TEMPORAL AND SPATIAL ANALYSIS OF CANCER RATES IN THE UNITED STATES

by

Yueh-Ying Han

B.S., Taipei Medical University, Taiwan, 2002

M.S., National Taiwan University, Taiwan, 2004

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Yueh-Ying Han

It was defended on

July 17, 2007

and approved by

Committee Chair

Joel L. Weissfeld, MD, MPH
Associate Professor
Department of Epidemiology
Graduate School of Public Health
University of Pittsburgh

Committee Co-chair

Devra L. Davis, PhD, MPH
Professor
Department of Epidemiology
Graduate School of Public Health
University of Pittsburgh

Committee Members

Gregg E. Dinse, ScD
Senior Investigator
Biostatistics Branch
National Institute of Environmental Health Sciences, North Carolina

Bernard D. Goldstein, MD
Professor
Department of Environmental and Occupational Health
Graduate School of Public Health
University of Pittsburgh

Evelyn O. Talbott, DrPH, MPH
Professor
Department of Epidemiology
Graduate School of Public Health
University of Pittsburgh

J.L. Weissfeld, MD, MPH

TEMPORAL AND SPATIAL ANALYSIS OF CANCER RATES IN THE UNITED STATES

Yueh-Ying Han, Ph.D.

University of Pittsburgh, 2007

ABSTRACT

Introduction: Spatial, temporal and racial patterns of cancer remain largely unexplained in the United States. Time trends of cancer incidence and mortality can be used to estimate the current cancer burden, anticipate clinical care needs, and suggest hypotheses regarding possible etiologic explanations for underlying trends. Methods: Using U.S. 1979-2003 cancer incidence and 1969-2003 cancer mortality data, age-period-cohort and Joinpoint regression models were fit to summarize gender- and race-specific temporal trends for three broad cancer categories that include tobacco-related cancer, screen-detectable cancer, and cancer unrelated to tobacco and screening. Demographic patterns and time trends of non-Hodgkin's lymphoma (NHL) incidence between Pennsylvania and the U.S. from 1985 to 2004 were compared. Using Idaho cancer incidence, 1990-2005, and arsenic levels in ground water, 1990-2005, spatial analysis was conducted to identify geographic patterns of cancer incidence and to evaluate the relationship between arsenic exposure in ground water and cancer incidence in Idaho. Results: Over the last three decades, tobacco-related cancer incidence declined among men and increased among women. Screen-detectable cancer incidence increased, more rapidly among men than women. For cancer

unrelated to tobacco and screening, incidence increased in every gender-and-race group. Though not identical, NHL incidence patterns, with substantial increases, were similar in the U.S. and Pennsylvania. NHL incidence was higher in Pennsylvania counties with a greater percentage of urban residents. Although spatial clustering was demonstrated in Idaho cancer incidence, no relationship was found between arsenic exposure in ground water and Idaho cancer incidence. Conclusion: NHL and other cancers unrelated to smoking or screening have increased in the U.S. in the past two decades in white and black men and women. Etiologic research should attempt to identify modifiable risk factors, including environmental exposures, responsible for the increasing incidence of NHL and cancer unrelated to tobacco and screening. The ecologic association observed in Pennsylvania between NHL incidence and urban residence may be relevant to NHL risk in the entire United States. Additional environmental and demographic information should be evaluated in order to clarify the arsenic-related cancer risk in Idaho counties where ground water has been found to contain higher levels of arsenic. Public Health Significance: Age, period and cohort modeling of cancer incidence and mortality provides important indications of current and future health care needs and also suggests hypotheses for future research. The results of this analysis provide health professionals, researchers, and policy-makers with detailed information and an understandable overview of cancer patterns in the United States. Hypotheses should be generated about these unexplained patterns of cancer so that avoidable cancer risks can be identified that will decrease the cancer burden and associated requirements for health care.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENT

I would like to thank my committee members, Dr. Gregg Dinse, Dr. Bernard Goldstein, and Dr. Evelyn Talbott for all of their supportive suggestions and comments on this work. I would like to thank my supervisor, Dr. Devra Davis, for all the supports during my doctoral study and research. She enables me to challenge myself and continuously provides her encouragements. I would especially like to thank Dr. Joel Weissfeld who not only serves as my advisor but also devotes himself to guiding me all the way through my dissertation process and never accepting less than my best efforts. He is a model of scientific excellence and a wisdom mentor that I will always look up to. With my best regard and thankfulness, I am being here to present this dissertation work.

1.0 INTRODUCTION

With an estimated 559,650 deaths for the year 2007 in the United States (U.S.), cancer is the second leading cause of death in the U.S. and accounts for nearly one-quarter of national deaths, exceeded only by heart diseases (American Cancer Society, 2007b). In middle-aged Americans, cancer is the leading cause of death, and, after accidents, the disease is also the leading cause of death in children (Davis et al., in press). Compared to 1950, cancer mortality decreased slightly in 2004 (193.9 per 100,000 in 1950 and 185.8 per 100,000 in 2004, adjusted to the U.S. 2000 standard population), while mortality for other major chronic diseases including heart disease, cerebrovascular disease, and pneumonia/influenza decreased noticeably from 1950 to 2004 (American Cancer Society, 2007c). American Cancer Society (ACS) estimates that almost 1.4 million new cancer cases will be diagnosed in 2007. In addition, approximately one out of every two American men and one out of every three American women will be diagnosed with some form of cancer during their lifetime, with 77% of all cancers occurring in people aged 55 or older (American Cancer Society, 2007b). Although cancer is a complex and multiply caused disease, the National Cancer Institute (NCI) and the World Health Organization (WHO) advise that many cancer occurrences or cancer deaths can be prevented by improving access to screening and early treatment, promoting the adoption of healthier lifestyles and improved nutrition, and controlling, reducing or avoiding exposures to workplace and other environmental carcinogens (National Cancer Institute, 2005a; World Health Organization, 2007).

Descriptive epidemiologic studies are indispensable for characterizing the cancer burden and for evaluating cancer control strategies, such as tobacco cessation and screening policies. Annually the U.S. spends about $100 billion in direct treatment costs for cancer. Time trend analysis of overall and site-specific cancer rates in population-based surveillance systems can indicate how variations in cancer occurrence and cancer deaths are associated with age, time period and birth cohort. Thus, programs to screen and treat cancers can be evaluated in terms of the time period when they were introduced and the birth cohorts that first adopted such programs. Cancer trend analysis can also assist policy makers in anticipating the demand for health care. Finally, the evaluation of cancer trends can prompt researchers to generate hypotheses about underlying factors that account for patterns of the disease. These analyses ultimately will lay the groundwork for etiologic studies on unexplained patterns of cancer and suggest opportunities for the private sector and government to formulate primary and secondary cancer prevention policies.

Using data from the Surveillance, Epidemiology, and End Results (SEER) Program of the U.S. National Cancer Institute, organized in terms of birth cohorts, time periods and age, we examined cancer incidence and mortality in three broad cancer categories: cancer related to tobacco smoking, cancer associated with screening technology, and all cancers excluded in the smoking and screening categories. We also focused on non-Hodgkin's lymphoma (NHL), a once rare cancer that is now common whose incidence has been strikingly increasing (National Cancer Institute, 2007b). Additional NHL cancer incidence data was obtained from the Pennsylvania Cancer Registry (PCR). These data were used to estimate the spatial-temporal trends in Pennsylvania, to compare the patterns between Pennsylvania and the entire U.S., and to explore demographic and other potential exogenous explanations for these patterns.

Another area of interest is the causative relationship between arsenic exposure and cancer risks. Arsenic, a metalloid, is considered a human carcinogen for more than forty years, although its mechanism of action remains poorly characterized. Populations exposed to arsenic from drinking water at levels of fifty parts per billion or higher have been found to face major public health risks (World Health Organization, 2001). By using a Geographic Information System framework, we evaluated the relationship between cancer incidence data from Cancer Data Registry of Idaho and arsenic data in ground water from the Idaho Department of Environment Quality. However, we were unable to evaluate the role of other potential carcinogenic exposures to this population.

1.1 CANCER STATISTICS

1.1.1 Cancer Data

Studies on cancer statistics mainly use data from population-based cancer registries that code ascertained cancer cases according to the International Classification of Disease for Oncology (ICD-O) rules.(World Health Organization, 2000) Cancer data in the U.S. are usually provided by the SEER program of the NCI, (National Cancer Institute, 2005c). The SEER 9 registries have been collecting and reporting cancer incidence and survival data since 1973 for various parts of the U.S. In 1975, the SEER 9 registries have covered 9.5% of total U.S. population, including the states of Connecticut, Iowa, New Mexico, Utah, and Hawaii and the metropolitan areas of Detroit, San Francisco-Oakland, and Atlanta, and the 13-county of Seattle-Puget Sound (National Cancer Institute, 2005c). The current SSER system collects and reports cancer

incidence and survival data from 15 population-based central cancer registries that cover 26 percent of the U.S. population. The U.S. racial/ethnic population coverage in SEER includes 23 percent of African Americans, 40 percent of Hispanics, 42 percent of American Indians and Alaska Natives, 53 percent of Asians, and 70 percent of Native Hawaiian and other Pacific Islanders (United States Department of Health and Human Services (DHHS), 2006). State cancer registries have been also used to estimate cancer rates on a smaller scale. Since 1994, information from statewide cancer registries, the National Program of Cancer Registries (NPCR), has recently been combined with SEER data to provide nearly complete coverage of the U.S. population with respect to cancer incidence (Centers for Disease Control and Prevention, 2007).

Age-adjusted and age-specific rates can each be used in the statistical analysis of cancer research. An age-adjusted rate, also known as an age-standardized rate (ASR), is a weighted average of the age-specific crude rates, where the weights are the proportions of persons in the corresponding age groups of a standard million population. To compare results from different populations or studies, cancer rates are generally standardized against ages distributed either into 5-year age groupings or single years of age (NCHS) (United Stated Bureau of the Census, 1996). The rates adjusted by the U.S. 2000 population are higher than the rates adjusted by the 1970 standard population because the proportion of older people, among whom cancer is more common, has been increasing over time in the U.S. Cancer rates adjusted for U.S. 2000 population can provide useful and comparable results on the subject of cancer statistics in the U.S.

1.1.2 Cancer Statistics Modeling

There is no perfect method or universally accepted model to describe temporal changes in cancer rates. Limitations exist for every method or model and interpretations under alternative models can be inconsistent. Two major approaches to the evaluation of cancer trends are widely used: an age-period cohort model that examines the effects singly and together of age, period and cohort, and a joinpoint model that expresses age-adjusted rates as a piecewise function of time with or without adjustments for reporting delay.

A major question for cancer epidemiology research is whether age-period-cohort (APC) analyses or joinpoint trend modeling with or without adjustments for reporting delay provides a better way to display cancer trends over time. Decisions about which method to use in evaluating cancer trends can vary depending on the goals of the evaluation. As this thesis will demonstrate, there are strengths and weaknesses of each approach. If the interest is chiefly in predicting short-term future demand for health care, then a simple joinpoint evaluation seems sensible. If the goal is to evaluate past and future causes of cancer, then more sophisticated APC modeling can be appropriate. Further methodological development is required to clarify the values and limitations of each method. While further research is needed to advance methods for evaluating cancer patterns and trends, analysis of cancer data such as that carried out here provides an important contribution to that effort.

1.1.2.1 Age-Period-Cohort Model

Fitting cancer data by a statistical model can yield useful summaries in terms of parameters in the model. An age-period-cohort (APC) model assumes that the logarithm of the expected rate is

a sum of three components representing the effects of age group, time period, and birth cohort. Various combinations of these three factors may be considered, and sometimes one combination might provide a particularly clear summary of the data (Holford, 1983). Because cancer is well known to increase with advancing age, age has been associated with changes in cancer rates; consequently, assessment of changes in cancer rates with advancing age over time has become the center of attention. A time period effect represents the factor that affects persons of all ages at essentially the same time. A cohort effect occurs when people born in the same year are exposed to a causative agent that renders them comparatively different from those born at another time (Arbeev, Ukraintseva, Arbeeva, & Yashin, 2005). The APC model has been frequently used to characterize cancer incidence in terms of temporal patterns associated with age, calendar year, and birth cohort.

1.1.2.2 Joinpoint Regression Model

Joinpoint regression is used to summarize long-term cancer trends and determine whether there are characteristic breaks in the slope over different time periods. This approach fits a linear regression function on a log scale with unknown join points, assuming constant variance and uncorrelated errors (Kim, Fay, Feuer, & Midthune, 2000). The program starts with the minimum number of joinpoint (e.g. 0 joinpoints, which is a straight line) and tests whether more joinpoints are statistically significant and must be added to the model (up to a maximum number of joinpoints specified by a user) (National Cancer Institute, 2005b). Using joinpoint models, one can determine the number of significant joinpoints and estimate rates of change of cancer incidence or mortality during segments of time between adjacent joinpoints.

1.1.2.3 Reporting-delay Adjusted Model

Reporting delay refers to the time elapsed before a diagnosed cancer case is reported to the cancer registry. According to NCI, the standard time delay between cancer diagnosis and the first report of cancer incidence data to the public is about 2 years. The information on cases from prior diagnosis years is updated in subsequent releases of the SEER data (Clegg, Feuer, Midthune, Fay, & Hankey, 2002). However, reporting delay and reporting error hamper the accuracy and timing of reporting cancer cases. Delayed reports have been a concern for persons diagnosed in recent years since cancer rates from these years correspond to the largest group of underreported cancer cases. Cancer incidence adjusted for reporting delay can correct current cancer counts and anticipate future corrections. For each one of cancer sites, several possible covariates (including age, race, and gender) have been considered by NCI investigators in prediction models of delay probabilities. Generally, reporting delay has been greatest for non-whites, and for screening detected cancers, such as prostate and breast (Clegg, Feuer, Midthune, Fay, & Hankey, 2002). NCI investigators evaluated reporting delay within the SEER system by fitting covariates to annual data between 1983 and 2005 and retrospectively correcting these with the counts recorded two years later. The model that minimized the sum of squared prediction errors is chosen as the final model (National Cancer Institute, 2007a).

1.1.3 Geographic Information System

Geographic Information System (GIS) is a system used for input, storage, query, manipulation, and output of geographic information. It is a practical illustration of combining software, data, and user to solve a problem, support a decision, and help to develop a plan. GIS is being used

with increasing frequency in public health, especially environmental health research. GIS can be applied to epidemiologic studies to define study population and geographic units, to estimate environmental levels of target contaminants, and to identify source and potential routes of exposure that relate to health consequences (Nuckols, Ward, & Jarup, 2004).

1.1.3.1 Mapping and Visualization

Mapping is the most widely used GIS technique in public health studies. GIS mapping techniques attempt to describe diseases or estimated outcomes on geographical maps which can identify different values by geographic units. Cloropeth (graduated colored) maps, proportional circles, quartile or percentile maps are generally used to describe the distributions of data. Continuous cartogram provides a distorted or weighted map with geographic regions sized according to a statistical parameter, for instance size of the population, but in a way that keeps the map recognizable.

1.1.3.2 Rate Smoothing

Smoothing techniques aim to remove the effects of the varying population sizes across the areas and to maintain stability of estimated rates in small areas. There are two methods of smoothing. One is Empirical Bayes smoothing. Another method uses a spatial filter (ratio smoothers) to estimate the rate at a given point by including observations at neighboring points within a fixed distance.

1.1.3.3 Spatial Autocorrelation Analysis

Spatial autocorrelation analysis includes tests and visualization of both global (test for clustering) and local (test for clusters) Moran's I statistic. The Moran's I statistic is used to find significant patterning but without designating any particular location as clustered. As the formal test for spatial dependence, Moran's I determines if the data are randomly distributed throughout a given region, or if there some nonrandom clustering is evident. Significance within Moran's I estimate indicates that there is a spatial dependence within the data. A test for cluster is designed to identify the locations of the clusters (or the spatial outliers) (Anselin, 2004). The global test is visualized by means of a Moran scatter plot, in which the slope of the regression line corresponds to Moran's I. The significance test is based on a permutation test (Anselin, Syabri, & Kho, 2006).

The local indicator of spatial autocorrelation (LISA) is used to identify local spatial clusters and outliers with respect to the measure of global association. It also includes several options for sensitivity analysis. For each location, the LISA value allows for the computation of its similarity with its neighbors and also to test its significance. The output of LISA analysis includes a Moran scatter plot and box plot depicting the distribution of the local statistics. Local Moran I estimates can also be displayed on a map to identify clusters (Anselin, 1995).

1.1.3.4 Spatial Data Modeling

Methods of identifying geographic relationships include spatial correlation analysis and spatial regression models. A commonly used method is Pearson's correlation coefficient. The Pearson's correlation coefficient between two spatial variables (which are both normally distributed) and

the correct significance can be calculated to identify two factors that vary in a similar manner with respect to geographic location. Spearman's rank coefficient correlation is used when the data are not normally distributed. Spatial regression analysis modifies linear regression, multiple regression, or logistical regression techniques in order to take account of spatial autocorrelation. Techniques are available to take account of the spatial structure of either the dependent or independent variables specified by a model.

1.2 CANCER LITERATURE REVIEW

It was recently reported that the cancer mortality continues to decline in the U.S. while the rate of newly diagnosed cancer is holding steady. In the Annual Report to the Nation on the Status of Cancer, 1975-2002, the death rate for all cancers combined dropped 1.1 percent annually from 1993 to 2002 (National Cancer Institute, 2005a). According to the report, declines in death rates reflect progress in prevention, early detection, and treatment; however, not all racial and ethnic segments of the U.S. population benefited equally. Delay-adjusted incidence for all cancers combined was on the rise until 1992, and began to decline between 1992 and 1995 and then stabilized in 1995 (H. L. Howe et al., 2006). It was stable in men from 1995 to 2003 and it increased 0.3 percent annually in women from 1987 to 2003. Changes in cancer incidence may result from changes in the prevalence of health behaviors and from changes in detection practices due to the introduction or increased use of screening and/or diagnostic techniques. Based on ACS estimates in 2007, cancers of the prostate and breast are the most frequently diagnosed cancers in men and women, respectively, followed by lung and colorectal cancer (American Cancer Society, 2007b). These four combined cancers are related to screening technologies and

tobacco smoking. Therefore, analyzing time trends of these cancer sites can provide us with tools to assess whether primary prevention, early detection, and treatment have helped to reduce cancer incidence and mortality.

1.2.1 Cancer Related to Tobacco Smoking

Based on the article "Cigarette Smoking Among Adults-United States, 2004", the rates of smoking among men (56.9%) were much higher than among women (28.4%) in 1955, but the gender difference began to diminish as the rates among men declined and the rates among women rose or declined slower (Centers for Disease Control and Prevention, 2005a). In 2005, regardless of race or ethnicity, prevalence of current cigarette smoker was 23.9% among men and 18.9% among women. Specifically, current smoking is higher in non-Hispanic black men (26.7%) than non-Hispanic white men (24.0%) and is higher in non-Hispanic white women (20.0%) than in non-Hispanic black women (17.3%) (Centers for Disease Control and Prevention, 2006c). The report also indicated that prevalence of cigarette smoking among U.S. adults did not change from 2004 to 2005. The adult prevalence might represent a stall in the decline in current cigarette smoking during the preceding 8 years and mirrors a lack of decline in smoking among adolescents since 2002 (Centers for Disease Control and Prevention, 2006a) .

Cancer of the oral cavity and pharynx, esophagus, pancreas, larynx, lung and bronchus, urinary bladder, and kidney and renal pelvis have been associated with tobacco smoking (Centers for Disease Control and Prevention, 2006b; Dinse, Umbach, Sasco, Hoel, & Davis, 1999).Data from the SEER program show that cancer incidence for lung, larynx, oral cavity, pharynx, and bladder has decreased dramatically in white men from 1991 through 1995. Incidence for these cancer sites also declined in white women with the exception of lung cancer. Black men and

women generally experienced similar decreased patterns, but oral cavity cancer continued to increase in men, as did larynx cancer in women (McKean-Cowdin, Feigelson, Ross, Pike, & Henderson, 2000). According to the estimates by ACS in 2007, among the top ten sites with respect to cancer mortality, are cancer of lung and bronchus, pancreas, esophagus, urinary bladder, and kidney, which together account for 47% of cancer deaths in men. Cancer of lung and bronchus and pancreas accounts for 32% of cancer deaths in women (American Cancer Society, 2007b). Many cancer deaths could be reduced further because smoking is a leading factor among preventable cancer causes.

1.2.2 Cancer Related to Screening Techniques

Prostate, breast, cervix uteri, and colon, and rectal are four recommended sites for early cancer detection in asymptomatic people (American Cancer Society, 2007a). Among men, screen-detectable cancers, prostate cancer and colorectal cancer, account for 18% of cancer deaths; in women, colorectal cancer and breast cancer account for 25% of all cancer deaths. Meanwhile, these cancer sites account for 39% of cancer cases in men and 37% in women (American Cancer Society, 2007b). It has been documented that among men, incidence for prostate cancer increased during the 1990s, while that of colorectal cancer decreased (Edwards et al., 2005). Among women for all races combined, incidence for cancer of the cervix uteri and colon-rectum has decreased between 1992 and 2002.

Regarding female breast cancer, the incidence rate increased by 0.4% annually from 1987 through 2002, which is less than the 3.7% annual increase from 1980 to 1987 (Edwards et al., 2005). Within the updated SEER data, breast cancer incidence for all races combined has decreased 2.7% annually between 2000 and 2004 (National Cancer Institute, 2006c). However,

in recent five years, breast cancer incidence declined among white women (3.1% per year) but increased among black women (1.2% per year). Elucidating racial disparity of female breast cancer trend needs further work. The last 5 years data of the Women's Health Initiative (WHI) clinical trial demonstrated that there were excess risks of breast cancer and cardiovascular disease among Hormone Replacement Therapy (HRT) users. Study has shown that HRT use has decreased significantly immediately post-WHI study since 2002 (Kim, Fay, Feuer, & Midthune, 2000) .Whether declined breast cancer incidence in recent years associate with decreased HRT prescription, and whether the decline differs by race requires evaluation by the epidemiology data existing today and in the future.

In general, mortality for cancer of the prostate among men, breast and cervix uteri among women, and colon and rectum in both men and women have decreased in recent years. Studying changes in cancer related to screening technologies will help us to evaluate the effectiveness of cancer screening program and provide insight into cancer prevention and control.

1.2.3 Cancer Unrelated to Smoking or Screening

Studying a grouping of cancers unrelated to smoking and screening is important because any observed effects are presumably uncorrupted by changes of smoking behavior or screening practice. Not all of the cancers we have treated as smoking-related are in fact completely caused by smoking. After excluding cancers linked to tobacco smoking (oral cavity and pharynx, esophagus, larynx, lung and bronchus, pancreas, bladder, and kidney and renal pelvis) and screening technologies (breast, prostate, cervix uteri, and colon-rectum), incidence for the remaining cancer sites (mainly are non-Hodgkin's lymphoma, melanoma of the skin, leukemia, thyroid and corpus and uterus) changed little in white women, but it has risen 0.5% per year in

black women, and 1.7% per year in white men and 2.1% per year in black men between 1975 through 1994 (Dinse, Umbach, Sasco, Hoel, & Davis, 1999). Different cancer patterns between men and women have been observed, with men showing larger increases in cancer incidence than women, Gender-specific risk factors, particularly those related to occupation or the environment, require further exploration.

1.2.4 Non-Hodgkin's Lymphoma

As cancer occurrence and death rates continue to decrease for all cancer sites combined and for most individual cancers in recent decades in the United States, NHL has been documented a substantial increase of incidence and mortality (National Cancer Institute, 2007b). NHL is now the fifth most common malignant neoplasm among males and females (American Cancer Society, 2007b). High rates of lymphoma have been observed among individuals with autoimmune disease, organ transplant, and primary or acquired immunodeficiencies (Fisher & Fisher, 2004). Infectious agents are being regarded as established causes of NHL (Engels, 2007). Epstein-Barr virus (EBV) is an important agent in the etiology of NHL. EBV infection is associated with increasing risk of NHL, especially Burkitt's lymphomas, central nervous system (CNS) lymphomas, T-cell NHL and NHL in individuals with immune dysfunction. Infections with Human T-cell leukemia/lymphoma virus type 1 (HTLV-1), Hepatitis C virus (HCV), Human herpes virus 8 (HHV8), and Helicobacter pylori have also been related to NHL risk (Baris & Zahm, 2000; Grulich & Vajdic, 2005). However, the relationship has been restricted to specific NHL subtypes and only account for a limited proportion of the NHL increases.

The role of environmental exposure may in part explain NHL incidence trends. Exposure to pesticides, herbicides, insecticides, and fungicides are also related to increased risk of NHL

(De Roos et al., 2003; Fritschi et al., 2005; Hardell & Eriksson, 1999; Rafnsson, 2006; Schroeder et al., 2001). Epidemiologic studies have shown that polychlorinated biphenyl (PCB) levels in peripheral blood or adipose tissue are associated with increased risk of NHL (De Roos et al., 2005; Quintana et al., 2004; Rothman et al., 1997) Use of hair dyes among women has been suggested as a risk factor for NHL, especially follicular type, B-cell and low-grade lymphomas. This is consistent with an extensive animal literature testing various components of hair dye toxicologically (Zhang et al., 2004). However, the evidence of a limited increase in the risk of NHL among personal hair dye users is inconsistent (Takkouche, Etminan, & Montes-Martinez, 2005). Elucidation of clear etiologic factors and their mechanistic role in NHL pathogenesis has been challenging and needs further investigations.

1.2.5 Arsenic and Cancer Risk

Arsenic is a metalloid found in water, soil, and air from natural sources and anthropogenic activities such as mining and agriculture applications. Populations exposed to arsenic from geologic contamination of water are found around the world, including locations in Taiwan, India, Inner Mongolia, Chile, Mexico, Argentina, and California and other parts of the western U.S. (Cantor, 1997). Arsenic can cause acute and chronic adverse health effects including multiple cancers (World Health Organization, 2001). Arsenic carcinogenesis has been developed by the animal model. Possible modes of action of arsenic carcinogenesis may include chromosomal abnormalities, oxidative stress, altered DNA repair, altered DNA methylation patterns, altered growth factors, enhanced cell proliferation, promotion/progression, gene amplification, and suppression of p53 (Kitchin, 2001). Other potential mechanisms also include genetoxicity, co-carcinogenesis, and tumor promotion (Hughes, 2002). Although the exact

mechanisms of arsenic action are not well known, arsenic contaminated drinking water is one of the major concerns in public health.

1.2.5.1 Arsenic Exposure Measurement

Among methods to test arsenic levels in human body, urinary arsenic can be a conclusive biomarker of recent exposure. In addition, recent systemic absorption can be measured by amounts of arsenic in hair and toenail or fingernail clippings (Robson, 2003). Quantification of inorganic arsenic and its metabolites in urine can provide important information on detoxification mechanisms that influence the actual doses to tissue and potential risk for chronic health effects (Chappell et al., 1997). It has also shown that toenails can provide an integrated measure of internal inorganic arsenic exposure and reflect all sources of exposure, including drinking water, diet, and occupation (Garland et al., 1993; Michaud et al., 2004). However, data for long-term arsenic exposure need to combine exposure duration, dose, exposure route and history, arsenic source assessment, and individual health history. Accurate measurement of arsenic in drinking-water at levels relevant to health also requires sophisticated laboratory analytic techniques. Analytical quality control and external validation are critical. However, reliability of field methods is yet to be fully evaluated (World Health Organization, 2001). Accurate measurement of arsenic exposures remains challenging.

1.2.5.2 Epidemiology Study

The latency period in humans of arsenic-related carcinogenesis is considered to be 30-50 years (Tapio & Grosche, 2006). Due to the long latency, studies on dose-response relationship between arsenic and internal or external cancers were based on areas with a long exposure history such as

Taiwan and Chile (Chiou et al., 1995; Ferreccio et al., 2000; Wu, Kuo, Hwang, & Chen, 1989). The relationship between arsenic exposure in drinking water and cancer risk has been documented from these studies. However, US-based epidemiologic studies have not been able to provide convincing evidence that ingested arsenic through contaminated water is associated with cancer risk (Bates, Smith, & Cantor, 1995; Lamm et al., 2004; Lewis, Southwick, Ouellet-Hellstrom, Rench, & Calderon, 1999). More recently, elevated cancer rates from high levels of arsenic in drinking-water in Bangladesh is attracting much attention to the issue of arsenic-related cancer risk. The number of people consuming arsenic-rich water in Bangladesh has increased dramatically since the 1970s due to well-drilling and population growth (Rahman et al., 2001). The most commonly manifested disease so far is skin lesions. Increased risk of skin and internal cancers has not been observed, but they are likely to become a principal human health concern arising from arsenic exposure.

In 2001, after considering the analyses and risk assessment on arsenic from the National Research Council, the U.S. Environmental Protection Agency lowered the maximum contaminant level (MCL) for arsenic in drinking water from 50 μg/l to 10 μg/l (United Stated Environmental Protection Agency, 2001). Health risks associated with inorganic arsenic exposures; particularly at relatively low exposure levels in the U.S. population has generated discussion and debate. More studies are needed to estimate the contributions of arsenic exposure to cancer risk.

1.3 SPECIFIC AIM

This research uses statistical models and Geographic Information System techniques to analyze population-based cancer incidence and mortality data. The research encompasses three discrete, but related projects (below):

1.3.1 Project#1

Smoking-associated cancer includes primary site-specific cancer where incidence is typically higher among persons with a personal smoking history. Screening-detected cancer includes primary site-specific cancer where shifts in cancer incidence can be attributed, at least in part, to variation in the use of medical technologies for detecting early or sub-clinical cancer. The third cancer category includes all primary site-specific cancer not belonging to either the smoking-or screening-related cancer categories which constitutes a collection where variation in incidence cannot easily be attributed to changes in tobacco use or to changes in medical care. Therefore, any variation observed in the occurrence of cancer in the other cancer category can be conceived as a possible consequence of general environmental influences on cancer risk, such as NHL.

Descriptive analyses of U.S. cancer rates using Joinpoint models have been annually conducted and reported by NCI researchers. However, by applying APC models, an alternative statistical approach, and by categorizing multiple site-specific cancers into three mutually exclusive cancer groups, Project #1 aims to distinguish itself from previous reports by providing a potential means for isolating cancer risk effects due to general environmental exposures. Project#1 specific aims are:

a) Perform APC analysis to summarize U.S. cancer trends for three broad categories that include smoking-related cancer, screening-related cancer, and cancer unrelated to smoking and screening

b) Analyze recent U.S. NHL cancer rates by characterizing race- and gender-specific incidence and mortality in terms of temporal patterns associated with age, calendar year (period), and year of birth (cohort).

1.3.2 Project#2

Previous analyses of SEER data and studies in other countries have documented time-related increases in NHL incidence in the United States. Using NHL incidence data from Pennsylvania, which is not included in SEER registry, one can identify whether the trends observed elsewhere also hold true in a statewide cancer registry. Associating the characteristics with county-specific NHL incidence, ecologic study is used to identify possible NHL risk factors in Pennsylvania. Project #2 specific aims are:

a) Apply APC and Joinpoint regression models to characterize race- and gender-specific NHL incidence in terms of temporal patterns in Pennsylvania and SEER registry during the 1985-2004 time periods.

b) Compare demographic patterns and temporal trends of NHL incidence between Pennsylvania and SEER registry during the 1985-2004 time periods according to the results from APC and Joinpoint models.

c) Use multiple linear regression models to evaluate factors associated with NHL incidence in Pennsylvania during the 1985-2004 time periods.

1.3.3 Project #3

Ground water contamination is the main route of arsenic exposure. Ground water in Idaho contains elevated levels of naturally occurring arsenic by geographic conditions and also by human agriculture activities. In Idaho, 95% of drinking water comes from ground water sources and arsenic exposure through ground water has been a concern. Geographic Information System (GIS) techniques are being used to visualize the geographic distribution of an exposure or disease. Spatial analytic method applied to the field of public health, though relatively rare, can help to develop information on interpreting exposure and health consequences. The methods we identify here are used to summarize gender- and county-specific cancer incidence and whether it is related to the arsenic level in ground water in Idaho. Project #3 specific aims are:

a) Apply GIS and spatial analytic technologies to describe geographic and temporal patterns in Idaho cancer incidence during the 1990-2005 time periods.

b) Use GIS techniques to visualize geographic distributions of arsenic levels in ground water in Idaho during the 1990-2005 time periods.

c) Use spatial analytic methods to evaluate risk factors associated with cancer incidence in Idaho during the 1990-2005 time periods.

2.0 PROJECT#1: AGE-PERIOD-COHORT ANALYSIS OF TOBACCO-RELATED CANCER, SCREEN-DETECTABLE CANCER, AND CANCER UNRELATED TO TOBACCO OR SCREENING

Manuscript in Preparation

Yueh-Ying Han [1,2], Joel L. Weissfeld [1,2], Devra L. Davis [1,2], Gregg E. Dinse [3],

Bernard D. Goldstein [4], Evelyn O. Talbott [1]

[1] Department of Epidemiology, Graduate School of Public Health, Pittsburgh, University of Pittsburgh, Pittsburgh, Pennsylvania
[2] Center for Environmental Oncology, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania
[3] Biostatistics Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina
[4] Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

Address correspondence to Joel L. Weissfeld, MD, MPH

5150 Centre Ave, Pittsburgh, UPMC Cancer Pavilion, 4th floor, PA15232

Phone: 412-623-3313

Email: weissfeldjl@upmc.edu

2.1 ABSTRACT

Introduction: Studies of cancer incidence and mortality trends can be used to characterize current and future demands for health care and to suggest etiologic hypotheses about avoidable causes of any unexplained patterns. **Methods:** Using United States (U.S.) 1979-2003 cancer incidence and 1969-2003 cancer mortality data, we fit gender- and race-specific age-period-cohort models to summarize age-specific rates for three broad cancer categories, tobacco-related cancer, screen-detectable cancer, and cancer in neither the tobacco-related or screen-detectable categories. **Results:** Tobacco-related cancer incidence declined among men and increased among women. Screen-detectable cancer incidence increased, more rapidly among men than women. For cancer unrelated to tobacco or screening, incidence increased in every gender-race group, with average annual percentage changes (AAPC) of 0.99 and 0.72 percent per year in white and black men and 0.92 and 0.79 percent per year in white and black women. Generational risks of these cancers are between 1.2 and 1.3 fold higher when one generation compared to previous generation. AAPC estimates for incident non-Hodgkin's lymphoma (NHL), a cancer unrelated to tobacco or screening, were 2.33 and 3.17 percent per year in white and black men and 1.92 and 3.29 percent per year in white and black women. Except for NHL and female tobacco-related cancer, mortality decreased over time. **Conclusions:** Underlying rates of cancer not related to smoking or screening are projected to increase. Persons over age 65 are among the fastest growing segment of the population today, a segment in which four out of every five cases of cancer will occur. The demand for cancer treatment will increase as a result of this demographic shift alone. Hypotheses should be generated and tested regarding underlying,

avoidable or modifiable risk factors, including environmental factors, that may account for some of these unexplained patterns of cancer unrelated to tobacco or screening.

2.2 INTRODUCTION

Analyses of cancer trends have been used to evaluate progress against cancer (H. L. Howe et al., 2006). A recent analysis showed cancer incidence increasing between 1975 and 1992, cancer incidence stable between 1992 and 2003, and cancer mortality declining since the early 1990s (H. L. Howe et al., 2006). Analyses of cancer trends involving specific organ sites or population subgroups have been used to study etiology and to measure health disparities (Edwards et al., 2005). However, the interpretation of population-based cancer statistics is complicated by variable case reporting, changes in medical practice, and a multiplicity of risk factors, including those related to personal behaviors and those related to carcinogens in air, water, and food.

One simplifying approach divides primary site-specific cancer into three broad categories, tobacco-related cancer, screen-detectable cancer, and cancer unrelated to tobacco or screening. Tobacco-related cancer includes 100% of all primary cancers having causal association with current and former tobacco use. Excluding tobacco-related cancer, screen-detectable cancer includes primary cancers where shifts in incidence can be attributed, at least in part, to variation in the use of medical technologies for early cancer detection. By definition, cancer unrelated to tobacco or screening is a residual category constituting a disease collection where variation in incidence can not easily be attributed to changes in tobacco use or in medical care. Compared to all-sites of cancer combined, occurrence of cancer unrelated to tobacco or screening may be a more specific indicator of general environmental influences on cancer risk.

An similar analysis but less extensive was performed by Dinse et al. (Dinse, Umbach, Sasco, Hoel, & Davis, 1999), using Poisson regression techniques to model U.S. 1975 through 1994 age-specific incidence in terms of age group, calendar time period, and birth cohort. They calculated temporal average annual percentage changes (AAPC) in incidence for cancer unrelated to tobacco or screening (Dinse, Umbach, Sasco, Hoel, & Davis, 1999). Estimates of AAPC were higher among white and black men (1.7 and 2.1 percent per year), than white and black women (0.1 and 0.5 percent per year). Gender-dependent occupational, environmental, or behavioral risk factors may account for this differential trend in cancer unrelated to tobacco or screening.

U.S. incidence rates for non-Hodgkin's lymphoma (NHL), an important example of cancer unrelated to tobacco or screening, are increasing dramatically (H. L. Howe et al., 2006; Muller, Ihorst, Mertelsmann, & Engelhardt, 2005), 2.4 percent and 4.2 percent per year among white and black women and 4.7 percent and 6.4 percent per year among white and black men, according to the analysis of Dinse *et al*. (Dinse, Umbach, Sasco, Hoel, & Davis, 1999). Possible explanations for this trend include improved diagnosis, changes in disease classification, use of immune-compromising medical therapies, and human immunodeficiency virus (HIV) infection (Fisher & Fisher, 2004; Grulich & Vajdic, 2005; Hennessy, Hanrahan, & Daly, 2004). However, in the opinion of some analysts, these factors only partially explain recent NHL trends (Clarke & Glaser, 2002; Muller, Ihorst, Mertelsmann, & Engelhardt, 2005).

Using methods described by Dinse *et al*. (Dinse, Umbach, Sasco, Hoel, & Davis, 1999) , this study summarizes U.S. temporal trends for tobacco-related cancer, screen-detectable cancer, and cancer unrelated to tobacco or screening. This study extends the work of Dinse *et al*. by

including up-to-date incidence data (1979-2003), corrected for reporting delay, and by examining mortality, as well as incidence.

2.3 MATERIALS AND METHODS

2.3.1 Cancer Data

SEER*Stat (version 6.2.4) (Surveillance Epidemiology and End Results (SEER) Program, 2006b, 2006c) and SEER*Stat databases were used to obtain race-, sex-, age-, year-, and cancer site-specific cancer incidence and mortality counts. SEER also provided race-, sex-, age-, year-, and county-specific population estimates (National Cancer Institute, 2006a) and race-, sex-, age-, year-, and cancer site-specific correction factors for reporting delay (Surveillance Epidemiology and End Results (SEER) Program, 2006a).

The current SEER Program collects and reports on cancer incidence and survival data from 15 population-based central cancer registries that cover 26 percent of the U.S. population and includes 23 percent of African Americans, 40 percent of Hispanics, 42 percent of American Indians and Alaska Natives, 53 percent of Asians, and 70 percent of Native Hawaiian and other Pacific Islanders (United States Department of Health and Human Services (DHHS), 2006). Analyses of cancer incidence were restricted here to the original nine SEER populations, which included five states (Connecticut, Iowa, New Mexico, Utah, and Hawaii), three metropolitan areas (Detroit, San Francisco-Oakland, and Atlanta), and a 13-county Seattle-Puget Sound region. Selections according to cancer primary site used the "SEER Incidence Site Recode ICD-O-3 (1/27/2003)" (http://seer.cancer.gov/siterecode), a coding scheme that maps ICD-O-3

primary site and histology data (converted from ICD-O-2, where necessary) into conventional primary site groupings, groupings which are consistent over time. Similar to Dinse *et al*. (Dinse, Umbach, Sasco, Hoel, & Davis, 1999), the tobacco-related cancer category included 100 % of all of the following sites: invasive cancers of the oral cavity and pharynx (ICD-O-3 primary site C000-148), esophagus (C150-159), pancreas (C250-259), larynx (C320-329), lung and bronchus (C340-349), urinary bladder (C670-679), and kidney and renal pelvis (C649, C659). The screen-detectable category included invasive cancers of the prostate (C619), female breast (C500-C509), cervix (C530-C539), and colon-rectum (C180-189, C199, C209, C260). Both categories excluded lymphoma (ICD-O-3 histology code 9590-9989). Cancer unrelated to tobacco or screening included all instances of invasive cancer not in the tobacco-related or screen-detectable cancer categories. NHL, a subgroup within the category of cancer unrelated to tobacco or screening, included cancers of ICD-O-3 histology 9590-9596 and 9670-9729, regardless of primary site, and cancers of ICD-O-3 histology 9823 (B-cell chronic leukemia) and 9827 (adult T-cell leukemia/lymphoma) at any ICD-O-3 primary site except C024, C098-C099, C111, C142, C379, C420-C422, C424, or C770-C779.

Analyses of cancer mortality covered the entire U.S., including nearly all 50 states and the District of Columbia. Selections according to cause of death used the "SEER Cause of Death (COD) Recode 1969+ (9/17/2004)" (http://seer.cancer.gov/codrecode), a scheme that maps ICD-9 (converted from ICD-8, where necessary) and ICD-10 cause of death codes into conventional cause of death categories, categories which are consistent over time.

Cancer incidence and mortality analyses were restricted to black and white races. Analyses of cancer incidence and mortality included only cancers diagnosed or deaths occurring, respectively, between ages 20 and 84 years. We did not attempt to model cancer endpoints

occurring after age 84 years, a setting where the recognition and classification of cancer are particularly vulnerable to variability in medical care. Also, we did not attempt to model cancer endpoints before age 20 years, because of the relative infrequency of cancer in these youngest age groups. To avoid age-period tabulations with empty cells, it also became necessary to exclude 20-24 year-old black men from the analysis of screen-detectable cancer incidence.

For every possible combination of cancer category (tobacco-related cancer, screen-detectable cancer, cancer unrelated to tobacco or screening, and NHL), sex (men and women), race (black and white), and 5-year age grouping between 20-24 and 80-84 years, we calculated age-specific cancer incidence for each of five 5-year time periods between 1979-1983 and 1999-2003 and age-specific cancer mortality for each of seven 5-year time periods between 1969-1973 and 1999-2003. To account for reporting delay, we multiplied each directly calculated cancer incidence rate by the appropriate correction factor (see above and reference (Clegg, Feuer, Midthune, Fay, & Hankey, 2002).

2.3.2 Age-Period-Cohort Analysis

We fit separate age-period-cohort (APC) models to age-specific cancer incidence and to age-specific cancer mortality rates, for every possible combination of cancer category, sex, and race (see Statistical Appendix). The APC models estimated the natural logarithm of age-specific cancer incidence or mortality rate as a function of thirteen 5-year age groups (between 20-24 and 80-84 years), five-year time periods (five time periods for incidence, seven for mortality), and overlapping 10-year birth cohorts (17 birth cohorts for incidence, 19 for mortality). Expressing the natural logarithm of the population at risk as an offset variable, we used Poisson regression (a generalized linear model with a log link function and Poisson response probability distribution,

implemented in PROC GENMOD, SAS Version 9.1, Cary, N.C.) to fit cancer incidence or mortality counts to variables representing age group, time period, and birth cohort. Using the method described by Dinse *et al*. (Dinse, Umbach, Sasco, Hoel, & Davis, 1999), we parameterized the APC model in terms of separate linear effects due to age group, time period, and birth cohort (each expressed as an integer ordinal variable) and age group-specific, time period-specific, and birth cohort-specific deviations added to the linear effects due to age group, time period, and birth cohort, respectively (see Statistical Appendix). Using the iterative approach described by Breslow (Breslow, 1984), we fit the extra-Poisson model if, according to the Pearson chi-square goodness of fit test ($p<0.05$), the simple Poisson model produced over-dispersed residuals. To test the statistical significance of the collective contribution made by a particular parameter set (age group, time period, or birth cohort), we dropped the relevant parameter set from the full age-period-cohort model and constructed the appropriate F-test from the change produced in the Pearson chi-square goodness of fit test statistic.

The APC model is overdetermined, limiting direct interpretation of the individual parameters estimated by the model. There is no way to determine whether any single one of the parameters—age, period, or birth cohort-- could on its own be a sufficient determinant of the outcome. We view age, period, and cohort effects as being smooth, continuous functions of time. Although these smooth functions could have very complex shapes, each set of effects still can be described by a straight line and a collection of deviations about that line. It turns out that the individual slopes of the 3 straight lines are not identifiable, but the deviations about those lines are identifiable. Furthermore, even though we can not uniquely estimate the individual slopes associated with the age, period, and cohort effects, we can uniquely estimate certain combinations of them, such as the sum of the age and period slopes and the sum of the period

and cohort slopes. Thus, we can uniquely estimate the drift (D^{drift}) (i.e. the sum of the period and cohort slopes) and use this to summarize long-term time trends. Drift generally involves contributions from time period and birth cohort effects, but we can not estimate the proportions attributable to each. In addition, we can use the patterns of deviation to attribute certain information to age, period and cohort effect individually.

Thus, our parameterization of the APC model (see Appendix A: Statistical Appendix) uniquely identified the sum D^{drift} of the two parameters representing the linear effects due to time period and birth cohort. Having modeled age-specific rates for 5-year age groups over contiguous 5-year time periods, we used the transformation $100\left(e^{D^{drift}/5}-1\right)$ to estimate the average annual percentage change (AAPC). The AAPC represents the age-independent relative change in cancer risk associated with a temporal change of 1 year in duration. The Wald 95 percent confidence interval for the AAPC was estimated at $100\left(e^{\left(D^{drift}\pm 1.96\sigma_{D^{drift}}\right)/5}-1\right)$, where $\sigma_{D^{drift}}$ is the estimate of the standard error of D^{drift}.

As a function of calendar year, we plotted, for each APC model, the time period-specific deviations added to the linear effects due to time period. As a function of year of birth, we also plotted the birth cohort-specific deviations added to the linear effects due to birth cohort. As noted above, the linear effect due to time period can be incorporated into the linear effect due to birth cohort. Therefore, the F-test applied to the time period parameter set can be interpreted as a statistical test of the curvature observed in the time period-specific deviation plot (H_0: $\boldsymbol{D}^{per}_{N_{per}-1}=0$, see Statistical Appendix). Similarly, the F-test applied to the birth cohort parameter set can be interpreted as a statistical test of the curvature observed in the birth cohort-specific deviation plot (H_0: $\boldsymbol{D}^{coh}_{N_{coh}-1}=0$, see Statistical Appendix). Regions of progressive increase and decrease were

interpreted as sequential time periods or birth cohorts exerting positive and negative pressure, respectively, on cancer risk (see Statistical Appendix). A plot with deviations decreasing over recent time periods or later birth cohorts was interpreted as an indicator of a time-related factor that possibly predicts diminishing risk for future time periods or birth cohorts.

2.4 RESULTS

2.4.1 Effect of Reporting Delay

As expected, estimates of average annual percentage change (AAPC) were increased by adjusting incidence for reporting delay (Appendix Table B-1). When examined according gender and race, increases in AAPC were small (≤0.07 percent per year) for tobacco-related and screen-detectable cancer, intermediate (between 0.11 and 0.13 percent per year) for cancers unrelated to tobacco or screening, and substantial (between 0.17 and 0.19 percent per year) for NHL. Adjustments for reporting delay did not affect inferences regarding the statistical significance of time period or birth cohort parameter sets (data not shown). All results which follow adjusted for reporting delay.

2.4.2 Tobacco-related Cancer

Age-adjusted tobacco-related cancer incidence and mortality were higher in men than women and higher in blacks than whites (Appendix Figure B-1). Age-adjusted incidence and mortality rates have been decreasing in men and increasing in women. However, beginning in 1989, the

rate of increase slowed in women. Tables 2-1 and 2-2 and Figures 2-1 to 2-4 contain the results from APC models. APC models confirmed generally decreasing tobacco-related cancer risk in men (AAPC estimates ranging between -0.89 and -1.47 percent per year) and increasing risk in women (AAPC estimates ranging between 0.18 and 0.83 percent per year). Except for incidence in black women and mortality in white men, both birth cohort and time period parameters significantly ($p<0.05$) improved the fit of APC models. Except for incidence in white women and black men, parameter deviations added to linear effects appeared either flat or downward sloping over recent birth cohorts and time periods. With respect to the tobacco-related cancer incidence, birth cohort parameter deviations sloped upward beginning with the 1949-1958 white female and the 1964-1973 black male birth cohorts (Figure 2-1).

2.4.3 Screen-detectable Cancer

Age-adjusted screen-detectable cancer mortality was highest in black men, intermediate in black women, and lowest in whites (Appendix Figure B-2). Recent times showed decreasing age-adjusted mortality in each sex-race group, but, in black men, this decline followed a period of prominent increase. Generally increasing over time, age-adjusted incidence peaked prominently in men between 1989 and 1995. APC models confirmed generally increasing incidence, more so in men (AAPC 2.51 to 3.02 percent per year) than women (AAPC 0.10 to 0.37 percent per year; Table 2-1) and decreasing mortality, more so in whites (AAPC -1.10 to -1.44 in whites) than blacks (AAPC -0.34 to -0.83 percent per year; Table 2-2). In women, birth cohort and time period parameters improved APC model fit, both for incidence and mortality. In men, for blacks only, only birth cohort parameters improved fit. In both incidence and mortality APC models,

parameter deviations added to linear effects appeared either flat or downward sloping over recent birth cohorts and time periods (Figures 2-1 to 2-4).

2.4.4 Cancer Unrelated to Tobacco or Screening

Age-adjusted mortality due to cancer unrelated to tobacco or screening was higher in men than women and higher in blacks than whites (Appendix Figure B-3). Age-adjusted incidence was lowest in black women and, after 1997, intermediate in black men. Over time, age adjusted incidence and mortality generally trended upward and downward, respectively. APC models confirmed generally increasing incidence (AAPC 0.72 to 0.99 percent per year; Table 2-1) and decreasing mortality (AAPC -0.35 to -0.80 percent per year; Table 2-2). For mortality, birth cohort and time period parameters improved fit, regardless of sex and race. The fit of APC incidence models were not significantly worsened by ignoring birth cohort effects among white men or by ignoring time period effects among white or black women. For women only, positive slopes covering recent birth cohorts were observed in birth cohort parameter deviation plots from incidence models (Figure 2-1) and in time period parameter deviation plots from mortality models (Figure 2-4).

2.4.5 Non-Hodgkin's Lymphoma

Age-adjusted NHL incidence was highest in white men, intermediate in black men and white women, and lowest in black women (Appendix Figure B-4). Age-adjusted incidence generally increased over time, although incidence flattened in men after 1990. Age-adjusted mortality increased with time before 1995 and, except for black women, decreased thereafter. APC models

confirmed generally increasing incidence (AAPC 1.92 to 3.29 percent per year; Table 2-1) and increasing mortality, more so in blacks (AAPC 1.39 to 1.54 percent per year) than whites (AAPC 0.28 to 0.62 percent per year; Table 2-2). For both incidence and mortality, time period parameters improved fit, except for the incidence model in black women. In incidence (Figure 2-2) and mortality (Figure 2-4) models, respectively, time period parameter deviation plots sloped downward after the 1989-1993 and 1994-1998 time periods. Birth cohort parameters improved fit only for mortality models in white men and in women, both white and black. The strongly positive slope connecting the 1969-1978 and 1974-1983 birth cohort parameter deviations in the incidence model for black women (Figure 2-1) may represent random variability in the single age-specific rate used to estimate the last (1974-1983) birth cohort parameter.

2.5 DISCUSSION

Results from APC models can be summarized, as follows, 1) tobacco-related cancer risk appears to decrease in men and increase in women, with recent birth cohorts of white women, and possibly black men, experiencing increasing pressure on incidence, 2) screen-detectable cancer incidence increased and mortality decreased, and 3) all other cancer incidence increased, with women in recent birth cohorts experiencing increasing pressure on incidence, and mortality decreased, but with recent time periods exerting increasing pressure on mortality in women. APC analysis of NHL, a subgroup of cancer unrelated to tobacco or screening, showed increasing risk, both incidence and mortality, with recent time periods exerting decreasing pressure on risk.

2.5.1 Tobacco-related Cancer

Other analysts have noted the impact of tobacco-related cancer, particularly lung and bronchus cancer, on declining U.S. male cancer incidence and mortality (McKean-Cowdin, Feigelson, Ross, Pike, & Henderson, 2000). In a recent update, Howe *et al*. (H. L. Howe et al., 2006) observed continuously decreasing incidence and mortality in U.S. men for cancers of the lung and bronchus, oral cavity and pharynx, bladder and pancreas, each a tobacco-related cancer. In our analyses, tobacco-related cancer trends were similar in white and black men, a pattern not highlighted by Howe *et al*. (H. L. Howe et al., 2006).

Featuring birth cohort- and time period-related effects, APC mortality analyses confirmed a slowing of the tobacco-related cancer epidemic in women. APC incidence analyses showed increasing tobacco-related cancer risk among white women born after 1949-1958, a pattern consistent with proportionally more younger women adopting cigarette smoking despite declines in the general population (Centers for Disease Control and Prevention, 2005a; Jemal, Ward, & Thun, 2005).

2.5.2 Screen-detectable Cancer

At a population level, a temporally circumscribed increase in incidence should result from the relatively recent and widespread use of mammographic screening for breast cancer, prostate-specific antigen (PSA) blood testing for prostate cancer, and endoscopic screening or fecal occult blood testing for colorectal cancer. If screening improves survival, then reduced late stage incidence and mortality should follow. Reduced cervical cancer incidence and mortality are

expected outcomes from Papanicolaou and human papilloma virus (HPV) testing for pre-malignant cervical dysplasia (National Cancer Institute, 2005a).

Our APC trend analyses confirmed increases in screen-detectable cancer incidence and decreases in mortality. If anything, APC parameter deviation patterns suggested waning incidence and mortality pressure from factors possibly affecting recent birth cohorts or recent time periods. Declining enthusiasm for cervical and breast cancer screening among certain population subgroups may explain recent birth cohort- or time period-specific reductions in screen-detectable cancer incidence (H. L. Howe et al., 2006; National Cancer Institute, 2005a). Examining results from mathematical models, Berry *et al*. (Berry et al., 2005) concluded that both screening and treatment-related factors are lowering U.S. breast cancer mortality. Other factors possibly affecting recent breast cancer trends include changes in hormone therapy use (Clarke et al., 2006), xenobiotic estrogen exposures (Clemons & Goss, 2001), and reproductive practices. Analysts (Chu, Tarone, & Freeman, 2003; Etzioni et al., 1999) offer conflicting opinions regarding the effects of PSA testing on recent prostate cancer mortality trends.

2.5.3 Cancer Unrelated to Tobacco or Screening

In 2007, the category created here of cancer unrelated to tobacco or screening is projected to include, in men and women, respectively, 27 and 32 percent of all-cancer incidence and 36 and 36 percent of all-cancer mortality (American Cancer Society, 2007b). Replicating methods used by Dinse *et al*. (Dinse, Umbach, Sasco, Hoel, & Davis, 1999), we also observed a generally increasing incidence of cancer unrelated to tobacco or screening. The analysis reported by Dinse *et al*. used 1975-1994 incidence data from the 9-registry SEER program. Adding corrections for reporting delay, our analysis used 1979-2003 incidence data, also from the 9-registry SEER

program. Relative to the Dinse *et al*. AAPC findings, we observed lower AAPC estimates in men and higher estimates in women, a result that eliminated the previously described gender difference with respect to AAPC.

Recent birth cohorts of women experienced increased pressure on the incidence of cancer unrelated to tobacco or screening. For white and black women, some have speculated that this phenomenon, appearing after the 1939-1948 birth cohort (Figure 2-1), may be possibly related to occupational exposures. According to Toossi (Toossi, 2002), the percentage of adult women working outside the home increased from 34 percent in 1950 to 60 percent in 2000. By 2004, women comprised 44 percent of the U.S. labor force (United States Office of Personnel Management, 2006). Exposures to benzene, chemical solvents, radiation, and pesticides and employment in food processing, textile, health case and garment industries have been linked to cancer risks in women (Zahm & Blair, 2003). Other work has found that blacks disproportionately work in blue-collar jobs, which may place them at higher risk from occupational cancers. In light of these observations, previous notions regarding the occupational attributable risk for cancer in women may need revision (Zahm & Blair, 2003).

Table B-2 (from Davis et al, in press in Biopharmacology, that comes from NCI) indicates that much of the decline in overall cancer mortality comes from reductions in cancers related to smoking or advances in early detection and treatment for screening-related cancer. It should be noted that the incidence of cancer unrelated to tobacco or screening is increasing against a backdrop of declining mortality. Improving treatments for cancers unrelated to tobacco or screening may be overcoming the expected effects of increasing incidence on mortality. However, if one assumes an average time period trend affecting cancer mortality favorably,

results from APC models (time period deviation plots; Figure 2-4) suggest that in recent time periods this favorable trend may be attenuating in women.

Analyses of U.S. NHL trends generally describe increasing incidence, with mortality first increasing until the late 1990s and decreasing thereafter, a favorable trend possibly related to new treatments for aggressive disease (Hennessy, Hanrahan, & Daly, 2004). Our APC analyses also showed generally increasing NHL incidence and mortality, although recent time period effects portend (Figures 2-2 and 2-4) lower risks in the future. Explanations for increasing NHL incidence have included improved ascertainment, changes in diagnostic practices, infection with human immunodeficiency virus (HIV), organ transplantation, and medical immunosuppression (Ansell & Armitage, 2005; Grulich, Vajdic, & Cozen, 2007). Environmental factors possibly contributing to NHL trends include occupational and agricultural exposures to pesticides, solvents, dusts, polychlorinated biphenyls, and bacteria (Band, Le, Fang, & Gallagher, 2004; Keller-Byrne, Khuder, Schaub, & McAfee, 1997; Lynge, Anttila, & Hemminki, 1997; Nelson, 2005; Pearce & Bethwaite, 1992; Rothman et al., 1997; Scherr, Hutchison, & Neiman, 1992; Zheng, Blair, Zhang, Weisenburger, & Zahm, 2002). Nevertheless, the continuing increases in NHL incidence remain poorly understood and are not likely to be due to improved ascertainment alone.

2.5.4 Methodology Issues

The use of population-based registries to study cancer incidence trends assumes uniform case ascertainment over time. Modeling past experience with delayed cancer case reporting, the SEER program has produced correction factors that can be used to adjust current cancer counts for case ascertainment expected to occur in the future (Midthune, Fay, Clegg, & Feuer, 2005). Our APC

analyses used these correction factors. However, other unrecognized biases in case ascertainment may exist such as fads in diagnoses.

Because it is difficult to examine a large array of age-specific rates, a variety of different methods for comparing changes in rates over time have been devised. Averaging disease rates across age groups is commonly used to simplify study of disease occurrence over time. However, study of age-adjusted rates assumes that time period-related factors that affect risk similarly regardless of age. In contrast, the APC model considers age-specific risks to be affected by factors related to birth cohort membership, in addition to factors related to calendar time. The APC model assumes that age-specific rates are affected equally by birth cohort-specific factors, regardless of time period, and equally by time period-specific factors, regardless of birth cohort membership. The APC model may fail to detect interesting time-limited factors having unique effects on specific age groups. For example, widespread prostate cancer screening started suddenly in the late 1980s and early 1990s. Prostate cancer screening was endorsed primarily for middle-aged men. As a result, the PSA test introduced a powerful time period effect that would not be expected to affect every birth cohort equally, but primarily those cohorts of men in their fifties and sixties during the late 1980s and early 1990s. APC models are not designed to detect prostate cancer risk effects related to the interaction between age group and time period.

2.6 CONCLUSION

In conclusion, APC analyses showed increasing incidence of NHL and cancer unrelated to tobacco or screening. Relative to earlier birth cohorts, white women belonging to recent birth cohorts may be facing an increased incidence of cancer unrelated tobacco or screening.

Determining whether the category of cancer unrelated to tobacco or screening constitutes a valid measure of general environmental influences on cancer risk requires more research.

It should be noted that the category of smoking-related cancers that we employed here provides a conservative indication of the impact of smoking on cancer. While about 85% of all lung cancer is believed to be associated with smoking, less than half of all bladder, kidney, esophageal and kidney cancer are thought to be smoking-related. As a result, the estimates provided in this exercise will overestimate the impact of smoking and underestimate the impact of other avoidable causes of these same tumors. Thus, occupational epidemiology has identified these same sites as elevated in groups employed with solvents, pesticides, and other workplace toxic agents.

Table 2-1 Results from cancer category-, gender-, and race-specific APC models of 9-registry SEER 1979-2003 age-specific incidence (thirteen 5-year age groups between 20-24 and 80-84 years), corrected for reporting delay

							Statistical significance[1]	
Cancer category	Gender	Race	Poisson model[2]	AAPC	95% C.I.	GR_{25}	cohort parameters	period parameters
Tobacco-related	Men	White	Extra	-0.92	-1.10, -0.73	0.79	****	**
		Black	Simple	-1.47	-1.87, -1.07	0.69	****	***
	Women	White	Extra	0.83	0.64, 1.02	1.23	****	*
		Black	Extra	0.18	-0.29, 0.66	1.05	****	
Screen-detectable	Men	White	Extra	2.51	1.93, 3.10	1.86		
		Black[3]	Extra	3.02	2.22, 3.83	2.11	*	
	Women	White	Extra	0.37	0.23, 0.50	1.10	****	****
		Black	Simple	0.10	-0.15, 0.36	1.03	****	**
Unrelated to tobacco or screening	Men	White	Extra	0.99	0.70, 1.29	1.28		****
		Black	Extra	0.72	0.28, 1.16	1.20	*	***
	Women	White	Extra	0.92	0.82, 1.01	1.26	****	
		Black	Simple	0.79	0.59, 0.99	1.22	**	
NHL	Men	White	Extra	2.33	1.78, 2.88	1.78		****
		Black	Extra	3.17	2.28, 4.06	2.18		***
	Women	White	Simple	1.92	1.70, 2.14	1.61		****
		Black	Simple	3.29	2.53, 4.05	2.24		

AAPC - Average annual percentage change, C.I. - confidence interval, GR_{25} - 25 year generational risk

1. p-value: <0.0001****, <0.001***, <0.01**, <0.05*

2. Poisson model fitted: simple Poisson model or extra-Poisson model

3. Analysis restricted to twelve 5-year age groups between 25-29 and 80-84 years

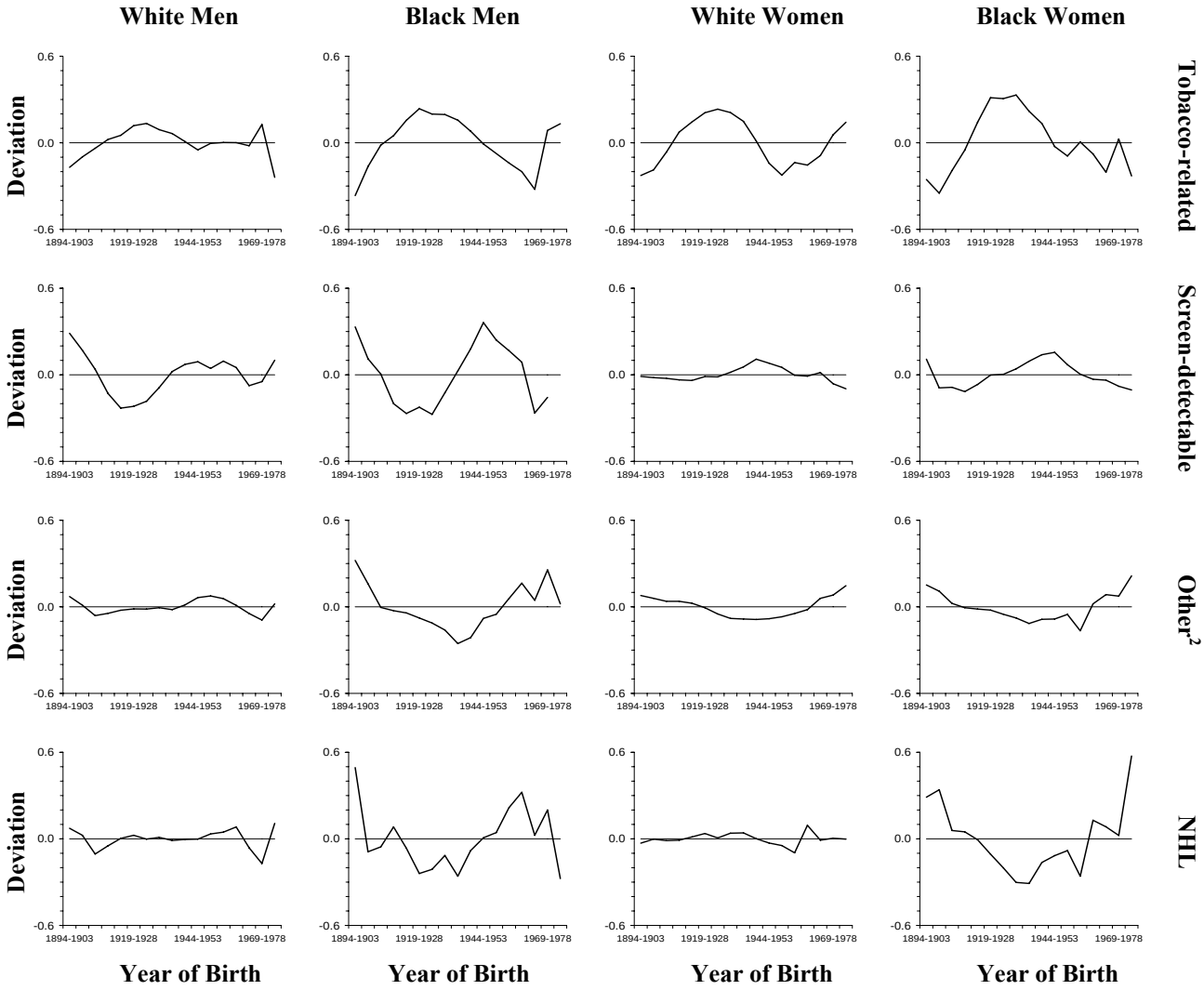
Table 2-2 Results from cancer category-, gender-, and race-specific APC models of U.S. 1969-2003 age-specific mortality (thirteen 5-year age groups between 20-24 and 80-84 years)

							Statistical significance[1]	
Cancer category	Gender	Race	Poisson model[2]	AAPC	95% C.I.	GR_{25}	cohort parameters	period parameters
Tobacco-related	Men	White	Extra	-0.89	-0.98, -0.80	0.80	****	
		Black	Extra	-0.92	-1.07, -0.77	0.79	****	****
	Women	White	Extra	0.80	0.68, 0.93	1.22	****	***
		Black	Extra	0.35	0.15, 0.55	1.09	****	****
Screen-detectable	Men	White	Extra	-1.10	-1.27, -0.93	0.76		
		Black	Extra	-0.34	-0.57, -0.11	0.92	****	
	Women	White	Extra	-1.44	-1.51, -1.36	0.70	****	****
		Black	Extra	-0.83	-0.95, -0.71	0.81	****	****
Unrelated to tobacco or screening	Men	White	Extra	-0.55	-0.61, -0.49	0.87	****	****
		Black	Extra	-0.35	-0.44, -0.26	0.92	****	****
	Women	White	Extra	-0.80	-0.84, -0.76	0.82	****	****
		Black	Extra	-0.68	-0.76, -0.60	0.84	****	****
NHL	Men	White	Extra	0.62	0.42, 0.82	1.17	****	****
		Black	Extra	1.39	1.09, 1.68	1.41		****
	Women	White	Extra	0.28	0.17, 0.40	1.07	****	****
		Black	Simple	1.54	1.27, 1.80	1.46	****	****

AAPC - Average annual percentage change, C.I. - confidence interval, GR_{25} - 25 year generational risk

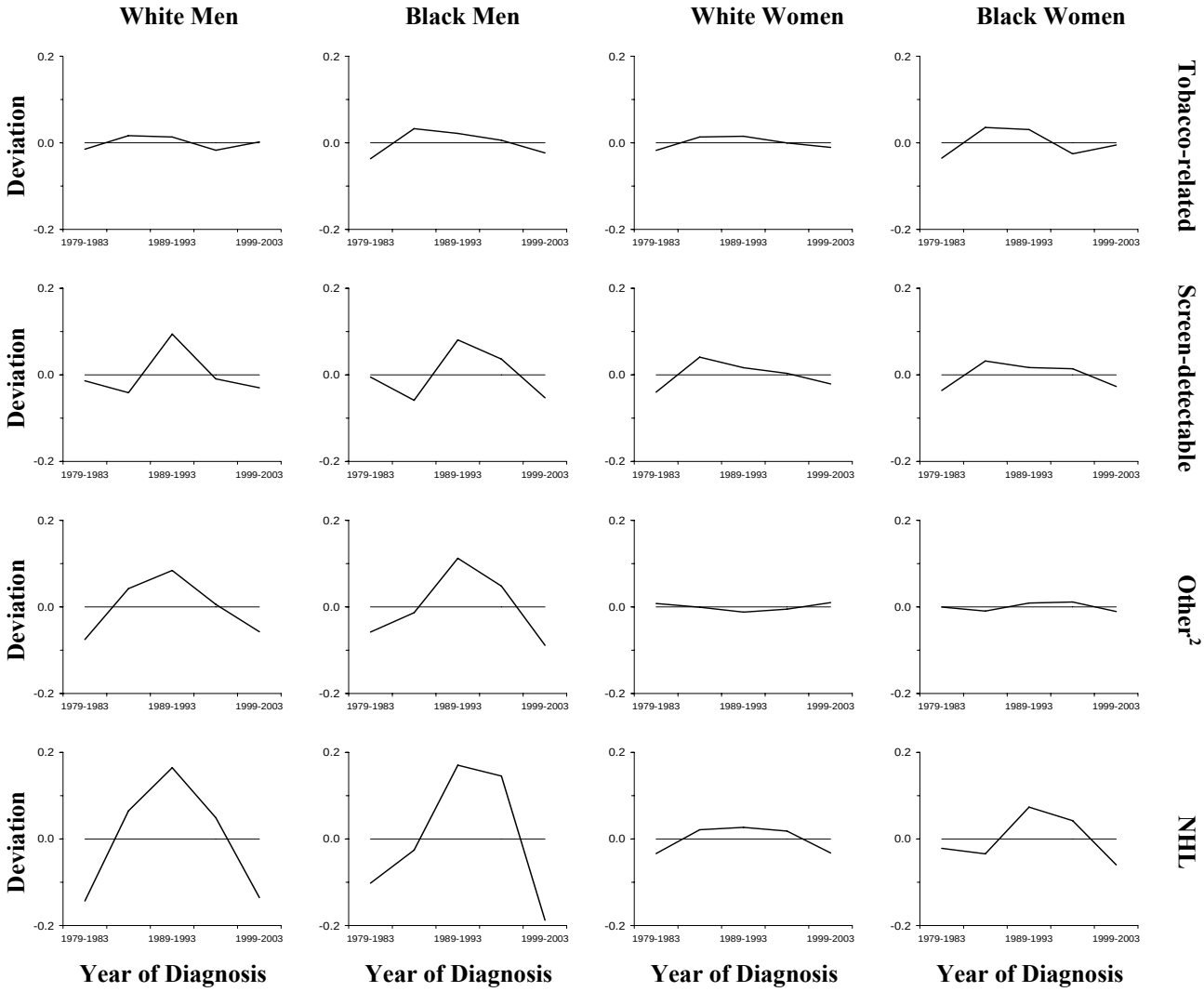
1. p-value: <0.0001****, <0.001***, <0.01**, <0.05*

2. Poisson model fitted: simple Poisson model or extra-Poisson model



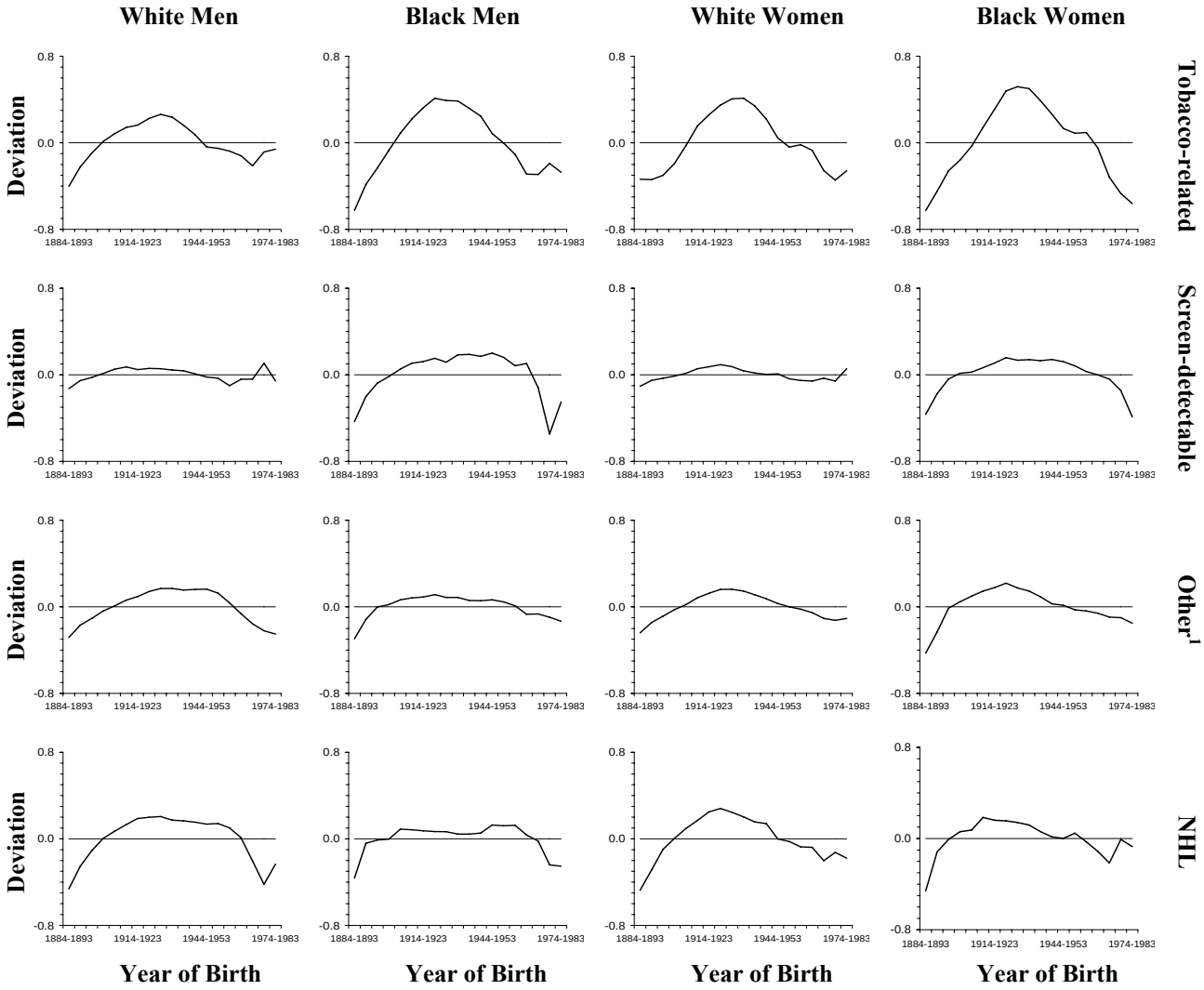
1 The APC analysis of screen-detectable cancer incidence in black men was restricted to twelve 5-year age groups between 25-29 and 80-84 years.
2 Cancer unrelated to tobacco or screening.

Figure 2-1 Results from cancer category-, gender-, and race-specific APC models of 9-registry SEER 1979-2003 age-specific incidence (thirteen 5-year age groups between 20-24 and 80-84 years), corrected for reporting delay; birth cohort-specific deviations added to the linear effect due to birth cohort



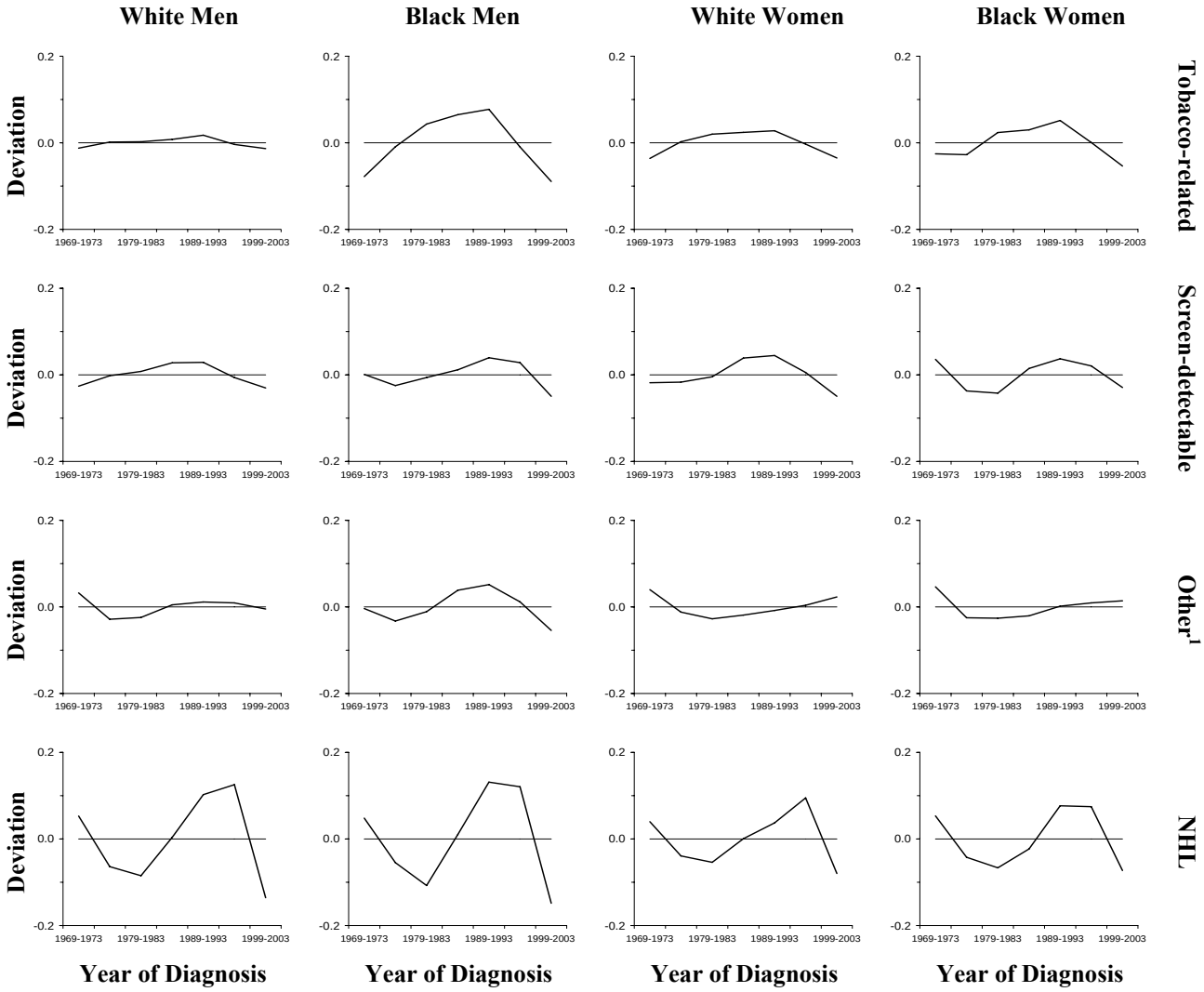
1. **The APC analysis of screen-detectable cancer incidence in black men was restricted to twelve 5-year age groups between 25-29 and 80-84 years.**
2. **Cancer unrelated to tobacco or screening.**

Figure 2-2 Results from cancer category-, gender-, and race-specific APC models of 9-registry SEER 1979-2003 age-specific incidence (thirteen 5-year age groups between 20-24 and 80-84 years), corrected for reporting delay; time period-specific deviations added to the linear effect due to time period



1. Cancer unrelated to tobacco or screening.

Figure 2-3 Results from cancer category-, gender-, and race-specific APC models of U.S. 1969-2003 age-specific mortality (thirteen 5-year age groups between 20-24 and 80-84 years); birth cohort-specific deviations added to the linear effect due to birth cohort



1. Cancer unrelated to tobacco or screening

Figure 2-4 Results from cancer category-, gender-, and race-specific APC models of U.S. 1969-2003 age-specific mortality (thirteen 5-year age groups between 20-24 and 80-84 years); time period-specific deviations added to the linear effect due to time period

3.0 PROJECT#2: NON-HODGKIN'S LYMPHOMA INCIDENCE: A COMPARISON OF DEMOGRAPHIC PATTERNS AND TEMPORAL TRENDS BETWEEN PENNSYLVANIA AND SEER REGISTRIES

Manuscript in Preparation

Yueh-Ying Han [1,2], Joel L. Weissfeld [1,2], Devra L. Davis [1,2], Gregg E. Dinse [3],

Bernard D. Goldstein [4], Evelyn O. Talbott [1]

[1] Department of Epidemiology, Graduate School of Public Health, Pittsburgh, University of Pittsburgh, Pittsburgh, Pennsylvania
[2] Center for Environmental Oncology, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania
[3] Biostatistics Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina
[4] Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

Address correspondence to Joel L. Weissfeld, MD, MPH

5150 Centre Ave, UPMC Cancer Pavilion, 4th floor, Pittsburgh, PA15232

Phone: 412-623-3313

Email: weissfeldjl@upmc.edu

3.1 ABSTRACT

Objective: Previous analyses of Surveillance Epidemiology and End Results (SEER) data have documented striking time-related increases in non-Hodgkin's lymphoma (NHL) incidence in the United States. Using data from the Pennsylvania Cancer Registry (PCR), which is not part of SEER, this study evaluated whether patterns of NHL differ in these two areas. **Methods:** Joinpoint and age-period-cohort (APC) regression models were applied to summarize gender- and race-specific NHL incidence time trends in Pennsylvania and the SEER registries between 1985 and 2004. Multiple regression analysis was used to identify demographic factors associated with age-adjusted county-specific NHL patterns in Pennsylvania. **Results:** According to APC analyses, Pennsylvania NHL incidence increased, on average, 1.5% and 2.5% annually in white and black men and 1.6% and 3.2% annually in white and black women. Average annual increase of gender- and race- specific NHL incidence in the SEER registries was similar to the PCR, except for white men (0.3% per year). Diffuse lymphoma appeared to be the most important contributor to the increase of NHL incidence in PCR and SEER registries. In Pennsylvania, NHL incidence was higher in counties with a greater percentage of urban residents ($p=0.005$). **Conclusion:** Though not identical, NHL incidence patterns in Pennsylvania and the SEER registries were generally parallel. The ecologic association observed in Pennsylvania NHL incidence and urban residence may be relevant to NHL patterns in the entire United States. Future work should explore what aspects of urbanization may account for the association with increased NHL risk.

3.2 INTRODUCTION

While the incidence of all cancers combined started to decrease in 1992 in the United States, non-Hodgkin's lymphoma (NHL) incidence rates have continued to demonstrate substantial increases for men and women since 1975 (H. L. Howe et al., 2006). With an estimated 63,190 new cases and 18,660 deaths projected for the year 2007, NHL is the fifth most common malignant neoplasm among men and women in the United States (American Cancer Society, 2007b). NHL occurrence has increased dramatically and internationally over the past few decades. Studies have shown that NHL incidence increased in Canada (Liu, Semenciw, & Mao, 2003), Australia (Grulich & Vajdic, 2005), Sweden, Denmark and Finland (Sandin et al., 2006) and other countries in Europe (Adamson et al., 2007). Increases in NHL incidence were also observed on a smaller scale within state cancer registries. Among Pennsylvania residents, 3330 new cases and 1140 deaths from NHL are expected for the year 2007. Based on the report from Pennsylvania Department of Health, in 1999-2003, the age-adjusted NHL incidence rates in Pennsylvania (24.5 per 100,000 in men and 17.0 per 100,000 in women) were slightly higher than rates in the United States (22.6 per 100,000 in men and 16.0 per 100,000 in women) (American Cancer Society, 2007b). Between 1994 and 2003, annual age-adjusted incidence was increasing among white and black men and women in Pennsylvania (Pennsylvania Department of Health, 2004).

Reasons for the rise in NHL incidence are not fully understood. Regardless of histologic subtypes, the impact of changes in case ascertainment, classification and diagnostic practice on overall NHL incidence trends is thought to be small (Clarke et al., 2004). NHL risk has been related to altered immunity, especially Human Immunodeficiency Virus (HIV) infection or

acquired immunodeficiency syndrome (AIDS) (Grulich, Vajdic, & Cozen, 2007). NHL cases are also attributed to certain environmental exposures, such as pesticides and solvents (Boffetta & de Vocht, 2007; Hartge & Smith, 2007). Phenoxy herbicide and 2,4-Dichlorophenoxyacetic acid (2,4-D), two widely utilized agricultural chemicals, have been consistently associated with higher NHL risk (Fisher & Fisher, 2004). Research has indicated that the causes for NHL vary according to histologic subtypes (Groves, Linet, Travis, & Devesa, 2000). It was documented that the distributions of NHL incidence by major subtypes varied across geographical regions related to host, racial, and environmental differences (Muller, Ihorst, Mertelsmann, & Engelhardt, 2005). Given the uncertainty of NHL etiology, researchers continue to investigate possible external causes of NHL, including environmental and occupational exposures.

To generate possible etiologic hypotheses, it is relevant and important to learn whether patterns of NHL occurrence observed in a nationwide cancer registry (Surveillance Epidemiology and End Results, SEER) agree with those observed in a statewide cancer registry. This study describes NHL time trends in the Pennsylvania Cancer Registry (PCR) and in the SEER registries by characterizing gender- and race-specific NHL incidence associated with age, calendar year, birth cohort, and histologic subtype. An ecologic analysis is performed to determine variables associated with county-specific NHL incidence in Pennsylvania.

3.3 MATERIALS AND METHODS

3.3.1 Data Sources

Information on newly diagnosed NHL cases in Pennsylvania from 1985 to 2004 was based on PCR data obtained from the Pennsylvania Department of Health. With the University of Pittsburgh Institutional Review Board exempting the study from review, we requested anonymous patient ID, along with gender, race, age at diagnosis, year of diagnosis, primary site, ICD-O-3 histology code, and individual FIPS county codes from the PCR Expanded Incidence File (EIF). PCR used the rules from SEER to determine multiple primaries. PCR converted ICD-O and ICD-O-2 to ICD-O-3 and used ICD-O-3 histology codes to recode the NHL cases. We used SEER*Prep software (version 2.3.5) (National Cancer Institute, 2006b) to convert case-level data in EIF to the SEER*Stat (version 6.3) database format (National Cancer Institute, 2007c), allowing data analysis by using SEER*Stat software. Using SEER*Stat, NHL cases and population counts were aggregated into 5-year age groups (thirteen groups from 20-24 to 80-84 years of age) and 5-year time periods (four groups from 1985-1989 to 2000-2004) according to gender and race (white and black). Each incidence rate was expressed as the number of cancers per 100,000 persons at risk and was age-adjusted relative to the U.S. 2000 standard population. Similar to the classification used by Groves et al. (Groves, Linet, Travis, & Devesa, 2000), NHL subtypes were classified according to the Working Formulation (WF), which was developed for prognostic purposes. We categorized NHL ICD-O-3 codes into six major subtypes: small lymphocytic NHL, follicular NHL, diffuse NHL, high-grade NHL, peripheral T-cell NHL, and NHL not otherwise specified (NOS). Information on newly diagnosed NHL cases in the U.S. was obtained from SEER*Stat (National Cancer Institute, 2006c) and analyses similar to PCR

data were conducted. We selected NHL data based on SEER 9 registries, which include states of Connecticut, Hawaii, Iowa, New Mexico, Utah, metropolitan areas of Atlanta, Detroit, San Francisco-Oakland, and 13 counties of Seattle-Puget Sound. The coverage includes 9.5% of the total U.S. population and does not include Pennsylvania or any cities in Pennsylvania.

Adjustment for reporting delay in cancer trend analyses could correct underreported cancer cases especially for persons diagnosed in recent years. To adjust current cancer cases accounting for delayed reports, delay-adjusted models produced by the National Cancer Institute (NCI) were used to correct cancer case counts in the SEER registries (National Cancer Institute, 2007a). In conformity with PCR regulations, the EIF we received excluded cancer reports from non-Pennsylvania cancer registries and federal (Department of Veterans Affairs, VA) medical care facilities. The PCR continually updates the EIF to include all available information. The EIF includes delayed reports, cancer reports received after publication of PCR annual cancer incidence and mortality summaries (Pennsylvania Department of Health, 2007). Comparing calendar year-, gender-, and race-specific NHL incidence counts in our restricted PCR EIF with incidence counts calculated and published annually (Pennsylvania Department of Health, 2007), we estimated correction factors that accounted for out-of-state or VA reporting and delayed reporting. Estimating correction factors of Pennsylvania reporting delays and out-of-state reports could diminish the magnitude of underestimation of recent NHL cases in cancer trend analysis.

Pennsylvania county population estimates were obtained from a publicly available internet resource (National Cancer Institute, 2006a) to estimate age-adjusted county-specific NHL incidence. County-specific U.S. 1990 and 2000 census estimates, including the percentage of white population, the number of males per 100 females, the percentage of employed civilians aged 16 years and over who work in the farming, fishing and forestry industries, and the

percentage of the population living in urban areas, were retrieved and averaged (United States Buerau of Census). We obtained cumulative cases of acquired immune deficiency syndrome (AIDS, not including human immunodeficiency virus (HIV) infection without AIDS) in Pennsylvania from 1980 to 2004 by county of residence from the Pennsylvania Department of Health (Pennsylvania Department of Health, 2005). Average annual AIDS incidence was calculated according to county-specific cumulative cases and population size.

3.3.2 Data Analysis

We modeled age-adjusted NHL incidence (corrected for delayed reports) in the Pennsylvania and SEER registries using joinpoint regression (National Cancer Institute. Statistical Research and Applications Branch, 2005) to describe long-term time trends for gender and race subgroups. This approach uses weighted least squares regression to fit multiple line segments, with unspecified join points, to the data on a log scale (Kim, Fay, Feuer, & Midthune, 2000). The program can determine the necessary number of joinpoints and their locations, and thus can assess whether trends in incidence rates have changed over time.

Using SAS software (version 9.1, SAS Institute Inc., Cary, NC, USA), we conducted age-period-cohort (APC) analyses to characterize gender- and race-specific NHL incidence in terms of temporal patterns associated with age, time period, and birth cohort. Using 5-year age and period intervals for each combination of race and gender, APC analysis was based on the methods reported by Dinse et al. (Dinse, Umbach, Sasco, Hoel, & Davis, 1999). An extra-Poisson model was fitted to account for the excess variation if a Poisson model did not fit well as judged by a 0.05-level chi-square goodness-of-fit test. We summed linear effects due to time period and birth cohort to form the drift (D), which represents long-term changes in cancer rates

over time. Cohort and period deviations from linearity were plotted to indicate departures from a trend line due to effects of birth cohort or time period. We estimated the Average Annual Percentage Change (AAPC) based on 5-year time intervals, where AAPC = 100* [exp (D/5)-1] and exp indicates an antilog. The generational risk represented as GR25, is the relative risk of cancer incidence in one generation versus the previous generation. Assuming that a generation spans 25 years, we estimated the 25-year generational risk by $GR_{25} = [\exp (D/5)]^{25}$.

We used county-specific NHL incidence in the 67 Pennsylvania counties to examine NHL geographic patterns between 1985 and 2004. Using GeoDa (version 0.95i) software, we created a shapefile by combining NHL data with Pennsylvania county Cartographic Boundary Files (United States Census Bureau, 2001). Gender-specific quartile maps (box maps, which focus on highlighting outliers in the data) were constructed to summarize the distribution of NHL incidence across the 67 counties. Multiple linear regression by the method of ordinary least squares was performed using SPSS software (version 12.0.1, SPSS Inc., Chicago, IL, USA) to evaluate whether county variations of race, gender, occupation, urbanization, and AIDS incidence associated with Pennsylvania county-specific NHL incidence between 1985 and 2004 time period.

3.4 RESULTS

According to joinpoint analysis, similar but not identical patterns of gender- and race-specific NHL age-adjusted incidence (corrected for delayed reports) were observed in Pennsylvania and the SEER registries (Figure 3-1). In Pennsylvania, NHL incidence was highest in white men, with a 2.6% estimated annual percentage change (EAPC) between 1985 and 1997 and a smaller

increase since 1997. Estimated NHL incidence among SEER white men rose 2.1 % annually from 1987 to 1995, leveled off between 1995 and 1998, and started to increase 1.1% annually in most recent years. NHL incidence in Pennsylvania black men rose aggressively, with a 7.4% increase per year during 1985-1994, and appeared to decrease slightly since 1994. In the SEER registries, NHL incidence in black men erupted in the late-1980s (EAPC=7.8%), began to level off in 1994, and peaked between 2002 and 2004. Female NHL incidence rates increased overtime from 1985 to 2004 (EAPC was 2.0% in white women and 2.6% in black women in Pennsylvania; EAPC was 1.5% in white women and 2.4% in black women in the SEER registries). However, NHL incidence rate was higher in white women than in black women. In Figure 3-2, age-specific incidence rates showed temporal increases among white men and women ages 65 years and older. Among white men younger than 45 years-old, NHL incidence appeared to be higher in SEER registries than in Pennsylvania. SEER white men younger than 55 years-old have experienced a decreasing pattern of NHL incidence since 1990-1994, except for the youngest age group. NHL incidence was decreasing consistently across the age groups in Pennsylvania white men.

APC model estimates of long-term increases in NHL incidence, 1985-2004, were given by the average annual percentage change (AAPC): 1.6% and 2.5% for white and black men, and 1.6% and 3.2% for white and black women in Pennsylvania, compared to 0.3% and 2.2% for white and black men, and 1.4% and 3.2% for white and black women in the SEER registries (Table 3-1). The generational risk (GR25) indicated that the four gender-race subgroups in Pennsylvania have between 49% and 119% greater risks of developing NHL than did the previous generation, compared to 9%-117% in the SEER registry. NHL incidence among Pennsylvania and SEER white men exhibits deviations from linearity due to birth cohort. In the

SEER registries, time period parameters significantly ($p<0.05$) improved the fit of APC models for NHL incidence in men and white women. Deviation added to linear effects has shown a downward sloping over recent birth cohorts and time periods among Pennsylvania and SEER white and black men (Figure 3-3)

In Pennsylvania, estimated incidence of small lymphocytic, follicular, and diffuse NHL was highest in white men, whereas high-grade NHL incidence was highest in black men (Appendix Table B-3). Estimated incidence of follicular NHL was higher in whites than in blacks, whereas the incidence of peripheral T-cell NHL was higher in blacks than in whites. Incidence of diffuse lymphoma was the highest among NHL subtypes and the rate increased overtime across Pennsylvania and SEER registries (Figure 3-4). In Pennsylvania, for two races combined, incidence of diffuse lymphoma has increased 2.8% annually in men and 2.7% annually in women between 1985 and 2004. Incidence of follicular and peripheral T-cell lymphoma has increased over time among men and women. Incidence of small lymphocytic lymphoma appeared to be stable during the study period. Incidence of high-grade NHL and not otherwise specified NHL has decreased in the 1990s. In the SEER registries, patterns of NHL incidence by subtype are similar to those in Pennsylvania.

In Pennsylvania, age-adjusted NHL incidence appeared to be higher in counties located in central areas (Appendix Figure B-5). According to multivariate regression analysis (Table 3-2), counties with higher percentages of residents living in urban areas had higher NHL incidence rates. Estimated NHL incidence was not associated with percentage of whites, ratio of males per 100 females, percentage of adults working in the farming, fishing and forestry industries, and AIDS incidence between 1980 and 2004.

3.5 DISCUSSION

Remarkable changes in NHL incidence in Pennsylvania and the SEER registries have been demonstrated. Overall, APC analyses confirmed the substantial increase of NHL incidence in the past two decades. Constant increase of NHL incidence was observed in women and in older age groups. In addition, estimated incidence of diffuse and follicular lymphoma has increased overtime. However, in the SEER registries, recent increase of NHL incidence in white men is unexplained. A striking increase in black men in 2002 may due to variation of case ascertainment. Further data including case ascertainment and population variation are needed to identify the pattern.

NHL is one of the major AIDS-associated malignancies highly impacted by the AIDS epidemic in the U.S. (Engels et al., 2006). In Pennsylvania, annual AIDS incidence increased rapidly since 1982 and started to level off in 1995. AIDS incidence was highest among people aged 30-49, and 79% of AIDS cases were males (Pennsylvania Department of Health, 2005). Featuring birth cohort and time period effects, APC analyses estimated a slowing NHL epidemic in Pennsylvania and SEER white and black men, an observation which is consistent with the decreased HIV/AIDS epidemic and improved treatment in the 1990s. Nevertheless, AIDS epidemic could not explain NHL patterns in women or elderly. Higher NHL incidence rate has been associated with altered immunity, including transplantation, blood transfusion, and immunosupression (Chow & Holly, 2002; Fisher & Fisher, 2004). (Grulich, Vajdic, & Cozen, 2007). Infectious agents, such as Epstein-Barr virus (EBV) are being regarded as established causes of NHL, especially in individuals with immune dysfunction (Engels, 2007) However, how these factors contribute to the NHL time trend are either unknown or thought to be small.

NHL cases are often attributed to uncertain environmental exposures. Relationship between occupational benzene exposure and NHL has been suggested (Goldstein & Witz, 1999; M. T. Smith, Jones, & Smith, 2007; Vineis, Miligi, & Costantini, 2007). Exposures to agricultural chemicals and solvents are associated with increased NHL risk (De Roos et al., 2003; Fritschi et al., 2005; Hardell & Eriksson, 1999; Rafnsson, 2006; Schroeder et al., 2001). According to data from National Agriculture Statistics Services, central Pennsylvania has the most acres of farmland in the state (United States Department of Agriculture, 2007). In Pennsylvania, counties located in central areas appear to have higher NHL incidence. Study on farming activity in central Pennsylvania should be considered to identify whether application of agricultural chemicals is related to NHL risks.

The study indicated that NHL risk differs by gender. Continuously increased NHL incidence among women is unexplained. Use of hair dyes among women has been associated with NHL development but the evidence is controversial (Takkouche, Etminan, & Montes-Martinez, 2005). Increased female workforce and occupational exposure in recent decades may be contributing. Study on the relationship between environmental exposures and gender-specific NHL incidence, especially among older groups with long-term exposure could provide directions of NHL etiology study.

Regarding NHL histologic subgroups, diffuse and follicular lymphoma appeared to be the main contributor of increased NHL incidence in Pennsylvania and the SEER registries. Occurrence of NHL subtypes may vary by risk factors. Industrial workers exposed to electromagnetic field and metals have been related to development of diffuse lymphoma (Band, Le, Fang, & Gallagher, 2004; Zheng, Blair, Zhang, Weisenburger, & Zahm, 2002). It has been suggested that exposures to agriculture chemicals exert stronger effects on diffuse and follicular

lymphoma (Band, Le, Fang, & Gallagher, 2004; Fritschi et al., 2005). In the United States, total pesticide use rose from 647 million pounds in 1964 to 1144 million pounds in 1979 and then declined to 971 million pound in 1995 (United States Environmental Protection Agency, 1997). The role of occupation and pesticide exposure in the increase of diffuse and follicular lymphoma obliges further evaluation. Lifestyle including diet, obesity, and physical activity has drawn attention to subtype NHL etiology (Chang et al., 2006; Chang et al., 2005; Kelemen et al., 2006; Zheng et al., 2004). Obesity-related NHL risk increase is greater for diffuse and follicular subtypes (Pan, Mao, Ugnat, & Canadian Cancer Registries Epidemiology Research, 2005; Skibola et al., 2004). In Pennsylvania, 16% of adults were obese in 1995 and it rose to 25 % in 2005 (Pennsylvania Department of Health, 2006a). The emerging obesity epidemic may contribute to the increasing incidence of NHL. Studies of the association between obesity, diet, and lymphoma subtype are warranted to confirm the associations.

High-grade NHL, a subtype related to 90% of AIDS-related NHL, was the most rapidly increasing subtype in 1978-1995, particularly among men (Groves, Linet, Travis, & Devesa, 2000). In recent years, observation of decreased high-grade NHL in Pennsylvania and SEER registries may reflect the decline of AIDS epidemic. Recent decrease of not otherwise specified NHL may due to improved histologic diagnosis of lymphoma. Findings from our population-based study provide evidence that the distribution of NHL subtypes differs in demographic patterns and incidence trends. Although changing classifications of NHL histologic subtypes and improved diagnosis may misclassify NHL histologic subtypes, future investigations should evaluate NHL risks according to subtype, as defined by histology and classification criteria.

In Pennsylvania, populations living in urbanized counties have been found to have higher NHL incidence. Devesa *et al* has found that incidence of NHL was lower in rural areas (Devesa

& Fears, 1992; Muller, Ihorst, Mertelsmann, & Engelhardt, 2005). Although interpreting the ecologic relationship between urbanization and increased NHL risk is limited, disparity of urbanization might be relevant to NHL risk in the entire US. Percent of total population living in urban area has increased 23% from 1950 to 2000 (64% in 1950 and 79% in 2000) (United States Bureau of Census). Characteristics of urban living, such as population density and structure, lifestyle, development process and socioeconomic factors, should be studied to identify potential NHL risk factors and to account for NHL geographic distributions in Pennsylvania.

3.6 CONCLUSION

Although the time course of NHL occurrence is not well explained, increased NHL incidence patterns observed in Pennsylvania generally paralleled with those observed in the SEER registries. Analyzing cancer data within Pennsylvania have provided insights of NHL epidemic and helped to identify candidacy of NHL etiologic factors. In conclusion, gender-specific risk factors should be investigated to explain continuous increase of NHL among women. Environmental exposure and changes of lifestyle provide clues of NHL etiologic study by evaluating patterns of histologic NHL subgroups. Future work should explore what factors and aspects of urbanization may account for association with increased NHL occurrence.

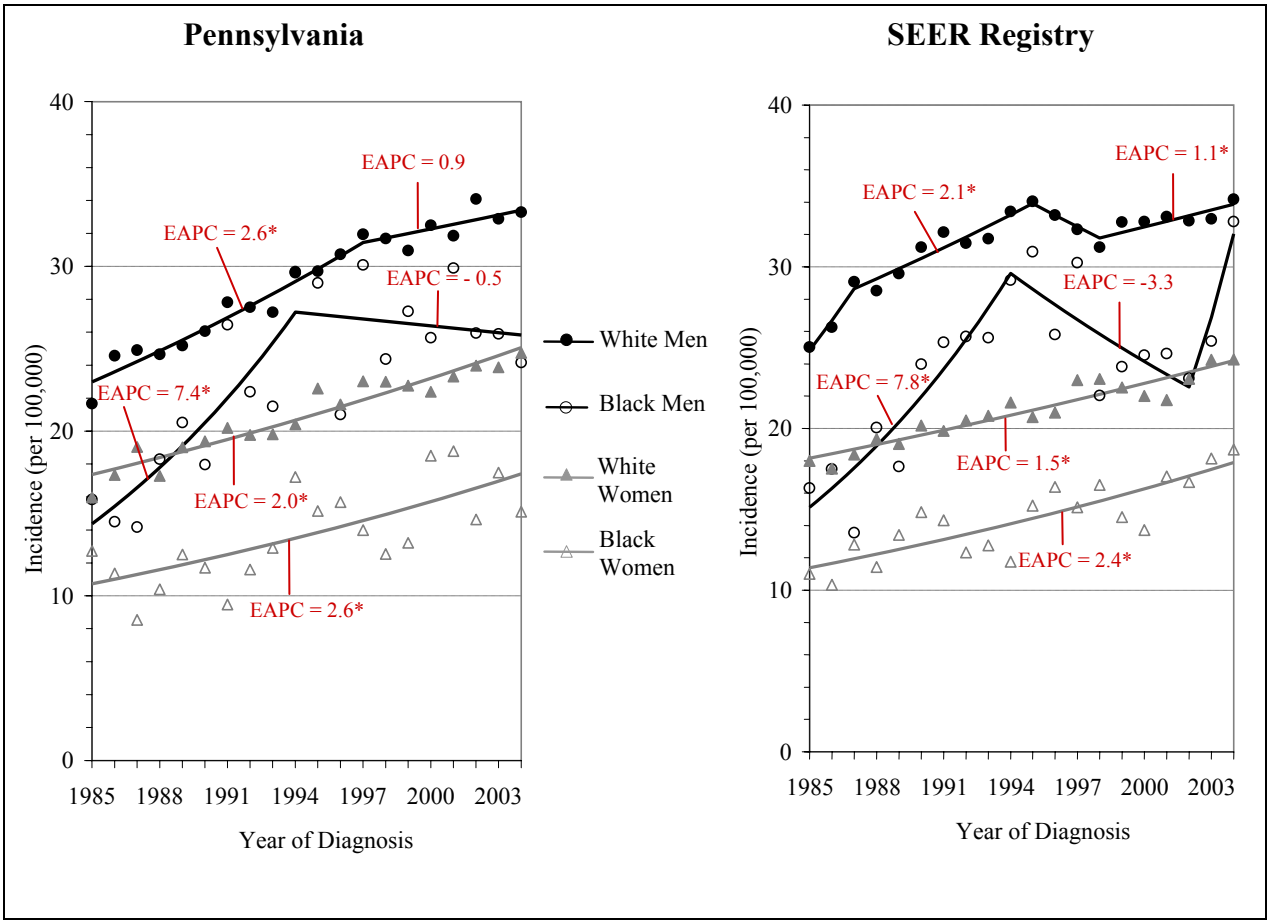


Figure 3-1 Joinpoint estimates of NHL incidence in the Pennsylvania and SEER registry by gender and race for persons diagnosed from 1985 to 2004 at ages 20 to 84, adjusted for age and reporting delays

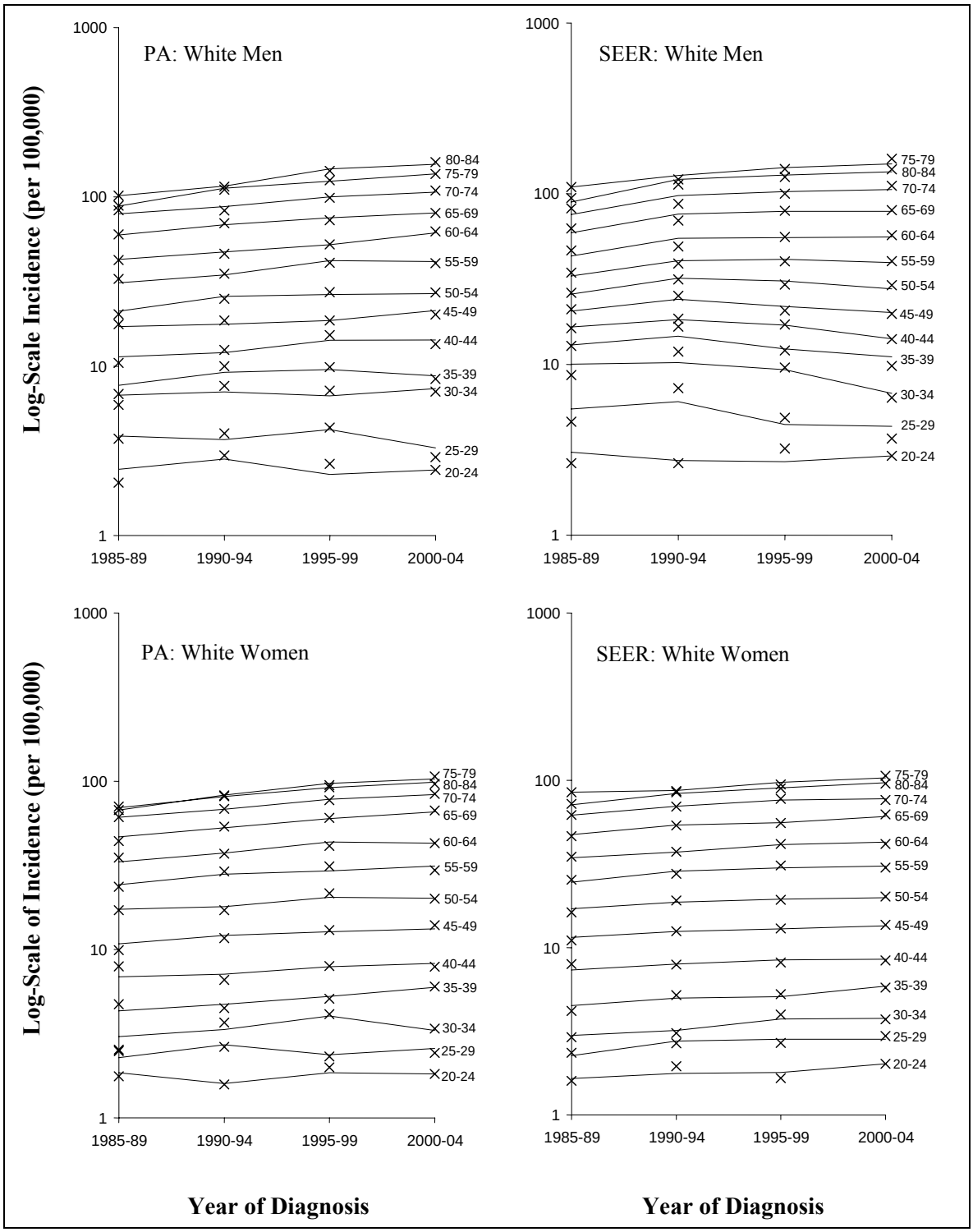


Figure 3-2 NHL incidence (Log-Scale) in the Pennsylvania (PA) and SEER registry by age group and time period for white men and women diagnosed from 1985 to 2004 at ages 20 to 84, adjusted for reporting delays. The X's are observed values and the lines are predicted values based on an age-period-cohort model with 5-year age and period interval

Table 3-1 Results from separate age-period-cohort models for NHL incidence in the Pennsylvania and SEER registry by gender and race for persons diagnosed from 1985 to 2004 at ages 20 to 84, adjusted for reporting delays. The APC analyses used 5-year age and period intervals

							Significance level [1]	
Registry	Gender	Race	Poisson model fits	AAPC	95% C.I.	GR25	cohort curvature	period curvature
PCR	Men	White	Yes	1.55	1.16, 1.94	1.47	0.0189*	0.4477
		Black	No	2.49	1.02, 3.99	1.85	0.4719	0.1199
	Women	White	Yes	1.61	1.18, 2.05	1.49	0.1969	0.1165
		Black	Yes	3.18	1.77, 4.61	2.19	0.7114	0.9596
SEER	Men	White	No	0.33	-0.23, 0.89	1.09	0.0159*	0.0077**
		Black	No	2.23	1.20, 3.27	1.74	0.4131	0.0005***
	Women	White	Yes	1.42	1.17, 1.67	1.42	0.1476	0.0262*
		Black	No	3.15	2.05, 4.26	2.17	0.4660	0.4485

PCR – Pennsylvania Cancer Registry

AAPC - Average annual percentage change, C.I. - confidence interval, GR25 - 25 year generational risk

[1] <0.001***, <0.01**, <0.05*

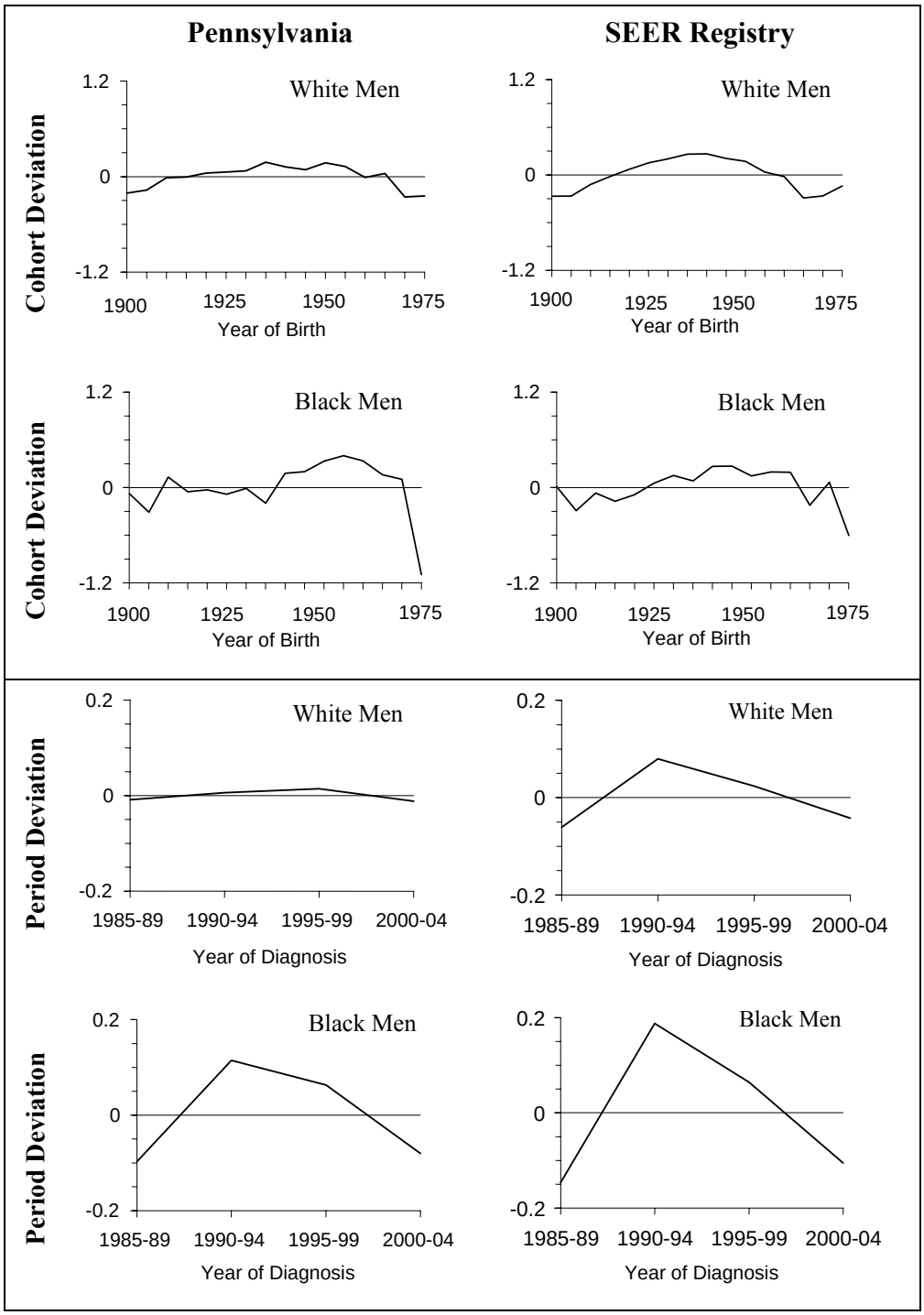


Figure 3-3 Cohort and period deviations from linearity as estimated from separate age-period-cohort models for NHL incidence in the Pennsylvania and SEER registry in men by race e for persons diagnosed from 1985 to 2004 at ages 20 to 84, adjusted for reporting delays

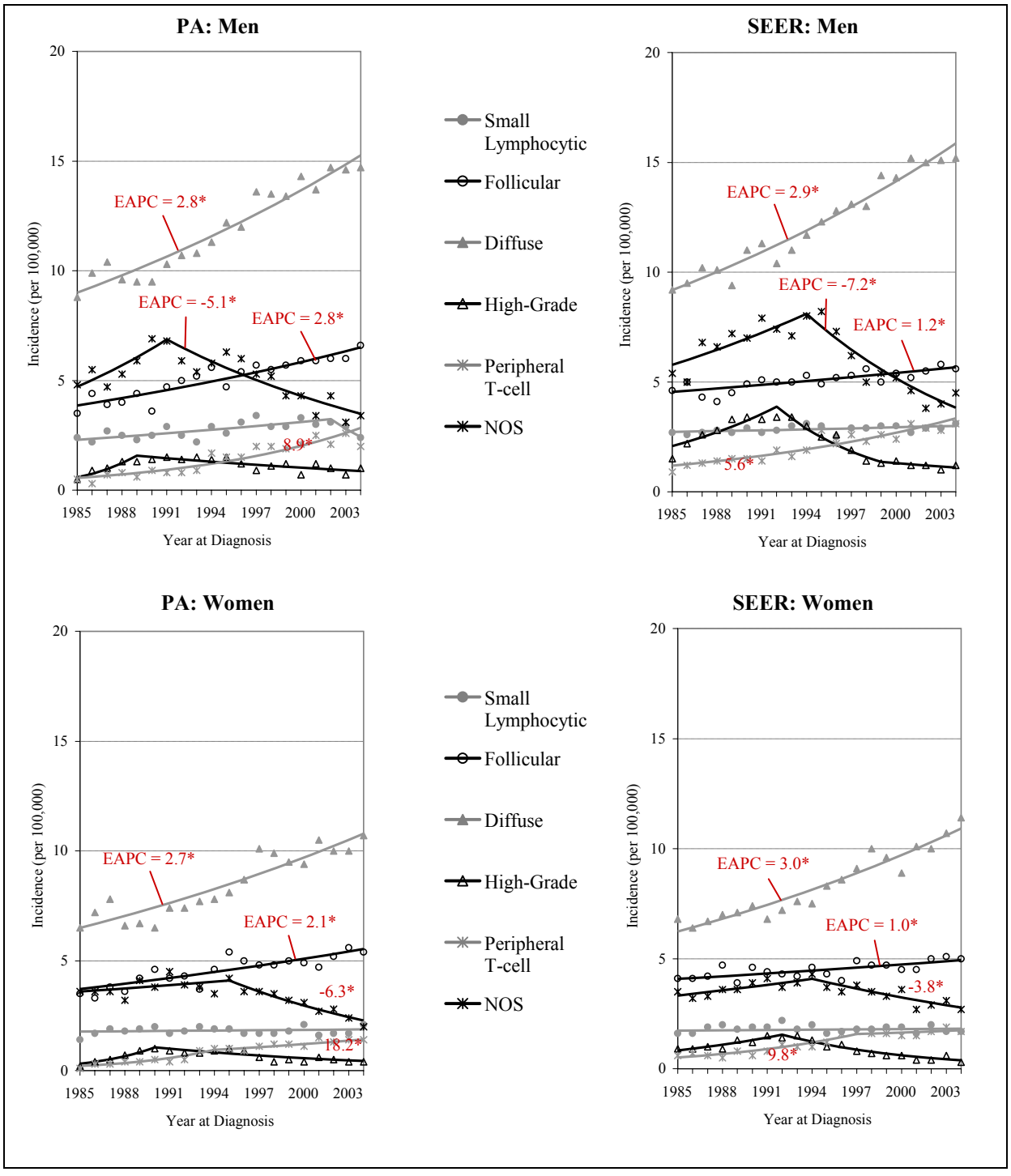


Figure 3-4 Joinpoint estimates of NHL incidence in the Pennsylvania and SEER 9 registries by gender, year of diagnosis, and histologic subtype of the Working Formulation (WF) classification for white and black combined diagnosed from 1985 to 2004 at ages 20 to 84, age adjusted to the 2000 U.S. standard population

Table 3-2 Results from multiple linear regression analysis for assessing which factors are important predictors of NHL incidence in Pennsylvania among persons diagnosed from 1985 to 2004 at ages 20 to 84, adjusted for age

Independent variables	Coefficient	Std. Error	t-statistic	P-value
Constant	-4.899	17.718	-0.277	0.783
White (%)	0.223	0.165	1.351	0.182
Men/100Women	0.022	0.029	0.751	0.456
FarmOccupPop (%)	0.108	0.448	0.241	0.811
Population in Urban (%)	0.070	0.024	2.950	0.005
AIDS Incidence[2]	0.190	0.187	1.020	0.312

1. R-squared from regression model equals 0.22.
2. AIDS incidence was calculated on the basis of county-specific accumulated AIDS cases and population sizes, 1980-2004

4.0 PROJECT#3: EVALUATION OF ARSENIC LEVLES IN GROUND WATER AND CANCER INCIDENCE IN IDAHO: AN ECOLOGIC STUDY

Manuscript in Preparation

Yueh-Ying Han [1,2], Joel L. Weissfeld [1,2], Devra L. Davis [1,2], Gregg E. Dinse [3],

Bernard D. Goldstein [4], Evelyn O. Talbott [1]

[1] Department of Epidemiology, Graduate School of Public Health, Pittsburgh, University of Pittsburgh, Pittsburgh, Pennsylvania
[2] Center for Environmental Oncology, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania
[3] Biostatistics Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina
[4] Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

Address correspondence to Joel L. Weissfeld, MD, MPH

5150 Centre Ave, UPMC Cancer Pavilion, 4th floor, Pittsburgh, PA15232

Phone: 412-623-3313

Email: weissfeldjl@upmc.edu

4.1 ABSTRACT

Objective: Long-term exposure to arsenic in drinking water has been related to multiple types of cancers. The long latency of cancer has challenged arsenic-related cancer research. This ecologic study evaluates the relationship between arsenic level in ground water and cancer incidence in Idaho. **Method:** Using cancer incidence data, 1990-2005, from the Cancer Data Registry of Idaho and arsenic data from the Idaho Department of Environmental Quality, we conducted spatial analysis to identify geographic patterns in incidence for cancers of the bladder, kidney, liver, lung, non-Hodgkin's lymphoma (NHL), and all sites combined. Spatial regression was applied to evaluate the relationship between arsenic level in ground water and cancer incidence. **Results:** Spatial clustering of lung cancer incidence was demonstrated in males and females. Incidence for kidney cancer and all sites combined was spatially clustered in females but not in males. Higher incidence for lung cancer and all cancer sites combined were clustered in north Idaho, while lower rates clustered in southeastern Idaho. No relationship was found between arsenic level in ground water and cancer incidence. However, smoking prevalence is significantly associated with cancer incidence for bladder, kidney, lung, and all sites combined. **Conclusion:** Exposure to arsenic in ground water is not associated with cancer incidence in Idaho. Although this study was not able to identify other environmental risks, smoking is contributing to excess cancer occurrence among Idaho counties. To clarify relationship between arsenic exposure and cancer risk, individual risk assessment adjusted for smoking and occupation is crucial in southwest counties where ground water has been found to contain relatively higher levels of arsenic.

4.2 INTRODUCTION

Arsenic is a metalloid found in water, soil, and air from natural sources and anthropogenic activities such as mining and agriculture applications. Arsenic is also used to manufacture agricultural products such as insecticides, herbicides, and fungicides (ASTDR, 1993). Populations exposed to arsenic from geologic contaminated water are found around the world, including locations in Taiwan, Chile, Mexico, Argentina, Bangladesh, and the United States (Kapaj, Peterson, Liber, & Bhattacharya, 2006; Marshall et al., 2007; Tapio & Grosche, 2006; Yoshida, Yamauchi, & Fan Sun, 2004). These studies have also demonstrated that arsenic exposure is associated with chronic health effects including cancer of the bladder, lung, liver, kidney, and skin. However, the latency of arsenic-related cancer may be as high as 50 years. The mode of arsenic action may relate to oxidative stress, altered DNA methylation, altered DNA repair, enhanced cell proliferation, gene amplification, and chromosomal abnormalities (Huang, Ke, Costa, & Shi, 2004; Hughes, 2002; Kitchin, 2001). Limitations on the individual exposure assessment and potential confounders hamper the etiologic research on arsenic-related cancer risk (Chappell et al., 1997; Gebel, 2000). In 2001, after considering the analyses and risk assessment on arsenic from National Research Council (NRC), the United States Environmental Protection Agency (EPA) lowered the maximum contaminant level (MCL) for arsenic in drinking water from 50 μg/l to 10 μg/l (United States Environmental Protection Agency, 2001). The change of MCL for arsenic in drinking water and cancer risks associated with arsenic exposures, particularly at relatively lower exposure levels in the United States population, has generated discussion and debate.

In order to provide an understanding of water-quality conditions and how they may vary locally, regionally, and nationally, the United States Geological Survey (USGS) has collected and analyzed several contaminants, including arsenic, in potable water from 18,850 wells in 595 counties across the United States since the 1970s. In 24 percent of the counties where data were available, at least 10% of the samples had arsenic concentration exceeding 10 µg/l. According to the geographic map that USGS developed, widespread and elevated arsenic concentrations in the United States were found in the northeast, west and the midwest, including the state of Idaho (Welch, A.Watkins, Helsel, & Focazio, 2000). Ground water samples from the Idaho Statewide Ambient Ground Water Quality Monitoring Program have been analyzed for arsenic since 1990. Approximately 15 percent of the 1,900 statewide monitored sites have arsenic concentration exceeding the MCL of 10 µg/L (Hagan, 2004; Neely, 2002). Ground water in southwest and south central Idaho has the highest density of arsenic concentrations that exceed the MCL. Since 95% of the population relies on ground water as a source of drinking water, arsenic has become a concern for many ground water users in Idaho. Recently, geographic information system (GIS) technology has been applied to analyze the geographic distribution of causative agents and their relationships to health outcomes in exposed populations. Applying GIS technology, this ecologic study examines the spatial patterns of county-specific cancer incidence in Idaho and evaluates the relationship between arsenic level in ground water and cancer incidence in Idaho.

4.3 MATERIALS AND METHODS

4.3.1 Data Collection

The Idaho Department of Water Resources (IDWR) provided 6,019 arsenic concentration measurements (range 0-950 µg/l) for approximately 1,900 Idaho ground water sources (wells and springs) sampled during summer months between 1990 and 2005. Laboratory analysis of arsenic in ground water is based on IDWR standard operating procedures according to EPA recommended methods for drinking water contaminants (EPA 200.9) (Neely, 2002). County rankings based on calendar year-specific arsenic concentration measurements were similar from year to year (data not shown). Therefore, for each of the 44 counties in Idaho, we averaged arsenic concentration values over all water sources and time points to estimate personal arsenic exposures for county residents. The number of measurements by county ranged between 35 (Camas County) and 570 (Canyon County). According to the IDWR (Neely, 2002), two-thirds of sampled ground water sources are designated for domestic use.

We requested county- and gender-specific newly diagnosed cases for all ages combined, from 1990 to 2005, for cancer of the bladder, kidney and renal pelvis, liver and bile duct, lung and bronchus, non-Hodgkin's lymphoma (NHL), and all sites combined from Cancer Data Registry of Idaho (CDRI). We obtained data aggregated in 5-year intervals from 1990 to 2004 and an overall data aggregated from 1990 to 2005. Reporting of cases was based on county of residence at the time of diagnosis. The cancer rates were age adjusted to the U.S. 2000 standard population. County-specific smoking prevalence and the percentage of population classified as not obese (body mass index (BMI) < 25) were retrieved from Geographic Report by CDRI, which is based on the Behavioral Risk Factor Surveillance System (BRFSS), 1997-2005.

Overall, 44,872 respondents were identified from BRFSS in the State of Idaho. Data were weighted by probability of selection, and poststratified to 2004 Idaho population estimates by age group, sex, and county. A minimum of 30 respondents was required to generate county-level statistics (Cancer Data Registry of Idaho, 2006). County-level demographic data on percentage of white population, number of males per 100 females, and population density per square mile of land area were retrieved from 2000 U.S. census summary files (United States Bureau of Census, 2005). Power calculation is indicated in Appendix C.

We also grouped Idaho counties into three categories according to arsenic concentration in ground water sources: low arsenic county ($\leq$2μg/l, n=23), intermediate arsenic county (2-9μg/l, n=16), and high arsenic county ($\geq$10 μg/l, n=5). Gender and cancer site-specific incidence was calculated by these three groups. Confidence interval is 95% for incidence rate.

4.3.2 Spatial Autocorrelation Analysis

Using ArcGIS software (ESRI, Redlands CA, version 9.1), we created a shapefile, a digital vector format for storing geometric location and associated attribute information, by combining cancer data and independent variables with the Idaho county map from the US census 2000 Cartographic Boundary Files (United States Bureau of Census, 2001). The shapefile was imported into GeoDa (University of Illinois, Urbana, IL, version 0.95i), a free program which provides a graphical interface for methods of descriptive spatial data analysis, to conduct spatial analysis for four different time period: 1990-1994, 1995-1999, 2000-2004, and 1990-2005. Based on the categorized incidence rates, quartile box maps, focused on highlighting outliers in the data, were constructed to check the cancer distribution of 44 counties between 1990 and

2005. An outlier was defined as any value that lies above 1.5 times of interquartile range (the difference in value between the 75% and 25% observation) added to the 75th percentile or below 1.5 times the interquartile range subtracted from the 25th percentile. Using Empirical Bayes Smoothing procedure to remove the effect of the varying population across the counties, we constructed smoothed quartile maps.

Contiguity weights file was first created to define neighbors based on queen contiguity which classifies areas as neighbors even if their borders only touch each other (Anselin, 2005). Global spatial autocorrelation analysis with queen-weighted matrix was used for testing clustering of age-adjusted cancer incidence in Idaho counties by gender and time period. In the global spatial autocorrelation test, Moran's I is the correlation value in a location that is calculated by the spatially weighted average of the neighboring values. Similar to the correlation coefficient, the values of Moran's I range from +1 meaning positive spatial autocorrelation (clustering) to -1 indicating negative spatial autocorrelation (dispersed); with values close to 0 indicating a random pattern (Moran, 1950). Local Indicator Spatial Autocorrelation (LISA) was used to test clusters of age-adjusted cancer incidence in Idaho counties by gender and time periods. Five scenarios may emerge in a LISA cluster map: (1) locations with high values and similar neighbors: high-high, (2) locations with low values and similar neighbors: low-low, (3) locations with high values and low value neighbors: high-low, (4) locations with low values and high-value neighbors: low-high, and (5) locations with no significant local autocorrelation (Anselin, 2005).

4.3.3 Spatial Regression Analysis

Using a queen-weighted matrix file, ordinary least squares regression was first performed to evaluate the spatial dependence in the model. County was used as the geographic unit and age-adjusted cancer incidence was the dependent variable. Arsenic levels, percentage of white population, number of men per 100 women, population density on land, percentage of current smokers, and percentage of residents with a BMI less than 25 were included in the regression model as covariates. Spatial lag and spatial error are two types of spatial dependence that were considered in this study. Spatial lag is caused by the autocorrelation present in the variables that are not included in the model. Spatial error is due to the autocorrelation present in the dependent variable. In the regression output of spatial autocorrelation diagnostics, Lagrange Multiplier (LM-lag) tests for a missing spatially lagged dependent variable were used in the linear models. Lagrange Multiplier (LM-error) tests for spatial error dependence were also used in the linear models. If neither LM-Lag nor LM-Error statistics rejected the null hypothesis, indicating no spatial lag or spatial error in the model; results from the ordinary least squares (OLS) model was selected for interpretation. The selection of spatial regression model is based on the rules documented in the Workbook of GeoDa (Anselin, 2005).

4.4 RESULTS

County average arsenic levels in ground water are highest in southwest Idaho (Figure 4-1). The counties of Owyhee, Washington, Twin Fall, Payette, and Canyon have relatively higher percentages of samples containing arsenic levels higher than 10 μg/l (64%, 56%, 49%, 38% and

37%, respectively, data not shown). Compared to SEER program, gender-specific cancer incidence for lung, liver and all sites combined in males and females is significantly lower in Idaho (Table 4-1). In Figure 4-2, box maps indicate that northern and western Idaho has a higher incidence for bladder, kidney, liver, and lung cancer, and for cancer in all sites combined. Counties in southeast Idaho have a lower incidence for cancer of the bladder, liver, lung, and all sites combined. NHL incidence is higher in Idaho's central and southern counties. Empirical Bayes smoothed maps show that no outliers were observed in Idaho county-specific cancer incidence, although more counties located in northern and central Idaho fall into the third and forth quartile in each cancer site (map not shown).

Female county-specific incidence rates for kidney cancer and for all sites combined showed significant ($p<0.05$) spatial autocorrelation (Figure 4-3). Incidence for lung cancer has the highest Moran's I statistics: 0.42 in males and 0.57 in females. The LISA cluster map illustrates the significant locations by coloring different types of spatial autocorrelation. Shown in Figure 4-3, high-high clusters of cancer incidence for all sites combined in females were identified in five northern counties. In males, counties in southeast Idaho had a lower bladder cancer incidence surrounding counties that had similar bladder cancer incidence. Low-low cluster of cancer incidence for all sites combined were also found in males and females. In males, significant high-high clusters of kidney cancer incidence were found in Kootenai County and Boise County. In both genders, higher incidence for lung cancer was clustered in northern Idaho and lower incidence for lung cancer was clustered in southeastern Idaho. No specific patterns were observed in incidence for liver cancer. Butte, Madison, Bonneville and Washington Counties had higher NHL incidence and were surrounded by counties with higher NHL incidence (Figure 4-3). Cancer incidence for all sites combined was spatially autocorrelated

in males and females from 1990 to 1994, and in females from 1995 to 1999 (Table 4-2). Spatial clustering of lung cancer incidence was observed with a highest Moran's I value in both genders from 1990-1994 to 2000-2004, except for males between 1995 and 1999.

Gender-specific Pearson correlation coefficient is presented in Table 4-3. No relationship was found between arsenic exposure in ground water and cancer risk in Idaho; however, incidence for lung cancer had the highest correlation coefficient. For males, compared to the rates in the low arsenic counties, the intermediate and high arsenic counties have significantly higher incidence for all sites combined and lung cancer (Table 4-4). Among females, cancer incidence for all sites combined, lung and bladder cancer was higher in the intermediate arsenic counties compared to the low arsenic counties. Based on spatial regression analysis, ordinary least squares (OLS) estimation was selected for cancer of the bladder, kidney, liver, NHL, and all sites combined. According to maximum likelihood estimation, the spatial lag model was selected for lung cancer (Table 4-5 and Appendix Table B-4). Between 1990 and 2005, cancer incidence was not associated with arsenic level in ground water and population density on land. Higher prevalence of current smokers was associated with higher incidence for cancer of the bladder, kidney, lung, and all sites combined. Incidence for kidney cancer was higher among counties having a higher proportion of white residents. Increased incidence for liver cancer was associated with a higher proportion of male residents and a higher proportion of population with higher BMI value (Table 4-5). Figure 4-4 demonstrates that smoking prevalence was higher in counties located in northern and southwestern Idaho, while counties in eastern Idaho had a lower smoking prevalence (Moran's I=0.34, p=0.004).

4.5 DISCUSSION

In Idaho, lower incidence for cancer of bladder, lung, and all sites combined was observed in the eastern area, a similar result to the IDCR reports (Christopher J. Johnson & Carson, 2000). No evidence was found that arsenic levels in ground water is associated with cancer incidence for bladder, kidney, liver, lung, NHL, and all sites combined in Idaho. Smoking is significantly attributed to the occurrence for cancer of the bladder, kidney, and lung, and all sites combined in Idaho residents. Patterns of smoking prevalence may account for the clustering in cancer incidence, especially lung cancer. Some of the information presented on cancer risk, including ethnicity, gender and obesity may also account for multiple causes of cancer in Idaho.

The role of arsenic in developing cancer has been studied. Due to the long latency of arsenic-related carcinogenesis, historically considered to be 30-50 years, ecologic studies on the relationship between arsenic exposure and cancer risk were strong and consistent only in the areas with high arsenic exposures. In southwest Taiwan, where arsenic in drinking water was as high as 600 μg/l, a dose-response effect was observed between arsenic-contaminated well water and age-adjusted mortality rates (1973-1986) for cancer of the bladder, kidney, skin, and lung in both genders (Wu, Kuo, Hwang, & Chen, 1989). Similar results were found in Cordoba, Argentina (Hopenhayn-Rich, Biggs, & Smith, 1998) and in Chile (Marshall et al., 2007; A. H. Smith, Goycolea, Haque, & Biggs, 1998). However, results from ecologic studies on cancer risk among populations exposed to lower arsenic level were controversial (Ayotte et al., 2006; Lamm et al., 2004; Lewis, Southwick, Ouellet-Hellstrom, Rench, & Calderon, 1999). In Idaho counties, average arsenic level in ground water is lower than 50 μg/l, ranging from 0.09 μg/l in Lewis County to 43.06 μg/l in Washington County. Moreover, only four counties have an average

arsenic level higher than 10 μg/l. It is difficult to identify exposure-disease associations in areas of generally low-level arsenic exposure. Similarly, Johnson and Mitchell recently evaluated the relationship between arsenic level in ground water and cancer incidence in Idaho. Adjusting for county-level cancer risk factors, no evidence links Idaho cancer incidence to arsenic exposure at the ZIP code level (Christopher J. Johnson & Mitchell, 2007).

Other US-based epidemiologic studies have been conducted in Utah (Bates, Smith, & Cantor, 1995), in the western Nevada and California (Steinmaus, Yuan, Bates, & Smith, 2003) and in New Hampshire (Karagas et al., 2004). Overall, these case-control studies have failed to demonstrate a direct relationship between arsenic exposure in drinking water and cancer risk in nonsmokers. However, increased cancer risk was observed among smokers. In Idaho, arsenic levels in ground water and smoking prevalence are both higher in northern and southwestern counties, whereas cancer incidence is higher. The interaction between smoking and arsenic genotoxicity and whether smoking may induce arsenic-related cancer should be further explored.

It has been reported that arsenic exposure is associated with increased incidence for non-leukemia childhood cancer, including bone cancer in 5-14 year-old children and lymphomas in 14-19 year-old young adults (Moore, Lu, & Smith, 2002). Arsenic carcinogenicity may vary by susceptibility in populations. We used cancer incidence of all ages combined to link cancer risk and arsenic exposure. Age-specific incidence rates would be helpful to identify arsenic-related cancer risk in vulnerable subgroups. Considering potential environmental caner risk in Idaho, the extent of radiation release from the Idaho National Laboratory (INL) and the exposure to public has been regarded. The INL is on the upper Snake River Plain that spans Butte, Bingham, Bonneville, Clark, and Jefferson Counties (Centers for Disease Control and Prevention, 2005b). Although no excess cancer mortality risk was observed in counties with nuclear facilities

(Forman et al., 1987; G. R. Howe, 1991), we have identified that higher NHL incidence clustered in Butte County and Bonneville County. Specific question regarding radiation exposure and NHL risk should be discussed.

Although exposure status and health outcomes are unknown among individuals, this ecologic-based study preliminarily evaluated suspected arsenic-related cancer risks in Idaho. Spatial analysis has confirmed that smoking contributes to cancer risks of the bladder, kidney, and lung. However, several limitations and biases can be raised. Because of the long latency of cancer and contemporary of exposure and cancer data, we have assumed that the exposure estimates and the demographic data in the 1990s and 2000s could provide indication of past exposures. Secondly, ecologic fallacy has limited the generalizability of interpretation. Averaged county-specific arsenic level could not capture a precise exposure for Idaho residents and lead results toward null. Calculating county-specific arsenic level in ground water by weighting site-specific utilizing population could provide better estimations for arsenic exposure. In addition, using residence at time of cancer diagnosis might not account for migration variation and misclassification could occur.

Cancer is a disease with multiple causes, looking only at arsenic in water may be a limited indicator of total carcinogenetic exposures. Arsenic toxicity could be altered by nutrients like selenium and zinc or water co-contaminants like antimony, which either suppress or enhance in vivo genotoxicity and carcinogenicity (Gebel, 2000). Arsenic toxicity also depends on exposure route and dose, frequency and duration, as well as on individual susceptibilities (Tchounwou, Centeno, & Patlolla, 2004), which are difficult to measure as a whole. Recently, a space-time information system was introduced which enables visualization and analysis of space-time data to assess individual lifetime exposure to inorganic arsenic (Meliker et al., 2007).

It's flexibility accounts for a number of combinations of parameters, including water and food consumption, residential and occupational histories, temporal changes in the magnitude and location of exposure, life-course for susceptibility changes across years and age. The space-time information system may be a useful tool for exposure and risk assessment compared to conventional GIS technologies.

4.6 CONCLUSION

In Idaho, geographic disparity of arsenic-contaminated ground water might not explain spatial patterns of cancer incidence. We validated that smoking has contributed critical proportions to cancer occurrence among Idaho residents. We also demonstrated the usefulness of GIS techniques on the epidemiological study. However, this study was not able to evaluate the role of occupational risks or of other environmental risk factors. Smoking only partly explains spatial patterns of cancer incidence in Idaho. Other environmental exposures should also be considered. Although arsenic levels in ground water is not shown to be related to cancer incidence in Idaho, the proportion of cancer cases attributable to arsenic exposure is not conclusive. Concerning populations exposed to relatively high arsenic-contaminated ground water in southwestern Idaho, conducting an individual-based study, adopting better tools on arsenic exposure assessment, and adjusting for smoking and other potential confounders would help to identify arsenic-related cancer risk in this population.

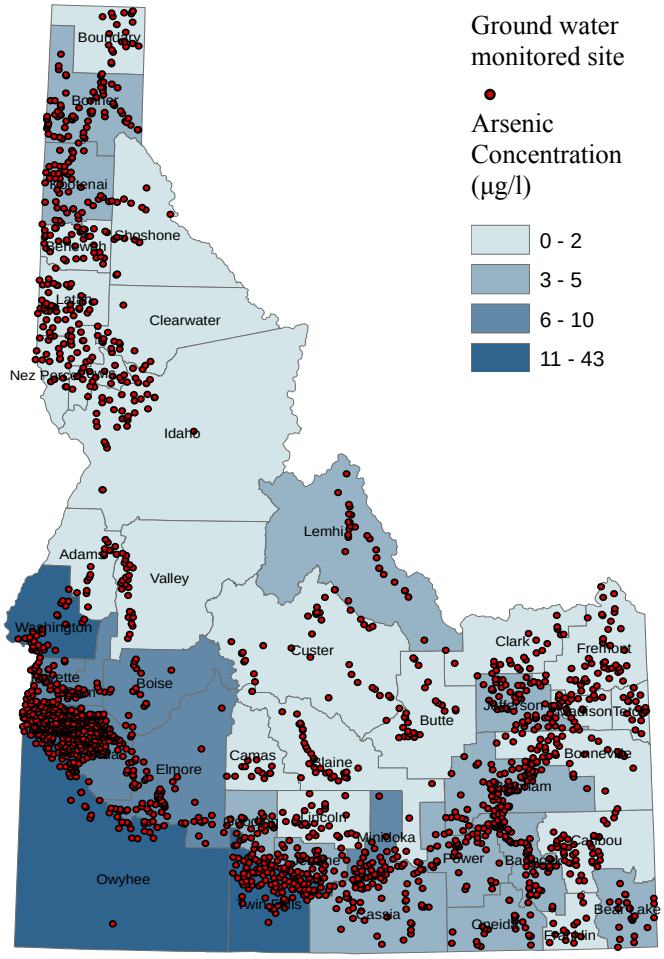


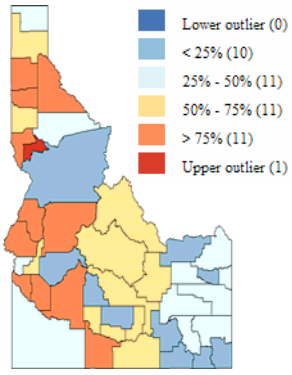
Figure 4-1 County average arsenic concentration in ground water, data was based on the Idaho Statewide Program, 1991-2004. Individual points identify locations of sampled wells

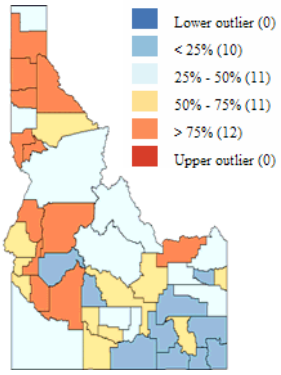
Table 4-1 Idaho and SEER 9 registries all cancer and selected cancer gender and site-specific incidence rates for person diagnosed between 1990 and 2004 (age-adjusted U.S. 2000 standard)

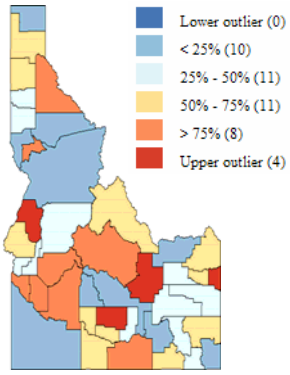
	Men							
	Idaho				SEER 9 Registries			
Cancer site	Rate[1]	SE	LCI	UCI	Rate	SE	LCI	UCI
All Malignant Cancers	534.8*	2.7	529.6	540.1	581.4	0.6	580.2	582.6
Lung and Bronchus	73.4*	1.0	71.5	75.4	86.8	0.2	86.3	87.3
Urinary Bladder	37.8	0.7	36.4	39.3	37.3	0.2	37.0	37.6
Kidney and Renal Pelvis	14.6*	0.4	13.8	15.5	16.2	0.1	16.0	16.4
Liver and bile duct	3.9*	0.2	3.5	4.4	8.0	0.1	7.8	8.1
Non-Hodgkin Lymphoma	20.6*	0.5	19.5	21.6	24.0	0.1	23.8	24.3
	Women							
	Idaho				SEER 9 Registries			
Cancer site	Rate	SE	LCI	UCI	Rate	SE	LCI	UCI
All Malignant Cancers	389.3*	2.1	385.2	393.4	417.3	0.5	416.4	418.2
Lung and Bronchus	42.8*	0.7	41.4	44.2	50.8	0.2	50.5	51.1
Urinary Bladder	8.4*	0.3	7.8	9.0	9.5	0.1	9.4	9.7
Kidney and Renal Pelvis	8.1	0.3	7.6	8.8	8.1	0.1	8.0	8.2
Liver and bile duct	1.9*	0.1	1.7	2.3	2.9	0	2.9	3.0
Non-Hodgkin Lymphoma	16.0	0.4	15.2	16.9	15.9	0.1	15.7	16.1

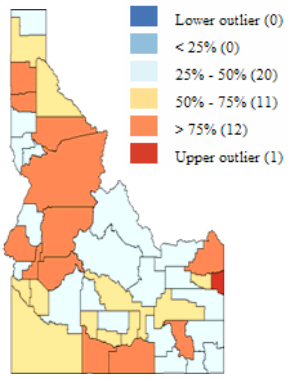
1. Rates are per 100,000 and age-adjusted to the 2000 US Standard Population (19 age groups - Census P25-1130) standard; CI: Confidence intervals (Tawari mod) are 95% for rates
* Incidence rate is significantly different from Idaho relative to SEER 9 registries

Males	Females	Males	Females
All Sites Combined		Liver and Bile Duct Cancer	

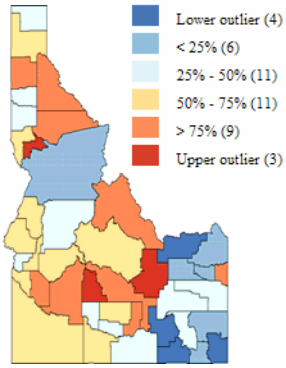


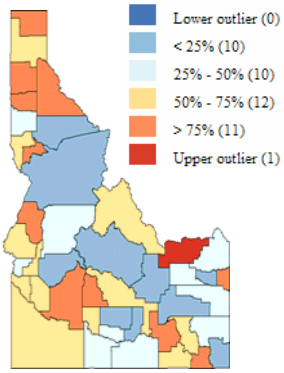


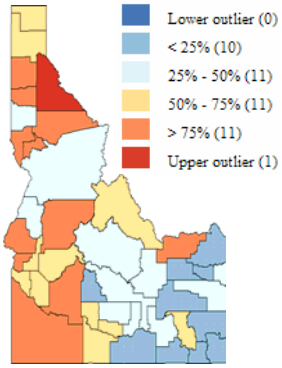


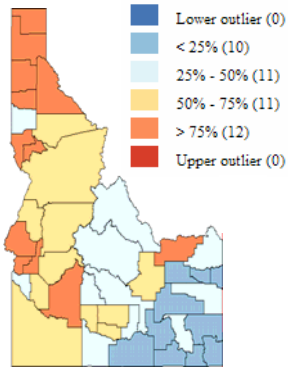


Bladder Cancer		Lung and Bronchus Cancer	

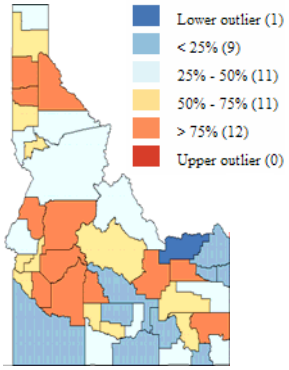


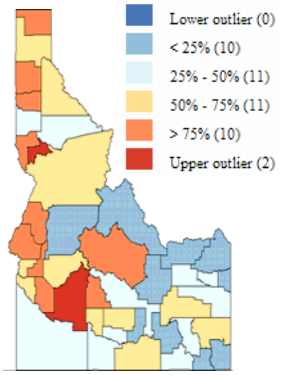


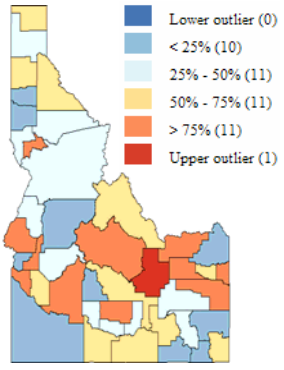




Kidney and Renal Pelvis Cancer		Non-Hodgkin's Lymphoma	







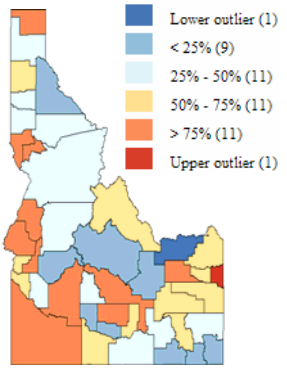


Figure 4-2 Box maps for cancer incidence in the Idaho cancer registry by gender and cancer site, for person diagnosed between 1990 and 2005, adjusted for age

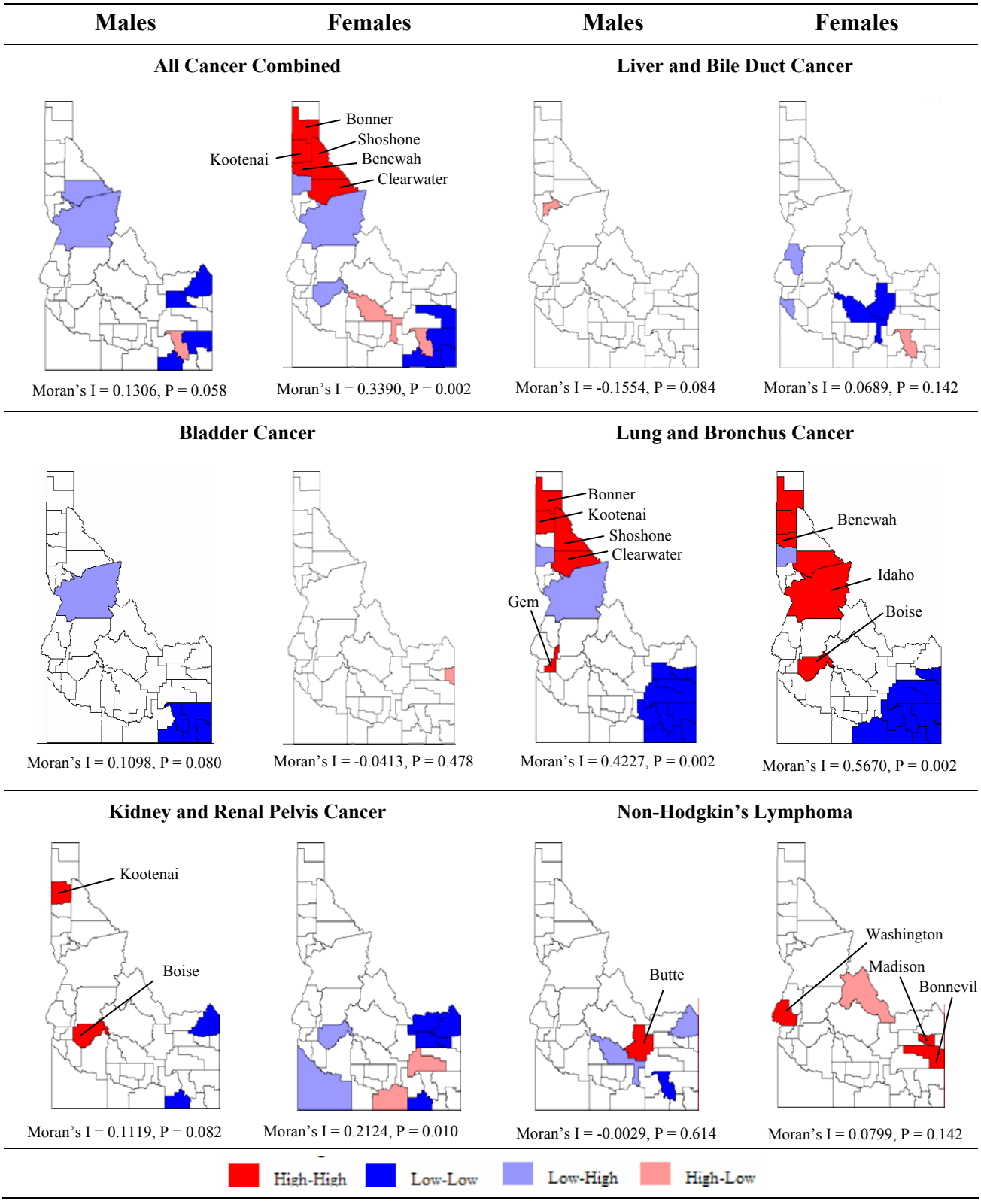


Figure 4-3 Cluster maps for cancer incidence in the Idaho cancer registry by gender and cancer site, for person diagnosed between 1990 and 2005, adjusted for age

Table 4-2 Moran's I value for cancer incidence in the Idaho cancer registry by gender and cancer site, for person diagnosed between 1990 and 2004, adjusted for age

		Males		Females	
Cancer Site	**Year**	**Moran's I**	**p-value**	**Moran's I**	**p-value**
All cancer combined	1990-1994	0.205	**0.024**	0.226	**0.008**
	1995-1999	0.035	0.274	0.239	**0.004**
	2000-2004	0.108	0.086	0.113	0.104
Bladder	1990-1994	0.015	0.308	0.026	0.266
	1995-1999	0.018	0.348	-0.070	0.258
	2000-2004	0.132	**0.042**	-0.095	0.220
Kidney & Renal Pelvis	1990-1994	0.058	0.112	-0.034	0.496
	1995-1999	-0.119	0.168	-0.011	0.622
	2000-2004	-0.088	0.250	0.220	**0.028**
Liver & Bile Duct	1990-1994	-0.137	0.090	-0.019	0.542
	1995-1999	-0.109	0.080	-0.006	0.594
	2000-2004	-0.008	0.634	0.059	0.218
Lung & Bronchus	1990-1994	0.326	**0.004**	0.512	**0.002**
	1995-1999	0.070	0.174	0.397	**0.002**
	2000-2004	0.394	**0.004**	0.312	**0.002**
NHL	1990-1994	0.132	0.078	-0.121	0.146
	1995-1999	-0.107	0.180	-0.076	0.316
	2000-2004	0.048	0.206	0.114	0.060

Table 4-3 Pearson Correlation coefficient for arsenic level in ground water and cancer incidence by gender and cancer site, for persons diagnosed from 1990 to 2005, adjusted for age

	All cases		Male		Female	
	β	p-value	β	p-value	β	p-value
All cancer sites	0.04	0.823	0.05	0.733	0.02	0.909
Bladder	-0.02	0.909	0.03	0.859	-0.01	0.963
Kidney	0.02	0.895	-0.10	0.513	0.15	0.339
Liver	0.06	0.684	-0.08	0.611	0.22	0.160
Lung	0.25	0.101	0.28	0.061	0.20	0.184
NHL	0.09	0.555	0.06	0.719	0.10	0.532

Data was analyzed by SPSS 12.1

Table 4-4 Idaho all cancer, and selected cancer site-specific incidence rates, by sex and county grouping according to arsenic ground water levels for person diagnosis between 1990 and 2004

	Low Arsenic Counties (≤ 2µg/L, N=23)			Intermediate Arsenic Counties (2-9µg/L, N=16)			High Arsenic Counties (≥ 10µg/L, N=5)		
	Rate[1]	LCI	UCI	Rate[1]	LCI	UCI	Rate[1]	LCI	UCI
Men and Women									
Average population	354,970			640,588			220,728		
All Malignant Cancers	422.7	417.0	428.5	457.8	453.3	462.3	440.7	433.5	448.0
Lung and Bronchus	50.2	48.2	52.2	57.7	56.1	59.4	58.3	55.7	61.0
Urinary Bladder	19.7	18.5	21.0	22.4	21.4	23.5	20.4	18.9	22.0
Kidney and Renal Pelvis	10.5	9.6	11.4	11.8	11.1	12.6	10.3	9.2	11.4
Liver and Intrahepatic Bile Duct	2.3	1.9	2.7	3.0	2.7	3.4	3.3	2.7	4.0
Non-Hodgkin Lymphoma	17.7	16.5	18.9	18.0	17.1	18.9	18.3	16.9	19.8
Men									
Average population	178,496			320,441			109,298		
All Malignant Cancers	500.9	491.6	510.4	537.7	530.3	545.1	529.9	518.1	541.9
Lung and Bronchus	65.0	61.7	68.4	74.0	71.2	76.8	80.0	75.5	84.8
Urinary Bladder	35.2	32.7	37.8	39.3	37.2	41.4	37.2	34.1	40.5
Kidney and Renal Pelvis	13.6	12.1	15.2	15.7	14.4	16.9	13.1	11.3	15.1
Liver and Intrahepatic Bile Duct	3.2	2.5	4.0	4.0	3.4	4.6	4.8	3.7	6.1
Non-Hodgkin Lymphoma	19.0	17.3	20.9	20.8	19.4	22.3	21.0	18.7	23.5
Women									
Average population	176,474			320,146			111,430		
All Malignant Cancers	364.9	357.6	372.4	400.0	394.3	405.8	376.8	367.7	386.2
Lung and Bronchus	38.3	36.0	40.8	45.2	43.2	47.2	41.5	38.6	44.7
Urinary Bladder	7.2	6.2	8.3	9.2	8.4	10.1	7.6	6.4	9.0
Kidney and Renal Pelvis	7.8	6.8	9.0	8.5	7.6	9.3	7.8	6.5	9.2
Liver and Intrahepatic Bile Duct	1.5	1.1	2.1	2.1	1.7	2.6	2.0	1.4	2.7
Non-Hodgkin Lymphoma	16.6	15.0	18.2	15.6	14.5	16.8	16.1	14.3	18.1

Rates are per 100,000 and age-adjusted to the 2000 US Standard Population (19 age groups - Census P25-1130) standard; Confidence intervals are 95% for rates (Tawari mod)

Table 4-5 Results from spatial regression analysis for assessing predictors of cancer incidence in Idaho among persons diagnosed from 1990 to 2005, adjusted for age

	All sites [1]		Bladder [1]		Kidney & Renal Pelvis [1]	
R-square	0.396		0.243		0.394	
Independent variable	β	p-value	β	p-value	β	p-value
Arsenic (ug/ml)	-0.227	0.858	0.005	0.970	0.014	0.848
White (%)*	1.914	0.240	0.114	0.464	0.245	**0.014**
Men/Women (%)*	1.718	0.334	0.232	0.178	0.111	0.291
Population Density/Land*	0.308	0.065	0.005	0.737	0.014	0.159
Current Smoker (%)	6.939	**0.004**	0.545	**0.017**	0.330	**0.018**
BMI<25 (%)	-0.992	0.598	0.164	0.367	-0.097	0.385
	Liver & Bile Duct [1]		Lung & Bronchus [2]		NHL [1]	
R-square	0.239		0.641		0.078	
Independent variable	β	p-value	β	p-value	β	p-value
Arsenic (ug/ml)	0.016	0.658	0.203	0.471	0.045	0.706
White (%)*	0.072	0.123	-0.462	0.186	0.013	0.933
Men/Women (%)*	0.114	**0.030**	0.155	0.685	0.165	0.319
Population Density/Land*	0.007	0.141	0.010	0.784	0.012	0.427
Current Smoker (%)	-0.021	0.748	0.226	**<0.001**	-0.217	0.309
BMI<25 (%)	-0.122	**0.028**	0.067	0.871	-0.260	0.142

1. Ordinary least squares estimation
2. Spatial lag model- Maximum likelihood estimation

Quartile Map	Box Map	LISA cluster map

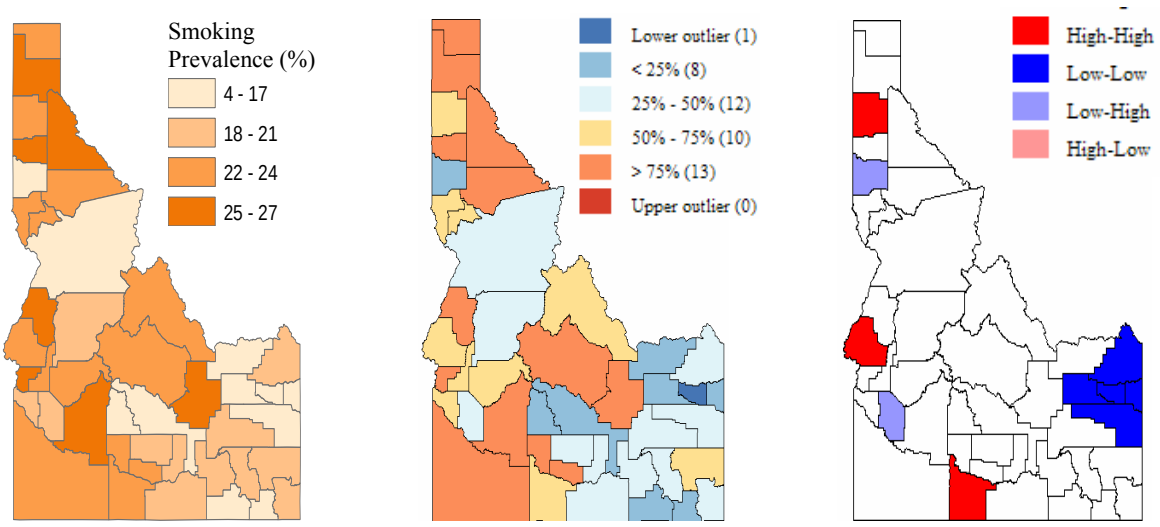


Figure 4-4 Quartile map, box map, and LISA cluster map for Idaho county smoking prevalence, Behavior Risk Factor Surveillance System (BRFSS), 1997-2005

.

5.0 DISCUSSION AND CONCLUSION

William Farr, one of the progenitors of epidemiology, advocated that medical statistics should be carefully studied, because these provide important clues about underlying causes of population health. His death in 1883 was lauded in the Lancet for his pioneering work laying the statistical tools that are the foundation of public health knowledge. This dissertation has avoided the florid writing of Farr, but is consistent with his notion that the trend is all we can examine: all else is conjecture. This three-part dissertation evaluates patterns of cancer at the national and state level, in an effort to clarify the current burden of cancer and to promote the development of etiologic hypotheses about avoidable cancer causes. The first part evaluates national patterns of cancer for various groupings of the disease for the past quarter century. The second part conducts an assessment of age-specific patterns of NHL in the State of Pennsylvania, compares these with the U.S. overall, and speculates about some possible underlying causes; and the third part carries out an exploratory Geographic Information System study employing cancer incidence data aggregated at the county level and arsenic levels reported in groundwater in the State of Idaho.

Cancer is a disease of multiple causes and stages. As a result the temporal and spatial analysis of national and state patterns of cancer cannot pinpoint specific causes. But, such analysis can form the backbone upon which to base future research and, under some circumstances where demographic, environmental and behavioral information are available, can also clarify the impact of cancer control policies.

Many studies have been conducted to identify cancer pathology, to understand cancer occurrence, and to propose and evaluate better cancer detection and treatment. The analysis of temporal and spatial patterns of cancer remains a complex and challenging enterprise that has recently been neglected. The purposes of the three related but independent reports presented in this dissertation are to increase awareness of available medical, environmental and demographic data in population-based state and national cancer registries, to provide innovative examples of modeling cancer data analysis that takes advantage of new developments in statistical thinking and Geographic Information Systems approaches, and to stimulate the need for additional research regarding environmental and other avoidable cancer risks.

According to our analysis, in the U.S., for the past quarter century, tobacco-related cancer incidence declined among men and increased among women. During this same time period, screen-detectable cancer incidence (breast, prostate, colon-rectal and cervix) increased, more rapidly among men than women. For the category of cancer that we created that was unrelated to tobacco or screening, incidence increased in four gender-race groups. Throughout this time period, the growth in, NHL--a cancer classified as unrelated to tobacco or screening-- has been strong and consistent in all groups nationwide. In the State of Pennsylvania, NHL incidence increases paralleled those of the nation in every gender-race group. The diffuse lymphoma histologic subgroup appeared to be the most important contributor to the NHL increase. NHL incidence was higher in Pennsylvania counties with a greater percentage of urban residents. In Idaho, using state of the art tools of Geographic Information System analyses, we found that county average arsenic levels in ground water are highest in southwestern counties. Incidence for lung cancer and all cancers combined was spatially clustered in men and women, with higher incidence clustered in northern Idaho and lower rates clustered in southeastern Idaho. Smoking

prevalence contributes to higher incidence for cancer of the bladder, kidney, lung, and all sites combined. No association was found between arsenic levels in ground water and cancer incidence in Idaho. We were not able, within the constraints of this analysis, to examine other potential causes of these geo-spatial variations in cancer in Idaho.

The cancer trends noted here indicate major shifts in the types of cancer in the U.S. within the past two decades. For men, smoking-related cancers have declined, while screening-detected and cancers not related to smoking or screening has increased. For women, smoking-related cancers generally increased, although the increase may be attenuating, and screening-detected and cancers not related smoking or screening have also increased. Improved diagnosis and access to health care certainly play a role in the reported increases; however, they should not be assumed to account completely for these trends. Evaluating time trends by gender may avoid diagnostic artifacts since it is not likely that diagnosis would substantially vary with gender over the years. The appropriate degree of aggregating or disaggregating cancer data has been discussed (Bailar, 1990). The categories used will reflect the purposes of the analysis. Studying cancer mortality by combing cancer sites is useful for assessing overall progress on cancer prevention and treatment. Many cancer cases are preventable. Early diagnosis and improved treatment for breast and colon-rectal cancer (and childhood cancer, not considered here) have been shown to reduce deaths.

The evaluation of time trends of specific types of cancer can be useful, as these can suggest hypotheses for future research regarding underlying causes. We have confirmed the nationwide decline in cancer deaths in recent decades, but also indicated some important trends in various groupings of cancer that merit serious examination. In addition, it should be noted that the overall decrease is not shared equally across subgroups within the population. As had

been widely reported, African Americans tend to fare more poorly at every stage of cancer, even adjusted for access to care. Whether these differences may be related to their differential workplace and residential exposures to degraded environments are a matter that should be explored systematically in future work. Ethnic disparity, health accessibility, or effectiveness of public health practice should be carefully inspected to determine their contribution to these disparities and propose policy interventions to ameliorate them. Determining why incidence for cancers not linked to smoking or screening is increasing, while all sites combined cancer incidence is decreasing, presents an important and challenging question for cancer researchers and public policy makers.

At this point, we cannot determine the proportional contribution to increased risk of cancer that stems from changes in lifestyle, medical practice, environmental carcinogens (i.e. contaminated drinking water or polluted air), occupational hazards (i.e. organic solvents, asbestos, or pesticides), and medical or background radiation (i.e. radon). These exposures, which concentrate in certain subgroups of the population, are involuntary but, to a large extent, avoidable, and often are completely beyond an individual's control or knowledge. As a cohort, the generation born in the middle of the twentieth century experienced several factors that could have increased their cancer risk compared to those born at the turn of that century, including newly introduced pesticides and other industrial materials and poor control of occupational carcinogens (Davis, Hoel, Fox, & Lopez, 1990) Estimating a relative cancer rate in ever-smokers versus never-smokers for different time points of interest would be more helpful in demonstrating a role for other environmental cancer causes not attributable to smoking.

Cancer registries provide data and serve as resources for conducting epidemiologic studies within large populations. It should be understood that there are clear limits to this work,

because the data on which it is based do not reflect the heterogeneity of the American population. Although routinely collected statistics on cancer occurrence are commonly reported among racial and ethnic populations, the SEER population is predominantly white, and the black population is primarily urban. The white population is represented well in each of the SEER areas but this is not true for blacks. Approximately 77% of the black population in SEER is represented by the metropolitan areas of Atlanta (34%), Detroit (33%), and San Francisco (13%) (National Cancer Institue (SEER)). The system does not include any estimates of cancer in black, rural-poor populations, especially in the southern part of the country, nor does it include urban African Americans in parts of the south, excepting Atlanta. Estimates of the number of cancer cases in black, rural-poor populations based on SEER incidence rates is not reflective of the cancer experience in these populations. As a result, interpretation of racial differences based on this study cannot be generalized to all populations in the United States.

In addition, quality of cancer case ascertainment has been a concern, although the U.S. maintains one of the highest quality cancer registries in the world. Changes of ICD-O coding regulation and changes in detection practices may affect records of diagnostic stage and histologic subtypes of cancer. However, the SEER program sets the standard for data quality worldwide. Quality control has been an integral part of the SEER Program activities since its inception in 1973 (National Cancer Institute, 2005c), and histopathologic confirmation rate has always been above 90% and close to 100% in recent years. Case ascertainment from PCR follows the definitions and manual used by the National Cancer Institute's SEER program (Pennsylvania Department of Health, 2006b). Although concerns of variation associated with the small number of events can occur when using county statistics, the data are comparable between PCR and SEER.

Meanwhile, data from population-based cancer registries, supplemented by linkage with administrative databases, are important and accessible resources for monitoring potential causes of cancer, or say, environmental epidemiologic/ecologic study. Taking Project #3 as an example, by linking cancer incidence from CDRI with arsenic levels in ground water monitored by Idaho Department of Environmental Quality, we evaluated the relationship between arsenic exposure in ground water and cancer risk. Although we cannot identify personal arsenic-related cancer risk, nor prove causal relationship, the study emphasizes the importance of smoking factors and other environmental exposures. Moreover, cancer is a disease with long latencies. We are aware that exposure estimates and demographic data that we obtained for Idaho are nearly contemporaneous with cancer data and that does not allow for the consideration of latency. Therefore, we assumed that the exposure estimates can provide some indication of past exposures that may have occurred decades earlier. Existing data has limited our ability to prove this assumption. Following the trail of environmental monitoring programs in the coming decades, the cohort data could provide refined causative relationships regarding environmental exposures and cancer outcomes that account for the long latency of cancer.

Efforts to project future cancer patterns and evaluate past ones are of vital importance for three major reasons. First of all, these projections will be critical to providing fair policies to provide care for the fastest growing segment of many industrial populations that is age sixty and older in whom most cancer can be expected to arise. Second, these analyses can highlight the often unappreciated value of national and state registries of cancer statistics and stimulate the development of novel statistical models to explore these registries more fully. Finally, this work, also lays the foundation upon which additional research can be based to propose, identify and evaluate underlying causes of the patterns and trends we have found. This study utilized existing

data to design a complex series of cancer trend modeling activities that provides an important addition to the literature and to public health inquiry. As resources to conduct epidemiologic studies remain limited, the importance of devising inventive approaches to existing data systems is clear. This dissertation affirms the value and importance of conducting analyses of descriptive data and sets the stage for additional research to be done in the U.S. and internationally on the patterns we have described here. As a concluding point, I anticipate this work will help to generate hypotheses on cancer etiology with regard to environmental and other avoidable causes of cancer; as well as assist public health officials to effectively plan services, prioritize health resource allocations, and develop and measure prevention and intervention strategies.

APPENDIX A

STATISTICAL APPENDIX FOR PROJECT#1

Consider the following age-period-cohort (APC) model with N_{age} age-specific rates in each of N_{per} time periods,

$$\ln(R_{ijk}) = \alpha + A_i + P_j + C_k + \varepsilon_{ijk}, \quad (1)$$

where R_{ijk} is the rate (cancer incidence or mortality) in the *i*th age group (indexed between 1 and N_{age}) during the *j*th time period (indexed between 1 and N_{per}) for the *k*th birth cohort (indexed between 1 and N_{coh}). The ε_{ijk} are error terms. The model includes an intercept (α) and separate terms for age group (A_i), time period (P_j), and birth cohort (C_k). Note, $k = N_{age} - i + j$ and $N_{coh} = N_{age} - 1 + N_{per}$, where N_{coh} is the number of distinct birth cohorts specified by the model.

For example, separately for each combination of gender, race, and cancer category, we fit an APC model to 91 age-specific cancer mortality rates (13 age-specific rates in each of 7 five-year time periods). Not including the intercept, a typical APC model would be fit against these 91 age-specific rates through use of 39 parameters, 13, 7, and 19 parameters for age, time period, and birth cohort, respectively. However, because of the interdependency of age, calendar time, and birth cohort (as given by $k = N_{age} - i + j$), not all parameters can be uniquely estimated.

To handle this problem, Dinse *et al*. (Dinse, Umbach, Sasco, Hoel, & Davis, 1999) parameterized the APC model, as follows,

$$A_i = \left[i - \frac{N_{age} + 1}{2} \right] \beta_{alin} + \sum_{m=1}^{N_{age}-1} H_{im}^{age} \beta_m^{age} , \quad (2)$$

$$P_j = \left[j - \frac{N_{per} + 1}{2} \right] \beta_{plin} + \sum_{m=1}^{N_{per}-1} H_{jm}^{per} \beta_m^{per} , \text{ and} \quad (3)$$

$$C_k = \left[k - \frac{N_{coh} + 1}{2} \right] \beta_{clin} + \sum_{m=1}^{N_{coh}-1} H_{km}^{coh} \beta_m^{coh} , \quad (4)$$

where the β_{alin}, β_{plin}, β_{clin}, $\boldsymbol{\beta}^{age}_{N_{age}-1}$, $\boldsymbol{\beta}^{per}_{N_{per}-1}$, and $\boldsymbol{\beta}^{coh}_{N_{coh}-1}$ are model parameters. $\mathbf{H}^{age}_{N_{age},N_{age}-1}$, $\mathbf{H}^{per}_{N_{per},N_{per}-1}$, and $\mathbf{H}^{coh}_{N_{coh},N_{coh}-1}$ are orthogonal Holford matrices (Holford, 1983). A bolded notation signifies a matrix or vector, with dimensions indicated as subscripts. Equations 2-4 effectively decompose the APC model (equation 1) into two components. The first component uses three parameters to represent linear effects due to age group (β_{alin}), time period (β_{plin}), and birth cohort (β_{clin}). The second component specifies age group-specific ($\sum_{m=1}^{N_{age}-1} H_{im}^{age} \beta_m^{age} = D_i^{age}$), time period-specific ($\sum_{m=1}^{N_{per}-1} H_{jm}^{per} \beta_m^{per} = D_j^{per}$), and birth cohort-specific ($\sum_{m=1}^{N_{coh}-1} H_{km}^{coh} \beta_m^{coh} = D_k^{coh}$) deviations added to the linear effects due to age group, time period, and birth cohort, respectively.

Although the three parameters representing the linear effects due to age group, time period, and birth cohort are not uniquely estimable, it has been shown that certain functions of any two of these parameters are uniquely estimable (Holford, 1983). We chose to define a uniquely estimable drift parameter D^{drift} as the sum of β_{plin} and β_{clin}, the linear effects due to time period and birth cohort, respectively. The drift parameter can be conceived as the linear effect on the natural logarithm of any age-specific rate (cancer incidence or mortality) associated with one time period increment, in the absence of any linear effects due to birth cohort. Or, equivalently, the drift parameter can be conceived as the linear effect associated with one birth cohort

increment, in the absence of any linear effects due to time period. These two examples represent the two extreme situations bounding an infinite number of possibilities. In general, the AAPC represents the age-independent relative change in cancer risk associated with a temporal change of 1 year in duration.

With respect to deviations added to the linear effects, equations 2-4 specify $N_{age}-1$, $N_{per}-1$, and $N_{coh}-1$ uniquely estimable parameters relevant to age group, time period, and birth cohort, respectively. Our APC models of age-specific cancer mortality, for example, defined 12, 6, and 18 deviations relevant to age group, time period, and birth cohort, respectively. Unfortunately, the individual parameters, $\boldsymbol{\beta}^{age}_{\mathrm{N}age-1}$, $\boldsymbol{\beta}^{per}_{\mathrm{N}per-1}$, and $\boldsymbol{\beta}^{coh}_{\mathrm{N}coh-1}$, and the deviations, $\boldsymbol{D}^{age}_{\mathrm{N}age-1}$, $\boldsymbol{D}^{per}_{\mathrm{N}per-1}$, and $\boldsymbol{D}^{coh}_{\mathrm{N}coh-1}$, calculated with these parameters, lack direct or meaningful interpretation. However, the predicted risks at two time periods j and j', $j \neq j'$, for two age groups, i and i', belonging to the same birth cohort, are related, as follows, where $i' = i + (j' - j)$.

$$\ln\left(\frac{\mathrm{R}_{i'j'k}}{\mathrm{R}_{ijk}}\right) - (\mathrm{A}_{i'} - \mathrm{A}_i) = \mathrm{P}_{j'} - \mathrm{P}_j = (j' - j)\beta_{plin} + \left(D^{per}_{j'} - D^{per}_{j}\right) \tag{5}$$

$$\frac{\mathrm{R}_{i'j'k} \big/ \mathrm{R}_{ijk}}{e^{\mathrm{A}i' - \mathrm{A}i}} = e^{(j'-j)\beta_{plin}} e^{D^{per}_{j'} - D^{per}_{j}} \tag{6}$$

Equation 6 enables a direct interpretation of $e^{D^{per}_{j'} - D^{per}_{j}}$, the exponentiated difference between any two time period parameter deviations. The value $e^{D^{per}_{j'} - D^{per}_{j}}$ can be interpreted as the cohort-independent (k fixed) risk at time period j' relative to time period j, adjusted for both age ($e^{\mathrm{A}i' - \mathrm{A}i}$) and the relative risk increase ($e^{(j'-j)\beta_{plin}}$) expected from the linear effect due to time period. The value $e^{D^{per}_{j'} - D^{per}_{j}}$ can be uniquely estimated and interpreted, even though $e^{\beta_{plin}}$ is unknowable.

For presentation purposes, we chose to plot the time period deviations D_j^{per} as a function of time period j. The time period parameter deviation change from time period j to time period j+1 can be visualized as the slope of the plot between times j and j+1. The exponentiated value of this slope can be interpreted as a birth cohort-independent time period-specific factor multiplied against $e^{\beta_{plin}}$ (the unknowable and constant linear effect from each time period increment) to estimate age-adjusted risk at time j+1 relative to time j. Regions of positive and negative slope can be viewed to cover sequential time periods exerting positive and negative pressure, respectively, on disease risk. Here, positive and negative pressure refers to generally undesired and desired trends, respectively, that serve to increase and decrease disease risks. Therefore, from a public health perspective, we desire time period deviations showing progressive decrease over the most recent time periods.

The same formulation described above can be used to guide the interpretation of plots of birth cohort parameter deviations *vs*. birth cohort. The exponentiated value of the birth cohort parameter deviation change from birth cohort k to birth cohort k+1 can be interpreted as a time period-independent birth cohort-specific factor multiplied against $e^{\beta_{clin}}$ (the unknowable and constant linear effect from each birth cohort increment) to estimate age-adjusted risk in birth cohort k+1 relative to birth cohort k. Regions of positive and negative slope can be viewed to cover sequential birth cohorts experiencing positive and negative pressure, respectively, on disease risk. From a public health perspective, we desire birth cohort deviations showing progressive decrease over the most recent birth cohorts.

APPENDIX B

SUPPLEMENTAL TABLES AND FIGURES

Table B-1 Results from cancer category-, gender-, and race-specific APC models of 9-registry SEER 1979-2003 age-specific incidence (thirteen 5-year age groups between 20-24 and 80-84 years), not adjusted and adjusted for reporting delay

			Annual average percent change (AAPC)		
Cancer category	Gender	Race	Not adjusted	Delay adjusted	Difference
Tobacco-related	Men	White	-0.98	-0.92	0.06
		Black	-1.53	-1.47	0.06
	Women	White	0.76	0.83	0.07
		Black	0.11	0.18	0.07
Screen-detectable	Men	White	2.44	2.51	0.07
		Black[1]	2.95	3.02	0.07
	Women	White	0.30	0.37	0.07
		Black	0.04	0.10	0.06
Unrelated to tobacco or screening	Men	White	0.88	0.99	0.11
		Black	0.59	0.72	0.13
	Women	White	0.80	0.92	0.12
		Black	0.66	0.79	0.13
NHL	Men	White	2.16	2.33	0.17
		Black	2.99	3.17	0.18
	Women	White	1.74	1.92	0.18
		Black	3.10	3.29	0.19

1. Analysis restricted to twelve 5-year age groups between 25-29 and 80-84 years

Table B-2 Summary of changes in cancer and mortality, 1950-2001 and 5-year relative survival rates, 1950-2001: men and women, by primary cancer site

Primary Site	All Races		Whites					
	Estimated Cancer Cases in 2001[a]	Actual Cancer Deaths in 2001[b]	Percent Change 1950 - 2001[c]				5-Year Relative Survival Rates (Percent)[e]	
			Incidence[d]		U.S. Mortality[b]			
			Total	APC	Total	APC	1950-54	1995-00
Oral Cavity and Pharynx	30,100	7,701	-33	-0.5	-48.9	-1.2	46	60.9
Esophagus	13,200	12,529	21.2	0.7	27.2	0.6	4	15.8
Stomach	21,700	12,319	-75.8	2.2	-84	-3.5	12	21.5
Colon and Rectum	135,400	56,887	17.5	0.1	-40.9	-1	37	64.1
Colon	98,200	17,860	38.2	0.4	-28.6	-0.6	41	63.1
Rectum	37,200	9,027	-16.7	-0.5	-68.7	-2.6	40	64.6
Liver and Intrahep	16,200	16,952	234.3	2.5	27.4	0.6	1	8
Pancreas	29,200	29,802	35.9	0.3	21.2	0.1	1	4.2
Larynx	10,000	3,797	17.8	0	-25.8	-0.5	52	67.4
Lung and Bronchus	169,500	156,380	290.1	2.2	265.4	2.2	6	15.4
Males	90,700	90,660	203.5	1.2	197.5	1.6	5	13.7
Females	78,800	65,720	685	4.2	608.1	4.3	9	17.4
Melanoma of the Skin	51,400	7,542	690.2	4.3	165.6	1.7	49	90.7
Breast (Females)	192,200	41,394	90	1.5	-22.2	-0.3	60	88.9
Cervix uteri	12,900	4,092	77.7	-2.3	-79.5	-3.5	59	74
Corpus and Uterus, NOS	38,300	6,783	14.9	-0.2	-68	-2	72	86.1
Ovary	23,400	14,800	2.6	0.1	3.8	-0.2	30	44
Prostate	198,100	30,719	286.2	3.3	8.8	0.3	43	100
Testis	7,200	335	143.4	2.1	73.3	3	57	96.2
Urinary Bladder	54,300	12,538	97.1	1.2	31.6	-0.9	53	82.7
Kidney and Renal Pelvis	30,800	12,372	181.7	2.1	42.8	0.7	34	63.9
Brain and Other Nervous	17,200	12,609	135.6	1.3	54.9	0.8	21	32.2
Thyroid	19,500	1,323	257.6	2.2	-44.3	-1.5	80	96.8
Hodgkin Lymphoma	7,400	1,323	13.2	0.1	-75.3	-3.3	30	85.9
Non-Hodgkin Lymphoma	56,200	22,123	248.9	3	140.9	1.7	33	60.3
Myeloma	14,400	10,795	272.6	1.9	252	1.9	6	31.9
Leukemia	31,500	21,451	33	0.4	7.9	0.2	10	47.8
Childhood 90-14) years	8,600	1,494	67.1	0.9	-70	-2.8	20	80.1
All States	1,268,000	553,760	85.9	1.5	1.5	0.1	35	65.5

The APC is the Annual Percent Change over the time interval. Rates used in the calculation of the APC are age adjusted to the 1970 U.S. Standard population.

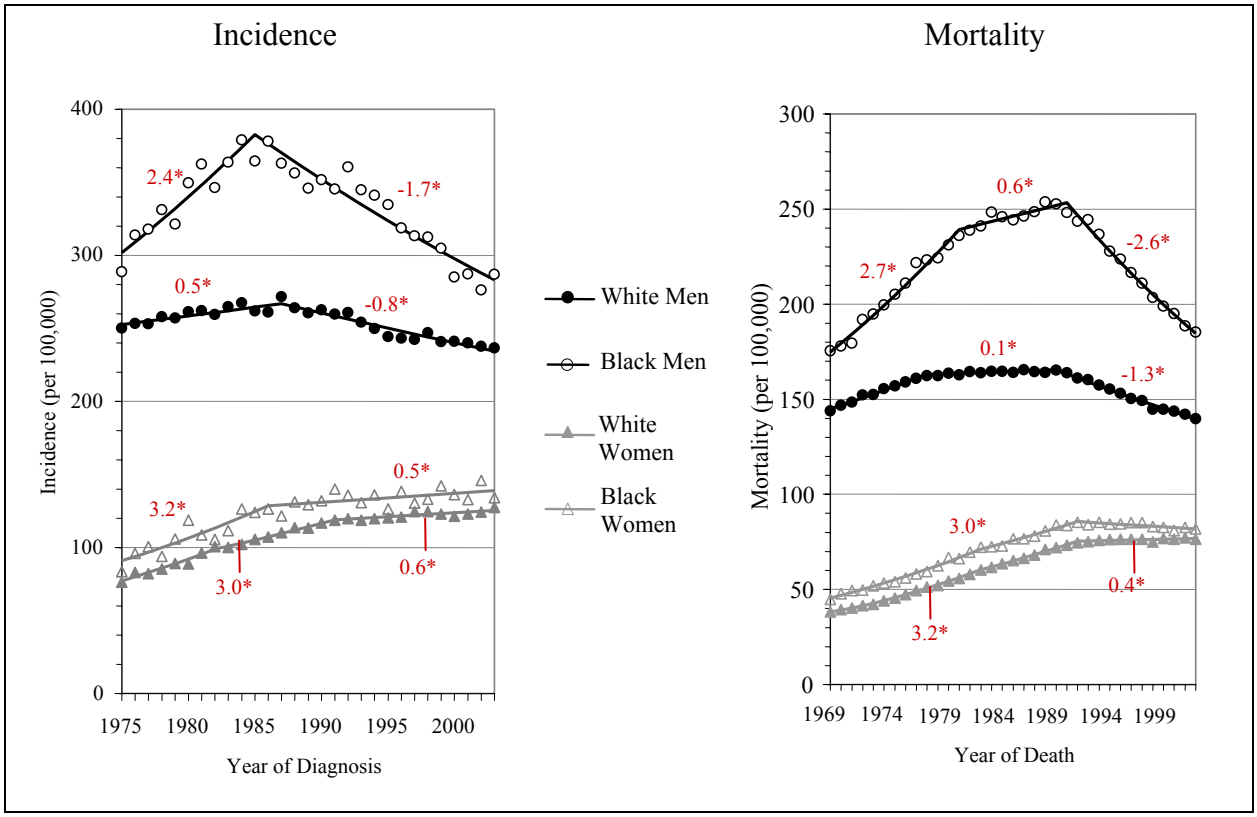
a. Facts and Figures, 2001. American Cancer Society. Atlanta, Georgia, 2001

b. NCHS public use data file for the total US. Due to coding changes throughout the years: Colon excludes other digestive tract; Rectum includes anal canal; Liver & Intrahep includes gallbladder & biliary tract., NOS: Lung & Bronchus includes trachea& pleura; Ovary includes fallopian tube; Urinary bladder includes other urinary organs; Kidney & Renal pelvis includes ureter: NHL and myeloma each include a small number of leukemias; NHL includes a small number of ill-detinmd sites.

c. All Sites, liver & Intrahop, Brain & Other nervous and Childhood cancers are for all races as opposed to whites.

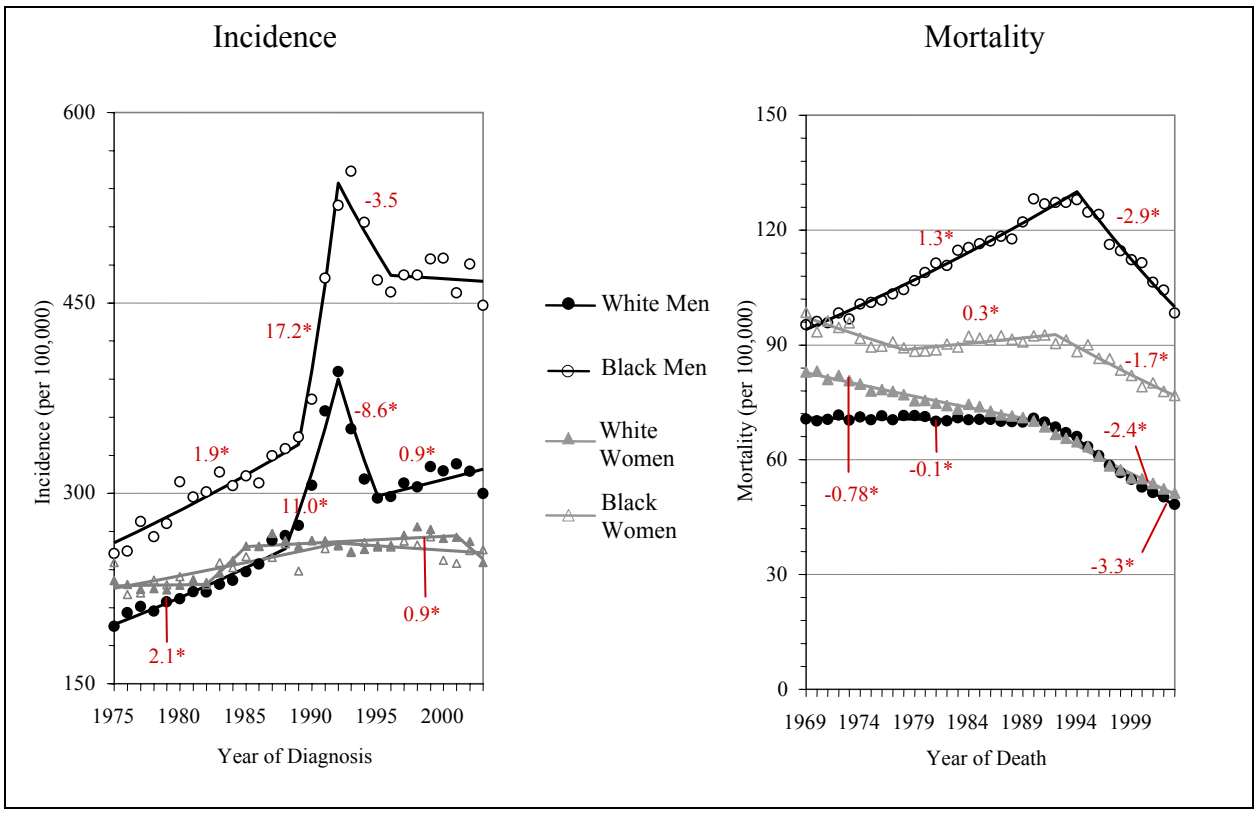
d. Data prior to 1973 are from Devesa, Silverman, Young, et al. Cancer Incidence and Mortality Trends Among Whites in the United States, 1947-48. JNCI 1987; 79:701-770 with the exception of All Sites, Liver& Intrahep, Brain & Other nervous and childhood cancers which come from historical Connecticut data. Data for 1973-2001 are from the same area used in Devena or the Connecticut registry of the SEER Program.

e. Rates for 1950-54 are from NCI Survival Report 5 with the exception of All Sites, Oral cavity & Pharynx, Colon & Rectum. Non-Hodgin lymphomas and Childhood cancers which come from historical Connecticut data. Rates for 1995-2000 are from the SEER Program with the exception of the sites just listed which come from the Connecticut registry of the SEER Program.



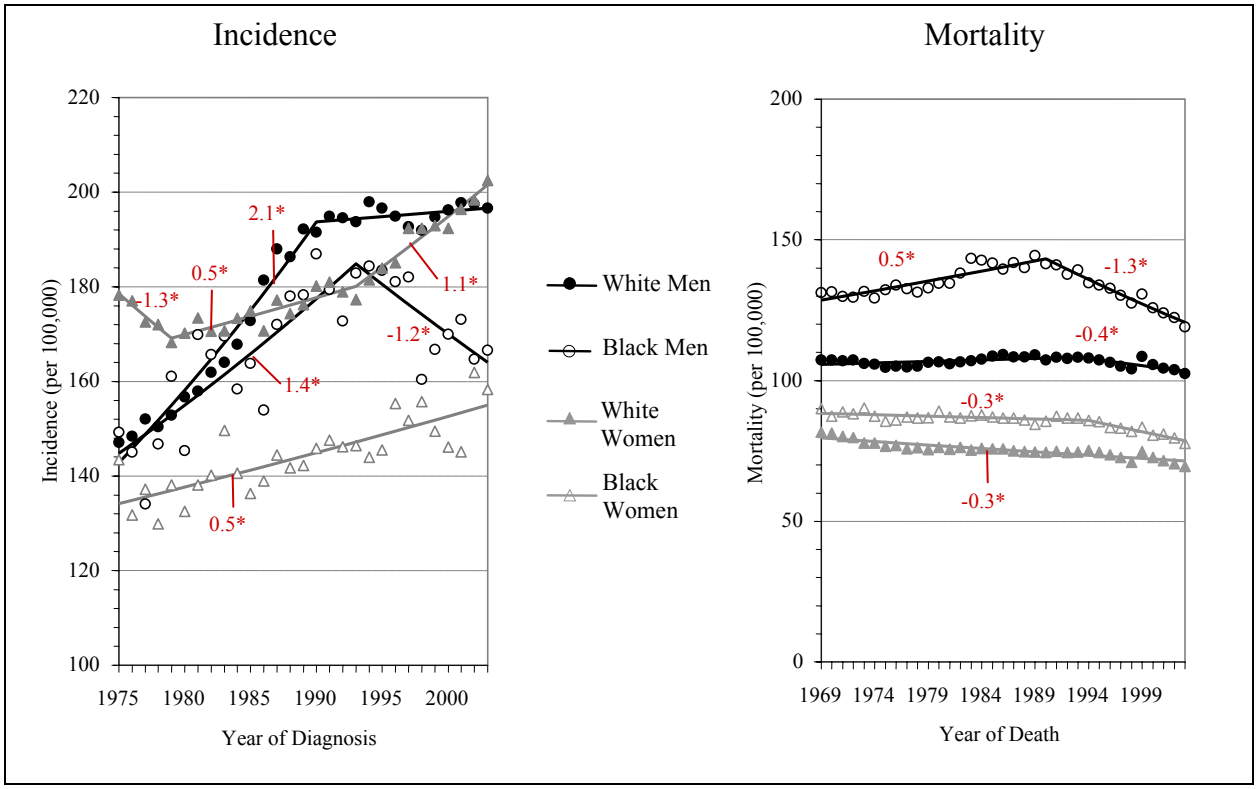
Rates age-adjusted to the 2000 US standard population
The number next to each line segment is the Estimated Annual Percent Change (EAPC, %); * indicates p<0.05

Figure B-1 Tobacco-related cancer incidence and mortality in the SEER 9-registries, by gender, race, and year of diagnosis or death (ages 20 to 84), adjusted for age and reporting delay. Lines show fit from Joinpoint regression



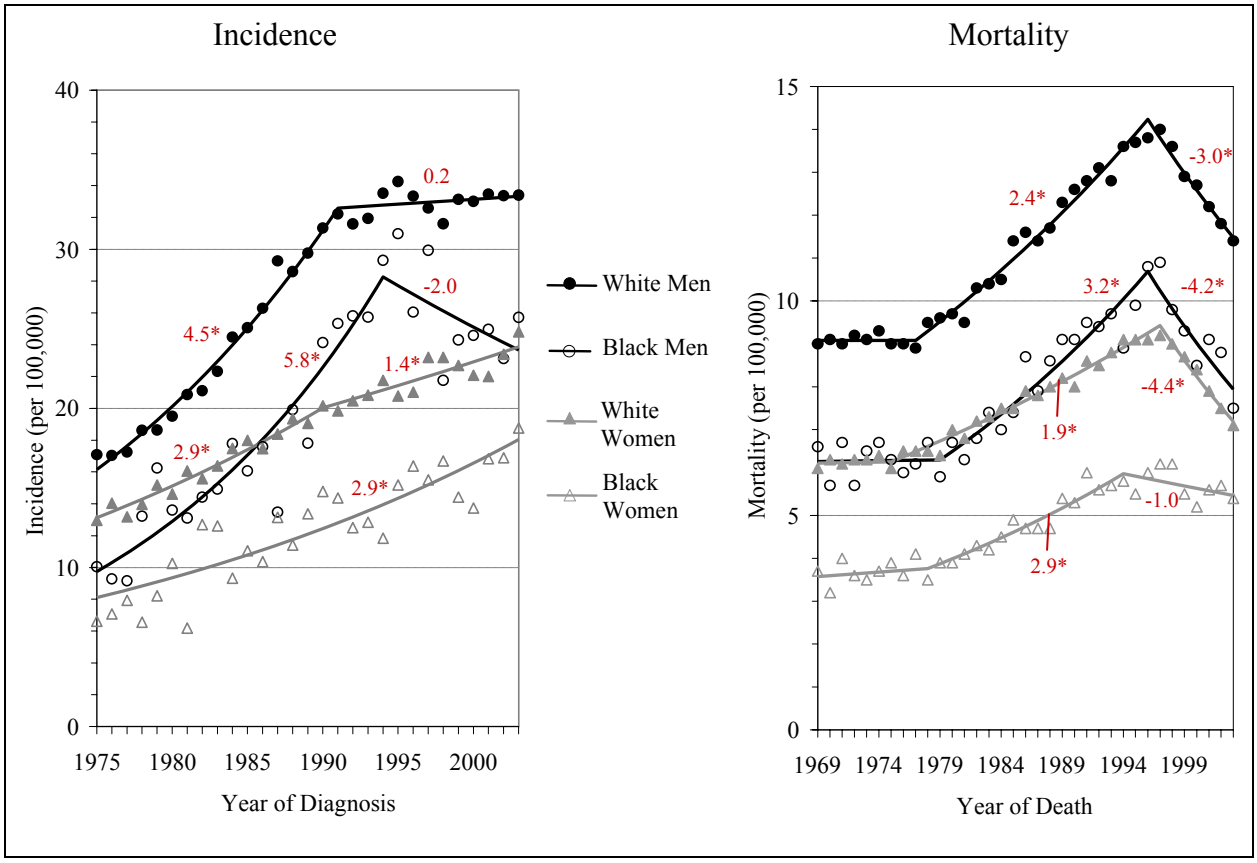
Rates age adjusted to the 2000 US standard population
The number next to each line segment is the Estimated Annual Percent Change (EAPC, %); * indicates p<0.05
Three joinpoints were observed for screen-detectable cancer incidence in white women: EAPC1=0.08 (1975-1982), EAPC2=4.18 (1982-1985), EAPC3=0.20 (1985-2001), and EAPC4=-3.55 (2001-2003)

Figure B-2 Screen-detectable cancer incidence and mortality in SEER 9-registries, by gender, race, and year of diagnosis or death (ages 20 to 84), adjusted for age and reporting delay. Lines show fit from Joinpoint regression



Rates age adjusted to the 2000 US standard population
The number next to each line segment is the Estimated Annual Percent Change (EAPC, %); * indicates p<0.05

Figure B-3 Incidence and mortality for cancer unrelated to tobacco or screening in SEER 9-registries, by gender, race, and year of diagnosis or death (ages 20 to 84), adjusted for age and reporting delay. Lines show fit from Joinpoint regression



Cancer rate was age adjusted to the 2000 US standard population
The number next to each line segment is the Estimated Annual Percent Change (EAPC, %); * indicates p<0.05

Figure B-4 Incidence and mortality for NHL in SEER 9 registries, by gender, race, and year of diagnosis or death (ages 20 to 84), adjusted for age and reporting-delay. Lines show fit from Joinpoint regression

Table B-3 Estimates of average NHL incidence in the Pennsylvania and SEER registry by gender, race, and histologic subtype of the Working Formulation (WF) classification for persons diagnosed from 1985 to 2004 at ages 20 to 84, age-adjusted to the 2000 U.S. standard population

	White Men		Black Men		White Women		Black Women	
Histology	PA	SEER	PA	SEER	PA	SEER	PA	SEER
Small Lymphocytic								
A: Small Lymphocytic	2.8	2.9	2.1	2.4	1.8	1.9	1.5	1.4
Follicular	5.4	5.3	1.9	2.4	4.8	4.8	1.8	1.6
B: Follicular small	2.0	2.1	0.6	0.9	1.7	1.9	0.5	0.6
C: Follicular mixed	1.4	1.5	0.4	0.6	1.4	1.4	0.6	0.5
D: Follicular large	0.9	0.8	0.4	0.5	0.8	0.8	0.3	0.3
(Follicular, NOS)	1.0	0.8	0.5	0.4	0.9	0.7	0.4	0.3
Diffuse	12.1	12.6	9.2	9.0	8.7	8.7	5.4	6.0
E: Diffuse small	0.8	1.0	0.5	0.5	0.3	0.4	0.1	0.3
F: Diffuse mixed	1.4	0.9	0.9	0.7	1.2	0.8	0.7	0.6
G: Diffuse large	9.9	10.8	7.8	7.9	7.2	7.5	4.5	5.1
High-Grade	1.1	2.2	1.4	1.7	0.7	0.9	0.4	0.7
H: Immunoblastic	0.7	1.6	0.8	1.2	0.5	0.7	0.2	0.5
I: Lymphoblastic	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1
J: Small noncleaved	0.2	0.5	0.4	0.4	0.1	0.1	0.1	0.1
Peripheral T-cell	1.4	2.1	2.2	2.6	0.8	1.1	1.4	1.7
NOS	5.1	6.0	5.3	5.8	3.5	3.5	2.9	3.0

Rates are per 100,000 and age-adjusted to the 2000 US Std Million (19 age groups) standard.
NOS=otherwise specified

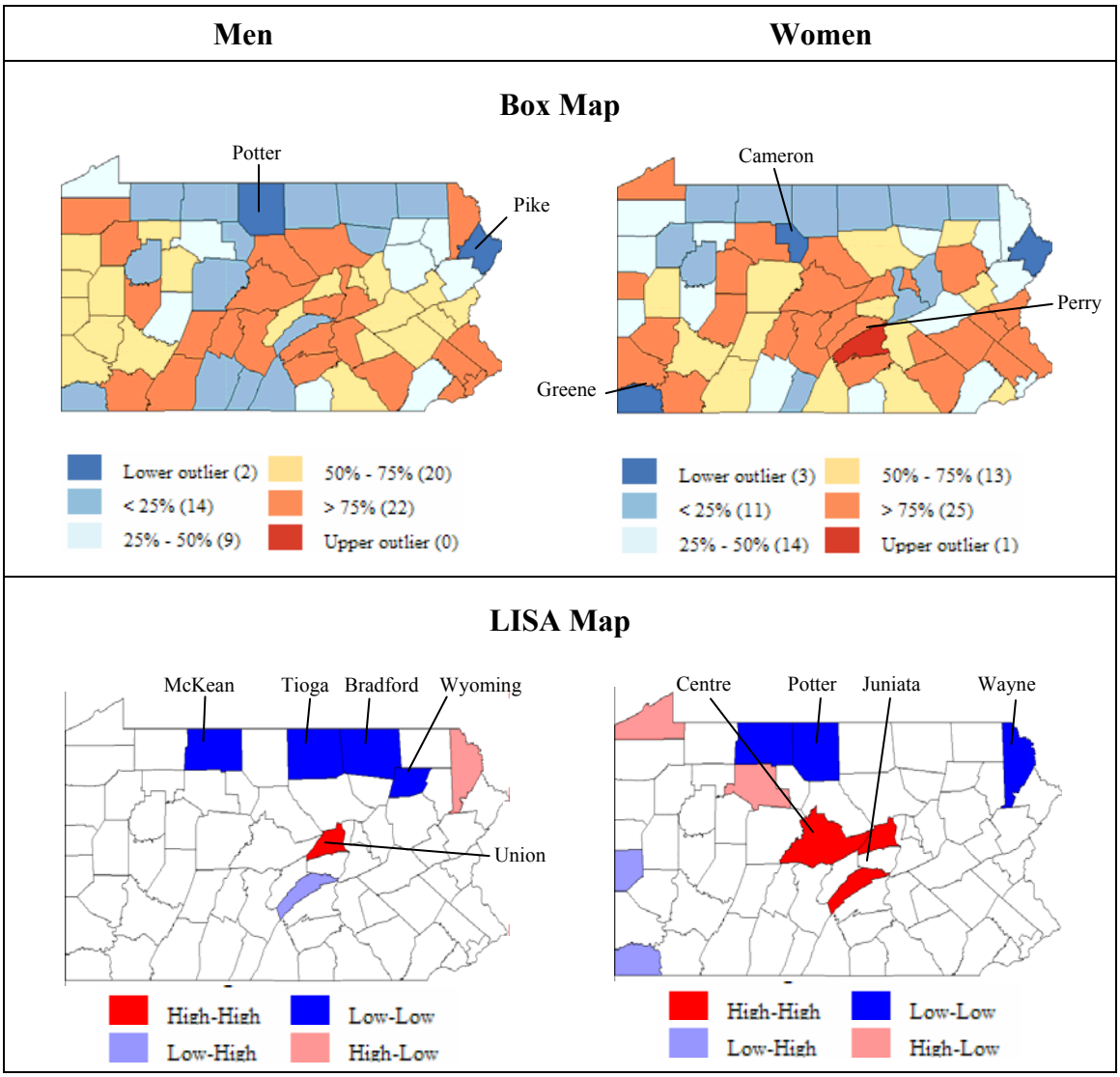


Figure B-5 Box maps and LISA cluster maps for NHL incidence in the Pennsylvania registry by gender and county for persons diagnosed from 1985 to 2004 at ages 20 to 84, adjusted for age

Table B-4 Results from spatial regression analysis for assessing predictors of cancer incidence in Idaho among persons diagnosed from 2000 to 2004, adjusted for age

	All sites [1]		Bladder [1]		Kidney & Renal Pelvis [1]	
R-square	0.396		0.243		0.394	
Independent variable	β	p-value	β	p-value	β	p-value
Arsenic (ug/ml)	0.922	0.575	-0.047	0.832	0.093	0.458
White (%)*	4.541	**0.035**	0.055	0.846	0.280	0.086
Men/Women (%)*	1.988	0.388	0.077	0.804	0.113	0.518
Population Density/Land*	0.314	0.143	-0.012	0.683	0.014	0.375
Current Smoker (%)	6.886	**0.024**	0.913	**0.026**	0.318	0.163
BMI<25 (%)	-2.473	0.314	0.183	0.577	-0.379	**0.048**
	Liver & Bile Duct [1]		Lung & Bronchus [2]		NHL [1]	
R-square	0.239		0.641		0.078	
Independent variable	β	p-value	β	p-value	β	p-value
Arsenic (ug/ml)	0.017	0.762	0.357	0.306	0.146	0.397
White (%)*	0.032	0.660	-0.252	0.561	-0.136	0.534
Men/Women (%)*	0.033	0.677	0.021	0.965	0.285	0.238
Population Density/Land*	0.005	0.519	0.012	0.778	0.008	0.732
Current Smoker (%)	0.121	0.237	2.359	**<0.001**	-0.237	0.443
BMI<25 (%)	-0.048	0.570	0.086	0.866	-0.082	0.748

[1] Ordinary least squares estimation
[2] Spatial lag model- Maximum likelihood estimation

APPENDIX C

POWER CALCULATION FOR PROJECT#3

We used the Risk Assessment Information System (RAIS, (United States Department of Energy, 2006) to calculate the Excess Lifetime Cancer Risks (ELCR) for the average man and woman. ELCR calculations used default settings for the water ingestion rate (2 L/day), exposure time (0.58 hours per day), exposure frequency (350 days/year), permeability constant (0.001 cm/hr), oral slope factor (1.50 kg-day/mg), and dermal slope factor (3.66 kg-day/mg). ELCR calculations considered cancer risks from both ingested and dermal routes of exposure to inorganic arsenic and assumed 70 and 60 kg body weights, 1.8 and 1.6 m^2 body surface areas, and 75.2 and 80.4 year exposure durations for men and women, respectively (Table C-1). The exposure durations represent the life expectancies at birth for men and women in 2004 (National Center for Health Statistics, 2006).

The Idaho Department of Water Resources (IDWR) provided 6,019 arsenic concentration measurements for approximately 1,900 ground water sources (wells and springs) distributed across the state of Idaho. Sampling occurred during summer months between 1990 and 2005. We calculated the average arsenic concentration in ground water sources from each of 23 counties (low risk counties) with low average arsenic concentrations (average arsenic concentration $\leq$2 µg/L) and from each of five counties (high risk counties) with high average arsenic concentrations (average arsenic concentration $\geq$10 µg/L). Using 1990-2004 county-specific

census counts as weighting factors, we averaged the county-specific average arsenic concentrations to produce weighted average arsenic concentrations in water (CW) for low and high risk counties, 1.41 and 13.93 µg/L, respectively.

Table C-1 shows, for men and women, the carcinogenic chronic daily intakes (ingested and dermal, CDI, in mg/kg-day) and ELCR at arsenic CW values of 1.41 and 13.93 µg/L. According to the Department of Energy RAIS, arsenic exposures in high risk Idaho counties are expected to increase lifetime cancer risks, relative to low risk Idaho counties, by 55.9 and 69.7 per 100,000 men and women, respectively. Analysis of historical U.S. census data confirmed significant growth in the Idaho resident population, amounting to a doubling in population size between 1970 and 2000, both for the state as a whole and for the five high risk counties (Figure C-1). According to the U.S. 2000 Census Summary File 3, 47.2 percent of residents of Idaho in calendar year 2000 were born in Idaho (United Stated Census Bureau).

Table C-2 shows results from power calculations. Power calculations assume a simple comparison between two independent (binomial) proportions, represented by the all malignant cancer incidence rates expected among residents of low risk and high risk Idaho counties (Hintze, 2005). In analyses that compare gender-specific 15-year cancer incidence rates between low and high risk Idaho counties, population sizes are sufficiently large to detect (with power >80%) an arsenic effect (consistent with inorganic arsenic concentrations in ground water), as long as ≥40% of the anticipated ELCR manifests in the at-risk populations. In analyses that compare gender-specific 5-year cancer incidence rates between low and high risk Idaho counties, population sizes are sufficiently large to detect (with power >80%) an arsenic effect (consistent with inorganic arsenic concentrations in ground water), as long as ≥60% of the anticipated ELCR manifests in the at-risk populations.

Table C-1 Results from application of the Department of Energy risk model for inorganic arsenic

Risk calculation	Men	Women
Life expectancy (yrs)	75.2	80.4
Body weight (kg)	70	60
Adult surface area	1.8	1.6
Carcinogenic CDI at CW=1.41 µg/L	4.19E-05	5.22E-05
Carcinogenic CDI at CW=13.93 µg/L	4.12E-04	5.14E-04
ELCR at CW=1.41 µg/L	6.33E-05	7.88E-05
ELCR at CW=13.93 µg/L	6.23E-04	7.76E-04
Difference (per 100,000)	55.9	69.7

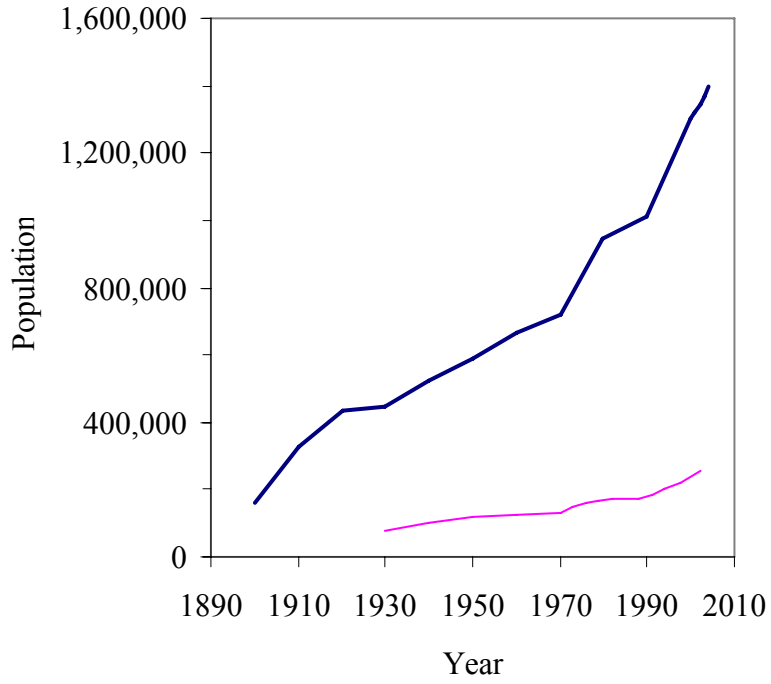


Figure C-1 Resident population for the state of Idaho (blue line) and in five Idaho counties (pink line, Canyon, Owyhee, Payette, Twin Falls, and Washington counties) with average ground water arsenic concentrations ≥10 µg/L (Census Of Population And Housing , http://www.census.gov/prod/www/abs/decennial/index.htm)

Table C-2 Power Calculation [1]

Power (1-β), difference between two 15-year average cancer incidence rates [2]

		Fractional effect [4]			
Sex	Incidence (per 100,000) [3]	0.20	0.30	0.40	0.50
Men	500 [5]	0.36	0.66	0.89	0.98
	520	0.34	0.65	0.87	0.97
	540	0.33	0.63	0.86	0.97
Women	360 [5]	0.64	0.94	1.00	1.00
	370	0.63	0.93	1.00	1.00
	380	0.62	0.92	0.99	1.00

Power (1-β), difference between two 5-year average cancer incidence rates [6]

		Fractional effect [4]			
Sex	Incidence (per 100,000) [3]	0.40	0.60	0.80	1.00
Men	500 [5]	0.48	0.82	0.97	1.00
	520	0.47	0.80	0.96	1.00
	540	0.46	0.79	0.96	1.00
Women	360 [5]	0.80	0.99	1.00	1.00
	370	0.79	0.98	1.00	1.00
	380	0.78	0.98	1.00	1.00

1.Based on two-sided Chi-square test with continuity correction and with a significance level of alpha=0.05

2. Based on total at-risk population counts for the 1990-2004 15-year time period: 2,677,438 men and 2,647,115 women in 23 low risk counties and 1,639,465 men and 1,671,457 women in 5 high risk counties

3.All malignant cancer incidence in the unexposed population

4.Assumed proportion of the Excess Lifetime Cancer Risk (ELCR, estimated at 55.9 and 69.7 per 100,000 men and women, respectively; see Table 1) manifesting in the exposed population (residents of high risk counties)

5.Observed 1990-2004 age-adjusted all malignant cancer incidence rates in low risk Idaho counties (personal communication from the Idaho Cancer Data Registry)

6.Based on total at-risk population counts for the 2000-2004 five-year time period: 934,663 men and 924,388 women in 23 low risk counties and 629,358 men and 637,956 women in 5 high risk counties

BIBLIOGRAPHY

Adamson, P., Bray, F., Costantini, A. S., Tao, M. H., Weiderpass, E., & Roman, E. (2007). Time trends in the registration of Hodgkin and non-Hodgkin lymphomas in Europe. *Eur J Cancer, 43*(2), 391-401.

American Cancer Society. (2007a). American Cancer Society Guidelines for the Early Detection of Cancer. Retrieved June 20, 2007, from http://www.cancer.org/docroot/ped/content/ped_2_3x_acs_cancer_detection_guidelines_36.asp

American Cancer Society. (2007b). Cancer Facts & Figures 2007. Retrieved May 08, 2007, from http://www.cancer.org/downloads/STT/CAFF2007PWSecured.pdf

American Cancer Society. (2007c). Cancer Statistics 2007 Presentation. Changing in the US death rates by cause, 1950 & 2004. Retrieved June 12, 2007, from http://www.cancer.org/docroot/PRO/content/PRO_1_1_Cancer_Statistics_2007_Presentation.asp

Anselin, L. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis, 27*, 93-115.

Anselin, L. (2004). Review of Cluster Analysis Software. Retrieved May 20, 2007, from http://www.schs.state.nc.us/NAACCR-GIS/pdfs/clustersoftwareFinal.pdf

Anselin, L. (2005). Center for Spatially Integrated Social Science. Exploring Spatial Data with GeoDa: A Workbook. Retrieved April 12, 2007, from https://www.geoda.uiuc.edu/pdf/geodaworkbook.pdf

Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis, 38*, 5-22.

Ansell, S. M., & Armitage, J. (2005). Non-Hodgkin lymphoma: diagnosis and treatment. *Mayo Clinic Proceedings, 80*(8), 1087-1097.

Arbeev, K. G., Ukraintseva, S. V., Arbeeva, L. S., & Yashin, A. I. (2005). Decline in human cancer incidence rates at old ages: Age-period-cohort considerations. *Demographic Research, 12*, 273.

ASTDR. (1993). *Toxicological Profile for Arsenic. TP-92/02*. Atlanta, Georgia: Agency for Toxic Substances and Disease Registry.

Ayotte, J. D., Baris, D., Cantor, K. P., Colt, J., Robinson, G. R., Jr., Lubin, J. H., et al. (2006). Bladder cancer mortality and private well use in New England: an ecological study. *Journal of Epidemiology & Community Health, 60*(2), 168-172.

Bailar, J. C. (1990). Death from all cancers. Trends in sixteen countries. *Ann N Y Acad Sci, 609*, 49-56; discussion 56-47.

Band, P. R., Le, N. D., Fang, R., & Gallagher, R. (2004). Identification of occupational cancer risks in British Columbia: a population-based case-control study of 769 cases of non-Hodgkin's lymphoma analyzed by histopathology subtypes. *Journal of Occupational & Environmental Medicine, 46*(5), 479-489.

Baris, D., & Zahm, S. H. (2000). Epidemiology of lymphomas. *Curr Opin Oncol, 12*(5), 383-394.
Bates, M. N., Smith, A. H., & Cantor, K. P. (1995). Case-control study of bladder cancer and arsenic in drinking water. *American Journal of Epidemiology, 141*(6), 523-530.
Berry, D. A., Cronin, K. A., Plevritis, S. K., Fryback, D. G., Clarke, L., Zelen, M., et al. (2005). Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med, 353*(17), 1784-1792.
Boffetta, P., & de Vocht, F. (2007). Occupation and the risk of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev, 16*(3), 369-372.
Breslow, N. E. (1984). Extra-Poisson Variation in Log-linear Models. *Applied Statistics, 33*(1), 38.
Cancer Data Registry of Idaho. (2006). Geographic Reports: County Cancer Profile. Retrieved April 09, 2007, from http://www.idcancer.org/special/CountyProfiles/CountyMap2006.htm
Cantor, K. P. (1997). Drinking water and cancer. *Cancer Causes Control, 8*(3), 292-308.
Centers for Disease Control and Prevention. (2005a). Cigarette smoking among adults--United States, 2004. *MMWR - Morbidity & Mortality Weekly Report, 54*(44), 1121-1124.
Centers for Disease Control and Prevention. (2005b). Project Profiles - Idaho National Engineering and Environmental Laboratory (INEEL): Fact Sheet: CDC Research Involving Radiation Releases from the Idaho National Laboratory. Retrieved July 02, 2007, from http://www.cdc.gov/nceh/radiation/factsheets/ineel_factsheet.pdf
Centers for Disease Control and Prevention. (2006a). Cigarette smoking among higher school students--United States *MMWR Morbidity and Mortality Weekly Report, 55*, 724-726.
Centers for Disease Control and Prevention. (2006b). *The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General*: National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
Centers for Disease Control and Prevention. (2006c). Tobacco Use Among Adults-United States, 2005. *MMWR Morbidity and Mortality Weekly Report, 55*(42), 1145-1148.
Centers for Disease Control and Prevention. (2007). National Program of Cancer Registries (NPCR). Retrieved June 30, 2007, from http://www.cdc.gov/cancer/npcr/
Chang, E. T., Balter, K. M., Torrang, A., Smedby, K. E., Melbye, M., Sundstrom, C., et al. (2006). Nutrient intake and risk of non-Hodgkin's lymphoma. *Am J Epidemiol, 164*(12), 1222-1232.
Chang, E. T., Smedby, K. E., Zhang, S. M., Hjalgrim, H., Melbye, M., Ost, A., et al. (2005). Dietary factors and risk of non-hodgkin lymphoma in men and women. *Cancer Epidemiology, Biomarkers & Prevention, 14*(2), 512-520.
Chappell, W. R., Beck, B. D., Brown, K. G., Chaney, R., Cothern, R., Cothern, C. R., et al. (1997). Inorganic arsenic: a need and an opportunity to improve risk assessment. *Environmental Health Perspectives, 105*(10), 1060-1067.
Chiou, H. Y., Hsueh, Y. M., Liaw, K. F., Horng, S. F., Chiang, M. H., Pu, Y. S., et al. (1995). Incidence of internal cancers and ingested inorganic arsenic: a seven-year follow-up study in Taiwan. *Cancer Research, 55*(6), 1296-1300.
Chow, E. J., & Holly, E. A. (2002). Blood transfusions and non-Hodgkin's lymphoma. *Epidemiologic Reviews, 24*(2), 269-279.
Chu, K. C., Tarone, R. E., & Freeman, H. P. (2003). Trends in prostate cancer mortality among black men and white men in the United States. *Cancer, 97*(6), 1507-1516.

Clarke, C. A., & Glaser, S. L. (2002). Changing incidence of non-Hodgkin lymphomas in the United States. *Cancer, 94*(7), 2015-2023.

Clarke, C. A., Glaser, S. L., Dorfman, R. F., Bracci, P. M., Eberle, E., & Holly, E. A. (2004). Expert review of non-Hodgkin's lymphomas in a population-based cancer registry: reliability of diagnosis and subtype classifications. *Cancer Epidemiology, Biomarkers & Prevention, 13*(1), 138-143.

Clarke, C. A., Glaser, S. L., Uratsu, C. S., Selby, J. V., Kushi, L. H., & Herrinton, L. J. (2006). Recent declines in hormone therapy utilization and breast cancer incidence: clinical and population-based evidence. *J Clin Oncol, 24*(33), e49-50.

Clegg, L. X., Feuer, E. J., Midthune, D. N., Fay, M. P., & Hankey, B. F. (2002). Impact of reporting delay and reporting error on cancer incidence rates and trends. *J Natl Cancer Inst, 94*(20), 1537-1545.

Clemons, M., & Goss, P. (2001). Estrogen and the risk of breast cancer. *N Engl J Med, 344*(4), 276-285.

Davis, D. L., Hoel, D., Fox, J., & Lopez, A. (1990). International trends in cancer mortality in France, West Germany, Italy, Japan, England and Wales, and the USA. *Lancet, 336*(8713), 474-481.

De Roos, A. J., Hartge, P., Lubin, J. H., Colt, J. S., Davis, S., Cerhan, J. R., et al. (2005). Persistent organochlorine chemicals in plasma and risk of non-Hodgkin's lymphoma. *Cancer Research, 65*(23), 11214-11226.

De Roos, A. J., Zahm, S. H., Cantor, K. P., Weisenburger, D. D., Holmes, F. F., Burmeister, L. F., et al. (2003). Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. *Occup Environ Med, 60*(9), E11.

Devesa, S. S., & Fears, T. (1992). Non-Hodgkin's lymphoma time trends: United States and international data. *Cancer Res, 52*(19 Suppl), 5432s-5440s.

Dinse, G. E., Umbach, D. M., Sasco, A. J., Hoel, D. G., & Davis, D. L. (1999). Unexplained increases in cancer incidence in the United States from 1975 to 1994: possible sentinel health indicators? *Annu Rev Public Health, 20*, 173-209.

Edwards, B. K., Brown, M. L., Wingo, P. A., Howe, H. L., Ward, E., Ries, L. A., et al. (2005). Annual report to the nation on the status of cancer, 1975-2002, featuring population-based trends in cancer treatment. *J Natl Cancer Inst, 97*(19), 1407-1427.

Engels, E. A. (2007). Infectious agents as causes of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev, 16*(3), 401-404.

Engels, E. A., Pfeiffer, R. M., Goedert, J. J., Virgo, P., McNeel, T. S., Scoppa, S. M., et al. (2006). Trends in cancer risk among people with AIDS in the United States 1980-2002. *Aids, 20*(12), 1645-1654.

Etzioni, R., Legler, J. M., Feuer, E. J., Merrill, R. M., Cronin, K. A., & Hankey, B. F. (1999). Cancer surveillance series: interpreting trends in prostate cancer--part III: Quantifying the link between population prostate-specific antigen testing and recent declines in prostate cancer mortality. *J Natl Cancer Inst, 91*(12), 1033-1039.

Ferreccio, C., Gonzalez, C., Milosavjlevic, V., Marshall, G., Sancha, A. M., & Smith, A. H. (2000). Lung cancer and arsenic concentrations in drinking water in Chile. *Epidemiology, 11*(6), 673-679.

Fisher, S. G., & Fisher, R. I. (2004). The epidemiology of non-Hodgkin's lymphoma. *Oncogene, 23*(38), 6524-6534.

Forman, D., Cook-Mozaffari, P., Darby, S., Davey, G., Stratton, I., Doll, R., et al. (1987). Cancer near nuclear installations. *Nature, 329*(6139), 499-505.
Fritschi, L., Benke, G., Hughes, A. M., Kricker, A., Turner, J., Vajdic, C. M., et al. (2005). Occupational exposure to pesticides and risk of non-Hodgkin's lymphoma. *American Journal of Epidemiology, 162*(9), 849-857.
Garland, M., Morris, J. S., Rosner, B. A., Stampfer, M. J., Spate, V. L., Baskett, C. J., et al. (1993). Toenail trace element levels as biomarkers: reproducibility over a 6-year period. *Cancer Epidemiol Biomarkers Prev, 2*(5), 493-497.
Gebel, T. (2000). Confounding variables in the environmental toxicology of arsenic. *Toxicology, 144*(1-3), 155-162.
Goldstein, B. D., & Witz, G. (1999). Chapter 4: Benzene. In M. Lipman (Ed.), *Environmental Toxicants: Human Exposure and Their Health Effects* (Second ed., pp. 121-150): John Wiley and Sons.
Groves, F. D., Linet, M. S., Travis, L. B., & Devesa, S. S. (2000). Cancer surveillance series: non-Hodgkin's lymphoma incidence by histologic subtype in the United States from 1978 through 1995. *Journal of the National Cancer Institute, 92*(15), 1240-1251.
Grulich, A. E., & Vajdic, C. M. (2005). The epidemiology of non-Hodgkin lymphoma. *Pathology, 37*(6), 409-419.
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.
Hagan, E. F. (2004). Idaho Department of Water Resources, Ground Water Quality Technical Brief, Statewide Ambient Ground Water Quality Monitoring Program Arsenic Speciation Results (2002 & 2003). Retrieved March 19, 2007, from http://www.idwr.idaho.gov/hydrologic/info/pubs/gwq/Arsenic_technical_Brief-final.pdf
Hardell, L., & Eriksson, M. (1999). A case-control study of non-Hodgkin lymphoma and exposure to pesticides. *Cancer, 85*(6), 1353-1360.
Hartge, P., & Smith, M. T. (2007). Environmental and behavioral factors and the risk of non-Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev, 16*(3), 367-368.
Hennessy, B. T., Hanrahan, E. O., & Daly, P. A. (2004). Non-Hodgkin lymphoma: an update. *Lancet Oncology, 5*(6), 341-353.
Hintze, J. (2005). NCSS and PASS. Number Cruncher Statistical Systems. Kaysville, Utah.
Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics, 39*(2), 311-324.
Hopenhayn-Rich, C., Biggs, M. L., & Smith, A. H. (1998). Lung and kidney cancer mortality associated with arsenic in drinking water in Cordoba, Argentina. *Int J Epidemiol, 27*(4), 561-569.
Howe, G. R. (1991). Risk of cancer mortality in populations living near nuclear facilities. *Jama, 265*(11), 1438-1439.
Howe, H. L., Wu, X., Ries, L. A., Cokkinides, V., Ahmed, F., Jemal, A., et al. (2006). Annual report to the nation on the status of cancer, 1975-2003, featuring cancer among U.S. Hispanic/Latino populations. *Cancer, 107*(8), 1711-1742.
Huang, C., Ke, Q., Costa, M., & Shi, X. (2004). Molecular mechanisms of arsenic carcinogenesis. *Mol Cell Biochem, 255*(1-2), 57-66.
Hughes, M. F. (2002). Arsenic toxicity and potential mechanisms of action. *Toxicol Lett, 133*(1), 1-16.

Jemal, A., Ward, E., & Thun, M. J. (2005). Contemporary lung cancer trends among U.S. women. *Cancer Epidemiology, Biomarkers & Prevention, 14*(3), 582-585.

Johnson, C. J., & Carson, S. L. (2000). Cancer Data Registry of Idaho, Geographic Reports: Cancer Trend in Idaho: 1971-1998. Retrieved May 01, 2007, from http://www.idcancer.org/special/trend_report/cancer%20trends%20interactive.pdf

Johnson, C. J., & Mitchell, T. S. (2007). An evaluation of potential association between arsenic concentrations in ground water and 2000-2004 cancer incidence rates in Idaho by Zip code. Cancer Data Registry of Idaho and Department of Environmental Quality. Retrieved May 01, 2007, from http://www.idcancer.org/special/arseniccancerzipcodev2.pdf

Kapaj, S., Peterson, H., Liber, K., & Bhattacharya, P. (2006). Human health effects from chronic arsenic poisoning--a review. *J Environ Sci Health A Tox Hazard Subst Environ Eng, 41*(10), 2399-2428.

Karagas, M. R., Tosteson, T. D., Morris, J. S., Demidenko, E., Mott, L. A., Heaney, J., et al. (2004). Incidence of transitional cell carcinoma of the bladder and arsenic exposure in New Hampshire. *Cancer Causes & Control, 15*(5), 465-472.

Kelemen, L. E., Cerhan, J. R., Lim, U., Davis, S., Cozen, W., Schenk, M., et al. (2006). Vegetables, fruit, and antioxidant-related nutrients and risk of non-Hodgkin lymphoma: a National Cancer Institute-Surveillance, Epidemiology, and End Results population-based case-control study. *Am J Clin Nutr, 83*(6), 1401-1410.

Keller-Byrne, J. E., Khuder, S. A., Schaub, E. A., & McAfee, O. (1997). A meta-analysis of non-Hodgkin's lymphoma among farmers in the central United States. *American Journal of Industrial Medicine, 31*(4), 442-444.

Kim, H. J., Fay, M. P., Feuer, E. J., & Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med, 19*(3), 335-351.

Kitchin, K. T. (2001). Recent advances in arsenic carcinogenesis: modes of action, animal model systems, and methylated arsenic metabolites. *Toxicol Appl Pharmacol, 172*(3), 249-261.

Lamm, S. H., Engel, A., Kruse, M. B., Feinleib, M., Byrd, D. M., Lai, S., et al. (2004). Arsenic in drinking water and bladder cancer mortality in the United States: an analysis based on 133 U.S. counties and 30 years of observation. *J Occup Environ Med, 46*(3), 298-306.

Lewis, D. R., Southwick, J. W., Ouellet-Hellstrom, R., Rench, J., & Calderon, R. L. (1999). Drinking water arsenic in Utah: A cohort mortality study. *Environ Health Perspect, 107*(5), 359-365.

Liu, S., Semenciw, R., & Mao, Y. (2003). Increasing incidence of non-Hodgkin's lymphoma in Canada, 1970-1996: age-period-cohort analysis. *Hematological Oncology, 21*(2), 57-66.

Lynge, E., Anttila, A., & Hemminki, K. (1997). Organic solvents and cancer. *Cancer Causes Control, 8*(3), 406-419.

Marshall, G., Ferreccio, C., Yuan, Y., Bates, M. N., Steinmaus, C., Selvin, S., et al. (2007). Fifty-year study of lung and bladder cancer mortality in Chile related to arsenic in drinking water. *J Natl Cancer Inst, 99*(12), 920-928.

McKean-Cowdin, R., Feigelson, H. S., Ross, R. K., Pike, M. C., & Henderson, B. E. (2000). Declining cancer rates in the 1990s. *Journal of Clinical Oncology, 18*(11), 2258-2268.

Meliker, J. R., Slotnick, M. J., Avruskin, G. A., Kaufmann, A., Fedewa, S. A., Goovaerts, P., et al. (2007). Individual lifetime exposure to inorganic arsenic using a space-time information system. *Int Arch Occup Environ Health, 80*(3), 184-197.

Michaud, D. S., Wright, M. E., Cantor, K. P., Taylor, P. R., Virtamo, J., & Albanes, D. (2004). Arsenic concentrations in prediagnostic toenails and the risk of bladder cancer in a cohort study of male smokers. *Am J Epidemiol, 160*(9), 853-859.

Midthune, D. N., Fay, M. P., Clegg, L. X., & Feuer, E. J. (2005). Modeling reporting delays and reporting corrections in cancer registry data. *Journal of the American Statistical Association, 100*(469), 61.

Moore, L. E., Lu, M., & Smith, A. H. (2002). Childhood cancer incidence and arsenic exposure in drinking water in Nevada. *Arch Environ Health, 57*(3), 201-206.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*(1-2), 17-23.

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

National Cancer Institue (SEER). Number of Persons by Race and Hispanic Ethnicity for SEER Participants (2000 Census Data). Retrieved July 31, 2007, from http://seer.cancer.gov/registries/data.html

National Cancer Institute. (2005a). Cancer Trend Progress Report - 2005 update. Retrieved March 25, 2007, from http://progressreport.cancer.gov/

National Cancer Institute. (2005b). Statistical Research and Applications. Joinpoint regression program (version 3.0). Retrieved January 05, 2007 from http://srab.cancer.gov/joinpoint/

National Cancer Institute. (2005c). Surveillance, Epidemiology, and End Results (SEER) Program. Retrieved March 07, 2007, from http://seer.cancer.gov/about/SEER_brochure.pdf

National Cancer Institute. (2006a). Single Year of Age County Population Estimates, 1969-2003. Retrieved February 02, 2007, from http://seer.cancer.gov/popdata/singleages.html

National Cancer Institute. (2006b). Surveillance, Epidemiology, and End Results (SEER) Program, SEER*Prep software, released May 2006. Retrieved January 18, 2007, from http://seer.cancer.gov/seerprep/

National Cancer Institute. (2006c). Surveillance, Epidemiology, and End Results (SEER) Program. SEER 1973-2004 public use data. SEER*Stat Database: Incidence-SEER 9 Regs, Nov 2005 sub (1973-2004), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2007, based on the November 2006 submission. Retrieved November 28, 2006, from www.seer.cancer.gov

National Cancer Institute. (2007a). Covariates Used in the Final Delay Distributions, By Cancer Site, and Race. Retrieved February 02, 2007, from http://srab.cancer.gov/delay/covariates.html

National Cancer Institute. (2007b). SEER Cancer Statistics Review, 1975-2004, based on November 2006 SEER data submission. Retrieved June 20, 2007, from http://seer.cancer.gov/csr/1975_2004/

National Cancer Institute. (2007c). Surveillance, Epidemiology, and End Results (SEER) Program, SEER*Stat software, Latest Release: April 2007. Retrieved May 01, 2007, from http://seer.cancer.gov/seerstat/

National Cancer Institute. Statistical Research and Applications Branch. (2005). Joinpoint Regression Program. (Version 3.0). Bethesda, Maryland.

National Center for Health Statistics. (2006). U.S. Life expectancy at birth in 2004: Health, United States, 2006 With Chartbook on Trends in the Health of Americans. Retrieved July 30, 2007, from http://0-www.cdc.gov.mill1.sjlibrary.org/nchs/data/hus/hus06.pdf

Neely, K. W. (2002). Arsenic Results from the Statewide Program, 1991-2001. Retrieved May, 08, 2007, from http://www.idwr.idaho.gov/hydrologic/info/pubs/wib/wib50p5-arsenic_1991-2001.pdf

Nelson, N. J. (2005). Studies examine whether persistent organic agents may be responsible for rise in lymphoma rates. *Journal of the National Cancer Institute, 97*(20), 1490-1491.

Nuckols, J. R., Ward, M. H., & Jarup, L. (2004). Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environmental Health Perspectives, 112*(9), 1007-1015.

Pan, S. Y., Mao, Y., Ugnat, A. M., & Canadian Cancer Registries Epidemiology Research, G. (2005). Physical activity, obesity, energy intake, and the risk of non-Hodgkin's lymphoma: a population-based case-control study. *American Journal of Epidemiology, 162*(12), 1162-1173.

Pearce, N., & Bethwaite, P. (1992). Increasing incidence of non-Hodgkin's lymphoma: occupational and environmental factors. *Cancer Res, 52*(19 Suppl), 5496s-5500s.

Pennsylvania Department of Health. (2004). Pennsylvania Cancer Incidence and Mortality 2003. Retrieved May 05, 2007, from http://www.dsf.health.state.pa.us/health/site/default.asp

Pennsylvania Department of Health. (2005). HIV/AIDS Surveillance Biannual Summary. Retrieved April 04, 2007, from http://www.dsf.health.state.pa.us/health/lib/health/hivaids_biannuals/preliminary_annual_report_december_2005.pdf

Pennsylvania Department of Health. (2006a). 2005 Behavioral Risks of Pennsylvania Adults. Retrieved April 20, 2007, from http://www.dsf.health.state.pa.us/health/lib/health/BRFSS/2005/PABRFSS2005OvrOb.pdf

Pennsylvania Department of Health. (2006b). Cancer Facts and Figures, Pennsylvania 2006: Introduction and Technical Notes. Retrieved June 27, 2007, from http://www.dsf.health.state.pa.us/health/lib/health/FactsnFigures/01.pdf

Pennsylvania Department of Health. (2007). Cancer Incidence and Mortality. Retrieved March 18, 2007, from http://www.dsf.health.state.pa.us/health/cwp/view.asp?a=175&Q=20151

Quintana, P. J., Delfino, R. J., Korrick, S., Ziogas, A., Kutz, F. W., Jones, E. L., et al. (2004). Adipose tissue levels of organochlorine pesticides and polychlorinated biphenyls and risk of non-Hodgkin's lymphoma. *Environ Health Perspect, 112*(8), 854-861.

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.

Rahman, M. M., Chowdhury, U. K., Mukherjee, S. C., Mondal, B. K., Paul, K., Lodh, D., et al. (2001). Chronic arsenic toxicity in Bangladesh and West Bengal, India--a review and commentary. *J Toxicol Clin Toxicol, 39*(7), 683-700.

Robson, M. (2003). Methodologies for assessing exposure to metals: human host factors. *Ecotoxicology and Environmental Safety, 56*, 104.

Rothman, N., Cantor, K. P., Blair, A., Bush, D., Brock, J. W., Helzlsouer, K., et al. (1997). A nested case-control study of non-Hodgkin lymphoma and serum organochlorine residues. *Lancet, 350*(9073), 240-244.

Sandin, S., Hjalgrim, H., Glimelius, B., Rostgaard, K., Pukkala, E., & Askling, J. (2006). Incidence of non-Hodgkin's lymphoma in Sweden, Denmark, and Finland from 1960 through 2003: an epidemic that was. *Cancer Epidemiol Biomarkers Prev, 15*(7), 1295-1300.

Scherr, P. A., Hutchison, G. B., & Neiman, R. S. (1992). Non-Hodgkin's lymphoma and occupational exposure. *Cancer Res, 52*(19 Suppl), 5503s-5509s.

Schroeder, J. C., Olshan, A. F., Baric, R., Dent, G. A., Weinberg, C. R., Yount, B., et al. (2001). Agricultural risk factors for t(14;18) subtypes of non-Hodgkin's lymphoma. *Epidemiology, 12*(6), 701-709.

Skibola, C. F., Holly, E. A., Forrest, M. S., Hubbard, A., Bracci, P. M., Skibola, D. R., et al. (2004). Body mass index, leptin and leptin receptor polymorphisms, and non-hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev, 13*(5), 779-786.

Smith, A. H., Goycolea, M., Haque, R., & Biggs, M. L. (1998). Marked increase in bladder and lung cancer mortality in a region of Northern Chile due to arsenic in drinking water. *American Journal of Epidemiology, 147*(7), 660-669.

Smith, M. T., Jones, R. M., & Smith, A. H. (2007). Benzene exposure and risk of non-Hodgkin's lymphoma. *Cancer Epidemiology Biomarkers and Prevention, 16*(3), 385-391.

Steinmaus, C., Yuan, Y., Bates, M. N., & Smith, A. H. (2003). Case-control study of bladder cancer and drinking water arsenic in the western United States. *American Journal of Epidemiology, 158*(12), 1193-1201.

Surveillance Epidemiology and End Results (SEER) Program. (2006a). Cancer Query Systems: Delay-Adjusted SEER Incidence Rates. Retrieved November 01, 2006, from http://srab.cancer.gov/delay/canques.html

Surveillance Epidemiology and End Results (SEER) Program. (2006b). SEER 1973-2003 public use data. SEER*Stat Database: Incidence-SEER 9 Regs, Nov 2004 sub (1973-2003), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2006, based on the November 2005 submission. Retrieved April 30, 2007, from www.seer.cancer.gov

Surveillance Epidemiology and End Results (SEER) Program. (2006c). SEER*Stat Database: Mortality - All COD, Public-Use With State, Total U.S. (1969-2003), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2006. Underlying mortality data provided by National Center for Health Statistics. Retrieved January 05, 2007, from www.seer.cancer.gov

Takkouche, B., Etminan, M., & Montes-Martinez, A. (2005). Personal use of hair dyes and risk of cancer: a meta-analysis. *Jama, 293*(20), 2516-2525.

Tapio, S., & Grosche, B. (2006). Arsenic in the aetiology of cancer. *Mutat Res, 612*(3), 215-246.

Tchounwou, P. B., Centeno, J. A., & Patlolla, A. K. (2004). Arsenic toxicity, mutagenesis, and carcinogenesis--a health risk assessment and management approach. *Mol Cell Biochem, 255*(1-2), 47-55.

Toossi, M. (2002). A century of change: the U.S. labor force from 1950 to 2050. *Monthly Labor Review, 125*, 15-28.

United Stated Bureau of the Census. (1996). Population Projections of the United States by Age, Sex, Race, and Hispanic Origin: 1995 to 2050, U.S Bureau of the Census, Current Population Reports. Retrieved March 27, 2007

United Stated Census Bureau. American FactFinder. Retrieved July 30, 2007, from http://factfinder.census.gov

United Stated Environmental Protection Agency. (2001). Arsenic in drinking water. Final rule. Fed. Regist. 66(14), 7000-7010 (January 22).
United States Buerau of Census. Retrieved March 04, 2007, from http://www.census.gov/
United States Bureau of Census. (2001). Cartographic Products Management Branch. Retrieved March 03, 2007, from http://www.census.gov/geo/www/cob/bdy_files.html
United States Bureau of Census. (2005). United States Census 2000 Summary File. Retrieved May 5, 2007, from http://www.census.gov/
United States Census Bureau. (2001). Cartographic Products Management Branch. Retrieved March 03, 2007, from http://www.census.gov/geo/www/cob/bdy_files.html
United States Department of Agriculture. (2007). National Agriculture Statistics Services. Retrieved April 25, 2007, from http://www.nass.usda.gov/index.asp
United States Department of Energy. (2006). The Risk Assessment Information System (RAIS). Retrieved July 31, 2007, from http://rais.ornl.gov/
United States Department of Health and Human Services (DHHS). (2006). Data on Health and Well-being of American Indians, Alaska Natives, and Other Native Americans, Data Catalog. Surveillance, Epidemiology, and End Results (SEER) Retrieved August 01, 2007, from http://aspe.hhs.gov/hsp/06/Catalog-AI-AN-NA/SEER.htm
United States Environmental Protection Agency. (1997). Pesticides industrial sales and usage: 1994-1995 Pesticide market estimates: Tables and Charts. Retrieved July 07, 2007, from http://www.epa.gov/oppbead1/pestsales/95pestsales/tables_charts1995_4.htm
United States Environmental Protection Agency. (2001). Arsenic in drinking water. Final rule. Fed. Regist. 66(14), 7000-7010 (January 22).
United States Office of Personnel Management. (2006). Central Personnel Data File (CPDF). Executive Branch (non-Postal) Employment by gender, Rave/National Origin, Disability Status, Veterans Status, Disabled Veterans, 1994-2004. Retrieved April 04, 2007, from http://www.opm.gov/feddata/html/cpdfinfo.asp
Vineis, P., Miligi, L., & Costantini, A. S. (2007). Exposure to solvents and risk of non-Hodgkin lymphoma: clues on putative mechanisms. *Cancer Epidemiol Biomarkers Prev, 16*(3), 381-384.
Welch, A. H., A.Watkins, S., Helsel, D. R., & Focazio, M. J. (2000). U.S. Geological Survey. Arsenic in Ground-Water Resources of the United States. USGS Fact Sheet 063-00. Retrieved March 02, 2007, from http://water.usgs.gov/nawqa/trace/pubs/fs-063-00/
World Health Organization. (2000). International Classification of Diseases for Oncology, 3rd Edition (ICD-O-3). Retrieved June 15, 2007, from http://www.who.int/classifications/icd/adaptations/oncology/en/index.html
World Health Organization. (2001). Arsenic in drinking water Retrieved April 20, 2007, from http://www.who.int/mediacentre/factsheets/fs210/en/index.html
World Health Organization. (2007). The world health organization's fight against cancer: Strategies that prevent, cure and care. Retrieved June 30, 2007, from http://www.who.int/cancer/publicat/WHOCancerBrochure2007.FINALweb.pdf
Wu, M. M., Kuo, T. L., Hwang, Y. H., & Chen, C. J. (1989). Dose-response relation between arsenic concentration in well water and mortality from cancers and vascular diseases. *Am J Epidemiol, 130*(6), 1123-1132.
Yoshida, T., Yamauchi, H., & Fan Sun, G. (2004). Chronic health effects in people exposed to arsenic via the drinking water: dose-response relationships in review. *Toxicol Appl Pharmacol, 198*(3), 243-252.

Zahm, S. H., & Blair, A. (2003). Occupational cancer among women: where have we been and where are we going? *Am J Ind Med, 44*(6), 565-575.

Zhang, Y., Holford, T. R., Leaderer, B., Boyle, P., Zahm, S. H., Flynn, S., et al. (2004). Hair-coloring product use and risk of non-Hodgkin's lymphoma: a population-based case-control study in Connecticut. *American Journal of Epidemiology, 159*(2), 148-154.

Zheng, T., Blair, A., Zhang, Y., Weisenburger, D. D., & Zahm, S. H. (2002). Occupation and risk of non-Hodgkin's lymphoma and chronic lymphocytic leukemia. *J Occup Environ Med, 44*(5), 469-474.

Zheng, T., Holford, T. R., Leaderer, B., Zhang, Y., Zahm, S. H., Flynn, S., et al. (2004). Diet and nutrient intakes and risk of non-Hodgkin's lymphoma in Connecticut women. *American Journal of Epidemiology, 159*(5), 454-466.