Guidelines for Examining Unusual Patterns of Cancer and Environmental Concerns

Centers for Disease Control and Prevention and the Agency for Toxic Substances and Disease Registry

Stephanie L. Foster[1,2], Amy M. Lavery[1], Suzanne K. Condon[3], Alisha A. Etheredge[1], Brian S. Kennedy[1], Patrick Breysse[3]

Acknowledgments List:
Leadership – Pamela Protzel Berman, Erik Svendsen, William Cibulas
Internal Steering Committee Members – Johnni Daniel, Vivi Siegel, Nicole F. Dowling, Elizabeth Irvin, Jessica Rinsky, Sophia Chiu, Doug Trout, Tegan K. Boehmer
ORISE Fellows – Lauren Byrne, Diana Diaz
Geospatial Research Team – Caitlin Mertzlufft, Angela Werner Liora Sahar, Marissa Grossman, Brian Lewis, D. Aaron Vinson, Frank Curriero, Trieste Musial, Michel Conn, Amir Mirsajedin
Visuals – Traci Augustosky, Preston Burt

[1]Division of Environmental Health Science and Practice, National Center for Environmental Health, Centers for Disease Control and Prevention
[2]Geospatial Research, Analysis, and Services Program, Agency for Toxic Substances and Disease Registry
[3]Office of the Director, National Center for Environmental Health, Centers for Disease Control and Prevention

Contents

Introduction

Quick Summary

The Centers for Disease Control and Prevention (CDC) National Center for Environmental Health (NCEH) and the Agency for Toxic Substances and Disease Registry (ATSDR) produced these guidelines to provide an update to the 2013 publication, "Investigating Suspected Cancer Clusters and Responding to Community Concerns: Guidelines from the CDC and the Council of State and Territorial Epidemiologists (CSTE)"(1). This update provides state, tribal, local, and territorial health departments guidance for an expanded approach to evaluating concerns about unusual patterns of cancer in communities, including those associated with local environmental concerns.

Notable Enhancements

Notable enhancements to the previous guidelines (referred to hereafter as the 2013 Guidelines) in this revised guidance include the following:

- Expanding the name of the guidance document to include examining patterns of cancer and environmental concerns
- Revising the definition of a cancer cluster
- Including specific and standardized approaches to better engage community advocates
- Providing a standardized template to better document the nature and extent of cancer and environmental concerns
- Updating approaches to identify and investigate unusual patterns of cancer, including the suggestion for proactive evaluation and routine monitoring
- Suggesting what information to share with the CDC
- Enhanced appendices describing statistical and geospatial methods supporting the evaluation of unusual patterns of cancer

While the enhancements offered in the revised guidance include new methods to better engage with community members and advocates, statistical and other scientific challenges may make it difficult to directly associate factors that may play a role in the cause(s) of unusual patterns of cancer. The revised guidelines propose an approach to identifying and investigating unusual patterns of cancer as part of routine surveillance activities as well as new criteria and decision trees for responding to cancer and environmental concerns.

Not Included

The new guidelines do not apply to clusters associated with occupational concerns with one exception: some guidance is offered for addressing cancer concerns in a school setting. These guidelines do not discuss diseases other than cancer that people might suspect have occurred in clusters in their communities, such as infectious disease. 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.

Future Enhancements

As advancements in science and technology occur, related materials (e.g., templates and calculator for county-level cancer rates) will be developed. These materials will be posted on the CDC Cancer Cluster website.

Historic Timeline of the CDC Guidelines

CDC Guidelines for investigating clusters have been transformed over the past few decades. The timeline summarizes the history and highlights events that helped prompt revisions.

1990

Original Guidelines Published

CDC publishes Guidelines for Investigating Clusters of Health Events (2). These original guidelines focus on noninfectious health events such as chronic diseases, injuries, and birth defects.

2010

Workgroup Convened to Revise Guidelines

CDC and CSTE convene a workgroup to revise the guidelines to assist state and local health officials better address community concerns associated with cancer. This workgroup includes experts (e.g., epidemiologists, environmental health specialists, health department staff, communicators) with experience responding to concerns about unusual patterns of cancer.

2013

Revised Guidelines Published

CDC/ATSDR publishes revised guidelines in the Morbidity and Mortality Weekly Report (1). The revisions focus on cancer clusters.

2016

Trevor's Law Signed into Law

President Barack Obama signs into law the Frank Lautenberg Chemical Safety for the 21st Century Act. Section 33—Trevor's Law—calls for the Department of Health and Human Services to provide periodic updates to guidelines for investigating potential cancer clusters (3).

2019 – 2021

Guidelines Updated

In 2019, CDC receives initial funding to update the 2013 Guidelines in accordance with Trevor's Law. NCEH/ATSDR staff develop a plan to ensure the most current information and scientific methods as well as community engagement and risk communication are considered for the next updates. A core team of NCEH/ATSDR subject matter experts in cancer research, environmental health, and geospatial science was established to gather inputs and draft the guidelines. The core team convened a CDC-wide steering committee and other subject matter experts to inform the guidelines update.

2022

Updates Published

CDC (through NCEH/ATSDR) publishes updated guidelines.

Methods for Updates

The core team gathered a variety of inputs to inform the update:

- Literature reviews and a media scan
- Subject matter expert input from academic partners, non-governmental organizations, the steering committee, and other federal partners
- Input from state, tribal, local, and territorial (STLT) partners
- Input from members of the public and advocacy groups that have been involved with cancer concerns in their communities

The core team also evaluated advances in the field of environmental epidemiology (e.g., geospatial methods, cancer genomics) and community engagement strategies to determine the feasibility of incorporating these elements to enhance the guidelines.

Literature Review and Media Scan

Literature Review

The literature review methodology included:

1. **Design**. The core team designed the literature review and search criteria with help from a librarian at CDC's Stephen B. Thacker library. We reviewed peer-reviewed articles published since the literature review for the 2013 Guidelines, January 2010 – April 2021. Because not all publications of cancer investigations conducted by state and local health departments appear in peer-reviewed literature, we also reviewed reports published on state or local health websites but not in peer-reviewed journals (gray literature).
2. **Focus Areas**. The literature review focused on the following focus areas:

    - Epidemiologic investigations of cancer clusters in community and residential settings
    - Geospatial and temporal methods to evaluate clusters
    - Rare event and small area estimation statistical methods
    - Novel approaches for grouping cancers by molecular characteristics
    - Approaches for engaging, educating, and communicating with affected communities

3. **Workgroup**. Spatial statisticians and geospatial epidemiologists from ATSDR's Geospatial Research Analysis and Service Program (GRASP) and the NCEH National Environmental Public Health Tracking Program (Tracking Program) established a geographic information systems (GIS) workgroup to focus on literature specific to spatial cluster methods and GIS cluster resources.

Media Scan

The core team conducted a media scan to assess the communication landscape related to environmental hazards and excess cancer. The core team used a two-phase approach that included a comprehensive retrospective media monitoring review and a content analysis of the media around specific cancer clusters identified in the review. Results from the media scan were also reviewed to identify potential participants for community focus groups in addition to those proposed by the stakeholder meeting participants.

Subject Matter Expert Input

CDC/ATSDR convened a panel of scientific experts to:

- Provide perspective on how to update the 2013 Guidelines
- Address any gaps in the guidelines
- Incorporate new approaches, particularly regarding methods of statistical analysis and community engagement

The experts included representatives from STLT public health agencies, public health partner organizations, and academia. Subject matter expertise included the following: assessing unusual patterns of cancer, environmental epidemiology, environmental risk assessment, geospatial methods and statistics, cancer registry data, pediatric oncology, community outreach, and risk communication.

Also, experts in cancer genomics from academia and the National Cancer Institute provided input on the current state of the science in cancer genomics and applicability to evaluations of unusual patterns of cancer. Results from subject matter expert discussions and the literature review indicate the need for ongoing research in cancer genomics before these advances may be feasibly integrated into the guidelines. To support this, a National Academies of Science, Engineering and Medicine workshop will be held in Fiscal Year 2022 that may provide additional guidance.

State, Tribal, Local, and Territorial Input

Survey

We developed a survey instrument (similar to the one used to develop the 2013 guidelines) to request information on STLT public health agency's approach, best practices, and capacity for addressing local cancer inquiries. We pilot tested the instrument with members of the expert panel who currently or formerly served as state health officials. After finalizing the survey, we discussed distribution approaches with STLT health officials. The survey was sent to:

- Members of the Council of State and Territorial Epidemiologists (CSTE)
- State Environmental Health Directors (SEHD) of the Association of State and Territorial Health Officials (ASTHO)
- Tribal Epidemiology Centers

The survey also asked STLT partners to submit cancer cluster protocols and reports if available to include in the gray literature review.

The goal of the survey was to understand current and best practices associated with the existing guidelines. We wanted to identify strengths, weaknesses, suggested revisions, and resource needs for STLT public health officials to conduct investigations associated with local cancer concerns.

STLT Focus Groups

We conducted a series of focus groups with STLT public health agency professionals who respond to unusual patterns of cancer in their communities. The purpose of the focus groups was to obtain additional feedback for improving the guidelines and allow discussion amongst STLT officials regarding approaches to cancer cluster inquiries. Questions primarily focused on best practices, limitations, and common needs when addressing cancer cluster inquiries.

CSTE and ASTHO Workgroups

CSTE and ASTHO convened workgroups to review the 2013 Guidelines and provide feedback and recommendations on facilitators and barriers to implementing the guidelines and specific tools, trainings, and non-financial resources to enable STLT agencies to better implement the guidelines. Workgroups were composed of individuals who were either previous or current state public health officials routinely involved in addressing community concerns about unusual patterns of cancer. CDC and ATSDR worked with CSTE and ASTHO in an effort to gain broad geographic representation among members.

Community and Public Input

Federal Register Notice Commentary

A request for public comment was released as a Federal Register Notice to solicit input from the public, including individuals, community groups, and scientific and medical professionals(4). The notice was open from May 15, 2019, through July 15, 2019, for public comments. Public comments are available in the federal docket.

Stakeholder Meeting

NCEH/ATSDR conducted a stakeholder meeting in April 2021 to obtain input from non-government organizations, academicians, clinicians, and community stakeholders regarding the role their organizations, connections, and similar organizations might play in responding to concerns about unusual patterns of cancer. Stakeholder meeting participants were also asked to suggest names of individuals that might be willing to participate in community focus groups held in the summer/fall of 2021.

Community Focus Groups and Key Informant Interviews

We used the following to identify participants for focus groups and interviews:

- Results of the media scan
- Input from the stakeholder-meeting attendees
- Suggestions from the STLT survey respondents
- Community members that have been involved in or are aware of excess cancer concerns or investigations in their communities
- Community members representing environmental justice communities and tribes
- Academicians

The goal of the focus groups and key informant interviews was to gather feedback for ways to improve public health and/or environmental officials' communication and engagement with individuals and communities concerned about environmental hazards and unusual patterns of cancer. We held seven virtual community focus groups and three key informant interviews throughout the summer and fall of 2021.

Cancer Cluster Definitions, Characteristics, and Recent Investigations

Definition of a Cancer Cluster

The 2013 Guidelines defined 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." CDC/ATSDR revised the definition to recognize some cancers may be similar etiologically (in terms of risk factors, causes, or origin). The definition has been revised to: "A greater than expected number of the **same or etiologically related cancer** cases that occurs within a group of people in a geographic area over a defined period of time."

This definition can be further understood as:

- **a greater than expected number**: When the number of observed cases is greater than typically observed in a similar setting.
- **of the same or etiologically related cancer cases**: Cases are of the same type, are within a family of tumors (e.g., Ewing's family of tumors), or have a known or suggested link to the same specific environmental or chemical exposures. It is possible to consider multiple cancer types when such 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/or sex).
- **in a geographic area:** The geographic area may be based upon pre-existing geopolitical boundaries (e.g., census tract, county, or zip code). It may be defined according to the nature and extent of potential exposures that may cross multiple or partial boundaries. For example, air pollution from a hazardous waste incinerator which may cross multiple counties or census tracts. These geographic boundaries are used to determine the number of cancer cases as they relate to the total population in this predefined area. It is possible to "create" or "obscure" a cluster inadvertently by modifying the area of interest.
- **over a period of time**: The timeframe used to establish the beginning and end dates for analysis. 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.

Characteristics of Cancer

Cancer is a single term describing different diseases that share a similar characteristic: uncontrollable cell growth and division(5,6). As a group, cancers are very common. Cancer is not one disease, but rather many different diseases with different causal mechanisms (5). Cancers are the second leading cause of death in the United States, exceeded only by diseases of the heart and circulatory system. According to the American Cancer Society, one of every two men and one of every three women will be diagnosed with some form of cancer at some time in their life (6,7).

Given the frequency with which cancers are diagnosed, situations may arise where an unusual number of primary site cancers (i.e., place in the body where the cancer started) are diagnosed among people in a particular location. It is possible this may be due to chance. These unusual numbers may also result from the following:

- Differential screening practices

- Access to health care, which may be more reflective of other social and economic factors (e.g., limited access to optimal healthcare services)
- Behavioral risks and social determinants of health, occupational exposures, and in some cases, exposures to environmental sources
- Genetic susceptibility to a particular cancer

Although the causes of many cancers are unknown, some causal relationships have been shown between environmental exposures and development of cancer in specific organs (e.g., inhalation of asbestos and mesothelioma)(8). According to inputs received for these guidelines, health departments respond to inquiries that typically stem from concerns that are predominately environmental. More research is needed to understand these relationships further.

Recent Cancer Cluster Inquiries and Investigations

To learn about more recent cancer cluster investigations, the STLT survey asked health agencies to estimate the number of inquiries they received about excess cancer in the last 7 years (2013-2019):

- 53% received 1-5 inquiries each year
- 23% received 6-10 inquiries
- 15% received 11-25 inquiries
- 9% received more than 25 each year

Quick Fact

On average, agencies received around 9 inquiries (range 1–62) about excess cancer in 2019, and of these, about 7 (range 1-31) were from individual residents.

Respondents reported that inquiries were received from the following (respondents could select more than one):

- Individual residents (75%)
- Physicians and healthcare providers (36%)
- Community advocacy groups (23%)

In addition to the STLT survey, we conducted a comprehensive literature review to identify published studies investigating non-occupational or non-disaster potential cancer clusters in the United States within the last 10 years (January 1, 2010–April 8, 2020). The articles discussed epidemiologic investigations of cancer clusters and/or elevated cancer incidence in communities, and most of these investigations were prompted by a suspected environmental exposure. A brief summary of the articles is provided below.

Population

The populations within these investigations varied in size, age group, and by cancer types studied.

Study Design

Ecological study designs were most often implemented, followed by case-control and cross-sectional study designs.

Data Sources

Cancer registry data were used for cancer outcome data in the majority of studies; alternately, some studies used medical records as the source of cancer outcome data.

Methods
Potential exposures were ascertained using a variety of methods including surveys, environmental sampling, biological sampling, interviews, surrogate exposures (e.g., using the distance from a nuclear power plant to a ZIP Code as a proxy for individual exposure (9)), and pre-existing data from the Environmental Protection Agency and/or state health departments. The majority of studies used geospatial methods to identify spatial patterns or to generate standardized incidence ratio (SIR) maps.

Results
Results of the studies varied. While some were unable to identify associations between cancer and environmental exposures (9–11), several demonstrated statistically significant associations (12–18).

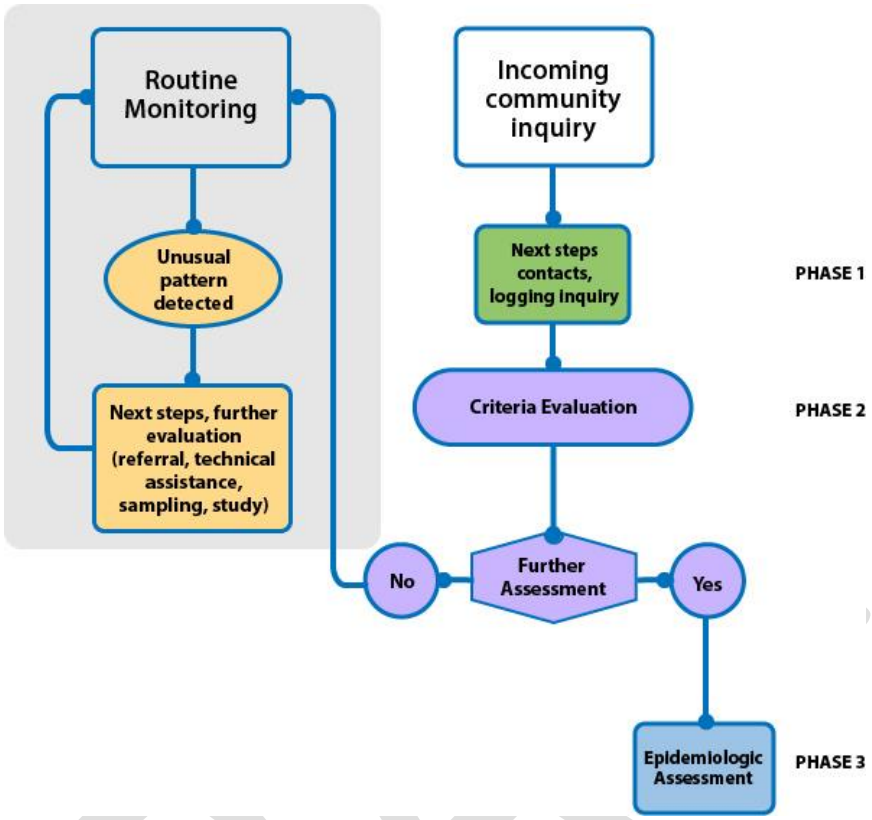
From this literature review, several articles reported a statistically significant association between the specified cancer and the exposure measure of interest but cited some limitations with interpreting the data. Of note, in epidemiology an association does not equate to causation. Many of these limitations are inherent of an ecological study design (see Phase 3 for more details). Limitations include the inability to measure the exact exposure biologically and instead relying on proximity to sites as an opportunity for exposure, the inability to measure lifestyle or behavioral risk factors that are associated with increased cancer risk, and challenges with modeling water or air movements that impact potential exposures.

Recognizing that the work of state health partners may be released as reports and not as published literature, a search of the gray literature was conducted to supplement the articles identified in the peer-reviewed literature. These reports aligned with feedback from the STLT and community focus groups and provided information relative to techniques and methodologies being used across public health agencies.

Proactive Evaluation and Routine Monitoring of Cancer Data

Typically, the approach to evaluating unusual patterns of cancers most commonly begins with an inquiry originating outside of a STLT partner organization and mostly from community concerns. While this is an important process, STLT partners have access to cancer data that can be evaluated in a proactive manner. Therefore, in addition to describing responsive approaches for addressing concerns raised by communities we encourage proactive evaluation of cancer registry data to monitor cancer trends and identify unusual patterns. Figure 1 outlines the enhanced process for evaluating patterns of cancer routinely and evaluating community inquiries about unusual patterns of cancer and environmental concerns. More details associated with the process are provided later in this document.

Figure 1. Summary of the revised guidelines and process.



Strengthening Partnerships

All states have a population-based cancer registry and conduct routine reviews of cancer incidence data. While cancer registry staff routinely review data, these reviews could be strengthened through partnerships with epidemiologists, statisticians, environmental health scientists, and other health officials who conduct community cancer investigations, if they are not cancer registry staff. These partnerships can facilitate the following:

- Awareness and attention to unusual patterns of cancer
- A more routine and robust evaluation of cancer data
- A more proactive (as opposed to responsive) approach to the evaluation of cancer data

The Value of Making Cancer Rates Publicly Available

Strong partnerships between population-based cancer registries and other public health professionals can also help more readily address community concerns and questions about unusual patterns of cancer within certain areas. One critical way to do this is for cancer registry, environmental health, and other staff to work together to make sub-state level cancer rates publicly available. For example, some states have provided data at a more geographically granular level, while using methods to preserve privacy and confidentiality, (e.g., county and census tract level cancer rates) as part of their CDC-funded

Important Reminders

CDC/ATSDR are available to provide technical support and provide assistance throughout all phases of the cancer investigation process.

state Environmental Public Health Tracking Program (19,20). Other states may not have the same resources and expertise, and they may not be able to publish their cancer data similarly. CDC's National Environmental Public Health Tracking Program publishes rates and is working to make standardized incidence ratios available on their Tracking Network Data Explorer (21), which utilizes interactive mapping tools. Having these data readily available, along with environmental data, such as on the Tracking Network, may allow states to proactively evaluate and routinely monitor potential issues that may warrant further assessment.

Evaluating Patterns of Cancer

Partnerships with cancer registry staff and access to registry data enable local health officials to calculate cancer-specific crude and adjusted rates for predetermined geographic locales such as counties and census tracts. Health officials can use these data to proactively evaluate and routinely monitor estimates of expected rates for cancers and explore how the rates compare to what is observed.

Cancer-specific, age-adjusted county rates should generally be compared to the same cancer-specific, age-adjusted rate estimated for that state (more details in Appendix A). Depending upon the number of cancers, rates calculated at sub-county levels, such as the census tract level, can similarly be compared to county rates and/or state rates. Evaluating these "observed" as compared to "expected" rates is also known as calculating a standardized incidence ratio (SIR) (22,23).

If the observed number of cases is the same as the expected number of cases, the value of an SIR is 1.0, sometimes multiplied by 100 for presentation (see Appendix A for additional guidance). The interpretation of the SIR is not straightforward. For routine evaluation, SIRs that deviate from 1.0 (or 100) may be examined further, with consideration of the confidence intervals (see Appendix A for additional guidance). In addition to the SIR, programs may consider visualization methods and techniques to examine cancer rates relative to the distribution of population characteristics and/or potential environmental risk factors (e.g., locations of National Priorities List sites)(24). Programs may apply spatial, temporal, or spatiotemporal methods designed to identify unusual distributions or unusual patterns to evaluate the distribution of cancers either in supplement to, or in parallel with, routine evaluation of the SIR (see Appendices A & B for more details).

If an unusual cancer pattern is identified, STLT programs responsible for proactive evaluation and routine monitoring of cancer rates should actively seek out other internal partners as appropriate (such as environmental public health programs or cancer control programs) to discuss findings, gather additional data and develop a follow-up plan (see Phase 1: Initiating Lines of Communication and Phase 2: Assessment of Criteria below). Processes should be in place in advance to inform communities when an unusual pattern is observed and include them in these discussions about how to move forward.

Additional Data Sources

Mortality data may be useful as a supplemental data source to the cancer registry data. State vital statistics registries provide access to cancer mortality rates. Additionally, the National Vital Statistics System publishes mortality data (25). Review of mortality data may be particularly helpful in addressing or understanding issues associated with disease burden or issues related to health equity such as differential access to care. In addition, evaluating mortality data along with cancer incidence data may provide insights regarding potential elevations in cancer that warrant further evaluation.

Electronic health record (EHR) data are becoming more readily available and may serve as an additional data resource in the future. EHR data may provide access to timelier data that are able to reflect more real-time reporting of cancer data. However, evaluations of the strengths and limitations of EHR as surveillance tools are warranted.

Responding to Community Concerns about Excess Cancer

STLT health officials often receive inquiries from the public about cancer occurrence within neighborhoods or communities. Community members may contact the health department with concerns about local environmental conditions that may or may not be related to the cancer(s) of concern. Initial contact with the person making the inquiry is a critical opportunity to understand their concerns. At this early stage, listening skills are paramount. This early interaction also allows the health official and community member to explore optimal community engagement strategies. In some cases, the inquirer is seeking information of a personal nature, often due to a recent cancer diagnosis in a family member or friend. In these situations, the health official can provide information directly over the phone or through email communication. In other situations, the inquirer may be seeking information on behalf of the community at large or a neighborhood or other community-based organization. Gathering additional information will likely be necessary.

Communicating with and Engaging Communities

According to the American Cancer Society more than 1,000 suspected cancer clusters are reported to state health departments within the United States each year (26,27). Although the specifics of each inquiry may vary, STLT public health agencies should be prepared for these recurring events.

Due to limits of epidemiologic and statistical methods, many cancer cluster investigations will be unable to establish a relationship between a specific environmental exposure and health outcome (27). In other cases, there may be no relationship. This presents communication and community engagement challenges.

Important Reminders

Establishing clear and ongoing communication channels about activities and challenges associated with the evaluation of unusual patterns of cancer and environmental concerns are important.

Establishing and Maintaining Trust

Generally, early interactions between a community member and a health official will determine the level of trust and credibility likely to follow throughout the assessment of health and environmental concerns. Establishing and maintaining trust throughout an inquiry depend on the following:

- Listening and understanding community concerns
- Ensuring equitable access to information
- Involving community members in the decision-making process

Different inquiries will require varying levels of community engagement. In many cases, an inquiry may satisfactorily end after an initial conversation with the concerned inquirer. The inputs received as part of the updates to this document also highlighted the need for trust between communities and health departments investigating community concerns about cancer.

Developing Communication Plans

STLT public health agencies should use cancer cluster inquiries as a broader opportunity to 1) engage with the communities they serve, and 2) support community efforts to implement policies, plans, and laws that impact health (including the removal of environmental burdens). Successful cancer cluster inquiries should aim for optimal community engagement and participation in order to build trust in the public health system.

To accomplish this, public health agencies should develop a cancer and environmental health communication plan and integrate risk communication principles throughout all steps of an inquiry. The plan should be clear and establish roles of health department staff including but not limited to identifying a lead for further assessment, a point of contact for community members seeking progress information and communications staff that can more broadly assist with disseminating information through social media and other communication channels established.

Using the Guidelines and Related Tools

This document provides suggestions for enhancing communication throughout the various phases of the assessment process. The 2013 Guidelines included a communication appendix encouraging proactive communication, community involvement, and transparency. In 2013, the National Public Health Information Coalition and CDC published a set of accompanying tools for state health departments to assist in engaging with communities. As with other tools and templates being developed, updates to the communication tools will be added to the CDC Cancer Cluster website. These tools can be used or adapted by STLT public health agencies to assist in their response. Additional resources are also provided on the CDC's Crisis and Emergency Risk Communication manual and ATSDR's website including the Community Stress Resource Center and the Community Engagement Playbook.

Using a Phased Approach to Respond to Community Inquiries

A phased approach to communicating with and engaging communities will help with the following:

- Initiating lines of communication with the inquirer and community members
- Examining data and criteria aimed at evaluating the occurrence of cancer
- Exploring further epidemiologic studies

Information about the phased approach is provided in the sections below, and additional details can be found in the Appendices.

Phase 1: Initiating Lines of Communication

The following section describes the information to collect (from the inquirer) and evaluate (using a decision-making template and other information).

Gather Information

1. Record the inquirer's information in a standard log when a call or email is first received. The log helps document if this is the first time the person is calling, if the geographic area or cancer of concern has been mentioned previously, and if there are multiple people raising concerns about the same geographic area or cancer type(s).
2. Use the Cancer Inquiry Intake Form, an electronic form provided on the CDC Cancer Clusters website, to gather the following key information:

- **Inquirer information**: name, residential address, email address, telephone number, length of residence at current location, organization affiliation (if any).
- **Inquirer information about the patterns of cancer**: types of cancer and number of cases of each type, age of people with cancer, geographic area of concern, time period over which cancers were diagnosed, and how the inquirer learned about concerns about cancer patterns
- **Other information**: any specific environmental concerns, other risk factors (e.g., occupation, behavioral risk factors, and family history of cancer), concerns in the affected area (e.g., the likely period of environmental contaminant exposures)

Important Reminders

If the inquirer requests anonymity, you should comply. But explain that the inability to follow up with the caller might hinder further investigation.

Keep in mind that the inquirer might not have information to differentiate between primary site and metastatic cancers and will most likely not be aware of all cases of cancer in the area or during the time period of concern.

3. Work with the inquirer to fully understand their range of concerns. In many cases the concerns associated with cancer may be prompted by other factors that warrant separate and/or related follow up (e.g., the presence of environmental hazards within a neighborhood or community).

4. States may elect to share information from the form (excluding personally identifiable information) with NCEH's Health Studies Section. Sharing this information may allow for evaluation of any potential regional or national trends requiring follow up and may allow federal officials to report the number of cancer cluster inquiries occurring nationally.

Establish Points of Contact

1. Establish the communication process (e.g., email, phone contact, or some other communication channel) as well as the timeframe for future contact.
2. Determine the primary community Point of Contact (cPOC) and gather that individual's contact information. The cPOC may be the inquirer. Alternatively, another individual identified by the inquirer may serve as the cPOC if they are viewed as a trusted representative of the inquirer and/or the community.
3. Determine the agency Point of Contact (aPOC) and the best method to reach them directly.

Important Reminders

Listen to the inquirer and ask questions. Don't dismiss the inquiry before gathering information.

The aPOC should have strong empathy skills and experience in cancer epidemiology and/or experience in environmental science.

Do not forget to log and record all actions taken regarding the inquiry. Also, relay this information to the inquirer. This communication about the information evaluated and the actions taken should be open, transparent, and thorough.

4. Provide an estimated timeline of when the aPOC will re-contact the inquirer/cPOC.

Collaborate with Others

1. Establish a team of colleagues to help address concerns over cancer patterns prior to receiving inquiries.
2. Have the aPOC reach out to collaborators to plan initial discussions about the cancer inquiry. Collaborators could include other staff within the health department (i.e., chronic disease and cancer control programs, environmental health, cancer registries, and geospatial specialists).

Important Reminders

Phase 2 is meant to build on the information gathered in Phase 1.

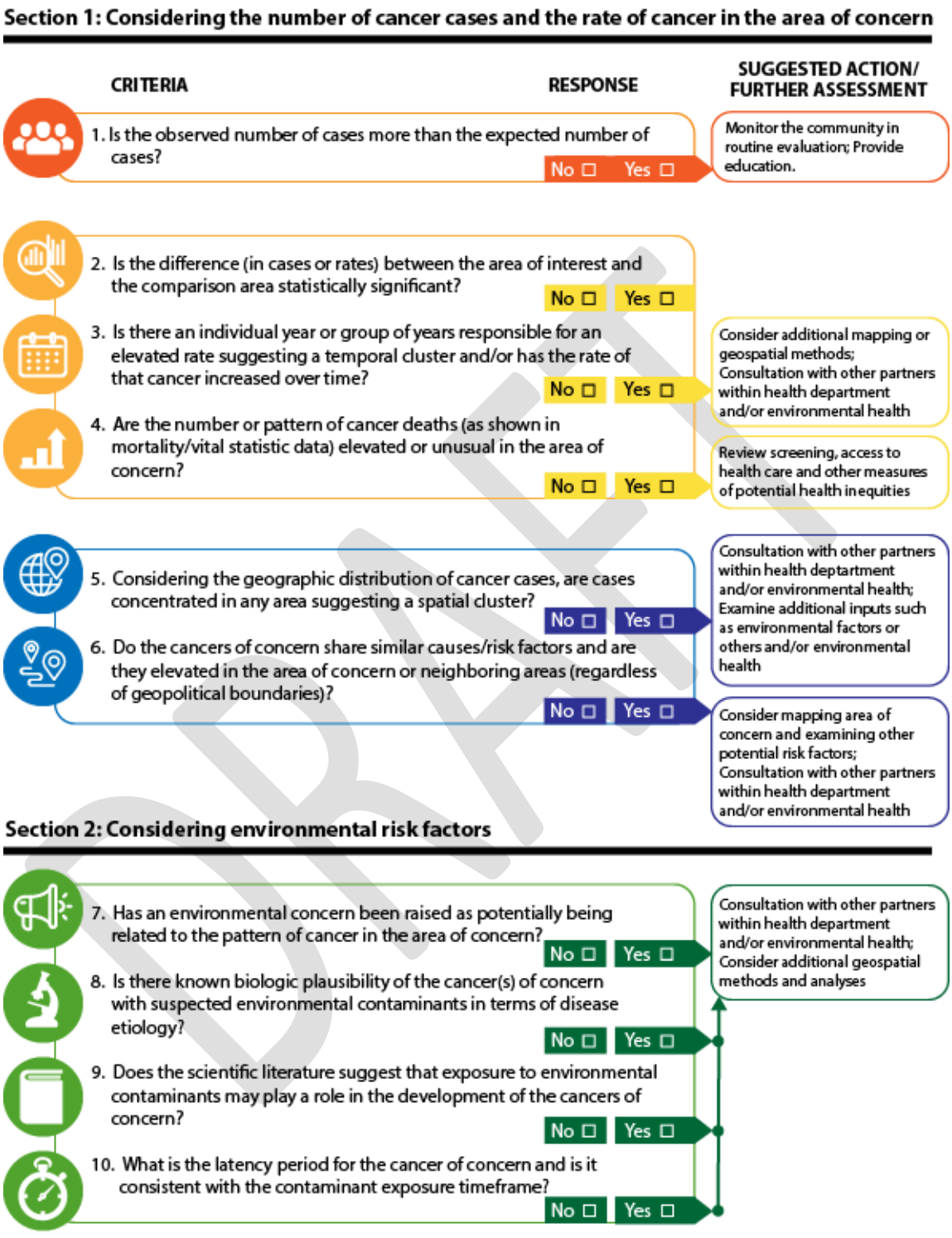
In many cases, Phase 1 inquiries will move on to Phase 2 assessment. But, in some cases, communication with the inquirer over the phone or through email will address the inquirer's concern.

3. Reach out to the state environmental regulatory agency if the inquiry is primarily regarding an environmental contaminant.
4. Request technical assistance, if necessary. The following groups may provide technical assistance:
    - CDC/NCEH Health Studies Section
    - ATSDR Office of Community Health and Hazard Assessment including regional staff
    - Regional Pediatric Environmental Health Specialty Units (PEHSUs)

Phase 2: Criteria to Determine Continued Assessment of a Report of an Unusual Pattern of Cancer

This section outlines the process for reviewing data and criteria for decision making to determine continued assessment. The following criteria are meant to assess the cancer of concern and/or related environmental risk factors. The form below (Form 2) is included on the CDC Cancer Cluster website to be used as a decision-making guide for assessing the criteria and a discussion on how to evaluate the criteria follows.

Form 2- Cancer Cluster Decision Making Form



Proceeding based on Responses to Criteria

Use the tables below to determine your next actions.

If ALL answers to sections 1 and 2 above are "No"

1. **No further assessment needed** at this time.

2. **Summarize collected information** in a written report or letter; provide summary to the inquirer/cPOC. The summary should include: ❒
   - Background information on patterns of cancer observed (rates and geography)
   - An explanation of how the agency investigated the inquiry about unusual patterns of cancer
   - A review of findings regarding the cancer of concern
   - A discussion of risk factors for the cancer(s) mentioned in the original inquiry
   - Agency plans or next steps based on the findings
   - A note or reference about routine monitoring and follow up

3. **Continue routine monitoring**, and plan to monitor this cancer and area of concern in routine evaluations of cancer data (to determine if the pattern changes): ❒
   - Conduct routine monitoring aimed at identifying unusual patterns of cancer through the use of geospatial/ statistical tools (and data that may be routinely available at geographic levels lower than the state as a whole).
   - Maintain the feedback loop with the original inquirer and establish procedures for future updates (e.g., making more geographically granular data available on the health department website).

If answer to number 1 is "Yes," AND answer to any other question in section 1 is "Yes"

Further assess the cancer pattern. This assessment may include consultation and/or referral to a cancer prevention and control program for consideration of intervention activities or an epidemiologic case series analysis. Further assessment could also include defining the geographic area or timeframe of concern using spatial and temporal methods. ❒

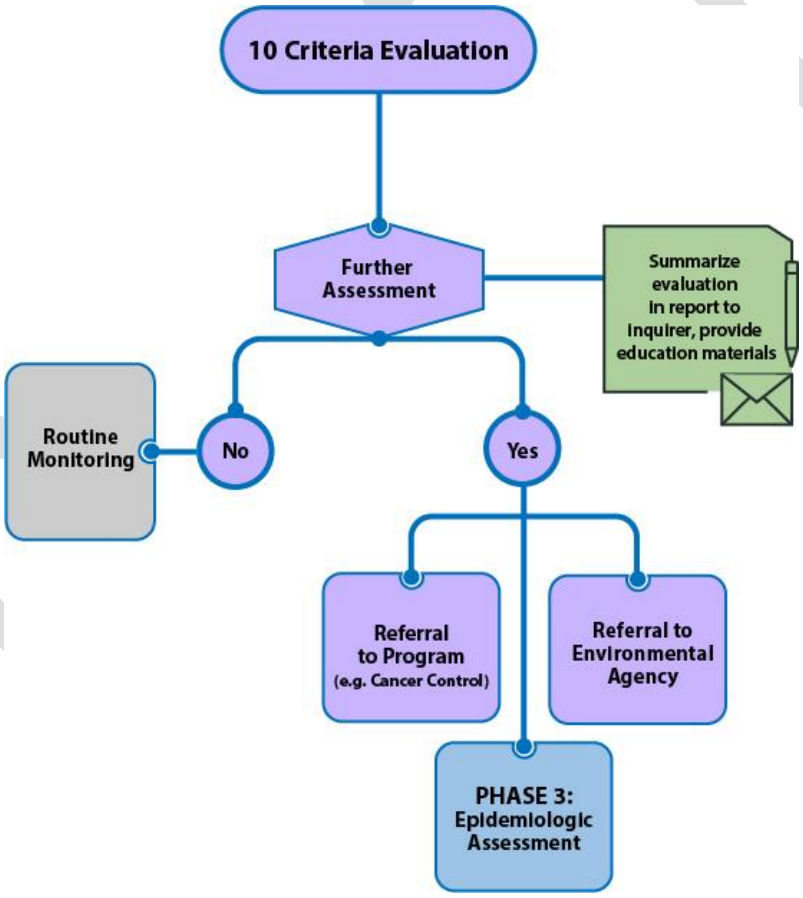
If answer to any question in section 2 is "Yes"

Consider information from section 2 (environmental data and environmental risk factors) during the additional assessment and/or referral to another agency (e.g., the local water department, the state environmental regulatory agency, ATSDR). ❒

In all of the scenarios presented above, you should also do the following:

- **Communicate with the inquirer** (about types of cancer and area of concern) to help raise cancer awareness and provide a state/local health official the opportunity to explain how cancer rates are calculated.
- **Summarize the information from the decision template form** (after completing the Phase 2). This summary should describe the information evaluated and what type of follow up steps might be taken. The summary should also commit the responding STLT program to continue monitoring the pattern of the cancer(s) of interest in the area of concern to determine whether the pattern of cancer changes.
- **Consider presenting the information in a community meeting** or public availability session, prior to issuing any type of media release (if, for example, the inquiry was made on behalf of a neighborhood or organized group).

Figure 2. Summary of Phase 2 Process.



Considerations when Reviewing Data

1. **Use the statistical significance of the cancer rate (standard incidence ratio) as part of the criteria discussed below to help determine next steps** in the investigation. While the statistical significance should not be the deciding factor, it does provide some measure of whether cancer seems to be occurring unusually in time, space, or both. Additional information for interpretation of results, including a discussion of confidence intervals, is provided in Appendix A.
2. **Evaluate the geographic distribution** of cancer cases within the area of concern to determine the likelihood of a spatial cluster. Qualitative and quantitative evaluation of the data is important in this process.
3. **Look at incident cases for each year** (if possible, for the cancer type) within the 5- or 10-year period used to calculate the SIR or standardized mortality ratio (SMR) to determine the potential for a temporal cluster. For example, are the cases over that given time period represented by cases within a 1- or 2-year part of the overall period suggesting a temporal trend? This type of evaluation of time trends can also be instructive in better understanding the potential for some factor (or set of factors) to have played a role in cancer occurrence.
4. **Evaluate the pattern of cancers of concern in neighboring areas** (e.g., census tracts or neighborhoods) and border communities within and across state lines. This may identify regional trends and patterns. This is of particular importance if an environmental concern is also expressed in the neighboring area. This approach provides an opportunity to evaluate geographic patterns that may emerge without reliance on geopolitical boundaries. Ensure that areas that are similar demographically are not also similar in terms of environmental hazards and contamination. Ensuring such appropriate comparisons can help address issues of environmental justice and health equity. See Appendix B for more about spatial and temporal methods to quantitatively evaluate unusual patterns.
5. **Collaborate with internal and external partners,** as needed, if an environmental concern is identified (or suggested). Collaborate with these partners to acquire or review environmental data. Remember, the community member/inquirer may provide useful information associated with environmental concerns, such as historic land use of the area and/or concerns about potential contamination related to former use of a property. Census data can indicate population changes over time, but the inquirer may also have qualitative, historic information relative to population change. This information on population change and land use can have significant impacts on decisions associated with the likelihood that exposure to environmental hazards might be associated with cancer occurrence, particularly as it relates to cancer latency/development periods as well as biologic plausibility. Understanding the biologic plausibility of the cancer(s) of concern being associated with the environmental factors is important for investigatory purposes as well as for risk communication. Consideration of the possible exposure routes is a principal factor in understanding whether cancer and exposure to environmental agents are potentially connected.

Communication Reminders

- Review with the inquirer what data are included in both the numerator and the denominator and explain how these data are used to calculate an observed rate and an expected rate of cancer (resulting in a standardized incidence or mortality ratio [SIR/SMR]).
- Describe the process used to compare cancer rates to address concerns about cancer patterns in a given area. Discuss why rates are generally calculated in 5- or 10-year intervals to enhance statistical stability due to the small number of annual events in smaller geographic areas.
- Comparing the cancer rate within a particular census tract/neighborhood, community, or county to the state as a whole is standard practice for STLT health officials addressing concerns over cancer patterns in a given area.
- Explain how comparing cancer patterns across communities and/or counties is important.

Concerns associated with occupational exposures should be discussed with state and federal occupational health officials.

6. **Periodically, community inquiries may involve concerns about unusual patterns of cancer(s) in a school setting.** Work closely with school officials to determine the cancer types of concern, the population at risk (e.g., students, staff), and the time period during which the cancers were diagnosed. To calculate a cancer rate among this population, determine the total population that should be included in the calculation. Determine the school population (specific to age and sex) for each year during the period of concern (e.g., beginning the year of the first cancer diagnosis and ending with the year of the most recent cancer diagnosis). Consider also any spatial or temporal patterns that might be identified (i.e., were all cases diagnosed among individuals spending time in one area or floor of a school building, were diagnoses more frequent in a particular group of years?). Once this information is available, the approach to assessment is the same as for addressing community-based concerns over unusual patterns of cancer.
7. **Community members may have a list of individuals diagnosed with cancer or a map presenting individual addresses.** While health officials cannot publicly confirm whether individuals on such a list or map have cancer (due to patient confidentiality and privacy laws), the information can be used to help identify the types of cancer(s) of concern as well as the time period and geographic area of interest. This data can be used to supplement information obtained through the cancer registry. This type of information must be confirmed through the cancer registry data or outreach to medical providers.

Response Associated with a Need for Further Assessment

As mentioned above, a report should be prepared that summarizes all of the steps taken to evaluate the information and concerns initially reported and why further assessment is advised. The report should include the following:

- Description of criteria used to determine whether or not the pattern of cancer appears unusual
- Assessment of environmental concerns (if any) that have been expressed and/or explored
- Clear description of next steps, if any
- An estimated timeframe for when the inquirer might learn of new information

The report should also include the following information, as appropriate.

If this is true…		Then include this information
If the next steps are associated with actions the responding program must undertake (e.g., a case series analysis, case-control study, etc.),	→	then describe the actions that must be taken to undergo such an analysis.
If the further assessment involves action by another program within the health agency,	→	then include the program and a contact person.
If the action(s) to be taken involve an agency external to the health department,	→	then provide the agency, program name, and a point of contact.

Establishing Strong Relationships and Partnerships with the Community

When further assessment activities are warranted, develop and implement a community engagement plan. Developing a plan together with community partners to share information on a regular basis regarding cancer and environmental health investigations will enhance trust and credibility. Consider including the following people:

- Staff within the health department with expertise in establishing strong relationships, such as public health nurses and/or social workers (i.e., to serve as aPOCs)
- Local medical professionals, such as Pediatric Environmental Health Specialty Unit (PEHSU) clinicians
- A trusted community member (i.e., to serve as a cPOC)

Important Reminders

CDC/ATSDR are available to provide technical support and assistance throughout all phases of the cancer investigation process.

Phase 3: Considerations for Epidemiologic Studies

Getting Started

1. **Schedule meetings** once the decision has been made to obtain more data to consider epidemiologic studies. Use these meetings to discuss the initial findings and proposed approach for further assessment with the initial inquirer and other community members.
2. **Provide a detailed description of the next steps** for evaluation. These next steps should be based on summary findings from the analysis of the 10 criteria in Phase 2.
3. **Develop a summary** that includes a rationale for continuing the investigation. For example, the assessment identified a potential spatial cluster or environmental factors that warrant further exploration.

Establish a Community Advisory Committee

A community advisory committee (CAC) can:

- Contribute valuable sources of information and liaise with other community members:
    - Provide specific details on the community, including help with cultural sensitivities or knowledge of current and/or historical environmental concerns in the area
    - Serve as a communication link back to the community at large

Open communication will be needed to set goals, establish timelines, and discuss other issues (e.g., possible study limitations) so that expectations are clear. Every community is different; however, agencies may be able to work with some existing organizations (e.g., neighborhood or religious organizations) to establish a CAC.

CAC and other community members can provide important insight and assist investigations:

- Help promote participation in studies and conduct outreach
- Identify specific tasks for community members that would contribute to the success of the investigation

- Identify additional data sources within the community (e.g., local lists that may be maintained which would identify previous residents)
- Identify population mobility patterns and housing development considerations

Consider the Following for Study Design and Protocol Development

The following sections highlight the most important considerations for study design and protocol development.

Consider developing a peer review committee with external partners to assess study design issues. Committee members should have a breadth of related skillsets in epidemiology, cancer, biostatistics, toxicology, and environmental health. The final draft of the protocol should be shared with the CAC in advance of initiating the study.

Case Definition

Generate a formal case definition ahead of further analyses. Include the type of cancer, the population, the geographic area of concern, and the time period of interest. Review the study area and revise (if necessary) before starting additional analyses. For example, early assessments may have indicated a particular contaminant of concern and more populations were found to be at potential risk for exposure to this contaminant. Thus, the study area would be expanded to include the other populations who were at potential risk for the exposure of concern.

Hypothesis

A hypothesis regarding cancer cases or potential environmental contaminants was likely already generated during the assessment phase; however, hypothesis generation can be an ongoing process, particularly if a potential exposure source has not been identified (23). Re-review and update hypotheses regarding the cancer cases and contaminant of interest (i.e., based on observations within the assessment phase to formulate analyses a priori). The hypothesis helps to guide the collection of data, the analysis plan, and interpretation of results.

Study Population

Define the population of interest for the study. Ask the following questions to help define the study population as well as the time period of interest:

- How are the boundaries defined for the geographic area of interest?
- What are the characteristics of the population? Are there specific characteristics that are noted for the cancer cases (i.e., age group, race/ethnicity)?
- What is the latency period for the particular cancers of interest? What years of residence would need to be included to reflect the latency/development period assumed for the cancers of concern?
- How long have the residents lived in the area? What is the mobility/migration of former residents?

Latency

Latency and change of residence add to the complexity of these investigations. Because residential history data are generally not available for the vast majority of cancer registries, collecting such data are critical as part of any epidemiologic study. Given the long latency period associated with cancers in

adults, behaviors and exposures that might have contributed to the development of cancer in a person typically occur years to decades before the diagnosis. For example, malignant mesothelioma, a tumor of the lining of the lung, is associated with asbestos exposure and the latent period between first exposure to asbestos and death from mesothelioma is often 30 years or longer (28).

Latency in an epidemiologic investigation influences the exposure period relevant to the investigation. For example, 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 would not likely be related to an exposure in the area of concern. 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 may 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.

Because childhood cancers generally have shorter latency periods than adults, changes of residence might be less of an issue in the investigation of unusual patterns of childhood cancers. However, childhood cancer cluster investigations may have the same limitations as adult counterparts. 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 (29), indicating that even among cancers with short latency periods, migration might be an important factor. Account for latency when designing any additional analyses or studies.

Additional environmental data

If additional environmental data are needed, partner with state or local environmental regulatory agencies. ATSDR regional offices can also provide technical assistance, such as, review of environmental sampling plans. Outline the types of sampling needed as well as resources available or necessary to conduct such sampling. Also, consider the following:

- Potential sources and routes of exposure (e.g., through air inhalation, water ingestion)
- Contaminant transport (e.g., if exposure via ambient air, meteorological factors such as wind direction and speed are important)
- Other contributing sources of exposure
- Boundaries of potential spread/movement (important to understand the potential at-risk population)
- Role of participating agencies in data collection and/or identification of resources to support the collection of environmental data.

Existing health data for case finding

Primarily, cases of cancer among the study population are identified from a state's cancer registry, using the case definition. Consider multiple existing data sources for use to help identify all possible cases. Form partnerships with Health Information Exchange or Network entities that may provide access to EHR.

Contributing risk factors

An environmental factor may have been identified as a potential risk factor for the cancer of interest; however, cancers may be caused by several different risk factors or through multiple causal mechanisms

(30). Review additional risk factors for the type(s) of cancer under investigation, such as social determinants of health, behavioral risks, occupational exposures, and in some cases genetic factors.

Types of Epidemiologic Studies to Consider

Methods used in environmental epidemiology are observational, not experimental. This makes the process of identifying a causative agent more difficult because researchers may not have access to information about other exposures or other variables that can create bias in the analysis (22,23,31). However, additional studies can potentially help make inferences about cancer cases and potential associations. Consider the following types of studies:

- **Case-series study**. A case-series study is a descriptive analysis of persons with a similar cancer diagnosis. This study can be particularly helpful when there are small numbers of cases and it is not feasible to pursue other studies (e.g., statistical challenges associated with a small number of individuals diagnosed with the disease). Case-series studies are often designed to collect more information about each person to identify any commonalities. For these studies, identify as many relevant cases as possible to avoid selection bias. Depending on the findings, a further analytical study may be possible/warranted (see case-control or cohort below).

- **Ecological study**. An ecological study compares aggregated environmental data to aggregated cancer data to examine general associations. Exposures and outcomes are generally grouped to a geographic area such as a census tract for comparison. Ecological studies are most often done initially to explore potential associations but alone cannot determine causality. Epidemiologists 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.

- **Case-control study**. Consider a case-control study when the etiology of the cancer is unknown. This type of study can also be used to assess the association between cancer and a presumed exposure while being able to control for confounding factors. A case-control study collects information from cancer cases meeting the case-definition and controls within the same study area. Exposures and risk factors are compared between the two groups, for example using regression analysis methods. Case-control studies are suggested for rare cancer outcomes. Unlike the ecological study, case-control studies allow for the collection of individual level data and risk factors to assess within the analyses. The primary disadvantages of a case-control study are the inability for individuals to recall historic events (e.g., exposures) and to provide a direct estimate of risk. Risk is often estimated as an odds ratio, showing the odds for a particular cancer to have occurred given a particular exposure, compared to odds of the same cancer without the exposure. For this type of study, the sample size needed should be calculated ahead of time considering the power needed to detect statistical differences between the populations (Appendix A).

- **Cohort (retrospective or prospective)**. Consider a cohort study when the exposure source is known or being investigated. Select people who have varying levels of exposure for the study. Prospective cohort studies, while one of the strongest types of studies to examine exposure-disease relationships, are very time consuming and expensive. Depending on the latency of the cancer of concern, a prospective cohort study may need to be 10-20 years or longer to collect enough data for analyses. The length of time needed often must account for study attrition and

thus require a very large sample size. A retrospective cohort study may be feasible, but also is reliant on time-related information. For example, a cohort study may want to compare a population exposed to chemicals from Factory X, which was operating in the early 1990's. To establish a study cohort, a population within that community would have to have been in the study area ~30 years previously. Because of the length of time needed for these studies, some challenges may arise that need to be accounted for within the study design. For example, a community may have experienced a substantial percentage of movement or migration. Also, as mentioned above, the latency of the cancer will also need to be accounted for within the study design. For example, the average latency for mesothelioma is generally thought to be decades (32). More recently, data from the World Trade Center Health Program suggests a minimum latency for mesothelioma of about 11 years and the minimum latency for thyroid cancer is 2.5 years (30). Studies assessing these types of cancers would have to account for the number of years that specific types of cancers develop following exposures. Additionally, limitations exist with quantifying historic exposures for retrospective cohort studies.

Other Considerations for Your Study

When evaluating the feasibility of different epidemiologic studies, consider the following:

- **Funding**. Funding may be needed to help with the investigation. Seek state and federal opportunities to assist with funding the investigation and/or providing technical assistance. In addition, academic partnerships may help to support the investigation.
- **Outreach**. For more robust analyses, reach out to other states to discuss case-finding and/or environmental sampling activities, specifically neighboring states if suspected environmental contaminants cross state borders.
- **Communication**. Ultimately, the goal of an epidemiologic investigation of unusual patterns or excesses of cancer is to understand the potential relationship between environmental risk factors identified and observed cases of cancer. Some epidemiologic investigations have identified associations between exposure to certain chemicals and certain cancers (33,34). However, methodological limitations and data limitations (such as unknown levels of exposure) often limit our ability to demonstrate that such relationships exist. Regardless of the outcome of an epidemiologic investigation, continuing to communicate with individuals most concerned about exposures and health outcomes is important.

CDC/ATSDR can provide technical assistance and guidance:

- ✓ Contact us at CCGuidelines@cdc.gov for guidance on developing and conducting studies.
- ✓ Visit the CDC Cancer Cluster Guidelines website (www.cdc.gov/nceh/clusters) to access tools and templates, cancer resources and education materials, and trainings for staff.

Appendix A: Statistical Considerations

This section provides general guidance regarding the use of epidemiological and statistical analysis methods associated with addressing concerns related to reports of unusual patterns of cancer. This section focuses on the descriptive statistics and epidemiologic methods most commonly used to assess occurrences of excess cancer. Frequencies, proportions, rates, and other descriptive statistics are useful first steps in evaluating the suspected cancer cluster. These statistics can be calculated by geographical location (e.g., census tracts) and by demographic variables such as age category, race and ethnicity, and sex. Comparisons can then be made across different stratifications using statistical summaries such as ratios.

Standardized Incidence Ratio

The standardized incidence ratio (SIR) is often used to assess whether there is an excess number of cancer cases, considering what is "expected" to occur within an area over time given existing knowledge of the type of cancer and the local population at risk. Simply stated, the SIR is a ratio of the number of observed cancer cases in the study population compared to the number that would be expected if the study population experienced the same cancer rates as a selected reference population (typically the state as a whole is used as a reference population). The CDC/NCEH National Environmental Public Health Tracking Program will make these rates available at the county level. The general equation is as follows:

$$SIR = \frac{Observed\ Cancer\ Cases\ (O)}{Expected\ Cancer\ Cases\ (E)}$$

Adjusting for Factors

The SIR can be adjusted for factors such as sex, race, and/or ethnicity, but it is most commonly adjusted for differences in age between two populations. In cancer analyses, adjusting for age is important because age is a risk factor for many cancers and the population in an area of interest could be on average younger or older than the reference population (35,36). In these instances, comparing the crude counts or rates would present a biased comparison.

For more guidance, 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) (22,23,37–41). There are two ways that are generally used to adjust via standardization, an indirect and a direct method. An example of one method is shown below, but a discussion of other methods is provided in several epidemiologic textbooks (22) and reference manuals (42).

An example is provided in the table below, adjusting for age groups. The second column, denoted with an "O", is the observed number of cases in the area of interest, which in this example is a particular county within the state. The third column shows the population totals for each age group within the county of interest, designated as "A." The state age-specific cancer rates are shown in the fourth column, denoted as "B." To get the expected number of cases in the fifth column, A and B must be multiplied for each row. The total observed cases and the total expected cases are then summarized.

Age group	Observed Number of Cases in County* (O)	County of Interest Population (A)	State Age-Specific Cancer Rate† (B)	Expected Cancer Cases (A x B= E)
40-49	50	20,000	0.001	20
50-59	150	23,000	0.007	161
60-69	200	25,000	0.010	250
70+	250	15,000	0.015	225
Total	**650**			**656**

*Number of cases in a specified timeframe.

† Number of cases in the state divided by the state population for the specified timeframe. Rates are typically expressed per 1000 or 100,000 population.

The number of observed cancer cases can then be compared to the expected. The SIR is calculated using the formula below.

$$SIR = \frac{Observed\ Cancer\ Cases\ (O)}{Expected\ Cancer\ Cases\ (E)}$$

$$SIR = \frac{650}{656} = 0.99$$

Confidence Intervals

A confidence interval (CI) is one of the most important statistics to be calculated, as it helps to provide understanding of both statistical significance and precision of the estimate. The narrower the confidence interval, the more precise the estimate (23).

A common way of calculating confidence intervals for the SIR is shown below (23):

$$95\%\ CI = \frac{(\sqrt{Observed}\ \pm 1.96/2)^2}{\text{Expected}}$$

Example from above:

$$95\%\ CI = \frac{(\sqrt{650}\ \pm 1.96/2)^2}{656} = (0.92\ , 1.07)$$

If the confidence interval for the SIR includes 1.0, the SIR is not considered statistically significant. However, there are many considerations when using the SIR. Because the statistics can be impacted by small case counts, or the proportion of the population within an area of interest, and other factors, the significance of the SIR should not be used as the sole metric to determine further assessment in the investigation of unusual patterns of cancer. Additionally, in instances of a small sample, exact statistical methods, which are directly calculated from data probabilities such as a chi-square or Fisher's exact test can be considered. These calculations can be performed using software such as R, Microsoft Excel, SAS, and STATA (41). A few additional topics regarding the SIR are summarized below.

Reference Population

The reference population used for the SIR could be the surrounding census tracts, other counties in the state, or the state as a whole. Selecting the appropriate reference population is dependent upon the hypothesis being tested and should be large enough to provide relatively stable reference rates. Decisions about including the state as a whole as the reference population, should be made prior to calculating the SIR. Another issue to consider is the size of the study population relative to the reference population. If the study population is small relative to the overall state population, including the study population in the reference population calculation will not yield substantially different results. However, excluding the study population from the reference population can yield more precise estimates. If the reference population is smaller than the state as a whole (such as another county), the reference population should be "similar" to the study population in terms of factors that could be confounders (like age distribution, socioeconomic status and environmental exposures other than the exposure of interest). However, the reference population should not be selected to be similar to the study population in terms of the exposure of interest. Ensuring such appropriate comparisons can help address issues of environmental justice and health equity.

Limitations and Further Considerations for the SIR

One difficulty in cancer cluster investigations is that the population under study is generally a community or part of a community, leading to a relatively small number of individuals comprising the total population (e.g., small denominator for rate calculations). Small denominators frequently yield wide confidence intervals, meaning that estimates like the SIR are not as precise as desired (37). Other methods, such as qualitative analyses or geospatial/spatial statistics methods, can provide further examination of the cancer and area of concern to better discern associations. Further epidemiologic studies may help calculate other statistics, such as logistic regression or Poisson regression. These methods are described in Appendix B.

Alpha, Beta, and Statistical Power

Another important consideration in cancer cluster investigations is the types of errors that can occur during hypothesis testing and the related alpha, beta, and statistical power for the investigation. A type I error occurs when null hypothesis is rejected but actually true (e.g., concluding that there is a difference in cancer rates between the study population and the reference population when there is actually no difference). The probability of a type I error is often referred to as alpha or α (43).

$$\alpha = \text{Probability}(\text{reject Ho} | \text{ Ho is true})$$

A type II error occurs when the null hypothesis is not rejected and it should have been (e.g., concluding that there is no difference in cancer rates when there actually is a difference). The probability of a type II error is often referred to as beta or β.

$$\beta = \text{Probability}(\text{do not reject Ho} \mid \text{Ho is false})$$

Power is the probability of rejecting the null hypothesis when the null hypothesis is actually false (e.g., concluding there is a difference in cancer rates between the study population and reference population when there actually is a difference). Power is equal to 1-beta. Power is related to the sample size of the

study—the larger the sample size, the larger the power. Power is also related to several other factors including the following:

- the size of the effect (e.g., rate ratio or rate difference) to be detected,
- the probability of incorrectly rejecting the null hypothesis (alpha), and
- other features related to the study design, such as the distribution and variability of the outcome measure.

As with other epidemiologic analyses, in cancer cluster investigations, a power analysis can be conducted to estimate the minimum number of people (sample size) needed in a study for detection of an effect (e.g., rate ratio or rate difference) of a given size with a specified level of power (1-beta) and a specified probability of rejecting the null hypothesis when the null hypothesis is true (alpha), given an assumed distribution for the outcome. Typically, a power value of 0.8 (equivalent to a beta value of 0.2) and an alpha value of 0.05 are used. An alpha value of 0.05 corresponds to a 95% confidence interval. Selection of an alpha value larger than 0.05 (e.g., 0.10: 90% confidence interval) can increase the possibility of concluding that there is a difference when there is actually no difference (Type I error). Selection of a smaller alpha value (e.g., 0.01: 99% confidence interval) can decrease the possibility of that risk and is sometimes considered when many SIRs are computed. The rationale for doing this is that one would expect to see some statistically significant apparent associations just by chance. As the number of SIRs examined increases, the number of SIRs that will be statistically significant by chance alone also increases (if alpha is 0.05, then we expect 5% of the results to be statistically significant by chance alone). However, one may consider this fact when interpreting results, rather than using a lower alpha value (44). Decreasing the alpha value used will also decrease power for detection of differences between the population of interest and the reference population.

In many investigations of suspected cancer clusters, the number of people in the study population is determined by factors that may prevent the selection of a sample size sufficient to detect statistically significant differences. In these situations, a power analysis can be used to estimate the power of the study for detecting a difference in rates of a given magnitude. This information can be used to decide if or what type of statistical analysis is appropriate. Therefore, the results of a power calculation can be informative regarding how best to move forward.

Author Acknowledgements:
Andrea Winquist, Angela Werner

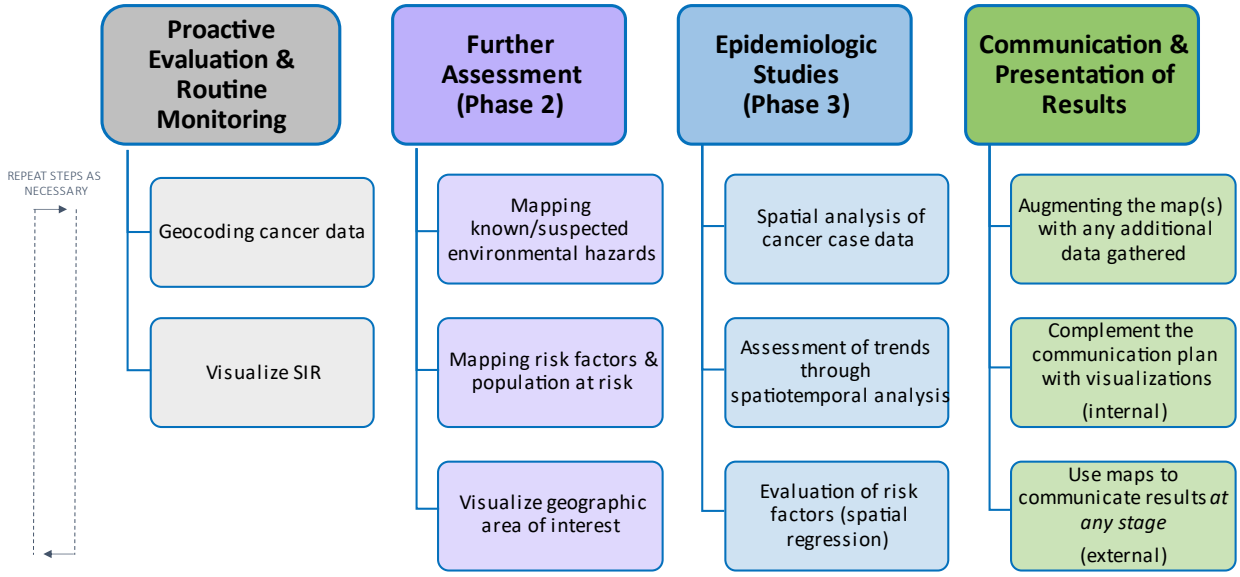
Appendix B: Mapping and Spatiotemporal Methods

Geospatial Visualization and Analysis

Geographic information systems (GIS) can be useful for all stages of evaluating unusual patterns of cancer. GIS may be used as part of proactive evaluation of cancer registry data and during Phase 2 assessments. Spatiotemporal regression methods are particularly useful for identifying and quantifying the relationships between risk factors and cancer cases during epidemiologic investigations (Phase 3). These processes are summarized in Figure 1.

Visualization, or mapping, can be used as an important communication tool to both internal and external stakeholders. As part of the routine evaluation of cancer data, maps can be shared with partners in programs such as comprehensive cancer control and environmental health for decision making. Additionally, maps can convey important information about the distribution of cases and potential environmental risk factors when engaging with the community.

Figure 1. GIS Activities Throughout a Cancer Cluster Investigation



Often, a first step in visualization and spatial analysis involves translating addresses collected as text in cancer registry data into coordinates that can be mapped. This process is known as geocoding, and the quality of the result is crucial as it is the basis for visualizations and analyses. Resources are available that provide detailed descriptions of this process (45–47) and tools are available at no cost for health departments from the National Cancer Institute as well as the North American Association of Central Cancer Registries (NAACCR) Geocoder.

Cancer registries and state health agencies typically have criteria related to release of data for small geographic areas. Because of confidentiality and privacy concerns, some data cannot be released to the public, unless the concerns are addressed. For example, a map of a small geographic area that identifies

the residences of cancer patients as points should not be made public (35). Similarly, many health agencies are prohibited from publicly releasing a table for a small geographic area with a small population, since each table cell might have only a few cases and could be used to identify individuals.

Once data are geocoded, they can be mapped along with other geographic data, such as suspected environmental risk factors, for crude assessments of their proximity to the cases. Different spatial (e.g., 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. This practice is more useful when longer periods of time are under study, as well as when there are larger numbers of cases (e.g., >10 cases). Mapping and analyzing the data over space and time can help reveal whether changes in incidence or mortality rates are observed and may suggest risk factors that warrant further consideration.

Varying the geographic scale or the geographic unit of aggregation can produce different patterns and results. This is known as the modifiable areal unit problem (48,49), which has also been identified relative to temporal aggregations (50). Multiple methods have been proposed to account for these issues (51–53); a common solution is to use differing scales. In doing so, variations in results can be identified if they exist.

The following section provides basic information on the principles of clustering methods and statistical considerations when working with spatially structured data. The National Cancer Institute (NCI) provides a variety of methods and tools for analysis of cancer statistics. Some of the tools are more statistical in nature, such as tools to calculate incidence and mortality rates and trends (available in SEER*Stat), while other visualization and analysis tools are geospatially focused. Links to the tools and information for further exploration are available at https://surveillance.cancer.gov/tools/. In addition to these freely available tools, the sections below detail other spatial and statistical methods and software that can be used for these analyses.

Clusters

Clusters can be detected by a variety of techniques that evaluate whether similar features, values, or observations are “close” or in close proximity to one another. These techniques can be divided into global, local, and focused methods. A table with methods and associated applications for those categories is available on the CDC Cancer Cluster website. The table is not meant to be a comprehensive review of applications but rather to provide initial guidance.

Global clustering statistics can be used to determine if there are patterns of clustering anywhere in the study area. Once clustering is deemed likely from global statistics, local clustering methods, including scan statistics, can help to identify clusters within the area of interest. It is worth noting that it is possible to detect statistically significant global clustering without evidence of local clustering and vice versa (54). In cases where there is a known point-source, focused tests can be considered. Regression analysis can then be used to understand the association between potential environmental risk factors and cases or to adjust for confounding factors such as latency in cases, mobility, and demographic variables (such as age and race). These methods are further described below along with example use cases and locations of available software for analysis.

Global Clustering Methods

Global clustering statistics detect patterns of spatial clustering that occur 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. One measure of global clustering is spatial autocorrelation, which is the degree of similarity of nearby features. Positive spatial autocorrelation means that features nearby one another have similar values, while negative autocorrelation signifies nearby features have dissimilar values.

Commonly used methods for testing global clustering are Geary's C (55), Moran's I (56), and the Oden's Ipop (57) that adjusts Moran's I for differences in population. Global clustering can also be assessed using the K-function (Ripley's) when point-level data are available (35,58). GeoDa† and R† packages are publicly available, and several global statistics are available within other proprietary software packages, such as ClusterSeer® † (BioMedware, Ann Arbor, MI), (see CDC Cancer Cluster website for more information).

Local Clustering Methods

Local clustering statistics, such as local indicators of spatial autocorrelation (LISA) (59), identify the locations of clusters or spatial outliers. Some global clustering statistics have local clustering statistic counterparts such as global and local Moran's I statistics and the Besag-Newell R (60). Another statistic, Getis-Ord Gi*, identifies hot and cold spots based on where features with high (hot) or low (cold) values are in close proximity to one another. The Getis-Ord Gi* provides estimates of statistical significance while identifying the locations of hot and cold spots that are not confined to a specific shape.

Local versions of Moran's I and Geary's C are available for free within R† packages such as usdm and spdep (61). Other programs also have Moran's I statistics and the Getis-Ord Gi* statistic, such as ArcGIS™ † tools (Esri, Redlands, CA), ClusterSEER†, and GeoDa† (62).

Spatial scan statistics can be used to scan a study region using a series of moving windows with increasing radii to identify areas where the observed cases included inside the window are greater than expected. This method can be expanded to incorporate time as an added dimension, allowing a scan for spatiotemporal clusters. To properly interpret the results of the spatial scan statistic, it is extremely important to identify the appropriate radius for the spatial scan window to avoid clusters that are too large or too small. Normally, the upper limit of the circle should not include more than 50 percent of the dataset or the study area (35,63).

One of the most popular scan statistics is Kulldorff's scan statistic for spatial, temporal, and space-time analysis (64,65), freely available within SaTScan™ †software (66). The SaTScan™ † software includes analyses for different data types including case counts (65), rates (67), case/control data (65), and even survival data (68). However, cancer clusters can appear in irregular shapes, which prompted Tango and Takahashi (69) to develop a flexible space-time scan statistic, implemented in the FleXScan† software. The flexscan methodology is also available as an R† package, rflexscan. This package implements both Kulldorff's and Tango & Takahashi's scan statistics.

An alternative scan statistic is that proposed by Besag-Newell (70), which is useful for regional data with small population sizes. It is available in the free software ClusterSeer® † This test gives results for both global and local clustering.

Focused Clustering Tests

A growing interest in recent years has been in the detection of clusters around a specific point-source, such as a single identifiable source of air, water, thermal, noise or light pollution (70). These focused tests are usually designed to identify a particular spatial pattern of clustering around the point-source or

specific geographic location. The location of the point-source of interest needs to be identified prior to the assessment, recognizing that different factors (meteorological, topographical and others) can influence the spatial pattern of potential exposures from the point-source (71). The size, shape, and scale of the analysis can also influence the results. For example, below are five focused cluster shapes and corresponding fitted models that can be considered (71):

1. **Distance Decline** (DD), where risk declines symmetrically in all directions with distance from the point-source
2. **Peaked Distance Decline** (PDD), where risk peaks closest to the point-source and then declines with distance
3. **Direction** (D), characterizing increased risk in a specific angle/direction from the point-source
4. **Distance Decline combined with Directional effect** (DDIR)
5. **Peaked Distance Decline combined with Directional effect** (PDDIR)

Widely known focused cluster tests prove to have higher relative power for different models:

1. **Lawson-Waller Score Test** (72,73). Provides robust results across different models. This score test is powerful against small deviations from the null in the direction of a specific alternative.
2. **Bithell's Linear Risk Score** (LRS) Test (74). Distance version used for DD, PDD; direction version used for D, DDIR, and PDDIR.
3. **Cuzick and Edwards' Test** (75). Performs well for large sample sizes (N>500) and also often used for PDD, DDIR, and PDDIR.
4. **Stone's Maximum Likelihood Test** (76). Used for DD.
5. **Tango's Focused Test** (77). Used for DD.
6. **Besag and Newell's Test** (70). Used for PDD.

Regression Analyses

Regression methods provide a complementary analysis and set of tools to cluster detection analysis. Regression analyses are commonly used in public health for two main reasons: (1) to predict an outcome, and (2) to understand the association between at least two variables (78). For example, after identifying spatial clusters of cancer cases, it may be useful to understand the relationship between potential environmental exposures and the cancer of interest while controlling for demographic and behavioral factors associated with increased risk (79,80). Alternatively, it may be of interest to predict the risk of a specific cancer across a wide geographic region if there are data on known environmental exposures (80).

Special considerations must be made when applying regression techniques to spatially structured data. Spatially structured data violate a key assumption of independence among observations due to inherent autocorrelation, where a given value is to some degree predicted by the values of its neighbors (35,81,82). General steps to overcome issues of spatial autocorrelation in regression, drawing primarily from Waller & Gotway (35) and Fotheringam & Rogerson (83), can be found on the CDC Cancer Cluster website.

Comparison of Methods

The choice of a statistical cluster detection method should take into consideration the strengths and weaknesses of the methods. Several criteria can be considered, such as the type of data (e.g., point-level data or areal data), the ease of use and availability of data or software, the transparency of the methods employed in a particular software, statistical power of the method to detect the cluster of interest, and the desired output (76). Multiple comparisons of methods and reviews of techniques have been published over the years (55,76,84–92), and a table with methods and associated applications is available on the CDC Cancer Cluster website.

Summary

Cluster detection and other advanced spatial analysis methods are available via proprietary and free applications. Such analysis often requires specialized knowledge about the data, appropriate use of the methods, and careful interpretation of the results. Specifically, the choice of which models to use depends on the type of data, the underlying assumptions based on the distribution of the data, as well as geospatial considerations such as the size of the study area, spatial scale of the data, aggregation, and masking. For example, results of analysis may greatly differ when implemented at the county or the census tract level, and different models should be implemented if both case and area-level risk factors are evaluated. Specific methods may also require additional considerations such as the type and size of the spatial scan window.

The results of the GIS and spatiotemporal analysis can be used internally for decision making, can inform actions, and can be used to communicate with the public. Therefore, collaborations with GIS professionals and spatial statisticians equipped with specialized skills can help to ensure proper methods are employed and interpretation of results are appropriate. If these experts are not available within the health agency, we highly encourage seeking consultation and technical assistance from CDC/ATSDR's Geospatial Research Analysis and Services Program (GRASP) by emailing the CCG mailbox (CCGuidelines@cdc.gov).

[†] Software noted are examples of package that are available freely or for purchase and do not represent an endorsement of any specific product by the Centers for Disease Control and Prevention or the Agency for Toxic Substances and Disease Registry.

Author Acknowledgements:
Liora Sahar, Marissa Grossman, Brian Lewis
ATSDR; Geospatial Research, Analysis, and Services Program

Appendix C. List of Acronyms

aPOC – Agency Point of Contact
ASTHO – Association of State and Territorial Health Officials
ATSDR – Agency for Toxic Substances and Disease Registry
CAC – Community Advisory Committee
CDC – Centers for Disease Control and Prevention
CSTE – Council of State and Territorial Epidemiologists
cPOC – Community Point of Contact
EHR – Electronic Health Records
GIS – Geographic Information Systems
GRASP – Geospatial Research, Analysis, and Services Program (ATSDR)
NCEH – National Center for Environmental Health (NCEH)
PEHSU – Pediatric Environmental Health Specialty Unit (ATSDR)
SEHD – State Environmental Health Directors
SIR – Standardized Incidence Ratio
SMR – Standardized Mortality Ratio
STLT – State, Tribal, Local, and Territorial

References

1. Abrams B, Anderson H, Blackmore C, Bove FJ, Condon SK, Eheman CR, et al. Investigating suspected cancer clusters and responding to community concerns: Guidelines from CDC and the council of State and Territorial Epidemiologists. MMWR Recomm Reports. 2013;62(1).

2. CDC. Guidelines for Investigating Clusters of Health Events. MMWR. 1990;39(RR-11).

3. Public Law 114–182: Frank R. Lautenberg Chemical Safety for the 21st Century Act. Washington D.C: 114th Congress ; Jun 22, 2016.

4. Federal Register. Updating federal guidelines used by public health agencies to assess and respond to potential cancer clusters in communities [Internet]. 2019. Available from: https://www.federalregister.gov/documents/2019/05/15/2019-09998/updating-federal-guidelines-used-by-public-health-agencies-to-assess-and-respond-to-potential-cancer?form=MY01SV&OCID=MY01SV

5. National Cancer Institute. Defining cancer. [Internet]. 2013. Available from: http://cancer.gov/cancertopics/cancerlibrary/what-is-cancer

6. American Cancer Society. Cancer Figures and Facts [Internet]. 2021. Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf

7. National Institute of Health, National Cancer Institute, Surveillance, Epidemiology and End Results Program. Cancer Stat Facts: Cancer of Any Site [Internet]. 2021 [cited 2022 Apr 7]. Available from: https://seer.cancer.gov/statfacts/html/all.html

8. ATSDR (Agency for Toxic Substances and Disease Registry). Toxicological profiles [Internet]. Available from: https://www.atsdr.cdc.gov/toxprofiledocs/index.html

9. Ma F, Lehnherr M, Fornoff J, Shen T. Childhood cancer incidence in proximity to nuclear power plants in Illinois. Arch Environ Occup Health [Internet]. 2011 Apr 29 [cited 2022 Jan 7];66(2):87–94. Available from: https://pubmed.ncbi.nlm.nih.gov/24484365/

10. Yaffee AQ, Scott B, Kaelin C, Cambron J, Sanderson W, Christian WJ, et al. Collaborative response to arsenic-contaminated soil in an Appalachian Kentucky neighborhood. J Toxicol Environ Health A [Internet]. 2019 [cited 2022 Jan 7];82(12):697–701. Available from: https://pubmed.ncbi.nlm.nih.gov/31307340/

11. Jacqueline F Moreau, Jeanine M Buchanich, Jacob Z Geskin, Oleg E Akilov, Larisa J Geskin. Non-random geographic distribution of patients with cutaneous T-cell lymphoma in the Greater Pittsburgh Area - PubMed. Dermatol Online J [Internet]. 2014 Jul 15 [cited 2022 Jan 7];20(7). Available from: https://pubmed.ncbi.nlm.nih.gov/25046454/

12. Messier KP, Serre ML. Lung and stomach cancer associations with groundwater radon in North Carolina, USA. Int J Epidemiol [Internet]. 2017 [cited 2022 Jan 7];46(2):676–85. Available from: https://pubmed.ncbi.nlm.nih.gov/27639278/

13. Gallagher LG, Webster TF, Aschengrau A, Vieira VM. Using residential history and groundwater

modeling to examine drinking water exposure and breast cancer. Environ Health Perspect [Internet]. 2010 Jun [cited 2022 Jan 7];118(6):749–55. Available from: https://pubmed.ncbi.nlm.nih.gov/20164002/

14. Wheeler DC, Ward MH, Waller LA. Spatial-temporal Analysis of Cancer Risk in Epidemiologic Studies with Residential Histories. Ann Assoc Am Geogr [Internet]. 2012 Sep [cited 2022 Jan 7];102(5):1049–57. Available from: https://pubmed.ncbi.nlm.nih.gov/30956280/

15. Fortunato L, Abellan JJ, Beale L, LeFevre S, Richardson S. Spatio-temporal patterns of bladder cancer incidence in Utah (1973-2004) and their association with the presence of toxic release inventory sites. Int J Health Geogr [Internet]. 2011 Feb 28 [cited 2022 Jan 7];10. Available from: https://pubmed.ncbi.nlm.nih.gov/21356086/

16. Liu-Mares W, MacKinnon JA, Sherman R, Fleming LE, Rocha-Lima C, Hu JJ, et al. Pancreatic cancer clusters and arsenic-contaminated drinking water wells in Florida. BMC Cancer [Internet]. 2013 Mar 12 [cited 2022 Jan 7];13. Available from: https://pubmed.ncbi.nlm.nih.gov/23510413/

17. Parikh PV, Wei Y. PAHs and PM2.5 emissions and female breast cancer incidence in metro Atlanta and rural Georgia. Int J Environ Health Res [Internet]. 2016 Jul 3 [cited 2022 Jan 7];26(4):458–66. Available from: https://pubmed.ncbi.nlm.nih.gov/26983363/

18. Levin RJ, De Simone NF, Slotkin JF, Henson BL. Incidence of thyroid cancer surrounding Three Mile Island nuclear facility: the 30-year follow-up. Laryngoscope [Internet]. 2013 Aug [cited 2022 Jan 7];123(8):2064–71. Available from: https://pubmed.ncbi.nlm.nih.gov/23371046/

19. New York State Cancer Registry. Cancer Data for New York State Counties [Internet]. 2021. Available from: https://www.health.ny.gov/statistics/cancer/registry/vol1.htm

20. Massachusetts Environmental Public Health Tracking. Standardized Incidence Ratios for Cancer [Internet]. 2021. Available from: https://matracking.ehs.state.ma.us/Health-Data/Cancer/sirs.html#MyPopup

21. Centers for Disease Control and Prevention. National Environmental Public Health Tracking Network Data Explorer [Internet]. [cited 2022 Apr 22]. Available from: https://ephtracking.cdc.gov/DataExplorer/

22. Gordis L. Epidemiology. Philadelphia, PA: Elsevier Saunders; 2014.

23. Merrill R. Environmental Epidemiology, Principles and Methods. Sudbury, MA: Jones and Bartlett Publishers, Inc.; 2008.

24. U.S. Environmental Protection Agency. Superfund: National Priorities List (NPL) [Internet]. 2022 [cited 2022 Apr 21]. Available from: https://www.epa.gov/superfund/superfund-national-priorities-list-npl

25. CDC. National Vital Statistics System: Mortality Data [Internet]. 2021. Available from: https://www.cdc.gov/nchs/nvss/deaths.htm

26. American Cancer Society. Cancer Clusters [Internet]. 2021. Available from:

https://www.cancer.org/cancer/cancer-causes/general-info/cancer-clusters.html

27. Thun MJ, Sinks T. Understanding cancer clusters. CA Cancer J Clin. 2004 Sep 1;54(5):273–80.

28. Lanphear B, Buncher C. Latent period for malignant mesothelioma of occupational origin - PubMed. J Occup Med [Internet]. 1992 [cited 2022 Jan 7];34(7):718–21. Available from: https://pubmed.ncbi.nlm.nih.gov/1494965/

29. 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 [Internet]. 2009 Nov [cited 2022 Jan 7];19(11):834–40. Available from: https://pubmed.ncbi.nlm.nih.gov/19364662/

30. CDC. World Trade Center Health Program [Internet]. Available from: https://www.cdc.gov/wtc/pdfs/policies/WTCHP-Minimum-Cancer-Latency-PP-01062015-508.pdf

31. Dicker R, Coronado F, Koo D, Parrish R. Principles of Epidemiology in Public Health Practice: An Introduction to Applied Epidemiology and Biostatics. CDC Office of Workforce and Career Development; 2006.

32. Frost G. The latency period of mesothelioma among a cohort of British asbestos workers (1978–2005). Br J Cancer. 2013 Oct 29;109(7):1965–73.

33. Costas K, Knorr RS, Condon SK. A case-control study of childhood leukemia in Woburn, Massachusetts: The relationship between leukemia incidence and exposure to public drinking water. Sci Total Environ. 2002;300(1–3):23–35.

34. Massachusetts Department of Public Health. The Wilmington Childhood Cancer Study: An Epidemiologic Investigation of Childhood Cancer from 1990-2000. 2021.

35. Waller LA, Gotway CA. Applied spatial statistics for public health data. New York : John Wiley and Sons; 2004.

36. National Cancer Institute. Cancer Incidence Statistics [Internet]. 2021 [cited 2022 Jan 7]. Available from: https://surveillance.cancer.gov/statistics/types/incidence.html

37. Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology. 2nd ed. New York, NY: Oxford University Press; 1996.

38. Sahai H, Khurshid A. Statistics in epidemiology: methods, techniques, and applications. Boca Raton: CRC; 1996.

39. Selvin S. Statistical analysis of epidemiologic data. New York, NY: Oxford University Press; 1996.

40. Breslow NE, Day NE. Statistical methods in cancer research. Volume I - The analysis of case-control studies. IARC Sci Publ. 1980;(32):5–338.

41. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.

42. United Kingdom and Ireland Association of Cancer Registries. Standard Operating Procedure: Investigating and Analysing Small-Area Cancer Clusters [Internet]. 2015. Available from: http://www.ukiacr.org/sites/ukiacr/files/file-

43. Pagano M, Gauvreau K. Principles of Biostatistics. 2nd ed. Pacific Grove, CA: Duxbury Thomson Learning; 2000.

44. Rothman KJ. No Adjustments Are Needed for Multiple Comparisons. Vol. 1. 1990.

45. CDC. Geography and Locational Referencing Subgroup of the Standards and Network Development Workgroup of the National Environmental Public Health Tracking Program. Environmental Public Health Tracking Version 1.0 (A resource for EPHT managers and a tool for their technical staff). 2005.

46. Goldberg DW, Swift JN, Wilson JP. Geocoding Best Practices: Reference Data, Input Data and Feature Matching.

47. Sahar L, Foster SL, Sherman RL, Henry KA, Goldberg DW, Stinchcomb DG, et al. GIScience and cancer: State of the art and trends for cancer surveillance and epidemiology. Vol. 125, Cancer. John Wiley and Sons Inc.; 2019. p. 2544–60.

48. Gehlke CE, Biehl K. Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material. J Am Stat Assoc. 1934 Mar;29(185A):169–70.

49. Openshaw S, Taylor PJ. A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. Stat Appl Spat Sci. 1979;127–44.

50. Cheng T, Adepeju M. Modifiable Temporal Unit Problem (MTUP) and Its Effect on Space-Time Cluster Detection. PLoS One [Internet]. 2014;9(6):1–10. Available from: www.plosone.org

51. Wong DW. Modifiable Areal Unit Problem. Int Encycl Hum Geogr. 2009;

52. Su MD, Lin M, Wen T. Spatial Mapping and Environmental Risk Identification. 2011.

53. Yoo E-H. GIS Methods and Techniques. Compr Geogr Inf Syst. 2018;

54. Huang L, Pickle LW, Das B. Evaluating spatial methods for investigating global clustering and cluster detection of cancer cases. Stat Med. 2008 Nov 1;27(25):5111–42.

55. Geary RC. The contiguity ratio and statistical mapping. Inc Stat [Internet]. 1954;5(3):115–46. Available from: https://about.jstor.org/terms

56. Moran PAP. Notes on continuous stochastic phenomena. Biometrika [Internet]. 1950;37(1/2):17–23. Available from: https://www.jstor.org/stable/2332142

57. Oden N. Adjusting Moran's I for population density. Stat Med. 1995;14(1):17–26.

58. Ripley BD. Spatial Statistics. New York, NY: John Wiley and Sons; 1981.

59. Anselin L. Local indicators of spatial association—LISA. Geogr Anal. 1995;27(2):93–115.

60. Costa MA, Assunção RM. A fair comparison between the spatial scan and the Besag–Newell Disease clustering tests. Environ Ecol Stat. 2005;12(3):301–19.

61. Naimi B, Hamm NAS, Groen TA, Skidmore AK, Toxopeus AG. Where is positional uncertainty a problem for species distribution modelling? Ecography (Cop). 2014;37(2):191–203.

62. Bivand RS, Wong DWS. Comparing implementations of global and local indicators of spatial association. Test [Internet]. 2018;27(3):716–48. Available from: https://doi.org/10.1007/s11749-018-0599-x

63. Kulldorff M, Nagarwalla N. Spatial Disease Clusters: Detection and Inference. Vol. 14, STATISTICS IN MEDICINE. 1995.

64. Kulldorff M. A spatial scan statistic. Commun Stat methods. 1997;26(6):1481–96.

65. Kulldorff M, Heffernan R, Jacobs J, Martins A, Mostashari F. A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. PLoS Med. 2005;2:e59.

66. Kulldorff M. Information management services, Inc. SaTScanTM v9. 2009;4.

67. Huang L, Tiwari RC, Zou Z, Kulldorff M, Feuer EJ. Weighted Normal Spatial Scan Statistic for Heterogeneous Population Data. J Am Stat Assoc [Internet]. 2009;104(487):886–98. Available from: https://doi.org/10.1198/jasa.2009.ap07613

68. Huang L, Kulldorff M, Gregorio D. A spatial scan statistic for survival data. Biometrics. 2007;63(1):109–18.

69. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. Int J Health Geogr. 2005;4(1):1–15.

70. Besag J, Newell J. The detection of clusters in rare diseases. J R Stat Soc Ser A (Statistics Soc. 1991;154(1):143–55.

71. Puett RC, Lawson AB, Clark AB, Aldrich TE, Porter DE, Feigley CE, et al. Scale and shape issues in focused cluster power for count data. Int J Health Geogr. 2005;4(1):1–16.

72. Lawson AB. Statistical methods in spatial epidemiology. Wiley; 2006.

73. Waller LA, Turnbull BW, Clark LC, Nasca P. Chronic disease surveillance and testing of clustering of disease and exposure: Application to leukemia incidence and TCE-contaminated dumpsites in upstate New York. Environmetrics. 1992;3(3):281–300.

74. Bithell JF. The choice of test for detecting raised disease risk near a point source. Stat Med. 1995;14(21-22):2309–22.

75. Cuzick J, Edwards R. Methods for investigating localized clustering of disease. Clustering methods based on k nearest neighbour distributions. IARC Sci Publ. 1996;(135):53–67.

76. Stone RA. Investigations of excess environmental risks around putative sources: statistical

problems and a proposed test. Stat Med. 1988;7(6):649–60.

77. Tango T. A class of tests for detecting 'general'and 'focused'clustering of rare diseases. Stat Med. 1995;14(21-22):2323–34.

78. Kleinbaum DG, Kupper LL, Nizam A, Rosenberg ES. Applied regression analysis and other multivariable methods. 5th ed. Boston, MA, USA: Cengage Learning; 2013. 1074 p.

79. Cardoso D, Painho M, Roquette R. A geographically weighted regression approach to investigate air pollution effect on lung cancer: A case study in Portugal. Geospat Heal. 2019/05/18. 2019;14(1).

80. Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. Env Heal Perspect [Internet]. 2004/06/17. 2004;112(9):998–1006. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15198920

81. Fotheringham AS. "The Problem of Spatial Autocorrelation" and Local Spatial Statistics. Geogr Anal [Internet]. 2009;41(4):398–403. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4632.2009.00767.x

82. Legendre P. Spatial Autocorrelation: Trouble or New Paradigm? Ecology [Internet]. 1993;74(6):1659–73. Available from: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1939924

83. Fotheringham AS, Rogerson PA. The SAGE handbook of spatial analysis [Internet]. Thousand Oaks, CA, USA: Sage; 2009. p. 528. Available from: https://methods.sagepub.com/Book/the-sage-handbook-of-spatial-analysis

84. Goujon S, Kyrimi E, Faure L, Guissou S, Hémon D, Lacour B, et al. Spatial and temporal variations of childhood cancers: Literature review and contribution of the French national registry. Cancer Med. 2018;7(10):5299–314.

85. Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. Stat Med. 2006;25(22):3929–43.

86. Lin H, Ning B, Li J, Ho SC, Huss A, Vermeulen R, et al. Lung cancer mortality among women in Xuan Wei, China: a comparison of spatial clustering detection methods. Asia Pacific J Public Heal. 2015;27(2):NP392–401.

87. Hanson CE, Wieczorek WF. Alcohol mortality: a comparison of spatial clustering methods. Soc Sci Med. 2002;55(5):791–802.

88. Kim J, Lee M, Jung I. A comparison of spatial pattern detection methods for major cancer mortality in Korea. Asia Pacific J Public Heal. 2016;28(6):539–53.

89. Chen J, Roth RE, Naito AT, Lengerich EJ, MacEachren AM. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of US cervical cancer mortality. Int J Health Geogr. 2008;7(1):1–18.

90. Kulldorff M, Song C, Gregorio D, Samociuk H, DeChello L. Cancer map patterns: are they random or not? Am J Prev Med. 2006;30(2):S37–49.

91. Jackson MC, Huang L, Luo J, Hachey M, Feuer E. Comparison of tests for spatial heterogeneity on data with global clustering patterns and outliers. Int J Health Geogr. 2009;8(1):1–14.

92. Tango T. Spatial scan statistics can be dangerous. Stat Methods Med Res. 2021;30(1):75–86.