

# Investigation of Non-infectious Disease Clusters

*A disease cluster is the occurrence of more than the expected number of persons diagnosed with a certain disease within a specific group, a geographic area, or a period of time.*

**January 2020**

## Cancer/Cluster Analysis Work Group (CAWG)

Cancer Data Registry of Idaho (CDRI)  
Idaho Hospital Association

Bureau of Communicable Disease Prevention  
Bureau of Community and Environmental Health  
Bureau of Vital Records and Health Statistics  
Division of Public Health  
Idaho Department of Health and Welfare

Idaho Department of Environmental Quality

## Table of Contents

<b>Introduction .....</b>	<b>2</b>
<b>PHASE I: RECEIPT OF REPORT AND INITIAL EVALUATION .....</b>	<b>5</b>
<b>PHASE II: CASE VERIFICATION AND RATE COMPARISON .....</b>	<b>8</b>
<b>PHASE III: DETERMINING FEASIBILITY OF CONDUCTING AN EPIDEMIOLOGIC STUDY .....</b>	<b>14</b>
<b>PHASE IV: TARGETED SURVEILLANCE OR EPIDEMIOLOGIC STUDY.....</b>	<b>16</b>
<b>APPENDIX A: INITIAL INQUIRY REPORT FORM .....</b>	<b>18</b>
<b>APPENDIX B: TALKING POINTS ON CLUSTERS AND CLUSTER FACT SHEET ...</b>	<b>20</b>

# Introduction

## Scope of this Document

Each year, more than 1,000 suspected disease clusters are reported to state health departments; most of them are cancer clusters. CDC defines a cancer cluster as a greater than expected number of cancer cases that occurs within a group of people in a geographic area over a defined time period.

Most reports to the Division of Public Health of non-infectious illness clusters involve cancer. However, the tools, methods, and decision points involved in the investigation of cancer clusters are also applicable to investigation of other non-infectious diseases and conditions, such as miscarriages and birth defects. These procedures are written to address non-infectious disease investigations as a whole, with an emphasis on cancer.

These procedures identify several decision points in the investigation process. These points will ensure that further resources are directed consistently and responsibly, and that the extent of investigations is based on sound epidemiologic principles.

These procedures are directed mainly toward the investigation of illnesses occurring in residential settings (e.g., neighborhood or community). When the workplace or other site is the target of the investigation, the methods described herein may be modified or additional investigative methods and partnerships may be employed (see resources below).

The procedures provide a rough blueprint for an investigation but cannot prescribe exactly what to do in every situation. Consideration of the local resource base, including staff skills, and professional judgment at decision making points need to be used in planning a response. Other resources may be used to help determine the appropriate response, including:

- “Investigating Suspected Cancer Clusters and Responding to Community Concerns: Guidelines from CDC and the Council of State and Territorial Epidemiologists” (MMWR Recomm Rep. 2013 Sep 27;62(RR-08):1-24), available at <https://www.cdc.gov/mmwr/pdf/rr/rr6208.pdf>.
- “Guidelines for Investigating Clusters of Health Events” (MMWR July 27, 1990/Vol. 39/No. RR-11), available at <https://www.cdc.gov/mmwr/preview/mmwrhtml/00001797.htm>.
- National Institute for Occupational Safety and Health. Occupational cancer. Washington, DC: National Institute for Occupational Safety and Health; 2010. Available at <https://www.cdc.gov/niosh/topics/cancer/default.html>.
- Brown AM. Investigating clusters in the workplace and beyond. *Occup Med (Lond)* 1999;49:443–7.
- “Cancer Clusters: A Toolkit for Communicators.” A Collaboration of The Centers for Disease Control and Prevention and The National Public Health Information Coalition, September 2013. Available at: <https://www.nphic.org/toolkits/cancer-cluster>.

Use of the procedures alone cannot guarantee a timely resolution of the problem under investigation, nor will it guarantee finding an answer to why a cluster may be occurring in an area of concern.

There are numerous non-infectious illnesses that are difficult to validate and for which there is

little information about “background” rates of illness in the population. These are difficult to investigate using this manual and include: subjective symptoms (e.g., headache); conditions such as “chemical allergy” for which there are no standard diagnostic criteria and population rates; and minor conditions, such as skin rash, for which a physician may not have been consulted.

## **Phases of the Investigative Procedure**

The investigative procedure is broken into several distinct phases, with decision points at each. Each phase is described in detail, beginning on page three. They are:

- Receipt of Report and Initial Evaluation (Phase I);
- Case Verification and Rate Comparison (Phase II);
- Determining Feasibility of Conducting an Epidemiologic Study (Phase III); and
- Targeted Surveillance or Epidemiologic Study (Phase IV).

## **Conclusion of the Investigation**

If an investigation is terminated at any point, there are steps that must be taken to conclude the investigation. Depending on the scope and complexity of the investigation, these steps are:

1. Write a report with a summary and conclusion. The format, length, and scope of the report depend on the investigation.
2. Obtain appropriate review. This can range from internal CAWG and Department of Health and Welfare review to review from other agencies such as the Centers for Disease Control and Prevention (CDC) and the Agency for Toxic Substances and Disease Registry (ATSDR). The decision to seek outside review is at the discretion of CAWG and depends upon the availability of subject matter or methodological expertise from outside agencies.
3. Communicate results to the public. This could range from writing a letter to releasing a public announcement to convening public meetings.

## **Final Report**

For each request, a final written report or letter will be produced, which will be distributed in a timely manner to the inquirer initially reporting the cluster, to the local public health district, and others as deemed appropriate by CAWG. This report will also be made available to any member of the public who requests it. If elevated rates of disease were found and findings suggest that preventable or avoidable causes (such as smoking or environmental exposure) have played a role in the increase, the report will include recommendations to the individual and the affected community regarding protection of their health. CAWG will work with IDHW public information officers on additional communication mechanisms, as appropriate, for the media, local public health districts, and community groups.

In addition, the lead investigator will be available during the investigation to answer questions from the public and the media and will draft press releases as needed for

consideration by CAWG members, the Administrator for the Division of Public Health and the CDRI and IDHW public information officers.

### **Protecting the Public if a Cause is Discovered**

Statistics indicate that it is difficult to determine specific causes of non-infectious disease clusters in community settings; however, it is important to note that there have been cluster investigations which have led to recommendations of value to the public. Should an investigation uncover a possible cause of a reported cluster, CAWG will ensure, through consultation with other health experts and agencies, that the findings are made available to key responding agencies as appropriate to reduce exposure, and to educate the public. In addition, communication regarding cancer cluster investigations provides the opportunity to educate about and evaluate other public health actions, such as smoking cessation programs, cancer screenings, public health consultations, and public health assessments.

# PHASE I: RECEIPT OF REPORT AND INITIAL EVALUATION

Phase I is divided into two sections: Cancer Reports and Non-cancer Reports

## CANCER REPORTS

1. The public health epidemiology staff receiving a report of suspected clusters involving cancer should complete the cluster form (Appendix A), which shall then be forwarded to the Cancer Data Registry of Idaho (CDRI) Epidemiologist. If a letter or email is received, it may be simply forwarded to CDRI. Alternatively, if a written request is received, it may be evaluated further to determine if the needed information is contained in the request.

### CDRI Contact information:

Email:	<a href="mailto:cjohnson@teamiha.org">cjohnson@teamiha.org</a>
Phone:	208-489-1380
Fax:	208-344-0180
Mailing address:	615 N. 7th Street P.O. Box 1278 Boise, Idaho 83701
Web site:	<a href="http://www.idcancer.org">www.idcancer.org</a>

2. The CDRI Epidemiologist will evaluate the form to see if the cluster report meets the following criteria:
  - A. The geographic area is a health district, county, ZIP code, or other small area such as neighborhood or workplace. (Facility-specific data requested by hospitals are not eligible under these criteria.)
  - B. The statistics are not previously published. For example, state, district-specific, and county cancer statistics are published annually by CDRI.
3. If the report does not meet the above criteria, a written response will be given by the CDRI Epidemiologist to the inquirer, with cancer fact sheets, statistical information, or other information as appropriate; additional information may be requested as well.
4. If the report meets the above criteria, an initial case definition will be formed by the CDRI Epidemiologist. It may be necessary to contact the inquirer to develop the case definition. The case definition will include:
  - Types of cancer and/or organ sites believed to be in excess
  - Location of index cases (geographic area, population group, place of work)
  - Time period of concern (for diagnosis of cases)
  - Suspected environmental exposures and likely period of exposures (if any)
  - Other risk factors (e.g., diet, infections, and family history)
  - Demographic characteristics of cases
  - How the person learned about the supposed cluster

During this phase, the CDRI Epidemiologist will also provide reference materials such as a County Cancer Profile and Cancer Cluster Fact Sheet to the inquirer.

5. Decision to Close the Investigation at Phase I.

Based on the information presented by the inquirer, the CDRI Epidemiologist will make an initial judgment about the advisability of pursuing an inquiry into the suspected cancer cluster. The decision might require discussions with other CAWG members and/or additional subject matter experts. Multiple factors bear on this decision, but it is primarily based on whether the evidence as presented fits the definition of a cluster and the biologic plausibility that the cancers could share a common etiology. Such factors as reports involving a rare cancer or an atypical demographic distribution of a certain type of cancer (e.g., multiple cases of breast cancer in men) support the decision to investigate further and should be considered. If exposure to a specific environmental contaminant is a concern in the community, the consensus in the scientific literature regarding an association between the environmental contaminant and the cancer of concern should be considered. Factors that do not support the need for further investigation include:

- Cancer cases within family members who are linked genetically (especially cancers known to be strongly genetically related)
- Reported disease that might not be cancer
- Different types of cancers not known to be related to one another
- A few cases of very common cancers (e.g., breast, lung)
- Cancer cases among persons who did not live in the same geographic location during the relevant timeframe based on latency, and thus could not have experienced a common carcinogenic exposure, and
- The lack of a plausible environmental cause

The CDRI Epidemiologist should clearly and accurately explain the rationale used to determine if an investigation will or will not be pursued based on the information provided about the cases and this protocol. If an inquirer is reporting an event that is not a suspected cancer cluster, but rather one involving a known or possible environmental contamination, they should be referred to the Idaho Department of Environmental Quality.

A decision at Phase 1 to not pursue an investigation is based on the determination that the reported cases are unlikely to comprise a cancer cluster; therefore, conducting a statistical assessment to determine whether an excess of cancer cases exists might be unsuccessful because the cancers are not likely to share a common, environmental etiology. This determination might involve multiple communications with the inquirer, as well as additional data-gathering. If the inquirer acknowledges and is satisfied with the decision to not move forward, the inquiry can be closed at Phase 1. If the inquirer is not satisfied with the decision and the verbal explanation, then CAWG shall provide a written explanation and include resources related to the decision. Regardless of the decision, CAWG shall document in a permanent log all information about the inquiry and the decision.

6. Decision to Proceed to Phase II.

If the information provided supports the decision to investigate the cancer concerns further, the CDRI Epidemiologist shall notify the inquirer, explain what that entails and

outline how CAWG will follow up with the inquirer and provide results. The CDRI Epidemiologist should ask the inquirer if there are others in the community (e.g., other residents with this cancer type) who would like to have a report on the results of the next step.

## NON-CANCER REPORTS

1. The public health epidemiology staff receiving a report of a suspected cluster of non-infectious disease should investigate the initial report as any other epidemiology investigation. If the public health district staff feels that state epidemiology assistance is needed, they should contact the state Bureau of Communicable Disease Control and Prevention at 208-334-5939.
2. The State Epidemiologist will discuss the suspected cluster with the district epidemiologist or epidemiology team, and together develop an initial plan for gathering more information about the suspected cluster.
3. Once initial information is gathered, the district and state epidemiology teams will develop an initial case definition. It may be necessary to contact the inquirer to develop the case definition. The case definition will include:
  - Disease or condition believed to be in excess
  - Location of index cases (geographic area, population group, place of work)
  - Time period of concern (for diagnosis of cases)
  - Suspected environmental exposures and likely period of exposures (if any)
  - Demographic characteristics of cases

Several criteria should be considered for making the decision to perform further investigation, including:

- Presence, magnitude and trend of excess observed cases
- Presence of an exposure scenario likely to cause the disease or condition
- Public concern.

The decision to proceed to Phase II is to be made by the district and State Epidemiologist. If further investigation is not warranted, the district epidemiologist, State Epidemiologist or other designated person will write a report and mail or email it to the person who reported the cluster. If it is necessary to cease the investigation because of a lack of information from the inquirer, this should be communicated in writing with an offer to follow up if further information becomes available.

4. If further investigation is warranted, proceed to Phase II.

## PHASE II: CASE VERIFICATION AND RATE COMPARISON

Phase II begins by presentation of the apparent cluster to CAWG. The initial presentation should result in a group discussion as to whether further evaluation is warranted, and determination of what additional information is needed.

### 1. Verification of "Index" Cases

If the initial information gathered based upon the inquirer's report reveals a need to investigate the matter further, the next step is to verify the "index" cases that have been reported. Because CDRI maintains a population-based cancer registry, it is often possible to verify cancer cases for certain levels of geography, including state, health district, county, ZIP code area, and census tract. For other levels of geography, including neighborhood, and for non-infectious diseases other than cancer, it is necessary to verify cases using other sources of data, including physician records, hospital records, and vital records. The availability of information about specific health problems can be limited because of confidentiality and access to records. Case verification for cancer will be performed by CDRI. Case verification for non-infectious diseases other than cancer will be performed under the direction of the State Epidemiologist.

Generally, the following information is needed to verify a case:

- Name
- Gender
- Date of birth
- Residence (preferably street address) at diagnosis
- Social Security number, if available

Full case ascertainment means finding all additional unreported cases of the disease in question which occurred in the location during the time period of interest and meet the case definition. If there is no registry or vital records system covering the disease in question, full ascertainment can be very arduous. When available, a registry can easily provide full ascertainment of cases or can be used to match case names against the roster of cohort members from a time-cohort cluster.

*To protect confidentiality, at no time during the case verification process, or thereafter, will individual-specific case data, including whether or not the reported case was confirmed as a cancer case, be shared with the inquirer, even if the inquirer initially provided the case names.*

In many instances, the original reports of illness are not supported by medical records. Sometimes, cases reported to public health officials occurred in persons who developed the disease or condition prior to moving into the area under study, and therefore should not be counted (in accordance with national cancer registry reporting standards). Reported clusters of one kind of disease may turn out to be a mixture of several conditions. Cases reported as a particular type of cancer may be found, when verified, to be several different types of cancer or not cancer at all. These findings should be reported to the inquirer in a general manner, without discussing specific cases.

## 2. Comparison of Observed and Expected Cases

As a first step, selection of the exact area of concern and selection of the area with which it is to be compared must be determined. These determinations are critical as they will guide the rest of the evaluation. Often, the comparison area is the remainder of the state of Idaho (for example, a certain county is the area of concern, and disease rates in that county are compared with disease rates in the other 43 counties in Idaho).

The expected number of cases will generally be calculated based upon age and sex-specific rates for the remainder of the state of Idaho during the same time period as the reported cluster. Five-year age categories and both sexes will be used in calculating stratum-specific rates for the remainder of Idaho. These rates will be applied to stratum-specific population estimates for the geographic area of interest to calculate the number of expected cases. Person-year estimates will be made by summing census population estimates over the time frame of interest. Relative risks and p-values will be calculated for tests of observed versus expected numbers of cases. The statistical calculations will be performed by the CDRI Epidemiologist, State Epidemiologist or designated Division of Public Health epidemiologist, or staff from the Bureau of Vital Records and Health Statistics.

The counting of observed cases and calculation of expected cases becomes more difficult as the geographic area of interest becomes smaller. Thus, CAWG can readily conduct analyses at the county level, but may be unable to perform an analysis for a particular street block due to the quality of address data available. Geocoding is the process of adding geographic location information to case records and is used to map case data and for small area analyses. The ability to geocode and map case data depends in part on the quality of the address information provided on hospital and laboratory records and death certificates.

As of November 2019, CDRI has geocoded data for the years 1990–2017. Cancer case records are assigned different geocode quality codes depending on the ability to match the case information to street locations and other map features. The percent of cases geocoded to various quality codes varies by county, as shown in Table 1. Analysis by geographic area is limited to those areas where  $\geq 90\%$  of cases are geocoded to a sufficient quality code. For example, for 2013–2017, analyses at the census tract level are available for all counties, because all counties have  $\geq 90\%$  of case data geocoded to support this level of analysis. Thirty-seven counties have  $\geq 90\%$  of case data geocoded to at least the block group level; for these counties, small area analyses may be conducted at the block group level. While nearly all CDRI cases have data on the ZIP Code and city of residence at the time of diagnosis, it is less desirable to conduct cancer cluster investigations at these levels of geography. For ZIP Code-level analyses, the population estimates are for generalized area representations of the United States Postal Service ZIP code service areas called ZIP Code Tabulation Areas (ZCTAs), but these are not the same as ZIP codes. Each ZCTA is constructed by aggregating the Census 2010 blocks for which most addresses use a given ZIP code. As a result, some addresses are assigned a ZCTA code that is different from their ZIP code. In addition, ZCTAs are not developed for ZIP codes that comprise only a small number of addresses. For city-level analyses, the city boundaries may be larger than the area of interest, and there may be misalignment of boundaries used for addresses of numerators and denominators because of city names being used for areas outside of city limits. In general, the population size of a typical census tract is the smallest denominator that will allow reliable results to be generated.

The highlighted cells in the Table 1 show in which counties small area analyses may be conducted at which geographic level. Small area analyses may also be conducted for the non-highlighted cells, but this will require additional resources to manually plot all cases in the area of interest, such that the remainder of the county may serve as the comparison area.

**Table 1: Cancer Data Registry of Idaho Geocoded Quality Codes by County, 2013 – 2017.**

County	Cumulative Percent of Geocoded Cases at Quality Code to Support Small Area Analysis at Given Level:				
	Block	Block Group	Tract/ZIP Code	City	County
ADA	100%	100%	100%	100%	100%
ADAMS	86%	87%	100%	100%	100%
BANNOCK	96%	97%	100%	100%	100%
BEAR LAKE	89%	90%	100%	100%	100%
BENEWAH	94%	95%	100%	100%	100%
BINGHAM	95%	96%	100%	100%	100%
BLAINE	80%	81%	100%	100%	100%
BOISE	93%	93%	100%	100%	100%
BONNER	95%	96%	100%	100%	100%
BONNEVILLE	99%	99%	100%	100%	100%
BOUNDARY	93%	95%	100%	100%	100%
BUTTE	82%	89%	100%	100%	100%
CAMAS	49%	54%	100%	100%	100%
CANYON	99%	99%	100%	100%	100%
CARIBOU	97%	97%	100%	100%	100%
CASSIA	96%	98%	100%	100%	100%
CLARK	57%	71%	100%	100%	100%
CLEARWATER	94%	95%	100%	100%	100%
CUSTER	73%	80%	100%	100%	100%
ELMORE	96%	97%	100%	100%	100%
FRANKLIN	96%	98%	100%	100%	100%
FREMONT	96%	97%	100%	100%	100%
GEM	99%	99%	100%	100%	100%
GOODING	97%	97%	100%	100%	100%
IDAHO	93%	95%	100%	100%	100%
JEFFERSON	98%	99%	100%	100%	100%
JEROME	98%	98%	100%	100%	100%
KOOTENAI	99%	99%	100%	100%	100%
LATAH	97%	97%	100%	100%	100%
LEMHI	91%	92%	100%	100%	100%
LEWIS	88%	90%	100%	100%	100%
LINCOLN	91%	92%	100%	100%	100%
MADISON	98%	98%	100%	100%	100%
MINIDOKA	96%	97%	100%	100%	100%
NEZ PERCE	99%	99%	100%	100%	100%
ONEIDA	99%	100%	100%	100%	100%
OWYHEE	93%	94%	100%	100%	100%
PAYETTE	98%	98%	100%	100%	100%
POWER	99%	99%	100%	100%	100%
SHOSHONE	94%	95%	100%	100%	100%
TETON	90%	90%	100%	100%	100%
TWIN FALLS	99%	99%	100%	100%	100%
VALLEY	86%	86%	100%	100%	100%
WASHINGTON	99%	99%	100%	100%	100%

Highlighted cells show in which counties small area analyses may be conducted at which geographic level. Note: Subject to change annually as additional cases are geocoded. When small area analyses are performed for cancer incidence with geocoded data, the comparison group(s) could be the remainder of Idaho, the remainder of the public health district, or the remainder of that county geocoded to the same or better quality code. The objective is to select a comparison population otherwise similar to the area of interest and to have a sufficient population size for statistically stable comparison rates.

In general, the statistical analysis will be conducted using the spatial scan statistic (SaTScan software) at the census tract level of geography on the entire state of Idaho to identify clusters of both high and low rates using the discrete Poisson model, with a maximum scanning window size to include up to 50% of the Idaho population. P-values will be derived from Monte Carlo replications under the null hypothesis of spatial randomness of the cancer of interest. Two separate analyses will be conducted: one adjusting for the age and sex distribution of the population, and a second that adjusts for age, sex, and additional area-based measures. In the second analysis, area-based measures of variables related to the cancer of interest will be combined with individual case characteristics including age at diagnosis, sex, and year of diagnosis. For example, county-level smoking rates are known to influence lung cancer incidence rates, and intensity of screening efforts may influence early-stage diagnosis of cancers amenable to screening. The purpose of this approach is to combine cluster detection analysis techniques with multilevel modeling of area-level influences on disease patterns, in order to examine the relationship between socioeconomic and behavioral influences and spatial patterning.

Properties that make the spatial scan statistic suitable for cluster analyses include:

- its ability to account for the uneven geographic distribution of cases and population densities
- its lack of assumptions about cluster size or location,
- its ability to adjust for multiple testing,
- its ability to identify the spatial locations where the null hypothesis is rejected, and
- its ability to detect multiple clusters.

Pertinent references for this approach include:

Kulldorff M. and Information Management Services, Inc. SaTScan: Software for the spatial and space-time scan statistics. <http://www.satscan.org/>

Kulldorff M. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26:1481-1496, 1997.

Klassen A, Kulldorff M, Curriero F. Geographical clustering of prostate cancer grade and stage at diagnosis, before and after adjustment for risk factors. *International Journal of Health Geographics*, 2005, 4:1.

In addition to this approach, other methods may be used depending on the quality of the geocoded data for the area of interest and the remainder of the state of Idaho. For example, if the area of interest is limited to a small area such as a census block, it would not be possible to utilize the spatial scan statistic on the entire state of Idaho because not all counties have

geocoded case data to support analysis at this level of geography.

### 3. Written Report

Written results of the previous step will be sent to the inquirer. For cancer, the CDRI Epidemiologist will be responsible for writing the report. The report will include, as an attachment, a Cancer Cluster Fact Sheet. For other non-infectious diseases, the district epidemiologist, State Epidemiologist or other principal investigator will be responsible for writing the report. A draft of the report will be submitted to CAWG members and the Division of Public Health Administrator. A period of one week (five working days) will be allowed for review. After comments are incorporated into the report, the final copy will be submitted to the inquirer, with copies to all CAWG members and the Division of Public Health Administrator. If the results of the inquiry show no excess, the cluster investigation is considered closed, unless continuing community concern is high; in this case, CAWG will discuss whether further action is needed.

Reported clusters will, upon investigation, fall into three categories: no excess; explained excess; unexplained excess:

*No excess.* Often an initial investigation reveals that no excess exists. This occurs when the observed number of cases for an area during a specific time frame is less than or equal to the expected number of cases for that area based upon rates in the remainder of Idaho. This also occurs when the observed number of cases is numerically greater than the expected number of cases, but not statistically significantly different from the expected number of cases (i.e.,  $p\text{-value} \geq 0.05$  or 95% confidence interval for rate ratio crosses 1.0). If an investigation shows no excess, and no further action is needed, CAWG will be copied on the response to the individual requesting the analysis.

*Explained excess.* Based upon experience to date in many states, concerns regarding non-infectious disease clusters arise because the public is not aware how common these conditions, such as cancer, spontaneous abortion, and birth defects, are. For example, an excess of lung cancer in a retirement community with a high percentage of smokers and no unusual environmental exposure is not likely to constitute a cluster. Typically, citizen concern subsides when they are adequately informed of the issue.

*Unexplained excess.* In some instances, however, the inquirer's concerns are confirmed. The number of cases may be more than expected based upon comparison rates (observed cases statistically greater than expected cases:  $p < 0.05$  or 95% CI  $> 1.0$ ), indicating that the concern warrants further investigation.

If an investigation shows any excess, or further action may be needed due to high community concern, CAWG will review the response to the individual requesting the analysis prior to releasing results.

An important consideration is the issue of practical versus statistical significance. If the case counts (observed and expected) are large enough, minor differences are more easily detected and may be statistically significant. However, the same difference may be of little practical or clinical significance (for example, a difference of 1% in a disease rate). Furthermore, rates based upon small numbers (i.e., fewer than 10 cases) are subject to substantial random variation. If the number of infant deaths in a county increased from 1 in 1994 to 2 in 1995, and

the number of births remained approximately constant, looking at the infant mortality rate would erroneously suggest that the problem had become twice as great. Examining the numbers behind rates is always a good idea, and in some cases just looking at the numbers makes more sense.

To address the problem of rates based on small numbers, all communications that contain rates or percentages should contain a caution about interpreting rates or percentages based on small numbers. An example is: "Rates based upon 10 or fewer cases (numerator) should be interpreted with caution, since they may vary greatly over time." In addition, for sub-county areas, a small cell suppression rule will be invoked to not present statistics when the number of cancer cases is less than 5, unless at least 5 years of data are aggregated. This rule will hold for statistics for all cancer sites combined and for individual primary sites. The rule may be disregarded with CAWG approval, such as in circumstances when rare diseases and conditions generate small numbers of cases that may be important for public health.

## PHASE III: DETERMINING FEASIBILITY OF CONDUCTING AN EPIDEMIOLOGIC STUDY

The purpose of Phase III is to assess the feasibility of performing an epidemiologic study to examine the association between the cancer cluster and a particular environmental contaminant. If further study is feasible, an outcome of this step should include a recommended study design. All activities in this step should be carried out in collaboration with community, environmental, and other partners. This step also provides the opportunity to evaluate additional public health actions, such as smoking cessation programs, cancer screenings, health risk assessments, removal of environmental hazards, or other activities that should be conducted. If beneficial to public health, these actions should not be delayed pending the decision to conduct or complete an epidemiologic study focused on assessing the association between the cluster of cases and a suspected environmental cause.

The first actions in determining the feasibility of further study of the identified cluster include determining the study hypothesis and reviewing the scientific literature. Investigators should share information about time, cost, goals, purpose, and limitations of a potential study with all partners and carefully communicate realistic expectations.

Investigators must assess potential study design issues as sample size, a small case number, and study power. Experienced scientists with appropriate skills should be included in the investigative team. The experts should include an epidemiologist, a toxicologist, a physician, an environmental protection specialist, and a community-nominated expert and/or local representative to provide advice on the assessment as needed. It is necessary to identify such parameters as study population and its characteristics, including what descriptive, health, and risk factor data should be collected and determine the feasibility of obtaining the data.

Investigators should:

- Confirm case diagnoses and determine which types of cancer and which cases meet the case definition
- Identify a comparison group that, depending on the study design, does not have the cancer of concern (i.e., a control group in a case-control study) or does not have the exposure of concern (i.e., unexposed group in a cohort study)
- Consider the willingness of persons to participate in interviews or studies for gathering data on health, possible exposures, the amount of time the affected persons have lived in the area, occupation, and other relevant risk factors and confounding variables
- Verify whether the environmental contaminants of concern are known carcinogens, consider possible and plausible routes of exposure, ascertain whether or not cases were exposed to an environmental contaminant in sufficient doses and for a sufficient time to make the association biologically plausible, and consider if the time sequence of exposure is consistent with the latency period and the causation of these particular cancers
- Determine whether residential and occupational histories for affected persons are obtainable
- Determine if it is possible to characterize exposure to suspected environmental hazards accurately at the individual level and in a way that reflects the period of concern

It is not recommended to engage in a general, open-ended inquiry to identify potential

contaminants in a community, in the absence of a suspected etiologic agent. Investigators should identify study design requirements and available resources to conduct the study. This process includes identifying the scope of the study and determining whether sufficient resources and data are available to complete meaningful work. Investigators should:

- Determine which parameters to use for geographic scope, study timeframe, and demographics and select a timeframe that allows for sufficient latency in cancers of concern
- Determine the study design, sample size, and the statistical tests necessary to study the association as well as the effect of a smaller sample size on statistical power
- Determine the appropriateness of the plan of analyses, including hypotheses to be tested as well as epidemiologic and policy implications, and
- Assess resource implications and requirements of the study and identify sources of funding

#### Decision to Close the Investigation at Phase III.

In some cases, despite the finding of a significantly elevated standardized incidence ratio (SIR), the feasibility assessment might indicate that further study will likely be unable to determine the cause of the elevated rate. In situations in which the types of cancers have no known association with an environmental contaminant, in which there are only a handful of cases, in which no suspected environmental hazard exists, or in which other factors explain the observed cancer excess (e.g., a substantial movement of residents during the study period), investigators might determine that data are insufficient or that insufficient justification exists for conducting further epidemiologic study.

If the feasibility assessment suggests that little will be gained from proceeding further, the investigator should close the inquiry and summarize the results of this extensive process in a report to the inquirer and all other concerned parties. In some circumstances, the public or the media might continue to demand further investigation, regardless of cost or biologic plausibility. Working with established community relationships, media contacts, and the advisory panel will be critical in managing and responding to expectations. If an extensive epidemiologic investigation is not carried out, it is critical to establish other possible options to support the community's health, depending on the information and resources available.

#### Decision to Continue to Phase IV.

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

## PHASE IV: TARGETED SURVEILLANCE OR EPIDEMIOLOGIC STUDY

If there is a decision to conduct additional analyses or studies, there are two possible routes: 1) targeted surveillance; and/or 2) epidemiologic study.

*Targeted Surveillance:* When the incidence of a non-infectious disease has been excessive in a small locality, the local officials and the public will want to know if this problem is ongoing or limited to a certain time period. For diseases or conditions covered by registries or vital records, a review of incidence can be conducted. Where a registry does not exist, it could be very difficult to monitor the disease; this may necessitate establishing a community reporting system to receive citizen reports about the disease and to monitor new cases in the community. This would be performed by the local Public Health District in consultation with the State Epidemiologist.

During implementation of targeted surveillance, it is important to gather accurate population estimates, as a census undercount in an area, and not an excess of cases, may be responsible for elevated disease rates.

*Epidemiologic Study:* This step involves a standard epidemiologic study that tests a hypothesis of the association between putative exposures and specific cancer types, for which all the preceding effort has been preparatory. Using the feasibility assessment as a guide, responders should develop a protocol and implement the study. The epidemiologic study will, at a minimum, be used to collect additional exposure history information about cases, and may include a case-control study, a cohort study, other study designs and possibly environmental sampling. The planning and implementation of such a study will be performed by members of CAWG with leadership by the State Epidemiologist, the district health departments, federal partners such as ATSDR, or the CDRI Epidemiologist, depending on the type of disease or condition of concern, and the complexity of the study, as determined by CAWG. For any epidemiologic study, it is important to have a robust working relationship with the community. With local knowledge about the hazards and risk factors in the community, investigators can make more informed decisions. CAWG will engage the inquirer in the selection of participants of a community panel for field studies that shall include, at a minimum, the inquirer, the local public health district, and membership from the local medical community.

The primary purpose of conducting an epidemiologic investigation of a statistically significant cancer cluster is to determine if exposure to a specific risk factor or environmental contaminant might be associated with the cluster. The results of the investigation are expected to contribute to epidemiologic and public health knowledge. It is acknowledged that because investigations of community-based cancer clusters rarely demonstrate a clear association with an environmental contaminant, they usually do not provide the resolution communities seek. Furthermore, an investigation can augment the existing fear and uncertainty in the community brought on by the perception that suspected cancer cluster exists, which might have negative social and economic impacts.

Epidemiologic studies are dependent in part upon the availability of funds and staff to properly implement the study. Often, multiple barriers exist which must be examined and overcome in order to proceed further. Barriers often include, but are not limited to:

- Persons identified as part of a cluster may be deceased or not locatable, and therefore unable to provide a detailed exposure history
- Persons identified as part of a cluster may not be willing to participate in a survey or health study
- Funding for additional studies is limited
- Federal assistance (e.g., EPA, ATSDR, CDC) is often needed for technical expertise, funding, and laboratory testing, and this assistance takes time to procure, or may not be available
- Privacy and confidentiality of all persons in the possibly affected community must be respected
- Collection and testing of environmental samples may cause significant delays
- Persons identified as part of a possible cluster may live or work outside of Idaho, and obtaining information on current whereabouts, disease information, or other information may be time-consuming or even at times not possible

Demonstrating a statistically significant association does not prove causation. The scientific rigor necessary for determining causation is difficult to achieve with an epidemiologic study alone; in addition, determining causation often relies on clinical and laboratory studies. Even if a cancer cluster is identified and environmental contamination is identified, an investigation might not demonstrate a conclusive association between the contamination and cancer. Other risk factors (e.g., smoking, personal behavior, occupational exposures and genetic traits) should also be explored. Conversely, even if the investigation does not identify an association between a particular suspected environmental exposure and cancer cluster, the exposure still might be linked to the cluster; however, in such a case more scientific information might be required (e.g., toxicologic and clinical data) to establish an association. Epidemiologic studies alone often are not able to detect small effects, particularly in small populations or when the number of cases is limited.

# APPENDIX A: INITIAL INQUIRY REPORT FORM

## Idaho Cancer and Non-Cancer Cluster Investigation Initial Inquiry Report Form

### Inquirer Information

First Name      Last Name      Phone Number

Street Address      City      State      Zip

E-mail Address

Affiliation of the Inquirer:

- Concerned citizen
- Private physician
- Employer representative
- Other \_\_\_\_\_ (specify)

### Area of Concern

Where has the reported cluster occurred (which city, county, neighborhood, etc.)?

Does the inquirer suspect a specific environmental exposure?

What is the time period during which people became ill?

What types of illnesses are being reported?

How many people (list ages, if known) are reported with illness?

If concern involves cancer, fax this form to: CDRI: fax 208-344-0180

## APPENDIX B: TALKING POINTS ON CLUSTERS AND CLUSTER FACT SHEET

**Here are some points which might be helpful when talking with a caller concerned about a disease cluster:**

- Usually clusters occur by chance alone and are not related to a specific exposure. That is to say, each case in the cluster probably has a different cause, even though the cases have clustered together in time and/or space.
- It's difficult to reconstruct exposure histories. This is especially true for diseases with long latency periods between exposure to a disease-causing agent and the onset of disease symptoms. What's in the air or water today may not be what was in the air or water several years or decades ago.
- It's difficult to detect subtle effects, especially when the number of cases is small. If the relationship were strong, we may have seen an association in other places, such as workers, where exposures are higher.
- For diseases of unknown etiology, we often don't know what to look for as a possible cause, unless there is a unique exposure of concern.

### **Regarding Cancer**

- Cancer is a term for a group of more than 100 different diseases in which abnormal cells divide without control and can invade nearby tissues. Cancer is very common: about 1 in 2 men and women will be diagnosed with cancer sometime in their life. About 1 in 4 deaths in the US is attributable to some form of cancer. Cancer is the leading cause of death in Idaho.
- The causes of many types of cancer are unknown.
- Cancer is almost always caused by a combination of factors that interact in ways that are not yet fully understood.
- Cancer is more likely to occur as people get older; because people are living longer, more cases of cancer can be expected in the future. This may create the impression that cancer is becoming much more common, when an increase in the number of cases of cancer is partly related to the aging of the population.
- There are many different types of cancer, which are caused by a wide variety of causal mechanisms. A variety of diagnoses speaks against a common origin.
- A cancer that spreads to another part of the body should not be considered a new case of cancer. For example, if a breast cancer spreads to the lung, this is not considered to be a new lung cancer.

- Some types of cancers may occur anywhere in the body. They should not be classified according to where they appear in the body. For example, non-Hodgkin lymphoma may manifest itself in the brain, but it is not brain cancer.
- Cancer involves a series of changes within cells that usually occur over the course of many years. More than 10 years can go by between the first cellular abnormality and the clinical recognition that cancer is present, which often makes it difficult to pinpoint the cause of the cancer.

#### **Regarding Birth Defects**

- Major birth defects occur in 1%–2% of live births.
- The causes of most birth defects are unknown.

#### **Useful Information for Reducing the Risk of Chronic Diseases**

- Don't smoke or chew tobacco.
- Eat at least 5 servings per day of fruits and vegetables.
- Limit the amount of fat — especially saturated fat — in your diet.
- Exercise regularly — one hour each day.
- Limit alcohol intake.
- Protect yourself from sunburn.
- Follow recommended guidelines for preventive services and screening for early detection and treatment, such as screening for colorectal, cervical and breast cancer, high blood pressure and high cholesterol.

#### **Useful Websites for the Consumer**

National Cancer Institute, "Cancer Clusters": <http://www.cancer.gov/cancertopics/factsheet/Risk/clusters>

CDC Cancer Web site: <http://www.cdc.gov/nceh/clusters/>

# CANCER CLUSTER FACT SHEET

Cancer is a term that includes more than 100 different diseases, each characterized by the uncontrolled growth and spread of abnormal cells. A **CANCER CLUSTER** is the occurrence of a greater than expected number of cases of a specific type of cancer within a small area or within a short period of time.

Cancer is one of America's greatest public health concerns. Nearly one in two men and women in the United States will be diagnosed with cancer sometime in their life. Cancer is the second leading cause of death in the United States. In Idaho, for the 5-year period of 2014–2018, cancer accounted for about 22% of deaths and was the second leading cause of death. When someone is diagnosed with or dies from cancer, family, friends, and neighbors sometimes learn of other cases of cancer in their community. This apparent clustering of cancers is often reported to health departments or the media. However, closer inspection usually reveals that these "suspected" clusters involve several different types of cancer among persons of different ages, sexes, and occupations. A "real" cancer cluster will usually involve one type or site of cancer.

When several cancers occur within a limited area, this may represent a real cluster, but it may not be the result of an increased community risk of cancer. For example, in Idaho there are 44 counties and every year, about half of the counties have rates of cancer that are above the average county value, and about half have rates that are below the average value. Counties may have above average rates one year and the next year the same counties may have rates below the average. This variation is expected and is more pronounced as the population being studied gets smaller (county, city, ZIP Code, neighborhood). Investigations of hundreds of reports of cancer clusters over many years by numerous states have shown approximately 15% of reported cancer clusters to be real clusters, based upon statistical evidence.

Cancer clusters that are a public health concern are the ones that represent a group of people at unusually high risk of cancer due to some factor or exposure that they have in common. Most commonly, cancer clusters are related to lifestyle factors, such as smoking or obesity, as opposed to a common environmental exposure. Sometimes, a study of these clusters can help prevent further cancers through targeted prevention or help us understand more about specific risks for cancer. Understanding the reasons for elevated cancer risk may take months or longer, and the reasons are not always resolved. Less than 5% of all cluster reports fall into this category of a meaningful cluster.

Cancer cluster investigations require data on the total number of residents and the number of diagnosed cancer cases in the area to be reviewed. At present time, the Cancer Data Registry of Idaho can investigate cancer incidence for several levels of geography: public health district, county, and census tract.

For more information regarding cancer clusters, contact:

Cancer Data Registry of Idaho  
615 North 7<sup>th</sup>  
PO Box 1278  
Boise, Idaho 83701  
(208) 489-1380  
[www.idcancer.org](http://www.idcancer.org)