

Investigating Suspected Cancer Clusters and Responding to Community Concerns

Guidelines from CDC and
the Council of State and Territorial Epidemiologists

Continuing Education Examination available at <http://www.cdc.gov/mmwr/cme/conted.html>.



U.S. Department of Health and Human Services
Centers for Disease Control and Prevention

CONTENTS

Introduction	1
Methods.....	2
Cancer Cluster Definitions, Characteristics, and Recent Investigations	3
Four-Step Process for Evaluating Suspected Clusters.....	5
Conclusion	13
APPENDIX A.....	15
APPENDIX B.....	19
APPENDIX C.....	22

Disclosure of Relationship

CDC, our planners, and our content experts disclose that they have no financial interests or other relationships with the manufacturers of commercial products, suppliers of commercial services, or commercial supporters. Presentations will not include any discussion of the unlabeled use of a product or a product under investigational use. CDC does not accept commercial support.

The *MMWR* series of publications is published by the Center for Surveillance, Epidemiology, and Laboratory Services (proposed), Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services, Atlanta, GA 30333.

Suggested Citation: Centers for Disease Control and Prevention. [Title]. *MMWR* 2013;62(No. RR-#):[inclusive page numbers].

Centers for Disease Control and Prevention

Thomas R. Frieden, MD, MPH, *Director*
 Harold W. Jaffe, MD, MA, *Associate Director for Science*
 Joanne Cono, MD, ScM, *Acting Director, Office of Science Quality*
 Chesley L. Richards, MD, MPH, *Deputy Director, Office of Public Health Scientific Services (proposed)*
 Pamela S. Diaz, MD, *Acting Director, Center for Surveillance, Epidemiology, and Laboratory Services (proposed)*

MMWR Editorial and Production Staff

Ronald L. Moolenaar, MD, MPH, <i>Editor, MMWR Series</i>	Martha F. Boyd, <i>Lead Visual Information Specialist</i>
Christine G. Casey, MD, <i>Deputy Editor, MMWR Series</i>	Maureen A. Leahy, Julia C. Martinroe,
Teresa F. Rutledge, <i>Managing Editor, MMWR Series</i>	Stephen R. Spriggs, Terraye M. Starr
David C. Johnson, <i>Lead Technical Writer-Editor</i>	<i>Visual Information Specialists</i>
Jeffrey D. Sokolow, MA, <i>Project Editor</i>	Quang M. Doan, MBA, Phyllis H. King
	<i>Information Technology Specialists</i>

MMWR Editorial Board

William L. Roper, MD, MPH, Chapel Hill, NC, <i>Chairman</i>	
Matthew L. Boulton, MD, MPH, Ann Arbor, MI	Timothy F. Jones, MD, Nashville, TN
Virginia A. Caine, MD, Indianapolis, IN	Rima F. Khabbaz, MD, Atlanta, GA
Barbara A. Ellis, PhD, MS, Atlanta, GA	Dennis G. Maki, MD, Madison, WI
Jonathan E. Fielding, MD, MPH, MBA, Los Angeles, CA	Patricia Quinlisk, MD, MPH, Des Moines, IA
David W. Fleming, MD, Seattle, WA	Patrick L. Remington, MD, MPH, Madison, WI
William E. Halperin, MD, DrPH, MPH, Newark, NJ	William Schaffner, MD, Nashville, TN
King K. Holmes, MD, PhD, Seattle, WA	

Investigating Suspected Cancer Clusters and Responding to Community Concerns

Guidelines from CDC and the Council of State and Territorial Epidemiologists

Prepared by

Beth (Vivi) Abrams, MPH,¹ Henry Anderson, MD,² Carina Blackmore, DVM, PhD,³ Frank J. Bove, ScD,⁴ Suzanne K. Condon, MSM,⁵ Christie R. Ehemann, PhD,⁶ Jerald Fagliano, PhD,⁷ Lorena Barck Haynes,⁸ Lauren S. Lewis, MD,¹ Jennifer Major, MPH,⁸ Michael A. McGeehin, PhD,⁹ Erin Simms, MPH,¹⁰ Kanta Sircar, PhD,¹ John Soler, MPH,¹¹ Martha Stanbury, MSPH,¹² Sharon M. Watkins, PhD,³ Daniel Wartenberg, PhD¹³
¹National Center for Environmental Health, CDC, Atlanta, Georgia; ²Wisconsin Department of Health Services, Madison, Wisconsin; ³Florida Department of Health, Tallahassee, Florida; ⁴Agency for Toxic Substances and Disease Registry, Atlanta, Georgia; ⁵Massachusetts Department of Public Health, Boston, Massachusetts; ⁶National Center for Chronic Disease Prevention and Health Promotion, CDC, Atlanta, Georgia; ⁷New Jersey Department of Health and Senior Services, Trenton, New Jersey; ⁸Haynes Ross Strategic, Seattle, Washington; ⁹Research Triangle Institute, North Carolina; ¹⁰Council of State and Territorial Epidemiologists, Atlanta, Georgia; ¹¹Minnesota Department of Health, St. Paul, Minnesota; ¹²Michigan Department of Community Health, Lansing, Michigan; ¹³Robert Wood Johnson Medical School, Piscataway, New Jersey

Summary

This report augments guidelines published in 1990 for investigating clusters of health events (CDC. Guidelines for investigating clusters of health events. MMWR 1990;39[No. RR-11]). The 1990 Guidelines considered any noninfectious disease cluster, injuries, birth defects, and previously unrecognized syndromes or illnesses. These new guidelines focus on cancer clusters. State and local health departments can use these guidelines to develop a systematic approach to responding to community concerns regarding cancer clusters. The guidelines are intended to apply to situations in which a health department responds to an inquiry about a suspected cancer cluster in a residential or community setting only. Occupational or medical treatment-related clusters are not included in this report. Since 1990, many improvements have occurred in data resources, investigative techniques, and analytic/statistical methods, and much has been learned from both large- and small-scale cancer cluster investigations. These improvements and lessons have informed these updated guidelines.

These guidelines utilize a four-step approach (initial response, assessment, major feasibility study, and etiologic investigation) as a tool for managing a reported cluster. Even if a cancer cluster is identified, there is no guarantee that a common cause or an environmental contaminant will be implicated. Identification of a common cause or an implicated contaminant might be an expected outcome for the concerned community. Therefore, during all parts of an inquiry, responders should be transparent, communicate clearly, and explain their decisions to the community.

Introduction

In 1990, CDC published guidelines for investigating clusters of health events (the 1990 Guidelines) (1). The 1990 Guidelines did not focus on any specific disease and considered any noninfectious diseases, injuries, birth defects, and previously unrecognized syndromes or illnesses. Many state, local, and tribal health departments have used the 1990 Guidelines as a basis for developing and implementing protocols to investigate suspected cancer clusters, employing the four-step approach (initial response, assessment, major

feasibility study, and etiologic investigation) identified in the 1990 Guidelines. Since the 1990 Guidelines were published, continued attention has been paid to suspected cancer clusters nationwide, leading CDC to publish additional details on the role of the guidelines in responding specifically to cancer clusters (2). Since 1990, many improvements have been made in the areas of data resources, investigative techniques, and analytic/statistical methods, and much has been learned from both large- and small-scale cancer cluster investigations.

This report augments the 1990 Guidelines by focusing specifically on cancer cluster investigations. The guidance provided in this report addresses additional subject areas that are deemed important by epidemiologists from state and local health departments (3). The additional subject areas include communications and resources for data and use of epidemiologic and spatial statistical methods. Useful websites, a resource not available in 1990, were added. The four-step process was retained, and more details were added.

The material in this report originated in the National Center for Environmental Health, Robin Ikeda, MD, Acting Director, and the Division of Environmental Hazards and Health Effects, Judith R. Qualters, PhD, Director.

Corresponding preparer: Kanta Sircar, Division of Environmental Hazards and Health Effects, National Center for Environmental Health, CDC. Telephone: 770-488-3384; E-mail: ddq0@cdc.gov.

Public health personnel in state and local health departments can use these guidelines to develop a systematic approach when responding to inquiries about suspected cancer clusters in residential or community settings. In addition, these guidelines might be helpful to a wider community of responders and epidemiologists who are concerned with such inquiries. These types of inquiries often are requested by community members or medical professionals concerned about what appears to be an unusually high number of diagnosed cases of cancer in a particular community, workplace, family, or school. Upon receiving an initial inquiry, health department personnel should respond rapidly to the caller's concerns, gather relevant information about the cancer cases, make a professional judgment on the likelihood that the reported situation could be an actual increase in cancer cases over those expected in a particular population, and determine whether further investigation is warranted. If appropriate, health department personnel then will need to provide resources for investigation of the suspected cluster, working with and involving members of the community as much as possible throughout the process.

Methods

In March 2010, the Council of State and Territorial Epidemiologists (CSTE) and CDC convened a workgroup (the authors of this report) to revise the 1990 Guidelines. The group comprised public health professionals selected by the leadership of CSTE's Environmental Epidemiology Subcommittee and by CDC's National Center for Environmental Health's (NCEH) Division of Environmental Hazards and Health Effects (EHHE). CSTE and CDC selected workgroup members with experience in responding to cancer cluster inquiries from communities and managing of cancer cluster investigations. Representatives included epidemiologists from state health departments who were selected in order to have input from states that represent a range of approaches to and capacities for cancer cluster investigations. In addition, CDC workgroup members included representatives from CDC organizations typically called upon to assist in cancer cluster investigations: NCEH/EHHE, the Agency for Toxic Substances and Disease Registry (ATSDR), and the National Center for Chronic Disease Prevention and Health Promotion's Division of Cancer Prevention and Control. CDC risk communications and statistical specialists, as well as epidemiologists at academic institutions experienced in cancer cluster investigations, participated in the workgroup.

The intent of the workgroup was to ensure a practical approach to the assessment, analysis, and investigation of response to cancer cluster concerns. Through regularly

scheduled conference calls and meetings from March 2010 to May 2011, the workgroup identified areas that warranted change from the 1990 Guidelines and sources of new information to incorporate in the revision of the guidelines. For these topics, the medical librarians at the CDC Public Health Library and Information Center conducted a comprehensive review of the published, peer-reviewed literature. To identify articles related to community cancer clusters, librarians conducted a structured literature search using multiple databases including PubMed (National Library of Medicine, National Institute of Health, Bethesda, Maryland, available at <http://www.ncbi.nlm.nih.gov/PubMed>), MEDLINE (available at <http://www.nlm.nih.gov/bsd/pmresources.html>), and CAB (available at <http://cabdirect.org>). English language peer-reviewed articles published between 1969 and 2010 were searched by using the following medical subject heading (MeSH) terms: "cluster analysis," "cancer cluster," "neoplasm," "environmental illness," and "not occupational diseases." Through this process, 166 articles were identified. In addition, members of the workgroup recommended 26 publications, including publications on communications and statistical analysis as well as nonscientific publications related to cancer clusters, and three unpublished cancer cluster investigation reports that were relevant to topics addressed in the guidelines. All articles and reports were reviewed by the workgroup members. Regarding topics on which no new published evidence was available, expert opinion was sought from workgroup members. In October 2010, an in-person meeting of the workgroup was held to begin writing these guidelines.

In addition to convening a technical workgroup, CSTE sent a survey to all state and territorial epidemiologists to assess the needs of public health professionals when responding to cancer cluster concerns in order to direct the focus and content of the guidelines (3). The survey included questions about the most common activities which states engage in when addressing a cancer cluster inquiry and what type of information would be useful. This survey identified areas (e.g., communications, resources for data, and epidemiologic methods) in which more details would be useful. After discussion, review, and incorporation of the findings from the survey, the workgroup decided to retain and update the four-step approach first described in the 1990 Guidelines. Updates included incorporating new technological advances (e.g., use of the Internet and websites) for information on relevant data resources, statistical tests, and mapping techniques as well as lessons learned from recent cancer cluster investigations. One important update is the emphasis on the importance of developing a robust working relationship with the community as soon as possible, including clear two-way communication

and transparency in all aspects of the response process, while maintaining scientific rigor.

The revised guidelines address questions about the availability of data, limitations associated with understanding cancer clusters, and decision-making about the extent to which inquiries can be followed up. For specificity, the revised guidelines are limited in scope to include only those situations in which a health department responds to an inquiry about a suspected cancer cluster in a residential or community setting. These guidelines do not address workplace cancer clusters or those related to medical treatment (e.g., cancers associated with pharmaceuticals). Workplace or occupational clusters and medically related clusters each present unique sets of circumstances, have unique and clearly defined populations at risk, and generally call for specific investigative methods, agencies, and partnerships (4,5). Similarly, these guidelines do not discuss diseases other than cancer that persons might suspect have occurred in clusters in their communities. However, some of the principles of risk communication, data analysis, and community involvement discussed in this report might be applicable to noncancer cluster investigations as well. Finally, the revised guidelines do not address routine surveillance conducted by cancer registries and programs to assess trends.

This report is divided into two sections and three appendices:

- The first section explains cancer cluster definitions, characteristics and lessons learned from recent investigations;
- The second section outlines a systematic, four-step process for evaluating potential cancer clusters;
- Appendix A provides an overview of sources of data and other resources useful for cancer cluster investigations;
- Appendix B describes considerations for developing effective communication strategies; and
- Appendix C highlights some useful statistical and epidemiologic approaches for investigating suspected cancer clusters.

Cancer Cluster Definitions, Characteristics, and Recent Investigations

Definition of a Cluster

CDC defines a cancer cluster as a greater than expected number of cancer cases that occurs within a group of people in a geographic area over a defined period of time (6). This definition can be broken down as follows:

- *a greater than expected number*: Whether the number of observed cases is greater than one typically would observe

in a similar setting (e.g., in a cohort of a similar population size and within demographic characteristics) depends on a comparison with the incidence of cancer cases seen normally in the population at issue or in a similar community.

- *of cancer cases*: The cancer cases are all of the same type. In rare situations, multiple cancer types may be considered when a known exposure (e.g., radiation or a specific chemical) is linked to more than one cancer type or when more than one contaminant or exposure type has been identified.
- *that occurs within a group of people*: The population in which the cancer cases are occurring is defined by its demographic factors (e.g., race/ethnicity, age, and sex).
- *in a geographic area*: The geographic boundaries drawn for inclusion of cancer cases and for calculating the expected rate of cancer diagnoses from available data are defined carefully. It is possible to “create” or “obscure” a cluster inadvertently by selection of a specific area.
- *over a period of time*: The time period chosen for analysis will affect both the total cases observed and the calculation of the expected incidence of cancer in the population.

When a health agency is investigating a suspected cancer cluster, it can use these parameters to help determine whether the reported cancer cases represent an increase in the ratio of observed to expected cases. The health agency also can use the parameters to identify characteristics that indicate whether cases might be related to each other and to determine whether the cases warrant further investigation. In the sections that follow, guidelines are provided to outline how to make this determination, including the appropriate information to collect, the necessary deliberations, the factors to take into account, and the analyses to perform.

Characteristics of Cancer and Clusters

The National Cancer Institute (NCI) of the National Institutes of Health (NIH) defines cancer as a term for a group of diseases in which abnormal cells divide without control and can invade nearby tissues (7). As a group, cancers are very common. Cancers are the second leading cause of death in the United States, exceeded only by diseases of the heart and circulatory system (8). One of every four deaths in the United States is attributable to some form of cancer. In 2009, approximately 1.47 million persons in the United States received a cancer diagnosis, and approximately 568,000 persons died from cancer (9).

Because cancer is common, cases might appear to occur with alarming frequency within a community even when the number of cases is within the expected rate for the population. As the U.S. population ages, and as cancer survival rates continue to improve, in any given community, many residents will have had some type of cancer, thus adding to the perception of an

excess of cancer cases in a community. Multiple factors affect the likelihood of developing cancer, including age, genetic factors, and such lifestyle behaviors as diet and smoking. Also, a statistically significant excess of cancer cases can occur within a given population without a discernible cause and might be a chance occurrence (10,11).

Three considerations are important for suspected cancer cluster investigations. First, types of cancers vary in etiologies, predisposing factors, target organs, and rates of occurrence. Second, cancers often are caused by a combination of factors that interact in ways that are not fully understood. Finally, for the majority of cancers, the long latency period (i.e., the time between exposure to a causal agent and the first appearance of symptoms and signs) complicates any attempt to associate cancers occurring at a given time in a community with local environmental contamination. Often decades intervene between the exposures that initiate and promote the cancer process and the development of clinically detectable disease (12).

Communicating effectively about the frequency and nature of cancer in explaining suspected cancer clusters can be difficult for public health agencies, and many of the scientific concepts involved (e.g., random fluctuation, statistical significance and latency period) might not be easy to explain to the community (13). Any number of community members, friends, or relatives with cancer is alarming and is too many from a personal perspective (11). When persons are affected personally by a case of cancer, they naturally seek an explanation of the cause of the cancer (13).

Cancer Cluster Investigations

As the 1990 Guidelines noted, finding a causal association between environmental contaminants and cancer is rare in a community cancer cluster setting (1). Evidence reported by state and local health agencies and federal agencies since 1990 that would suggest otherwise is limited, and most investigations of suspected cancer clusters do not lead to the identification of an associated environmental contaminant (10).

State and local health agencies receive approximately 1,000 inquiries per year regarding suspected cancer clusters (14). The majority of these inquiries can be resolved during the initial response, which consists of the initial contact and follow-up contact with the caller, if needed. The resulting health education can be an important public service (14). Even if inquiries concern events that meet the statistical criteria for a cancer cluster, investigations of suspected cancer clusters are unlikely to find an associated environmental contaminant (1,11). For example, one of the largest suspected cancer clusters investigated by CDC's NCEH and by other agencies concerned cases of childhood leukemia in Fallon, Nevada. Although initial analysis demonstrated a statistically significant ($p < 0.05$)

increase in the number of cases, subsequent epidemiologic investigations did not identify a statistically significant association with environmental contaminants (15).

Suspected cancer clusters that consist of cases of one type of cancer, a rare type of cancer, or a type not identified usually in a certain demographic group are thought to be more likely to have a common cause (10). Even if these factors are present, the suspected cluster might not be associated with an environmental exposure and in fact might be a chance occurrence. A type of cancer under investigation might not be associated biologically with any environmental contaminants of concern in the community. In other words, a suspected environmental contaminant might not be in the causal pathway for a certain type of cancer. One common but false assumption held by persons not familiar with the scientific study of cancer is that a single environmental contaminant is likely to cause any or all kinds of cancer. Toxicologic and epidemiologic studies do not support this assumption. Cancer is not one disease, but rather many different diseases with different causal mechanisms (16).

In addition, two statistical issues influence the ability of the health agency to determine an association between the cancer(s) in question and environmental exposures. First, a suspected cancer cluster investigation with a small number of cases (e.g., one that involves a rare cancer type comprising only a few cases) might result in a lack of statistical power to detect an association. Second, because of the substantial number of cancer patients who might live in a community, a spurious association with an environmental contaminant can occur by chance alone, without the contaminant being a causal factor (17).

The health agency should avoid imprecise and post hoc definitions of such concepts as case, population, geographic area, or exposure period because such definitions might bias or limit an investigation. For example, case definitions that include different cancers generally are not useful, unless the environmental contaminant under consideration has been associated with multiple cancer types.

Latency and change of residence add to the complexity of these investigations. Because of the long latency period associated with cancers, behaviors and exposures that might have contributed to the development of cancer in a person typically occur years to decades before the diagnosis (e.g., malignant mesothelioma, a lung tumor, is associated with asbestos exposure). The latent period between first exposure to asbestos and death from mesothelioma is often 30 years or longer (18). Latency needs to be considered in an investigation of a suspected cancer cluster because it influences the exposure period relevant to the investigation. If a person with cancer did not live in the suspected cancer cluster area during the relevant exposure period (possibly 20 years previously), then

that person's cancer cannot be related to an environmental contaminant of concern or to any exposure in the suspected cancer cluster area. Conversely, the latency period might limit the ability to detect a cancer cluster or identify cancers related to an environmental exposure that occurred in the past. In a mobile population, a cancer cluster resulting from an environmental contamination occurring years or even decades earlier might go undetected because exposed residents have moved away from the community before the cancer develops. Thus, as persons move in and out of different communities, their cumulative exposure profile will change.

Suspected childhood cancer cluster investigations have the same limitations as adult counterparts (19). However, because childhood cancers generally have shorter latency periods, changes of residence might be less of an issue in the investigation of suspected clusters involving childhood cancers (20). For example, in one California study of 380 children with a diagnosis of leukemia, approximately 65% of the study participants changed residence between birth and diagnosis (21), indicating that even among cancers with short latency periods, migration might be an important factor.

Because investigations rarely demonstrate a clear association with an environmental contaminant, investigations of community-based cancer clusters usually do not provide the resolution communities seek (11). Furthermore, a suspected cancer cluster investigation can have unintended consequences. An investigation can augment the existing fear and uncertainty in the community brought on by the perception that a suspected cancer cluster exists, which might have a negative social and economic impact (22). Therefore, during all stages of an inquiry or investigation, responders should not only be transparent and receive community input but also explain their decisions to the community.

Four-Step Process for Evaluating Suspected Clusters

Because major investigations require substantial resources and might not identify the cause of cancer cases, a stepwise approach is recommended. Both the likelihood of identifying a causal factor and the feasibility of studying the relationship should be considered before proceeding to the next step. Regardless of the extent of the investigational response, the process of responding to community concerns provides opportunities to increase communities' knowledge about cancer and to encourage participation in cancer screenings and healthy behaviors. For this reason, education and consultation are advised at all steps of the process.

Four steps are recommended to respond to a report of a suspected cancer cluster, including procedures, guidance on, and considerations for closing the inquiry or proceeding to the next step:

- Step 1: Initial contact and response
- Step 2: Assessment
- Step 3: Determining the feasibility of conducting an epidemiologic study
- Step 4: Conducting an epidemiologic study to assess the association between cancers and environmental causes.

These steps update the four-stage process discussed in the 1990 Guidelines but should be implemented with two qualifiers. First, the extent to which a health agency is able to follow these guidelines depends on existing resources and infrastructure. Second, the delineation between the steps is not necessarily fixed. Often, a health agency might choose to combine steps or to pursue a problem with several approaches. The four-step process is intended to be flexible, so that health agencies and their partners may use it as model guidelines and adapt it to their own existing protocols, resources, staffing, organizational systems, and policies.

Step 1. Initial Contact and Response

Description

The purpose of Step 1 is to collect information from the inquirer (i.e., the person calling, writing, or emailing the report of a suspected cancer cluster) so as to determine whether the inquirer's concern warrants further follow up. On the basis of the information collected, the health agency will need to decide whether to pursue the inquiry further. This step focuses on obtaining and evaluating whatever information the inquirer can provide as well as relevant data available to the health agency (e.g., data from cancer registry, census, and environmental databases [see Appendix A]).

The inquirer should be referred quickly to the responsible unit in the health agency, and the problem should not be dismissed prematurely (i.e., before information is collected). Although the majority of reports of potential clusters will be closed at the time of initial response because the inquirer's concerns are not consistent with a potential cancer cluster, the first encounter is often the health agency's best opportunity to educate the inquirer about the nature of cancer and suspected cancer clusters.

To be an effective initial responder and a successful manager of reports of a local suspected cancer cluster, the health agency needs to understand the context of the inquirer's concern, the nature of the perceived problem, the history of how it has or has not been reported to authorities, and if applicable

how authorities have responded to date. In addition, other necessary background information should be gathered including demographic characteristics (e.g., age group) of the persons with cancer and the population group of which they are members. Not only is this information essential to the scientific investigation, but it is also important for effective information-gathering, communication and coordination with the inquirer or community. In addition, it is essential for the effective management of a suspected cancer cluster inquiry to be open, transparent, and thorough with respect to the evaluation of information and actions taken. It is also important to be sensitive and responsive to the inquirer's concerns.

Procedures

- The health agency responder (the responder) should be empathetic, listen to the inquirer's concerns, and record the information received.
- The responder should gather identifying information on the inquirer: name, address, telephone number, length of residence at current location, and organization affiliation, if any. However, the responder should comply with requests for anonymity and explain that the inability to follow up with the caller might hinder further investigation.
- The responder should gather initial data on the potential cluster from the inquirer: types of cancer and number of cases of each type, age of people with cancer, geographic area of concern, period over which cancers were diagnosed, and how the person reporting learned about the supposed cluster. Keep in mind that the inquirer might not know the true primary cancer diagnoses and will most likely not be aware of all cases of cancer in this area or during the period of concern.
- The responder should gather information from the inquirer about any specific environmental hazards or concerns, other risk factors (e.g., diet, infections, and family history) and other concerns in the affected area (e.g., the likely period of environmental contaminant exposures).
- On the basis of the information presented, the responder should make an initial judgment about the advisability of the health agency's pursuing an inquiry into the suspected cancer cluster. The decision might require discussions with other people in the health agency.
 - Multiple factors bear on this decision, but it is primarily based on whether the evidence as presented fits the definition of a cluster and the biologic plausibility that the cancers could share a common etiology. Such factors as reports involving a rare cancer or an atypical demographic distribution of a certain type of cancer (e.g., multiple cases of breast cancer in men) support

the decision to investigate further and should be considered. If exposure to a specific environmental contaminant is a concern in the community, the consensus in the scientific literature regarding an association between the environmental contaminant and the cancer of concern should be considered.

- Factors that do not support the need for further investigation include:
 - cancer cases within family members who are linked genetically (especially cancers known to be strongly genetically related);
 - reported disease that might not be cancer;
 - different types of cancers not known to be related to one another;
 - a few cases of very common cancers (e.g., breast, lung);
 - cancer cases among persons who did not live in the same geographic location during the relevant timeframe based on latency, and thus could not have experienced a common carcinogenic exposure; and
 - the lack of a plausible environmental cause.
- The responder should clearly and accurately explain the rationale used to determine if an investigation will be pursued based on the information provided about the cases as well as the health agency's procedures. For example, the rationale for not pursuing an investigation could be that the reported cancers are unlikely to be related to plausible environmental exposure.
- If an inquirer is reporting an event that is not a suspected cancer cluster but rather one involving a known or possible environmental contamination, the caller should be referred to the appropriate environmental resources agency. The responder should work with the health agency's communication experts to assess the potential community concern and impact, and prepare a plan to address concerns.
 - The health agency should provide responders with talking points about the nature of cancer, its frequency and occurrence, how different types of cancers reported are related to separate causes, that rates of disease do somewhat increase and decrease in a population over time (random fluctuations), and so forth. These points can be used to educate inquirers about cancer and to provide them with further resources that address their concerns.
 - If the information provided supports the decision to investigate the cancer concerns further, the health agency responder should notify the inquirer, explain what that entails and outline how the agency will follow-up with the inquirer and provide results. The

responder should ask the inquirer if there are others in the community (e.g., other residents with this cancer type) who would like to have a report on the results of the next step.

Recommendations for Step 1

- The health agency responder should have expertise or training in cancer and/or environmental epidemiology.
- The responder should have training and experience in risk communication because, understandably, community residents can be extremely distressed by the perception of an excess amount of cancer in their community (22). The ability to make a judgment on the facts presented and to communicate the factors in that judgment clearly depends on having both scientific expertise and experience in communication.
- The responder should be knowledgeable about cancer, cancer prevention, and guidelines on investigating suspect cancer clusters. The responder also should be able to offer the inquirer easily accessible resources, such as the health agency's or CDC's cancer website (<http://www.cdc.gov/cancer>).
- If one person in the health agency with comprehensive expertise (i.e., in all areas described) is not available, the responder should collect initial information and tell the inquirer to expect a follow-up call. The responder should then discuss the case with colleagues who have the necessary expertise before responding to the caller with an initial judgment.
- The health agency and responder can access, at minimum, county-level cancer statistics from the state cancer registry to understand and explain the reported cases in an appropriate context. A list of state cancer registries is available at <http://www.cdc.gov/cancer/npcr>.
- If possible, the responder should be, or become, relatively familiar with the geographic area of concern, its demographic profile, and its history (e.g., industrial and residential development) in order to understand the health and environmental concerns of the community.

Decision to Close the Investigation at Step 1

A decision at Step 1 not to pursue an investigation is based on the determination that the reported cases are unlikely to comprise a cancer cluster; therefore conducting a statistical assessment to determine whether an excess of cancer cases exists might be unsuccessful because the cancers are not likely to share a common, environmental etiology. This determination might involve multiple communications with the inquirer, as well as additional data-gathering. If the inquirer acknowledges and is satisfied with the decision not to move forward, the inquiry can be closed at this point. If the inquirer is not satisfied with

the decision and the verbal explanation, then the health agency should consider providing a written explanation and include resources related to the decision. Regardless of the decision, the health agency should document in a permanent log all information about the inquiry and the decision.

Decision to Continue to Step 2

The data gathered at this point might suggest the need for further evaluation. If so, the health agency should elect to proceed to Step 2 to determine whether an excess of cancer cases exists.

Step 2. Assessment

Description

The primary purpose of Step 2 is to determine whether the suspected cancer cluster is a statistically significant excess. Several components of the follow-up investigation are necessary to determine if an excess of cancer cases exists in the community. These important components include the study design, as well as the collection, analysis, and interpretation of relevant data. Decisions must be made concerning the case definition, how the population of concern (the study population) is defined, the choice of comparison cancer rates, and the choice of statistical methods. To address these components, the health agency investigation team (the investigators) leading the follow-up investigation of Step 2 (and subsequent steps) will need to have epidemiologic expertise or collaborate with an epidemiologist. The time needed to complete Step 2 varies, depending on the complexity of the suspected cancer cluster.

This step also includes identification of local environmental concerns. Depending on the circumstances, communicating with partners and identifying and communicating with key community members about the assessment might be appropriate as a part of this step. Creating a comprehensive communication plan is important, in order to identify audiences, communication needs, and communication channels. More detailed information is provided elsewhere (see Appendix B).

Calculating a standardized incidence ratio (SIR) (23,24) is recommended at this step. The SIR is generally calculated to provide an estimate of the likelihood that an excess of cases exists in the population of concern (the study population) compared to the general or reference population. The SIR is a ratio of the number of observed cases to the number of expected cases. The observed cases are the cases that actually occurred in the study population within a specific timeframe. The expected number of cases is the number that would occur in the study population if the occurrence of cancer in that population was the same as the reference population. Since

cancer rates vary with age, the expected number takes into account the age distribution in the study population. It is calculated by multiplying the age specific cancer incidence rates of the reference population by the corresponding age specific group in the study population. In the calculation of the SIR, factors that must be considered include:

- the type(s) of cancer and number of cases,
- the period of concern,
- the geographic area of concern,
- the background cancer incidence in the larger reference population (available through the cancer registry), and
- the demographic characteristics of the cases and the reference population.

More detailed information is provided (see Appendix C).

Procedures

- The investigators should define the study population, by demographic characteristics, geographical area and time period of concern. These factors, in addition to cancer type, are also included in the case definition:
 - Demographic characteristics might include age, sex, race/ethnicity, or residential location. The study population could be all-inclusive, or it could be limited to a specific demographic group. For example, the study population could include females only, adults only, or children only.
 - The appropriate geographic area (study area) and time period need to be selected. Privacy issues should be considered when collecting, analyzing, and presenting data on a few cases in a small geographic area. Statistical analysis of neighborhood level data or data from sparsely populated areas might not be possible because of limited numbers. Limited numbers might lead to a lack of statistical power and therefore to an instability of rates. Decisions about timeframe and geographic boundaries should take into consideration the concerns of the caller or community, as well as any known or suspected environmental contamination and pathways of contamination (19).
 - The case definition includes information on the type of cancer (e.g., primary site, histology, and grade). Cancer registries collect cancer diagnoses based on the ICD-O codes (International Classification of Diseases for Oncology, 3rd Edition [ICD-O-3]) (25). Cases might be limited to a specific age or sex (e.g., limiting small cell lung cancer to women because the hypothesis is that there is an increase in the number of cases in women). Cases of cancer among the study population generally are identified from a state's cancer registry, using the case definition.
- An all-cancer SIR (i.e., one calculated for all types of cancers combined) might be useful for communication and educational purposes, but it is not useful for explaining or exploring potential etiologies. If an all-cancer SIR is presented with the results, a discussion of its limitations for investigating etiologies and its usefulness for cancer education should be included.
- The case definition, study population, study area, and period of interest will require justification. The definitions and the justification should be transparent to the community so that they understand the rationale behind the approach taken. This means sharing information that is consistent, timely, and expressed in a manner that the lay public is able to understand. Otherwise, these decisions might be seen as arbitrary and thus be rejected by the community.
- The investigators should calculate incidence rates, the SIR, and 95% confidence interval (CI) and other descriptive statistics. Procedural steps are discussed in detail elsewhere (see Appendix C).
 - An SIR of >1.0 indicates that the observed number of cases is greater than the number that would be expected for the population. The SIR increases when the number of observed cases in excess of the number of expected cases increases.
 - The CI is an indication of the statistical precision of the SIR value.
 - In addition to whether the SIR is statistically significant, the investigators should consider the suspected cluster in the context of the plausibility that the cancers could share a common etiology based on the latency, on community patterns of migration in and out, known risk factors for the cancer of concern, and the potential for exposure to a contaminant of concern, as well as other factors (see “Decision to close the investigation at Step 2”).
- The investigator should understand community concerns and identify facts about local environmental factors by:
 - reviewing the literature on risk factors for the types of cancers in question, investigating both the human and animal studies using PubMed and/or other sources (e.g., <http://www.ncbi.nlm.nih.gov/pubmed>);
 - reviewing literature on possible associations between the types of cancers and known or suspected environmental exposures because results from the literature review might affect the case definition (e.g., the types of cancers considered for study); and
 - identifying whether there is concern in the community about known or suspected environmental exposures or other factors that the community suspects are

related to health problems. By using the community members' "local knowledge" (i.e., understanding of the community, its history, and its members as distinct from the scientific/technical expertise that is provided by health agencies) about the hazards and risk factors in their community, as well as data from environmental and other databases, investigators can make more informed decisions.

- The investigator should communicate with the inquirer and the community as indicated (see Appendix B). In many cases, communication about the activities in this step is only with the inquirer. However, if communitywide concerns exist about the cancer cases or environmental conditions, early involvement of other community members might be appropriate.
 - The investigator should share the SIR calculation with the inquirer and other community residents and describe the process, the results, and the implications of the results.
 - The investigator should consider who else should be notified after the SIR determination that there is, or is not, an excess number of cases (e.g., the local health agency, the state environmental agency, community residents).

Recommendations for Step 2

- Because of the variety of issues involved in this phase of the investigation, a team approach involving epidemiologists, toxicologists, communicators, and other experts might be necessary.
- Health agencies should document all decisions, communications, and processes.
- It might also be useful to examine the trend of a cancer type that is documented to be completely unrelated to the cancer type and/or exposure of concern. The purpose of this examination is to identify other factors that might affect trends or excess cancer cases detected. If all cancers appear elevated or depressed in a similar time frame (including those that are not related), other factors ought to be considered. These factors include the possibility that the estimated denominator might be incorrect or that the community has an unusually high proportion of persons with high-risk health behaviors (e.g., smoking).

Decision to Close the Investigation at Step 2

The decision to close the inquiry at this step or to move forward to Step 3 is based on multiple factors. The decision to move forward is best made on the basis of a review of the statistical analysis as well as an understanding of the scientific facts presented. To interpret the SIR, the health agency must answer these questions:

- Are there enough cases and a large enough population for statistical stability (17,23,26)? In general, the population size of a typical census tract (27) is the smallest denominator that will allow reliable results to be generated.
- If there is a large enough numerator for statistical stability, how likely is it that this SIR might have occurred by chance, assuming that the underlying incidence rates were not elevated (for example, does the 95% CI exclude 1.0)?
- Are there environmental contaminants and/or events that could be related to the cases?
- Are there any population-related issues (e.g., a substantial number of persons moving into the community) that might in part explain the observed cancer excess?

Other information in addition to the SIR is required to allow estimation of the likelihood that the observed cancers represent an excess, could potentially be related to one another, and share a common etiology. These questions include the following:

- Has there been an increase in the incidence rate of the specific cancer over time?
- How many more observed cases are there than expected (the number of excess cases)?
- Are the demographic characteristics of these cases unusual for the type of cancer (e.g., in a younger age group for a cancer [such as lung cancer] that usually occurs only in older age groups)?

The investigator needs the complete picture in order to determine the likelihood that the observed cancers represent an actual excess, could potentially be related to one another, and share a common etiology. An SIR of limited magnitude that is not statistically significant, coupled with a lack of known association with an environmental contaminant and no trend of increasing incidence over time, justifies closing the inquiry at Step 2. However, a statistically significant SIR of great magnitude and an increasing trend in incidence rate, together with a known environmental contaminant would argue for continuing to Step 3.

The following examples illustrate how these data can be synthesized. For example, an SIR of <2.0 with CIs surrounding or overlapping 1.0 and/or a small number of cases (e.g., <10), coupled with limited evidence of an etiologic relationship between the type of cancer and the suspected environmental contaminant, might justify a decision to close the inquiry at Step 2. As another example, an SIR of >4 with CIs that do not overlap 1.0, and ≥10 cases that might be etiologically linked, should encourage advancing to Step 3. Moderate elevations in a SIR, involving small numbers of cases, and instances of rare cancers pose the most difficult decisions for health agencies. Additional information might be needed to assist in the decision to continue the investigation to the next step.

Once it is decided to close the inquiry at Step 2, it is important to respond to the caller in writing, explaining the process and results, including the determination that the cases likely do not comprise a cancer cluster. The inquiry should then be closed, with appropriate documentation in system logs.

Even if continuing with a cluster investigation is not indicated, the inquiry might have raised other concerns, including known or suspected environmental contamination. In that case, the health agency should work with partners to facilitate other public health actions or interventions, as warranted (e.g., health screening, health risk assessments, or education on cancer prevention). In these circumstances, it is important to communicate clearly with the community about the scientific basis for the actions, being careful to set realistic expectations for the community.

Some scientific experts have recommended implementing guidelines that use resources for cancer education and larger, long-term population based studies to determine risk factors for cancer rather than proceeding beyond Step 2 into cluster studies. This is because cluster studies almost never yield definitive answers regarding the cause of any specific cluster (28,29). Each health agency makes a decision as to what resources are available.

Decision to Continue to Step 3

Step 3 consists of gathering more information to assess the feasibility of conducting an epidemiologic study to determine whether the cases are associated with a common etiological factor. This process will engage additional resources and be more visible to the community. If a decision is made to move forward, the health agency should provide a written report to the caller, as well as to any partners contacted. This report should include a description of the results of the preliminary analyses and circumstances, carefully articulating what is known and not known at this point. Finally, the report should describe the health agency's plan (i.e., next steps).

Step 3. Determining Feasibility of Conducting an Epidemiologic Study

Description

The purpose of Step 3 is to assess the feasibility of performing an epidemiologic study to examine the association between the cancer cluster and a particular environmental contaminant. If further study is feasible, an outcome of this step should include a recommended study design.

All activities in this step should be carried out in collaboration with community, environmental, and other partners. Decisions should reflect the concerns, interests, and expertise of all

partners. The health agency should follow the communication plan created in the previous step. This communication plan needs to be tailored to the community, and it should proactively address the information needs of stakeholders. It may be adapted as needed.

Additionally, this step provides the opportunity to evaluate additional public health actions, such as smoking cessation programs, cancer screenings, health risk assessments, removal of environmental hazards, or other activities that should be conducted. If beneficial to public health, these actions should not be delayed pending the decision to conduct or complete an epidemiologic study focused on assessing the association between the cluster of cases and a suspected environmental cause.

Procedures

- The first actions in determining the feasibility of further study of the identified cluster include determining the study hypothesis and reviewing the scientific literature and past health agency reports.
 - The investigators should identify hypotheses. Past agency reports and logs should be reviewed to determine whether the same type of cancer has led to other inquiries and investigations. On the basis of preliminary findings, another literature search might be conducted (librarians at a public health library can assist with literature searches). A list of helpful resources is available at <http://phpartners.org>. The hypotheses should focus on known causes of the cancer in question.
 - The investigator should involve environmental, community, local public health, and other partners. The investigator should learn what the community needs and wants and assess the feasibility of addressing these concerns. The investigator should ask if partners know of any possible environmental risk factors in the area. The investigator also should discuss with the community whether there are issues of cultural sensitivity to consider in planning an investigation. A community panel is one method to involve the community. The investigators should define the panel's specific purpose and maintain open and frequent communication with the members. Advantages of a community panel also include gaining valuable information and diverse perspectives from members. However, it might be difficult to achieve consensus with such diverse membership. Finally, investigators should share information about time, cost, goals, purpose, and limitations of a potential study, with all partners and carefully communicate realistic expectations.

- The health agency and its investigators should consider establishing an expert advisory panel to assess potential study design issues as sample size, a small case number, and study power. Health agencies should consider whether the panel contains internal experts from within the health agency or needs help from state government or external experts. Experienced scientists with appropriate skills should be included. The experts should include an epidemiologist, a toxicologist, a physician, an environmental protection specialist, and a community-nominated expert and/or local representative to provide advice on the assessment as needed. The panel's specific role, objectives, and deliverables will need to be defined.
- The investigators should identify such parameters as study population and its characteristics, including what descriptive, health, and risk factor data should be collected and determine the feasibility of obtaining that data. Investigators should:
 - confirm case diagnoses and determine which types of cancer and which cases meet the case definition,
 - identify a comparison group that depending on the study design does not have the cancer of concern (i.e. a control group in a case-control study) or does not have the exposure of concern (i.e. unexposed group in a cohort study),
 - consider the feasibility of obtaining data on individuals in the comparison group, and
 - explore the willingness of persons to participate in interviews or studies for gathering data on health, possible exposures, the amount of time the affected persons have lived in the area, and occupation. Investigators should also consider the ability to assess other relevant risk factors and confounding variables.
- The investigators should ascertain the plausibility that the cases and contaminants could potentially be associated. Specifically, they should:
 - verify whether the environmental contaminants of concern are known carcinogens,
 - consider possible and plausible routes of exposure to affected persons;
 - ask whether persons with cancer actually were exposed to an environmental contaminant in sufficient doses and for a sufficient time to make the association biologically plausible. On the basis of the consensus of the published literature, agencies should determine whether the exposure dose and duration were sufficient to have caused cancer, and they should ask whether the time sequence of exposure is consistent with the latency period and the causation of these particular cancers;
 - consider the possibility that historical records of chemical use or contamination at the particular location might be more important than more recent environmental testing that might not reflect past contamination;
 - determine whether residential and occupational histories for affected persons are obtainable. Information on a person's residential and occupational histories is of critical importance in constructing a complete exposure history. Investigators also should consider whether the person's occupation has a bearing on the total exposure.
- The investigators should identify the available data on the environmental contaminant of concern.
 - Environmental data can be helpful in cancer cluster investigations as part of the determination of the possibility of exposure to an environmental contaminant. However, it is generally not recommended to engage in a general, open-ended inquiry to identify potential contaminants in a community, in the absence of a suspected etiologic agent. Additional environmental testing should be carried out only when there is a clear scientific rationale and all the factors discussed below are considered.
 - Investigators should determine whether information on previous exposures is available. Because of the long latency of cancer, an historical exposure assessment might be more important than consideration of current exposures. Some exposures might have occurred >20 years previously. The suspected environmental contamination exposure period might span decades, and it might have changed during that time. In light of these factors, current environmental testing rarely provides accurate data on historical exposure.
 - Investigators should determine whether they can characterize exposure to suspected environmental hazards accurately at the individual level and in a way that reflects the period of concern.
 - Investigators should identify known current, ongoing, and historical environmental concerns in the community. The process of collecting such concerns also can have the benefit of enhancing their reliability and credibility with the community.
 - Investigators should partner with state and regional environmental protection specialists and toxicologists to secure aid in reviewing and interpreting historical exposure assessments. The investigator and subject matter experts should review historical environmental

sampling data to determine whether there are any known or suspected area environmental contaminants that could potentially be related to cancer cases or are known or suspected carcinogens and whether there are any known or suspected exposure pathway(s).

- Health agencies and their investigators should communicate with the public. At this point, the finding of an elevated rate of cancer in the community will likely be public knowledge, and the community might expect extensive environmental testing. There might also be pressure from the media or elected officials to perform new testing. Clear communication is critical to explaining why environmental testing is not feasible or not appropriate.
- Investigators should assess information on possible exposures, not only for informing the design of a possible epidemiologic study but also for providing information for public health education, actions, or interventions that might be appropriate independently from an epidemiologic study (e.g., environmental testing to guide remediation efforts unrelated to the cancer of concern).
- Investigators should identify study design requirements and available resources to conduct the study. This process includes identifying the scope of the study and determining whether sufficient resources and data are available to complete meaningful work. Investigators should:
 - determine which parameters to use for geographic scope, study timeframe, and demographics and select a timeframe that allows for sufficient latency in cancers of concern;
 - determine the study design, sample size, and the statistical tests necessary to study the association as well as the effect of a smaller sample size on statistical power;
 - determine the appropriateness of the plan of analyses, including hypotheses to be tested as well as epidemiologic and policy implications; and
 - assess resource implications and requirements of the study and identify sources of funding.

Recommendations for Step 3

- Investigators should maintain communication with the community.
- Investigators should support the community through acting on an understanding that members might have valuable information about hazards in the area.
- Investigators should use a data-driven process for decision-making.

- Investigators should be proactive in maintaining interagency coordination and involving needed experts in advisory roles.
- Investigators should carry out the feasibility assessment as broadly as possible with existing information, including assessment of previous efforts in environmental or clinical testing.

Decision to Close the Investigation at Step 3

In some cases, despite the finding of a significantly elevated SIR, the feasibility assessment might indicate that further study will likely be unable to determine the cause of the elevated rate. In situations in which the types of cancers have no known association with an environmental contaminant, in which there are only a handful of cases, in which no suspected environmental hazard exists, or in which other factors explain the observed cancer excess (e.g., a substantial movement of residents during the study period), investigators might determine that data are insufficient or that insufficient justification exists for conducting further epidemiologic study. If the feasibility assessment suggests that little will be gained from proceeding further, the investigator should close the inquiry and summarize the results of this extensive process in a report to the initial caller and all other concerned parties.

In some circumstances, the public or the media might continue to demand further investigation, regardless of cost or biologic plausibility. Working with established community relationships, media contacts, and the advisory panel will be critical in managing and responding to expectations. If an extensive epidemiologic investigation is not carried out, it is critical to establish other possible options to support the community's health, depending on the information and resources available.

Decision to Continue to Step 4

If the activities in Step 3 to assess the feasibility of an epidemiologic study suggest that it is warranted, the responders should proceed to Step 4. Further outreach, health assessment, interventions, or other public health actions also might be appropriate. Conducting epidemiologic investigations can take several years; the health agency should consider what can be done in the interim to help protect the community's health and keep its members informed. This level of investigation often can be seen as research rather than public health response to a community concern. Providing periodic progress reports to keep the community involved can help overcome this perception.

Step 4. Conducting an Epidemiologic Investigation

Description

The primary purpose of conducting an epidemiologic investigation of the suspected cancer cluster is to determine if the exposure to a specific risk factor or environmental contaminant might be associated with the suspected cancer cluster. Demonstrating a statistically significant association does not prove causation. The scientific rigor necessary for determining causation is difficult to achieve with an epidemiologic study alone; in addition, determining causation often relies on clinical and laboratory studies (28). This distinction should be communicated to an audience not familiar with these methodologies.

Considerations

This step involves a standard epidemiologic study that tests a hypothesis of the association between putative exposures and specific cancer types, for which all the preceding effort has been preparatory. Using the feasibility assessment as a guide, responders should develop a protocol and implement the study. The circumstances of most epidemiologic studies tend to be unique. More specific guidelines are provided (see Appendix C).

The results of an investigation are expected to contribute to epidemiologic and public health knowledge. This contribution might 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. It could take many years for such studies to be completed, and even then the result often provides an incomplete picture.

However, even if a cancer cluster is identified and environmental contamination is identified, an investigation might not demonstrate a conclusive association between the contamination and cancer. Other risk factors (e.g., smoking, personal behavior, occupational exposures and genetic traits) also should be explored. Conversely, even if the investigation does not identify an association between a particular suspected environmental exposure and cancer cluster, the exposure still might be linked to the cluster; however, in such a case more scientific information might be required (e.g., toxicologic and clinical data) to establish an association. Epidemiologic studies alone often are not able to detect small effects, particularly in small populations or when the number of cases is limited.

Sometime in advance of beginning an extensive investigation, it is important that health agency responders and the investigation team be clear with the community, the media, and others about the inherent difficulties in undertaking such studies. Every effort should be made to set realistic expectations

about the information an epidemiologic investigation will likely provide. Regardless of how exhaustive or comprehensive an investigation, few provide definitive answers and address the community's concerns. Even when expectations are established before the investigation begins, such circumstances can be disappointing to all, and particularly worrisome to the potentially affected persons. Thus after the investigation concludes, the health agency response often persists. Continuous interaction and relationship with the community, along with transparency of process, continue to be vital in such circumstances. Ongoing open communication, information sharing and public awareness efforts might be needed in order help the community overcome frustrating circumstances.

Conclusion

Public health agencies, including cancer registries, continue to receive hundreds of inquiries about suspected cancer clusters every year (10,14,29). Since publication of the 1990 Guidelines, many changes have taken place in data quality, technology, and communication. Data resources have become richer and statistical methods more refined, and many lessons have been learned from 2 decades of cancer cluster investigations.

Cancer cluster investigations continue to present many challenges. Populations at risk continue to be difficult to define, related environmental contaminants might have been in place many years earlier than the contaminant under investigation, and epidemiologic methods to provide strong evidence of association in large studies have limited value in community settings (14). Only a small fraction of cancer cluster inquiries might meet the statistical and etiological criteria to support a cluster investigation through all the steps outlined in this report. Because of the continuing challenges involved in investigating suspected cancer clusters, state and local health agencies continue to place an important emphasis on transparent and effective communication. The purpose of the revised guidelines contained in this report are to provide needed decision support to public health agencies in order to promote sound public health approaches, facilitate transparency and build community trust when responding to community cancer cluster concerns.

Acknowledgments

This report was reviewed by Kurt Snipes, PhD, California Department of Public Health. Other reviewers and editors of this report include Brian Lewis, HP Enterprise Services (contractor) and Anne S. Venner, MA, Federico Feldstein, MA, and Gregory Kearney, DrPH, East Carolina University.

References

1. CDC. Guidelines for investigating clusters of health events. *MMWR* 1990;39(No. RR-11).
2. Kingsley BS, Schmeichel KL, Rubin CH. An update on cancer cluster activities at the Centers for Disease Control and Prevention. *Environ Health Perspect* 2007;115:165–71.
3. Council of State and Territorial Epidemiologists. A synopsis of the 2010 national assessment of state cancer cluster investigations and protocols. Atlanta, GA: Council of State and Territorial Epidemiologists; 2010.
4. National Institute for Occupational Safety and Health. Occupational cancer. Washington, DC: National Institute for Occupational Safety and Health; 2010. Available at <http://www.cdc.gov/niosh/topics/cancer>.
5. Brown AM. Investigating clusters in the workplace and beyond. *Occup Med (Lond)* 1999;49:443–7.
6. CDC. Cancer clusters. Atlanta, GA: US Department of Health and Human Services, CDC; 2012. Available at <http://www.cdc.gov/nceh/clusters/about.htm>.
7. National Cancer Institute. NCI dictionary of cancer terms. Bethesda, MD: National Cancer Institute; 2013. Available at <http://www.cancer.gov/dictionary?cdrid=45333>.
8. Heron, Melonie. Deaths: leading causes for 2009. *National Vital Statistics Reports*, Vol. 61, No. 7. Available at http://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_07.pdf.
9. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2009 incidence and mortality data. Atlanta, GA: US Department of Health and Human Services, CDC, National Cancer Institute; 2013. Available at <http://www.cdc.gov/uscs>.
10. Thun MJ, Sinks T. Understanding cancer clusters. *CA Cancer J Clin* 2004;54:273–80.
11. Aldrich T, Sinks T. Things to know and do about cancer clusters. *Cancer Invest* 2002;20:810–16.
12. Schottenfeld D, Fraumeni JF, Samet JM, Colditz GA, Whittemore AS, eds. *Cancer epidemiology and prevention*. New York, NY: Oxford University Press; 2006.
13. Bender AP, Williams AN, Johnson RA, Jagger HG. Appropriate public health responses to clusters: the art of being responsibly responsive. *Am J Epidemiol* 1990;132(1 Suppl):S48–52.
14. Trumbo CW. Public requests for cancer cluster investigations: a survey of state health departments. *Am J Public Health* 2000;90:1300–2.
15. Rubin CS, Holmes AK, Belson MG, et al. Investigating childhood leukemia in Churchill County, Nevada. *Environ Health Perspect* 2007;115:151–7.
16. National Cancer Institute. *Defining cancer*. Bethesda, MD: National Cancer Institute; 2013. Available at <http://cancer.gov/cancertopics/cancerlibrary/what-is-cancer>.
17. Neutra R, Swan S, Mack T. Clusters galore: insights about environmental clusters from probability theory. *Sci Total Environ* 1992;127:187–200.
18. Lanphear BP, Buncher CR. Latent period for malignant mesothelioma of occupational origin. *J Occup Me* 1992;34:718–21.
19. Grufferman S. Methodologic approaches to studying environmental factors in childhood cancer. *Environ Health Perspect* 1998;106(Suppl 3):881–6.
20. Devita Jr. VT, Hellman S, Rosenberg SA, eds. *Cancer, principles and practice of oncology*. 7th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2005.
21. Urayama KY, Von Behren J, Reynolds P, Hertz A, Does M, Buffler PA. Factors associated with residential mobility in children with leukemia: implications for assigning exposures. *Ann Epidemiol* 2009;19:834–40.
22. Guidotti TL, Jacobs P. The implications of an epidemiological mistake: a community's response to a perceived excess cancer risk. *Am J Public Health* 1993;83:233–9.
23. Kelsey JL. *Methods in observational epidemiology*. New York, NY: Oxford University Press; 1996.
24. Khurshid A. *Statistics in epidemiology: methods, techniques, and applications*. Boca Raton, FL: CRC Press; 1996.
25. Fritz A, Jack A, Parkin DM, Percy C, Shanmugarathan S, Sobin L, Whelan S. *International Classification of Diseases for Oncology (ICD-O)*. 3rd ed. Geneva, Switzerland: World Health Organization; 2000.
26. Selvin S. *Statistical analysis of epidemiologic data*. New York, NY: Oxford University Press; 1996.
27. US Census Bureau. *Dicennial Management Division glossary*. Washington, DC: US Census Bureau. Available at <http://www.census.gov/dmd/www/glossary.html#C>.
28. Rothman, K J. A sobering start for the cluster busters' conference. *Am J Epidemiol* 1990;132(Suppl):S6–13.
29. Neutra RR. Counterpoint from a cluster buster. *Am J Epidemiol* 1990; 132:1–8.

APPENDIX A

Data and Other Resources

Since the 1990 guidelines for investigating clusters of health events (1) were published, a substantial increase has occurred in the number of sources of available data that can help public health agencies respond to cancer cluster inquiries and conduct cancer cluster investigations. These sources include data on cancer diagnoses, demographics, and environmental quality.

Cancer Registries

The state cancer registry is a vital data source for suspected cancer cluster investigations. The state central cancer registry, which receives reports of all new cancer cases from clinical facilities in the state, will have numerator data (i.e., the number of new cancer cases) for calculating the SIR as well as data for the appropriate comparison measures for reference populations. In 1990, many states did not have a cancer registry, and the majority of states with registries lacked resources to gather complete data. Today, every state has a statewide central cancer registry for collecting, managing, and analyzing high-quality data on incident (i.e., newly diagnosed) cases of cancer and cancer mortality among residents.

Two federal programs support central cancer registries which compile data on cancer incidence: CDC's National Program of Cancer Registries (NPCR) supports central registries covering 96% of the U.S. population with registries in 45 states, the District of Columbia, Puerto Rico, and the Pacific Islands (2) and NCI's Surveillance, Epidemiology and End Results (SEER) Program includes five state registries and a number of regional and special population registries (3). Together, these programs collect data for the entire US population (3). Uniform national data standards for all registries are developed and promoted by the North American Association of Central Cancer Registries (NAACCR) (4).

The state and national registries have data on cancer type (e.g., organ site, histology, and many other fields) as well as detailed demographic information on the individuals with cancer. Although state registries most often group cancer statistics by county, many registries are also able to characterize data on individual cases by geographic location (geocoding). Age, sex, and race/ethnicity geocoded information permits researchers to calculate the SIR at various geographic levels. The majority of states have internet sites for cancer statistics; SEER (available at <http://seer.cancer.gov>) and NPCR (available at <http://www.cdc.gov/cancer/npcr>) also present cancer statistics.

Completeness of the NPCR and SEER registries varies by state, although in general they have a high level of completeness and accuracy. NAACCR certifies registries annually based on completeness overall (95% and 90% for Gold and Silver, respectively) and for specific data items such as race, age, and gender (5). There might be a delay (≥ 1 year) between cancer diagnosis and the availability of complete data in the cancer registry. Preliminary data might be available for more recent years; however, these data might not contain all cancer cases from these years. The state registry will have information on which years have complete information.

Limitations and cautions to the use and interpretation of data from cancer registries include the following:

- Registry information generally contains patient address at date of diagnosis only.
- The majority of registries do not collect information on possible risk factors (e.g., smoking history). Cancer registries do have fields for usual occupation and industry, but the data are often incomplete.
- The types of cancer that are most likely to be underreported occur in persons with late-stage cancers that are treated with palliative care (e.g., persons who might not be hospitalized for surgery or treatment). Other likely underreported types include those who have been diagnosed in a physician's office without hospitalization (e.g., early stage melanoma). Many hospitals routinely collect cancer data for their own purposes and for most hospitals reporting to central registries is routine. However, reporting from nonhospital facilities is less reliable. Consequently data for cancer patients who are never hospitalized for diagnosis and treatment tend to be less complete and might be reported later than other cases (6).
- Codes and rules for counting cancer cases do change. Some histology classifications change from benign to malignant and vice-versa, depending on the coding edition. Ovarian cancers and hematopoietic cancers are prominent examples. These are for the most part exceptions, and they will be known by the cancer registry personnel.
- Occasionally, changes in diagnostic criteria might change how a cancer is diagnosed, possibly creating changes in the frequency in which the cancer is detected and reported. These types of changes are adopted at different rates by physicians and hence in reports to the registries.
- Data on race and ethnicity are captured in registry data; however this data is collected inconsistently with some

providers relying on a patient's self-report and others assessing race based on observation.

- Many registries are aware of “quirks” or “anomalies” in possible mismatching of numerator and denominator data of their regions as a result of rapidly growing or shrinking areas or large population centers that straddle county or other borders.

These limitations notwithstanding, the existence of population-based cancer registries has greatly reduced the resource intensity of determining how many and of what type of cancers have occurred in a given area in a state. These registries thus present efficient opportunities for answering questions that the public has about cancer concerns, including suspected cancer clusters.

State Cancer Profiles

Cancer incidence and mortality data, compiled largely from registry data, are also available on State Cancer Profiles (available at <http://statecancerprofiles.cancer.gov>), a collaborative effort between CDC and NCI (7). The data on this site include state- and county-level cancer incidence and death rates. Statistical assessments are provided for upward and downward trends in rates by county and comparisons to state rates. A mapping capability also is provided; however, the maps do not reflect statistical differences in cancer incidence or death rates. Although the target audiences for this information are health planners, policy makers, and cancer information providers engaged in cancer control planning, the media as well as members of the public also use this site. In addition to data on cancer incidence and mortality, the site provides risk behavior data based on CDC's Behavioral Risk Factor Surveillance System (8).

Data on Deaths

In addition to incidence data from cancer registries, data on deaths compiled by state vital records offices might be a useful supplement in identifying data on cancer cases. Death records are most useful for cancer with high mortality and a short survival period such as pancreatic, liver, lung, and some types of brain cancer. However, death records are not very useful for cancers with lower mortality, such as breast, thyroid, prostate, or colon cancers, from which patients are likely to survive. Death records increasingly are submitted to state health agencies online, and they are often available within weeks or even days after death. When survival is likely to be short (within 2 years), death records can help to fill in gaps

in the cancer registry case count, since registries might have a 1–2 year lag in ascertaining complete records.

Limitations and cautions in the use of death records in cancer cluster investigations include the following:

- Death records might be limited by the requirement that the residence of the deceased is recorded as the address at the time of death; this address might or might not be the place where the individual resided at the time of the cancer diagnosis.
- Death records are not necessarily completed by the physician who best knew the patient's medical history, meaning that the given cause(s) of death might not always be accurate.

U.S. Census Bureau

The U.S. Census Bureau's American FactFinder (available at <http://factfinder.census.gov>) can provide valuable data for use in determining the denominator for incidence calculation (9). State, county, census tract, and census block level data are available. Census data include total population figures, along with socioeconomic status, race/ethnicity, age, sex, and many other useful characteristics of a population.

Limitations and cautions about the use of census data include the following:

- Census numbers might be inaccurate for intercensal years when substantial population changes (rapid growth, shrinkage, or aging changes) occur.
- Census boundaries occasionally change, most often in rapidly growing areas that are often subdivided, making comparison between years or combining data from different years difficult. American FactFinder allows a user to see the changes between census years (e.g., between 2000 and 2010).
- The census tract is defined by the U.S. Census Bureau, and it is a relatively homogeneous unit with respect to population characteristics. A census tract generally contains between 1,000 and 8,000 persons, with an optimum size of 4,000 persons (10). Cancer clusters of concern frequently are confined to areas smaller than a census tract. Because census tracts are subdivided into census blocks and block groups, blocks and block groups might be combined if a census tract does not give the needed geographic boundaries. The number of cases occurring within a block or a block group might be far too small to allow reporting of cancer cases without privacy concerns or creating statistically unstable rates. Registries often will not release data at the block group level or even the census tract level because of privacy concerns.

- Census units might not be similar to contamination boundaries.
- The state demographer is the best resource for information regarding changes in population size.

Zip codes can be and often are used as geographic areas for cluster investigations, especially if they are a better fit for communities at issue. There are two major limitations to using zip codes for cancer cluster investigations: 1) zip code boundaries might change more often than census boundaries, and 2) zip codes cross county and census boundaries. Moreover, a person might have a post office box or a rural route address that is in a different zip code than the actual residence. Real estate sites, such as City-Data.com (available at <http://www.city-data.com>), often can be useful for researching population changes and demographic information.

National Environmental Public Health Tracking Network

One resource that was not available during the development of the 1990 Guidelines is CDC's National Environmental Public Health Tracking Network (Tracking Network), a nationwide surveillance network that provides health, environmental hazard, and exposure data and information to better inform and protect communities (11). The Tracking Network (<http://www.cdc.gov/nceh/tracking>) is a web-based system of integrated data and information derived from a variety of sources, including federal, state, and local agencies and registries.

Along with other selected health outcomes, the Tracking Network offers data and health messaging on several categories of cancers, including leukemia (by subtype), pediatric cancers, brain cancer, and other cancer types. The website will include additional types of cancers in the future. The cancer data are derived from a compilation of registry data, including NPCR and NCI's SEER programs. Cancer health outcomes data available for many states can be viewed in map, table, or graph format. Annual age-adjusted rates and annual number of cases are available for each selected cancer category for each state, and 5-year average annual rates are available by county. Other information, including demographic and socioeconomic characteristics, health behaviors, and biomonitoring data are also available. Because of a limited or low number of case counts and data confidentiality and human protection laws, health data are protected from being viewed on the Tracking Network at a higher geographic resolution, such as by census tract. In some cases, a request for individual or identifiable data might be granted by state cancer registries directly.

Environmental data primarily derived from federal, state, and local regulatory environmental protection departments (or agencies) are available on the Tracking Network. However, state and local jurisdictions might provide more detailed environmental data, along with staff members who are knowledgeable about issues surrounding a particular situation.

Data from State and Territorial Environmental Agencies

State and local environmental protection agencies routinely collect environmental data. Because these data are collected in places and at times according to regulatory purposes, they might be useful in identifying environmental hazards in cancer cluster investigations, or they might only approximate the environmental conditions at the site of the potential cancer cluster. Environmental agencies regularly collect data on water quality and air quality for compliance with air and water quality standards. These agencies also often permit and regulate industrial or other facilities that generate, transport, or store hazardous waste or other chemicals. The agencies will therefore have records of compliance and noncompliance that might indicate emissions into the environment. The state agencies are also involved, along with the Environmental Protection Agency (EPA), in monitoring pollution and in the oversight of the cleanup of contaminated sites. Although some states conduct surveillance on pesticide-related illness and injury, not all states regularly collect and maintain data on pesticide use or exposure; if collected, the data are usually kept at the state department of agriculture and sometimes by the state environmental protection agencies.

EPA collects environmental data for regulatory purposes, and the agency publishes the data on its website. A viewer can use tools on the EPA website to view information on air quality or water quality or to see if there are local Superfund sites, brownfields (12), or releases from manufacturing facilities (14). The information is available at the zip code level and can be displayed on a map.

The staff located within state or local environmental protection departments can be a helpful resource for providing information about local environmental conditions that might lead to exposure to contamination. The staff's assistance should be engaged in evaluating available environmental data for relevance to a cancer cluster inquiry or investigation because the data collection areas are determined by regulatory requirements and might not provide information specific to a particular site of public health interest. EPA's list of State and Territorial Environmental Agencies is available at <http://www.epa.gov/epahome/state.htm>.

Sources of information on the association between specific environmental contaminants and cancer are available. Weight-of-evidence-evaluations of carcinogens are published by the International Agency for Research on Cancer (IARC) (IARC cancer classifications are available at <http://www.iarc.fr>) and the National Toxicology Program (NTP's Report on Carcinogens is available at <http://ntp.niehs.nih.gov/go/roc>). These evaluations tend to focus on exposures that have been of concern for some time and therefore on which there are substantial data. Not all potential carcinogens have been evaluated by these organizations. Other sources of information include PubMed (available at <http://www.ncbi.nlm.nih.gov/pubmed>), the ATSDR Toxic Substance Portal (available at <http://www.atsdr.cdc.gov/substances/index.asp>), and the ATSDR series of Toxicological Profiles on various chemicals (available at <http://www.atsdr.cdc.gov/toxprofiles/index.asp>).

By using the community members' local knowledge about the hazards and risk factors in their community as well as data from environmental and other databases, the investigator can make more informed decisions during the investigation process. For example, information provided by the concerned community members and by available databases can be useful in defining the geographic area and time period for the population at risk, increasing the accuracy and precision of the population definition. Readily available information on environmental hazards in the area of interest can be reviewed to determine if any of the hazards have a space and/or time pattern that can be related to the suspected cancer cluster. A thorough evaluation of environmental hazards with input from the community is appropriate because it might suggest some relevant public health interventions that turn out to be valuable, independent of any suspected cancer cluster. For example, in a community concerned about contaminants in private well systems, proper maintenance of private well systems might be an appropriate public health education program, regardless of whether contaminants are found, particularly if residents express confusion over how to maintain these wells.

Biomonitoring

Biomonitoring is the measurement, usually in blood or urine, of chemical compounds, elements, or their metabolites in the body. Although biomonitoring indicates exposure to a substance at some level, it might not indicate when the

exposure occurred or what effects the exposure might have on health in the future. Because of the long latency period associated with the development of cancer, the limitations of current environmental data also apply to using or collecting current biomonitoring data. The relevant exposure might have occurred years before and might not be detectable at the time that samples for biomonitoring are collected. Although a substance is detected in the body, it might not be a carcinogen or it might not be at levels known to cause the disease. For the U.S., CDC's National Health and Nutrition Examination Survey (NHANES) provides reference data for over 200 chemicals in the blood and urine for a selection of the survey's participants (14). Biomonitoring is a relatively new field, and there is a need for more research to permit an understanding of which substances at what concentrations in the body contribute to cancer.

References

1. CDC. Guidelines for investigating clusters of health events. *MMWR* 1990;39(No. RR-11).
2. CDC. National Program of Cancer Registries. Available at <http://www.cdc.gov/cancer/npcr>.
3. National Cancer Institute. Surveillance, Epidemiology and End Results (SEER) Program. Available at <http://seer.cancer.gov>.
4. North American Association of Central Cancer Registries. NAACCR Standards for Cancer Registries. Available at <http://www.naacrr.org>.
5. North American Association of Central Cancer Registries. Certification levels. Available at <http://www.naacrr.org/Certification/CertificationLevels.aspx>.
6. Penberthy L, McClish D, Peace S, Gray L, Martin J, Overton S, Radhakrishnan S, Gillam C, Ginder G. Hematologic malignancies: an opportunity to fill a gap in cancer surveillance. *Cancer Causes Control* 2012;23:1253–64.
7. National Cancer Institute. State cancer profiles, 2011. Available at <http://statecancerprofiles.cancer.gov>.
8. CDC. Behavioral Risk Factor Surveillance System survey data. Atlanta, GA: US Department of Health and Human Services, CDC; 2013. Available at <http://www.cdc.gov/brfss>.
9. US Census Bureau. American fact finder. Available at <http://factfinder.census.gov/home/saff/main.html>.
10. US Census Bureau. Census bureau glossary. Available at <http://www.census.gov/dmd/www/glossary.html#C>.
11. Tango T. *Statistical methods for disease clustering*. New York, NY: Springer; 2010.
12. Environmental Protection Agency. My environment query. Available at <http://www.epa.gov/myenvironment>.
13. Environmental Protection Agency. Toxic release inventory. Available at <http://factfinder.census.gov/home/saff/main.html>.
14. CDC. Fourth national report on human exposure to environmental chemicals, updated tables, 2012. Available at <http://www.cdc.gov/exposurerePort/pdf/FourthReport.pdf>.

APPENDIX B

Communication

An overriding goal throughout the process of a cancer cluster investigation, beginning with the initial contact, is to communicate with transparency and to embrace community involvement. The health department and its process should be accessible to the community. This section provides guidance and resources on communicating during a cancer cluster response.

Developing Communication Plans

Before responding to any inquiries concerning a possible cancer cluster, the health agency should develop a one-on-one communication strategy. Key points in such a strategy should include the importance of listening and how to ask questions that will help determine the nature of the caller's concerns. If possible, responders should try to ascertain in the first call, the level of concern across the larger community. A basic communication plan should be created for answering initial inquiries about possible excess cancer cases. Such a plan will include anticipated characteristics of possible callers, questions to employ to gather the appropriate information, and talking points about cancer, clusters, and the scientific evaluation process. The plan also should define commonly used terms (e.g., cluster) in a clear and accessible way and emphasize that when speaking to a caller, a responder should use such terms in a consistent manner. Statistical concepts such as small samples size, random fluctuation, and statistical significance are difficult concepts for the general public audience to understand, and having consistent, clear, talking points that address these concepts is helpful.

If and when the investigators determine that the entirety of the evidence (e.g., an elevated SIR and an environmental contaminant that is linked to the cancer of concern in the published literature) supports proceeding with an investigation, they should make a concerted effort to establish a solid communication plan within the health agency's communications office. Components of such a plan should include identification of audience and messages, stakeholder groups, types of meetings, communications with the media, social networking possibilities, proactive versus reactive communication, and a commitment to a transparent approach.

Communication Audience

The communication audience throughout the process of inquiry or investigation will include the initial caller, other concerned community members, community leaders, public health partners, government officials, media, physicians, real estate agents, and other groups, depending on how far the inquiry progresses. The media might approach the health agency with questions at any time, and the health agency will need to be prepared with clear statements for publication. At all stages of the process, the primary concern is the community. If community concerns include a known or suspected industrial contamination, those in the health agency taking the inquiry or handling community and media relations should interact with the community before or at the same time as with the company responsible for the contamination, not after. The media can be important partners in conveying information to community members. However, the health agency should not underestimate the importance of meeting face-to-face with individuals with cancer, their families and impacted community members. This is especially important for sharing information about the health agency's actions or findings. The particular persons who comprise the "community" and the nature of community involvement will change during the steps of cancer cluster inquiries and investigations. The appropriate partners and stakeholders should be identified and involved.

In the initial contact, communication generally is aimed at the person reporting a concern about cancer in the community. The person might be a medical professional or a legislator or community resident with little or no medical expertise. After the health agency responder takes the call, the responder should communicate with agency partners (in the health agency(s) and, if necessary, in the appropriate environmental protection agency) to alert them to the community's concerns.

After the initial response and as a part of the inquiry, communication might extend to the inquirer's family and friends as part of the information gathering and sharing process. If the inquiry progresses past Step 1, the intended audience for communications will broaden to include community residents, members of the media, other agencies (state, local, or federal), and possibly elected officials. Once anyone beyond the initial inquirer is involved, the local health agency should be included in any communications, regardless of whether a statistical excess of cases can be determined.

If an excess of cancer cases is identified (Step 2) and an epidemiologic study is being considered (Step 3), two-way communication with community members is important. One method to accomplish such communication is to convene a community panel. This entity should include individuals who represent the community and, if possible, those with specific expertise that might be helpful during the process. The health agency should hold regular meetings with the panel. The panel should be well organized and have an agenda to keep the discussion on track and to conduct a useful dialogue. Participants in meetings might include concerned residents, residents with expertise, and local health, media, and elected officials. Such meetings provide a useful way to learn about the community and to build trust, credibility, and transparency. They are also useful for keeping the investigation's activities appropriate, focused, and on track. The community panel should be established early in an investigation; otherwise, other models might need to be considered. In communities where trust in government has eroded, it is particularly important to engage the community in the selection of participants of a community panel.

Health agency officials should use their best judgment and assess through personal interactions with community members, media, and internet postings whether a community panel (set up to facilitate communication around the community's cancer cluster concerns) is warranted. If not, the health agency and its investigators should work to establish relationships with existing, trusted community groups and suggest regular, structured, two-way communication with those groups.

Communicating in Uncertain and Stressful Situations

Because of the perception of health and environmental risk, persons can feel uncertain, worried, and less trusting. Accordingly, principles of risk communication should be part of the training for anyone dealing with the process of cancer cluster inquiries or investigations (1). A few key communication concepts at any step of the inquiry include the following, adapted for cancer clusters from previous guidance (2):

- be a credible and consistent source,
- create realistic expectations,
- raise awareness of other credible sources,
- be empathetic and have patience,
- be supportive and receptive to the information reported, and
- listen clearly and consistently.

Proactive Community Involvement

During Step 2 (the process of determining whether an excess of cancer cases exists), obtaining community input might be useful but not vital. However, once the decision is made to proceed to Step 3, proactive community involvement is critical, not only for gathering information but also for sharing the investigation parameters and process with the community and other affected or collaborating partners.

One way to involve the community broadly is to establish advisory groups, such as a community panel (See Step 3, Procedures). Another way is to hold public meetings. If, during the process of investigation, a need is identified to have public meetings, a clear agenda and goal should be set for each meeting, including discussions of major milestones (e.g., completion of the feasibility assessment). The format and atmosphere of a public meeting can have great influence on its outcome. For example, town hall-type public meetings can allow community members to express frustrations and feelings to officials. Health agency personnel who listen well can establish credibility with the community in such meetings. However, some agencies might have difficulty in communicating well in this format. In these cases, an agency should use trained facilitators who understand the local culture. In such meetings, the health agencies should keep presentations short and use plain language. An alternative is to conduct public meetings with a series of "stations," at which data (e.g., maps) can be presented and discussed in one-to-one or small-group communication. This is one way to involve partners such as environmental agencies and community groups in this type of meeting.

Depending on the community's unique needs, one of these approaches or a combination might work best. For each type of meeting, the health agency should include resources for community members who attend, such as educational materials about cancer. Because dealing with a suspect cancer cluster can bring great stress to members of the community, potentially causing additional stress-related illness, resources about stress management also might be useful in promoting public health.

Other options for communicating on a regular basis with the community include establishing a toll-free telephone number for use by members of the community to ask questions during the entire process, providing regular (e.g., monthly) written updates between meetings, creating a website with all relevant information (including a compilation of questions and answers) or, if necessary, establishing a community office. The local health agency will be a valuable partner at this stage of communications.

Another avenue is to work with the state communications department and/or public affairs office to use social media as

a communication forum about the investigation. Community members are likely to use social media to obtain information. Putting information out on social media sites and inviting questions has advantages and disadvantages. It is similar to having a toll-free number available, but it also allows for two-way communication that can be viewed by and shared with others. Members of the community also might use their own social media sites, including blogs, to ask questions and express their own opinions. Monitoring such sites provides a valuable opportunity for the health agency to be aware of community concerns and to address misconceptions (3,4,5).

Resource for State and Local Health Agencies

CDC and the National Public Health Information Coalition (NPHIC) have published a useful resource which is currently available to state and local health agencies, providing detailed guidelines on communicating in cancer cluster investigations (available at <http://www.nphic.org>). *Cancer Clusters: A Toolkit*

for Communicators (6) includes information on working through a suspected cancer cluster scenario. It provides suggested outreach techniques for various audiences and offers answers to commonly asked questions about suspected cancer clusters. It also provides literature resources, a glossary of cancer cluster terms, a guide to education by use of social media, and case studies.

References

1. Krimsky S, Plough AL. Environmental hazards: communicating risks as a social process. Dover, MA: Auburn House; 1988.
2. Covello VT, Allen F. Seven cardinal rules of risk communication. Washington, DC: US Environmental Protection Agency; 1988. Available at http://www.epa.gov/care/library/7_cardinal_rules.pdf.
3. CDC. CDC guide to writing for social media. Atlanta, GA: US Department of Health and Human Services, CDC; 2012. Available at <http://www.cdc.gov/socialmedia/Tools/guidelines/pdf/GuideToWritingForSocialMedia.pdf>.
4. Fox S. The social life of health information, 2011. Available at <http://www.pewinternet.org/Reports/2011/Social-Life-of-Health-Info.aspx>.
5. Dredze M. How social media will change public health. IEEE Intelligent Systems, 2012;27:81–4.
6. National Public Health Information Coalition. Cancer clusters: a toolkit for communicators. Available at <http://www.nphic.org>.

APPENDIX C

Statistical and Epidemiologic Approaches

A suspected cancer cluster investigation attempts to answer two questions: 1) is there an actual “excess” (that meets statistical and biological plausibility criteria) and 2) is this excess associated with an environmental contaminant? Addressing these questions begins by defining the study population and locating relevant cases and then determining the appropriate geographic boundaries and time period.

This section provides an outline of the basic epidemiological and statistical analysis methods that are recommended for investigating a cancer cluster. This section focuses on the methods most relevant and most commonly used in cancer cluster investigations: the SIR and confidence interval, mapping, and descriptive and spatial statistical and epidemiologic methods.

Standardized Incidence Ratio and Confidence Interval

The measure typically used to assess whether there is an excess number of cancer cases is the SIR. This measure is explained in many epidemiologic textbooks (sometimes under standardized mortality ratio, which uses the same method but measures mortality instead of incidence rates) (1–5). Simply stated, the SIR is a ratio of the number of observed cancer cases in the study population to the number that would be observed (often called “expected”) if the study population experienced the same cancer rates as an underlying population (often called the “reference” population). The reference population could be the surrounding census tracts, other counties in the state, or the state as a whole (not including the community under study).

The SIR can be adjusted for factors such as sex, race, and/or ethnicity, but it is most commonly used to adjust for differences in age between two populations. Various techniques can be used to account for these factors. For example, stratification, which is calculating an SIR by groups (e.g., by calendar year), is a commonly employed technique. (6)

Confidence Interval

A confidence interval is calculated to determine the precision of the SIR estimate and the statistical significance. If the confidence interval includes 1.0, the SIR is not statistically significant. The narrower the confidence interval, the more confidence one has in the precision of the SIR estimate. One difficulty in cancer cluster investigations is that the population

under study is generally a community or part of a community, typically resulting in a small denominator, and such small denominators frequently yield wide confidence intervals, meaning that the SIR is therefore not as precise as desired (1).

Considering Alpha and Beta Level Values

The alpha is the probability of rejecting the null hypothesis when the null hypothesis is true (no difference in cancer rates between the study population and reference population). Although there are no absolute cut-points, responders often use an alpha value of 0.05 (or equivalently a 95% confidence interval).

Selection of an alpha value larger than 0.05 (e.g., 0.10: 90% confidence interval) will increase the risk of false positive results. Selection of a smaller alpha value (e.g., 0.01: 99% confidence interval) may be considered when many SIRs are computed because the number of SIRs that will be statistically significant by chance alone increases (in other words, with a 95% confidence interval, one expects to see five statistically significant results in a group of 100 results).

Beta and power are related to each other. Both are related to the sample size of the study—the larger the sample size, the larger the power. Power, or $1 - \beta$ (beta), is the probability of rejecting the null hypothesis when the null hypothesis is actually false. Like alpha, the beta has no absolute cut-points; however, responders often use a beta value of 0.20 or less (or equivalently a power of 0.8 or more) (1).

Power Analysis

Power analysis is useful in determining the minimum number of people (sample size) needed in a study in order to test the hypothesis and detect a possible association. In most suspected cancer cluster investigations, the cases and study population are defined prior to the analysis. Therefore, a power analysis can be used to determine if the number of cases in the investigation is sufficient, usually a power of 0.8 or greater (3).

Mapping the Cancer Cluster

When considering the geographic distribution of cases, responders have various methods they can use. For example, they might develop a visual representation showing the location

of each case superimposed on the underlying population density to get an approximation of the distribution of the relative rates of cancer.

It also can be useful to plot the location of suspected environmental risk factors on the map for the purpose of making a crude assessment of their proximity to the cases. However, to avoid the “Texas Sharpshooter fallacy” (i.e., a situation in which cases are noticed first and then the “affected” area is selected around them, thus making there appear to be a geographical relationship, similar to an instance in which the sharpshooter shoots the side of the barn first and then draws the bull’s-eye around the bullet holes), responders must first outline their definitions, assumptions, and methods (7). Often, a few different spatial (e.g., spatial: census block, census tract, zip code, municipality, or county) or temporal scales (e.g., week, month, year, or several years) can be mapped to look for possible patterns related to specific space and/or time units that merit more careful investigation. This process is systematic, and procedures are outlined a priori. The patterns in such maps often differ dramatically, and they might suggest specific exposures that warrant further consideration. This practice is more useful when longer periods of time are under study, as well as larger numbers of cases (e.g., >10 cases).

Cancer registries and state health agencies typically have criteria related to release of data for small geographic areas. Because of privacy concerns, some data cannot be released to the public, unless the privacy concerns are addressed. For example, a pin-point map of a small geographic area that identifies the residence of a cancer patient should not be made public (8). Similarly, many health agencies are prohibited from publicly releasing a table for a small geographic area with a small population, for each table cell might have only a few cases.

Descriptive and Spatial Statistical and Epidemiologic Methods

Frequencies, rates, and descriptive statistics are useful first steps in evaluating the suspected cancer cluster. Confidence intervals can also be calculated for rates. Epidemiologic references can explain these methods (9). Other statistical approaches include Poisson regression. Often, the number of cases is limited, therefore limiting the type of analysis. If an investigation progresses to a case-control study, the odds ratio can be calculated. These study designs have been discussed in detail elsewhere (1,3,4).

Since the publication of the 1990 Guidelines, the field of spatial epidemiology has grown, especially in environmental health. This growth is influenced by the increased availability of geocoded data and statistical software. Space/time cluster

analysis methods are often used to provide evidence about the existence of a suspected cluster and to define more precisely the extent of the suspected cluster in space and time.

As with any other epidemiologic analysis, there might be methodological issues with the use of clustering tools. Many of these concerns (e.g., limitations associated with small populations, environmental data quality, disease latency periods, and population migration) have been described in this report. Census data can provide the denominators for this type of analysis, and all the limitations associated with rapidly changing populations and intercensal year estimates also apply to these spatial/time cluster methods. In addition, when exposure or outcome analysis uses aggregate data and not data collected on an individual level, responders must use caution when interpreting this type of analysis, because the association with a particular environmental contaminant might not be true for individual cases, especially if there is heterogeneous distribution of the exposure over the geographic area. The related bias is known as ecological inference fallacy. Detailed information regarding methodological issues has been published previously (10).

Many methods have been developed to facilitate what is termed “space/time cluster analysis.” These methods assess whether cases are closer to one another than would be observed if the cases had been distributed at random. The concept of “close” might mean closer geographically, closer in time, or closer both geographically and in time. The numeric value of “close” is determined by the responder. For a responder to make a determination of clustering, the space-time distances have to be summarized and then evaluated with any of a variety of statistical techniques. This task can be performed by summarizing where and when each case occurred, typically using the individuals’ residence and the reported date of incidence. Some of the simplest methods merely compare the average distances between nearby cases to the average distances between cases and nearby noncases (or controls). If, on average, the cases are sufficiently closer to other cases (in space, time, or both space and time) than they are to noncases, the situation may be described as a cluster. Clusters can be detected by use of spatial autocorrelation techniques. Global clustering statistics, such as Geary’s *C* (11), detect spatial clustering that occurs anywhere in a study area. They do not identify where the cluster(s) occur, nor do they identify differences in spatial patterns within the area. Local clustering statistics, such as Local Indicators of Spatial Autocorrelation (LISA) (12), identify potential clustering within smaller areas inside a study area. Often, global techniques are used first to identify potential clustering; then, local methods are used to pinpoint the clusters in the sample area. Many global statistics have local counterparts. For example, global Moran’s *I* is the summation of local Moran’s *I* statistics (13). Clusters reported to health agencies most often are local. It is beyond the scope of this

report to describe more than a few of the most commonly used methods, and even then, these methods are described only briefly.

A useful summary of these techniques has been published recently (14). One of the most popular techniques for detecting clusters is called the spatial scan statistic. Its most commonly used implementation is the SaTScan software (15) (available at <http://www.satscan.org>). The underlying concept for this approach is the scan statistic, which considers both spatial areas and time intervals (16). Other implementations include the nearest neighbor test (17) and the Small Area Health Statistical Unit (SAHSU)'s "Rapid Inquiry Facility" (RIF) (18). Additional, statistical cluster methods have been discussed elsewhere (19). All of these methods have strengths and weaknesses. In a choice of a statistical cluster method, it might be useful to consider several criteria, such as ease of use and availability, the clarity and transparency of the method, its statistical power to detect the cluster of interest, and the method's ability to produce the desired output (20). Comparisons and reviews have been published (21). In addition, the Appendix of the 1990 Guidelines describes additional spatial statistical methods.

References

1. Kelsey JL. *Methods in observational epidemiology*. New York, NY: Oxford University Press; 1996.
2. Khurshid A. *Statistics in epidemiology: methods, techniques, and applications*: CRC; 1996.
3. Selvin S. *Statistical analysis of epidemiologic data*. New York, NY: Oxford University Press; 1996.
4. Breslow NE, Day NE. *Statistical methods in cancer research*. In: International Agency for Research on Cancer. *The design and analysis of cohort studies*. Lyon, France: International Agency for Research on Cancer Scientific Publications; 1980.
5. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.
6. New Jersey Department of Health & Senior Services Cancer Epidemiology Services. Fact sheet: explanation of standardized incidence ratios. Available at http://www.state.nj.us/health/eohs/passaic/pompton_lakes/pompton_lakes_fs_sir.pdf.
7. Gawande A. The cancer-cluster myth. *The New Yorker* 1999;8:34–7.
8. CDC. National Programs of Cancer Registry United States Cancer Statistics technical notes: statistical methods: suppression of rates and counts. Available at http://www.cdc.gov/cancer/npcr/uscs/technical_notes/stat_methods/suppression.htm.
9. Gordis L. *Epidemiology*. Philadelphia, PA: Saunders Elsevier; 2009.
10. Beale L, Abellan JJ, Hodgson S, Jarup L. Methodologic issues and approaches to spatial epidemiology. *Environ Health Perspect* 2008;116:1105–10.
11. Geary RC. The contiguity ratio and statistical mapping. *The Incorporated Statistician* 1954;5:115–46.
12. Anselin L. Local indicators of spatial association: LISA. *Geographical Analysis* 1995;27:93–115.
13. Moran P. Notes on continuous stochastic phenomena. *Biometrika* 1950;37:17–23.
14. Tango T. *Statistical methods for disease clustering*. New York, NY: Springer; 2010.
15. Kulldorff M. A spatial scan statistic. *Commun Stat-Theor M* 1997;26:1481–96.
16. Naus J. The distribution of the size of maximum cluster of points on the line. *J Am Stat Assoc* 1965;60:532–8.
17. Cuzick J, Edwards R. Spatial clustering for inhomogeneous populations. *J Roy Stat Soc B Met* 1990;52:73–104.
18. Aylin P, Maheswaran R, Wakefield J, et al. A national facility for small area disease mapping and rapid initial assessment of apparent disease clusters around a point source: the UK Small Area Health Statistics Unit. *J Public Health Med* 1999;21:289–98.
19. Rogerson PA. *Statistical methods for geography: a student's guide*. Thousand Oaks, CA: Sage Publications Ltd; 2010.
20. Robertson C, Nelson TA, MacNab YC, Lawson AB. Review of methods for space-time disease surveillance. *Spatial and Spatio-temporal Epidemiology* 2010;1:105–16.
21. Waller LA, Gotway CA. *Applied spatial statistics for public health data*. New York: John Wiley and Sons; 2004.

The *Morbidity and Mortality Weekly Report (MMWR)* Series is prepared by the Centers for Disease Control and Prevention (CDC) and is available free of charge in electronic format. To receive an electronic copy each week, visit *MMWR*'s free subscription page at <http://www.cdc.gov/mmwr/mmwrsubscribe.html>. Paper copy subscriptions are available through the Superintendent of Documents, U.S. Government Printing Office, Washington, DC 20402; telephone 202-512-1800.

Address all inquiries about the *MMWR* Series, including material to be considered for publication, to Editor, *MMWR* Series, Mailstop E-90, CDC, 1600 Clifton Rd., N.E., Atlanta, GA 30333 or to mmwrq@cdc.gov.

All material in the *MMWR* Series is in the public domain and may be used and reprinted without permission; citation as to source, however, is appreciated.

Use of trade names and commercial sources is for identification only and does not imply endorsement by the U.S. Department of Health and Human Services.

References to non-CDC sites on the Internet are provided as a service to *MMWR* readers and do not constitute or imply endorsement of these organizations or their programs by CDC or the U.S. Department of Health and Human Services. CDC is not responsible for the content of these sites. URL addresses listed in *MMWR* were current as of the date of publication.

ISSN: 1057-5987