

Statewide Leukemia Study
November 18, 2016

CANCER INCIDENCE STUDY

**Utah Statewide Investigation
of Leukemia
for Spatiotemporal Clustering Patterns
Between 1980 to 2013**

November 18, 2016

Prepared by the

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

TABLE OF CONTENTS

ACKNOWLEDGMENT.....	1
EXECUTIVE SUMMARY	2
INTRODUCTION	3
DATA AND METHODS	7
FINDINGS	11
DISCUSSION	13
CONCLUSIONS AND RECOMMENDATIONS	16
AUTHORSHIP, REVIEW AND CITATION	18
CERTIFICATION	19
REFERENCES	20
FIGURES AND TABLES	36
DEFINITIONS.....	40
RESOURCES	45

ACKNOWLEDGMENT

Cancer data used for this investigation was obtained from the Utah Cancer Registry (UCR). The UCR is funded by contract number HHSN261201300017I from the National Cancer Institute (NCI) Surveillance, Epidemiology and End Results (SEER) Program with additional support from the Utah Department of Health (UDOH) and the University of Utah (UCR 2015).

Other data and analytical tools used for this investigation were obtained from the Utah Environmental Public Health Tracking Network (UEPHTN). In addition, the UEPHTN provides geocoding services to UCR data. The UEPHTN is funded by a grant from the Centers for Disease Control and Prevention (CDC), Environmental Public Health Tracking Branch. The current UEPHTN award is number 1U38EH000954 (UEPHTN 2016).

EXECUTIVE SUMMARY

Cancer is a dominating environmental public health concern (CDC 2016). A function of epidemiology is to investigate cancer incidence, starting with a statistical review of cancer cases. The Environmental Epidemiology Program (EEP), a program within the Utah Department of Health (UDOH), conducts statistical reviews of cancer in Utah. During the past ten years, the EEP has received a number of requests from concerned citizens to investigate leukemia incidence in Utah.

This report presents a statistical review of the spatial and temporal distribution of leukemia in Utah from 1980 to 2013 using a spatiotemporal scan methodology. The purpose of this review was to identify regions of Utah with an historical or ongoing excess occurrence of leukemia. No areas of elevated leukemia rates were found in Utah during this timeframe.

The rate of leukemia in Utah (13.3 cases per 100,000 people) is rising and is similar to the current national rate (13.4 cases per 100,000). A comprehensive literature review of known risk factors for leukemia did not reveal any significant environmental risk other than exposure to strong ionizing radiation and benzene. Other risks have been studied and the available evidence was found to be generally inconclusive. This report can be used by local and state public health officials to formulate a response to concerned citizens who perceive increased leukemia rates in their communities.

INTRODUCTION

Blood Constituents and Functions: Blood is a fluid that performs essential physiological functions to maintain life. The main components of blood are blood cells and blood plasma. The cellular components of blood are comprised of erythrocytes (red blood cells), leukocytes (white blood cells), and thrombocytes (platelets). Erythrocytes transport inhaled oxygen to tissue throughout the body and carbon monoxide to the lungs for exhalation. Leukocytes play an important role in the body's immune system response to pathogens and other foreign substances. There are five main types of leukocytes: lymphocytes, monocytes, basophils, neutrophils, and eosinophils. Thrombocytes help facilitate hemostasis to mitigate blood loss when blood vessel damage occurs. Blood plasma is the extracellular matrix that facilitates the flow of blood cells throughout the body (Silverthorn 2013a).

Blood Cell Creation: Hematopoiesis is the process of blood cell creation and differentiation. Blood cells originate from pluripotent hematopoietic stem cells, which are created in bone marrow. These stem cells differentiate into either myeloid or lymphoid progenitor stem cells, which determine what type of blood cell is ultimately produced. Myeloid stem cells can produce erythrocytes, thrombocytes, or certain types of leukocytes (monocytes, basophils, neutrophils, or eosinophils). Basophils, neutrophils, and eosinophils may also be referred to collectively as granulocytes, due to their grain-like appearance. The lymphoid stem cells produce lymphocytes (Silverthorn 2013b; NCI 2013). Figure 1 displays pathways of hematopoiesis.

Leukemia: Leukemia refers to cancer of the blood cells; it arises as a result of mutations that develop in the blood stem cells during hematopoiesis. Instead of differentiating and maturing, the mutated blood stem cells replicate and proliferate (Hoffman et al. 2000). These cancerous cells eventually leave the bone marrow, enter the blood stream, and impact other organs in the body. The overabundance of these cancerous cells interferes with the proper function of normal blood cells (NCI 2013).

Leukemia is classified by the type of stem cell from which the abnormalities originate. Lymphocytic leukemia arises from the lymphoid progenitor cells, and myeloid (or myelogenous) leukemia arises from the myeloid progenitor cells (Jaffe et al. 2001; Knowles 1992). These classifications are further designated by the rate at which the mutated cells proliferate (acute or chronic). This results in four primary classifications of leukemia (Linnet et al. 2006): acute lymphocytic leukemia (ALL), chronic lymphocytic leukemia (CLL), acute myeloid leukemia (AML), and chronic myeloid leukemia (CML).

The nationwide incidence of leukemia has remained stable between 1975 and 2013 (NCI 2016a). However, there are variations in trends among the different classifications of leukemia. From 2004 to 2013, rates of AML (4.1 cases per 100,000 people) increased an average of 3.4% annually, and rates of ALL (1.7 cases per 100,000 people) have increased an average of 0.6% annually (NCI 2016b, c). During this same timeframe, rates of CLL (4.6 cases per 100,000 people) have decreased an average of 1.3% annually, and rates of CML (1.8 cases per 100,000 people) have remained stable (NCI, 2016d, e). In Utah, the rate of all leukemia is 13.3 cases per 100,000 people; the nationwide rate is 13.4 cases per 100,000 people (NCI 2016f).

Known Risk Factors for Leukemia: In addition to the known risk factors discussed below, differences in screening and detection practices may contribute to variations in leukemia incidence across geographies and over time. Some risk factors may only apply to certain types of leukemia.

Sex: Leukemia is slightly more common in men than women (Ries et al. 2003). Nationally, approximately 1.5% of both men and women will be diagnosed with a form of leukemia in their lifetime (NCI 2016g).

Radiation: Ionizing radiation is a frequently studied risk factor for leukemia (Linet et al. 2006). X-rays and gamma rays are examples of ionizing radiation. While AML is most frequently associated with ionizing radiation exposure, AML, CML, and ALL are frequently grouped together when describing radiation-related risks for leukemia (Boice et al. 1991; Linet et al. 2006). Studies of atomic bomb survivors demonstrated the effects of ionizing radiation on leukemia mortality and AML incidence (Nakashini et al. 1999; Shimizu et al. 1990; UNSCEAR 1994; Little and Muirhead 1998; Preston et al. 1994). The highest risks of ALL incidence were among children younger than age 10 (Preston et al. 1994). The risk of developing therapy-induced AML increases when an individual is exposed to radiation for treatment purposes (Boice et al. 1987; Curtis et al. 1992, 1994; Travis et al. 1994, 1999; Kuttesch et al. 1996; Inskip 1999). Children who were in utero when their mothers were exposed to radiation from atomic bombs did not have an increased risk of ALL (DeLongchamp et al. 1997). Some evidence suggests that children whose fathers are employed in nuclear facilities experience increased risks for ALL (Gardner et al. 1990; Roman et al. 1993). However, this conclusion is not always consistent (Kinlen et al. 1993; McLaughlin et al. 1993). There is no evidence that environmental radioactivity increases the risk of leukemia in adults (Forastiere et al. 2002; Law et al. 2000), or children (Steinbuch et al. 1999; UK Childhood Cancer Study Investigators, 2002; Lubin et al. 1998; Kaletsch et al. 1999). Nonionizing radiation sources are not associated with increased risk of leukemia (Feychting & Ahlbom 1994; Severson et al. 1988).

Benzene: Benzene (C₆H₆) is a flammable, liquid hydrocarbon that is widely used in chemical, plastic, rubber, pesticide, drug, dye, lubricant, and detergent manufacturing (ATSDR 2007). Unleaded gasoline and cigarette smoke are common sources of low-level benzene exposure (Brugnone et al. 1994; Kok and Ong 1994; Melikian et al. 1993; Ong and Lee 1994). Benzene is a known carcinogen that is typically associated with AML (ATSDR 2007; IARC 2012; Delore and Borgomano 1928; Ott et al. 1978; Rinsky et al. 1987; Wong 1987; Yin et al. 1987; Hayes et al. 1997; Rinsky et al. 2002). Benzene has also been associated with CML, ALL, myelodysplastic syndrome, and non-Hodgkin lymphoma (Yin et al. 1987; Rinsky et al. 1987; Schnatter et al. 1992; 1996a,b; Hayes et al. 1997; Huebner et al. 1997; Ireland et al. 1997; Rushton and Romaniuk 1997; Savitz and Andrews 1997).

Genetic disorders: Genetic disorders are predominantly associated with AML and ALL, accounting for an estimated 5% of these types of leukemia (Taylor and Birch 1996; Birch 1999). Down syndrome (trisomy 21) and Bloom syndrome are both associated with an increased risk of AML and ALL (Bloom et al. 1966; German 1997; Hasle et al. 2000; Malkin et al. 2000; Hill et al. 2003; Linet et al. 2006). Other syndromes associated with AML include GATA1 mutations, Fanconi anemia, Shwachman-Diamond syndrome, amegakaryocytic thrombocytopenia, dyskeratosis congenita, and Kostmann syndrome (Wechsler et al. 2002; Mundchau et al. 2003; Groet et al. 2003; Alter 2003; Dror

and Freedman 2001; Alter 1996).

Familial History: Family history is an important risk factor for CLL, which is the most commonly shared subtype among family members with leukemia (Vidabaek 1947; Sgambati et al. 2001; Pottern et al. 1991; Linet and Pottern 1992). Conversely, evidence suggests that familial history presents a lower risk in developing AML and ALL. While familial AML patterns have been documented, the specific mechanisms of pathogenesis are more varied and often not shared among family members (Horwitz et al. 1997; Kwong et al. 2000; Grimwade et al. 1993; Olopade et al. 1996; Mandla et al. 1998). Evidence for familial ALL patterns have been documented in cases of inter-family marriage and siblings with certain genetic mutations (Kende et al. 1994; Rischewski et al. 2000).

Tobacco Use: Tobacco use has been associated with a 20-50% increased risk of AML in adults (Brownson et al. 1993; Doll 1996). Some evidence suggests that tobacco use may increase the risk of CLL, but these findings have not always been consistent (Kinlen and Rogot 1988; Garfinkel and Boffetta 1990; Brown et al. 1992a; Friedman 1993; Adami et al. 1998). Among children, maternal smoking during pregnancy is not associated with an increased risk of AML (Magnani et al. 1990; Van Duijn et al. 1994; Brondum et al. 1999; Schuz et al. 1999; Sasco and Vainio 1999) or ALL (Sorahan et al. 1995; Shu et al. 1996). Paternal smoking prior to conception is not associated with an increased risk of AML among children (Magnani et al. 1990; Sorahan et al. 1997). However, it has been associated with an increased risk of ALL (Ji et al. 1997; Sorahan et al. 1997; Sorahan et al. 2001; Shu et al. 1996).

Alcohol Use: Alcohol use is not associated with increased risk of AML in adults (Williams and Horn 1977; Blackwelder et al. 1980; Hinds et al. 1980; Cartensen et al. 1990; Brown et al. 1992b). There is some evidence that maternal alcohol use while pregnant may increase the risk of AML in children (Severson et al. 1993; Shu et al. 1996), but not ALL (Van Steensel-Moll et al. 1985b; Nishi and Miyake 1989; Shu et al. 1996; Schuz et al. 1999; Infante-Rivard et al. 2002).

Previous Studies of Leukemia Risk in Utah: The United States started conducting test detonations of nuclear weapons in July 1945. The first test took place at the Alamogordo Bombing Range (now known as the White Sands Missile Range) in New Mexico. Within the United States, nuclear testing was conducted at ten different tests sites in Alabama, Alaska, Colorado, and Nevada (Beck & Bennett 2002; DOE 2000).

The Nevada Test Site (now the Nevada National Security Site), located in Nye County, Nevada, was used to conduct nuclear detonation tests from 1951 to 1992. One thousand twenty-one (1,021) detonations occurred at the site, including 100 detonations that were above ground. The above ground detonations all occurred between 1951 and 1958 (Beck & Bennett 2002; DOE 2000; Fehner & Gosling 2000).

Several studies have evaluated the effects of radioactive fallout from nuclear detonation testing on leukemia incidence in southwestern Utah. While the majority of these studies found statistically significant elevated risks of leukemia, increased risk was typically observed in individuals younger than age 20 (Weiss 1965; Lyon et al. 1979; Johnson 1984; Machado et al. 1987; Stevens et al. 1990). One study (Land et al. 1984) did not observe a statistically significant association. Another study attempted to evaluate improved methodologies of estimating fallout exposure levels in southwestern Utah (Simon et al. 1995).

Cancer Incidence Statistical Reviews: A core function of epidemiology is to track and evaluate disease patterns. This function helps public health officials and policymakers identify and assess communities with public health challenges; define public health priorities; develop and implement informed public health policy; monitor and evaluate public health actions; discover knowledge about public health concerns; and guide public health outreach, education and intervention activities (Dicker 2002; Lawson & Kulldorff 1999; Stanbury et al. 2012; Thacker 2000; Thacker et al. 2012). Cancer is a dominating environmental public health concern (CDC 2016; Goujon-Bellec et al. 2011; Morrone 2011; Wakefield et al. 2000). Public concerns about excess cancer risk often result in requests to public health agencies to conduct investigations. Public health agencies conduct investigations of cancer incidence using several methods. The first is a cancer incidence statistical review, which focuses on determining if a particular community is experiencing more cancer than would be expected. A cancer statistical review is usually conducted by linking cancer registry and population data and evaluating trends. From the public health perspective, a cancer incidence statistical review is most useful in identifying community needs about cancer-related health education and awareness building, public health screening services, and other public health interventions. These kinds of studies empower the community to make improvements in governmental policymaking and health care services (Bell et al. 2006; Kingsley et al. 2007).

One of the outcomes of a statistical review is identifying the probable patterns of disease clustering. A spatial cluster (also called a “hot spot”) is defined as a limited area within a general study area with a significant and meaningful increase in the incidence of disease. A temporal cluster is a defined period of time within a larger range of time with a significant and meaningful increase in disease incidence. A spatiotemporal cluster is a cluster defined in both the geographic and temporal dimensions (Aamodt et al. 2006; Hinrichsen et al. 2009; Lawson & Kulldorff 1999; Wakefield et al. 2000; Wheeler 2007).

The discovery of a possible disease cluster usually warrants additional action, either as continued monitoring or a more aggressive investigation. However, disease clusters may not always be a public health concern. When evaluating a long period of time, historical clusters may be discovered that have already resolved themselves. Clustering may be the natural result of the distribution of residential or demographic population patterns or may be a function of wall-to-wall analytical units that do not properly accommodate disease patterns. For example, there are no areas within the boundaries of Utah that are not also part of a census tract geographic area. Thus, some census tracts include areas of geography (e.g., the Bonneville Salt Flats, or upper elevations of Utah’s mountain ranges) that are uninhabited. Clusters also may occur because of chance or due to the presence of factors that are immeasurable or highly variable (Wakefield et al. 2000). Furthermore, clusters may be reported due to improper application of statistical analytical methods (Tango 1999).

Another method available to public health practitioners is a cancer cluster investigation. Cancer cluster investigations focus on characterizing the size and extent of a population with known cancer excess and determining potential causal factors. The cancer cluster methodology involves linking many causal variables, usually collected by medical record review and individual surveys or interviews, followed by a complex statistical analysis to identify the few variables that seem to explain the risk (Kingsley et al. 2007). However, cluster investigations rarely result in important

discoveries of causality (Goodman et al. 2012; Kingsley et al. 2007).

Public Statement of Concern: The Environmental Epidemiology Program (EEP), within the Utah Department of Health (UDOH), receives concerns from the public about perceived excess cases of cancer (or clusters). During the past ten years, the EEP has received a number of concerns about excess leukemia in communities in Daggett, Emery, Salt Lake, and Utah counties. Those investigations examined the rate of leukemia compared to the rest of the state for a prior defined small area (i.e., a neighborhood or community) using a retrospective statistical review methodology (CDC 1990; Jekel et al. 1996; Mann 2003). None of the statistical reviews found a statistically significant increased risk for leukemia in the communities investigated.

Study Objectives: This report presents a statistical review of the spatial and temporal distribution of all types of primary leukemia in Utah from 1980 through 2013 using a spatiotemporal scan methodology. The purpose of this review was to identify statewide trends of leukemia and regions of Utah with an historical or ongoing excess incidence of leukemia.

Authority and Funding: This study was conducted as part of the UDOH Executive Director's responsibility to investigate public health concerns within Utah. The Executive Director delegates responsibility for cancer investigations to the EEP. Cancer, population, and geographic data for this investigation are collected, maintained, and made available by the Utah Environmental Public Health Tracking Network (UEPHTN). The UEPHTN also funds the SAS[®] and ArcGIS[®] analytical software application licenses that were used to conduct this investigation. The UEPHTN is funded by a grant from the Centers for Disease Control and Prevention (CDC) (UEPHTN 2016). Personnel time used to conduct this investigation was charged against state-funded Environmental Health Administrative funds. No federal funds were directly used to conduct this investigation.

DATA AND METHODS

Study Design: This investigation is a retrospective statistical review of cancer using spatiotemporal scanning methodology to identify spatial clusters in the data. Statistical reviews are not cancer cluster investigations and lack the power to link cancer incidence to putative risk factors (Jekel et al. 1996; Kingsley et al. 2007; Mann 2003). A statistical review is a tool used by the EEP to evaluate the health status of a population, identify public health needs, and assess public health activities. A good study design includes determining the underlying spatiotemporal epidemiologic theory; selecting appropriate scales of analysis; selecting an appropriate analytical methodology; defining risk and exposure; and determining how to manage locational and attribution uncertainty (Meliker & Sloan 2011). The appropriate analytics model for this investigation applies binomial statistics, since an individual in the population either is or is not a cancer case. Since cancer incidence is relatively rare, the Poisson distribution (a special case of the binomial distribution) is appropriate. For this investigation, the smallest, consistently available scale for case and population data is the census tract geographic unit for the spatial dimensions, and diagnosis year for the temporal dimension. This investigation used the spatiotemporal scan statistic to look for current and historic clusters. The spatiotemporal scan method creates many different aggregations of contiguous spatial and temporal analytical units (e.g., northern Utah County from 2008 to 2012). The method then compares the incidence of cancer inside each

aggregation to the incidence of cancer outside the aggregation to identify spatiotemporal areas of excess cancer. The study's null hypothesis is that the incidence of cancer is randomly dispersed in both the geographic and temporal dimensions. Age is an important risk factor for cancer and was controlled for in this investigation.

Cancer Data: Cancer incidence data on people diagnosed with invasive cancer (behavior type 3) between 1973 and 2013 were obtained from the Utah Cancer Registry (UCR). The EEP receives cancer data for all reported CIS and invasive cancers on an annual basis (UCR 2015). For this investigation, only invasive leukemia cancers were considered. Population data aggregated by age group and sex at the 2010 U.S. census tract geography for Utah is only available from 1980 forward. Therefore, cancer data from 1980 through 2013 were used for this investigation.

The primary site of a cancer is the body site, organ, or tissue in which the tumor originates. Metastasis is the spread of cancer cells originating in one site or tissue to another organ or tissue within the body (e.g., ovarian cancer cells spreading to the liver). Metastatic cancer cells carry some characteristics of tissue type from where they originated often allowing the primary site of metastatic tumors to be identified (Higginson et al. 1992). The UCR data categorizes a metastatic cancer according to the organ or tissue that was the site of the primary tumor, regardless of the location of metastasis. Additionally, the UCR does not report benign tumors (behavior type 0) or borderline tumors (behavior type 1) to the EEP.

The UCR completes a rigorous data review for completion and quality before data are released to the EEP. The most recent years of data are not made available to the EEP until the review has been finalized. The UCR data includes diagnostic information, patient demographics, and residential addresses of the cases, as well as information about the behavior of the cancer (UCR 2015).

Individuals with multiple primary invasive cancers have multiple records in the UCR data set in sequential order. These cancers are distinguished by unique cancer registry tracking numbers and a cancer sequence number. The sequence number allows discrimination between the first cancer diagnosis and subsequent diagnoses (NCI 2012; UCR 2015). Diagnostic coding of cancers includes the International Classification of Disease Oncology, 3rd Edition (ICD-O-3) codes for site, histology, and behavior (WHO 2012). For convenience to the UDOH, the UCR groups cancer into 42 major cancer types by site following the guidance provided by the National Cancer Institute (NCI) Surveillance, Epidemiology, and End Results (SEER) Program (NCI 2012). These site codes are a convenient grouping for conducting surveillance analyses (UCR 2015). Leukemia cases were identified using histology codes 9733 to 9964. Table 1 provides details about the site codes. Due to the similar nature of environmental-related risk factors shared by leukemia subtypes, all subtypes were combined for analysis.

Between 1980 and 2013, 6,997 cases of leukemia were reported to the UCR. The residential address information provided by the UCR includes the patient's street address, city, and ZIP code at the time of diagnosis (UCR 2015). The EEP geocodes each registry record's residential address data to obtain an x- and y-coordinate for that address. Most addresses are automatically geocoded using address locator data obtained from the Utah State Geographic Information Database (SGID), maintained by the Utah Automated Geographic Reference Center (AGRC) (AGRC 2016). Addresses not found due to land reutilization, street name changes, address

realignment, or are newer than the address locator file, were researched using historic street maps and references or using online street maps, integrated aerial photographs, and other references to locate the address and manually geocode it. Using the geocoded x- and y-coordinates, the EEP was able to geo-reference 6,725 (96.1%) of the cancer cases to their respective U.S. 2010 census geographic tracts. A few (272 cases or 3.9%) case addresses could not be geocoded (e.g., postal box addresses, etc.). These cases were geo-referenced to the most populated census tract for the smallest known geographical area (the ZIP code or municipal boundary) indicated by the address. It is highly unlikely that this geo-referencing method affected the final outcome. The final count for leukemia included in this statistical review was 6,725 cases.

Population Data: The 2010 U.S. census divides Utah into 588 census tracts (USCB 2002, 2012a) with a median population of 4,700 (range 0 to 21,591) persons per census tract in the year 2010. These small area geographies range in size between 0.2 to 6,108 square miles (average = 144 square miles, standard deviation = 608 square miles) and have a population density ranging from less than one person per square mile in the sparsely populated areas to more than 29,400 persons per square mile in the urbanized Wasatch Front. The average density was 3,443 people per square mile (standard deviation = 3,358). Commercially available U.S. census population data for Utah for the 1980, 1990, 2000 and 2010 censuses (Geolytics 2014) were used to estimate annual five-year age-group population counts in each census tract for each intercensal year between 1980 and 2013. These estimates were made by applying annual population growth rates derived from the previous and subsequent decennial data following national population estimation guidelines (USCB 2012b).

Census Tract Data: Geographic Information System data for the 2010 U.S. Census Bureau (USCB) geographies in the form of shape files for Utah census counties, Utah census tracts, Utah census block groups and Utah census blocks were obtained from the SGID (AGRC 2016). The 2010 U.S. Census data applies a “wall-to-wall” geographic coverage, meaning that there are no areas within state boundaries that are not accounted for within a census level geography. In Utah, some census geographies include a mix of residential neighborhoods, commercial/industrial areas and uninhabited areas. The data for each census geographic unit includes a geographic centroid (the center point of the area geography). The scan statistical methodology uses the centroid point to represent the “average” location of the population and disease cases for each geographic analytical unit. This study used a Cartesian projection of the data using the North American 1983 Universal Transverse Mercator (UTM) datum for zone 12N, which results in the centroid x- and y-coordinates being expressed in meters. For this investigation, a population-weighted centroid was preferable for representing the geographic location of the study population and cases. The population-weighted centroid was determined using census blocks. The U.S. census enumerates population at all geographic levels, but provides age and sex stratification only at the census block group level and larger. A population-weighted centroid was calculated by importing census tract and census block GIS data attribute tables containing fields for the census tract level standard federal identifier (STFID) and the geographic x- and y- coordinates into SAS[®] for Windows version 9.3 (SAS 2011) as follows:

$$Coord_T = \frac{\sum(Coord_B Pop_B)}{\sum Pop_B}$$

Where: *Coord* is the x- or y-coordinate
T is the target census tract
B is the source census block contained within the target census tract
Pop_B is the census block total population

On average, the population-weighted centroid differed from the geographic centroid by 472 (range 0 to 16,346) meters.

The STFID is a unique label applied to census geographic units. The STFID was used as the key to link census geography to population data and cancer case data.

Data Linkage: Census tract case and population data were tabulated by census tract using the STFID, by year, and five-year age group using SAS[®] for Windows version 9.3 (SAS 2011). Coordinate data, case data, and population data referenced by the STFID were exported to a database file that is compatible for import into SaTScan.

SaTScan: The SaTScan[™] version 9.4.2 computer application applies spatiotemporal scanning methodology (Kulldorff 2010; Kulldorff & IMS 2011; SaTScan 2015). SaTScan implements a class of statistics known as “scan statistics” originally developed to scan through the spatial and temporal dimensions of interest, looking for anomalies in the incidence of events of interest (Wakefield et al. 2000). Cases and the underlying population are represented by a three-dimensional space-time point. This study uses the census tract centroid as the geographic component coordinates. The scan statistical method creates many cylindrical windows, where the base represents geography and the height represents time. These cylindrical windows are centered on each census tract and unit of time. Each cylinder is expanded incrementally to include multiple contiguous census tracts and units of time. The incidence of cancer represented inside the cylinder is compared to the incidence of cancer outside the cylinder to identify areas and time periods of statistically elevated cancer incidence. Many thousands of overlapping cylinders are evaluated and ranked for the likelihood of a cancer cluster. For this evaluation both circular shaped and elliptical shaped geography bases were used. The elliptical-based scan included all orientations and shapes of ellipses (Jones & Kulldorff 2012; Kulldorff 1997, 2010; Kulldorff & Nagarwalla 1995; Kulldorff et al. 2006).

The SaTScan application features a number of models that can be used. For this study, the discrete Poisson model for space-time cluster detection was used (Amin & Burns 2014; Kulldorff 1997, 2010; Wagner et al. 2013). Age was added to the model as a covariate. Model parameters are decisions or limitations applied within the application to “tune” the model. Examples of model parameters include limits on the shape and size of the windows and the inclusions of various adjustments for spatial nonstationarity. For this investigation, the model

used an elliptic spatial window shape with medium noncompactness penalty. In most cases, the choice of the penalty does not dramatically change the findings (Goujon-Bellac et al. 2011). The temporal maximum cluster size was set at 90% of the study period (Hsu et al. 2004; Van Meter et al. 2008). The incidence rate was adjusted using an automatically calculated log-linear trend because some areas of Utah have experienced population growth at unequal rates (Kulldorff 2010). No geographic overlapping of clusters was allowed. Scans were run with other model parameters (e.g., more or less compactness, population size limits, with or without stationarity adjustment, etc.) with little difference in the findings. A more liberal p-value of less than or equal to 0.10 was used instead of the typical 0.05 threshold to determine statistical significance. This decision was allowed because of the small case count (“the rarity”) for the clusters (Dietz et al. 2011; Hsu et al. 2004; Park 2010; Wagner et al. 2013; Wheeler 2007). The SaTScan application implements methodology as part of the likelihood calculation to control for the many calculations. Only areas with higher-than-expected rates were considered during the scan. Cluster data was output as a data file that was joined to the attribution table of a geographic data file (shapefile) of Utah census tracts for symbolization and visualization. Relative risk is one of the measures SaTScan generates to quantify the disease burden for a likely cluster. SaTScan only reports cluster areas that have a statistically elevated relative level.

Cluster Homogeneity and Cluster Confirmation: Several discrete and noncontiguous areas with slightly elevated rates that individually are not statistically powerful enough to be distinguished from the random variation may, when combined within an aggregated area, result in the delineation of a cluster area. This kind of false-cluster would be represented visually by a heterogeneous presentation of small area rates. True clusters would have a homogenous presentation of high rates compared to the surrounding small areas (Chen et al. 2008). A direct age-standardized incidence rate for the aggregated cluster period for each of the 588 census tracts within the state was calculated using the overall state rate as the standard (Anderson & Rosenberg 1998; Besag & Newell 1991; Breslow & Day 1987; Esteve et al. 1994; Jekel et al. 1996; Mann 2003; Selvin 1996). The census tracts within each cluster area were isolated using a spatial clipping technique; the global Moran’s I test was used to determine the level of spatial homogeneity of rates within the cluster area (Cromley & McLafferty 2012; Marshall 1991; Moore & Carpenter 1999; Moran 1950; Wakefield et al. 2000; Waller & Gotway 2004). The local Moran’s I and the local Getis-Ord G tests were used to confirm the cluster locations determined by the scan test (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. 2000; Waller & Gotway 2004).

FINDINGS

Statewide Descriptive Assessment: Between 1980 and 2013, 6,725 geocode-able cases of primary leukemia were reported to the UCR. For 5,426 (80.7%) of those individuals, leukemia was their first cancer experience. The other 1,299 (19.3%) cases had experienced a previous primary cancer of another type. Of the cases diagnosed during the 34-year study period (1980-2013), 4,505 (66.9%) of cases had died by the end of the study period (2013).

The distribution of cases by age at the time of diagnosis is shown in Figure 2. The median age of cases is 65 years (range 0 years to 102 years). Fifty-eight percent (3,893) of cases were male. The

majority (6,159 or 92%) of the cases are non-Hispanic Caucasian.

Leukemia is typically classified from the type of blood stem cell from which the mutation originates (myeloid or lymphocytic) and the rate at which mutated cells proliferate (acute or chronic). Less common types of leukemia are grouped into different categories. The following table presents the distribution of leukemia by general classification. Each class of cancer may have several subclasses.

Leukemia Classification	Case Count	Percent
Acute Lymphocytic	966	14.4%
Chronic Lymphocytic	2,184	32.5%
Other Lymphocytic	283	4.2%
Acute Myeloid	1,823	27.1%
Chronic Myeloid	865	12.9%
Other Myeloid	66	0.9%
Monocytic	143	2.1%
Other Acute	148	2.2%
Aleukemic, Subleukemic, and NOS*	247	3.7%
Total	6,725	100.0%

* = not otherwise specified

The number of primary leukemia cancer diagnoses in Utah increased from 112 cases in 1980 to 288 cases in 2013. This increase is correlated ($R^2 = 0.93$) with Utah's statewide population growth. This correlation was used to inform the approach for the temporal trend adjustments applied to the final model. The Utah population has increased from approximately 1.47 million people in 1980 to 2.77 million in 2010 (GOMB 2015).

According to the Utah state cancer profile provided by the NCI, the rate of leukemia among Utah's population changed around 1990 (NCI 2016a). Figure 3 presents a graphic taken from the NCI cancer profile. The standardized leukemia incidence rate in Utah in 1975 was 11.5 cases per 100,000 people. By 1990, the incidence rate in Utah was 11.1 cases per 100,000 people. The average rate of decrease was -0.8 cases per 100,000 people per year from 1975 to 1990. By 2012, the incidence rate was 14.1 cases per 100,000 people. The average rate of increase was 0.7 from 1990 to 2012 (NCI 2016a).

SaTScan Results: SaTScan is a tool that scans data using all possible permutations of contiguous geography and time up to the maximum limits set by the user to identify likely spatio-temporal clusters. The tool quantifies the burden of these likely clusters with a relative risk measure and the significance of the clusters with a probability or p-value (Kulldorff & IMS 2011).

The population is presented in person-years units. To understand this unit, a cluster in a community of 1,000 persons lasting 10 years represents 10,000 person-years (1,000 persons x 10 years = 10,000 person-years). Relative risk is a ratio of the risk (incidence rate) of cancer in the cluster area population to the state's risk. If the cluster area's level of risk equals the state's level of risk, the relative risk ratio will equal one, which is interpreted as no increased burden of disease. Values greater than one are interpreted as higher risk than expected burden of disease. Conversely,

values lower than one are interpreted as lower risk than expected. SaTScan only reports likely clusters when the relative risk ratio is statistically elevated, however, for convenience of interpretation, the 95% confidence intervals (95% CIs) are included (Frumkin & Kantrowitz 1987). The 95% CI ranges that almost include 1.0 (for example, an interval range of 1.1 – 1.5) are less meaningful than those that do not (for example, an interval range of 2.0 – 2.5). Both possible cluster areas had meaningfully increased relative risk values.

SaTScan generates an estimate of the likelihood of the cluster being a real spatiotemporal cluster and not just an artifact of the variability in the data. The likelihood is presented as a measure (probability) of randomness (or p-value). High p-values indicate a high degree of probability that the pattern is a result of the random variability in the data and not a real cluster. Low p-values indicate a higher likelihood of a real cluster. For this investigation, a p-value less than or equal to 0.10 was used to identify the significance of clustering (Dietz et al. 2011; Hsu et al. 2004; Wagner et al. 2013; Wheeler 2007).

No historical or current leukemia clusters were identified by scanning the data.

DISCUSSION

Cancer: Cancer is second only to heart disease as a leading cause of death and public health concern in the U. S. (CDC 2016; Goujon-Bellec et al. 2011; Morrone 2011; Wakefield et al. 2000). 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 style 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 (NCI 2016h).

Performance of the SaTScan Application: SaTScan is widely used and well accepted as a tool for discovering spatiotemporal clusters of cancer (Aamodt et al. 2006; Almeida et al. 2011; Chen et al. 2008; Cromley & McLafferty 2012; Oliveira et al. 2011; Robertson & Nelson 2010; Van Meter et al. 2008). The discrete Poisson model performs well over a wide range of disease burden levels and geographic or temporal scales and is the preferred model (Cromley & McLafferty 2012; Neill 2009). Because the tool is easy to use and the results are easy to interpret, SaTScan is particularly popular for use by state and local public health agencies with responsibility to carry out cancer surveillance and cluster assessment. However, the SaTScan tool and its application in this study are not without limitations.

- SaTScan uses simple circular or elliptical shaped geographic filters to identify study areas that might be clusters. These study areas consist of aggregations of small area geographies, which are in turn represented to the application by centroid points. SaTScan is unable to consider the true geography of the small areas or the aggregations of those small areas. Rather, it only considers the location of the centroid points. Because of this limitation, SaTScan responds best when the small area geographies used to represent populations are somewhat uniform in size and simple in shape. SaTScan is most able to detect circular- or elliptical-shaped clusters and

may not be able to detect irregularly (e.g., “S” or “U”) shaped clusters (Aamodt et al. 2006; Goujon-Bellec et al. 2011; Oliviera et al. 2011; Wheeler 2007).

- Because of the use of simple circular or elliptical filters to identify clusters, consideration of the potential shape of clusters in the study area is an important concern (Cromley & McLafferty 2012; Wheeler 2007). For this study, elliptical filters were used.
- Related to the above limitations is the tendency of SaTScan to merge several small, strong irregularly shaped clusters that do not fit well in a circular or elliptical filter into one larger, less significant cluster that fits better (Oliveira et al. 2011; Van Meter et al. 2008).
- Clusters are detected by capturing census tract population-weighted centroids inside a scanning filter, but are visualized (and thus interpreted) by the census tract boundaries, which are wall-to-wall and may include unpopulated areas (e.g., mountain ranges, salt flats, etc.). The visualization of the cluster does not exactly reflect the true location and shape of the cluster (Read et al. 2011).
- Additionally, the SaTScan tool is not capable of considering geographical features other than the census tracts represented by their centroid point. As a result, the aggregating process may result in the combining of distinct communities that may have natural barriers (i.e., lakes, mountain ranges, etc.) that tend to isolate them from each other.
- Census tracts are politically-derived boundaries designed to segregate populations based on political habits, not necessarily on health risk.
- SaTScan identifies those contiguous census tract areas with sufficient levels of elevated risk that are most likely to be part of the cluster. The clusters are presented as the boundary that includes all of the census tracts that were part of the cluster. This boundary is an epidemiological tool for presentation and not a definitive location of where excess risk starts or ceases to exist. A census tract may include one or more unique local neighborhoods. Local neighborhoods within a census tract in a cluster may have a normal level of risk that is lost due the averaging with other neighborhoods in the census tract or cluster. Correspondingly, neighborhoods outside of the cluster may experience an excess level that is lost in the averaging process. Thus, individuals may not experience what this report is presenting with respect to the rate of leukemia in their neighborhoods.
- Another concern about using census tract geographies is one of sensitivity to scaling parameters and the “modifiable areal unit problem (MAUP).” SaTScan is sensitive to the boundary effects described by MAUP (Chen et al. 2008; Ozonoff et al. 2007). There are two kinds of issues; both are examples of the zonation problem associated with MAUP. One is the ability to aggregate neighboring census tracts bounded in this study by the state boundary (Parenteau & Sawada 2011). Census tract boundaries that include the state boundary are limited to including only those neighboring census tracts that are in the direction toward the inside of the state or are also along the boundary. In Utah, 28 (5.6%) of the 496 census tracts are on the state boundary, but these census tracts represent 54% of the total state land area. The other issue is described by the number of neighboring census tracts (Parenteau & Sawada

2011). On average, each census tract has six (range = 1 to 17, standard deviation = 2.2) neighbors. Five census tracts have only one neighbor.

- At the census tract level, the statistical sensitivity of the SaTScan (the ability to detect a cluster location) decays quickly when the relative risk is below 2.5 or when the expected case count is small. The expected case count can be small when either the population in the suspected cluster area is small or when attempting to find clusters of a rare disease. The population can be small because the suspected cluster area is small or the population density is low. However, the SaTScan tool has good specificity (ability to distinguish between true clusters and random variation) at all levels of risk (Aamodt et al. 2006; Cromley & McLafferty 2012; Goujon-Bellec et al. 2011; Jackson et al. 2009).
- When performing spatial-temporal statistical analysis to detect disease clusters, how the local geographies are conceptualized is one of the weakest theoretical aspects of these kinds of studies. For this investigation, the 2010 U.S. census tract geographies were used. The EEP has not yet obtained the required data to make good estimates of the intercensal population sizes for the 2010 census. The EEP was not able to determine which census block groups the cancer cases belonged in for a high enough percentage of cases to consider using the census block group geographies. Census tracts are designed with the intent of election consistency and without consideration of health concerns, health status, or health risks. Many census tracts include large areas within the tract that are not residential (either commercial/industrial or uninhabited) (Parenteau & Sawada 2011). The spatial scan statistic factors in uneven geographical population densities and conditions as part of the analysis for hot spots (Hsu et al. 2004). Although only a small number (3.9%) of the cancer cases could not be accurately geocoded, those cases may result in a geographic selection bias (Dietz et al. 2011).
- With respect to the population estimates used by the EEP, the case count more accurately reflects the true dynamic population growth in a region than the steady state growth represented by a straight line interpolation between the decennial census tabulations. The true population growth trend is not linear. As a result, there will be periods of time where the true population sizes are significantly different from the estimated population sizes. This may be occurring in the western Salt Lake/Utah county area.

Methodology Limitations: The public often wants public health investigations to determine if cancer risk can be linked to a putative environmental concern. The methods (the indirect standardized incidence ratio and the spatiotemporal scanning for clusters) used in this investigation do not have the capability to definitively link the findings of elevated cancer risk to any inherent or external risk factors including environmental exposures. There are a number of limitations that impede this linkage. These kinds of cancer statistical reviews are based on annual incidence data reported to the UCR, and the incidence of cancer per year is dependent on diagnosis of clinically-manifested cancer. There is seldom any knowledge about the frequency, duration, or intensity of cancer victims' exposure to a putative environmental concern. Cancer can have a variable length latency period between the time of exposure to the actual manifestation and diagnosis of cancer. Cancer can be present for some time before an individual seeks medical assistance that leads to a diagnosis. There is seldom sufficient information available to statistically control for the many non-environmental factors that contribute to cancer risk, let alone exposure to other potential environmental risks that are not the putative environmental concern. Chance also plays a role in

the distribution of cancer and is often the dominating causal factor in small populations or for diseases that occur rarely. Often, a few types of cancer may be statistically elevated for disparate periods, but that conclusion may change if the analytical periods are changed. Overcoming these limitations usually requires a comprehensive assessment of individual risk supported by a clear and consistent trend of elevated rates for a population.

This investigation used data from the UCR and U.S. Census. In Utah, the diagnosis of cancer for all site categories is reportable to the UCR. When a Utah resident seeks diagnosis, a report is generated and the UCR follows-up on the report to confirm information and collect additional factors about the case. This process occurs when cases are diagnosed in Utah, but may not occur if a case is diagnosed outside of the state. The UCR may contain records of incidence of cancer in people who recently moved to the study area prior to their diagnosis, and conversely may lack records on individuals who lived most of their life in the study area but moved elsewhere before seeking diagnosis and treatment. These situations create ascertainment biases. For the purposes of diagnosis, the EEP assumes that the ascertainment bias is non-systematic, meaning that the “move-in” and “move-out” situations balance each other. It is highly unlikely that this assumption is true in all cases and can be a significant limitation when the study population is small.

The EEP uses U.S. census data purchased from a commercial vendor. The vendor has re-tabulated 1980, 1990, 2000, and 2010 data for the 2010 census block groups in Utah. Re-tabulation involves population-distribution weighting based on census blocks that may not be consistent through time. The EEP estimates intercensal population counts using linear regression between the known census tabulations. This methodology does not account for short-term population growth dynamics such as the zoning and development of a new subdivision, which can occur in just a few years.

A limitation of these kinds of investigations is that inferences leading to public health meaning are based on the snapshots of reality generated by data analysis (Meliker & Sloan 2011). An investigation that uses population-based summary data rather than individual-level data is called an ecologic study by epidemiologists. This investigation is an ecologic study. An interpretation error commonly associated with ecologic investigations is to apply population-level risk findings to the individual. This kind of interpretation error is called an “ecologic fallacy.” This risk metric should not be applied to individuals. 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 resiliency to cancer (Greenland 2001; Greenland & Robins 1994; Izquierdo & Schoenbach 2000; Morgenstern 1982, 1995; Rockhill 2005).

CONCLUSIONS AND RECOMMENDATIONS

The study failed to identify any current or historic leukemia clusters. A comprehensive review of the literature did not reveal any significant environmental risk other than exposure to ionizing radiation and benzene. Potential causal factors identified by literature review were presented, but not investigated as part of this investigation. Random variation (chance) is an important element in any investigation involving cluster, and can play a dominant role when the investigation involves rare diseases or small population units.

People who are afflicted with leukemia are best served by their health care team. Concerned citizens, who think they may have leukemia, or may be at risk for developing leukemia, should be referred to their health care provider. This report can be used as a tool when formulating a response to concerned citizens who perceive increased leukemia rates in their communities.

AUTHORSHIP, REVIEW, AND CITATION

This report was prepared by:

Matt McCord, MSPH
Environmental Epidemiology Program
Bureau of Epidemiology
Utah Department of Health

Mail: PO Box 142104, Salt Lake City, Utah 84114-2104
Street: 288 North 1460 West, Salt Lake City, Utah 84116
Phone: (801) 538-6191
Fax: (801) 538-6564
Email: mmccord@utah.gov

Certifying Reviewers:

Allyn K Nakashima, MD
Director, Bureau of Epidemiology
State Epidemiologist
Utah Department of Health

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

Heather Borski
Director, Division of Disease Control and Prevention
Utah Department of Health


Recommended Citation:

Environmental Epidemiology Program. 2016. *Utah Statewide Investigation of Leukemia for Spatiotemporal Clustering Patterns Between 1980 to 2013*. Salt Lake City, UT: Utah Department of Health.


CERTIFICATION

This report titled “Utah Statewide Investigation of Leukemia for Spatiotemporal Clustering Patterns Between 1980 to 2013” was prepared by the Environmental Epidemiology Program, Utah Department of Health. This report covers an investigation of cancer incidence using standard and approved methodology and procedures existing at the time the investigation herein reported was begun. Editorial and technical review was completed by UDOH certifying reviewers and program partners.


Approved by:

 11/16/16

Allyn K. Nakashima, MD
Director, Bureau of Epidemiology
State Epidemiologist
Utah Department of Health

 11/14/16

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

 11/17/16

Heather Borski
Director, Division of Disease Control and Prevention
Utah Department of Health

REFERENCES

Web links for citations of government or organization websites may wrap onto multiple lines.

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

Adami J, Nyren O, Bergstrom R, et al. 1998. Smoking and the risk of leukemia, lymphoma, and multiple myeloma (Sweden). *Cancer Causes Control* 9:49-56.

AGRC (Automated Geographic Reference Center). 2016. Utah's state geographic information database: 2000 demographic data files. Available <http://gis.utah.gov> [accessed August 12, 2016].

Almeida ACL, Duarte AR, Duczmal LH, Oliveira FLP, Takahashi RHC. 2011. Data-driven inference for the spatial scan statistic. *International Journal of Health Geographics* 10:47.

Alter BP. 1996. Fanconi's anemia and malignancies. *Am J Hematol* 53:99-100.

Alter BP. 2003. Cancer in Fanconi anemia 1927-2001. *Cancer* 97:425-440.

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

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.

ATSDR (Agency for Toxic Substances and Disease Registry). 2007. Toxicological Profile for Benzene. US Department of Health and Human Services, Atlanta, GA.

Beck HL, Bennett BG. 2002. Historical overview of atmospheric nuclear weapons testing and estimates of fallout in the continental United States. *Health Physics* 82(5):591-608.

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.

Besag J, Newell J. 1991. The detection of clusters of rare disease. *Journal of the Royal Statistical Society, Part A* 154:143-155.

Birch JM. 1999. Genes and cancer. *Arch Dis Child* 80:1-3.

Blackwelder WC, Yano K, Rhoads GG, et al. 1980. Alcohol and mortality: The Honolulu Heart Study. *Am J Med* 68:164-169.

Bloom GE, Warner S, Gerland PS, et al. 1966. Chromosome abnormalities in constitutional aplastic anemia. *N Engl J Med* 274:8-14.

Boice JD Jr, Blettner M, Kleinerman RA, et al. 1987. Radiation dose and leukemia risk in patients treated for cancer of the cervix. *J Natl Cancer Inst* 79:1295-1311.

Boice JD Jr, Morin MM, Glass AG, et al. 1991. Diagnostic x-ray procedures and risk of leukemia, lymphoma, and multiple myeloma. *J Am Med Assoc* 265: 1290-1294.

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

Brondum J, Shu XO, Steinbuch M, et al. 1999. Parental cigarette smoking and the risk of acute leukemia in children. *Cancer* 85:1380-1388.

Brown LM, Gibson R, Blair A, et al. 1992a. Smoking and risk of leukemia. *Am J Epidemiol* 135: 763-768.

Brown LM, Gibson R, Burmeister LF, et al. 1992b. Alcohol consumption and risk of leukemia, non-Hodgkin's lymphoma, and multiple myeloma. *Leuk Res* 16:979-984.

Brownson RC, Novotny TE, Perry MC. 1993. Cigarette smoking and adult leukemia. A meta-analysis. *Arch Intern Med* 153:469-475.

Brugnone F, Perbellini L, Giuliani C, et al. 1994. Blood and urine concentrations of chemical pollutants in the general population. *Med Lav* 85: 370-389.

Cartensen JM, Bygren LO, Hatschek T. 1990. Cancer incidence among Swedish brewery workers. *Int J Cancer* 45:393-396.

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). 2016. Leading causes of death. Available: <http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm> [accessed August 10, 2016].

Chen J, Roth RE, Naito AT, Lengerich EJ MacEachren AM. 2008. Geovisual analytics

to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *International Journal of Health Geographics* 7:57.

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.

Curtis RE, Boice JD, Stovall M, et al. 1992. Risk of leukemia after chemotherapy and radiation treatment for breast cancer. *N Engl J Med* 326:1745-1751.

Curtis RE, Boice JD, Stovall M, et al. 1994. Relation of leukemia risk of radiation dose after cancer of the uterine corpus. *J Natl Cancer Inst* 86:1315-1324.

Delongchamp RR, Mabuchi K, Yoshimoto Y, et al. 1997. Cancer mortality among atomic bomb survivors exposed in utero or as young children, October 1950 – Nov 1992. *Radiat Res* 147:385-395.

Delore P, Borgomano C. 1928. Leukemia aigue un cour de l'intoxication benzeique, sur l'origine toxique de certains leukemias aigutes et leurs relations avec les anemies graves. *J Med Lyon* 9:227-233.

Dicker RC. 2002. A brief review of the basic principles of epidemiology. In: *Field Epidemiology*, 2nd ed. (Greg 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.

DOE (Department of Energy). 2000. United States nuclear tests: July 1945 through September 1992, DOE/NV-2009-REV 15. Las Vegas, NV: Nevada Operations Office. Available: <https://nnsa.energy.gov/sites/default/files/nnsa/inlinefiles/doe%20nv%202000e.pdf> [accessed August 12, 2016].

Doll R. 1996. Cancers weakly related to smoking. *Br Med Bull* 52:35-49.

Dror Y, Freedman MH. 2001. Shwachman-Diamond syndrome marrow cells show abnormally increased apoptosis mediated through the Fas pathway. *Blood* 97:3011-3016.

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.

Fehner TR, Gosling FG. 2000. Origins of the Nevada test site, DOE/MA-0518. Washington, DC: Department of Energy, History Division. Available: <http://www.dd.anl.gov/ddtraining/50yrsNTSHistory.pdf> [accessed August 12, 2016].

Feychting M, Ahlbom A. 1994. Magnetic fields, leukemia, and central nervous system tumors

in Swedish adults residing near high-voltage power lines. *Epidemiology* 5:501-509.

Forastiere F, Sperati A, Cherubini G, et al. 2002. Adult myeloid leukaemia, geology, and domestic exposure to radon and gamma radiation: A case-control study in central Italy. *Occup Environ Med* 55:106-110.

Friedman GD. 1993. Cigarette smoking, leukemia, and multiple myeloma. *Ann Epidemiol* 3: 425-428.

Frumkin H, Kantrowitz W. 1987. Cancer clusters in the workplace: an approach to investigation. *Journal of Occupational Medicine* 29(12):949-952.

Gardner MJ, Snee MP, Hall AJ, et al. 1990. Results of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plants in West Cumbria. *BMJ* 300: 423-429.

Garfinkel L, Boffetta P. 1990. Association between smoking and leukemia in two American Cancer Society prospective studies. *Cancer* 65:2356-2360.

Geolytics, Inc. 2014. Census digital optical disks (CDs) for the 1980, 1990, 2000 and 2010 US census counts in the 2010 census boundaries. Information: <http://www.GeoLytics.com> [accessed August 12, 2016]. The following disks were used to generate intercensal estimates used by EEP:

- 1980 Census in 2010 Boundaries, Long (SF3) and Short Form (SF1), version 1.0
- 1990 Long Form in 2010 Boundaries, The complete 1990 U.S. Census Data normalized to 2010 boundaries, Release 1.0
- 2000 Long Form in 2010 Boundaries, The complete 2000 U.S. Census Data normalized to 2010 boundaries, Release 1.0
- American Community Survey 2011

German J. 1997. Bloom's syndrome. XX. The first 100 cancers. *Cancer Genet Cytogenet* 93:100-106.

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.

GOMB (Governor's Office of Management and Budget). 2015. Population Estimates by State and County. Available: <http://gomb.utah.gov/budget-policy/demographic-economic-analysis/> [accessed August 15, 2016].

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.

Goujon-Bellec S, Demoury C, Guyot-Goubin A, Hemon D, Clavel J. 2011. Detection of clusters of a rare disease over a large territory: performance of cluster detection methods. *International Journal of Health Geographics* 10:53.

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.

Grimwade DJ, Stephenson J, De Silva C, et al. 1993. Familial MDS with 5q-abnormality. *Br J Haematol* 84:536-538.

Groet J, McIwaine S, Spinelli M, et al. 2003. Acquired mutations in GATA1 in neonates with Down syndrome with transient myeloid disorder. *Lancet* 362:1617-1620.

Hasle H, Clemmensen IH, Mikkelsen M. 2000. Risks of leukaemia and solid tumours in individuals with Down's syndrome. *Lancet* 355:165-169.

Hayes RB, Yin SN, Dosemeci M, et al. 1997. Benzene and dose-related incidence of hematologic neoplasms in China. Chinese Academy of Preventive Medicine-National Cancer Institute Benzene Study Group. *J Natl Cancer Inst* 89:1065-1071.

Higginson J, Muir CS, Munoz N. 1992. Thyroid and other endocrine glands. In: *Human cancer: epidemiology and environmental causes*. (Higginson J, Muir CS, Munoz N. eds.). London, UK: Cambridge University Press. pp.445-448.

Hill DA, Gridley G, Cnattingius S, et al. 2003. Mortality and cancer incidence among individuals with Down syndrome. *Arch Intern Med* 163:705-711.

Hinds MW, Kolonel LN, Lee T, et al. 1980. Associations between cancer incidence and alcohol and cigarette consumptions among five ethnic groups in Hawaii. *Br J Cancer* 41: 929-940.

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.

Hoffman R, Benz EJ Jr, Shattil SJ, et al. 2000. *Hematology. Basic Principles and Practice*. 3rd ed. New York: Church Livingstone.

Horwitz M, Benson KF, Li FQ, et al. 1997. Genetic heterogeneity in familial acute myelogenous leukemia: evidence for a second locus at chromosome 16q21-23.2. *Am J Hum Genet* 61:873-881.

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.

Huebner WW, Schnatter AR, Nicolich MJ, et al. 1997. Mortality experience of a young petrochemical industry cohort. 1979-1992 follow-up study of US-based employees. *J Occup Environ Med* 39:970-982.

IARC (International Agency for Research on Cancer). 2012. Benzene. In: *Chemical Agents and Related Occupations, Volume 100F, A Review of Human Carcinogens, IARC Monographs of Carcinogenic Risks to Humans*. Lyon, France: IARC Press.

Infante-Rivard C, Krajcinovic M, Labuda D, et al. 2002. Childhood acute lymphoblastic leukemia associated with parental alcohol consumption and polymorphisms of carcinogen-metabolizing genes. *Epidemiology* 13:277-281.

Inskip PD. 1999. Second cancer following radiotherapy. In: Neugut AI, Meadows AT eds. *Multiple Primary Cancers*. Philadelphia: Lippincott Williams & Wilkins, pp. 91-135.

Ireland B, Collin JJ, Buckley CF, et al. 1997. Cancer mortality among workers with benzene exposure. *Epidemiology* 8:318-320.

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.

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.

Jaffe ES, Harris NL, Stein H, et al. 2001. *Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon, France: IARC Press.

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

Ji BT, Shu XO, Linet MS, et al. 1997. Paternal cigarette smoking and the risk of childhood cancer among offspring of nonsmoking mothers. *J Natl Cancer Inst* 89:238-244.

Johnson CJ. 1984. Cancer incidence in an area of radioactive fallout downwind from the Nevada Test Site. *J Am Med Assoc* 251: 230-236.

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

- Kaletsch U, Kaatsch P, Meinert R, et al. 1999. Childhood cancer and residential radon exposure – results of a population-based case-control study in Lower Saxony (Germany). *Radiat Environ Biophys* 38:211-225.
- Kende G, Toren A, Mandel M, et al. 1994. Familial leukemia: Description of two kindreds and a review of the genetic aspects of the disease. *Acta Haematol* 92:208-211.
- Kinlen LJ, Rogot E. 1988. Leukaemia and smoking habits among United States veterans. *Br Med J* 297:657-659.
- Kinlen LJ, Clarke K, Balkwill A. 1993. Paternal preconceptional radiation exposure in the nuclear industry and leukaemia and non-Hodgkin's lymphoma in young people in Scotland. *BMJ* 306:1153-1158.
- 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.
- Knowles DM, ed. 1992. *Neoplastic Hematopathology*. Baltimore, MD: Williams & Wilkins.
- Kok PW, Ong CN. 1994. Blood and urinary benzene determined by headspace gas chromatography with photoionization detection: Application in biological monitoring of low-level nonoccupational exposure. *Int Arch Occup Environ Health* 66:195-201.
- Kulldorff M. 1997. A spatial scan statistic. *Communications in Statistics: Theory and Methods*. 26(6):1481-1496.
- Kulldorff M. 2010. SaTScan™ 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. SaTScan™ software for spatial and space time scan statistics, for MS Windows, version 9.1.1. Information Management Services, Inc., Silver Spring, MD. Information: <http://www.satscan.org>.
- Kulldorff M, Nagarwalla N. 1995. Spatial disease clusters: detection and inference. *Statistics in Medicine* 14(8):799-810.
- Kuttesch JF Jr, Wexler LH, Marcus RB, et al. 1996. Second malignancies after Ewing's sarcoma: Radiation dose dependency of secondary sarcomas. *J Clin Oncol* 14:2818-2825.
- Kwong YL, Ng MH, Ma SK. 2000. Familial acute myeloid leukemia with monosomy 7: Late onset and involvement of a multipotential progenitor cell. *Cancer Genet Cytogenet* 116:170-173.

- Land CE, McKay FW, Machado SC. 1984. Childhood leukemia and fallout from the Nevada nuclear tests. *Science* 223: 139-144.
- Law GR, Kane EV, Roman E, et al. 2000. Residential radon exposure and adult acute leukemia. *Lancet* 355:1888.
- 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.
- Linnet MS, Pottern LM. 1992. Familial aggregation of hematopoietic malignancies and risk of non-Hodgkin's lymphoma. *Cancer Res* 52 (19 Suppl): 5468s-5473s.
- Linnet MS, Devesa SS, Morgan GJ. 2006. The Leukemias. In: *Cancer Epidemiology and Prevention*, 3rd Ed. (Schottenfeld D, Fraumeni JH, eds.). New York, NY: Oxford University Press pp. 841-871.
- Little MP, Muirhead CR. 1998. Curvature in the cancer mortality dose response in Japanese atomic bomb survivors: Absence of evidence of threshold. *Int J Radiat Biol* 74:471-480.
- Lubin JH, Linnet MS, Boice JD Jr, et al. 1998. Case-control study of childhood acute lymphoblastic leukemia and residential radon exposure. *J Natl Cancer Inst* 90: 294-300.
- Lyon JL, Klauber M, Gardner JW, et al. 1979. Childhood leukemias associated with fallout from nuclear tests. *N Engl J Med* 300: 397-402.
- Machado SG, Land CE, McKay FW. 1987. Cancer mortality and radioactive fallout in southwestern Utah. *Am J Epidemiol* 125: 44-61.
- Magnani C, Pastore G, Luzzatto L, et al. 1990. Parental occupational and other environmental factors in the etiology of leukemias and non-Hodgkin's lymphomas in childhood: A case-control study. *Tumori* 76:413-419.
- Malkin D, Brown EJ, Zipursky A. 2000. The role of p53 in megakaryocyte differentiation and the megakaryocytic leukemias of Down syndrome. *Cancer Genet Cytogenet* 116:1-5.
- Mandla SG, Goobie S, Kumar RT, et al. 1998. Genetic analysis of familial myelodysplastic syndrome: Absence of linkage to chromosomes 5q31 and 7q22. *Cancer Genet Cytogenet* 105: 113-118.
- Mann CJ. 2003. Observation research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency Medicine Journal* 20(1):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.

- McLaughlin JR, King WD, Anderson TW, et al. 1993. Paternal radiation exposure and leukaemia in offspring: The Ontario case-control study. *BMJ* 307:959-966.
- Melikian AA, Prahalad AK, Hoffman D. 1993. Urinary trans-muconic acid as an indicator of exposure to benzene in cigarette smokers. *Cancer Epidemiol Biomarkers Prev* 2:47-51.
- Meliker JR, Sloan CD. 2011. Spatio-temporal epidemiology: principles and opportunities. *Spatial and Spatio-temporal Epidemiology* 2(1):1-9.
- 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(12):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.
- Mundchau G, Gurbunxani S, Gaines AS, et al. 2003. Mutagenesis of GATA1 is an initiating event in Down syndrome leukemogenesis. *Blood* 101:4298-4300.
- Nakashini M, Tanaka K, Shintani T. et al. 1999. Chromosomal instability in acute myelocytic leukemia and myelodysplastic syndrome patients among atomic bomb survivors. *J Radiat Res (Tyokyo)* 40:159-167.
- NCI (National Cancer Institute). 2012. SEER program coding and staging manual 2012. Available: http://seer.cancer.gov/archive/manuals/2012/SPCSM_2012_maindoc.pdf [accessed August 12, 2016].
- NCI (National Cancer Institute). 2013. What you need to know about leukemia. Available: <http://www.cancer.gov/publications/patient-education/leukemia.pdf> [accessed August 1, 2016].
- NCI (National Cancer Institute). 2016a. State Cancer Profiles: Nationwide. Historical Trends 1975-2013, Incidence, SEER 9 Registries, Leukemia, All Races (including Hispanic), All Ages, Both Sexes. Available: <http://www.statecancerprofiles.cancer.gov> [accessed August 2, 2016].
- NCI (National Cancer Institute). 2016b. SEER Stat Fact Sheets: Acute Myeloid Leukemia. Available: <http://www.seer.cancer.gov/statfacts/html/amyl.html> [accessed August 1, 2016].
- NCI (National Cancer Institute). 2016c. SEER Stat Fact Sheets: Acute Lymphocytic Leukemia.

Available: <http://www.seer.cancer.gov/statfacts/html/aly1.html> [accessed August 1, 2016].

NCI (National Cancer Institute). 2016d. SEER Stat Fact Sheets: Chronic Lymphocytic Leukemia. Available: <http://www.seer.cancer.gov/statfacts/html/clyl.html> [accessed August 1, 2016].

NCI (National Cancer Institute). 2016e. SEER Stat Fact Sheets: Chronic Myeloid Leukemia. Available: <http://www.seer.cancer.gov/statfacts/html/cm1l.html> [accessed August 1, 2016].

NCI (National Cancer Institute). 2016f. State Cancer Profiles: Utah (age-adjusted incidence rates). Available: <http://www.statecancerprofiles.cancer.gov> [accessed August 2, 2016].

NCI (National Cancer Institute). 2016g. SEER Stat Fact Sheets: Leukemia. Available: <http://www.seer.cancer.gov/statfacts/html/leuks.html> [accessed August 1, 2016].

NCI (National Cancer Institute). 2016h. Surveillance, Epidemiology and End Results (SEER) Program. Available: <http://seer.cancer.gov> [accessed August 15, 2016].

Neill DB. 2009. An empirical comparison of spatial scan statistics for outbreak detection. *International Journal of Health Geographics* 8:20.

Nishi M, Miyake H. 1989. A case-control study of non-T cell acute lymphoblastic leukaemia of children in Hokkaido, Japan. *J Epidemiol Community Health* 43:352-355.

Oliveira FLP, Duczmal LH, Cancado ALF, Tavares R. 2011. Nonparametric intensity bounds for the delineation of spatial clusters. *International Journal of Health Geographics* 10:1.

Olopade OI, Roulston D, Baker T, et al. 1996. Familial myeloid leukemia associated with loss of the long arm of chromosome 5. *Leukemia* 10:669-674.

Ong CN, Lee BL. 1994. Determination of benzene and its metabolites: application in biological monitoring of environmental and occupational exposure to benzene. *J Chromatogr B Biomed Appl* 660:1-22.

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

Ott MG, Townsend JC, Fishbeck WA, et al. 1978. Mortality among individuals occupationally exposed to benzene. *Arch Environ Med* 33:3-10.

Ozonoff A, Jeffery C, Manjourides J, Forseberg White L, Pagano M. 2007. Effect of spatial resolution on cluster detection: a simulation study. *International Journal of Health Geographics*

6:52.

Parenteau M-P, Sawada MC. 2011. The modifiable areal unit problem (MAUP) in the relationship between exposure to NO₂ and respiratory health. *International Journal of Health Geographics* 10:58.

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

Pottern LM, Linet M, Blair A, et al. 1991. Familial cancers associated with subtypes of leukemia and non-Hodgkin's lymphoma. *Leuk Res* 15:305-314.

Preston DL, Kusumi S, Tomonaga M, et al. 1994. Cancer incidence in atomic bomb survivors. Part III. Leukemia, lymphoma and multiple myeloma, 1950-1987. *Radiat Res* 137:S68-97.

Read S, Bath P, Willett P, Maheswaran R. 2011 Measuring the spatial accuracy of the spatial scan statistic. *Spat Spatio-temporal Epidemiol* 2(2):69-78.

Ries LAG, Eisner MP, Kosary CL, et al. 2003. SEER Cancer Statistics Review, 1975-2000. Hankey BF, Miller BA, Clegg L, Mariotto A, Fay MP, Feuer EJ, Edwards BK, eds. Bethesda, MD: National Cancer Institute. Available at: http://seer.cancer.gov/csr/1975_2000.

Rinksy RA, Smith AB, Hornung R, et al. 1987. Benzene and leukemia. An epidemiologic risk assessment. *N Engl J Med* 316:1044-1050.

Rinksy RA, Hornung RW, Silver SR, et al. 2002. Benzene exposure and hematopoietic mortality: A long-term epidemiologic risk assessment. *Am J Ind Med* 42:474.

Rischewski JR, Clausen H, Leber V, et al. 2000. A heterozygous frameshift mutation in the Fanconi anemia C gene in familial T-ALL and secondary malignancy. *Klin Paediatr* 212: 174-176.

Robertson C, Nelson TA. 2010. Review of software for space-time disease surveillance. *International Journal of Health Geographics* 9:16.

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.

Roman E, Watson A, Beral V, et al. 1993. Case-control study of leukaemia and non-Hodgkin's lymphoma among children aged 0-4 years living in west Berkshire and north Hampshire health districts. *BMJ* 306:615-621.

Rushton L, Romaniuk H. 1997. A case-control study to investigate the risk of leukaemia associated with exposure to benzene in petroleum marketing and distribution workers in the

United Kingdom. *Occup Environ Med* 54:152-166.

SAS. 2011. SAS[®] statistical analytical software for MS Windows desktop version 9.3. SAS Institute, Inc., Cary, NC.

Sasco AJ, Vainio H. 1999. From in utero and childhood exposure to parental smoking to childhood cancer: A possible link and the need for action. *Hum Exp Toxicol* 18:192-201.

SaTScan. 2015. Software for the spatial, temporal, and space-time scan statistics for MS Windows desktop version 9.4.2. Information: <http://www.satscan.org/> [accessed August 12, 2016].

Savitz DA, Andrews KW. 1997. Review of epidemiologic evidence on benzene and lymphatic and hematopoietic cancers. *Am J Ind Med* 31:287-295.

Schnatter AR, Theriault G, Katz AM, et al. 1992. A retrospective mortality study within operating segments of a petroleum company. *Am J Ind Med* 22:209-229.

Schnatter AR, Armstrong TW, Nicolich MJ, et al. 1996a. Lymphohaematopietic malignancies and quantitative estimates of exposure to benzene in Canadian petroleum distribution workers. *Occup Environ Med* 53:773-781.

Schnatter AR, Armstrong TW, Thompson LS, et al. 1996b. The relationship between low-level benzene exposure and leukemia in Canadian petroleum distribution workers. *Environ Health Perspect* 104:1375-1379.

Schuz J, Kaletsch U, Meinert R, et al. 1999. Association of childhood leukemia with factors related to the immune system. *Br J Cancer* 80:585-590.

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.

Severson RK, Stevens RG, Kaune WT, et al. 1988. Acute nonlymphocytic leukemia and residential exposure to power frequency magnetic fields. *Am J Epidemiol* 128:10-20.

Severson RK, Buckley JD, Woods WG, et al. 1993. Cigarette smoking and alcohol consumption by parents of children with acute myeloid leukemia: An analysis within morphological subgroups – a report from the Children's Cancer Group. *Cancer Epidemiol Biomarkers Prev* 2:433-439.

Simon SL, Till JE, Lloyd RD, Kerber RL, Thomas DC, Preston-Martin S, Lyon JL, Stevens W. 1995. The Utah Leukemia Case Control Study: Dosimetry Methodology and Results. *Health Physics* 68(4): 460-471.

Sgambati M, Linet M, Devesa S. 2001. Chronic lymphocytic leukemia: Epidemiological,

familial, and genetic aspects. In: Cheson B, ed. *Chronic Lymphocytic Leukemia*. 2nd ed. New York: Marcel Dekker, Inc., pp. 33-62.

Shimizu Y, Kato H, Schull W, et al. 1989. Studies of mortality of A-bomb survivors. Mortality, 1950-1985. Part 1. Comparison of the risk coefficients for site-specific cancer mortality based on DS86 and T65DR shielded kerma and organ doses. *Radiat Res* 118:502-524.

Shu X-O, Ross JA, Pendergrass TW, et al. 1996. Parental alcohol consumption, cigarette smoking, and risk of infant leukemia: A Children's Cancer Group study. *J Natl Cancer Inst* 88: 24-31.

Silverthorn DU. 2013a. Plasma and the Cellular Elements of Blood. In: *Human physiology: an integrated approach*, 6th Ed. (Silverthorn DU, Johnson BR, Ober WC, Garrison CW, Silverthorn AC. eds.). Glenview, IL: Pearson Academic, Inc.

Silverthorn DU. 2013b. Blood Cell Production. In: *Human physiology: an integrated approach*, 6th Ed. (Silverthorn DU, Johnson BR, Ober WC, Garrison CW, Silverthorn AC. eds.). Glenview, IL: Pearson Academic, Inc.

Sorahan T, Lancashire RJ, Temperton DH, et al. 1995. Childhood cancer and paternal exposure to ionizing radiation: A second report from the Oxford Survey of Childhood Cancers. *Am J Ind Med* 28:71-78.

Sorahan T, Prior P, Lancashire RJ, et al. 1997. Childhood cancer and parental use of tobacco: Deaths from 1971 to 1976. *Br J Cancer* 76:1525-1531.

Sorahan T, McKinney PA, Mann JR, et al. 2001. Childhood cancer and parental use of tobacco: Findings from the inter-regional epidemiological study of childhood cancer (IRESCC). *Br J Cancer* 84:141-146.

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.

Steinbuch M, Weinberg CR, Buckley JD, et al. 1999. Indoor residential radon exposure and risk of childhood acute myeloid leukemia. *Br J Cancer* 81:900-906.

Stevens W, Thomas DC, Lyon JL, Till JE, Kerber, RA, Simon SL, Lloyd RD, Elghany NA, Preston-Martin S. 1990. Leukemia in Utah and Radioactive Fallout from the Nevada Test Site: A Case-Control Study. *J Am Med Assoc* 264(5): 585-591.

Tango T. 1999. 8 Comparison of general tests for spatial clustering. 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.

Taylor GM, Birch JM. 1996. The hereditary basis of human leukemia in: Hendersen ES, Lister TA, Greaves MJ, eds. *Leukemia*. 6th ed. Philadelphia: WB Saunders, pp. 210-245.

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* 61:3-9.

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.

Travis LB, Curtis RE, Stovall M, et al. 1994. Risk of leukemia following treatment for non-Hodgkin's lymphoma. *J Natl Cancer Inst* 86:1450-1457.

Travis LB, Holowaty EJ, Bergfeldt K, et al. 1999. Risk of leukaemia after platinum-based chemotherapy for ovarian cancer. *N Eng J Med* 340:351-357.

UCR (Utah Cancer Registry). 2015. 2015 cancer dataset for the Utah Environmental Public Health Tracking Network: containing public use data records for primary in-situ Utah resident cancers from 1973 to 2013. Electronic data transfer. Information: <http://ucr.utah.edu> [accessed August 1, 2016].

UCR (Utah Cancer Registry). 2015. Utah Cancer Registry. Available: <http://ucr.utah.edu> [accessed July 20, 2016].

UEPHTN (Utah Environmental Public Health Tracking Network). 2016. Utah environmental public health tracking network. Information: <http://epht.health.utah.gov> [accessed August 1, 2016].

UK Childhood Cancer Study Investigators. 2002. The United Kingdom Childhood Cancer Study of exposure to domestic sources of ionizing radiation: 1: Radon gas. *Br J Cancer* 86: 1721-1726.

UNSCEAR (United Nations Scientific Committee on the Effects of Atomic Radiation). 1994. Sources and effects of ionizing radiation. United Nations Publ No E. 94. IX. 11. Annex A. Epidemiological evaluation of radiation-induced cancer. New York: United Nations Scientific Committee on the Effects of Atomic Radiation, pp. 297-431.

USCB (U.S. Census Bureau). 2002. Appendix A. Geographic terms and concepts in summary file 3: 2000 census population and housing, technical documentation. SF3/14 RV. Available: <http://www.census.gov/prod/cen2000/doc/sf3.pdf> [accessed August 12, 2016].

USCB (U.S. Census Bureau). 2012a. 2010 Geographic terms and concepts: census tract. Available: https://www.census.gov/geo/reference/gtc/gtc_ct.html [accessed August 12, 2016].

- USCB (U.S. Census Bureau). 2012b. Method for intercensal population estimates: 2000 to 2010. Available: http://www.census.gov/popest/methodology/2000-2010_Intercensal_Estimates_Methodology.pdf [accessed August 12, 2016].
- Van Duijn SM, Van Steensel-Moll HA, Coebergh JW, et al. 1994. Risk factors for childhood acute non-lymphocytic leukemia: An association with maternal alcohol consumption during pregnancy? *Cancer Epidemiol Biomarkers Prev* 3:457-460.
- 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.
- Vidabek A. 1947. Familial leukemia. A preliminary report. *Acta Med Scand* 127:26-52.
- 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.
- Wechsler J, Greene ME, McDevitt MA, et al. 2002. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nat Genet* 32:148-152.
- Weiss ES. 1965. Leukemia mortality in southwestern Utah 1950 to 1964. Washington, DC: US Public Health Service. Unpublished manuscript.
- 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.
- WHO (World Health Organization). 2012. International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3) website. Available: <http://www.who.int/classifications/icd/adaptations/oncology/en/> [accessed August 12, 2016].
- Williams RR, Horm JW. 1977. Association of cancer sites with tobacco and alcohol consumption and socioeconomic status of patients: Interview study from the Third National Cancer Survey. *J Natl Cancer Inst* 58:525-547.
- Wong O. 1987. An industry-wide mortality study of chemical workers occupationally

exposed to benzene. II. Dose response analyses. *Br J Ind Med* 44:382-395.

Yin S-N, Li G-L, Tain F-D, et al. 1987. Leukaemia in benzene workers: A retrospective cohort study. *Br J Ind Med* 44:124-128.

FIGURES AND TABLES

Figure 1. Hematopoiesis pathways (NCI 2013).

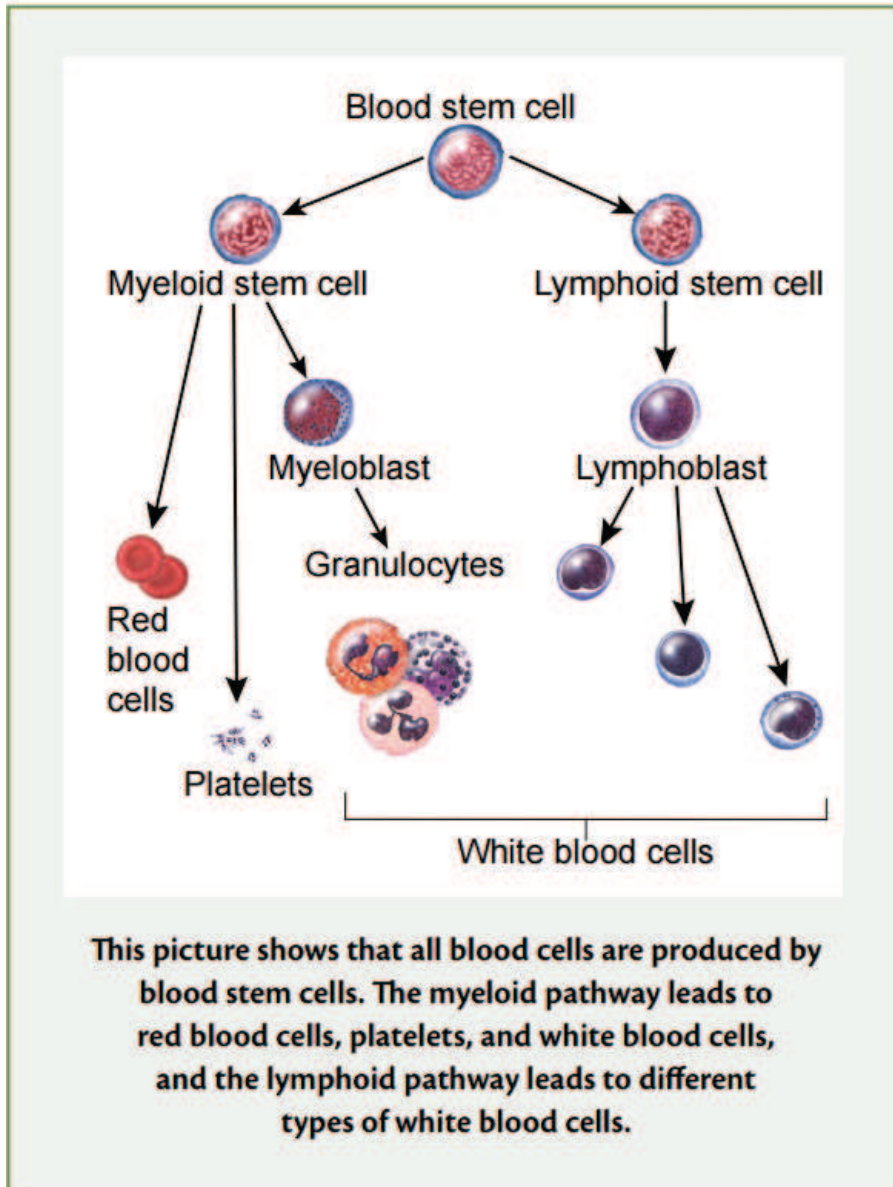


Figure 2. Distribution of primary leukemia cases (n = 6,725) by 5-year age group, Utah.

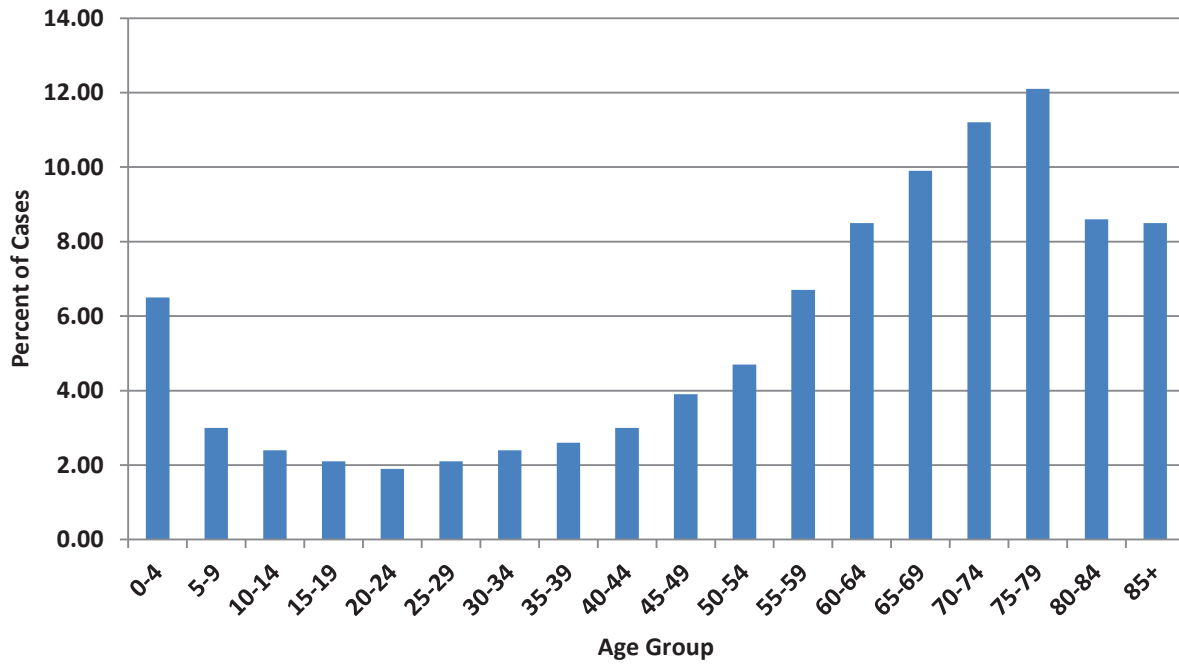


Figure 3. Standardized annual incidence rate of primary leukemia in Utah from 1975 to 2012. Figure was obtained from the Utah State Cancer Profile provided by the National Cancer Institute (NCI 2015b).

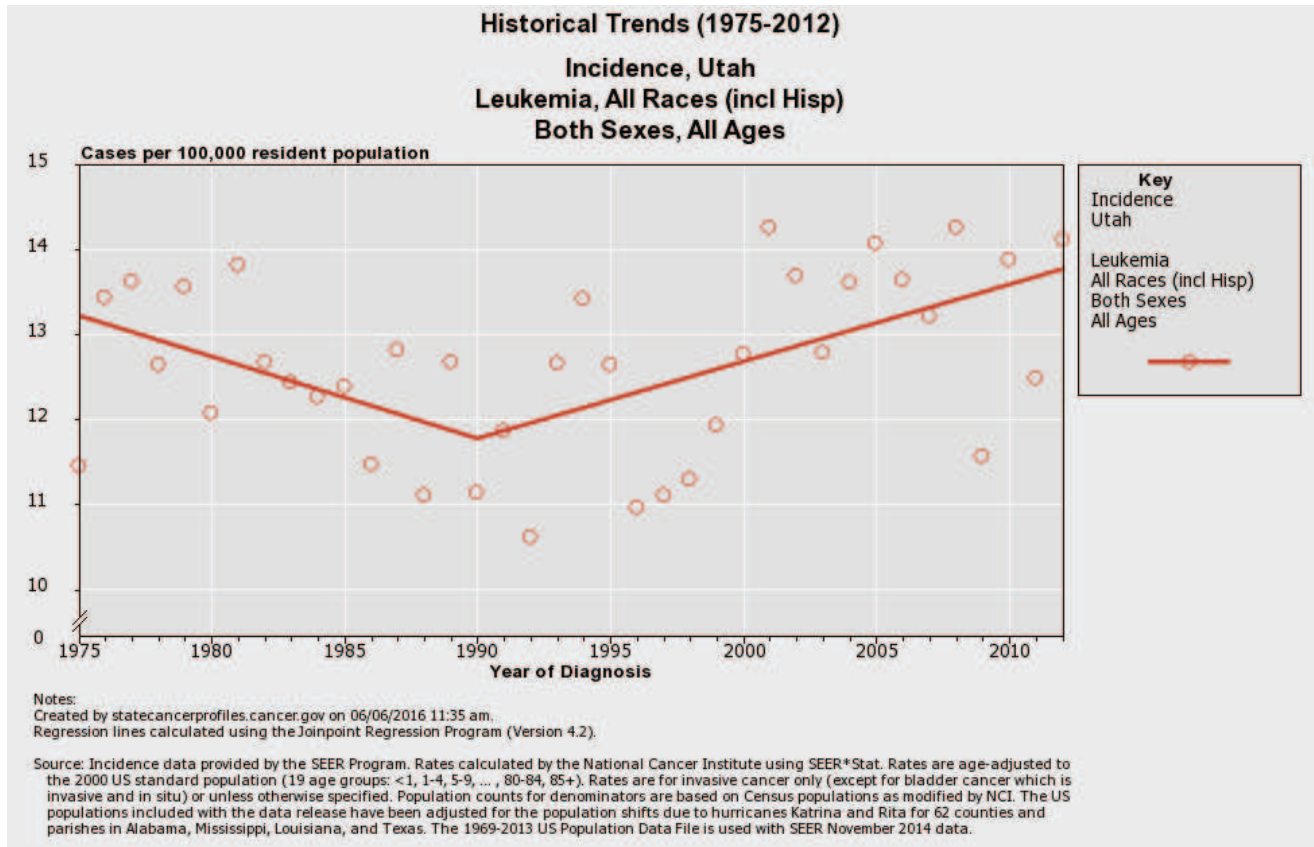


Table 1. UCR Site Codes for Leukemia

<u>Type</u>	<u>ICD-O-3 Code</u>	<u>SEER Site Recode Variable</u>
Acute Lymphocytic	M9826 M9835-M9837	35011
Chronic Lymphocytic	C420, C421, C424 M9823	35012
Other Lymphocytic	M9820, M9832-M9834 M9940	35013
Acute Myeloid	M9840, M9861, M9866, M9867, M9871-M9874, M9895-M9897, M9910,	35021
Chronic Myeloid	M9863, M9875, M9876, M9945, M9946	35022
Acute Monocytic	M9891	35031
Other Acute	M9801, M9805, M9931	35041
Aleukemic, subleukemic, and NOS	M9733, M9742, M9800, M9831, M9870, M9948, M9963, M9964	35043

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 more than 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>.
- 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/>.
- ArcGIS** A complete desktop GIS software application for producing maps and conducting spatial analysis. This application is developed and distributed by ESRI. EEP uses version 10.3. For more information see: <http://www.esri.com/software/arcgis>.
- CDC** Centers for Disease Control and Prevention. The CDC is a federal agency under the U.S. Department of Health and Human Services responsible for protection and promoting public health at the national level. For more information see: <http://www.cdc.gov/>.
- CI** Confidence Interval. Because there is some error in estimating a population parameter, and that error increases as the population size decreases, the confidence interval is used to indicate the reliability of the parameter estimate. The way a 95% confidence interval is interpreted along with the estimated parameter is that the measured value of the parameter is the reported value and one can be assured with 95% confidence (or 1 in 20 chances of being wrong) that the real parameter value is within the reported confidence interval.
- CIS** Carcinoma in-situ is an early form of cancer that is defined by the absence of an invasion of tumor cells into the surrounding tissue. Instead, the lesion is flat or follows the existing architecture of the organ. In this state, CIS seldom cause clinical symptoms sufficient to prompt the person with CIS to seek medical assistance and are generally undetected. CIS can progress to invasive tumors and are therefore considered a precursor or incipient form of cancer.
- EEP** Environmental Epidemiology Program. A program within the Bureau of Epidemiology, Division of Disease Control and Prevention, UDOH. The EEP was established in 1996 and is responsible for investigating diseases related to the environment. The program has two sections. One section conducts surveillance and data management activities including managing the UEPHTN. The other section conducts health hazards risk assessment, including cancer investigations. The program is staffed by personnel with experience and expertise in environmental epidemiology, environmental sciences, toxicology, statistics,

public health informatics and geomatics, and health education. For more information see: <http://health.utah.gov/enviroepi/>.

- ESRI ESRI is a leading developer and supplier of GIS software and geographically referenced data. ESRI is headquartered in Redlands, California. The EEP uses the ArcGIS software application developed by ESRI. For more information see: <http://www.esri.com>.
- GeoLytics GeoLytics is a commercial vendor of census and demographic data calibrated to the 2000 census boundaries. The EEP purchased 1970, 1980, 1990, 2000, and 2010 census data from GeoLytics to be the basis for estimating intercensal population counts for each of the 1,481 census block group boundaries in Utah. Population counts are aggregated into 5-year age groups for each sex. For more information see: <http://www.geolytics.com>.
- GIS Geographic Information Systems. A GIS includes computer software and geographically referenced data. The EEP uses ArcGIS as the computer software, and obtains data from ESRI or AGRC.
- ICD-O-3 International Classification of Disease - Oncology, 3rd Edition. The ICD-O-3 is one of a number of internationally established coding standards for coding site (topography) and histology (morphology) of neoplasms (cancers). For more information see: <http://www.who.int/classifications/icd/adaptations/oncology/en/>.
- LISA Local Indicator of Spatial Autocorrelation is a measure of how distinct adjacent small areas are to one another. When the disease rate value is used as the test criteria, a LISA is one method that can be used to identify cluster areas.
- MAUP Modifiable Areal Unit Problem is a source of statistical bias that can affect the results of statistical hypothesis testing. There are two concerns associated with MAUP. One is a scale problem associated with the number and size of arbitrary geographic units. The other is the zonation effect which is associated with the arrangement and the establishment of boundaries of the geographic units.
- NAACCR North American Association of Central Cancer Registries. NAACCR was established in 1987 as a collaborative professional organization for cancer registries, governmental agencies, and professional associations that work with cancer registries. All central cancer registries in the United States and Canada are members. The purpose of NAACCR is to promote standards and enhance the quality of cancer registry data. The NAACCR also promotes training, epidemiologic research, public health activities, and patient care improvement policies related to cancer. For more information see: <http://www.naacr.org>.
- 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/>.

- p-Value Probability value. A measure of probability of randomness. The range of the p-Value is between 0 and 1. This measure associated with a parameter is interpreted as the probability that the particular parameter's value could occur randomly. Thus a significant parameter value is one that has little random probability or a small p-Value. Typically a p-Value less than 0.05 is considered significant.
- R^2 Correlation Coefficient. The R^2 is a measure of the degree of agreement between two or more parameters. The R^2 value range is between negative one (-1) and positive one (+1). An R^2 value close to zero (0) means that considered parameters are uncorrelated. Their relationships are completely random. R^2 values close to negative one (-1) are inversely correlated meaning that as one parameter increases the other parameters decrease. R^2 values close to positive one (+1) are correlated meaning that the parameters increase or decrease together. The R^2 is significant when it is large (close to either negative or positive one).
- RR Relative Risk. The RR is a statistical measure used in epidemiology to quantify how the risk of an event (such as developing a disease) is related to the presence of a causal factor (e.g., exposure or spatial-temporal location). Relative risk is a ratio of the risk (or probability of the event occurring) among a target population, compared with a control population. If the risks of the two populations are equal then RR will equal 1.0. If the risk in the target population is greater than the comparison population then RR will be greater than 1.0 and the degree of increase is reflective of the magnitude of the increased risk in the target population. Thus an RR of 2.0 indicates that the target population has twice the risk compared to the comparison population. An RR less than 1.0 indicates that the target population has less risk than the comparison population. The ability of a risk estimate to quantify the true risk increases as the population increases. Thus it is common practice to report the estimated relative risk along with a 95% confidence interval.
- SAS SAS[®] (originally from “Statistical Analysis System”) is a globally-recognized system of integrated computer software products provided by SAS Institute Inc. The SAS system includes a large variety of data manipulation and statistical analysis processes. The EEP uses the desktop version 9.2. For more information see: <http://www.sas.com>.
- SaTScan SaTScan[™] (from “Space and Time Scan”) is an internationally recognized computer application that applies various space, time, or space-time scanning techniques to data that contains geographic and temporal locating variables. This tool is designed to aggregate data, test for disease, and determine the likelihood of a cluster. For more information see: <http://www.satscan.org/>.
- 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/>.

- SGID Utah State Geographic Information Database. The SGID is a central data warehouse of digital mapping information established by Utah Code 63F-1-507. The SGID contains a variety of state and state agency data used for epidemiologic investigations. For more information see: <http://gis.utah.gov/data/>.
- STFID Standard Federal Identifier. The STFID is a unique code for each census unit that can be used as a primary link key for tabular linking of GIS-enabled data. The code consists of a chained sequence of state (2 numerals), county (3 numerals), census tract (6 numerals), census block group (1 numeral) and census block (4 numerals) identifiers. The state and county identifiers are the state and county federal information processing standards (FIPS) codes. The state of Utah is 49, thus for all Utah STFID, the first two numerals are 49.
- 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 that cancer diagnoses be reported 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 provides cancer data to the EEP through the UEPHTN. For more information, see: <http://ucr.utah.edu/>.
- 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 exposure and hazards data, and supporting data. Data from the UCR and population data derived from the USCB is warehoused in the UEPHTN. For more information see: <http://epht.health.utah.gov/>.
- 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/>.
- UTM Universal Transverse Mercator. The UTM is a geographic coordinate system that uses a series of zones and the Cartesian x- and y-coordinates to represent the

location of a point on the Earth. The value of the UTM system is that the final component of the UTM x- and y-coordinates are given in meters as an offset from the zonal reference point. Thus it is easy to accurately measure the distance between two points within the same zone.

- The EEP uses the 1983 North American Datum. This refers to a geodetic network of reference points (typically a brace marker anchored in a concrete block) that were placed in 1983. This survey was conducted to update a 1927 survey using the latest satellite and remote sensing technology and has a much higher level of precision.
- The UTM system defines the Earth into 60 zones, each of which is 6° of longitude in width. The zone boundary provides a reference point for the east-west measurement. Utah is situated completely within Zone 12.

RESOURCES

Web links for websites may wrap onto multiple lines.

American Cancer Society

<http://www.cancer.org/cancer/leukemia/>

Huntsman Cancer Institute

<http://healthcare.utah.edu/huntsmancancerinstitute/patient-care/clinics-and-care-teams/hematology-cancer-program/>

Intermountain Healthcare Cancer Services

<http://intermountainhealthcare.org/services/cancer-care/>

Leukemia & Lymphoma Society

<http://www.lls.org/leukemia>

Mayo Clinic

<http://www.mayoclinic.org/diseasesconditions/leukemia/basics/definition/con-20024914>

National Cancer Institute

<http://www.cancer.gov/types/leukemia>

Utah Cancer Action Network

<http://www.ucan.cc/>

Utah Cancer Control Program

<http://cancerutah.org/>

Utah Cancer Specialists

<http://www.utahcancer.com/>