Protocol for Investigating Cancer Cluster Concerns in Utah

May 2, 2016

Environmental Epidemiology Program
Bureau of Epidemiology
Division of Disease Control and Prevention
Utah Department of Health

Mailing: PO Box 142104
Salt Lake City, Utah 84114-2104

Delivery: 288 North 1460 West
Salt Lake City, Utah 84116

Email: eep@utah.gov

Phone: (801) 538-6191
Fax: (801) 538-6564

This Report prepared by the Cancer Cluster Workgroup:

Chairpersons

Kim Herget, MStat, Utah Cancer Registry

Jenny Robertson, MSPH, Salt Lake County Health Department

Secretary

Matt McCord, MSPH, Environmental Epidemiology Program

Workgroup Members

Allyn Nakashima, MD, State Epidemiologist

Greg Williams, MPH, CPM, Environmental Epidemiology Program

Nathan LaCross, PhD, MPH, Environmental Epidemiology Program

Lynette Philips, MPA, Utah Cancer Control Program

Titus Oyeladun, MPH, Utah Cancer Control Program

Heather Sarin, MPA, MBA, Utah Cancer Control Program

Michael Friedrichs, MStat, Bureau of Health Promotion

Neng Wan, PhD, University of Utah Department of Geography

Table of Contents

Maintenance, Review, and Clearance Process:

1. Review Cycle: This protocol should be reviewed and updated every 3 years or when major edits or corrections are needed.
2. Maintenance: Minor corrections that do not change definitions, roles and responsibilities, scope, or methods implemented by this protocol (such as adding current references or making corrections to the document style) do not require review and clearance until the next review cycle. Major edits (i.e., implementation of changes in definitions, roles and responsibilities, scope, or methodology) require that the protocol go through the review and clearance process before those changes can be implemented.
3. Required Reviews: Prior to clearance, the following agencies will have an opportunity to review the protocol and provide comment.
    a. Utah Cancer Registry
    b. Utah Cancer Control Program
    c. Utah's 13 local health departments
    d. Utah Department of Health Public Information Officer (PIO)
4. Comments: Electronic comments (i.e., MS Word tracked changes, Google Docs) are preferred, but reviewers are allowed to return comments in any convenient form. The Environmental Epidemiology Program manager will collect comments, which will be reviewed by the workgroup.
5. Clearance of the protocol requires certification by the Director of the Bureau of Epidemiology, the State Epidemiologist, and the Environmental Epidemiology Science Advisor. This protocol will contain the current certification.

CERTIFICATION

This protocol titled "Protocol for Investigation Cancer Cluster Concerns in Utah" was prepared by the Cancer Cluster Workgroup for the Environmental Epidemiology Program, Utah Department of Health. Editorial and technical review was completed by stakeholder and clearance partners. This protocol is effective on certification.

Approved by:

Cristie Chesler
Director, Bureau of Epidemiology
Utah Department of Health

Allyn K Nakashima, MD
State Epidemiologist
Utah Department of Health

Wu Xu, PhD
Director, Center for Health Data
Utah Department of Health

Scope

Responding to and investigating non-infectious disease cluster concerns in Utah is the responsibility of the local health departments (LHDs), supported by the Environmental Epidemiology Program (EEP) within the Utah Department of Health (UDOH) (Langeberg et al. 2004; Stanbury et al. 2012; Thacker 2000; Thacker 2012). Throughout this protocol, the terms disease and cluster will be used to refer to non-infectious diseases and clusters often associated with environmental exposure concerns. Cancer clusters are the type of cluster concerns most frequently addressed by the EEP. While this protocol will focus on cancer clusters, most of the guiding principles and methodologies can be applied to other kinds of non-infectious disease cluster concerns communicated to the EEP.

The EEP, in conjunction with the Utah Cancer Registry and the Salt Lake County Health Department, and supported by other UDOH programs and LHD representatives, agreed to develop this protocol to standardize the process to be used by the EEP for responding to and investigating cluster inquiries and concerns in Utah.

The EEP responds to or investigates disease cluster inquiries or concerns for the following reasons (Bell et al. 2006; Dicker 2002; Kingsley et al. 2007; Lawson & Kulldorff 1999; Stanbury et al. 2012; Thacker 2000; Thacker et al. 2012):

- to educate the public about risk factors that may contribute to the occurrence of disease;
- to respond to the concerns of citizens;
- to help the public become informed about and focus on the environmental problems in their communities which may underlie their concern about disease;
- to further understanding of the etiologies of disease;
- to empower communities to make improvements in governmental policymaking and health care services; and
- to respond to disease risk concerns associated with other environmental or public health investigations, actions, or responses.

Protocol Goals

The goals of this protocol are to:

- outline the process used by the UDOH to respond to a cluster-related inquiry from the public;
- describe communication and coordination between the LHD(s) and the UDOH;
- describe the record keeping process; and
- provide useful information that can be incorporated as part of an investigation.

Guiding Principles

The following topics are guiding statements of facts and principles considered in developing this protocol and should be used when considering education and responses to public inquiries about perceived clusters.

1. ***Cancer***
    a. Cancer biology: Cancer is a broad group of more than 100 diseases that involve uncontrollable cell replication and growth. Often these cells are "undifferentiated," meaning they have lost their tissue-specific characteristics. As these cells grow to form tumor tissue, they invade nearby healthy tissue or spread through metastasis to other tissues. This invasion or spread disrupts the functions of the affected healthy tissues. Cancer cells may also produce metabolic products that can be transported to other parts of the body resulting in adverse health effects. Different kinds of cancer have different physiological characteristics, causal risk factors, prognoses, and treatment (ACS 2013; Goodman & Samet 2006; Kings & Robins 2006; NCI 2012a; NCI 2012b; Stein & Colditz 2004; Weinberg 2006).

    b. Relevance in Public Health Priority: The American Cancer Society (ACS) estimates that about one in two men and one in three women will develop cancer (all invasive sites) sometime in their life (lifetime risk) (ACS 2009; NCI 2011a, 2011b). In the U.S., cancer is the second leading cause of death (CDC 2012). Among all causes of death, approximately one in four men and one in five women will die of cancer (ACS 2009; NCI 2011a, 2011b). On average, about one in nine people will develop two or more cancers in his or her lifetime (Wilkins & Woodgate 2008).

    c. Public perceptions: Cancer is a dominating environmental public health concern. Public fear of cancer resulting from environmental hazards is reinforced by U.S. environmental regulatory actions that use cancer as a mechanism for making regulatory decisions (CDC 2012; Goujon-Bellec et al. 2011; Morrone 2011; Warburton 1999; Wakefield et a. 2000).

    d. Cancer risks: Risk factors that contribute to the development of cancer include both inherent and external factors. Inherent factors include a variety of genetic susceptibilities. External factors include life choices and behaviors (e.g., tobacco use, alcohol use, poor diet, obesity, lack of physical activity, excessive sunlight exposure, and sexual behavior), medical conditions and medications, oncogenic pathogens, and chemical or radiological environmental exposures. Cancer may be the result of several factors interacting together with a triggering event (ACS 2013; Goodman & Samet 2006; NCI 2012b).

2. ***Public health responsibilities***: A core function of epidemiology is to track and evaluate disease patterns. This duty helps public health officials and policy-makers identify and assess communities with public health challenges, define public health priorities, monitor and evaluate public health actions, and discover knowledge about public health concerns (Dicker 2002; Stanbury et al. 2012; Thacker 2000; Thacker et al. 2012).

Definition of Investigative Methods

1. ***Statistical Review***: A retrospective ecologic study based on registry and census data. These are also called a small area analysis, a disease cluster analysis, or a pre-epidemiologic study cluster (dos Santos Silva 1999; Esteve et al. 1994; Jekel et al. 1996; Kingsley et al. 2007; Mann 2003). The main objective of a statistical review is to verify that a population has an ongoing or emerging cluster of disease and to characterize the magnitude and temporal activity of the cluster. A statistical review is the typical outcome of Phase II in the methodology described in this protocol.

2. ***Statewide Statistical Scan***: An ecologic study based on registry and census data. The main objective of a statewide statistical scan is to identify previously undefined areas with higher than expected disease rates, or "hot spots." A statewide scan is an alternative outcome of a Phase II investigation (Kulldorff 2010; Kulldorff & IMS 2011).

3. ***Environmental and Community Investigation***: An ecologic assessment of data from the U.S. census, the American Community Survey, Utah Indicator-Based Information System for Public Health (IBIS-PH), the Utah Environmental Public Health Tracking Network (UEPHTN), U.S. Environmental Protection Agency (EPA) data systems, and/or other relevant data sources that provide information about the magnitude of known risks and exposures. The purpose of this type of investigation is to characterize and quantify known potential risk factors to which cancer is attributed for the community. An environmental and community investigation can be conducted concurrent with a statistical review. This method is useful in determining the feasibility of conducting an epidemiologic investigation (Abrams et al. 2014; Bender et al. 1990; CDC 1990).

4. ***Epidemiologic Investigation***: A study that involves a population of case and control persons (Abrams et al. 2014; Bender et al. 1990; CDC 1990). These may also be called risk assessment studies, case-control studies, or case-cohort studies (which is a variant of the case-control design). The main objective of the epidemiologic investigation is to quantify and apportion risk to various factors, including behaviors and exposures. An epidemiologic investigation report is the outcome of Phase IV in the methodology described in this protocol.

Definition and Characteristics of a Cluster

A cluster is the occurrence of a greater than expected number, real or perceived, of cases of one type of a disease, or a group of diseases related by the same causal factors, occurring within a group of people in a defined geographic area over a defined period of time, and reported to a health agency (Aamodt et al 2006; Abrams et al. 2013; CDC 1990; Hinrichsen et al. 2009; Kingsley et al. 2007; Lawson & Kulldorff 1999; Wakefield et al. 2000; Wheeler 2007). Perceived clusters are usually reported when individuals learn that friends, family members, neighbors, or coworkers have been diagnosed with cancer. It is important to determine whether perceived clusters are truly the occurrence of more cases than would be expected. Moreover, cluster reports may help health professionals in identifying public needs regarding public health knowledge and identifying community social and economic disparities that lead to increased concern. Responding to a cluster concern requires not only an understanding of epidemiology, but also an appreciation for the public's concern and understanding of the principles of effective risk communication.

For the purposes of this protocol, a report of a cluster is considered suspect until it is statistically confirmed. A reported cluster may be confirmed by the results of the statistical review.

Process for Responding to Cluster Inquiries

The EEP uses the four levels of response as outlined by CDC (Abrams et al. 2014; Bender et al. 1990; CDC 1990). The four levels are 1) initial contact, 2) a statistical review, 3) feasibility investigation, and 4) an epidemiologic investigation. All responses start at the first level and progress sequentially through levels until the concern has been resolved.

PHASE	***METHODS***
1. Initial contact	• Gather data about concern • Educate • Connect with local health department
2. Statistical Review	• Defined population statistical review • Statewide scan for hotspots • Environmental and community investigation (available data)
3. Feasibility	• Environmental and community investigation (community needs) • Community consultation • Subject matter expert consultation
4. Epidemiologic Investigation	• Data collection • Exposure assessment • Risk assessment

1. ***Initial Contact***:

    a. PURPOSE: During the initial contact, the goals of EEP are to:
        i. Collect information about the cluster concern.
        ii. Provide information about disease cluster risks.

    b. PROCESS:
        i. Collect the following information about the cluster reporter: name, phone number, email address, mailing address, and affiliation (concerned citizen, physician, employer representative, etc.)
        ii. Collect information on the perceived cluster:
            1. What are the cancer types being observed and how many of each type?
            2. What are the ages or age ranges of the cancer cases?
            3. What is the time period that these cancers are occurring?
            4. What is the geographic area where these cancers are occurring?
        iii. Collect information on what the reporter suspects are environmental exposures or causes (if any).
        iv. Based on the information provided, provide the reporter with information about cancer (e.g., risk factors, the frequency of cancer, etc.).
        v. Provide the reporter with links to the EEP cancer website for additional information about cancer.
        vi. Describe the process for initiating additional investigation through the LHD and provide a contact for the LHD.
        vii. Report to the LHD the contact information for the reporter and the nature of the conversation.

    c. INITIATING AN INVESTIGATION:
        i. If the reporter is a concerned citizen, physician, employer representative, or a local government agency, the EEP will:
            1. Respond to questions at the time of contact as outlined above.
            2. Refer the reporter to the LHD.
            3. Notify the LHD of the contact.
            4. Initiate further investigation on request from the LHD.
            5. Report all findings to the LHD.
            6. Support the LHD in communicating the findings to the reporter.
            7. With the approval of the LHD, post reports of investigations to the EEP cancer website.

ii. If the reporter is a state official (acting in the capacity of a state official) or a federal official (acting in the capacity of a federal official), the EEP will:
    1. Respond to questions at the time of contact as outlined above.
    2. Initiate an investigation.
    3. Notify the LHD that an investigation has been initiated.
    4. Collaborate and coordinate with the LHD when appropriate.
    5. Report to the requesting official or agency.
    6. Provide a copy of the report to the LHD.
    7. Support the LHD or requesting agency in communicating the findings to the affected community.
    8. Post the investigation report to the EEP cancer website.

2. ***Defined Population Statistical Review***:

a. PURPOSE: The purposes of a statistical review are to:
    i. Confirm or validate an observation that cancer rates in a specified population are higher than expected.
    ii. Quantify the magnitude of cancer rates in a specified population.
    iii. Assess the temporal trend in cancer rates in a specified population.

b. PROCESS:
    i. The EEP will work with the LHD(s) with jurisdiction of the study population to:
        1. Define the geographic boundaries of the study population. For linkage purposes, the geography should be defined as an aggregation of census block groups or census tracts, using the 2010 U.S. census geographic identification labels. The population should include at least 20,000 persons at any point of time.

        2. Define which cancer site categories will be reviewed. The Utah Cancer Registry (UCR) provides cancer site category codes for anatomically-based cancer sites as follows:

1) Oral cavity and pharynx	2) Esophagus	3) Stomach
4) Small intestines	5) Colon	6) Rectum and recto-sigmoid junction
7) Anus, anal cancer, and anorectum	8) Liver and interhepatic bile duct	9) Gallbladder and biliary ducts
10) Pancreas	11) Other digestive system	12) Larynx
13) Lung and bronchus	14) Other respiratory system	15) Bones and joints
16) Soft tissue (including heart)	17) Cutaneous melanoma	18) Other non-melanoma skin
19) Breast	20) Cervix	21) Uterus

22) Ovary	23) Other female genital	24) Prostate
25) Testis	26) Other male genital	27) Bladder
28) Kidney and renal pelvis	29) Other urinary	30) Eye and orbit
31) Brain	32) Other central nervous system	33) Thyroid
34) Other endocrine	35) Hodgkin lymphoma	36) Non-Hodgkin lymphoma
37) Multiple myeloma	38) Lymphocytic leukemia	39) Myeloid leukemia
40) Monocytic leukemia	41) Other leukemia	42) Other sites or types not listed above

3. The request can be for some or all of these types.
4. Define the temporal study period and analytical periods to be used. Typically the study period should be from 1980 to the latest available data. Starting at 1980 is dictated by the availability of population data apportioned by the 2010 U.S. Census boundaries. The latest available data typically as a 2-3 year lag behind the current year. This lag is because the UCR is unable to release case data until it is fully vetted. The UCR releases data to the EEP annually. The analytical periods should be in aggregates of 1 to 5 years. Typically each analytical period should have at least 100,000 person-years.
5. Define statistical significance and actionable findings (Bender et al. 1990; Caldwell 1990; Langeberg et al. 2004; Thun & Sinks 2004).

ii. Statistical significance is determined by the 99% confidence limits. The 99% confidence limits are used because of the multiple analyses conducted (Anderson et al. 2012). Additional criteria may include screening for magnitude and trends, such as:

1. Any last analytical period where the risk ratio is three or more standard errors above 1.0.
2. A sequence of two or more analytical periods that includes the last analytical period where the risk ratios of each analytical period are at least two standard errors above 1.0.
3. A sequence of four or more analytical periods that includes the last analytical period where the risk ratios of each period are at least one standard error above 1.0.
4. A persistent trend of elevated risk ratios in all analytical periods, with at least the last analytical period having a statistically significant elevated risk ratio.

5. Situations where some of these criteria are met, but not the last period criteria, are considered historic clusters that have resolved, and are not actionable.

iii. For actionable findings, EEP should work with the LHD and community to identify possible actions each entity can take based on the findings. Possible actions may include:
   1. Providing community education about cancer risks and risk avoidance.
   2. Identifying and connecting community members to health care resources available for screening and early intervention.
   3. A report empowering the community to take political action.
   4. Conducting an epidemiologic investigation.

iv. Confidentiality and suppression of data: During analysis, any analytical unit (cancer site group x analytical period) that has three or fewer cases will be suppressed. Suppression will include analytical units with zero cases (Langeberg et al. 2004).

v. Cumulative study period analysis (i.e., 1980 through 2014) or the aggregation of all cancer types will not be done.

vi. The EEP may coordinate with the LHD to present a study plan to the LHD board of health and/or to the study community. The purpose of this presentation is to:
   1. Describe the study plan and seek input on decisions about definitions of the geography, time elements, case definitions, and definitions of significance and actionable clusters.
   2. Describe the outcomes, challenges, and limitations of a statistical review.
   3. Coordinate community involvement in the study.

vii. Statistical methods:
   1. The total Utah population (excluding the study area population) will be used as the comparison population. Alternatively, in coordination with the LHD, a subset of the Utah population (i.e., other areas of similar composition and size to the study area, or the county or LHD jurisdictional area) may be used. The national population or the population of another state or set of states will not be used for the following reasons:
      a. Utah typically has lower cancer rates than the nation, and the national rates will dominate the investigation resulting in not finding significant clusters.

   b. National and state population information is difficult to obtain at the detail necessary for the statistical review.

2. Records of cancer cases that are georeferenced to another state or are reported from the U.S. Veteran's Administration Hospital will be excluded.

3. Records of myeloid leukemia (site code 39) that are not the first cancer (sequence is 0 or 1) will be excluded because of the potential of being therapy-induced leukemia (Godley & Larson 2008; Leone et al. 1999, 2011; Sill et al. 2011; Wilkins & Woodgate 2008).

4. The UCR sometimes provides the EEP with records that are still in the vetting process so that the EEP can geocode and georeferenced those records. Only records approved by the UCR for analytical purposes will be used. The UCR provides a status flag (IBIS = 1) for approved records.

5. Cases and population will be stratified into the following age-groups:
   a. 0 to 19 years
   b. 20 to 34 years
   c. 35 to 49 years
   d. 50 to 64 years
   e. 65 to 74 years
   f. 75 years and older

6. For non-sex-specific cancers, rates will be calculated for each sex individually and for both sexes combined (i.e., male, female, and both). For sex-specific cancers, rate will be calculated for the appropriate sex only.

7. Cancer rates for each cancer site group by each sex group for each analytical period will be calculated as raw ('crude') rates per 100,000 person-years.

8. Cancer rate ratios for each cancer site by each sex group for each analytical period will be calculated using the indirect age-adjusted rate method (Anderson & Rosenberg 1998; dos Santos Silva 1999; Esteve et al. 1994; Jekel et al. 1996; Selvin 1996).

9. Confidence limits for rate ratios will be calculated using the Byers method (Breslow & Day 1987; Rothman & Boice 1979, 1982; Sahai and Khurshid 1983, 1996).

10. The rate ratio standard error will be calculated (Freund and Williams 1966).

11. Temporal trends for the rates will be calculated using the Kendal-Tau (preferably B) method (Kendall 1938).
12. For quality control purposes, the following comparisons should be reviewed:
    a. The number of potentially misassigned cases can be identified from records that are in the ZIP codes that include the study area, but are not georeferenced to an appropriate level of census geography. If more than 5% of the cancer cases are potentially misassigned, the data should be reviewed to see if additional records can be geocoded or georeferenced.
    b. The relative frequencies of cancer cases by type (by UCR site code), age group, and sex group between the study population and the comparison population. There should be at least a 90% level of correlation.
    c. The relative population magnitude by time, age group, and sex group between the study population and the comparison population. There should be at least a 90% level of correlation.
13. Results of the statistical review for each investigated cancer site group will be categorized into one of the following categories for reporting purposes.
    a. No excess: No statistically significant or actionable findings were found for any of the analytical periods.
    b. Historic excess: One or more analytical periods, but not the last analytical period, had a statistically significant and actionable finding.
    c. Explained excess: The findings for the last analytical period were statistically significant and actionable, there is no unusual environmental exposure potential, and the cancer type seems likely given the characteristics of the study population.
    d. Unexplained excess: The findings for the last analytical period were statistically significant and actionable, and the cancer type seems unlikely given the characteristics of the study population.

c. REPORTING:
    i. The EEP will provide a report to requesting agency. The EEP will coordinate with the requesting agency on which of two report options to be generated. If the requesting agency is not the local health department in whose jurisdiction the investigation occurred, the EEP will provide a courtesy copy of the report to the

local health department as well. The two reporting options are a data table or a comprehensive report.

ii. ***Data table***. The EEP will generate an MS Excel file (.xls or .xlsx) of results for the investigation. Requested cancer site category data will be presented as separate worksheets with the file. This typically can be generated within 5 working days of the request for the report, depending on how complete georeferencing of the current cancer data is.

1. This reporting option will not include any environmental or community investigation data.
2. The data reported will include an interpretation of results as described above.
3. Data tables do not go through a formal peer-review process.
4. Data tables will not be posted to the EEP cancer information website.

iii. ***Comprehensive report***. The EEP will prepare a comprehensive written report that describes the methodology used as part of the investigation, including the data query components of the environmental and community investigation, interpretation of the data, and a brief discussion of the known risk factors associated with each cancer site group found to have explained or unexplained excesses. This report will take up to four months to generate depending on the readiness of the data and the extent of the findings.

1. Comprehensive reports go through a formal peer-review and clearing process.
2. Some comprehensive reports, with approval of the LHD, may be posted on the EEP cancer information website.

3. ***Statewide Scan***:

a. PURPOSE: The purposes of a statewide scan are to:

i. Respond to a number of expressed concerns from the public about an excess of the same cancer type.
ii. Explore state small area cancer rates for locations with excessive cancer rates (also called “hot spots”).
iii. Respond to needs for statewide government policy making.

b. PROCESS:

i. Prior to implementing a statewide hotspot investigation, the EEP will convene an advisory group to develop an investigation plan. Typically, this advisory group

would consist of representation from the UDOH Cancer Control Program, the local health departments, and the Utah Cancer Registry.

1. The EEP will explore the case data to identify the smallest small area unit (i.e., census block groups, census tracts, ZIP codes) that 95% of the cancer cases can be referenced to, either by geocoding or by assignment from other georeferencing labels. The EEP will also determine if a scan at that geographic resolution will be meaningful.
2. A statewide scan should address only one site cancer type.
3. The EEP should be prepared to address the following questions:
    a. What are the environmental exposure risks?
    b. What are the lifestyle risks?
    c. What are the inherited risks?

ii. The spatial scan statistic using a discrete retrospective Poisson space or space-time model (i.e., SaTScan or the SpatialEpi package for R) will be used.

iii. **SaTScan** should be set up as follows (Kulldorff 2010; Kulldorff & IMS 2011; SaTScan 2015):

1. Data files:
    a. An aggregated file of case counts and population counts stratified by small area, year, and age/sex developed by linking case and population data from UEPHTN data warehouse.
    b. A population-weighted centroid table of the geographic small area units prepared from available GIS tools.

Geographic ID	Time ID	Age/Sex Covariate(s)	Case Count	Population Count

Geographic ID	X-Coordinate	Y-Coordinate

2. Setup:
    a. If more than five years of data are available, the preferred type of analysis is a retrospective space-time discrete Poisson probability model (Amin & Burns 2014; Kulldorff 1997, 2010; Wagner et al. 2013). If five or fewer years are used, then a purely spatial retrospective discrete Poisson model should be used.
    b. The maximum spatial window can be determined by using a cluster information criterion statistic (CLIC) (Han et al. 2011).

c. The maximum temporal window for space-time analyses should be set at 90% (Hsu et al. 2004; Van Meter et al. 2008).

d. No spatial or temporal adjustments are needed.

e. The data should be tested with a circular spatial window shape, but the final analysis should use an elliptic window shape with a medium non-compactness penalty (Goujon-Bellac et al. 2011. Jones & Kulldorff 2012; Kulldorff 1997, 2010; Kulldorff & Nagarwalla 1995; Kulldorff et al. 2006).

f. The Oliveira's F adjustment for borders should be enabled.

g. The GINI optimized cluster collection should be enabled.

h. The model should allow no geographic overlapping.

i. All other settings should be left at default.

3. Model evaluation: SaTScan will identify many potential clusters. For each potential cluster identified, the following should be used to determine if the cluster is reportable and actionable.
    a. The model fit p-value should 0.10 or less. This value is appropriate because of the small case counts associated with clusters (Dietz et al. 2011; Hsu et al. 2004; Park 2010; Wagner et al. 2013; Wheeler 2007).
    b. The individual small area rates and the magnitude of the variance on the rates should be calculated and analyzed for homogeneity using the local spatial autocorrelation (i.e., the LISA or Getis-Ord tests) (Chen et al. 2008).
        i. Neighborhood proximity should be defined as "Queen's Case" adjacency.
        ii. Check for general homogeneity of rates and variance within the cluster area using a global test (Cromley & McLafferty 2012; Marshall 1991; Moore & Carpenter 1999; Moran 1950; Wakefield et al. 2001; Waller & Gotway 2004).
        iii. Identified hotspot boundaries using local spatial autocorrelation of the rates should correspond with the SaTScan clusters (Anselin 1995; Cromley & McLafferty 2012; Getis & Ord 1992, 1996; Jackson et al. 2009; Jacquez & Greiling 2003; Ord & Getis 1995; Tiefelsdorf & Boots 1997; Wakefield et al. 2001; Waller & Gotway 2004). The population included in a cluster for each method should include at least 90% of the population

contained in the respective cluster found by the other method.

c. The confidence limits and risk ratio standard error can be calculated and the actionable state of the cluster can be categorized as described for the statistical review.

d. SaTScan clusters not meeting the model fit requirements, or the local spatial autocorrelation fit requirements, should not be reported.

iv. The data review components of an environmental and community investigation may be included as part of a statewide scan investigation.

v. Review: The results should be reviewed with the advisory group to develop the interpretation and recommendations.

c. REPORTING:

i. The report should describe the purpose for the investigation, background for the cancer type being evaluated (including known risk factors and epidemiology), and the scan methodology. The report should provide a map and descriptive statistics about the cluster. The report may include an interpretation of the results and recommendations as developed by the advisory group.

ii. The EEP will develop a risk communication strategy that includes a very brief synopsis of the report, 2-3 key messages, and contacts for media questions.

iii. Prior to reporting, the EEP will coordinate with the UCR, the Risk Assessment Coordination Committee (RACC), the Department of Environmental Quality (UDEQ), and all local health departments. This coordination will include providing an embargoed copy of the report and the risk communication plan.

iv. The EEP will post an electronic copy of the report to the UEPHTN public portal cancer website.

4. ***Environmental and Community Investigation***:

a. PURPOSE: The purposes of an environmental and community investigation are to:

i. Identify potential environmental exposure concerns in the community.

ii. Identify behavioral and socioeconomic concerns in the community.

iii. Explore the feasibility of conducting an epidemiologic investigation.

b. PROCESS:

i. Conduct a literature review of risks associated with those cancer types that are statistically elevated.

ii. Conduct a data review:

1. The data review will be conducted as part of the statistical review if a comprehensive report was requested.
2. Collect demographic and housing information about the study area and the comparison area(s) from the U.S. census and/or the American Community Survey. This information may include characterization of:
    a. Population size and median age of population.
    b. Percent of population that are of minority race, or of Hispanic/Latino ethnicity.
    c. Percent of population that are not native to the study area (born out of state or foreign born).
    d. Percent of population that are not U.S. Citizens or do not use English as their primary language (for access to health care concerns).
    e. Percent of the adult population who lack at least a high school education.
    f. Number of households and average household size.
    g. Median household income.
    h. Percent of population living below the poverty level.
    i. Percent of population older than 65 years living alone.
    j. Percent of adults that are employed.
    k. Percent of employed adults that are working in occupations that are at high risk for chemical exposure jobs (i.e., agriculture, construction, manufacturing, military, or transportation).
    l. Percent of occupied housing units that are owner occupied.
    m. Percent of occupied housing units that are heated with dirty fuels (coal, oil, wood, etc.).
3. Collect behavioral information about the study area (or as close an approximation as available) and the comparison population from the Utah Behavioral Risk Factors Survey (UBRFS) data on the IBIS-PH. This information may include:
    a. Prevalence of smoking and tobacco use.
    b. Prevalence of alcohol use – especially excessive alcohol use.
    c. Prevalence of obesity.
    d. Percent of population with access to health care.
    e. Percent of population who routinely use wellness checks.
4. Collect data about potential environmental exposures. This information may include:
    a. Total length of high density traffic routes (freeways, highways, etc.) standardized by area for both the study and comparison

populations. This data can be obtained from the Utah Automated Geographic Resource Center (AGRC).

b. Total agricultural land (a proxy for pesticide use) within the study area. This data can be obtained from AGRC.

c. Percent of population within 1 kilometer through 5 kilometers of a Toxic Release Inventory (TRI) site, as well as characteristics associated with those sites (lists of chemicals and release quantities). This data can be obtained from the U.S. EPA TRI website.

d. Percent of population within 1 kilometer through 5 kilometers of a Comprehensive Environmental Response, Compensation and, Liability Act (CERCLA) site, and characteristics of those sites. (CERCLA sites are also known as Superfund sites.) This data can be obtained from the U.S. EPA Envirofacts Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS) website.

5. The standard z-test can be used to compare significant differences between percentage measures of the study area and comparison populations.
6. A 2-tailed t-test (with 1 degree of freedom) is useful for quantifying the significance of differences in average or median measurements.

iii. Work with the LHD to conduct community meetings and listening sessions. These meeting may collect information about:

1. Community environmental perceptions and concerns.
2. Community understanding and expectations about completed and continuing investigations.
3. Community and political will for conducting an epidemiologic investigation.

c. REPORTING: The EEP will prepare a summary of the findings from the data queries and community meeting, along with an interpretation of those findings and recommendations based on the interpretation. This report will take up to two months to generate. The report will not go through a formal peer-review and clearing process.

5. ***Epidemiologic Investigation***:

a. PURPOSE: The purposes of an epidemiologic investigation are to:

i. Explore and better define the attribution of various risk factors associated with cancer for a particular population.
ii. Further characterize a cancer cluster.

iii. Provide individual level education about cancer risk and risk mitigation.
iv. Evaluate knowledge about public health programs and services.
v. Discover new associations to the risk for cancer.

b. PROCESS:
   i. The following are basic guidelines. The exact nature of each part will vary depending on the situation. In conducting this investigation, EEP staff will seek input and advice from the state epidemiologist and other knowledgeable partners.
   ii. Feasibility study: The EEP will collaborate with the LHD and community to conduct a feasibility study.
      1. Identify key stakeholders and a process to organize or inform stakeholders. These may include:
         a. The local health board.
         b. Community or neighborhood leaders or councils.
         c. If biomonitoring will be conducted, the laboratory that will be conducting the analysis will be a necessary part of the stakeholder group.
         d. UDEQ and other involved state agencies.
         e. Involved federal agencies.
         f. Local health care organizations.
         g. Local academic partners.
         h. Local advocacy groups.
         i. Local business or industry leaders.
      2. Utilize or complete an Environmental and Community Investigation to identify possible risks that need further investigation.
      3. Identify a specific environmental concern or set of concerns that have the potential to increase cancer risk. Open-ended inquiries in the absence of suspected environmental concerns are not recommended.
      4. Verify the temporal natural biology (i.e., latency) of the cancers of concern.
      5. Verify the carcinogenicity of the environmental exposure concerns.
      6. Document the temporal history of the sources for the environmental exposures of concern.
      7. Identify and evaluate potential routes of exposure. This can be done using Agency for Toxic Substances and Disease Registry (ATSDR) exposure assessment protocols.
      8. Identify processes for estimated environmental exposures. This may be based on modeling or the ATSDR exposure assessment protocols.
      9. Identify the control group.

10. Identify resources that can be used to conduct the investigation. Resources will be needed for data gathering (e.g., risk assessment surveys) and data analysis.
11. Identify processes for gathering information about deceased or move-out cases.
12. For current and continuing environmental exposure concerns, identify and characterize potential biomarkers, including the specificity and sensitivity of those biomarkers. Identify the biological media that could be used to analyze for those biomarkers. Identify laboratories with capacity to analyze for those biomarkers.
13. Identify federal or state assistance that may be available to support some or all of this investigation. Assistance may include funding or consultation support.
14. With the LHD, the EEP will make a decision on the feasibility of conducting additional investigation.

iii. Study design:

1. The EEP will work with the LHD to request federal and or state assistance.
2. The EEP will work with the LHD to develop a study plan that will include:
    a. A participant recruitment process. This may include developing a participant screening process, a statement for informed consent, a statement for participant rights, concurrent health education materials, and methods to identify and enroll participants.
    b. An exposure and risk assessment process that may include a risk assessment questionnaire and a process to collect, store, and transport biological materials for exposure assessment.
    c. A process for communicating individualized risk assessments to participants.
    d. A process for digitizing and securing data. This process will include data linkage protocols.
    e. A process for analyzing the data.
    f. A timeframe for conducting the investigation and reporting results.
    g. This type of investigation may require an Institutional Review Board (IRB). The UDOH IRB provides consultation on whether an IRB is necessary. EEP will consult with the UDOH IRB and if necessary prepare a study plan for IRB approval.
    h. A process for training the assessment team.
3. The LHD will host a stakeholder coordination meeting. At that meeting:

a. The findings of previous phases of the investigation will be reviewed.
b. The study protocol will be reviewed and explained.
c. The stakeholders will have opportunity to provide feedback.

iv. Study implementation:
1. Seek IRB approval if required.
2. EEP will conduct training for the assessment team, which may include:
a. The process for selecting and recruiting participants.
b. The process for conducting a standardized risk and exposure survey.
c. The process for collecting, storing, and transporting biological samples.
3. Coordinate with the LHD to inform community leaders and other stakeholders that the study plan is being implemented.
4. Conduct the study. The EEP will monitor the progress of the study and work with the LHD to resolve challenges that emerge.

v. Data analysis:
1. Data linkage connects personal exposure and risk assessment information to exposure modeling results, biomonitoring results, and environmental sampling results.
2. In most cases, logistic regression will be the analytical method.

c. REPORTING: The EEP will prepare a comprehensive written report that summarizes the findings of the previous phases of the investigation, describes the methodology for data collection and analysis, interprets the results, discusses the dominant risk factors, and recommends targeting those risk factors. This report will take up to eight months to generate.
i. Comprehensive reports go through a formal peer-review and clearing process. The peer-review process includes an internal review (EEP staff), followed by a partner review (which may include UCR staff, LHD staff, or UDEQ staff), and finally certification by the director of the Bureau of Epidemiology, the state epidemiologist, and the EEP science advisor. The comprehensive report will include a clearance certificate.
ii. Some comprehensive reports, with approval of the LHD, may be posted on the EEP cancer information website.

Coordination and Record Keeping

1. Peer Review: All formal or comprehensive reports will go through a peer-review process.

a. Internal Review: At least two EEP staff members.
b. Partner peer review:
    i. The requesting agency, or the advisory group, and/or all LHDs.
    ii. UDEQ through the RACC.
    iii. UCR.
    iv. Utah Cancer Control Program.
    v. The UDOH PIO office.

2. Certification: All formal or comprehensive reports will be certified by:
    a. The director of the Bureau of Epidemiology.
    b. The state epidemiologist.
    c. The EEP science advisor.

3. Investigation Records: The EEP will maintain a registry of investigations that includes the requesting agency, the study name (typically by location), the type of study (data table or report), the request date, and the completion date. The UCR may request this registry as part of their data use accountability.

4. Coordination:
    a. Local Health Departments: In addition to the direct coordination previously described in the protocol, the EEP may provide updates and a summarized discussion of recently completed or ongoing investigations at the Council of Local Environmental Health Administrators (CLEHA) quarterly meetings and the Epidemiology Affiliate Group (AFG) meetings.
    b. Utah Department of Environmental Quality: The EEP briefs the RACC of all risk assessment activities it is conducting, including investigations of cancer clusters.
    c. Other Stakeholders: The EEP will work with the RACC and PIO office to identify other stake holders (e.g., transportation, commerce, natural resources, etc). that may be impacted by the findings of the investigation. The EEP will keep those stakeholders apprised of the progress of the investigation and of the final results prior to public release of the report.
    d. Integration with the Department of State All Hazard Operations Plan: Typically cancer clusters investigations do not require activation of an all hazards operation plan. In the event that the investigation is associated with an event that activates an all hazards operation plan, the EEP will coordinate through EDO to integrate into those operations.

5. Posting to the EEP Website:
    a. The UEPHTN maintains a cancer topics section as part of the network's public portal, found at epht.health.utah.gov/epht-view/topic/Cancer.html. On the landing page for cancer topics is a link to posted cancer studies. Reports that are not associated with an

ATSDR cooperative agreement site-specific health assessment will be posted here. Typically, these will include investigations conducted at the request of local health departments.

b. Cancer assessments conducted as part of an ATSDR cooperative agreement or state-directed site-specific health assessment are typically incorporated into a larger public health assessment. Those reports are maintained on the Utah "ATSDR's Partnership to Promote Localized Efforts to Reduce Environmental Exposure" (Utah APPLETREE) website.

References

Aamodt G, Samuelsen SO, Skrondal A. 2006. A simulation study of three methods for detecting disease clusters. *International Journal of Health Geographics* 5:15.

ACS (American Cancer Society). 2009. Lifetime risk of developing or dying from cancer web site. Available: http://www.cancer.org/Cancer/CancerBasics/lifetime-probability-of-developing-or-dying-from-cancer [accessed October 1, 2013].

ACS (American Cancer Society). 2013. Learn about cancer. Available: http://www.cancer.org/cancer/index [accessed October 1, 2013].

Amin R, Burns JJ. 2014. Cluster of adolescent and young adult thyroid cancer in Florida counties. *BioMed Research International* 2014:832573

Anderson H, Blackmore C, Stanbury M, Simms E, Tai C, Tran A. 2012. A synopsis of the 2010 national assessment of state cancer cluster investigations and protocols. Atlanta, Georgia: Council of State and Territorial Epidemiologists.

Anderson RN, Rosenberg HM. 1998. Age standardization for death rates: implementation of the year 2000 standard. *National Vital Statistics Report* 47(3):1-17.

Anselin L. 1995. Local indicators of spatial autocorrelation: LISA. *Geographic Analysis* 27(2):93-115.

Bell BS, Hoskins RE, Pickle LW, Wartenberg D. 2006. Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. *International Journal of Health Geographics* 5:49.

Bender AP, Williams AN, Johnson RA, Jagger HG. 1990. Appropriate public health response to clusters: the art of being responsibly responsive. *American Journal of Epidemiology* 132(Suppl 1):S48-S52.

Breslow NE, Day NE. 1987. Rates and Rate Standardization. In: The Design and Analysis of Cohort Studies, Vol 2. IARC Scientific Publication No 82. (BreslowNE, Day NE, eds). Lyon, France: International Agency for Research on Cancer.

Caldwell GG. 1990. Twenty-two years of cancer cluster investigations at the Centers for Disease Control. *American Journal of Epidemiology* 132(Suppl 1):S43-S47.

CDC (Centers for Disease Control and Prevention). 1990. Guidelines for investigating clusters of health events. *Morbidity and Mortality Weekly Report – Recommendations and Reports* 39(RR-11):1-16.

CDC (Centers for Disease Control and Prevention). 2012. Leading causes of death. Available: http://www.cdc.gov/nchs/fastats/lcod.htm [accessed October 1, 2013].

Chen AY, Jemal A, Ward EM. 2009. Increasing incidence of differentiated thyroid cancer in the United States, 1988-2005. *Cancer* 115(16):3801-3807.

Cromley EK, McLafferty SL. 2012. 5 Analyzing spatial clustering of health events. In: GIS and public health, 2nd ed. (Cromley EK, McLafferty SL, eds.). New York, NY: The Guilford Press.

Dicker RC. 2002. A brief review of the basic principles of epidemiology. In: Field Epidemiology, 2nd Ed. (Gregg MB, ed.). New York, NY: Oxford University Press.

Dietz NA, Sherman R, MacKinnon J, Fleming L, Arheart KL, Wohler B, Lee DJ. 2011. Toward the identification of communities with increased tobacco-associated cancer burden: application of spatial modeling techniques. *Journal of Carcinogenesis* 10:22.

dos Santos Silva, I. 1999. Cancer epidemiology: principles and methods. Lyon, France: International Agency for Research on Cancer.

Dutzik T, Bauman J. 2002. Health tracking and disease clusters: the lack of data on chronic disease incidence and its impact on cluster investigations. Washington DC: US Public Interest Research Group.

Esteve J, Benhamou E, Raymond L. 1994. Statistical methods in cancer research: IV descriptive epidemiology (IARC Scientific Publication Number 128). Lyon, France: International Agency for Research on Cancer.

Freund JE, Williams FJ. 1996. Dictionary/outline of basic statistics. New York, NY: Dover Publications, Inc.

Getis A, Ord JK. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24(3): 189-206.

Getis A, Ord JK. 1996. Local spatial statistics: an overview. In: Spatial analysis: modelling in a GIS environment. (Longley P, Batty M, eds.). Cambridge, UK: GeoInformation International, 261-277.

Godley LA, Larson RA. 2008. Therapy-related myeloid leukemia. *Seminars in Oncology* 35(4):418-429.

Goodman M, Naiman JS, Goodman D, LaKind JS. 2012. Cancer clusters in the USA: what do the last twenty years of state and federal investigation tell us. *Critical Reviews in Toxicology* 42(6):474-490.

Goodman SN, Samet JM. 2006. Cause and cancer epidemiology. In: Cancer epidemiology and prevention, 3rd ed. (Schottenfeld D, Fraumeni JF, eds). New York City, NY: Oxford University Press

Greenland S. 2001. Ecologic versus individual-level sources of bias in ecologic estimates of contextual health effects. *International Journal of Epidemiology* 30(6):1343-1350.

Greenland S, Robins J. 1994. Ecologic studies – biases, misconceptions, and counter examples. *American Journal of Epidemiology* 139(8):747-760.

Han J, Feuer R, Stinchcomb D, Tatalovich Z, Lewis D, Zhu L. 2011. Optimizing maximum window size for scan statistics. North American Association of Central Cancer Registries, 2011 Annual Conference: Louisville, KY.

Hinrichsen VL, Klassen AC, Song C, Kulldorff M. 2009. Evaluation of the performance of tests for spatial randomness on prostate cancer data. *International Journal of Health Geographics* 8:41.

Hsu CE, Jacobson H, Mas FS. 2004. Evaluating the disparity of female breast cancer mortality among racial groups - a spatiotemporal analysis. *International Journal of Health Geographics* 3:4.

Jackson MC, Huang L, Luo J, Hachey M, Feuer E. 2009. Comparison of tests for spatial heterogeneity on data with global clustering patterns and outliers. *International Journal of Health Geographics* 8:55.

Jacquez GM, Greiling DA. 2003. Local clustering in breast, lung and colorectal cancer in Long Island, New York. *International Journal of Health Geographics* 2:3.

Jekel JF, Elmore JG, Katz DL. 1996. Epidemiology, biostatistics and preventive medicine. Philadelphia, PA: WB Saunders Co.

Jones SG, Kulldorff M. 2012. Influence of spatial resolution on space-time disease cluster detection. *PLOS One* 7(10):e48036.

Kendall M. A new measure of rank correlation. 1938. *Biometrika* 30(1-2):81-93.

King RJB, Robins MW. 2006. What is cancer? In: Cancer biology, 3rd Ed. (King RJB, Robins MW, eds.) San Francisco, California: Benjamin Cummings Publishing.

Kingsley BS, Schmeichel KL, Rubin CH. 2007. An update on cancer cluster activities at the Centers for Disease Control and Prevention. *Environmental Health Perspectives* 115(1):167-171.

Izquierdo JN, Schoenbach VJ. 2000. The potential and limitations of data from population-based state cancer registries. *American Journal of Public Health* 90(5):695-698.

Kulldorff M. 1997. A spatial scan statistic. Communications in Statistics: *Theory and Methods* 26(6):1481-1496.

Kulldorff M. 2010. SaTScanTM user guide for version 9.0 (also supplied with version 9.1). Information Management Services, Inc., Silver Spring, MD. For more information visit http://www.satscan.org

Kulldorff M, Huang L, Pickle L, Duczmal L. 2006. An elliptical spatial scan statistics. *Statistics in Medicine* 25(22):3929-3943.

Kulldorff M, IMS (Information Management Services, Inc.). 2011. SaTScanTM software for spatial and space time scan statistics, for MS Windows, version 9.1.1. Silver Spring, MD: Information Management Services, Inc. For more information visit http://www.satscan.org.

Kulldorff M, Nagarwalla N. 1995. Spatial disease clusters: detection and inference. *Statistics in Medicine* 14(8):799-810.

Langeberg W, Contreras J, Hatch M, Kinney G, Sukhan S, Williams G. Cancer Cluster Workgroup: 2004. Protocol for investigating cancer clusters in Utah. Salt Lake City, UT: Utah Department of Health.

Lawson AB, Kulldorff M. 1999. 7 A review of cluster detection methodology. In: Disease mapping and risk assessment for public health (Lawson A, Biggeri A, Bohning D, Lesaffre E, Viel J-F, Bertollini R, eds.). Chichester, West Sussex: John Wiley & Sons, Inc.

Leone G, Mele L, Pulsoni A, Equitani F, Pagano L. 1999. The incidence of secondary leukemia. *Haematologica* 84(10):937-945.

Mann CJ. 2003. Observation research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency Medicine Journal* 20:54-60.

Marshall RJ. 1991. A review of methods for statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society A* 154(3):421-441.

Moore DA, Carpenter TE. 1999. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiologic Reviews* 21(2):143-161.

Moran PAP. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37:17-23.

Morgenstern H. 1982. Uses of ecologic analysis in epidemiologic research. *American Journal of Public Health* 72:1336-1344.

Morgenstern H. 1995. Ecologic studies in epidemiology: concepts, principles, and methods. *Annual Reviews of Public Health* 16:16-81.

Morrone M. 2011. From cancer to diarrhea: the moving target of public concern about environmental health risks. *Environmental Health Insights* 5:87-96.

NCI (National Cancer Institute). 2011a. SEER cancer statistics review 1975-2008. Lifetime risk (percent) of being diagnosed with cancer by site and race/ethnicity: males, 17 SEER areas, 2006-2008 (Table 1.15) and females, 17 SEER areas, 2006-2008 (Table 1.16). Available: http://seer.cancer.gov/csr/1975_2008/results_merged/topic_lifetime_risk_diagnosis.pdf [accessed October 1, 2013].

NCI (National Cancer Institute). 2011b. SEER cancer statistics review 1975-2008. Lifetime risk (percent) of dying from cancer by site and race/ethnicity: males, total US, 2006-2008 (Table 1.18) and females, total US, 2006-2008 (Table 1.19). 2011b. Available: http://seer.cancer.gov/csr/1975_2008/results_merged/topic_lifetime_risk_death.pdf [accessed October 1, 2013].

NCI (National Cancer Institute). 2012a. Surveillance, Epidemiology and End Results (SEER) Program. Available: http://seer.cancer.gov [accessed November 26, 2012]. Use this index to navigate to the relevant booklet from NCI.

NCI (National Cancer Institute). 2012b. What you need to know about cancer. Available: http://www.cancer.gov/cancertopics/wyntk/cancer [accessed March 5, 2013]. Use this index to navigate to the relevant booklet from ACS.

Ord JK, Getis A. 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis* 27(4):286-306.

Park HM. 2010. Hypothesis testing and statistical power of a test. Bloomington, IN: Indiana University; Center for Statistical and Mathematical Computing.

Rockhill B. 2005. Theorizing about causes at the individual level while estimating effects at the population level: implications for prevention. *Epidemiology* 16(1):124-129.

Rothman KJ, Boice JD. 1979. Epidemiologic analysis with a programmable calculator. NIH Publication 79-1649. Washington, DC: Government Printing Office.

Rothman KJ, Boice JD. 1982. Epidemiologic analysis with a programmable calculator, New Edition. Boston, MA: Epidemiology Resources, Inc.

Sahai H, Khurshid A. 1983. Confidence intervals for the mean of a Poisson distribution: a review. *Biometrical Journal* 35:857-867.

Sahai H. Khurshid A. 1996. Statistics in Epidemiology: Methods, Techniques and Applications. Boca Raton, FL: CRC Press, Inc.

SaTScan. 2015. Software for the spatial, temporal, and space-time scan statistics for MS Windows desktop version 9.4.1. http://www.satscan.org/ [Accessed June 15, 2015].

Selvin S. 1996. Chapter 1: Measures of risk: rates and probabilities. In: Monographs in epidemiology and biostatistics, Vol 25: Statistical analysis of epidemiologic data. (Selvin S, ed.). Oxford, UK: Oxford University Press.

Sill H, Olipitz W, Zebisch A, Schulz E, Wolfler A. 2011. Therapy-related myeloid neoplasms: pathobiology and clinical characteristics. *British Journal of Pharmacology* 162(4):792-805.

Stanbury M, Anderson H, Blackmore C, Fagliano J, Heumann M, Kass D, McGeehin M. 2012. Functions of environmental epidemiology and surveillance in state health departments. *Journal of Public Health Management and Practice* 18(5):453-460.

Thacker SB. 2000. Historical development. In: Principles and practice of public health surveillance, 2nd Ed. (Teutsch SM, Churchill RE, eds.). New York, NY: Oxford University Press.

Thacker SB, Qualters JR, Lee LM. 2012. Public health surveillance in the United States: evolution and challenges. *Morbidity and Mortality Weekly Report – Surveillance Supplement* 61(3):3-9.

Thun MJ, Sinks T. 2004. Understanding cancer clusters. *CA Cancer Journal for Clinicians* 54(5):273-280.

Tiefelsdorf M, Boots B. 1997. A note on the extremities of local Moran's I and their impact on global Moran's I*. *Geographical Analysis* 29(3):248-257.

Van Meter KC, Christiansen LE, Hertz-Picciotto I, Azari R, Carpenter TE. 2008. A procedure to characterize geographic distributions of rare disorders in cohorts. *International Journal of Health Geographics* 7:26.

Wagner SE, Bauer SE, Bayakly AR, Vena JE. 2013. Prostate cancer incidence and tumor severity in Georgia: descriptive epidemiology, racial disparity, and geographic trends. *Cancer Causes and Control* 24(1):153-166.

Wakefield JC, Kelsall JE, Morris SE. 2000. 8 Clustering, cluster detection, and spatial variation in risk. In: Spatial epidemiology: methods and application (Elliott P, Wakefield JC, Best N, Briggs D, eds.). New York, NY: Oxford University Press, Inc.

Waller LA, Gotway CA. 2004. Spatial clustering of health events: regional count data. In: Applied spatial statistics for public health data (Waller LA, Gotway CA, eds.). Hoboken, NJ: John Wiley & Sons, Inc. 200-271.

Wartenberg D. 1999. Using disease-cluster and small-area analysis to study environmental justice. In: Toward Environmental Justice (Gavin Jr, Mattison DR, & Committee on Environmental Justice, IOM). Washington DC: National Academy Press. pp 79-102.

Weinberg RA. 2006. The nature of cancer. In: Biology of cancer. (Weinberg RA, ed.) New York, New York: Garland Science, Taylor & Francis Group.

Wheeler DC. 2007. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996 – 2003. *International Journal of Health Geographics* 6:13.

Wilkins KL, Woodgate RL. 2008. Preventing second cancers in cancer survivors. *Oncology Nursing Forum* 35(2):E12-E22.

Statistical Review Analytical Methodology

Indirect Standardized Incidence Rate. The raw (sometimes called "crude") disease incidence rate (number of incident case per time period divided by the person-years per period) reflects reality. The raw rate is the simplest and most straightforward summary of the population experience. Interpretation of a disease incidence rate involves a comparison of that rate with some acceptable comparison rate to determine if the rate in question is high or low. Because the interpretation of rates will almost always involve comparing two populations with two different age distributions, comparison of a raw disease incidence rate with a comparison rate is problematic. It does not make sense to compare the rate of disease of a relatively young population to the rate in a relatively old population for a disease that is more common in the elderly, as it will not be possible to state with confidence that the disease rate is higher or lower than expected. For this reason, when the objective is to compare two rates, age standardized rates are preferable. However, it should be noted that the rate itself, once standardized, is not the exact disease burden. The standardized rate should be of the same magnitude as the raw rate.

The indirect standardization method is preferred when the disease count in each age group is small or zero. A disadvantage of the indirect method is that while the adjusted rate is comparable to the comparison population used in its computation, it is not comparable to other population rates. For example, if a study area's cancer rates are adjusted using the Utah state population, they are comparable to the Utah state rates. However, they are not comparable to the county rates or to national rates as those will be adjusted using different criteria (or not adjusted at all).

The indirect standardized rate for the study area (ISR_M) is calculated by:

$$ISR_M = \frac{C_M}{\sum_{age}\left(\frac{C_{U,age}}{P_{U,age}} P_{M,age}\right)} \times \left(\frac{C_U}{P_U}\right) \times 100{,}000$$

Where:

ISR_M is the indirect standardized incidence rate for the study area.

C_M is the total cancer incidence count for the study area for a specific analytical period (e.g., 1990 - 1994).

$C_{U,age}$ is an age-group (e.g., 0 to 19 year in age, etc.) specific cancer incidence count for the comparison population (e.g., the state of Utah) for a specific analytical period.

$P_{U,age}$ is the age-group specific count of person-years (e.g., number of 0-19 year olds in 1990 plus number of 0-19 year olds in 1991 plus number of 0-19 year olds in 1992 ..., etc.) for the comparison population (e.g., the state of Utah) for a specific analytical period.

$P_{M,age}$ is the age-group specific count of person-years for the study area for a specific analytical period.

C_U is the total cancer incidence count for the comparison population (e.g., the state of Utah) for a specific analytical period.

P_U is the total count of person-years for the comparison population (e.g., the state of Utah) for a specific analytical period.

For purposes of presentation, it is standard practice to present rates per a population of 100,000 people. For example, 60 cases per 100,000 people is easier to understand than 0.00006 cases per person.

$$E_M = \sum_{age}\left(\frac{C_{U,age}}{P_{U,age}} P_{M,age}\right)$$

E_M is the expected case count of cancer incidence for the study area for a specific analytical period. This is the denominator factor of the first term of the rate formula.

Standardized Incidence Rate Ratio. The standardized incidence ratio (SIR) is a way of comparing two rates. When using the indirect standardized rate method, the SIR is the first term of the formula to compute the rate.

$$SIR = \frac{C_M}{\sum_{age}\left(\frac{C_{U,age}}{P_{U,age}} P_{M,age}\right)} = \frac{C_M}{E_M}$$

The Byar's confidence limits (CL) can be calculated for the SIR by:

$$\overline{\underline{SIR}} = \frac{(C_M + k)}{E_M} \times \left[1 - \left(\frac{1}{9 \cdot (C_M + k)}\right) \pm \left(\frac{Z_\alpha}{3 \cdot \sqrt{C_M + k}}\right)\right]^3$$

Where:

SIR is the standardized incidence ratio. The bar over and under indicates the upper and lower confidence limits of the SIR.

C_M is the total case count of incident cancers for the study area for a specific analytical period.

E_M is the expected case count of incident cancers for the study area for a specific analytical period.

k is a constant for symmetry. For the upper confidence limit, *k* = 1. For the lower confidence limit, *k* = 0.

Z_α is the normal distribution function for a confidence interval. For the upper confidence interval it is a positive value, while it is negative for the lower. Z-values are given by the following table:

Confidence level	Z– value
70%	0.842
80%	1.282
90%	1.645
95%	1.960
98%	2.326
99%	2.576

Risk Ratio Standard Error. The SIR and confidence limits of the SIR provide a measure of burden and statistical significance. For example, an SIR of 2.0 (CL 1.1 – 2.9) is significant and suggests that the study population has twice as much burden as the comparison population. The standard error is used to assess how significant (without reference to the magnitude of burden) and is useful to determine the need for action.

The standard error (SE) is:

$$se = e^{\left(\sqrt{\frac{\left(1-{C_M}/{P_M}\right)}{C_M}+\frac{\left(1-{E_M}/{P_M}\right)}{E_M}}\right)}$$

Example (1)

Observed Cases =	50
Expected Cases =	25
Study Population =	50,000
Rate Ratio =	2
Lower 99% Confidence Limit =	1.3
Upper 99% Confidence Limit =	2.9
Interpretation:	Statistically Significant
Standard Error =	1.06
1 SE above 1.0 (1.0 + SE)	2.06
Interpretation:	The rate ratio (2.0) is less than 1 SE above the base (2.0 < 2.06). Would need 4 consecutive analytical periods of similar results for the results to be actionable.

Example (2)

Observed Cases =	50
Expected Cases =	10
Study Population =	50,000
Rate Ratio =	5
Lower 99% Confidence Limit =	3.3
Upper 99% Confidence Limit =	7.2
Interpretation:	Statistically Significant
Standard Error =	1.13
3 SE above 1.0 (1.0 + 3*SE)	4.39

Interpretation:	The rate ratio (5.0) is more than 3 SE above the base (5.0 > 4.39). A single analytical period with this result would be actionable.

Commonly used Acronyms and Definitions

ACS	American Cancer Society. The ACS, first established in 1913, is a nationwide voluntary health organization dedicated to eliminating cancer. The society, headquartered in Atlanta, Georgia, has over 900 offices throughout the United States. ACS funding is used for patient support services, research, prevention, detection and treatment and society operations. For more information, see: http://www.cancer.org.
ACS	American Community Survey. The ACS is an ongoing survey that provides annual updates to population and demographic estimates derived from census data. The ACS is operated by the U.S. Census Bureau. For more information, see: http://www.census.gov/programs-surveys/acs.
AGRC	Automated Geographic Reference Center. An agency within the Utah Department of Information Technology responsible for maintaining a repository of geographic information system (GIS) data files and GIS functionality. For more information, see: http://gis.utah.gov.
ATSDR	The Agency for Toxic Substances and Disease Registry. A federal public health agency within the U.S. Department of Health and Human Services. The agency focuses on minimizing human health risks associated with exposure to hazardous substances. For more information, see: http://www.atsdr.cdc.gov.
CDC	The Centers for Disease Control and Prevention. A federal public health agency within the U.S. Department of Health and Human Services. The agency focuses on protecting public health by conducting and supporting health promotion, prevention, and preparedness activities. For more information, see: http://www.cdc.gov.
EEP	Environmental Epidemiology Program. A program within the Bureau of Epidemiology, Division of Disease Control and Prevention, Utah Department of Health. The EEP was established in 1996 and is responsible for investigating diseases related to the environment. The program has three sections. One section conducts surveillance and data management activities, including managing the UEPHTN. The second section conducts health hazards risk assessment, including cancer investigations. The third section is responsible for environmental sanitation activities. The program is staffed by personnel with experience and expertise in environmental epidemiology, environmental sciences, toxicology, statistics, public health informatics and geomatics, sanitation, and health education. For more information, see: http://health.utah.gov/enviroepi.

EPA	U.S. Environmental Protection Agency. The U.S. EPA is one of the executive agencies within the federal government. The U.S. EPA is responsible for regulatory actions that protect human health and the environment from environmental health hazards. For more information, see: http://www.epa.gov.
GIS	Geographic Information System. A GIS includes computer software and geographically referenced data. The EEP uses QGIS or ArcGIS as the computer software, and obtains data from ESRI or AGRC.
IBIS-PH	The Indicator-Based Information System for Public Health. IBIS-PH is the primary online repository of public health and other population-level data and information in Utah. For more information, see: http://www.ibis.health.utah.gov.
NCI	National Cancer Institute. The NCI is one of the National Institutes of Health and part of the U.S. Department of Health and Human Services. The NCI was established under the National Cancer Act of 1937 and is primarily responsible for conducting surveillance and research about cancer incidence, diagnosis, prevention, treatment, and rehabilitation. The SEER program is operated by the NCI. For more information, see: http://www.cancer.gov.
SEER	Surveillance, Epidemiology and End Results Program. The SEER program is an agency within the NCI. The SEER program works with state cancer registries to develop and disseminate incidence and mortality statistics about cancer in the United States. The SEER program also establishes standards for the analysis of cancer data and interpretation of cancer statistics. For more information, see: http://seer.cancer.gov.
UBRFS	Utah Behavioral Risk Factors Survey. The UBRFS is an ongoing telephonic survey conducted by the Office of Public Health Assessment, UDOH. This survey collects data about health-related behaviors in the non-institutionalized Utah adult population. For more information, see: http://health.utah.gov/opha/OPHA_BRFSS.htm.
UCR	Utah Cancer Registry. The UCR is operated under authority from the UDOH by the University of Utah. The UCR was established in 1966 to be a statewide population-based cancer registry. Utah administrative rule requires the reporting of cancer diagnoses to the UCR. The UCR collaborates with the NCI, SEER and the North American Association of Central Cancer Registries to implement data standards for cancer data. The UCR provide cancer to the EEP through the UEPHTN. For more information, see: http://ucr.utah.edu.

UDEQ — Utah Department of Environmental Quality. The UDEQ is one of the executive agencies within Utah state government. The UDEQ strives to safeguard public health and quality of life by protecting and enhancing the environment through the implementation, compliance monitoring and enforcement of environmental laws. For more information, see: http://deq.utah.gov.

UDOH — Utah Department of Health. The UDOH is one of the executive agencies within Utah state government. The UDOH strives to improve health in Utah through promoting healthy lifestyles, evidence-based interventions, creating healthy and safe communities and eliminating health disparities. The EEP is a program within the UDOH. For more information, see: http://health.utah.gov.

UEPHTN — Utah Environmental Public Health Tracking Network. The UEPHTN is a data warehouse that contains health outcomes, environmental and supporting data. Data from the UCR and population data derived from the USCB are warehoused in the UEPHTN. For more information, see: http://epht.health.utah.gov/epht-view.

USCB — U.S. Census Bureau. Officially the "Bureau of the Census," the USCB is an agency authorized by Federal law, within the U.S. Department of Commerce, that is charged with preparing and conducting regular surveys and censuses of the U. S. population. In addition to the decennial population survey, the USCB conducts a number of other surveys and has recently implemented the ACS. For more information, see: http://www.census.gov.

Cancer Incidence: The term incidence refers to new cases occurring in a period of time, usually annually. Cancer incidence is the number of new cases that occurred in a year. New cancer cases occur when a diagnosis is made. For example, the 2009 national age-adjusted incidence rate is 4.64 cancer cases per 1,000 people per year. For more information, see: http://www.cancer.gov/statistics/glossary/incidence.

Cancer Incidence Rate: This is a ratio of the cancer incidence (the number of new cancer diagnoses in a population) divided by the total time experienced by that population. When computing a multi-year rate, the total population added from each year of the rate period is used to get the rate. For more information, see: http://www.cancer.gov/statistics/glossary/incidence.

Cancer Prevalence: The term prevalence refers to the number of cases that exist either at a moment in time or during a period of time (e.g., annual, lifetime, etc.). When using this term, the time should be included. For example, the 2009 national lifetime cancer prevalence rate

is approximately 414.65 cases of cancer per 1,000 people. For more information, see: http://www.cancer.gov/statistics/glossary/prevalence.

Opportunities, Challenges, and Limitations

Listed below are some, but not all, of the considerations that may influence the response to a potential cancer cluster.

Opportunities

- A response can instill the public with confidence about government action to their concern and empowers the public for their own responses.
- The findings of a statistical review can help public health officials and policymakers identify and assess communities with public health challenges and can inform public health priorities.
- The findings may guide and empower communities to appropriately direct efforts and take actions regarding their concerns.
- Statistical review results (especially those conducted as a follow-up to an initial investigation) may be useful in monitoring and evaluating public health actions, programs, and policies.
- Statistical review results may provide evidence and impetus for further epidemiologic investigation.
- The environmental and community investigation helps inform possible actions that may reduce risks and improve health.
- The environmental and community investigation may help inform the community on potential political actions for policy development.
- The environmental and community investigation will help inform the feasibility study for conducting an epidemiologic investigation.
- Epidemiologic investigations are able to address and quantify some causal risks.

Challenges and Limitations

- The public often wants the initial study (the statistical review) to address specific or nonspecific environmental concerns (e.g., a specific industry, air quality in general, etc.). Statistical reviews are unable to support identification and characterization of risks such as environmental exposures. There is seldom sufficient information available to statistically control for the many non-environmental factors that contribute to cancer risk (e.g., population-level behavior risks, genetic risks, etc.).
- There is often considerable variability in the time between the triggering event and the emergence of cancer. Since these kinds of investigations rely on the clinically-manifested

incidence of cancer, there may be some temporal misclassification of the cases in each analytical period.

- Statistical reviews are limited by power constraints of statistics (the sensitivity or ability to correctly find a real effect). It is not possible to evaluate areas with population counts below certain thresholds and maintain statistical veracity. Thus, statistical reviews often aggregate data across space and time as part of the methodology. Aggregation may result in the study including a larger area than where the cluster is perceived to be at and/or longer analytical periods than when the cluster is perceived to have occurred.
- The methodology is dependent on registry data and census estimates. For example, persons who lived most of their lives outside of Utah but moved to Utah just before being diagnosed with cancer, or *vice versa*, persons who lived in Utah for most of their lives but moved out before being diagnosed, are misclassified with respect to their exposure to causal events. Similarly, the census-based population estimates do not adequately account for local population trends and growth.
- Statistical reviews are ecologic studies in nature. The results should only apply to the population and cannot be used to address individual risk concerns. An individual may have no risk or a risk several times higher than the population risk based on the individual's genetic makeup, behaviors, exposure history, and susceptibility or resilience to various risk factors.
- This investigation provides some assessment of potential risks factors that are not controlled for in the statistical review. As such, it is not possible to specifically link the risks to the cancers of concern, except as indicated by literature review.
- In the literature, these kinds of assessments have been rarely successful in discovering new knowledge about risk and environmental causes.
- In most cases, comprehensive data about all of the known risk factors is absent. Some data (e.g., behaviors, family histories, and exposure histories) can be collected by individual risk assessments. Other data can be generated through exposure reconstruction or modeling. However, some data, such as genetic risks, may not be available or be very difficult and expensive to acquire.
- The data may be limited by lack of willingness to participate amongst the study cohort.
- Data collection will involve some costs and unintended consequences to the participants. At a minimum, there will be a time cost. Depending on the probable risk factors, there may also be medical costs and sharing of privately held information.
- Typically, epidemiologic investigations require a long time to complete.