2016

Incidence of Non-Hodgkin's Lymphoma by Residential Proximity to Superfund Sites in Kentucky: A Multivariate Analysis

William Brent Webber
University of Kentucky

STUDENT AGREEMENT:

I represent that my capstone and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's capstone including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

William Brent Webber, Student

David Mannino, MD, Committee Chair

Dr. Wayne Sanderson, Director of Graduate Studies

ABSTRACT OF CAPSTONE

William Brent Webber

The College of Public Health

University of Kentucky

2016

INCIDENCE OF NON-HODGKIN'S LYMPHOMA BY RESIDENTIAL PROXIMITY TO SUPERFUND SITES IN KENTUCKY: A MULTIVARIATE ANALYSIS

ABSTRACT OF CAPSTONE

A Capstone project submitted in partial fulfillment of the requirements for the degree of Doctor of Public Health in the College of Public Health
at the University of Kentucky

By: William Brent Webber

Student Name

Lexington, Kentucky

Director: David M. Mannino, MD
Lexington, Kentucky
Co-Director: Ramona Stone, PhD, MPH
Lexington, Kentucky
Committee Member: W. Jay Christian, PhD, MPH
Lexington, Kentucky

ABSTRACT OF CAPSTONE

INCIDENCE OF NON-HODGKIN'S LYMPHOMA BY RESIDENTIAL PROXIMITY TO SUPERFUND SITES IN KENTUCKY: A MULTIVARIATE ANALYSIS

Non-Hodgkin's Lymphoma (NHL) is a category of cancers that arise from the lymphocytes of the immune system. The rates of NHL in the United States and Kentucky began to rise in the mid-20th Century, shortly after the manufacture, use, and disposal of numerous chemical substances began to increase during and after the Second World War. While the etiology of NHL is not fully known, there are several chemical substances for which evidence exists of a possible link between exposures and development of NHL and other cancers. Several of these substances are also present in sites within Kentucky designated by the US Environmental Protection Agency as hazardous waste sites under the Superfund program. The present investigation sought to determine whether residential proximity to Superfund sites in Kentucky was a significant risk factor for NHL. Geospatial coordinates for all Superfund sites in Kentucky were obtained, along with US Census 2010 population data at the census tract level, and de-identified data from the Kentucky Cancer Registry for all NHL cases between 1995 and 2012, including residential geospatial coordinates. Incidence data was calculated at the level of census tract, except for <5km buffer rings and 5-10km buffer circles around each Superfund site, whose NHL incidence data was calculated separately. Residence within the <5km and 5-10km buffer zones were the exposure variables, and other potentially relevant covariates were considered for the models, and tested for multicollinearity and significance.

Because of spatial autocorrelation of NHL incidence data and non-stationarity uncovered during exploratory regression and diagnostics, geographically weighted regression was used in addition to ordinary least squares regression. Using the best-fitting models, it was determined that residence less than 5km and between 5-10km from the nearest Superfund site were both significant factors in elevated cumulative NHL incidence rates. The Beale Code for rural/urban characteristics of the census tract was another significant predictor, with more rural areas having higher NHL incidence rates. Directions for future research, public health implications, and potential strategies for distal and proximal interventions are presented based on the results of this study.

KEYWORDS: Superfund, environmental exposures, non-Hodgkin's lymphoma

(Student's Signature) William Brent Webber
(Date) February 11, 2016

INCIDENCE OF NON-HODGKIN'S LYMPHOMA BY RESIDENTIAL PROXIMITY TO SUPERFUND SITES IN KENTUCKY: A MULTIVARIATE ANALYSIS

By
William Brent Webber
2016

__David M. Mannino_____
(Signature of Capstone Director)

(Date)___February 11, 2016_____

__David M. Mannino_____
(Signature of Director of Doctoral Studies)

(Date)___February 11, 2016_____

RULES FOR THE USE OF CAPSTONE PROJECTS

Unpublished capstone projects submitted for the Doctor's degree and deposited in the University of Kentucky library are as a rule open for inspection but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgements.

Extensive copying or publication of the capstone project in whole or in part also requires the consent of the Dean of the College of Public Health of the University of Kentucky.

A library that borrows this capstone project for use by its patrons is expected to secure the signature of each user.

Name **Date**

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

INCIDENCE OF NON-HODGKIN'S LYMPHOMA BY RESIDENTIAL PROXIMITY TO SUPERFUND SITES IN KENTUCKY: A MULTIVARIATE ANALYSIS

William Brent Webber

College of Public Health

University of Kentucky

TABLE OF CONTENTS

LIST OF TABLES

PAGE

LIST OF FIGURES

PAGE

ACKNOWLEDGEMENTS

I am grateful for the support and encouragement of so many people, which has resulted in the final product you are reading. Though my name is on the front page, this work was truly a collaborative effort.

Dr. Ramona Stone initially developed the idea that became this capstone, and was the best faculty advisor and research mentor that I could ever ask for. Dr. David Mannino (as chair) and Dr. Jay Christian completed my Capstone Committee, providing thoughtful guidance along the way.

Dr. Eric Durbin and Jaclyn Nee from the Kentucky Cancer Registry provided the cancer incidence data that was critical to this project, and were both thoughtful and willing to answer questions that I inevitably had about the data.

My supervisors at both UK EHS (Lee Poore, David Hibbard) and UNC-Chapel Hill EHS (Ray Hackney, Pete Reinhardt) always encouraged professional development and education, and gave me permission to take courses during work hours that made it possible to complete the curriculum requirements.

Many family members (particularly mom and dad) and friends believed in me and supported me in this endeavor, even when they did not understand why I wanted to do it. Thanks to the Lexington Friends Meeting for helping me to stay centered and calm during the storms, and several cohorts of DrPH classmates I had the pleasure to know.

I learned from many great faculty members at the University of Kentucky, and the University of North Carolina. I am particularly grateful to Professor Pete Andrews at UNC, whose American Environmental Policy course I took back in 2006 cemented my desire to pursue graduate education in environmental health, and was the first official step toward this degree.

The preceptors for my three practicum experiences (Luke Mathis at Lexington-Fayette County Health Department, Gary Shaver at NC State University, and Genia McKee with the Kentucky Injury Prevention and Research Center) were gracious, kind, and provided quality experiences that fulfilled the essential functions of applied public health.

Finally, I have to acknowledge my favorite professor of all, Dr. Kelly Webber. Her love and support make me the most fortunate man imaginable. Hope you don’t mind having another Dr. Webber in the house!

CHAPTER 1
INTRODUCTION

Background – Non-Hodgkin's Lymphoma

Non-Hodgkin's lymphoma (NHL) is a category of cancers that arise from the lymphocytes (white blood cells) of the immune system. NHL can arise from B-cell, T-cell, or natural killer (NK) lymphocytes, and is differentiated from Hodgkin's lymphoma (also known as Hodgkin's disease) by the absence of a particular type of abnormal cell called the Reed-Sternberg cell, which is present in Hodgkin's lymphoma[1]. The incidence and prevalence of NHL far exceeds that of Hodgkin's lymphoma in the United States (U.S.); in 2014, NHL accounted for 88.5% of all estimated new lymphoma cases, and was responsible for sixteen times more cancer deaths than Hodgkin's lymphoma[2-3]. Lymphomas differ from leukemia, another type of cancer that can manifest in the lymphatic system, in that leukemia is a cancer of the blood-forming cells in bone marrow, which can develop into myeloid or lymphoid variants[4]. By contrast, lymphomas arise from the abnormal transformation and growth of already differentiated B-cells (and, less frequently, T-cells) in the lymphatic system, often resulting in solid tumors[5].

Different types of NHL are most often categorized by the types of cells affected, the location(s) of solid tumors, or both. Approximately 85% of NHL cases arise from B-cells, 15% from T-cells, and less than one percent from NK cells[6-7]. The most common forms of NHL in the U.S. are the diffuse large B-cell lymphomas (approximately 33% of cases)[8] and B-cell follicular lymphomas (approximately 20% of cases)[6]. The most common type of T-cell NHLs are peripheral T-cell lymphomas, which account for

approximately 14% of U.S. NHL cases[6]. Cases of NHL where tumors arise from lymph nodes or other lymphatic tissues such as the spleen or thymus are referred to as intranodal NHL, whereas extranodal NHL arises from lymphatic cells in other organs such as the small intestine, stomach, and skin[9-10]. Approximately 25% of U.S. NHL cases are of the extranodal type[10].

In 2012, NHL was the 8th most common cancer in the overall U.S. population, the 6th most common cancer among males, and the 7th most common among females[11]. Males have a higher incidence rate for NHL compared to females in the U.S. (22.5 vs. 15.3 per 100,000)[11] and worldwide, for reasons that are not fully understood but which could involve protective effects from estrogen or other hormones in females[12-13]. NHL incidence rates in the U.S. increased markedly during the middle of the 20th Century, before stabilizing in the mid-1990s but remaining among the highest in the world to the present day[2,14-16]. While most other forms of cancer either showed a decline in the incidence rate during the 20th Century, or an increase that could be directly tied to known causal factors (e.g. lung cancer and tobacco smoking) or improved screening and early detection, the increase in NHL incidence defies simple explanations. Therefore, the contribution of several factors including exposures from the external environment must be considered.

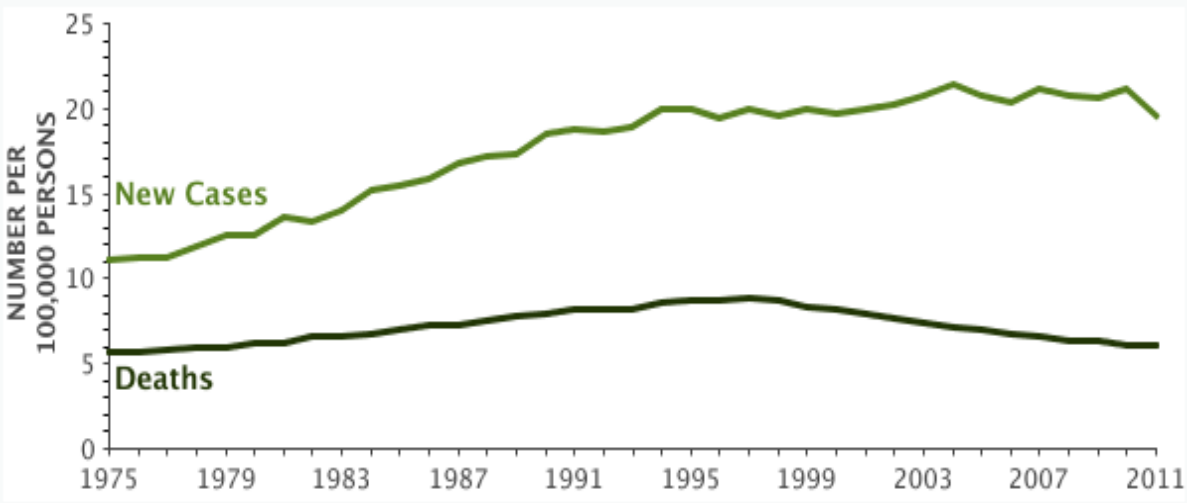


Figure 1: Hon-Hodgkin's Lymphoma Incidence rate and death rate per 100,000, United States, 1975-2011 (Source: National Cancer Institute, Surveillance, Epidemiology, and End Results (SEER) Program[2])

Background – Hazardous Waste Management Practices: United States

The demands of the Second World War (1939-1945) led to the proliferation of chemical investment and innovation in both Allied and Axis nations. This set up a postwar technological trajectory that led to the rapid expansion of fields such as petrochemicals, pesticides, and pharmaceuticals[17]. New products were being introduced to the market and into the environment before all the known and suspected human and environmental toxicities could be fully known. The first reports of human and aquatic toxicities from new agricultural pesticides, which had been originally developed as wartime agents, began emerging as early as the late 1940s[18].

In the early 1950s, at approximately the same time when NHL incidence began to increase in the U.S.[19], new questions began to arise about the effects that widespread chemical usage, contamination, and waste disposal practices might be having on human health and the environment, and whether these might be contributing

to specific disease outcomes. The tremendous increase in the production and use of new and established chemical products following the Second World War led to the generation of both hazardous and non-hazardous waste streams at previously unseen levels, and led to greater potential for community or ecosystem exposures[20]. One touchstone in the development of the modern grassroots American environmental movement was the publication of *Silent Spring* by Rachel Carson in 1962, which detailed the negative effects of non-selective pesticides on avian species[21].

The ways in which various jurisdictions and industries chose to deal with wastes frequently resulted in environmental costs being borne by the disadvantaged, minorities, and persons other than those in the generating industry or municipality[22]. Greater environmental consciousness and demands for justice led to federal standards as a baseline for state and local regulations, and the eventual establishment of the U.S. Environmental Protection Agency (EPA) in 1970. The Clean Air Act and Clean Water Act were the first major regulations adopted by EPA; however, solid waste and hazardous waste was considered a unique local problem that resisted federal regulation until 1976[23].

The Resource Conservation and Recovery Act (RCRA) of 1976 set forth national standards for sanitary landfills and other disposal methods of municipal solid wastes, and “cradle-to-grave” transport, storage, and ultimate disposal of hazardous wastes[24]. However, while RCRA dealt with waste disposal and reduction practices going forward, it did not deal with past dump sites which could continue to affect their surrounding communities for decades[23]. The need for additional federal regulations to deal with past hazardous waste sites was vividly illustrated in the late 1970s by the Love Canal

disaster in Niagara Falls, New York and the discovery of the Valley of the Drums in Bullitt County, Kentucky[25-26].

The Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA) established mechanisms for determining which sites constituted the greatest threat to human health and the environment, and designated a tax on petrochemical industries to generate a trust fund known as “Superfund” to clean up these sites[27]. When potentially responsible parties could not be found, or did not have the resources necessary to adequately clean-up a site, money from the Superfund would be used to pay for these activities. In general, all sites on which CERCLA-covered activities have occurred or are being investigated are known as Superfund sites and maintained in EPA databases, whereas the most hazardous sites requiring greater and long-term remediation activities are put on the National Priorities List (NPL) and are often additionally referred to as NPL sites[28]. As of 2015, there were a total of 234 Superfund sites in Kentucky, twenty of which are currently or formerly designated as NPL sites.

Statement of the Problem

Long before the official designation of NPL and Superfund sites, both the research community and those who live near these sites wondered the extent to which human health effects might have resulted from exposures to materials at these sites. The public frequently submitted requests for “cancer cluster” investigations to the Centers for Disease Control and Prevention (CDC) around these sites, partially due to a

tendency among the public to attribute negative health outcomes to external environmental factors to a greater extent than the scientific community[28-29].

Nevertheless, there are also scientific reasons to investigate causality of environmental factors for health outcomes. Several contaminants present at Superfund sites are of special concern due to their persistence, bioaccumulation, acute toxicity, and likelihood of exposure due to localized accumulation[30]. However, multiple spatial and temporal issues complicate these investigations, such as on-site process variability, residential mobility, latency factors for chronic diseases such as cancer, geologic and meteorological variance, transformation of waste products over time, and delineation of exposure vs. non-exposure areas[31]. Because human exposures to environmental insults from Superfund sites are usually low-level and variable, most physiological-based pharmacokinetic modeling strategies used in classical toxicology are not a good fit[32]. Furthermore, there are numerous possible confounders with environmental exposure outcomes research in human subjects, such as race, socio-economic status, along with smoking, diet, exercise, recreational activities, and occupational exposures which must be considered[32].

With these caveats in mind, it remains critically important to evaluate the potential causality of environmental factors in negative health outcomes, particularly those such as NHL whose increased incidence in the U.S. and several other developed nations tracks reasonably closely with the wider industrial use, disposal, and dispersion of chemical substances into the environment.

Purpose and Significance of the Study

Kentucky has 20 NPL sites and 234 total Superfund sites, but to date no analyses with geospatial tools have been published on the possible relationship between health status and residential proximity to these sites. Kentucky is also presently ranked 47th nationally for overall health and is the state with the highest cancer death rate[33]. Though several studies have examined possible cancer clustering around hazardous waste sites, many used crude and relatively large areas of "exposure" and "non-exposure", and were not able to demonstrate possible gradient effects. This study will use statistical and geospatial tools to model the extent to which residential proximity to Superfund sites in Kentucky might explain prevalence of NHL, while controlling for covariates.

NHL was chosen as the outcome variable due to its possible association with environmental exposures, and its unique historic incidence trends in the United States. For the years 2007-2011, Kentucky was ranked 4th nationally for age-adjusted non-Hodgkin's lymphoma death rate[34]. The NHL rates in Kentucky also parallel the national and international Western trends of increased incidence in the mid-20th Century across all genders and age groups[35], with the highest overall rates seen in white males[8,35]. This pattern is seen in both intranodal and extranodal forms of NHL, with extranodal NHL exhibiting similar demographic patterns to intranodal NHL[9,36].

Geospatial analysis tools can be used to determine potential gradient effects if data for both exposures and outcomes can be geocoded and modeled. Ultimately, if it is determined that residential proximity to NPL/Superfund sites increases NHL risk, and/or "hot spots" are uncovered, public health strategies for prevention and early detection

can be directed to these areas in order to save lives and prolong quality of life[37]. In addition, if these risks are demonstrated to disproportionately affect disadvantaged persons, the study can provide supporting evidence for programs designed to foster improvements in social justice and equality.

The current study examined incidence rates from 1995 through 2012 (the most recent year from which full data was available) from the Kentucky Cancer Registry for all forms of non-Hodgkin's lymphoma (NHL), intranodal and extranodal. While all types of NHL could have heritable and lifestyle factors for risk, they also have known or suspected etiologies from the external environment as described earlier.

The goal of the project was to perform a spatial analysis to examine the relationship between NHL cancer incidence and proximity to hazardous sites. Data sources included the following:

- US EPA Superfund location data for Kentucky. Geospatial coordinates are available for all twenty NPL sites in Kentucky. For the non-NPL sites, 113 of the 214 were unique, non-duplicative sites with geospatial coordinates available; thus, the total number of EPA sites available for inclusion is 133;
- Kentucky Cancer Registry 1995-2012 case data for incidence of non-Hodgkin's lymphoma, intranodal and extranodal;
- 2010 US Census demographic data at the census tract level. This is a unit of census data that can be tabulated while maintaining subject confidentiality. Census tracts represent between 1500 and 8000 people, and are intended to represent neighborhoods that are relatively stable and homogenous[38].

From these data, the next steps were to use methods of geospatial analysis, modeling, and outcome measurement to determine whether a significant effect exists between residential distance from Superfund sites and the incidence of NHL. The software utilized included ArcGIS, SPSS, and SAS in addition to MS Excel. Results could have a significant impact on public health if local hot spot areas are found where beneficial activities could be focused, such as cancer screening, nutritional interventions that could reduce vulnerability to stressors[39], or built environment interventions.

CHAPTER 2

LITERATURE REVIEW

The following review of the literature is a summary of key concepts foundational to understanding the types of health effects that can cluster near hazardous waste sites, the use of geospatial tools to investigate these clusters, and the possible environmental causes of non-Hodgkin's lymphoma (NHL) and other cancer types. It represents theoretical and empirical knowledge gathered from the disciplines of epidemiology, medicine, environmental health, geography, and governance. The works cited are collected from peer-reviewed journal articles, book chapters, conference papers, symposium proceedings, and governmental agencies. Because the first Superfund sites were designated by the US EPA in 1981, an examination was conducted of the literature from 1981 to 2015 in order to capture the earliest evaluations of possible health effects at these sites. The databases and sources used to identify the scholarly literature included EBSCO Academic Search Complete, NCBI PubMed, OCLC WorldCat, and LexisNexis. The key words and phrases for the searches included "Superfund", "environment", "cancer", "geospatial analysis", "lymphoma", "non-Hodgkin lymphoma", and "non-Hodgkin's lymphoma". A secondary review of writings referenced in the bibliographies of key works augmented the process.

The literature review also focused on the most common types of cancer in the study population for which it was possible that environmental exposures could play a role in their initiation, promotion, or progression. The primary focus was on NHL, but literature on breast cancer and bladder cancer was also reviewed, as these were three

of the ten most frequent cancers in the population study basin. The other common cancer types in the study basin had predominant risk factors that were genetic or dietary (colon cancer), or suffered from the presence of overwhelmingly powerful confounders (e.g. lung cancer and high rates of smoking and indoor radon exposure, skin cancer and ultraviolet exposure).

One of the first epidemiologic studies of a designated Superfund National Priorities List (NPL) site found that white males in the county hosting the site had a significantly elevated odds of developing bladder cancer (OR = 1.7, $p<0.025$)[40]. The authors measured exposure at county level, using national averages for comparison, and multiple outcomes were evaluated, so the elevated risk might have been due to multiple comparisons[40]. A study from two Superfund sites in Texas found that residents designated as "high-exposure" due to residential proximity to the sites self-reported more neurological symptoms compared to low-exposure populations[41], and that incidence rates for multiple cancers were elevated in the vicinity of a Department of Defense Superfund site in Massachusetts[42]. Serum immunoglobulin A levels were found in one meta-analysis to be consistently but *not* significantly elevated for residents near Superfund sites compared to matched controls at least five miles away from sites[43]. Another study estimated that multi-state Superfund site cleanup activities reduced the rate of infant congenital abnormalities by 20 to 25 percent for mothers who resided 5 km or less from the sites[44].

Studies have also examined the degree to which contaminants could migrate from Superfund sites into the surrounding ecosystems and communities. Tree bark samples within 10 km of an NPL site in Michigan showed 10- to 100-fold increases in

dichlorodiphenyltrichloroethane (DDT), hexabromobenzene, and polybrominated biphenyls compared to sites >10 km distant[45]. Passive sampling devices that mimicked the way living organisms accumulate lipophilic contaminants were deployed near an NPL site in Portland, Oregon, and contaminant levels that would result in an excess cancer risk greater than the EPA limit of $1x10^{-6}$ were found[46]. Residents near former uranium mining NPL sites had drinking-water ionizing radiation levels that exceeded public limits[47]. Researchers have also investigated social justice concerns with the siting of Superfund/NPL sites and found that poor and/or minority populations tend to be disproportionately affected[48-52].

The addition of geospatial information and tools in public health research have increased precision for examining spatial patterns within data, understanding relationships between outcomes and environmental variables, and inferring exposure patterns[53]. When precise address information is available for cases, geospatial analysis can provide sharp, precise boundaries of a cluster or area of exceedance to most efficiently deploy public health resources[54]. As evaluated areas get smaller (e.g. county, census tract, census block, geospatial coordinates), there is less variability in exposures, and ecological fallacy becomes less likely[55].

A geographic distribution analysis showed that blood levels of dieldrin (an organochlorine insecticide) increased by 1.6 ng/g for each one mile of closer residential proximity to a Superfund site in Maryland[56]. Another study of an NPL site contaminated with polychlorinated biphenyls (PCBs) found that residential proximity to the site was not a significant factor in cord serum PCB levels, but being born before or during dredging activities to remove PCBs from the site was significant[57]. Geospatial analysis has also

been used to identify clusters of childhood cancer near NPL sites in Dade County, Florida[58], very-low birth weight near multiple NPL sites in Harris County, Texas[59], and to investigate and confirm the unequal burden of NPL/Superfund sites among specific racial, ethnic, and socioeconomic demographics[60-64].

Outside the context of U.S. Superfund/NPL sites, exposures to e-waste dismantling sites in China at the village level were found to significantly elevate serum levels of thyroid-stimulating hormone and polybrominated diphenyl ethers, along with micronucleated binucleated cells[65]. In Taiwan, spatial autocorrelation analysis identified hot spots for various cancers in females in areas with high levels of environmental exposures to arsenic, nickel, and chromium[66]. And, in Australia, excess cancer risk and elevated soil arsenic from historic gold-mining activities were both found in economically disadvantaged areas[67].

Some primary and secondary contaminants frequently found at Superfund sites are suspected of initiating or promoting specific types of cancer such as breast, bladder, and NHL. Superfund site contaminants were shown to initiate or promote breast tumorigenesis through endocrine disruption[68]. Organochlorine compounds such as PCBs and DDE can act as estrogen mimics and partition into adipose tissues[69-70]. Organic solvents such as halogenated hydrocarbons and aromatic amino/nitro compounds exhibit mammary tumorigenic activity in rodent models[71], and women who were occupationally exposed to solvents prior to first full-term birth had a significantly elevated breast cancer risk[72]. Additional animal studies have revealed more than 200 chemicals and heavy metals that were mammary carcinogens or estrogen mimics[73-74].

While the predominant risk factor for bladder cancer is smoking, accounting for approximately 50% of all cases[75], environmental exposures to aromatic amines and polycyclic aromatic hydrocarbons are also risk factors for bladder cancer[76-79]. Arsenic that leaches into drinking water is another known bladder cancer risk factor, though the mechanisms are not well understood[80].

NHL incidence rates in the United States increased dramatically during the middle of the 20th Century and plateaued in the mid-1990s, but incidence rates still remain well above the levels seen prior to the post-Second World War chemical age. Persistent organochlorine compounds that became prevalent in the early- and mid-20th Century have been suspected as a causal factor, and numerous studies have shown associations between these compounds and NHL[81-88]. Meta-analysis of multiple case-control studies has demonstrated a significant association between occupational exposures to pesticides and NHL[89]. Non-occupational exposures to two specific types of organochlorines, chlordanes and DDT, have been repeatedly associated with NHL in multiple studies[89-93]. Occupational exposures to pentachlorophenol has been associated with increased risk of NHL[94]. Polychlorinated biphenyls (PCBs) are another broad category of organochlorines that are persistent organic pollutants with high patterns of usage in the early 20th Century, and which show consistent causal associations with NHL[83,95-97].

Other chemicals that have been positively associated with NHL and which can be present at Superfund sites include phenoxy herbicides[87,98], carbamate insecticides[87], organophosphorus insecticides[87], benzene and benzyl compounds[99-101], trichloroethylene[16], perchloroethylene[102], polychlorinated dibenzo-*p*-dioxins and

dibenzofurans[103-104], 1,3-butadiene[105], cadmium[106], and high nitrate levels in community water supplies[107]. The incidence and mortality rates of NHL were elevated for persons with non-occupational exposures to herbicides[108]. Elevated risk of NHL was also found in residents living near the Italian equivalent to NPL sites[109], residential areas where exposures to traffic noise consistently exceeded 65 decibels[110], lumber and wood products facilities[111], pulp and paper industries[112-113], copper smelters[113], refineries that emit lead and cadmium[114], and residences where geothermal hot water is used[115].

Numerous mechanisms have been proposed for how environmental exposures could lead to increased rates of NHL. Immune system suppression, which can be triggered by xenobiotics, is one of the primary known risk factors for NHL[95,116-118]. Widespread exposures to lymphomagenic substances can trigger immunosuppressive conditions[36]. Conversely, persons with a history of allergies, other hyperimmune disorders, or asthma appear to have a reduced risk of developing NHL[119-121]. Overexpression of cellular protein Exportin-1 which mediates the transport of other proteins between the nucleus and cytoplasm has also been associated with increased risk of NHL, and Exportin-1 inhibitors have shown early promise in treatment of NHL[122].

While the Human Immunodeficiency Virus (HIV) has been suspected as one of the causal factors in rising NHL incidence due to its profound immunosuppressive effect, NHL incidence rates have also risen among the HIV-uninfected[123-124]. In addition, HIV and the associated illness of Acquired Immune Deficiency Syndrome (AIDS) did not rise to a level detectable by public health surveillance in the U.S. until 1981[125], and the NHL incidence spike started decades earlier.

Some xenobiotics can have a directly toxic effect on the hematopoietic system, which can lead to various forms of lymphoma or leukemia. Examples of this include benzene[126], cadmium[127], and lead[128]. Another possible mechanism of toxicity is the generation of reactive oxygen species following xenobiotic exposures, resulting in damage to cellular DNA[129], or chronic antigen stimulation resulting in inflammatory cascades[36]. Alternately, the site contaminants themselves can be transformed into toxic free radicals capable of direct cytotoxicity. Two examples of this are the transformation of pentachlorophenol into free radicals that can persist in the environment for decades[130], and the generation of environmentally persistent free radicals that can inhibit cytochrome p450-based xenobiotic metabolism[131]. An Algerian study indicated that pesticide exposures altered the ratio of T-helper 1 to T-helper 2 cells via proliferation of nicotinamide adenine dinucleotide phosphate hydrogen (NADPH), significantly increasing the risk of NHL[132].

It is also probable that gene-environment (GxE) interactions lead to the development of some NHL cases[133-134]. Multiple subtypes of NHL are associated with the t(14;18) chromosomal translocation and oncogenic activation, which can be triggered by environmental toxicants[129,135-138]. For some cases of NHL, the relationship between organochlorine exposures and outcome appears to be modified by variations of genes for numerous interleukins[139]. Genetic variance in xenobiotic metabolism and DNA repair pathways have also been shown to likely modify the relationship between NHL and chlorinated hydrocarbon exposure[140]. Single-nucloetide polymorphisms in the Ataxia-Telangiectasia Mutated and Tumor Necrosis Factor-alpha loci are associated with elevated risk of multiple B-cell NHL subtypes[141-142].

For the twenty sites in Kentucky on the National Priorities List, information on the contaminants of concern at each site is available from the US Library of Medicine TOXMAP webpage[143]. The following seven categories of contaminants associated with increased risk of NHL were present at Kentucky NPL sites: benzene and benzyl compounds (15 of 20 NPL sites, 75%), lead (14/20, 70%), polychlorinated biphenyls (11/20, 55%), cadmium (9/20, 45%), trichloroethylene (7/20, 35%), organochlorines other than PCBs (6/20, 30%), and perchloroethylene (2/20, 10%). Information on contaminants of concern was not available for the non-NPL Superfund sites in Kentucky.

In summary, the literature review shows that NHL incidence in the U.S. and Kentucky has risen in a temporal pattern that appears commensurate with the greater use and dispersion of multiple chemical substances into the environment. The literature also demonstrated there are feasible ecological and biological mechanisms by which substances from Superfund and other hazardous chemical sites can enter the community environment and exert toxic effects, including those that could trigger NHL. The methodology in the next chapter details the research strategy employed to investigate whether residential proximity to Superfund sites in Kentucky could be at least partially responsible for an increased incidence of NHL. This question has not been previously investigated, and the results could point to the need for greater screening, awareness, and other interventions that might save lives, prolong the quality of life, or increase environmental justice in affected regions.

CHAPTER 3
METHODOLOGY

This is an observational population study, using data from three sources, to examine the relationship between the residential proximity of patients diagnosed with non-Hodgkin's lymphoma (NHL) to environmentally hazardous sites and the likelihood for developing NHL cancer, while controlling for individual characteristics. The hypothesis is that residential proximity to Superfund sites in Kentucky is a significant factor in an increased risk of NHL, even after adjusting for other covariates.

Data Sources

The NHL cancer data from the Kentucky Cancer Registry (KCR) was obtained for 18 years, from 1995 to 2012, following approval of the University of Kentucky Institutional Review Board (Appendix 1). All individual identifying information was removed from the data by KCR before it was given to investigators, other than the geospatial coordinates for their residential address. Each patient was a assigned a random unique identification number, which was used to unduplicate the data to retain the first cancer diagnosis and eliminate any subsequent ones.

The environmental exposure data came from the US EPA Superfund website for Region 4 (which includes the state of Kentucky)[146]. The EPA sites were categorized based on whether they were presently or formerly on the National Priorities List, and the geospatial coordinates of the area where the contamination occurred or is occurring. In Kentucky, there are 20 current or former NPL sites, all of which had full geospatial

coordinates available. There were a total of 214 additional, non-NPL Superfund sites, for which only 113 had geospatial coordinates available (Figure 2). The 133 total sites with geospatial data were treated as point-sources of environmental exposure.

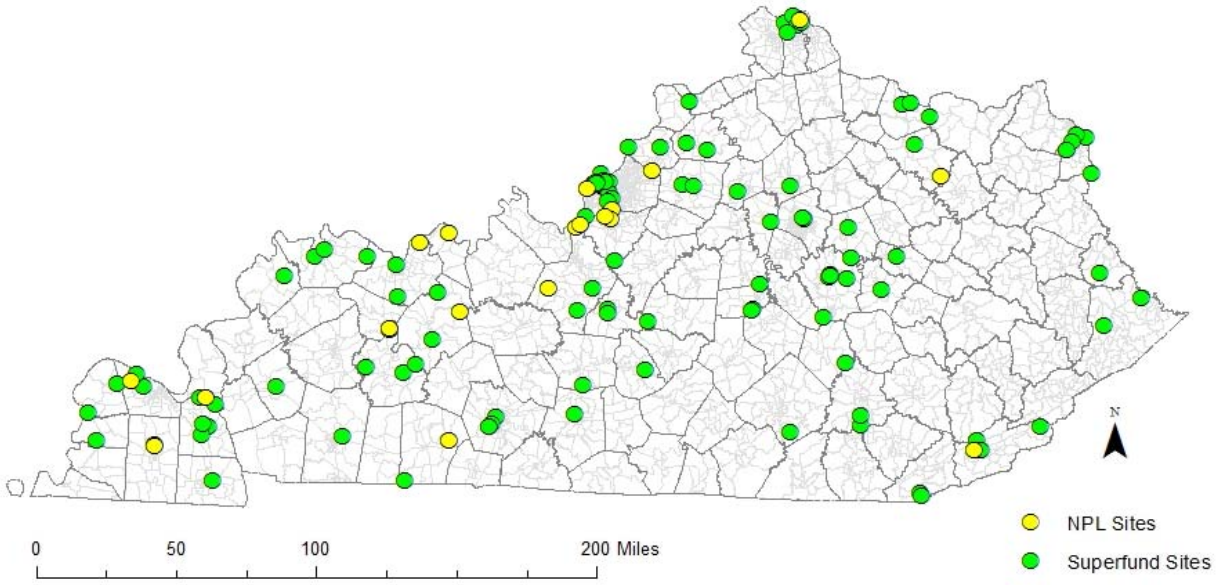


Figure 2. Location of 133 NPL/Superfund sites in Kentucky for analysis

Census tract Tiger file was obtained from the 2010 US Census website. There are a total of 1,115 census tracts in Kentucky of which 734 had reported cases of NHL between 1995 and 2012.

Independent Variables

The independent variables include categorical residential proximities to Superfund sites and individual patient data from the KCR. The 2010 census tract counts

of all people by gender and age categories (5-year increments) was used for the computation of age-standardized incidence rates (e.g., overall, for males, for females).

The patient level data from the KCR included NHL cases with both types of this cancer: intranodal and extranodal. Other variables available for the NHL cases included sex, race, ethnicity, tobacco usage (categorical), age at diagnosis, family history of NHL, county of residence, Appalachia region designation, and Beale Code that categorizes areas by their level of urbanization. The exposure, or patient's residential proximity to Superfund sites, was measured by an ordinal variable with three categories: 0= exposure beyond 10 km, 1= exposure within 10 km, but beyond 5 km, and 2= exposure within a radius of 5 km. For the multivariate analysis the exposure variable was recoded into two dummy variables; exposure within 5km (yes/no), and exposure between 5km and 10 km (yes/no), with exposure beyond 10km as the reference group for the analysis.

Dependent Variables

The dependent variables in this study are the age adjusted rates: overall, by gender, and by cancer type (extranodal, intranodal). The age standardization of rates were computed using the direct method with the 2000 US Census population as a weighting factor, per current recommendations from the Centers for Disease Control and Prevention[144-145]. Table 1 shows the crude rate for NHL incidence in Kentucky, 1995-2012, the age-adjusted rate, and the weighting factors applied.

Table 1: Weighting Factors for Age Standardization of Kentucky NHL Data, 1995-2012, based on 2000 US Census

Age Group	Population (A)	Number of NHL cases (B)	Age-Specific NHL incidence Rate (per 100,000) (C)	Weights for 2000 U.S. Standard Pop. (D)	Weighted Rate (E)
0 – 9	565255	61	10.79158964	0.141668548	1.528828834
10 – 19	580949	127	21.86078296	0.145200521	3.174197075
20 – 29	575264	221	38.41714413	0.131007086	5.032918105
30 – 39	566331	583	102.9433317	0.151805676	15.62738206
40 – 49	614893	1300	211.4188973	0.153968555	32.55186211
50 – 59	607482	2392	393.7565228	0.111169775	43.77382405
60 – 69	436630	3546	812.1292628	0.073057233	59.33191678
70+	392563	6143	1564.844369	0.092122607	144.1575428
Total	4,339,367	14373	331.223425	1.00000	305.1784718

Note: C=B / A; E=C*D

The patient data was geocoded at census tract level; each census tract has a 12-digit Federal Information Processing Standards (FIPS) Code[147]. Each case was placed wtihin a census tract, based on its geographic coordinates. The age-adjusted incidence rates of NHL were estimated for each exposure area, and for all census tracts outside the exposure areas, by using the 2010 Census census tract population as the denominator and the 1995-2012 NHL cases as the numerator, along with the 2000 US Standard population weighting factors. Specifically, the exposure areas were developed in ArcMap by drawing 5km and 10km buffers around each Superfund site and by identifying which census tracts and how much of their geographical areas fall within each exposure area. When buffers of neighboring Superfund sites intersected, they were dissolved into a single area of exposure, and the perimeter of all of the conjoined buffers became the boundary of the newly created exposure areas. Therefore, the 5km exposure areas have different sizes and shapes, including different number of census

tracts (or fragments of census tracts), and different numbers of Superfund sites within their boundaries.

There were 71 areas within 5km of one or more Superfund sites, and each was assigned a unique ID. The 5km dissolved areas were removed (erased) from the 10km disolved areas, to form a secondary area of exposure (that sometimes looks like a donut) that was outside 5km of any Superfund site, but within 10km of at least one Superfund site. There were 45 areas of exposure between 5km and 10km away from any Superfund site; these areas were also assigned a unique ID. Finally, the remaining areas of the state, outside the 5km and 10km exposure areas, formed the third area of interest, the “unexposed” areas of the state, for which the incidence rates were computed at census tract level.

To account for the distribution of population across the census tract fragments that fall within the 5km radius, between 5km and 10km “donut”, and beyond 10km, the proportion of each census tract that falls within a specific area of exposure was calculated. This calculated percentage was applied to the computation of the census population with specific characteristics (e.g., in terms of age and gender) within each fragment. A multiple exposure variable was created to account for the differences in the number of Superfund sites within the boundaries of different exposure areas.

The numerator in the formulae for the age-adjusted rates was the number of cases with specific gender and age characteristics within each exposure area, and was calculated using the spatial location of each patient’s residence. Specifically, cases were placed into exposure categories based on whether their residential geospatial coordinate was within a 5km exposure area, a 5km to 10km “donut” exposure area, or

outside all of the exposure areas. Cases with residence located outside of all the exposure areas (e.g. lived more than 10km from any Superfund site) were classified as “unexposed”. At the latitude where Kentucky is located on the globe, 103.44 km equals one decimal degree.

Univariate analysis included counts, proportions, means, medians and standard deviations for the following characteristics by level of analysis:

a) For the patient or case level data, univariate analysis is provided for the following variables: gender, race, age at diagnosis, current tobacco use, tumor type, family history of NHL, residence in Appalachian regions, Beale Code urban or rural designation, and residential proximity to the nearest Superfund site.
b) For the census tract level, univariate analysis is provided for the following variables: number of NHL cases, total population, and number of Superfund sites within the tract.

Bivariate analysis was conducted using statistics such as chi-square tests, t-tests, and the one-way analysis of variance (ANOVA). The underlying population was naturally skewed, but the sampled data set was large enough to offset this.

Spatial regression analyses were performed for each dependent variable. The dependent variables were the incidence rates of NHL at the census tract level, overall and stratified by gender and SEER tumor type classification. In each of these regression models, the principal predictor variable was the type of exposure area (within 5km, over 5km to 10km, and over 10km) measuring the residential proximity to Superfund sites as a proxy for possible exposures to contaminants at these sites. Other independent

variables that were identified in the literature as possibly related to NHL incidence and also considered for the regression models were smoking status and family history of cancer. Race was not included in the analyses due to the small number of cases that were of other race than Caucasian/White.

Due to the relatively large size of the study basin (the entire Commonwealth of Kentucky), and its underlying regional, cultural, and socioeconomic diversity, it was determined that Geographically Weighted Regression (GWR) would most likely be necessary for analysis. With large study areas such as an entire state or region, it is often not prudent to use global or aspatial regression because the impact of covariates can vary across the area[148-149]. Diagnostic tools were used on the data to detect the presence of spatial autocorrelation and clustering, and thus to confirm the choice to use GWR in addition to ordinary least squares regression modeling.

CHAPTER 4
RESULTS

The results of the descriptive analyses are shown in Table 2 for case data. There were a total of 14,373 new NHL cases in Kentucky between 1995 and 2012. Per the 2010 US Census, Kentucky has a total of 1,115 census tracts within its 120 counties. While 82.3% of the NHL cases could be assigned to census tracts based on high-quality residential geospatial coordinates, for the remaining 17.7% the geospatial coordinate was the centroid of their residential ZIP code. This often occurs when a case lists their address as a rural route or post-office box.

Univariate Analysis

The caseload of 14,373 patient population included 51.5% males, 94.7% of all cases were white, and 39.1% were current users of tobacco products. Intranodal NHL accounted for 70.8% of all cases, 71.7% of male cases, and 69.9% of female cases. Only 3.7% of the cases had a known prior family history of NHL. For most cases, there was either no prior family history (52.2%) or an unknown prior family history (44.1%). Only 28.1% of cases lived in counties that were part of the designated region of Appalachia, and only 9.6% of cases lived in Beale Code designated rural regions. Of all cases that were of other than white race, only 0.6% were Hispanic or Latino of any race (data not shown) and 4.4% were African American; due to the low proportion of non-white cases, analyses by race could not be completed. In accordance with national NHL

statistics, 67.4% of all diagnoses occurred in patients age 60 or older. Nearly 30 percent of cases lived within 5km of a Superfund site.

Table 2: Descriptive Statistics for the Patient Data (N=14,373)

Demographic Variable	Category	Number	% of Total
Gender	Female	6978	48.5
	Male	7395	51.5
Race	White	13617	94.7
	Black	632	4.4
	Other/Unknown	124	0.9
Age at Diagnosis	0-9	61	0.4
	10-19	127	0.9
	20-29	221	1.5
	30-39	583	4.1
	40-49	1300	9.0
	50-59	2392	16.6
	60-69	3546	24.7
	70 and above	6143	42.7
Tobacco Use	Non-User	5715	39.8
	Cigarette Smoker	5237	36.4
	Cigar-Pipe Smoker	138	1.0
	Smokeless Tobacco User	136	0.9
	Multiple Types of Tobacco Used	116	0.8
	Not Recorded/Unknown	3031	21.1
Tumor Type	Intranodal NHL	10181	70.8
	Extranodal NHL	4192	29.2
Family History of NHL	No	7495	52.2
	Yes	533	3.7
	Unknown	6345	44.1
Appalachia Region	No	10337	71.9
	Yes	4036	28.1
Beale Code Classification	Urban	12997	90.4
	Rural	1376	9.6
Residential Proximity to Nearest Superfund site	<5 kilometers	4225	29.4
	5-10 kilometers	3570	24.8
	>10 kilometers	6578	45.8

The total number of new cases per year showed an upward trend between 1995 and 2007, after which rates plateaued but stayed elevated through 2012 (Figure 3). The highest total number of cases was in the last year of collected data (2012), and the

lowest number was in the first year collected (1995). When cases were stratified by race into white and non-white categories, this trend was more pronounced in non-white populations (Figure 4), though the peak year for non-white NHL cases was 2009.

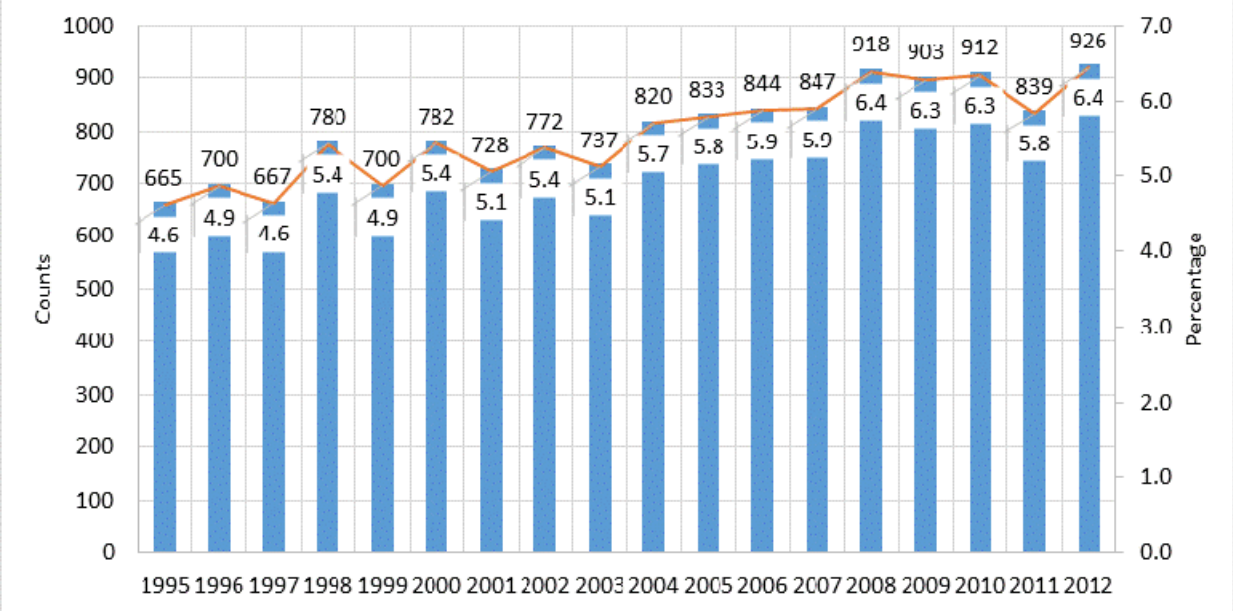


Figure 3. New non-Hodgkin's Lymphoma Cases in Kentucky, 1995-2012

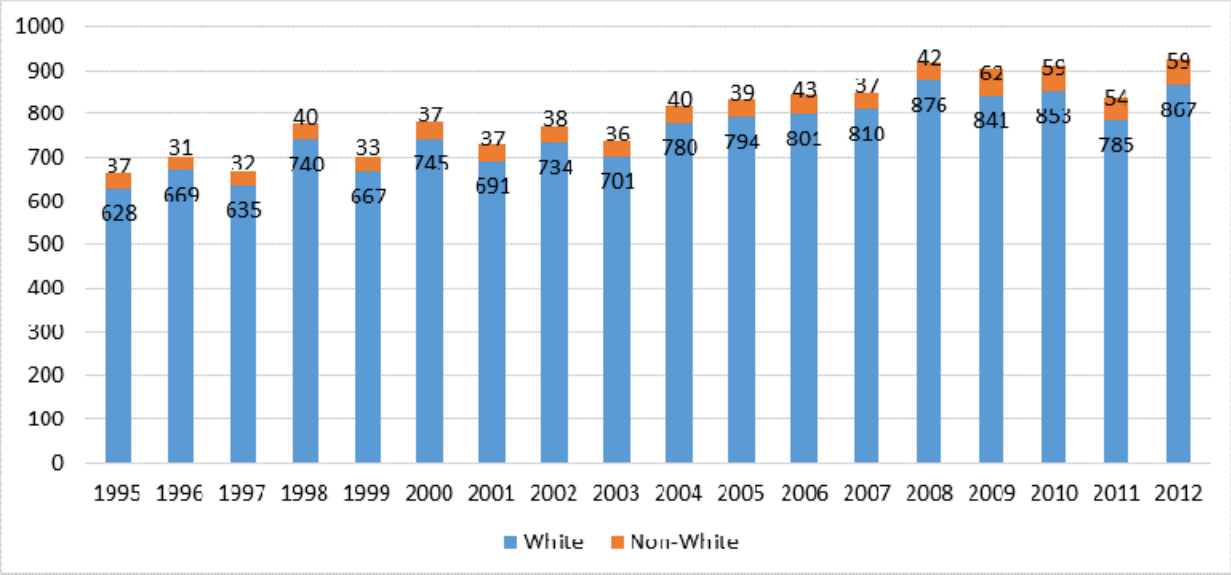


Figure 4. Number of new non-Hodgkin's Lymphoma Cases in Kentucky, 1995-2012, by Year and by Race

Figures 5 and 6 show the distribution of cases by age categories, separated by gender and tumor classification of intranodal (SEER classification 33041) and extranodal (SEER classification 33042) types. As expected, an age-related increase in NHL incidence was observed for both males and females, and for both SEER classifications, with a sharp increase in NHL for females ages 60-69. Intranodal NHL cases were consistently more than double the extranodal cases, across all age groups, and both showed sharp increases in the 60-69 age group.

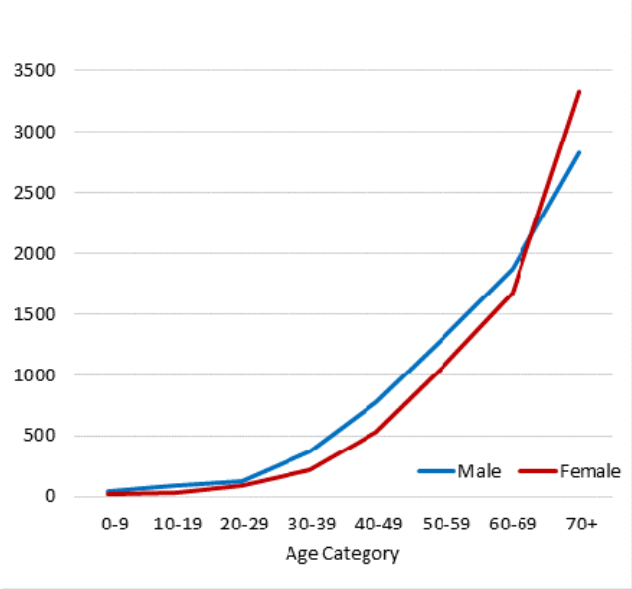


Figure 5. NHL by Sex and Age at Diagnosis

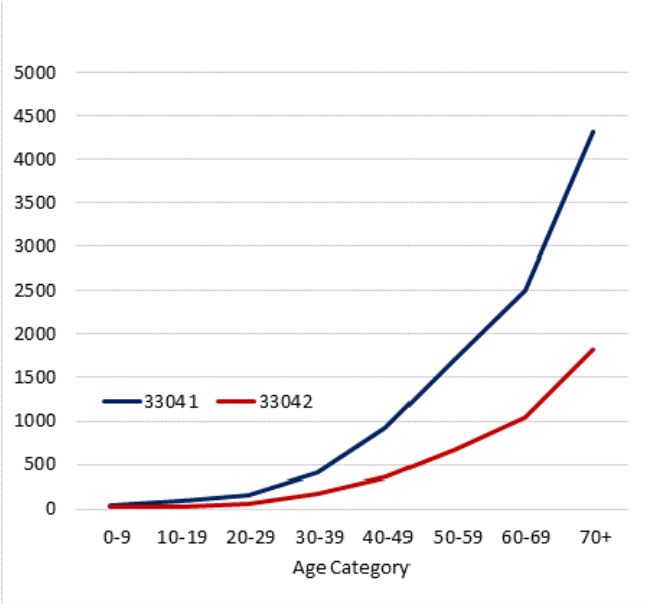


Figure 6. NHL by SEER Classification and Age at Diagnosis

Univariate analysis at the census tract level is summarized below. The mean population per 2010 Census tract in Kentucky was 4104.8 (standard deviation 1721.0), with a median value of 3920. The number of NHL cases per 100,000 in each census

tract had a mean value of 209.8 (standard deviation 335.7), and a median value of 28.5. It is apparent that some influential census tracts are driving the mean number of NHL cases higher, as evidenced by the stark difference between the mean and median values. Eighty-seven percent of all census tracts in Kentucky did not have a Superfund site within their borders, and the remaining 13% had between one and five sites per tract.

Bivariate analyses

Table 3 presents descriptive analyses for gender, race, residence in Appalachian regions, Beale Code, family history of NHL, primary SEER tumor type, and tobacco use, stratified into three categories of residential proximity to the nearest Superfund site. Although analyses by race were not conducted due to the low proportion of non-white cases, it is worth mentioning that non-white NHL patients were more likely to live within 5km of the Superfund sites, whereas residents of Appalachia and Beale Code designated rural areas were less likely to live near them. These results are not surprising given the distribution of the Superfund sites shown earlier in Figure 2; note that there is a dense concentration of sites in the area of the state with the highest percentage of African-American residents, and relatively few sites in the eastern and southern regions of Kentucky that are designated as Appalachian, and more likely to be classified as rural. Significant differences in categorical percentages also exist for tobacco use among cases, and in family history of NHL. The percentage of NHL cases with no family history of NHL, or without a known family history, were significantly higher for the cases residing within 5km of Superfund sites.

Table 3: Bivariate Analysis for the Case Data by Exposure

Demographic Variable		Residential Proximity to Nearest Superfund/NPL Site			Chi-square value (df)	P-value
		<5 km	5-10 km	>10 km		
Gender	Male	2170 (29.4%)	1793 (24.2%)	3432 (46.4%)	3.54 (2)	.170
	Female	2055 (29.4%)	1777 (25.5%)	3146 (45.1%)		
Race	White	3826 (28.1%)	3400 (25.0%)	6391 (46.9%)	234.04 (4)	<.001
	Non-white	351 (55.4%)	133 (21.0%)	150 (23.7%)		
Appalachia	No	3459 (33.5%)	3070 (29.7%)	3808 (36.8%)	1198.44 (2)	<.001
	Yes	766 (19.0%)	500 (12.4%)	2770 (68.6%)		
Beale Code Classification	Urban	4157 (32.0%)	3497 (26.9%)	5343 (41.1%)	1186.59 (2)	<.001
	Rural	68 (4.9%)	73 (5.3%)	1235 (89.8%)		
Family History of NHL	Yes	133 (25.0%)	130 (24.4%)	270 (50.7%)	9.94 (4)	<.001
	No	2234 (29.8%)	1817 (24.2%)	3444 (46.0%)		
	Unknown	1858 (29.3%)	1623 (25.6%)	2864 (45.1%)		
SEER Type	Intranodal	2969 (29.2%)	2547 (25.0%)	4665 (45.8%)	1.12 (2)	.572
	Extranodal	1256 (30.0%)	1023 (24.4%)	1913 (45.6%)		
Tobacco Use Status	Non-User	1716 (30.0%)	1479 (25.9%)	2520 (44.1%)	12.49 (4)	.014
	Tobacco User	1653 (29.4%)	1347 (23.9%)	2627 (46.7%)		
	Not recorded	856 (28.2%)	744 (24.5%)	1431 (47.2%)		

Because census tracts can vary in both population and area, it is important to compute and use in the analyes the rate of NHL incidence per 100,000 persons, rather than the number of cases. Also, to account for aging effects on health and for the differences in the number of and ages of residents across block groups, data was age-adjusted. The cumulative incidence rates for NHL per 100,000 persons from 1995-2012 in each census tract, age-adjusted using the 2000 US Census standard population, are depicted in Figure 7. Exposure and age-adjusted outcome data was aggregated at the levels of census tract, plus 5km buffers and 5-10km buffers around Superfund sites that

could contain parts of several census tracts. There are some areas of noticeably higher NHL incidence, scattered mostly in the western and south-central regions of Kentucky.

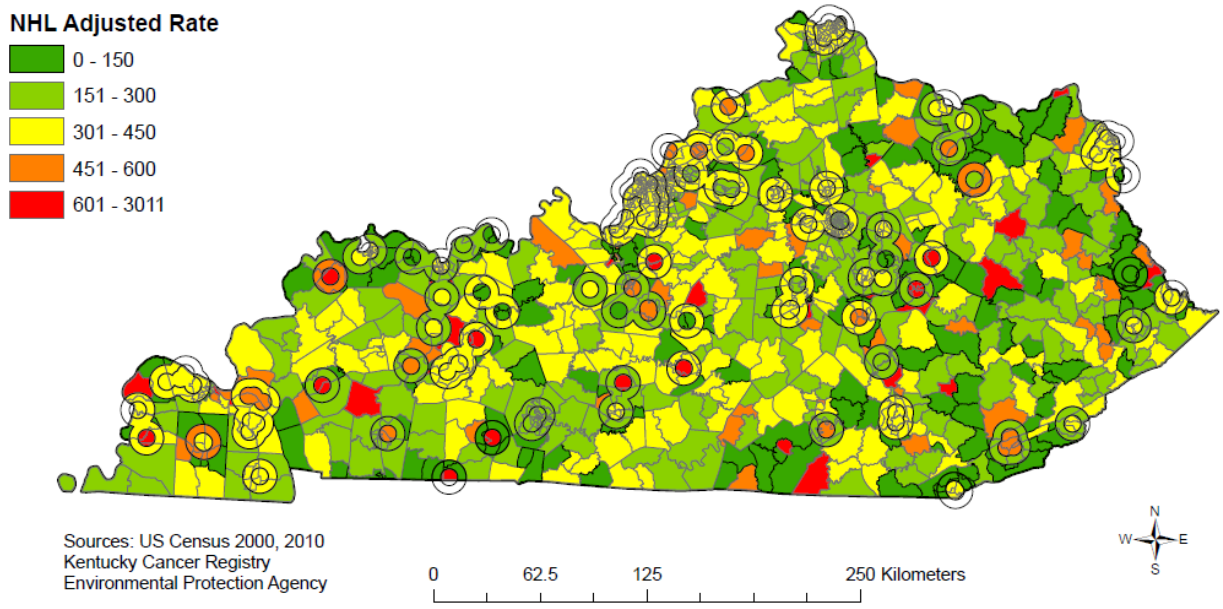


Figure 7. Overall Cumulative NHL Incidence Rate per 100,000 People (1995-2012)

One-way analysis of variance (ANOVA) was used to compare the average age-adjusted cumulative NHL incidence rates by exposure groups (<5km, 5-10km, over 10km). The incidence rates (overall, for males, females, for each SEER tumor classification, and for each SEER classification by gender) were significantly greater within 5km exposure areas than in the other two groups; further, the rates within 5km and 10km from the Superfund sites were significantly greater than the rates in the unexposed areas. In almost all strata, the means in the unexposed group are significantly lower than those in the exposed groups; at a significance level of $p<0.05$,

the only category in which the exposed and unexposed group NHL rates did *not* differ significantly was for the average incidence rate for the intranodal NHL in females (p=.064) . The data reflects the national trends, in that the male patients have a higher incidence rate than females for both intranodal and extranodal NHL (Table 4).

Table 4: Age-Adjusted 1995-2012 Cumulative NHL Incidence Rates by Exposure Group

	Age-adjusted NHL Incidence Rates	Incidence Rate by Exposure Group Mean (std dev)			ANOVA	
		<5km	5-10km	>10km	F statistic	p-value
Overall		457.0 (244.7)	308.6 (100.6)	290.9 (215.7)	17.8	<.001
Gender	Male	542.4 (341.2)	338.3 (113.3)	325.8 (249.5)	21.6	<.001
	Female	382.9 (240.2)	285.3 (116.7)	262.4 (303.6)	5.1	.006
SEER Type	Intranodal	323.4 (200.2)	218.7 (73.3)	208.5 (180.6)	12.3	<.001
	Extranodal	133.7 (82.8)	89.9 (49.6)	82.5 (76.6)	13.4	<.001
Gender * SEER Type	Intranodal -Males	384.1 (294.8)	239.7 (89.6)	235.8 (196.6)	15.8	<.001
	Intranodal - Females	267.7 (215.1)	202.4 (84.6)	185.9 (281.1)	2.8	.064
	Extranodal - Males	158.3 (154.3)	98.6 (60.2)	90.0 (102.1)	12.3	<.001
	Extranodal - Females	115.2 (83.5)	82.8 (60.3)	76.5 (97.4)	5.0	.007

In addition, independent samples t-tests were used to compare the mean age-adjusted incidence rates between Appalachia and Non-Appalachia regions (Table 5). None of the incidence rates were significantly different between Appalachia and non-Appalachia. This finding is important because, for most types of cancer, Appalachia is generally known to have significantly higher incidence rates.

Table 5: Age-Adjusted 1995-2012 Cumulative NHL Incidence Rates by Region

	Age-Adjusted NHL Incidence Rates	Mean Number of Cases (std dev)		Independent t-test	
		Appalachia	Non-Appalachia	T-statistic	p-value
Overall		305.8 (184.0)	308.0 (261.2)	-0.137	.891
Gender	Male	338.1 (275.4)	350.9 (249.5)	0.654	.513
	Female	283.3 (388.3)	268.4 (203.6)	-0.676	.499
SEER Type	Intranodal only	222.4 (226.0)	217.1 (142.5)	-0.393	.695
	Extranodal only	85.6 (79.8)	88.7 (75.4)	0.533	.594
Gender * SEER Type	Intranodal for Males	246.0 (219.6)	251.4 (197.7)	0.344	.731
	Intranodal for Females	202.3 (371.2)	188.6 (167.6)	-0.675	.500
	Extranodal for Males	92.1 (103.3)	99.5 (110.2)	0.921	.357
	Extranodal for Females	81.0 (102.1)	79.8 (89.9)	-0.172	.863

Multivariate Analyses

Spatial regression models were developed using the ArcMap software. For multivariate analysis, the exposure was measured with two dummy variables: exposure within 5km (yes=1/no=0), and exposure between 5km and 10km (yes=1/no=0), with exposure beyond 10km as the reference group for the analysis. Because the outcome and exposure data have a spatial dimension, it was important to determine whether spatial autocorrelation existed in the data. Exploratory regression and diagnostic tests were conducted to verify whether geographically weighted regression (GWR) would be necessary, or if standard ordinary least squares (OLS) regression modeling alone would suffice. Testing was also performed to determine whether predictor effects on the outcome were consistent across the studied area (stationarity), as an additional determination whether GWR would be necessary.

Data characteristics from the individual level such as race, smoking status, and family history of NHL, which were compared across exposure categories in bivariate

analysis, were not available at the census-tract level and thus were not included in multivariate models. Two other individual-level data characteristics (gender and SEER tumor type) were controlled at the multivariate level by calculating age-standardized NHL incidence rates by gender and SEER tumor type, along with the overall rate. Exposure categories, Appalachian status, Beale Code, and a series of population and housing characteristics at the census tract level were considered for inclusion in multivariate models. Multicollinearity coefficients obtained during the exploratory regression steps showed that most of the population and housing characteristics at the census tract level exhibited significant multicollinearity and could not be included in the same model (Variance Inflation Factors >10). Furthermore, none of these tract-level characteristic variables had a significant effect on NHL incidence rates, so they were not considered for subsequent regression modeling. There was no multicollinearity between Appalachian status and Beale Code, so both variables were considered for regression modeling. See Appendix 2: Exploratory Regression Output.

The Global Moran's I tool in ArcGIS was used to quantify the presence of spatial autocorrelation among residuals (Table 6). Overall rates were examined, along with rates by gender, SEER tumor classification, and by both gender and SEER classfication. All showed significant and positive Z-scores, and thus significant autocorrelation and clustering of similar residual values.

The Anselin's local Moran's I tool (also known as the Local Indicators of Spatial Association or LISA) in ArcGIS also confirmed the presence of autocorrelation,

Table 6: Tests for Spatial Autocorrelation of Outcome Data Using Global Moran's I

Stratum		Global Moran's I	Z-score	P-value
Overall		0.033592	7.303229	<.001
	Intranodal	0.023284	5.168179	<.001
	Extranodal	0.058884	12.61339	<.001
Male		0.038898	8.384161	<.001
	Intranodal	0.028624	6.203308	<.001
	Extranodal	0.051379	11.03733	<.001
Female		0.02186	5.121112	<.001
	Intranodal	0.010848	2.720941	0.007
	Extranodal	0.053696	11.53571	<.001

clustering, and spatial outliers. Figure 8 depicts, for the overall NHL incidence data, the existence of multiple geographic areas where significant high and low clustering of NHL rate data exist, along with areas where significant spatial outliers occur (e.g. low-NHL local areas adjacent to high-NHL areas, or vice versa). The pattern is similar to what was observed in the overall incidence data from Figure 7, with high clusters in the western and central regions, and low clusters and high-low outliers predominate in the eastern and southern regions that are also designated as Appalachia.

The clustering and outlier areas of NHL incidence data stratified by gender, SEER classification, and both gender and SEER classification are collectively depicted in Figures 9 through 11. Each of these maps also confirm that spatial autocorrelation and clustering exist in this data. For the most part, these maps show a profile similar to the overall NHL incidence data, with the notable exception that a cluster of extranodal NHL incidence in females was observed in the Appalachian region of Ashland-Boyd County in northeastern Kentucky, an area with abundant petrochemical industries and Superfund sites.

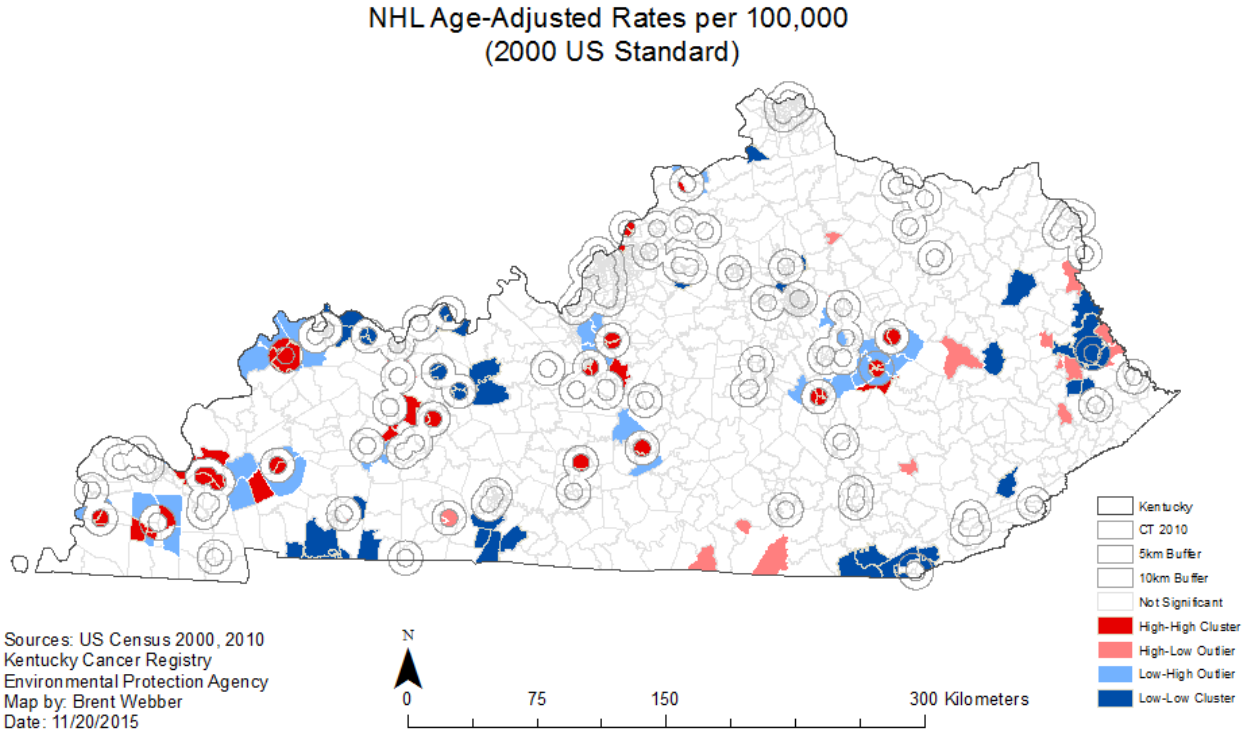


Figure 8. Anselin's Local Indicators of Spatial Association (LISA) for Overall Cumulative NHL Incidence Data

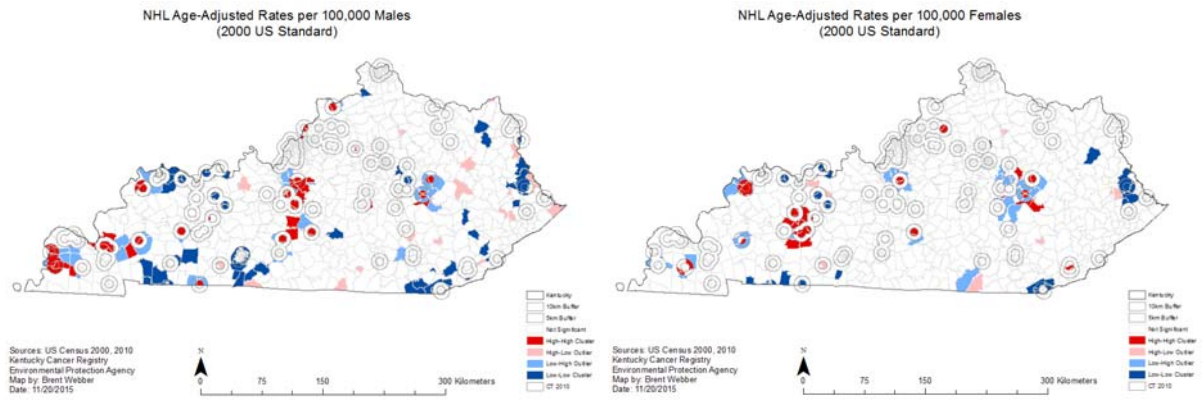


Figure 9. Anselin's LISA for Cumulative NHL Incidence Data by Gender Strata

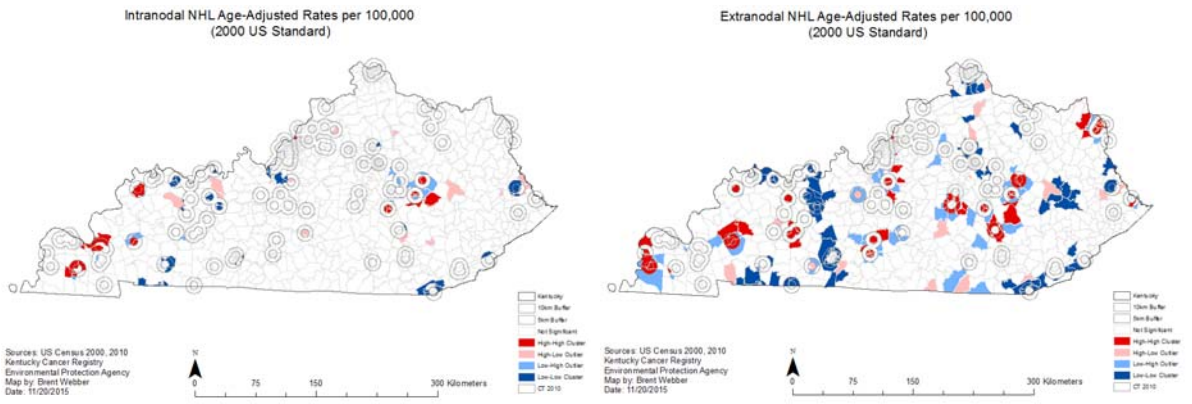


Figure 10. Anselin's LISA for Cumulative NHL Incidence Data by SEER Tumor Classification Code

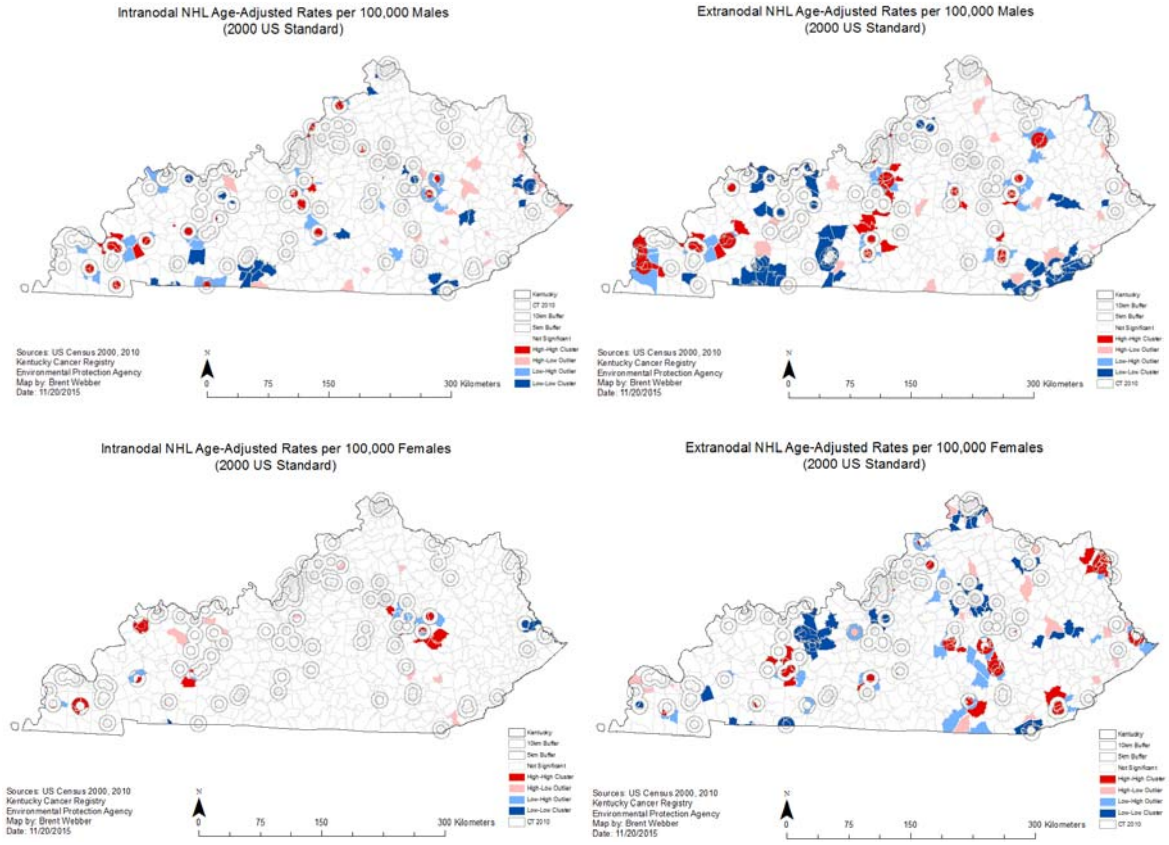


Figure 11. Anselin's LISA for Cumulative NHL Incidence Data by Gender and SEER Tumor Classification Code

The hot spot analysis was conducted to identify areas of significant high or low spatial clustering of NHL incidence data using the Getis-Ord G_i^* statistic. This differs from the Anselin's LISA analysis in that each feature (e.g. census tract/buffer zone) and its neighboring features are compared to the average of all features on the map, whereas Anselin's LISA only compares values from each feature to contiguous neighboring features. Hot and cold spots were mapped at the 99%, 95%, and 90% confidence limits. Results for overall data are depicted in Figure 12, and results by gender and SEER tumor classification are shown in Figures 13 and 14, respectively. There are interesting differences between the maps, as the profile of hot and cold spots is not necessarily uniform between overall data, data by gender, and data by SEER classification. As with the Anselin LISA tests, the overall data map shows the hot spots being located predominantly in the western and central parts of the state. When data is split by gender, the hot spots are more prominent for male cases, particularly the spots in the western region of the state, and the central region of the state corresponding to metro Louisville and Hardin County. When data is split by SEER classification, the western and metro Louisville hot spots become less prominent, and for extranodal NHL a hot spot emerges in Ashland-Boyd County. Collectively, these maps present further evidence of the need to utilize GWR in modeling the effect.

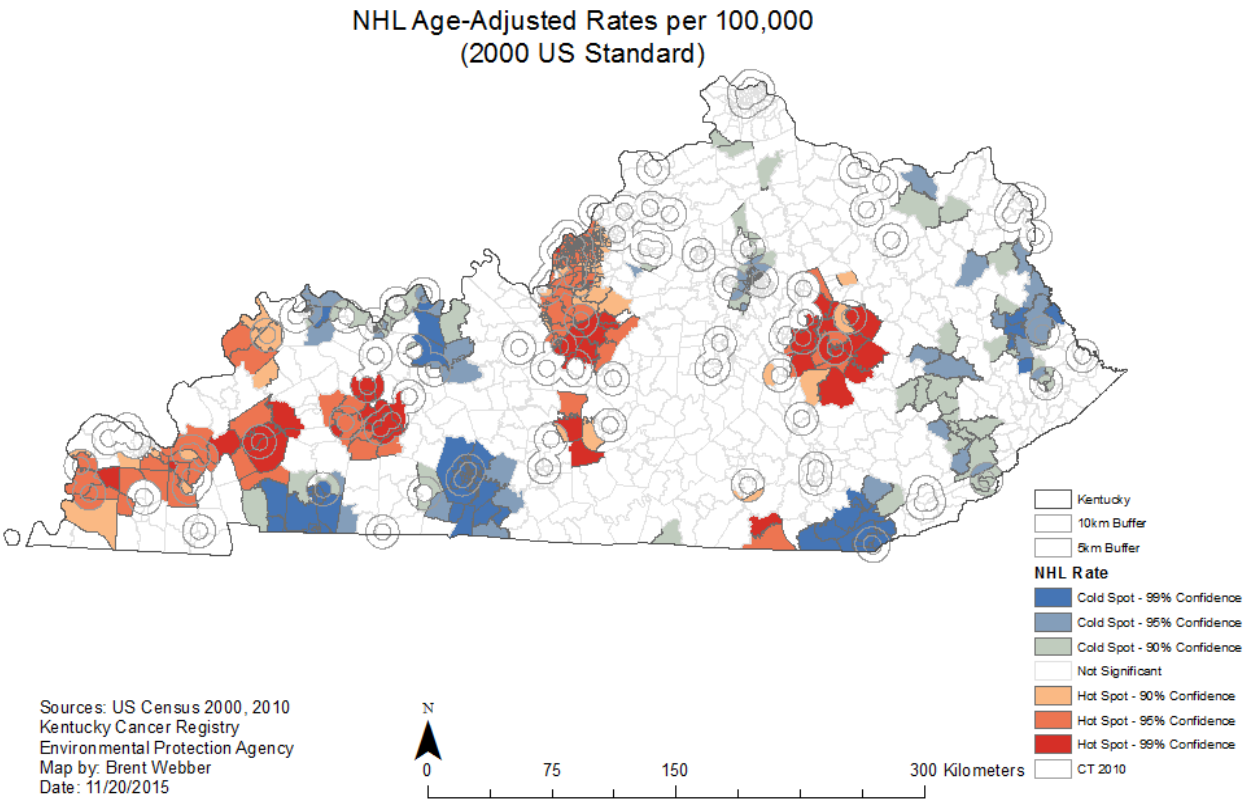


Figure 12. Hot Spot Analysis for Cumulative NHL Incidence Data – Overall

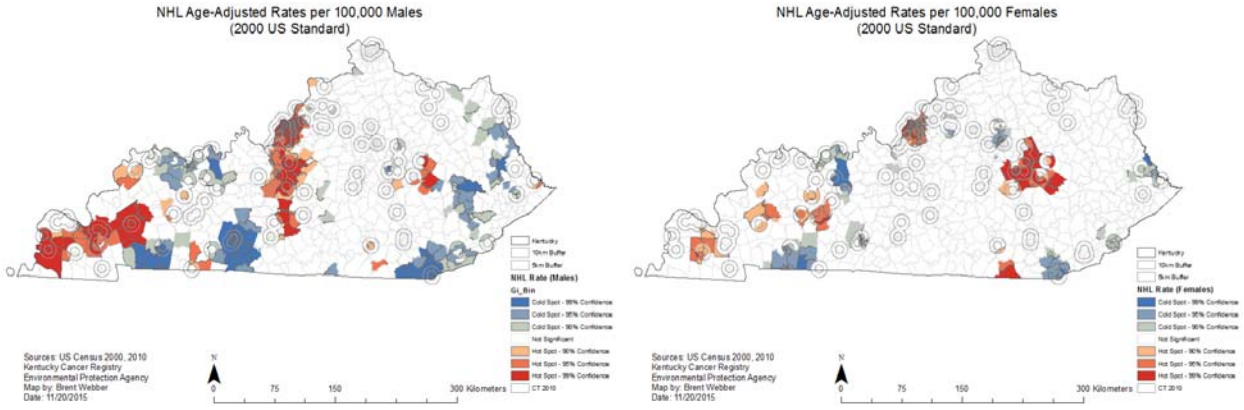


Figure 13. Hot Spot Analysis for Cumulative NHL Incidence Rates by Gender

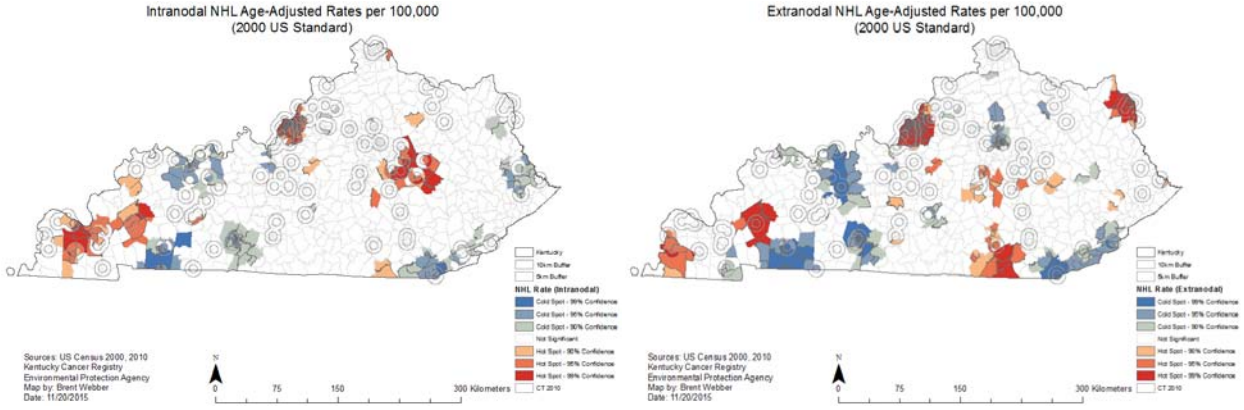


Figure 14. Hot Spot Analysis for Cumulative NHL Incidence Rates by SEER Tumor Classification Code

Regression Analysis

The ordinary least squares (OLS) regression predicts the age-adjusted NHL cumulative incidence rates per 100,000 people using the two variables measuring exposure (<5km from nearest Superfund site, 5-10km from nearest Superfund site), along with Appalachian region (yes=1/no=0), and Beale Code (1 to 9) as independent variables. OLS modeling was performed for the cumulative incidence rate per 100,000 people for total NHL, intranodal NHL (SEER code 33041), extranodal NHL (SEER code 33042), and incidence rates for total NHL per 100,000 males and females. For each category, a base model was calculated with only the two exposure variables, and a full model which added the Appalachia and Beale Code variables. This resulted in a total of ten OLS models. Key outputs are summarized in Table 7. For the full list of OLS outputs and diagnostics, see Appendix 3.

Specificially, the analysis showed that both exposure categories had significant and high coefficient values in all ten OLS models, with higher value in the category closer to the Superfund sites (Exposure <5km). Compared to the reference group (exposure beyond 10km), the <5km groups had significantly more NHL cases per 100,000, ranging from 35.9 additional cases/100k in Model 9 (the base extranodal NHL model) to 157.92 additional cases/100k in Model 4 (the full male cases NHL model), when all other variables were held constant. Compared to the reference group, the 5-10km exposure groups also had significantly more NHL cases/100,000 ranging from 15.60 in Model 9 to 65.03 in Model 4.

Table 7: OLS Regression Modeling Results

Model	Variables	Coefficient	Std. Error	Prob.	Wald ($Pr>\chi^2$)	Koenker (BP) ($Pr>\chi^2$)	AIC	Adj R-squared
Model 1:	Intercept	261.26	8.43	<.001				
Overall Cumulative	Exp <5km	113.60	10.76	<.001	121.04	14.20	25137.29	0.070
Incidence Rate (CIR)	Exp 5-10km	50.82	8.98	<.001	(<.001)	(<.001)		
	Intercept	239.54	9.93	<.001				
Model 2:	Appalachia	-2.55	13.36	0.85				
Overall CIR	Beale Code	5.80	1.93	0.003	125.89	20.08	25127.17	0.076
	Exp <5km	123.82	11.60	<.001	(<.001)	(<.001)		
	Exp 5-10km	57.19	9.11	<.001				
Model 3:	Intercept	292.56	9.70	<.001				
Male CIR	Exp <5km	147.43	13.01	<.001	137.81	29.59	25825.84	0.082
	Exp 5-10km	59.19	10.36	<.001	(<.001)	(<.001)		
	Intercept	269.53	11.61	<.001				
Model 4:	Appalachia	-23.89	14.36	0.10				
Male CIR	Beale Code	8.28	2.38	<.001	141.04	47.17	25814.09	0.089
	Exp <5km	157.92	14.17	<.001	(<.001)	(<.001)		
	Exp 5-10km	65.03	10.52	<.001				
Model 5:	Intercept	235.61	11.30	<.001				
Female CIR	Exp <5km	85.05	13.10	<.001	50.01	4.20	25938.92	0.027
	Exp 5-10km	44.61	11.80	<.001	(<.001)	(0.12)		
	Intercept	215.19	13.13	<.001				
Model 6:	Appalachia	16.51	18.60	0.38				
Female CIR	Beale Code	3.56	2.49	0.15	66.92	7.25	25933.66	0.030
	Exp <5km	94.96	13.91	<.001	(<.001)	(0.12)		
	Exp 5-10km	51.42	11.80	<.001				
Model 7:	Intercept	187.19	6.91	<.001				
Intranodal CIR	Exp <5km	77.70	8.67	<.001	89.27	7.81	24294.79	0.052
	Exp 5-10km	35.21	7.28	<.001	(<.001)	(0.02)		
	Intercept	173.59	8.27	<.001				
Model 8:	Appalachia	-6.91	11.37	0.54				
Intranodal CIR	Beale Code	4.17	1.57	0.01	92.62	12.30	24289.26	0.055
	Exp <5km	84.01	9.42	<.001	(<.001)	(0.02)		
	Exp 5-10km	38.98	7.39	<.001				
Model 9:	Intercept	74.07	2.91	<.001				
Extranodal CIR	Exp <5km	35.90	3.64	<.001	107.84	39.73	21117.27	0.057
	Exp 5-10km	15.60	3.27	<.001	(<.001)	(<.001)		
	Intercept	65.95	3.52	<.001				
Model 10:	Appalachia	4.35	4.17	0.30				
Extranodal CIR	Beale Code	1.64	0.72	0.02	119.96	71.29	21104.74	0.064
	Exp <5km	39.81	3.95	<.001	(<.001)	(<.001)		
	Exp 5-10km	18.21	3.37	<.001				

The Beale Code was also a significant predictor for the NHL cumulative incidence in most of the full models: the only full model in which it was not significant at p=0.05 was Model 6 (female cases, p=0.15). Being a resident of the Appalachia region was not a significant predictor in any of the OLS models; furthermore, the effect of living in the Appalachia region was not consistent across models, as the regression coefficients were negative in Models 2, 4, and 8 (the all-cases, male, and intranodal case models, respectively), and positive in Models 6 and 10 (female and extranodal case models, respectively). Full interpretation of the OLS model regression coefficients is provided below.

Interpretation of OLS Model Coefficients: Model 4

For the OLS model which had the highest coefficient of determination (Model 4, with an adjusted R^2 value of 0.089) and thus represented the best-fitting of the OLS models, the regression equation is:

*18-year Cumulative Incidence of NHL /100,000 males = 269.53 - 23.89*Appalachia + 8.28*BealeCode + 157.92*Exp<5km + 65.03*Exp5-10km.*

The reference group is made up of male cases who reside in Kentucky, outside the Appalachia region, and within an urban metro area with a population of one million or more (Beale Code = 0), and at least 10km away from all Superfund sites. Thus, the intercept shows that the overall NHL incidence rate (DV) for the reference group is 269.53/100,000 males.

The variable *Appalachia* measures the difference in the NHL cumulative incidence rate per 100,000 persons between the people residing in the Appalachia

region and the people residing elsewhere in the state. For males, the regression coefficient is -23.89 indicating that Appalachia region has on average lower NHL cumulative incidence rates than the rest of the state for males. The 18-year cumulative incidence rate for males was lower than in the reference group, and equaled (269.53 - 23.89) = 245.64 cases/100,000 males. Thus, the regression coefficient (b=-23.89, p=.010) shows that the Appalachian region in Kentucky has on average an overall NHL cumulative incidence rate for males that is 23.89 cases/100,000 lower than the non-Appalachian Kentucky, when all other variables in the model are held constant; however, this regression coefficient was not significantly different from zero, and can be approximated to be zero in the calculations. In addition, the Appalachia variable was not significantly different from zero in any of the other full OLS models, and had a positive coefficient value in the female and extranodal case models (Models 6 and 10, respectively).

The variable *Beale Code* measures urbanicity and rurality. Values range from 0 to 9; the smaller the code value, the more urban an area. The Beale Code value of 0 is assigned for residence within a metropolitan area of population 1,000,000 or more, and Beale Code 9 is assigned for residence within a rural area (population <2500) not adjacent to any urban areas (population >2500). The regression coefficient (b=8.28, p=<.001) in the best-fitting OLS model (Model 4) shows that as the Beale Code increases by 1 unit, the overall cumulative male NHL incidence rate increases by 8.28 NHL cases/100,000, when all other variables in the model are held constant; thus, the more rural an area is the higher the overall cumulative NHL incidence rate. For the most rural areas, where Beale Code = 9, the incidence rate numerator equals [269.53 +

(8.28*9)] = 344.05 NHL cases/100,000, an increase of 74.52 cases per 100,000 males above the rate for the reference group. The most urban Beale Code value in the study basin was 1, as there were no central metropolitan areas with population >1 million in Kentucky. When Beale Code = 1, the incidence rate numerator equals [269.53 + (8.28*1)] = 277.81 NHL cases/100,000, an increase of 8.28 cases/100,000 males above the reference group. The Beale Code regression coefficient was significantly different from zero ($p<.05$) in the best-fitting model and in all other OLS models except for the female cases model (Model 6, $p=0.15$).

The regression coefficients for the exposure variables show the difference in the NHL cumulative incidence rate per 100,000 between the reference group and cases who reside within 5km from one or more Superfund sites (Exp<5km), or more than 5km away but within 10km of one or more Superfund sites (Exp5-10km). For the best-fitting OLS model (Model 4), when examining residence within 5km of Superfund sites ($b=157.92$, $p<.001$), the NHL incidence rate per 100,000 males equals (269.53 + 157.92) = 427.45 NHL cases/100,000. For residence within 5km and 10km of one or more Superfund sites ($b=50.82$, $p<.001$), the male NHL cumulative incidence rate equals (269.53+ 65.03) = 334.56 NHL cases/100,000. Both coefficients are significantly different from zero, but the Exp<5km exposure variable has a much larger effect on NHL incidence rates.

Interpretation of OLS Model Coefficients: Model 2

For all the OLS models, the exposure variables have significant and positive regression coefficients ($p<.001$), with the <5km variables having higher values than the

5-10km variables. For Model 2, which is the best-fitting model that includes **all** NHL cases in Kentucky from 1995-2012, rather than subdividing by gender or SEER type, the regression equation is:

*18-year Cumulative Incidence of Intranodal NHL /100,000 = 239.54 + (-2.55)*Appalachia + 5.80* Beale Code + 123.82*Exp<5km + 57.19*Exp5-10km.*

The reference group is made up of NHL cases who do not reside in the Appalachia region, and reside within an urban metro area with a population of one million or more (Beale Code = 0), and at least 10km away from all Superfund sites. Thus, the intercept shows that the cumulative NHL incidence rate per 100,000 people (DV) for the reference group is 239.54.

As with Model 4, the coefficient *Appalachia* in Model 2 is negative but not significant (b=-2.55, p=.85). The incidence rate was lower than in the reference group, and equaled (239.54 – 2.55) = 236.99 cumulative NHL cases/100,000, or 2.55 cases/100,000 lower than non-Appalachian Kentucky residents, when all other variables in the model are held constant.

The *Beale Code* coefficient is positive and significant in Model 2 (b=5.80, p=.003). For the most rural areas (Beale Code = 9), the cumulative intranodal NHL incidence rate is [239.54 + (5.80*9)] = 291.74 cases/100,000, an increase of 52.2 cases/100,000 above the reference group. For the most urban areas in the study basin (Beale Code = 1), the increase above the reference group is 5.80 cases/100,000, for a total of 245.34 cases/100,000, when all other variables are held constant.

Both of the exposure variables are positive and highly significant in Model 2, but the Exp<5km variable has a much larger effect on cumulative NHL incidence rates. For residence within 5km of Superfund sites (b=123.82, p<.001), the cumulative NHL incidence rate per 100,000 equals (239.54 + 123.82) = 363.36 NHL cases/100,000. For residence >5km but within 10km of one or more Superfund sites (b=57.19, p<.001), incidence rate equals (239.54 + 57.19) = 296.73 NHL cases/100,000.

Other OLS Models

The remaining OLS models (which include every base model, plus the full models that separately evaluated female, intranodal, and extranodal cases) had lower coefficients of determination compared to the best-fitting models, indicating a poorer fit around the regression line. None of the OLS models in Table 7 explained a large amount of the variability around the fitted regression line, with the coefficients of determination ranging from 2.7% to 8.9% (Appendix 3). The OLS models had acceptable levels for the variance inflation coefficients, but significant Koenker (BP) statistics in almost all models indicate non-consistent relationships between the dependent and independent variables (non-stationarity); thus, a geographically weighted regression (GWR) is more likely to be appropriate than the OLS models.

Geographically Weighted Regression

GWR was the final stage of analysis. As with OLS models, the explanatory variables included the two exposure groups, Appalachia region (1=yes, 0=no), and Beale Code (numeric code, higher for rural areas). Adaptive kernel density estimation

was utilized, along with the Akaike Information Criterion (corrected) to estimate bandwidth.

Base models (exposure groups only) and full models (exposure groups plus Appalachia plus Beale Code) were generated for five different age-adjusted NHL incidence rates: all cases, male cases, female cases, intranodal cases, and extranodal cases, for a total of ten GWR models. These ten models were labeled as Models 11 through 20, and the outputs for each of these GWR models are listed in Table 8.

The Akaike's Information Criterion (AIC) values were compared between each GWR models and their analagous OLS models from Table 7. For example, Model 11 in GWR was compared to Model 1 in OLS, since both used the same subset of data and predictor variables. In each case, for all ten model pairs, the AIC values were lower for the GWR models compared to their OLS counterparts, indicating that the GWR models were a better fit for the data. Comparing the adjusted R-squared values from the OLS models (Table 7) to the unadjusted R-squared values in GWR models (Table 8) also makes it apparent that the GWR models represent a better fit around the regression line, and explain a larger percentage of the variability. For the OLS models, the coefficients of determination ranged from 2.7% to 8.9%, whereas for the GWR models, these ranged from 6.6% to 24.6%. Adjusted R-squared values should not be used to make inferences about the proportion of variance explained by GWR models, since these values are sensitive to bandwidths used to calculate degrees of freedom[150]. The R-squared results are in agreement with the AIC and confirm that, when looking at all NHL cases in the data set, the best-fitting model of the set is Model 11 (the GWR base model), which explains approximately 23.1% of the variability in the overall NHL

incidence rate. When looking at subsets of NHL cases, the best-fitting model is Model 13 (the GWR base model for males), which explains approximately 24.6% of the variability in NHL incidence rate for male subjects, and represents the overall best-fitting model of all GWR and OLS models.

Table 8: GWR Modeling Results

Model	Variables	Number of Neighbors	Sigma	Akaike's Information Criterion	R-square
Model 11: Overall CIR	Exp <5km Exp 5-10km	241	155.81	24893.80	0.231
Model 12: Overall CIR	Appalachia Beale Code Exp <5km Exp 5-10km	834	163.24	25047.16	0.134
Model 13: Male CIR	Exp <5km Exp 5-10km	241	185.75	25569.11	0.246
Model 14: Male CIR	Appalachia Beale Code Exp <5km Exp 5-10km	834	194.50	25720.15	0.152
Model 15: Female CIR	Exp <5km Exp 5-10km	241	196.84	25791.99	0.154
Model 16: Female CIR	Appalachia Beale Code Exp <5km Exp 5-10km	836	204.10	25905.25	0.066
Model 17: Intranodal CIR	Exp <5km Exp 5-10km	241	125.65	24067.25	0.209
Model 18: Intranodal CIR	Appalachia Beale Code Exp <5km Exp 5-10km	834	131.76	24224.08	0.107
Model 19: Extranodal CIR	Exp <5km Exp 5-10km	241	55.10	20900.00	0.210
Model 20: Extranodal CIR	Appalachia Beale Code Exp <5km Exp 5-10km	834	57.01	21005.11	0.132

When non-stationarity is present in the data, GWR is designed to allow local variation in the explanatory variable coefficients. For Model 11, the all-cases best-fitting GWR base model, the coefficient values for the Exp <5km and Exp 5-10km variables are listed and depicted in Figure 15. The mean value for the Exp <5km coefficient was 120.67 (standard deviation 84.48, t-statistic=62.59, p<.001), and for Exp 5-10km it was 45.94 (standard deviation 66.35, t-statistic=30.37, p<.001). Therefore, the average increase in cumulative NHL incidence rate per 100k above the baseline condition (Exp >10km) in areas within 5km of Superfund sites in Kentucky was 121 cases/100k, and in areas between 5km and 10km from Superfund sites it was 46 cases/100k.

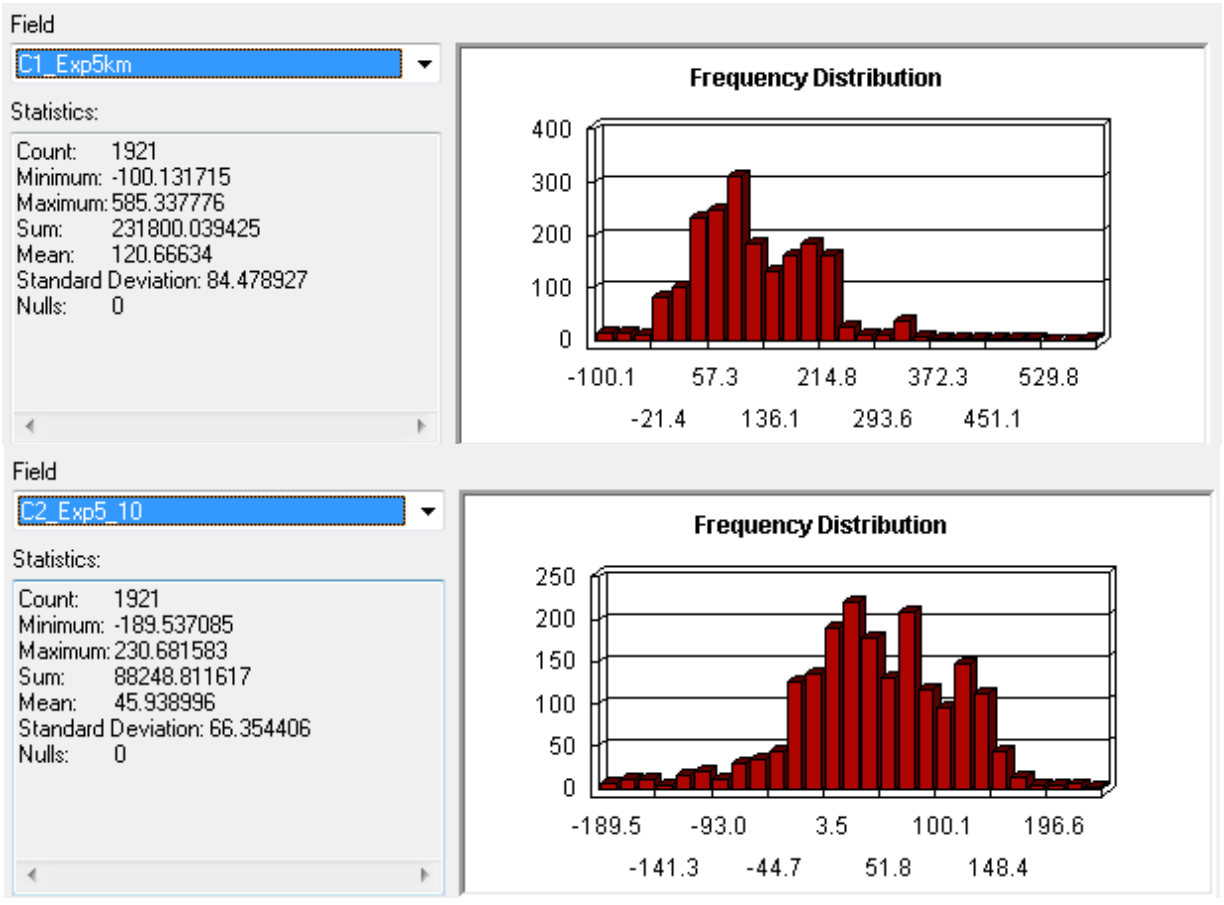


Figure 15: GWR Coefficient Values for All-Cases Best-Fitting Model

For the overall best-fitting model (Model 13, the GWR base model for all NHL cases in male subjects), the coefficient values for the Exp <5km and Exp 5-10km variables are listed and depicted in Figure 16. The mean value for the Exp <5km coefficient was 159.93 (standard deviation 104.01, t-statistic=67.39, p<.001), and for Exp 5-10km it was 58.59 (standard deviation 69.70, t-statistic=36.84, p<.001). Residence within 5km of a Superfund site in Kentucky increased the cumulative NHL cases per 100,000 males above the baseline by 160, and residence 5-10km away increased cases/100,000 males by 59, when all other variables were held constant.

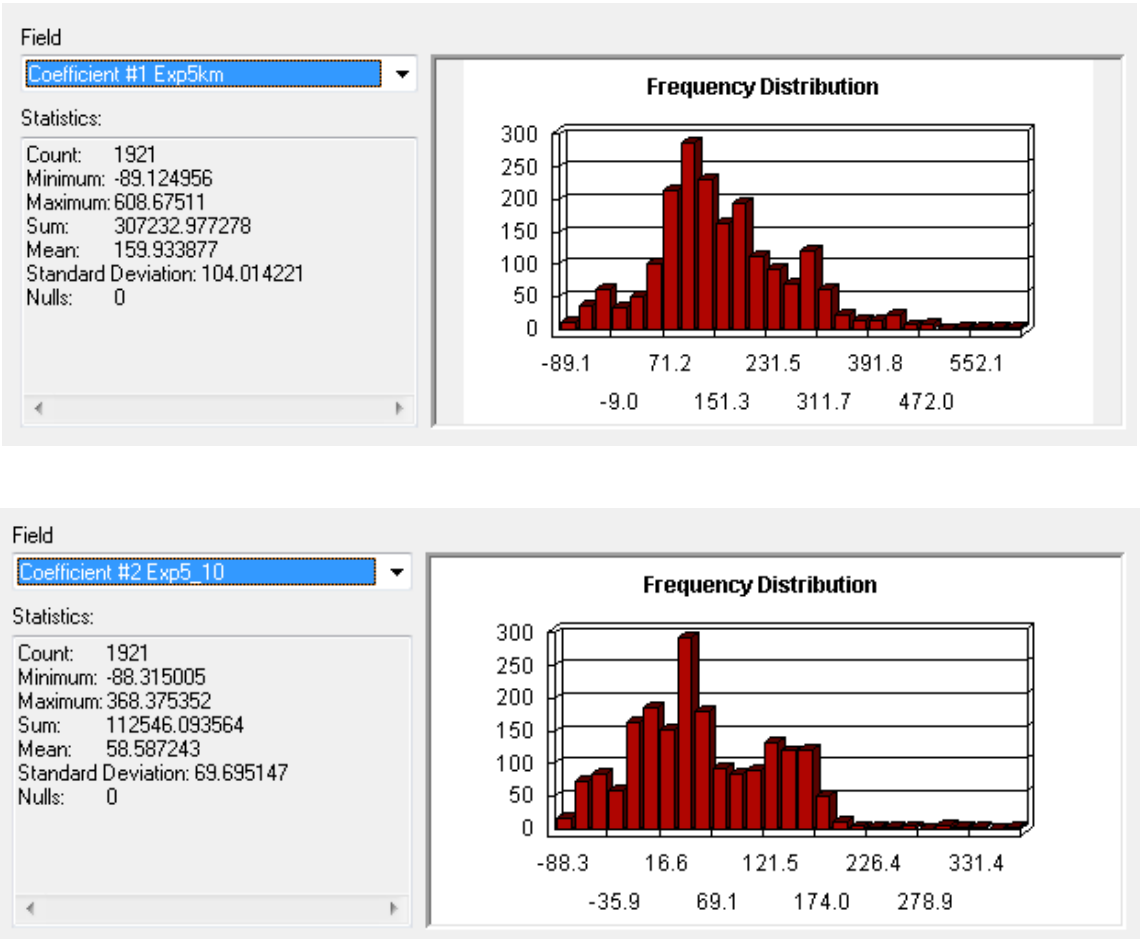


Figure 16: GWR Coefficient Values for Male Cases (and Overall) Best-Fitting Model

The coefficient values for the best-fitting models that examined only female, intranodal, and extranodal NHL cases are depicted below in Figures 17 through 19. Only the “base” models are shown, since in each case the “full” models had higher AIC values and lower R^2 values. In all three models, both of the exposure variables were positive and significantly different than zero.

Figure 17 shows that, for female cases (Model 15), the mean value for the Exp <5km coefficient was 85.86 (standard deviation 88.33, t-statistic=42.60, p<.001), and for Exp 5-10km it was 33.95 (standard deviation 80.67, t-statistic=18.45, p<.001). Figure 18 shows that, for intranodal cases (Model 17), the mean value for the Exp <5km coefficient was 80.95 (standard deviation 67.01, t-statistic=52.95, p<.001), and for Exp 5-10km it was 30.04 (standard deviation 55.60, t-statistic=23.68, p<.001). Figure 19 shows that, for extranodal cases (Model 19), the mean value for the Exp <5km coefficient was 39.71 (standard deviation 24.84, t-statistic=70.07, p<.001), and for Exp 5-10km it was 15.90 (standard deviation 25.15, t-statistic=27.71, p<.001). The results depicted in Figures 15 through 19 demonstrate that, in both the overall NHL data set and subsets of data by gender and SEER tumor classification, residential proximity of less than 5km from the nearest Superfund site, or between 5km and 10km from the nearest Superfund site, is a significant predictor of NHL incidence.

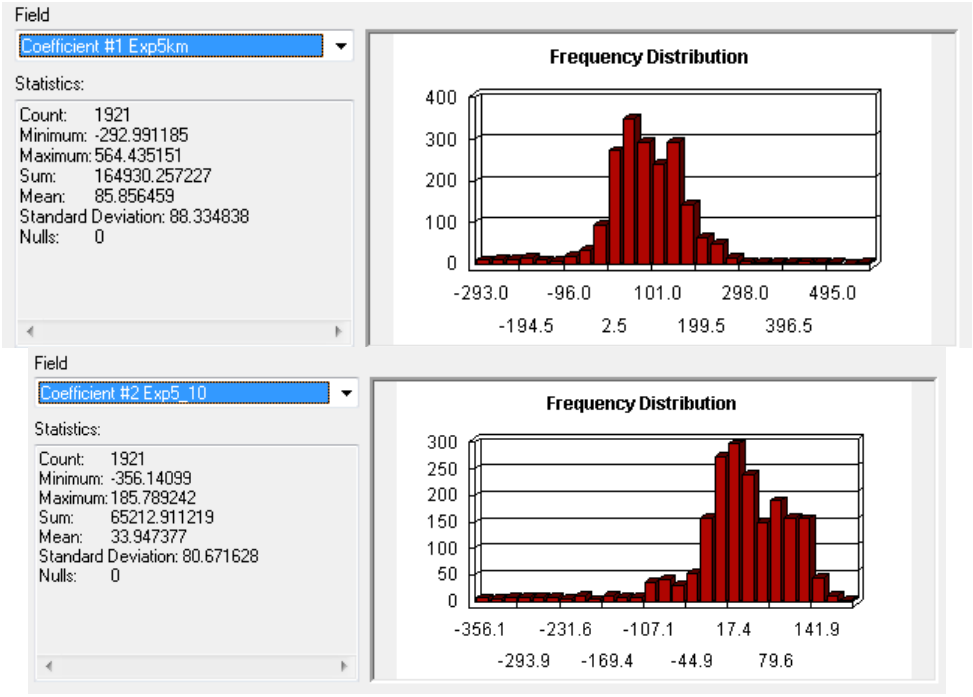


Figure 17: GWR Coefficient Values for Female Cases Best-Fitting Model

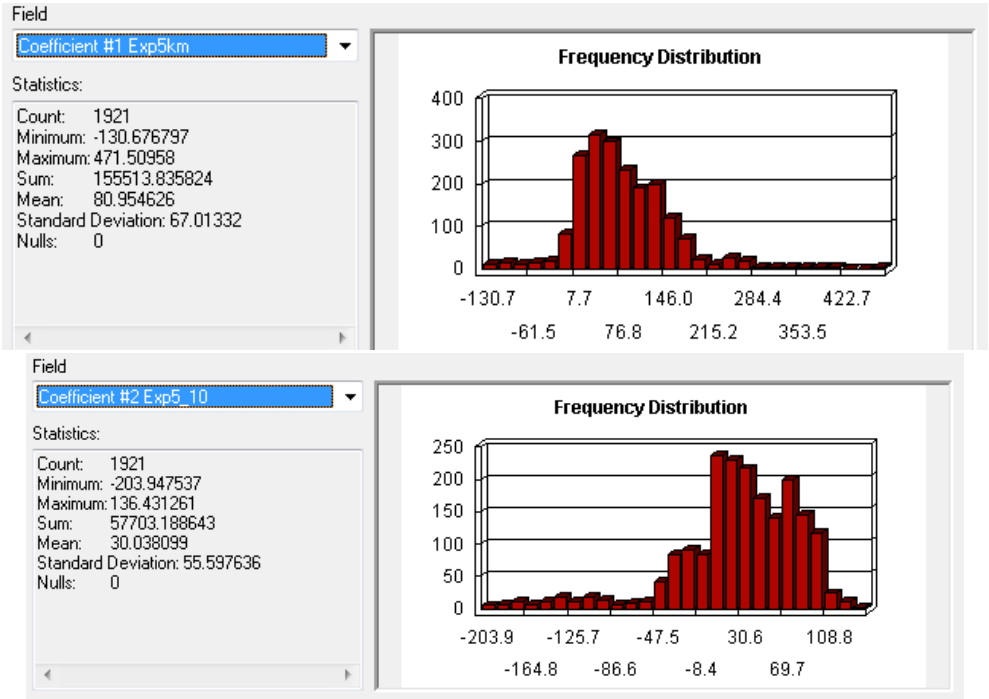


Figure 18: GWR Coefficient Values for Intranodal Cases Best-Fitting Model

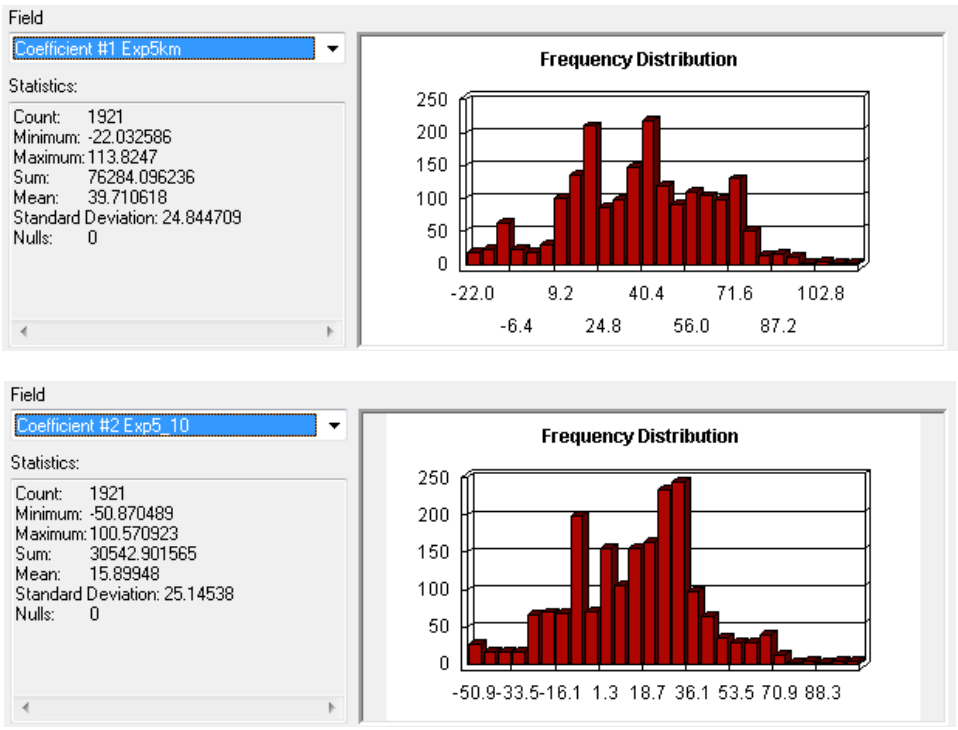


Figure 19: GWR Coefficient Values for Extranodal Cases Best-Fitting Model

The mapped outputs of the GWR model residuals are depicted below. Figure 20 shows the residuals for Model 11 (the all-cases base model), Figure 21 shows residuals for Model 13 (male cases base model, and the overall best-fitting model) and Model 15 (female cases), and Figure 22 shows the residuals for intranodal and extranodal NHL base models (Models 17 and 19, respectively). The GWR residuals for "full" models (which include Appalachia and Beale Code) are not shown because in every instance they had higher AIC values and lower R^2 values than their base counterparts.

The patterns and magnitudes of residuals by census tract, 5-10km buffer zone, and 5km buffer zone depicted in Figures 20 through 22 are not surprising, given that the best-fitting GWR model still only explained 24.6% of the variability in the dependent

variable. An additional 75.4% or more, depending on the particular model, is therefore not explained, and these errors between the observed incidence rates and the rates predicted by models do not show any readily apparent pattern in any model. It can be noted that in all models, there appear to be more areas of “high” standardized residuals than “low” standardized residuals. The highest magnitude areas, where the observed incidence rates exceed the predicted rates by more than 2.5 standard deviations, are most prominent in the central and western areas of Kentucky. Low areas, where the observed incidence rates are lower than the predicted rates, are randomly scattered throughout the state. Across all residual maps, the tracts/buffer zones with high and low residual values tend to hold up across all models, though the color shading might change slightly to indicate lower-magnitude residuals in the 1.5 to 2.5 standard deviations range.

The combination of standardized GWR regression residual diagrams and the adjusted R^2 values strongly suggest that there are likely other explanatory variables which might contribute to NHL incidence, and which we were unable to capture in the current investigation. However, the data collected and analyzed in the present investigation supported the hypothesis that residential proximity to Superfund sites in Kentucky is a significant factor in elevated incidence of non-Hodgkin’s lymphoma. These results are similar to those documented in Georgia by investigators who looked at NHL risk and residential proximity to areas where benzene was released and documented in the EPA Toxics Release Inventory[151].

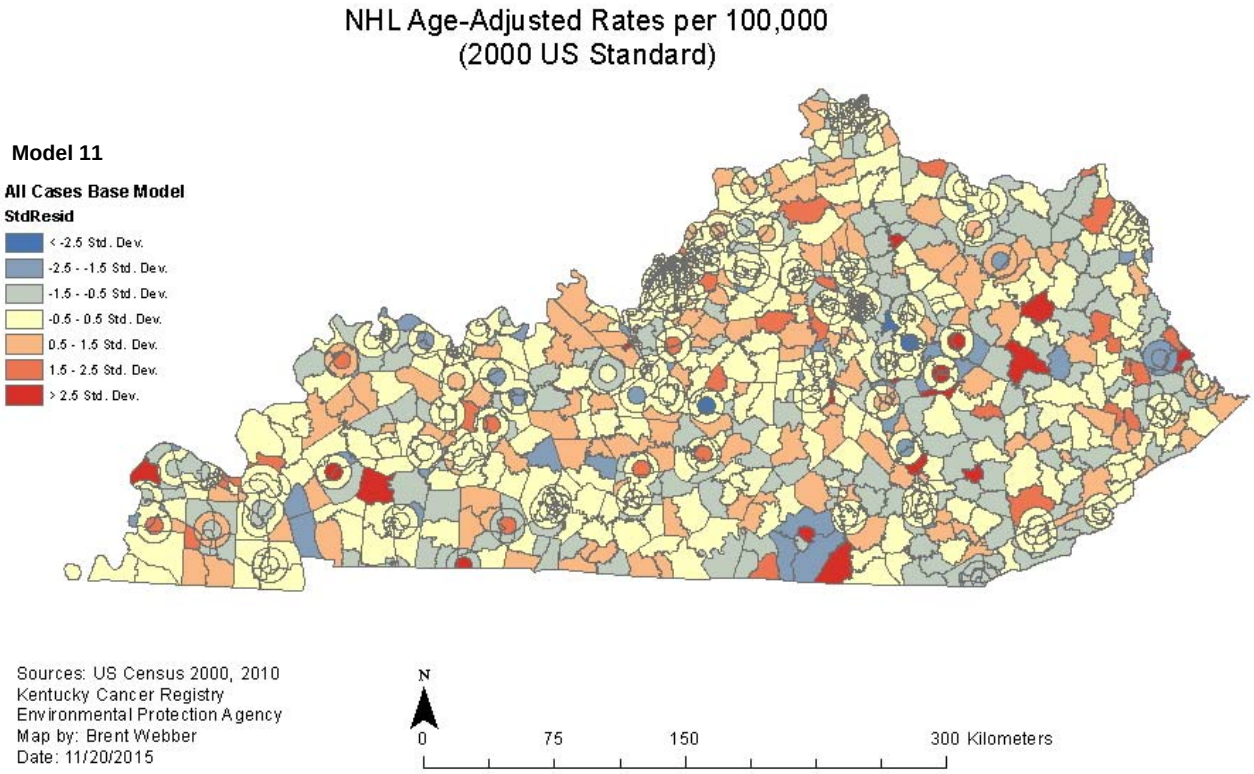
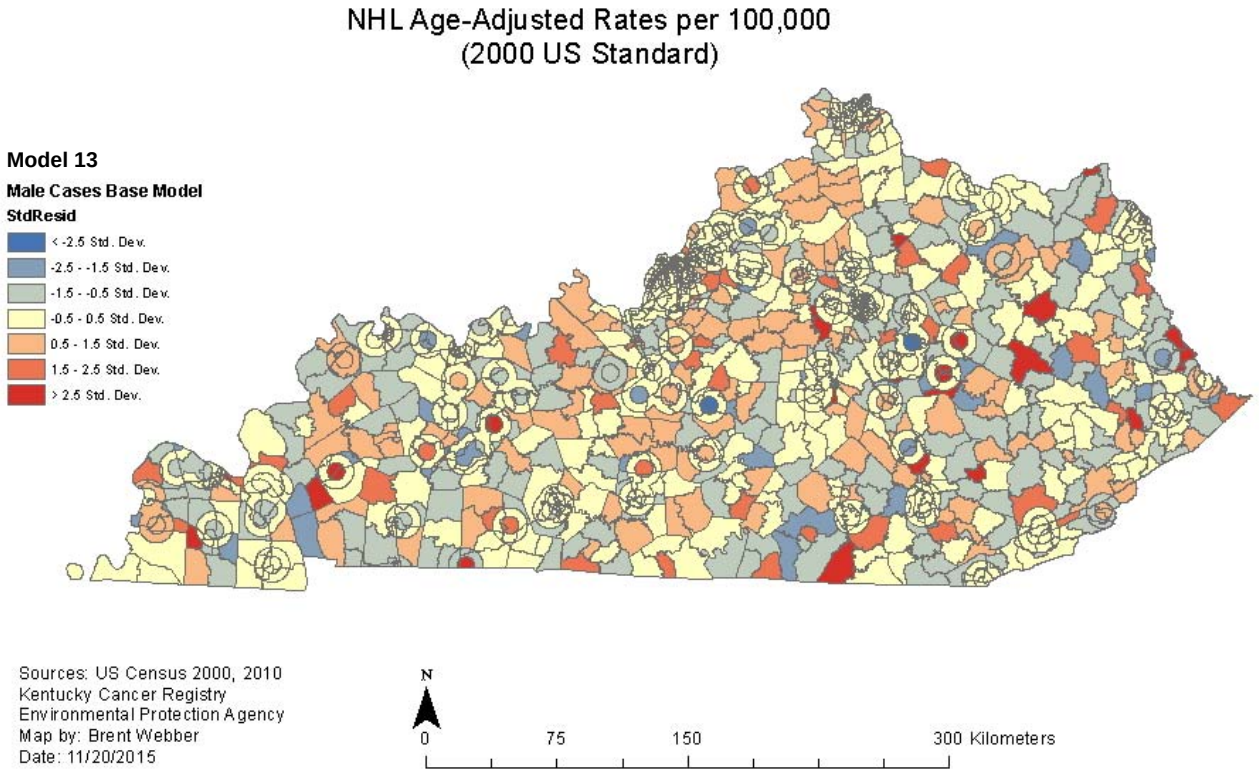


Figure 20. Geographically Weighted Regression Residuals, Cumulative NHL Incidence Data for All Cases, Kentucky, 1995-2012



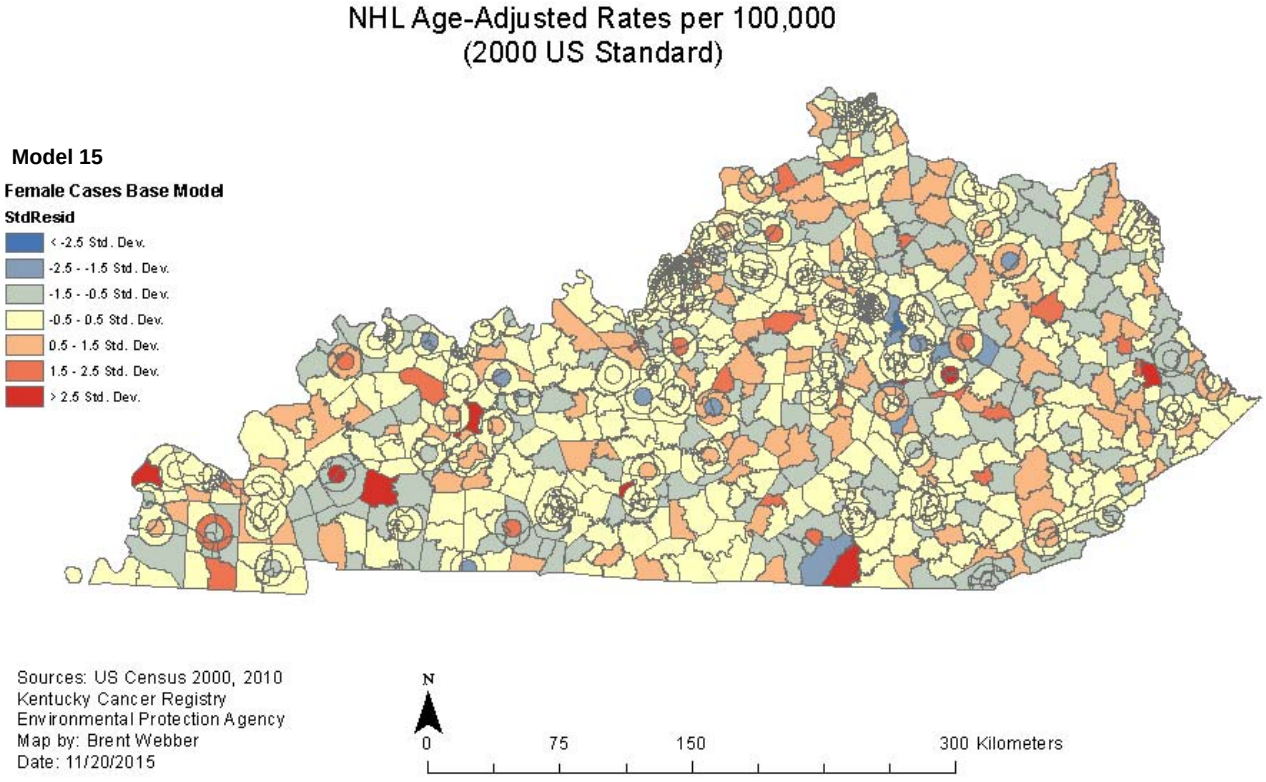


Figure 21. Geographically Weighted Regression Residuals, Cumulative NHL Incidence Data for Male and Female Cases, Kentucky, 1995-2012

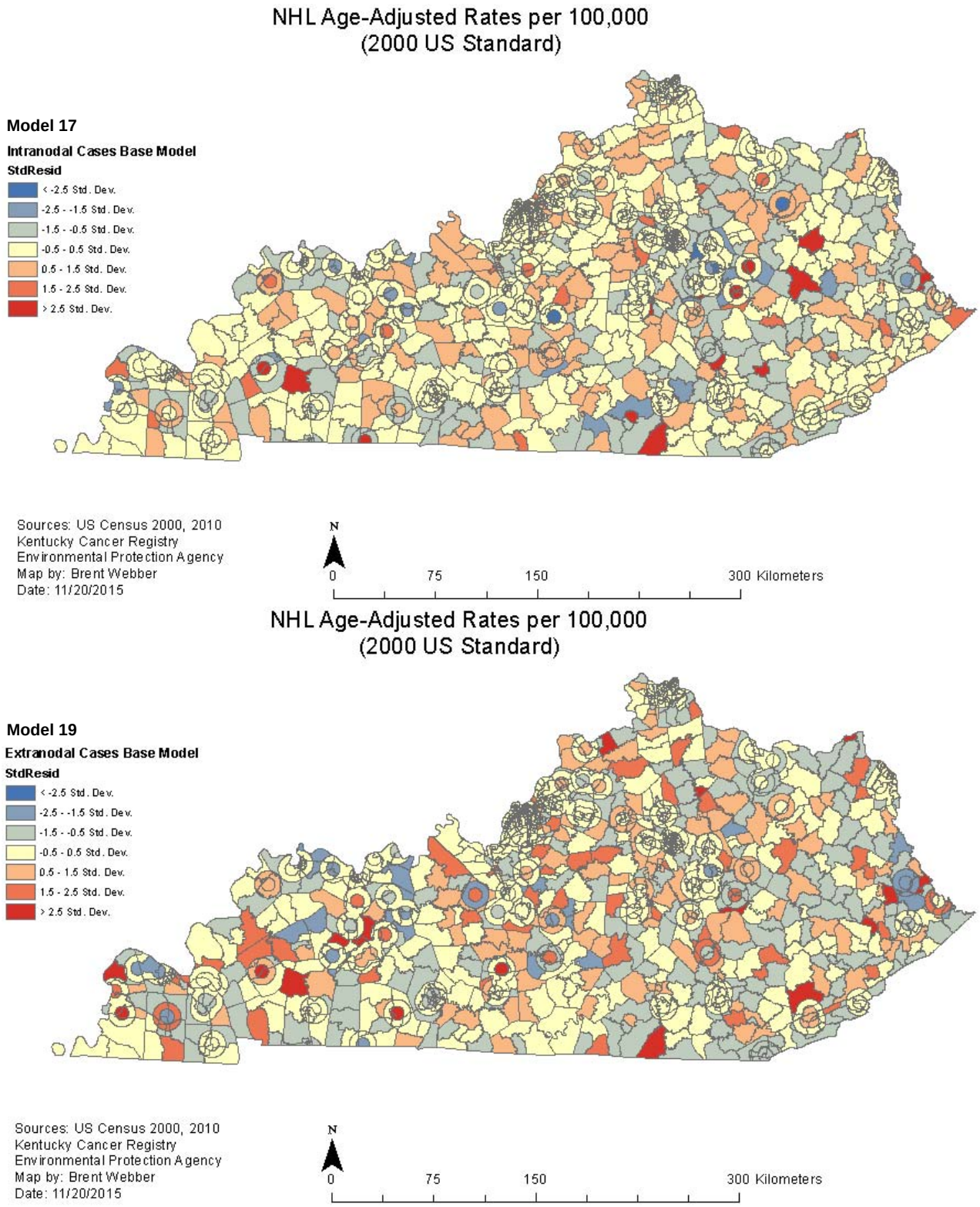


Figure 22. Geographically Weighted Regression Residuals, Cumulative NHL Incidence Data for Intranodal and Extranodal Cases, Kentucky, 1995-2012

CHAPTER 5
IMPLICATIONS FOR PUBLIC HEALTH

Non-Hodgkin's lymphoma is one of the most prevalent cancers in Kentucky, the United States, and many Western countries. Incidence and prevalence rates both increased markedly during the latter half of the 20th Century, in a time-lagged concurrency with the marked increase in the manufacture, use, and disposal of chemical products, which strongly links the two phenomena together. Research is ongoing to discover ways to diagnose, treat, and prevent NHL, and the current five-year survival rate following NHL diagnosis in the United States is 69%[152]. This research is urgently needed, because persons are more susceptible to NHL as they age, and demographic trends point toward the aging of the U.S. population between 2015 and 2050[153].

The present study investigated whether residential proximity to NPL/Superfund sites in Kentucky could be at least partially responsible for the increasing and then plateauing incidence of NHL from the mid-20th Century to the present. Data representing all cases of NHL reported to the Kentucky Cancer Registry between 1995 and 2012 were acquired and cleaned. This data was used to calculate 18-year incidence rates for NHL at the 2010 U.S. Census tract level. Because NHL risk is not evenly distributed across the lifespan, incidence data was age-adjusted using a standard population from the 2000 U.S. Census. These age-adjusted NHL rates were

the outcome data, which were evaluated through both conventional and geospatial statistical methods.

The total number of cases of NHL in Kentucky showed an upward trend from 1995 to 2012, which was more pronounced in non-white subjects. Bivariate analysis revealed that non-white cases, and cases with no prior known family history of NHL, were significantly more likely to live within 5km of Superfund sites. Rates were significantly higher in the exposed (10km or less) regions than in the unexposed regions, across every outcome stratum except intranodal cases in females. Rates of NHL were not higher in the Appalachian (eastern and southern) regions of Kentucky, making NHL unlike many other forms of cancer which show clustering or higher rates in Appalachian compared to non-Appalachian regions of the United States[154-157].

The outcome data was spatially autocorrelated, clustered, and exhibited non-stationarity, all indicators that geographically weighted regression (GWR) techniques would be necessary. Both Ordinary Least Squares (global) regression and localized GWR showed that NHL rates were significantly elevated in geographic areas within 5km of Superfund sites, and in areas between 5km and 10km of Superfund sites, when controlling for covariates which were significant in bivariate analysis. The effect size was greater in the population with more exposure (<5km). Therefore, the investigation supported the hypothesis that closer residential proximity to Superfund sites in Kentucky was a significant factor in NHL risk.

The results from the present study point to disparities in rate of NHL across Kentucky due to living near recognized areas of environmental toxicity. These areas, by virtue of being on the U.S. EPA Superfund list, were formally designated as needing

evaluation and some degree of clean-up. The results raise questions about environmental justice. Persons in disadvantaged positions often do not have the political or economic power to organize resistance to nearby hazardous or noxious land uses, whereas more affluent regions are able to mount NIMBY (Not in my Backyard) responses and keep them away[158]. Unequal burdens of exposure and outcome along racial, ethnic, and socioeconomic lines are often the product of facility siting decisions that make initial economic sense due to lower property costs, lower required compensation levels for adjacent landowners, or a desire to reduce local levels of unemployment or underemployment[159]. While multivariate analysis by race could not be conducted in the present study due to the small proportion of non-white NHL cases, it must be noted that non-white cases were significantly more likely to live within 5km of Superfund sites compared to white cases. This is consistent with patterns that exist nationwide, particularly among African-Americans[160].

Are persons who live closer to Superfund sites, and have higher rates of NHL, actually exposed to higher levels of contaminants which can trigger NHL? The nature and pathways of exposure to potentially hazardous substances from Superfund sites in Kentucky are not fully known, but reasonable hypotheses can be drawn from limited available data. Of the twenty Kentucky sites that scored highest in the EPA's Hazard Ranking System and were thus on the National Priorities List, a majority had one or more on-site contaminants known or suspected to increase NHL incidence. The most commonly found of these contaminants were benzene, lead, PCBs, cadmium, trichloroethylene, organochlorines other than PCBs, and perchloroethylene[143]. The most likely exposure pathways in the current context would be through ground water or

surface water, whereas offsite exposure through air or contaminated soil would probably be limited. Air exposures from Superfund sites are generally due to gaseous pollutants, secondary pollutants generated by reactions, or wind erosion from contaminants deposited onto surfaces, which are not as likely with the sites in the present study. While contaminated soil itself does not migrate off-site, plumes of pollutants can migrate with the water table and move laterally through soils.

Recommendations

Persons who live close to Superfund sites in Kentucky appear to have a significantly higher NHL risk. As with many public health issues, there are a variety of target areas where efforts can be directed. These will be described below, starting with the upstream (distal, societal) factors, then working down to community, neighborhood, and individual-level approaches.

Upstream approaches that prevent the problem from occurring can be a possible target, but these approaches are often expensive, beyond the scope of traditional public health practice, and require extensive inter-agency involvement. For example, mandatory evictions and buyouts of residences near the most contaminated sites can prevent residential exposure, and have occurred in areas such as Niagara Falls, NY[161], Times Beach, MO[162], and Houston, TX[163]. However, this can lead to numerous legal and ethical challenges. Land values might be depressed due to the adjacent activities that necessitated the buyouts, or persons might resist buyouts and leaving their ancestral domiciles. It is not likely that the U.S. EPA or any governmental agency with

eminent domain powers would deem it feasible or desirable to evict/buyout residents living near Superfund sites to reduce NHL risk, based on current evidence.

American environmental policy and siting decisions are slowly transitioning from NIMBY to “Not in Anyone’s Backyard” as the knowledge of both the direct (toxicity, illness, reduced property values) and indirect (stress, allostatic load) effects of residential proximity to hazardous materials and processes become more known[164]. In the United States, the burden for absorbing the externality of pollution has shifted from the surrounding community to the generating entities through federal environmental laws such as SARA Title III (also known as the Emergency Planning and Community Right-to-Know Act), the Hazardous and Solid Waste Amendments of RCRA, and the Ground Water Rule. These efforts have resulted in improvements to new and existing hazardous sites, but vigilant surveillance from both regulatory and community entities is still necessary to ensure that environmental outcomes are fair and just. Upstream interventions, education, and advocacy could improve justice in future siting, monitoring, clean-up, and closure activities, but would be slow to impact those currently affected.

Another factor curtailing the ability to clean existing sites and reduce possible NHL risk is that the actual “Superfund” from CERCLA no longer exists. The tax on petrochemical industries that created the fund to perform cleanups ceased in 1995 following significant pressure on the Legislature, and the fund officially went bankrupt in 2003[165]. Since that time, Superfund site activities have been funded by annual appropriations to the EPA, which have been stagnant or declining for years in a highly competitive environment for Federal dollars[166]. While there has been an effort to reinstate the Superfund tax during the Obama administration[166], as of 2015 it has still

not been enacted and is not likely to be considered soon. Unless grassroots advocacy can mobilize a groundswell of support from citizens to bolster the Superfund program, the current slow pace of assessment and cleanup will continue, and persons who live near contaminated sites will continue to be at higher risk for NHL and other negative effects to health and well-being. To the extent that public health can advocate for those who share unequal health burdens due to hazardous sites, and for increased funding for CERCLA and other beneficial programs that tackle the problem upstream, it should do so. However, in the present climate, it is more prudent to apply public health efforts downstream at the regional, community, or individual level.

Research on the potential causal factors of NHL has already provided several interesting clues, but must continue until definitive linkages are found. Community and regional research on NHL incidence is also critical. The present study provides an important example of possible causal linkage between hazardous waste site exposures and NHL; however, studies similar to this one should be replicated in other settings, particulary in areas with greater racial and ethnic diversity. Also, more research is needed to bring exposure pathways into clearer view so that exposure misclassification is minimized. Water would be the most likely media for transfer of toxicants from Superfund sites to adjacent residential areas, with soil and air as less likely paths. Tracking drinking-water sources for NHL cases and controls, and comparing levels of possible NHL-triggering toxicants in surface water and groundwater samples, would be an important next step, as would examination of groundwater hydrography data to know which direction(s) from the site would be most likely affected. For sites where soil/surface contamination is the primary hazard, examining wind-rose data might show

where airborne/dustborne contaminants are most likely to have an impact. Misclassification due to residential mobility could not be evaluated in the present study, as only the residential coordinates at the time of diagnosis were available. Studies that are able to include model components for residential change would be helpful, and have shown promise[167].

Community and individual level efforts at education and screening are critical, even though these are downstream actions that do not directly address the upstream factors driving NHL incidence. Education and awareness campaigns about NHL, its risk factors, and symptoms could lead to earlier diagnosis and better outcomes in affected communities. At the present time, there is not a simple, inexpensive screening test for NHL that makes it amenable to a “mobile clinic” setting like other cancers such as breast and colon. Early detection relies on techniques such as lymph node biopsy, blood cell chemistry and morphology tests, or imaging scans which can detect not just NHL but other hematological malignancies[168]. Encouraging persons who score high on risk factor and symptom surveys for NHL to seek medical advice and screening could save lives and improve outcomes, if combined with funding for tests and specialized medical knowledge in the communities most affected by NHL. Medical research should continue to investigate simple, low-cost, sensitive and specific methods for detecting NHL, as it will most likely continue to be a cancer of high-incidence as the population ages.

Summary

Non-Hodgkin's lymphoma (NHL) incidence in the United States and many other Western nations increased throughout the 20th Century, in a pattern that suggests greater exposure to chemicals might be a causal factor. Mechanistic research suggests many pathways by which chemicals and xenobiotics can trigger NHL. The present study demonstrated that residential proximity to hazardous waste sites in Kentucky was a significant risk factor for NHL. Additional research, advocacy, and education should focus on mechanisms of NHL incidence, replicating the present study in other contexts and with monitoring data, addressing upstream factors that lead to unequal burdens of hazardous material exposures and NHL, downstream education and awareness, and better methods for NHL screening and early detection.

REFERENCES

1. Küppers, R., & Hansmann, M.L. (2005). The Hodgkin and Reed-Sternberg cell. *Int J Biochem & Cell Biology, 37*, 511-517.

2. National Cancer Institute: Surveillance, Epidemiology, and End Results Program (SEER). *SEER stat fact sheets: non-Hodgkin's lymphoma*. Retrieved February 28, 2015, from http://seer.cancer.gov/statfacts/html/nhl.html.

3. National Cancer Institute: Surveillance, Epidemiology, and End Results Program (SEER). *SEER stat fact sheets: Hodgkin lymphoma*. Retrieved February 28, 2015, from http://seer.cancer.gov/statfacts/html/hodg.html.

4. Lymphoma Association, United Kingdom. (2011, June 21). *Leukaemia and lymphoma: what's the difference?* Retrieved February 28, 2015, from http://www.nhs.uk/ipgmedia/national/Lymphoma%20Association/Assets/LeukaemiaandLymphoma-thedifference.pdf.

5. National Cancer Institute (NCI). (2007, September). *What you need to know about non-Hodgkin lymphoma*. Retrieved February 28, 2015, from http://www.cancer.gov/publications/patient-education/non-hodgkin-lymphoma.pdf

6. American Cancer Society. (2015, January 28). *Types of non-Hodgkin lymphoma*. Retrieved February 28, 2015, from http://www.cancer.org/cancer/non-hodgkinlymphoma/detailedguide/non-hodgkin-lymphoma-types-of-non-hodgkin-lymphoma.

7. American Society of Clinical Oncology. (2014, November). *Lymphoma - non-Hodgkin: subtypes*. Retrieved February 28, 2015, from http://www.cancer.net/cancer-types/lymphoma-non-hodgkin/subtypes.

8. Al-Hamadani, M., Habermann, T.M., Cerhan, J.R., *et al*. (2015). Non-Hodgkin lymphoma subtype distribution, geo-demographic patterns, and survival in the US: a longitudinal analysis of the National Cancer Data Base from 1998-2011. *Am J Hematology, 90*(9), 790-795.

9. Newton, R., Ferlay, J., Beral, V., & Devesa, S.S. (1997). The epidemiology of non-Hodgkin's lymphoma: comparison of nodal and extra-nodal sites. *Int J Cancer, 72*, 923-930.

10. Zucca, E., & Cavalli, F. (2000). Extranodal lymphomas. *Ann Oncol, 11*(Suppl 3), 219-222.

11. Centers for Disease Control and Prevention (CDC): U.S. Cancer Statistics Working Group. (2015). *United States cancer statistics: 2012 top ten cancers*. Retrieved December 8, 2015, from https://nccd.cdc.gov/uscs/toptencancers.aspx.

12. Horesh, N. & Horowitz, N.A. (2014). Does gender matter in non-Hodgkin's lymphoma? Differences in epidemiology, clinical behavior, and therapy. *Rambam Maimonides Med J, 5*(4), open access, retrieved from http://www.rmmj.org.il/userimages/429/1/PublishFiles/450Article.pdf.

13. Lee, J.S., Bracci, P.M., & Holly, E.A. (2008). Non-Hodgkin lymphoma in women: Reproductive factors and exogenous hormone use. *Am J Epidemiol, 168*(3), 278-288.

14. Harris, R.E. (2013). *Epidemiology of chronic disease: Global perspectives*. Burlington, MA: Jones & Bartlett Learning, LLC.

15. Clarke, C.A. & Glaser, S.L. (2002). Changing incidence of non-Hodgkin lymphomas in the United States. *Cancer, 94,* 2015-2023.

16. Bassig, B.A., Lan, Q., Rothman, N., *et al*. (2012). Current understanding of lifestyle and environmental factors and risk of non-Hodgkin lymphoma: an epidemiological update*. J Cancer Epidemiol*, open access: article ID 978930, doi:10.1155/2012/978930.

17. Smith, J.K. (2007). The American chemical industry since the petrochemical revolution. In L. Galambos, T. Hikino, & V. Zamagni (Eds.), *The Global Chemical Industry in the Age of the Petrochemical Revolution*. Cambridge: Cambridge University Press.

18. Nash, L. (2004). The fruits of ill-health: Pesticides and workers' bodies in post-World War II California. *Osiris,19*, 203-219.

19. Burton C, Jack A, Adamson P, & Roman E. (2010). *Descriptive epidemiology*. In I.T. Magrath (Ed.), *The Lymphoid Neoplasms* (3rd ed.). London: Hodder Arnold.

20. Salhotra, A.M. (2012). Human health risk assessment for contaminated properties. *Progress in Molecular Biology and Translational Science: Toxicology and Human Environments*, *112*, 285–306.

21. Dunn, R. (2012). In retrospect: Silent Spring. *Nature, 483*(31), 578-579.

22. Grant, D., Trautner, M.L, Downey, L., & Thiebaud, L. (2010). Bringing the polluters back in: Environmental inequality and the organization of chemical production*. American Sociological Review, 75*(4), 479-504.

23. Andrews, R.N.L. (2006) *Managing the environment, managing ourselves*. New Haven, CT: Yale University Press.

24. U.S. Environmental Protection Agency (EPA). (2016, January 5). *EPA History: Resource Conservation and Recovery Act*. Retrieved January 8, 2016, from http://www.epa.gov/aboutepa/epa-history-resource-conservation-and-recovery-act.

25. Beck, E.C., (1979). *The Love Canal Tragedy*. EPA Journal, January 1979. Retrieved February 28, 2015, from http://www2.epa.gov/aboutepa/love-canal-tragedy.

26. U.S. Environmental Protection Agency (EPA). (2016, January 6), *A.L. Taylor (Valley of Drums), Brooks, KY*. Retrieved January 8, 2016, from http://cumulis.epa.gov/supercpad/cursites/csitinfo.cfm?id=0402072.

27. U.S. Environmental Protection Agency (EPA). (2015, September 30). *CERCLA Overview*. Retrieved January 8, 2016, from http://www.epa.gov/superfund/superfund-cercla-overview.

28. Kingsley, B.S., Schmeichel, K.L., & Rubin, C.H. (2006). An Update on Cancer Cluster Activities at the Centers for Disease Control and Prevention*. Environ Health Perspect, 115*(1), 165-171.

29. Dumalaon-Canaria, J.A., Hutchinson, A.D., Prichard, I., & Wilson, C. (2014). What causes breast cancer? A systematic review of causal attributions among breast cancer survivors and how these compare to expert-endorsed risk factors*. Cancer Causes Control, 25*(7), 71-85.

30. Ela, W.P., Sedlak, D.L., Barlaz, M.A., *et al*. (2011). Toward identifying the next generation of superfund and hazardous waste site contaminants*. Environ Health Perspect, 119*(1), 6-10.

31. Najem, G.R., & Cappadona, J.L. (1991). Health effects of hazardous chemical waste disposal sites in New Jersey and in the United States: a review*. Am J Prev Med, 7*(6), 352-362.

32. Kamrin, M.A., Fischer, L.J., Suk. W.A., *et al*. (1994). Assessment of Human Exposure to Chemicals from Superfund Sites*. Environ Health Perspect Supplements, 102 Suppl*(1), 221-228.

33. United Health Foundation. (2014). *America's health rankings: A call to action for individuals and their communities* (25th ed.). Retrieved February 28, 2015, from http://cdnfiles.americashealthrankings.org/SiteFiles/Reports/Americas%20Health%20Rankings%202014%20Edition.pdf.

34. National Cancer Institute: Surveillance, Epidemiology, and End Results Program (SEER). (2014). *Table 19.22: Non-Hodgkin Lymphoma Age-adjusted Cancer Death Rates by State, All Races, 2007-2011, Males and Females Fact Sheet.* Retrieved February 28, 2015, from http://seer.cancer.gov/csr/1975_2011/browse_csr.php?sectionSEL=19&pageSEL=sect_19_table.22.html.

35. Devesa, S.S., & Fears, T. (1992). Non-Hodgkin's lymphoma time trends: United States and international data*. Cancer Res, 52*(Suppl), 5432s-5440s.

36. Fisher, S.G., & Fisher, R.I. (2004). The epidemiology of non-Hodgkin's lymphoma. *Oncogene, 23*, 6524-6534.

37. Sherman, R.L., Henry, K.A., Tannenbaum, S.L., *et al*. (2014). Applying spatial analysis tools in public health: an example using SaTScan to detect geographic targets for colorectal cancer screening interventions. *Prev Chronic Dis, 11*, 130264. doi: http://dx.doi.org/10.5888/pcd11.130264.

38. Iceland, J., & Steinmetz, E. (2003, July). *The effects of using census block groups instead of census tracts when examining residential housing patterns*. Retrieved February 28, 2015, from http://www.census.gov/hhes/www/housing/housing_patterns/pdf/unit_of_analysis.pdf.

39. Petriello, M.C., Newsome, B.J., Dziubla, T.D., *et al*. (2012). Modulation of persistent organic pollutant toxicity through nutritional intervention: Emerging opportunities in biomedicine and environmental remediation. *Sci Total Environ, 491-492*, 11-16.

40. Budnick, L.D., Logue, J.N., Sokal, D.C., *et al*. (1984). Cancer and birth defects near the Drake Superfund site, Pennsylvania. *Arch Env Hlth, 39*(6), 409-413.

41. Dayal, H., Maierson, D., Gupta, S., *et al*. (1995). Symptom clusters in a community with chronic exposure to chemicals in two Superfund sites. *Arch Env Hlth, 50*(2), 108-111.

42. Ozonoff, D., Aschengrau, A., & Coogan, P. (1994). Cancer in the Vicinity of a Department of Defense Superfund Site in Massachusetts. *Toxicology and Industrial Health, 10*(3), 119-141.

43. Williamson, D.M., White, M.C., Poole, C., *et al*. (2006). Evaluation of Serum Immunoglobulins among Individuals Living Near Six Superfund Sites. *Environ Health Perspect, 114*(7), 1065-1071.

44. Currie, J., Greenstone, M., & Moretti, E. (2011). Superfund cleanups and infant health. *American Economic Review: Papers & Proceedings, 101*(3), 435-441.

45. Peverly, A.A., Salamova, A., & Hites, R.A. (2014). Air is still contaminated 40 years after the Michigan Chemical Plant disaster in St. Louis, Michigan. *Environ Sci Technol, 48*, 11154-11160.

46. Allan, S.E., Sower, G.J., & Anderson, K.A. (2011). Estimating risk at a Superfund site using passive sampling devices as biological surrogates in human health risk models. *Chemosphere, 85*(6), 920-927.

47. Dias da Cunha, K.M., Henderson, H., Thompson, B.M., & Hecht, A.A. (2014). Groundwater contamination with ^{238}U, ^{234}U, ^{235}U, ^{226}Ra, and ^{210}Pb from past uranium mining: cove wash, Arizona. *Environ Geochem Health, 36*, 477-487.

48. Szasz, A., & Meuser, M. (1997). Environmental inequalities: literature review and proposals for new directions in research and theory. *Current Sociology 45*(3), 99–120.

49. Szasz, A., & Meuser, M. (2000). Unintended, inexorable: the production of environmental inequalities in Santa Clara County, California. *American Behavioral Scientist 43*(4), 602–632.

50. Ringquist, E.J. (2005). Assessing the evidence of environmental inequities: a meta-analysis. *J Policy Analysis and Management 24*(2), 223–247.

51. Smith, C.L. (2009). Economic deprivation and racial segregation: Comparing Superfund sites in Portland, Oregon and Detroit, Michigan. *Soc Sci Research, 38*, 681-692.

52. Greenstone, M., & Gallagher, J. (2008). Does hazardous waste matter? Evidence from the housing market and the Superfund program. *Quarterly Journal of Economics, 123*(3), 951-1003.

53. Brewer, C.A. (2006). Basic mapping principles for visualizing cancer data using Geographic Information Systems (GIS). *Am J Prev Med, 30*(Suppl 2), S25-S36.

54. Copeland, G. (2010). The role of public health and how boundary analysis can provide a tool for public health investigations: the public health perspective*. Spat Spatiotemporal Epidemiol, 1*(4), 201-205.

55. English, D. (1996). Geographical epidemiology and ecological studies. In P. Elliott, J. Cuzick, D. English, & R. Stern (Eds.), *Geographical and environmental epidemiology: Methods for small-area studies* (pp. 3-21). Oxford: Oxford University Press.

56. Gaffney, S.H., Curriero, F.C., Strickland, P.T., *et al*. (2005). Influence of geographic location in modeling blood pesticide levels in a community surrounding a U.S. Environmental Protection Agency Superfund site*. Environ Health Perspect, 113*(12), 1712-1716.

57. Choi, A.L., Levy, J.I., Dockery, D.W., *et al*. (2006). Does living near a Superfund site contribute to higher polychlorinated biphenyl (PCB) exposure? *Environ Health Perspect, 114*(7), 1092-1098.

58. Kearney, G. (2008). A procedure for detecting childhood cancer clusters near hazardous waste sites in Florida*. J Env Health, 70*(9), 29-34.

59. Thompson, J.A., Bissett, W.T., & Sweeney, A.M. (2014). Evaluating geostatistical modeling of exceedance probability as the first step in disease cluster investigations:

very low birth weights near toxic Texas sites. *Environ Health, 13*(47), Open Access, retrieved January 8, 2016, from http://www.ehjournal.net/content/13/1/47.

60. Burwell-Naney, K., Zhang, H., Samantapudi, A., *et al*. (2013). Spatial disparity in the distribution of Superfund sites in South Carolina: an ecological study*. Environ Health, 12*(96), Open Access, retrieved January 8, 2016, from http://www.ehjournal.net/content/12/1/96.

61. Maranville, A.R., Ting, T.F., & Zhang, Y. (2009). An environmental justice analysis: Superfund sites and surrounding communities in Illinois. *Environmental Justice, 2*(2), 49-59.

62. Maantay, J. (2002). Mapping environmental injustices: Pitfalls and potential of geographic information systems in assessing environmental health and equity. *Environ Health Perspect, 110*(Suppl 2), 161-171.

63. Pais, J., Crowder, K., & Downey, L. (2013). Unequal trajectories: Racial and class differences in residential exposure to industrial hazard. *Social Forces, 92*(3), 1189-1215.

64. Heitgerd, J.L., & Lee, C.V. (2003). A new look at neighborhoods near National Priorities List sites. *Social Science & Medicine, 57*(6), 1117-1126.

65. Yuan, J., Chen, L., Chen, D., *et al*. (2008). Elevated serum polybrominated diphenyl ethers and thyroid-stimulating hormone associated with lymphocytic micronuclei in Chinese workers from an e-waste dismantling site*. Environ Sci Technol, 42*, 2195-2200.

66. Chiang, C.T., Lian, I.B., Chang, Y.F., & Chang, T.K. (2014). Geospatial disparities and the underlying causes of major cancers for women in Taiwan*. Int J Environ Res Public Health, 11*(6), 5613-5627.

67. Pearce, D.C., Dowling, K., & Sim, M.R. (2012). Cancer incidence and soil arsenic exposure in a historical gold mining area in Victoria, Australia: a geospatial analysis*. J Expo Sci Environ Epidemiol, 22*(3), 248-257.

68. Macon, M.B., & Fenton, S.E. (2013). Endocrine disruptors and the breast: Early life effects and later life disease*. J Mammary Gland Biol Neoplasia, 18*(1), 43-61.

69. Bulger, W.H. & Kupfer, D. (1983). Estrogenic action of DDT analogs*. Am J Internal Med, 4*, 163-173.

70. DeBruin, L.S. & Josephy, P.D. (2002). Perspectives on the chemical etiology of breast cancer*. Environ Health Perspect Supplements, 110*(1), 119-128.

71. Dunnick, J.K., Elwell, M.R., Huff, J., & Barrett, J.C. (1995). Chemically induced mammary gland cancer in the National Toxicology Program's carcinogenesis bioassay. *Carcinogenesis, 16*(2), 173-179.

72. Ekenga, C.C., Parks, C.G., D'Aloisio, A.A. *et al*. (2014). Breast cancer risk after occupational solvent exposure: the influence of timing and setting. *Cancer Res, 74*(11), 3076-3083.

73. Brody, J.G., Rudel, R.A., Michels, K.B., *et al*. (2007). Environmental pollutants, diet, physical activity, body size, and breast cancer: where do we stand in research to identify opportunities for prevention? *Cancer, 109*(Suppl 12), 2627-2634.

74. Hays, K.A., & Breshears, M.A. (2011). Presence of hyperplastic pectoral mammary glands in a white-footed mouse (*Peromyscus leucopus*) from a Superfund site in Oklahoma, USA. *J Wildlife Diseases, 47*(1), 255-258.

75. Burger, M., Catto, J.W.F., Dalbagni, G., *et al*. (2013). Epidemiology and risk factors of urothelial bladder cancer. *Eur Urol, 63*(2), 234-241.

76. Golka, K., Wiese, A., Assennato, G., & Bolt, H.M. (2004). Occupational exposure and urological cancer. *World J Urol, 21*(6), 382-391.

77. Carreon, T., Hein, M.J., Hanley, K.W., *et al*. (2014). Bladder cancer incidence among workers exposed to o-toluidine, aniline and nitrobenzene at a rubber chemical manufacturing plant. *Occup Environ Med, 71*(3), 175-182.

78. Porru, S., Pavanello, S., Carta, A., *et al*. (2014). Complex relationships between occupation, environment, DNA adducts, genetic polymorphisms and bladder cancer in a case-control study using a structural equation modeling. *PLoS One, 9*(4), Open Access, retrieved January 8, 2016, from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094566.

79. Frederickson, S.M., Messing, E.M., Reznikoff, C.A., *et al*. (1994). Relationship between in vivo acetylator phenotypes and cytosolic N-acetyltransferase and O-acetyltransferase activities in human uroepithelial cells. *Cancer Epidemiol Biomarkers Prev, 3*, 25-32.

80. Letasiova, S., Medvedova, A., Sovcikova, A., *et al*. (2012). Bladder cancer, a review of the environmental risk factors. *Environ Health, 11*(Suppl 1), Open Access, retrieved January 8, 2016, from http://www.ehjournal.net/content/11/S1/S11.

81. Cantor, K.P., Strickland, P.T., Brock, J.W., *et al*. (2002). Risk of non-Hodgkin's lymphoma and prediagnostic serum organochlorines: ß-hexachlorocyclohexane, chlordane/heptachlor-related compounds, dieldrin, and hexachlorobenzene. *Environ Health Perspect, 111*(2), 179-183.

82. Rothman, N., & Cantor, K.P. (1997). A nested case-control study of non-Hodgkin lymphoma and serum organochlorine residues. *Lancet, 350*(9073), 240-244.

83. Maifredi, G., Donato, F., Magoni, M., *et al*. (2011). Polychlorinated biphenyls and non-Hodgkin's lymphoma: a case-control study in northern Italy. *Environ Res, 111*(2), 254-259.

84. Engel, L.S., Lan, Q., & Rothman, N. (2007). Polychlorinated biphenyls and non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev, 16*(3), 373-376.

85. Rafnsson, V. (2006). Risk of non-Hodgkin's lymphoma and exposure to hexachlorocyclohexane, a nested case-control study*. Eur J Cancer, 42*(16), 2781-2785.

86. Viel, J.F., Fournier, E., & Danzon, A. (2010). Age-period-cohort modelling of non-Hodgkin's lymphoma incidence in a French region: a period effect compatible with an environmental exposure*. Environ Health, 9*(47), Open Access, retrieved January 8, 2016, from http://www.ehjournal.net/content/9/1/47.

87. Schinasi, L., & Leon, M.E. (2014). Non-Hodgkin lymphoma and occupational exposure to agricultural pesticide chemical groups and active ingredients: a systematic review and meta-analysis*. Int J Environ Res Public Health, 11*(4), 4449-4527.

88. Spinelli, J.J., Ng, C.H., Weber, J.P., *et al*. (2007). Organochlorines and risk of non-Hodgkin lymphoma. *Int J Cancer, 121*, 2767-2775.

89. Merhi, M., Raynal, H., Cahuzak, E., *et al*. (2007). Occupational exposure to pesticides and risk of hematopoietic cancers: Meta-analysis of case-control studies. *Cancer Causes Control, 18*, 1209-1226.

90. Viel, J.F., Floret, N., Deconinck, E., *et al*. (2011). Increased risk of non-Hodgkin lymphoma and serum organochlorine concentrations among neighbors of a municipal solid waste incinerator. *Environ Int, 37*, 449-453.

91. Hardell, K., Carlberg, M., Hardell, L., *et al*. (2009). Concentrations of organohalogen compounds and titres of antibodies to Epstein-Barr virus antigens and the risk for non-Hodgkin's lymphoma. *Oncol Rep 21*, 1567-1576.

92. Quintana, P.J., Delfino, R.J., Korrick, S., *et al*. (2004). Adipose tissue levels of organochlorine pesticides and polychlorinated biphenyls and risk of non-Hodgkin's lymphoma. *Environ Health Perspect, 112*, 854-861.

93. Brauner, E.V., Sorensen, M., Gaudreau, E., *et al*. (2012). A prospective study of organochlorines in adipose tissue and risk of non-Hodgkin lymphoma. *Environ Health Perspect, 120*(1), 105-111.

94. Zheng, R., Zhang, Q., Zhang, Q., et al. (2015). Occupational exposure to pentachlorophenol causing lymphoma and hematopoietic malignancy for two generations. *Toxicology and Industrial Health, 31*(4), 328-342.

95. Freeman, M.D., & Kohles, S.S. (2012). Plasma levels of polychlorinated biphenyls, non-Hodgkin lymphoma, and causation. *J Environmental and Public Health*, Article ID 258981, doi:10.1155/2012/258981.

96. Kramer, S., Hikel, S.M., Adams, K., *et al*. (2012). Current status of the epidemiologic evidence linking polychlorinated biphenyls and non-Hodgkin lymphoma, and the role of immune dysregulation. *Environ Health Perspect, 120*(8), 1067-1075.

97. Strauss, H.S., & Heiger-Bernays, W. (2012). Methodological limitations may prevent the observation of non-Hodgkin's lymphoma in bioassays of polychlorinated biphenyls. *Toxicologic Pathology, 40*, 995-1003.

98. Müller, A.M.S., Ihorst, G., Mertelsmann, R., & Engelhardt, M. (2005). Epidemiology of non-Hodgkin's lymphoma (NHL): trends, geographic distribution, and etiology*. Ann Hematol, 84*, 1-12

99. Mehlman, M.A. (2006). Causal relationship between non-Hodgkin's lymphoma and exposure to benzene and benzene-containing solvents*. Ann N.Y. Acad Sci, 1076*, 120-128.

100. Cocco, P., T'Mannetje, A., Fadda, D., *et al*. (2010). Occupational exposure to solvents and risk of lymphoma subtypes: results from the Epilymph case-control study. *Occup Environ Med, 67*(5), 341-347.

101. Smith, M.T., Jones, R.M., & Smith, A.H. (2007). Benzene exposure and risk of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev, 16*(3), 385-391

102. Vlaanderen, J., Straif, K., Pukkala, E., *et al*. (2013). Occupational exposure to trichloroethylene and perchloroethylene and the risk of lymphoma, liver, and kidney cancer in four Nordic countries. *Occup Environ Med, 70*, 393-401.

103. Brown, T. & Rushton, L. (2012). Occupational cancer in Britain. Haematopoietic malignancies: leukaemia, multiple myeloma, non-Hodgkin's lymophoma*. Brit J Cancer, 107*, s41-s48.

104. Floret, N., Lucot, E., Badot, P.M., *et al*. (2007). A municipal solid waste incinerator as the single dominant point source of PCDD/Fs in an area of increased non-Hodgkin's lymphoma incidence*. Chemosphere, 68*, 1419-1426.

105. International Agency for Research on Cancer (IARC) Working Group. (2012, June 25). *IARC Monograph: 1,3-Butadiene*. Retrieved March 1, 2015, from http://monographs.iarc.fr/ENG/Monographs/vol100F/mono100F-26.pdf.

106. Kelly, R.S., Lundh, T., Porta, M., *et al*. (2013). Blood erythrocyte concentrations of cadmium and lead and the risk of B-cell non-Hodgkin's lymphoma and multiple myeloma: a nested case-control study. *PLoS One, 8*(11), Open Access, retrieved

January 8, 2016, from
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0081892.

107. Cocco, P., Broccia, G., Aru, P., *et al*. (2003). Nitrate in community water supplies and incidence of non-Hodgkin's lymphoma in Sardinia, Italy. *J Epidemiol Community Health, 57*(7), 510-511.

108. Fazzi, R., Manetti, C., Focosi, D., *et al*. (2010). Areas with high soil percolation by herbicides have higher incidence of low-grade non-Hodgkin lymphomas. *Ann Hematol, 89*, 941-943.

109. Comba, P., Ricci, P., Iavarone, I., *et al*. (2014). Cancer incidence in Italian contaminated sites*. Ann Ist Super Sanità, 50*(2), 186-191.

110. Sorensen, M., Poulsen, A.H., Ketzel, M., *et al*. (2015). Residential exposure to traffic noise and risk for non-Hodgkin lymphoma among adults*. Environ Res, 142*, 61-65.

111. De Roos, A.J., Davis, S., Colt, J.S., *et al*. (2010). Residential proximity to industrial facilities and risk of non-Hodgkin lymphoma. *Environ Res, 110*(1), 70-78.

112. Ramis, R., Vidal, E., Garcia-Perez, J., *et al*. (2009). Study of non-Hodgkin's lymphoma mortality associated with industrial pollution in Spain, using Poisson models*. BMC Public Health, 9*(26). Open Access, retrieved January 8, 2016, from http://www.biomedcentral.com/1471-2458/9/26.

113. Johnson, K.C., Pan, S., Fry, R., *et al*. (2003). Residential proximity to industrial plants and non-Hodgkin lymphoma. *Epidemiology, 14*(6), 687-693.

114. Ramis, R., Diggle, P., Boldo, E., *et al*. (2012). Analysis of matched geographical areas to study potential links between environmental exposure to oil refineries and non-Hodgkin lymphoma mortality in Spain. *Int J Health Geogr, 11*(4). Open Access, retrieved January 8, 2016, from http://www.ij-healthgeographics.com/content/pdf/1476-072X-11-4.pdf.

115. Kristbjornsdottir, A., & Rafnsson, V. (2015). Cancer mortality and other causes of death in users of geothermal hot water*. Acta Oncologica, 54*, 115-123.

116. Grulich, A.E., Vajdic, C.M., & Cozen, W. (2007). Altered immunity as a risk factor for non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev, 16*(3), 405-408.

117. Engels, E.A., Cerhan, J.R., Linet, M.S., *et al*. (2005). Immune-related conditions and immune-modulating medications as risk factors for non-Hodgkin's lymphoma: a case-control study. *Am J Epidemiol, 162*(12), 1153-1161.

118. Vajdic, C.M., Falster, M.O., de SanJose, S., *et al*. (2009). Atopic disease and risk of non-Hodgkin lymphoma: an InterLymph pooled analysis. *Cancer Research, 69*(16), 6482-6489.

119. Hoffman, J., Hoppin, J., Blair, A., *et al*. (2014). Presentation 0105: Farm exposures, allergy symptoms and risk of non-Hodgkin lymphoma in the Agricultural Health Study. *Occup Environ Med, 71*(Suppl 1), A11.

120. Pahwa, M., Harris, S.A., Hohenadel, K., *et al*. (2012). Pesticide use, immunologic conditions, and risk of non-Hodgkin lymphoma in Canadian men in six provinces. *Int J Cancer, 131*, 2650-2659.

121. Zhou, M.H., & Yang, Q.M. (2015). Association of asthma with the risk of acute leukemia and non-Hodgkin lymphoma. *Molecular Clin Oncol, 3*, 859-864.

122. Han, X., Wang, J., Shen, Y., *et al*. (2015). CRM1 as a new therapeutic target for non-Hodgkin lymphoma. *Leukemia Res, 39*, 38-46.

123. Shiels, M.S., Engels, E.A., Linet, M.S., *et al*. (2013). The epidemic of non-Hodgkin lymphoma in the United States: disentangling the effect of HIV. *Cancer Epidemiol Biomarkers Prev, 22*(6), 1069-1078.

124. Chihara, D., Ito, H., Matsuda, T., *et al*. (2014). Differences in incidence and trends of haematological malignancies in Japan and the United States*. Brit J Haematology, 164*, 536-545.

125. U.S. Department of Health & Human Services. *A timeline of AIDS*. Retrieved March 1, 2015, from https://www.aids.gov/hiv-aids-basics/hiv-aids-101/aids-timeline/.

126. Wang, L., He, X., Bi, Y., & Ma, Q. (2012). Stem cell and benzene-induced malignancy and hematotoxicity. *Chem Res Toxicol, 25*, 1303-1315.

127. Waalkes, M.P., Rehm, S., Sass, B., & Ward, J.M. (1992). Induction of tumours of the hematopoietic system by cadmium in rats. *IARC Sci Publ, 118*, 401-404.

128. Demir, C., Demir, H., Esen, R., *et al*. (2011). Altered serum levels of elements in acute leukemia cases in Turkey. *Asian Pac J Cancer Prev, 12*, 3471-3474.

129. Alavanja. M.C.R., Ross, M.K., & Bonner, M.R. (2013). Increased cancer burden among pesticide applicators and others due to pesticide exposure. *CA Cancer J Clin, 63*, 120-142.

130. dela Cruz, A.L.N., Cook. R.L., Dellinger, B., *et al*. (2014). Assessment of environmentally persistent free radicals in soils and sediments from three Superfund sites. *Environ Sci Processes Impacts, 16*, 44-52.

131. Reed, J.R., dela Cruz, A.L.N., Lomnicki, S.M., & Backes, W.L. (2015). Inhibition of cytochrome p450 2b4 by environmentally persistent free radical-containing particulate matter. *Biochemical Pharmacology, 95*, 126-132.

132. Zahzeh, M.R., Luokidi, B., Meziane, W., *et al*. (2015). Relationship between NADPH and Th1/Th2 ratio in patients with non-Hodgkin lymphoma who have been exposed to pesticides. *J Blood Med, 6*, 99-107.

133. Wang, S.S. & Nieters, A. (2010). Unraveling the interactions between environmental factors and genetic polymorphisms in non-Hodgkin lymphoma risk. *Expert Rev Anticancer Ther, 10*(3), 403-413.

134. Liu, J., Huang, J., Zhang, Y., *et al*. (2013). Identification of gene-environment interactions in cancer studies using penalization. *Genomics, 102*, 189-194.

135. Chiu, B.C., Dave, B.J., Blair, A., *et al*. (2006). Agricultural pesticide use and risk of t(14;18)-defined subtypes of non-Hodgkin lymphoma. *Blood, 108*, 1363-1369.

136. Schroeder, J.C., Olshan, A.F., Baric, A., *et al*. (2001). Agricultural risk factors for t(14;18) subtypes of non-Hodgkin's lymphoma. *Epidemiology, 12*, 701-709.

137. Chiu, B.C., & Blair, A. (2009). Pesticides, chromosomal aberrations, and non-Hodgkin's lymphoma. *J Agromedicine, 14*, 250-255.

138. Ambinder, A.J., Shenoy, P.J., Malik, N., *et al*. (2012). Exploring risk factors for follicular lymphoma. *Adv Hematol, 626035*. Open Access, retrieved January 8, 2016, from http://www.hindawi.com/journals/ah/2012/626035/.

139. Colt, J.S., Rothman, N., Severson, R.K., *et al*. (2009). Organochlorine exposure, immune gene variation, and risk of non-Hodgkin lymphoma. *Blood, 113*, 1899-1905.

140. Kelly, R.S. & Vineis, P. (2014). Biomarkers of susceptibility to chemical carcinogens: the example of non-Hodgkin lymphomas. *Brit Med Bulletin, 111*, 89-100.

141. Wang, S.S., Vajdic, C.M., Linet, M.S., *et al*. (2015). Associations of non-Hodgkin's lymphoma (NHL) risk with autoimmune conditions according to putative NHL locl. *Am J Epidemiol, 181*(6), 406-421.

142. Rendleman, J., Antipin, Y., Reva, B., *et al*. (2014). Genetic variation in DNA repair pathways and risk of non-Hodgkin's lymphoma. *PLoS One, 9*(7), Open Access, retrieved January 8, 2016, from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0101685.

143. U.S. National Library of Medicine, TOXMAP. Retrieved August 27, 2015 from http://toxmap-beta.nlm.nih.gov/toxmap/flex/.

144. Anderson, R.N., & Rosenberg, H.M. (1998). Report of the second workshop on age adjustment: National Center for Health Statistics. *Vital Health Stat, 4*(30), 1-35.

145. Klein, R.J., & Schoenborn, C.A. (2001). Age adjustment using the 2000 projected U.S. population. *Healthy People Statistical Notes* (no. 20). Hyattsville, Maryland: National Center for Health Statistics.

146. United States Environmental Protection Agency (EPA). (2015, December 11). *Search for Superfund Sites Where You Live.* Retrieved January 8, 2016, from http://www.epa.gov/superfund/search-superfund-sites-where-you-live.

147. United States Census Bureau. *2010 Geographic Terms and Concepts: Block Groups*. Retrieved July 14, 2015, from https://www.census.gov/geo/reference/gtc/gtc_bg.html.

148. Goovaerts, P., Xiao, H., Adunlin, G., *et al.* (2015). Geographically-weighted regression analysis of percentage of late-stage prostate cancer diagnosis in Florida. *Applied Geography, 62*, 191-200.

149. Legg, R., & Bowe, T. (2009). Applying geographically weighted regression to a real estate problem. *ArcUser, 12*(2), 44-45.

150. ESRI Resources (2009). ArcGIS Desktop Help 9.3: Interpreting GWR Results. Retrieved December 3, 2015, from http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Interpreting_GWR_results.

151. Bulka, C., Nastoupil, L.J., McClellan, W., *et al*. (2013). Residence proximity to benzene release sites is associated with increased incidence of non-Hodgkin lymphoma. *Cancer, 119*(18), 3309-3317.

152. American Cancer Society. (2015, March 11). *Survival Rates and Factors that affect Prognosis (Outlook) for non-Hodgkin Lymphoma*. Retrieved October 6, 2015, from http://www.cancer.org/cancer/non-hodgkinlymphoma/detailedguide/non-hodgkin-lymphoma-factors-prognosis.

153. Ortman, J.M., Velkoff, V.A., & Hogan, H. (2014). An aging nation: the older population in the United States. *Current Population Reports P25-1140*. Retrieved October 6, 2015, from https://www.census.gov/prod/2014pubs/p25-1140.pdf.

154. Christian, W.J., Huang, B., Rinehart, J., & Hopenhayn, C. (2011). Exploring geographic variation in lung cancer incidence in Kentucky using a spatial scan statistic: Elevated risk in the Appalachian coal-mining region. *Public Health Reports, 126*, 789-796.

155. Hendryx, M., Fedorko, E., & Anesetti-Rothermel, A. (2010). A geographical information system-based analysis of cancer mortality and population exposure to coal mining activities in West Virginia, United States of America. *Geospatial Health, 4*(2), 243-256.

156. Appalachian Community Cancer Network. *The Cancer Burden in Appalachia, 2009*. Retrieved October 26, 2015, from http://www.accnweb.com/docs/CancerBurdenAppalachia2009.pdf.

157. Anderson, R.T., Yang, T.C., Matthews, S.A., *et al*. (2014). Breast cancer screening, area deprivation, and later-stage breast cancer in Appalachia: Does geography matter? *Health Services Res, 49*(2), 546-567.

158. McGurty, E.M. (1997). From NIMBY to civil rights: the origins of the environmental justice movement. *Env History, 2*(3), 301-323.

159. Hermansson, H. (2007). The ethics of NIMBY conflicts. *Ethic Theory Moral Prac, 10*(1), 23-34.

160. Ash, M., & Fetter, T.R. (2004). Who lives on the wrong side of the environmental tracks? Evidence from the EPA's Risk-Screening Environmental Indicators Model. *Soc Sci Quarterly, 85*(2), 441-462.

161. New York State Department of Health. *Love Canal: A Special Report to the Governor & Legislature: April 1981*. Retrieved November 15, 2015, from https://www.health.ny.gov/environmental/investigations/love_canal/lcreport.htm.

162. Associated Press. (1983, September 7). *Federal buyout of Times Beach, MO, begins*. New York Times. Retrieved November 15, 2015, from http://www.nytimes.com/1983/09/07/us/federal-buyout-of-times-beach-mo-begins.html.

163. Frantz, D. (1992, June 20). *Landmark toxic site deal struck : Environment: A developer and six chemical firms will buy the Houston neighborhood and pay more than $207 million in damages to families*. Los Angeles Times. Retrieved November 15, 2015, from http://articles.latimes.com/1992-06-20/business/fi-556_1_toxic-waste.

164. Lake, R.W. (1996). Volunteers, NIMBYs, and environmental justice: Dilemmas of democratic practice. *Antipode, 28*(2), 160-174.

165. Eilperin, J. (2004, November 25). *Lack of funding slows cleanup of hundreds of Superfund sites*. Washington Post. Retrieved November 15, 2015, from http://www.washingtonpost.com/wp-dyn/articles/A11246-2004Nov24.html.

166. United States Environmental Protection Agency (March 2014). *FY 2015 EPA Budget in Brief*. Retrieved November 15, 2015, from http://www2.epa.gov/sites/production/files/2014-03/documents/fy15_bib.pdf.

167. Wheeler, D.C., Waller, L.A., Cozen, W., & Ward, M.H. (2012). Spatial-temporal analysis of non-Hodgkin lymphoma risk using multiple residential locations. *Spat Spatiotemporal Epidemiol, 3*(2), 163-171.

168. Univerisity of Texas, M.D. Anderson Cancer Center (2015). *Non-Hodgkin's Lymphoma Diagnosis*. Retrieved November 15, 2015, from http://www.mdanderson.org/patient-and-cancer-information/cancer-information/cancer-types/non-hodgkins-lymphoma/diagnosis/index.html.

APPENDIX 1: INSTITUTIONAL REVIEW BOARD APPROVAL LETTER

Office of Research Integrity
IRB, IACUC, RDRC
315 Kinkead Hall
Lexington, KY 40506-0057
859 257-9428
fax 859 257-8995
www.research.uky.edu/ori/

Initial Review

Approval Ends
September 28, 2015

IRB Number
14-0644-P3H

TO: William Brent Webber, MS
Environmental Health & Safety, Health Behavior
College of Public Health
201 Environmental Health & Safety Bldg.
Speed Sort 0314
PI phone #: (859)257-7600

FROM: Chairperson/Vice Chairperson
Medical Institutional Review Board (IRB)

SUBJECT: Approval of Protocol Number 14-0644-P3H

DATE: October 1, 2014

On September 29, 2014, the Medical Institutional Review Board approved your protocol entitled:

Geospatial Analysis for Incidence of Breast Cancer, Bladder Cancer, and non-Hodgkin's Lymphoma by Residential Proximity to Superfund Sites in Kentucky

Approval is effective from September 29, 2014 until September 28, 2015 and extends to any consent/assent form, cover letter, and/or phone script. If applicable, attached is the IRB approved consent/assent document(s) to be used when enrolling subjects. **[Note, subjects can only be enrolled using consent/assent forms which have a valid "IRB Approval" stamp unless special waiver has been obtained from the IRB.]** Prior to the end of this period, you will be sent a Continuation Review Report Form which must be completed and returned to the Office of Research Integrity so that the protocol can be reviewed and approved for the next period.

In implementing the research activities, you are responsible for complying with IRB decisions, conditions and requirements. The research procedures should be implemented as approved in the IRB protocol. It is the principal investigators responsibility to ensure any changes planned for the research are submitted for review and approval by the IRB prior to implementation. Protocol changes made without prior IRB approval to eliminate apparent hazards to the subject(s) should be reported in writing immediately to the IRB. Furthermore, discontinuing a study or completion of a study is considered a change in the protocol's status and therefore the IRB should be promptly notified in writing.

For information describing investigator responsibilities after obtaining IRB approval, download and read the document "PI Guidance to Responsibilities, Qualifications, Records and Documentation of Human Subjects Research" from the Office of Research Integrity's IRB Survival Handbook web page [http://www.research.uky.edu/ori/IRB-Survival-Handbook.html#PIresponsibilities]. Additional information regarding IRB review, federal regulations, and institutional policies may be found through ORI's web site [http://www.research.uky.edu/ori]. If you have questions, need additional information, or would like a paper copy of the above mentioned document, contact the Office of Research Integrity at (859) 257-9428.

Roger Heischman, MD / jh
Chairperson/Vice Chairperson

APPENDIX 2: EXPLORATORY REGRESSION OUTPUT

```
**************************************************************************
Choose 1 of 22 Summary
   Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)  VIF   SA   Model
 0.06 25165.51 0.00  0.77 1.00 0.00  +EXP5KM***
 0.00 25271.59 0.00  0.17 1.00 0.00  +H4**
 0.00 25273.79 0.00  0.89 1.00 0.00  +H10*
     Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
**************************************************************************
Choose 2 of 22 Summary
       Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)  VIF   SA   Model
 0.07 25137.29 0.00  0.00 1.28 0.00  +EXP5KM***  +EXP5_10***
 0.06 25159.76 0.00  0.49 1.01 0.00  +H17***  +EXP5KM***
 0.06 25160.08 0.00  0.62 1.01 0.00  +H14***  +EXP5KM***
     Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
**************************************************************************
Choose 3 of 22 Summary
           Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)  VIF   SA   Model
 0.08 25125.05 0.00  0.00 1.32 0.00  +H17***  +EXP5KM***  +EXP5_10***
 0.08 25125.22 0.00  0.00 1.39 0.00  +EXP5KM***  +EXP5_10***  +BEALE_R***
 0.07 25129.02 0.00  0.00 1.36 0.00  -H3***  +EXP5KM***  +EXP5_10***
     Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
**************************************************************************
Choose 4 of 22 Summary
             Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)  VIF   SA   Model
 0.08 25122.72 0.00  0.00 1.39 0.00  +H14**  +EXP5KM***  +EXP5_10***  +BEALE_R***
 0.08 25122.73 0.00  0.00 1.32 0.00  +H4**  +H17***  +EXP5KM***  +EXP5_10***
 0.08 25122.77 0.00  0.00 1.51 0.00  +H17*  +EXP5KM***  +EXP5_10***  +BEALE_R**
     Passing Models
AdjR2 AICc JB K(BP) VIF SA   Model
**************************************************************************
Choose 5 of 22 Summary
               Highest Adjusted R-Squared Results
AdjR2    AICc   JB K(BP)  VIF   SA   Model
 0.08 25119.56 0.00  0.00 1.38 0.00  +H4***  -H11***  +H17***  +EXP5KM***  +EXP5_10***
```

0.08 25121.71 0.00 0.00 1.56 0.00 +H4* +H17* +EXP5KM*** +EXP5_10*** +BEALE_R*
0.08 25121.94 0.00 0.00 1.39 0.00 +H4 +H14** +EXP5KM*** +EXP5_10*** +BEALE_R***
Passing Models
AdjR2 AICc JB K(BP) VIF SA Model
****
********** Exploratory Regression Global Summary (R_CT_ADJ) **********

Percentage of Search Criteria Passed

Search Criterion	Cutoff	Trials	# Passed	% Passed
Min Adjusted R-Squared	> 0.50	24509	0	0.00
Max Coefficient p-value	< 0.05	24509	86	0.35
Max VIF Value	< 7.50	24509	22413	91.45
Min Jarque-Bera p-value	> 0.10	24509	0	0.00
Min Spatial Autocorrelation p-value	> 0.10	18	0	0.00

Summary of Variable Significance

Variable	% Significant	% Negative	% Positive
EXP5KM	100.00	0.00	100.00
H4	50.39	0.00	100.00
H5	36.12	85.02	14.98
H2	25.26	14.62	85.38
H17	24.22	3.32	96.68
EXP5_10	19.48	76.30	23.70
H14	18.84	0.00	100.00
H9	15.10	0.24	99.76
H10	15.10	0.24	99.76
H3	13.57	72.24	27.76
H13	12.61	0.26	99.74
BEALE_R	11.20	45.77	54.23
H8	7.97	83.39	16.61
H1	4.68	58.18	41.82
H7	4.68	58.18	41.82
H11	2.50	36.29	63.71
H15	2.39	35.82	64.18
P77	0.00	14.79	85.21
H12	0.00	0.09	99.91
H16	0.00	0.59	99.41
APPAL	0.00	85.54	14.46
H6	---	---	---

Summary of Multicollinearity*

Variable	VIF	Violations	Covariates
P77	2.83	0	--------
H1	412.67	679	H8 (50.22), H15 (2.07)
H2	624.13	105	H8 (7.77), H5 (7.77), H15 (0.15)
H3	247.02	13	H8 (0.96), H4 (0.96), H5 (0.96)
H4	115.62	13	H3 (0.96), H8 (0.96), H5 (0.96)
H5	489.10	118	H8 (8.73), H2 (7.77), H3 (0.96), H4 (0.96), H15 (0.15)
H6	--------	ALL	INTERCEPT (100.00)
H7	412.67	679	H8 (50.22), H15 (2.07)
H8	361.82	1476	H1 (50.22), H7 (50.22), H5 (8.73), H2 (7.77), H15 (4.29), H3 (0.96), H4 (0.96)
H9	393.81	310	H15 (15.01), H17 (0.44), H11 (0.15), H13 (0.07)
H10	393.81	310	H15 (15.01), H17 (0.44), H11 (0.15), H13 (0.07)
H11	37.42	4	H15 (0.30), H10 (0.15), H9 (0.15), H13 (0.15), H17 (0.15)
H12	1.38	0	--------
H13	9.01	2	H11 (0.15), H15 (0.15), H17 (0.15), H10 (0.07), H9 (0.07)
H14	1.65	0	--------
H15	226.36	464	H9 (15.01), H10 (15.01), H8 (4.29), H1 (2.07), H7 (2.07), H17 (0.89), H11 (0.30), H2 (0.15), H13 (0.15), H5 (0.15)
H16	1.10	0	--------
H17	44.50	12	H15 (0.89), H10 (0.44), H9 (0.44), H11 (0.15), H13 (0.15)
EXP5KM	1.45	0	--------
EXP5_10	1.35	0	--------
APPAL	1.48	0	--------
BEALE_R	2.31	0	--------

* At least one model failed to solve due to perfect multicollinearity.
Please review the warning messages for further information.

Summary of Residual Normality (JB)

JB	AdjR2	AICc	K(BP)	VIF	SA	Model
0.000000	0.000557	25274.824611	0.030913	1.000000	0.000000	+H2
0.000000	-0.000136	25276.155398	0.159671	1.000000	0.000000	+H1
0.000000	0.000288	25275.340849	0.158239	1.000000	0.000000	+P77

Summary of Residual Spatial Autocorrelation (SA)

SA	AdjR2	AICc	JB	K(BP)	VIF	Model
0.000000	0.000557	25274.824611	0.000000	0.030913	1.000000	+H2
0.000000	0.002236	25271.593816	0.000000	0.171381	1.000000	+H4**
0.000000	0.055842	25165.510295	0.000000	0.771149	1.000000	+EXP5KM***

Table Abbreviations
AdjR2 Adjusted R-Squared
AICc Akaike's Information Criterion
JB Jarque-Bera p-value
K(BP) Koenker (BP) Statistic p-value
VIF Max Variance Inflation Factor
SA Global Moran's I p-value
Model Variable sign (+/-)
Model Variable significance (* = 0.10, ** = 0.05, *** = 0.01)

APPENDIX 3: OLS OUTPUTS AND DIAGNOSTICS

Model 1: Base Model for all NHL

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	261.259684	6.358816	41.086216	0.000000*	8.429543	30.993338	0.000000*	--------
EXP5KM	113.604440	9.379363	12.112171	0.000000*	10.761866	10.556203	0.000000*	1.280278
EXP5_10	50.817410	9.213760	5.515383	0.000000*	8.984044	5.656407	0.000000*	1.280278

OLS Diagnostics

Input Features:	MyCTfile	Dependent Variable:	R_CT_ADJ
Number of Observations:	1921	Akaike's Information Criterion (AICc) [d]:	25137.290706
Multiple R-Squared [d]:	0.071067	Adjusted R-Squared [d]:	0.070098
Joint F-Statistic [e]:	73.367082	Prob(>F), (2,1918) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	121.038705	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	14.200823	Prob(>chi-squared), (2) degrees of freedom:	0.000825*
Jarque-Bera Statistic [g]:	237324.616058	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

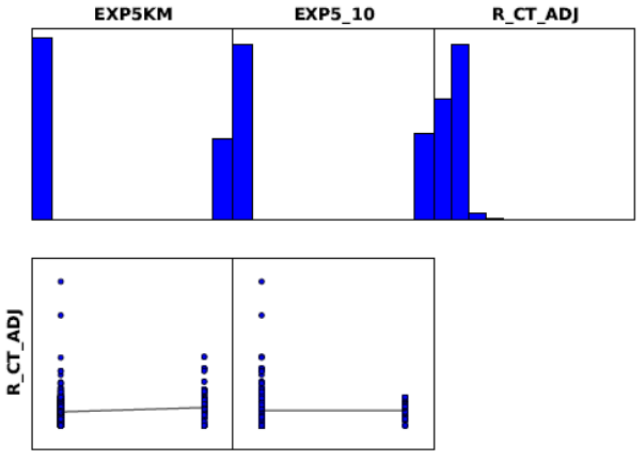
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

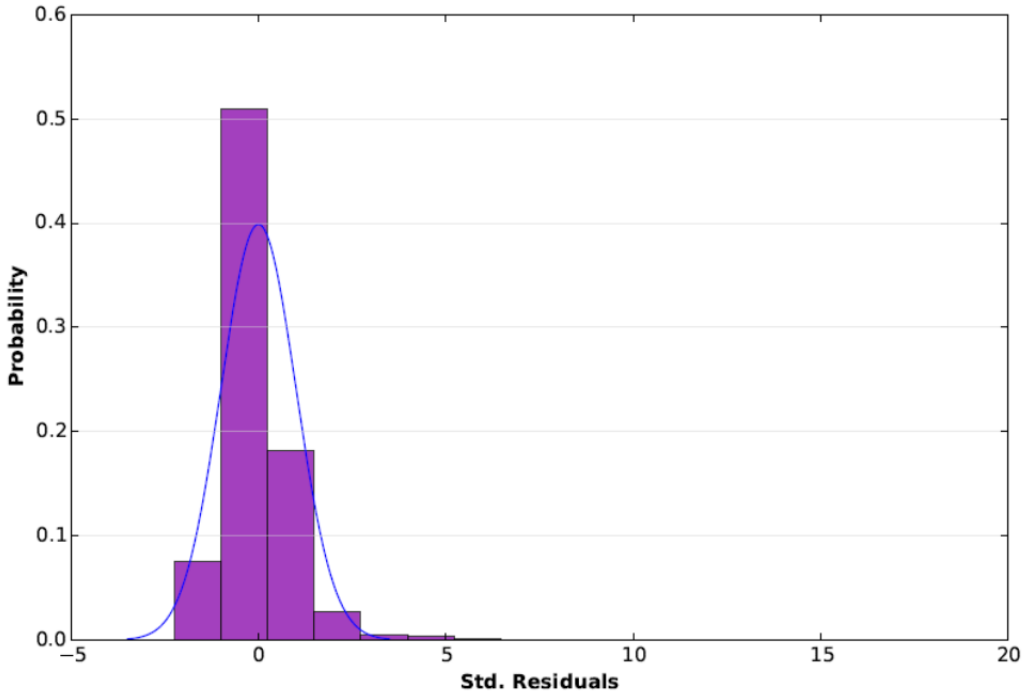
Variable Distributions and Relationships



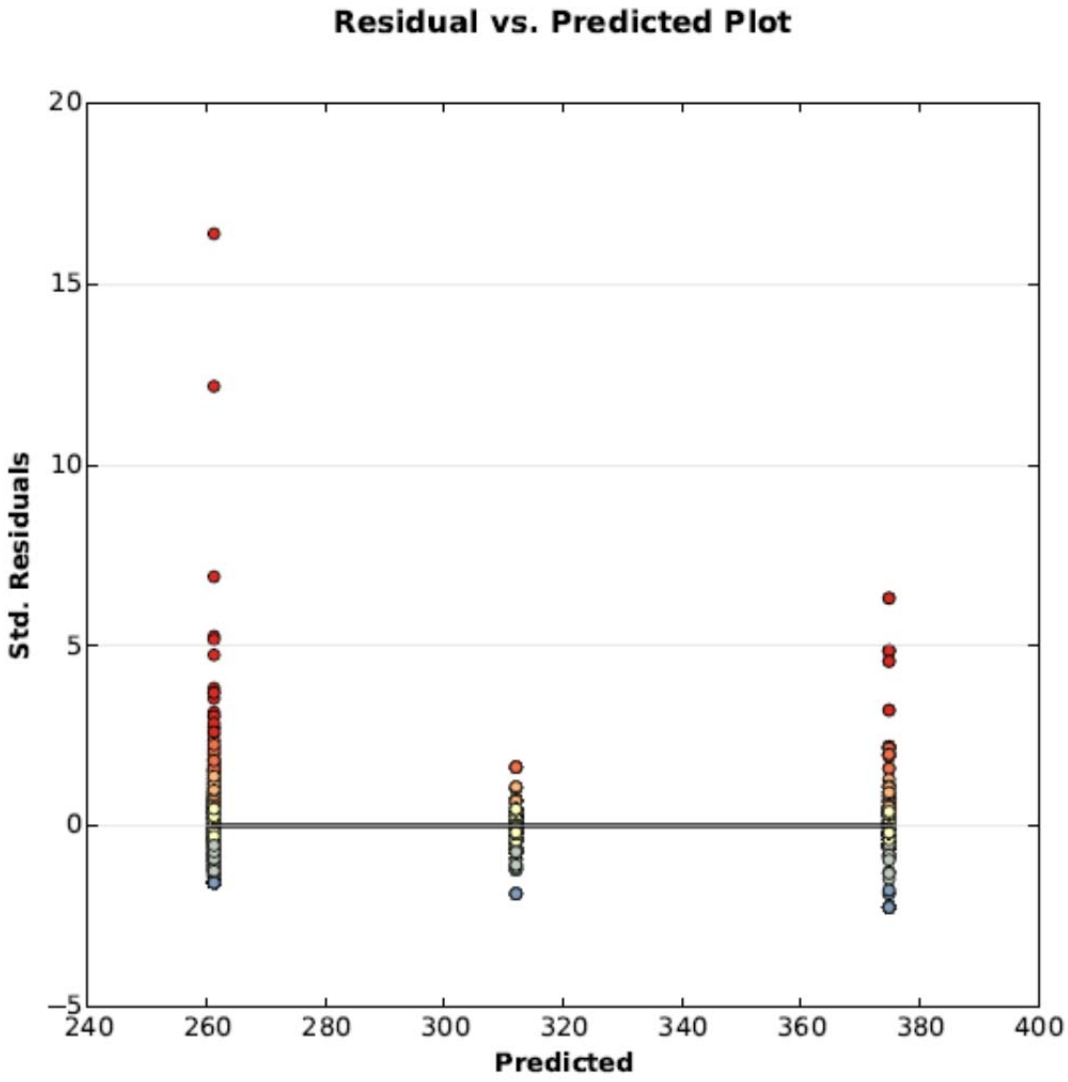
The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.

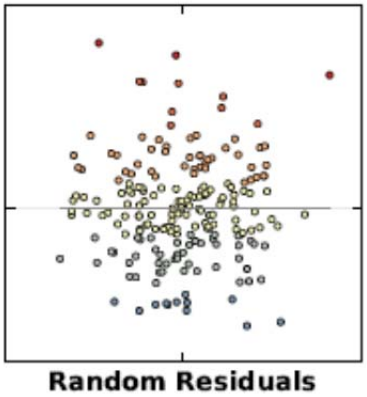
Histogram of Standardized Residuals



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Model 2: Full Model for all NHL

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	239.541128	8.611948	27.814977	0.000000*	9.928792	24.125909	0.000000*	--------
EXP5KM	123.817718	9.751359	12.697483	0.000000*	11.600214	10.673744	0.000000*	1.392617
EXP5_10	57.190135	9.364325	6.107235	0.000000*	9.113008	6.275660	0.000000*	1.330845
APPAL	-2.550922	10.199197	-0.250110	0.802532	13.362985	-0.190895	0.848622	1.432529
BEALE_R	5.804889	1.743195	3.330029	0.000901*	1.930053	3.007632	0.002678*	1.500482

OLS Diagnostics

Input Features:	MyCTfile	Dependent Variable:	R_CT_ADJ
Number of Observations:	1921	Akaike's Information Criterion (AICc) [d]:	25127.172445
Multiple R-Squared [d]:	0.077880	Adjusted R-Squared [d]:	0.075955
Joint F-Statistic [e]:	40.455172	Prob(>F), (4,1916) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	125.891550	Prob(>chi-squared), (4) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	20.076643	Prob(>chi-squared), (4) degrees of freedom:	0.000482*
Jarque-Bera Statistic [g]:	243426.330551	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

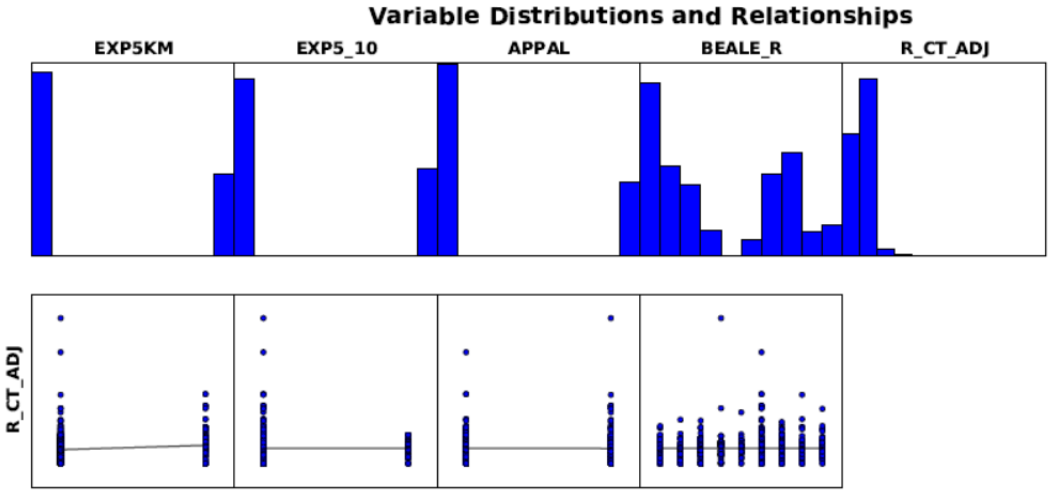
[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

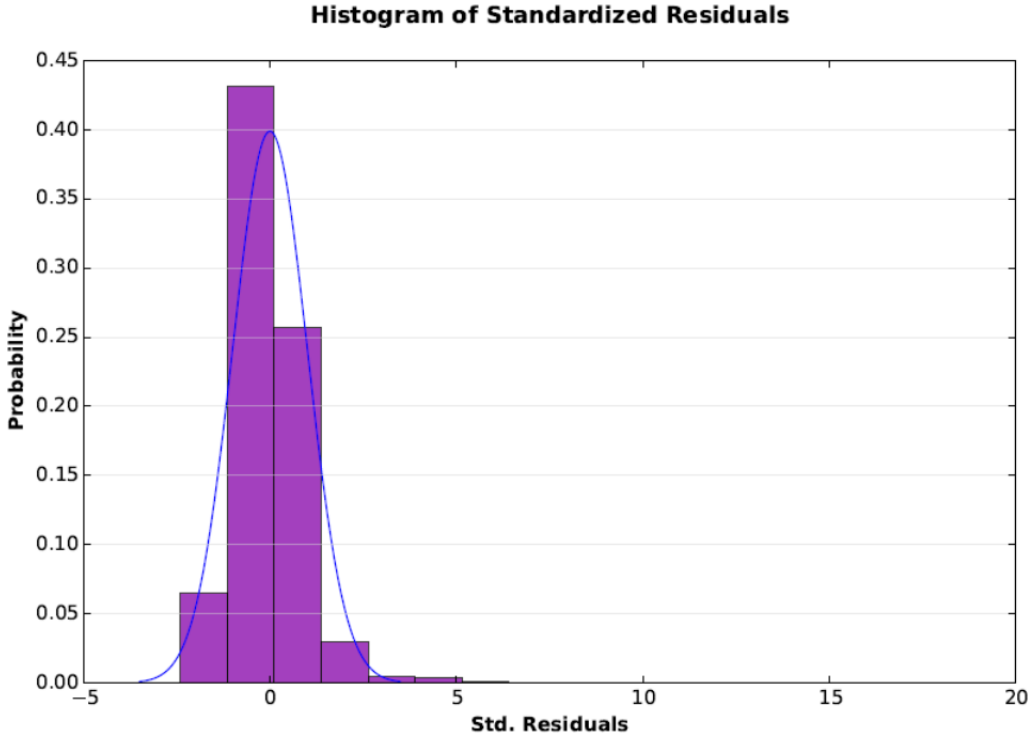
[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

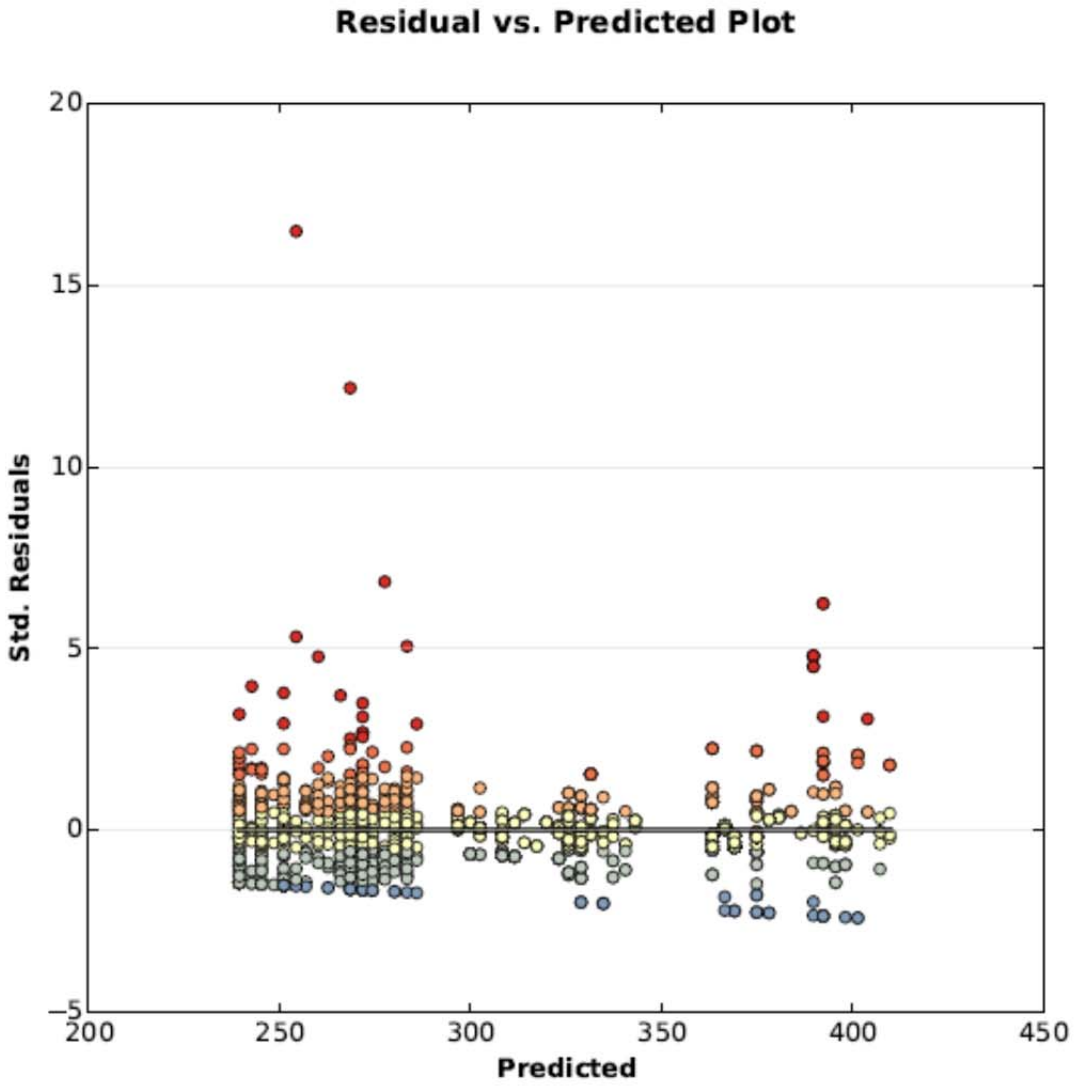


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

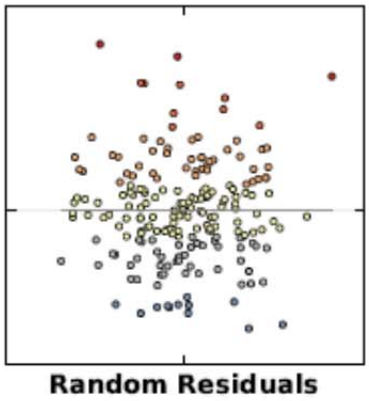
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Model 3: Base Model for Male NHL

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	292.563577	7.606924	38.460167	0.000000*	9.703136	30.151445	0.000000*	-------
EXP5KM	147.433972	11.220344	13.139880	0.000000*	13.012880	11.329850	0.000000*	1.280278
EXP5_10	59.185899	11.022236	5.369681	0.000000*	10.360951	5.712400	0.000000*	1.280278

OLS Diagnostics

Input Features:	MyCTfile	Dependent Variable:	R_CTM_ADJ
Number of Observations:	1921	Akaike's Information Criterion (AICc) [d]:	25825.841414
Multiple R-Squared [d]:	0.082925	Adjusted R-Squared [d]:	0.081969
Joint F-Statistic [e]:	86.716004	Prob(>F), (2,1918) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	137.806895	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	29.594910	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	37576.340457	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

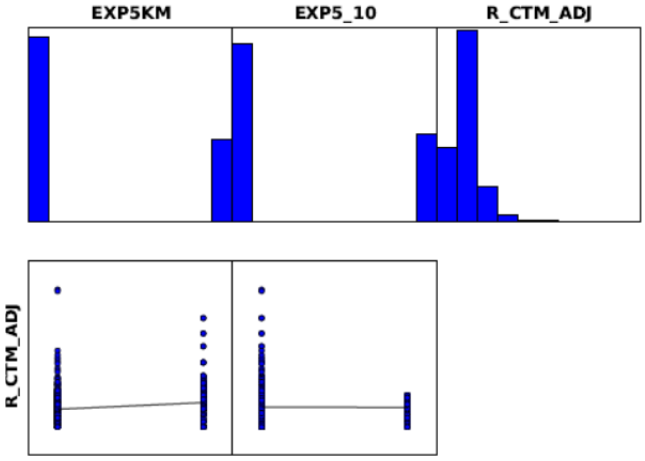
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

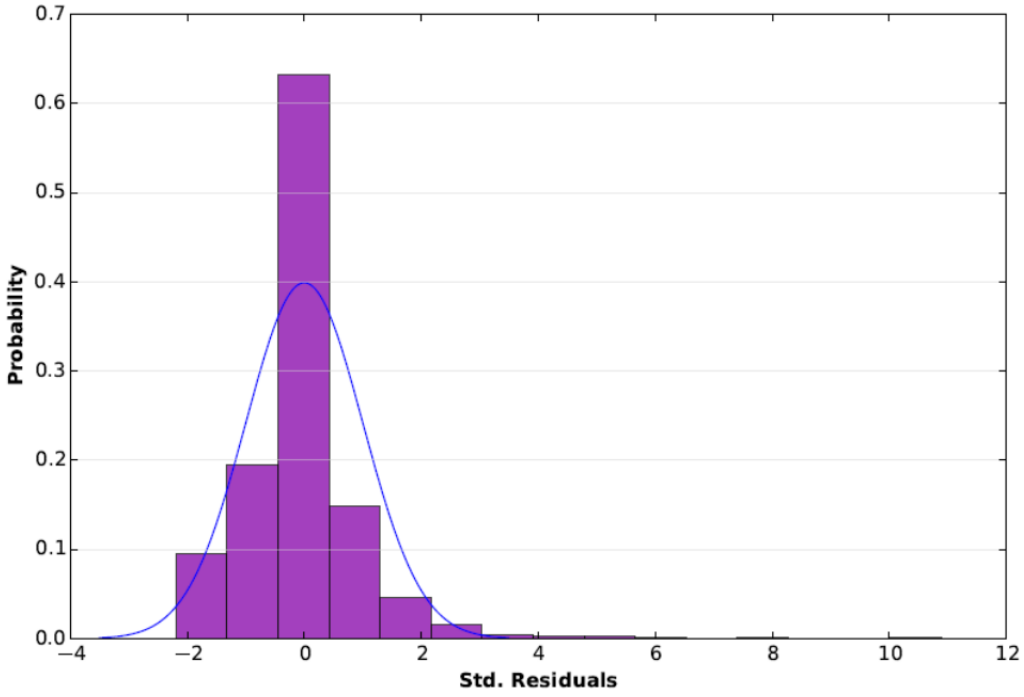
Variable Distributions and Relationships



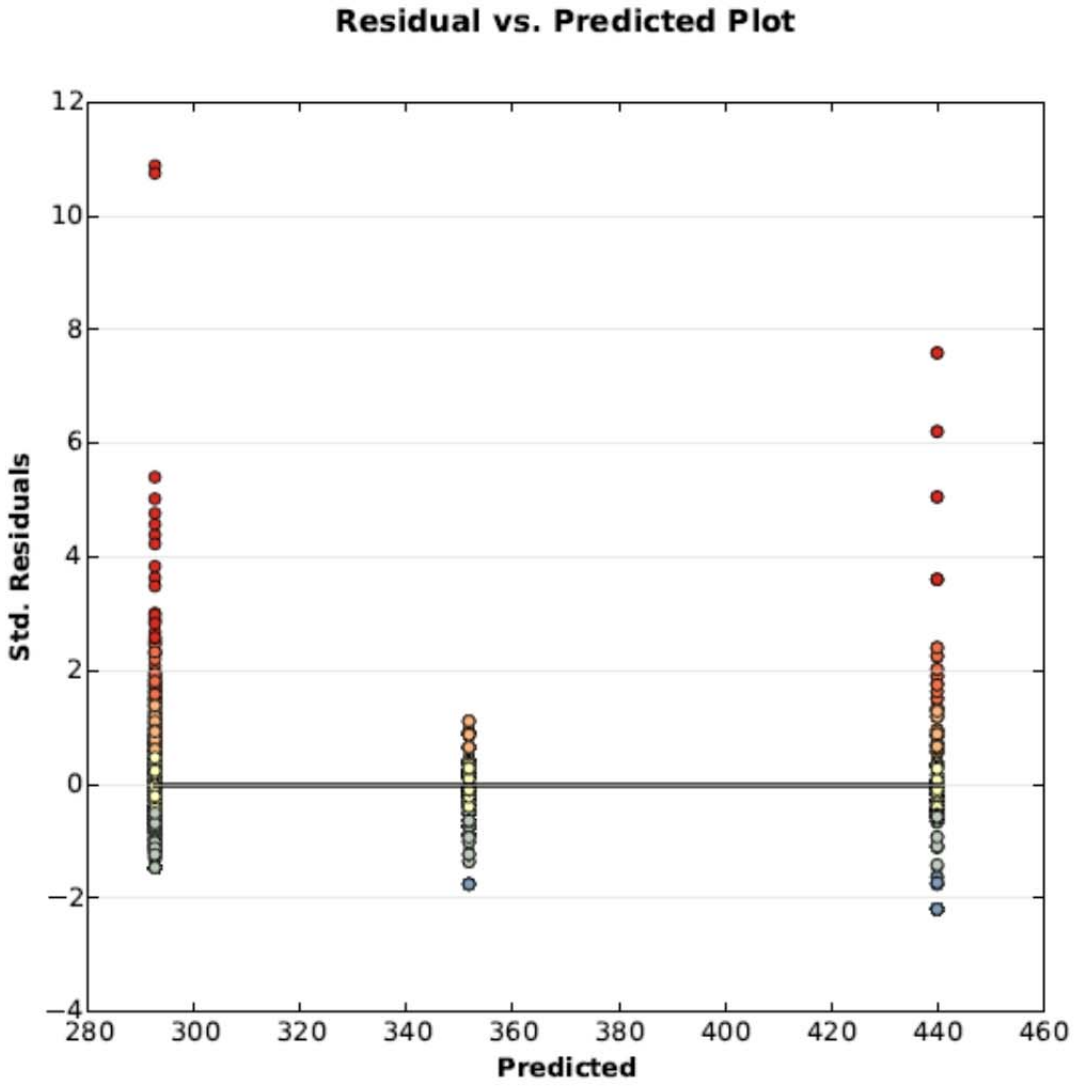
The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.

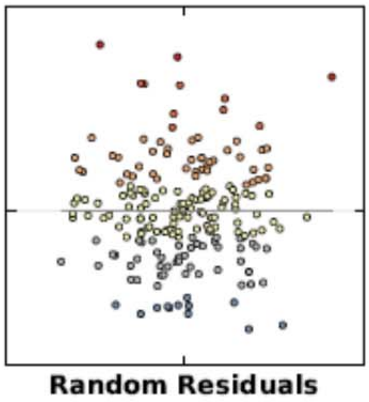
Histogram of Standardized Residuals



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Model 4: Full Model for Male NHL

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	269.529715	10.297926	26.173203	0.000000*	11.613269	23.208773	0.000000*	--------
EXP5KM	157.923351	11.660403	13.543559	0.000000*	14.169999	11.144909	0.000000*	1.392617
EXP5_10	65.034862	11.197598	5.807930	0.000000*	10.522964	6.180280	0.000000*	1.330845
APPAL	-23.889477	12.195915	-1.958810	0.050274	14.359151	-1.663711	0.096345	1.432529
BEALE R	8.279514	2.084463	3.972012	0.000082*	2.383221	3.474085	0.000540*	1.500482

OLS Diagnostics

Input Features:	MyCTfile	Dependent Variable:	R_CTM_ADJ
Number of Observations:	1921	Akaike's Information Criterion (AICc) [d]:	25814.091519
Multiple R-Squared [d]:	0.090424	Adjusted R-Squared [d]:	0.088525
Joint F-Statistic [e]:	47.619040	Prob(>F), (4,1916) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	141.038310	Prob(>chi-squared), (4) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	47.174387	Prob(>chi-squared), (4) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	36017.265368	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

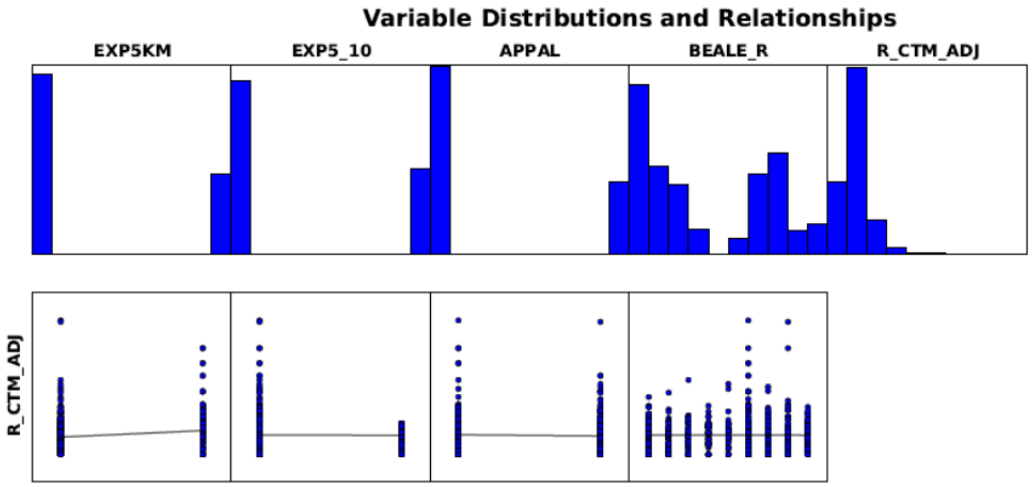
[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

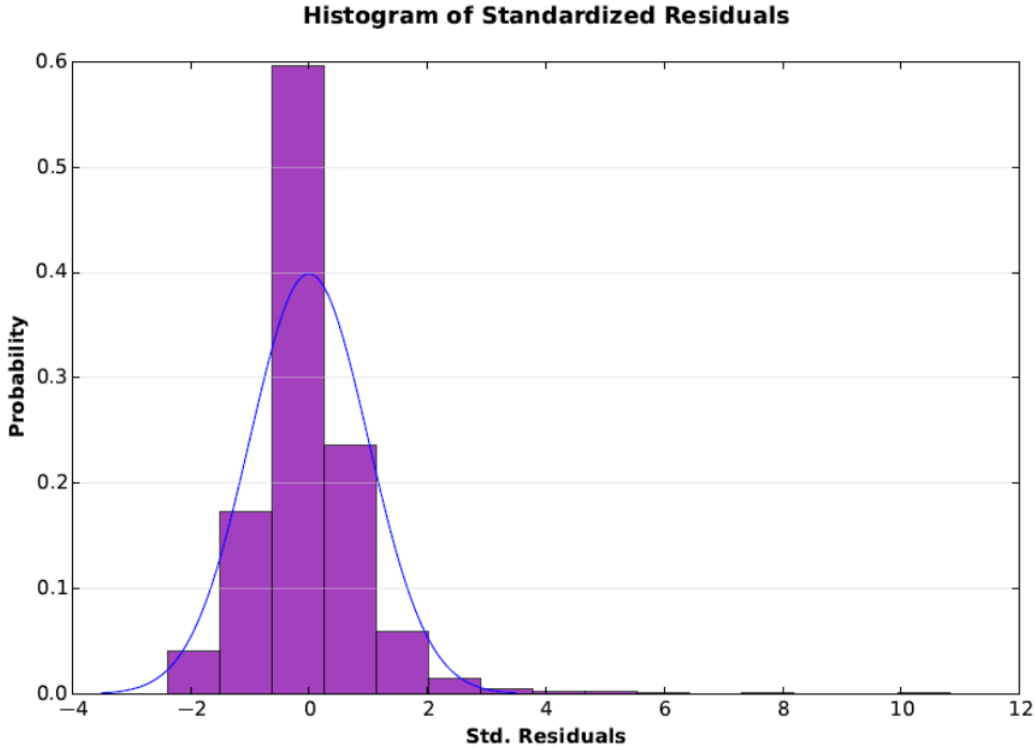
[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

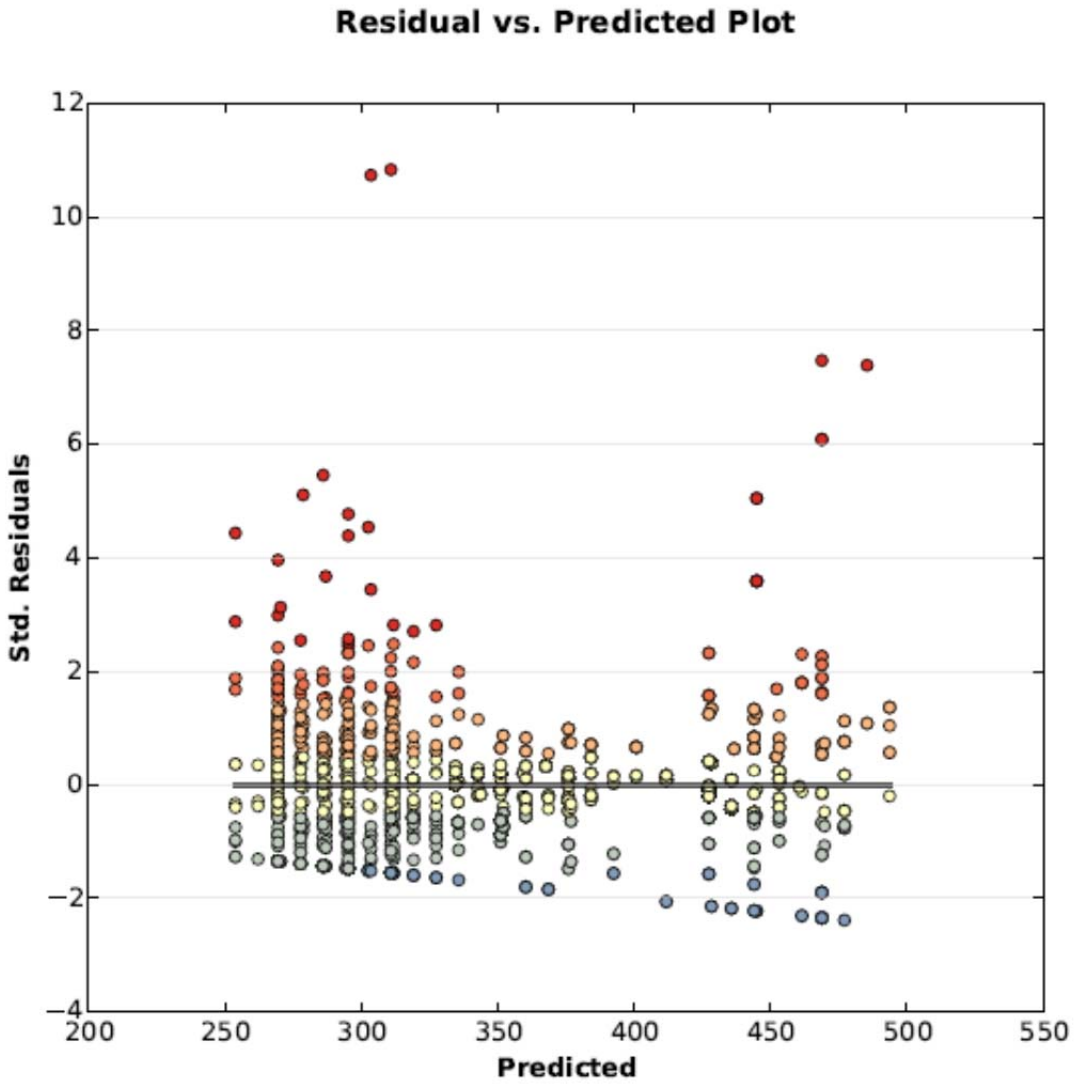


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

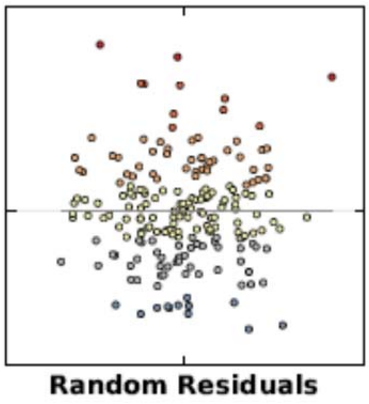
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Model 5: Base Model for Female NHL

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	235.606768	7.834149	30.074327	0.000000*	11.303290	20.844088	0.000000*	-------
EXP5KM	85.051495	11.555505	7.360258	0.000000*	13.103441	6.490776	0.000000*	1.280278
EXP5_10	44.613264	11.351480	3.930172	0.000097*	11.804499	3.779344	0.000173*	1.280278

OLS Diagnostics

Input Features:	MyCTfile	Dependent Variable:	R_CTF_ADJ
Number of Observations:	1921	Akaike's Information Criterion (AICc) [d]:	25938.924443
Multiple R-Squared [d]:	0.027618	Adjusted R-Squared [d]:	0.026604
Joint F-Statistic [e]:	27.238140	Prob(>F), (2,1918) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	50.006922	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	4.198043	Prob(>chi-squared), (2) degrees of freedom:	0.122576
Jarque-Bera Statistic [g]:	8145891.162318	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

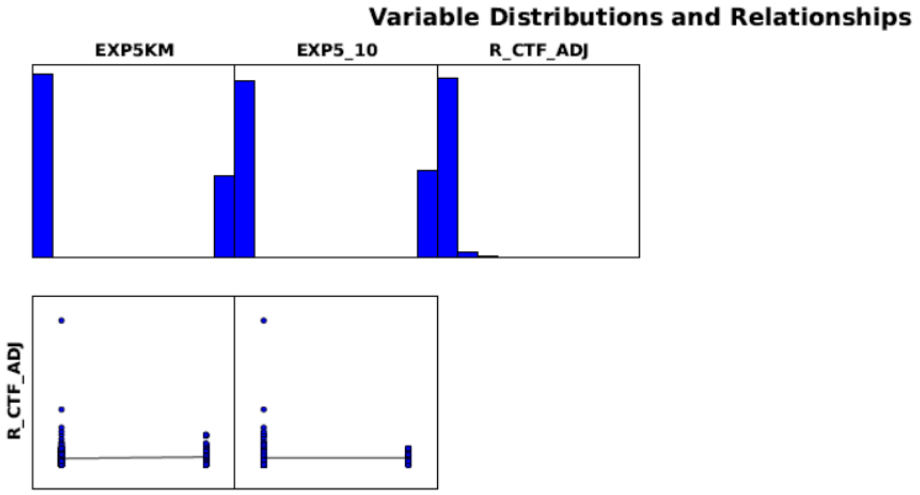
[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

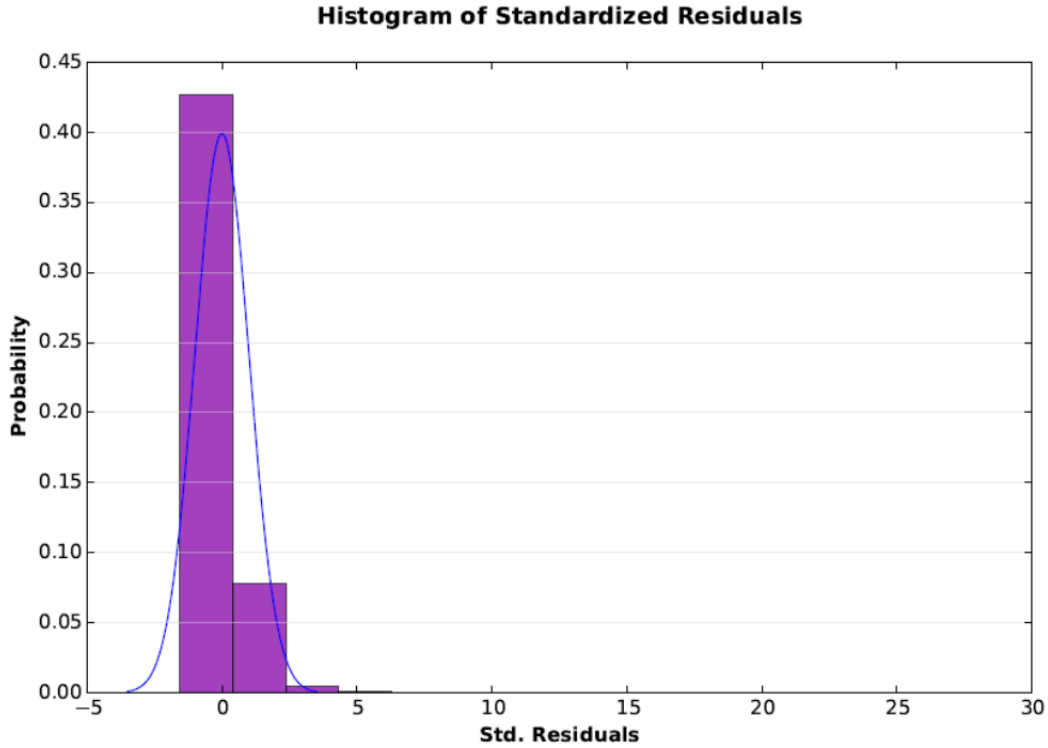
[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

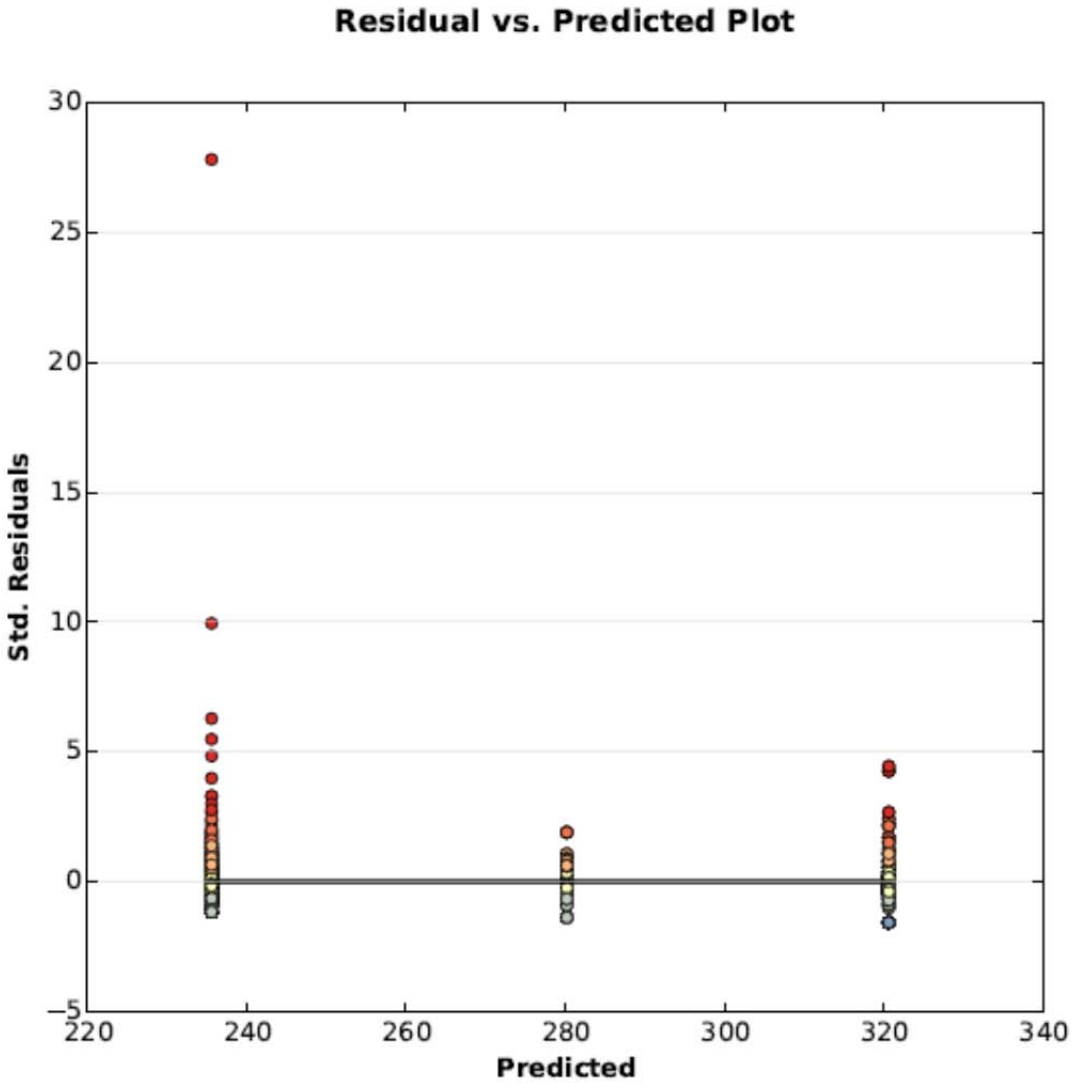


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

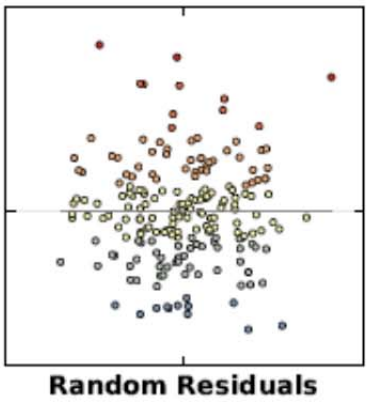
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Model 6: Full Model for Female NHL

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	215.187104	10.623452	20.255854	0.000000*	13.127248	16.392400	0.000000*	-------
EXP5KM	94.959646	12.028998	7.894228	0.000000*	13.905735	6.828812	0.000000*	1.392617
EXP5_10	51.416951	11.551564	4.451082	0.000012*	11.802920	4.356291	0.000017*	1.330845
APPAL	16.512507	12.581438	1.312450	0.189535	18.604187	0.887569	0.374869	1.432529
BEALE_R	3.562471	2.150355	1.656690	0.097758	2.494085	1.428368	0.153363	1.500482

OLS Diagnostics

Input Features:	MyCTfile	Dependent Variable:	R_CTF_ADJ
Number of Observations:	1921	Akaike's Information Criterion (AICc) [d]:	25933.660294
Multiple R-Squared [d]:	0.032308	Adjusted R-Squared [d]:	0.030288
Joint F-Statistic [e]:	15.992155	Prob(>F), (4,1916) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	66.917051	Prob(>chi-squared), (4) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	7.246854	Prob(>chi-squared), (4) degrees of freedom:	0.123404
Jarque-Bera Statistic [g]:	8223202.316077	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

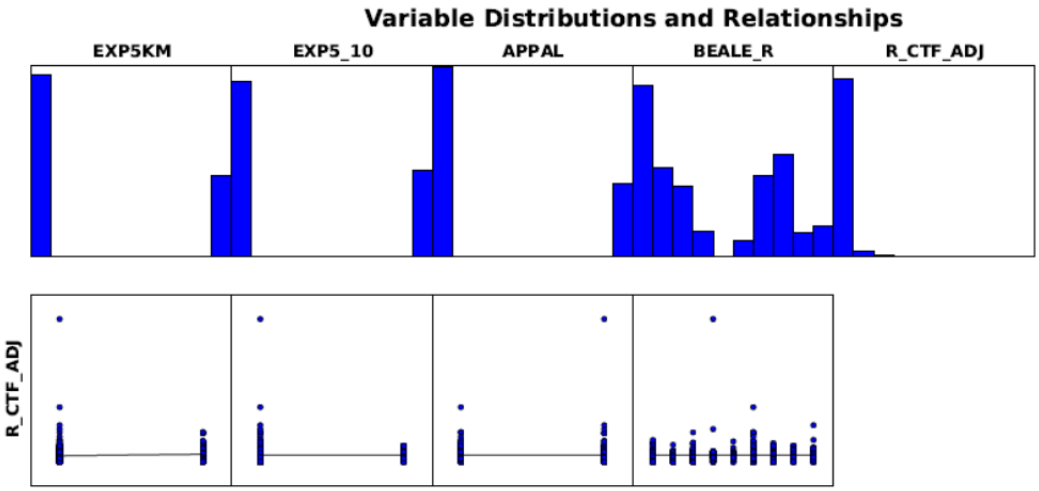
[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

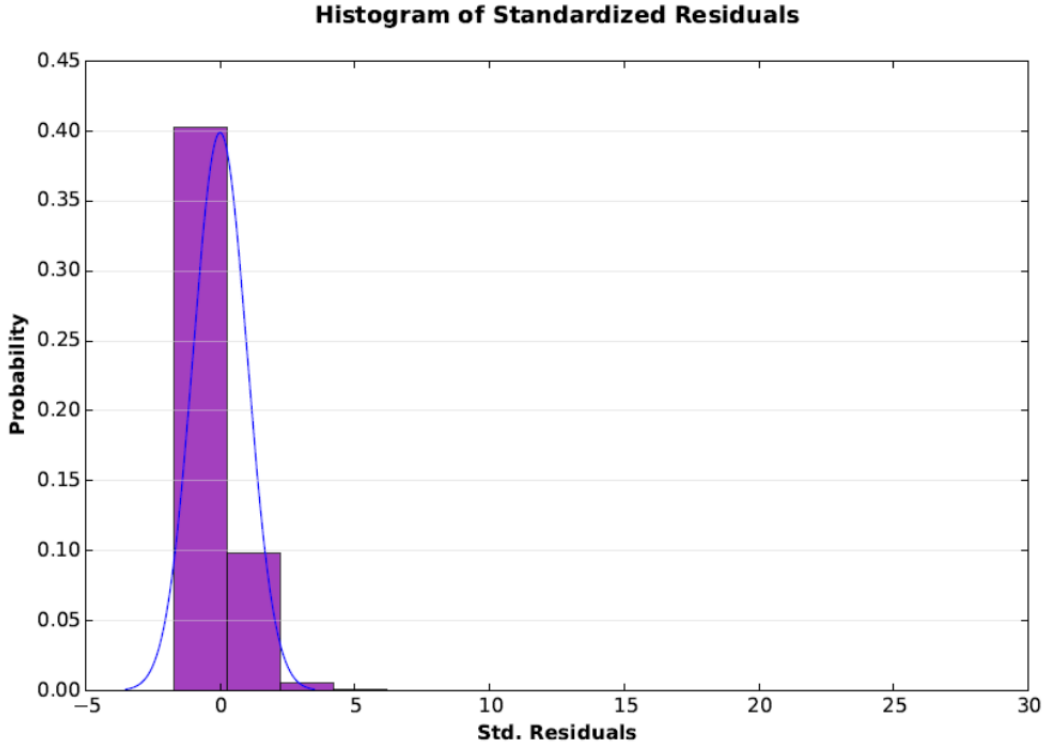
[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

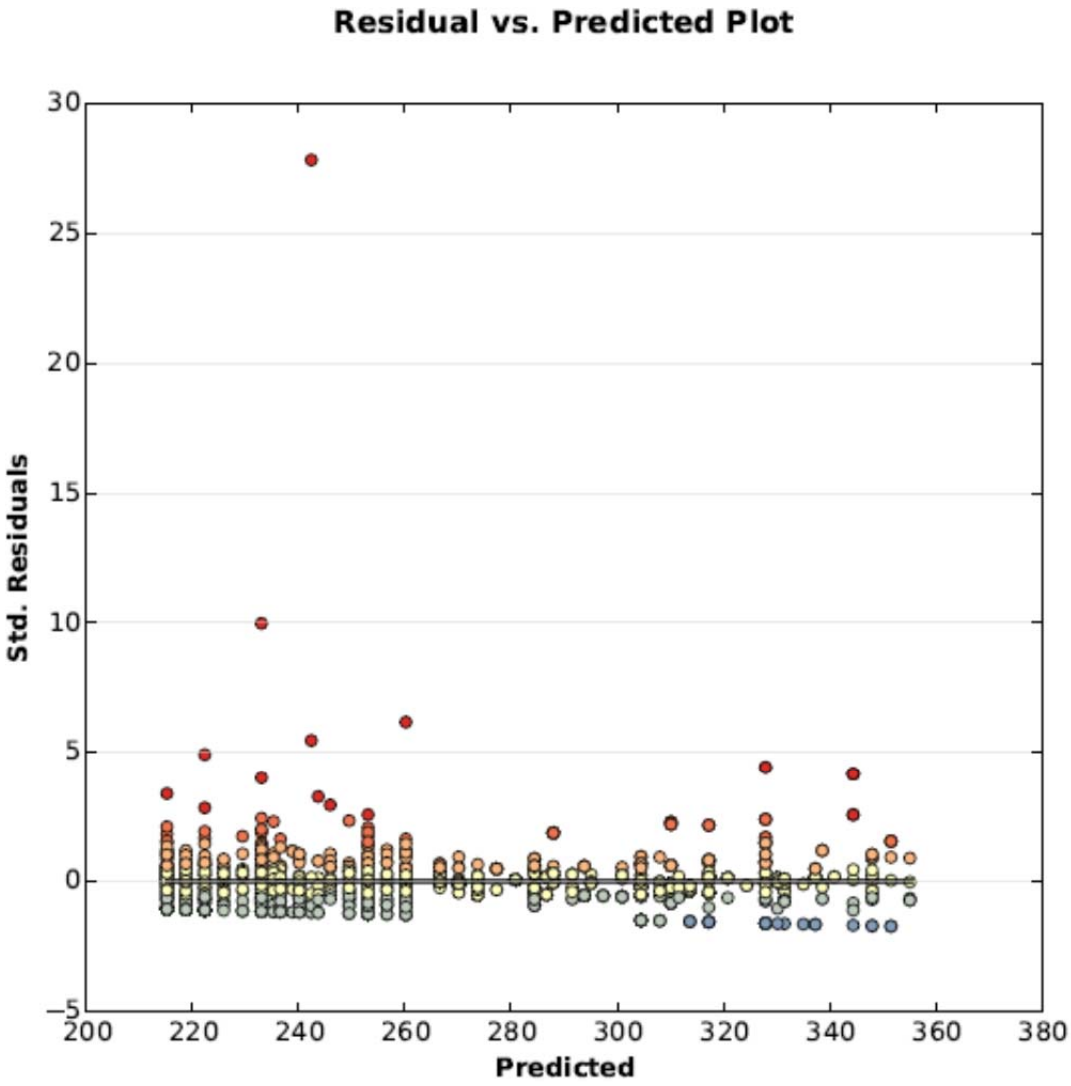


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

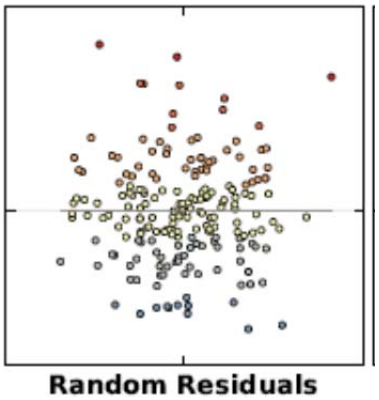
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Model 7: Base Model for Intranodal NHL

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	187.190115	5.106706	36.655748	0.000000*	6.910504	27.087765	0.000000*	--------
EXP5KM	77.697995	7.532479	10.315064	0.000000*	8.666739	8.965079	0.000000*	1.280278
EXP5_10	35.211482	7.399484	4.758640	0.000003*	7.278428	4.837787	0.000002*	1.280278

OLS Diagnostics

Input Features:	MyCTfile	Dependent Variable:	R_CTS41_AD
Number of Observations:	1921	Akaike's Information Criterion (AICc) [d]:	24294.787477
Multiple R-Squared [d]:	0.052562	Adjusted R-Squared [d]:	0.051574
Joint F-Statistic [e]:	53.203199	Prob(>F), (2,1918) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	89.273631	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	7.805278	Prob(>chi-squared), (2) degrees of freedom:	0.020189*
Jarque-Bera Statistic [g]:	1075281.853900	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

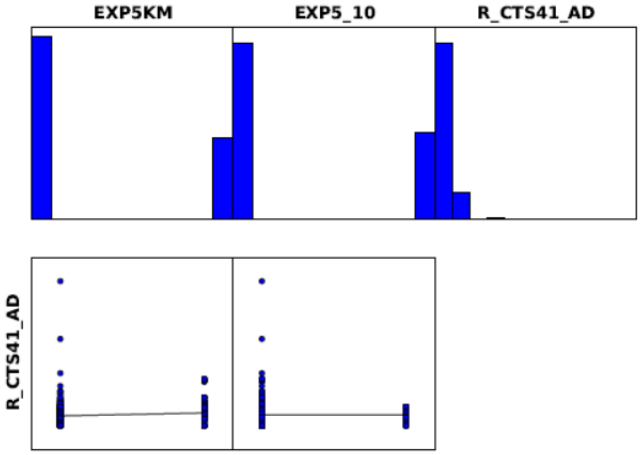
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

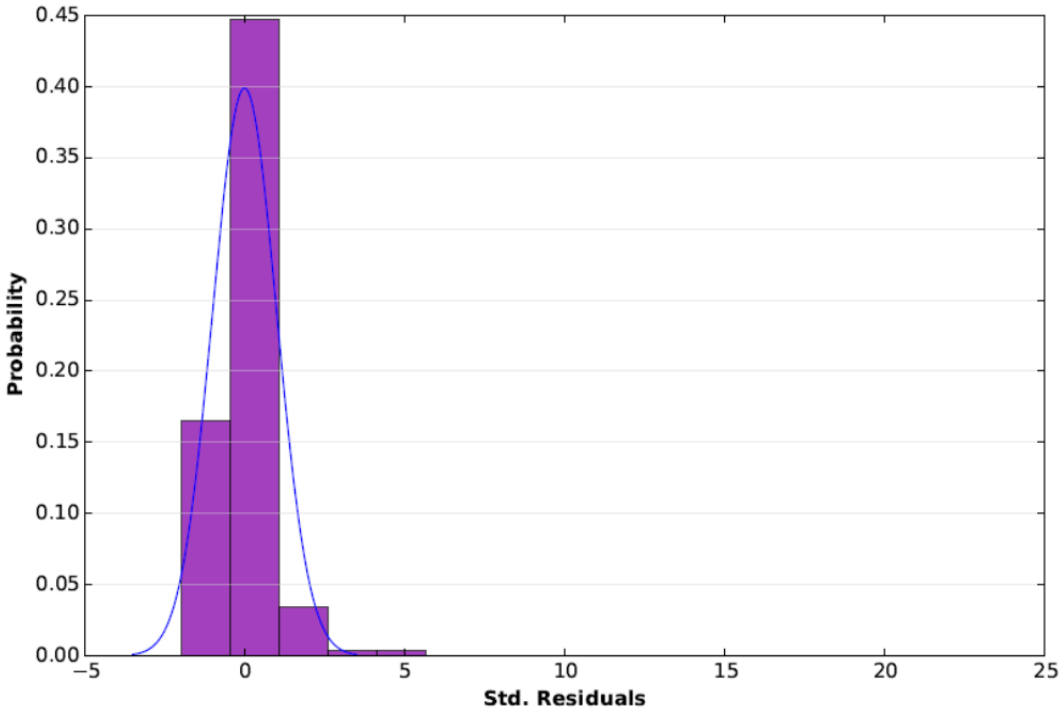
Variable Distributions and Relationships



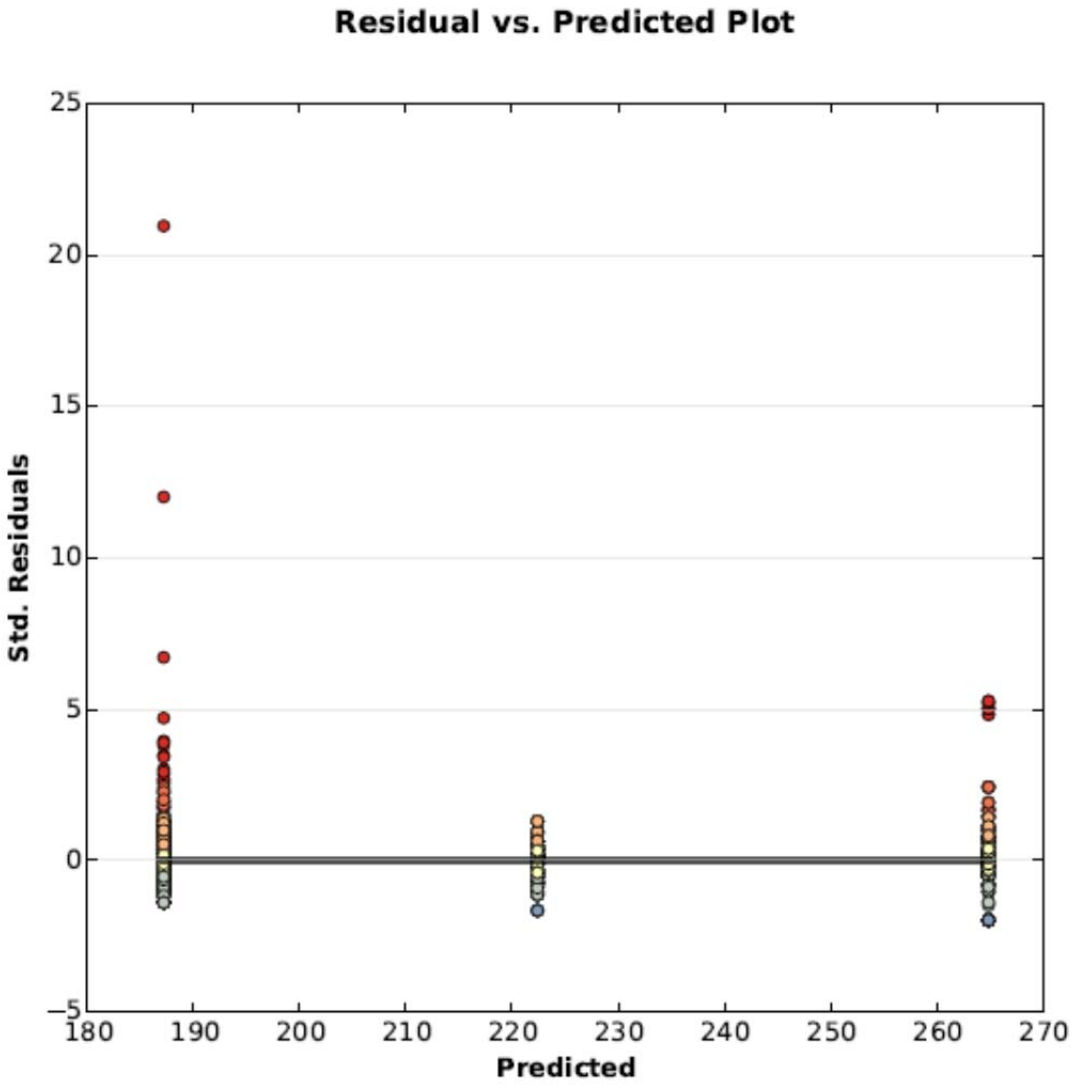
The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.

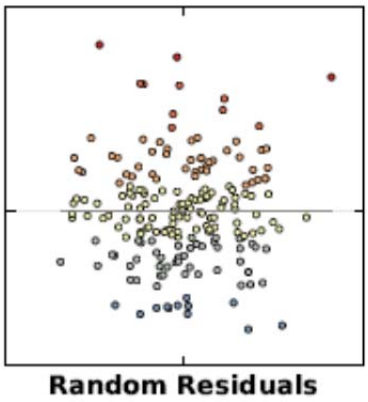
Histogram of Standardized Residuals



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Model 8: Full Model for Intranodal NHL

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	173.586664	6.924453	25.068648	0.000000*	8.274226	20.979204	0.000000*	--------
EXP5KM	84.009294	7.840598	10.714653	0.000000*	9.419607	8.918556	0.000000*	1.392617
EXP5_10	38.975101	7.529403	5.176387	0.000001*	7.385097	5.277534	0.000000*	1.330845
APPAL	-6.905924	8.200684	-0.842116	0.399814	11.369236	-0.607422	0.543646	1.432529
BEALE_R	4.167885	1.401619	2.973622	0.002990*	1.571583	2.652029	0.008062*	1.500482

OLS Diagnostics

Input Features:	MyCTfile	Dependent Variable:	R_CTS41_AD
Number of Observations:	1921	Akaike's Information Criterion (AICc) [d]:	24289.264959
Multiple R-Squared [d]:	0.057258	Adjusted R-Squared [d]:	0.055290
Joint F-Statistic [e]:	29.092330	Prob(>F), (4,1916) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	92.616297	Prob(>chi-squared), (4) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	12.295990	Prob(>chi-squared), (4) degrees of freedom:	0.015281*
Jarque-Bera Statistic [g]:	1112476.526260	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

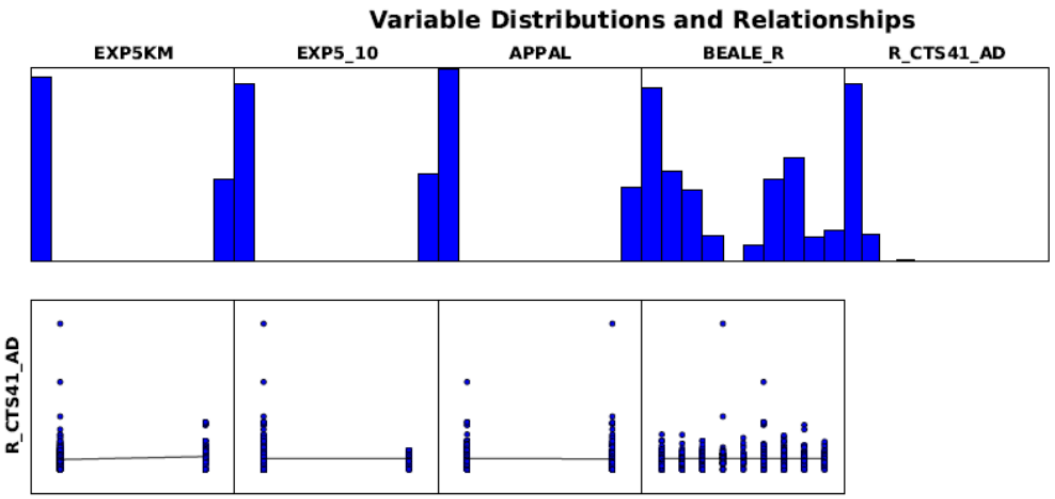
[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

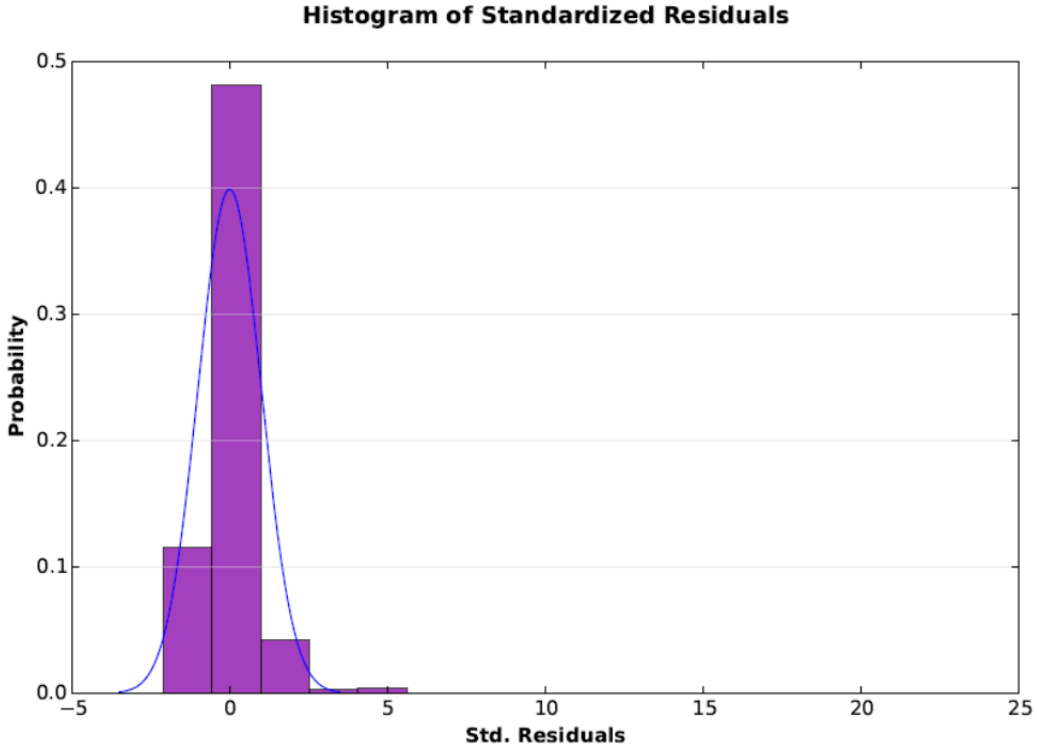
[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

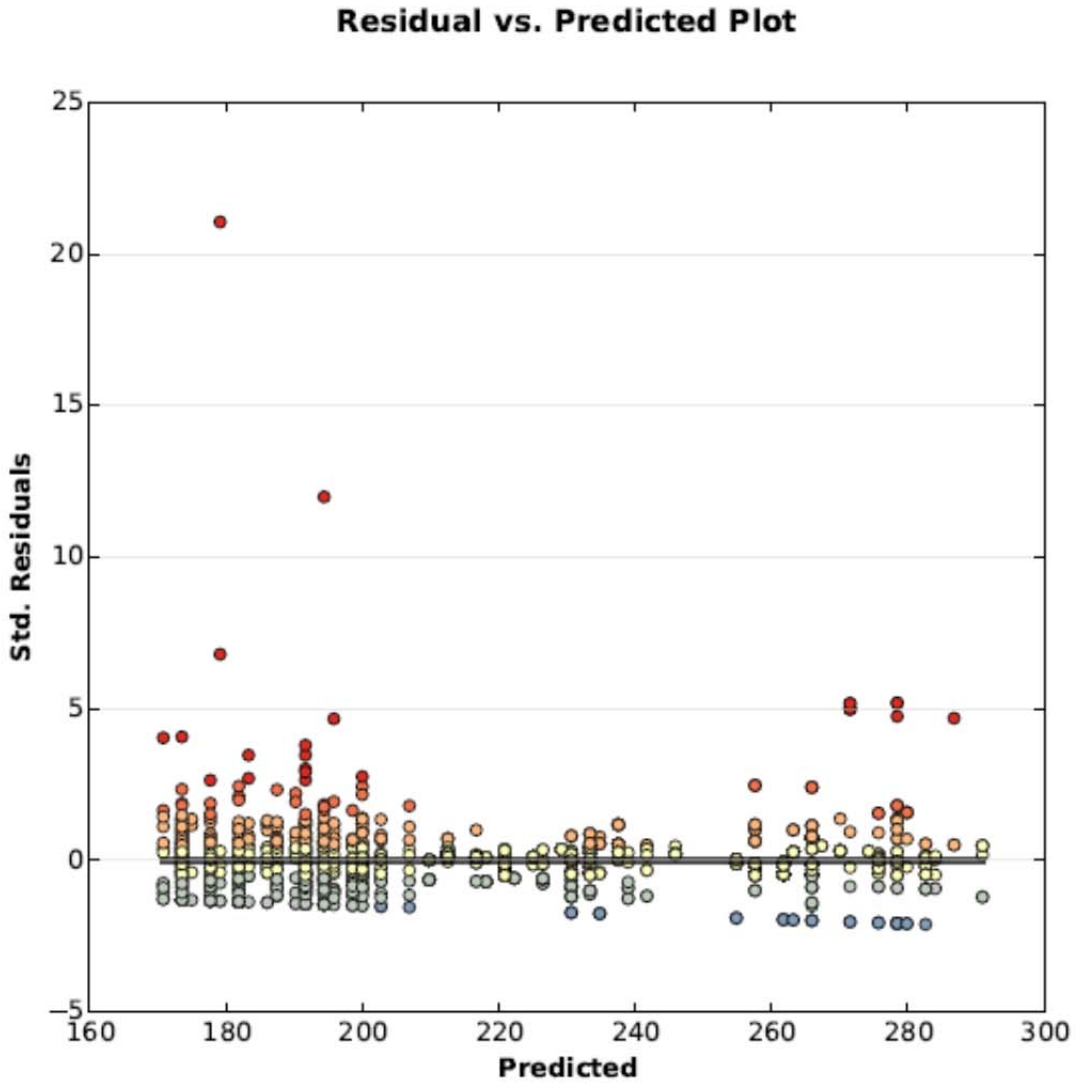


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

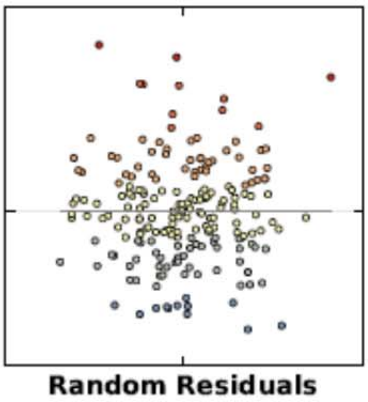
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Model 9: Base Model for Extranodal NHL

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	74.069885	2.233360	33.165219	0.000000*	2.908544	25.466315	0.000000*	-------
EXP5KM	35.904827	3.294245	10.899259	0.000000*	3.635693	9.875649	0.000000*	1.280278
EXP5_10	15.604554	3.236081	4.822052	0.000002*	3.269209	4.773189	0.000003*	1.280278

OLS Diagnostics

Input Features:	MyCTfile	Dependent Variable:	R_CTS42_AD
Number of Observations:	1921	Akaike's Information Criterion (AICc) [d]:	21117.272213
Multiple R-Squared [d]:	0.058370	Adjusted R-Squared [d]:	0.057388
Joint F-Statistic [e]:	59.446247	Prob(>F), (2,1918) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	107.840408	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	39.733059	Prob(>chi-squared), (2) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	14498.933131	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

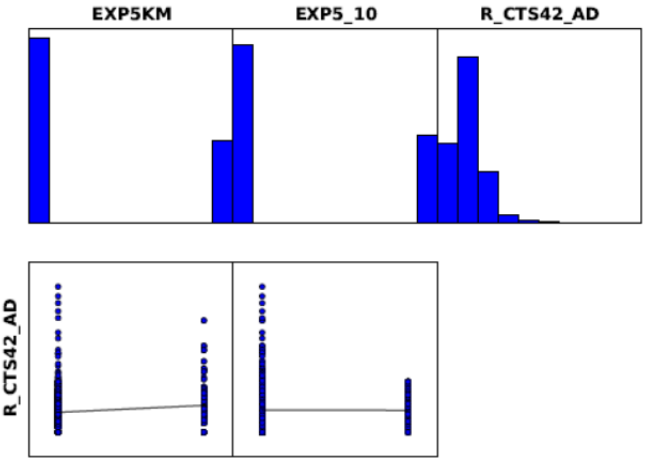
[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

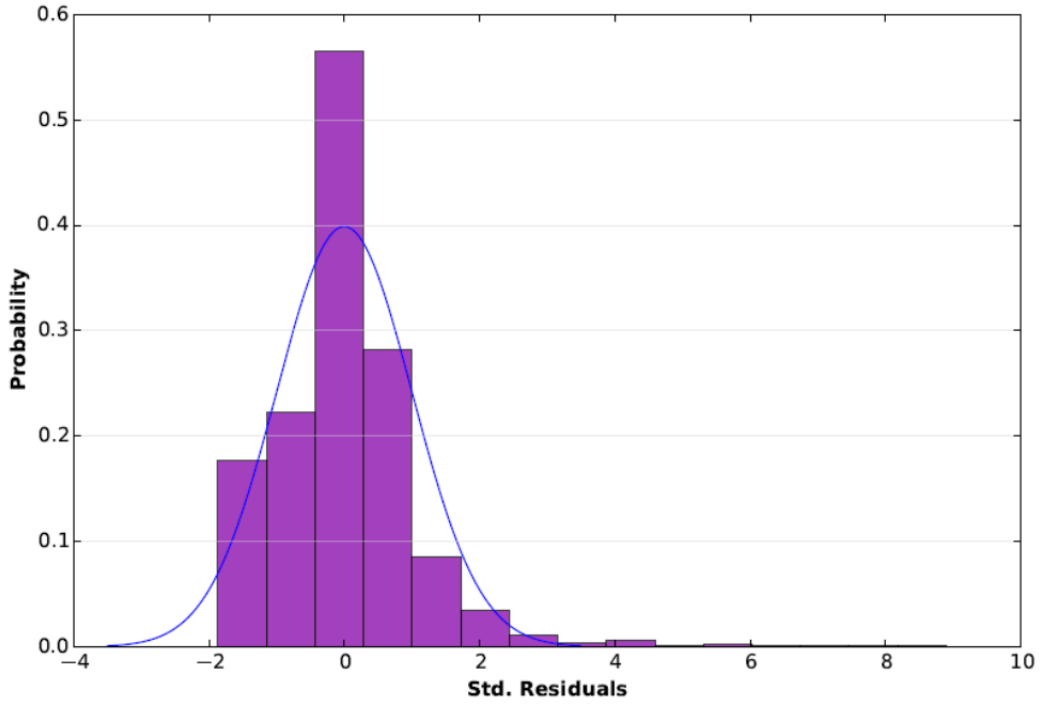
Variable Distributions and Relationships



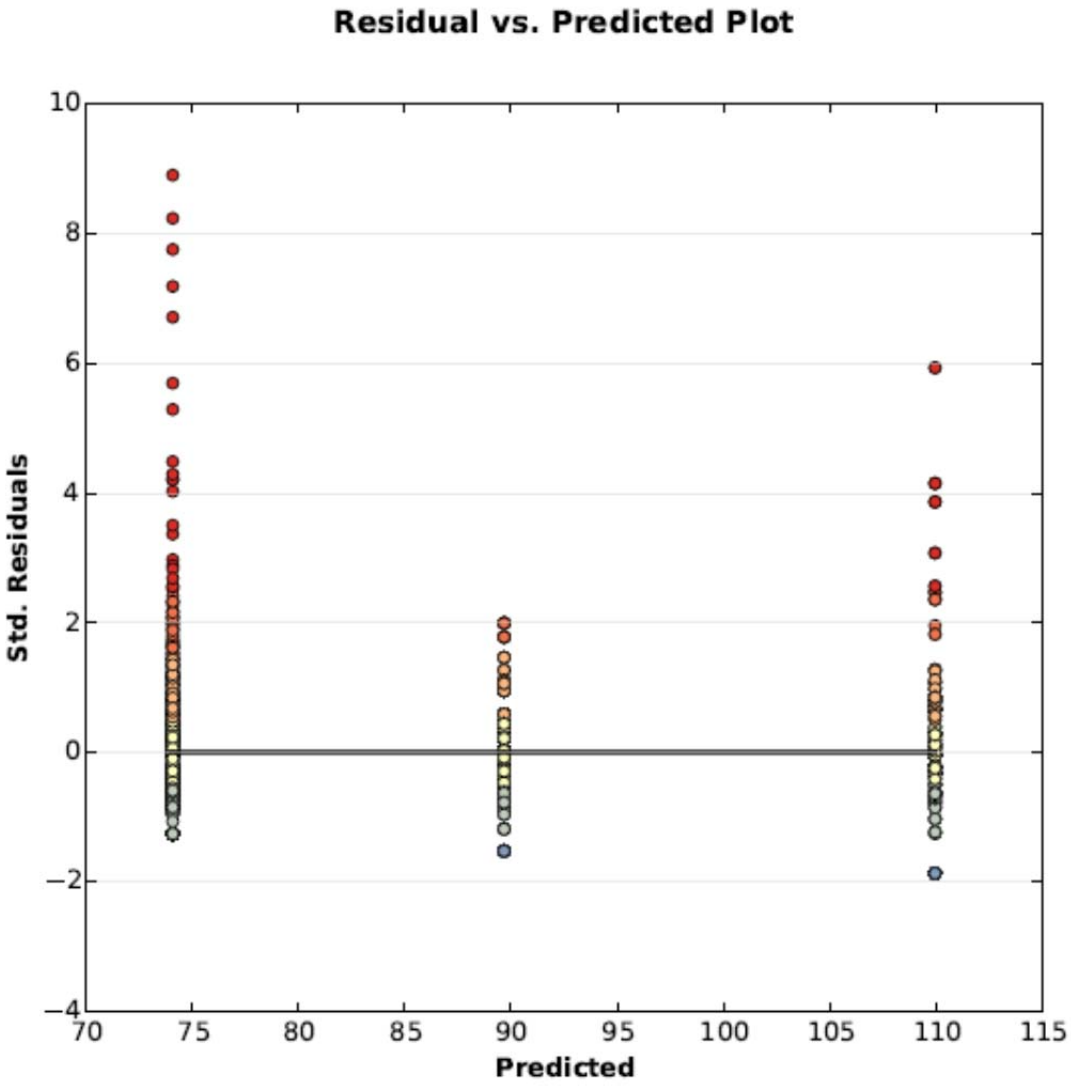
The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.

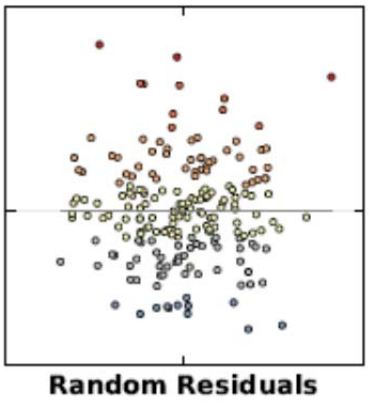
Histogram of Standardized Residuals



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



Model 10: Full Model for Extranodal NHL

Summary of OLS Results - Model Variables

Variable	Coefficient [a]	StdError	t-Statistic	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]	VIF [c]
Intercept	65.954339	3.022809	21.818891	0.000000*	3.520187	18.736031	0.000000*	--------
EXP5KM	39.806987	3.422744	11.630137	0.000000*	3.946930	10.085558	0.000000*	1.392617
EXP5_10	18.213716	3.286895	5.541314	0.000000*	3.368130	5.407665	0.000000*	1.330845
APPAL	4.353241	3.579936	1.216011	0.224133	4.167813	1.044490	0.296380	1.432529
BEALE_R	1.637293	0.611864	2.675908	0.007513*	0.729817	2.243430	0.024967*	1.500482

OLS Diagnostics

Input Features:	MyCTfile	Dependent Variable:	R_CTS42_AD
Number of Observations:	1921	Akaike's Information Criterion (AICc) [d]:	21104.736845
Multiple R-Squared [d]:	0.066451	Adjusted R-Squared [d]:	0.064502
Joint F-Statistic [e]:	34.095838	Prob(>F), (4,1916) degrees of freedom:	0.000000*
Joint Wald Statistic [e]:	119.957375	Prob(>chi-squared), (4) degrees of freedom:	0.000000*
Koenker (BP) Statistic [f]:	71.284226	Prob(>chi-squared), (4) degrees of freedom:	0.000000*
Jarque-Bera Statistic [g]:	13826.971716	Prob(>chi-squared), (2) degrees of freedom:	0.000000*

Notes on Interpretation

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

[a] Coefficient: Represents the strength and type of relationship between each explanatory variable and the dependent variable.

[b] Probability and Robust Probability (Robust_Pr): Asterisk (*) indicates a coefficient is statistically significant ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Robust Probability column (Robust_Pr) to determine coefficient significance.

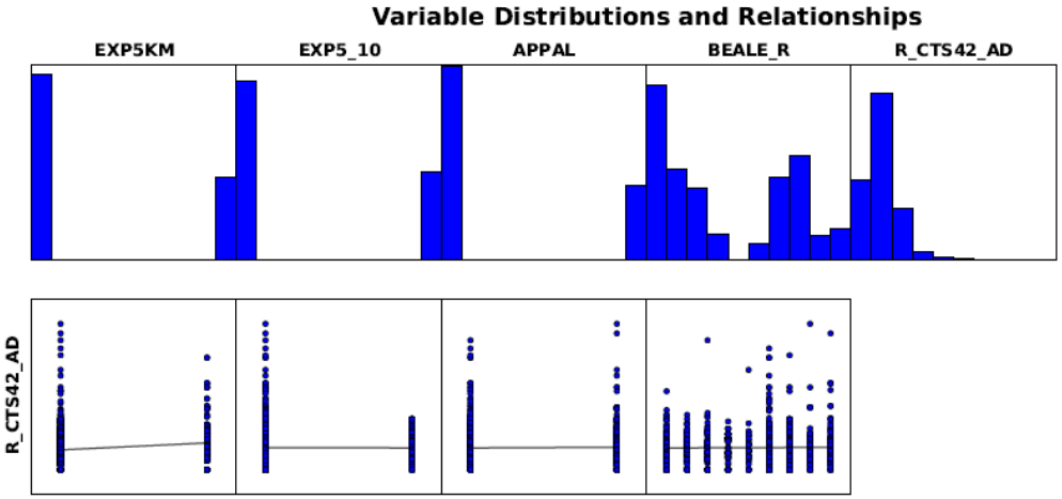
[c] Variance Inflation Factor (VIF): Large Variance Inflation Factor (VIF) values (> 7.5) indicate redundancy among explanatory variables.

[d] R-Squared and Akaike's Information Criterion (AICc): Measures of model fit/performance.

[e] Joint F and Wald Statistics: Asterisk (*) indicates overall model significance ($p < 0.01$); if the Koenker (BP) Statistic [f] is statistically significant, use the Wald Statistic to determine overall model significance.

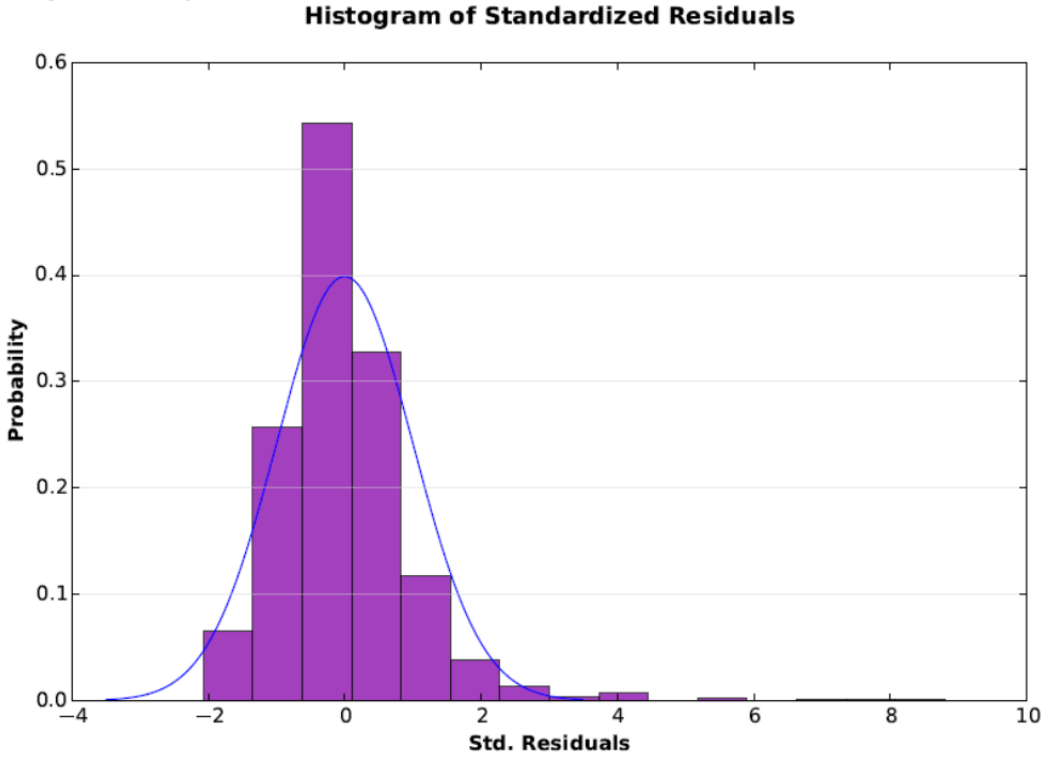
[f] Koenker (BP) Statistic: When this test is statistically significant ($p < 0.01$), the relationships modeled are not consistent (either due to non-stationarity or heteroskedasticity). You should rely on the Robust Probabilities (Robust_Pr) to determine coefficient significance and on the Wald Statistic to determine overall model significance.

[g] Jarque-Bera Statistic: When this test is statistically significant ($p < 0.01$) model predictions are biased (the residuals are not normally distributed).

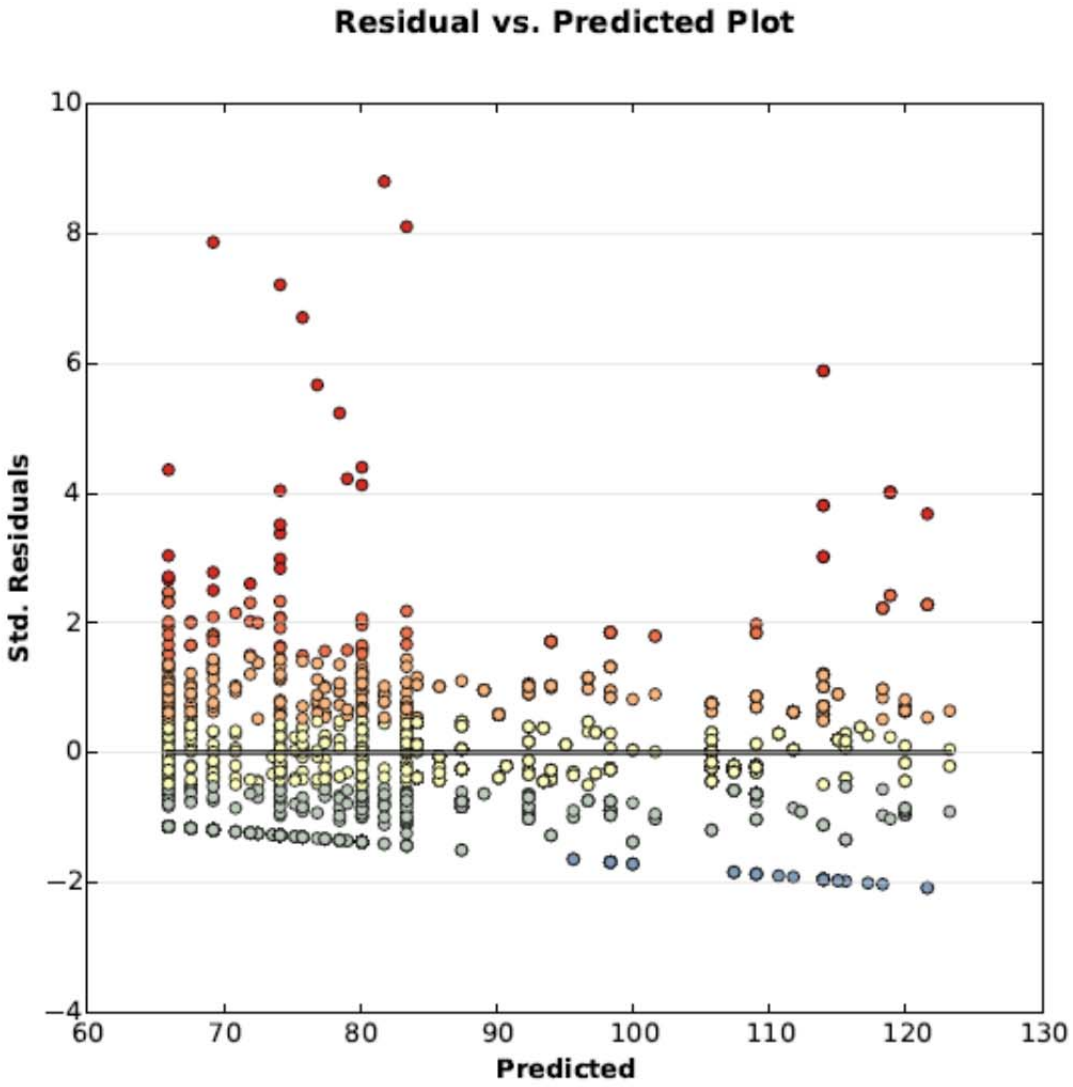


The above graphs are Histograms and Scatterplots for each explanatory variable and the dependent variable. The histograms show how each variable is distributed. OLS does not require variables to be normally distributed. However, if you are having trouble finding a properly-specified model, you can try transforming strongly skewed variables to see if you get a better result.

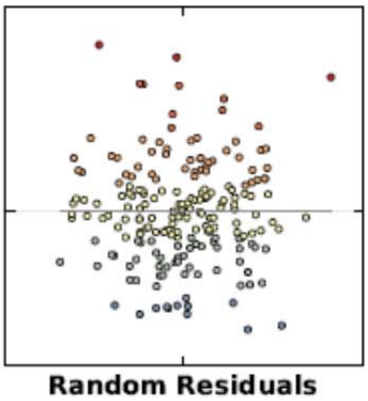
Each scatterplot depicts the relationship between an explanatory variable and the dependent variable. Strong relationships appear as diagonals and the direction of the slant indicates if the relationship is positive or negative. Try transforming your variables if you detect any non-linear relationships. For more information see the Regression Analysis Basics documentation.



Ideally the histogram of your residuals would match the normal curve, indicated above in blue. If the histogram looks very different from the normal curve, you may have a biased model. If this bias is significant it will also be represented by a statistically significant Jarque-Bera p-value (*).



This is a graph of residuals (model over and under predictions) in relation to predicted dependent variable values. For a properly specified model, this scatterplot will have little structure, and look random (see graph on the right). If there is a structure to this plot, the type of structure may be a valuable clue to help you figure out what's going on.



VITA

Name: William Brent Webber

Date and place of birth: October 29, 1970, Mount Vernon, Illinois, USA

Educational Institutions Attended and Degrees Awarded:

University of Kentucky, Lexington, KY. Master of Science in Public Health, Industrial Hygiene Curriculum. September 1997.

University of Illinois, Urbana-Champaign, IL. Bachelor of Science in Microbiology. May 1993.

Professional Positions Held:

Senior Industrial Hygienist, University of Kentucky Environmental Health & Safety, Lexington, KY, 2008-present.

Safety Coordinator, Kentucky Cabinet for Health & Family Services, Frankfort, KY, 2007-2008.

Industrial Hygienist / Chemical Hygiene Officer, University of North Carolina, Chapel Hill, NC, 2004-2007.

Industrial Hygiene Inspector and Consultant, North Carolina Department of Labor, Raleigh, NC, 1998-2004.

Scholastic and Professional Honors:

National Institute of Environmental Health Sciences, training grant awardee, 1995-1997.

Certified Industrial Hygienist, American Board of Industrial Hygiene. Certificate Number 8548. Obtained 6/2003.

Certified Safety Professional. Board of Certified Safety Professionals. Certificate Number 17963. Obtained 1/2004.

Professional Publications:

Webber WB and Fotopulos CP. Establishing a radon management program for public university facilities. *Facilities* 2016. Accepted in final form, awaiting online ahead of print publication.

Webber WB, Ernest LJ, Vangapandu S. Mercury exposures in university herbarium collections. *Journal of Chemical Health & Safety* 2011 18(2): 11-14.

Signature: