of the American Statistical Association* 1978;73:323-7. 29. Cliff AD, Ord JK. Spatial processes, models and application. London: Pion, 1981. 30. Diggle PJ. Statistical analysis of spatial patterns. London: Academic Press, 1983. 31. Stroup DF, Williamson GD, Herndon JL, Karon JM. Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Statistics in Medicine* 1989;8:323-9.

Disclaimer All MMWR HTML documents published before January 1993 are electronic conversions from ASCII text into HTML. This conversion may have resulted in character translation or format errors in the HTML version. Users should not rely on this HTML document, but are referred to the original *MMWR* paper copy for the official text, figures, and tables. An original paper copy of this issue can be obtained from the Superintendent of Documents, U.S. Government Printing Office (GPO), Washington, DC 20402-9371; telephone: (202) 512-1800. Contact GPO for current prices.

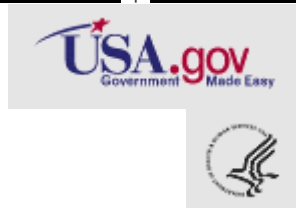
**Questions or messages regarding errors in formatting should be addressed to mmwrq@cdc.gov.

Page converted: 08/05/98

[HOME](#) | [ABOUT MMWR](#) | [MMWR SEARCH](#) | [DOWNLOADS](#) | [RSS](#) | [CONTACT](#)
[POLICY](#) | [DISCLAIMER](#) | [ACCESSIBILITY](#)

SAFER • HEALTHIER • PEOPLE™

Morbidity and Mortality Weekly Report
Centers for Disease Control and Prevention
1600 Clifton Rd, MailStop E-90, Atlanta, GA 30333,
U.S.A



[Department of Health
and Human Services](#)

This page last reviewed 5/2/01