Investigation of Non-infectious Disease Clusters in Occupational Groups, Worksites and Other Cohorts

January 2026

Non-Communicable Disease Analysis Working Group (nCAWG)

Idaho Hospital Association (IHA)

Cancer Data Registry of Idaho (CDRI)

Idaho Time-Sensitive Emergency Registry (TSE Registry)

Idaho Department of Health and Welfare (IDHW), Division of Public Health

Bureau of Environmental Health and Communicable Disease

Bureau of Prevention and Community Health

Bureau of Vital Records and Health Statistics

Idaho Department of Environmental Quality

Table of Contents

Introduction

A disease cluster is the occurrence of more than the expected number of people diagnosed with a certain disease within a specific population, geographic area, and time. Among disease cluster concerns raised by the public and health care providers, cancer cluster concerns are the most common. Clusters involving other non-infectious diseases also warrant investigation, such as neurological disorders, reproductive outcomes, and sensory impairments. Examples of other types of cluster concerns previously raised in Idaho include Creutzfeldt-Jakob disease, heart defects in infants, miscarriage, hair loss, and acute, severe hearing loss. Although the focus of this document is responding to occupational cancer clusters, methods described in this document may be used to investigate a wide variety of non-infectious disease cluster concerns arising in occupational or other cohorts.

Cancer clusters of public health concern are those that represent a group of people who are at unusually high risk of cancer due to a common factor or exposure. Studying clusters may help us prevent future cancers by intervening on an exposure or better understanding risks for specific cancers. Cancer cluster concerns may be at the community level or may be related to a specific occupation or work site. Investigations of hundreds of reports of community cancer clusters over many years by numerous states have shown that approximately 15% of reported cancer cluster concerns show statistically elevated rates of cancer, meaning there is statistical evidence for the cluster.

An estimated that 4%–10% of cancers in the US are caused by occupational exposures.[1] Investigations of cancer in occupational settings have contributed critically to our understanding of cancer risks; about one third of the factors recognized as human carcinogens were first documented via worksite studies.[2] In addition, studies of worksite exposures have led to the recognition non-cancer clusters involving carcinogens, neurotoxins, and reproductive hazards.

Central cancer registries are sometimes asked to respond to reports of disease clustering within occupational groups or worksites. In Idaho, the Non-Communicable Disease Analysis Working Group (nCAWG), with membership from the Idaho Hospital Association, Idaho Department of Health and Welfare Division of Public Health, and the Idaho Department of Environmental Quality,[*] is responsible for responding to cancer cluster concerns. Understanding the reasons why cancer risk is elevated may take several months, and the causes are not always resolved.

Purpose and Scope of this Document

This document serves to establish a protocol by which nCAWG will respond to concerns of disease clusters within worksites, occupational groups, or other defined cohorts which—especially from a statistical methods standpoint—have distinct differences from community, i.e., residential, cluster concerns (see: Investigation of Non-infectious Disease Clusters available at https://idcancer.org/statistical-data/cancer-clusters.php). It is designed to guide investigations

[*] nCAWG charter available here: https://idcancer.org/pdfs/resources/cancer-clusters/nCAWG_Charter_2022_v20220612_w20230818amendment.pdf

of both cancer and non-cancer conditions, recognizing that each type of disease may require distinct approaches to case definition, data collection, verification, and analysis. Consideration of the local resource base, including staff skills and professional judgment at decision-making points, must be used in planning a response.

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 a particular population. Other resources may be used to help determine the appropriate response, including:

- Brown AM. Investigating clusters in the workplace and beyond. Occup Med (Lond) 1999; 49:443–7. doi: 10.1093/occmed/49.7.443.
- 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/cancer/about/.
- "Guidelines for Examining Unusual Patterns of Cancer and Environmental Concerns" (2022) https://www.cdc.gov/cancer-environment/php/guidelines/summary.html

Phases of the Investigative Protocol

The phases of a cancer or other non-infectious disease cluster investigation in the worksite, occupational, or other cohort setting are:

- Phase I: Receipt of Report and Initial Evaluation;
- Phase II: Determining Feasibility of Conducting an Epidemiologic Study; and
- Phase III: Epidemiologic Study.

The investigation may not include all three phases outlined above, as at each phase the investigation may reach a natural stopping point. Irrespective of achieved phase, however, at the end of a worksite cancer investigation, a report summarizing steps taken, results, and conclusions will be written by Idaho Hospital Association staff, specifically Cancer Data Registry of Idaho epidemiologists. All participating agencies will have the opportunity to review the report prior to distribution to the inquirer initially reporting the cluster concern and local public health district staff. nCAWG may determine that the report requires outside review prior to distribution. If only Idaho Hospital Association staff were involved in the response to the inquirer, the report will be distributed to members of nCAWG for their information.

Throughout the investigation, the lead investigator will be available to answer questions from the worksite or cohort contact, the public, and the media, and will draft press releases as needed for consideration by nCAWG members, Division of Public Health Administrator, IHA President, and IDHW public information officers.

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 the verbal report of a suspected cancer cluster should complete the Cluster Form (Appendix A), which shall then be forwarded to the Cancer Data Registry of Idaho (CDRI) epidemiologists along with any additional information that may support the request. If the report is received in the form of Appendix A via letter or email, it may be simply forwarded to CDRI. If an unstructured written request is received, it may need to be evaluated further to determine if the needed information is contained in the request; Appendix A can guide this determination of additional information.

   CDRI Contact information:

Name:	Chris Johnson or Bozena Morawski
Email:	cjohnson@teamiha.org or bmorawski@teamiha.org
Phone:	208-338-5100
Fax:	208-344-0180
Mailing address:	P.O. Box 1278 Boise, ID 83701
Web site:	https://idcancer.org/

2. CDRI epidemiologists will verify that the cancer cluster concern being reported is associated with a worksite, occupational, or other cohort, and that related statistics are not previously published elsewhere. Reference materials such as a County Cancer Profile and Cancer Cluster Fact Sheet will be provided to the inquirer. In addition to information provided to the inquirer, similar information may be provided to the site contact from the CDRI epidemiologist. Establishing a relationship with the worksite manager or other leadership for a cohort is important, as information may be requested to support the investigation.

3. An initial cancer case definition will be formed by a CDRI epidemiologist, based on information from the inquirer. The case definition must include:

   - Type(s) of cancer believed to be in excess. This may reflect the same primary site or cancers with known shared etiologies;
   - Location of index cases (e.g., place of work, area within worksite, shared space); and
   - Time period of concern (for diagnosis of cases).

   Additional information about the site and/or cases may be requested and include:

- Suspected environmental exposure(s) and likely period of exposures (if any)
- Other risk factors (e.g., diet, infections, and family history)
- Personal identifying information for cases (first name, middle name, last name, date of birth, sex, social security number, start and end dates of employment)
- The relationship between the inquirer and the cluster concern

4. A CDRI epidemiologist will make an initial judgment about the advisability of pursuing an inquiry into the suspected worksite, occupational, or other cancer cluster. The decision to pursue an investigation will primarily be based on whether the evidence presented fits the definition of a cluster, whether the cases can be verified using cancer surveillance data, and if there is biologic plausibility that the cancers could share a common etiology. Case verification for cancer will be performed by CDRI. To protect confidentiality, at no time during the case verification process, or thereafter, will individual-specific case data, including whether the reported case was confirmed as a cancer case, be shared with the inquirer, even if the inquirer initially provided personal identifying information for the cases.

   After the initial review, a presentation of the cluster concern will be made to nCAWG, and the CDRI epidemiologist will make a recommendation to nCAWG on how to proceed. Should nCAWG agree that further evaluation is warranted, the group will identify additional information needed to compare observed and expected cases of cancer or non-infectious diseases in the cohort to a comparable population. nCAWG shall document information about the inquiry and the decision.

   The decision to close the investigation at this point might require discussions with additional subject matter experts, as well as multiple communications with the inquirer to gather additional data. If an inquirer is reporting an event that is not a suspected cancer cluster but rather one involving a known or possible environmental contaminant, they should be referred to the Idaho Department of Environmental Quality and IDHW Bureau of Environmental Health and Communicable Disease (Environmental Health Program). If the inquirer is not satisfied with the decision to not pursue the investigation, nCAWG shall provide a written explanation and include resources related to the decision.

   The decision to pursue the investigation shall prompt the CDRI epidemiologist leading the investigation to notify the inquirer, explain what that entails, and outline how nCAWG will follow up with the inquirer and provide results. The CDRI epidemiologist should ask the inquirer if there are others at the cluster concern location (e.g., other workers or other staff onsite with the cancer(s) under investigation) 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 among persons at a worksite, within an occupational group, or among other cohort should investigate the initial report as any other epidemiologic investigation (per IDAPA 16.02.10, "extraordinary occurrence of illness.") If the public health district staff feels that state epidemiology assistance is needed, they should contact the state Bureau of Environmental Health and Communicable Disease at 208-334-5939.

2. Once initial information is gathered, the district (and state epidemiology teams, if needed) will develop an initial case definition. It may be necessary to contact the inquirer to develop the case definition. The case definition must include:

    - Disease or condition believed to be in excess;
    - Location of index cases (geographic area, population, place of work);
    - Time period of concern (for diagnosis of cases); and
    - Suspected worksite, occupational, or cohort environmental exposures and likely period of exposures, if any.

    Case verification for non-infectious diseases will be performed under the direction of the State Epidemiologist. To protect confidentiality, at no time during the case verification process, or thereafter, will individual-specific case data be shared with the inquirer, even if the inquirer initially provided the case names.

    The decision to pursue the investigation further will be made by the local public health district, and involves considering:

    i. the concern from persons at the worksite, in occupational group, or in other cohort;
    ii. presence of worksite, occupational, or cohort exposure(s) likely to cause the disease or condition; and
    iii. presence, magnitude and trend of excess observed cases at the site or within the group.

    The Division of Public Health has limited technical capacity to conduct preliminary worksite or occupational cluster investigations. A decision to pursue further investigation will also depend on the technical capacity of staff to conduct an in-depth investigation. If further investigation is warranted, resources available for local public health districts to refer to include: the National Institute for Occupational Safety and Health (NIOSH) Health Hazard Evaluation Program (http://www.cdc.gov/niosh/hhe/), and the Idaho Occupational Safety and Health Consultation Program (https://www.boisestate.edu/oshcon/). Please note that both of these evaluation services require referrals to come from employees or the employer of the worksite.

    The decision to close the investigation shall prompt a written report to be sent 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.

Phase II: Determining Feasibility of Conducting an Epidemiologic Study

The purpose of Phase II is to assess the feasibility of performing an epidemiologic study to compare the number of observed cases among the worksite, occupational group, or other cohort with expected numbers of cases from general population rates and/or examine the association between the cancer or non-infectious disease 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 worksite or cohort partners and may also include community, environmental, and other partners. This step provides the opportunity to evaluate additional public health actions, such as smoking cessation programs, cancer screenings, health risk assessments, removal of environmental hazards, or other activities that should be conducted. If beneficial to public health, these actions should not be delayed pending the decision to conduct or complete an epidemiologic study.

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 including sample size, a small case number, and study power. Experienced scientists with appropriate skills should be included in the investigative team. The investigative team may include an epidemiologist, a toxicologist, a physician, an environmental health specialist, and representative(s) from the worksite or cohort to provide advice on the assessment as needed. It is necessary to identify such parameters as the 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 or non-infectious disease meet the case definition.
- Verify whether suspected 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 the cancers of interest.
- Identify a comparison group that, depending on the study design, does not have the disease of concern. Examples include:
    - Control group in a case-control study, or does not have the disease of concern,
    - Unexposed group in a cohort study, or is otherwise representative of the population.
- 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 worked

at the worksite of interest, occupation, and other relevant risk factors and confounding variables.
- 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 worksite or other cohort 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 impact of the sample size on statistical power;
- Determine the appropriateness of the planned 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.

In situations in which the types of cancers or non-infectious condition have no known association with an environmental contaminant, in which no suspected environmental hazard exists, or in which other factors may explain the observed excess, (e.g., cases diagnosed among newly hired workers), 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 cohort, public, or media might continue to demand further investigation, regardless of cost or biologic plausibility. Working with established cohort 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 health of those at the worksite, occupational, or other cohort cluster location, depending on the information and resources available.

If the feasibility assessment suggests an epidemiologic study is warranted, further outreach, health assessment, interventions, or other public health actions also might be appropriate. Conducting epidemiologic investigations can take several years; the health agency should consider what can be done in the interim to help protect the community's health and keep its

members informed. This level of investigation often can be seen as research rather than public health response to a worksite, occupational, or cohort concern. Providing periodic progress reports to keep the cohort of concern involved can help overcome this perception.

Phase III: Epidemiologic Study

If a study is warranted, the CDRI epidemiologist will conduct a study using the preparatory effort from Phase II. 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 take the form of a case-control study, a cohort study, or cross-sectional with possible environmental sampling. The design will depend on the nature of the disease, the availability of data, and the suspected exposure pathways. The planning and implementation of such a study will be performed by members of nCAWG with leadership by the State Epidemiologist, the local public health districts, federal partners (such as the Agency for Toxic Substances and Disease registry [ATSDR]), or the CDRI epidemiologist(s), depending on the type of disease or condition of concern, and the complexity of the study. nCAWG will engage with the public health district and the inquirer to engage appropriate parties to support the investigation.

In the case of a worksite or occupation-based investigation, CDRI and the public health district will engage the worksite or professional association(s) in the selection of members to participate in the investigation; these may include a worksite/occupational group lead contact, local public health district staff, and membership from the local medical community, among others. It is important to involve local partners for many reasons, not the least of which is that an investigation may augment existing fear and uncertainty in the site brought on by the perception that a suspected disease cluster exists, which might have negative social and economic impacts.

Epidemiologic studies are dependent upon type of disease or condition, availability of data, feasibility of identifying appropriate comparison groups, ability to measure exposures and outcomes accurately as well as the availability of funds and staff for proper implementation. Multiple barriers may exist which must be examined and overcome to proceed further. These include:

- Persons identified as part of a cluster may be deceased, missing, or reside or work outside of Idaho and therefore unable to provide a detailed exposure history or disease information.
- Persons identified as part of a cluster may be unwilling to participate in a survey or health study.
- Federal assistance (e.g., Environmental Protection Agency, ATSDR, Centers for Disease Control and Prevention [CDC]) in terms of technical expertise, funding, and laboratory testing, is time-intensive to procure, or may not be available.
- Privacy and confidentiality of all persons in the possibly affected worksite or occupational cluster must be respected. This includes contract workers, former employees, and other

persons with substantial exposure to the site but who may not fit the definition of, e.g., volunteers.
- Collection and testing of environmental samples may cause significant delays.

Demonstrating a statistically significant association does not prove causation; indeed, causation is frequently impossible to determine, even when clinical and laboratory study information is available. Even if a disease cluster is identified and environmental contamination is identified, an investigation might not demonstrate a conclusive association between the two. Other risk factors (e.g., smoking, personal behavior, and genetic traits) should also be explored. Conversely, even if the investigation does not identify an association, the exposure still might be linked to the cluster; however, in such a case more scientific information might be required to establish an association (e.g., toxicological and clinical data). Epidemiologic studies alone often are not able to detect small effects, particularly in small populations or when the number of cases is limited.

1. Verification of "Index" Cases

   In many instances, reported diagnoses are not supported by medical records. For example, cases reported as a particular type of cancer may be found to be several different types of cancer or not cancer at all. There are also occasions where reported cases occurred in persons who developed the illness prior to entering the cohort (e.g., prior to employment)) and therefore should not be included in the analyses. For all these reasons, it is necessary to verify cancer cases and cases of non-infectious diseases using other sources of data, including worksite employee rosters, volunteer rosters, physician records, hospital records, and vital records. The availability of information about specific health problems can be limited because of confidentiality and access to such records.

   For cancer, case verification will be performed by CDRI.[†] For non-infectious diseases other than cancer, case verification will be performed under the direction of the local public health district when a single jurisdiction is involved performed, and with guidance and direction of the State Epidemiologist. Full case ascertainment means finding all cases of the disease in question which occurred in the location or occupational cohort during the period of interest and meet the case definition, not only those reported by an inquirer. To achieve this for cancer, a list of cohort members (e.g., current and former worksite employees, volunteers, contract workers), should be compared to the CDRI database to identify all cancer cases diagnosed while cohort members were Idaho residents. Cooperation with other states may be required to identify cancers diagnosed among non-state residents who are members of the cohort under study (e.g., persons who live in border communities and work across state lines or persons who move out of state after employment and are diagnosed with cancer in that state). Performing this comparison will allow for the confirmation of cancer primary site, date of diagnosis, and other information.

[†] Because CDRI was not population-based until 1971, cancer cases and person-time at risk should not be counted prior to 1971 in the resulting analytic dataset. This may result in the exclusion of cancer cases in the original employee or occupational cohort dataset, since they have not been ascertained as true cancer cases as per the registry. To compile person-time, the clock should start at either their initial or latest date of hire or licensure (censored to 1971).

It also allows for statistical assessment of the existence of a cancer cluster. To conduct the linkage between cohort members and the CDRI database, CDRI would request the following fields: first name, middle name, last name, date of birth (at the least, month and year), sex, home and/or employment address, and Social Security number. For employee cohorts, start and end dates of employment would be needed. For occupational rather than employee cohorts, license type and status (i.e., active, inactive, or historical), dates of license original issue and expiration, and latest year of license recertification are necessary.

2. Comparison of Observed and Expected Numbers of Cases

i. Selection of Comparison Group

As a first step, selection of the exact area of concern (e.g., one or several buildings in a worksite, where an occupational cohort practices, neighborhood), and selection of the area with which it is to be compared must be determined. Once the area of concern is defined, the comparison/reference population will be selected. The comparison area may be the state of Idaho, the county in which the worksite is located, or the national cancer rates in the registries comprising the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program or CDC's National Program of Cancer Registries. Selecting additional characteristics within the comparison group, such as high or low socioeconomic status, may offer more comparable reference rates of disease or health condition with which to compare the observed cases. The expected number of cases will generally be calculated based upon age-specific and sex-specific rates for the state of Idaho during the same period as the reported cluster. These rates will be multiplied by person time in the cohort to calculate the number of expected cases.

ii. Data Management

Certain initial data cleaning steps may be necessary to maximize the utility of the dataset from the worksite, occupational group, or cohort of interest—in particular with respect to matching sensitivity and specificity with the CDRI population-based cancer registry.

- Match*Pro, a probabilistic record linkage program developed at the National Cancer Institute in support of the SEER Program (https://seer.cancer.gov/tools/matchpro/), may be used de-duplicate records in the group of interest.
- LexisNexis Accurint, an online service (https://secure.accurint.com/app/bps/main), may be used to populate missing data, e.g., birth date, or residency changes.
- TransUnion, an online service (https://www.transunion.com/), may be used to populate missing data, e.g., birth date, or residency changes.
- In instances where a registry exists for the disease in question, investigators may want to engage the registry for additional information about cases. This includes registries with self-reported information.

- Sex is important when calculating observed to expected ratios, and if sex is missing or completely unavailable in the data from the group of interest, a file of name frequencies by sex from the National Center for Health Statistics can be used to impute values for sex. If a name is commonly (i.e., 45%–65% of the time) used for both males and females, further research should be conducted to assign the proper sex to the individual.
- Idaho Health Data Exchange may be used to identify or confirm cases in the absence of records in cancer registry or other data. This becomes particularly important when concerned citizens report cancer concerns that are more recent than our current cancer data or for clusters of conditions that are not specifically reportable in Idaho..
- Social Security Death Index[‡] may be used to help ascertain vital status among cohort members and date of death (https://www.ssa.gov/dataexchange/request_dmf.html).

iii. Linkage

Linkage between the worksite, occupational group, or cohort of interest and the CDRI population-based cancer registry can subsequently be performed using Match*Pro, where probabilistic record linkage scores are computed based on the theoretical framework developed by Fellegi and Sunter.[3] Investigators should decide which cancer cases will be included in the linkage results as part of their investigative protocol (e.g., all cancers linked to a particular person, only those during a particular time period, only a first primary).

A carefully considered linkage approach will maximize sensitivity and specificity, and minimize processing time. In instances where data sets are large, comparison of all possible matches may be impractically computationally intensive, in which case "blocking," or only submitting pairs that meet certain basic criteria to a full match, is an important strategy. Records are "blocked" on a subset of fields (e.g., birth date, social security number, and Soundex functions[§] of last and first name). Candidate matching pairs are identified from instances where at least N number of blocking fields match; match scores will only be computed between these potential matches. Careful manual review of matching results should be made prior to finalizing an analytic dataset.

iv. Person-time Contribution

Person-time is the sum of time each individual in the cohort is at risk during the study period. Vital status or date of last contact should be determined to inform the person-time contribution of each person. The Social Security Death Index may be used to help ascertain vital status among cohort members and date of death, informing right-hand censoring (death, loss to follow-up, end of study). Additionally, for cancer clusters, the end

[‡] Social Security Death Index Master File information: https://www.ssa.gov/dataexchange/request_dmf.html; https://dmf.ntis.gov/
[§] Soundex Indexing System information: http://www.archives.gov/research/census/soundex.html

of the study period should be the latest date for which CDRI can be reasonably sure that all cases in the cohort have been reported to CDRI.

Disease latency needs to be considered in the analyses, in particular when investigations include cancer, many types of which have long latency periods (https://www.cdc.gov/wtc/pdfs/policies/WTCHP-Minimum-Cancer-Latency-PP-01062015-508.pdf). If a latency period is assumed, cancer cases and person-time at risk will be accrued starting at a pre-determined time point equal to the latency period after the start of the study period (usually the date of hire at the worksite or licensure for the occupation of interest). Latency periods assume that no case diagnosed within the latency period could be related to a site hazard. If a person ended employment prior to the end of the study period, cancer cases and person-time at risk should still be accrued until the end of the study, given that the person still had the potential exposure of interest. A mock-up of case and person-time contributions is shown in Figure 1.

Figure 1. Examples of Case and Person-Time Contributions to a Worksite or Occupational Cohort Cancer Cluster Investigation.

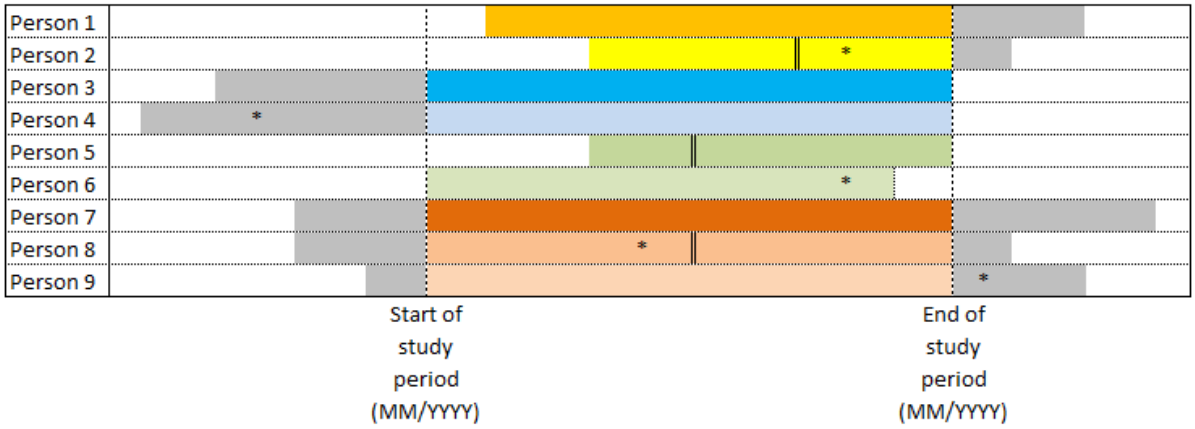

Figure 1. Mock-up data are for illustration purposes only and not sourced from any analytic dataset. Asterisk (*) indicates cancer diagnosis. Grey bars indicate no contribution to person-time. Persons 2, 6 and 8 would each contribute cases to the investigation, as denoted by asterisks. Persons 1, 3, 4, 5, 7 and 9 would contribute person-time, but not cases (no cancer diagnosis). Aside from person 6, who died shortly before the end of the study period, each person would contribute person-time up until the end of the study period, even if they left employment sometime between the start and end of the study period (i.e., persons 2, 5 and 8; denoted by double vertical lines). Person 9 would not contribute to the case count because the cancer diagnosis occurred after the end of the study.

v. Analyses

Many types of statistical software can be used for data management and analyses that are required to investigate occupational, worksite, or other cohorts. SAS (Statistical Analysis System) is one example of a software suite developed by SAS Institute (Cary, NC) for data management and advanced analytics (https://www.sas.com/en_us/home.html).

IHA recommends conducting analyses for these types of cancer incidence investigations in SEER*Stat software (https://seer.cancer.gov/seerstat/) using a custom database.

Standardized incidence ratios (SIR) and p-values will be calculated for tests of observed versus expected numbers of cases using the SEER*Stat MP-SIR session. Results will typically be tabulated using CDRI annual report primary site categories (https://idcancer.org/statistical-data/annual-reports.php). For non-cancer cluster investigations using mortality data, IHA has developed SAS code to accrue person time by achieved age to compare to age-specific mortality referent rates. Reported clusters will, upon investigation, fall into three categories:

1. ***No excess.*** This occurs when the observed number of cases for a worksite, occupational group, or other cohort is less than or equal to the expected number of cases, based on general population reference rates. 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-value ≥ 0.05 or 95% confidence interval [CI] crosses 1.0).

2. ***Explained excess.*** Based upon the experience in many jurisdictions, concerns regarding non-infectious disease clusters arise because the public is not aware of how common conditions such as cancer, spontaneous abortion, and birth defects are. For example, an excess of lung cancer in a worksite with a high percentage of smokers and no unusual environmental exposure is not likely to constitute a cluster caused by an environmental toxin. Clear and thoughtful communication with concerns inquirers may alleviate concerns about worksite- or occupation-related risks. This is also an opportunity to provide more actionable and impactful information to inquirers (e.g., awareness of smoking cessation programs, lung cancer screening for eligible populations).[**]

3. ***Unexplained excess.*** In some instances, an 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 that is entirely above 1.0), indicating that the concern warrants further investigation.

An important consideration is the issue of practical versus statistical significance. If observed and expected case counts are large enough, minor differences are more easily detected and may be statistically significant. However, this difference may be of little practical or clinical significance (e.g., 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 2014 to 2 in 2015, 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.

[**] U.S. Preventive Services Task Force Lung Cancer Screening Recommendations: https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/lung-cancer-screening

Another important consideration to consider is confounding. When designing the study, investigators must proactively identify and control for confounding variables (factors associated with both exposure and the outcome). Common confounders include age, sex, socioeconomic status, smoking status, occupational history, residential history and other environmental exposures.

To address the problem of rates based on small numbers, all communications containing rates or percentages should include a caution about interpretation. 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, rate and cell suppression rules will be invoked to align with the CDRI Data Release Policy.[††] Aspects of the Data Release Policy may be disregarded with nCAWG approval, such as in circumstances when rare diseases and conditions generate small numbers of cases that may be important for public health.

3. Written Report

Written results will be sent to the inquirer, other interested parties as discussed above, and the lead contact for the worksite, occupational group, or other cohort. For more information about communications of cancer clusters to the public, please visit https://www.cdc.gov/cancer-environment/php/guidelines/responding-community-concerns.html.

For cancer cluster investigations, the CDRI epidemiologist or designated Division of Public Health epidemiologist will write the report and include, as an attachment, a Cancer Cluster Fact Sheet. For other non-infectious diseases, the local public health district epidemiologist, state-level epidemiologist, or other principal investigator will be responsible for writing the report. A draft will be submitted to nCAWG members and, depending on the requirements of the investigation as determined by nCAWG, the Division of Public Health Administrator. A period of two weeks (10 working days) will be allowed for review. After comments are incorporated into the report, a final copy will be submitted to the inquirer, all nCAWG members, and the local public health district(s). If the results of the inquiry show no excess, the cluster investigation is considered closed unless continuing concern is high; in this case, nCAWG will discuss how to best address continuing concerns.

[††] CDRI Data Release Policy: https://idcancer.org/pdfs/data-users/CDRI_Data_Release_Policy.pdf

References

1. Fischman ML, Rugo HS. Occupational cancer. In: LaDou J, Harrison RJ, eds. *CURRENT Diagnosis & Treatment: Occupational & Environmental Medicine, 5e*. McGraw-Hill Education; 2013.
2. Blair A, Marrett L, Beane Freeman L. Occupational cancer in developed countries. *Environ Health*. Apr 5 2011;10 Suppl 1(Suppl 1):S9. doi:10.1186/1476-069X-10-S1-S9
3. Fellegi IP, Sunter AB. A Theory for Record Linkage. *Journal of the American Statistical Association*. 1969;64(328):1183-1210. doi:10.1080/01621459.1969.10501049

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 ☐ 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 exposure or etiologic agent? If yes, what do they suspect and why?

What is the time period during which people became ill or experienced the health condition?
What types of illnesses, health conditions, main symptoms are being reported?
How many people (list ages, if known) are reported with illness or health condition?

If concern involves cancer, fax this form to the Cancer Data Registry of Idaho at 208-338-7800 or securely email it to cjohnson@teamiha.org and bmorawski@teamiha.org. Do not use unencrypted email such as Outlook or Gmail to transmit this form.

Appendix B: Talking Points on Clusters

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. 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 periods between the 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. We may have seen an association in other places where exposures are higher, if the relationship were strong.

- For diseases of unknown etiology or origin, 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 multiply without control and can invade nearby tissues. Cancer is very common: according to the American Cancer Society about 1 in 2 men and 1 in 3 women will be diagnosed with cancer sometime in their life. About 1 in 5 deaths in the US is attributable to some form of cancer. Cancer is the second leading cause of death in Idaho and the US.

- 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 cancer 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.

Useful Websites for the Consumer

National Cancer Institute, Cancer Clusters. Cancer.gov. https://www.cancer.gov/about-cancer/causes-prevention/risk/substances/cancer-clusters-fact-sheet

Centers for Disease Control and Prevention. Unusual Patterns of Cancer, the Environment, and Community Concerns: About Unusual Patterns of Cancer. CDC.gov. https://www.cdc.gov/cancer-environment/about/