Guidelines for Investigating Clusters of Chronic Diseases

Arkansas Department of Health

Appathurai Balamurugan, MD, DrPH, DipABLM, FAAFP, Deputy Chief Medical Officer, State Chronic Disease Director
Lori Simmons, MS, Epidemiology Branch Chief
Abby Holt, MPH, Cancer Registry Director
Robert Delongchamp, PhD MPH, Consultant and Professor of Epidemiology, UAMS College of Public Health

Table of Contents

Overview

The purpose of these guidelines is to ensure a standardized coordinated response from Arkansas Department of Health (ADH) colleagues receiving calls from the public, health professionals, or others about potential clusters of chronic diseases (cancer, birth defects or any unrecognized noninfectious syndromes or illnesses). The guidelines are also intended to be a reminder to ADH personnel about the importance of communicating and coordinating with appropriate entities, both within the central office and in the region where the potential cluster occurs.

This protocol only pertains to clusters of chronic diseases such as cancers, birth defects, or an unrecognized noninfectious syndrome or illness. It does not apply in emergency situations such as infectious disease outbreaks, bioterrorism events, or other hazardous events.

This document is based on the information from the Morbidity and Mortality Weekly Report (MMWR) report, "Investigating Suspected Cancer Clusters and Responding to Community Concerns. Guidelines from CDC and the Council of State and Territorial Epidemiologists", published on September 27, 2013, Vol. 62, No. 8: Pages 1-24.

This document has been reviewed and approved by ADH Science Advisory Committee on January 24, 2014.

Chronic Disease Cluster Investigation Team (CD-CIT)

The Arkansas Department of Health has formed a Chronic Disease Cluster Investigation Team (CD-CIT) to assure a coordinated response for any suspected chronic disease clusters or concerns. The members include:

1. Deputy State Medical Director, State Chronic Disease Director
2. Consultant Statistician/Epidemiologist
3. Epidemiology Branch Chief
4. Cancer Registry Director

Additionally, the following members of ADH will be involved in forming an expert advisory panel to resolve any issues as needed. The members of Expert Advisory Panel include:

1. Medical Director, Chronic Disease Prevention and Control Branch
2. State Epidemiologist
3. Director, Center for Health Advancement
4. Director, Center for Public Health Practice
5. Deputy Director for Public Health Programs
6. Deputy State Health Officer/Chief Science Officer
7. Director, Arkansas Department of Health

Outline for Evaluating Suspected Clusters

Step 1:

i. Suspect cluster is reported to ADH colleague.

ii. ADH colleague routes the report to CD-CIT member.

iii. CD-CIT member evaluates legitimacy of concern and, if needed, collects additional information to initiate an investigation.

iv. CD-CIT member decides whether to resolve the investigation and communicate the results to the caller, or if not resolved, to move to Step 2.

Step 2:

i. CD-CIT meets to begin to assess whether the cluster represents an excess of cases above that expected.

ii. CD-CIT defines the study population, conduct statistical analysis to calculate the standardized incidence ratio with 95% confidence intervals.

iii. CD-CIT reviews the literature on the risk factors and assesses possible associations between the risk factors and suspected excess.

iv. CD-CIT decides whether to resolve the investigation and communicate the results to the caller, or if not resolved, to move to Step 3.

Step 3:

i. Expert advisory panel informed and participates in hypotheses generation and evaluation and potential study design.

ii. Expert advisory panel ascertains the plausibility of an association, and feasibility to conduct definitive study.

iii. Expert advisory panel decides whether to resolve the investigation and communicate the results to the caller, or if not resolved, to move to Step 4.

Step 4:

i. CD-CIT and Expert advisory panel engages relevant stakeholders, external partners and community members to conduct an epidemiologic study.

ii. Share the results of the investigation with relevant stakeholders, partners and community members. Recommend interventions to address the issue as appropriate.

4-Step Process of Evaluating Suspected Clusters

The following steps are used to respond to reports of suspected excesses of chronic diseases, including procedures, guidance on, and considerations for closing the inquiry or proceeding to next steps. We have used cancer as an example to illustrate proceeding through the four-step process.

Step 1. Initial Contact and Response

1. CD-CIT member receiving call or report of suspected cluster will document the following information from the inquirer/reporter to determine whether the concern warrants further investigation.

    a) All identifying/contact information about the inquirer/reporter (hereafter referred to as the individual): name, address, telephone number, length of residence at current location, and organization affiliation, if any. Comply with requests for anonymity but explain that the inability to follow-up with the caller might hinder further investigation.
    b) Initial data on the potential cluster from the individual: types of cancer and number of cases of each type, age of the people with cancer, geographic area of concern, period over which the cancers were diagnosed, and how the individual learned about the suspected cluster. The individual might not know the true primary cancer diagnoses and will most likely not be aware of all cases of cancer in this area or during the period of concern.
    c) Information from the individual about any specific environmental hazards or concerns, other risk factors (e.g., tobacco use, diet, infections, and family history of disease) and other concerns in the affected area (e.g., likely period of environmental contaminant exposures). If the individual is reporting an event that is not a suspected cancer cluster, but rather one involving a known or possible environmental contamination, the individual should be referred to the environmental epidemiology section.

2. Initial Investigation
    a) CD-CIT member will review state and county-level incidence and mortality cancer statistics from the Cancer Registry in the appropriate context. Available from (http://www.cancer-rates.info/ar/) and (https://wonder.cdc.gov/).
    b) CD-CIT member will become familiar with the geographic profile within the area of concern (e.g., demographics, industrial and residential development) in order to

understand the health and environmental concerns of the community. Use resources noted in Appendix A.

c) CD-CIT member will consult with the CD-CIT to make an initial judgment about the advisability of the ADH pursuing an inquiry into the suspected cancer cluster.

3. Decisions

a) Decision to Close the Investigation at Step 1 –

i. Based on the review of cancer registry, risk, and geographic profile, the suspected cluster of cancer cases might be unrelated because the cancers are not likely to share a common, environmental etiology.

ii. CD-CIT member will notify the individual and explain the following: the nature of cancer (e.g., increase risk with age, other behavioral factors), its frequency and occurrence, how different types of cancers are related to different causes, that rates of disease do increase and decrease in a population over time (random fluctuations), and explain rates in the appropriate context, and provide easily accessible resources such as the Centers for Disease Control and Prevention (CDC)'s cancer website (http://www.cdc.gov/cancer), etc.

iii. If the individual is satisfied with the decision not to move forward, the inquiry will be closed. If the individual is not satisfied with the decision, then the CD-CIT member should provide a written explanation and include resources related to the decision.

b) Decision to Continue to Step 2 –

i. The data gathered might suggest the need for further investigation, as determined by the CD-CIT.

ii. CD-CIT member will notify the individual and provide initial investigative outcomes and outline how the ADH will follow-up. The CD-CIT member should ask if there are others in the community (e.g., other residents with this cancer type) who would like to have a report on the results of the next step.

iii. CD-CIT member will provide easily accessible resources such as the CDC's cancer website, (http://www.cdc.gov/cancer).

iv. Regardless of the decision, all documentation should be included in a permanent log.

Step 2. Assessment

The primary purpose of step two is to determine whether the suspected excess of cancer is a statistically significant excess. Because of the variety of issues involved in this phase, the entire CD-CIT may need to get involved.

1. CD-CIT will define the study population, including:
   a) Demographic characteristics: age, sex, race/ethnicity, or residential location. The study population could be all-inclusive.
   b) Geographic area: select study area (usually county-level) and time period. Privacy issues should be considered when collection, analyzing, and presenting analysis on cases at the neighborhood level. The Cancer Registry does not release information on < 6 cases at the state or county-level. Statistical analysis at the neighborhood level with sparsely populated areas might not be possible because of small, limited numbers. Limited numbers can lead to a lack of statistical power and therefore to an instability of rates.
   c) Case definition: information on the cancer type based on the ICD-O codes (e.g. primary site, histology, and behavior).

2. CD-CIT will consult with Arkansas Central Cancer Registry (ACCR) staff to confirm diagnoses of patients within the area of concern. This information is used to validate diagnoses among cases and may provide additional tumor characteristics (e.g., histology, behavior, grade). Personal health information about cases under investigation cannot be shared due to privacy concerns.

   Since 1996, through funding from the National Program of Cancer Registry (NPCR) at the CDC, the ACCR has been collecting population-based cancer incidence data among all residents in Arkansas. The ACCR collects high quality and complete data, and has been consistently certified as a gold-standard registry designated by the North American Association of Central Cancer Registries and as a Registry of Distinction by NPCR. A flow chart of the data processes are available in Appendix D.

   a) Verify cancer diagnosis.
      i. Verification is a multi-step process. The Cancer Registry collects pathology reports in real-time from hospitals and private or commercial laboratories. This information, along with collected medical records from hospitals, oncology clinics, radiation treatment facilities, specialty clinics, and hospice and nursing facilities are used for verification purposes.

Note: verification can only occur if the patient was a resident of Arkansas at the time of diagnosis, during the period between 1996 to present.

ii. Obtain histologic reevaluation, if needed.

iii. If warranted by the CD-CIT, a data linkage between the cases of concern and the cancer registry or other vital records databases can be performed. Data items needed to match records include name, date of birth, SSN, sex, and race/ethnicity.

iv. See Appendix A for limitations about using and interpreting cancer registry data.

3. CD-CIT will calculate a standardized incidence ratio (SIR) with 95% confidence intervals and other descriptive statistics (e.g., incidence/mortality trends) as appropriate.

   The SIR calculation provides an estimate of the likelihood that an excess of cases exists in the population of concern (study population) compared to the general or reference population (e.g., population of Arkansas residents). See Appendix B (page 18-21) for more information on statistical approaches.

Standard Incidence Ratio (SIR)

$$\text{SIR} = \frac{\text{observed cases}}{\text{expected cases}}$$

Standard Error for SIR

$$\text{SE}_{\text{SIR}} = \frac{\sqrt{\text{observed}}}{\text{expected}}$$

95% Confidence Intervals for SIR

$$\text{CI} = \frac{\left[\sqrt{\text{observed}} \pm (1.96 \times 0.5)\right]^2}{\text{expected}}$$

   An SIR > 1.0 indicates the observed number of cases is greater than the number that would be expected for the population. The 95% CI is an indication of the statistical precision and significance of the SIR value. If the 95% CI includes 1.0, the SIR is not statistically significant.

4. CD-CIT will review the literature on the risk factors for the types of cancers in question and on the possible associations between the types of cancer and known or suspected environmental exposures using PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) or other sources.

5. The CD-CIT may consider examining trends of a cancer type that is unrelated to the cancer and/or exposure of concern. If this cancer appears elevated or depressed in a similar time frame, other factors should be considered. These factors include the possibility that the community has an unusually high proportion of persons with high-risk health behaviors (e.g., smoking).

6. Decisions

    a) Decision to Close the Investigation at Step 2 –
        i. An SIR of limited magnitude that is not statistically significant, coupled with the lack of known association with an environmental contaminant and no trend of increasing incidence over time, justifies closing the inquiry at Step 2. For example, an SIR <2.0 with CIs surrounding or overlapping 1.0 and/or small number of cases (e.g., <10), between the type of cancer and the suspected environmental contaminant, might justify a decision to close the inquiry.
        ii. CD-CIT will communicate with the individual and share the SIR results and describe the process, results, and implications as indicated in Appendix C.

    b) Decision to Continue to Step 3 –
        i. An SIR >4 with CIs that do not overlap 1.0, and ≥10 cases that might be etiologically linked, should encourage advancing to Step 3.
        ii. Provide a written report to the inquirer, as well as to any partners contacted. This report should include a description of the results of the preliminary analyses and circumstances, carefully acknowledging what is known and unknown at this point. The report should outline the Department's next steps.
        iii. SIR calculations not specifically identified above will be handled on a case by case basis.

Step 3. Determining Feasibility of Conducting an Epidemiologic Study

All activities in this step should be carried out in collaboration with community, environmental, and other partners.

1. CD-CIT will engage the Expert Advisory Panel to identify hypotheses and potential study design.
    a) Efforts will focus on known causes of cancer in question.

i. CD-CIT will use past agency reports and logs to determine whether the same type of cancer has led to other inquiries and investigation.
ii. CD-CIT will conduct another literature search.

2. CD-CIT and Expert Advisory Panel will identify study parameters and proceed with investigation as follows:
    a) Confirm the diagnosis.
    b) Identify a comparison group that depending on the study design does not have the cancer of concern (i.e. control group in a case-control study) or does not have the exposure of concern (i.e. unexposed group in a cohort study).
    c) Explore feasibility of obtaining data on individuals in the comparison group, and explore the willingness of persons to participate in interviews or studies for gathering data on health, possible exposures, the amount of time the affected persons lived in the area, and occupation.
3. CD-CIT and Expert Advisory Panel will ascertain the plausibility that the cases and contaminants could potentially be associated, including:
    a) Verifying whether the environmental contaminants of concern are known carcinogens.
    b) Considering possible and plausible routes of exposure to affected persons.
    c) Asking whether persons with cancer actually were exposed to an environmental contaminant in sufficient doses for a sufficient time to make the association biologically plausible.
    d) Considering the possibility of historical records of chemical use or contamination at the particular location.
    e) Determining whether residential and occupational histories for affected persons are obtainable.
4. CD-CIT will assemble available information from standard sources on the environmental contaminant of concern.
    a) It is not recommended to engage in general, open-ended inquiry to identify potential contaminants in a community, in the absence of a suspected etiologic agent.
    b) Additional environmental testing should be carried out only when there is a clear scientific rationale, and;
        i. Because of the long latency of cancer, an historical exposure assessment might be more important than consideration of current exposures.
        ii. Investigators should determine whether they can characterize exposure to suspected environmental hazards accurately, at the individual level in a way that reflects the period of concern.
5. CD-CIT and Expert advisory panel will determine study design, including:

a) Geographic scope, study timeframe (allows for sufficient latency in cancers of concern), and demographics.
b) Study design, sample size, and statistical tests necessary to study the association as well as the effect of a small sample size on statistical power.
c) Resource implications and resource/funding requirements of the study.

6. Decisions
    a) Decision to Close the Investigation at Step 3 –
        i. If the feasibility assessment suggests that little will be gained from proceeding further, the investigation should be closed and summarized in a report to the initial caller and other concerned parties.
        ii. The public or media might continue to demand further investigation, regardless of cost or biological plausibility. Work with established community relationships, media contacts, and the advisory panel to manage the response.
    b) Decision to Continue to Step 4 –
        i. If the activities in Step 3 warrant further epidemiological study, proceed to Step 4.

Step 4. Conducting an Epidemiologic Investigation

Conducting epidemiologic investigations can take several years; the ADH will consider what should be done in the interim to help protect community health and keep members informed.

This step involves a detailed epidemiologic study that tests a hypothesis of the association between the suspected exposure and specific cancer types, for which all the preceding effort has been preparatory. See Appendix B for a guide to statistical and epidemiologic approaches for conducting investigations.

APPENDICES

APPENDIX A: Data and Other Resources

Cancer Registries

The Arkansas Central Cancer Registry (ACCR) receives reports of all new cancer cases from clinical facilities in the state, will have numerator data (i.e., the number of new cancer cases) for calculating the SIR as well as data for the appropriate comparison measures for reference populations.

- ACCR: http://www.cancer-rates.info/ar/index.php
  - Phone: 501-661-2463
- State Cancer Profiles: http://statecancerprofiles.cancer.gov/
- Surveillance, Epidemiology, and End Results (SEER): http://seer.cancer.gov/faststats/index.php
- United States Cancer Statistics: https://wonder.cdc.gov/cancer.html

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

- Registry information generally contains patient address at the date of diagnosis only.
- The ACCR collects information on possible risk factors (e.g., smoking history) and usual occupation, but the data are often incomplete.
- The types of cancer that are most likely to be underreported occur in persons with late-stage cancers that are treated with palliative care (e.g., persons who might not be hospitalized for surgery or treatment). Other likely underreported types include those who have been diagnosed in a physician's office without hospitalization (e.g., early stage melanoma). Many hospitals routinely collect cancer data for their own purposes and for most hospitals reporting to central registries is routine. However, reporting from nonhospital facilities is less reliable. Consequently, data for cancer patients who are never hospitalized for diagnosis and treatment tend to be less complete and might be reported later than other cases.
- Codes and rules for counting cancer cases change over time. Occasionally, changes in diagnostic criteria might change how a cancer is diagnosed, possibly creating changes in the frequency in which the cancer is detected and reported. These types of changes are adopted at different rates by physicians and hence in reports to the registries.
- Data on race and ethnicity are captured in registry data; however, this data is collected inconsistently with some providers relying on a patient's self-report and others assessing race based on observation.

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

Data on Deaths

Data on deaths compiled by the Health Statistics Branch might be a useful supplement in identifying data on cancer cases. Death records are most useful for cancer with high mortality and a short survival period such as pancreatic, liver, lung, and some types of brain cancer. However, death records are not very useful for cancers with lower mortality, such as breast, thyroid, prostate, or colon cancers, from which patients are likely to survive.

> CDC Wide-ranging Online Data for Epidemiologic Research (Wonder): http://wonder.cdc.gov/mortSQL.html

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

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

U.S. Census Bureau

The U.S. Census Bureau can provide valuable data for use in determining the denominator for incidence calculation. State, county, census tract, and census block level data are available.

> U.S. Census Bureau: https://data.census.gov/cedsci/

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

- Census numbers might be inaccurate for intercensal years when substantial population changes (rapid growth, shrinkage, or aging changes) occur.
- Census boundaries occasionally change, most often in rapidly growing areas that are often subdivided, making comparison between years or combining data from different years difficult.
- The census tract is defined by the U.S. Census Bureau, and it is a relatively homogeneous unit with respect to population characteristics. A census tract generally contains between 1,000 and 8,000 persons, with an optimum size of 4,000 persons. Cancer clusters of concern frequently are confined to areas smaller than a census tract. Because census tracts are subdivided into

census blocks and block groups, blocks and block groups might be combined if a census tract does not give the needed geographic boundaries. The number of cases occurring within a block or a block group might be far too small to allow reporting of cancer cases without privacy concerns or creating statistically unstable rates. Registries often will not release data at the block group level or even the census tract level because of privacy concerns.

- Census units might not be similar to contamination boundaries.

Zip codes can be and often are used as geographic areas for cluster investigations, especially if they are a better fit for communities at issue. There are two major limitations to using zip codes for cancer cluster investigations:

- Zip code boundaries might change more often than census boundaries, and
- Zip codes cross county and census boundaries. Moreover, a person might have a post office box or a rural route address that is in a different zip code than the actual residence.

National Environmental Public Health Tracking Network

CDC's National Environmental Public Health Tracking Network (Tracking Network) is a nationwide surveillance network that provides health, environmental hazard, and exposure data.

- Tracking Network: http://www.cdc.gov/nceh/tracking/

Data from State and Territorial Environmental Agencies

State and local environmental protection agencies routinely collect environmental data. Because these data are collected in places and at times according to regulatory purposes, they might be useful in identifying environmental hazards in cancer cluster investigations, or they might only approximate the environmental conditions at the site of the potential cancer cluster. Environmental agencies regularly collect data on water quality and air quality for compliance with air and water quality standards. These agencies also often permit and regulate industrial or other facilities that generate, transport, or store hazardous waste or other chemicals. The agencies will therefore have records of compliance and noncompliance that might indicate emissions into the environment. The state agencies are also involved, along with the Environmental Protection Agency (EPA), in monitoring pollution and in the oversight of the cleanup of contaminated sites. EPA collects environmental data for regulatory purposes, and the agency publishes the data on its website.

- EPA's list of State and Territorial Environmental Agencies: https://www.epa.gov/home/health-and-environmental-agencies-us-states-and-territories
- Arkansas Department of Environmental Quality (ADEQ) searchable databases: http://www.adeq.state.ar.us/compsvs/webmaster/databases.htm
- ATSDR: http://www.atsdr.cdc.gov/substances/index.asp

- ATSDR series of Toxicological Profiles: http://www.atsdr.cdc.gov/toxprofiles/index.asp

The staff located within state or local environmental protection departments can be a helpful resource for providing information about local environmental conditions that might lead to exposure to contamination. The staff's assistance should be engaged in evaluating available environmental data for relevance to a cancer cluster inquiry or investigation because the data collection areas are determined by regulatory requirements and might not provide information specific to a particular site of public health interest.

Sources of information on the association between specific environmental contaminants and cancer are available. Weight-of evidence-evaluations of carcinogens are published by the International Agency for Research on Cancer (IARC) (IARC cancer classifications are available at http://www.iarc.fr) and the National Toxicology Program (NTP's Report on Carcinogens is available at http://ntp.niehs.nih.gov/go/roc). These evaluations tend to focus on exposures that have been of concern for some time and therefore on which there are substantial data. Not all potential carcinogens have been evaluated by these organizations.

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

Biomonitoring

Biomonitoring is the measurement, usually in blood or urine, of chemical compounds, elements, or their metabolites in the body. Although biomonitoring indicates exposure to a substance at some level, it might not indicate when the exposure occurred or what effects the exposure might have on health in the future. Because of the long latency period associated with the development

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

- NHANES Report on Human Exposure to Environmental Chemicals: http://www.cdc.gov/exposurerePort/pdf/FourthReport.pdf

APPENDIX B: Statistical and Epidemiologic Approaches

Standardized Incidence Ratio and Confidence Interval

The measure typically used to assess whether there is an excess number of cancer cases is the SIR. The SIR is a ratio of the number of observed cancer cases in the study population to the number that would be observed (often called "expected") if the study population experienced the same cancer rates as an underlying population (often called the "reference" population). The reference population could be the surrounding census tracts, other counties in the state, or the state as a whole (not including the community under study).

Confidence Interval

A confidence interval is calculated to determine the precision of the SIR estimate and the statistical significance. If the confidence interval includes 1.0, the SIR is not statistically significant. The narrower the confidence interval, the more confidence one has in the precision of the SIR estimate. One difficulty in cancer cluster investigations is that the population under study is generally a community or part of a community, typically resulting in a small denominator, and such small denominators frequently yield wide confidence intervals, meaning that the SIR is therefore not as precise as desired.

Considering Alpha and Beta Level Values

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

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

Power Analysis

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

Mapping the Cancer Cluster

When considering the geographic distribution of cases, responders have various methods they can use. For example, they might develop a visual representation showing the location of each case superimposed on the underlying population density to get an approximation of the distribution of the relative rates of cancer.

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

Cancer registries and state health agencies typically have criteria related to release of data for small geographic areas. The ACCR does not release information on < 5 cases at the state or county-level. Limited numbers can lead to a lack of statistical power and therefore to an instability of rates. For example, a pin-point map of a small geographic area that identifies the residence of a cancer patient should not be made public. Similarly, many health agencies are prohibited from publicly releasing a table for a small geographic area with a small population, for each table cell might have only a few cases.

Descriptive and Spatial Statistical and Epidemiologic Methods

Frequencies, rates, and descriptive statistics are useful first steps in evaluating the suspected cancer cluster. Confidence intervals can also be calculated for rates. Other statistical approaches include Poisson regression. Often, the number of cases is limited, therefore limiting the type of analysis. If an investigation progresses to a case-control study, the odds ratio can be calculated.

As with any other epidemiologic analysis, there might be methodological issues with the use of clustering tools. Many of these concerns (e.g., limitations associated with small populations, environmental data quality, disease latency periods, and population migration) have been described in this report. In addition, when exposure or outcome analysis uses aggregate data and not data collected on an individual level, responders must use caution when interpreting this type of analysis, because the association with a particular environmental contaminant might not be

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

One of the most popular techniques for detecting clusters is called the spatial scan statistic. Its most commonly used implementation is the SaTScan software (available at https://www.satscan.org/). The underlying concept for this approach is the scan statistic, which considers both spatial areas and time intervals. Other implementations include the nearest neighbor test and the Small Area Health Statistical Unit (SAHSU)'s "Rapid Inquiry Facility" (RIF) in a choice of a statistical cluster method, it might be useful to consider several criteria, such as ease of use and availability, the clarity and transparency of the method, its statistical power to detect the cluster of interest, and the method's ability to produce the desired output.

APPENDIX C: Communication

Developing Communication Plans

Before responding to any inquiries concerning a possible cancer cluster, the health agency should develop a one-on-one communication strategy. Key points in such a strategy should include:

- The importance of listening and how to ask questions that will help determine the nature of the caller's concerns, and
- Trying to ascertain in the first call, the level of concern across the larger community.

Resource for State and Local Health Agencies

CDC and the National Public Health Information Coalition (NPHIC) have published a useful resource which is currently available to state and local health agencies, providing detailed guidelines on communicating in cancer cluster investigations.

Cancer Clusters: A Toolkit for Communicators: https://www.nphic.org/toolkits/cancer-cluster

It provides suggested outreach techniques for various audiences and offers answers to commonly asked questions about suspected cancer clusters. It also provides literature resources, a glossary of cancer cluster terms, a guide to education by use of social media, and case studies.

A basic communication plan should be created for answering initial inquiries about possible excess cancer cases. The plan also should define commonly used terms (e.g., cluster) in a clear and accessible way and emphasize that when speaking to a caller, a responder should use such terms in a consistent manner. Statistical concepts such as small samples size, random fluctuation, and statistical significance are difficult concepts for the general public audience to understand, and having consistent, clear, talking points that address these concepts is helpful.

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

Communication Audience

The communication audience throughout the process of inquiry or investigation will include the initial caller, other concerned community members, community leaders, public health partners, government officials, media, physicians, real estate agents, and other groups, depending on how

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

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

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

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

considered. In communities where trust in government has eroded, it is particularly important to engage the community in the selection of participants of a community panel.

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

Communicating in Uncertain and Stressful Situations

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

- Be a credible and consistent source,
- Create realistic expectations,
- Raise awareness of other credible sources,
- Be empathetic and have patience,
- Be supportive and receptive to the information reported, and
- Listen clearly and consistently.

Proactive Community Involvement

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

One way to involve the community broadly is to establish advisory groups, such as a community panel (See Step 3, Procedures.). Another way is to hold public meetings. If, during the process of investigation, a need is identified to have public meetings, a clear agenda and goal should be set for each meeting, including discussions of major milestones (e.g., completion of the feasibility assessment). The format and atmosphere of a public meeting can have great influence on its outcome. For example, town hall–type public meetings can allow community members to express frustrations and feelings to officials. Health agency personnel who listen well can establish credibility with the community in such meetings. However, some agencies might have difficulty in communicating well in this format. In these cases, an agency should use trained facilitators who

understand the local culture. In such meetings, the health agencies should keep presentations short and use plain language. An alternative is to conduct public meetings with a series of "stations," at which data (e.g., maps) can be presented and discussed in one-to-one or small-group communication. This is one way to involve partners such as environmental agencies and community groups in this type of meeting.

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

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

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

Office of Communications and Health Marketing

Arkansas Department of Health

The Arkansas Department of Health Office of Communications and Health Marketing staff serves as Department liaisons with external media-related and other public activities. The Department follows a coordinated response for media inquiries. Media relations are coordinated by the Public Information Officer, Director of Health Communications and Marketing, and the five Regional Media Liaisons.

The Director of the Office of Health Communications and Marketing can be reached at 501-661-2474 (office). The Public Information Officer can be reached at 501-661-2150.

If it is after hours or a holiday contact the Emergency Operations Center at 501-661-2136 and they will contact the Communications Office Staff on call.

APPENDIX D: ACCR Data Processes Flow Chart