PROXIMITY OF CANCER CASES TO AIRPORTS WITHIN THE STATE OF TEXAS

by

SALA NANYANZI SENKAYI

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2012

ACKNOWLEDGEMENTS

I would like to thank the Texas Department of State Health Services for providing the cancer data in addition to their help with calculating the expected cancer rates. In addition, thanks to Drs. Barbara S. Glenn and Thomas F. Bateson of U.S. Environmental Protection Agency (EPA), National Center for Environmental Assessment (NCEA), for their guidance and advice in the selection and application of appropriate environmental epidemiological methods to this study. They helped me to obtain an understanding of the environmental data sets available through the National Center for Environmental Assessment, Research Triangle Park, Environmental Protection Agency (RTP-EPA). Special thanks go to Dr. Nancy Rowe with the University of Texas at Arlington (UTA) for her help with running the statistical analysis, Joshua Been (UTA) for his GIS advice, EPA Region 6 GIS Group, EPA Region 6 Library Contractor, and the Region 6 Regional Science Council for assistance in obtaining the Regional Research Partnership Program grant. I would also like to thank my family (especially my father, Dr. Abu L. Senkayi; my mother, Sunajeh N. Senkayi; and my brother, Ali K. Senkayi) and all my friends for their support and encouragement over the years.

Special thanks to my major professor, Dr. Sattler, for listening to my ideas and providing guidance in addition to her tremendous support over the years. Additionally, I would like to thank Dr. Grover for encouraging me to apply to Graduate School. I also wish to thank Drs. Andrew Hunt, Victoria Chen, and Subhrangsu Mandal for their helpful suggestions as members of my committee.

May 3, 2012

ABSTRACT

PROXIMITY OF CANCER CASES TO AIRPORTS
WITHIN THE STATE OF TEXAS

Sala Nanyanzi Senkayi, Ph.D.

The University of Texas at Arlington, 2012

Supervising Professor: Melanie L. Sattler

Determining whether the incidence of cancer may be associated with pollutants emitted from airports is important in order to better plan cities to reduce such incidences. This research aims to answer the following questions: 1) Is there a trend between the incidence of various kinds of cancer and proximity to air emission sources, including airports, in Texas? and 2) Specifically, is there a relationship between childhood leukemia and airport benzene emissions? Texas has two airports ranked among the top 10 U.S. airports in terms of enplanements for Jan.-March 2010: Houston ranked 7th, while Dallas was 3rd.[1] State-wide cancer incidence data was obtained in July of 2009[2]; additionally, state-wide data on age, race and gender-specific cancer rates[3,4,5,6] were obtained from the Texas Department of State Health Services (DSHS)[7]. These rates were applied to the demographic make-up of each geographical unit to compute the number of cancer cases which would be expected in each area. The observed number of cases in each area was compared to the expected number for leukemia, lymphoma (both for children 9 years and under), colon and respiratory (both for all ages) cancers at the state-wide

block group level. Colon cancer was used as a negative control because benzene exposures are not expected to be associated with the incidence of colon cancer. The ratio of the number of observed cases was divided by the number of expected cases for each block group and was then plotted against the distance to major emission sources (railroads, airports, industrial facilities and roads), using Geographical Information Systems (GIS). In order to address the second question of a relationship between childhood leukemia and airport benzene emissions, a Poisson regression model was developed using county emissions as the predictor variables and childhood leukemia as the response variable. Additionally, distance to the emission sources for children under age 9 with and without leukemia was compared. The 3 analyses all suggest that airport benzene emissions contribute to incidences of childhood leukemia in Texas.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Statement of the Problem

Texas is home to nearly 20.9 million residents,[8] making Texas the second most populated state in the Union (according to the US Census). Texas comes in fourth behind New York, Florida and California (which has the highest cancer rates)[9] in terms of total cancer cases within the United States (incidences per 1,000,000 population for both males and females) for 2001-2005. According to the "Impact of Cancer on Texas," documents (5th and 6th Editions released in 1991 and 1995 respectively), cancer is the second leading killer after heart disease within Texas (and the United States).[10,11] More recent publications continue to rank cancer second to heart disease in terms of deaths within the US.[12,13] Cancer not only has a grave effect on those who suffer from it, it also has a devastating impact on the friends and family members of those who are diagnosed with the disease. In addition, cancer places a large burden on the economy, "costing millions of dollars for health care and resulting in more than two billion dollars in lost earnings".[14]

More recent articles demonstrate an increased economic burden due to the disease. Figure 1.1 represents the cost of cancer care (in millions of dollars) within regions of the State as of 2007.[15]

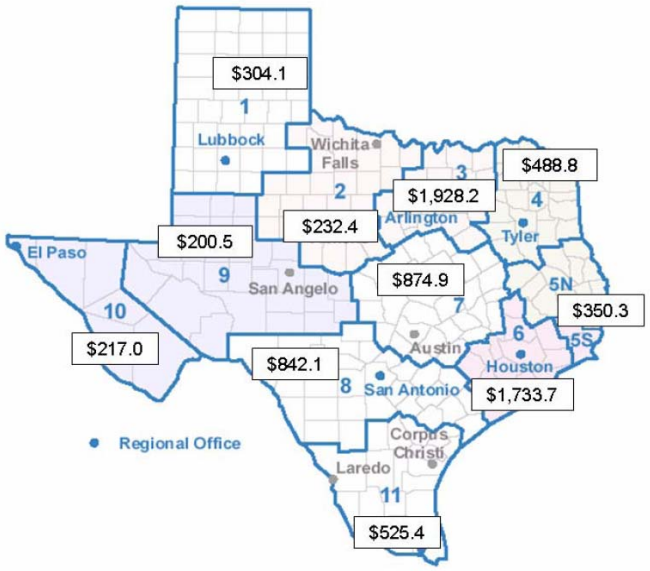


Figure 1.1. Cost of Cancer Care in Texas by HSR, 2007 (all figures are in millions of dollars)

Figure 1.2 shows the cost of cancer morbidity (in millions of dollars) within regions the State as of 2007.[16]

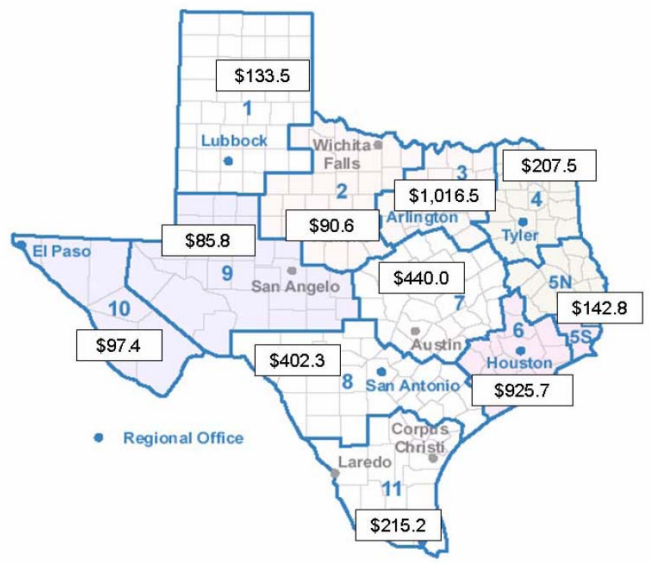


Figure 1.2. Cost of Cancer Morbidity in Texas by HSR, 2007 (all figures are in millions of dollars)

Finally Figure 1.3 provides a map of regions within the State that demonstrates the cost of cancer mortality (in millions of dollars) as of 2007.[17]

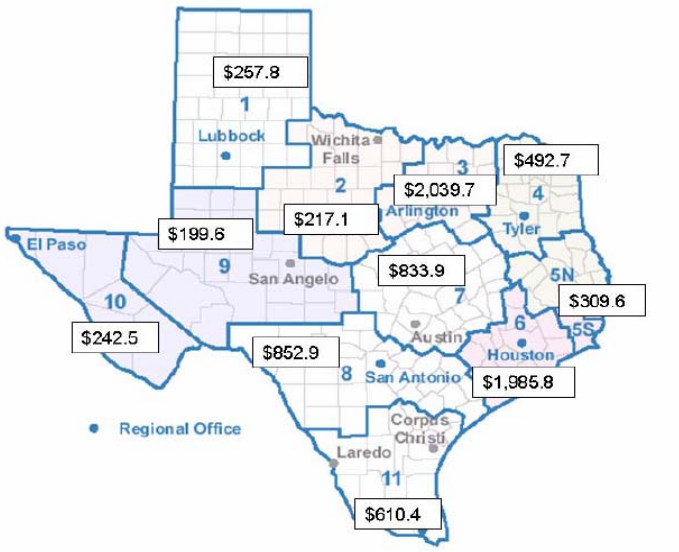


Figure 1.3. Cost of Cancer Mortality in Texas by HSR, 2007 (all figures are in millions of dollars)

Notice that all three maps clearly demonstrate that both the Dallas Fort Worth and Houston areas spend the most in terms of cost of cancer care, morbidity and mortality. These two areas are also the home of several relatively large airports that inevitably contribute to the air pollution within these locations. Texas has two airports ranked among the top 10 U.S. airports in terms of enplanements for Jan.-March 2010: Houston ranked 7th, while Dallas came in 3rd.[18]

Various studies have linked air pollutant emissions to cancer.[19] According to a presentation that was provided by Ted Palma of USEPA-QAQPS, benzene is the most significant carcinogen in the 2002 National-Scale Air Toxics Assessments (NATA) Pollutant Drivers Chart, as shown in Figure 1.4. A number of sources have linked benzene exposure to development of leukemia.[20,21,22] Airports are known sources of benzene emissions.[23,24,25]

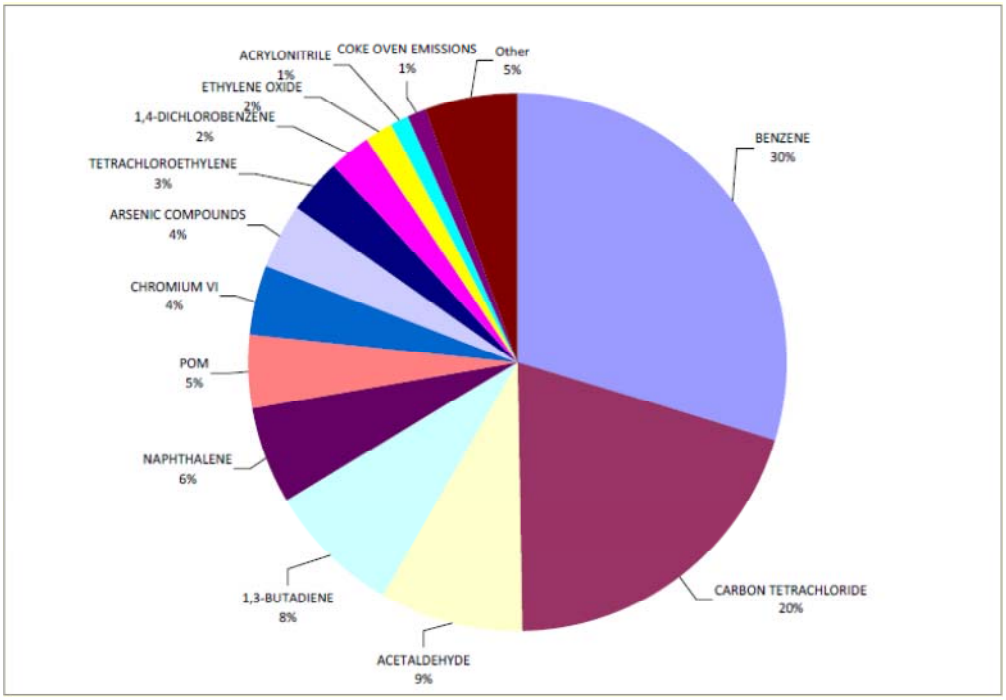


Figure 1.4. 2002 NATA Pollutant Drivers Chart via The US Air Toxic Program and Results from the National Evaluation of Air Toxics Presentation given by Ted Palma, August 2009.

.

1.2 Objectives of this Study

Determining if pollutants have an effect on public health in the form of cancer, and potentially minimizing the public's exposure to such pollutants, would be beneficial in order to lessen this economic and emotional burden suffered due to each newly diagnosed case of cancer. Specifically, determining whether cancer incidences may be associated with pollutants emitted from airports may lead to improved city planning or even infrastructure, where proximity to polluting facilities is better understood and accounted for.

This study thus aims to answer the following questions:

1. Is there a trend between the incidence of various kinds of cancer and proximity to air emission sources, including airports, in Texas?
2. Specifically, is there a relationship between childhood leukemia and airport emissions?

1.3 Dissertation Organization

The organization of the remaining dissertation chapters is as follows:

- Chapter 2 reviews previous literature concerning studies linking air emissions and cancer, studies linking benzene and leukemia, and studies of air pollution impacts of airports.
- Chapter 3 describes methods used to address the question “Is there a trend between incidence (or the observed incidence divided by the expected) of various kinds of cancer and proximity to air emission sources, including airports, in Texas?”, as well as results.
- Chapter 4 describes methods used to address the question “Specifically, is there a relationship between childhood leukemia and airport emissions?” as well as results.
- Chapter 5 summarizes important results and main conclusions associated with the research objectives. In addition, further studies and recommendations are suggested.

CHAPTER 2

LITERATURE REVIEW

2.1 Studies Linking Air Emissions and Cancer

Several studies have attempted to look at the effects of known air pollution[26] on cancer in general.

An article presented at the Mickey Leland National Urban Air Toxics Research Center in 2005 indicated that air pollution is one of the major pathways of humans being exposed to polycyclic aromatic hydrocarbons (PAHs) (which are made up of two or more benzene rings).[27,28]

A study entitled “geographic risk modeling of childhood cancer relative to county-level crops, hazardous air pollutants and population density characteristics in Texas” (2008) utilized geographic information systems (GIS) to link geographic factors with various diseases. The study found higher rates of germ cell tumors and "other" gliomas in areas of intense cropping (note these areas are known to use pesticides); higher rates of hepatic cancer near hazardous air pollutant (HAP) release facilities; and higher rates of Hodgkin lymphoma and malignant bone tumors in counties with rapidly growing population.[29] The study notes that there are conflicting results in reference to automobile exhaust and childhood cancer, in addition the report notes that a “critical review concluded that the weight of the epidemiological evidence indicates no increased risk for childhood cancer associated with exposure to traffic-related residential air pollution”.[30] This study, like many other studies related to cancer, focused on counties. County-level data maybe too broad to decipher minute trends associated with cancer incidences/rates and proximity to several polluting facilities.

A study titled “extended follow-up and spatial analysis of the American Cancer Society study of particulate air pollution and mortality” (2009) looked at several types of deaths for individuals and their relationship to pollutants. This study concluded that some areas (Los Angeles, for example) showed a relationship of particulate matter (PM) exposure and deaths, while other areas did not. The document hints that these “divergent results argue for caution in extrapolating from such studies in any one

metropolitan area to other areas".[31] A group of researchers lead by Dr. Daniel Krewski has conducted a very extensive follow-up study (2000), in which all causes of death, lung cancer and cardio-pulmonary disease (CPD) were looked at spatially (Los Angeles and New York City Intra-Urban Areas) in order to note any trends with PM2.5 and sulfur dioxide (SO_2). The follow-up study concluded that there may be an increased risk of lung cancer due to long-term exposure to PM2.5; however, timing of exposure could not be accounted for in this study.[32]

An investigation of the relationship between air emissions of volatile organic compounds (VOCs) and the incidence of cancer in Indiana counties (2006) is the last article linking air emissions and cancer that will be mentioned. This study utilized statistical linear regression modeling to address Toxic Release Inventory (TRI) emissions of VOC and county level incidence of some types of cancers and found correlations between VOCs and the incidence of some types of cancer. The most significant correlations in all three models were cancers of the brain and nervous system. Remaining significant correlations involve cancers almost exclusively related to the skin and endocrine systems (skin, melanoma, endocrine system, and thyroid cancers). The other cancers for which there were significant correlations in at least one model include urinary system, female genital system, lymphoma, leukemia, and oral cavity. Finally, cancers not significantly predicted by any TRI emission rates included pharynx, digestive, lung, respiratory, male genital, and breast cancers.

2.2 Studies Linking Benzene and Leukemia

Cancer among children is on the rise, and approximately one third of childhood cancers are leukemia, according to the National Cancer Institute. Benzene has been shown to trigger leukemia in young children, as noted by Dr. Barbara Glenn and by the US Department of Health and Human Services (DHH). Additionally, the World Health Organization (WHO) indicates that a lifetime exposure of 1 $\mu g/m^3$ will lead to an estimated six cases of leukemia per million inhabitants. A very interesting study was conducted in 2008 titled "Childhood Lymphohematopoietic Cancer Incidence and Hazardous Air Pollutants in Southeast Texas, 1995–2004," which documented that census tracts containing the highest levels of benzene also demonstrated higher rates of acute lymphocytic leukemia (ALL) and acute myeloid

leukemia (AML), which are the two most common types of leukemia in children. In addition, there was not a relationship between benzene or 1,3-butadiene and lymphoma incidences. The benzene and 1,3-butadiene concentrations were obtained via a U.S. EPA 1999 National-Scale Air Toxics Assessment (NATA) project. The project utilized a computer simulation model, the Assessment System for Population Exposure Nationwide (ASPEN), in order to estimate the pollutant levels for each census tract within the Unites States.

A study titled "Benzene and human health: A historical review and appraisal of associations with various diseases" indicates that there is not enough known about the actual measurement of different chemicals that form pollution. In addition, the author hypothesizes that not enough benzene is present, within the natural environment, to cause various diseases. In addition this article states that many studies have failed to demonstrate a relationship between traffic density and incidence of childhood cancers (note that in this study, childhood cancers refers to children under 5). The study recommends looking at genetic risk factors in addition to pollutants (within the environment) and they note that there have been some studies which have demonstrated a relationship between benzene and cancer while others studies do not.

2.3 Studies of Air Pollution Impacts of Airports

Based upon searches of various research databases, including the use of the Environmental Protection Agency's (EPA's) Library System Search, internet search engines, and personal discussions with experienced personnel and academia involved within the area of air pollution/air quality and cancer, it appears that very little research has been conducted to examine the health related impacts of airports revolving around a ten year timeframe for an entire state. However, the following studies have attempted to look at the air pollution impacts of airports. The study "personal, indoor, and outdoor exposure to VOCs in the immediate vicinity of a local airport" (2009) focused on measuring the ambient concentration of benzene, toluene, ethylbenzene, and *o*-, *m*-, and *p*-xylene (BTEX) at 15 homes located "close" to the airport. This study focused on Teterboro Airport, the oldest operating airport in the New York/New Jersey metropolitan area, which has also become one of the busiest local airports in the US. The study

demonstrated that individuals living near the airport are not at any greater risk of being exposed to benzene verses others within the entire US; however, personal benzene levels are higher than the World Health Organization (WHO) recommended lifetime levels of 1 $\mu g/m^3$.[33] The study also concluded that it is possible to measure airport activities within the "immediate vicinity"; however, these measurements were dependent on other variables such as wind. In addition, the study also suggested that further studies should be conducted in order to understand all of the relationships.[34]

A separate study (2000) conducted by ENVIRON International Corporation for the City of Chicago was based on estimated emissions of certain hazardous air pollutants (HAPs) that were provided by K.M. Chng. Dispersion modeling, exposure parameters and toxicity criteria were used in this study, which followed approaches provided by EPA. The goal of this study was to quantify the risks associated with the type of emissions generally found at airports. It is interesting to note that the emissions data used were lower than those found in the US EPA's National Emission Trends database (for VOC). Plus, not all of the pollutants typically associated with airports were included; in fact, pollutants such as acetaldehyde and naphthalene were not included in this risk evaluation. However, the list of pollutants was very extensive and did include pollutants such as benzene, 1,3-butadiene, formaldehyde, 7-PAH and others. A positive feature of this report is that a comprehensive list of sources of volatile organic compounds was included: aircraft; ground support service vehicles; motor vehicles on roadways, including traffic on airport roadways, queuing at the terminal curbsides and traffic in parking facilities; fuel storage and handling; and heating and a refrigeration plant. Finally, it is important to note that these risk estimates only include emissions from potential sources that are associated with only O'Hare Airport. Thus, based on the estimated emissions of the pollutants mentioned previously, the individual cancer risk due to operations at O'Hare Airport exceeded 10^{-6} within 1,000 square miles surrounding Chicago, Illinois O'Hare (ORD) airport, and the cancer risks exceed 1 in 100,000 (i.e., 1×10-5) over an area of approximately 40 square miles, assuming 70 years of exposure. Plus, the document quotes "the maximum hypothetical cancer risk at the airport property boundary is estimated to be approximately 1 in 10,000 (i.e., one in ten thousand, or 1×10^{-4}). These risk estimates do not include emissions from potential sources not associated with O'Hare Airport".[35]

Another study conducted in Southern California (2009) indicates that "Aircraft activity clearly results in markedly elevated ultra fine particulate (UFP) number concentrations".[36]

A study entitled "residential proximity to large airports and potential health impacts in New York State" (2000) was conducted in order to determine whether individuals residing near commercial airports demonstrated an increased rate of hospital admissions for respiratory conditions, verses individuals residing further from the airports. This particular study only showed a relationship with 2 of the 3 airports. The study indicates that other variables need to be accounted for, such as how many times one goes to the hospital or if sick persons go to a hospital at all.[37]

2.4 Importance of this Study

A number of studies, like those cited above, have linked air pollution and cancer incidence; and specifically benzene emissions and leukemia. Although a number of studies have looked at airport emissions, only one, focused on Chicago O'Hare, has investigated a potential link between airport emissions and cancer incidences. The Teterboro Airport study did look at benzene exposures surrounding the airport; however, this study had limitations such as finding the appropriate emissions data and trying to disaggregate or determine the exact source of the pollutants.

As mentioned in Chapter 1, Texas has two airports ranked among the top 10 U.S. airports in terms of enplanements: Houston (7^{th}) and Dallas (3^{rd}).[38] Even though invasive cancer incidence rates in Texas (with respect to leukemia from 1995 to 2005) have remained relatively the same (see Figures 2.1 and 2.2), Texas does rank high in the U.S. in terms of lymphocytic leukemia, as well as chronic myeloid leukemia, as shown in Figures 2.3 and 2.4.

Invasive Cancer Incidence Rates in Texas													
Leukemia, 1995-2005													
Year		**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**1995-2005**
Population at Risk		18958751	19340342	19740317	20157531	20558220	20945963	21332847	21710788	22057801	22418319	22798770	230019649
Total Cases		2012	2025	2169	2096	2301	2418	2656	2511	2663	2706	2690	26247
Crude Rate		10.6	10.5	11.0	10.4	11.2	11.5	12.4	11.6	12.1	12.1	11.8	11.4
Age-Adjusted Rate		12.4	12.3	12.8	12.1	13.0	13.5	14.5	13.4	13.9	13.8	13.3	13.2
95% Confidence Interval	Lower	11.8	11.7	12.2	11.6	12.5	12.9	13.9	12.8	13.4	13.3	12.8	13.0
	Upper	12.9	12.8	13.3	12.6	13.6	14.0	15.1	13.9	14.4	14.3	13.9	13.4
Note: All rates are per 100,000. Rates are age-adjusted to the 2000 U.S. Standard Population.													
Data accessed April 29, 2012. Cancer Incidence File, January 2012.													
Veterans Health Administration and military hospital reporting is incomplete for 2008-2009 Texas Cancer Registry (TCR) cancer cases. Therefore, case counts and incidence rates in 2008-2009 are underestimated and should be interpreted with caution.													

Figure 2.1. Invasive Cancer Incidence Rates in Texas via Texas Cancer Registry.

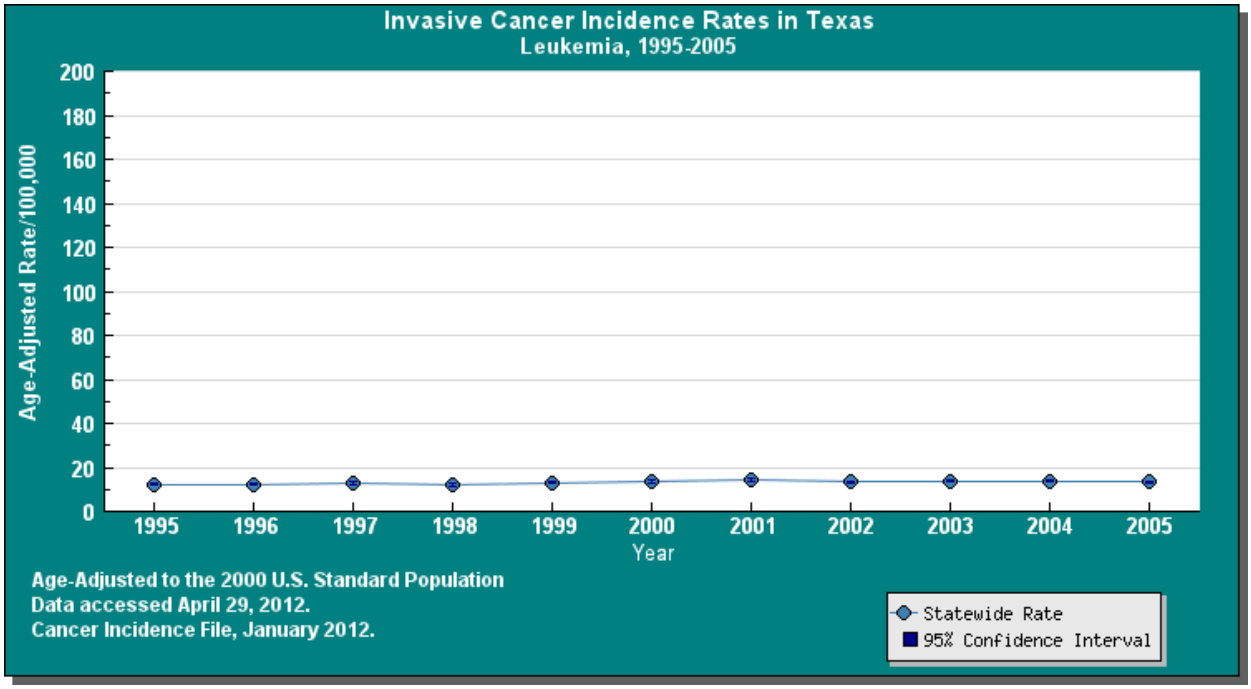
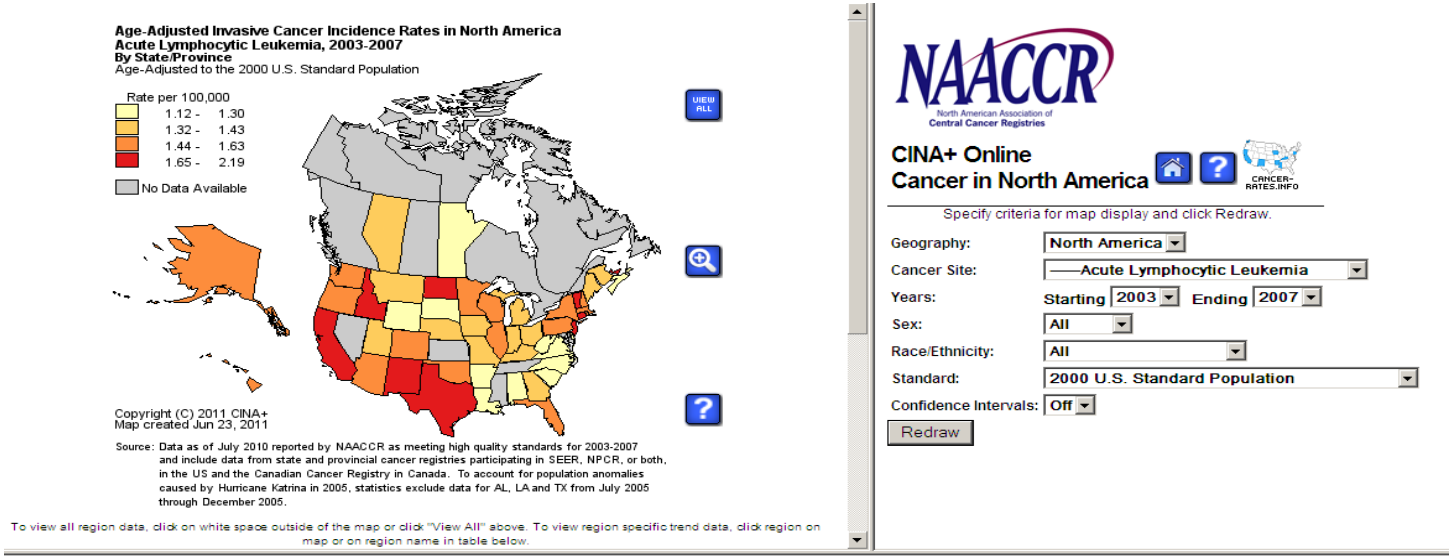


Figure 2.2. Invasive Cancer Incidence Rates Trend Graph in Texas via Texas Cancer Registry.



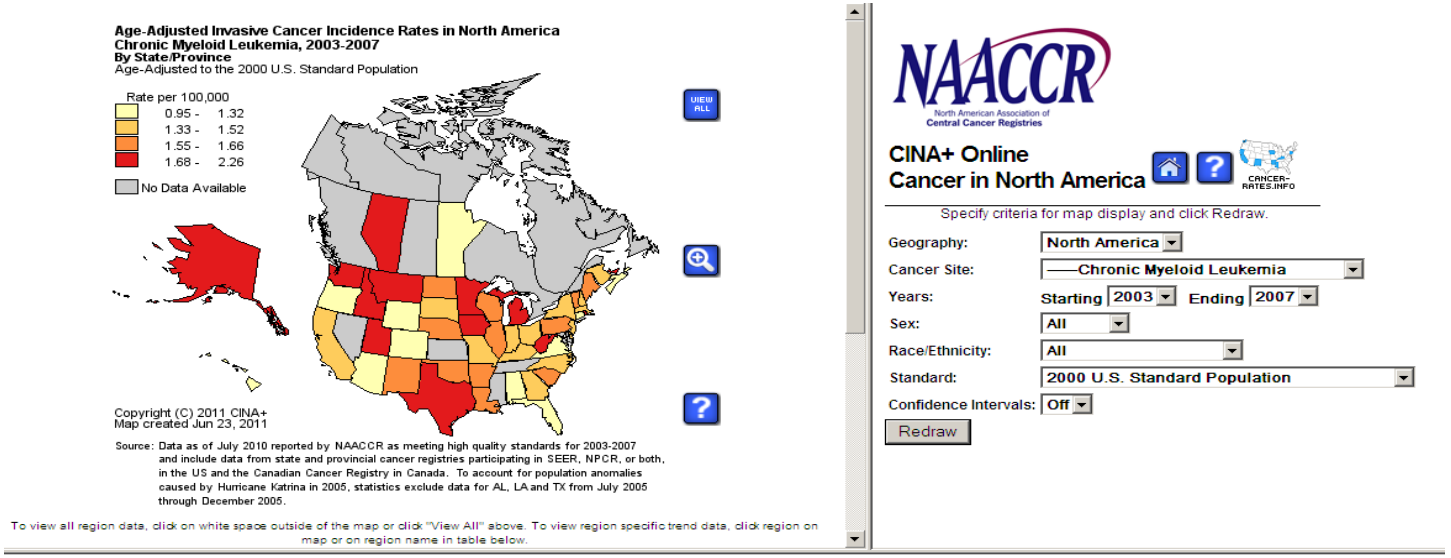
Click a column header to sort. Click the region name for more detailed information.
Sorting by State/Province will sort the bar graph by State/Province, otherwise the bar graph will be sorted by age-adjusted rate.

Invasive Cancer Incidence Rates by State/Province in North America

Acute Lymphocytic Leukemia, 2003-2007

State/Province	Population at Risk	Cases	Crude Rate	Age-adjusted Rate
Vermont	3095360	60	1.94	2.19
North Dakota	3178464	57	1.79	1.84
New Mexico	9572377	178	1.86	1.83
Prince Edward Island	688988	11	1.60	1.81
California	179321309	3247	1.81	1.80
New Jersey	43138333	716	1.66	1.70
Connecticut	17399761	280	1.61	1.66
Idaho	7134794	122	1.71	1.66
Texas	103103525	1766	1.71	1.65
Washington	31354466	494	1.58	1.63
Hawaii	6306953	101	1.60	1.63
Rhode Island	5318963	78	1.47	1.57
Wisconsin	27689353	414	1.50	1.56
Pennsylvania	61813165	924	1.49	1.56

Figure 2.3. Map of Acute Lymphocytic Leukemia Sites (In 2003 – 2007) via the North American Association of Central Cancer Registries (for some US States and Canadian Provinces).



Click a column header to sort. Click the region name for more detailed information.
Sorting by State/Province will sort the bar graph by State/Province, otherwise the bar graph will be sorted by age-adjusted rate.

Invasive Cancer Incidence Rates by State/Province in North America

Chronic Myeloid Leukemia, 2003-2007

State/Province	Population at Risk	Cases	Crude Rate	Age-adjusted Rate
Rhode Island	5318963	133	2.50	2.26
Minnesota	25555106	530	2.07	2.04
Idaho	7134794	134	1.88	1.97
Montana	4679663	96	2.05	1.88
Prince Edward Island	688988	15	2.18	1.88
Iowa	14778551	316	2.14	1.88
Alaska	3337438	46	1.38	1.85
West Virginia	9026215	185	2.05	1.76
Washington	31354466	545	1.74	1.76
Michigan	50383095	906	1.80	1.76
Texas	103103525	1599	1.55	1.76
Alberta	16677212	266	1.59	1.73
Utah	12575656	164	1.30	1.68
Illinois	63566310	1046	1.65	1.66

Figure 2.4. Map of Chronic Myeloid Leukemia Sites (In 2003 – 2007) via the North American Association of Central Cancer Registries (for some US States and Canadian Provinces).

Airports are known sources of benzene emissions.[39,40,41] Given that benzene has been linked to leukemia, that airports are significant sources of benzene emissions, and that Texas has several large airports, a study to determine whether a link between airports and leukemia exists in Texas would be valuable. Focusing on childhood leukemia in particular can eliminate several confounding factors, including smoking and workplace exposure, as discussed in later sections.

This study thus aims to answer the following questions:

1. Is there a trend between the ratio of observed to expected incidence of various kinds of cancer and proximity to air emission sources, including airports, in Texas?
2. Specifically, is there a relationship between childhood leukemia and airport emissions?

Chapter 3 will address question 1, and Chapter 4 will address question 2.

CHAPTER 3

METHODOLOGY AND RESULTS FOR QUESTION 1: IS THERE A TREND BETWEEN OBSERVED CANCER CASES OR THE OBSERVED TO EXPECTED RATIOS OF VARIOUS KINDS OF CANCER AND PROXIMITY TO AIR EMISSION SOURCES, INCLUDING AIRPORTS, IN TEXAS?

Methodology to address this question included 2 steps:

A) Calculate the observed to expected ratio for block groups state-wide for various types of cancer (leukemia, lymphoma, colon, and respiratory),

B) Plot the observed to expected ratio for each block group vs. distance to major emission sources (railroads, airports, industrial facilities, roads), using GIS.

Each of these steps will be discussed in turn.

3.1 Calculate the observed to expected ratio for block groups state-wide for various types of cancer (leukemia, lymphoma, colon, and respiratory)

Observed to expected (O/E) ratio can be calculated as follows:

$$\text{O/E ratio} = \frac{\text{Observed number of cancer cases per population in a given geographic area, per time period of interest}}{\text{Expected number of cancer cases per population in a given geographic area, per time period of interest}}$$

Data used to calculate observed to expected ratio is discussed below.

3.1.1 Obtain block group and census tract data

A shape file containing GIS information for Texas was obtained from SimplyMap.[42] Texas county spatial data was obtained via the EPA.

Block group and census tract shape files for the state of Texas[43] were obtained from Environmental Systems Research Institute also known as ESRI (this data was uploaded into GIS using the WSG 84 projection). This initial block group shape file included 14,463 block groups, and the census tract shape file consisted of 4388 census tracts. Block group spatial information coupled with demographic information was later obtained from EPA in order to perform the observed to expected calculations. It is important to note that the ESRI block group and census tract shape files are compatible with GIS. Since the ESRI shapefiles were already loaded into GIS early in the project, all that was needed was to join the EPA demographic data (for the year 2000) to this file after performing the observed to expected calculations. Therefore, instead of uploading the EPA block group shape files with the demographic data already attached, the demographics data (race, gender and age) was simply extracted for each of the block group area and later matched with the ESRI block groups already in GIS. Using the EPA demographics data and just matching it with the ESRI shape files also eliminated having to reproject this data shape file to match that of the GIS map already in use, as well as having to request demographic data from DSHS.

When performing the observed to expected ratio calculations, individual population counts were needed for age groups (grouped in 5 year sets), for both male and females within each of the three race categories (Black, Hispanic and White). In other words, population demographic data was needed in order to calculate the observed to expected ratios (as already explained); thus, the census population demographics data was obtained and matched to each of the 14,463 block groups already uploaded into GIS (again these ESRI shape files are files with spatial information attached making it easy to upload them into GIS). It is important to note that the 2000 block group polygon GIS shape file boundaries do not match the boundaries of the 2000 census tract or county polygon shape files; therefore, the following crucial steps were performed. The 2000 block group polygon GIS shape file was converted into points using a tool that converts features (such as these block group polygons) into centroid points (a point

within the middle of each of the block group polygons in order to assist in avoiding point overlap). Additionally, this new layer, consisting of both the total cancer cases of leukemia, lymphoma, respiratory and colon cases along with the demographic data (that had already been adjusted) for those respective groups, was aggregated into the census tract and county polygon layers. These census tract and county polygon layers now contained the observed cancer data along with the expected calculations that could then be divided in order to obtain the observed to expected ratios for those larger areas.

3.1.2 Obtain observed cancer incidence data

3.1.2.1 Cancer observed incidence data: strategies for addressing uncertainty in latency period and confounding factors.[44]

Carcinogenesis is a complex multistage process. Each stage of development may involve different lengths of time and be caused by a different set of etiologic factors. Most cancer types are diagnosed several years after they were initiated, promoted and then matured to a stage where they come to the attention of the individual or their doctor (e.g., lung, breast, prostate). In other words, these cancer types have a long latency. The relevant exposures that may have caused the initiation of the cancer were those that occurred prior to the initiation. Environmental contaminants also may cause the progression of a cancer to the next stage of development but prior to detection.

Therefore, models of residential proximity or emissions that exist today are not necessarily relevant to cancer diagnosed in recent years, unless the researcher can establish that the exposure conditions were the same during the relevant period when cancers were initiated or progressed through the stages. One strategy to minimize the uncertainty in considering the length of a latency period is to pick a cancer type known to have a shorter latency, such as lymphohematopoietic cancer. This would allow for better temporal targeting of the etiologically relevant window of time during which exposure may have caused a case of cancer. There are also several subtypes of lymphohematopoietic cancer and the epidemiologic literature contains many studies that suggest associations with specific hazardous substances for some of the subtypes.

Another approach to address the uncertainty in cancer latency periods is to restrict the analysis to childhood cases, which would very clearly define the maximum length of latency. For children less than 10 years old with leukemia diagnosed during 1995-2005, their etiologically relevant window of time would likely be from 1985 to 2000, assuming 5-10 year latency.

Limiting analysis to childhood cancers can also eliminate confounding factors, such as smoking and occupational exposure, associated with adults' exposure history. Exposure history linked to other locations (including moving from another city/state or country after having lived there for 50 years or so) is another confounding factor for adults, which is less of an issue for children.[45]

3.1.2.2 Cancer observed incidence data used in this study

The Department of Social and Health Services (DSHS) provided 10 years of cancer data (from 1995 – 2005) for Texas (visit the Data Dictionary to learn more about the provided cancer data).[46] The database included 925,781 total incidences of all types of cancer. This data provided the observed cancer incidences per area per time and can be obtained via a request to DSHS.[47]

Incidences of leukemia, lymphoma, respiratory cancer, and colon cancer were pulled from the database and were selected for this analysis for the following reasons. Since leukemia and lymphoma in children represent specific types of lymphohematopoietic cancers and have a relatively short latency period, it is acceptable to assume that current airport emissions and location are representative of the etiologically relevant exposure period. Respiratory cancer has been associated with air pollutants; however, according to the American Lung Association smoking is responsible for approximately 90% of lung cancer deaths. Colon cancer is not felt to be associated with air pollutants and is a test of the GIS methodology to assess cancer associations. This concept is important since a link between air pollutants and cancer is more likely to be able to be identified for a specific type of cancer, such as leukemia, and for a specific pollutant, such as benzene. In addition, aggregating exposures (i.e. the combined output of NATA emissions in pounds) will tend to blur any true causal effects of one chemical if other chemicals that do not play a role in cancer development are included. Since different chemicals cause different cancers, specificity is important.

The cancer incidence data contained the latitude and longitude (lats/longs) for each incidence, which were used to upload this shape file into GIS. The cancer data was then "intersected" with the State of Texas shape file in order to eliminate cancer indices that fell outside of the state; in addition, those indices with inconclusive lat/longs were also eliminated (for example, there were quite a few with lat/longs of 0/0). This task deleted about 200 points, leaving 925,510 incidences, which were utilized for the remainder of the project. The cancer leukemia density map, with cancer density shown in white in Figure 3.1, was overlain by an airport shape file obtained from EPA and consisting of 1966 airport points, shown in pink in Figure 3.1. Note (as documented earlier), the two largest airports are located in areas with the greatest leukemia density, which are roughly within the Houston and DFW areas.

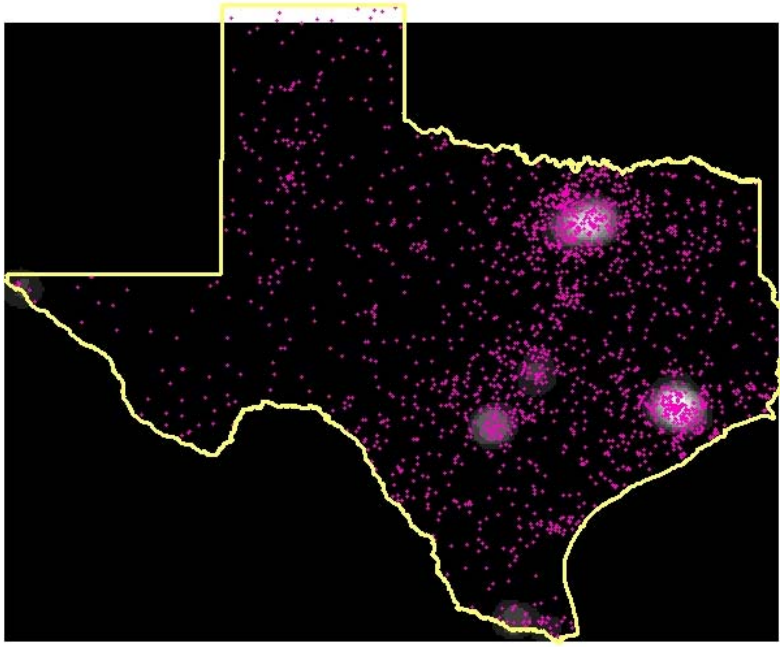


Figure 3.1. Map of Airports via EPA Dataset & Leukemia Cancer Density

The block group shape file was joined with the cancer shape file to generate a new count column representing all the aggregate cancer incidences within each block group. Since greater numbers of cancer cases are expected to be found in population centers, cancer cases in a given geographic area

are typically divided by the population of the area to provide an idea of how many incidences occurred within these areas for that timeframe (number of cases/population).

Block group population numbers range from very small to very large populations. It is important to note that migration patterns of individuals within these block groups cannot be accurately tracked for this study. Although epidemiological studies can be performed which take into account movement patterns, this kind of information was not available for this study. It is also important to note that these types of population movements will have a great effect on the observed to expected ratio for areas with a tiny population. Basically, if 1 incidence occurs in a population of 4 total individuals residing within a particular block group, then according to this type of calculation, one would expect the crude observed to expected ratio to be equal to the incidence (1 for this example) divided by the total population of 4. Thus, this cancer incidence has occurred 25% of the time within this example block group population, for a given year. However, this simple formula does not account for the rate at which individuals of different age, race and gender would acquire this type of disease within a given population; therefore, demonstrating the importance of adjusting the total population by the age, race and gender rates for each block group. This entails further dividing the example incidence of 1 by a total population that has already been adjusted by age, race and gender for a given time period. This is done by simply multiplying the total population by a rate adjusted for a specific age/race/gender of an area per 1,000,000 childhood population or 100,000 total population per time (please see the example provided in Section 4.1.2). Now imagine if that individual moved from this block group to another before the Census 2000 Population Data had been collected, or this individual may have just passed away. In reference to the previous example, this would mean that there is a cancer incidence of one; however, the population is now just made up of 3 (or fewer) individuals rather than the original 4. In addition, if this individual were to have gotten several different kinds of cancers (which happens), then it's feasible that 25% can quickly increase to even larger numbers. Because of the above example, the data demonstrated that areas with populations smaller than 1,000,000 children or 100,000 total population had higher observed to expected ratio (O/E) outcomes due to the lack of a large enough population to account for performing these calculations. The Department of State Health Services (DSHS) performs their expected calculations using counties with large

populations,[48] a list of these counties can be found in Appendix B. Figure 3.2 below demonstrates the Age-Adjusted Invasive Cancer Incidence Rates in Texas by county, rate per population of 100,000. Note that the counties with “Risk Population less than 1000” are highlighted in white to indicate that a rate per 100,000 could not be accurately provided at this time. For this project, observed to expected calculations were performed at block group levels, which are much smaller than census tracts and counties; therefore, we obtained some very large observed to expected ratio numbers. However these observed to expected calculations were very useful in that we were able to notice minute relationships between higher observed to expected ratios and visibly closer distances to certain emitters (see Appendix B for all scatter plots and maps of the observed to expected ratios in addition to Figures 3.6 to 3.21 of the observed to expected scatter plots). Observed to expected verses distance to emitter plots were made at the census tract and county levels as well (also located in Appendix B). It is important to note that there were a lot of zeros calculated for these observed to expected ratios due to the many block groups with no diagnosed cancer cases.

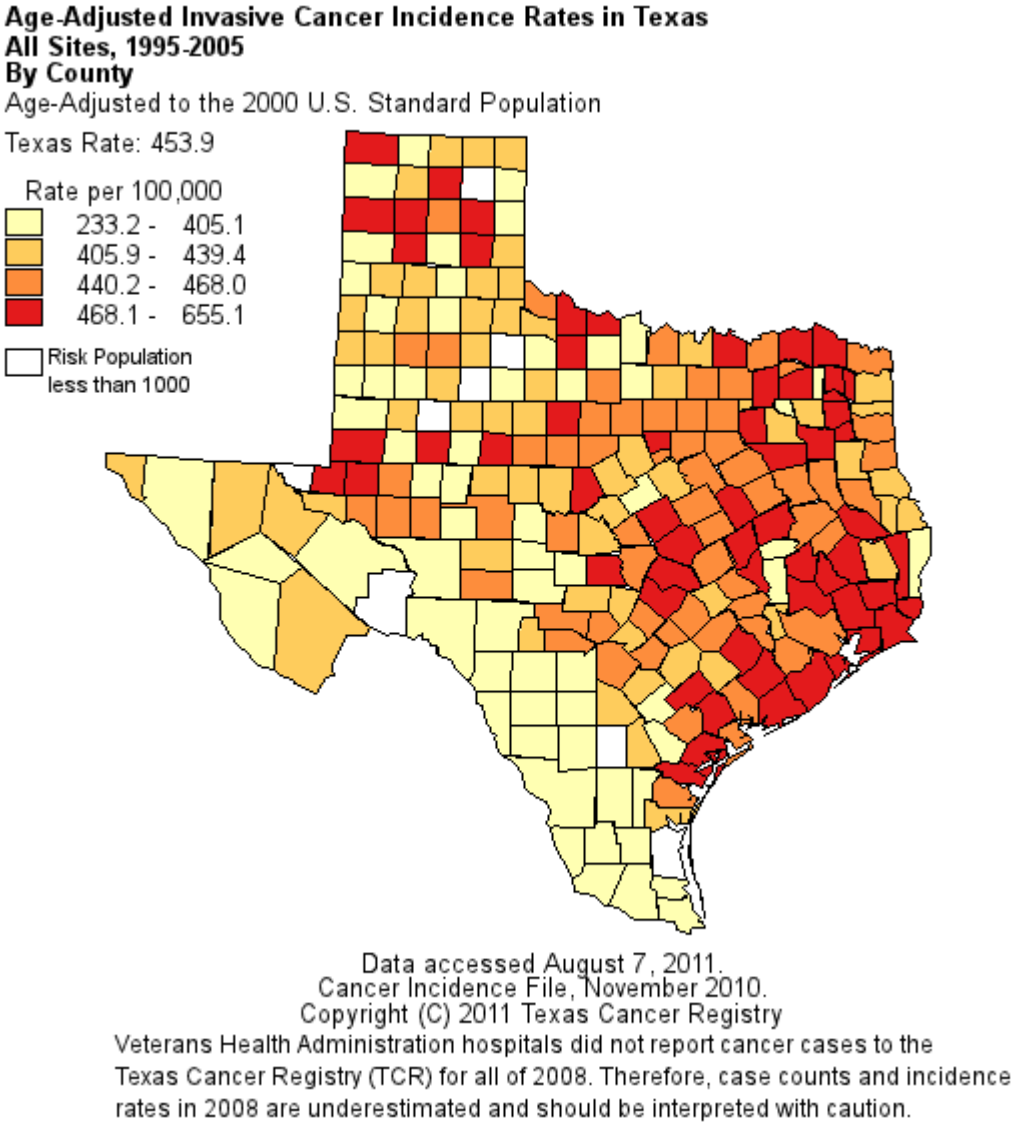


Figure 3.2. Map of Age-Adjusted (to the 2000 US Standard Population) Invasive Cancer Incidence Rates in Texas (All Sites 1995 – 2005) by county rate per population of 100,000[49]

Additionally, due to reasons explained in the previous paragraph, it was interesting to note that some block group areas had more cancer incidences than the total population, since it is very possible that one individual could get cancer repeatedly within several years. The cancer data spans 10 years, while the census data was only for the year 2000, thus demonstrating the importance of multiplying the Census 2000 Population Data by 10 in order to account for all 10 years of the provided cancer data time

period. In summary, if persons develop more than one kind of cancer during the 10 year period, the incidences could be higher than the total population for the year of 2000, as was demonstrated with this set of cancer incidence data. Therefore, multiplying the total population by 10 was very essential in order to account for the 10 year timeframe of when these cancers occurred; thus, all observed to expected calculations performed within this project included multiplying the total population by 10 (please see Figure 3.4 with respect to the observed to expected ratio calculations).

3.1.3 Calculate expected incidences

Expected cancer incidences are cancer incidences that are typically common for a given area during a defined length of time. In other words, the expected cancer cases (calculated using historical cancer data) are essentially estimates of anticipated cancer incidences; observed incidences greater than or less than an expected value would indicate a deviation from the norm. However, needless to say, observed cases equaling expected cases will yield a 1 to 1 ratio (or observed to expected ratio of 1 to 1), thus signaling that there was neither an increase nor decrease in new cases. As previously noted, expected cancer incidences are calculated using historical cancer incidence data thus resembling benchmarks; therefore, dividing the known, observed cancer incidences by what is expected will produce an observed to expected ratio that accounts for these historically cancer cases. In order to calculate the expected cancer incidences, the Census 2000 Population Data (from the previous sections) was multiplied by rates specific for the differing age, race and gender groups (as expressed earlier), thus adjusting for those inherent differences. The rates used within this study were obtained by the Texas Department of State Health Services (DSHS) and were for years between 1995 - 2000.[50,51,52,53] Rates can be found in Appendix A. In addition, an example of the leukemia rates used can be found in Figure 3.3 below. The Center for Disease Control documents how these rates are obtained.[54]

Childhood Leukemia Cancer Rates per 1,000,000
u:\Datareq\2010\10378tbl2.xls

Texas statewide Age-Specific Rates, Average Annual Rates, 1995-2000,
Selected Age Groups, and by race and ethnicty and sex (Data Request # 10378)
All Chioldhood Leukemias combined (ICCC-3 Group I)
Rates are per
1,000,000

	All Races		White Non-Hispanic		Black		Hispanic		Other Race	
	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate
00-09 years	69.6	51.7	66.0	51.2	27.7	27.7	86.2	59.1	69.7	51.0
00-04 years	91.3	71.9	90.0	74.4	41.9	40.1	107.4	77.4	70.4	67.5
05-09 years	47.5	31.3	43.1	29.2	14.4	16.3	62.5	38.9	68.9	33.2

*Rates are per 1,000,000.

Prepared by the Texas Department of State Health Services, Cancer Epidemiology and
Surveillance Branch, Texas Cancer Registry. Data Request # 10378 11/5/2010
Incidence Source: Texas Department of State Health Services, Cancer Epidemiology
and
Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09,
SEER*Prep 2.4.3, .

Figure 3.3. Childhood Leukemia Cancer Rates per 1,000,000 provided by the TX DSHS

3.1.4 Calculate observed to expected ratio

Observed to expected ratio is calculated using the following ratio (formula provided by Office of Research and Development Environmental Epidemiologists Drs. Barbara Glenn and Thomas Bateson):

$$\text{O/E ratio} = \frac{\text{Observed number of cancer cases per population in a given geographic area, per time period of interest}}{\text{Expected number of cancer cases per population in a given geographic area, per time period of interest}}$$

As expressed previously, the aggregate observed cases were simply divided by the aggregate expected cases (which had already been adjusted for age, race and gender) for each block group. This formula for calculating either Specific Relative Risk (the observed incidences for each age, race and gender group divided by the expected incidences that had been adjusted for each age, race and gender group) or the Standardized Relative Risk (the observed incidences for the different age, race and gender groups aggregated for each area divided by the expected incidences for the different age, race and gender groups aggregated for each area) was provided by Drs. Barbara Glenn and Thomas Bateson (Environmental Epidemiologists) of Office of Research and Development (ORD). Figure 3.4 demonstrates the spreadsheet used to calculate the expected incidences and ultimately the observed to expected ratios that were used in the scatter plots provided in Figures 3.6 to 3.21.

Relative Risk Calculations Provided by the Office of Research and Development (Drs Glenn & Bateson)

Calculation of relative risk as the ratio of Observed cases to Expected cases adjusting for Gender, Race and Age

Cancer	Zip/tract/ county	Gender	Race	Age	# in 2000 Census * 10	Gender, Race, and Age-specific Rate	Expected	Observed	Sum of Expected per Geo. Unit	Sum of Observed per Geo. Unit	Gender, Race, and Age-specific RR RR=O/E	Gender, Race, and Age-standardized RR RR=O/E
Leukemia	1	Female	W	0-9	Find these #'s	Use TX or CA as needed	=F5*G5		=sum(H5:H11)	=sum(I5:I11)	=I5/H5	=K5/J5
Leukemia	1	Female	B	0-9								
Leukemia	1	Female	H	0-9								
Leukemia	1	Male	W	0-9								
Leukemia	1	Male	B	0-9								
Leukemia	1	Male	H	0-9								
Leukemia	2	Female	W	0-9	ETC.....							

Figure 3.4. Relative Risk (also referred to as the Observed to Expected Ratio) Calculation Process provided by Drs. Barbara Glenn and Thomas Bateson of National Center for Environmental Assessment (NCEA) the Office of Research and Development, Environmental Protection Agency

Observed to expected ratios were calculated for each category shown in Table 3.1 below, for each block group within the State, using the General Relative Risk Calculation spreadsheet. Since colon cancer would not be expected to be impacted by airport emissions of benzene, it was used as a negative control.

Table 3.1. Observed to Expected Ratio Categories

Type of Cancer	Races	Age	Gender
Leukemia	Black, White and Hispanics	Ages 9 and under	Males and females
Lymphoma	Black, White and Hispanics	Ages 9 and under	Males and females
Respiratory	Black, White and Hispanics	All ages	Males and females
Colon	Black, White and Hispanics	All ages	Males and females

3.2 Plot the observed to expected ratio for each block group vs. distance to major emission sources (railroads, airports, industrial facilities, roads), using GIS

Sources of data for the major emission sources are shown in Table 3.2 below.

Table 3.2. Sources of data for major emission sources

Shape File		Source	Number (ex: points, segments, polygons, or items)
Airport benzene emissions		EPA	1966 points
Toxic Release Inventory (TRI) Facilities – benzene emissions		EPA	247 points
Roads	GIS -(Interstate, Major Road, State Highway and US Highways only)	EPA	25,131 segments
	Statistics: County Level Benzene Emissions	EPA	254 items
Railroads	GIS	EPA	47,287 segments
	Statistics: County Level Benzene Emissions	EPA	254 items

An example of the airport emission data provided via EPA is shown within Table 3.3. The top 50 Texas airports (in order from most benzene emissions to least) are presented in this table. The top 2 (as

expressed previously) are George Bush (IAH) with 7.4 tons/year and Dallas/Fort Worth (DFW) with 6.09 tons/year. No other airport comes close to these emissions. It is interesting to note that quite a few of the top 10 airports are located within the DFW or Houston areas. In addition, all other airports (starting with number 11 and greater) are under 0.9 tons/year and nowhere near the over 6 tons/year seen by the two leaders. The importance of using the EPA airport emissions data is that it already had emissions in tons/year associated with each airport (as demonstrated in the table). In addition, the EPA data included far more airports (1966 total points for hospital airports, farm airports, etc.) than those provided by Texas Department of Transportation (aka TxDOT which just had 387 points total); thus, airports such as those used for careflights were also analyzed. Using this file inevitably increased the validity of this project, thus demonstrating one of many major advantages of working with the EPA. The EPA airport, facility, road and railroad data will be used in the all of the following analysis: GIS Observed to Expected Ratios, Statistical and GIS Distance to Analyses.

Table 3.3. Top 50 Texas Airport Benzene Emissions (tons/year, in order from largest to smallest)

Facility	Facility Name	Type	Latitude	Longitude	Landing/Take off	Benzen	Emissions
IAH	George Bush Intercon	AIRPORT	29.9901	-95.363	33961222	71432	7.4044854
DFW	Dallas/Fort Worth In	AIRPORT	32.9094	-97.0833	37174309	71432	6.0890914
HOU	William P Hobby	AIRPORT	29.6408	-95.2869	3048647	71432	2.2633908
DAL	Dallas Love Field	AIRPORT	32.8512	-96.8636	2707080	71432	2.2374011
SAT	San Antonio Intl	AIRPORT	29.5418	-98.4863	4398445	71432	1.8341946
AUS	Austin-Bergstrom Int	AIRPORT	30.18	-97.6789	4025715	71432	1.6894803
DWH	David Wayne Hooks Me	AIRPORT	30.074	-95.5555	107224	71432	1.4376865
52F	Northwest Regional	AIRPORT	33.0498	-97.23224	83000	71432	1.1229041
SSF	Stinson Muni	AIRPORT	29.3384	-98.4808	84712	71432	1.1007393
ADS	Addison	AIRPORT	32.9778	-96.8407	73995	71432	1.0052453
AFW	Fort Worth Alliance	AIRPORT	32.9975	-97.3221	104116	71432	0.9956113
T41	La Porte Muni	AIRPORT	29.6692	-95.06419	70284	71432	0.9508698
GKY	Arlington Muni	AIRPORT	32.6624	-97.09391	67093	71432	0.9048621
ELP	El Paso Intl	AIRPORT	31.8017	-106.4	1420937	71432	0.8608799
FTW	Fort Worth Meacham I	AIRPORT	32.8248	-97.3666	62335	71432	0.8375401
HYI	San Marcos Muni	AIRPORT	29.8936	-97.86469	61520	71432	0.8189424
HQZ	Mesquite Metro	AIRPORT	32.7469	-96.53042	60364	71432	0.8166624
DTO	Denton Muni	AIRPORT	33.2007	-97.19798	58606	71432	0.7929486
RBD	Dallas Executive	AIRPORT	32.6872	-96.8766	55500	71432	0.7484381
TKI	Mc Kinney Muni	AIRPORT	33.1683	-96.592	52965	71432	0.7203811
LBB	Lubbock Intl	AIRPORT	33.6824	-101.8293	417240	71432	0.6885316
LVJ	Clover Field	AIRPORT	29.5213	-95.24217	50484	71432	0.6831286
GPM	Grand Prairie Muni	AIRPORT	32.6985	-97.04653	50539	71432	0.6831023
EFD	Ellington Field	AIRPORT	29.5936	-95.1663	73474	71432	0.6591979
IWS	West Houston	AIRPORT	29.8181	-95.67261	45121	71432	0.6110312
GGG	East Texas Rgnl	AIRPORT	32.3914	-94.7215	49509	71432	0.5993077
AXH	Houston-Southwest	AIRPORT	29.5061	-95.47692	43631	71432	0.5902823
CXO	Montgomery County	AIRPORT	30.3518	-95.41447	44877	71432	0.5563584
SGR	Sugar Land Rgnl	AIRPORT	29.6222	-95.65653	39882	71432	0.5403891
LBX	Brazoria County	AIRPORT	29.1086	-95.46208	38036	71432	0.5142711
MAF	Midland Internationa	AIRPORT	31.9474	-102.2166	190229	71432	0.5033091
TPL	Draughon-Miller Cent	AIRPORT	31.1525	-97.40778	38130	71432	0.4921949
TA90	GREEN ACRES	AIRPORT	29.9746	-95.818005	35785.22	71432	0.4841371
GTU	Georgetown Muni	AIRPORT	30.6788	-97.67938	35330	71432	0.4738477
GLS	Scholes Intl At Galv	AIRPORT	29.2653	-94.86041	30075	71432	0.4142133
ABI	Abilene Regional	AIRPORT	32.4202	-99.6953	69357	71432	0.4096476
CRP	Corpus Christi Intl	AIRPORT	27.7769	-97.5135	213492	71432	0.4093409
FWS	Fort Worth Spinks	AIRPORT	32.5652	-97.30808	29871	71432	0.4038165
AMA	Amarillo Intl	AIRPORT	35.2054	-101.733	280434	71432	0.4036318
TYR	Tyler Pounds Rgnl	AIRPORT	32.3468	-95.4115	33937	71432	0.388553
MFE	Mc Allen Miller Intl	AIRPORT	26.1828	-98.2465	146633	71432	0.3842018
PWG	Mc Gregor Executive	AIRPORT	31.4849	-97.31653	27436	71432	0.3705007
0TX1	PECAN PLANTATION	AIRPORT	32.3540	-97.676414	26706.09	71432	0.3613058
F39	Grayson County	AIRPORT	33.7141	-96.67367	27195	71432	0.3609709
ERV	Kerrville Muni/Louis	AIRPORT	29.9767	-99.08568	25900	71432	0.3499184
TX91	MADEIRA AIRPARK	STOLPOR	32.9076	-96.596936	25504.62	71432	0.3450511
RKP	Aransas Co	AIRPORT	28.0867	-97.04461	41110	71432	0.3365638
LRD	Laredo Intl	AIRPORT	27.5326	-99.4654	138961	71432	0.3250858
HRL	Valley Intl	AIRPORT	26.233	-97.6641	268143	71432	0.3191143
ODO	Odessa-Schlemeyer Fi	AIRPORT	31.9205	-102.38709	22878	71432	0.3095156

3.2.1 Near distance to GIS runs

In addition to airports, industrial facilities, roads and railroads were included in the analysis since they also contribute to air quality. A distance to the nearest emitter (in miles) was obtained (within GIS); this process was repeated five times (for the entire cancer data provided by the DSHS) in order to nearest emitter from airports, TRI facilities, road segments, railroad segments, and all 4 categories combined (All Emitters).

Near is one of the analysis tools used in calculating proximity. This tool computes the distance from each point in the input feature class or layer to the nearest polyline, or point, in the near feature class or layer, within the maximum search radius.[55] Figure 3.5 below demonstrates how nearest distance is determine within GIS.

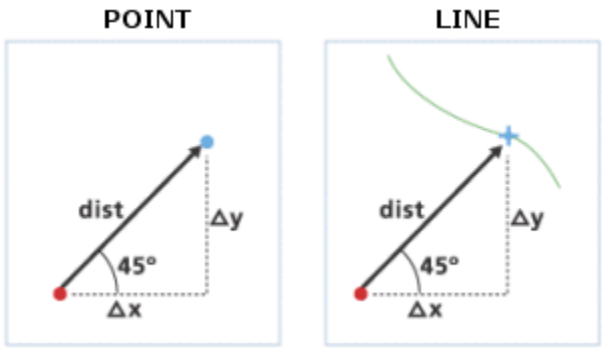


Figure 3.5. GIS

The average Near_Distance was calculated by summing the distance for each cancer incidence to the closest airport and then dividing by the total number of incidence distances. The near tool was also used to determine the average distance of the cancer incidences to roads, railroads and facilities. The average distance was used within the observed to expected ratio analysis, in addition to the distance to analysis presented within Section 4.2.2 of this report. The block group observed to expected ratio data was plotted verses the average distance of the cancer incidences within each block group to airports, roads, railroads and facilities. Figures 3.6 to 3.21 present the observed to expected ratios verses Distance scatter plots. Notice the difference between the negative control (Colon Cancer) and the other

three cancers that were added to this analysis in order to determine if any or all demonstrate a relationship with the previously described emitters.

3.2.2 Plot Results

The plots can be interpreted in the following manner: as observed to expected ratio increases (demonstrated by moving up the y axis), the distance should decrease (demonstrated by moving left on the x axis), indicating a close proximity to the source. Therefore, the higher observed to expected ratio should have a distance approaching zero on the x axis, and this is demonstrated by the leukemia observed to expected plots. It can be argued that respiratory and lymphoma somewhat demonstrate this attribute, but not to the extent of leukemia. Note it appears that block groups demonstrate a proximity to roads in the scatter plots, irrespective of high and low observed to expected ratios. Additionally, the road scatter plots tend to resemble the all emitter scatter plots, hinting that roads contribute generously to the total county benzene concentration. This will be apparent in the Poisson regression statistics section presented later within this report (Section 4.2). Such plots for airports, facilities, roads and railroads are shown in Figures 3.6 to 3.15 below. Note that Figures 3.6, 3.8, 3.10, 3.12, and 3.14 show data for only 447 of the 14,463 block groups; the other 14,016 block groups had 0 incidences, and thus were not plotted. These 14,016 block groups that do not have an observed lymphoma case are therefore represented by zeros with respect to the O/E ratio. Logging the y - axis eliminates these block groups (since the log of zero does not return any value in GIS). These points thus do not appear on the scatter plot. Additionally, for colon cancer, O/E ratios for 1,243 block groups were also eliminated in the same fashion, while 12,592 block groups with respect to leukemia and 463 block groups with respect to respiratory were also eliminated. This step is important in that it ensures block groups without cancer cases are not provided a zero distance to the emitters.

The five color classifications in Figures 3.6 to 3.15 were generated using Jenks natural breaks default classification within GIS. Note that GIS has several different methods of classifying data, each of which would yield differing results. Jenks is the default and seems to be the most widely used;[56,57] thus, Jenks was used for this study. Since GIS's Natural Breaks (Jenks) Classification was used to categorize these dataset within the provided scatter plots, it is important to note that this method of categorizing data

actually separates the dataset in places where large changes occur, thus ensuring that the variation within each category (there are five total categories) is small, while the variation between the 5 categories is large. This ensures that data within a category (or group) is more similar then other data outside of that classification category. Again, Jenks Natural Breaks is one of several univariate classification schemes within GIS and it utilizes the steps shown in Figure 3.6 in order to group the data.[58]

Step 1: The user selects the attribute, x, to be classified and specifies the number of classes required, k.

Step 2: A set of k-1 random or uniform values are generated in the range [min{x},max{x}]. These are used as initial class boundaries.

Step 3: The mean values for each initial class are computed and the sum of squared deviations of class members from the mean values is computed. The total sum of squared deviations (TSSD) is recorded.

Step 4: Individual values in each class are then systematically assigned to adjacent classes by adjusting the class boundaries to see if the TSSD can be reduced. This is an iterative process, which ends when improvement in TSSD falls below a threshold level, i.e. when the within class variance is as small as possible and between class variance is as large as possible. True optimization is not assured. The entire process can be optionally repeated from Step 1 or 2 and TSSD values compared.

Figure 3.6 Jenks Natural Breaks Algorithm[59]

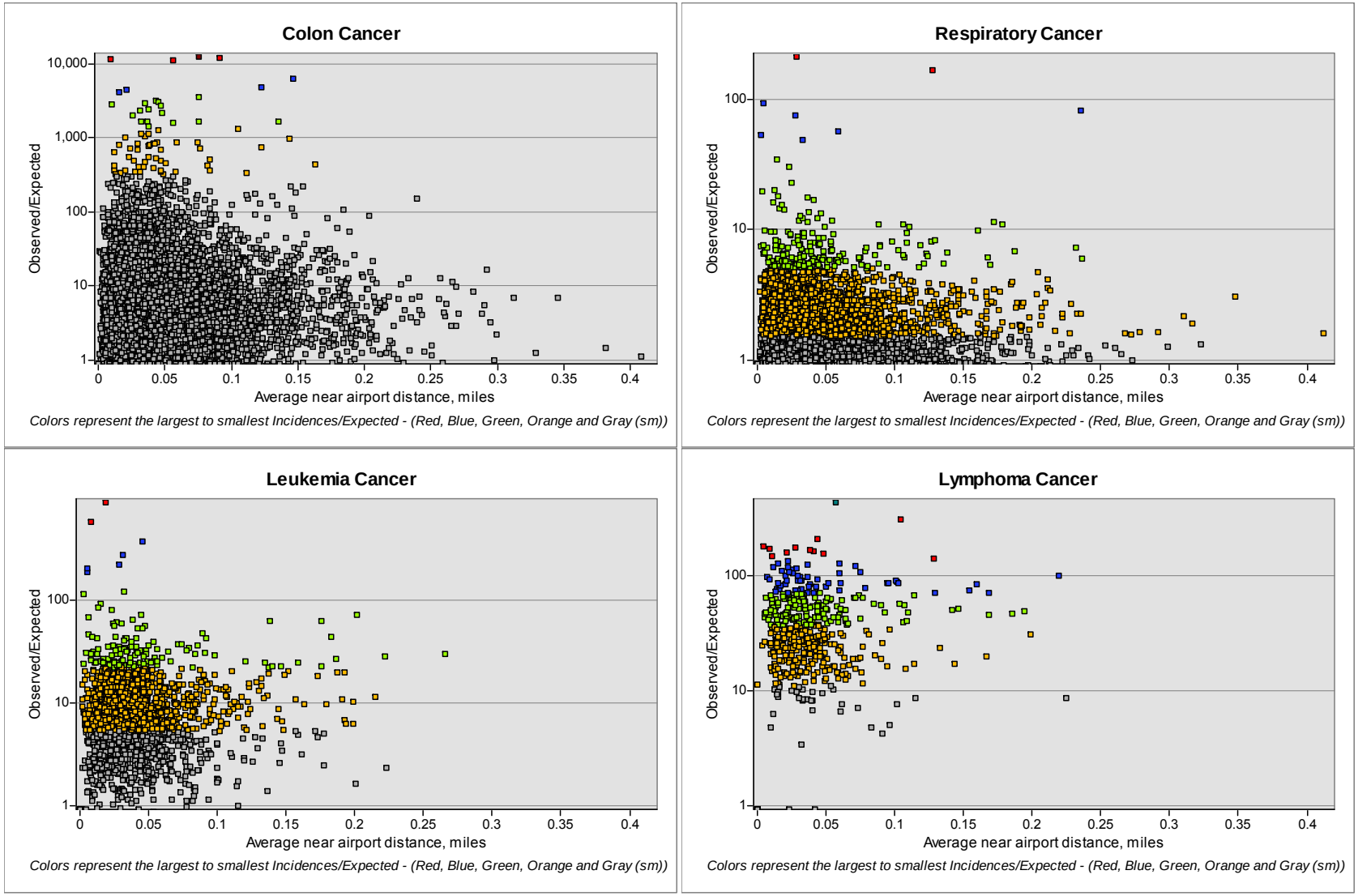


Figure 3.7. Scatter Plots of Ratio of Observed to Expected Incidences of Colon Cancer (all ages), Respiratory Cancer (all ages), Leukemia (age 9 and under), Lymphoma (age 9 and under), versus the Average Near Distance to Airports

.

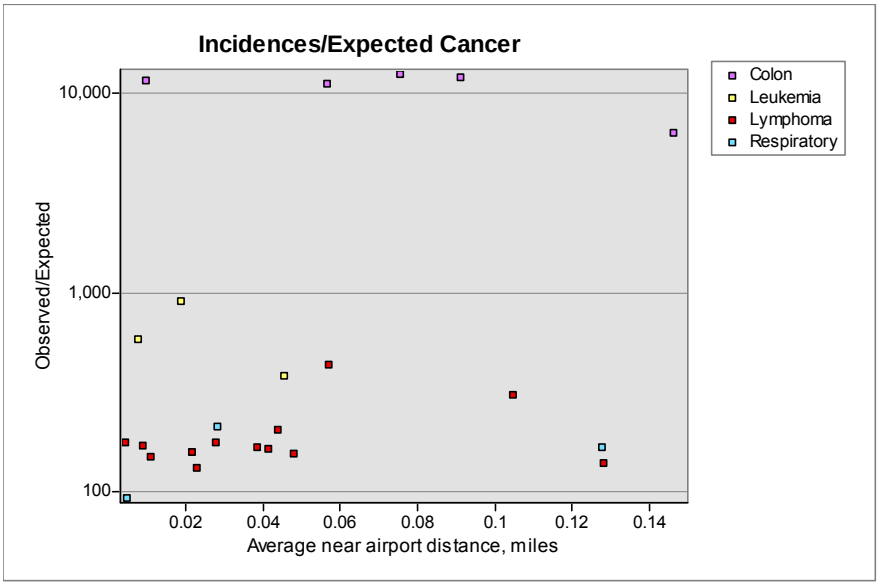


Figure 3.8. Scatter Plot of Highest Ratio of Observed to Expected Incidences of Colon Cancer (all ages), Respiratory Cancer (all ages), Leukemia (age 9 and under), Lymphoma (age 9 and under), versus the Average Near Distance to Airports

In Figure 3.7, if an emission source is causing cancer, one would expect the ratio of the observed to expected incidences to decrease as the distance from the emission source increases. Additional analysis using scatter plots focusing on the highest observed to expected ratio with respect to distance was performed. See Figures 3.8, 3.10, 3.12, 3.14 and 3.16 for the highest observed to expected ratio with respect to distance airports, facilities, roads, railroads and all emitters. These scatter plots have colon, leukemia, lymphoma and respiratory all on the same plot so the reader can note the distances between the each cancer type.

Figure 3.8 demonstrates what we will refer to as the highest of the observed divided by expected cancer incidences that were obtained using GIS's Natural Breaks (Jenks) Classification and plotted against an average distance with a maximum of 0.14 miles in order to obtain a zoomed in view of the points.

Note that the closer the points in Figure 3.8 are to the sources, the more likely that a relationship may exist. Leukemia demonstrates a closeness to the source (with a 0.05 miles furthest distance), while data for colon, lymphoma, and respiratory are dispersed to a distance of 0.15 miles. Thus, as noted in Figure 3.7, these scatter plots suggests a relationship between observed to expected ratio for leukemia

(more so then the other three cancer groups) and proximity to emission sources of airports in Texas. Additionally, Table 3.4 below provides the average distance from airports for the observed to expected ratios shown in Figure 3.8 and demonstrates that high observed to expected ratios for leukemia are closer to airports, with an average distance of 0.0239 miles, while colon ratios are furthest, with an average distance of 0.0757 miles. This suggests that a relationship between leukemia and airports exists.

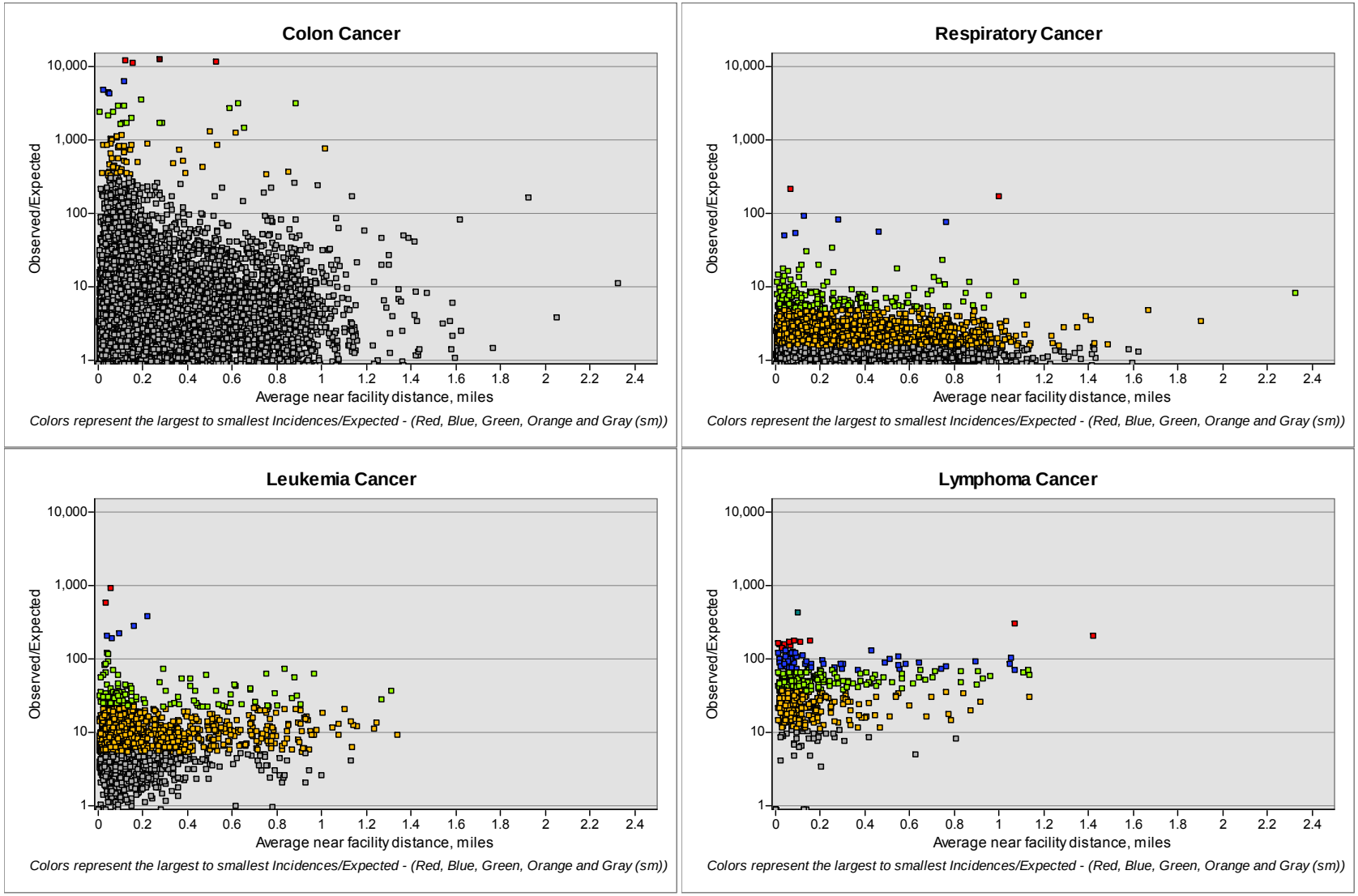


Figure 3.9.Scatter Plots of Ratio of Observed to Expected Incidences of Colon Cancer (all ages), Respiratory Cancer (all ages), Leukemia (age 9 and under), Lymphoma (age 9 and under), versus the Average Near Distance to Facilities.

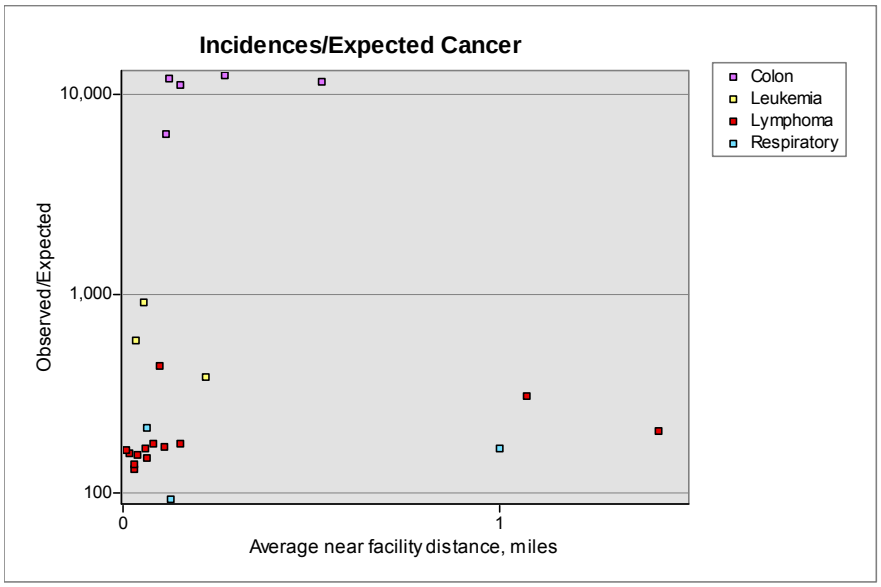


Figure 3.10. Scatter Plot of Highest Ratio of Observed to Expected Incidences of Colon Cancer (all ages), Respiratory Cancer (all ages), Leukemia (age 9 and under), Lymphoma (age 9 and under), versus the Average Near Distance to Facilities

According to Figure 3.9, it is apparent that leukemia demonstrates closeness to the source, while data for colon, lymphoma, and respiratory are dispersed to greater distances, up to 1.5 miles. Thus, these scatter plots suggest a relationship between observed to expected ratio for leukemia (more so then the other three cancer groups) and proximity to emission sources of facilities, in Texas. Additionally Table 3.4, located below, provides the average distance from facilities for the scatter plots located within Figure 3.10 and demonstrates that leukemia (for the highest observed to incidence ratio) is on average closer to facilities with a distance of 0.1023 miles, while respiratory is the furthest with a distance of 0.3970. The fact that colon (our negative control) demonstrates a closeness to facilities as well indicates that there is not a strong relationship with facilities emissions and these cancer cases.

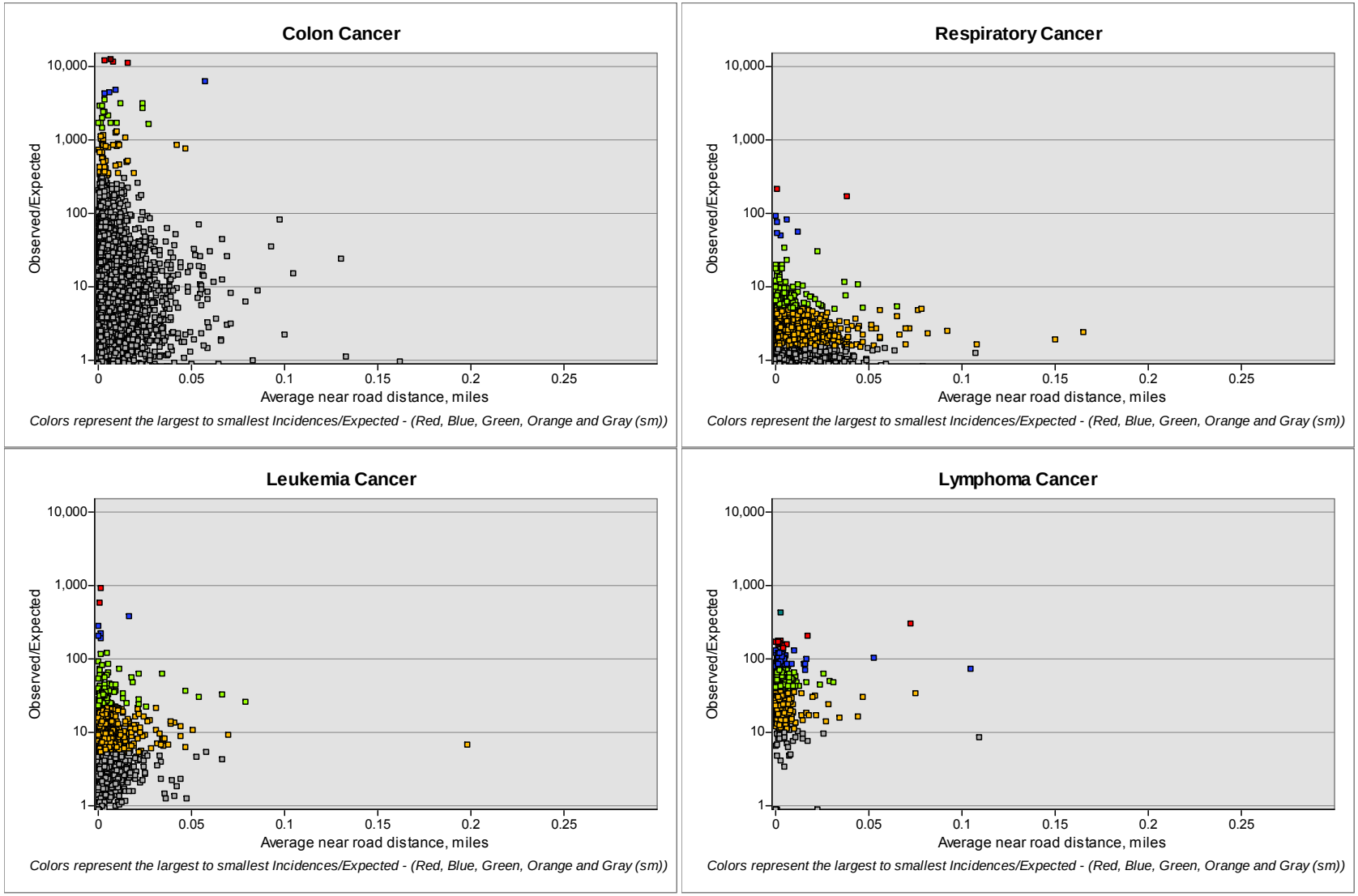


Figure 3.11. Scatter Plots of Ratio of Observed to Expected Incidences of Colon Cancer (all ages), Respiratory Cancer (all ages), Leukemia (age 9 and under), Lymphoma (age 9 and under), versus the Average Near Distance to Roads

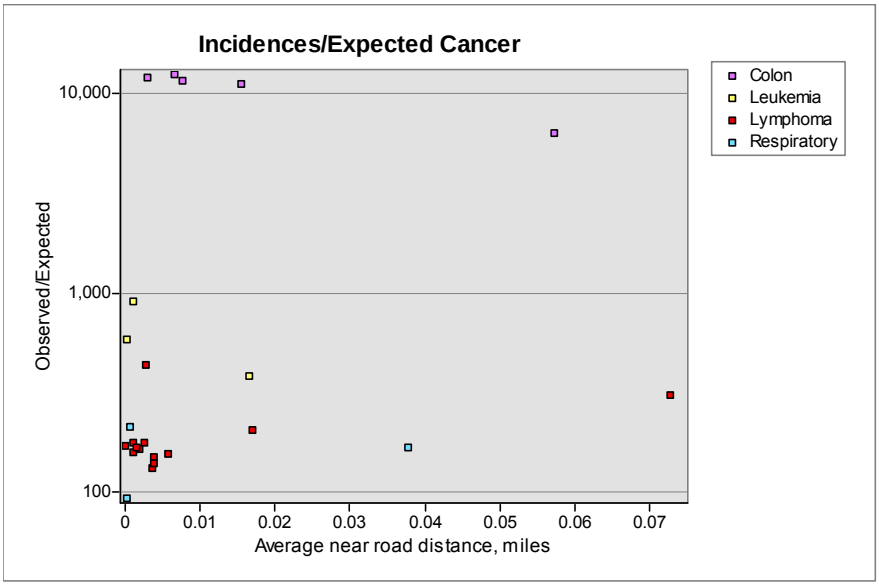


Figure 3.12. Scatter Plot of Highest Ratio of Observed to Expected Incidences of Colon Cancer (all ages), Respiratory Cancer (all ages), Leukemia (age 9 and under), Lymphoma (age 9 and under), versus the Average Near Distance to Roads

Figure 3.11 can be interpreted similarly as is described for Figure 3.7. However, it is important to note that from a qualitative inspection, the general trend of decreasing y-values with increasing x-values cannot be seen as clearly for these scatter plots as was demonstrated for scatter plots of airports. In fact, it appears that the distance is generally similar irrespective of high or low observed to expected ratios. Thus, from qualitative inspection, one can conclude that within the State of Texas, most persons reside near a road regardless of observed to expected ratios.

Figure 3.12 demonstrates the highest of the observed divided by expected cancer incidence ratios plotted against an average distance with a maximum of 0.75 miles in order to obtain a zoomed in view of the points. It is apparent that leukemia demonstrates a closeness to the source (with a 0.17 miles furthest distance), while data for colon, lymphoma, and respiratory are dispersed to greater distances of up to 0.75 miles. Thus, these scatter plots suggest a relationship between observed to expected ratio for leukemia (more so then the other three cancer groups) and proximity to emission sources of roads, in Texas. Additionally Table 3.4 below provides the average distance from roads for the scatter plots located within Figure 3.12 and demonstrates that leukemia (for the highest observed to incidence ratio) is on average closer to roads with a distance of 0.0060 miles, while colon is the furthest with a distance of 0.0181. Thus, a relationship between leukemia and roads may exist.

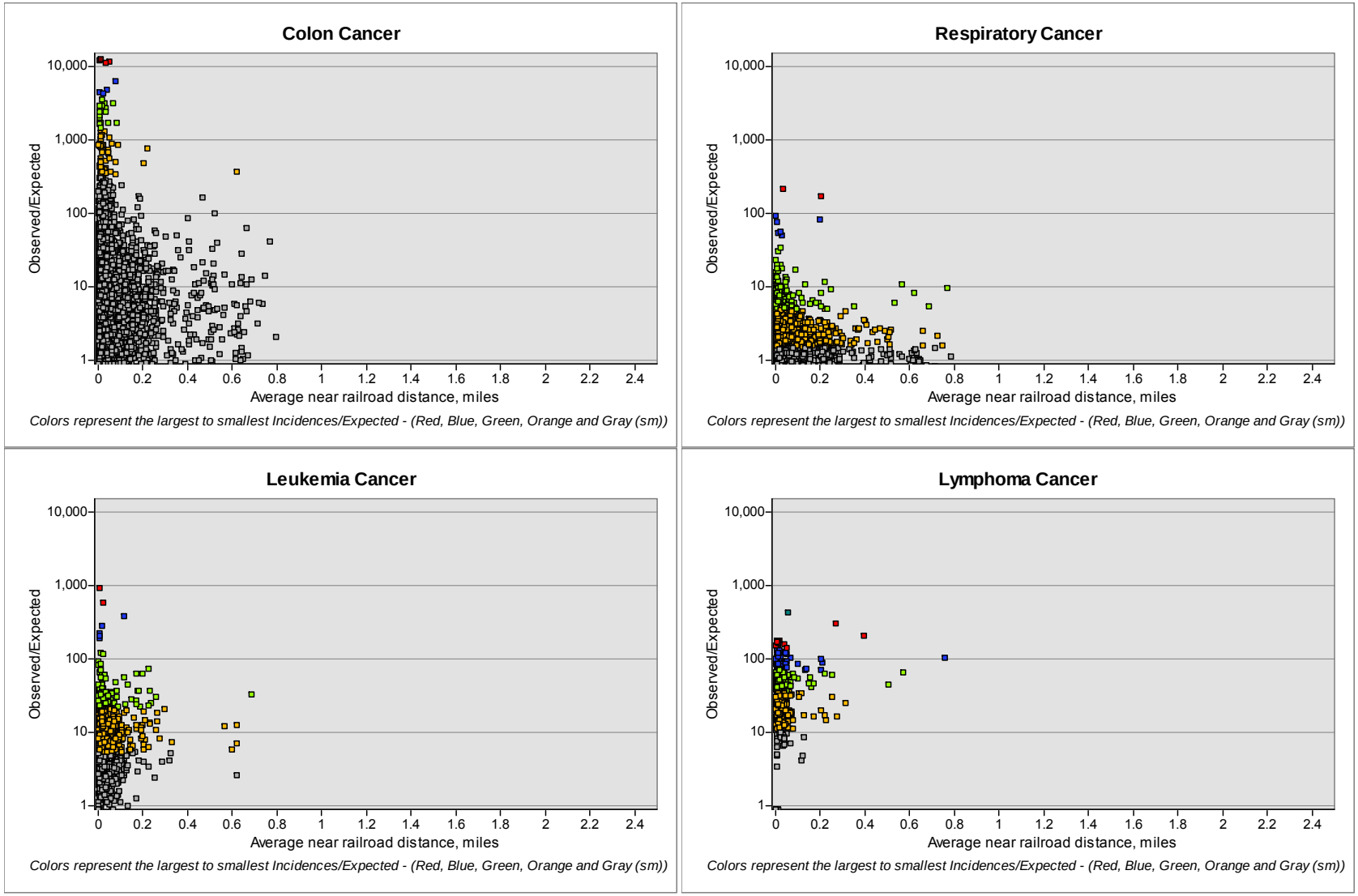


Figure 3.13. Scatter Plots of Ratio of Observed to Expected Incidences of Colon Cancer (all ages), Respiratory Cancer (all ages), Leukemia (age 9 and under), Lymphoma (age 9 and under), versus the Average Near Distance to Railroads

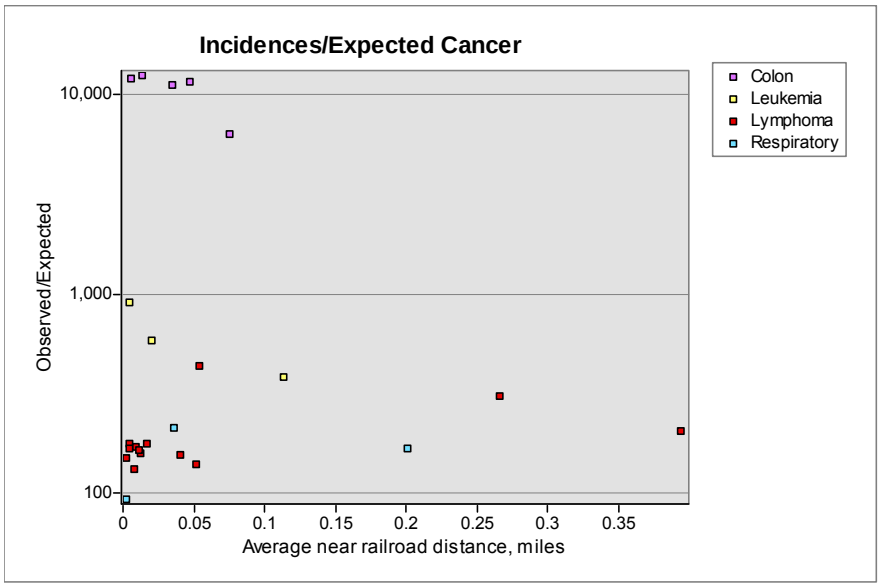


Figure 3.14. Scatter Plot of Highest Ratio of Observed to Expected Incidences of Colon Cancer (all ages), Respiratory Cancer (all ages), Leukemia (age 9 and under), Lymphoma (age 9 and under), versus the Average Near Distance to Railroads

Figure 3.14 demonstrates the highest of the observed divided by expected cancer incidence ratios against an average distance with a maximum of 0.4 miles in order to obtain a zoomed in view of the points. It is apparent that leukemia demonstrates closeness to the source (with a furthest distance around 0.12 miles) Respiratory and lymphoma are dispersed to greater distances of 0.2 and 0.4 miles. Colon, however, is closest to the source, dispersed only up to about 0.8 miles. Thus, these scatter plots suggest a relationship between observed to expected ratio for colon (more so then the other three cancer groups) and proximity to emission sources of railroads, in Texas. Additionally, Table 3.4 below provides the average distance from railroads for the scatter plots located within Figure 3.14 and demonstrates that colon (for the highest observed to incidence ratio) is on average closer to railroads with a distance of 0.0352 miles, while respiratory is the furthest with a distance of 0.0794 miles. This suggests that a relationship between leukemia and railroads does not exist; additionally, a relationship between railroads and the cancers documented within this study does not exist since colon cancer (our negative control) demonstrates a closer distance with respect to the highest observed to expected ratio to railroads.

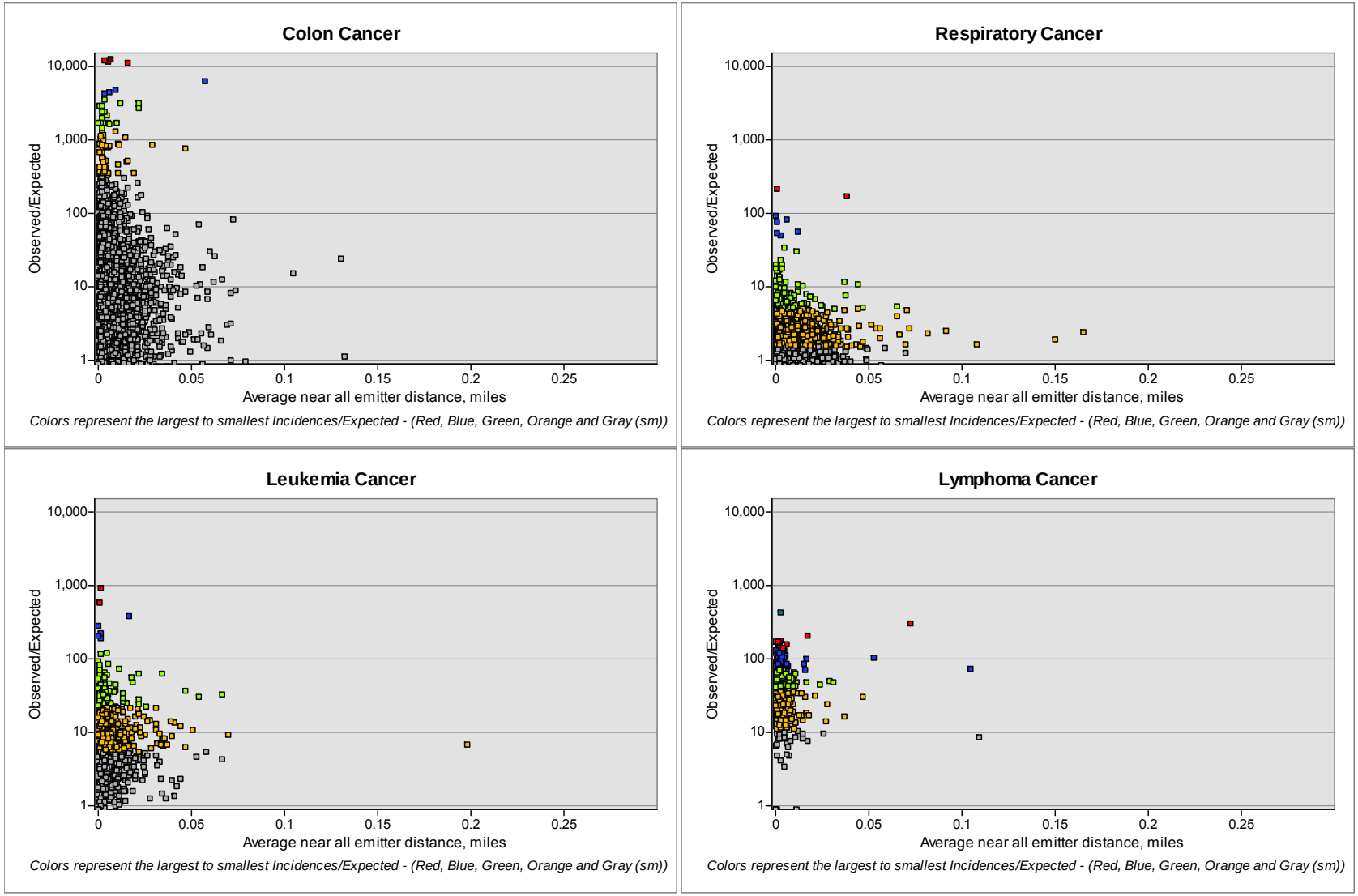


Figure 3.15. Scatter Plots of Ratio of Observed to Expected Incidences of Colon Cancer (all ages), Respiratory Cancer (all ages), Leukemia (age 9 and under), Lymphoma (age 9 and under), versus the Average Near Distance to All Emitters

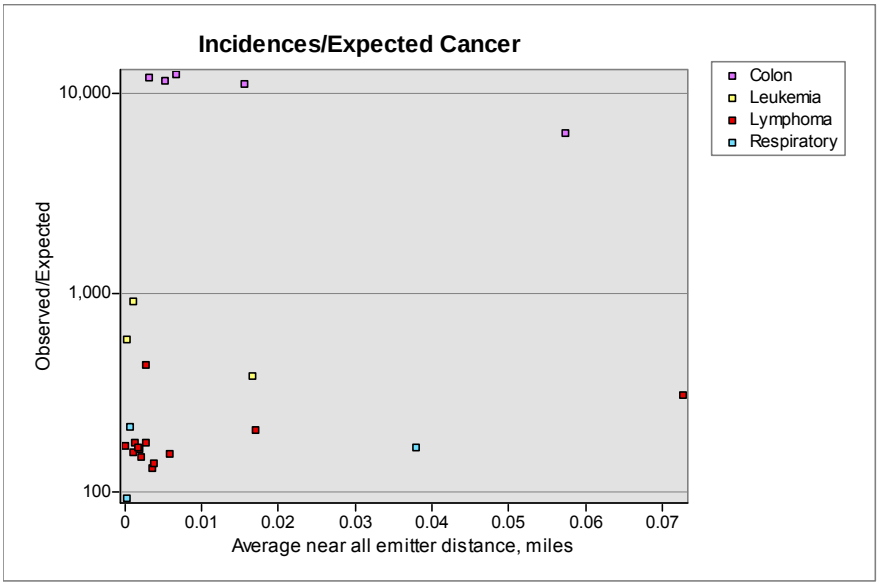


Figure 3.16. Scatter Plot of Highest Ratio of Observed to Expected Incidences of Colon Cancer (all ages), Respiratory Cancer (all ages), Leukemia (age 9 and under), Lymphoma (age 9 and under), versus the Average Near Distance to All Emitters

Figure 3.16 demonstrates the highest of the observed divided by expected cancer incidence ratios plotted against an average distance with a maximum of 0.075 miles in order to obtain a zoomed in view of the points. It is apparent that leukemia demonstrates a closeness to the source (with a 0.02 miles furthest distance), while data for colon, lymphoma, and respiratory are dispersed to greater distances up to 0.075 miles. Thus, these scatter plots suggest a relationship between observed to expected ratio for leukemia (more so then the other three cancer groups) and proximity to emission sources of all emitters, in Texas. Additionally Table 3.4 below provides the average distance from all emitters for the scatter plots located within Figure 3.16 and demonstrates that leukemia (for the Highest observed to incidence ratio) is on average closer to all emitters with a distance of 0.006 miles, while colon is the furthest with a distance of 0.0176. This suggests that a relationship between leukemia and all emitters may exist.

Table 3.4. Average distance of highest of cancer incidences to emission sources

	Leukemia	Lymphoma	Respiratory	Colon
Number of Observations	3	13	3	5
Average Distance to Emission Source, miles				
Airport	0.0239	0.0428	0.0535	0.0757
Facility	0.1023	0.2449	0.3970	0.2373

Table 3.4. – *Continued*

Road	0.0060	0.0091	0.0129	0.0181
Railroad	0.0459	0.0670	0.0794	0.0352
All Emitters	0.0060	0.0090	0.0129	0.0176

CHAPTER 4

METHODOLOGY AND RESULTS FOR QUESTION 2: SPECIFICALLY, IS THERE A RELATIONSHIP BETWEEN CHILDHOOD LEUKEMIA AND AIRPORT EMISSIONS?

There are several limitations in using an ecological type of analysis to evaluate associations of exposure to air pollutants with cancer risk because no individual-level data are available regarding exposure or other risk factors. Other risk factors for benzene and leukemia include smoking, radiation exposure, and exposure to other leukemia-causing chemicals. Childhood leukemia was selected for study because, given its relatively shorter latency period, we could reasonably assume that current airport locations and benzene emissions (for 2005) are relevant to those in existence in a 10-year time frame before the cancer cases were diagnosed. For example, for children less than 10 years old with leukemia diagnosed during 1995-2005, their etiologically relevant window of time would reasonably be from 1985 to 2000, assuming 5-10 year latency. In addition, risk factors such as smoking and occupational exposure to benzene would not be expected to be confounders of the analyses in children. Exposure history linked to other locations (including moving from another city/state or country after having lived there for 50 years or so) is another confounding factor for adults, which is less of an issue for children.[60] However, environmental tobacco smoke exposure could be a confounder; our analyses were not adjusted for this risk factor since this data was not provided in the dataset obtained from the Texas DSHS.

The methodology to address Question 2 included 2 steps:

A) Develop a Poisson regression model correlating childhood leukemia incidence to benzene emissions from various sources (railroads, airports, industrial facilities, roads) on a county level, and

B) Use GIS to compare average nearest distance to various sources (railroads, airports, industrial facilities, roads) for those with and without childhood leukemia state-wide.

Each of these steps will be discussed in turn.

4.1 Develop a Poisson regression model correlating childhood leukemia incidence to benzene emissions from various sources (railroads, airports, industrial facilities, roads) on a county level.

Using SPSS, the following regression equation relating childhood leukemia incidence to benzene emissions (NATA emissions from 2005) was developed:

$$L = -6.664 + 0.440 * \ln (P) + 0.497 * \ln (B_{roads}) + 0.230 * \ln (B_{airports})$$

where
L = number of leukemia cases in children 9 years and under within the county.
P = population of individuals 9 and under in the county in the year 2000,
B_{roads} = county-wide benzene emissions from roads (tons/year),
$B_{airports}$ = county-wide benzene emissions from airports (tons/year),

Data for 3 of Texas' 254 counties (1.2%) was excluded from the analysis since these counties contained zeros and missing values. It is important to note that TRI facilities did not demonstrate a strong relationship with leukemia for those 9 and under when viewed in a GIS scatter plot of dependent and independent variables, and thus was eliminated as a predictor variable. Benzene emissions from railroads were also included as a potential predictor variable, but its coefficient was not statistically significant; therefore, it was eliminated from the final model. Table 4 provides minimum, maximum, mean, and standard deviation values for the response and predictor variables, which indicate the range of values over which the regression equation applies. Logs were taken of the predictor variable values because the data did not fit linear assumptions in its raw form (as was demonstrated by the initial scatter plots, see Appendix C); thus the data needed to be transformed, thereby increasing linearity thus ensuring a better linear model.

Table 4.1 Response and predictor variable information

Variable	N	Min.	Max.	Mean	Std. Deviation
L	251	.0	408.0	8.49	34.08
ln P	251	8.42	20.13	14.52	1.77
ln B_{roads}	251	-.09	6.38	2.5	1.2
ln $B_{airports}$	251	-4.73	2.74	-1.82	1.48

A Pearson Chi-Square goodness of fit test was run and yielded a value of 1.872, which indicates a good fit for the model since it is much less than 5 and close to 1.

Finally, Table 4.1 indicates that population and benzene emissions from roads and airports are statistically significant predictor variables for childhood leukemia incidences.

Table 4.2 Regression model parameter estimates

Parameter	β	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-6.664	.6643	-7.967	-5.362	100.631	1	.000
ln P	.440	.0507	.340	.539	75.221	1	.000
ln B_{roads}	.497	.0712	.358	.637	48.811	1	.000
ln $B_{airports}$	.230	.0414	.149	.311	30.813	1	.000
(Scale)	1.765[a]						

Thus, according to the regression model developed, a relationship does exist between childhood leukemia and benzene emissions from roads and airports.

4.2 Use GIS to compare average nearest distance to various sources (railroads, airports, industrial facilities, roads) for those with and without childhood leukemia state-wide.

The near tool was used to determine the average distance of the cancer incidences to roads, railroads and facilities, as described in Section 4.1.2.1. Next, a synthetic population shape file was obtained from Research Triangle Institute (RTI International). RTI International states that the company provides health solutions by "offering innovative research and technical services to governments and businesses worldwide". The synthetic population file was created by "extending an iterative proportional fitting method to generate a synthesized, geospatially explicit, human agent database that represents the US population in the 50 states and the District of Columbia in the year 2000".[61] This file contained latitude and longitude coordinates for each individual (approximately 20 million) living within the entire State of Texas, but only data relating to those ages 9 and under (4.1 million individuals) was used for this project. Again, the "near tool" was used in order to determine the distance in miles from each of the

synthetic population points to the four kinds of emitters previously mentioned, as well as the nearest emitter, regardless of type (titled All Emitters).

Table 4.3 below compares:

1) the average distance between the 4 categories of emitters and all individuals 9 and under in Texas, and
2) the average distance between the 4 categories of emitters and individuals 9 and under with leukemia in Texas.

Table 4.3 Average distance from benzene emitters for persons 9 and under with/without leukemia in Texas

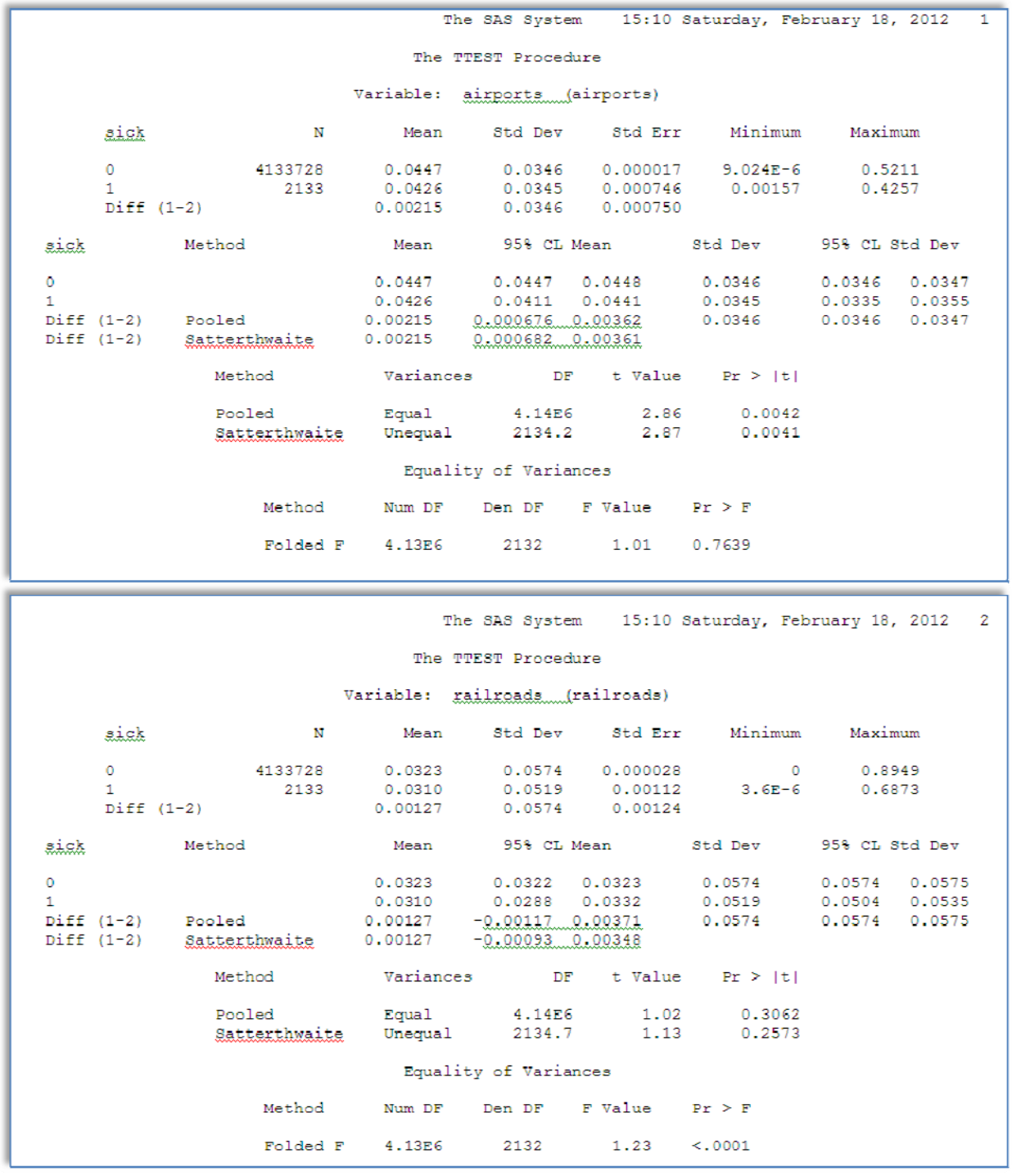
Emitter Category	Avg. distance (in miles) from persons 9 and under to nearest emitter		Difference between Dist 1&2*	Pr > |t|
	Persons without cancer	Persons with cancer		
Airports	0.045	0.042	0.00215	0.0042; Significant Difference
TRI Facilities	0.221	0.190	0.0222	0.0001; Significant Difference
Roads	0.0048	0.0050	-0.00034	0.1016; not a Significant Difference
Railroads	0.033	0.031	0.00127	0.2573; not a Significant Difference

* Please note that rounding occurs at different stages within the statistical program causing slight differences in the output for the Difference Between Dist 1&2 column.

All those average distances that are furthest from the emitters are highlighted in yellow. Notice that for 3 of the 4 cases, the average distance of those without cancer is higher than the average distance of those with cancer; this is to be expected. In other words, it is expected that those who are not inflicted with cancer should reside further from the emitters, or those with cancer should live on average closer to the source if the source is suspected of contributing to the illness. However, in the case of the roads, it is apparent that the individuals with cancer actually live further on average from the source (roads) then those who do not have the disease. This statement is supported by the scatter plots, where it was noted that most of the block groups (irrespective of having a high or low observed to expected ratio)

demonstrated a close proximity to roads. This simple observation minimizes the chance of the roads having any effect on those individuals sick with cancer due to the obvious fact that those individuals without cancer reside closer on average to the source.

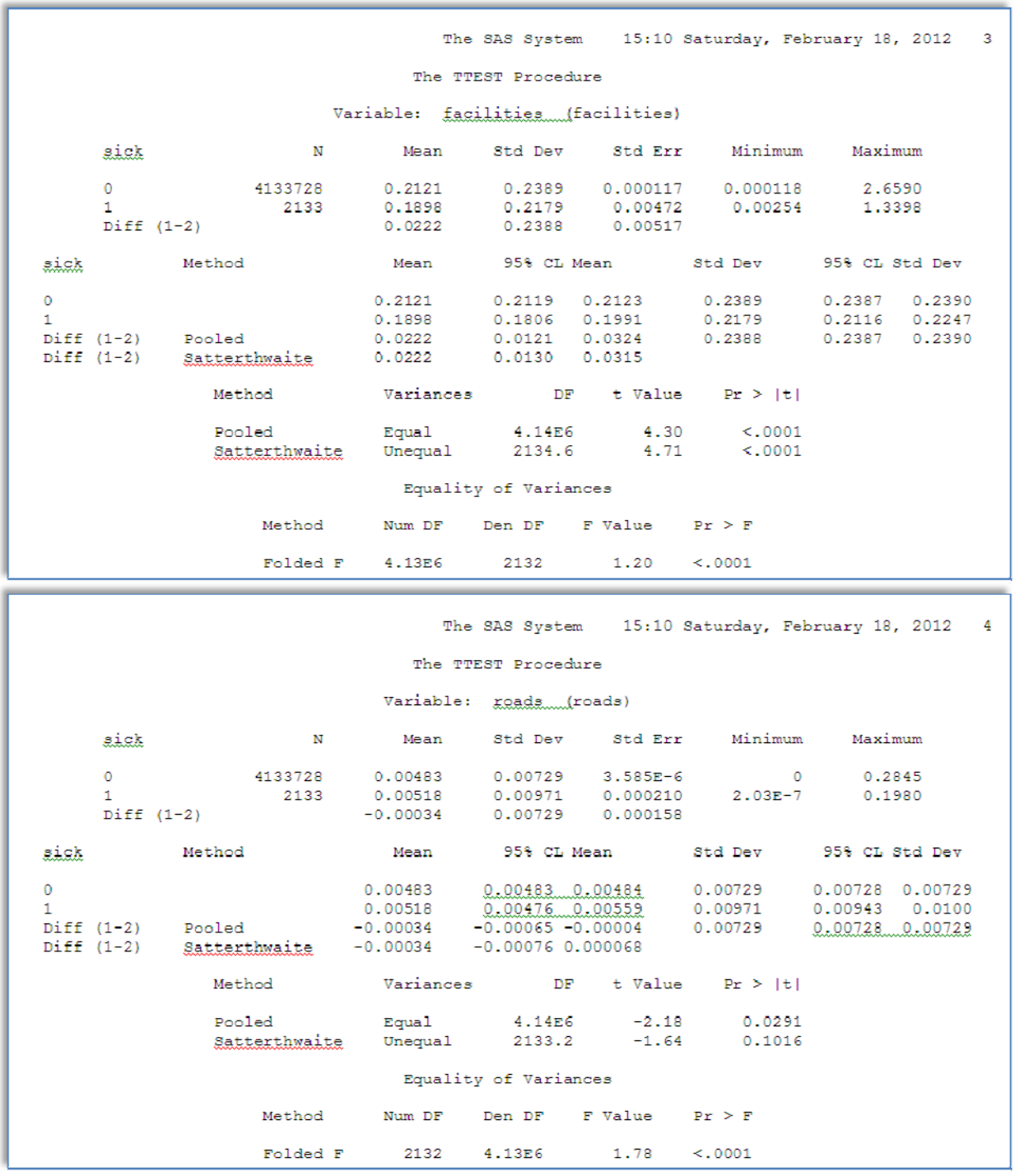
T-Tests were performed in order to determine whether the difference in mean distances were statistically significant. According to Table 4.3, for airports and TRI facilities, the average distance of persons with cancer is lesser than the average distance of those without cancer, and the difference is statistically significant to a 95% confidence level. This provides evidence that airports and TRI facilities may be contributing to cancer incidences. For roads and railroads, the difference between the average distance from the facilities for persons with and without cancer was not statistically significant. The T Test ANOVA output can be found in Figure 4.1.



```
                                    The SAS System    15:10 Saturday, February 18, 2012   1

                                    The TTEST Procedure

                                  Variable:  airports  (airports)

      sick                 N        Mean     Std Dev     Std Err     Minimum     Maximum

      0              4133728      0.0447      0.0346    0.000017   9.024E-6      0.5211
      1                 2133      0.0426      0.0345    0.000746     0.00157     0.4257
      Diff (1-2)                  0.00215      0.0346    0.000750

sick          Method               Mean       95% CL Mean         Std Dev      95% CL Std Dev

0                                 0.0447      0.0447   0.0448      0.0346      0.0346   0.0347
1                                 0.0426      0.0411   0.0441      0.0345      0.0335   0.0355
Diff (1-2)    Pooled              0.00215    0.000676  0.00362     0.0346      0.0346   0.0347
Diff (1-2)    Satterthwaite       0.00215    0.000682  0.00361

                Method           Variances        DF    t Value    Pr > |t|

                Pooled           Equal         4.14E6       2.86      0.0042
                Satterthwaite    Unequal       2134.2       2.87      0.0041

                                  Equality of Variances

                     Method      Num DF    Den DF    F Value    Pr > F

                     Folded F    4.13E6      2132       1.01    0.7639
```

```
                                    The SAS System    15:10 Saturday, February 18, 2012   2

                                    The TTEST Procedure

                                  Variable:  railroads  (railroads)

      sick                 N        Mean     Std Dev     Std Err     Minimum     Maximum

      0              4133728      0.0323      0.0574   0.000028           0      0.8949
      1                 2133      0.0310      0.0519     0.00112     3.6E-6      0.6873
      Diff (1-2)                  0.00127      0.0574     0.00124

sick          Method               Mean       95% CL Mean         Std Dev      95% CL Std Dev

0                                 0.0323      0.0322   0.0323      0.0574      0.0574   0.0575
1                                 0.0310      0.0288   0.0332      0.0519      0.0504   0.0535
Diff (1-2)    Pooled              0.00127    -0.00117  0.00371     0.0574      0.0574   0.0575
Diff (1-2)    Satterthwaite       0.00127    -0.00093  0.00348

                Method           Variances        DF    t Value    Pr > |t|

                Pooled           Equal         4.14E6       1.02      0.3062
                Satterthwaite    Unequal       2134.7       1.13      0.2573

                                  Equality of Variances

                     Method      Num DF    Den DF    F Value    Pr > F

                     Folded F    4.13E6      2132       1.23    <.0001
```

Figure 4.1. T-Test SAS output of the distance to airports and railroads for those with and without leukemia (9 years and under).



The SAS System 15:10 Saturday, February 18, 2012 3

The TTEST Procedure

Variable: facilities (facilities)

sick	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	4133728	0.2121	0.2389	0.000117	0.000118	2.6590
1	2133	0.1898	0.2179	0.00472	0.00254	1.3398
Diff (1-2)		0.0222	0.2388	0.00517		

sick	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		0.2121	0.2119	0.2123	0.2389	0.2387	0.2390
1		0.1898	0.1806	0.1991	0.2179	0.2116	0.2247
Diff (1-2)	Pooled	0.0222	0.0121	0.0324	0.2388	0.2387	0.2390
Diff (1-2)	Satterthwaite	0.0222	0.0130	0.0315			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	4.14E6	4.30	<.0001
Satterthwaite	Unequal	2134.6	4.71	<.0001

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	4.13E6	2132	1.20	<.0001

The SAS System 15:10 Saturday, February 18, 2012 4

The TTEST Procedure

Variable: roads (roads)

sick	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	4133728	0.00483	0.00729	3.585E-6	0	0.2845
1	2133	0.00518	0.00971	0.000210	2.03E-7	0.1980
Diff (1-2)		-0.00034	0.00729	0.000158		

sick	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		0.00483	0.00483	0.00484	0.00729	0.00728	0.00729
1		0.00518	0.00476	0.00559	0.00971	0.00943	0.0100
Diff (1-2)	Pooled	-0.00034	-0.00065	-0.00004	0.00729	0.00728	0.00729
Diff (1-2)	Satterthwaite	-0.00034	-0.00076	0.000068			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	4.14E6	-2.18	0.0291
Satterthwaite	Unequal	2133.2	-1.64	0.1016

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	2132	4.13E6	1.78	<.0001

Figure 4.2. T-Test SAS output of the distance to facilities and roads for those with and without leukemia (9 years and under).

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This study provides 3 lines of evidence that airport benzene emissions contribute to increased incidences of childhood leukemia in the State of Texas:

- First, ratios of observed to expected incidences were plotted for 4 kinds of cancer vs. distance to major emission sources (railroads, airports, industrial facilities, roads), for each block group in the State of Texas using GIS. According to this more general analysis, respiratory cancer, colon cancer and leukemia do demonstrate a decrease in observed to expected ratio with an increase in distance from the airport source. Thus, a separate analysis was conducted which focused on the highest observed to expected ratio. This analysis demonstrated a visual closeness to the source for leukemia (0.05 miles), while all the other cancers were dispersed to a further distance of 0.15 miles. Therefore, an average distance from emitters was calculated for these highest observed to expected ratios. Of this analysis, leukemia was the closest, while colon (the negative control) was the furthest. This suggests a relationship between leukemia (for those block groups with the highest observed to expected ratios) and proximity to emission sources, including airports, in Texas.

- Second, a separate (independent of the previous) analysis involving a Poisson regression model was developed for estimating cases of childhood leukemia and benzene emissions from roads and airports (all grouped at the county level). Since the coefficient for benzene emissions from airports was statistically significant in this model, it indicates a link between airport emissions and incidence of childhood leukemia, which supports our previous findings.

- Third, the average nearest distance to the 4 benzene emitter categories of persons with and without childhood leukemia in Texas was compared. For airports and TRI facilities, the average distance of children with leukemia was lesser than the average distance of children without leukemia, and the difference was statistically significant to a 95% confidence level. This final analysis (also independent of the previous analyses) provides additional evidence that airport benzene emissions contribute to childhood leukemia incidences.

5.2 Recommendations for Future Work

The intent of this project was to devise a relatively efficient, quick and easy way for agencies to provide evidence that polluting entities do or do not impact the health of the general public. Agencies in other states could conduct similar analyses for childhood leukemia and benzene emissions. Using an approach similar to that used in this study, other childhood diseases with environmental stimuli could be studied. As this study noted, focusing on children may eliminate important confounding factors that are almost impossible to account for with large sets of (state or nationwide) data. .If a relationship between a type of emission source and

disease is suggested, it would be wise to take steps to reduce the public's exposure to such sources.

APPENDIX A

TEXAS DEPARTMENT OF STATE HEALTH SERVICES (DSHS) RATES, IN ADDITION TO A RELATIVE RISK CALCULATION SPREADSHEET PROVIDED BY DRS. BARBARA GLENN AND THOMAS BATESON OF NCEA

Childhood Leukemia Cancer Rates per 1,000,000

u:\Datareq\2010\10378tbl2.xls

Texas statewide Age-Specific Rates, Average Annual Rates, 1995-2000,
Selected Age Groups, and by race and ethnicty and sex (Data Request # 10378)
All Chioldhood Leukemias combined (ICCC-3 Group I)
Rates are per
1,000,000

	All Races		White Non-Hispanic		Black		Hispanic		Other Race	
	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate
00-09 years	69.6	51.7	66.0	51.2	27.7	27.7	86.2	59.1	69.7	51.0
00-04 years	91.3	71.9	90.0	74.4	41.9	40.1	107.4	77.4	70.4	67.5
05-09 years	47.5	31.3	43.1	29.2	14.4	16.3	62.5	38.9	68.9	33.2

*Rates are per 1,000,000.

Prepared by the Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry. Data Request # 10378 11/5/2010
Incidence Source: Texas Department of State Health Services, Cancer Epidemiology and
Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3, .

Colon Cancer Rates per 100,000

Texas statewide Age-Specific Rates, Average Annual Rates, 1995-2000,
Five-Year Age Groups, and by race and ethnicty and sex (Data Request # 10378)
Colon Cancer (ICD-O-3), Rates are per 100,000

	All Races		White Non-Hispanic		Black		Hispanic		Other Race	
	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate
00-04 years	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
05-09 years	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10-14 years	0.0	0.1	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.0
15-19 years	0.2	0.3	0.2	0.4	0.0	0.0	0.2	0.2	0.0	0.0
20-24 years	0.5	0.5	0.6	0.7	0.5	0.7	0.4	0.2	0.5	0.6
25-29 years	1.3	1.0	1.6	1.0	2.0	0.8	0.7	1.0	1.4	1.0
30-34 years	1.9	2.4	1.8	2.3	3.2	3.0	1.7	2.1	1.5	2.1
35-39 years	4.6	4.6	4.8	4.7	7.4	6.5	3.4	3.8	2.7	3.1
40-44 years	8.9	9.1	9.3	8.2	11.3	12.8	7.4	9.8	4.3	6.1
45-49 years	18.5	18.5	17.5	17.0	28.0	31.7	17.1	16.8	15.8	10.8
50-54 years	35.5	27.9	34.5	27.0	57.2	49.1	29.2	21.6	28.7	17.3
55-59 years	63.0	44.5	63.2	43.8	98.9	71.8	48.0	37.2	45.9	18.2
60-64 years	98.8	64.2	100.4	65.2	142.4	92.0	77.2	48.3	53.3	53.9
65-69 years	154.9	105.2	159.7	106.5	188.7	154.8	124.1	76.3	94.0	85.6
70-74 years	206.0	148.6	213.8	152.0	258.5	217.6	154.9	102.2	110.8	112.4
75-79 years	271.3	199.2	283.1	203.1	307.7	268.3	199.9	143.7	187.9	133.4
80-84 years	323.9	243.6	331.1	248.9	398.2	291.4	240.1	185.3	287.2	182.7
85+ years	383.6	309.2	405.4	319.9	432.2	335.6	254.0	204.7	189.1	362.5

* Rates are per 100,000.

Prepared by the Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry. Data Request # 10378 11/5/2010
Incidence Source: Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3, .

Respiratory Cancer Rates per 100,000

u:\datareq\2010\10385tbl2.xls

Texas Statewide Age-Specific Rates, Average Annual Rates, 1995-2000,
Five-Year Age Groups, and by race and ethnicity and sex (Data request # 10385)
All Respiratory System Cancers (ICD-O-3), Rates per 100,000

	All Races		White Non-Hispanic		Black		Hispanic		Other Race	
	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate	Male Rate	Female Rate
00-04 years	0.5	0.3	0.8	0.5	0.3	0.2	0.2	0.3	0.6	0.0
05-09 years	0.2	0.1	0.3	0.1	0.0	0.0	0.1	0.1	0.0	0.0
10-14 years	0.2	0.0	0.2	0.1	0.2	0.0	0.2	0.0	0.0	0.0
15-19 years	0.3	0.3	0.2	0.3	0.2	0.5	0.3	0.2	0.6	0.0
20-24 years	0.9	0.4	0.9	0.6	0.9	0.4	0.9	0.2	0.5	0.0
25-29 years	0.9	0.9	1.4	1.0	0.5	0.8	0.6	0.6	0.5	1.5
30-34 years	2.2	1.6	2.8	1.9	2.2	1.7	1.1	1.1	3.0	1.0
35-39 years	6.4	5.4	6.8	6.4	12.7	7.1	3.1	2.6	4.8	4.2
40-44 years	18.8	12.7	20.8	14.7	34.1	18.0	8.3	5.8	8.6	7.8
45-49 years	42.5	24.8	44.1	28.4	90.7	35.0	19.1	11.1	20.3	16.9
50-54 years	95.2	54.4	101.4	63.8	196.7	70.1	40.2	21.7	29.7	28.2
55-59 years	190.3	104.3	203.3	123.0	333.5	129.6	93.2	39.2	80.0	42.0
60-64 years	327.9	165.9	356.0	196.2	523.6	195.2	154.7	62.8	121.6	61.4
65-69 years	496.5	248.5	535.2	291.7	679.8	263.2	274.0	96.6	209.2	95.4
70-74 years	646.1	311.0	676.0	359.1	919.5	287.1	416.1	138.3	360.1	119.2
75-79 years	685.7	315.9	705.5	350.6	860.1	272.4	507.8	158.5	524.6	266.8
80-84 years	665.9	262.8	674.3	279.9	784.5	211.9	558.8	182.9	590.4	333.2
85+ years	555.8	199.7	562.6	200.8	535.4	175.6	532.8	214.0	486.2	166.1

Rates are per 100,000.

Prepared by the Texas Department of State Health Services, Cancer Epidemiology
and
Surveillance Branch, Texas Cancer Registry. Data Request # 10385 (table 2) 11/19/2010
Incidence Source: Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3, .

Lymphoma Cancer Rates per 1,000,000

u:\datareq\2010\20385tbl.xls

Texas Statewide Age-Specific Rates, Average Annual Rates, 1995-2000
Selected Age Groups, and by race and ethnicity and sex (Data Release #10385)
All Childhood Lymphomas (ICCC-3: II Lymphomas and reticuloendothelial neoplasms)
Rates are per 1,000,000

	All Races		White Non-Hispanic		Black		Hispanic		Other Race	
	Male Rate*	Female Rate*	Male Rate*	Female Rate*	Male Rate*	Female Rate*	Male Rate*	Female Rate*	Male Rate*	Female Rate*
00-09 years	15.4	10.3	15.3	8.6	11.1	8.4	16.4	13.1	15.7	3.4
00-04 years	11.4	13.0	12.9	12.6	3.2	6.4	12.6	16.2	6.0	0.0
05-09 years	19.2	7.8	17.7	4.8	18.7	10.4	20.1	10.1	25.0	6.6

*Rates are per 1,000,000 and age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130) standard.

Prepared by the Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry. Data Request # 10385 11/16/2010
Incidence Source: Texas Department of State Health Services, Cancer Epidemiology and
Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3,
.

Relative Risk Calculations Provided by the Office of Research and Development (Drs Glenn & Bateson)

Calculation of relative risk as the ratio of Observed cases to Expected cases adjusting for Gender, Race and Age

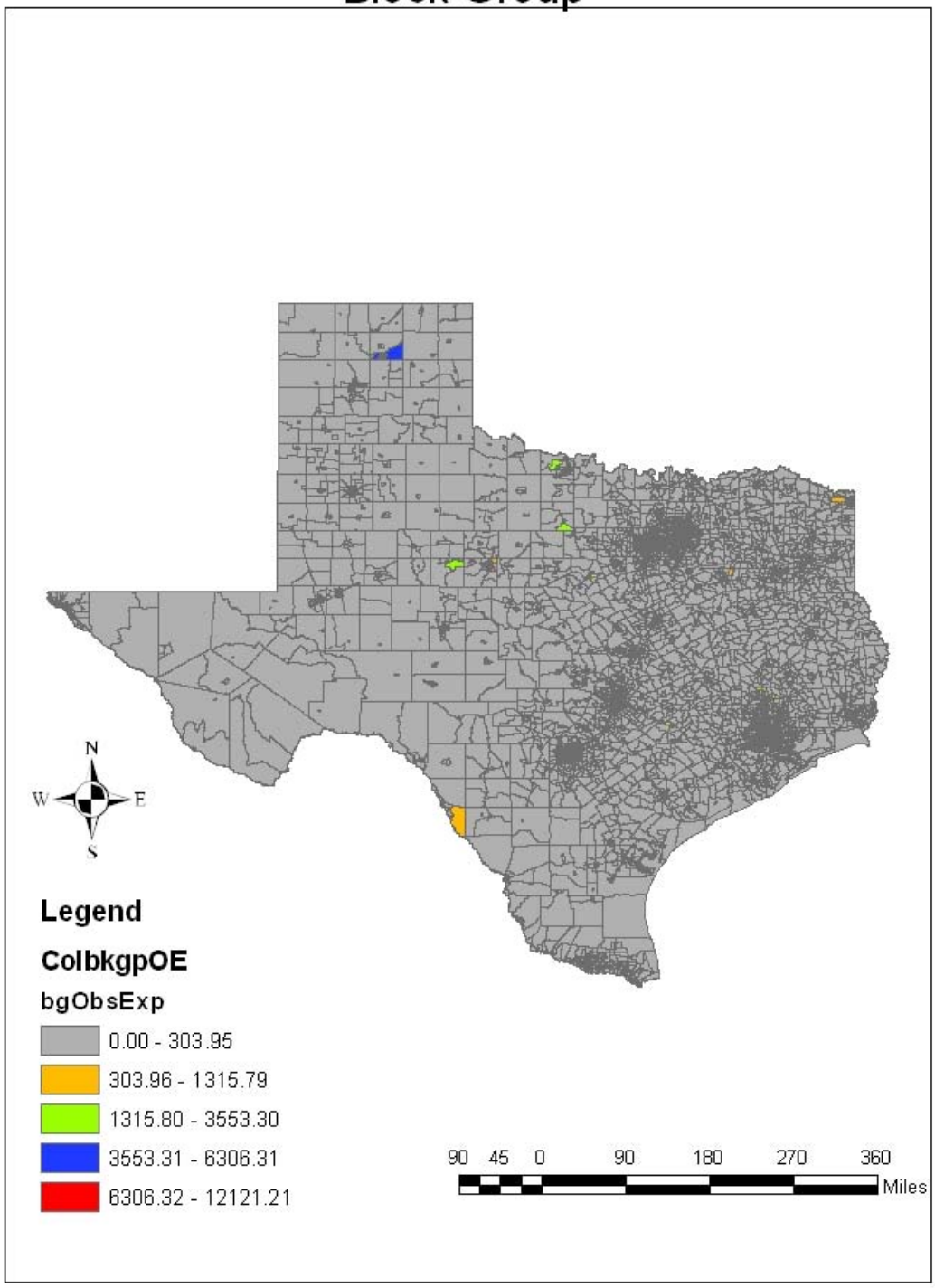
Cancer	Zip/tract/ county	Gender	Race	Age	# in 2000 Census * 10	Gender, Race, and Age-specific Rate	Expected	Observed	Sum of Expected per Geo. Unit	Sum of Observed per Geo. Unit	Gender, Race, and Age-specific RR RR=O/E	Gender, Race, and Age-standardized RR RR=O/E
Leukemia	1	Female	W	0-9	Find these #'s	Use TX or CA as needed	=F5*G5		=sum(H5:H11)	=sum(I5:I11)	=I5/H5	=K5/J5
Leukemia	1	Female	B	0-9								
Leukemia	1	Female	H	0-9								
Leukemia	1	Male	W	0-9								
Leukemia	1	Male	B	0-9								
Leukemia	1	Male	H	0-9								
Leukemia	2	Female	W	0-9	ETC……							

APPENDIX B

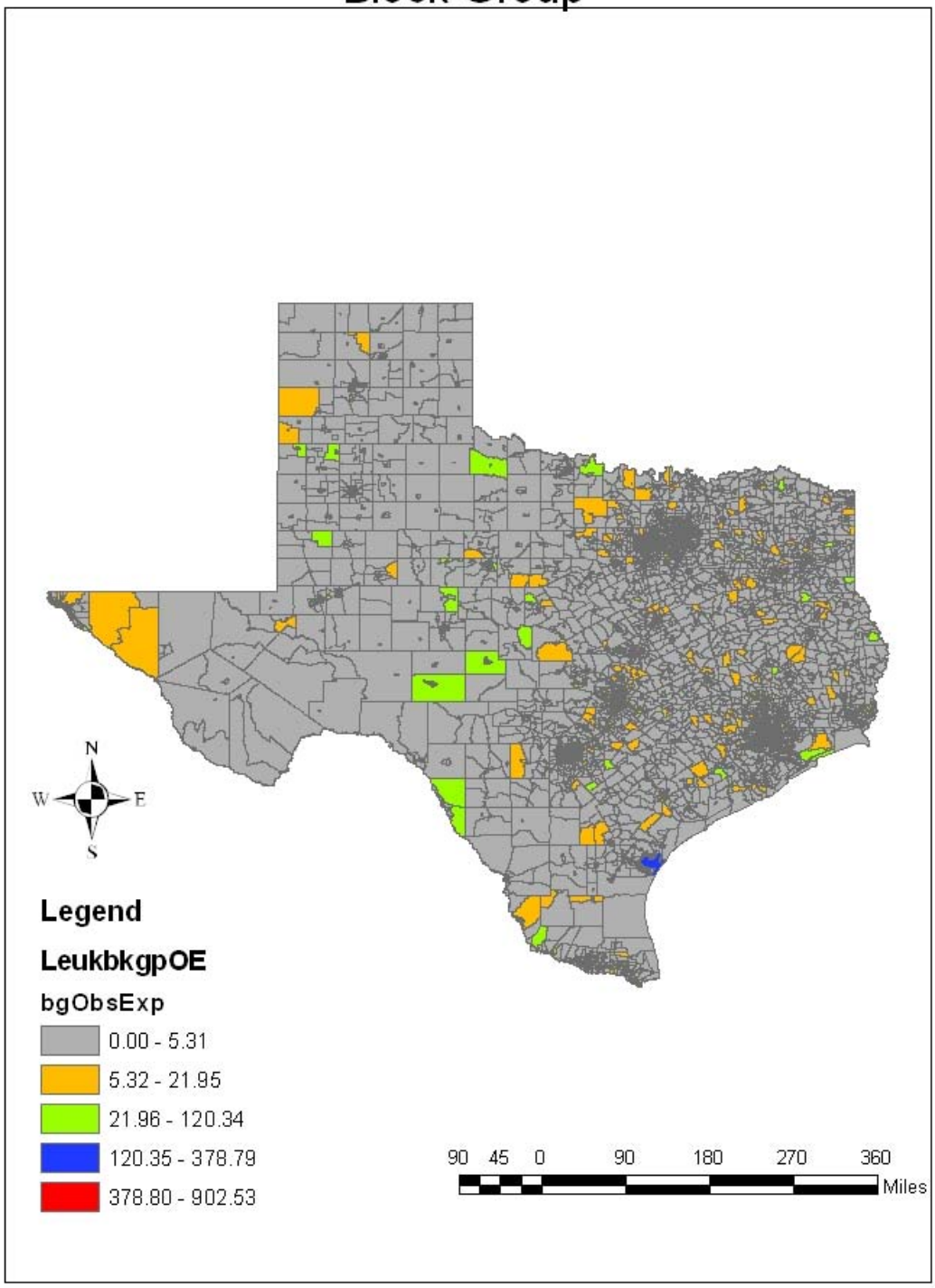
OBSERVED TO EXPECTED SCATTER PLOTS

The following are block group scatter plots of observe to expected ratios verses distance to emitters (in miles). These scatter plots can be found in Chapter 3 of the dissertation. The vertical axis has been logged in order to better visualize the data points. Additionally, maps that correspond with these scatter plots have been provided. Note that these maps correspond with all of the provided block group scatter plots of observe to expected ratios verses distance to emitters (in miles).

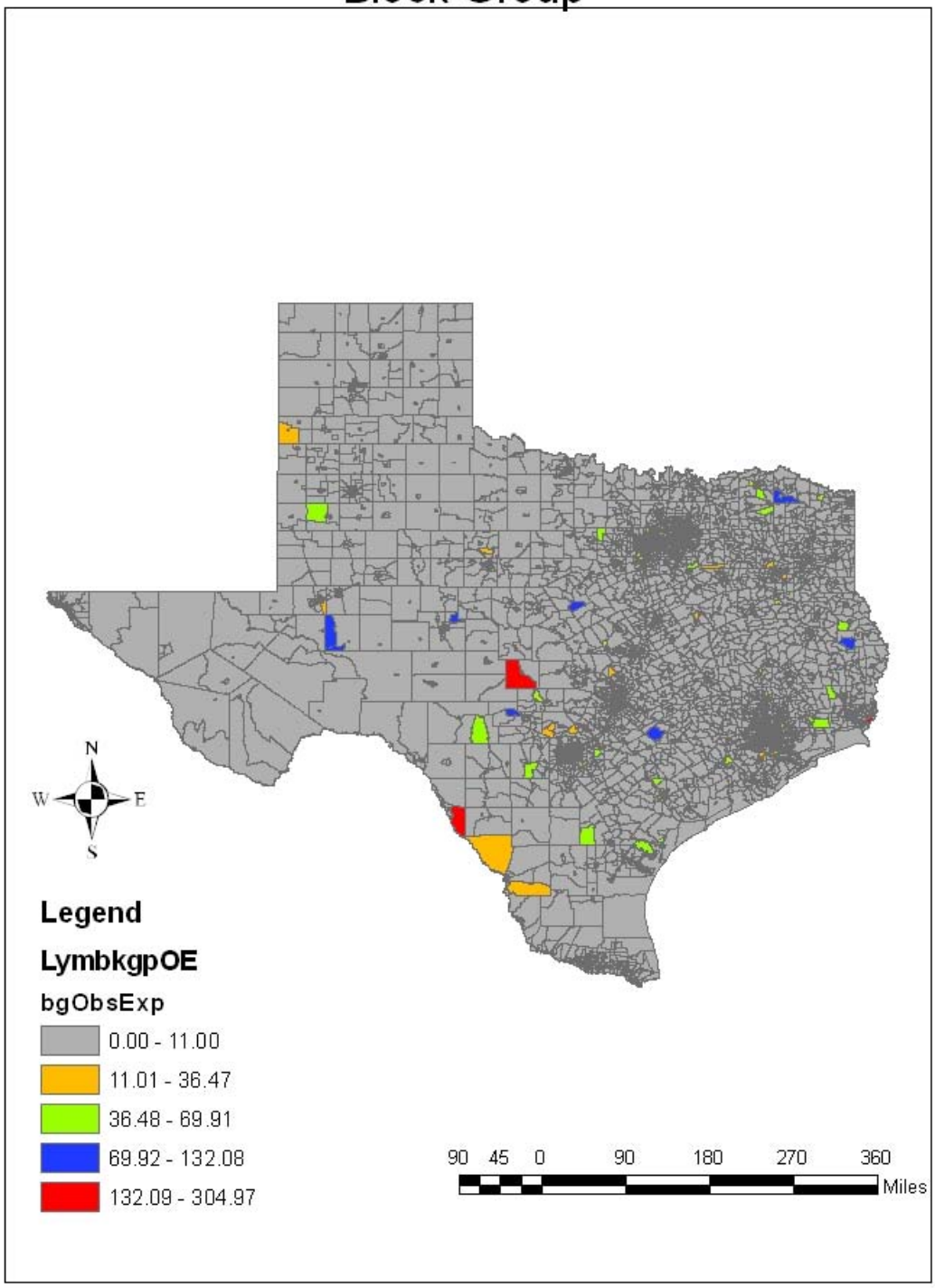
Observed to Expected Ratios of Colon Cases by Block Group



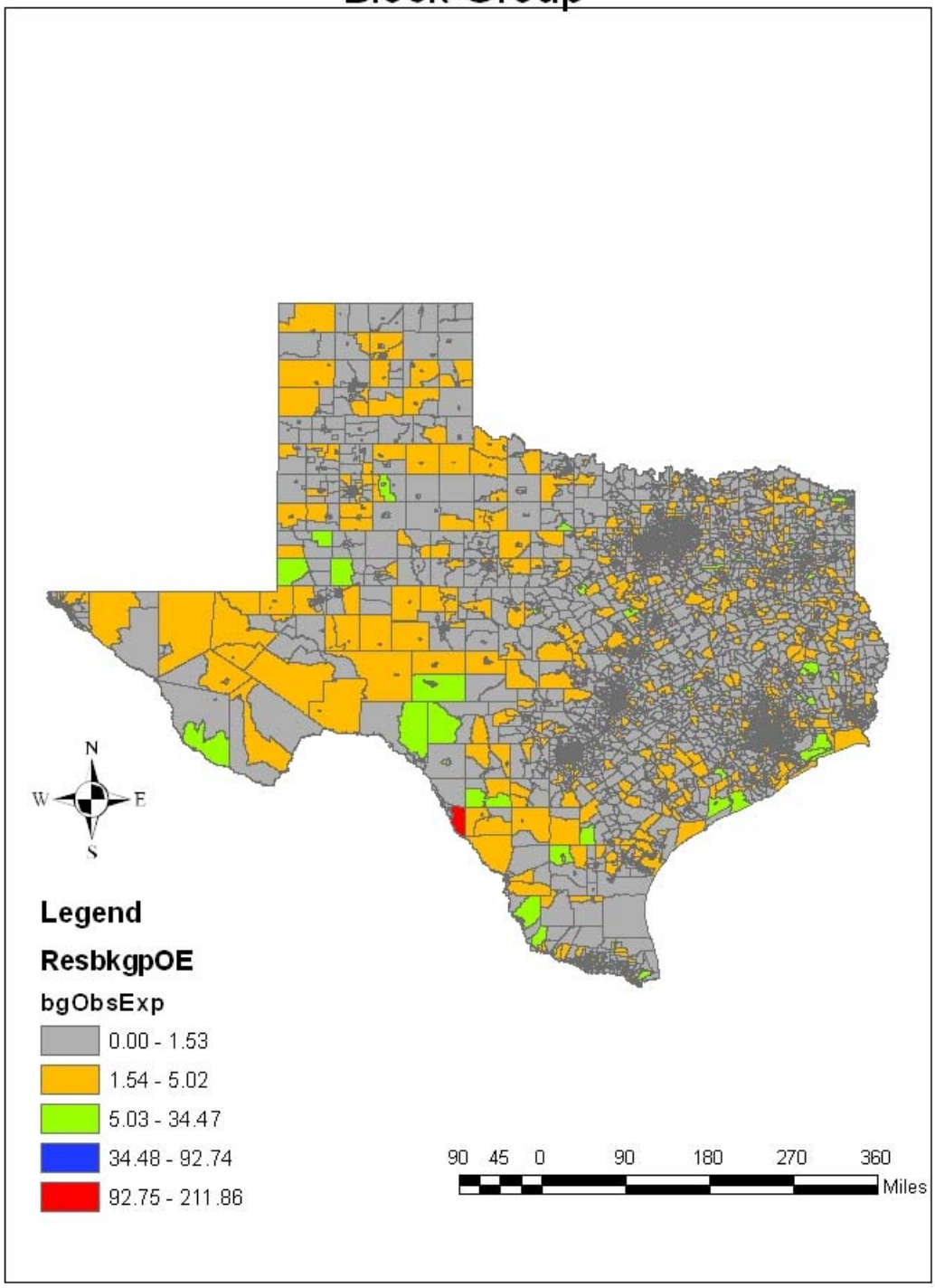
Observed to Expected Ratios of Leukemia Cases by Block Group



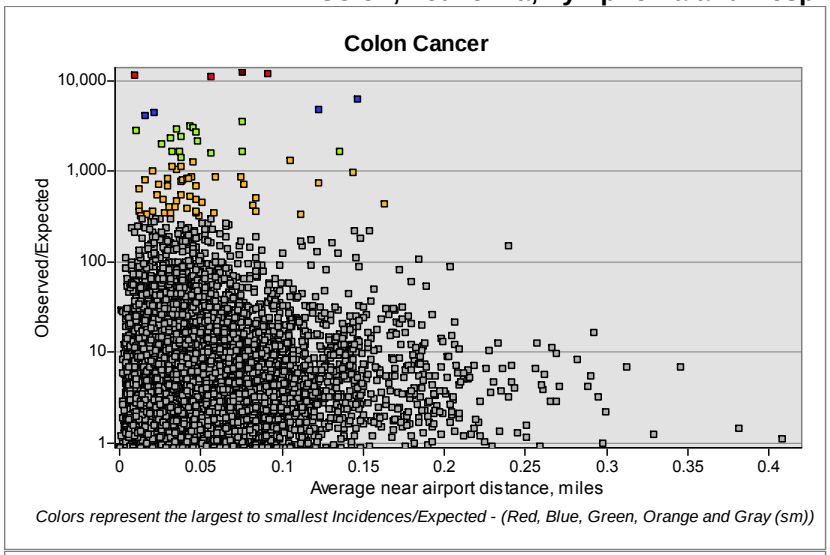
Observed to Expected Ratios of Lymphoma Cases by Block Group



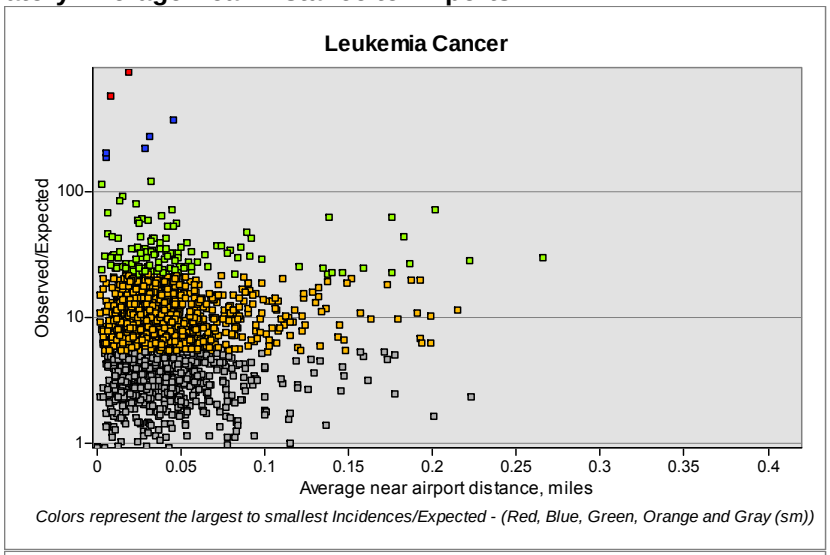
Observed to Expected Ratios of Respiratory Cases by Block Group



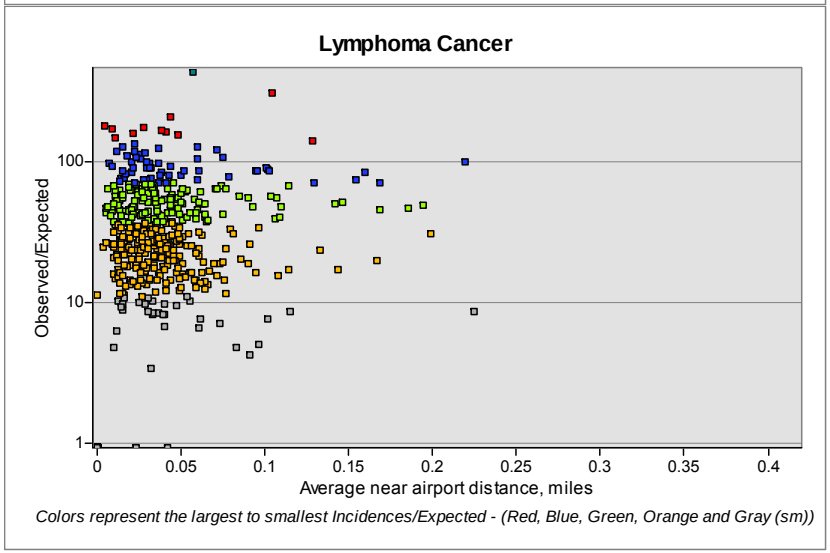
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Airports



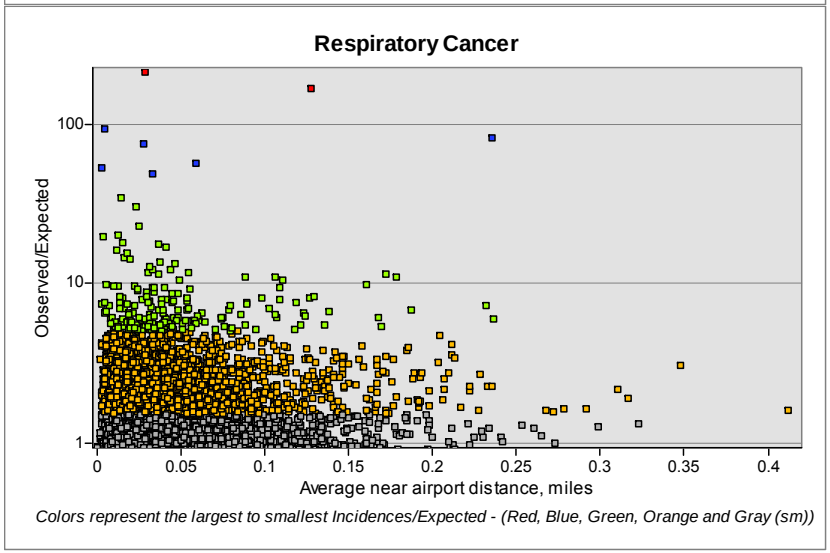
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))



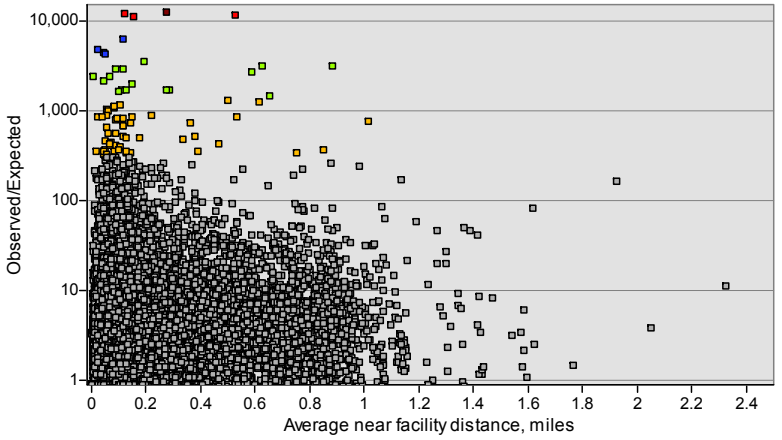
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

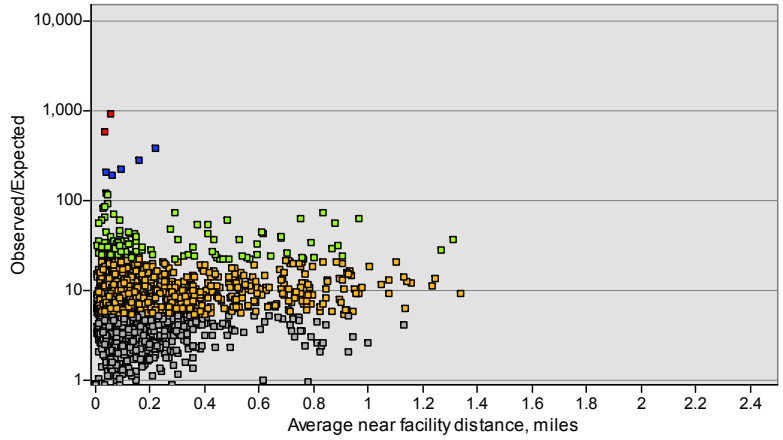
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Facilities

Colon Cancer



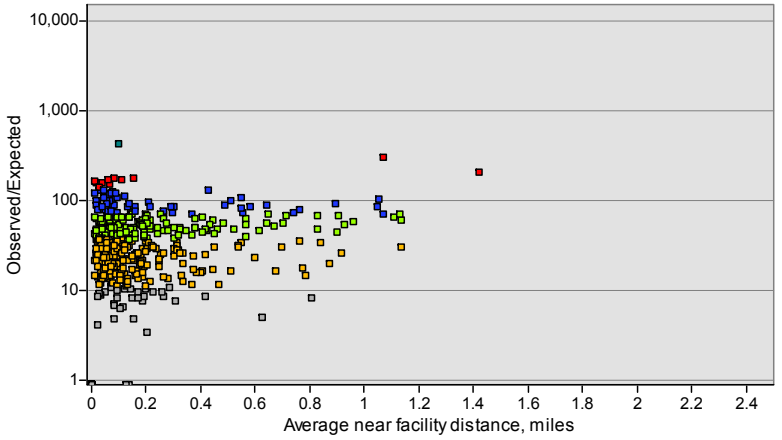
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



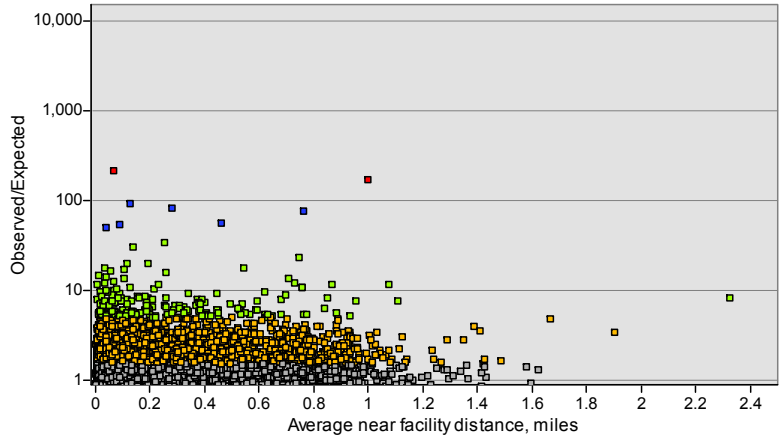
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

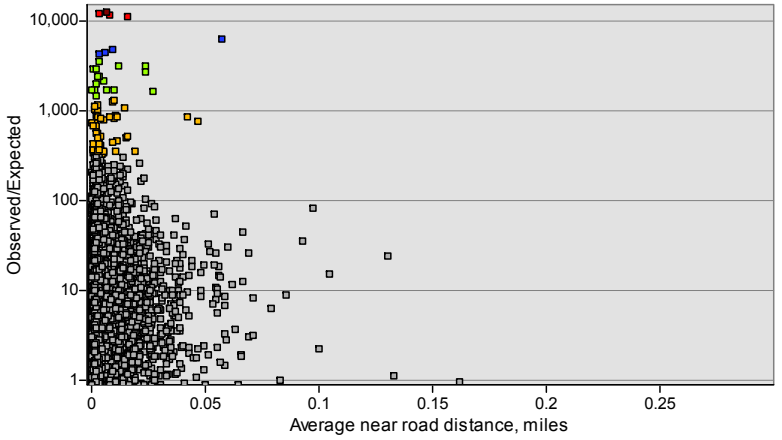
Respiratory Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

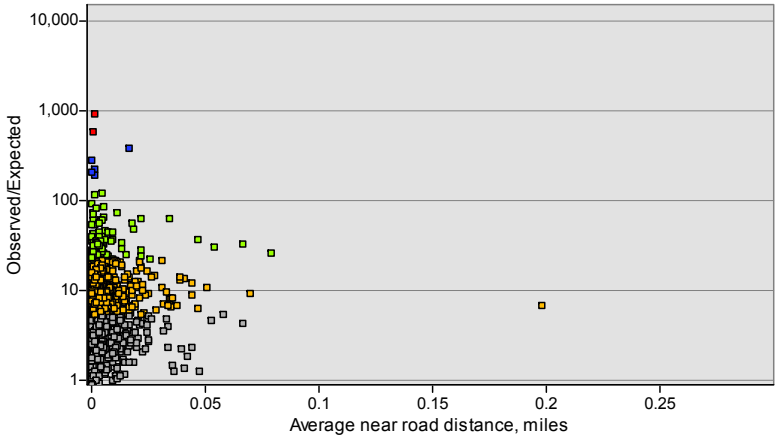
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Roads

Colon Cancer



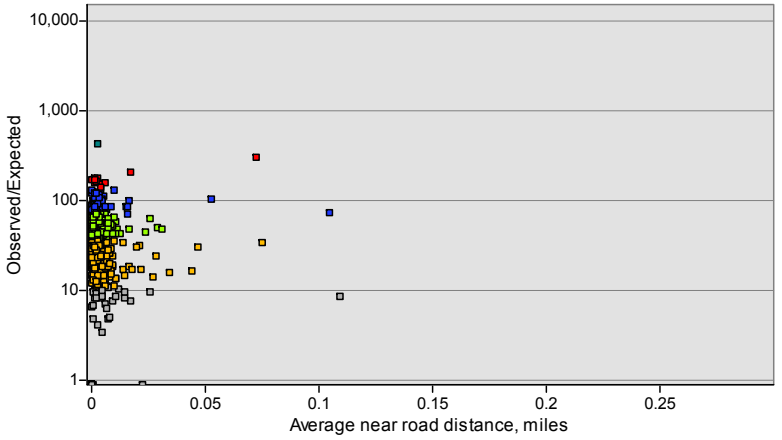
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



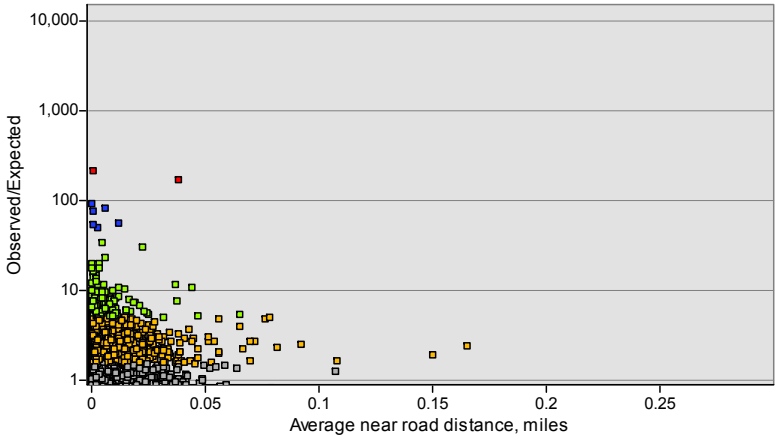
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

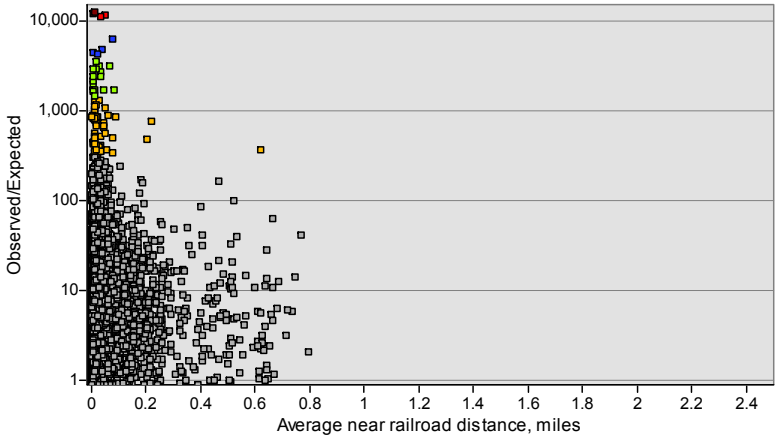
Respiratory Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

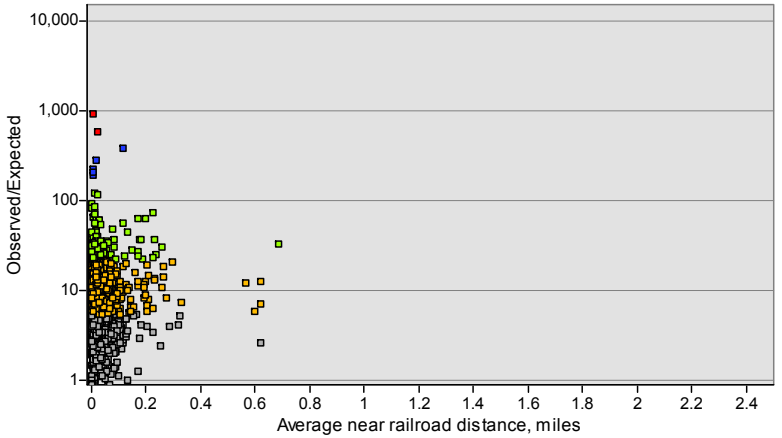
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Railroads

Colon Cancer



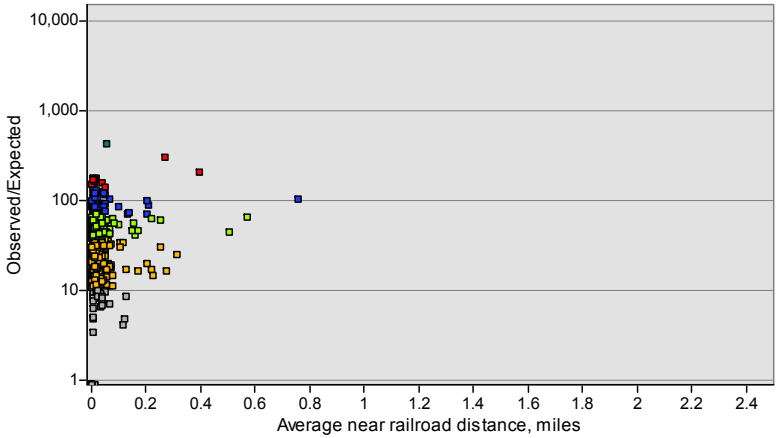
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



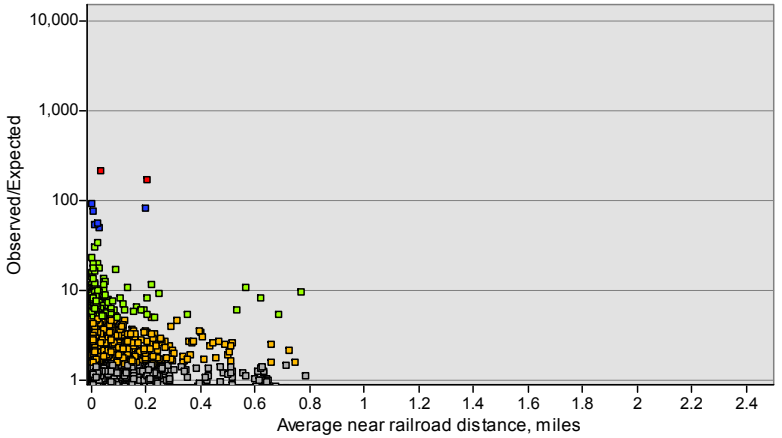
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

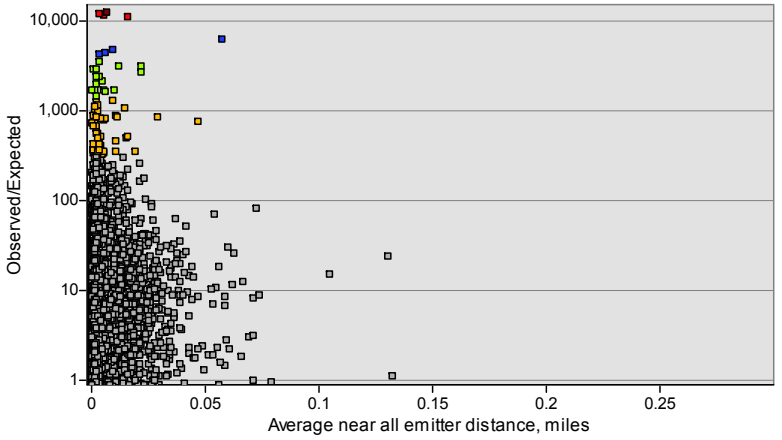
Respiratory Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

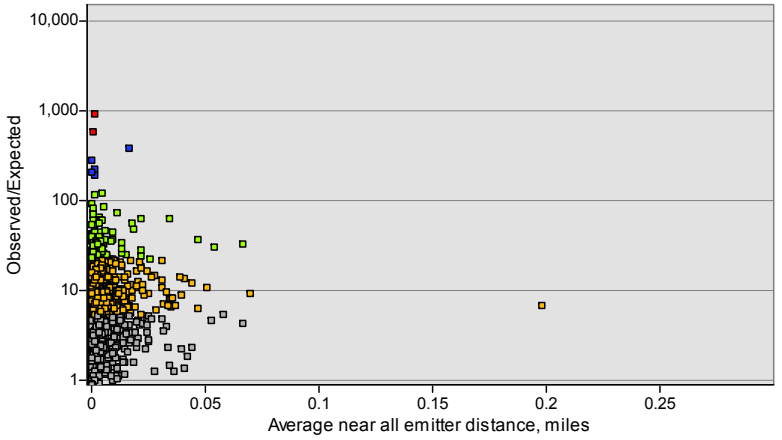
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to All Emitters

Colon Cancer



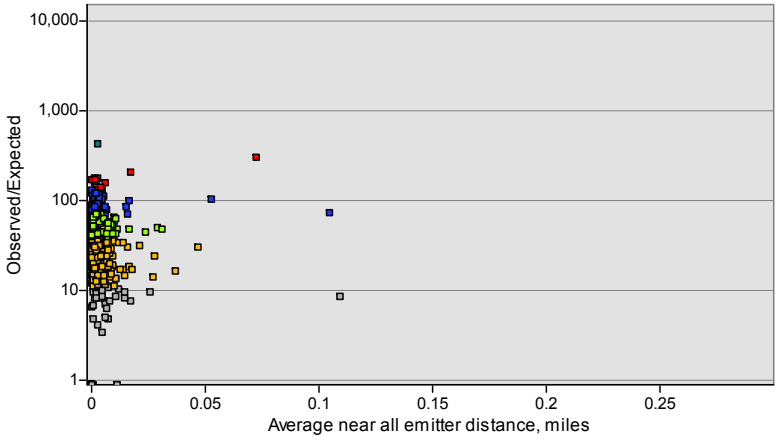
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



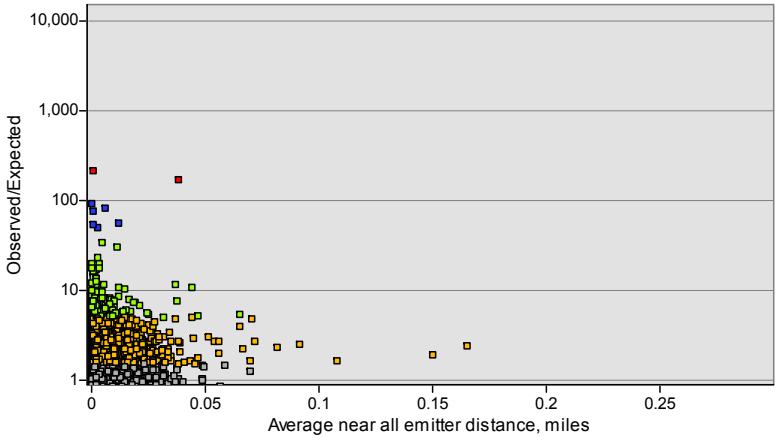
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



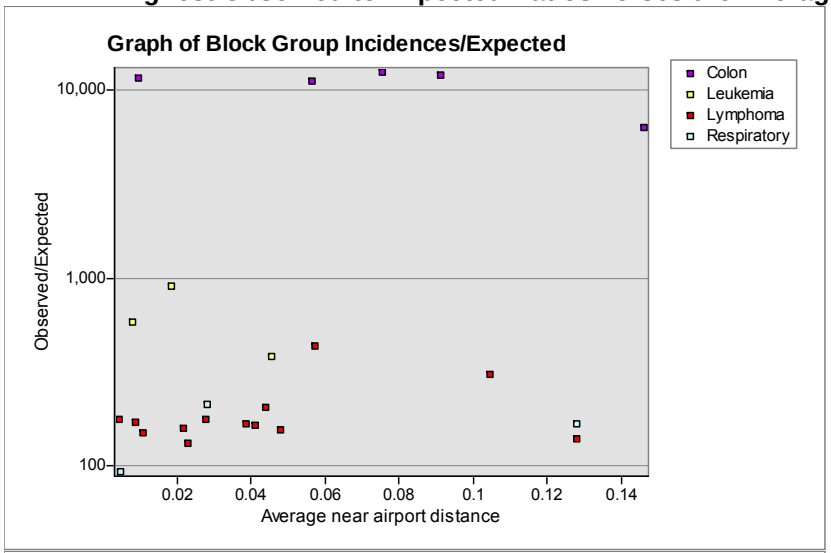
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

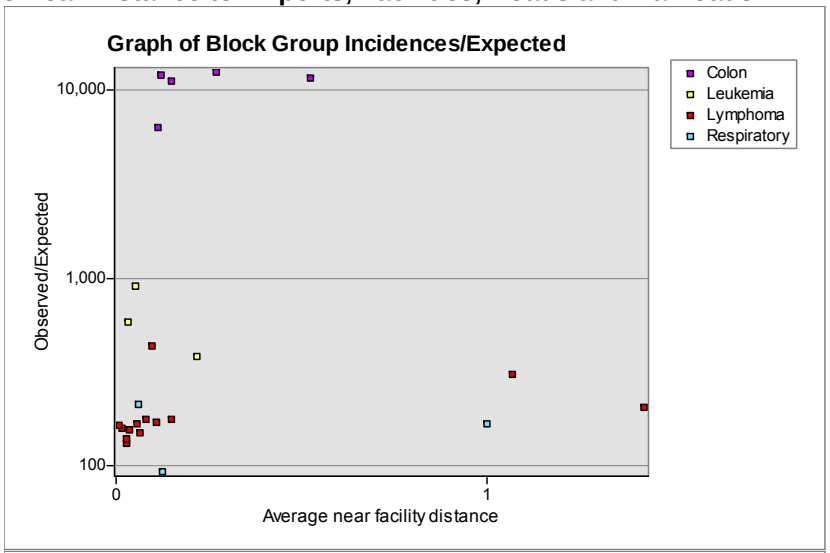
Respiratory Cancer

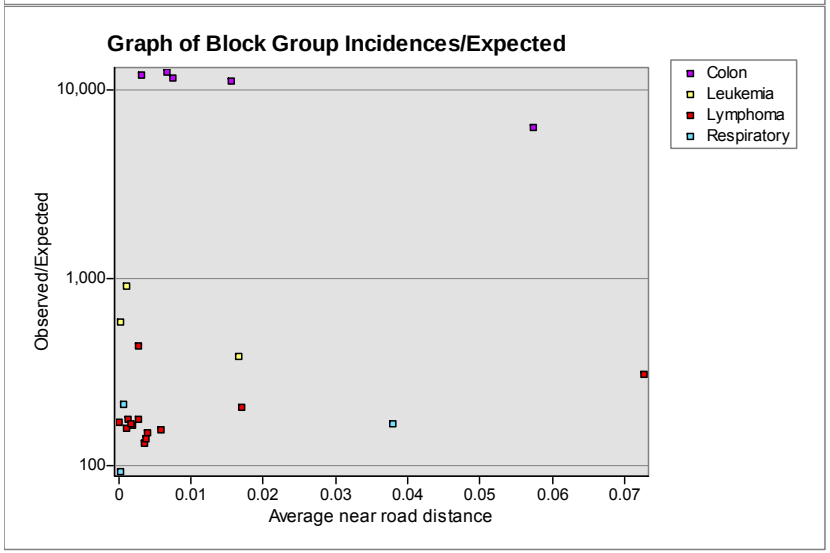


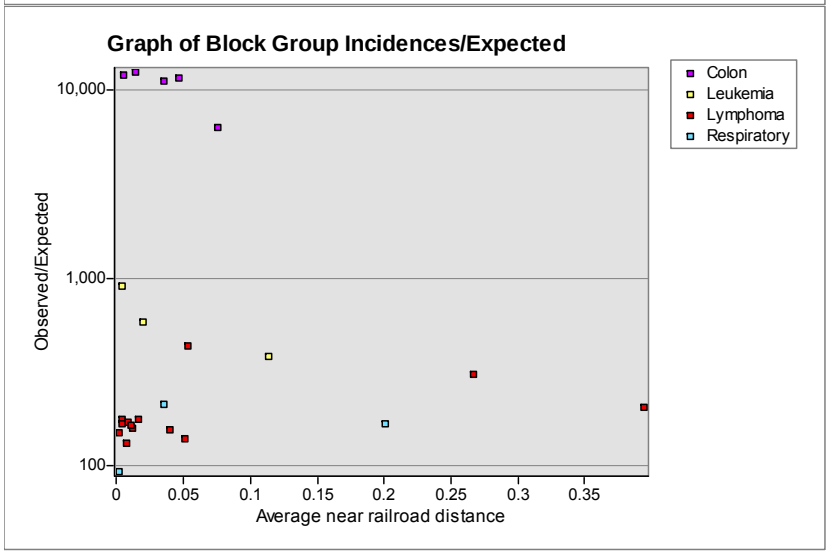
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Highest Observed to Expected Ratios verses the Average Near Distance to Airports, Facilities, Roads and Railroads









Highest Observed to Expected Ratios verses the Average Near Distance to All Emitters

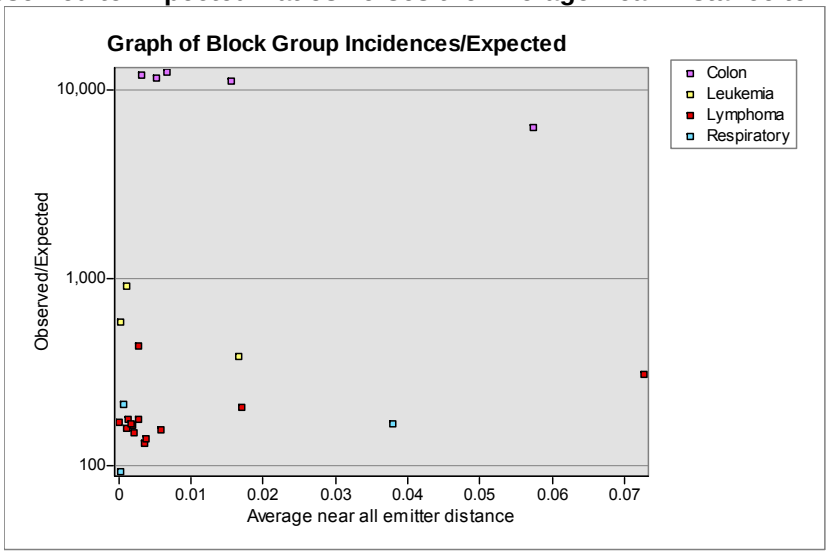
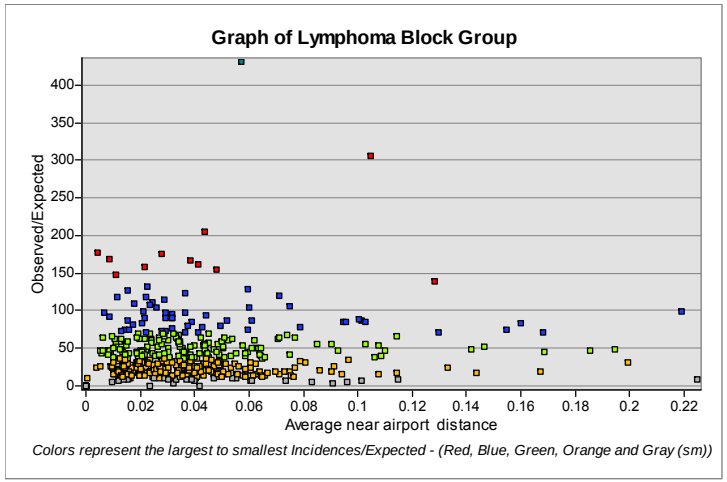


Chart with Highest Observed to Expected Ratio Averages verses the Average Near Distance to Airports, Facilities, Roads, Railroads and All Emitters

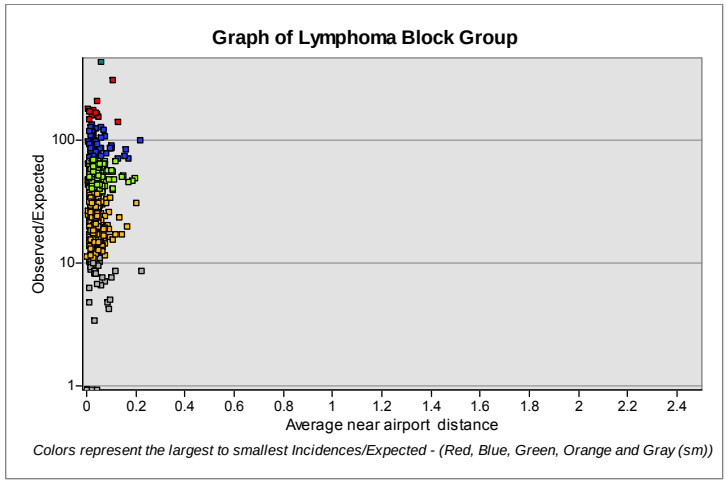
	Leuk-3 total	Lym-13 total	Res-3 total	Col-5 total
Ave_bgObsE	622.4175	194.1812	157.4899	10726.5665
Airport	0.0239	0.0428	0.0535	0.0757
Facility	0.1023	0.2449	0.3970	0.2373
Road	0.0060	0.0091	0.0129	0.0181
Railroad	0.0459	0.0670	0.0794	0.0352
All Emitters	0.0060	0.0090	0.0129	0.0176

The following are block group lymphoma scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

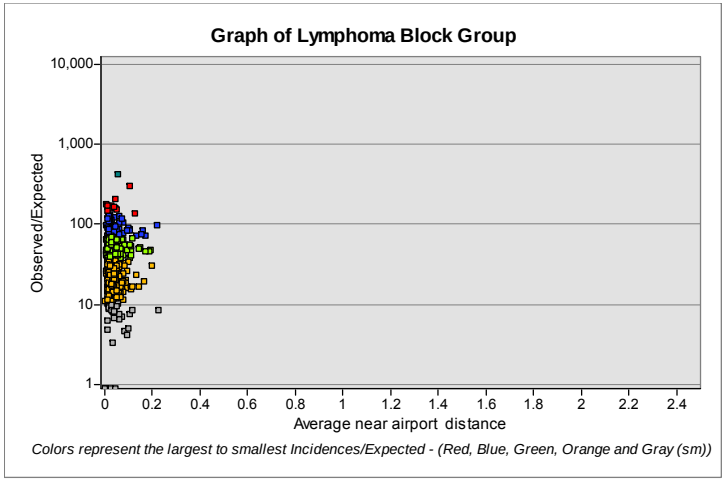
1)



2)

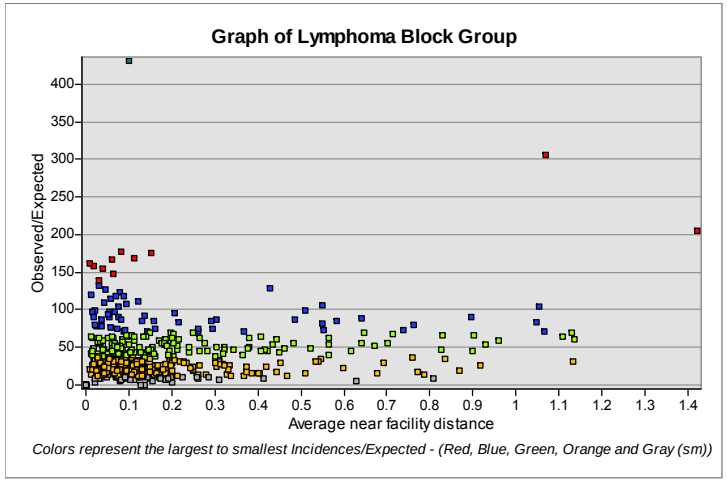


3)

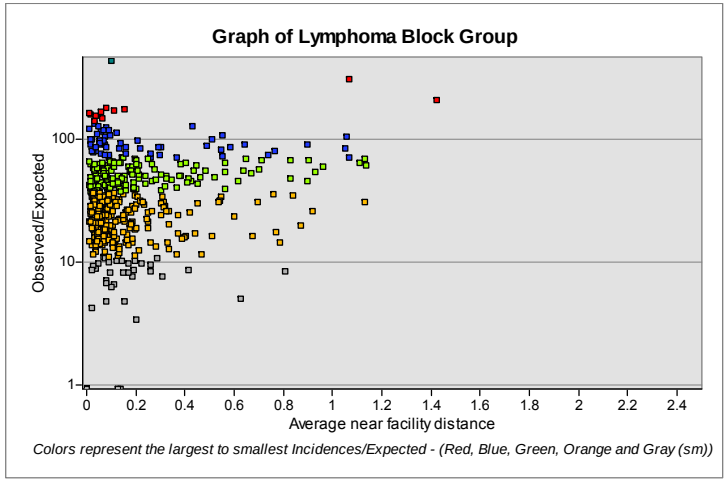


The following are block group lymphoma scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

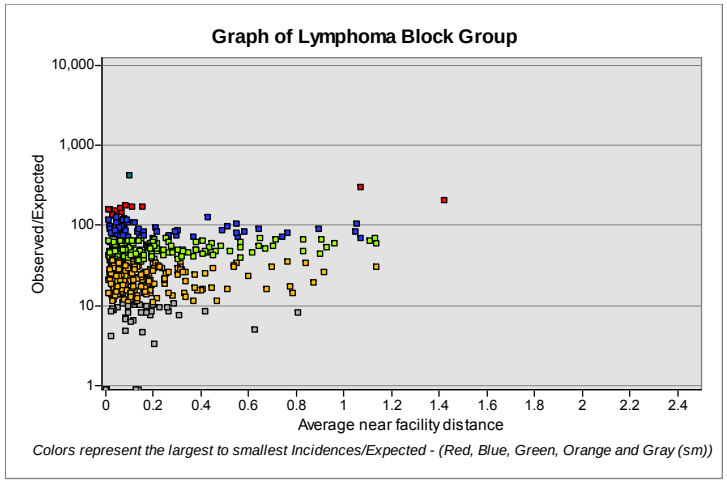
1)



2)

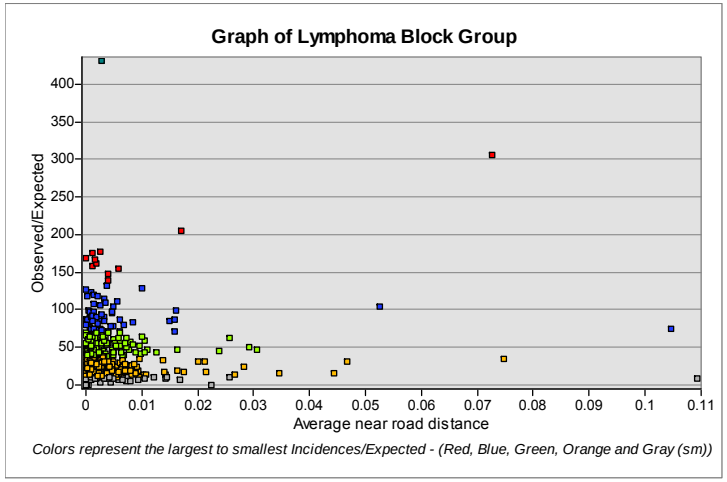


3)

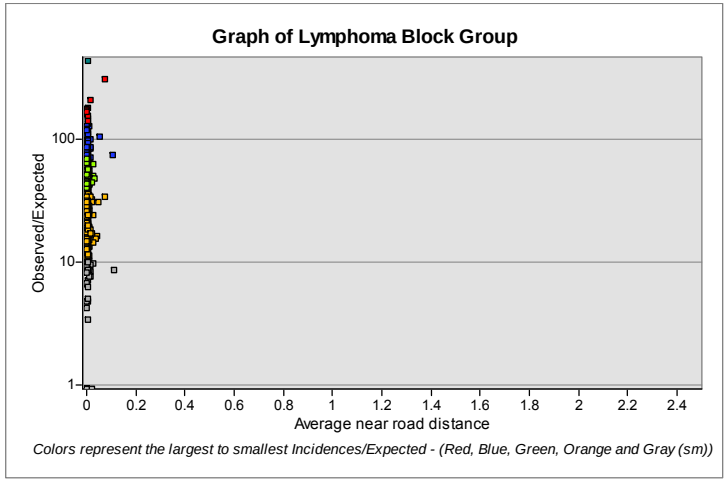


The following are block group lymphoma scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

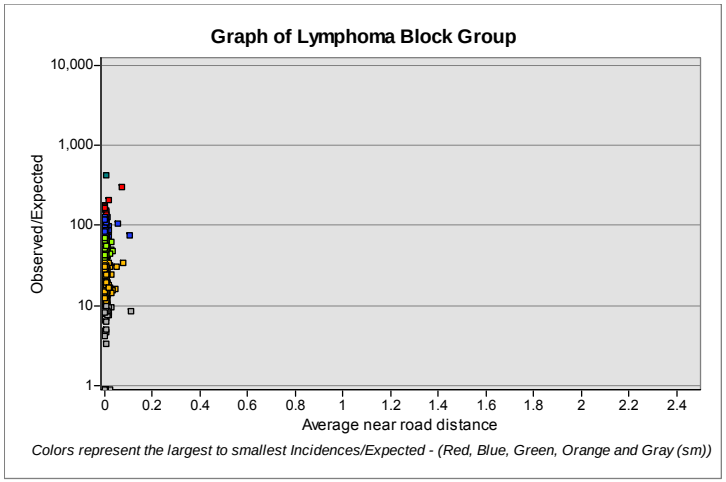
1)



2)

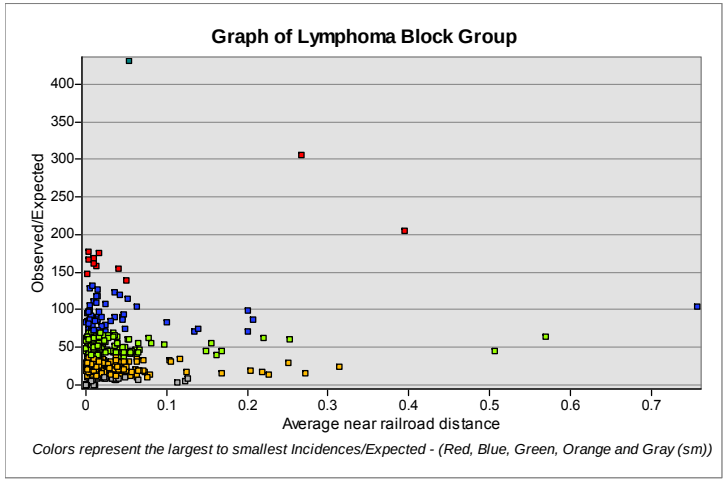


3)

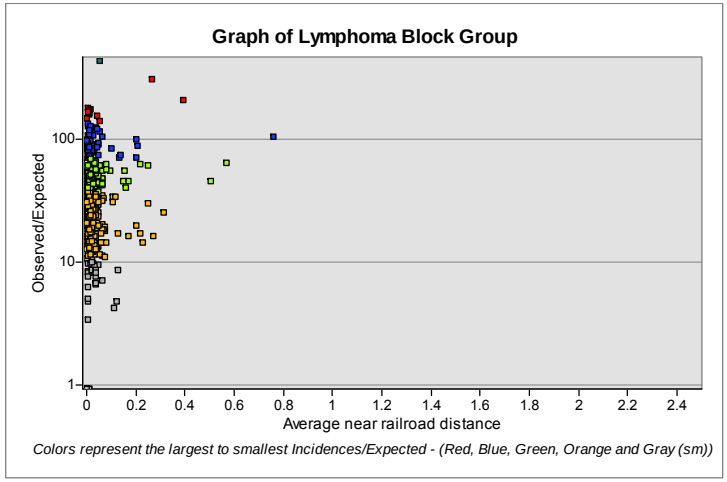


The following are block group lymphoma scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

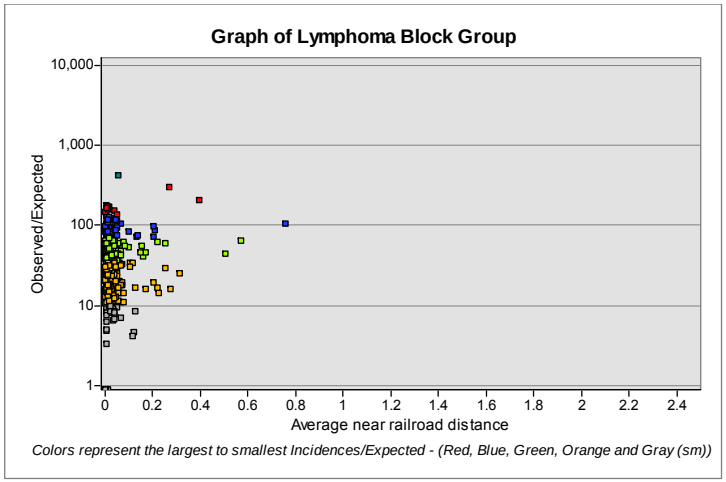
1)



2)

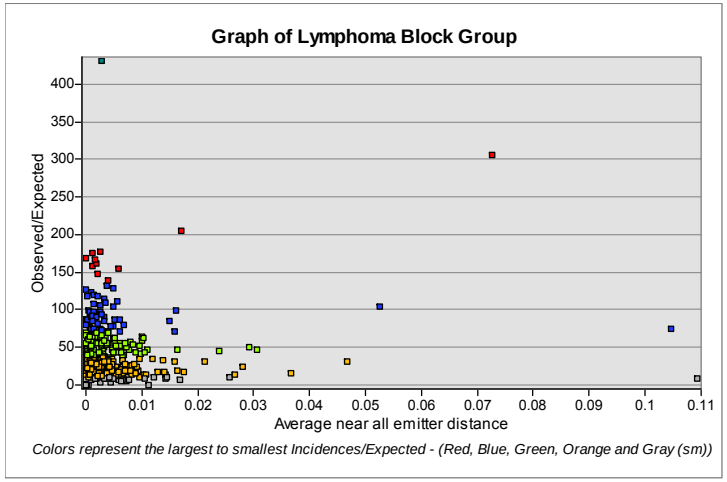


3)

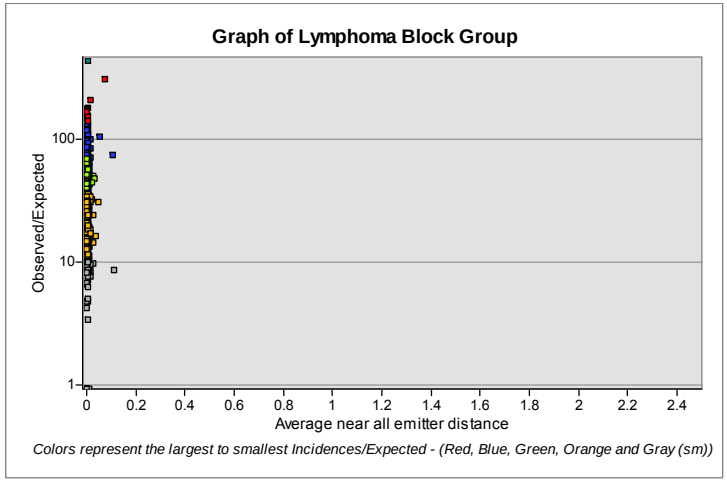


The following are block group lymphoma scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

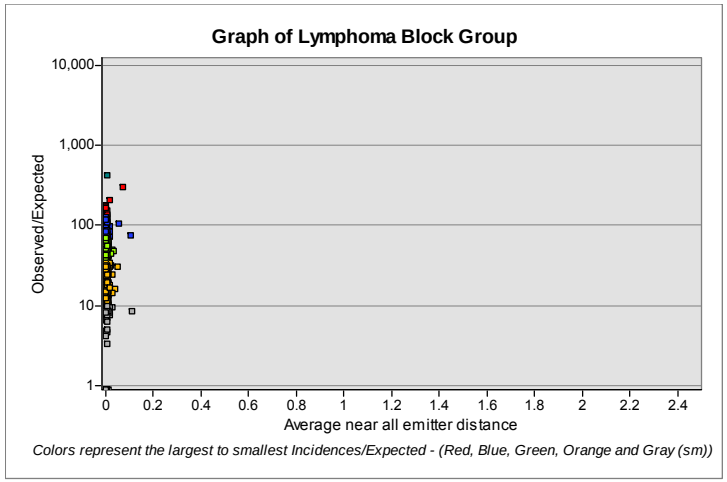
1)



2)

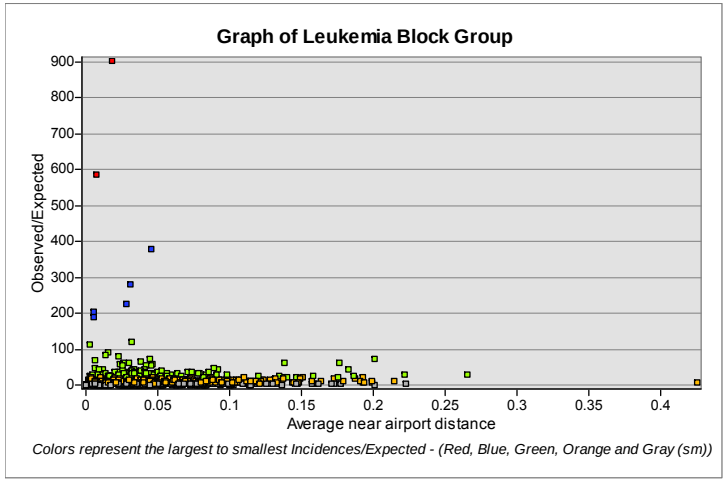


3)

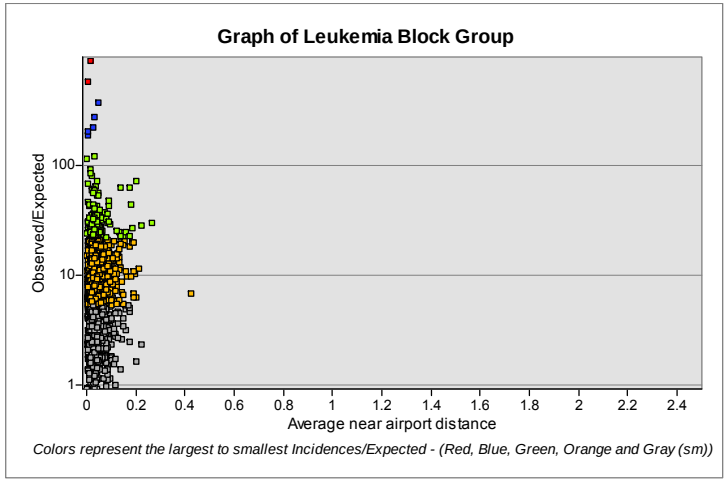


The following are block group leukemia scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

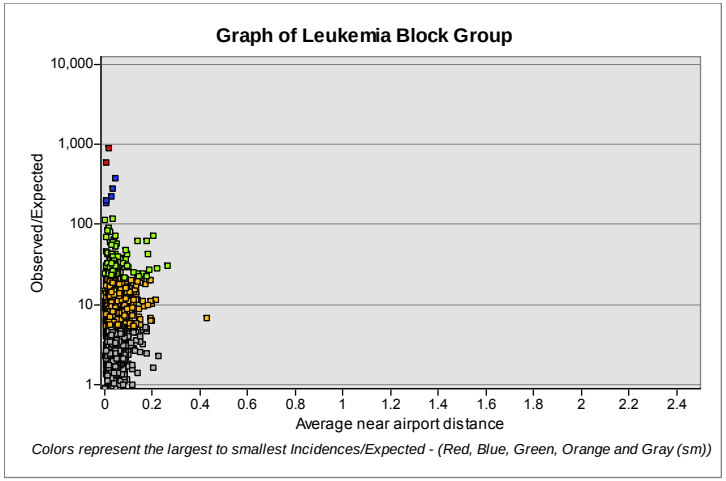
1)



2)

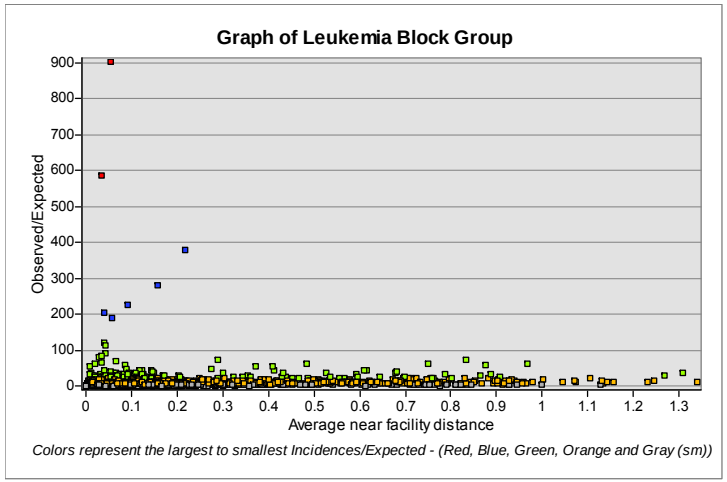


3)

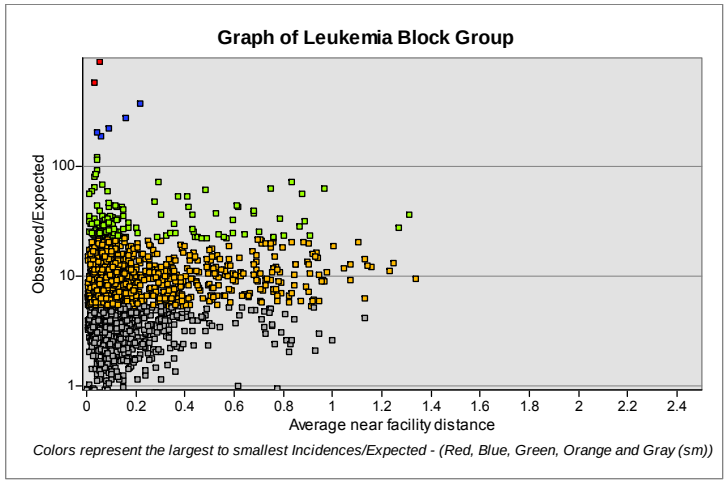


The following are block group leukemia scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

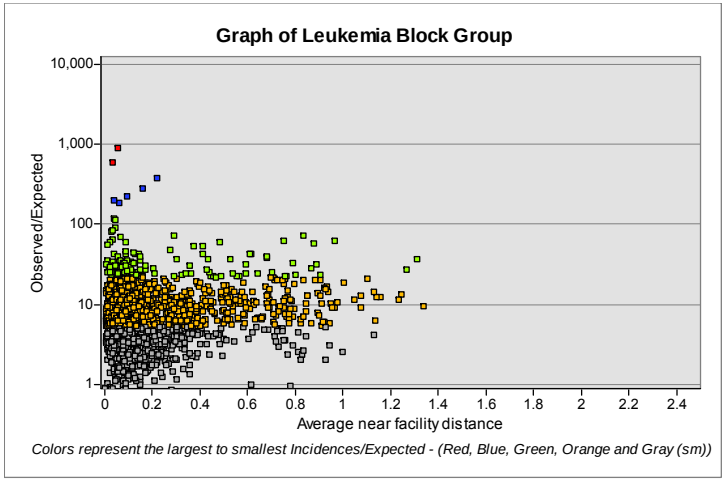
1)



2)

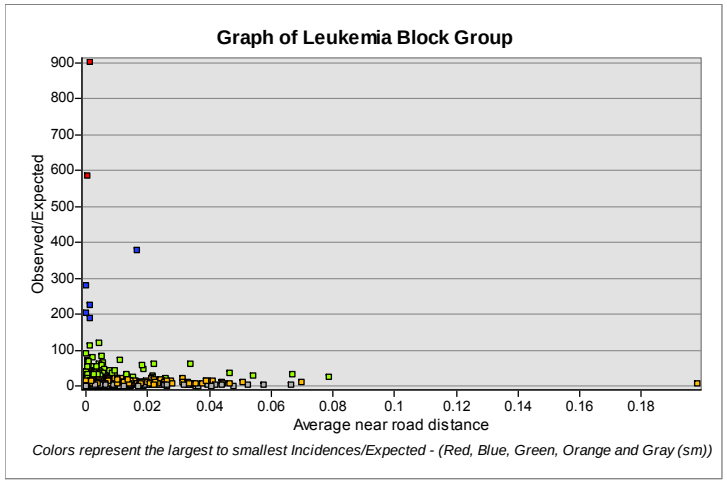


3)

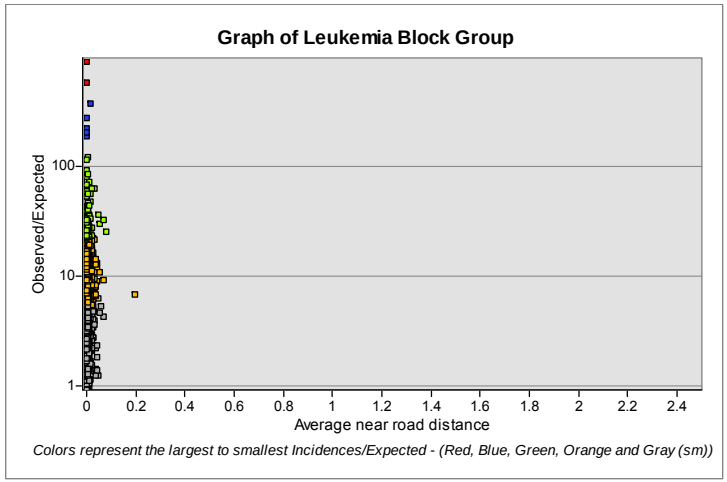


The following are block group leukemia scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

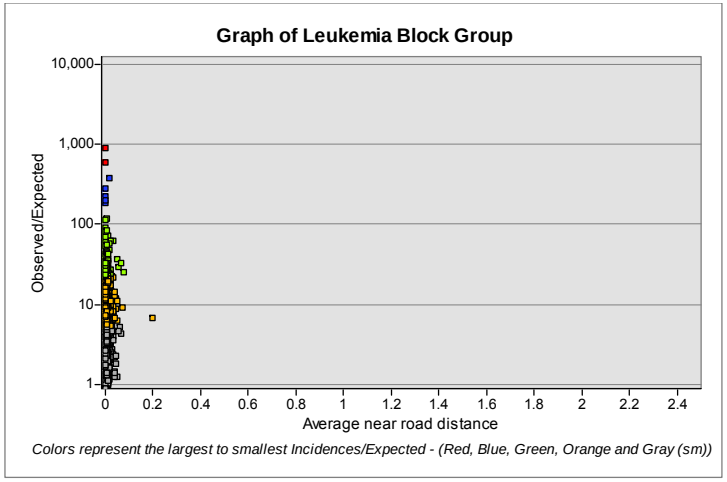
1)



2)

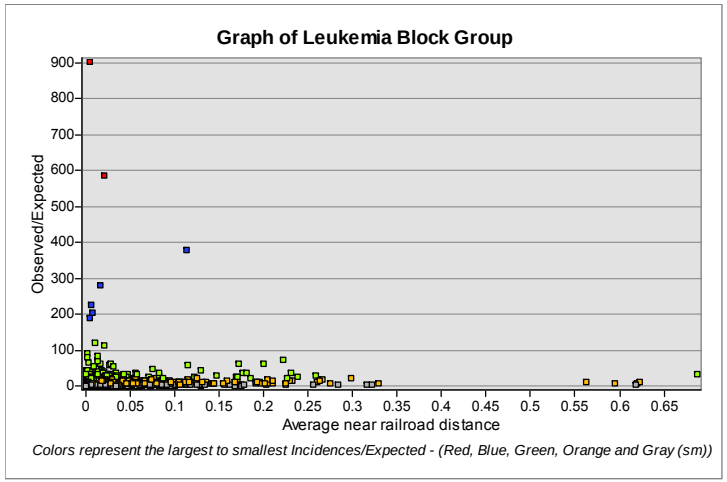


3)

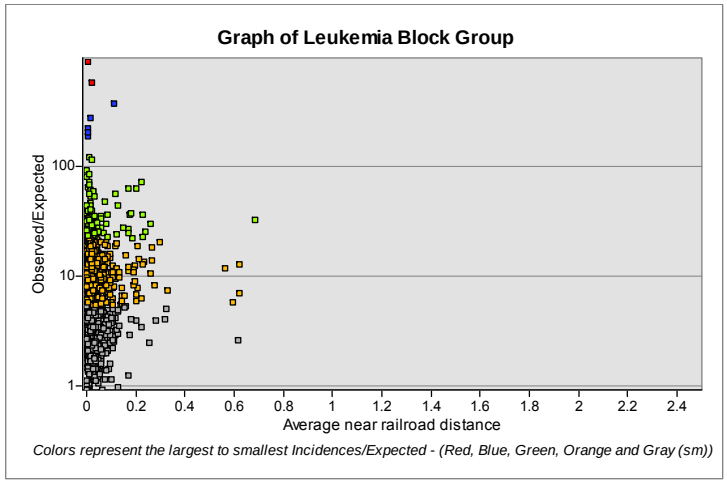


The following are block group leukemia scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

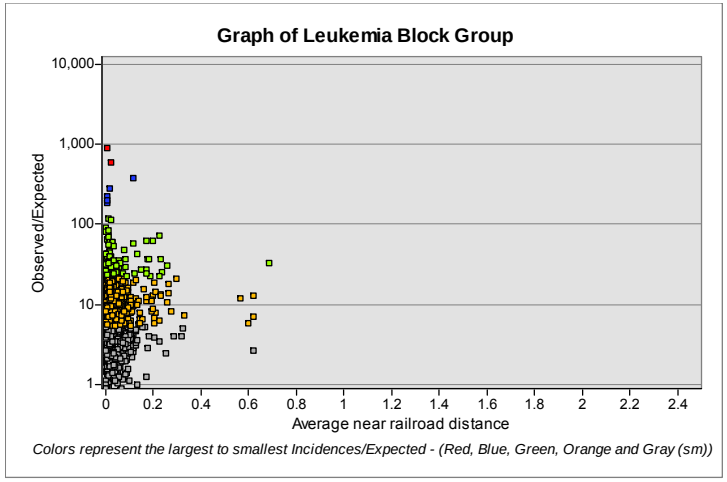
1)



2)

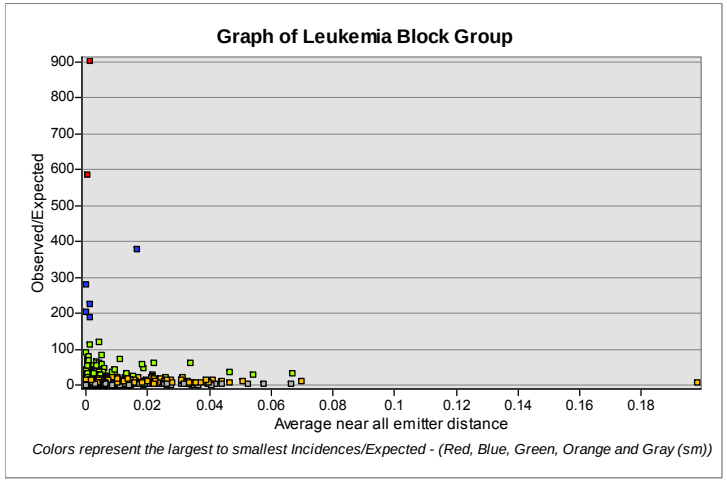


3)

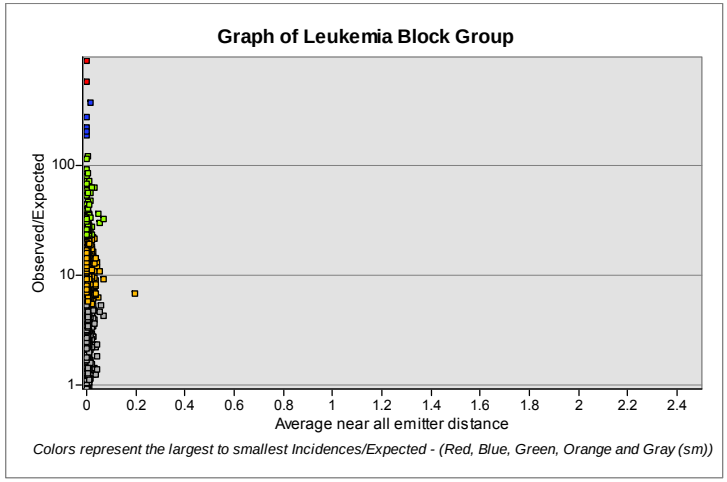


The following are block group leukemia scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

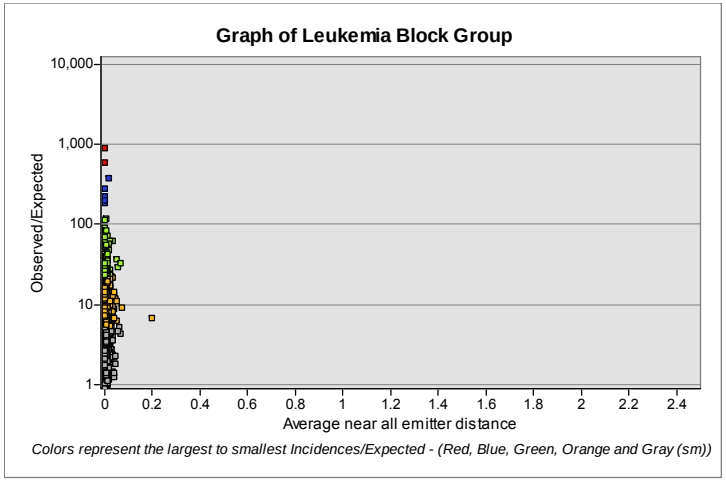
1)



2)

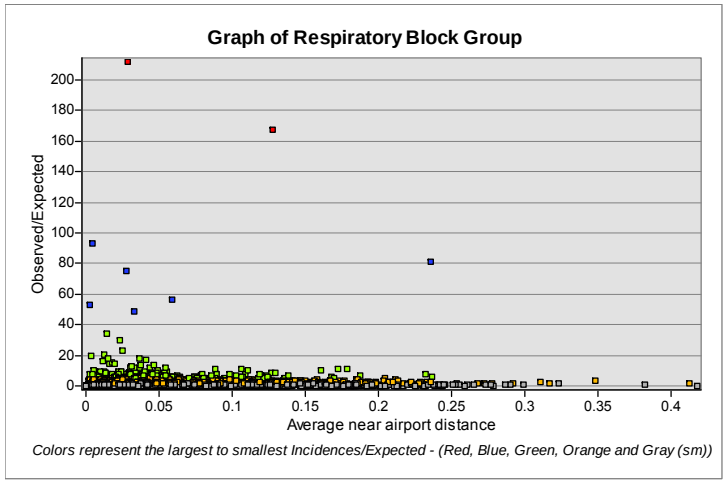


3)

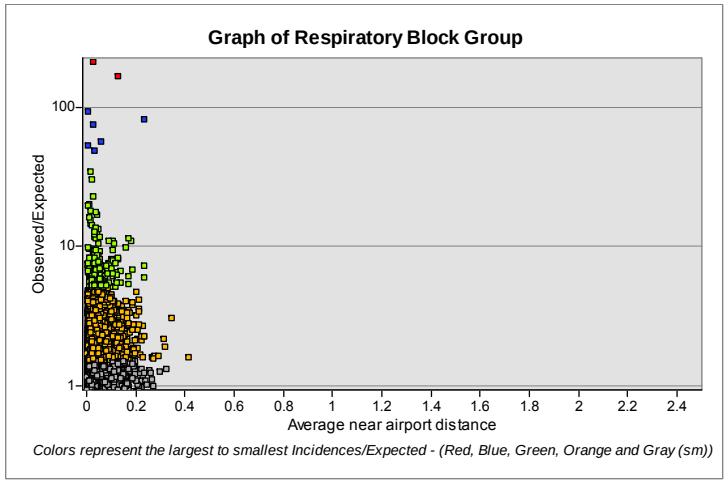


The following are block group respiratory scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

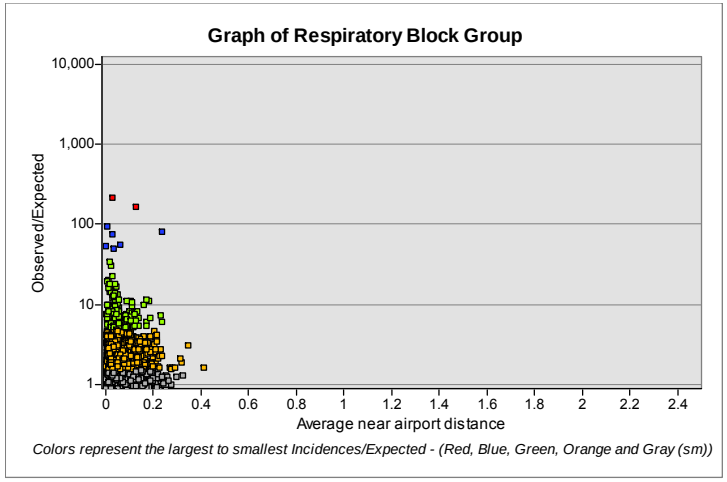
1)



2)

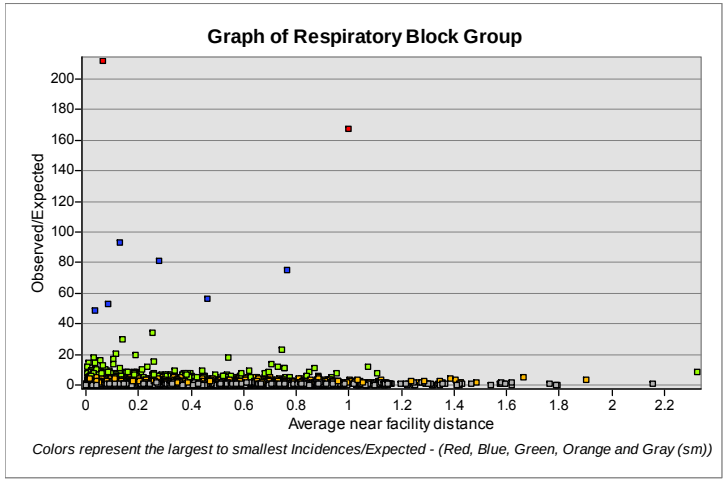


3)

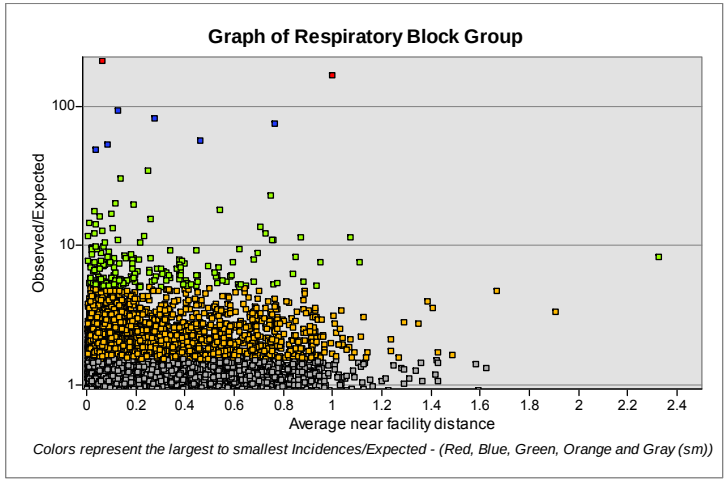


The following are block group respiratory scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

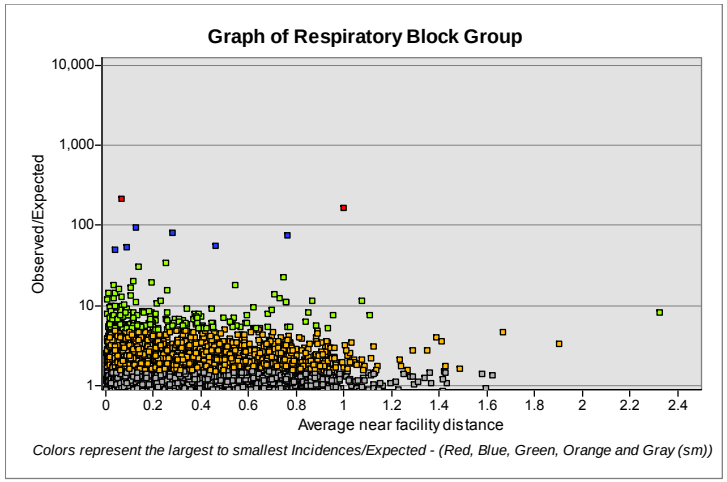
1)



2)

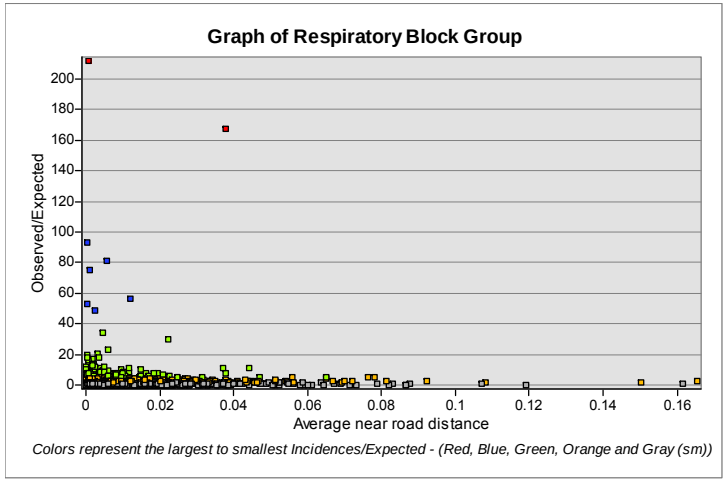


3)

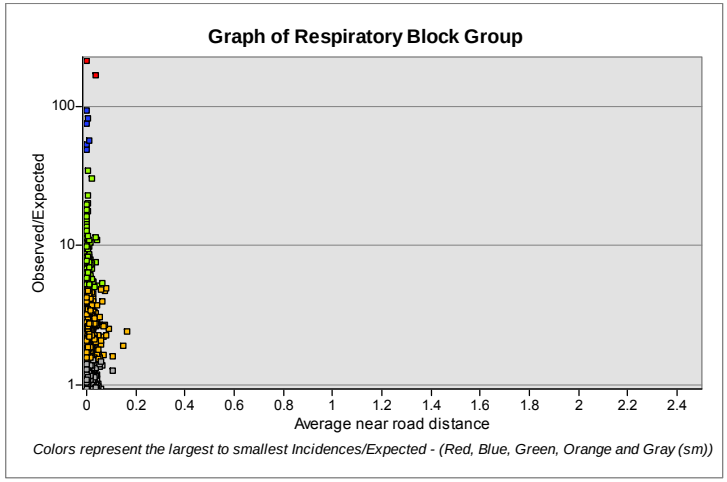


The following are block group respiratory scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

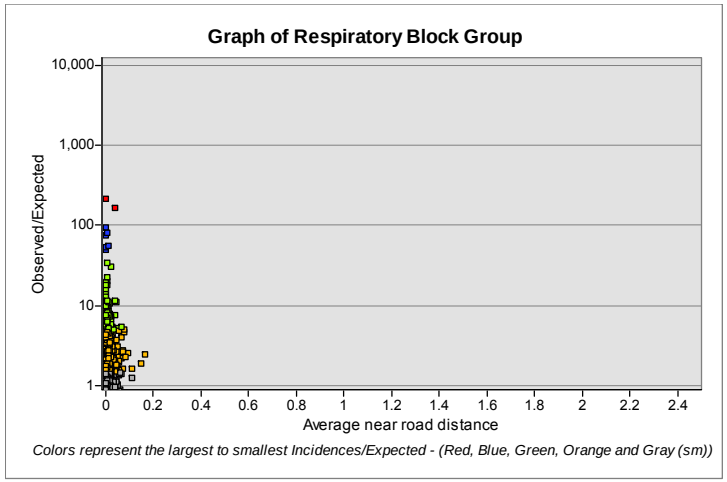
1)



2)

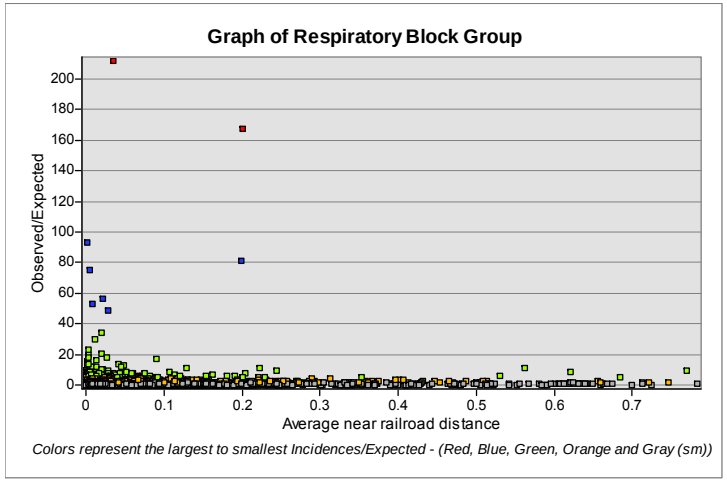


3)

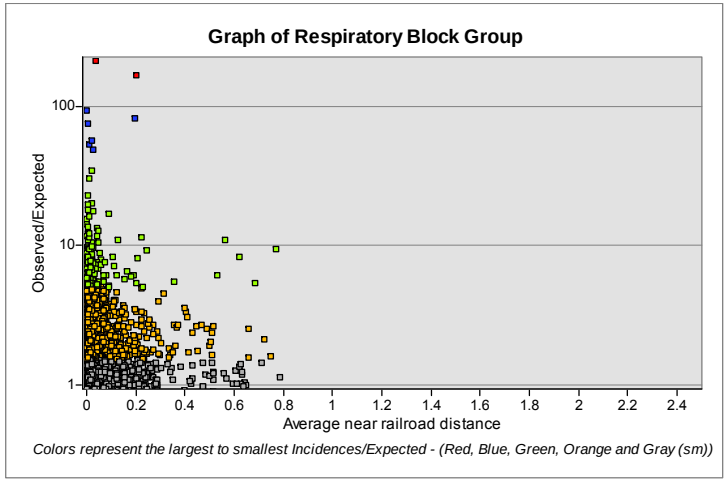


The following are block group respiratory scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

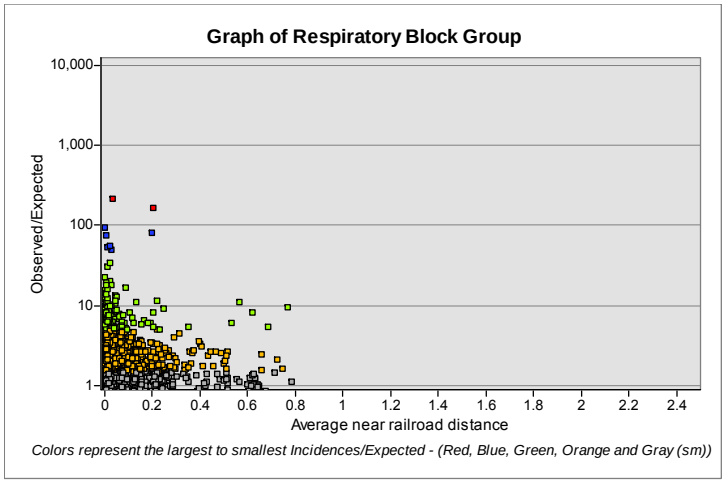
1)



2)

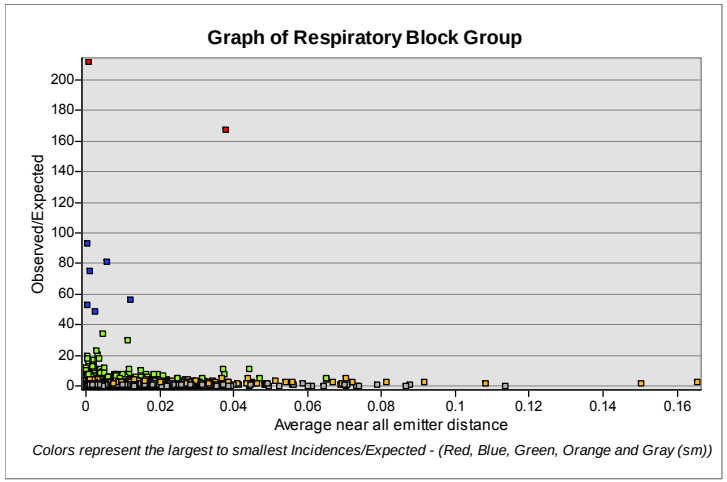


3)

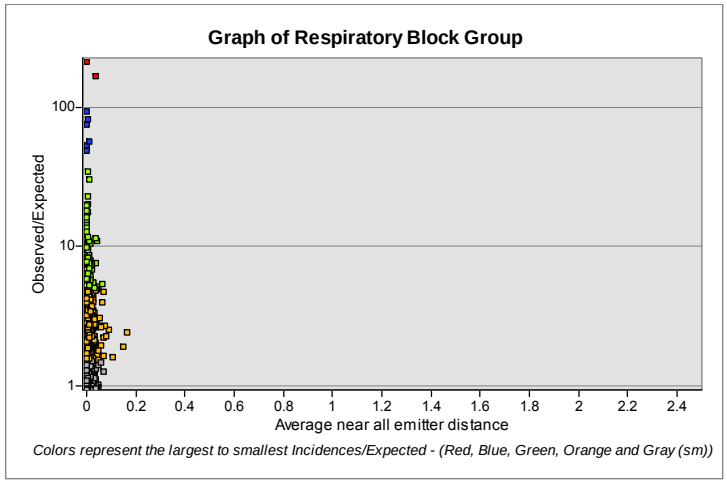


The following are block group respiratory scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have three scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged and 3) with the horizontal axis set to a distance of 2.5 miles.

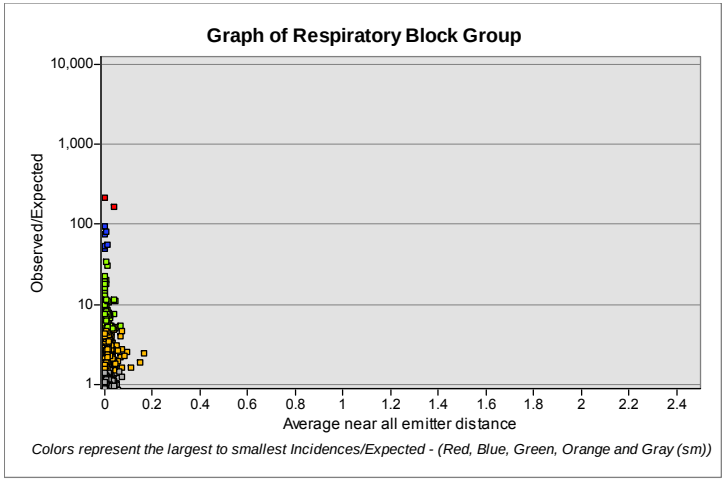
1)



2)

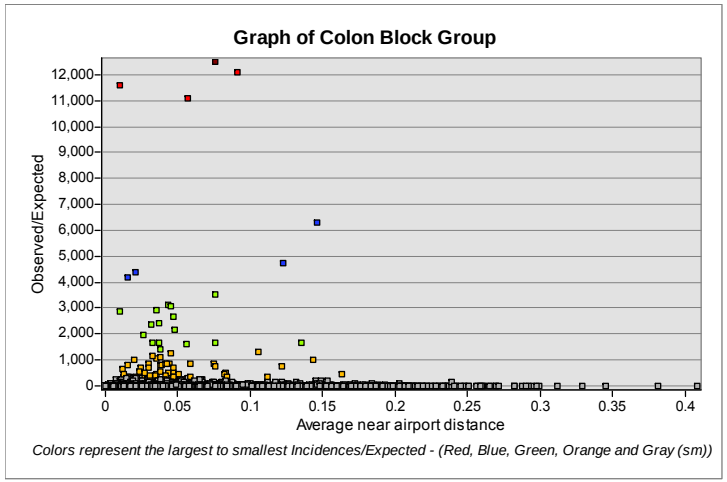


3)

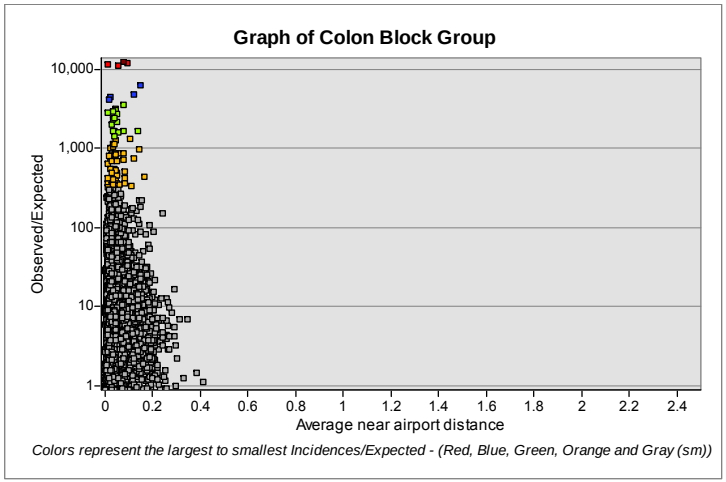


The following are block group colon scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have two scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged; also note that the horizontal axis is set to a distance of 2.5 miles.

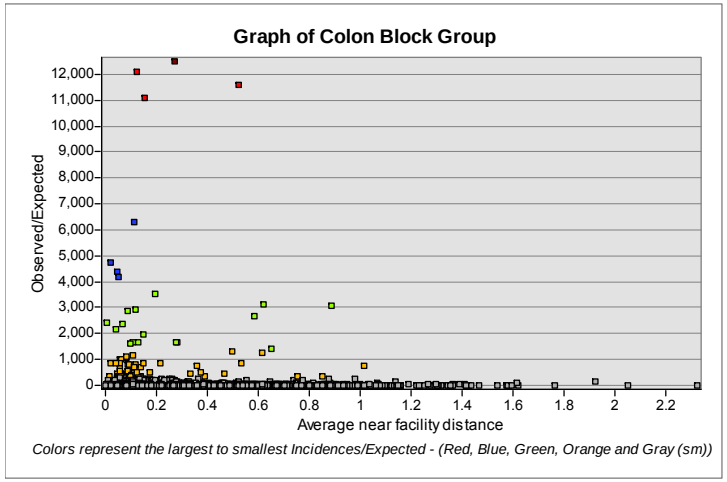
1)



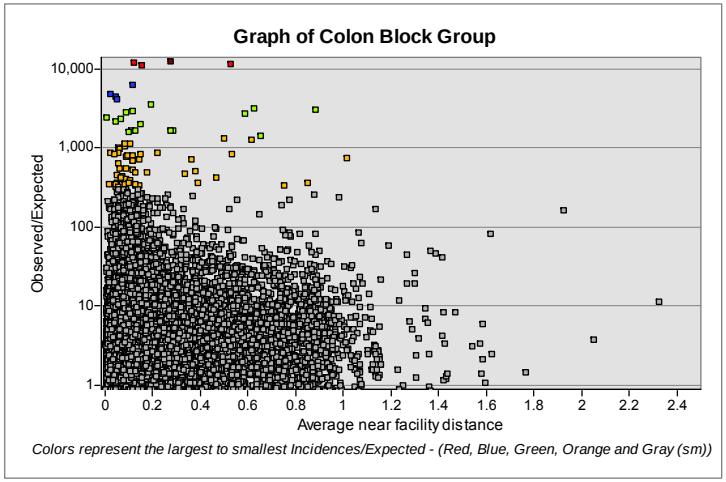
2)



1)

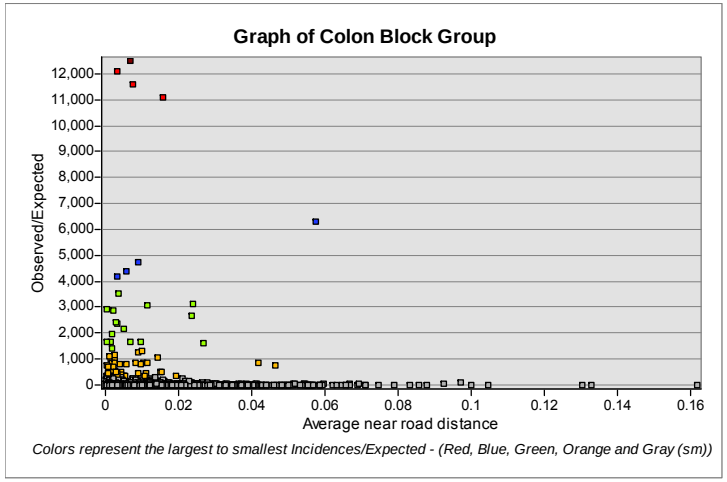


2)

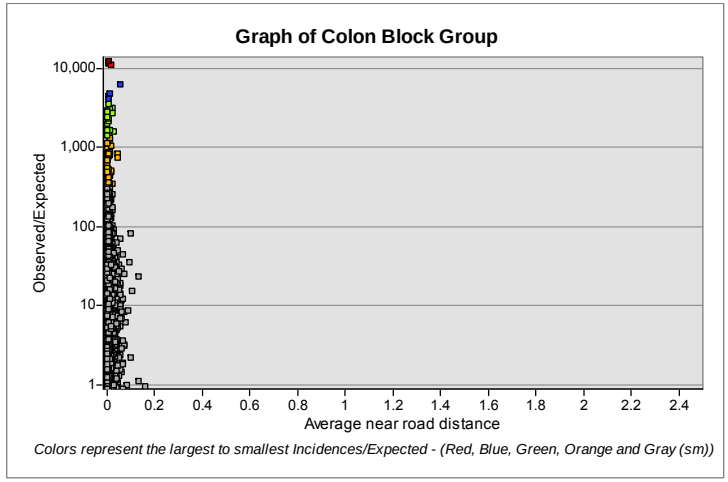


The following are block group colon scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have two scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged; also note that the horizontal axis is set to a distance of 2.5 miles.

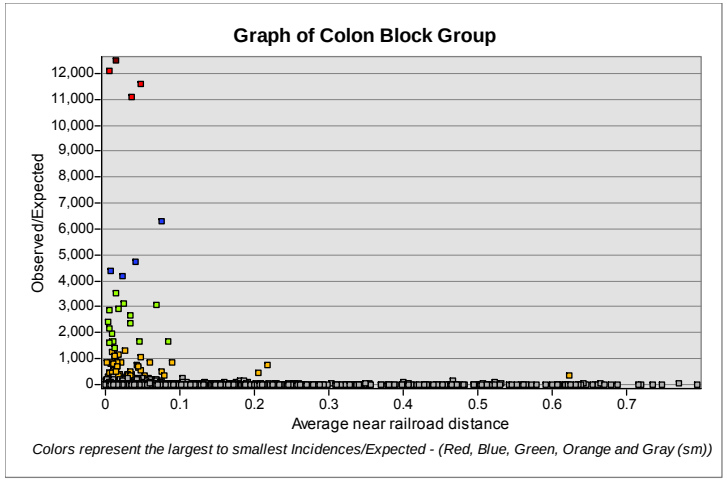
1)



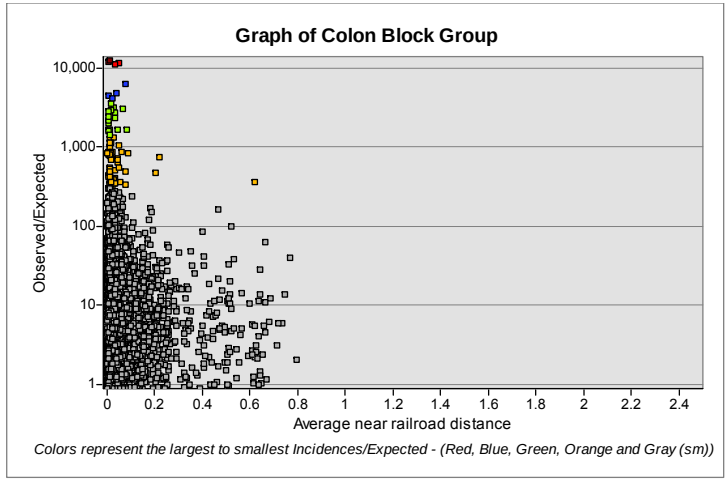
2)



1)

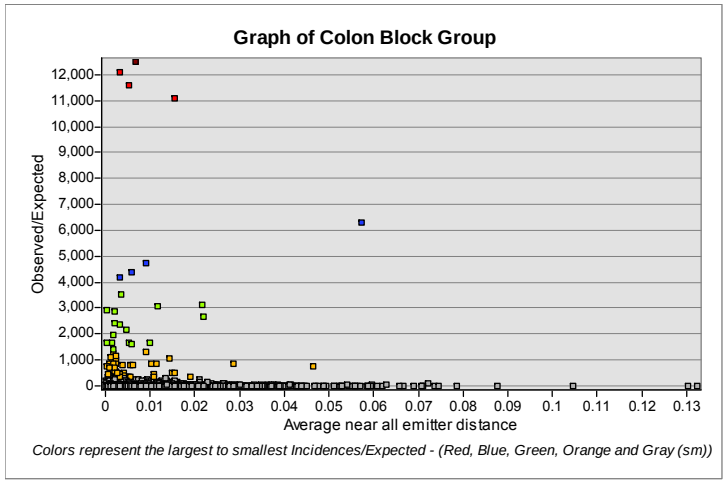


2)

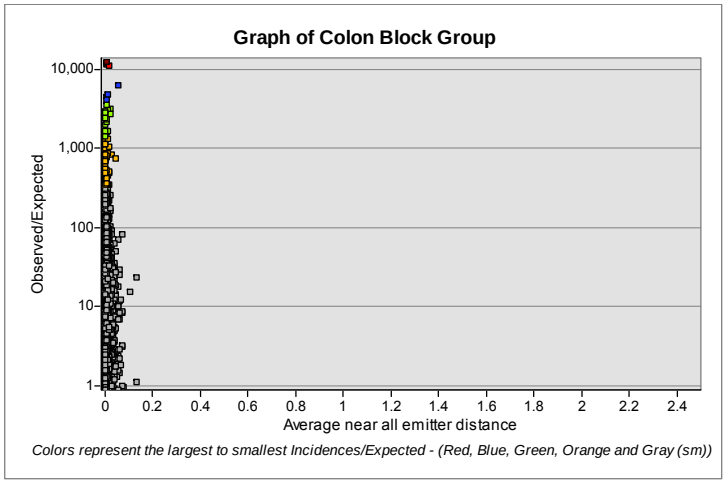


The following are block group colon scatter plots of observe to expected ratios verses distance to emitters (in miles). All rows have two scatter plots representing the data points 1) in their natural form, 2) with the vertical axis logged; also note that the horizontal axis is set to a distance of 2.5 miles.

1)

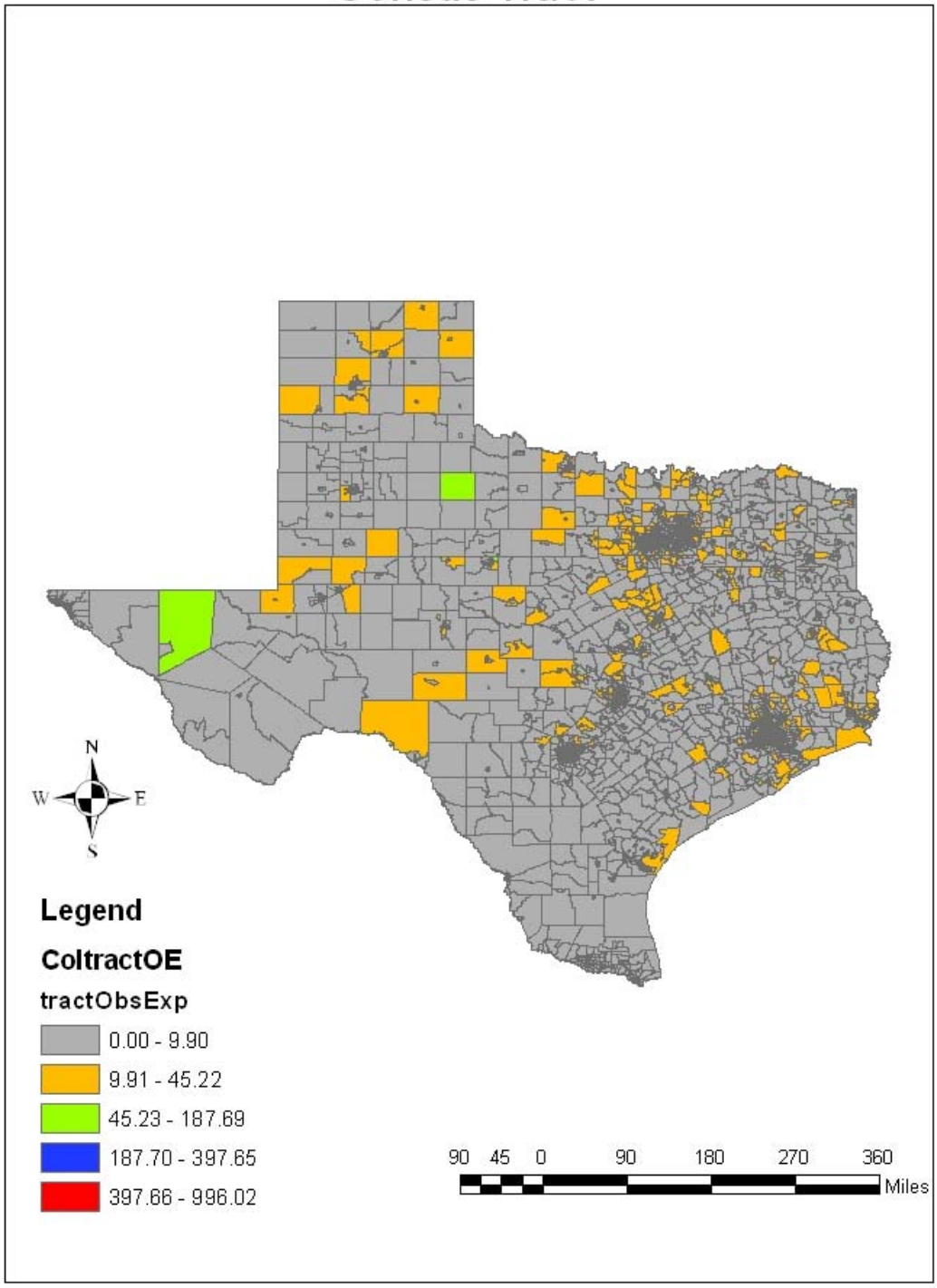


2)

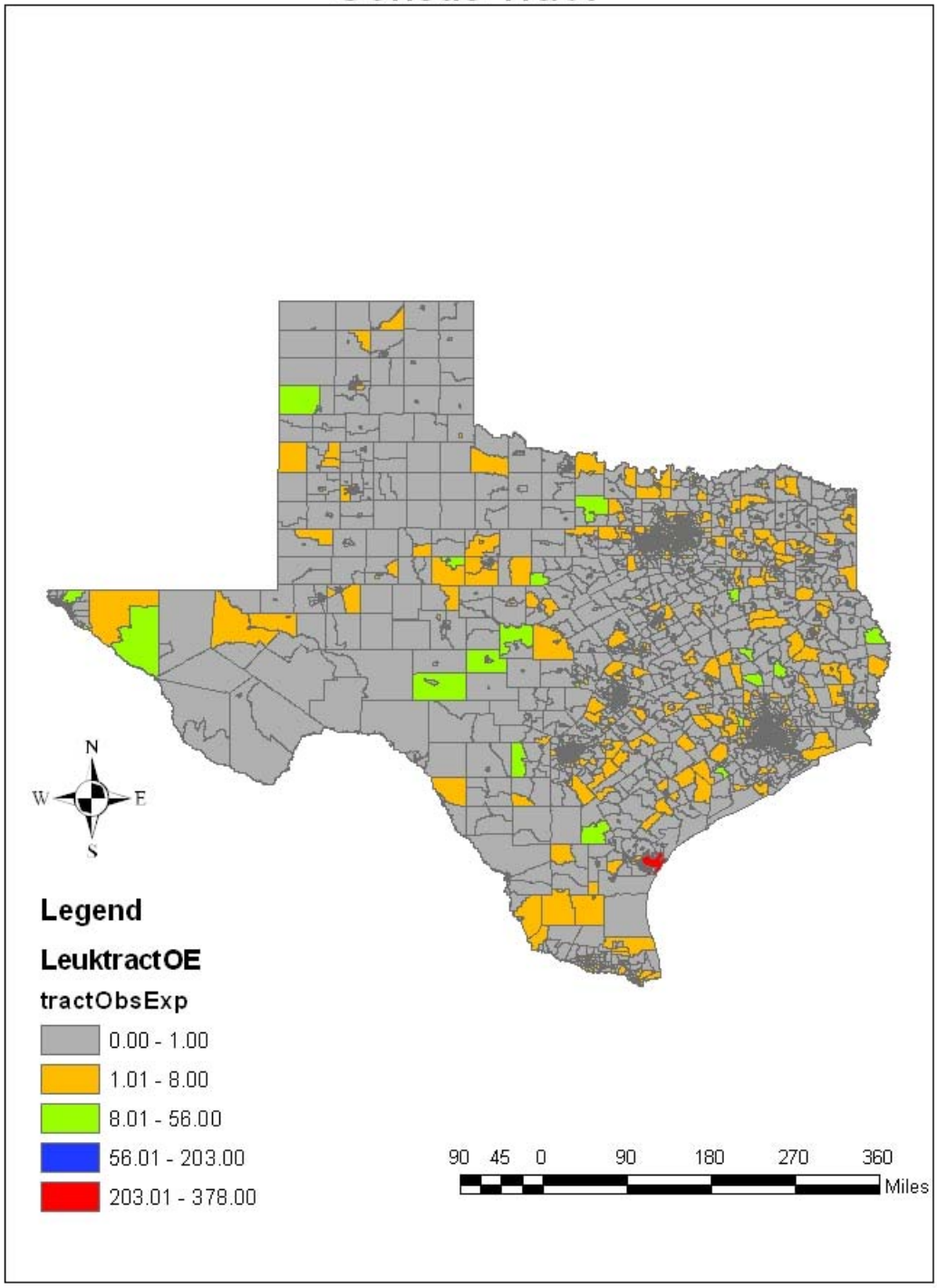


The following are census tract scatter plots of observe to expected ratios verses distance to emitters (in miles). The vertical axis has been logged in order to better visualize the data points. Additionally, maps that correspond with these scatter plots have been provided.

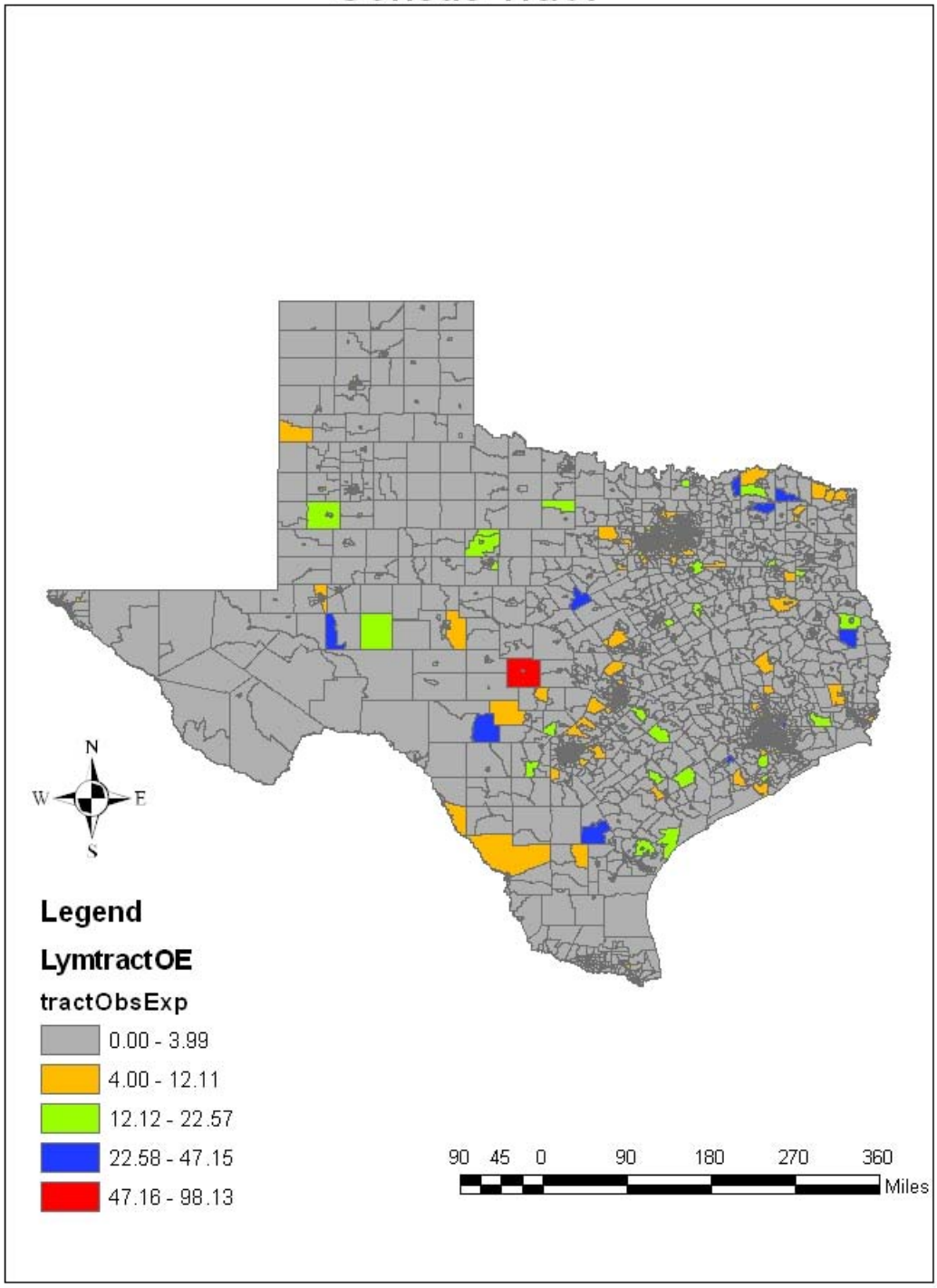
Observed to Expected Ratios of Colon Cases by Census Tract



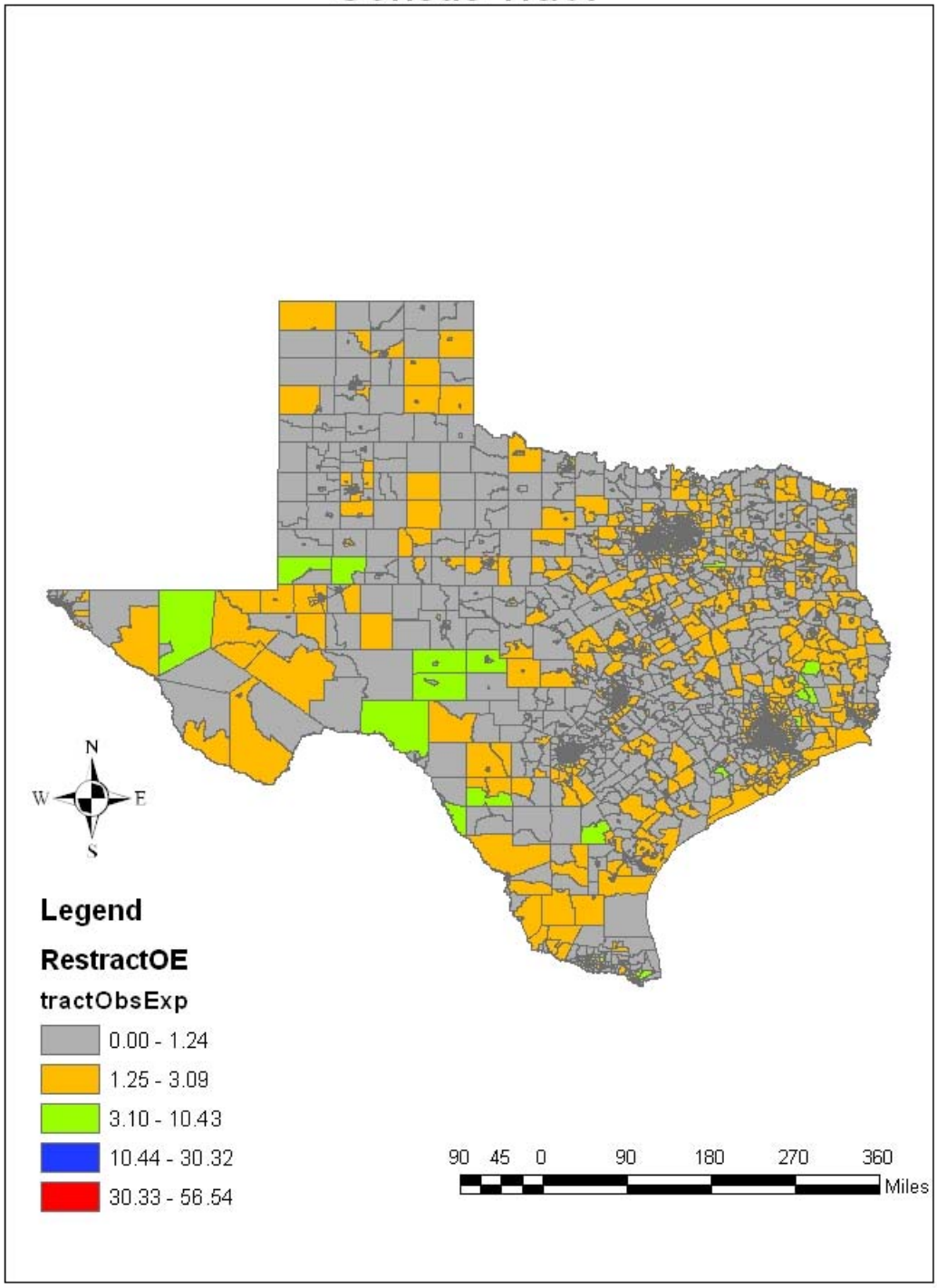
Observed to Expected Ratios of Leukemia Cases by Census Tract



Observed to Expected Ratios of Lymphoma Cases by Census Tract

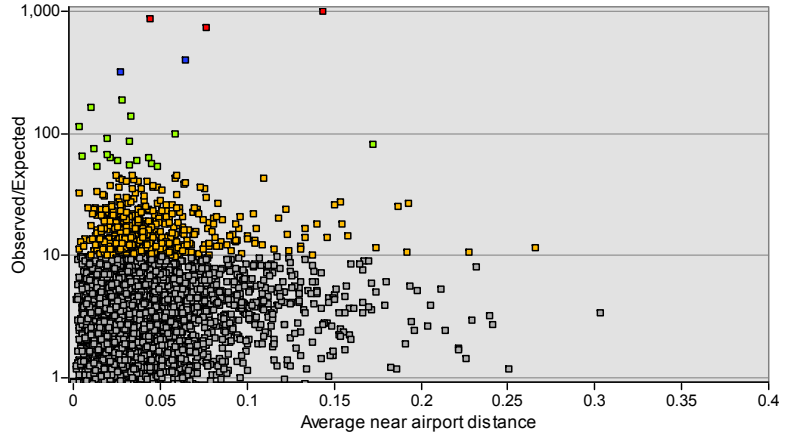


Observed to Expected Ratios of Respiratory Cases by Census Tract



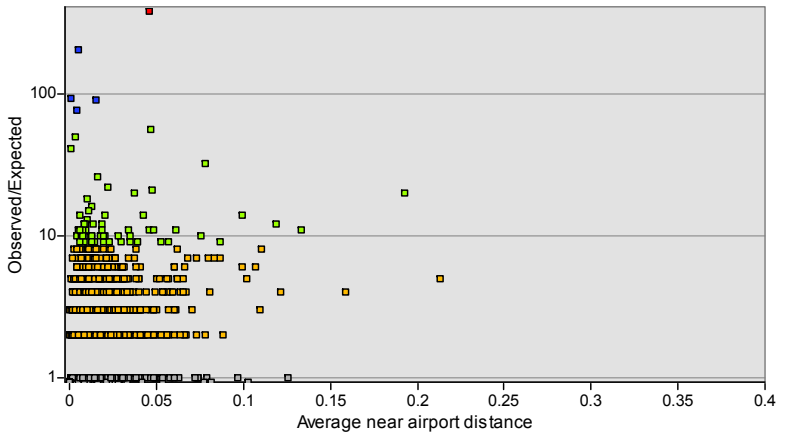
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Airports

Colon Cancer



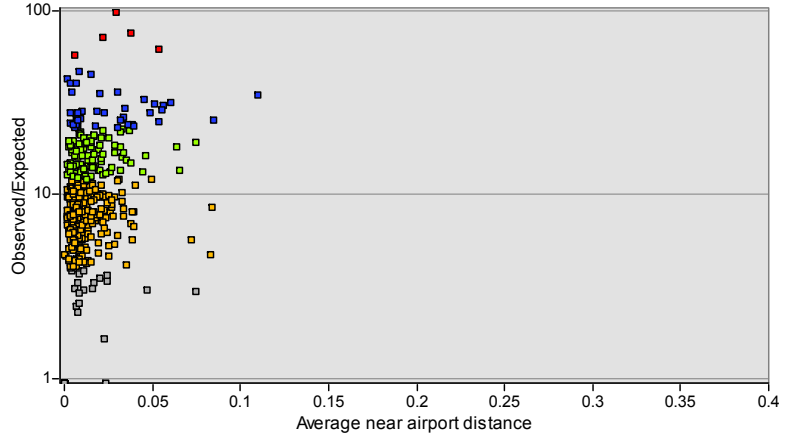
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



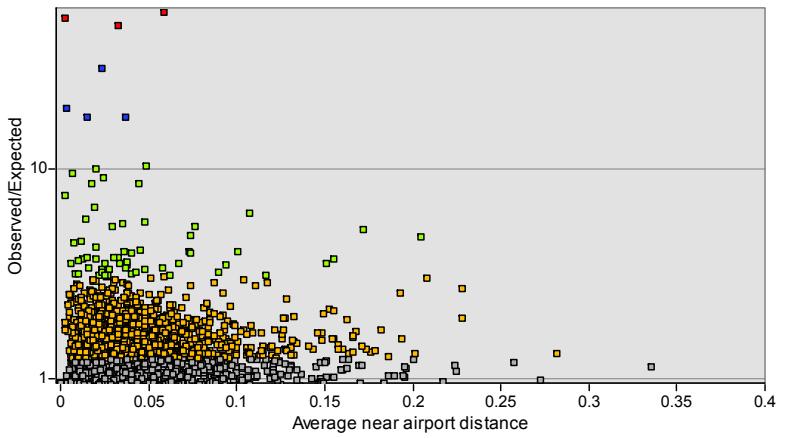
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

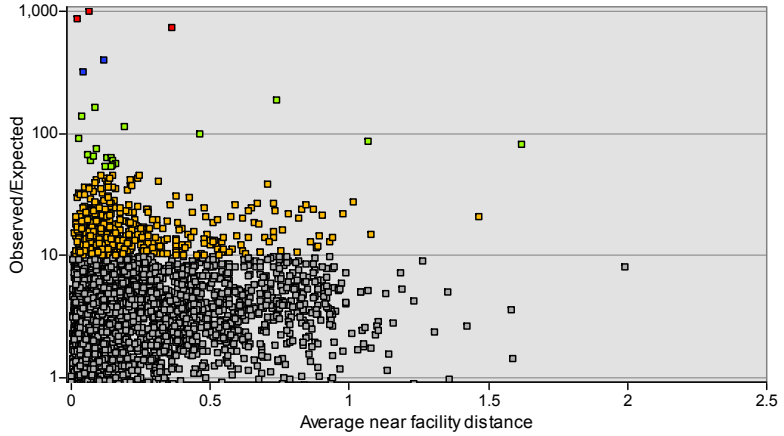
Respiratory Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

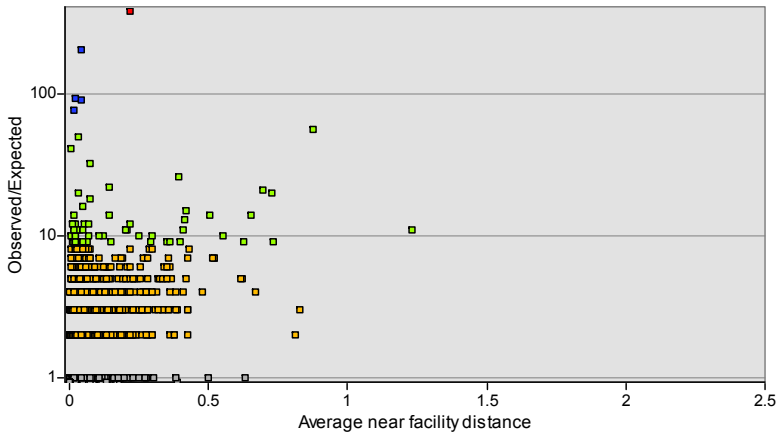
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Facilities

Colon Cancer



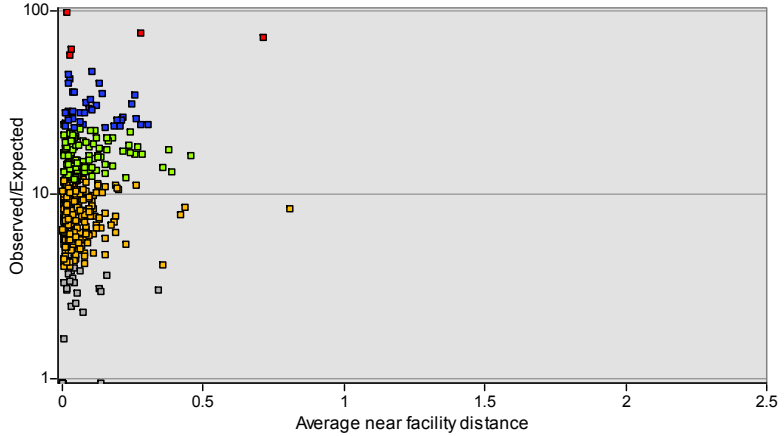
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



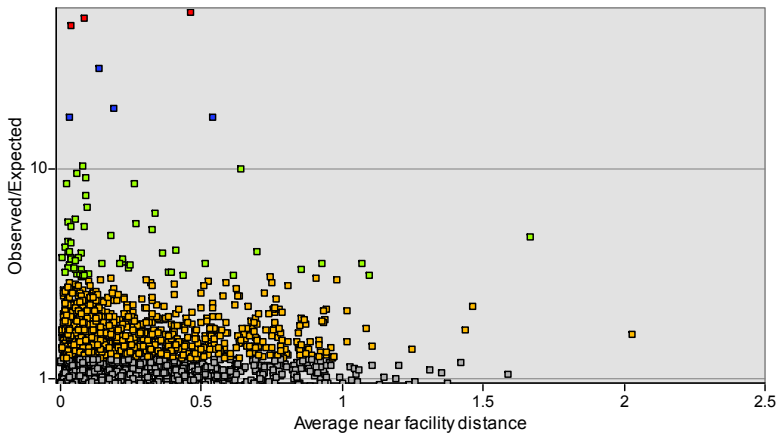
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

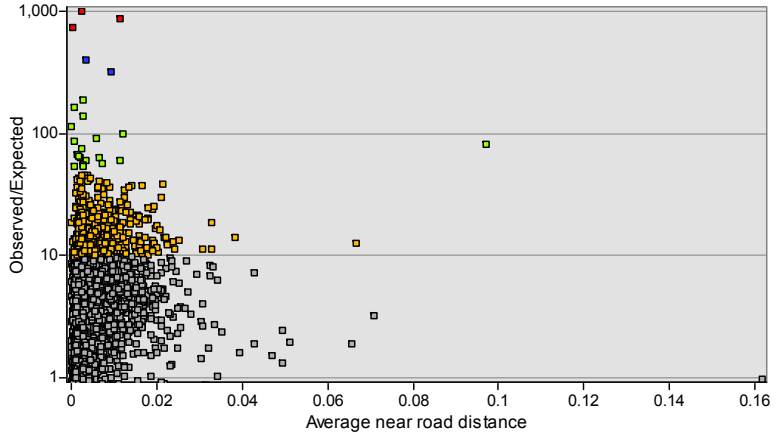
Respiratory Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

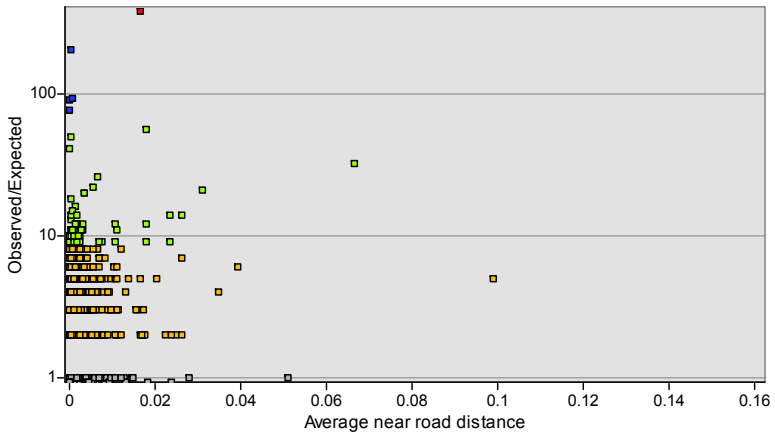
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Roads

Colon Cancer



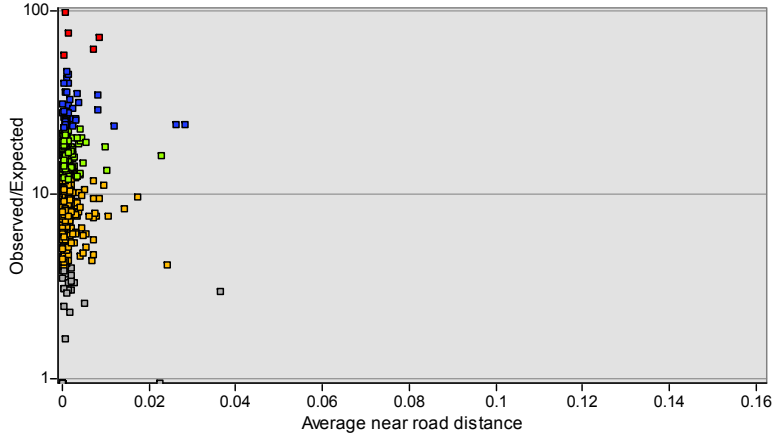
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



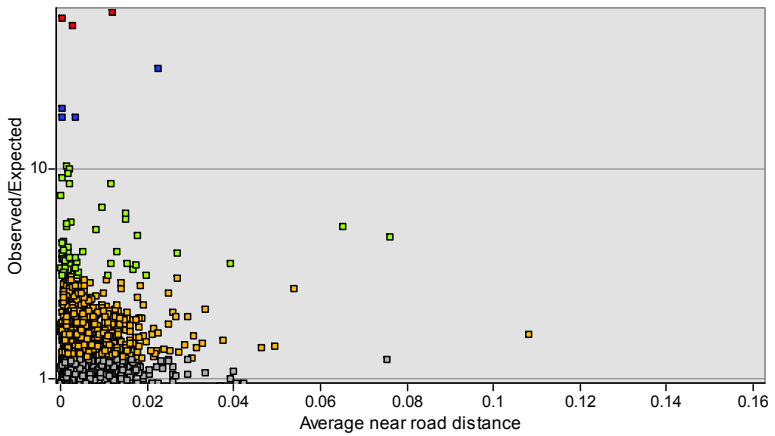
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

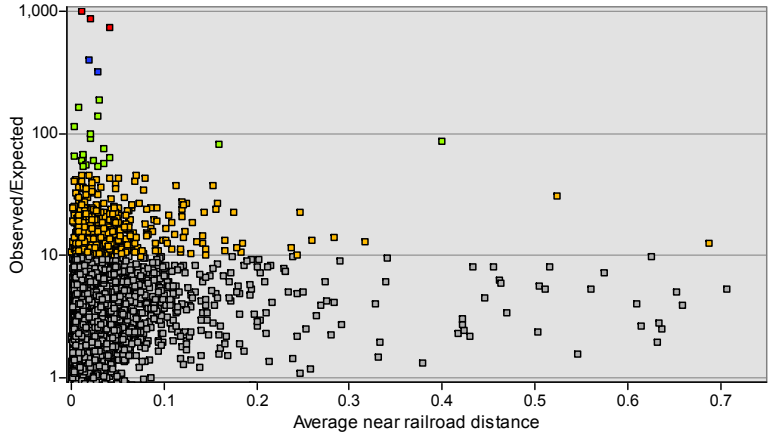
Respiratory Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

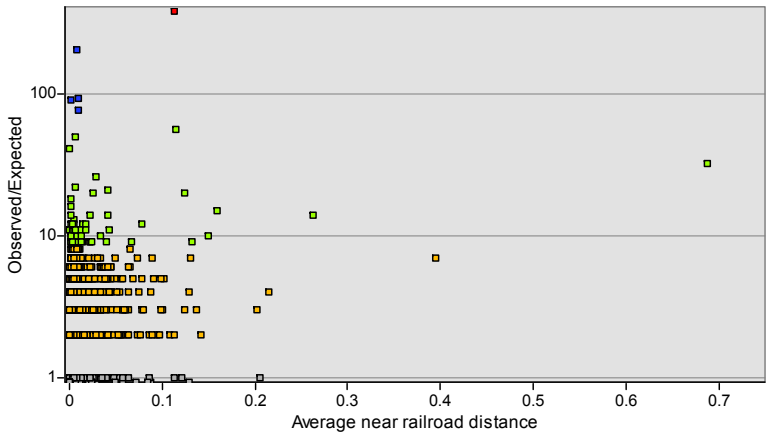
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Railroads

Colon Cancer



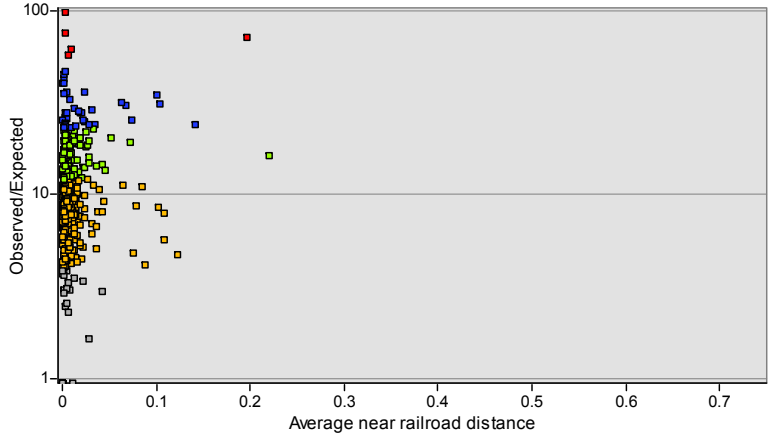
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



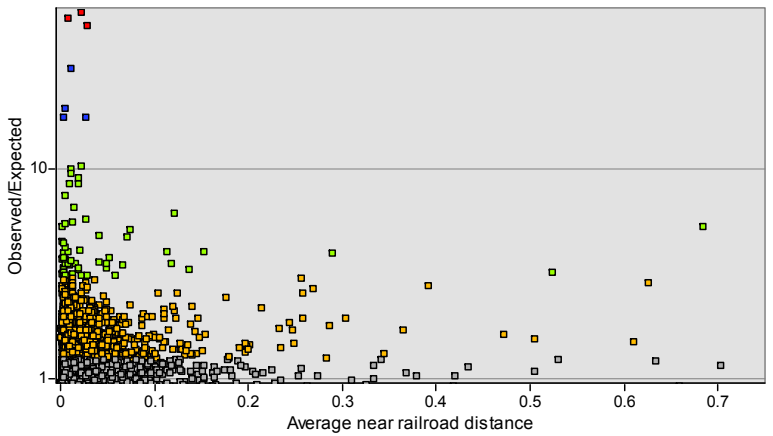
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



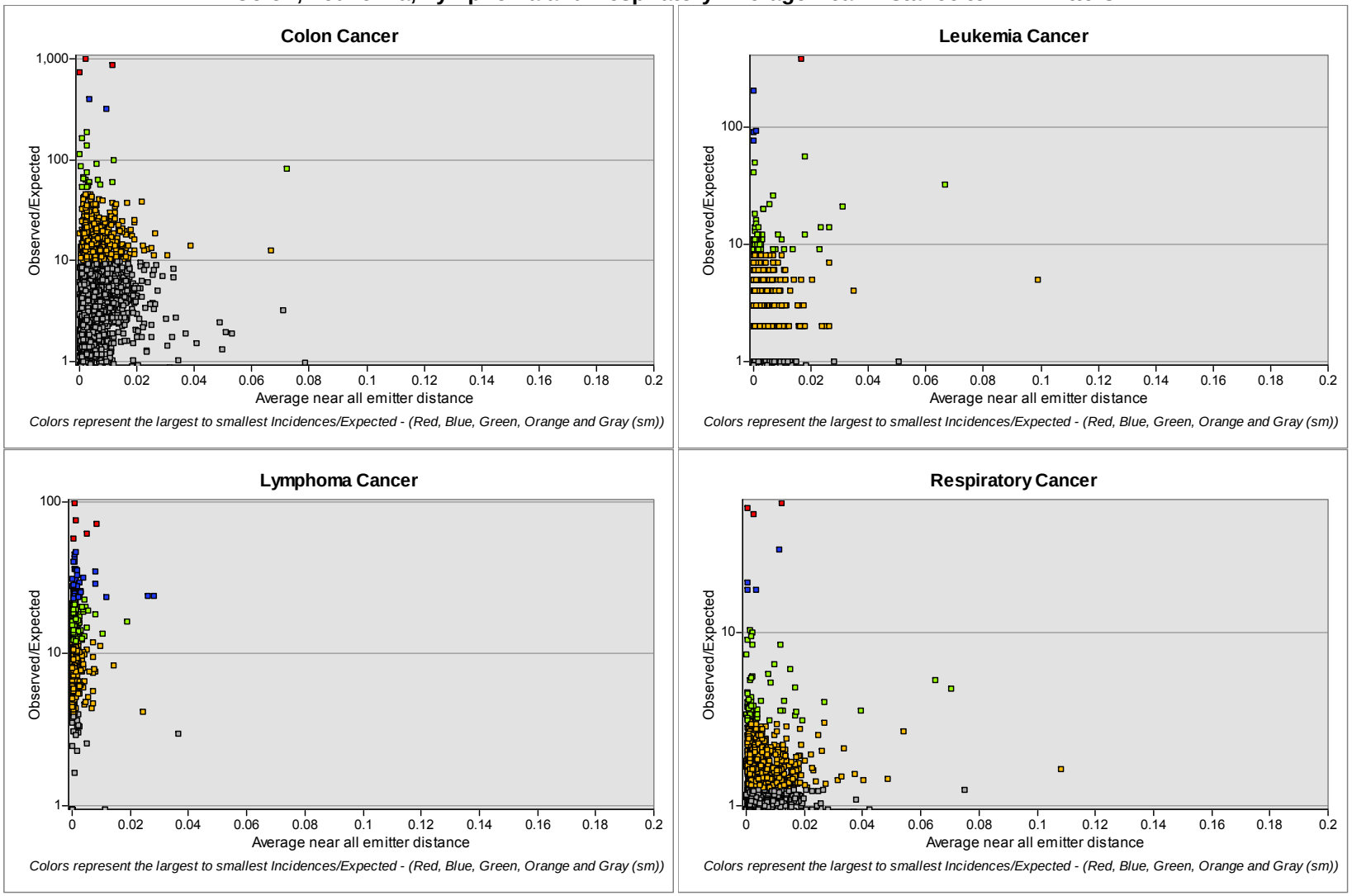
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Respiratory Cancer



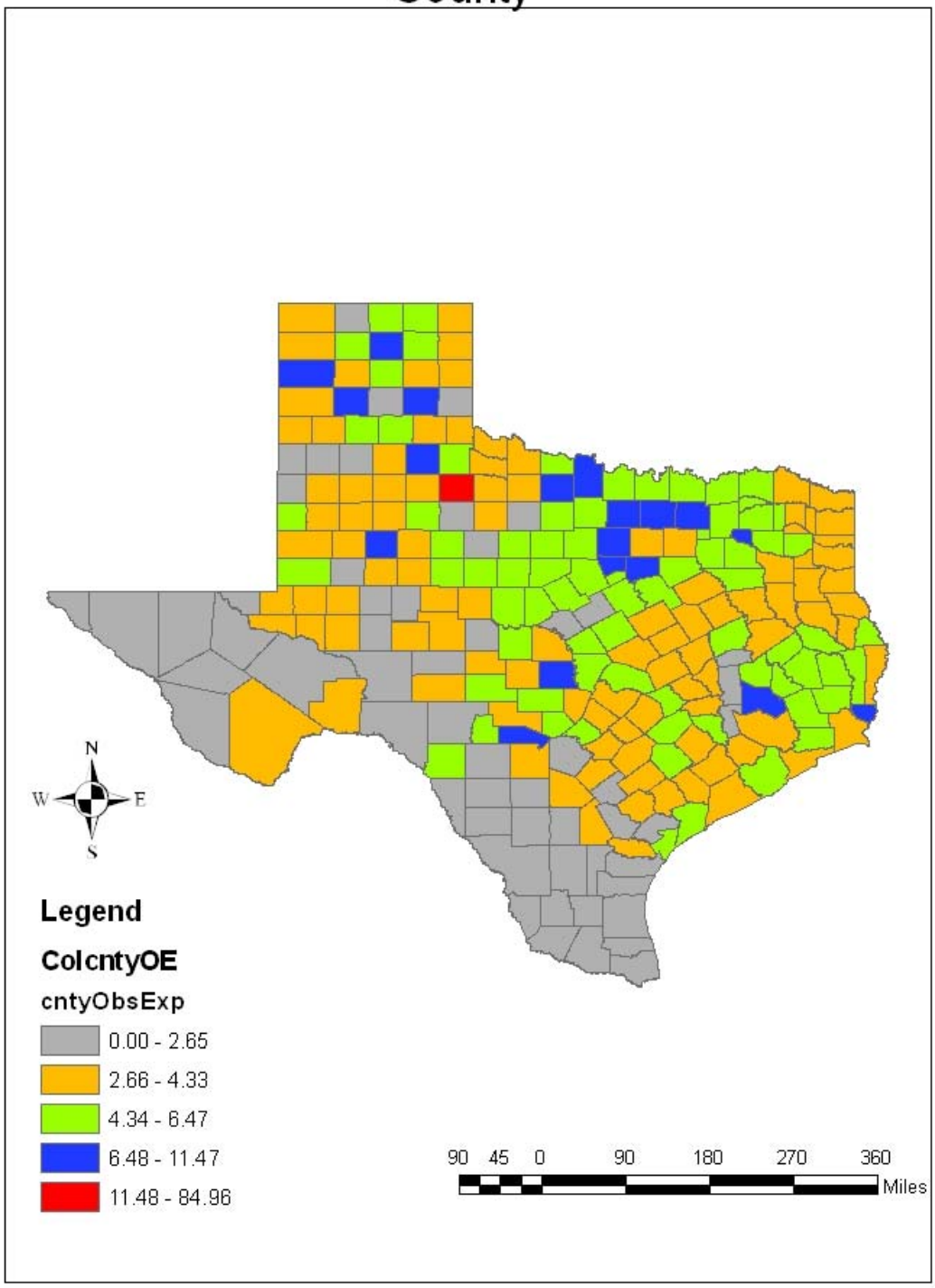
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to All Emitters

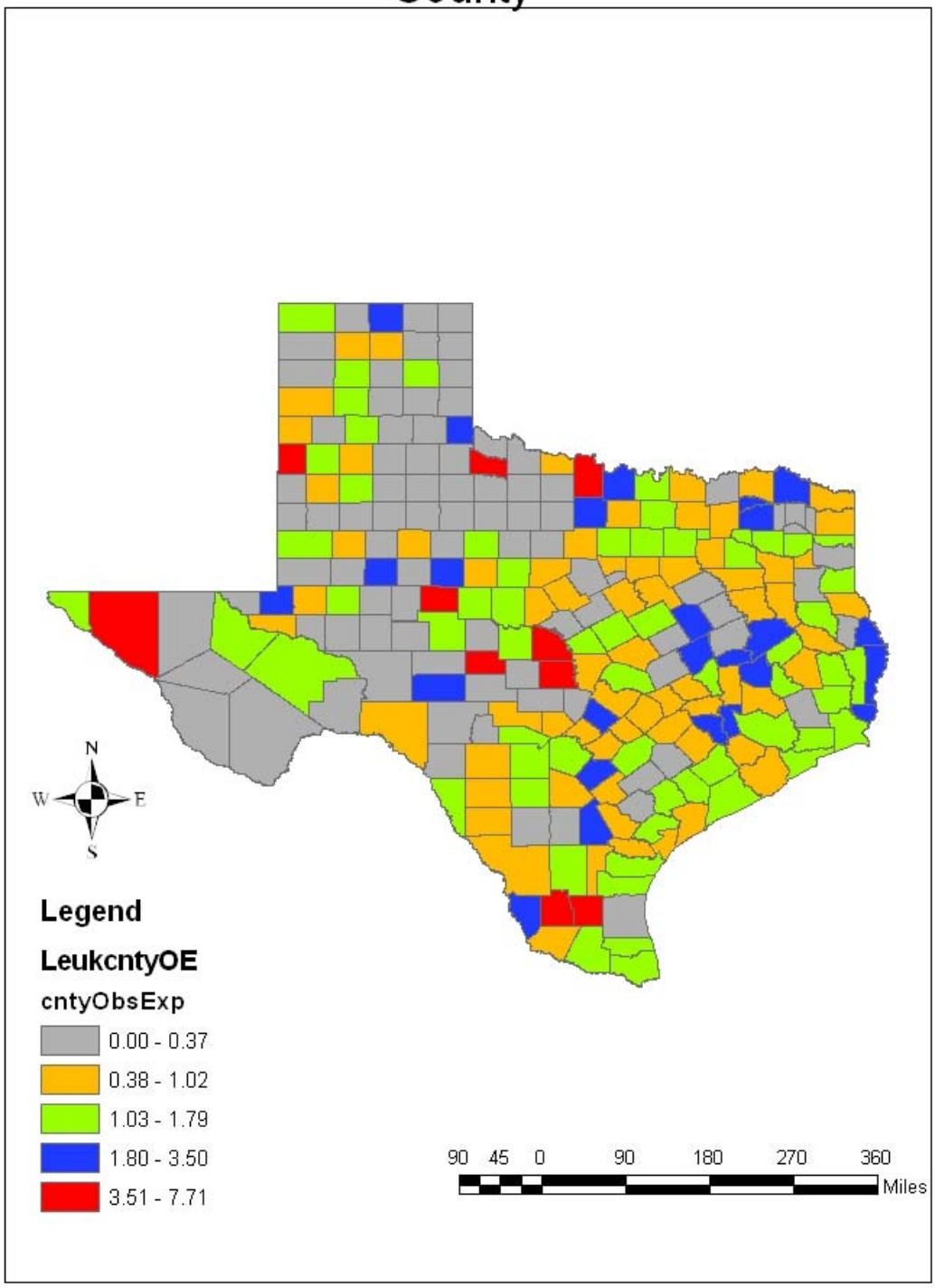


The following are county scatter plots of observe to expected ratios verses distance to emitters (in miles). Additionally, maps that correspond with these scatter plots have been provided.

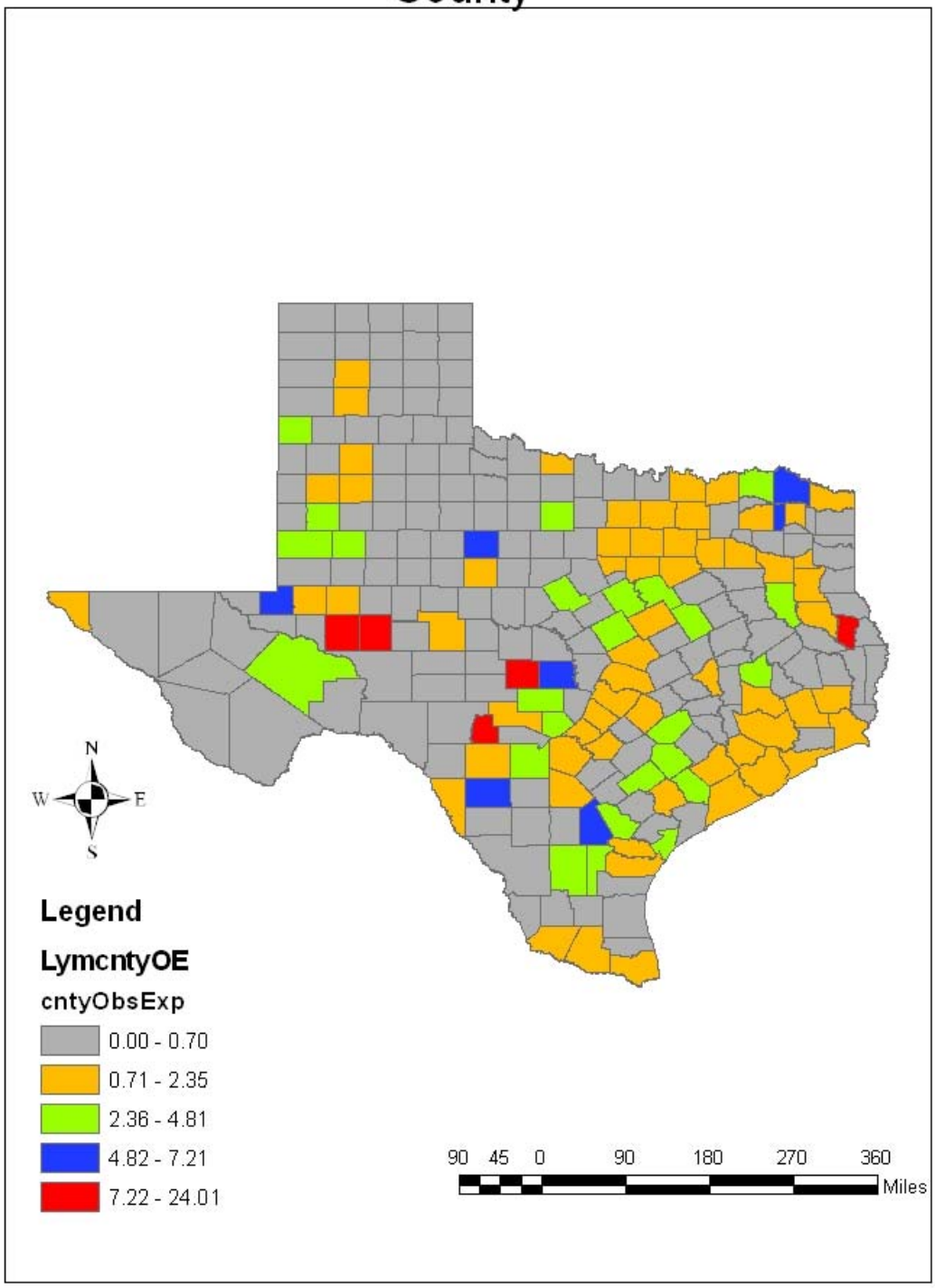
Observed to Expected Ratios of Colon Cases by County



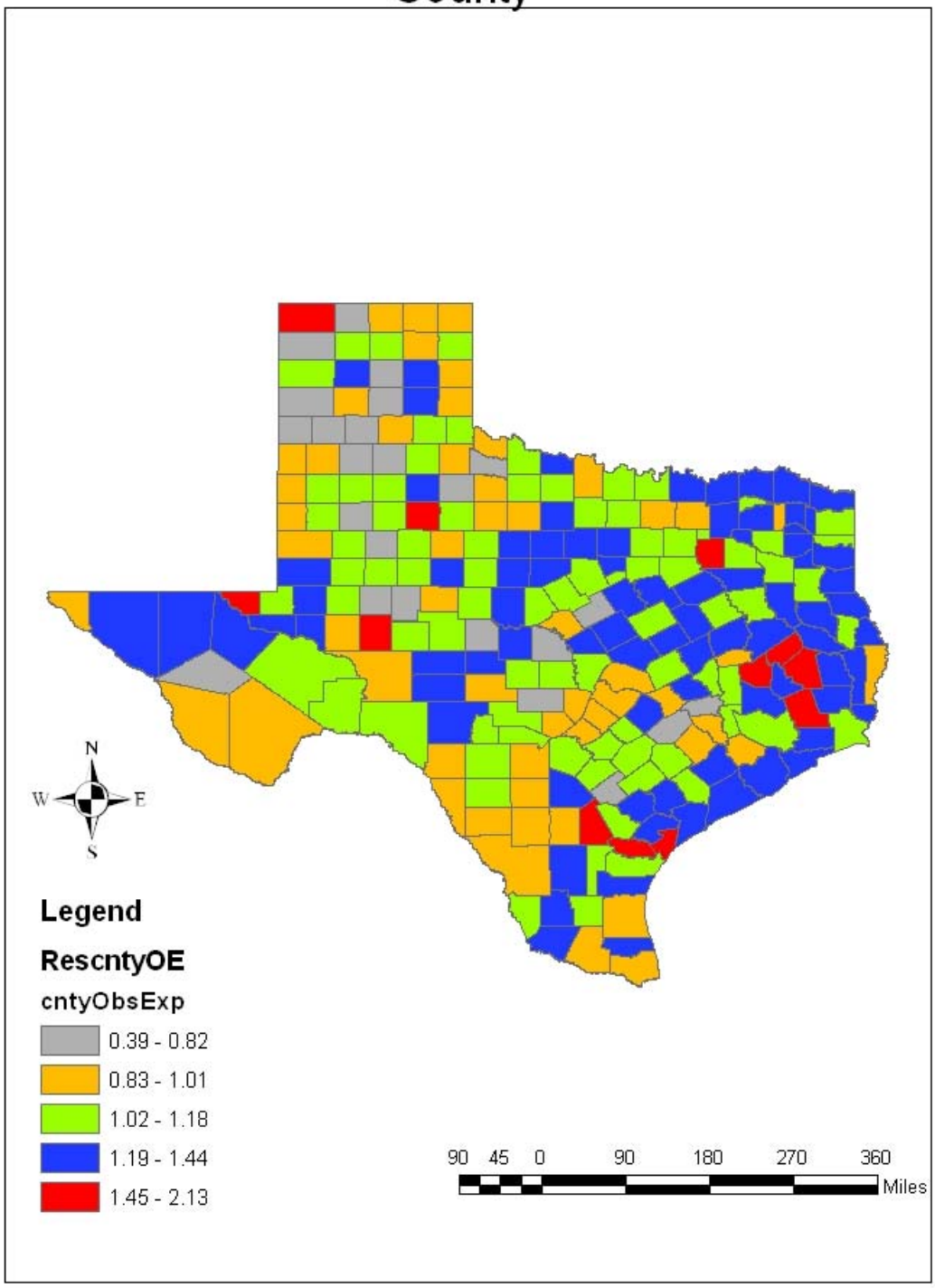
Observed to Expected Ratios of Leukemia Cases by County



Observed to Expected Ratios of Lymphoma Cases by County

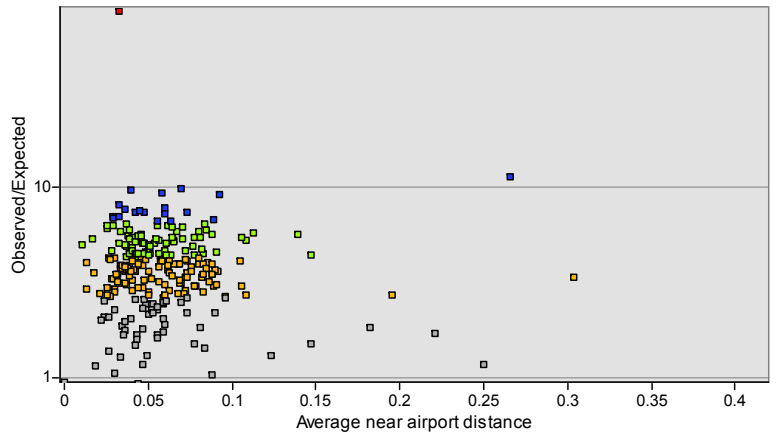


Observed to Expected Ratios of Respiratory Cases by County



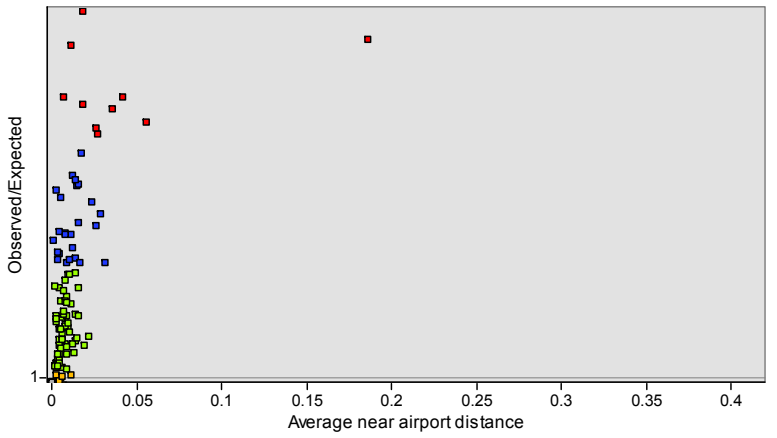
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Airports

Colon Cancer



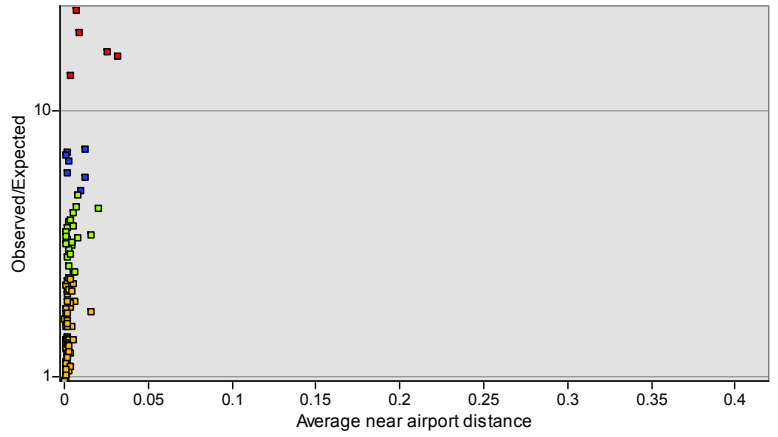
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



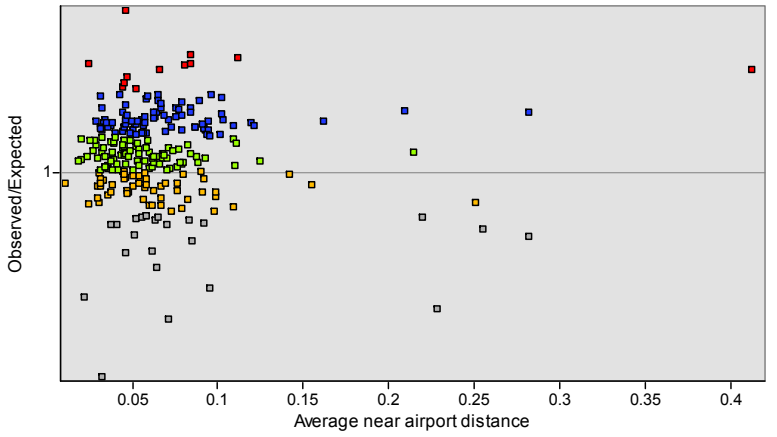
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



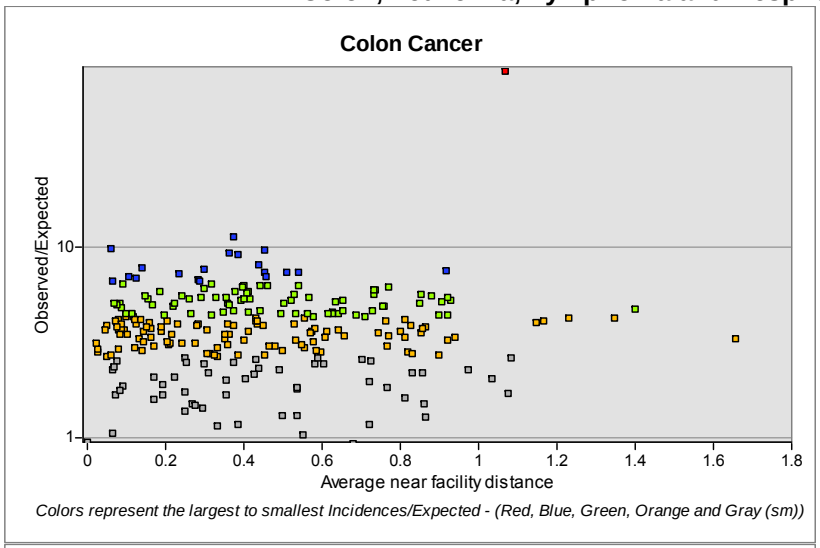
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Respiratory Cancer

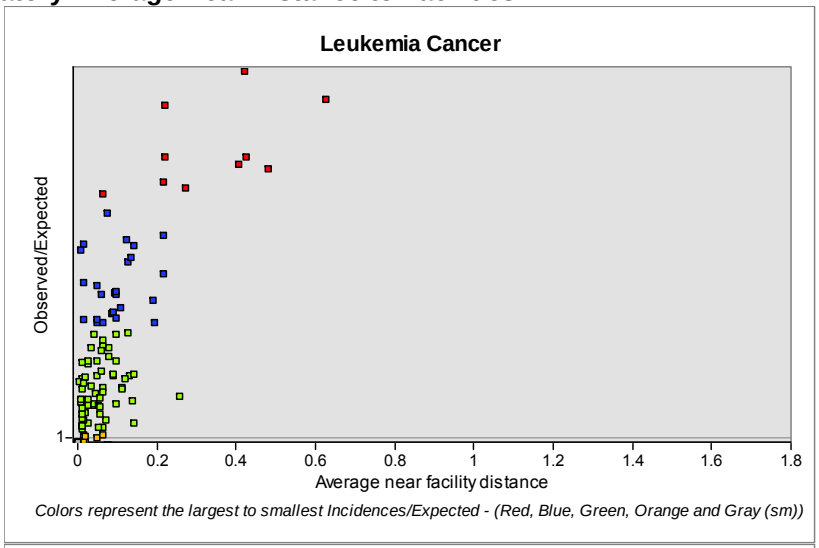


Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

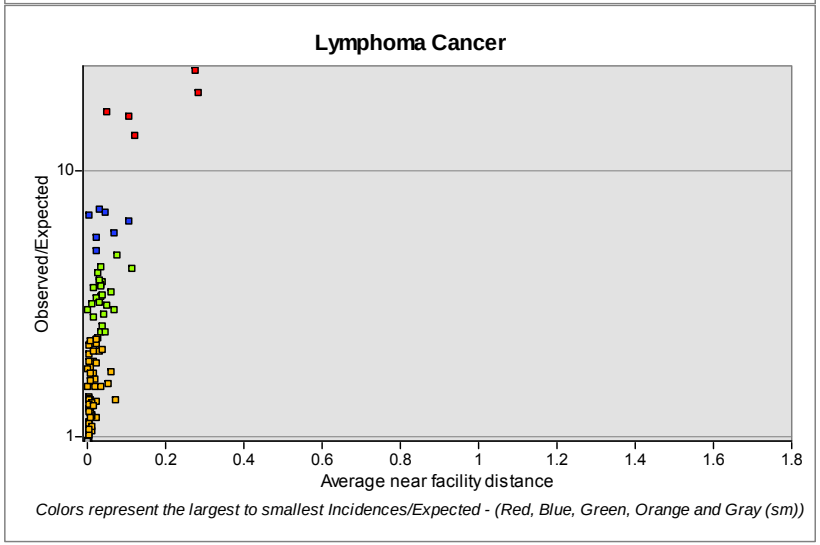
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Facilities



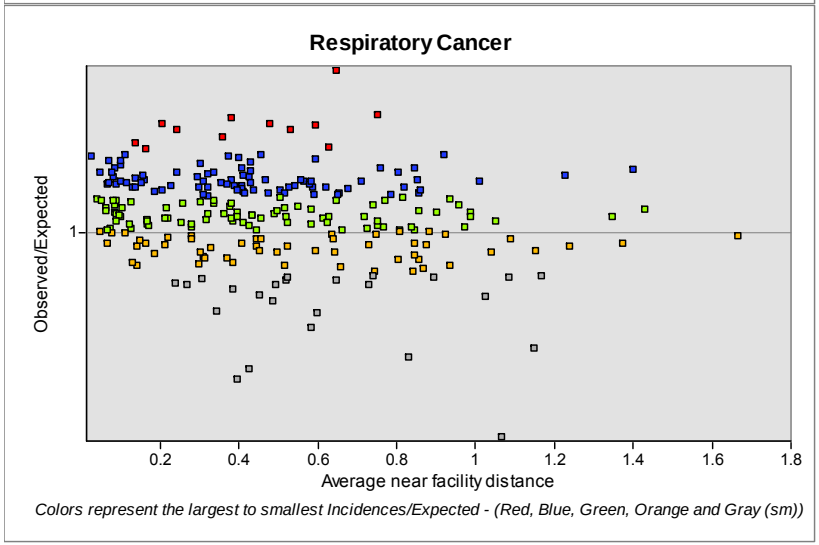
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

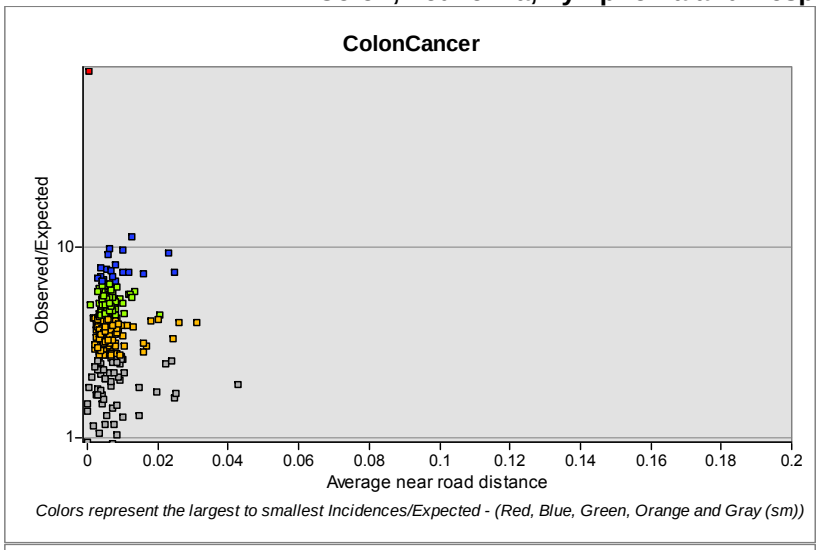


Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

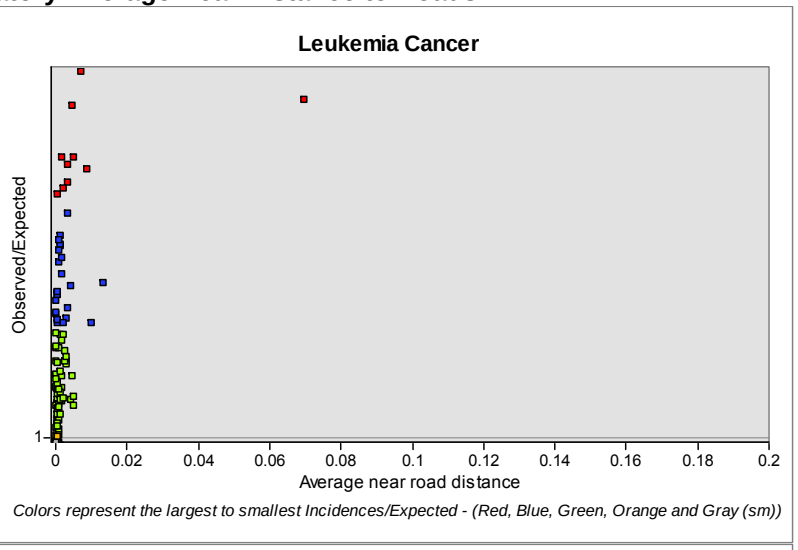


Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

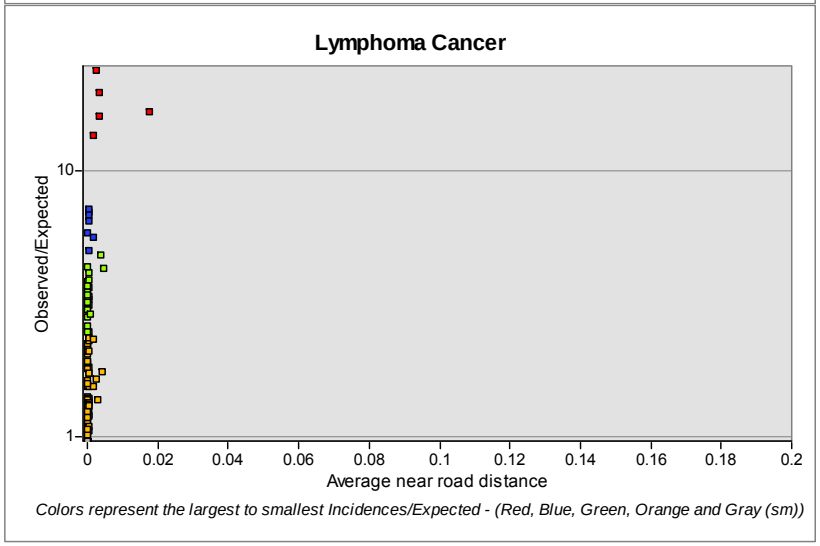
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Roads



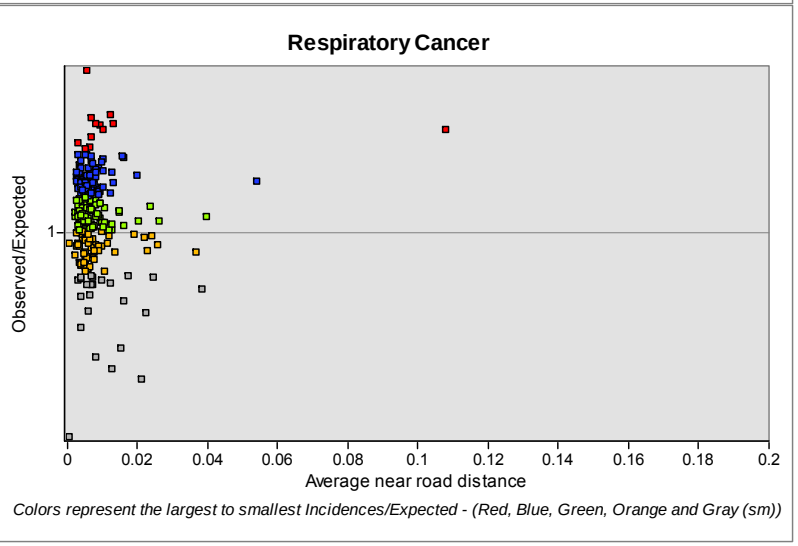
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

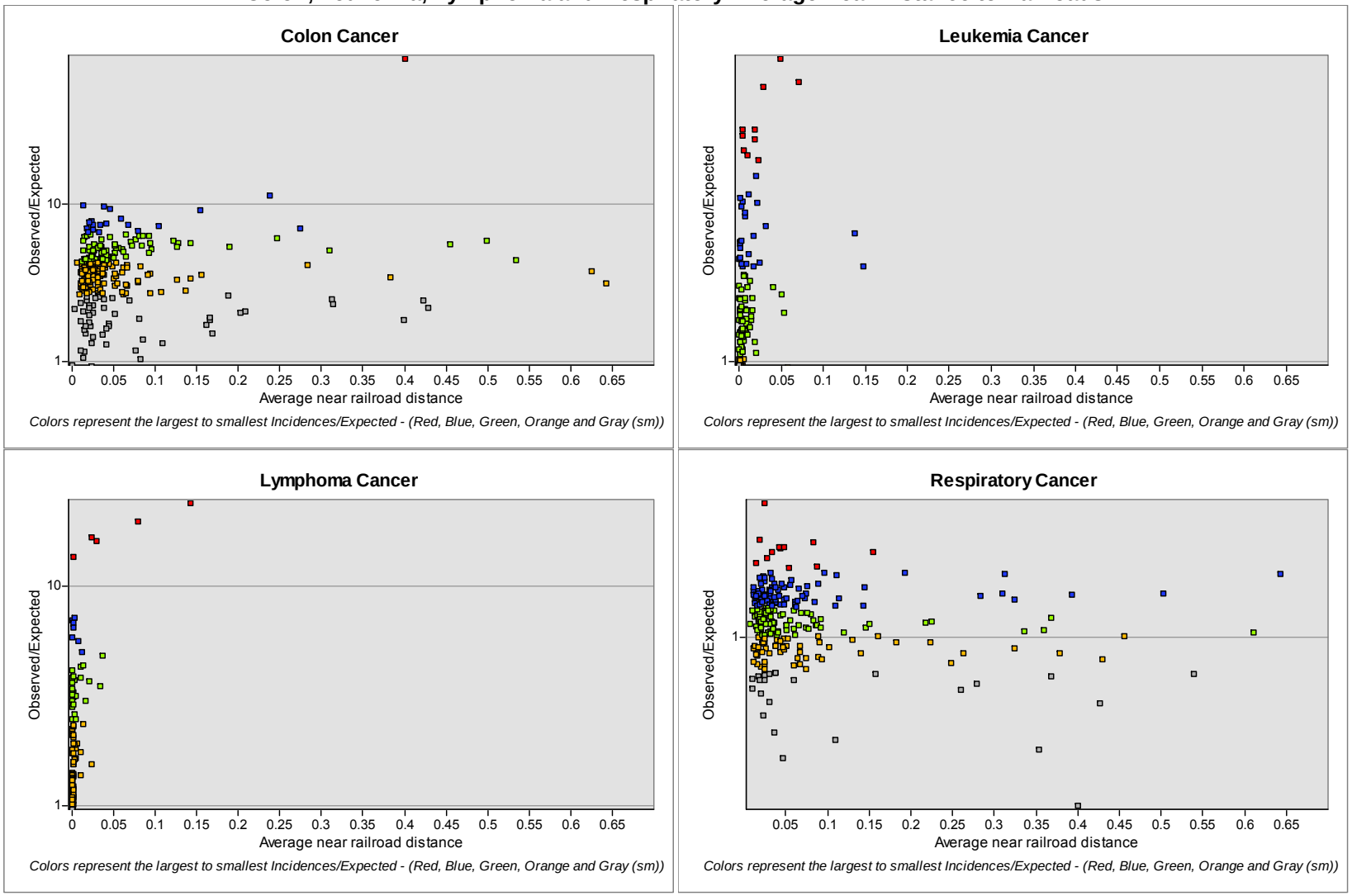


Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

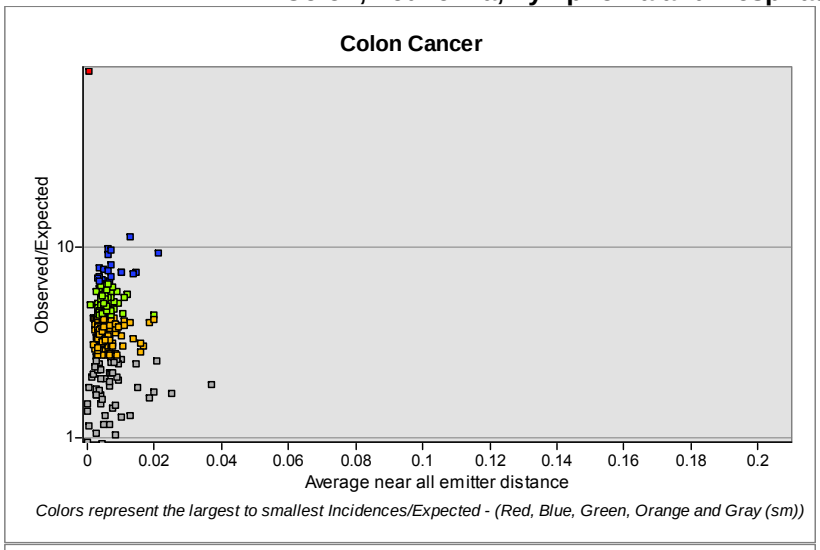


Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

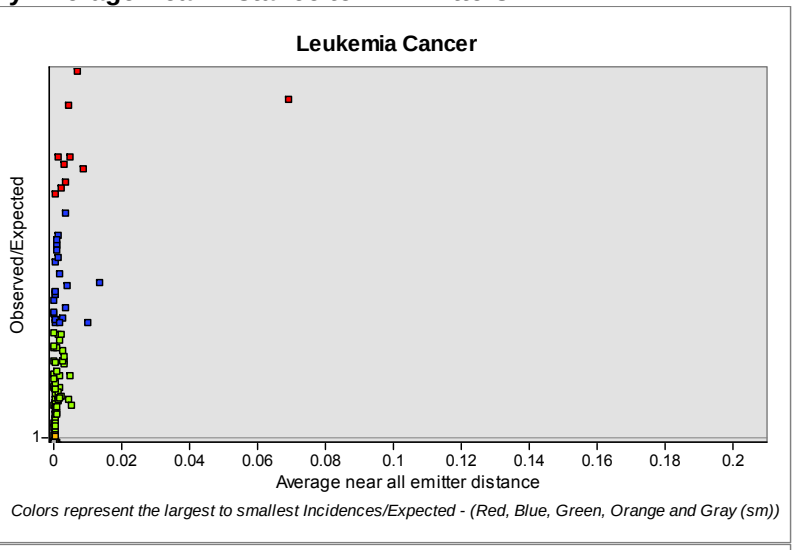
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Railroads



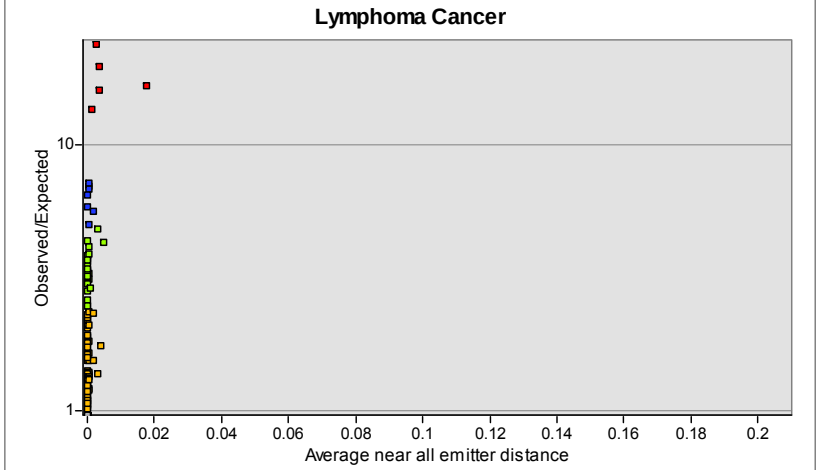
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to All Emitters



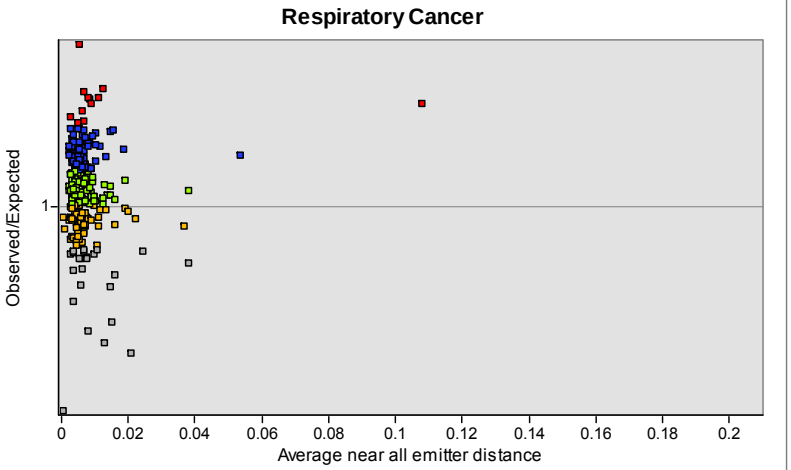
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))



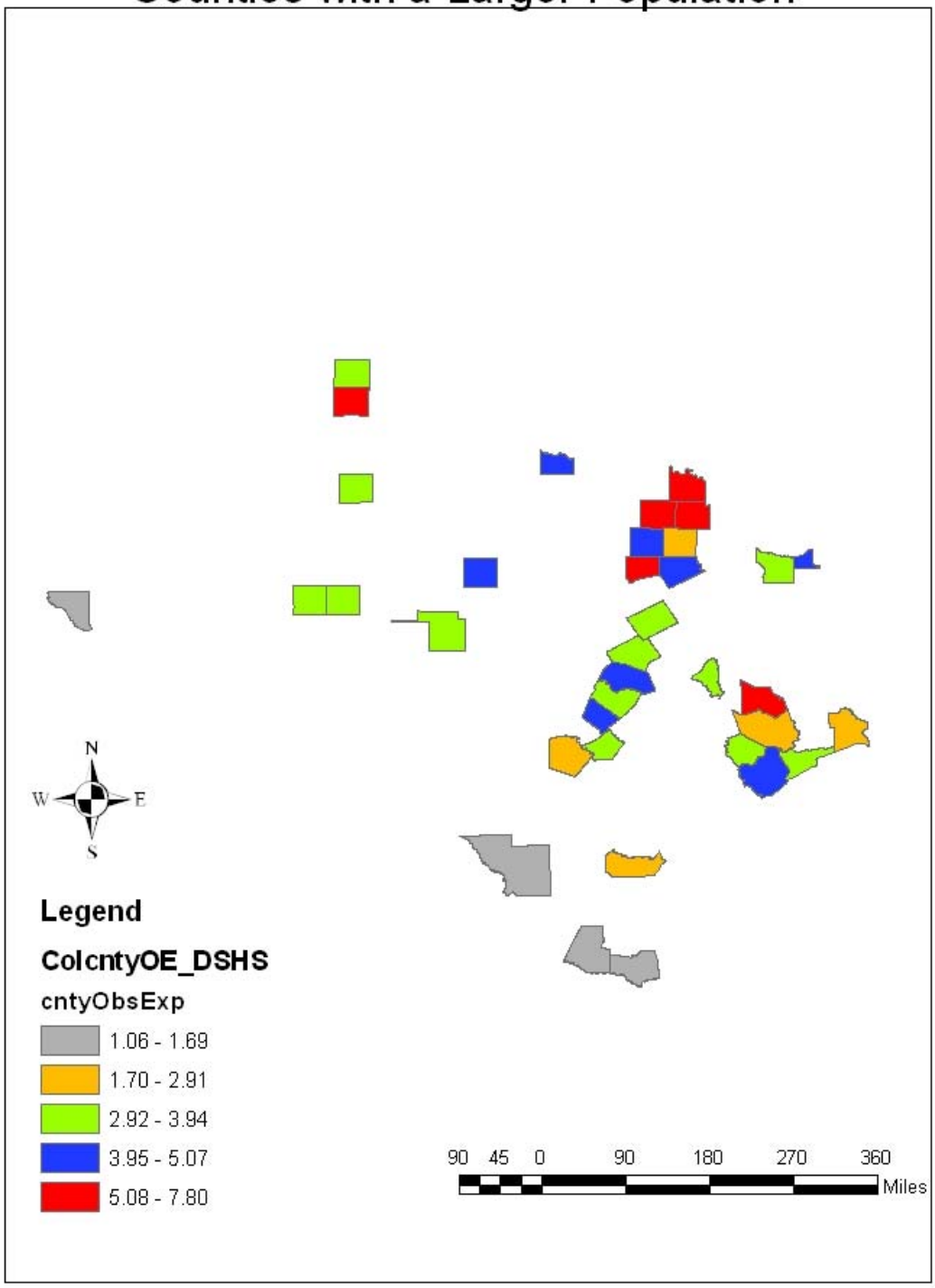
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

The following are scatter plots of observe to expected ratios verses distance to emitters (in miles) for counties with a relatively large population. Additionally, maps that correspond with these scatter plots have been provided.

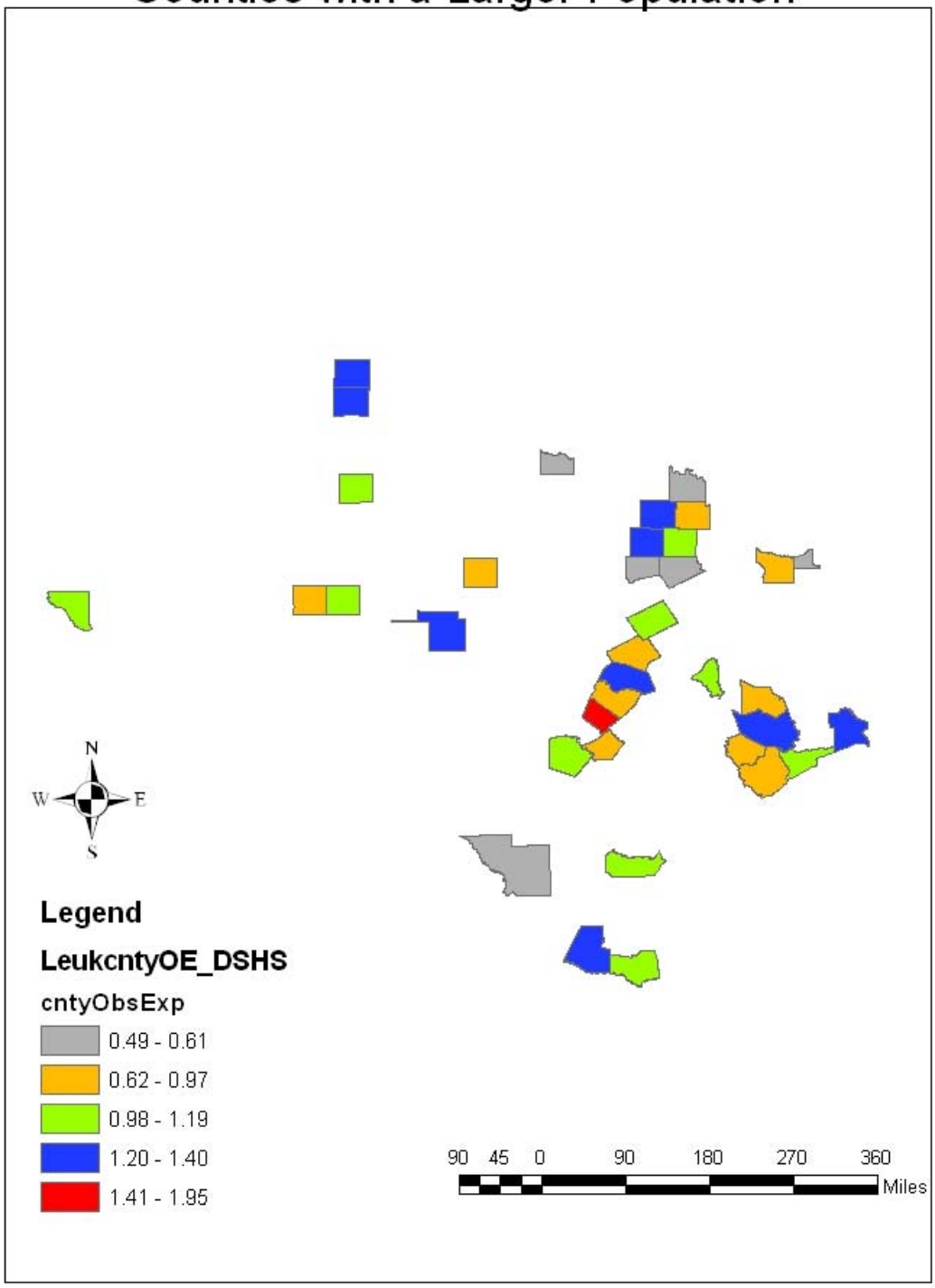
Directly below is a list of the 32 counties (with relatively large population sizes) used in the following scatter plot analysis.

1. Potter
2. Randall
3. Wichita
4. Grayson
5. Lubbock
6. Denton
7. Collin
8. Tarrent
9. Dallas
10. Smith
11. Brazos
12. Williamson
13. Montgomery
14. Travis
15. Hays
16. Jefferson
17. Harris
18. Guadalupe
19. Gregg
20. Johnson
21. Ellis
22. Taylor
23. Ector
24. Midland
25. El Paso
26. Mclennan
27. Tom Green
28. Bell
29. Fort Bend
30. Bexar
31. Brazoria
32. Galveston
33. Webb
34. Nueces
35. Hidalgo
36. Cameron

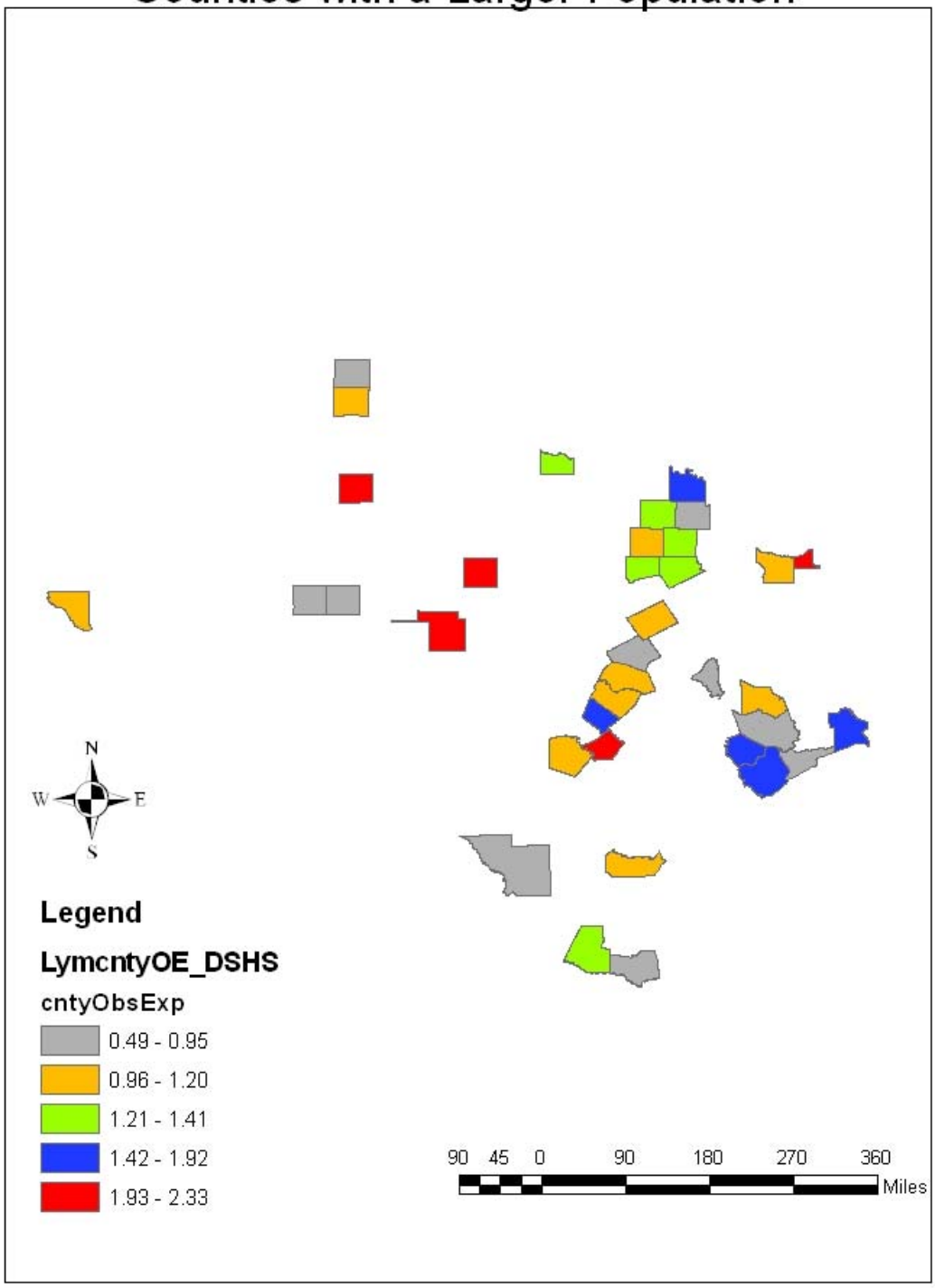
Observed to Expected Ratios of Colon Cases for Counties with a Larger Population



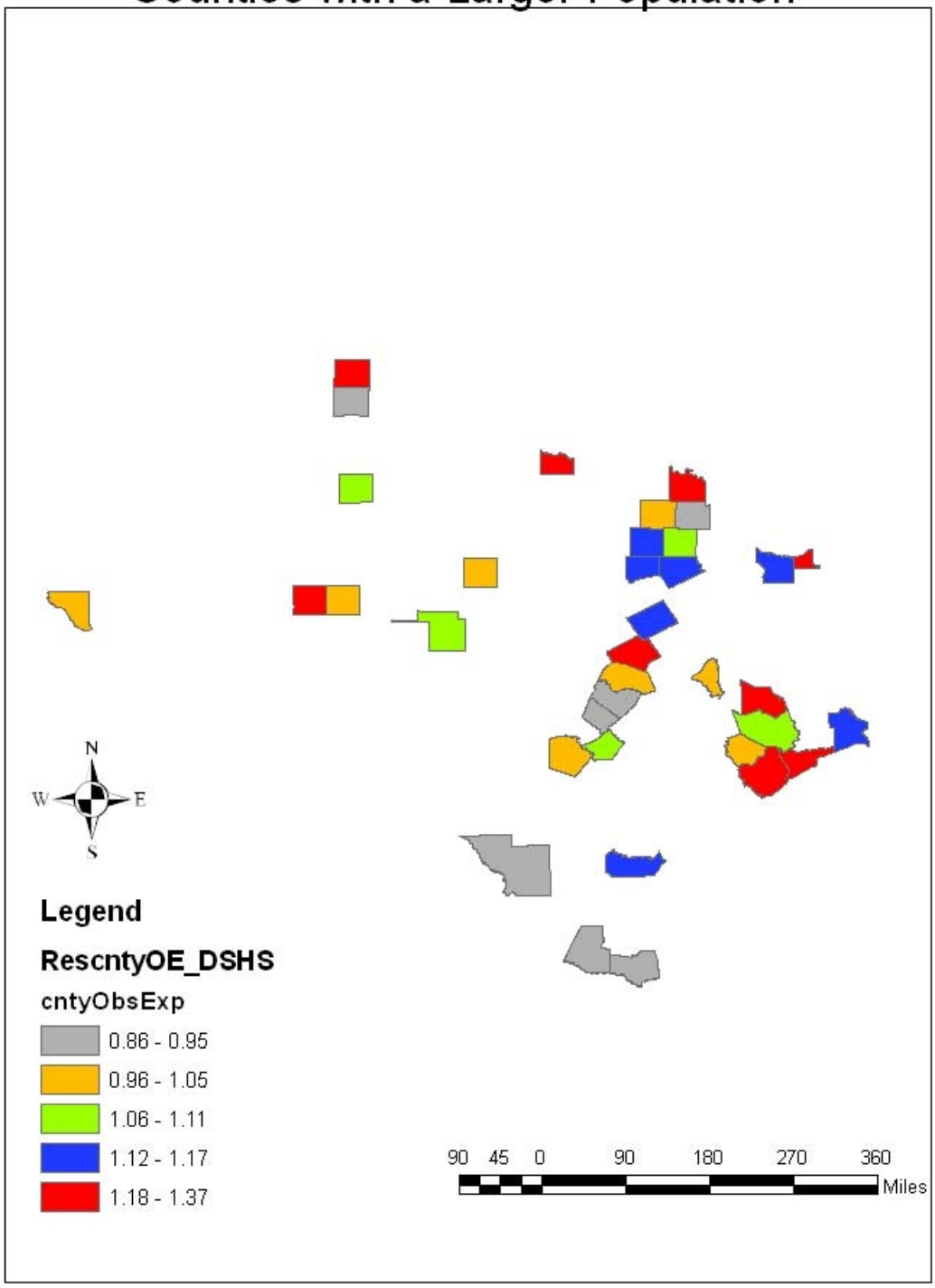
Observed to Expected Ratios of Leukemia Cases for Counties with a Larger Population



Observed to Expected Ratios of Lymphoma Cases for Counties with a Larger Population

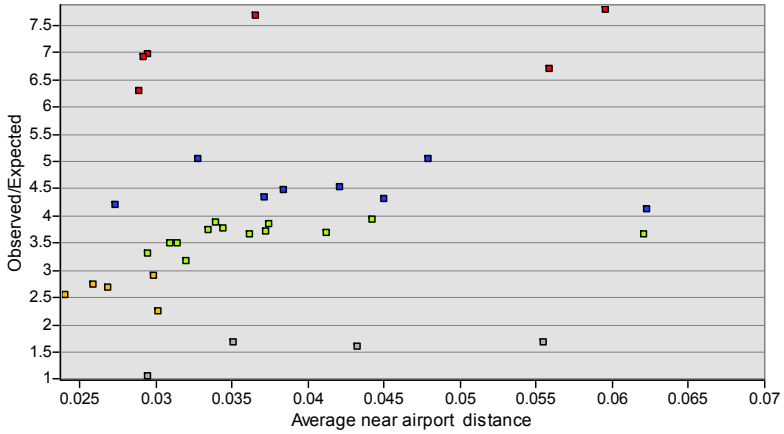


Observed to Expected Ratios of Respiratory Cases for Counties with a Larger Population



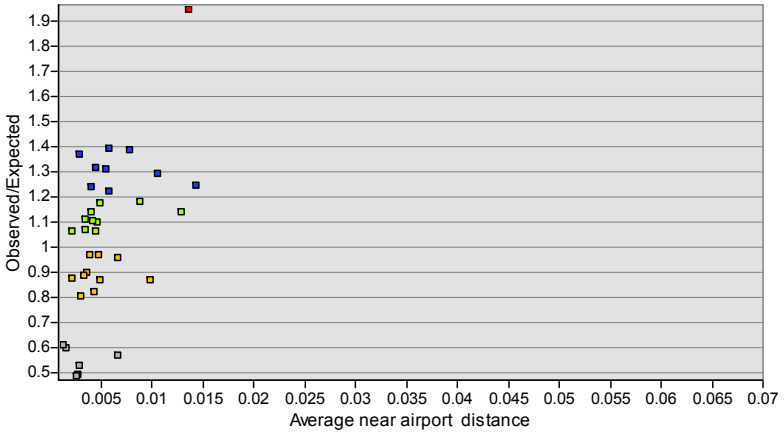
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Airports

Colon Cancer



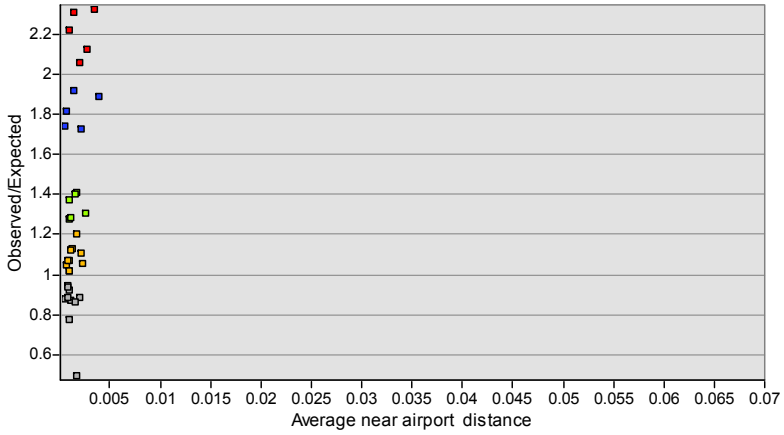
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



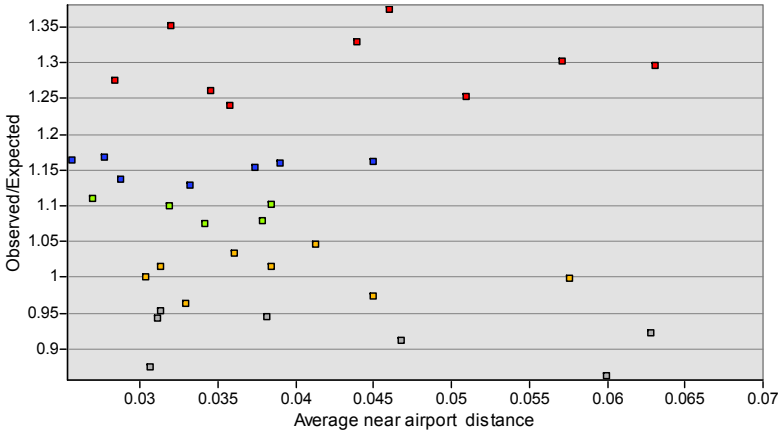
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

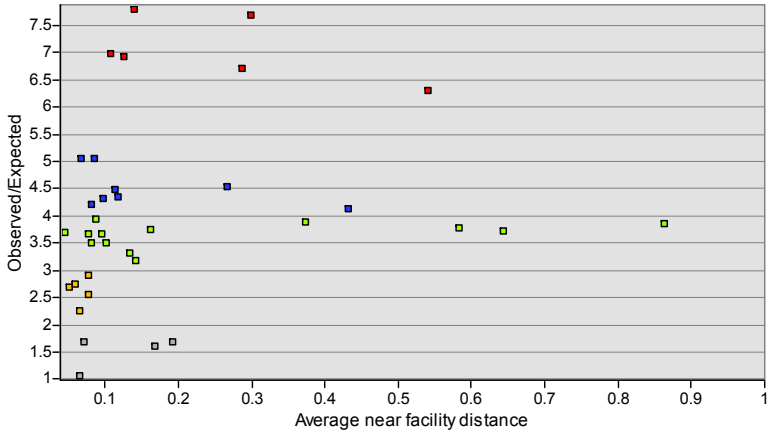
Respiratory Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

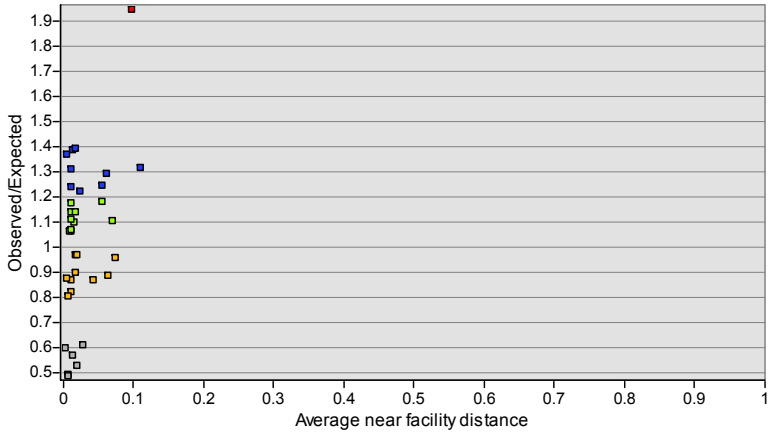
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Facilities

Colon Cancer



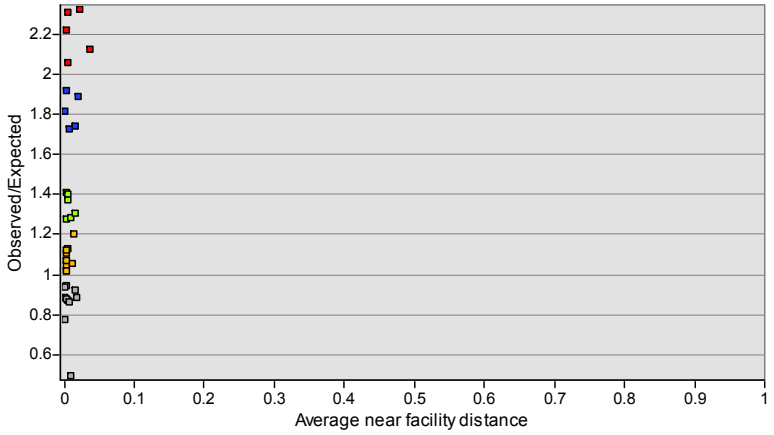
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



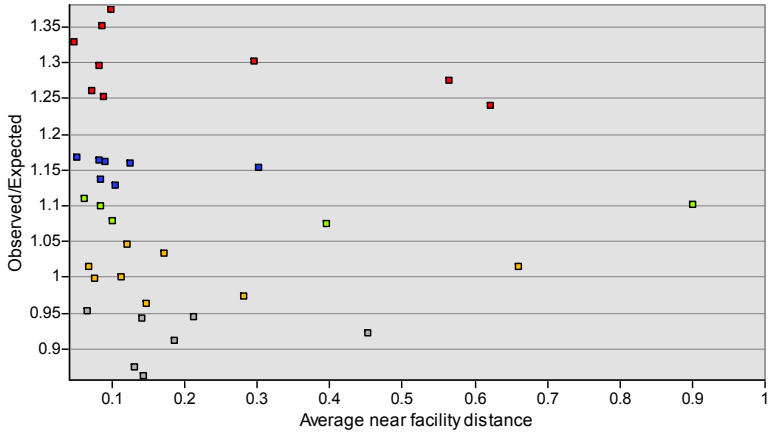
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

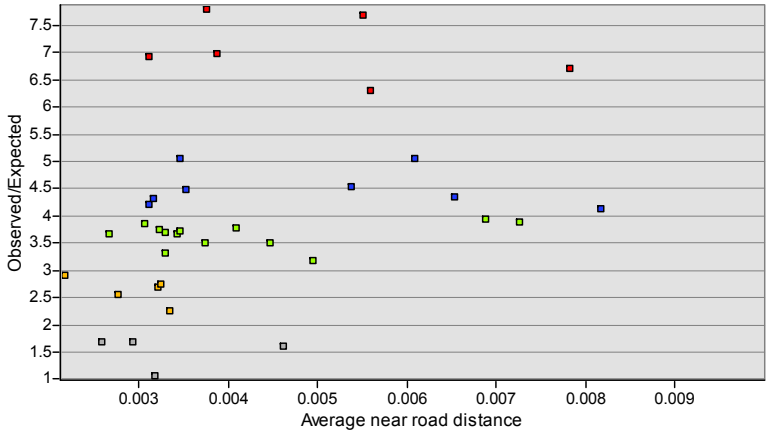
Respiratory Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

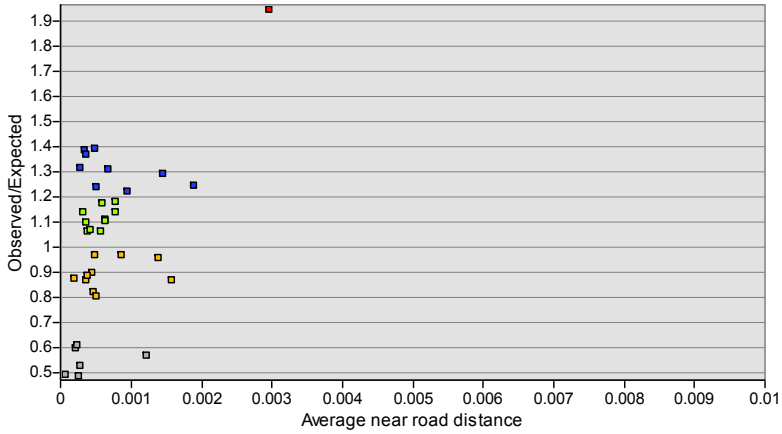
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Roads

Colon Cancer



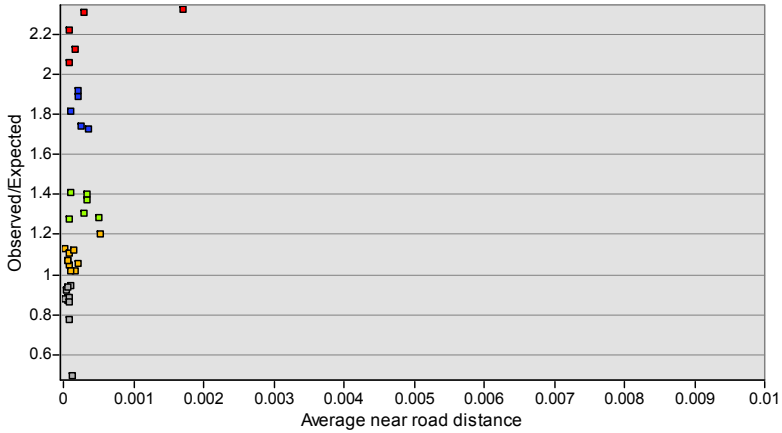
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



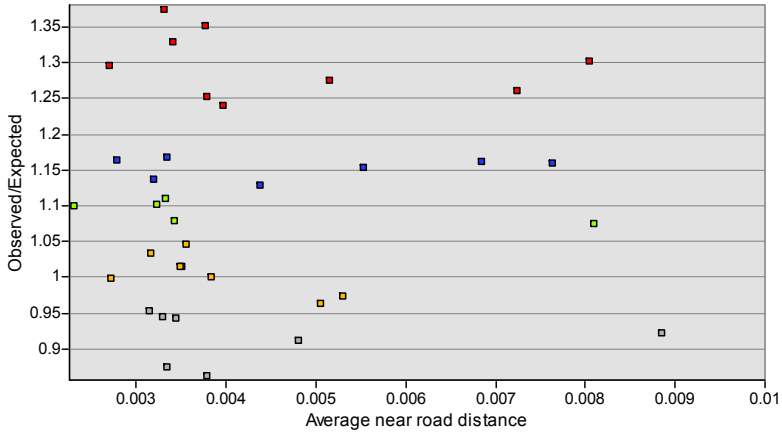
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



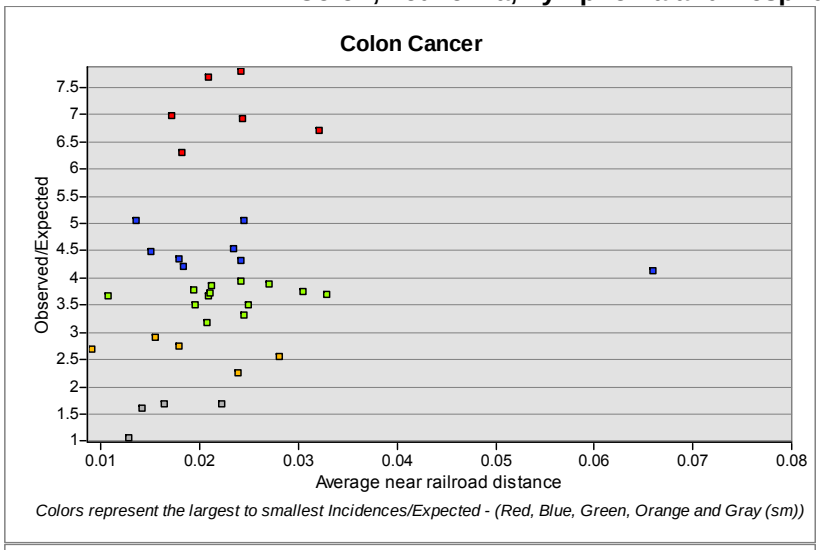
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

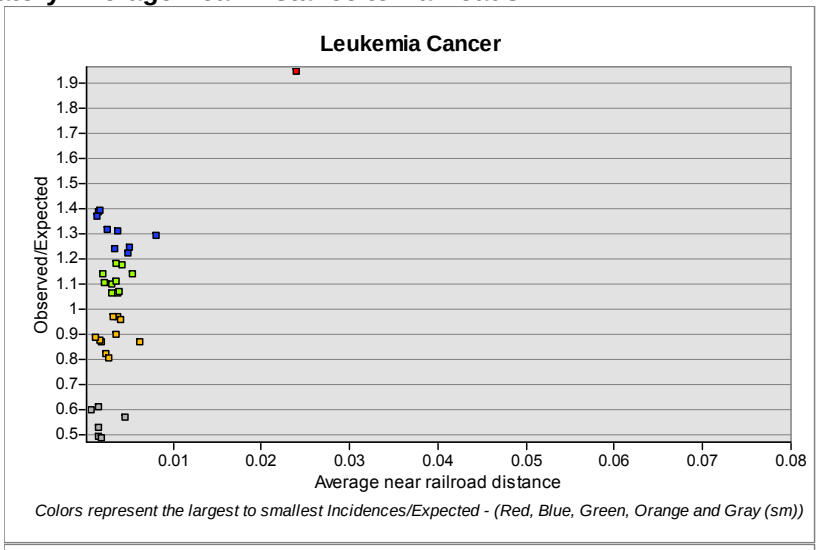
Respiratory Cancer

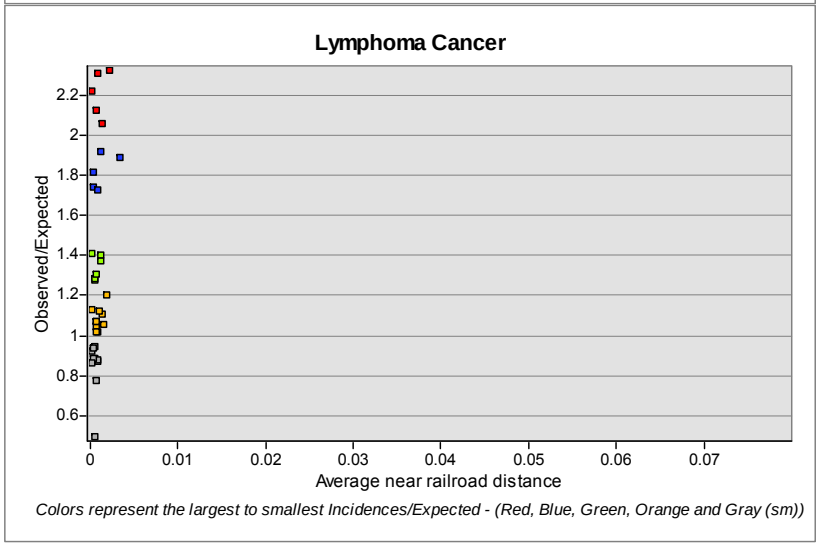


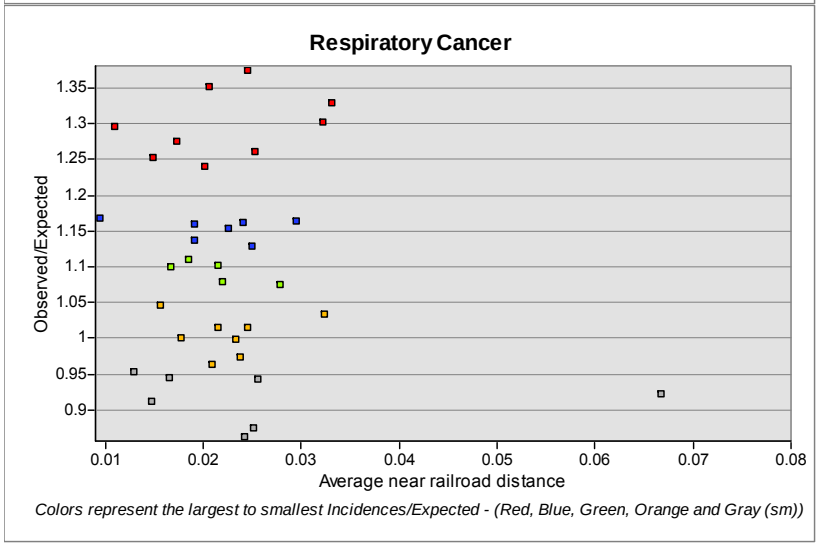
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to Railroads



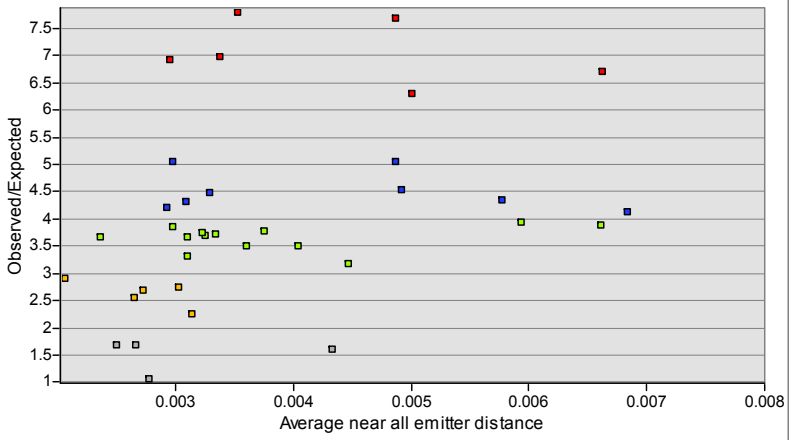






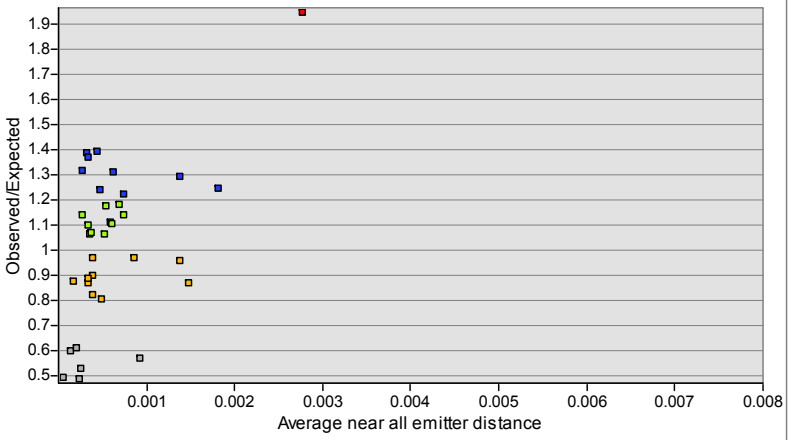
Colon, Leukemia, Lymphoma and Respiratory Average Near Distance to All Emitters

Colon Cancer



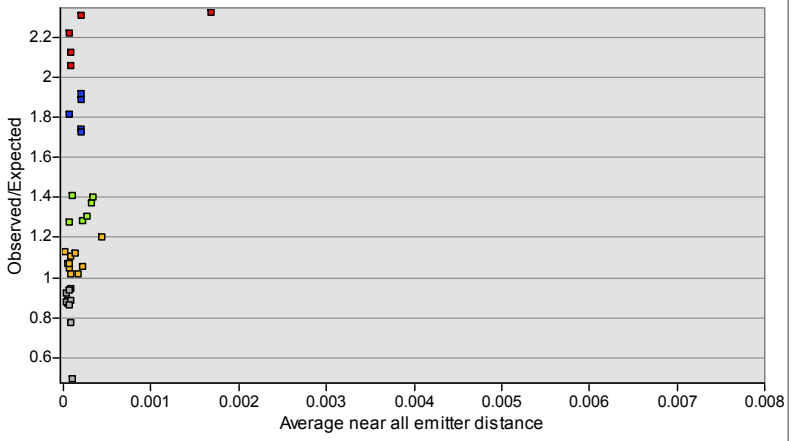
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Leukemia Cancer



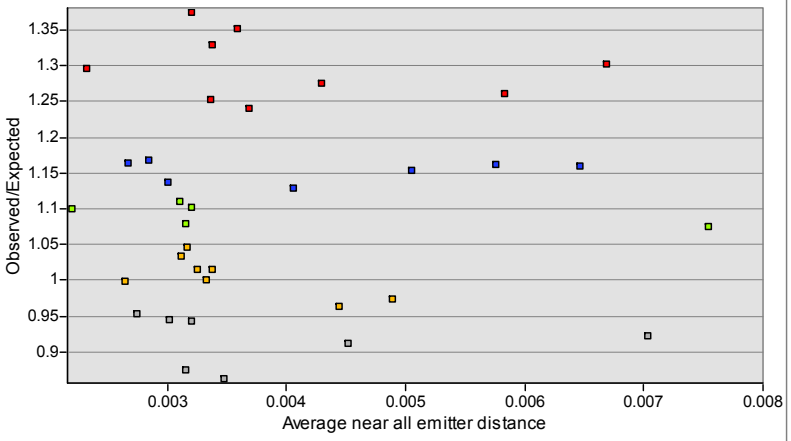
Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Lymphoma Cancer



Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

Respiratory Cancer

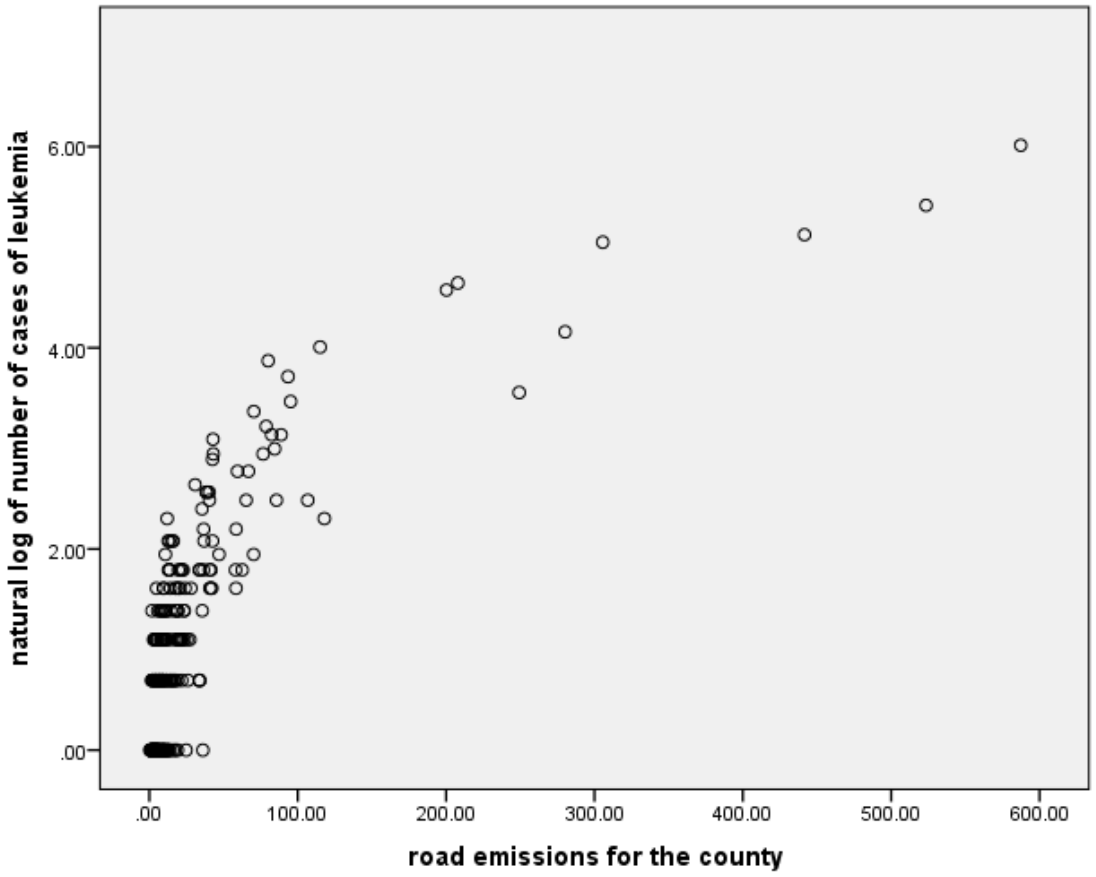


Colors represent the largest to smallest Incidences/Expected - (Red, Blue, Green, Orange and Gray (sm))

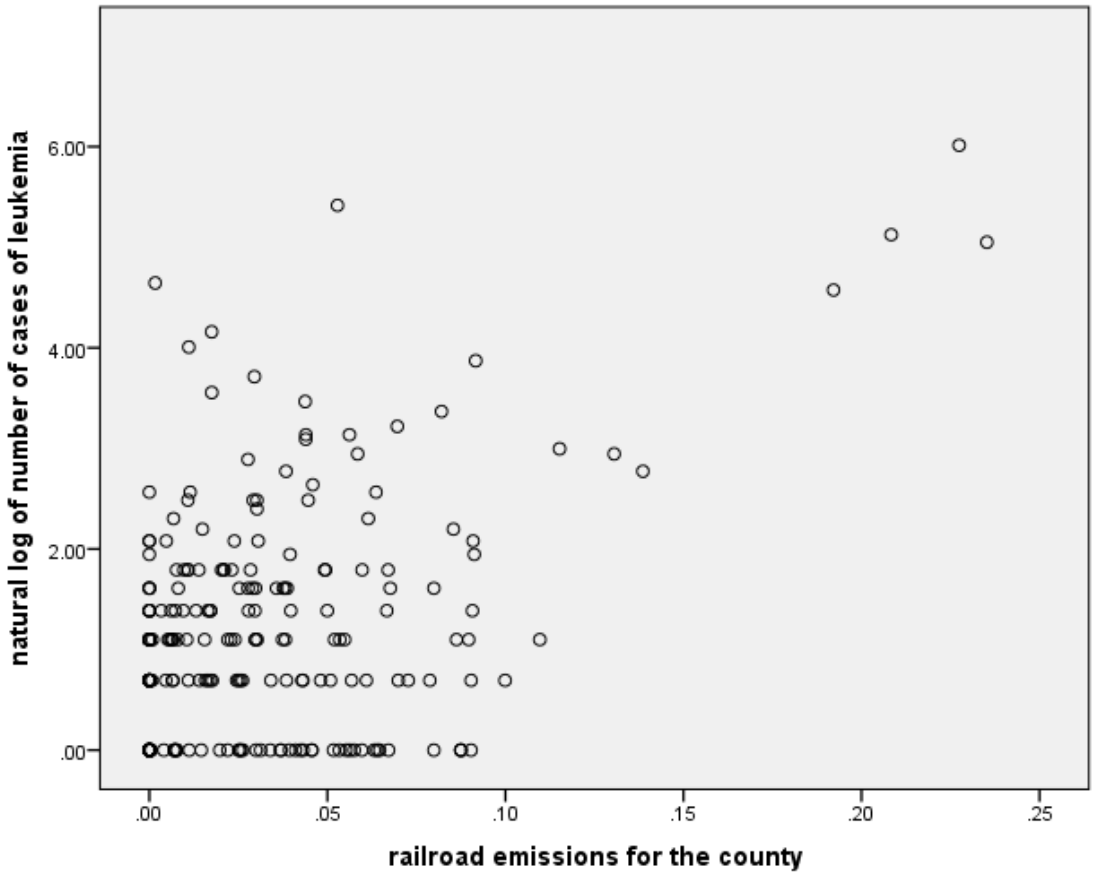
APPENDIX C

STATISTICS ITEMS

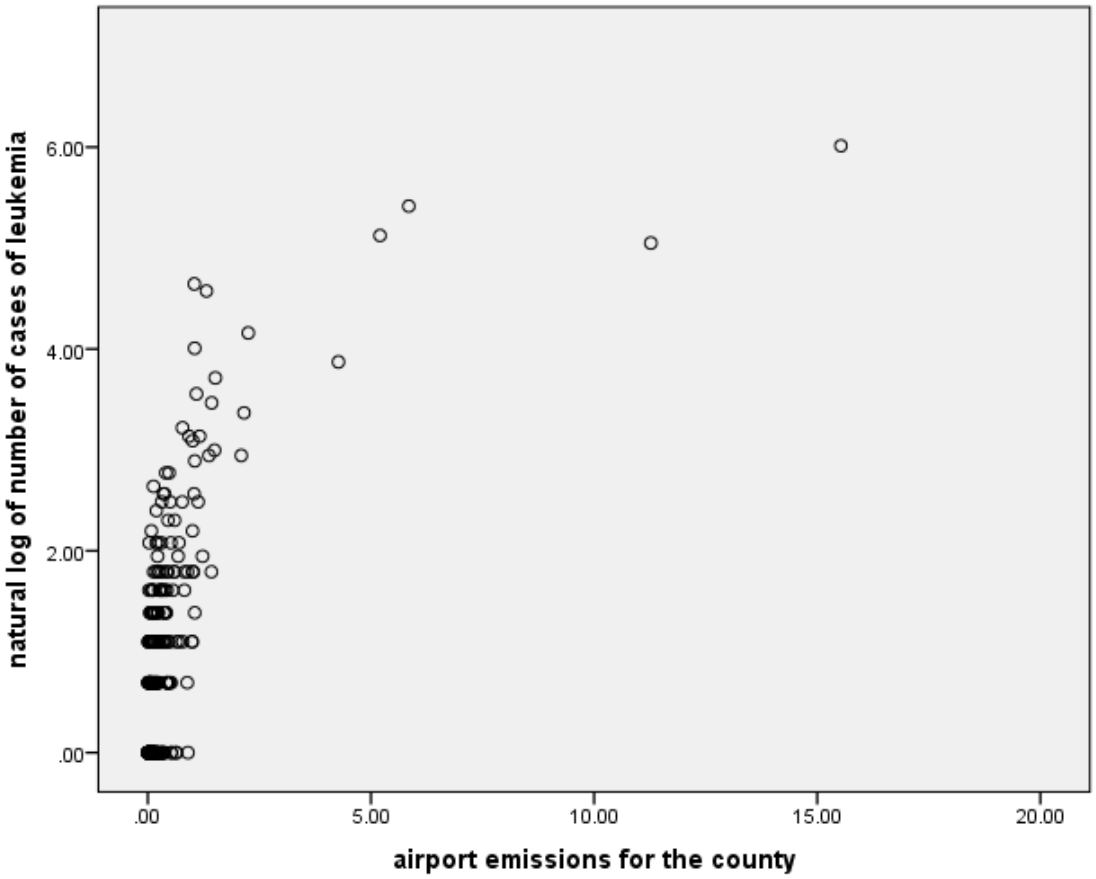
Graph of natural log of number of cases of leukemia per county verses county road emissions in tons/year. This graph demonstrates the need for transforming the data in order to increase linearity.



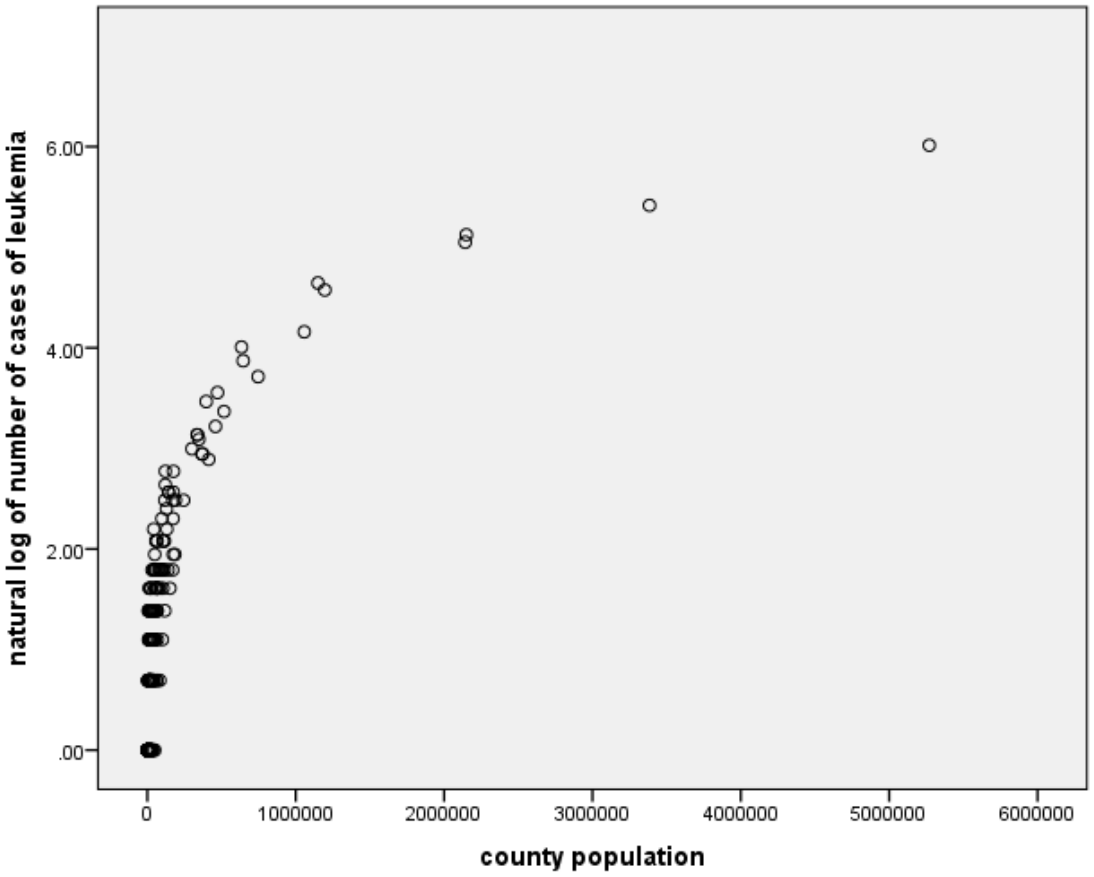
Graph of natural log of number of cases of leukemia per county verses county railroad emissions in tons/year. This graph demonstrates the need for transforming the data in order to increase linearity.



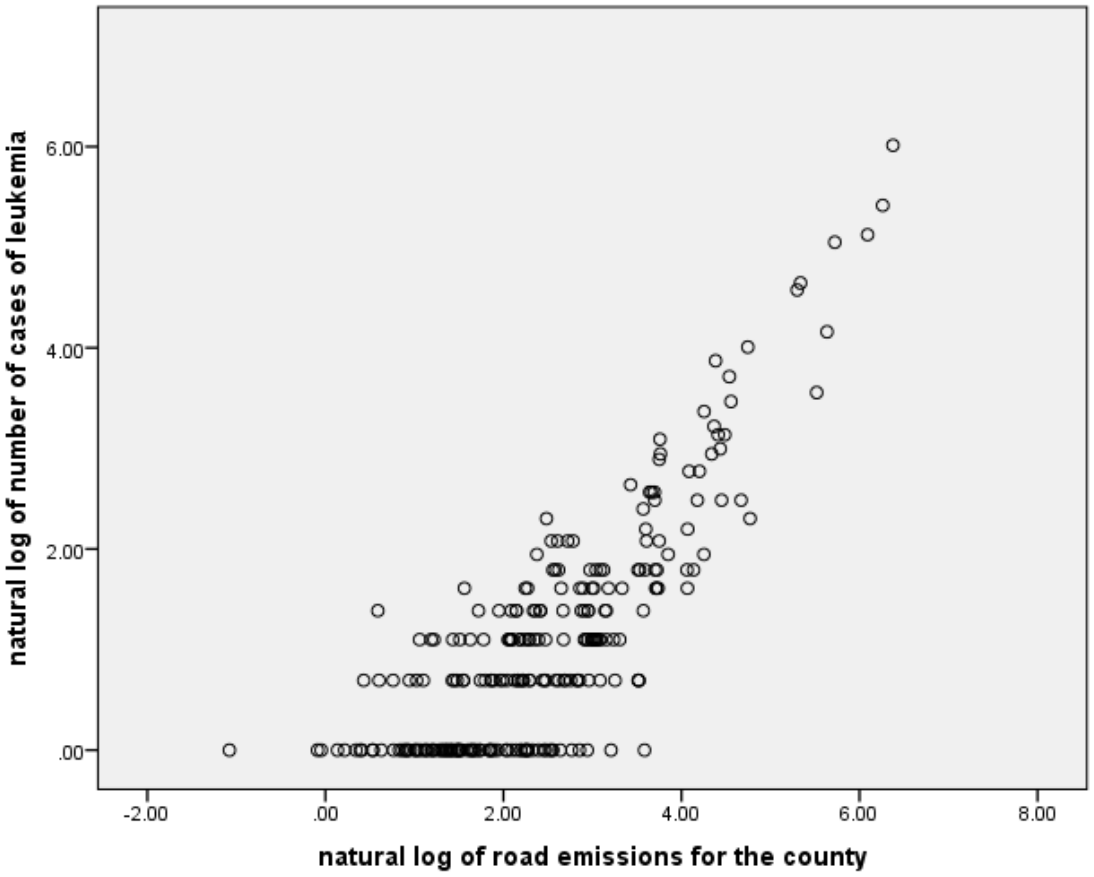
Graph of natural log of number of cases of leukemia per county verses county airport emissions in tons/year. This graph demonstrates the need for transforming the data in order to increase linearity.



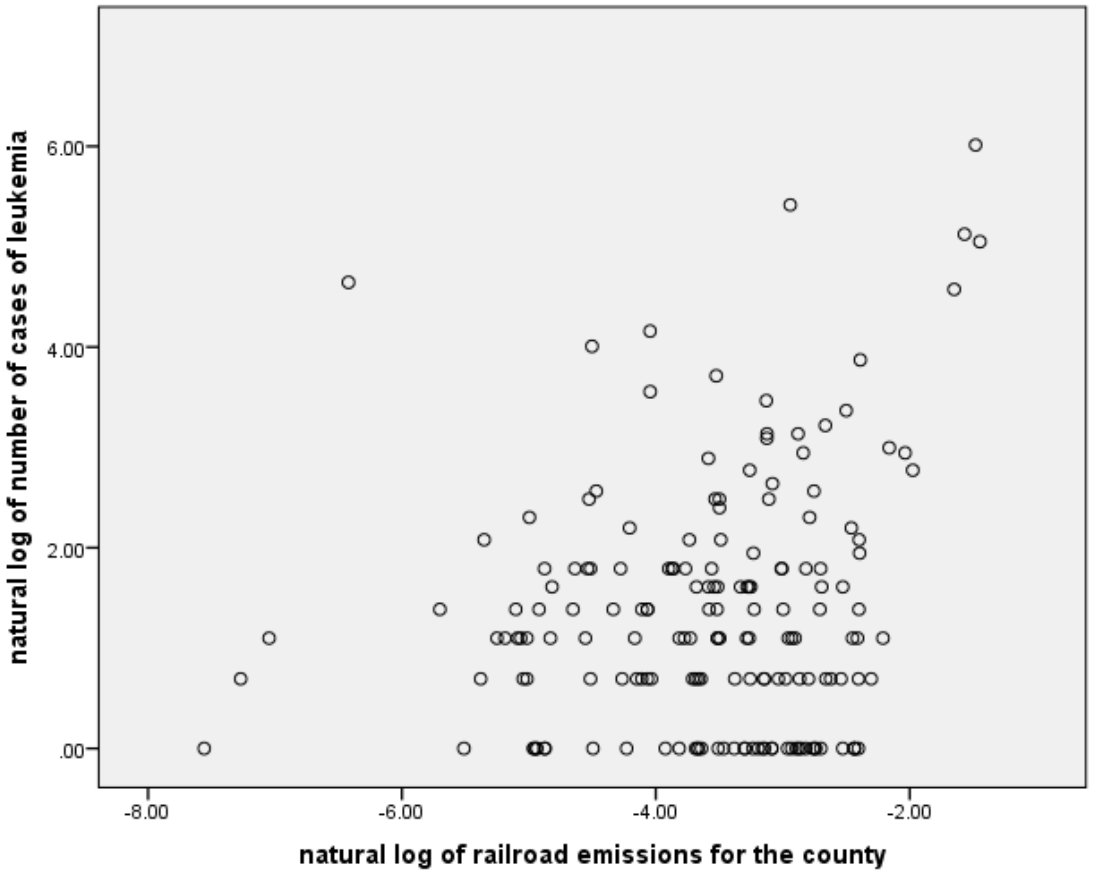
Graph of natural log of number of cases of leukemia per county verses county population. This graph demonstrates the need for transforming the data in order to increase linearity.



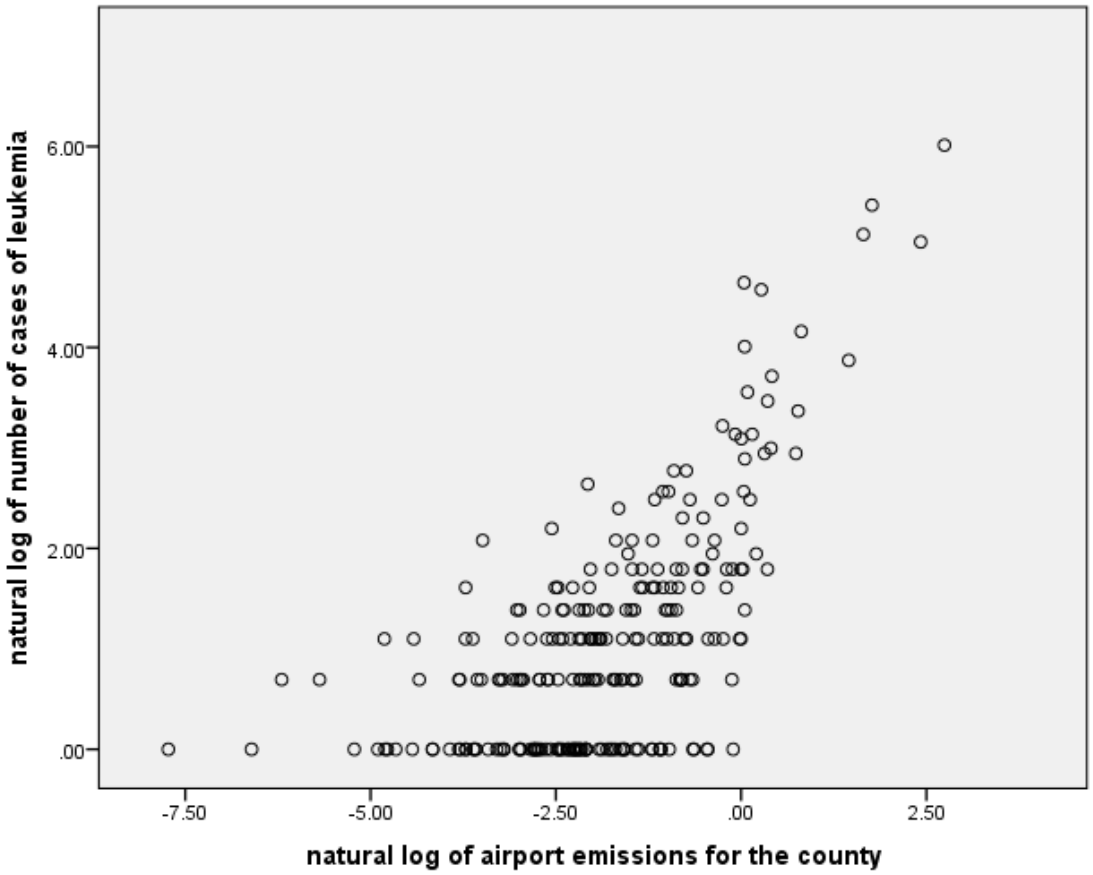
Graph of natural log of number of cases of leukemia per county verses natural log of county road emissions in tons/year. This graph demonstrates an improvement in linearity.



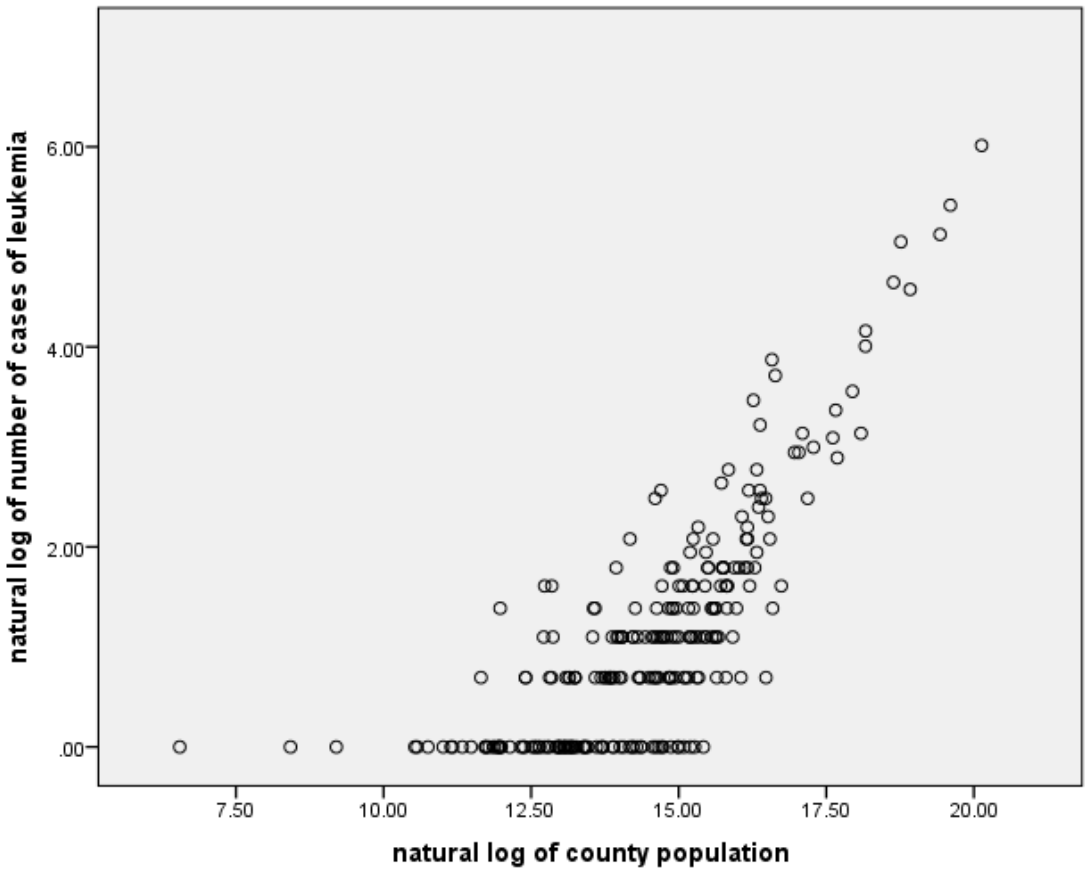
Graph of natural log of number of cases of leukemia per county verses natural log of county railroad emissions in tons/year. This graph demonstrates an improvement in linearity.



Graph of natural log of number of cases of leukemia per county verses natural log of county airport emissions in tons/year. This graph demonstrates an improvement in linearity.



Graph of natural log of number of cases of leukemia per county verses natural log of county population. This graph demonstrates an improvement in linearity.



Poisson Model Run Output

List of Model Terms

N = number of items
Sum_Count = number of leukemia cases (in children 9 years and under) within the county
lnpop = natural log of number of cases of leukemia (in children 9 years and under) per county
lnonroad = natural log of county-wide benzene emissions from roads (tons/year),
lnairport = natural log of county-wide benzene emissions from airports (tons/year),

Model Information

Dependent Variable	Sum_Count
Probability Distribution	Poisson
Link Function	Log

Case Processing Summary

	N	Percent
Included	251	98.8%
Excluded	3	1.2%
Total	254	100.0%

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Sum_Count	251	.00000000	408.00000000	8.4940239044	34.08194484098
Covariate	lnpop	251	8.42	20.13	14.5229	1.76924
	lnonroad	251	-.09	6.38	2.5039	1.19587
	lnairports	251	-7.73	2.74	-1.8157	1.47874

Goodness of Fit[d]

	Value	df	Value/df
Deviance	435.929	247	1.765
Scaled Deviance	247.000	247	
Pearson Chi-Square	462.482	247	1.872
Scaled Pearson Chi-Square	262.045	247	
Log Likelihood[a,b]	-490.311		
Adjusted Log Likelihood[c]	-277.813		
Akaike's Information Criterion (AIC)	988.622		
Finite Sample Corrected AIC (AICC)	988.784		
Bayesian Information Criterion (BIC)	1002.724		
Consistent AIC (CAIC)	1006.724		

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
4378.942	3	.000

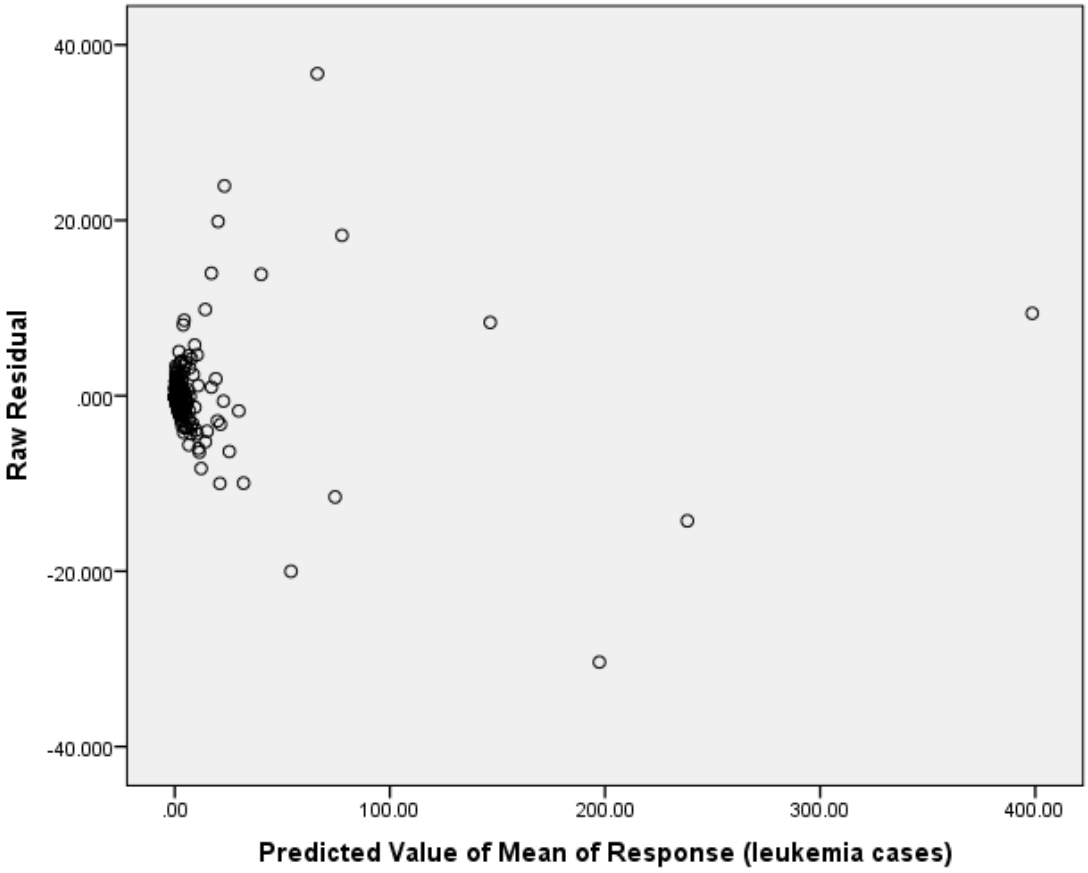
Tests of Model Effects

Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	100.631	1	.000
lnpop	75.221	1	.000
lnonroad	48.811	1	.000
lnairports	30.813	1	.000

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-6.664	.6643	-7.967	-5.362	100.631	1	.000
lnpop	.440	.0507	.340	.539	75.221	1	.000
lnonroad	.497	.0712	.358	.637	48.811	1	.000
lnairports	.230	.0414	.149	.311	30.813	1	.000
(Scale)	1.765[a]						

Graph of Residuals from model by predicted leukemia cases per county



T Test Output

The below T Test was performed in order to determine whether the difference in mean distance (in miles) to airports (of individuals 9 & under with or without leukemia) were statistically significant.

The SAS System 15:10 Saturday, February 18, 2012 1

The TTEST Procedure

Variable: airports (airports)

sick	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	4133728	0.0447	0.0346	0.000017	9.024E-6	0.5211
1	2133	0.0426	0.0345	0.000746	0.00157	0.4257
Diff (1-2)		0.00215	0.0346	0.000750		

sick	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		0.0447	0.0447	0.0448	0.0346	0.0346	0.0347
1		0.0426	0.0411	0.0441	0.0345	0.0335	0.0355
Diff (1-2)	Pooled	0.00215	0.000676	0.00362	0.0346	0.0346	0.0347
Diff (1-2)	Satterthwaite	0.00215	0.000682	0.00361			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	4.14E6	2.86	0.0042
Satterthwaite	Unequal	2134.2	2.87	0.0041

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	4.13E6	2132	1.01	0.7639

The below T Test was performed in order to determine whether the difference in mean distance (in miles) to railroads (of individuals 9 & under with or without leukemia) were statistically significant.

The TTEST Procedure

Variable: railroads (railroads)

sick	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	4133728	0.0323	0.0574	0.000028	0	0.8949
1	2133	0.0310	0.0519	0.00112	3.6E-6	0.6873
Diff (1-2)		0.00127	0.0574	0.00124		

sick	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		0.0323	0.0322	0.0323	0.0574	0.0574	0.0575
1		0.0310	0.0288	0.0332	0.0519	0.0504	0.0535
Diff (1-2)	Pooled	0.00127	-0.00117	0.00371	0.0574	0.0574	0.0575
Diff (1-2)	Satterthwaite	0.00127	-0.00093	0.00348			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	4.14E6	1.02	0.3062
Satterthwaite	Unequal	2134.7	1.13	0.2573

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	4.13E6	2132	1.23	<.0001

The below T Test was performed in order to determine whether the difference in mean distance (in miles) to facilities (of individuals 9 & under with or without leukemia) were statistically significant.

The TTEST Procedure

Variable: facilities (facilities)

sick	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	4133728	0.2121	0.2389	0.000117	0.000118	2.6590
1	2133	0.1898	0.2179	0.00472	0.00254	1.3398
Diff (1-2)		0.0222	0.2388	0.00517		

sick	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		0.2121	0.2119	0.2123	0.2389	0.2387	0.2390
1		0.1898	0.1806	0.1991	0.2179	0.2116	0.2247
Diff (1-2)	Pooled	0.0222	0.0121	0.0324	0.2388	0.2387	0.2390
Diff (1-2)	Satterthwaite	0.0222	0.0130	0.0315			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	4.14E6	4.30	<.0001
Satterthwaite	Unequal	2134.6	4.71	<.0001

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	4.13E6	2132	1.20	<.0001

The below T Test was performed in order to determine whether the difference in mean distance (in miles) to roads (of individuals 9 & under with or without leukemia) were statistically significant.

The TTEST Procedure

Variable: roads (roads)

sick	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	4133728	0.00483	0.00729	3.585E-6	0	0.2845
1	2133	0.00518	0.00971	0.000210	2.03E-7	0.1980
Diff (1-2)		-0.00034	0.00729	0.000158		

sick	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		0.00483	0.00483	0.00484	0.00729	0.00728	0.00729
1		0.00518	0.00476	0.00559	0.00971	0.00943	0.0100
Diff (1-2)	Pooled	-0.00034	-0.00065	-0.00004	0.00729	0.00728	0.00729
Diff (1-2)	Satterthwaite	-0.00034	-0.00076	0.000068			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	4.14E6	-2.18	0.0291
Satterthwaite	Unequal	2133.2	-1.64	0.1016

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	2132	4.13E6	1.78	<.0001

REFERENCES

[1] Administration Bureau of Transportation. March 2010 Airline Traffic Data: System Traffic Up 2.4 Percent from March 2009 (See table 5 titled Top 10 U.S. Airports (ranked by January-March 2010 System Scheduled Enplanements on U.S. Airlines)). http://www.bts.gov/press_releases/2010/bts029_10/html/bts029_10.html. Accessed June 12, 2010

[2]Cancer data have been provided by the Texas Cancer Registry statewide incidence file, 1995-2007, Department of State Health Services, created December, 2008.

[3]Texas statewide Age-Specific Rates, Average Annual Rates, 1995-2000, Selected Age Groups, and by race and ethnicity and sex for All Childhood Leukemias combined (ICCC-3 Group I) Rates are per 1,000,000. Data Request 11/5/2010 Incidence Source: prepared by Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3, .

[4]Texas Statewide Age-Specific Rates, Average Annual Rates, 1995-2000 Selected Age Groups, and by race and ethnicity and sex (Data Release #10385) All Childhood Lymphomas (ICCC-3: II Lymphomas and reticuloendothelial neoplasms) Rates are per 1,000,000. Data Request # 10385 11/16/2010 Incidence Source: prepared by Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3, .

[5]Texas Statewide Age-Specific Rates, Average Annual Rates, 1995-2000, Five-Year Age Groups, and by race and ethnicity and sex (Data request # 10385) All Respiratory System Cancers (ICD-O-3), Rates per 100,000. Data Request # 10385 (table 2) 11/19/2010 Incidence Source: prepared by Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3, .

[6]Texas statewide Age-Specific Rates, Average Annual Rates, 1995-2000, Five-Year Age Groups, and by race and ethnicity and sex Colon Cancer (ICD-O-3), Rates are per 100,000. Data Request 11/5/2010 Incidence Source: prepared by Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3, .

[7]Cancer data have been provided by the Texas Cancer Registry, Cancer Epidemiology and Surveillance Branch, Texas Department of State Health Services, 1100 W. 49th Street, Austin, Texas, 78756, http://www.dshs.state.tx.us/tcr/default.shtm, or (512) 458-7523. Viewed 2009.

[8]Texas State Library and Archives Commission. United States and Texas Populations 1850-2009. http://www.tsl.state.tx.us/ref/abouttx/census.html. Viewed June 12, 2010.

[9]Howe HL, Lake A, Firth R, Lehnherr M, Bayakly R, Copeland G, Wu XC, Stroup A, Roney D (eds). Cancer in
North America, 2001-2005. Volume One: Combined Cancer Incidence for the United States and Canada.
Springfield, IL: North American Association of Central Cancer Registries, Inc. May 2008. pageIII-6.

[10]The Texas Cancer Council, the Texas Department of Health, the University of Texas M. D. Anderson Cancer Center & Other Institutions. Impact of Cancer on Texas 5th Edition. The

University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, 1991. page 1.
[11]The Texas Cancer Council, the Texas Department of Health, the University of Texas M. D. Anderson Cancer Center & Other Institutions. Impact of Cancer on Texas 6th Edition. The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, 1995. page vii.
[12]American Cancer Society. Cancer Facts & Figures 2008. Atlanta: American Cancer Society; 2008. page 2. pdf
[13]American Cancer Society. Cancer Facts & Figures 2009. Atlanta: American Cancer Society; 2009. page 2. pdf
[14]The Texas Cancer Council, the Texas Department of Health, the University of Texas M. D. Anderson Cancer Center & Other Institutions. Impact of Cancer on Texas 5th Edition. The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, 1991. page v.
[15]Tan, A., Freeman, D.H., Freeman, J.L., Zhang, D.D., Dayal, H., and Philips, B.U. (March 2009).
The Cost of Cancer in Texas, 2007. Texas Cancer Registry, Texas Department of State Health Services, Publication
No. 10-13121.. page 12.
[16]Tan, A., Freeman, D.H., Freeman, J.L., Zhang, D.D., Dayal, H., and Philips, B.U. (March 2009).
The Cost of Cancer in Texas, 2007. Texas Cancer Registry, Texas Department of State Health Services, Publication
No. 10-13121.. page 16.
[17]Tan, A., Freeman, D.H., Freeman, J.L., Zhang, D.D., Dayal, H., and Philips, B.U. (March 2009).
The Cost of Cancer in Texas, 2007. Texas Cancer Registry, Texas Department of State Health Services, Publication
No. 10-13121.. page 18.
[18]Research and Innovative Technology Administration Bureau of Transportation. March 2010 Airline Traffic Data: System Traffic Up 2.4 Percent from March 2009 (See table 5 titled Top 10 U.S. Airports (ranked by January-March 2010 System Scheduled Enplanements on U.S. Airlines)). http://www.bts.gov/press_releases/2010/bts029_10/html/bts029_10.html. Accessed June 12, 2010
[19]University of Washington along with partial funding from EPA. The MESA Air Pollution Study, Strengthening the Scientific Foundation for Air Quality Management; 2006. page 2.
[20]U.S. Department of Health and Human Services Public Health Service Agency for Toxic Substances and Disease Registry, Toxicological Profile for Benzene (update), Prepared by Syracuse Research Corporation Under Contract No. 200-2004-09793. August 2007. page 11.
[21]World Health Organization (WHO) (1999). Air quality guidelines for Europe, WHO regional publication, European series.
http://www.euro.who.int/__data/assets/pdf_file/0005/74732/E71922.pdf. Viewed May 22, 2011
[22]Kristina W. Whitworth, Elaine Symanski, and Ann L. Coker. Childhood Lymphohematopoietic Cancer Incidence and Hazardous Air Pollutants in Southeast Texas, 1995–2004, Environmental Health Perspectives VOLUME 116 | NUMBER 11 | November 2008,
[23]Transportation Research Board of the National Academies - Sponsored by the Federal Aviation Administration, Ezra Wood, Scott Herndon, Richard Miake-Lye, David Nelson, Mara Seeley. Aircraft and Airport-Related Hazardous Air Pollutants: Research Needs and Analysis. Airport Cooperative Research Program (ACRP Report 7) May 2008.

[24]Teterboro Airport, Teterboro, New Jersey Detailed Air Quality Evaluation, Prepared for New Jersey Department of Environmental Protection Trenton, New Jersey, ENVIRON International Corporation, February 11, 2008
[25]Ying Zhou*1 and Jonathan I Levy2, "Between-airport heterogeneity in air toxics emissions associated with individual cancer risk thresholds and population risks," Environmental Health 2009, 8:22 doi:10.1186/1476-069X-8-22, May 2009
[26]University of Washington along with partial funding from EPA. The MESA Air Pollution Study, Strengthening the Scientific Foundation for Air Quality Management; 2006. page 2.
[27]An Li, Todd M. Schoonover et al. Plycyclic Aromatic Hydrocarbons in the Air of Ten Chicago Area Homes. Mickey Leland National Urban Air Toxics Research Center, NUATRC Research Report No. 6 2005. page 3.
[28]An Barbara J Turpin, Clifford P Weisel, et al, Relathionships of Indoor, Outdoor, and Personal Air (RIOPA). Mickey Leland National Urban Air Toxics Research Center, HEI Report 130, NUATRC Research Report No.106 2007. page 2.
[29]James A Thompson, Susan E Carozza and Li Zhu via Environmental Health Journal. Geographic risk modeling of childhood cancer relative to county-level crops, hazardous air pollutants and population density characteristics in Texas. http://www.ehjournal.net/content/7/1/45. Published September 25, 2008. Viewed 2011
[30]James A Thompson, Susan E Carozza and Li Zhu via Environmental Health Journal. Geographic risk modeling of childhood cancer relative to county-level crops, hazardous air pollutants and population density characteristics in Texas. http://www.ehjournal.net/content/7/1/45. Published September 25, 2008. Viewed 2011
[31]Daniel Krewski. Extended Analysis of the American Cancer Society Study of Particulate Air Pollution and Mortality, Statement Synopsis of Research Report 140. Health Effects Institute 101 Federal Street, Suite 500
Boston, MA 02110. 2001.
[32]Daniel Krewski. Extended Analysis of the American Cancer Society Study of Particulate Air Pollution and Mortality, Statement Synopsis of Research Report 140. Health Effects Institute 101 Federal Street, Suite 500
Boston, MA 02110. 2001.
[33]Kyung-Hwa Jung, Francisco Artigas, Jin Young Shin. Personal, indoor, and outdoor exposure to VOCs in the immediate vicinity of a local airport. In Environ Monit Assess, Publication Info: @ springer Science+Business Media B.V.2010. page 10.
[34]Kyung-Hwa Jung, Francisco Artigas, Jin Young Shin. Personal, indoor, and outdoor exposure to VOCs in the immediate vicinity of a local airport. In Environ Monit Assess, Publication Info: @ springer Science+Business Media B.V.2010. page 13
[35]Environ International Corporation: Preliminary Modeling Evaluation of Risks Associated with Emissions from Chicago O'Hare Airport. Prepared for City of Park Ridge, Illinois; 1999.
[36]Aircraft Emission Impacts in a Neighborhood Adjacent to a General Aviation Airport in Southern California. Shishan Hu, Scott Fryun, Kathleen Kozawa, Steve Mara, Arthur M Winer, and Suzanne E Paulson, Department of Atmospheric and Oceanic Sciences, University of California, 405 Hilgard Ave., Los Angeles, California 90095-1565, Environmental Health Sciences Department, School of Public Health, University of California, 650 Charles E. Young Drive South, Los Angeles, California 90095-1772,Preventive Medicine, Environmental Health Division, Keck School of Medicine, University of Southern California, 1540 Alcazar Street CHP-236 Los Angeles, California 90032, and California Air Resources Board, Research Division, 1001 I Street, Sacramento, California 95814, Accepted September 9, 2009. Environ. Sci. Technol. 2009, 43, 8039–8045

[37]Residential proximity to large airports and potential health impacts in New York State, S. Lin, J. P. Munsie, M. Herdt-Losavio, S. A. Hwang, K. Civerolo, K. McGarry, T. Gentile. Int Arch Occup Environ Health (2008) 81:797–804 DOI 10.1007/s00420-007-0265-1
[38]Research and Innovative Technology Administration Bureau of Transportation. March 2010 Airline Traffic Data: System Traffic Up 2.4 Percent from March 2009 (See table 5 titled Top 10 U.S. Airports (ranked by January-March 2010 System Scheduled Enplanements on U.S. Airlines)). http://www.bts.gov/press_releases/2010/bts029_10/html/bts029_10.html. Accessed June 12, 2010
[39]Transportation Research Board of the National Academies - Sponsored by the Federal Aviation Administration, Ezra Wood, Scott Herndon, Richard Miake-Lye, David Nelson, Mara Seeley. Aircraft and Airport-Related Hazardous Air Pollutants: Research Needs and Analysis. Airport Cooperative Research Program (ACRP Report 7) May 2008.
[40]Teterboro Airport, Teterboro, New Jersey Detailed Air Quality Evaluation, Prepared for New Jersey Department of Environmental Protection Trenton, New Jersey, ENVIRON International Corporation, February 11, 2008
[41]Between-airport heterogeneity in air toxics emissions associated with individual cancer risk thresholds and population risks, Ying Zhou*1 and Jonathan I Levy2, Environmental Health 2009, 8:22 doi:10.1186/1476-069X-8-22, May 2009
[42]SimplyMap. SimplyMap Login. http://simplymap.com/index.php. Downloaded in 2009.
[43]ESRI GIS Mapping Software. Download Census 2000 TIGER/Line Data. http://arcdata.esri.com/data/tiger2000/tiger_download.cfm. Downloaded May 15, 2010.
[44]Glenn, Barbara; Bateson, Thomas. Personal correspondence via e-mail, Sept. 13, 2010.
[45]U.S. Department of Health and Human Services Public Health Service Agency for Toxic Substances and Disease Registry, Toxicological Profile for Benzene (update), Prepared by Syracuse Research Corporation Under Contract No. 200-2004-09793. August 2007. page 11.
[46]Department of State Health Services (DSHS). Texas Cancer Registry limited use data files (Data Dictionary.) http://www.dshs.state.tx.us/tcr/researchers.shtm#datasets. Viewed 2009.
[47]Cancer data have been provided by the Texas Cancer Registry, Cancer Epidemiology and Surveillance Branch, Texas Department of State Health Services, 1100 W. 49th Street, Austin, Texas, 78756, http://www.dshs.state.tx.us/tcr/default.shtm, or (512) 458-7523. Viewed 2009.
[48]Department of State Health Services (DSHS). Texas Cancer Registry, Expected Numbers of Cancer Cases and Deaths, Texas, 2005. http://www.dshs.state.tx.us/tcr/expected-2005.shtm. Viewed August 7, 2011.
[49]Texas Cancer Registry. Cancer Rates in North America. http://www.cancer-rates.info/common/draw_copy.html. Viewed in August 7, 2011.
[50]Texas statewide Age-Specific Rates, Average Annual Rates, 1995-2000, Selected Age Groups, and by race and ethnicity and sex for All Childhood Leukemias combined (ICCC-3 Group I) Rates are per 1,000,000. Data Request 11/5/2010 Incidence Source: prepared by Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3,.
[51]Texas Statewide Age-Specific Rates, Average Annual Rates, 1995-2000 Selected Age Groups, and by race and ethnicity and sex (Data Release #10385) All Childhood Lymphomas (ICCC-3: II Lymphomas and reticuloendothelial neoplasms) Rates are per 1,000,000. Data Request # 10385 11/16/2010 Incidence Source: prepared by Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3, .
[52]Texas Statewide Age-Specific Rates, Average Annual Rates, 1995-2000, Five-Year Age Groups, and by race and ethnicity and sex (Data request # 10385) All Respiratory System Cancers (ICD-O-3), Rates per 100,000. Data Request # 10385 (table 2) 11/19/2010 Incidence

Source: prepared by Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3, .

[53]Texas statewide Age-Specific Rates, Average Annual Rates, 1995-2000, Five-Year Age Groups, and by race and ethnicity and sex Colon Cancer (ICD-O-3), Rates are per 100,000. Data Request 11/5/2010 Incidence Source: prepared by Texas Department of State Health Services, Cancer Epidemiology and Surveillance Branch, Texas Cancer Registry, Incidence - Texas, 1995-2007, Cut-off 11-19-09, SEER*Prep 2.4.3, .

[54]Center for Disease Control (CDC). CDC Wonder - United States Cancer Statistics Public Information Data: Incidence 1999 – 2006. http://wonder.cdc.gov/wonder/help/cancer-v2006.html#Population%20Data%20Sources. Viewed in August 7, 2011.

[55]ESRI GIS Mapping Software. ArcGIS Desktop Help 9.2. http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=An_overview_of_commonly_used_tools. Downloaded May 23, 2011.

[56]ESRI GIS Classifying numerical fields for graduated symbology. ArcGIS Resource Center Desktop Help 10. http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//00s50000001r000000.htm. Viewed April 4, 2012.

[57]Geospatial Analysis – Web edition. http://www.spatialanalysisonline.com/output/. April 4, 2012.

[58]Geospatial Analysis – Web edition. http://www.spatialanalysisonline.com/output/. April 4, 2012.

[59]Geospatial Analysis – Web edition. http://www.spatialanalysisonline.com/output/. April 4, 2012.

[60]U.S. Department of Health and Human Services Public Health Service Agency for Toxic Substances and Disease Registry, “Toxicological Profile for Benzene (update),” Prepared by Syracuse Research Corporation Under Contract No. 200-2004-09793. August 2007. page 11.

[61]Wheaton, W.D., Cajka, J.C., Chasteen, B.M., Wagener, D.K., Cooley, P.C., Ganapathi, L., Roberts, D.J., and Allpress, J.L. (2009). Synthesized Population Databases: A US Geospatial Database for Agent-Based Models. RTI Press publication No. MR-0010-0905. Research Triangle Park, NC: RTI International. Retrieved [October 20, 2010] from http://www.rti.org/rtipress

BIOGRAPHICAL INFORMATION

Sala Nanyanzi Senkayi was born in Kampala, Uganda and moved to the U.S. with her parents when she was a toddler. Sala graduated from Duncanville High School in Texas. She obtained a bachelor's degree in Biomedical Sciences from Texas A&M University in College Station, Texas. She also obtained two BS degrees in Microbiology and Biology from the University of Texas at Arlington. In 2010, she earned a Master of Science degree in Environmental and Earth Sciences from the same university. She is currently employed as an Environmental Scientist by the U.S. Environmental Protection Agency (EPA) Region 6 in Dallas, Texas. Sala worked as an EPA intern for several years before she became a full time EPA employee. Before joining EPA, she worked for Baylor College of Dentistry supervising a Biomedical Science Lab. While at Baylor College of Dentistry, she contributed to several papers published in peer-reviewed scientific journals. Her accomplishments at EPA include winning a Regional Research Partnership Program grant that enabled her to work with EPA's Office of Research and Development (ORD) scientists at the Research Triangle Park (RTP) located in North Carolina. This collaboration enabled her to gain knowledge about EPA's environmental datasets used in this study. In addition, this collaboration enabled her to develop an environmental epidemiological method that was used to determine whether or not relationships between airport emissions and cancer incidences exist within the State of Texas.