Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Nikhila Gandrakota Date

INCIDENCE OF CUTANEOUS T CELL LYMPHOMA AND AIR LEVELS OF BENZENE AND TRICHLOROETHYLENE

By

Nikhila Gandrakota, MBBS

Master of Public Health

Epidemiology

Dr Jeffrey M. Switchenko, PhD, MS

Faculty Thesis Advisor

Dr Kevin Ward, PhD, MPH

Faculty Thesis Advisor

Dr Mary Jo. Lechowicz, MD

Field Advisor

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University in partial
fulfillment of the requirements for the degree of
Master of Public Health in
Epidemiology
2020

INCIDENCE OF CUTANEOUS T CELL LYMPHOMA AND AIR LEVELS OF BENZENE AND TRICHLOROETHYLENE

By

Nikhila Gandrakota, MBBS

Master of Public Health

Epidemiology

Dr Jeffrey M. Switchenko, PhD, MS

Faculty Thesis Advisor

Dr Kevin Ward, PhD, MPH

Faculty Thesis Advisor

Dr Mary Jo. Lechowicz, MD

Field Advisor

An abstract of the thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University in partial
fulfillment of the requirements for the degree of
Master of Public Health in
Epidemiology
2020

Abstract

INCIDENCE OF CUTANEOUS T CELL LYMPHOMA AND AIR LEVELS OF BENZENE AND TRICHLOROETHYLENE

By Nikhila Gandrakota

Background: The incidence of CTCL has been increasing in the US and this might be due to increased exposure to industrial chemicals. Benzene and TCE are chemicals which are known carcinogens and have been associated with various malignancies. Increased incidence of CTCL has been linked to environmental exposures. This study explores the spatial association of these chemicals to the incidence of CTCL in various states across the US.

Methods: We used CTCL case data from the Surveillance, Epidemiology, and End Results (SEER) program for the years 2000-2016, population data from the U.S. Census and identified benzene and TCE exposure and concentration levels by county in 1996, 1999 and 2002 for all the states under review. We performed Poisson regression on observed cases of CTCL, using benzene and TCE levels and SES variables. We mapped standardized incidence ratios by county to examine spatial patterns. Cluster analyses was conducted at the county level for the states with more than 30 counties.

Results: Clusters of high standardized incidence ratios were identified in several of the states under review. This clustering of SIR was statistically significant in the states of California, Georgia, Iowa, Kentucky and Louisiana. Overall, Poisson regression models depicted an association between the benzene & CTCL incidence but not TCE.

Conclusions: CTCL incidence might be associated with exposure to toxic levels of Benzene and TCE. Further research should also investigate other industrial chemicals causing increased CTCL incidence as well as the spatial association in other states of the US.

INCIDENCE OF CUTANEOUS T CELL LYMPHOMA AND AIR LEVELS OF BENZENE AND TRICHLOROETHYLENE

By

Nikhila Gandrakota, MBBS

Master of Public Health

Epidemiology

Dr Jeffrey M. Switchenko, PhD, MS

Faculty Thesis Advisor

Dr Kevin Ward, PhD, MPH

Faculty Thesis Advisor

Dr Mary Jo. Lechowicz, MD

Field Advisor

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University in partial
fulfillment of the requirements for the degree of
Master of Public Health in
Epidemiology
2020

ACKNOWLEDGEMENTS

Dr. Jeffrey Switchenko, I am sincerely grateful for your support and guidance on this project.

Dr Mary Jo Lechowicz, your expertise and guidance are extremely valued.

Dr. Kevin Ward, thank you for your understanding and assistance.

Thanks to Mom, Dad, Grandmother, Grandfather, Saaketh, Sunil, Jyothi, Ankith, Krithi & Girija Atta for unending love and support!

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

NHL – Non-Hodgkin's lymphoma

CTCL – Cutaneous T cell Lymphoma

EPA – Environmental Protection Agency

NATA- National Air Toxics Assessment

NCI – National Cancer Institute

GIS – Geographic Information Systems

ASPEN – Assessment System for Population Exposure Nationwide

HAPEM – Hazardous Air Pollutant Exposure Model

ACS – American Community Survey

SEER – Surveillance, Epidemiology, and End Results Program

SIR – Standardized incidence ratio

TRI – Toxics Release Inventory

INTRODUCTION

Non-Hodgkin lymphomas (NHL) are a group of cancers that affect the immune system; the very mechanism which should protect our body from disease. NHLs originate in the lymph nodes and are composed of malignant lymphocytes (both B cells and T cells). In 2001, a systematic classification scheme for the 30 + different types of NHLs were created by the World Health Organization, which are then further separated by the cell type involved (either B cell or T cell). Cutaneous T-cell lymphoma (CTCL) is one of several types of non-Hodgkin's lymphoma where the skin is the primary site.

CTCLs have been recognized for over 200 years now, since one of its forms, Mycosis Fungoides was first described by Jean Louis Albert[1]. CTCL represents a variety of lymphomas with different clinical presentations, histological features and therapeutic considerations. CTCL patients may have a variety of clinical morphologies, depending on the disease subtype, including erythematous, hyperpigmented or hypopigmented patches with or without atrophy, or with thickened plaques, which may resemble mild, inflammatory or autoimmune disorders such as eczema, psoriasis, morphea, pityriasis lichenoides chronica, pityriasis rubra pilaris, drug eruptions, poikiloderma, panniculitis, vitiligo, and pigmented purpuric dermatoses. Therefore, in dermatology, CTCL is considered a "strong mimicker." In fact, CTCL is often hard to detect, especially in early and erythrodermic stages, and it takes an average of six years to achieve a definitive diagnosis after it is first presented[2].

CTCL predominantly manifests in the form of Mycosis fungoides (MF), primary cutaneous anaplastic large cell lymphoma (pcALCL) and Sezary syndrome (SS) accounting together for approximately 80% of CTCL. In the United States, the incidence has been increasing since 1970 until 2000, but stabilized in the current decade[3,4]. The incidence of CTCL is higher in African American population, especially males, and studies have shown their survival rate is lower when

compared to Caucasians[5,6,7]. The fascinating aspect about this cancer is that the proliferation of T cells takes place only in the skin. As the malignancy progresses in a subset of patients, the disease can spread to lymph nodes and other organs. There is no known cure for CTCL, though some patients have long-term remission with treatment and many more live symptom-free for many years.

BACKGROUND

Although majority of the rise in CTCL incidence can be attributed to increased detection and documentation of these cancers, this does not completely describe the population trends. Concomitant rise in CTCL cases has been noticed with industrial expansion, occurrence in non-blood-related relatives, and documented associations of occupational chemicals with other hematological malignancies. All these factors are pointing towards environmental exposures causing this disease. Clustering has been reported in a variety of studies across the globe including Sweden, Canada, Pittsburgh, Pennsylvania and Texas[7,8]. The study in Texas showed that three neighborhoods in metropolitan Houston had CTCL rates between five and twenty times the projected population rate[8]. Many such studies proposed environmental and occupational exposures as inciting factors, but there was limited analysis conducted on toxic exposures.

Benzene and Trichloroethylene

Benzene is a natural constituent of crude oil and is one of the elementary petrochemicals. It is used primarily as a precursor to the manufacture of chemicals with more complex structure, such as ethylbenzene and cumene, of which billions of kilograms are produced annually. Benzene is a colorless and highly flammable liquid with a sweet smell, and is responsible for the aroma around petrol (gasoline) stations. It has been limited to less than 1% in gasoline because it is a known human carcinogen. The major sources of benzene exposure are tobacco smoke, automobile service stations, exhaust from motor vehicles, and industrial emissions; however, ingestion and dermal absorption of benzene can also occur through contact with contaminated water. The effects of acute exposure to high concentrations of benzene (neurological, dermal, respiratory, gastrointestinal) can be evident immediately after exposure. Mild effects include headache, lightheadedness, dizziness, confusion, nausea, impaired gait, and blurred vision[9]. More severe effects include tremors, respiratory depression, confusion, loss of consciousness, coma, and death.

Trichloroethylene is a halocarbon commonly used as an industrial solvent. It is a clear non-flammable liquid with a sweet smell. Trichloroethylene is an effective solvent for a variety of organic materials. TCE is a solvent that is routinely used in "degreasing" and is used omni presently in dry cleaning products, and to a lesser degree in consumer products. It is also used for making other chemicals, especially the refrigerant, HFC-134a. Exposure to very high concentrations of trichloroethylene can cause dizziness headaches, sleepiness, incoordination, confusion, nausea, unconsciousness, and even death. Trichloroethylene has been found in at least 1,051 of the 1,854 National Priorities List sites identified by the Environmental Protection Agency (EPA)[10].

Carcinogenicity of Benzene and Trichloroethylene

Many occupational chemicals have carcinogenic properties, but benzene and trichloroethylene (TCE) are two regulated carcinogens with well-established links to hematological cancers, specifically NHL. The Department of Health and Human Services has determined benzene and trichloroethylene as known human carcinogens. The International Agency for Research on Cancer (IARC) has also classified benzene and trichloroethylene as carcinogenic to humans. The EPA has characterized trichloroethylene as carcinogenic to humans by all routes of exposure. Hematologic neoplasms such as acute myelogenous leukemia have been documented to occur with chronic exposures as low as 10 ppm benzene. Prior studies performed comprehensive analysis of the carcinogenic properties of benzene, and it was found that air levels as small as 1 ppm were toxic. There is strong evidence that trichloroethylene can cause kidney cancer in people and some evidence for trichloroethylene-induced liver cancer and malignant lymphoma.

We performed spatial analysis of CTCL cases in the states of California, Connecticut, Georgia, Hawaii, Iowa, Kentucky, Louisiana, New Jersey, New Mexico, Utah using geocoded incidences of CTCL from the SEER program. Only states with all counties reporting to SEER were included. Poisson regression analysis was performed on the demographic data extracted the from the U.S.

Census Bureau. This included all the states except Hawaii, which lacked information of SES data. In the regression model, CTCL incidence was correlated to ambient benzene and TCE levels, obtained through the EPA's National Air Toxics Assessment (NATA) database in each county from 1996, 2000 and 2002. Further an in-state clustering analysis and in-state Poisson regression analysis were performed on the states which had more than 30 counties.

MATERIALS AND METHODS

Patient Demographics and CTCL Data

We used three sets of secondary data in our analyses: National-Scale Air Toxics Assessment data of benzene and trichloroethylene exposure and concentration levels[11] by county from the years 1996, 1999 and 2002 for all the states under the study, Incidence of CTCL from the years 2000-2016 from the SEER cancer registry[12] for the states under the study, and population and demographic data from the United States Census from the year 2010[13].

We included patients over the age 15 with a new CTCL diagnosis for the states under review from the SEER cancer registry between 2000 and 2016. ICD-O-3 histology codes used for data extraction are: 9700 (mycosis fungoides (MF)), 9701 (Sezary syndrome (SS)), 9702, (mature T-cell lymphoma not otherwise specified (NOS)), 9705 (angioimmunoblastic T-cell lymphoma), 9708 (subcutaneous panniculitis-like T-cell lymphoma), 9709 (CTCL NOS), 9714 (Anaplastic large cell lymphoma (ALCL), CD30+), 9718 (primary cutaneous anaplastic large cell lymphoma), 9719 (extra nodal natural killer (NK)/T-cell lymphoma, nasal type (ENKL)), 9827 (adult T-cell leukemia/lymphoma), and we selected for "Primary site of skin", coded 44.0 through 44.9 as documented in prior studies[3].

Demographic and disease characteristics of patients were gathered including age, sex, race, year of diagnosis, histology, primary cancer site, census tract and county. A total of 12,556 cases were identified in the SEER cancer registry during this time period. Among these, we identified a total of 11,048 cases during this time period belonging to the states under review.

Analysis was performed on the data at the county level due to the low incidence of CTCL in

majority of the census tracts.

Standardized incidence ratios (SIRs) have been measured for all counties in the states under review to assess the risk of CTCL. The SIRs were determined by dividing the number of CTCL cases reported within each county between 2000 and 2016 by the number of CTCL cases expected between 2000 and 2016. National incidence rates for CTCL for each race-sex-age group have been multiplied by the number of individuals in each subgroup respectively to estimate the expected number of cases per year for all the states under review. The national incidence rates were obtained from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program, by age, sex and race subgroups {SEER*Stat version 8.3.6}. The number of individuals in each subgroup were obtained from the U.S. Census Bureau population data by county, age, race and sex subgroups. To estimate the expected number of cases from the years 2000-2016, we multiplied the expected cases in each county by 17 (total number of years of observed data). The subgroups included in the study were of total 20 combinations: age (15-59, ≥60), sex (male, female) and race groups (White/Caucasian, African American, Hawaiian and Pacific Islander, American Indian & Alaskan Native, Asian).

The calculation of SIR for each county i of the states under review was measured using the formula:

$$\text{SIR}_i = \frac{\text{Observed number of cases from 2000-2016 in county } i}{\text{Expected number of cases from 2000-2016 in county } i}$$

where the expected cases per year in county i =

$$\sum_{j=1}^{n} \text{Population in subgroup } j \text{ in county } i \text{ x National CTCL rate for subgroup } j,$$

and n = number of subgroups.

An SIR > 1 suggested an increased risk of CTCL than would have been predicted based on that county's demographic structure.

Benzene and TCE Data

We collected the concentration and exposure levels of benzene and TCE from NATA (National Air Toxics Assessment) database released by the EPA (Environmental Protection Agency). NATA presents an average estimate of chemical releases for each census tract and county based on many various models of dispersion and from the Toxics Release Inventory (TRI) database, which monitors toxic sites.

The Assessment System for Population Exposure Nationwide (ASPEN) is a model for each census tract and county depicting the collection of data from state, federal and local databases on industries, demographics, meteorology and housing. It consists of complex data to clearly define ambient toxic levels such as rate of release, wind speeds, breakdown, release location and settling of release. Hazardous Air Pollutant Exposure Model (HAPEM), a second model which provides an estimate of the estimate the exposure concentration of air toxics through ambient air concentration data, population data, indoor/outdoor microenvironment concentration relationships and human activity pattern data. An average of ASPEN and HAPEM data from the years 1996, 1999 and 2002 for each county of the states under review was used for our primary analysis. Both ASPEN and HAPEM estimate the levels in microgram per cubic meter (μg/m3). These averages of exposure and concentration levels for each county were the basis for our analyses discussed below.

Socio Economic Status Data

We have included socioeconomic data from US Census Bureau as covariates for the adjusted risk ratios for the following socioeconomic variables for all the states under review: Household Median Income, Percentage of High school graduates in the county population, Percentage of the population below the poverty level and percentage of the population uninsured. We obtained this data from the 2010 ACS (American Community Survey) 5-year estimates for each of the

above variables for all the counties of the states under review except Hawaii as the census data did not have complete information on SES data of Hawaii.

Spatial Analysis

We used two programs to perform county-specific benzene / TCE and SIR spatial analysis: ArcGIS 10.7.1 and GeoDa 1.14.0. We uploaded county shapefiles obtained from the US Census Bureau's 2000 TIGER/Line files via ArcGIS. These county shapefiles for all the states under review were overlaid with SIRs, benzene and TCE exposure and concentration levels, and expected and observed cases to obtain maps of these variables by county. We then exported the shapefiles with these maps to GeoDa to examine the spatial relationship and to assess for the clustering of the toxins and the SIRs. On GeoDa, we conducted global and local spatial analyses to classify the distribution of SIRs and toxins as either clustered, scattered, or random. The global analysis generated a Moran's *I* statistic, as well as a z-score and pseudo p-value to check for statistical significance; these statistics are based on 999 Monte Carlo simulations. Our null hypothesis assumed that the features are randomly distributed. The Moran's *I* statistic measures the spatial correlation and establishes if the data are clustered (positive) or dispersed (negative). Subsequently, a local Moran's *I* is used to construct cluster maps displaying "hot spots" reflecting high SIR or benzene / TCE clusters or "cold spots" with lower-value clusters.

Statistical Analysis

We have merged the data of all states under review (except Hawaii, as Hawaii did not have all the variables under consideration in the US census data) to estimate the associations between observed CTCL cases and the concentration and exposure data of benzene and TCE (ASPEN and HAPEM) using models of Poisson regression. Univariate and multivariable models were explored, where SES variables were controlled for as model covariates in the multivariable models. In all Poisson regression models, the log of expected cases served as an offset. The

analysis was conducted across all states, as well as by state for those with more than 30 counties. State-level models excluded the uninsured variable to level of missingness. Model hypotheses were tested and confirmed. For the analyzes, the program SAS 9.4 (SAS Institute Inc., Cary, NC) was used, and statistical significance at a level of 0.05 was assessed.

RESULTS

CTCL and SES Data

A total of 591 counties were present in the merged data of all the states except Hawaii. The mean of observed cases was 18.43 and the mean of expected cases was 17.75. The mean SIR was 0.77. Median household income reported by the census data was available for 590 counties out of 591 under consideration, for which the mean was $44,093.64. The data on percentage of Uninsured population was available for 343 counties from the census data, for which the mean was 15.83%. The data on the percentage high school graduates was available for 590 counties, and the mean was 83.36%. The data on the percentage of population below the poverty line was available for all the 591 counties, for which the mean was 12.82% (Table 1).

A total of 11,038 cases were identified in the states under review for this study; of this 4673 were female and 6365 were male. The mean age at diagnosis was 59.94 years. Of these, 8786 were White, 1530 were African Americans, 679 were Asian or Pacific Islanders and 43 were American Indian/Alaska Natives. Of the 11,038 cases, 5970 (54.09%) were mycosis fungoides, 3507 (31.8%) were CTCL NOS, 1058 (9.59%) were primary cutaneous anaplastic large cell lymphoma, and 503 (4.56%) were another subtype (Table 2).

Among all the states, the states with the highest number of cases were seen in California with 4918 cases followed by New Jersey with 1657 cases and Georgia with 1305 cases. The following table (Table 3) depicts the number of cases in each state by race. Among all the states under review, 5095 (46.16%) were in age group 15-59 and 5943 (53.84%) were older than 60 years. Among the cases in the age group 15-59, 2314 were female and 2781 were male. Among the cases aged over 60 years, 2359 were female and 3584 were male. A distribution of cases by sex and age group for all the states under the review is attached (Table:4).

Benzene and TCE data

The mean concentration level of benzene across all counties (n=591) was 0.703 μg/m3 and the mean exposure level was 0.659 μg/m3. The mean TCE concentration level was 0.072 μg/m3 and the mean exposure level was 0.059 μg/m3. We have also measured the Pearson correlation coefficients (Prob >|r| under $H_{0:}$ Rho=0) between the levels of Benzene and TCE exposure and concentration across the years 1996, 1999 and 2002.

Spatial Analysis

Maps were constructed depicting the SIR, benzene, and TCE maps with categories defined using quantiles for all the states under review. The maps depict the areas of high SIR, benzene and TCE concentration in specific counties in majority of the states under review. The Moran's I value and p-value for SIR for the states in which spatial analysis was performed is shown (Table:8). These Moran's I statistics were positive, statistically significant for states California, Georgia, Iowa, and Louisiana indicating that there is clustering in all those states. Further, we identified hot spots and cold spots in all those states, represented in Figures:11-17.

Maps were constructed for these states with "high-high" SIR indicating areas of high SIR surrounded by other areas of high SIR. The areas showing "low-low" SIR indicating areas of low SIR surrounded by other areas of low SIR are also depicted.

Correlations

Correlations of benzene exposures between 1996, 1999, and 2002 ranged from 0.82-0.91 (all $p<0.001$), while correlations of TCE exposures between 1996, 1999, and 2002 ranged from 0.28-0.58 (all $p<0.001$). Similar correlations were observed with the concentration data. This demonstrated consistency in toxic exposure estimates within counties over time, particularly for benzene. Thus, estimates were averaged for further analysis.

Poisson Regression Analysis

Poisson regression analysis was performed on the combined dataset of all the states under review the dependent variable was observed cases and either benzene in μg/m^3 or TCE in μg/m^3 (ASPEN and HAPEM), as the independent variables. The univariate β_1 estimate and p-value for CTCL risk and benzene ASPEN concentration (μg/m^3) and observed cases and TCE ASPEN concentration (μg/m^3) were 0.23 ($p<0.001$) and 1.03 ($p<0.001$) respectively. In addition, the β_1 estimate and p-value for observed cases and benzene HAPEM exposure (μg/m^3) and TCE HAPEM exposure (μg/m^3) were 0.22 ($p<0.001$) and 1.42 ($p<0.001$) respectively (Table 5). In addition, the β_1 estimate and p-value for observed cases and benzene HAPEM exposure (μg/m^3) and TCE HAPEM exposure (μg/m^3) were 0.34 ($p<0.001$) and 3.7 ($p<0.001$) respectively. The relative risk is equal to $\exp(\beta_1)$. The positive, significant values of β_1 indicate that for a 1 unit increase in μg/m^3 in benzene/TCE exposure and concentration, there is an increased in the risk of CTCL by the corresponding value of $1-\exp(\beta_1)$. Estimates from HAPEM exposure were more strongly associated with observed cases for TCE and estimates from ASPEN concentration were more strongly associated with observed cases for benzene.

Another Poisson regression model was performed on the combined dataset of all the states under review. The dependent variable was observed cases and either benzene in μg/m^3 or TCE in μg/m^3 (ASPEN and HAPEM), and the four SES variables: Percentage High School Education, Median Family Income, Percentage Uninsured and Percentage below the poverty level as the covariates. The β_1 estimate and p-value for observed cases and benzene ASPEN concentration (μg/m^3) and TCE ASPEN concentration (μg/m^3) were 0.21 ($p<0.001$) and 0.22 ($p=0.1401$) respectively. The β_1 estimate and p-value for observed cases and benzene HAPEM exposure (μg/m^3) and TCE HAPEM exposure (μg/m^3) were 0.19 ($p<0.001$) and 0.32 ($p=0.1093$) respectively. Our analysis showed that Benzene was found to be an independent predictor of CTCL risk, but not TCE.

Further, Poisson regression models were also run in the states which had more than 30 counties. The dependent variable was the observed number of cases, the independent variables being either benzene in $\mu g/m^3$ or TCE in $\mu g/m^3$ (ASPEN and HAPEM), and the four SES variables: Percentage High School Education, Median Family Income, and Percentage below the poverty level as the independent variables. As the data on percentage uninsured was missing for a significant number of counties in each state, the regression models were run excluding the variable: Percentage Uninsured. The results for those regression models for all the states under the review with more than 30 counties is shown (Table:9).

DISCUSSION

CTCL is a rare cancer, yet with high mortality and morbidity with median survival being less than 5 years. Even though geographic clustering has been described in several studies for CTCL, very limited environmental risk factors have not been widely focused. We have tried to establish the relationship if any between the benzene and TCE levels and CTCL incidence through this study. We found increased exposure to the environmental toxins, benzene and TCE, was associated with increased incidence of CTCL in majority of the states under review, pointing towards an environmental cause for CTCL. Geographic clustering has been identified in several regions nationally and internationally, particularly in urban and industrial regions[14-19]. Our analysis further adds strength to these previous studies.

We identified 11,038 new cases of CTCL diagnosed between 2000 and 2016 from the SEER registry. We have established the standardized incidence ratios for each county, considering race, sex, and age. In addition, cluster analysis identified several hot-spots and cold-spots in the states under review. Overall, we demonstrated geographic clustering of CTCL is correlated with exposure to the environmental toxins TCE and benzene in most of the states under review, suggesting a possible etiologic role of pollution. Identifying specific etiologic triggers for CTCL has significant clinical implications, as it may suggest a need for increased skin protection in certain high-risk exposure environments to reduce the risk of CTCL in the future or simply increased awareness of the risks of exposure. Our findings should be validated in larger national and international studies that include efforts to identify the remaining triggers such as infections, radioactivity, and other chemical agents.

Strengths and Weaknesses

The major strength of this study is using the publicly available data from NATA, SEER and the U.S. Census Bureau to conduct a descriptive study. Additionally, this is the first study of its kind to our knowledge to use a combination of spatial cluster statistics, statistical modeling, and visual

representation of data through GIS to analyze CTCL clustering across several US states. We have also used indirect standardization to eliminate the effects of age, race and sex on CTCL incidence.

One of the major limitations of this study is using the data at the county level. As the incidence of CTCL is very low with usually 1-10 cases per census tract, it has not been possible to analyze the data at the census tract level. Thus, the exposure and concentration levels of benzene and TCE may not hold true to the individual level. Additionally, all cases of CTCL may not have been captured if CTCL was misclassified as peripheral T-cell lymphoma or T-cell lymphomas not-otherwise specified, and a skin primary was not included in the diagnosis. Also, SEER registry does not include data from all the states of the US. Hence our study is limited to the data on the states available in the registry.

The use of U.S. Census data as denominators for incidence rates may have also introduced bias to this study. Race was recoded into five major categories, but in the 2010 census, respondents could have chosen more than one racial category. This may have affected population estimates and thus could have affected calculations for CTCL incidence standardized by race, age, and sex. Also, any significant changes in the population of the states under review from 2000 – 2016 might have affected the CTCL incidence. The missing data in the US Census pertaining to Percentage Uninsured population has led some of the regression models to be limited to lesser number of counties. Also, any unmeasured confounders could have affected the Poisson regression model. Residuals may also be spatially auto correlated. We did not address this in our analyses.

Future Directions

The etiology and risk factors for CTCL are likely multifactorial. To further assess the risk of these and other toxins not included in this report, our results should be confirmed though a larger nationwide assessment of CTCL incidence which includes a broader list of regulated environmental toxins. It is crucial to further investigate the etiology of CTCL as the incidence continues to increase.

REFERENCES

1. Lessin SR . Alibert lymphoma: renaming mycosis fungoides . Arch Dermatol 2009 ; 145 : 209 – 210 .
2. Kirsch IR, Watanabe R, O'malley JT, Williamson DW, Scott LL, Elco CP, et al. TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. Sci Transl Med. (2015) 7:308ra158. doi: 10.1126/scitranslmed.aaa9122
3. Korgavkar K, Xiong M, Weinstock M. Changing incidence trends of cutaneous T-cell lymphoma. *JAMA dermatology*. 2013;149(11):1295-1299.
4. Criscione VD, Weinstock MA. Incidence of cutaneous T-cell lymphoma in the United States, 1973-2002. Arch Dermatol. 2007;143(7):854-859.
5. Wilson LD, Hinds GA, Yu JB. Age, race, sex, stage, and incidence of cutaneous lymphoma. *Clin Lymphoma Myeloma Leuk.* 2012;12(5):291-296.
6. Imam MH, Shenoy PJ, Flowers CR, Phillips A, Lechowicz MJ. Incidence and survival patterns of cutaneous T-cell lymphomas in the United States. *Leuk Lymphoma.* 2013;54(4):752-759.
7. Litvinov IV, Tetzlaff MT, Rahme E, et al. Demographic patterns of cutaneous T-cell lymphoma incidence in Texas based on two different cancer registries. *Cancer Med.* 2015;4(9):1440-1447.
8. Litvinov IV, Tetzlaff MT, Rahme E, Habel Y, Phil M, Risser DR, Gangar P, Jennings MA, Sc. M, Pehr K, Prieto VG, Sasseville D, and Duvic M. Identification of geographic clustering and regions spared by the Cutaneous T-Cell Lymphoma (CTCL) in Texas using two distinct cancer registries. Cancer, 2015. 121(12): 1993–2003.
9. Toxic Substances Portal - Benzene. (n.d.). Retrieved from https://www.atsdr.cdc.gov/mmg/mmg.asp?id=35&tid=14
10. Toxic Substances Portal - Trichloroethylene (TCE). (n.d.). Retrieved from https://www.atsdr.cdc.gov/toxfaqs/tf.asp?id=172&tid=30

11. (n.d.). Retrieved from https://archive.epa.gov/nata2002/web/html/tables.html

12. Surveillance Epidemiology, and End Results (SEER) Program, SEER*Stat Database: Incidence - SEER 21 Registries Research Data, Nov 2018 Sub (2000-2016) <Katrina/Rita Population Adjustment> https://seer.cancer.gov/data/seerstat/nov2017/

13. Decennial Census of Population and Housing. U.S. Census Bureau https://www.census.gov/programs-surveys/decennialcensus/data/datasets.2010.html .

14. Gip L, Nilsson E. [Clustering of mycosis fungoides in the County of Vasternorrland]. *Lakartidningen.* 1977;74(12):1174-1176.

15. Ghazawi FM, Netchiporouk E, Rahme E, et al. Distribution and Clustering of Cutaneous T-Cell Lymphoma (CTCL) Cases in Canada During 1992 to 2010. *Journal of cutaneous medicine and surgery.* 2018;22(2):154-165.

16. Ghazawi FM, Netchiporouk E, Rahme E, et al. Comprehensive analysis of cutaneous T-cell lymphoma (CTCL) incidence and mortality in Canada reveals changing trends and geographic clustering for this malignancy. *Cancer.* 2017;123(18):3550-3567.

17. Moreau JF, Buchanich JM, Geskin JZ, Akilov OE, Geskin LJ. Non-random geographic distribution of patients with cutaneous T-cell lymphoma in the Greater Pittsburgh Area. *Dermatol Online J.* 2014;20(7).

18. Litvinov IV, Tetzlaff MT, Rahme E, Habel Y, Phil M, Risser DR, Gangar P, Jennings MA, Sc. M, Pehr K, Prieto VG, Sasseville D, and Duvic M. Identification of geographic clustering and regions spared by the Cutaneous T-Cell Lymphoma (CTCL) in Texas using two distinct cancer registries. Cancer, 2015. 121(12): 1993–2003.

19. Switchenko JM, Bulka C, Ward K, et al. Resolving uncertainty in the spatial relationships between passive benzene exposure and risk of non-Hodgkin lymphoma. *Cancer epidemiology.* 2016;41:139-151.

20. TIGER/Line® Shapefiles and TIGER/Line® Files. U.S. Census Bureau https://www.census.gov/geo/maps-data/data/tiger-line.html.

FIGURES

Figure 1: Standardized incidence ratio (SIR) of CTCL for each county in California, aggregating cases from 2000-2016.

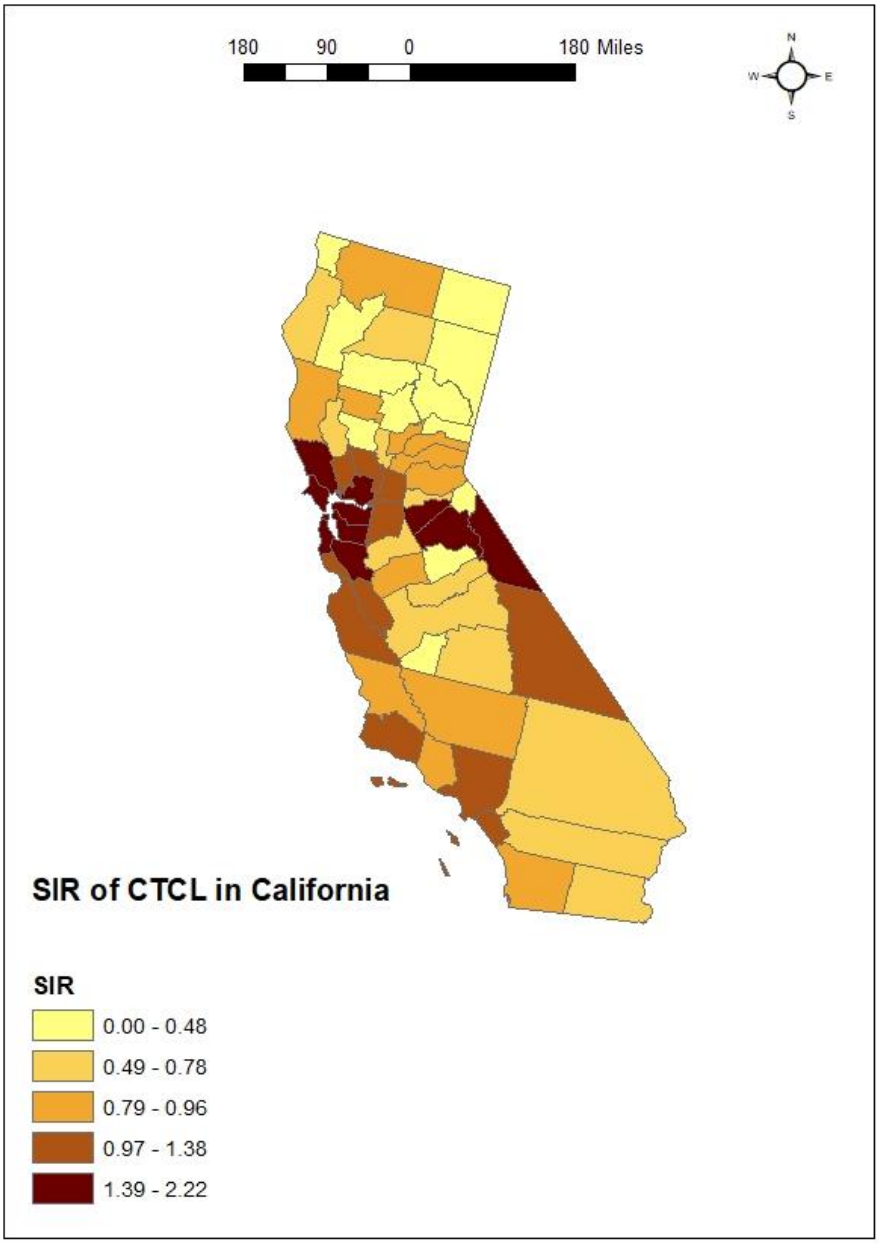


Figure 2: Standardized incidence ratio (SIR) of CTCL for each county in Connecticut, aggregating cases from 2000-2016.

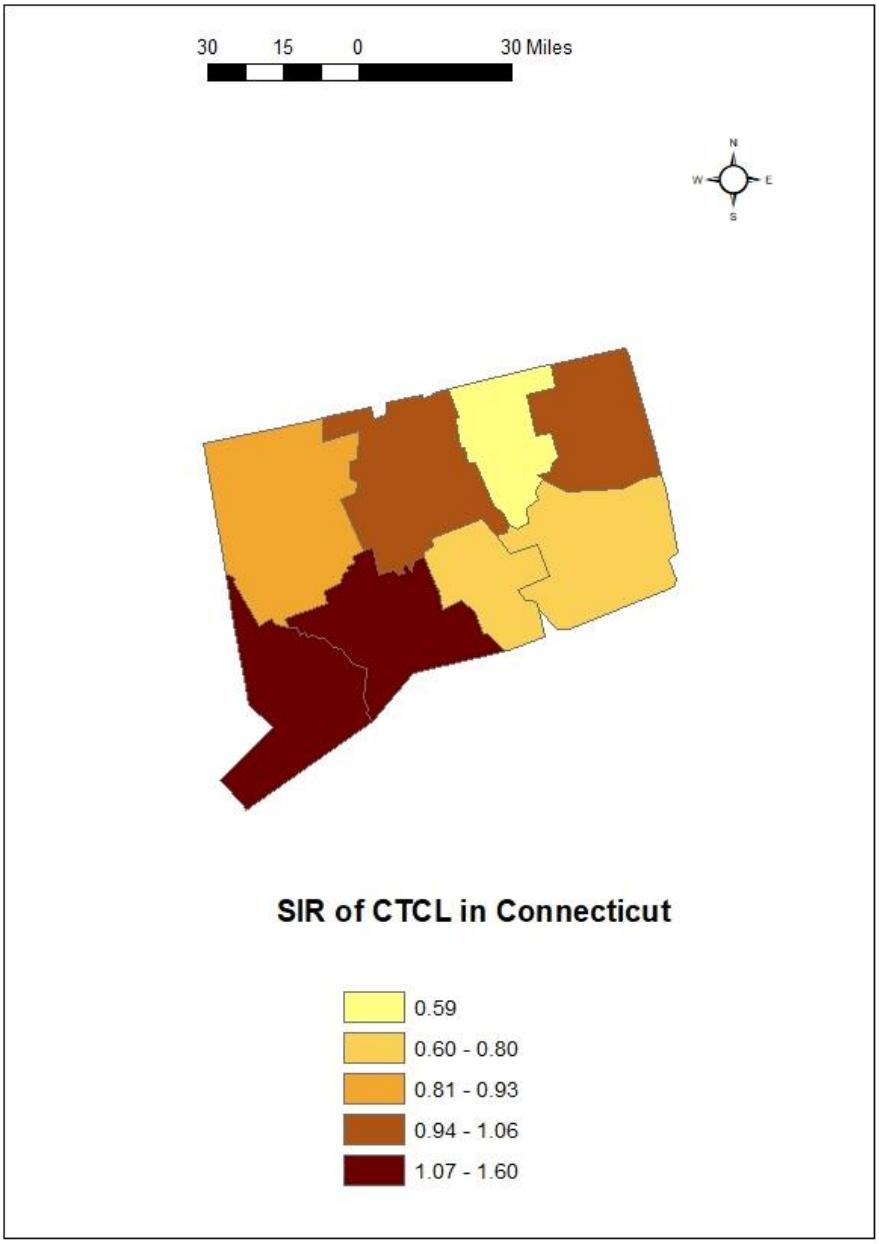


Figure 3: Standardized incidence ratio (SIR) of CTCL for each county in Georgia, aggregating cases from 2000-2016.

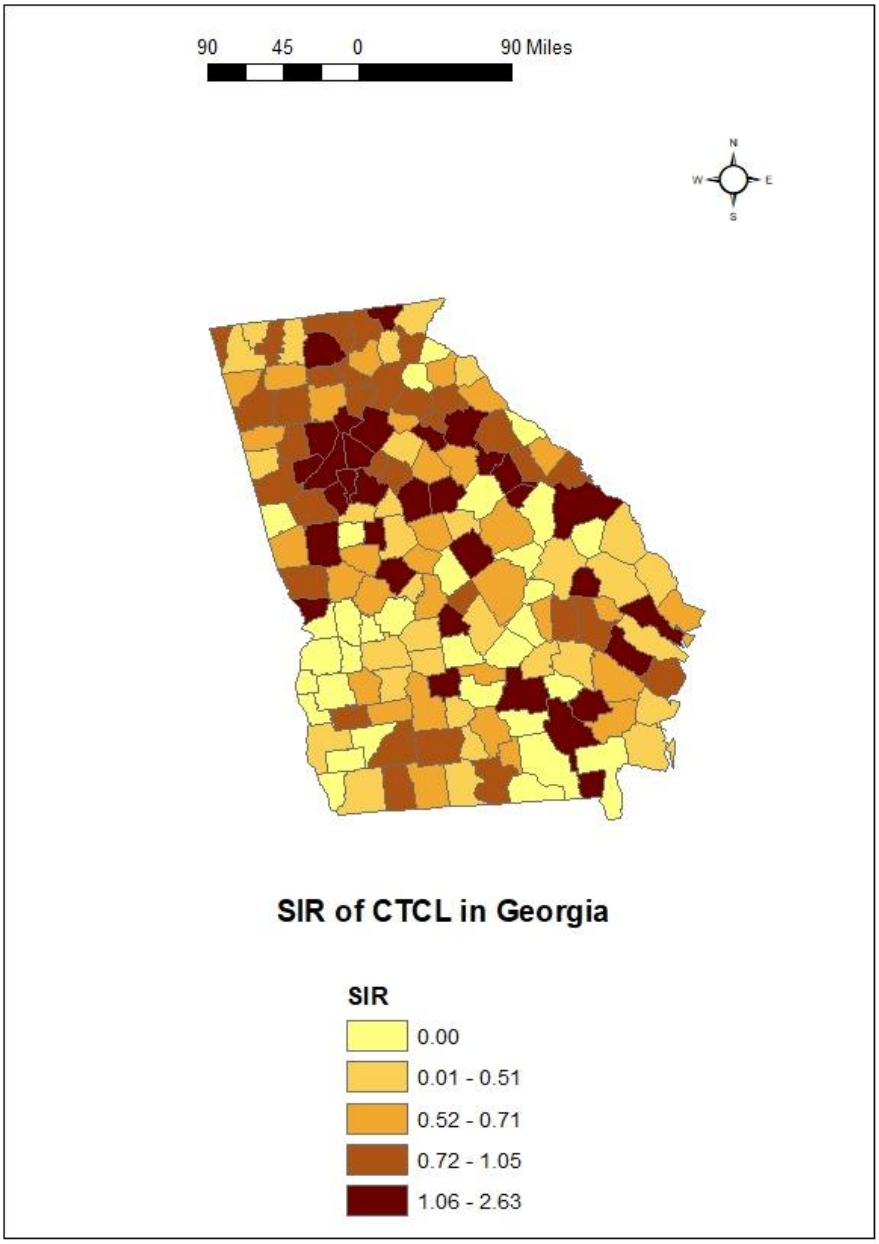


Figure 4: Standardized incidence ratio (SIR) of CTCL for each county in Hawaii, aggregating cases from 2000-2016.

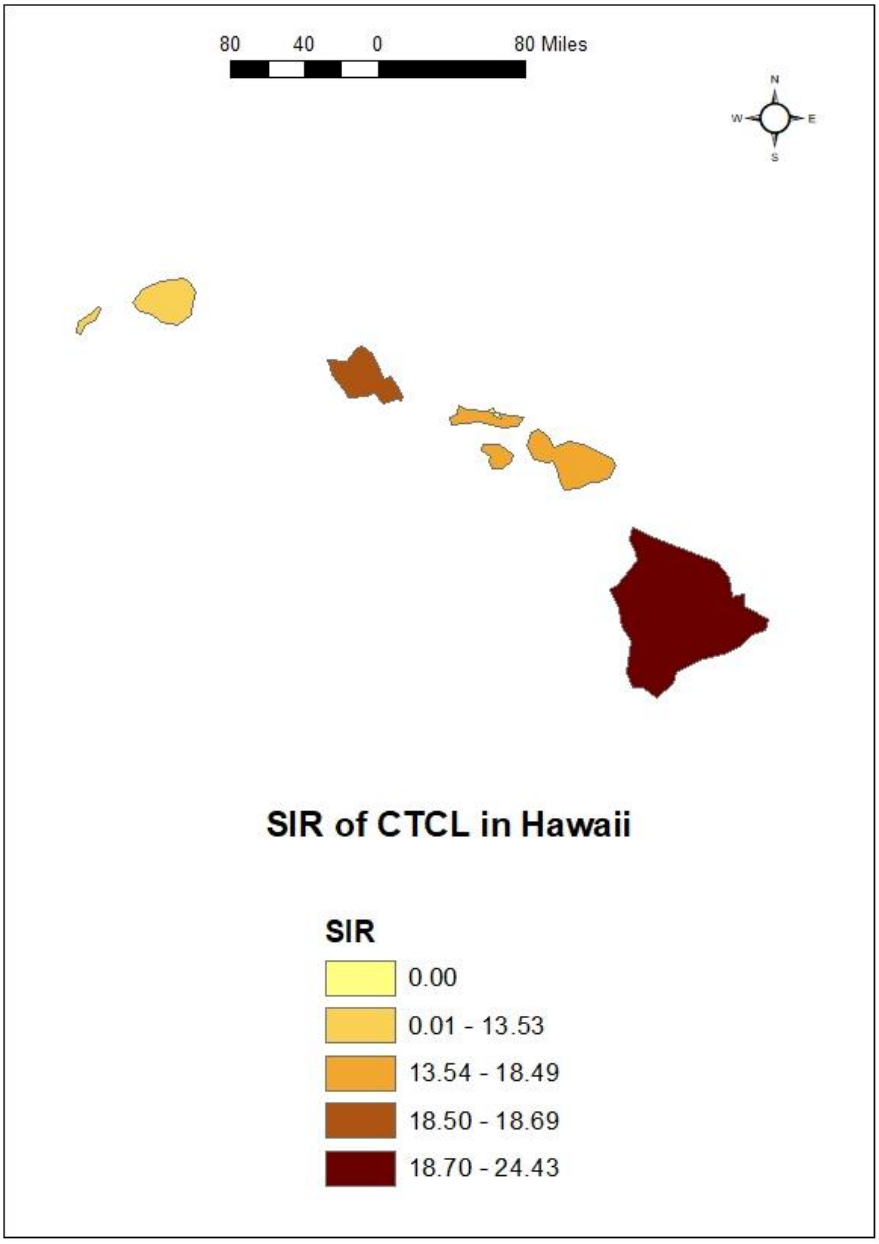


Figure 5: Standardized incidence ratio (SIR) of CTCL for each county in Iowa, aggregating cases from 2000-2016.

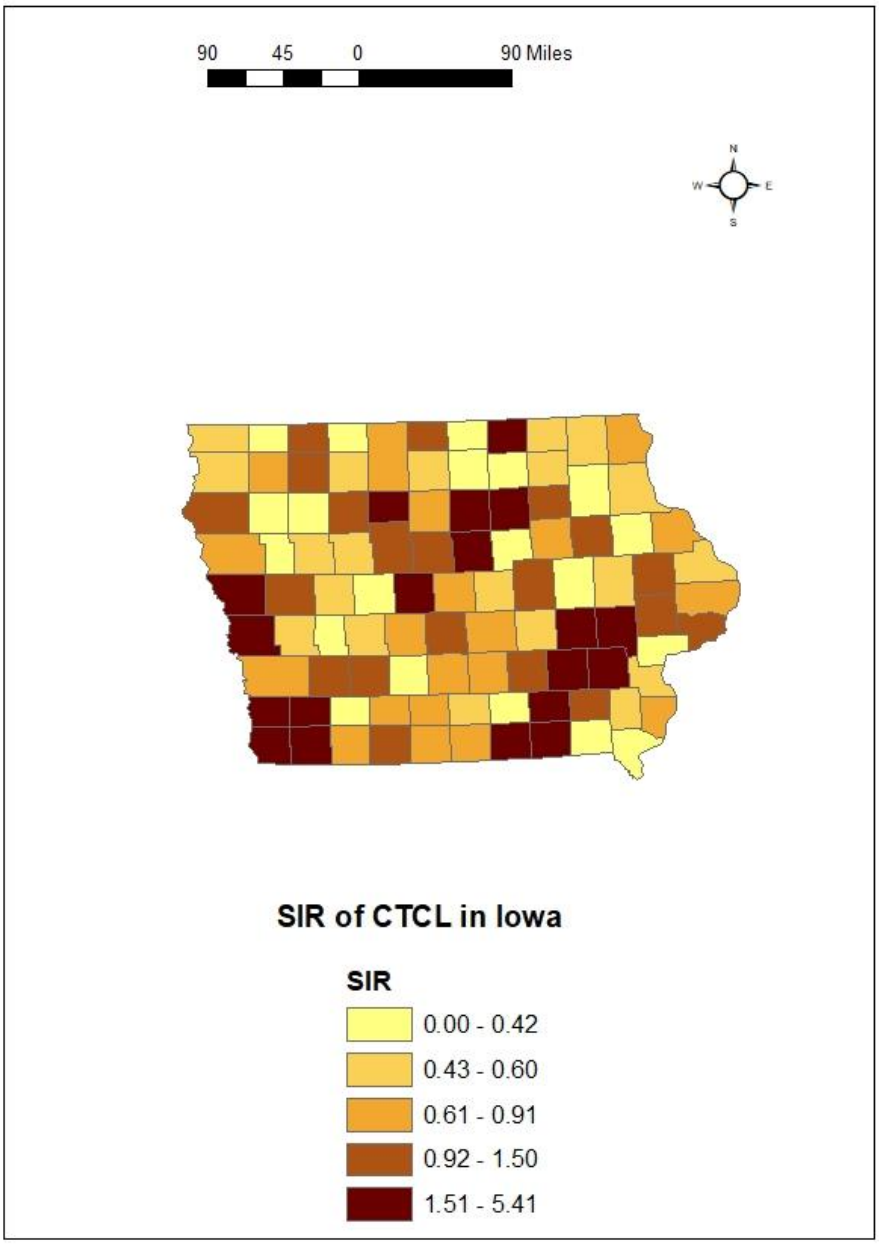


Figure 6: Standardized incidence ratio (SIR) of CTCL for each county in Kentucky, aggregating cases from 2000-2016.

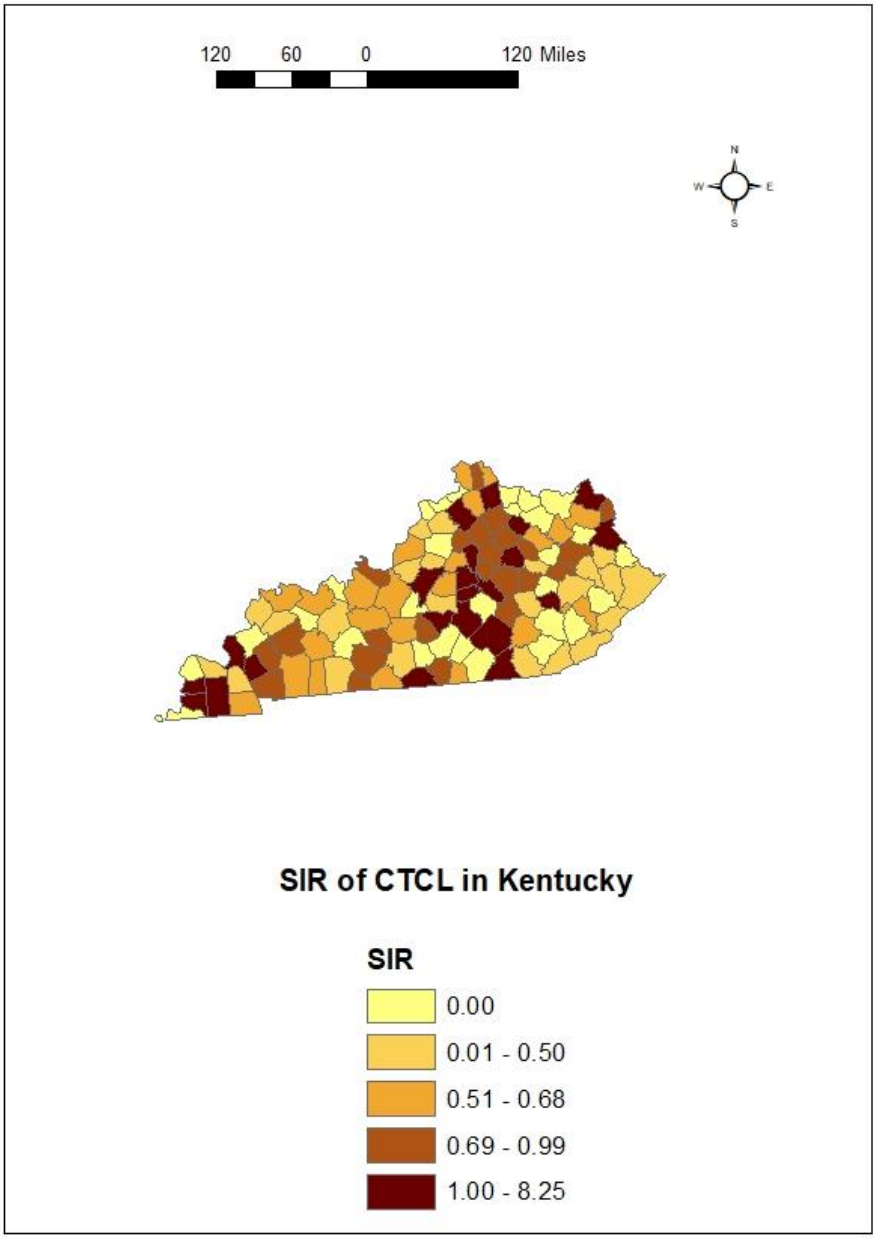


Figure 7: Standardized incidence ratio (SIR) of CTCL for each county in Louisiana, aggregating cases from 2000-2016.

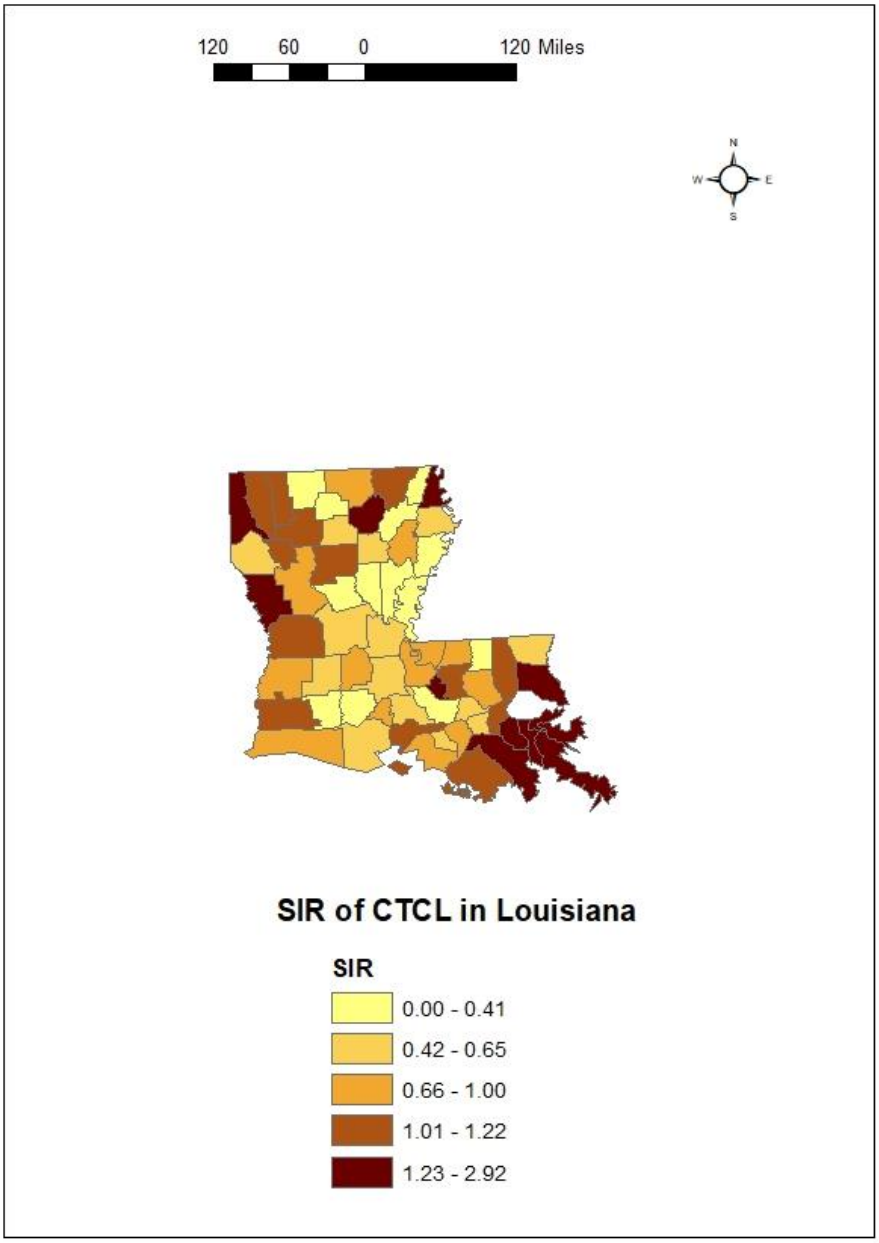


Figure 8: Standardized incidence ratio (SIR) of CTCL for each county in New Jersey, aggregating cases from 2000-2016.

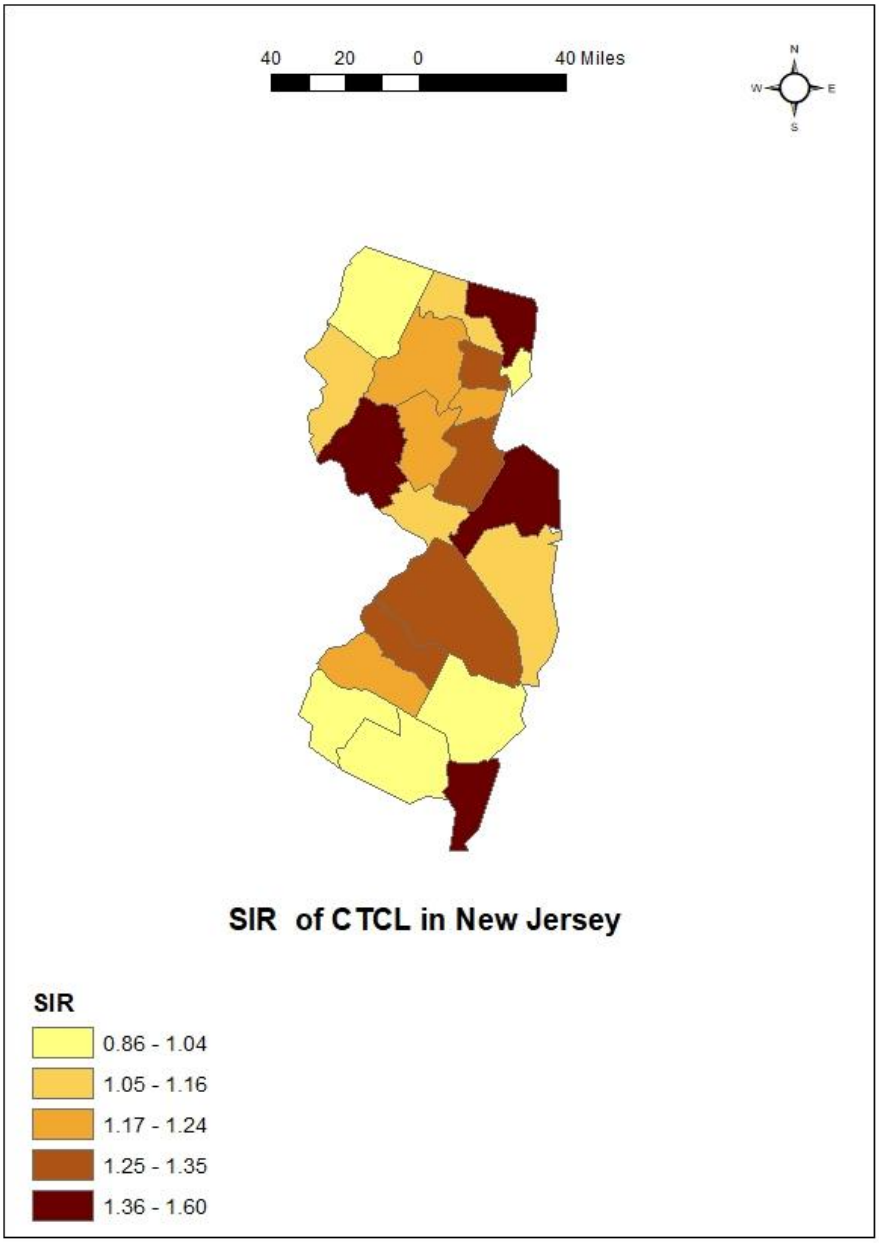


Figure 9: Standardized incidence ratio (SIR) of CTCL for each county in New Mexico, aggregating cases from 2000-2016.

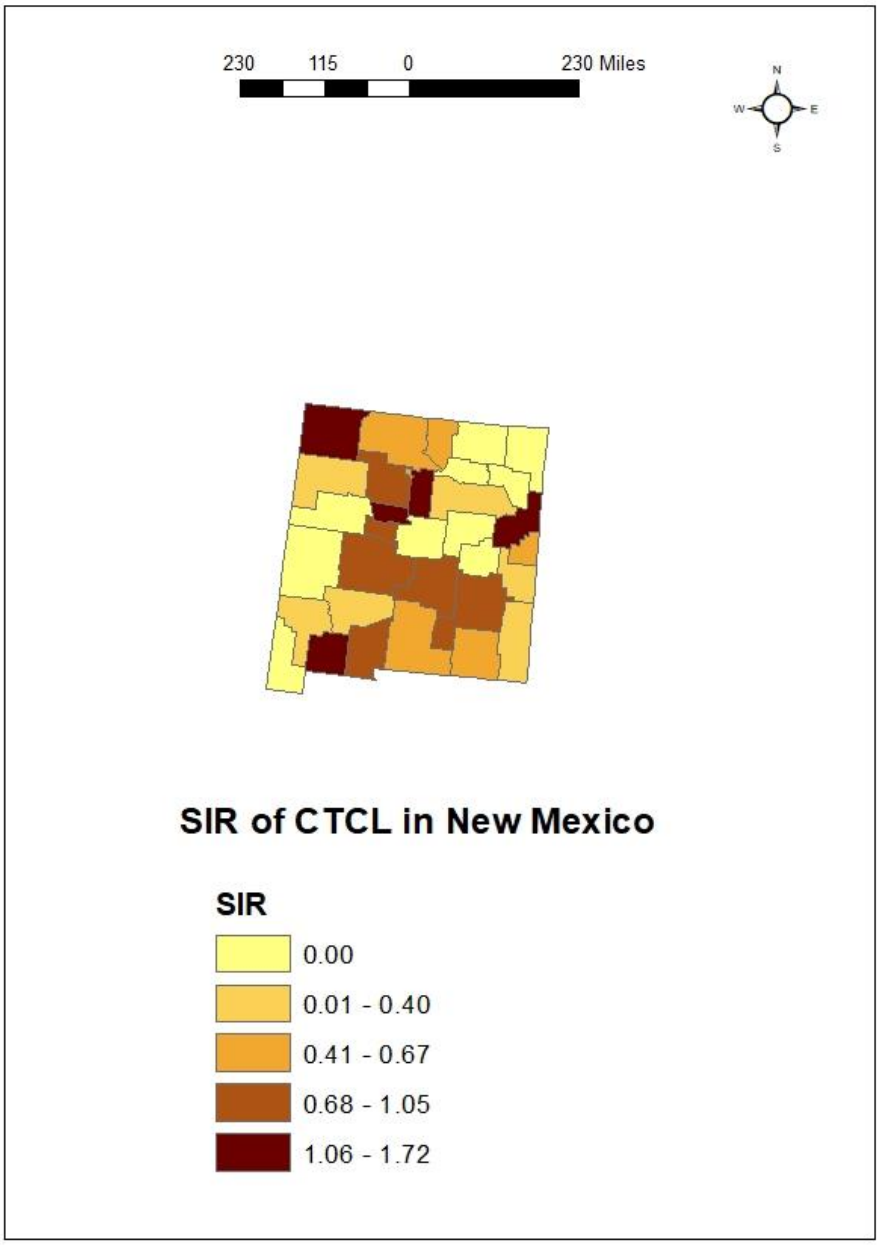


Figure 10: Standardized incidence ratio (SIR) of CTCL for each county in Utah, aggregating cases from 2000-2016.

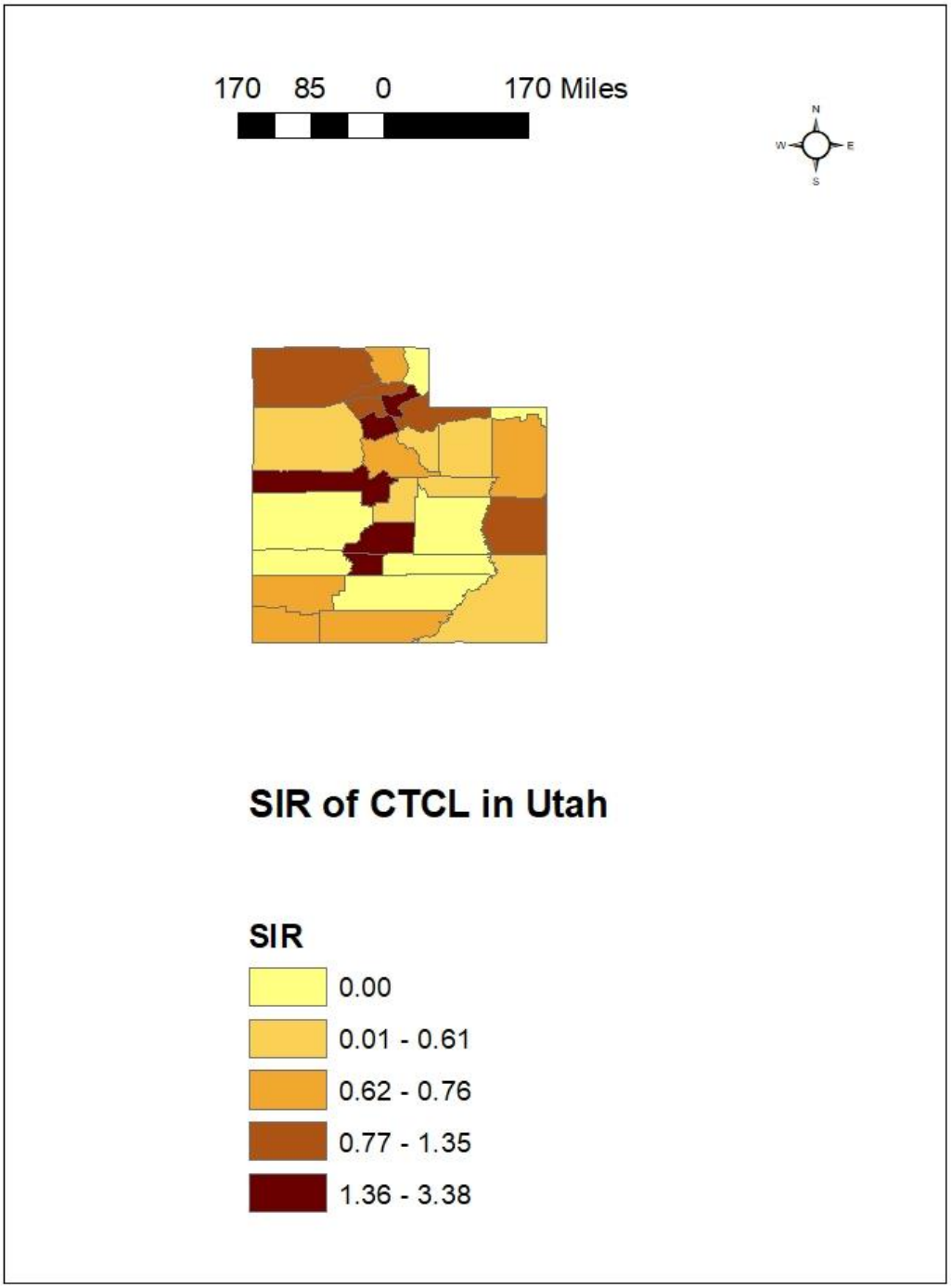


Figure 11: Cluster Map of SIR in California

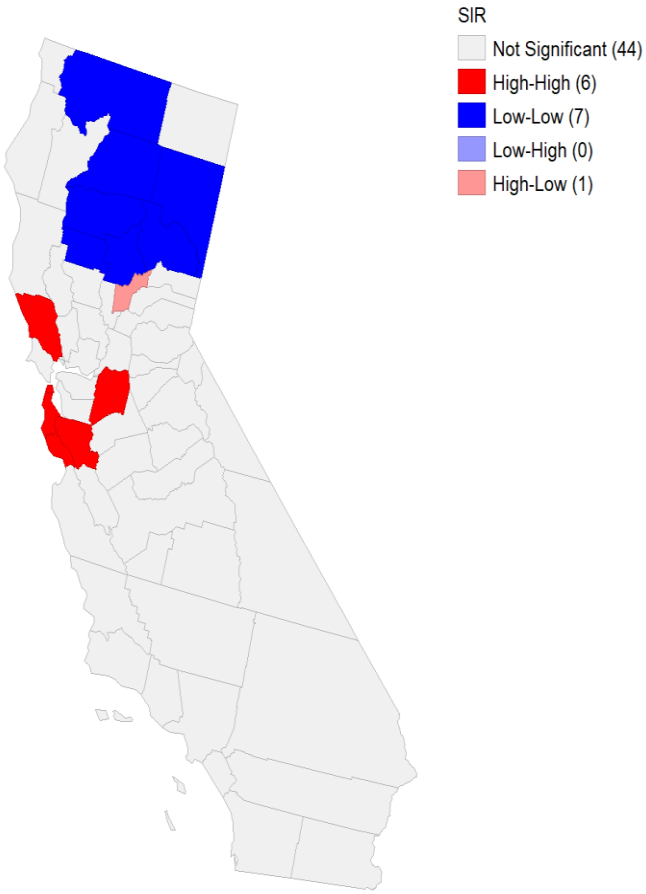


Figure 12: Cluster Map of SIR in Georgia

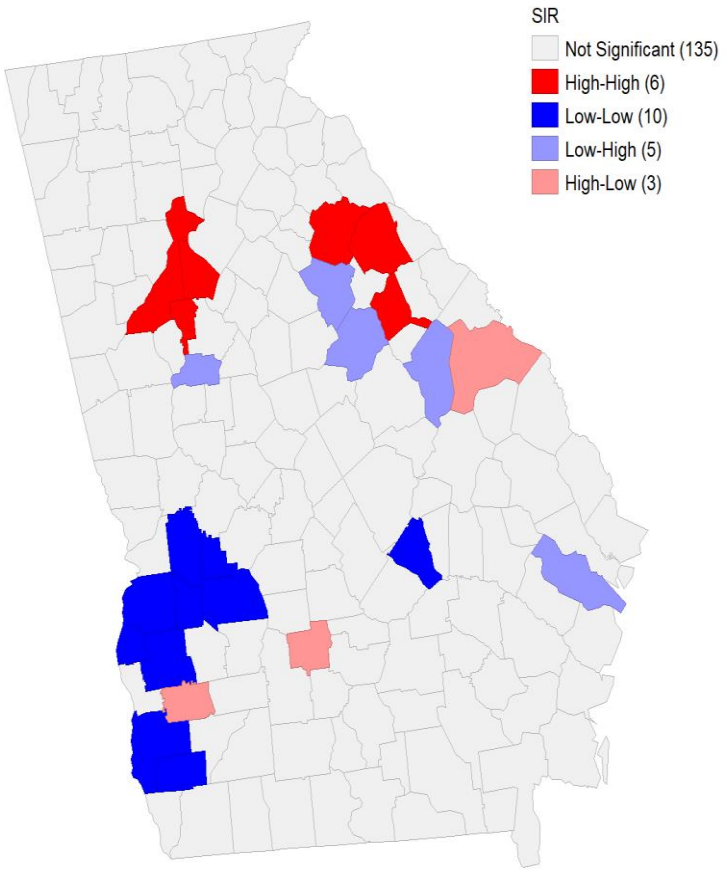


Figure 13: Cluster Map of SIR in Iowa

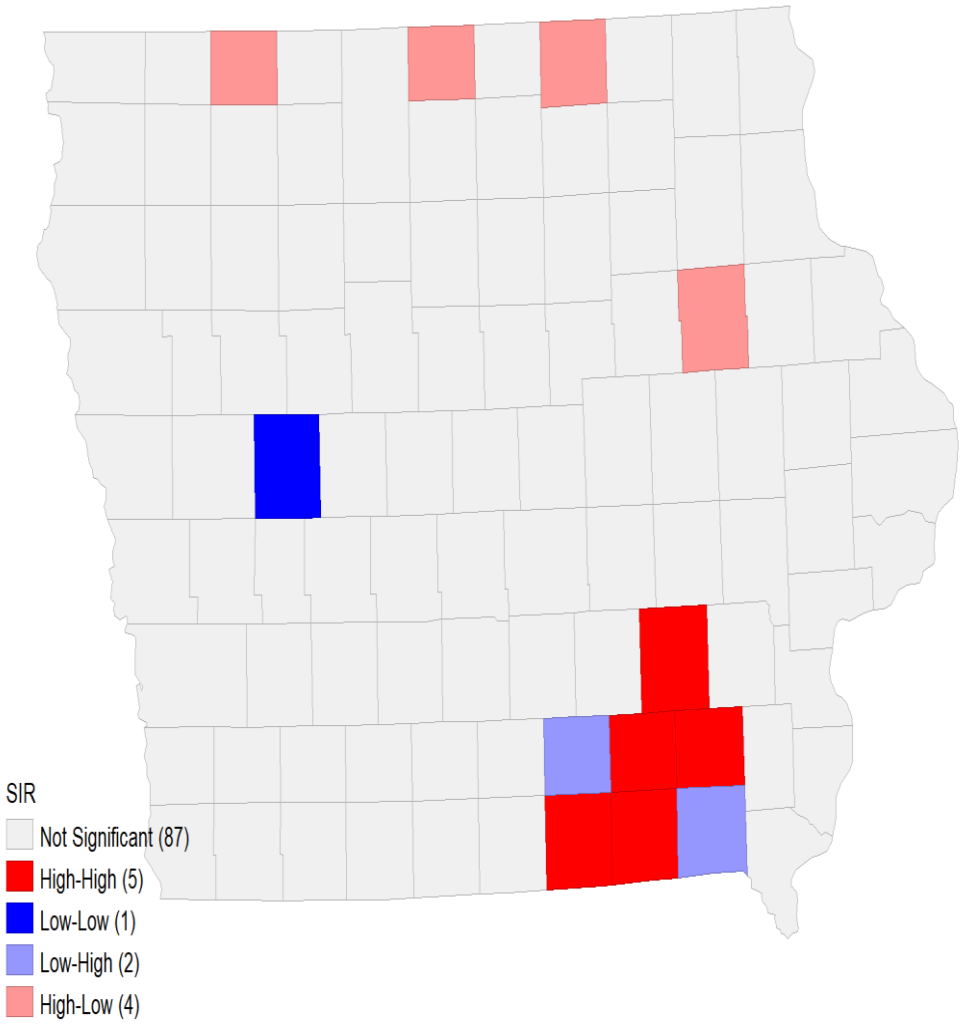


Figure 14: Cluster Map of SIR in Kentucky

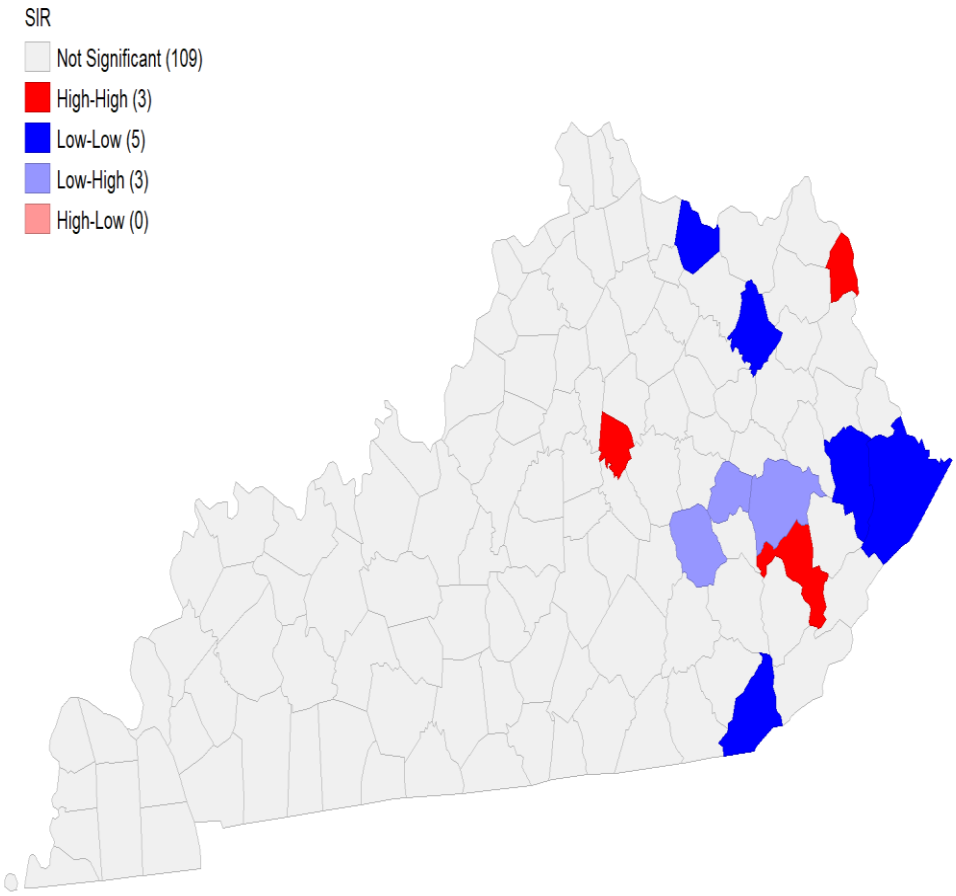


Figure 15: Cluster Map of SIR in Louisiana

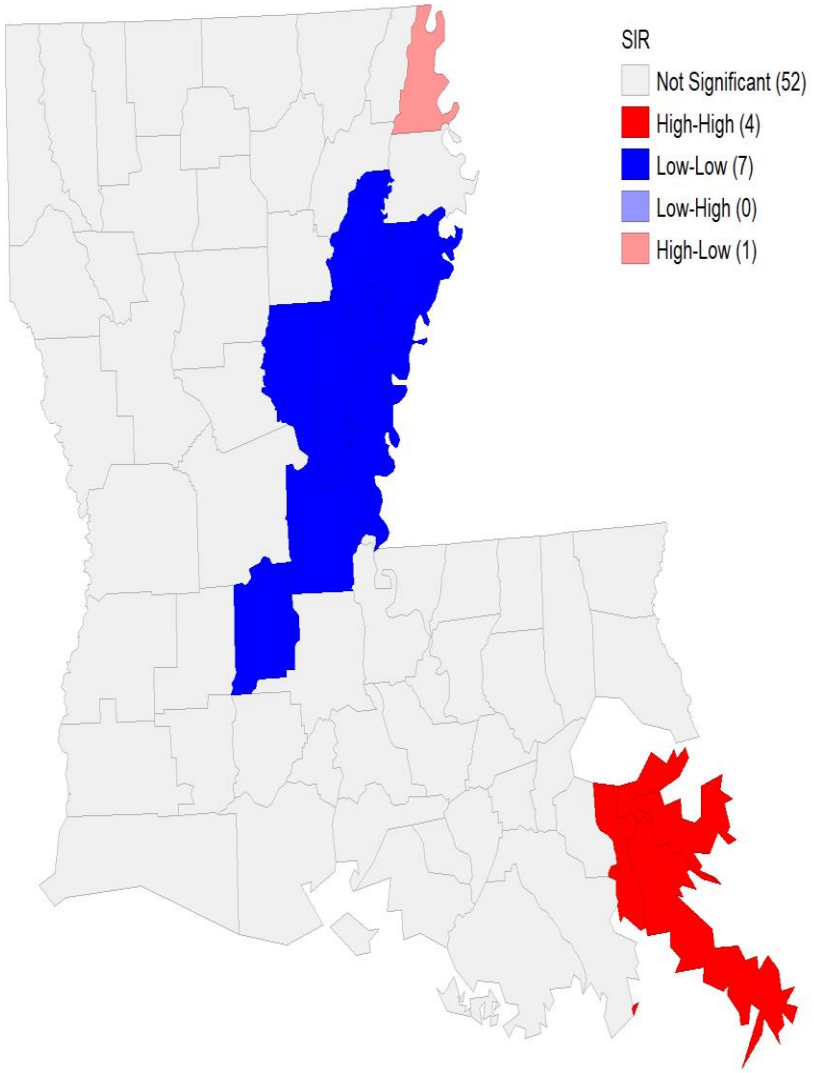


Figure 16: Cluster Map of SIR in New Jersey

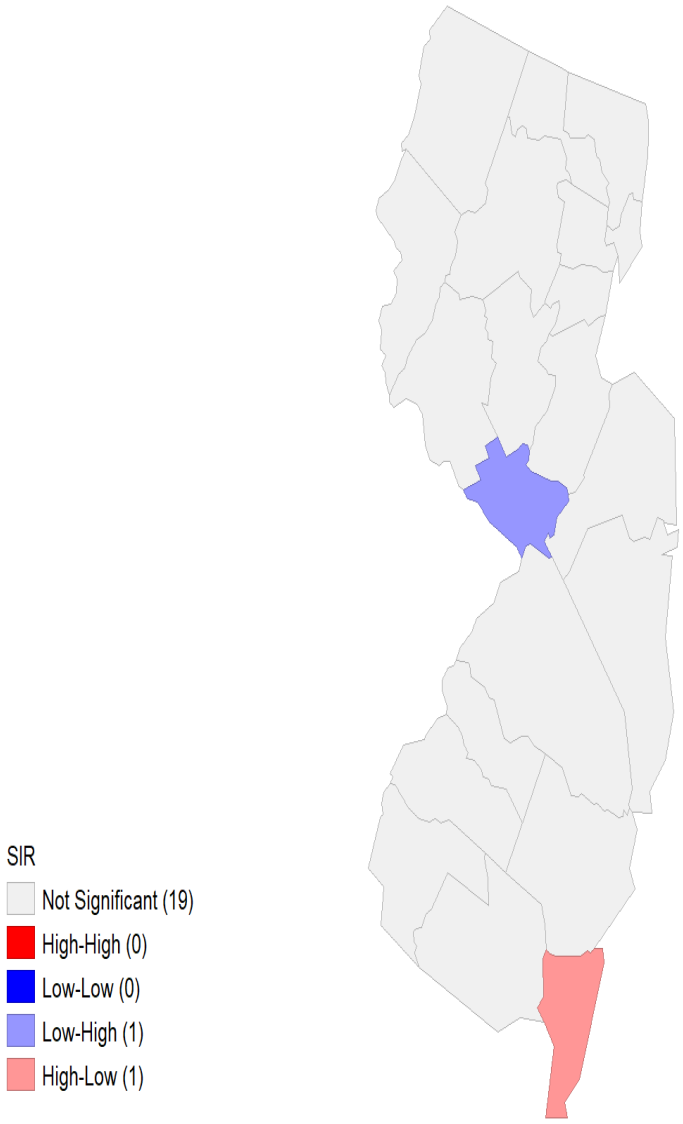
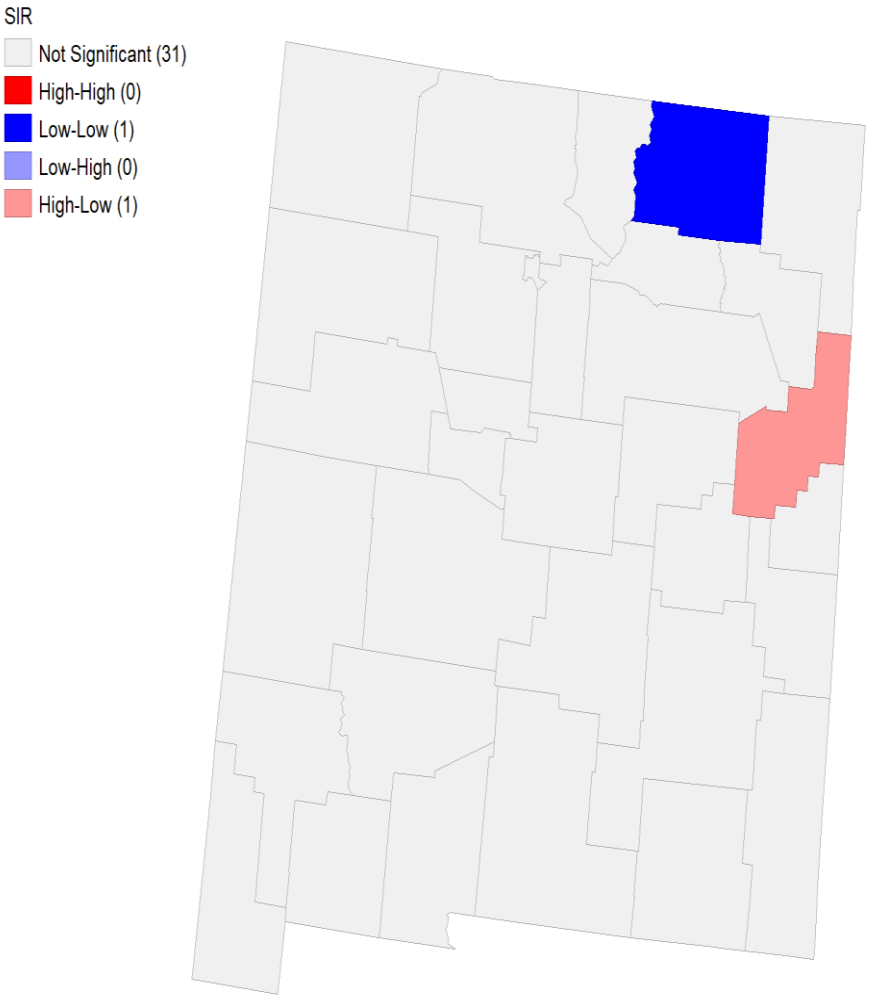


Figure 17: Cluster Map of SIR in New Mexico



TABLES

Characteristic	N	Mean	Minimum	Maximum
SIR	591.00	0.77	0.00	8.25
Observed Cases	591.00	18.43	0.00	1230.00
Expected Cases	591.00	17.75	0.16	1100.53
Benzene concentration	591.00	0.70	0.24	3.24
Benzene Exposure	591.00	0.66	0.16	3.33
TCE concentration	591.00	0.07	0.04	0.61
TCE exposure	591.00	0.06	0.03	0.45
Percentage of High School Graduates	590.00	83.36	58.40	100.00
Median Household Income	590.00	44093.64	19351.00	103643.00
Percentage below the poverty level	591.00	12.82	0.30	37.60
Percentage Uninsured	343.00	15.84	3.80	41.20

Table 1: Descriptive Statistics of observed cases, expected cases, Benzene, TCE and SES variables across all the states under the study.

Characteristic	Value	N (%) / Mean (SD)
Age	Mean	59.94 (16.73)
Sex	Female	4673 (42.34%)
	Male	6365 (57.66%)
Race	White	8786 (79.6%)
	African American	1530 (13.9%)
	Asian/Pacific Islander	679 (6.15%)
	American Indian /Alaska Native	43 (0.39%)
Subtype	MF (9701)	5970 (54.09%)
	CTCL NOS (9702, 9709)	3507 (31.8%)
	pcALCL (9718)	1058 (9.59%)
	All others	503 (4.56%)

Table 2: Descriptive statistics for the 11,038 CTCL patients.

State	Observed Cases	White	African American	American Indian/Alaska Native	Asian/Pacific Islander	Unknown
California	4918	3948	464	22	484	0
Connecticut	684	597	77	0	10	0
Georgia	1305	822	465	1	17	0
Hawaii	155	73	2	0	80	0
Iowa	450	437	8	0	5	0
Kentucky	436	389	44	0	1	2
Louisiana	813	547	260	2	4	0
New Jersey	1657	579	194	0	65	819
New Mexico	258	230	9	16	3	0
Utah	362	345	7	2	8	0
TOTAL	11,038	7,967	1,530	43	677	821

Table 3: Observed cases by Race in each of the states under review

	Total counties	Observed cases	Age 15-59		Age > 60	
			Female	Male	Female	Male
California	58	4918	1078	1265	964	1611
Connecticut	8	684	113	189	155	227
Georgia	159	1305	348	337	262	358
Hawaii	5	155	29	40	28	58
Iowa	99	450	64	88	105	193
Kentucky	120	436	76	101	106	153
Louisiana	64	813	181	204	189	239
New Jersey	21	1657	306	407	410	534
New Mexico	33	258	56	47	66	89
Utah	29	362	63	103	74	122
Total			2314	2781	2359	3584
			5095 (46.16%)		5943 (53.84%)	

Table 4: Distribution of Observed cases by Age & Sex in each of the states under review

Chemical	**Variable Type**	**N**	**β_1 (95% CI)**	**p-value**
Benzene ($\mu g/m^3$)	Concentration (ASPEN)	591	0.23 (0.20- 0.25)	<0.0001
	Exposure (HAPEM)	591	0.22 (0.19-0.24)	<0.0001
TCE ($\mu g/m^3$)	Concentration (ASPEN)	591	1.03 (0.79-1.27)	<0.0001
	Exposure (HAPEM)	591	1.42(1.11-1.72)	<0.0001

Table 5: Poisson Regression Parameters for Benzene and Trichloroethylene across all the states under review

Chemical	**Variable Type**	**N**	**β_1 (95% CI)**	**p-value**
Benzene ($\mu g/m^3$)	Concentration (ASPEN)	343	0.21 (0.17- 0.26)	<0.0001
	Exposure (HAPEM)	343	0.19 (0.15-0.24)	<0.0001
TCE ($\mu g/m^3$)	Concentration (ASPEN)	343	0.22 (-0.07-0.52)	0.1401
	Exposure (HAPEM)	343	0.32(-0.07-0.70)	0.1093

Table:6 Poisson Regression Parameters for Benzene and Trichloroethylene, controlling for SES variables across all the states under review

State	N	Benzene Concentration		Benzene Exposure		TCE concentration		TCE exposure	
		β_1 (95% CI)	P value	β_1 (95% CI)	P value	β_1 (95% CI)	P value	β_1 (95% CI)	P value
California	57	0.08 (0.03-0.13)	0.0034	0.08 (0.03-0.13)	0.0017	0.02(-0.71-0.76)	0.9561	-0.02 (-0.93-0.88)	0.9629
Georgia	159	0.34 (0.24-0.44)	<0.0001	0.34 (0.24-0.44)	<0.0001	0.48 (0.07-0.89)	<0.0208	0.75 (0.19 -1.31)	<0.0084
Iowa	99	0.002(-0.41-0.42)	0.9943	0.08 (-0.46-0.61)	0.7794	-1.48(-9.27-6.31)	0.7093	-2.85 (-13.12 - 7.42)	0.5865
Kentucky	120	0.15 (-0.07-0.38)	0.1926	0.15 (-0.07-0.37)	0.1960	0.57 (-0.79-1.94)	0.4111	0.61(-1.11–2.32)	0.4896
Louisiana	64	0.53 (0.35-0.71)	<0.0001	0.55 (0.37-0.74)	<0.0001	7.80 (3.93-11.65)	<0.0001	9.20 (4.81 - 13.60)	<0.0001
New Mexico	33	0.74 (0.46-1.03)	<0.0001	0.92 (0.57-1.28)	<0.0001	23.07 (13.67 - 32.46)	<0.0001	26.92 (16.05-37.79)	<0.0001
Utah	20	0.71 (0.38-1.05)	<0.0001	0.69 (0.37-1.00)	<0.0001	24.65 (13.01 - 36.30)	<0.0001	30.88 (16.53-45.24)	<0.0001

Table 7: State wise Poisson regression parameters for Benzene and TCE, controlling for SES variables (Excluding Percent Uninsured)

State	Moran's I Statistic for SIR	P-value
California	0.395	0.001
Georgia	0.107	0.001
Iowa	0.146	0.011
Kentucky	-0.035	0.257
Louisiana	0.313	0.001
New Jersey	-0.132	0.319
New Mexico	-0.019	0.440

Table 8: Moran's I statistics and p-values by state