2019

Hazardous Waste Site Proximity and Type 2 Diabetes: From Youths to Adults

Theresa Ann Johnson
Walden University

Walden University

College of Health Sciences

This is to certify that the doctoral dissertation by

Theresa A. Johnson

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Raymond Thron, Committee Chairperson, Public Health Faculty
Dr. Vasileios Margaritis, Committee Member, Public Health Faculty
Dr. Scott McDoniel, University Reviewer, Public Health Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2018

Abstract

Hazardous Waste Site Proximity and Type 2 Diabetes: From Youths to Adults

by

Theresa A. Johnson

MSHSA, Barry University, 2000

BS, Baylor College of Medicine, 1984

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

February 2019

Abstract

Type 2 diabetes has reached epidemic proportions in adults and youths. Persistent organic pollutants and endocrine-disrupting chemicals (EDCs), such as pesticides, dioxins, and organochlorines, are omnipresent and persist in the environment with potential for human exposure via contaminated air, waterways, soil, and human food supply. EDCs have been correlated with diabetes incidence and risks. Residential proximity to hazardous waste sites (HWS) has been correlated with increased hospital admission rates for diabetes. The study used a sample population (N = 1,724), ages ≥ 12 years from the 2005-2012 Continuous NHANES and HWS data from the National Priorities List of Superfund Sites. The ecosocial theory of disease distribution, and geocoordinates provided theoretical support. Mann-Whitney U test and binary logistic regression analysis were used to investigate the relationship between residence ≤ 1 mile compared with residence >1 mile from a HWS in NHANES surveyed counties of NJ, PA, NY, and CA on the outcome abnormal A1c ≥ 5.7% while controlling the effect of the moderators: abnormal body mass index (BMI), age, sex, and race/ethnicity on the relationship. Participants with a BMI ≥ 28.95 kg/m^2 were 1.8 times and persons ≥ 58 years of age were 2.1 times more likely to have an A1c ≥ 5.7%. Also, non-Hispanic Whites residing >1 mile of a HWS had 82.1 % reduced risk of abnormal A1c compared with the same group residing ≤ 1 mile of a HWS. The results forge opportunities for future studies to consider border distance between residence and HWS. In addition, the results may promote positive social change through diabetes risks education, environmental health education, and practices and raise dialogue about social justice and the geographic distribution of hazardous waste sites.

Hazardous Waste Site Proximity and Type 2 Diabetes: From Youths to Adults

by

Theresa A. Johnson

MSHSA, Barry University, 2000

BS, Baylor College of Medicine, 1984

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

February 2019

Dedication

To my mother (Valaria Odom Johnson), who traveled this journey with me on earth and in heaven. Your everlasting praise, “to God be the glory for the things He has done”, reigns in my heart and mind daily. This is your final degree! To my brother, Samuel K. Johnson, for always encouraging me and telling me that I was strong. Your legacy lives! To the both of you…sleep in heavenly peace.

Acknowledgments

A wealth of gratitude to my dissertation committee faculty, the Walden Academic Skills Center and Tutoring Service and the RDC Analysts at the Atlanta, GA NCHS Research Data Center. And a special thank you to my family and friends who supported me every computer key strike after another!

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Introduction

The adverse health effects of tainted water and contaminated soil within or near residential areas forms one investigative target in environmental epidemiology (Friis & Sellers, 2004; Hermanson & Johnson, 2007). Public health concerns regarding the risks to human health from exposure to environmental contaminants in the air, water, soil, and food supply either from HWS, current or historical manufacturing processes, years of vector-borne disease protection (mosquito spraying with dichlorodiphenyltrichloroethane [DDT]), or water run-off from toxic chemicals has increased in past decades (Environmental Work Group, 2011; Goncharov, Bloom, Pavuk, Birman & Carpenter, 2010). These environmental threats and contributors to the burden of human diseases in adults, children, and the unborn have not escaped the eye of researchers. Bijlsma and Cohen (2016) have not only stated that human exposure to environmental chemicals has increased "exponentially over the past decades" but that there is mounting research evidence that suggests these chemicals in "air, water, soil, food, building materials and household products are toxicants that contribute to many of the chronic diseases typically seen in clinical practice" (p. 3).

Human exposure to hazardous environmental waste, air pollutants, and endocrine-disrupting chemicals (EDCs), such as persistent organic pollutants (e.g., polychlorinated biphenyls [PCBs], dioxin, and dioxin-like compounds), bisphenol A, phthalates, pesticides (such as DDT and its metabolites), and arsenic have been associated with either the incidence of, increased risk of, or as contributors to chronic diseases. Some of

these chronic diseases include cancer, asthma, and other respiratory conditions (Andre et al., 2006; X. Liu, Lessner, & Carpenter, 2012). In addition, neurological and neurobehavioral disorders, learning disabilities, certain reproductive and developmental disorders, as well as cardiovascular disease, hypertension, and obesity have been associated with human exposure to environmental contaminants (Breton et al., 2016; Goncharov et al., 2010; Harari et al., 2010; Méndez-Gómez et al., 2008). Although Type 2 diabetes has been viewed in the medical establishment as an outcome of lifestyle choices and unwanted genetics, researchers have also concluded human exposure to persistent organic pollutants (POPs) and EDCs or chemical compounds within the environment as a strong contributor to the global prevalence of this chronic condition (Anderson et al., 2012; Codru, Schymura, & Negoita, 2007; Lee, Porta, Jacobs, & Vandenberg, 2014; Tseng et al., 2002). Although pesticides such as DDT, used for mosquito control, were banned from the United States in 1972, metabolites of DDT, dichlorodiphenyldichloroethylene (DDE) and 1,1-dichloro-2,2-bis(p-chlorophenyl) ethane (DDD), and other POPs, remain in the environment as potential threats to human health. In fact, the Endocrine Society in its second scientific statement since 2009 on EDCs summarized population based studies on human exposure to EDCs (which includes some POPs such as DDT metabolites), and the complex mechanisms by which EDCs alter the human endocrine system (Gore et al., 2015). The introduction of human exposure to EDCs persistent in the environment strongly suggests a potential environmental effect on the burden of Type 2 diabetes in the United States and globally (Færch et al., 2012; Lee et al., 2014).

Within the scientific and medical communities, the term *environmental influence on adverse health outcome(s)* has expanded through the years to include health risks from the internal environment (household) and external environment. More specifically, environmental influence as a potential contributor to chronic disease refers to known and unknown infusions of hazardous chemicals and chemical compounds into the organic environment (water, soil, food supply, etc.) and that which may persist in the organic environment for years, having the potential for adverse health effects on human populations exposed to these toxicants (Stockholm Convention, 2016a, 2016b).

As Type 2 diabetes mellitus (T2DM) has reached epidemic proportions throughout the globe, to include its prevalence in children, consideration of the environmental effects on the burden of T2DM in the United States and other developed and developing countries has been under investigation during the latter 20th century (World Health Organization [WHO], 2016). Although studies have been published on the association of residential exposure to environmental contaminants and T2DM, these studies have been limited to adult populations (Kouznetsova, Huang, Ma, Lessner, & Carpenter, 2006; Navas-Acien, Silbergeld, Pastor-Barriuso, & Guallar, 2008). I propose that prior researchers examining the relationship between remote (residential) exposure and the diabetogenic effects of environmental EDCs, such as dioxins, and dioxin-like compounds, PCBs, organochlorine pesticides, arsenic, and phthalates, limited the study population to adults due to the outcome of historical studies that examined the latency effect of POPs years after exposure and years after a large quantity of contaminant exposure. For example, Cranmer, Louie, Kennedy, Kern, and Fonseca (2000) reported on

the incidence of non-Hodgkin's lymphoma in Vietnam veterans years after their exposure to agent orange. Bertazzi et al., (2001) reported on the adverse health effects of a population exposed to a large quantity of dioxins 20 years following the 1976 Seveso, Italy, industrial accident. Although the study populations in the 20-year Seveso, Italy, study included adults and children, outcomes of T2DM years after exposure is known to have been investigated and reported in adults only.

However, human exposure to commercial and industrial chemicals is ubiquitous in the general population of the United States and other developed and developing countries (Bijlsma & Cohen, 2016). The ubiquity of environmental contaminants with potential for adverse human health outcomes makes all individuals, children and adults, susceptible to environmental toxicants and POPs via poor quality ambient air, proximity to HWS, lifestyle (e.g., use of plastic non-BPA bottles), water quality, and contaminated food supply (Environmental Work Group, 2011; Tyrrell, Melzer, Henly, Galloway & Osborne, 2013; United Nations Environmental Programme, 2013). Coupled with the ubiquitous nature of commercial and industrial chemicals is the uncertainty in the scientific community about the full effects of low levels of human exposure to POPs such as PCBs, PBDEs, DDT, and DDE on chronic disease (Codru et al., 2007; Stockholm Convention, 2016a). This raises concerns about the potential health effects from childhood exposure to EDCs/chemical compounds during the preteen and adolescent years when hormonal growth and change are on a natural surge, thereby subjecting persons younger than 20 years to the potential manifestations of an altered endocrine system, which could include signs and symptoms of T2DM. To the best of my knowledge, my research is the first

investigation aimed at investigating the relationship between residential proximity to HWS and T2DM that includes adolescents from a representative sample of the United States population. Specifically, I targeted the relationship between residential proximity to HWS and the abnormal hemoglobin A1c in a study population ages 12 years and older.

In this chapter, I briefly discuss the research literature relevant to the scope of my investigation, providing foundational information on the knowledge gap related to environmental influences and the disease burden of T2DM in the United States. Much of this has been stated in the aforementioned paragraphs, and is continued in the sections that follow. I provide in Chapter 1 the research problem and specific research questions for this investigation as well as the purpose and significance of my study to fulfill the knowledge gap concerning the T2DM epidemic in youths and adults of the United States and globally. I provide in this chapter the epidemiologic theory that framed this investigation as well as a rationale for the selection of the research design. Within Chapter 1, I define the study variables, exposure assessment, HWS, residential proximity to HWS, endocrine disruptors, POPs, and the study population. I provide the medical definition for the diagnosis of prediabetes and T2DM, outline assumptions relevant to my study, and discuss why I have chosen this study focus. Chapter 1 concludes with a description of the study limitations, study significance, and implications for social change. I then transition to Chapter 2 to explore the research literature that grounded my study.

Background

The potential association between human exposure to POPs and chronic diseases has been an ongoing conversation of epidemiologic studies and collaborative environmental health workgroups during the last 25 years (Collaborative on Health and the Environment [CHE], 2014; Suarez-Lopez, Lee, Porta, Steffes, & Jacobs, 2015). Dioxins, of which 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) is the most potent compound, and its metabolites, DDE and DDD, remain in the environment as byproducts of industrial processes such as hospital waste incineration, chemical synthesis of certain compounds, and pulp from paper mills and have the potential to affect animals and humans via contaminated waterways, soil, and food (Environmental Work Group, 2011; Gore et al., 2015; Kouznetsova et al., 2006). TCDD is a human carcinogen, having also non-carcinogenic effects. For example, Harari et al. (2010) described neurobehavioral deficits in children exposed to pesticides. Lind, Bavel, Salihovic, and Lind (2012) described incidence of carotid atherosclerosis in association with serum levels of POPs. These studies add to the historical presentations by researchers in the 1990s and early 2000s investigating the effects of human exposure to dioxin and dioxin-like compounds and chronic disease.

For several years, researchers have been concerned about the adverse health effects from long-term exposure to POPs at low levels of exposure as well as the potential relationship between distance-based human exposure to hazardous waste and environmental contaminants and adverse human health outcomes. In a 1998 study by Beretazzi et al., the researchers concluded an increase in cancer incidence in Seveso,

Italy, among a population with long-term dioxin exposure. Kramarova et al. (1998) discussed the prevalence of non-Hodgkin's lymphomas in Vietnam War veterans exposed to agent orange. Lawson et al. (2004) identified a significant relationship between occupational exposure to dioxins and adverse birth outcomes. Specific to my dissertation are historical investigations involving human exposure to environmental dioxins and dioxin-like compounds, or endocrine altering chemicals or chemical mixtures, and the potential effects or association with endocrine related conditions such as diabetes. Cranmer et al. (2000) identified an association between TCDD, hyperinsulinemia, and insulin resistance. Kouznetsova et al. (2006) described a positive relationship between hospitalization rate for diabetes and residential proximity to HWS. Thus, the curiosity between human exposure to environmental contaminants and ill health has a foundational existence in the scientific community and one that remains an area of exploration.

Dioxins are not the only POPs or hazardous environmental chemicals associated with adverse hormone actions, such as diabetes. POPs and environmental EDCs are omnipresent in the environment, and the lists of POPs and endocrine-disrupting environmental chemicals with potential for human disease continues to grow (Stockholm Convention, 2016a, 2016c). Other POPs such as PCBs and polybrominated diethyl ethers, as well as other EDCs such as organochlorine pesticides, arsenic, benzene, bisphenol A, phthalates, and mixtures of environmental EDCs, bioaccumulate and biogmagnify in the environment much like dioxins with potential for human exposure via contaminated air, waterways, soil, and the human food supply (Gore et al., 2015; Oluoch-Otiego et al., 2016; Stockholm Convention, 2016c). The last decade of environmental

research on chemical toxicants has revealed mammalian studies and some epidemiologic studies with increasing evidence of potential for adverse human health effects, such as diabetes, associated with exposure to environmental endocrine disrupting toxicants (Gore et al., 2015). However, the literature is also clear that continued research is necessary to ascertain the level of exposure risk of environmental endocrine disruptors with potential for adverse health effects in humans (Bijlsma & Cohen, 2016).

Since the May 1995 Stockholm Convention and the United Nations Environment Programme (UNEP) 2005 establishment of the POPs review committee (POPRC), PCBs and mixtures of environmental chemicals and POP compounds have been highlighted for their endocrine disrupting capabilities and diabetogenic potential in humans at exposure dose levels considered to be current or low levels of exposure (Gore et al., 2015; Stockholm Convention, 2016b). The term *current* exposure refers to the level of exposure that is consistent with the ubiquitous nature of POPs and EDCs (low-level exposure) and general population exposure risks. Codru et al. (2007) described a positive association between elevated serum PCBs, DDE, and hexachlorobenzene (HCB) and diabetes in an adult population of Native Americans. Lee et al. (2010) concluded that certain POPs may increase diabetes risk through a pathway of endocrine disruption.

With respect to the relationship between residential proximity to HWS and health risks in populations younger than 20 years, Liu et al. (2012) concluded an increased risk of hospitalization for respiratory illnesses in subpopulations residing in the same ZIP code as a polluting fuel-fired power plant facility. Members of these subpopulations included children and adults. Although studies such as this have done much to advance

knowledge concerning the relationship between remote environmental exposure to pollutants and chronic respiratory disease in persons younger than 20 years, the available research on remote residential exposure to a HWS and T2DM in youths is less than desirable. The aim of my study is to bridge this knowledge gap, adding to literary works on the epidemiology of T2DM in youth and children's environmental health, through the inclusion of a study population ages 12 years and older. In Chapter 2, I will further illustrate research literature related to the scope of my investigation and research design.

Problem Statement

Historically a disease among adults, T2DM is now also a disease in persons younger than 20 years. The emergence of T2DM among youth has been noted since studies by Rosenbloom, Young, Joe, and Winter (1999) and Pinhas-Hamel and Zeitler (2005). The prevalence of T2DM in youth (persons younger than 20 years) has continued to increase in the last 2 decades, rising to epidemic proportions and changing the global face of this chronic disease that once carried a clinical diagnosis code defined as adult-onset diabetes (Dabelea et al., 2007; WHO, 2016). The international classification of disease (ICD) diagnosis code for adult-onset diabetes is now history as T2DM is now not exclusive to adult populations.

The global epidemic of T2DM in youth has forced scholars involved in diabetes research to look at not only lifestyle and genetics as factors for T2DM incidence and prevalence, but also potential environmental exposures to dioxins, dioxin-like compounds, PCBs, arsenic, and other environmental EDCs and chemical mixtures (Lee et al., 2014; Liu, Ying, Harkema, Sun, & Sanjay, 2013; Navas-Acien et al., 2008).

Environmental measures of exposure such as ambient temperature, air pollutants, and microbial contaminants of water supply or soil have long been studied for their relationship to chronic diseases. Many researchers involved in investigating the relationship between environmental exposures and chronic diseases have used adult populations (Codru et al., 2007; Lind et al, 2012; Sergeev & Carpenter, 2011). Other investigators have focused on the association of paternal or maternal occupational or environmental exposures and outcomes of reproductive, fetal, and child health (Harari et al., 2010; Lawson et al., 2004). Thus, environmental exposures to chronic disease association type of studies are not a new phenomenon in epidemiology.

I assert that my study was the first investigational inquiry that examined a relationship between T2DM in a U.S. population inclusive of youths and adults and residential proximity to HWS. Specifically, I investigated the potential association between residential proximity to specific HWS in New Jersey, Pennsylvania, New York, and California, and the abnormal glycohemoglobin (A1c) of National Health and Nutrition Examination Survey (NHANES)-surveyed respondents ages 12 years and older residing within counties of New Jersey, Pennsylvania, New York, and California that were included in the Continuous NHANES surveys between 2005 and 2012. My study includes HWS registered on the National Priorities List (NPL), which includes sites under the federal Superfund cleanup program. The NPL contains HWS which may have been involved in the generation, management, and final disposition of hazardous substances, some of which may be registered as endocrine disrupting materials, potential EDCs, or POPs (Environmental Protection Agency [EPA], 2016; Gore et al., 2015). For

my study, *abnormal glycohemoglobin* is defined as an A1c ≥ 5.7 % and is inclusive of prediabetes and T2DM lab value diagnoses (National Institute of Diabetes and Digestive and Kidney Diseases, 2018).

Purpose of the Study

In this quantitative study, I aimed to examine the relationship between residential proximity to HWS in select counties of New Jersey, Pennsylvania, New York, and California and the abnormal A1c% of subjects ages 12 years and older residing within the same county of the HWS. According to the EPA's NPL (https://www.epa.gov/superfund/national-priorities-list-npl-sites-state), New Jersey, Pennsylvania, New York, and California contained the largest number of HWS in the United States in 2016. As of 2016 November, the EPA's reported number of HWS in these states were: New Jersey, 114; Pennsylvania, 95; New York, 86; and California, 98. A review of the site listings for each state identified NPL sites registered as early as 1983 and within the NHANES survey period of 2005-2012. Specifically, I investigated the relationship between residing less than or equal to 1 mile compared to residing greater than 1 mile from a HWS within the county of residency and the A1c values of ≥ 5.7 % for Continuous NHANES survey respondents ages 12 years and older during the 2005-2012 survey cycles. The counties and HWS included in my study were based on the specific counties from which NHANES selected their participants during the survey years 2005-2006, 2007-2008, 2009-2010, and 2011-2012. Due to disclosure risk (per Centers for Disease Control [CDC] research guidelines), data analysis was presented in aggregate (i.e., subjects ages 12 years and older). However, *youths* were defined in this study as

persons 12 to 20 years of age. I selected this age range of 12 to 20 years because the diagnosis of T2DM younger than 12 years of age is rare, and I wanted to be inclusive of the span of adolescent ages commonly considered by medical professionals when referencing this age group (Dabelea et al., 2007).

I centered my investigation around an independent variable, a dependent variable, and four confounding or moderating variables. The independent variable for this inquiry was remote environmental residential exposure (which I explain later in the definitions section and again in Chapter 3). The dependent variable was abnormal glucose metabolism, defined as a glycated hemoglobin or hemoglobin A1c level of 5.7 to 6.4 percent (prediabetes), and A1c greater than or equal to 6.5% (T2DM). The moderating (interaction) variables were *overweight* and *obesity* defined by the BMI for age percentiles in boys and girls 2 to 20 years of age and a BMI $\geq 25 kg/m^2$ for the study population ≥ 21 years of age, chronological age, sex, and race/ethnicity. In Chapter 3, I further detail the definitions and classifications of the independent, dependent, and moderating variables.

Research Questions

Research Question (RQ1)

Within select counties of New Jersey, Pennsylvania, New York, and California, as an aggregate, are there significant differences in the abnormal A1c% value (defined in my study as a glycohemoglobin or A1c $\geq 5.7\%$) between the study subjects ages 12 years and older residing within 1 mile of a HWS compared to the same population residing greater than 1 mile from a HWS within their county of residency?

H_{01}: For the study population ages 12 years and older there will be no significant difference in the A1c % between residential groups.

H_{a1}: For the study population ages 12 years and older there will be a significant difference in the A1c % between residential groups.

Independent variable (categorical): Residential groups, as an aggregate residing within NHANES surveyed counties of New Jersey, Pennsylvania, New York, and California, that contained one or more HWS within the 2005-2012 NHANES survey cycles. Defined as residence ≤ 1 mile from the HWS (remote exposure) or residence > 1 mile from the HWS (no exposure).

Dependent variable (continuous): A1c ≥ 5.7%.

Descriptive Statistics

The Disclosure Manual of the National Center for Health Statistics (NCHS) Research Data Center (https://www.cdc.gov/rdc/data/b4/disclosuremanual.pdf) provides several rules/guidelines to researchers for preventing disclosure. Some of these rules, relevant to my investigation, included:

- Any procedure that produces output on an individual or institution must be removed.
- Extreme values or values representing an individual must be removed. Examples include minima, maxima, medians, and modes.
- All cells with a frequency less than 5 are asterisked or removed from reviewed outputs.

Given these study limitations, only aggregate data (i.e., data from combined states) were permissible for publication and descriptive statistics or frequency tables on the continuous NHANES variables in the study dataset were limited. Thus, I present categorical descriptive statistics in the results tables of Chapter 4.

Inferential statistical test for analysis

I originally proposed using the independent samples t test. However, given the non-normal distribution of the dependent variable, I applied the Mann-Whitney *U* test to respond to RQ1 (see Evaluation of Statistical Assumptions in Chapter 4).

Variables Related to RQ1

For my study, I selected participant data from the Continuous NHANES survey years of 2005-2006, 2007-2008, 2009-2010, and 2011-2012. Specifically, the NHANES public use demographic file contained the study variables for age, sex, race, and ethnicity, as well as the pregnancy status at time of the NHANES clinical exam. The laboratory data file glycohemoglobin, also publicly available, contained the study variable hemoglobin A1c recorded at the time of the survey interview and clinical examination. The NHANES geocoding file provided, under restricted data access, the residential geocode (latitude and longitude coordinates) matched to each study participant. I provide further discussion on the utility of the NHANES data under Nature of the Study and in Chapter 3.

Geocodes are numerical values providing latitude (north–south directional coordinates) and longitude (east–west directional coordinates). I used the geocodes of the HWS and the residential geocode of the participants to calculate spatial distance between

the residence and the hazardous waste site(s) within the same county as the study subject. The proximity to HWS was categorized as ≤ 1 mile and > 1 mile residential distance from the HWS, formulating two aggregated groups. That is, I placed all study subjects regardless from state (New Jersey, Pennsylvania, New York, and California) categorized as residing ≤ 1 mile from a hazardous wastes site within the same county of residency as the study subjects, into one group and I placed the aggregate of study subjects residing > 1 mile from a within county HWS in a separate group. Each subject within each group had an observation for the scaled dependent variable.

I originally proposed the use of several web portals to identify HWS in New Jersey, Pennsylvania, New York, and California (specifically, the ATSDR Hazardous Substance Releases and Health Effects Database, the Toxic Substance Portal and Envirofacts web sites, and the Resource Conservation and Recovery Act Biennial Report for years 2005-2013) in addition to the NPL of Superfund sites. However, the NPL provided sufficient detail on each HWS that included the contaminants of concern on the site, geocoordinates, NPL registration date, geographic coverage of the HWS and site cleanup efforts. The ATSDR Hazardous Substance Releases and Health Effects Database however, were used to gain background information on the health effects of contaminants discussed in the narratives of the NPL Superfund sites. I originally proposed using the U.S. Census Geocorder as the tool to geocode the HWS. However, I discovered that geocoordinates were available for each HWS on the NPL. Therefore, it was not necessary for me to use the U.S. Census Geocorder. The Research Data Center (RDC) analyst calculated the distance between the HWS geocode and the participant residential

geocode, yielding a scaled residential proximity measurement that I categorized into two groups for use in my study: residence ≤ 1 mile from a within county HWS (remote exposure) or residence > 1 mile from a within county HWS (no exposure). In Appendix A, I provide a detailed explanation of how the remote exposure variable was created.

Research Question (RQ2)

For the study subjects ages 12 years and older residing within select counties of New Jersey, Pennsylvania, New York, and California, as an aggregate, what is the effect of the moderators, abnormal BMI, age, sex, and race/ethnicity, on the relationship between residential proximity to a HWS within the county of residency (categorized as residing ≤ 1 mile or residing > 1 mile from a HWS) and abnormal A1c% (i.e., A1c ≥ 5.7%)?

H_{02}: For the study population ages 12 years and older, the moderators abnormal BMI, age, sex, and race/ethnicity will have no significant effect on the relationship between residential proximity to a HWS and abnormal A1c%.

H_{a2}: For the study population ages 12 years and older, the moderators abnormal BMI, age, sex, and race/ethnicity will have a significant effect on the relationship between residential proximity to a HWS and abnormal A1c%.

Independent Variable/Predictor (categorical): Residential groups. Defined as residence ≤ 1 mile from a hazardous waste site (remote exposure) or residence > 1 mile from the HWS (no exposure).

Dependent Variable (categorical): Abnormal A1c value of ≥ 5.7%. Categorized as low abnormal A1c (<6.10%) and high abnormal A1c (≥ 6.10%).

Moderating Variables (MV) (categorical)

MV_1 Abnormal BMI: Categorized as abnormal BMI low (28.94 kg/m^2 or less) and Abnormal BMI high ($\geq$ 28.95 kg/m^2).

MV_2 Age of subjects: Categorized as low age < 58 years old and high age ($\geq$ 58 years).

MV_3 Sex of subjects: Male or Female.

MV_4 Race and Ethnicity as per NHANES survey data. These groups were categorized as: Mexican Americans, other Hispanics, non-Hispanic Whites, non-Hispanic Blacks, and other non-Hispanic race including non-Hispanic Multiracial.

Descriptive Statistics

The study limitations regarding disclosure risks and publication of descriptive statistics for RQ2 are the same as that discussed for RQ1. Thus, I present categorical descriptive statistics in the results tables of Chapter 4.

Inferential statistical test for analysis

Binary logistic regression with moderator interactions.

Variables Related to RQ2

In addition to the geocoding file, demographic file, and glycohemoglobin data files of NHANES, the body measures file from the NHANES public data survey years 2005-2012, provided the BMI, height, and weight of each participant. Children grow in height and weight. Therefore, the BMI reference standards for children 2 to 20 years of age is represented by percentiles within the growth charts for boys and girls (Figures 1 and 2). In children between the ages of 2 to 20 years, normal weight is characterized by a

BMI between the fifth and < 85th percentile for age and sex. A BMI between the > 85th and 95th percentile for age and sex is characterized as overweight. Obesity in this age group is characterized by a BMI ≥ 95th percentile for age and sex. The research literature identifies overweight (BMI 25.0-29.9) and obesity (BMI ≥ 30) as a risk factors for T2DM in adult and children. As children develop adjusting in both height and weight, the 85th and 95th percentiles for BMI approximate the 25 and 30 kg/m^2 thresholds for overweight and obesity in adults (Baker et al., 2005). In my sample population, the BMI of subjects 12 to 20 years of age was matched to the represented scale for age and sex on the CDC BMI for age percentiles charts (https://www.cdc.gov/healthyweight/assessing/BMI/childrens_BMI/about_childrens_BMI.html) for U.S. boys and girls ages 2 to 20 years (see Figures 1 and 2).

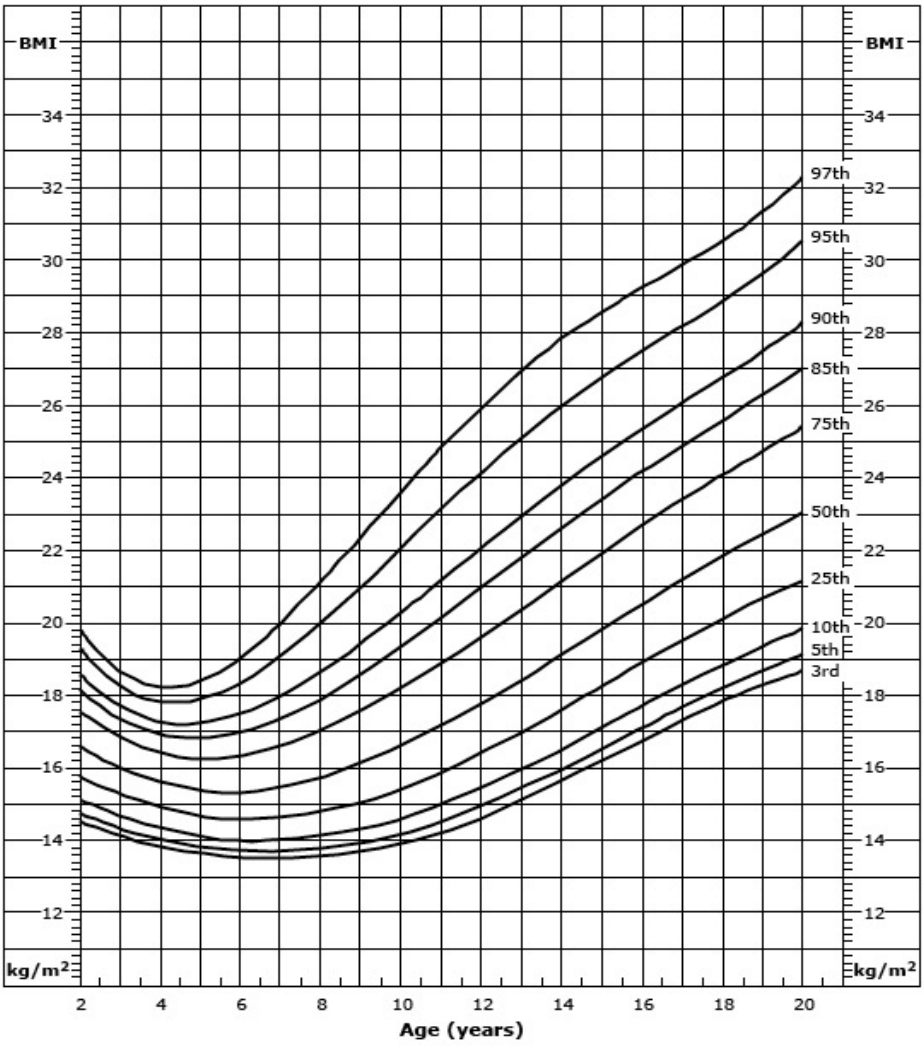

Figure 1. BMI percentiles for boys ages 2 to 20 years.

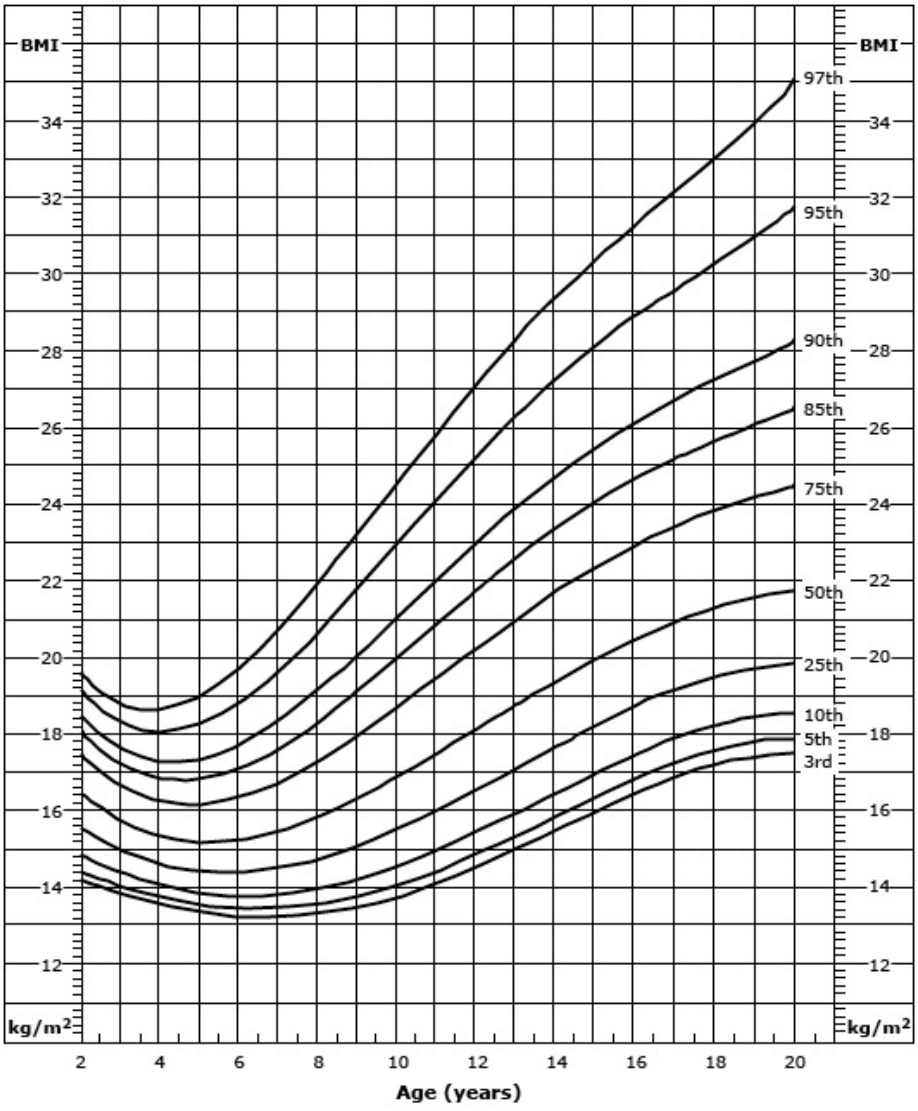

Figure 2. BMI percentiles for girls ages 2 to 20 years.

In Chapter 3, I detail the sample size, error of probability, effect size, and power for statistical analysis to answer each research question and to test my hypotheses.

Theoretical Foundation

The ecosocial theory of disease distribution and Hill's criteria of causation provided the theoretical framework for my study. The ecosocial theory of disease distribution was selected because its construct of embodiment of disease and core proposition of determining patterns of disease distribution align with the pathway of exposure investigated in my study, the environment or HWS exposure (Krieger, 2012). Hill's criteria of causation were selected because the biological plausibility criteria aligned with the proposed biological plausibility of elevated or abnormal A1c values (suggestive of pre-diabetes and T2DM) from human exposure to EDCs and POPs (Gore et al., 2015; Hill, 1965). In Chapter 2, I provide additional information supporting the rationale for utilizing these approaches in my study.

Nature of the Study

I approached this quantitative investigation using and ecological correlation study design and simple moderation analysis. This study design enabled me to compare two groups of type 2 diabetics (aka persons meeting the A1c diagnoses of pre-diabetes and T2DM) ages 12 years and older categorized according to their residential proximity to a within county HWS, where near proximity was defined as residing ≤ 1 mile from the HWS. Additionally, this design enabled me to investigate the relationship between residential proximity to a HWS and abnormal A1c and compare mean group differences in abnormal A1c values between two independent residential groups. The key study variables were residential proximity to HWS or remote environmental exposure (independent variable), abnormal glucose metabolism measurement defined as an A1c ≥

5.7% (dependent variable), and the moderating variables: overweight and obesity defined by the abnormal BMI for age percentiles in boys and girls 2 to 20 years of age and BMI ≥ 25kg/m2 for subjects age ≥ 21 years, chronological age, sex, and race/ethnicity.

Demographic Data

The unit of analysis was aggregated subpopulations within residential geocodes. Data on the study population (persons ages 12 years and older) was collected in datasets from the Continuous NHANES, which is sponsored by the United States NCHS (see the section, Variables Related to RQ1 and RQ2 in Chapter 1). The Continuous NHANES study program has been in existence since 1999 and collects interview, laboratory and examination data on a representative sample of the United States annually; approximately "5, 000 persons each year" and revisits "15 counties" each year (CDC/NCHS, 2016). For my study, Continuous NHANES data for years 2005-2012 (specifically the geocoding file, demographic, body measures and glycohemoglobin files) were used and included the following components: age, gender, race and ethnicity, body measurements (height, weight, BMI), pregnancy status at time of exam, hemoglobin A1c, and residential geocodes. Public data included all except residential geocodes. The process for obtaining residential geocode data of participants is detailed in Chapter 3.

Exposure Assessment

According to the NPL (https://www.epa.gov/superfund/national-priorities-list-npl-sites-state), as of November 2016 the states of New Jersey, Pennsylvania, New York and California, were the top four states in the country with the most HWS and known releases of environmental contaminants under surveillance by the EPA. A review of these states

on the NPL website identified HWS in existence prior to and during the NHANES survey years of 2005-2012 (the years from which my study population was drawn). Under my instructions, distance based calculations using the residential geocodes of the study subjects and the geocodes of the HWS facilities were performed by the RDC analyst to assign residential proximity or remote exposure and create the two independent residential groups. I provide details in Chapter 4 and Appendix A, about how the remote exposure variable was created.

I selected this ecological study design based on my review of previous research designs by Kouznetsova et al. (2006), Liese et al. (2010), and X. Liu et al. (2012), as well as a general review of the literature regarding residential proximity to environmental hazards and adverse health outcomes, and human exposure to POPs, EDCs and risk of diabetes. Kouznetsova and X. Liu, concluded an increased rate of hospitalization for diabetes and an increased rate of hospitalization for respiratory disease respectively based on association between residential proximity to HWS using residential ZIP codes of the study participants and the HWS. However, the use of ZIP codes alone has been scrutinized by other researchers as being too large of a unit for spatial aggregation without inflicting bias. Liese and colleagues (2010), minimized this ecological bias in their study on geographic variation in Type 1 and T2DM by using geocoding which narrowed their study cases to the street and census tract level rather than the common five or nine-digit ZIP code.

I proposed that collectively, my use of an ecological study design incorporating geocodes, coupled with: the growing body of evidence suggesting an association between

human exposure to environmental contaminants and diabetes, the unknown knowledge in the scientific community about low-dose exposure to POPs and health risks, and prior studies investigating the time between human exposure to POPs and the clinical presentation of chronic disease, enabled me to generate hypotheses to investigate the potential influence of HWS exposure in humans and the T2DM disease burden within a study sample inclusive of youths and adults. I also chose this research design for its analytical approach, cost (financial as well as human resources), time frame for completion, and its acceptance in epidemiologic research.

Definitions

Environmental residential exposure (remote exposure)/residential proximity to HWS (independent variable). This is the grouped classification and labeling of the study participant's residential proximity to a HWS. Using the geocoordinates of the participant's residence and the geocoordinates of the HWS located in the same county and closest to the participant's residence (within the states of New Jersey, Pennsylvania, New York, or California), under my direction, the distance between residential and HWS geocoordinates was calculated by the RDC analyst using the GEODIST function in SAS (see Chapter 4 and Appendix A). Study subjects residing ≤ 1 mile from a HWS were considered as being in near proximity and thus environmentally exposed (or having remote exposure) to the HWS. Participants residing > 1 mile from the HWS were considered as having no residential exposure.

Abnormal glucose metabolism (dependent variable). The clinical diagnosis of T2DM or pre-diabetes (National Diabetes Information Clearinghouse, 2012; Perreault &

Faerch, 2014). I proposed that I would use either fasting blood glucose (FBS), hemoglobin A1c or oral glucose tolerance (OGTT) to select study participants as either of these measures may be used to diagnose T2DM. However, given that there was a sufficient sample size of NHANES respondents with A1c values, A1c was the laboratory measurement I used to identify study subjects (see Chapters 3 and 4). For my study, serum hemoglobin A1c values of ≥ 6.5% (diabetic) and 5.7 to 6.4 percent (pre-diabetic) were used to identify study subjects from the NHANES public data files.

Hemoglobin A1c. Serum glycated hemoglobin measurement used in the diagnosis of T2DM, forecasting one's risk for the development of T2DM (i.e. pre-diabetes), and management of T2DM (Morris et al., 2013; Santaguida et al., 2005).

Assumptions

For this study, I assumed that the NHANES survey respondents supplied the NHANES interviewer with accurate information about years lived at current residence, country of origin, pregnancy status and history of T2DM (if diagnosis was *participant reported* only and not confirmed by a medical provider or diagnostic testing). I also assumed accuracy in height and weight of the participants if self-reported and not measured by the NHANES examiner. However, on my review of the secondary data, all observations for body measures (i.e., height, weight, BMI) and laboratory values (A1c%) for the study sample, were met under the NHANES medical examination.

In approaching this investigation, I assumed that HWS on the NPL list of Superfund sites (simply due to the registration of these sites on the NPL), contained known and unknown (not yet investigated) hazardous chemicals and chemical

compounds with potential endocrine altering properties to adversely affect human health and effect the burden of T2DM in the United States. I also assumed that some of these HWS had a history of toxic releases with potential human exposure (based on my review of the HWS narratives and records of site clean-up activities). I also assumed that study subjects would not live directly on a HWS.

Scope and Delimitations

Of the more than 300 million people worldwide with diabetes, the WHO reported that ninety percent of these people have T2DM (WHO 2016). And, while historically a disease of adults, the WHO also reported in 2013 and 2016 the increased occurrence of T2DM in children. Established in the United States in 2001, the CHE described five tenets forming the structure of its network: "1) We have an epidemic of chronic disease in this country; 2) 70% of these diseases are preventable; 3) Investing more resources into prevention rather than just treatment is not only prudent, but critical for a healthy, more equitable society; 4) Identifying and reducing environmental toxicants related to these diseases is essential; 5) Taking precautionary action must be an integral part of any plan to broadly improve health" (https://www.healthandenvironment.org/about/a-brief-history/founding-of-che).

Since its inception, the CHE has summarized the literature on environmental influences of T2DM and has formulated a toxicant and disease database of ten environmental contaminants with varying strength of association (strong, good, & limited) to T2DM (CHE 2014). Therefore, because of the global prevalence of T2DM in persons younger than 20 years, the literary works on environmental influences of T2DM,

and considering the uncertainty about the full effects of environmentally low-levels of dioxin exposure on chronic disease, I chose to include in my study focus a representative U.S. population sample ages 12 years and older (Codru et al., 2007).

Although Navas-Acien et al. (2008), investigated the association of T2DM using serum concentration measurements of POPs and volatile toxicants from the NHANES 2003-2004 participant data, this concept was not explored in my study. My rationale for excluding this type of data was because NHANES data for survey years 2005-2012 did not collect data on serum concentrations of POPs and volatile compounds every year. And, that which was collected included a 1/4 to 1/3 subsample of the study population, which was not always inclusive of participants between the ages of 12-19 years. However, based on information contained in the NHANES 1999-2016 *Survey Content Brochure,* BMI, hemoglobin A1c, FBG, and OGTT were collected on participants during the Continuous NHANES survey periods 2005-2012 (CDC, 2016a). Further, as addressed earlier, there is precedence for studies on residential HWS proximity and association of diabetes (Kouznetsova et al., 2006), which strengthened the external validity of my investigation.

Limitations

I selected an ecologic study design for its analytical approach, cost (financial as well as human resources), time frame for completion, and its acceptance in the field of epidemiological research. I was keenly aware of the disadvantages within this research model. A major limitation is the ecologic fallacy. The ecological fallacy involves making

erroneous inferences from aggregate research observations towards persons at the individual level (Trochim & Donnelly, 2007, p. 13). I projected that using geocodes and therefore fine-tuning the relationship between HWS and the study population would help to minimize bias. I investigated an exposure-disease relationship, where residential proximity to a HWS was the independent variable (remote exposure) and abnormal glucose metabolism (i.e. laboratory diagnosis of pre-diabetes and T2DM) was the dependent variable (disease outcome).

The Continuous NHANES archival data that I used for my study (years 2005 – 2012, public use glycohemoglobin file) provided the results for hemoglobin A1c% collected from serum samples of the study population (CDC, 2016a). This allowed me to select persons with pre-diabetes and T2DM who fit my study population (based on diabetes mellitus diagnostic criteria which I explain in Chapter 3). However, the exposure assessment for the study was based on the study participant's residential proximity to EPA assessed HWS in New Jersey, Pennsylvania, New York, and California. I defined residential proximity to a HWS using geocoding and distance based analysis/calculation. This process created a finer remote distance variable in comparison to using the 5-digit ZIP code. Liese et al. (2010), used geocoding in their study on geographic variation in Type 1 and T2DM and concluded that analysis using geocoding and census tract rather than ZIP code level data, was the most unbiased type of analysis.

The second limitation was that the exposure assessment was not measured independently, but based on residential proximity to a within county HWS. I did not test the serum of study participants for environmental contaminants such as TCDD or PCBs.

Thus, during my investigation of a relationship between hemoglobin A1c levels (of the study population, ages 12 years and older) and residential proximity to HWS, I was cautious not to assert a positive or negative correlation at the individual level based on data where the exposure assessment is aggregated and not measured for each participant independently as was the outcome measure, A1c.

Thirdly, my use of archival data had the potential of posing limitations for me to formulate research questions based on data availability. This limitation did arise. I originally assumed the investigation and comparison of groups within the same county (which could potentially have meant the comparison of eight or more groups in this investigation). For example, comparing the A1c outcome of two residential groups from Middlesex county in New Jersey (those residing ≤ 1mile of a HWS compared to those residing > 1 mile of a HWS). However, due to the risk of disclosing the precise geographic location of NHANES respondents as well as the risk of too small a cell sample for analysis, the name of the county that each NHANES respondent lived in was not released to me by the RDC and categorical residential groups were only considered for release in aggregate. That is, regardless of the state of residency, all study subjects categorized as residing ≤ 1mile of a HWS were sorted into one group and study subjects residing > 1 mile of a HWS within their county of residency were sorted into another group. While this limited how I framed the research questions for this study it did not alter the foundation of the research questions based on fulfillment of the knowledge gap in the literature (see Chapters 4 and 5).

Walden IRB notification of limitations impacting study. I was also limited in publishing study analysis that included descriptive statistics at the county or state level as well as limited in frequency or descriptive statistics on cells with 5 or less observations. These limitations were in accordance with the NCHS RDC Disclosure Manual (https://www.cdc.gov/rdc/data/b4/disclosuremanual.pdf). Although this limitation affected the display of results it did not hinder the meaningfulness of the study. I submitted to the Walden IRB a change form notifying of limitations which resulted in modifications of two of my research questions and elimination of another (from my original three research questions presented in the Walden proposal). These changes to my study were approved and once approved I commenced data collection at the RDC.

In addition to the above, I was aware that archival data might include missing, incomplete, or compromised data (Rudestam & Newton, 2007) and I charged myself with the full responsibility in research design to address this issue (see Chapters 4 & 5). Lastly, after reviewing the guidelines of the NCHS regarding the use of public archival data and the process for acquiring the use of restricted archival data (*http://www.cdc.gov/rdc/B3Prosal/PP300.htm*), I followed these guidelines precisely. Although the merged public and restricted datasets from my study must be retained in the NCHS Research Data Center (RDC) and while my methods, results and discussions chapters must be reviewed by an RDC analyst prior to publication (to assert that no disclosure risk exists), the CDC, RDC or NCHS do not assert authorship or control over my dissertation or publications that may extend from my dissertation.

Significance of the Study

Dioxins, of which TCDD is the most potent compound, remain in the environment as byproducts of industrial processes such as hospital waste incineration, chemical synthesis of certain compounds, and pulp from paper mills (Kouznetsova et al., 2006) and have the potential to adversely affect the health of animals and humans via contaminated waterways, soil and food. Likewise, PCBs, although banned from industrial use in the 1970s, persists in the environment and bioaccumulate and biomagnify (Codru et al., 2007) in the food supply of animals and humans. Cranmer et al. (2000) demonstrated an association between dioxin exposure and diabetes in the investigation of Vietnam Veterans exposed to Agent Orange. Codru et al. (2007) has also suggested an association between impaired glucose tolerance and PCBs and chlorinated pesticide contaminated air, water, and soil that enters the food chain. My review of the literature concerning the existence of POPs in today's environment coupled with my exploration of studies regarding: (1) the latency of TCDD and PCBs elimination from adipose tissue (Aylward & Hays, 2002; Stockholm Convention, 2016c); (2) the accumulation of POPs in serum lipids (Aylward et al., 2005); (3) the diabetogenic effects of POPs and EDCs and (4) the uncertainty about the full effects of low levels of dioxin exposure on chronic disease, helped form my foundation for this dissertation.

My investigation filled a knowledge gap left by Kouznetsova et al. (2006) concerning the potential association between residential proximity to HWS and T2DM, with a focus on a U.S. population inclusive of youths and adults. And, as Martuzzi, Mitis and Forastier (2010), and Mohai, Lantz, Morenoff, House and Mero (2009) demonstrated

significant disparities in human exposures to environmental burdens, the outcomes of this study may promote positive social change through the channels of environmental justice and further investigations about safe environmental limits of toxic endocrine disruptors.

Summary

During the latter part of the 20th century, T2DM, once considered a disease of adult populations has become a global concern in children's health. The WHO has reported that 90% of the more than 300 million people worldwide with diabetes are type 2 diabetics, and that an increasing number of type 2 diabetics globally are persons younger than 20 years (WHO, 2016). Though lifestyle risk factors such as obesity, an unbalanced diet, and lack of exercise are reported as significant contributors to the incidents of T2DM in the United States, the addition of potential human exposure to POPs and other hazardous environmental compounds with endocrine disrupting capabilities to the list of risk factors for T2DM has continued to be explored for more than a decade (Bertazzi et al., 2001; CHE 2014; Codru et al., 2007; Kouznetsova et al., 2006; Longnecker & Daniels, 2001; Navas-Acien et al., 2008; Remillard & Bunce, 2002; Sergeev & Carpenter, 2011; Tseng et al., 2002; WHO 2013).

As the full effects of low levels of dioxin exposure remains uncertain the knowledge gap on residential proximity to HWS and T2DM in persons 12 years of age and older, remains an area of exploration (Aylward & Hays, 2002; Codru et al., 2007). Such a knowledge gap in environmental health and T2DM of youths and adults was the foundation of my dissertation and research design. In Chapter 2, I discuss the literary

search strategy I used in this investigation and provide a review of the literature that helped shape the theoretical foundation, scope, and research design for my dissertation.

Chapter 2: Literature Review

Introduction

The curiosity of researchers about environmental influences related to adverse health outcomes charges explorations and investigations into environmental epidemiology with regard to chronic disease risks. For several decades, there have been concerns about the effects of environmental influences on the incidence and prevalence of T2DM. In addition to risk factors such as obesity and genetics, there has been an increased concern in the past 25 years about human exposure to endocrine disrupting environmental chemicals and chemical mixtures and the influence of these contaminants to alter (or disrupt) the endocrine systems of humans, thereby affecting the prevalence of chronic endocrine conditions such as diabetes (Lee et al., 2010).

POPs, of which some are EDCs, are widespread in the environment in low concentrations and remain in the environment years after their initial use (such as DDT pesticide spraying during the Vietnam War) or as byproducts of industrial processes (Bijlsma & Cohen, 2016). As aforementioned, sufficient animal and human studies have demonstrated the risk of diabetes from long-term human exposure to POPs and EDCs, such as dioxins, dioxin-like compounds, PCBs, and arsenic (Lee et al., 2014). Hazardous waste site programs at the state and federal level are human efforts to maintain control over environmental contaminants and protect communities directly and indirectly (human food chain: fish, livestock, etc.) from contaminated water and soil. However, environmental EDCs are continuously being discovered, giving rise to potentially more human exposure. In addition, there remains an unknown in the scientific literature about

what is considered the lowest concentration of threat from these environmental contaminants on human disease (Liu et al., 2013; Stockholm Convention, 2016b). This then raises the question about the vulnerability of any human endocrine system along the lifespan from prebirth, childhood, adolescence, or adulthood when exposed to these contaminants.

Once historically a disease of adult populations, T2DM is now diagnosed in populations younger than 20 years. Obesity plays a role as a risk factor in T2DM for children and adults. However, it is not farfetched to consider the risk of childhood as well as adult exposure to environmental EDCs and POPs as a potential risk factor for T2DM in adolescents given the ubiquitous nature of these contaminants and the unknown effects from low concentration exposure on human health. Further, although HWS programs are efforts to protect humans and wildlife from potentially threatening environmental contaminants, 100% containment of hazardous waste toxic releases is not guaranteed 100% of the time. Therefore, my aim in this study was to investigate the relationship between residential proximity to EPA reported HWS on the NPL list of Superfund Sites in NHANES surveyed counties of New Jersey, Pennsylvania, New York, and California (during the period 2005-2012) and the abnormal glucose metabolism measurements (defined in my study as an A1c $\geq$ 5.7%) among persons ages 12 years and older living within those surveyed counties.

Chapter 2 includes the literature search strategy that I used to investigate prior research related to my problem statement and study purpose. In this chapter, I summarize studies that provided the foundation for my investigation of HWS proximity and T2DM

in youths and adults. Using a theme approach, in Chapter 2, I review, analyze, and synthesize major theoretical and methodology approaches that were applicable to my investigation and provided a rationale for selection of theory choice. In this association study, I use an ecological study design for data analysis. I provide in Chapter 2 a justification for selection of study variables and research design. I conclude the chapter with a summarization of what is known and what is not known about HWS exposure and its potential for T2DM in humans. I also address how my investigation filled the literature gap related to environmental risks and T2DM, extends the knowledge of children's environmental health, and extends inquiries related to environmental justice and whose backyard may bear the burden of HWS.

Literature Search Strategy

I used the Walden University library network, Google Scholar, and Bing search engines to retrieve literary work on the research topic. Other search engines included government and peer-reviewed websites and organizations, predominantly the CDC, NCHS, National Institutes of Health (NIH), Human Genome Epidemiology Network (HuGE Net), WHO, United States EPA, Agency for Toxic Substances and Disease Registry (ATSDR), National Institute of Environmental Health Sciences (NIEHS), U.S. Department of Health & Human Services (DHHS), UpToDate, and others as listed within the references. Databases used included Academic Search Complete, CINAHL, PubMed, Medline, NHANES, CHE Toxicant and Disease Database, Scorecard, the Inter-University Consortium for Political and Social Research (IUCPSR), and others as listed within the references. I searched publications from 1975 to 2016.

Search Categories and Terms

I began the process of gathering information on the study topic by creating broad categories. From these categories, key search terms emerged. Search categories and key terms related to toxins and HWS included *dioxin(s)*, *TCDD*, *mechanism of dioxin toxicity*, *TCDD epidemiological studies*, *EDCs and hazardous waste sites*. Search terms related to *diabetes* included *type 2 diabetes*, *T2DM*, *etiology, clinical manifestations and diagnosis of T2DM*; *pre-diabetes*, *risk factors for T2DM*, *management of T2DM*, *environmental risk factors for diabetes*, *prevalence and incidence of T2DM in the US and globally*. Search terms specific to the study population included *T2DM in persons younger than 20 years*, *T2DM in children and adolescents,* and *diabetes prevention in adolescents and adults*.

Considering the study focus of exposure and chronic disease outcome, an exposure disease category included search terms such as *environmental exposure to toxicants and adverse health outcomes*, *POPs and disease risk*, *endocrine disrupting chemical exposure and disease*, *dioxin and diabe*tes, *insulin resistance and dioxin*, *human dioxin exposure*, *dioxin toxicity*, *children and dioxin exposure*, *hyperinsulinemia and TCDD*, *glucose tolerance and dioxin*, *adiposity and dioxins*, *hazardous waste site toxic releases*, *residential proximity to hazardous waste sites and disease (and diabetes)*, *EDCs*, *TCDD and insulin resistance*, *TCDD case-control studies*, *noncancerous effects of dioxins*, *endocrine disruptors*, *measuring dioxin and endocrine disrupting chemical levels in humans*, *serum lipid dioxin*, *dioxin in urine, and testing urine for dioxins*. The category environment generated search terms such as *hazardous waste sites*, *geography of dioxins*, *geographic distribution of dioxin (toxic substance) releases*, *POPs and water (and soil)*

contamination, *contaminated waters/waterways in the United States*, *low-dose dioxin exposure*, *safe levels of dioxin*, *safe levels of TCDD*, *environmental endocrine disruptors*, *hazardous waste and endocrine disruptors*, *occupational exposure to TCDD*, *dioxin environmental limits*, *PCBs*, *POPs*, and *environmental influences of chronic diseases*.

Research Design and Secondary Dataset Discovery

In my review of research designs and analysis applications for consideration in my investigation, I applied the key search terms *ecological studies*, *social epidemiology research*, *negative binomial regression*, *Poisson regression*, *linear regression*, *test of mean differences*, *moderation analysis*, and *spatial analysis*. I searched the IUCPSR and Continuous NHANES websites for secondary datasets that included the study population (persons ≥ 12 years with A1c lab values meeting the study definition), and contained the primary independent and dependent variables of my study. My search terms included *age*, *education level*, *diabetes diagnosis*, *hemoglobin A1c reports/results*, *serum blood glucose, impaired blood glucose*, *health and medication history*, *residential ZIP code*, *residential geocoordinates*, *race/ethnicity*, *gender*, *BMI*, *toxicant exposure history*, and *toxic chemical level (urine or serum)*. Only the Continuous NHANES study reports for survey cycles between 2005 and 2012 were useful from this search.

Dissertation Reviews

I searched the Walden University dissertation database and the global dissertation database through Proquest. I limited my dissertation reviews to published dissertations from 2001 to 2016. Search terms for archived dissertations included *EDCs*, *hazardous waste sites and endocrine disease*, *proximity to hazardous waste sites and*

disease/adverse health outcomes/endocrine disease, *environmental toxic releases and disease*, and *hazardous waste site and diabetes*. From this search, I selected five Walden university dissertations and five global dissertations with significant relevance to my dissertation topic and research design. I list dissertations relevant to my study topic in Appendix B and Appendix C.

Scope of Literature Review

To provide the background and a historical perspective which demonstrated decades of exploration on the association of human POPs, EDCs and HWS exposure and adverse health outcomes such as diabetes, I reviewed studies from 1975 to 2016. Researchers of earlier animal studies concluded the cancerous and non-cancerous effects of POPs such as TCDD. These early studies influenced researchers to develop human population-based and case-control studies, where researchers have through the years concluded outcomes such as cancer, pre-birth neurological defects, neurological disease, and cardiovascular disease related to human exposure to environmental contaminants (Hall et al., 1975; Lawson et al., 2004). In other studies, researchers have concluded outcomes of diabetes, hyperinsulinemia, and insulin resistance based on long-term human exposure (usually described in terms of 10 or more years) to POPs either via occupational exposure or exposure from contamination of the air, water, or soil by these substances (Bertazzi et al., 1998; Codru et al., 2007; Fierens et al., 2003; Harari et al., 2010; Kramarova et al., 1998).

With respect to previous studies on human exposure to POPs and EDCs and diabetes, the storage of these compounds in human adipose tissue (as evidenced by serum

lipid levels of these compounds in some studies) is identified by the investigator(s) of these studies as a significant role in the exposure risks-to-adverse health outcome concept (Lee et al., 2006). Thus, I suggest that the premise would be that the greater the liposity, the longer the exposure to the contaminate, and the greater the concentration of the contaminant in serum or adipose tissue, the greater the likelihood for adverse health risks. However, with the widespread presence of EDCs and chemical mixtures in the environment, even at low levels, what remained unknown at the time of my investigation was the determinants of long term exposure and the full, true, and safe limits of human exposure to EDCs and POPs. These known and unknown concepts about EDCs and POPs influenced me to generate questions such as: Is long-term exposure to POPs a measurement of months or years? Does one month or less than 10 years pass the test for long-term exposure? Are safe limits of environmental exposure measured by contaminant parts per million (or other denominator), and does this vary with chronological age? What role might other factors such as adiposity, gender, or age at time of initial exposure play in adverse health outcomes? Therefore, I concluded that it was prudent to extend investigations related to HWS proximity and T2DM to populations younger than 20 years to meaningfully explore the potential risks of environmental POPs and EDCs exposure in this younger population (X. Liu et al., 2012; Navas-Acien et al., 2008; Sergeev et al., 2011; Tseng, et al., 2002).

The studies I reviewed for my investigation demonstrated the relevance of my research to fulfill the knowledge gap concerning environmental hazardous waste exposure and chronic disease outcomes. As I embarked on this study there remained a

knowledge gap in the research literature relevant to residential proximity or remote exposure to HWS and T2DM in a study population that included pre-teens, adolescents and adults. The decades of peer-reviewed studies that I analyzed for my research spoke to the timeliness of my investigation to add to the discipline of public health and the body of knowledge concerning potential environmental risk factors for T2DM in children and adults (Codru et al., 2007; Gore et al., 2015; Kouznetsova et al., 2006; Mohai et al., 2009; Thayer, Heindel, Buchner & Gallo 2012).

Theoretical Foundation

The ecosocial theory of disease distribution and Hill's criteria of causation provided the dual theoretical framework for my study. I selected the ecosocial theory of disease distribution because its construct of embodiment of disease and core proposition of determining patterns of disease distribution aligned with the pathway of exposure being investigated in my study, the environment or HWS exposure (Krieger, 2012).

The Bradford-Hill or Hill's criteria of causation provided an additional foundation for my study. This theory incorporates the consideration of eight criteria (strength, consistency, specificity, temporal relationship, dose-response or biological gradient, biological plausibility, coherence, and experiment) for utilization by researchers as checks and balances towards association studies in search of factors related to the explanation, establishment, risks, incidence and prevalence of disease (Hill, 1965). Hill's criteria imply that hypotheses can be generated from association studies and that causal relationships may be established through repeated association studies. Within Hill's criteria, the elements strength, temporal relationship, dose-response relationship,

biological plausibility, and coherence served as the major backlight for my investigation of residential proximity to HWS and T2DM in a U.S. population of youths and adults. I brought to my study strong evidence from the literature that Hill's criteria of causation (aka Hill's considerations for causal inference) was applied in previous animal and human studies which concluded a relationship between human exposure to EDCs and POPs and adverse health effects, to include diabetes (Lee et al., 2006). EDCs and POPs are ubiquitous in the environment and have the potential to adversely influence the health of humans via contaminated soil, air, water and the human food chain (Environmental Work Group, 2011).

If EDCs are ubiquitous in the environment and if some of these chemicals are byproducts of industrial processes, then I suspected that HWS, established for the containment of toxic materials, might comprise these EDCs or POPs. Further, residents near HWS may potentially be subjected to toxic releases from HWS and potentially exposed to EDCs via contaminated air, water, soil or food supply. If then as the literature suggest, that the list of EDCs and POPs with potential to disrupt the human endocrine system continues to expand (Stockholm, 2016b) and if the impact from low levels of exposure to these chemicals remains in question, then I concluded that it was prudent to be inquisitive about the influence of low levels of exposure of EDCs and POPs on the T2DM disease burden of youths and adults residing within a 1-mile radius from a HWS. These parameters, which are a reflection of exposure risks (represented in my study as residential proximity/remote exposure) preceding an outcome (expressed in my study by the dependent variable A1c $\geq$ 5.7%) and biological plausibility (as represented by the

historical perspective of EDCs and POPs associated diseases), supported my rationale for selecting Hill's criteria of causation along with the ecosocial theory to guide my investigation of HWS proximity and T2DM in a U.S. population of adolescents and adults (Boberg, Lessner & Carpenter, 2011; Gore et al., 2015; Hill, 1965; Kouznetsova et al., 2006).

Literature Review Related to Key Variables and Concepts

Human Exposure to POPs, EDCs and Disease Outcomes

There exist in the literature several studies concluding an association between various disease risk from human exposure to EDCs/chemical mixtures such as bisphenol A (BPA), phthalates, pesticides, herbicides, and POPs such as dioxins, and dioxin metabolites, and industrial chemicals such as PCBs, and polybrominated diethyl ethers (Gore et al., 2015). The human disease burden as a result from exposure to these toxicants ranges from cancer to neurologic disorders and neurobehavioral deficits, respiratory diseases to cardiovascular diseases, and congenital disorders to endocrine disorders such as diabetes (Goncharov et al., 2010, Harari et al., 2010). The burden of disease may be expressed in human populations before birth and along the lifespan.

In 2013, researchers attending the National Toxicology Program released a summative report of 72 epidemiological studies which concluded a positive correlation between some POPs and T2DM, citing organochlorine compounds: *trans*-nonachlor, DDE, PCBs, dioxins, and dioxin-like compounds as the strongest associations to date (Taylor et al., 2013). The 72 epidemiological studies referenced in this summative report excluded populations younger than 20 years. However, POPs are ubiquitous in the

environment and the bioavailability of POPs in the food chain places populations across the life span susceptible to exposure, and thus, susceptible to potential adverse health outcomes related to exposure.

Various authorities have defined differently the maximum daily exposure limits and tolerable daily intake (TDI) limits of POPs. This left a burning question as to what minimal conditions of human exposure to POPs (i.e., POPs dose-response relationship, years of exposure to POPs) might influence the T2DM disease burden in the United States (Liem, Furst, & Rappe, 2000; Schecter et al., 2001; U.S. EPA, 2012). It was the unknown about the full adverse human health effects related to low-level exposure to POPs, the growing identification of chemicals/chemical mixtures classified as EDCs, and the ubiquitous nature of these substances that provided me with more support for my study concerning HWS proximity and T2DM in youths as well as adults.

Rationale for Chosen Methodology

The aim of my study was to investigate the relationship between the abnormal glucose metabolism measurement (i.e., A1c $\geq$ 5.7%) of persons 12 years and older who resided less than or equal to 1 mile from a HWS compared to the same population of youths and adults with A1c values $\geq$ 5.7% who resided greater than 1 mile from a HWS. Additionally, my study aimed to investigate the impact of the interaction remote exposure and abnormal BMI, age, sex, race/ethnicity on the outcome (A1c $\geq$ 5.7%). In my study, the analysis of disease outcome (A1c $\geq$ 5.7%) from exposure risk (based on residential proximity) was measured at the group level. Because the units of measure in my study were at the population or group level, I selected an ecological study design.

Ecological studies are useful in investigating causal processes at the group level and are often used in geographical epidemiological studies, such as my dissertation project. The results from ecological studies may reach valid causal relationships at the group level which may present valid causal inferences that might assist with hypotheses generation for future studies. Ecological studies are widely used and respected in epidemiology and the results from these type of studies may be helpful for public and private policy decisions at the state, county or regional levels. One of the most highlighted cautions of an ecological study, of which I remained keenly aware of, was avoidance of the ecological fallacy. The ecological fallacy implies the application of erroneous causal inferences to individuals based on the analysis of aggregate data alone (Portnov, Dubnov, & Barchana, 2007). I discuss the statistical inferences of my study results in Chapters 4 and 5.

I took great care in my planned methodological design and incorporated multiple units of analysis in a hierarchical fashion (Trochim & Donnelly, 2007, p. 13) using scaled and categorical measures. My methodological design included appropriate statistical tests to respond to each hypothesis with a sufficient study sample. In addition, my study included parameters to decrease the risk of Type I and Type II errors and included sufficient statistical power to enhance the meaningfulness of the results for each hypothesis. I maintained the reasoning of the results at the group level. My application of these measures in my study enabled me to steer-away from the ecological fallacy.

Relationship Between Residential Proximity and Disease

In general, the application of an ecological study design to investigate relationships between environmental contaminants and chronic disease risk, incidence, or prevalence is not a new phenomenon in epidemiology. Also, not new to epidemiology are investigations that study the geographic spatial relationships of disease between groups from different areas of a state or regions of a country. What has strengthened in population based studies of this type in the past few years are efforts by researchers to minimize the ecologic fallacy and improve upon variable selection and statistical analysis by making as finite as possible the geographic area from which the study subjects are identified. This means, for example, the application by researchers, of geocoordinates and census tracts rather than ZIP codes to identify study subjects. An investigator's use of ZIP codes in ecological studies is often viewed as a weaker approach (Liese et al., 2010).

There is evidence from previous studies of significant relationship between residential proximity to HWS and association with increased hospitalization rates for respiratory disease (primarily asthma) and diabetes (Kouznetsova, et al., 2006; X. Liu et al., 2012). For some studies the contents of the HWS were unknown, yet suspected due to the federal or state designation as a HWS (Boberg et al., 2011). For other studies with significant findings of cancer, cardiovascular disease and fetal defects, the contents of the HWS (such as benzene, arsenic and PCBs) may have been known though not necessarily measured (Navas-Acien et al., 2008). Thus, with respect to residential proximity as a surrogate for exposure there is precedent for conducting ecological studies based on the identification of a site as a federal or state designated HWS without a full identification

of the contents or full measurement of the contents or concentration of toxicants from the HWS.

Studies Related to the Independent Variable

I defined the *independent variable* for my study as residential proximity to a HWS (remote exposure) where proximity was a measurement between the geocoordinates of the study subject's residence and the closest HWS to the subject within the county of residency. The concept of investigating disease outcomes based on remote exposure or the distance between a study subject's residence and a HWS has been performed in other studies using ZIP code data (Kouznetsova et al., 2006; X. Liu et al., 2012; Mohai et al., 2009). In those studies, remote exposure was based on residency within a ZIP code containing or abutting a HWS. However, for my study, instead of traditional ZIP code data, residential proximity to a HWS was defined using the residential geocoordinates of the participants and the geocoordinates of the HWS. It is suggested from the literature that the application of geocoordinates enhances the strength of the remote exposure assessment (Liese et al., 2010).

Studies Related to the Dependent Variable

The dependent variable was abnormal glucose metabolism, defined in my study as a glycated hemoglobin level (also known as hemoglobin A1c, A1c, glycohemoglobin or HbA1c) of 5.7 to 6.4 percent (pre-diabetes) or ≥ 6.5% (diabetes). The A1c% is an expression of a three-month average of serum glucose concentrations and indicates the attachment of extra circulating glucose to the red blood cells. Although the glycated hemoglobin level is monitored in both type 1 (also known as juvenile diabetes) and

T2DM, the A1c in persons with T2DM is well correlated with the fasting blood glucose concentration, unlike that found in type 1 diabetes due to the greater variability of blood glucose concentrations in type 1 diabetics (Reinehr, 2013; Springer et al., 2013).

Studies Related to the Moderating Variables

As with adult populations, obesity is considered a risk factor for the development of T2DM in youth (Dabelea et al., 2007; Morris et al., 2013; WHO 2016). In addition, adipose tissue acts as a storage container for POPs (Codru et al., 2007; Merrill et al., 2013). Given this evidence in the literature it was prudent for me to consider obesity in the evaluation of study participants. As in adult populations, the incidence of T2DM in children and adolescents increases with chronological age and the disease distribution among males and females and racial/ethnic groups is similar to adult populations. Adolescent girls are nearly twice as likely as adolescent boys to develop T2DM and the disease is more prevalent in Native American, African American, Hispanics, Asian-Americans, and Pacific Islander children and adolescents (Dabela et al., 2014). Therefore, abnormal BMI, age, sex, race/ethnicity were used in my study as *moderating variables*.

Summary and Conclusion

T2DM is a condition involving several risk factors such as obesity, sex, race/ethnicity, and genetics. In the past two decades, there has been a growing concern about the role of environmental contaminants and diabetes risks. Potential human exposure to POPs and EDCs such as DDE (a derivative of the pesticide DDT), PCBs, bisphenol A (BPA), arsenic, herbicides, and other pesticides via contaminated air, water, soil, and the human food supply is a biologically plausible mechanism that threatens the

global burden of T2DM among adults and persons 12-20 years of age. These substances are ubiquitous in the environment and there is an uncertainty about the human exposure limits to POPs and EDCs that might impact diseases such as diabetes. As scientist continue to study the environment for additional POPs, my study brought awareness to the potential influence that these omnipresent chemicals and chemical mixtures might have on the environmental health of children and adults and environmental justice (Bijlsma & Cohen, 2016; Gore et al., 2015; WHO, 2016).

Previous studies investigating the relationship between residential proximity to HWS or POPs exposure and T2DM were conducted using adult populations (Kouznetsova et al., 2006; Lee et al., 2014). However, HWS exposure risk is not exclusive to adult populations and the effect of low-levels of exposure on adverse health outcomes continues to be explored. This highlighted a gap in the literature. Therefore, it was prudent to extend investigations related to HWS proximity and T2DM to populations less than 20 years to meaningfully explore the potential diabetes risks of environmental POPs and EDCs remote exposure in this younger population as well as adults. Chapter 3 details the methodology for my ecological quantitative study.

Chapter 3: Research Method

Introduction

I conducted a quantitative correlation study using independent samples *t* test and binary logistic regression moderation analysis to investigate the relationship between residential proximity to HWS in New Jersey, Pennsylvania, New York, and California and the A1c value of ≥ 5.7% in persons ages 12 years and older residing within 1 mile of a HWS compared to youths and adults of the same population type residing greater than 1 mile from a HWS within the same county. I used archival participant data from the Continuous NHANES, survey years between 2005 and 2012, (specifically the geocoding file, demographic, body measures, and glycohemoglobin files) and HWS data from the United States NPL of Superfund Sites. Most of the archival data were publicly available. Some NHANES data were available as restricted data. In Chapter 3, I provide a rationale for the research design and variable selection. In Chapter 3, I also define the target population and detail the methodology, archival data retrieval process, data organization, and the data analysis design and strategy to respond to the research questions. I address threats to validity as well as ethical considerations. The chapter concludes with a methodology summary and transition to Chapter 4.

Research Design and Rationale

This was a quantitative correlation study designed to answer questions at the group level using a representative sample of U.S. youths and adults from New Jersey, Pennsylvania, New York, and California surveyed within specific counties by NHANES during the 2005-2012 survey cycles. As such, the study and analysis were based on an

ecological design model where (a) *N* = residential groups of youths and adults (ages ≥ 12 years) with an A1c value of ≥ 5.7% residing either ≤ 1 mile from a HWS or > 1 mile from a HWS, (b) the pathway of exposure was the environment (HWS), and (c) the disease outcome (abnormal glucose metabolism expressed as pre-diabetes and T2DM) was measured at the group level. The independent variable for my study was residential proximity to the HWS/remote environmental exposure. This was a distance based measurement derived from the difference in geocoordinates between the HWS and the participant residential geocoordinates at the time of the Continuous NHANES survey. The dependent variable was abnormal glucose metabolism, expressed as an A1c ≥ 5.7%. *The moderating variables* were abnormal BMI (BMI > 85th percentile for age and sex for children and adolescents ages 2 to 20 years and BMI ≥ 25kg/m^2 for subjects > 20 years), chronological age, sex, and race/ethnicity.

Study Participant and Variable Selection Rationale

The population selection for my study was intricately related to my problem statement reflecting the growing national epidemic of T2DM in adolescents as it continues with high rates of disease among adults (Dabelea et al., 2007). Although a gap in the literature exists in identifying studies that have examined the relationship between residential proximity to HWS and T2DM in the adolescent population, there is precedent for epidemiologic studies examining diabetes and residential proximity to HWS in adult populations (Kouznetsova et al., 2006). Because obesity is considered a risk factor for T2DM in adolescent and adult populations (WHO, 2016) and POPs are lipophilic and may accumulate in human adipose tissue (Merrill et al., 2013), I used abnormal BMI in

my study as a moderating variable. The prevalence of T2DM in the U.S. population increases with age, is more prevalent among females, and is unequally distributed among racial/ethnic groups (Diabetes Report Card, 2017). I also included chronological age, sex, and race/ethnicity of participants as moderating variables in my study.

As aforementioned, studies using ZIP codes alone as a measure of remote environmental hazardous exposure have been scrutinized by other researchers as being too large of a unit for spatial aggregation without inflicting bias. Therefore, I used a smaller unit, geocoordinates, to aide in strengthening my independent variable and minimizing ecological bias. This approach was needed to advance the literature on childhood and adult remote exposure to HWS and the potential health threat from remote human exposure to EDCs and POPs that may be contained within or near HWS.

Research Design Selection

The ecosocial theory of disease distribution embraces an integrated approach to population health research. The central question for the ecosocial theory is "Who and what is responsible for population patterns of health, disease, and well-being as manifested in present, past and changing social inequalities in health?" (Krieger, 2001). I proposed that in association studies, the application of the ecosocial theory, as well as Hill's criteria of causation, enable the generation of hypotheses that integrate social and biological reasoning into the framework of the research design. It was from this vantage point that I framed my research questions and selection of study variables.

My research questions, research design, and data analysis aligned within the core constructs and core propositions of the ecosocial theory of disease distribution in that,

through my research questions, I sought to investigate a correlation between an exogenous pathway (remote HWS exposure) and disease expression (A1c ≥ 5.7%) in a population. I used data that were a representative sample of noninstitutionalized adolescents and adults of the United States, which aided in aligning my research design to the investigation of patterns of T2DM in the United States within this study population. Incorporating an ecological design for data analysis enabled (a) the comparisons of group means with respect to the dependent variable and (b) the presentation of aggregate or group-level data in regression analysis between the independent and dependent variables when moderated by abnormal BMI, chronological age, sex, and race/ethnicity. In addition, using archive data and an ecological design is an acceptable practice in epidemiologic research and for my study was cost effective, in both dollars and human resources, and time sensitive (adaptable to my time of study completion).

Methodology

Population

I used archival data that represented a U.S. sample of noninstitutionalized adolescents and adults (ages 12 years and older). Specifically, the respondents were from NHANES surveyed counties within New Jersey, Pennsylvania, New York, and California during the survey cycles between 2005 and 2012. The decision to use participant data from the Continuous NHANES survey was based on the acceptance in the epidemiologic community about the processes used to achieve subjective and objective population

health data from participants of NHANES. Later in this chapter, I describe the Continuous NHANES survey procedures.

Using G*Power statistical power analyses software (https://www.macupdate.com/app/mac/24037/g-power), I originally proposed a minimal sample size of 240 subjects to run a one-way ANOVA using eight groups. However, in my approved RDC proposal (i.e., RDC proposal ID No. p1578), due to disclosure risk (as per the NCHS Disclosure Manual, discussed in Chapter 1), I was allowed to present only aggregate data (explained further in Chapter 4), which narrowed my remote exposure groups to two (instead of the originally proposed eight groups). Thus, instead of using a one-way ANOVA, I considered using the independent samples *t* test. I again used G*Power to estimate the new minimal sample size under the same parameters as the initially proposed one-way ANOVA (i.e., error of probability $\alpha = 0.05$, standard power of $\beta = 0.80$ and a medium effect size $d = 0.5$). I estimate a new minimal sample size of 144 subjects to complete my investigation (see Figure 3 under Data Analysis). However, given the non-normal distribution of the dependent variable in the study sample, I applied the nonparametric alternative to the independent samples *t* test, the Mann-Whitney *U* test, to respond to RQ1.

After I aggregated the NHANES public use data for survey cycles between 2005 and 2012 (specifically, NHANES demographic, body measures, glycohemoglobin, and reproductive health file files), the RDC analyst merged my aggregated dataset with the restricted data (residential geocoordinates from the NHANES geocoding file) of NHANES surveyed respondents from all four states (New Jersey, Pennsylvania, New

York, and California). From the merged NHANES public and restricted data, there were a total of 10,942 cases from which to begin the sampling process and attain a dataset that included the study population and variables needed for my investigation (i.e., subjects ages ≥ 12 years, non-pregnant, having an A1c ≥ 5.7%, and having an observation for BMI and remote exposure). The final dataset consisted of a study sample $N = 1,724$ cases, which was more than sufficient to draw meaningful conclusions in response to both research questions.

Sampling

To begin the process of selecting the Continuous NHANES archival datasets for my study, I began with identifying NHANES survey years relevant to my study and searching for the population and variables pertinent to my research study within each two-year survey cycle between 2005 and 2012. Continuous NHANES public data was readily accessible and downloadable to SPSS®. From the home page of the Continuous NHANES (https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx) I identified the survey years needed for my study: Continuous NHANES datasets for 2005-2006, 2007-2008, 2009-2010, and 2011-2012 and identified, by viewing the restricted variables in the geocoding file, that each survey cycle included the residential geocoordinates for the respondents. I then searched for the study population needed by viewing the Demographic File Variables List from each survey cycle (https://wwwn.cdc.gov/Nchs/Nhanes/Search/DataPage.aspx?Component=Demographics&CycleBeginYear=2005). Next, I searched for the dependent and moderating study variables within the NHANES demographic file, examination data (body measures file),

and laboratory data (glycohemoglobin file). After identifying respondents matching the age parameters for my study subjects and measurements that match the operationalization of the dependent and moderating variables (A1c%, BMI, height and weight, chronological age, sex, and race/ethnicity), I concluded that this archival data was pertinent to respond to my research questions.

To create my *study sample* once all relevant Continuous NHANES public data, restricted data, and HWS data were merged, I commenced nonprobability sampling. This method of sampling was convenient, purposive, and of modal instance because the required study population were persons 12 years of age and older and it was necessary for all subjects to have an observation for remote exposure (REMEXP) as well as an observation for the other study variables (Trochim & Donnelly, 2007). After aggregating the public data files from NHANES survey cycles between 2005-2012 (NHANES demographic, body measures, glycohemoglobin and reproductive health data files (specifically the question coded for pregnancy), I submitted this data in an SPSS file format to the RDC analyst. Each of the study participants retained their respondent sequence number (SEQN) as provided in the NHANES study code books (CDC, 2016b) until the public data file was merged with the restricted data. The RDC analyst then created the variable "newID" for each case in the dataset which replaced the SEQN that was in the NHANES public data files. The RDC analyst merged my aggregated public dataset with the NHANES restricted data (residential geocoordinates) and identified surveyed respondents from New Jersey, Pennsylvania, New York, and California during

the 2005-2012 survey period. This provided me with a dataset of 10,942 cases from which to begin the sampling process.

Identifying Subjects With A Remote Exposure Observation

Given the requirements in my study for subjects to have examination, laboratory, and interview data (i.e., BMI, height, weight, A1c, & response to the pregnancy question), and an observation for the independent variable (remote exposure), my first stage of stratification included identifying subjects that had an interview and exam and that could be categorized as either residing LE1mile (≤ 1mile) or GTmile (> 1mile) from a HWS within their county of residency. After the RDC analyst merged the NHANES public and restricted data, my RDC analyst then merged this file with the HWS data that I submitted in an Excel spread sheet. Next the RDC analyst identified a state county match (i.e., respondent with residential geocoordinate data matched to a HWS within his/her county of residency). Where a subject lived in a county containing more than one HWS, that subject would later be categorized (on the independent variable) based on the HWS closest to the residence. At the completion of this process there were a total of 8,774 observations that had an interview and exam for which a distance measure (REMEXP) could be calculated. These observations were spread across the four states and a total of 21 counties.

In the final stage of sampling leading to the study population of $N = 1{,}724$, I acknowledged that my county representation might be less than 21 due to the narrowing of the population on other study parameters (i.e., abnormal BMI, age, abnormal A1c). A final count of county representation could not be released on the study population ($N =$

1,724) due to disclosure risk. However, it was noted (by my RDC analyst) that as an aggregate there remained REMEXP representation across the four states in the study population.

The age variable. Data within the Continuous NHANES surveys contained a few different age variables (i.e., age in years at screening, age in months at screening, age in months at exam, and age in months at exam for persons under 85 years at screening). For the most part there was not a significant difference with these age variables, although for persons older than 80, age in months was not collected after 2007 and some persons older than 80 did not always complete both an interview and exam. Those survey respondents not completing an interview and exam were filtered-out during the process for identifying subjects who did complete both an interview and exam (as I described in the preceding section). I calculated an updated age variable, age in months (later translated to age in years for display in the results table) for each respondent to ensure that no respondent meeting the population age parameters with a REMEXP observation would be missed during the filtering process to derive at the age for the study population. It was important to have an age variable that included the exam age in months as this age in months was necessary for calculating the BMI for adolescents (i.e., ages 12 to 20).

Filtering for pregnancy. Prior to the age stratification, I filtered the file to exclude respondents with a "yes" response for the pregnancy variable. In my approved RCD proposal (i.e., RDC proposal ID No. p1578), I proposed the use of the pregnancy variable in the reproductive health file to filter the study population responding to the interview question "Are you pregnant now?", which contained responses coded as yes,

no, refused, and don't know. However, when all public and restricted datasets were merged I recognized that a pregnancy variable was also included in the NHANES Demographic File which described pregnancy status at the time of the NHANES clinical exam. The pregnancy variable in the demographic file was coded as yes, positive lab pregnancy test, or self-reported at exam, not pregnant, cannot ascertain if the participant is pregnant at exam, or missing. For the 2005-2006 NHANES survey cycle the pregnancy status at time of exam was recorded for ages eight to 59. From 2007 to 2012, this variable was recorded for ages 20 to 44 only. NHANES pregnancy data for participants younger than age 20 and older than age 44 during survey cycles between 2007 and 2012 was restricted.

By the time that I discovered that I had overlooked the pregnancy status at time of exam variable in the NHANES Demographic file (i.e., RIDEXPRG), my RDC proposal was already approved. To have requested the use of restricted pregnancy information would have required that I submit an amendment to my RDC proposal and undergo a second review process (subject to the same time of processing, i.e., three months or more). Given project time restraints, I elected to include respondents with "missing" or "can't ascertain" pregnancy data observation (for subjects < 20 and > 44 years) if the participant met the threshold for abnormal A1c and all other study variable observations. While this may have added limitations to the discussion of results, I did not expect that the inclusion of these subjects with missing pregnancy status observation to greatly influence the study outcome given the size of the study sample. Of the 10,942 cases identified before the sampling process began, there were 7.3% with a positive response of

"yes" for the pregnancy variable. I excluded these cases. This retained subjects with a remote exposure observation, examination (i.e., BMI and laboratory values) and non-pregnancy status (N = 8,651). For the age stratification, I excluded persons less than 12 years. I scaled the outcome variable (A1c) beyond the threshold to include subjects with A1c values ≥ 5.7% creating a study sample of N = 1,724.

Rationale for Level of Statistical Significance Parameters

Because several EDCs and chemical mixtures are POPs and because POPs are ubiquitous in the environment, I was unable to control for all pathways of direct or remote exposures the respondents might have encountered in their lifetime. Also, my study did not include direct serum measurement of POPs or EDCs. Instead, exposure was suggestive (or remote) and based on residential proximity to a HWS. Therefore, I selected a significance level (α) of 0.05 and power (β) of 0.80, typically found in social epidemiology research (Trochim & Donnelly, 2007). These statistical parameters were found in studies by Kouznetsova et al. (2006) where researchers concluded a significant increase in hospitalization rates for diabetes in adult populations within residential proximity to HWS. A study by X. Liu et al. (2012) used similar statistical parameters in testing their hypothesis on hospitalization rates for respiratory disease and residential proximity to air pollution by fuel powered plants and HWS. In addition, a study by Lee et al. (2010) utilized similar statistical significance level parameters to conclude that several POPs at low levels of repeat exposure may increase risk of T2DM through endocrine disruption.

NHANES Archival Data

Survey procedures. Continuous NHANES surveys have been ongoing in the United States since 1999. The NHANES survey examines a nationally representative sample of the U.S. population, about "5, 000 persons each year" (CDC, 2016c) from all ages and various counties throughout the United States. NHANES survey participants are civilians, not in the armed forces, non-institutionalized (i.e. not living in a nursing home or prison), and not U.S. natives living abroad. NHANES data are obtained using a complex survey design that includes a multistage, probability sampling process. In doing so, every state, every county or every county within a state, or every household within a county is not surveyed during each survey cycle (CDC, 2016c). The code book for each survey year details the questions on the questionnaires, interview data, physical examination conducted, physiological measurements taken (e.g., BMI), and laboratory test performed for respondents. All interviews are conducted in the respondent's home by trained NHANES interviewers and physical examinations, body measurements, and laboratory tests are conducted by trained healthcare professions. Physical examinations and laboratory tests are completed on a mobile van outfitted as a health clinic (CDC, 2016d).

NHANES survey weights. Bearing in mind the complex survey design of NHANES, I worked closely with my RDC analyst once my RDC proposal was approved. I reviewed NHANES survey weights, primary sampling units (PSUs), and strata for utility in my study. Given that survey weights, PSUs, and strata are designed to assist with analysis of probability samples, I concluded that since I used a non-probability

convenience sample, it was not necessary for me to use these measures. My RDC analyst agreed. Thus, I only extrapolated statistical inferences in my study to the level of the study population.

NHANES Data access and RDC Proposal for Restricted Data

NHANES respondent demographic, laboratory, and body measurement data was unrestricted and available from the NHANES web site as public data (described above). Respondent residential geocoordinate information was restricted data. I requested permission for use of the restricted data from the NCHS Research Data Center (RDC) under a formal RDC proposal process that included detailing the NHANES survey years and specific variable requested (i.e., respondent residential geocoordinates) as well as my research purpose, research questions, and the public health benefit related to my study. As a requirement of the RDC proposal application process, my RDC proposal included my rationale for use of the restricted data, disclosure risk, outside data that would be merged with the NHANES data, and rationale for use of the outside data (in my case the HWS data), as well as the predicted data analysis results tables, figures and/or graphs that I planned to take out of the RDC (CDC, 2016e). I proposed that the timeline from submission of RDC proposal to approval would take approximately 6 weeks. However, the RDC approval process took a period of 8 to 9 months (I discuss this further in Chapter 4). Once my RDC proposal was approved by the NCHS proposal review committee, I was assigned the proposal approval ID No. p1578 and an RDC analyst.

Entry to the RDC

After completion of the required online NCHS RDC Confidentiality Orientation, submission of notarized confidentiality forms (i.e., Designated Agent and Access Agreement) (https://www.cdc.gov/rdc/b4confidisc/cfd400.htm), which were required of me and my dissertation Chair/mentor, and after payment of the administrative fee, I was scheduled a period of sessions to enter the RDC in Atlanta, GA. Part of my agreements with the RDC required that the RDC analyst review the methods, results and discussion chapters of my manuscript related to the project prior to any publications (to include dissertation publication) to ensure no disclosure risk have been breached. Information regarding the process for requesting NHANES restricted data is found at (https://www.cdc.gov/rdc/b3prosal/PP300.htm).

The RDC center is secured. As the researcher, I was assigned to a work station upon entry to the center. No laptops or phones were allowed at the work station. Documents (to include notes) coming into the center were required to be reviewed by the analyst before entry and no paper documents were allowed out of the Center. My review and analysis of my data file was only achieved while in the RDC on my scheduled dates. In accordance with the NCHS Disclosure Manual, no raw data from my approved RDC proposal datafile was every released to me from the RDC. Residential geocoordinate data, county and state of residency data of the study subjects were never released to me. Each time that I visited the RDC and completed my analysis, I placed my SPSS outputs in an electronic review folder. After my RDC analyst had the opportunity to review my statistical analysis for disclosure risk and approve its release, my analyzed data was

emailed to me, usually within 24-48 hours. Where there was a question in disclosure risk, some output data was not permitted for release (e.g., certain results and frequency tables - I discuss this further in Chapter 4). The RDC will retain my raw dataset (my pre-analysis and analytic files) as well as all output data (to include the non-released output date) in an electronic file for a period of five years.

Exposure Data

Hazardous waste site data applicable to this study was unrestricted. HWS data for New Jersey, Pennsylvania, New York and California was obtained from the NPL Superfund Site. The EPA's NPL site list (https://www.epa.gov/superfund/superfund-national-priorities-list-npl) provides the site name and address (including county), EPA site identification number, the date the site was listed, and an interactive map that includes the geocoordinates of the HWS. A narrative of the site is also available on the NPL and includes the land area covered by the site, the chemicals/chemical mixtures contained at the HWS, and if occurred, information about the effects of contaminates on human lives, wildlife, water, air, and soil and site clean-up efforts.

I initially proposed that details about the HWS would require the use of several federal registries. However, the NPL website provided complete details. And since the exact HWS matched to the county of residency for each study subject was unknown to me, it was not futile to compare the narratives of each of the hundreds of HWS across several federal databases. However, websites and databases such as the ATSDR Hazardous Substance Release and Health Effects Database (U.S. Department of Health and Human Services, 2016), the ATSDR Toxic Substance Portal, the EPA Envirofacts

(https://www3.epa.gov/enviro/), and the Resource Conservation and Recovery (RCR) Act Biennial Reports for years 2005-2013 (EPA, 2016) provided background information for my review and understanding of POPs, endocrine disruptors and the adverse effects of this contaminants on human health. I also proposed that the U.S. Census Bureau Census Geocorder (https://www.census.gov/geo/maps-data/data/geocoder.html), and Google Maps application program interface or geocoding API (https://www.programmableweb.com/api/google-maps-geocoding) would be used as primary and backup secondary geocoder for the HWS. However, this was obsolete given the data provided on the NPL site. In addition, the CDC SaTScan (software for the spatial, temporal, and space-time scan statistics (http://www.satscan.org) was not necessary as the residential geocoordinate data of the study subjects was not disclosed to me. Instead, the RDC analyst calculated the simple distance between residential geocoordinates (restricted data) and HWS geocoordinates using a program in SASS (see Chapter 4 and Appendix A).

Data Analysis

I used SPSS software to analyze data to respond to the research questions. I used G*Power software to estimate the minimal predicted sample size for my study. The *a priori* for RQ1(see Figure 3) indicates the minimal sample size required to demonstrate statistical significance, reduce the chance of a Type 1 error or rejecting the null hypothesis when it is true, demonstrate strength or effect between variable differences and relationships, and the minimal sample size needed for a given power sufficient to reduce the chance of a Type II error or failing to reject the null hypothesis when there

actually is an effect in the study population (Creswell, 2008; Cohen, Cohen, West & Aiken, 2003).

RQ1. Within select counties of New Jersey, Pennsylvania, New York, and California, as an aggregate, are there significant differences in the abnormal A1c% value (defined in my study as a glycohemoglobin or A1c ≥ 5.7%) between the study subjects ages 12 years and older residing within 1 mile of a HWS compared to the same population residing greater than 1 mile from a HWS within their county of residency?

H_{01}: For the study population ages 12 years and older there will be no significant difference in the A1c % between residential groups.

H_{a1}: For the study population ages 12 years and older there will be a significant difference in the A1c % between residential groups.

Independent variable (categorical): Residential groups, as an aggregate residing within NHANES surveyed counties of New Jersey, Pennsylvania, New York, and California, that contained one or more HWS within the 2005-2012 NHANES survey cycles. Defined as residence ≤ 1 mile from the HWS (remote exposure) or residence > 1 mile from the HWS (no exposure).

Dependent variable (continuous): A1c ≥ 5.7%.

Descriptive statistics

In Chapter 1, the section on research question RQ1, I detail the study limitations regarding disclosure risks and publication of descriptive statistics for RQ1. Thus, I present categorical descriptive statistics in the results tables of Chapter 4. 1). Thus, categorical descriptive statistics are presented in the results tables of chapter 4.

Inferential statistical test for analysis

I originally proposed using the independent samples t test. However, given the non-normal distribution of the dependent variable, I applied the Mann-Whitney *U* test to respond to RQ1 (see Evaluation of Statistical Assumptions in Chapter 4).

I used for my study a standard error of probability $\alpha = 0.05$, standard power of $\beta = 0.80$, and medium effect size ($d = 0.5$). Researchers have reported that any change in A1c percent is linked to one's risks for complications associated with diabetes. For example, according to a study by Eeg-Olofsson et al. (2010) reduction in A1c within one percent reduces the risk of complications (such as cardiovascular complications) up to 30 percent or more. More reduction in A1c further reduces the risk of adverse health effects or disease complication. Therefore, I proposed that any difference in A1c% between the two group means would be meaningful in a practical sense. Because I used secondary data I had no control of the number of participants in the NHANES dataset for the survey period 2005-2012, which is why I believed it was reasonable to utilize an industry standard effect size of medium range. The *a priori* for RQ1 (Figure 3) predicted that a minimum study sample of 144 subjects would be necessary. However, the final *N* for the study sample was sufficiently more than enough to respond to both research questions (see Chapter 4 results and discussion in Chapter 5).

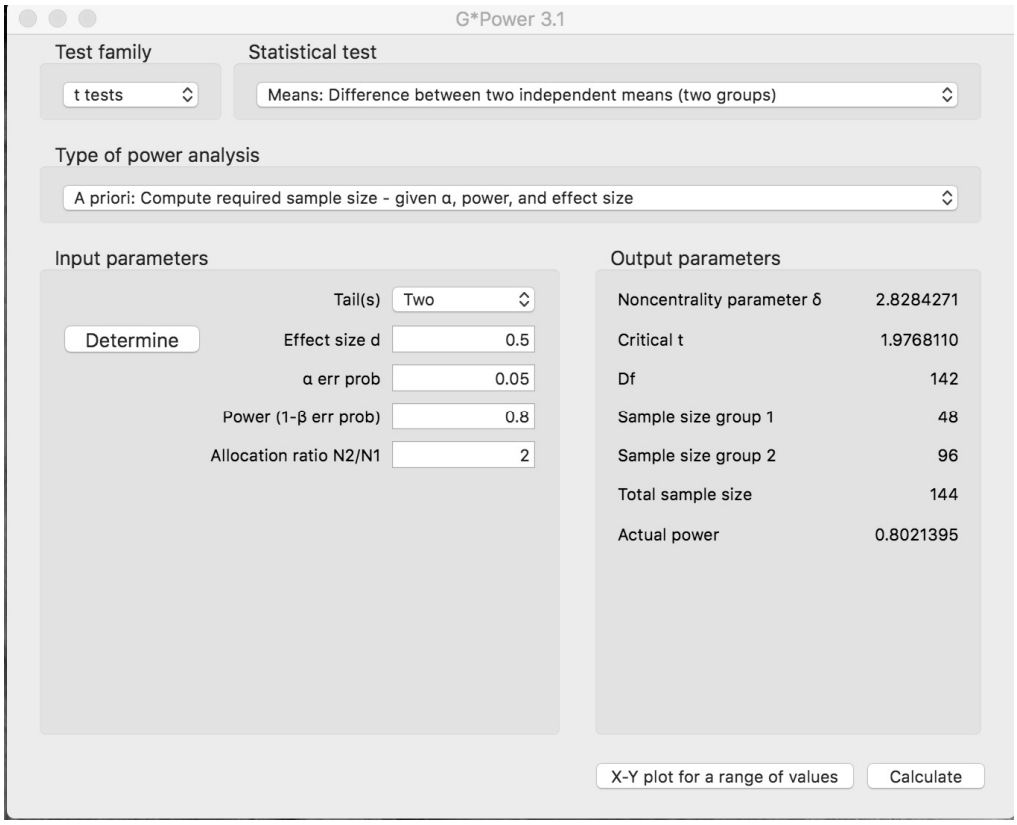

Figure 3. A prior RQ1.

Although I applied the test family *t* tests in the G*Power prediction to estimate the study sample required for RQ1, I applied the Mann-Whitney *U* to test the hypothesis of RQ1 given the non-normal distribution of the dependent variable that I identified during the Evaluation of Assumptions (see Chapter 4).

I predicted a skewed population and this is reported in the results. I anticipated that most people would not live near or on a HWS. I proposed that if a group lived near a HWS (LE1mile), the difference in the findings within the dependent variable in comparison to the group residing GTmile of a HWS would be meaningful in a practical sense. Chapters 4 and 5 discuss the results and the relevance of the results to the study purpose.

RQ2. For the study subjects ages 12 years and older residing within select counties of New Jersey, Pennsylvania, New York, and California, as an aggregate, what is the effect of the moderators, abnormal BMI, age, sex, and race/ethnicity, on the relationship between residential proximity to a HWS within the county of residency (categorized as residing ≤ 1 mile or residing > 1 mile from a HWS) and abnormal A1c% (i.e., A1c ≥ 5.7%)?

H_{02}: For the study population ages 12 years and older, the moderators abnormal BMI, age, sex, and race/ethnicity will have no significant effect on the relationship between residential proximity to a HWS and abnormal A1c%.

H_{a2}: For the study population ages 12 years and older, the moderators abnormal BMI, age, sex, and race/ethnicity will have a significant effect on the relationship between residential proximity to a HWS and abnormal A1c%.

Independent Variable/Predictor (categorical): Residential groups. Defined as residence ≤ 1 mile from a HWS (remote exposure) or residence > 1 mile from the HWS (no exposure).

Dependent Variable (categorical): Abnormal A1c value of ≥ 5.7%. Categorized as low abnormal A1c (<6.10%) and high abnormal A1c (≥ 6.10%).

Moderating Variables (MV) (categorical)

MV_1 Abnormal BMI: Categorized as abnormal BMI low (28.94 kg/m^2 or less) and Abnormal BMI high (≥ 28.95 kg/m^2).

MV_2 Age of subjects: Categorized as low age < 58 years old and high age (≥ 58years).

MV_3 Sex of subjects: Male or Female.

MV_4 Race and Ethnicity as per NHANES survey data. These groups were categorized as: Mexican Americans, other Hispanics, non-Hispanic Whites, non-Hispanic Blacks, and other non-Hispanic race including non-Hispanic Multiracial.

Descriptive Statistics

The study limitations regarding disclosure risks and publication of descriptive statistics for RQ2 are the same as that discussed for RQ1. Thus, I present categorical descriptive statistics in the results tables of Chapter 4.

Inferential statistical test for analysis

Binary logistic regression with moderator interactions.

As with research question 1, I used a standard error of probability and power $\alpha = 0.05$ and $\beta = 0.8$ respectively, as well as an industry standard medium effect size ($d = 0.15$) to express the magnitude and variability of the effect on the outcome (Baguley, 2009). The final study sample of $N = 1{,}724$ cases was more than sufficient to respond to RQ2.

RQ2 Analysis Modifications and Rationale

The results of the independent samples *t*-test (conducted for research question 1) indicated a non-linear relationship between residential proximity to a HWS and abnormal A1c. Additionally, the analysis for RQ1 resulted in no significant mean difference in abnormal A1c between residential groups (i.e., those residing ≤ 1 mile or residing > 1 mile from a HWS within the county of residency, see results Chapter 4). Given these results for RQ1, multiple linear regression moderation analysis could not be applied to

RQ2 as originally planned as the assumptions for such inferential analysis were not met. However, given that studies as discussed in the literature review (see the section on Studies Related to the Moderating Variables, in Chapter 2), describe obesity as a risk factor for T2DM (WHO 2016), and suggest adipose tissue as a storage container for POPs (Merrill et al., 2013), the influence of abnormal BMI on the relationship between the independent and dependent variable remained a rational inquiry for this study. In addition, the literature supports an increased incidence of T2DM with advancing age, disease predisposition among females, and increased disease prevalence among racial/ethnic groups of color (Dabela et al., 2014). Therefore, it remained relevant to this study that I assess the effect of the moderators: abnormal BMI, chronological age, sex, and race/ethnicity on the relationship between remote exposure to HWS and abnormal A1c. Thus, I applied binary logistic regression with moderator (covariate) interaction variables to respond to RQ2. In Chapter 4, I detail the statistical assumptions and process of preparing variables for the Binary Logistic Regression Analysis.

Threats to Validity

Threats to External Validity

External validity refers to the ability to generalize the constructs of cause and effect beyond the study subjects. Within my study this referred to the ability or inability to generalize causal relationship between residential proximity to HWS and abnormal A1c% values associated with pre-diabetes and T2DM, temporally or to populations and in different geographic locations beyond my study sample (Trochim & Donnelly, 2007). While previous studies reported a significant number of human studies in adult

populations on the chronic disease risks from exposure to POPs and EDCs, most based on long-time exposure (Lee et al., 2006; X. Liu et al., 2012; Navas-Acien et al., 2008; Sergeev et al., 2011), there remained a literature gap relevant to POPs and EDCs studies and chronic disease risks such as T2DM in populations during periods of critical growth beyond the fetal stage, such as my study population (which included participants ages 12-20 years). Yet, through my literature review, I could not find any evidence that theorized or concluded that children and adolescents were excluded from or immune to toxic or hazardous waste exposure and potential threats from POPs and EDCs that inhabit the environment (Benachour, Moslemi, Sipahutar, & Seralini, 2007; Bijlsma & Cohen, 2016; Stockholm Convention, 2016a, 2016b).

However, threats to external validity were contingent upon relying on the NPL reports of HWS contaminants and not direct environmental measurements by myself. Additionally, I used a convenience sample from archival data for the study subjects. I did not conduct individual exposure assessments nor did I gather subjective, physical examination, or laboratory data prospectively or longitudinally on individual subjects. Instead my exposure assessment was remote. These features of my study threatened the external validity of the results. Therefore, I took care during the discussion of my results to avoid the ecological fallacy and make conclusive inferences based only on group data from the study sample (N=1,724).

Threats to Internal Validity

My application of different statistical tests (i.e., non-parametric Mann-Whitney U and binary logistic regression with moderation), as well as my application of sufficient

statistical power and effect size (to detect the existence of a meaningful relationship between variables), sufficient sample size, and inclusion of moderating variables, strengthened the discussion of the study results. I anticipated that threats to concluding a causal relationship (within my study sample) would be effected by a lack of individually measured POPs or EDCs in the study subjects as well as unknown parameters such as: other volatile compound exposure history of the study subjects, comorbidities of the participants, as well as measurements of POPs/EDCs in the soil, water, air, or household.

Exclusion of volatile compound history. I did not use the history of volatile compound exposure or results of measurements of serum or urine volatile compounds of study subjects from the Continuous NHANES dataset. My primary rationale for exclusion of any NHANES data on volatile compound exposure or direct urine or serum measurements of volatile compounds was because this history and these measurements were not collected during every survey cycle and when collected it was not collected for every participant during the years 2005-2012 (CDC, 2016a, 2016b). Thus, a lot of missing data regarding volatile compound exposure history or test measurements was highly suspected as the years for NHANES collection of this data was randomized during 2005-2012 and again randomized among survey respondents and rarely collected on participants younger than 20 years.

My research design and strategy for statistical analysis aligned with the ecosocial theory of disease distribution. Although I assessed that there was no mean group difference (in response to RQ1), there was a statistically significant reduction in risk of abnormal A1c% of study participants of White race residing > 1 mile of a HWS

compared to the other White race population residing ≤ 1mile of a HWS or other racial/ethnic groups (see results Chapter 4). Given the ecosocial theory, concluding a causal relationship could not extend beyond *suggestive*, as the exposure was remote. Yet, Hill's criteria of causation retain applicability under the criteria of *biological plausibility* given the current evidence of possible increase risk of diabetes from POP exposure through a pathway of endocrine disruption (Lee et al., 2010). However, one study does not negate the ecosocial theory applied to this study or the biological plausibility given the current literature about POPs, EDCs and human health risks. I discuss this further in Chapters 4 and 5.

Threats to Construct/Statistical Conclusion Validity

Adverse exposure and accumulation of exposures to environmental hazards, such as toxic substances, formulates a pathway of embodiment (or pathway of disease exposure or risk) and susceptibility to disease that is captured in the constructs of the ecosocial theory of disease distribution (Kriger, 1994). Considering the independent variable of my study, residential proximity to HWS, the unit of exposure was remote. Proximity to HWS was operationalized using geocoordinates and calculating the difference between the residential geocoordinates of the study subjects and the geocoordinates of a HWS closest to the study subjects residing in the same county. The application of geocoordinates in determining proximity is considered a more precise measurement in comparison to census tract or ZIP code data (Liese et al., 2010).

To minimize threats to construct validity, in my research design I utilized the value of residential proximity ≤ 1 mile from a HWS to operationalize *remote exposure,*

which was similar to previous studies and holds acceptance in geo-spatial analysis and environmental epidemiology as aforementioned in the literature review. Remote exposure was operationalized and analyzed at the group level. The lack of inclusion of direct individual measurement of volatile compounds in the serum or lipid tissue of the study subjects threatened the validity of my independent variable. In addition, the time period that the subjects resided in the county containing the HWS and exclusion of other possible environmental exposures pathways (such as where the subjects attended school, or if exogenous exposures could have taken place in another country, pesticide use in the home, and plastic materials used in the home or school) also threatened the validity of the independent variable and its contribution to the outcome variable (abnormal glucose metabolism, operationalized as an A1c $\geq$ 5.7%). Threats to construct validity were handled by drawing conclusion at the group level and applying statistical inferences to the study sample only, thereby avoiding the ecological fallacy.

Ethical Procedures

I used archival data of the study participants and HWS as aforementioned. The study sample was derived from the Continuous NHANES survey cycles between the period of 2005-2012. Data on the subjects was processed through the IRB or research review board of the Centers of Disease Control/NCHS/Continuous NHANES (CDC, 2016d). My RDC proposal (requesting use of Continuous NHANES restricted data) was reviewed and approved by a research review committee at the NCHS (CDC, 2016e); RDC proposal ID No. p1578. I detail above, under data access, the process for creating and accessing the merged datafile containing the study variables as well as ongoing

processes during the study to ensure no disclosure risk. In addition, my study was approved by the Walden University IRB (approval No. 06-01-17-0108719).

Summary

I conducted a quantitative study using archival data and an ecological design statistical strategy to respond to two research questions regarding the relationship between residential proximity to HWS in NHANES surveyed counties of New Jersey, Pennsylvania, New York and California (between 2005-2012) and the abnormal A1c% of youths and adults (A1c $\geq$ 5.7%) residing $\leq$ 1 mile of a HWS compared to youths and adults with A1c $\geq$ 5.7% residing > 1 mile from a HWS located in the same county as the study subject. The unit of analysis for my study was at the group level. A minimum total sample size of 144 participants was proposed, however, after aggregation of data from New Jersey, Pennsylvania, New York, and California, there was a sample size of N = 1,724, which was sufficiently more than enough to respond to the research questions. Under the data analysis plan I detailed the specific sample size required to answer each question given a level of statistical significance, statistical power, and effect size to draw meaningful conclusion from the results. For RQ1, I applied non-parametric testing due to the skewed population on the dependent variable (i.e., Mann Whitney U). Given the non-linear relationship of the study sample, I modified my RQ2 methodology approach to include binary logistic regression analysis with moderator interactions to test the hypothesis of RQ2. I addressed in Chapter 3 the ethics and threats to internal and external validity surrounding my study and provided detail on how my study was conducted, how the study subjects were protected, and the processes applied to ensure no disclosure risk

in accordance with the CDC, NCHS and the RDC. I also detailed the process for acquisition of the restricted NHANES geocode data and my RDC proposal process. In Chapter 4, I present the results from my study.

Chapter 4: Results

Introduction

My purpose in this study was to examine the relationship between residential proximity to HWS in select counties of New Jersey, Pennsylvania, New York, and California and the abnormal glucose metabolism measurement (defined in my study as a glycohemoglobin or A1c ≥ 5.7%) of subjects ages 12 years and older residing within the same county of the HWS. I used the research questions and hypotheses to guide my study. For RQ1, I hypothesized that for the study population ages 12 years and older, there would be a significant difference in the abnormal A1c % between residential groups (those residing ≤ 1 mile compared to residents residing > 1 mile of a HWS within their counties of residency). For RQ2, I hypothesized that for study subjects ages 12 years and older, moderators abnormal BMI, age, sex, and race/ethnicity would have a significant effect on the relationship between residential proximity to a HWS and abnormal A1c%.

In Chapter 4, I discuss the data collection procedures, time frame for data collection, and any discrepancies in the data collection from that proposed as well as any limitations to data collection and/or limitations in display of statistical analyses. Also in Chapter 4, I provide the results of descriptive and inferential statistics related to my study sample and research questions. Prior to transitioning to Chapter 5, I conclude Chapter 4 with an assessment of my study hypotheses.

Data Collection

I used archival data from the NHANES, years 2005-2012 and HWS information from the 2016 NPL of Superfund Sites in this study (see Chapter 3).

Time Frame for Data Collection

The compilation of HWS data took approximately 4 months. Approval of my RDC proposal in which I requested use of restricted NHANES data took 8 months (RDC approval ID No. p1578). Merging of the HWS data with the restricted data and public use NHANES datasets and preparation of my dataset by the RDC analyst for onsite use at the RDC took approximately 4 weeks. Entry to the RDC is restricted to those who have completed an orientation process, paid the service fee, and scheduled an appointment. This orientation process (which had to be completed by myself and my dissertation committee chair/mentor), fee payment, and scheduled dates for the RDC onsite visit took an additional 2 to 3 weeks to complete once my dataset was approved as ready by the RDC analyst. Onsite data review, RDC analyst updates to the merged dataset (HWS, NHANES public and NHANES restricted data) and onsite statistical analysis at the RDC spanned 8 weeks. After my time at the RDC, it took approximately 2 weeks to receive the detailed description from the RDC analyst on the STRATA software process applied to compute the distance/remote exposure variable. Given that the raw data are retained at the RDC, the data analysis process could be completed only onsite. This period is included as part of my data collection timeline. Overall, it took 14 months to complete the data collection process.

Discrepancies in Data Collection and Rationale

Modifications in the collection of HWS data. At the time of my proposal, either the NPL website did not have an interactive map locating the HWS or this link was not active, or I did not notice it. When I began organizing the HWS data, an interactive map

was available and when I selected a site from the list it opened to a page having an interactive map that included the site name, county location, date site added to the NPL list, and the latitude and longitude coordinates for the site. From the map, one could also access the narrative document for the site, which included information about site contaminants and administrative details such as cleanup efforts and size of area (sometimes included). Given this, it was not necessary for me to use the U.S. Bureau Census geocoder (as proposed in Chapter 3) to determine the geocoordinates for the HWS. However, because of disclosure risk (in accordance with the NCHS RDC Disclosure Manual, discussed in Chapters 1 and 3), the RDC did not provide me with the actual names of the counties in which NHANES respondents lived during the aggregate survey period 2005-2012. Therefore, I had to include all HWS identified on the NPL list for each state (New Jersey, New York, Pennsylvania, and California) and present this in an Excel spread sheet to the RDC analyst.

As I began reviewing the HWS data, I became aware that the geocoordinates for the HWS were a midpoint of the site. However, each HWS expanded some areal distance. In my proposal, no NHANES respondent was projected to live on the HWS. However, a respondent could live beyond its borders. Therefore, I considered the inclusion of radius miles (distance in land miles from the HWS midpoint to its borders). This radius miles distance of the HWS (RADmiles) would be subtracted from the difference between the HWS geocoordinates and the residential geocoordinates of the study subjects (DISMILES). The goal was that RADmiles – DISMILES would create a more precise residential distance or remote exposure variable for my study. However, the

RDC analyst indicated that negative values for distance were reported using HWS radius calculations, which would indicate that survey respondents resided on or within the HWS. There was no documentation on the NPL list of Superfund sites that suggested that the study subjects resided on or within the HWS.

I made a query to my RDC analyst regarding the possibility of including a radial distance (RADmiles) for the study subjects. Theoretically, it would seem prudent that if radius miles were calculated for the HWS that the same type of calculation should be done for the study participants so that the variable (RADMILES) would be included in the distance calculations between the HWS and the respondents' place of residency. However, the area or square footage of the residential land occupied by survey respondents was not collected by NHNAES and, therefore, such calculations could not be provided.

Given that the residential area radius could not be calculated (as such data was not collected by NHANES) and that negative values were reported when the HWS radius miles were included in the distance computations, I decided to retain the proposed simple distance calculations (i.e., geocoordinates of the HWS – geocoordinates of the survey respondents place of residence) as the remote exposure variable (RemExp), categorized as residing ≤ 1mile (LE1mile) or > 1mile (GTmile) from a HWS located within the county of residency. I provide in Appendix A excerpts by the RDC analyst on the creation of the RemExp variable using simple distance.

Modifications in remote exposure calculation. I originally proposed using CDC SaTScan software to calculate the RemExp distance. However, this was completed

by the RDC analyst using the GEODIST function in SAS© (Statistical Analysis Software). According to the RDC analyst, "negative values were not an issue for the simple distance calculation" (see Appendix A).

Descriptive and Demographic Characteristics of the Study Sample

My study sample was one of convenience. However, it provided a fair representation of participants by sex and ethnicity and was reflective of the historical presentation of T2DM in adults as well as the growing epidemic of T2DM being addressed in the adolescent population. For example, in the Diabetes 2017 Report Card (CDC, 2018), the CDC indicated that the prevalence of diabetes continues to be highest among adult populations; reporting that of the 30.3 million people of the U.S. population diagnosed with diabetes in 2015, 30.2 million (or 12.2% of all U.S. adults) were adults aged 18 or older (Figure 4). The Diabetes 2017 Report also stated an estimated 34% of all U.S. adults in 2015 had prediabetes (defined in my study as an A1c of 5.8% to 6.4%), and that most were not aware. With respect to children and adolescents, the Diabetes 2017 Report described that more than 5,000 youths aged 10-19 years were newly diagnosed with T2DM during 2011-2012, with the highest incidence rates among people of color (Figure 5). While my statistical inferences are limited to the study population (given the non-probability sampling), the complexity of the NHANES survey design and sampling methods (discussed in Chapter 3), strengthens the characteristics of my study sample and how proportional my study sample is to the non-institutionalized U.S. population of adults and adolescents.

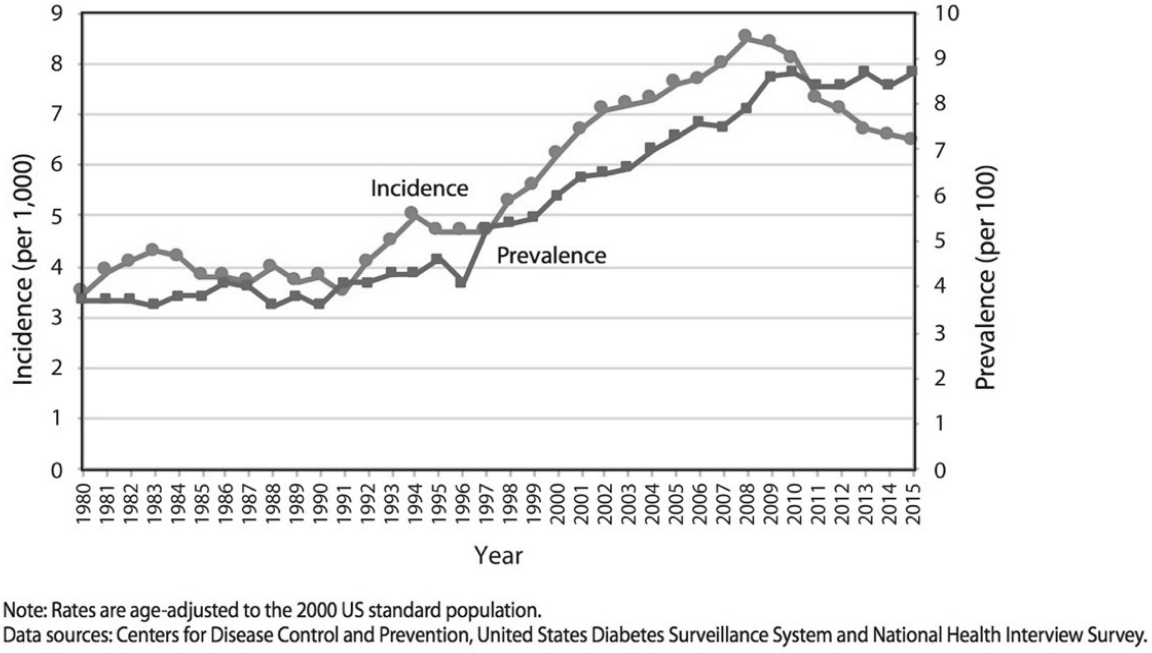

Figure 4. Trends in diabetes, adults ages 18 years and older, 1980-2015.

From Centers for Disease Control and Prevention (2018). Diabetes Report Card 2017, p. 2.

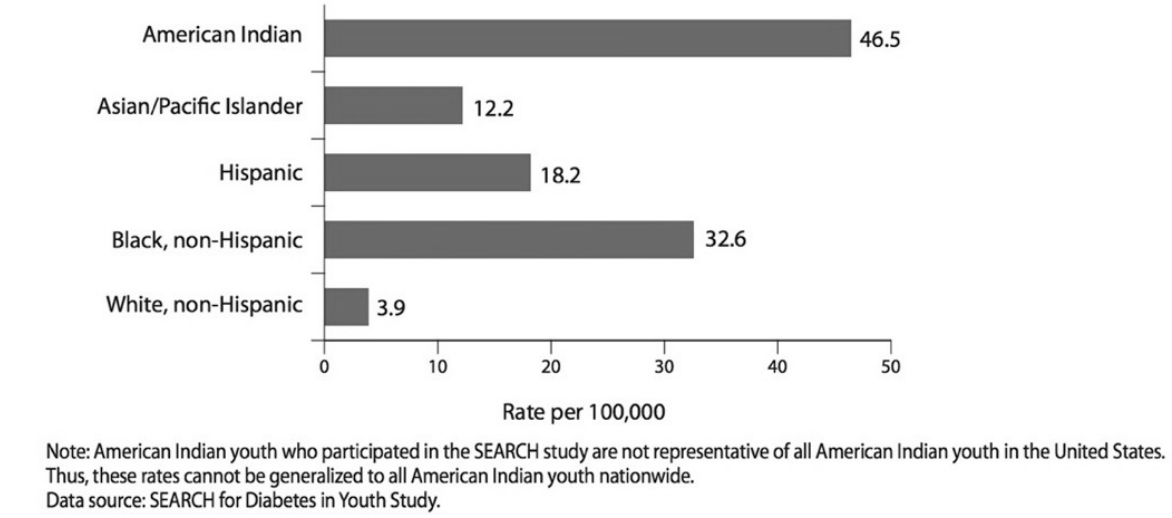

Figure 5. Incidence of Type 2 diabetes among U.S. children and adolescents ages 10 to 19 years by race/ethnicity, 2011-2012.

From Centers for Disease Control and Prevention (2018). Diabetes Report Card 2017, p. 8.

Results

Descriptive Statistics

The study sample ($N = 1,724$), was a non-probability sample stratified from the aggregate Continuous NHANES datasets for years 2005-2012. I used the NHANES public files: demographic, body measures, and laboratory; and restricted geocoding files, The study sample included non-pregnant subjects meeting the study criteria for age (≥ 12 years), the study definition for abnormal A1c (≥ 5.7%), and having a count for remote exposure to a HWS (i.e., a simple distance measure from home residence to a HWS within the county of residency). The study sample included a near even distribution of males (50.9%) and females (49.1%), with the predominance of the sample falling within the 21 years and older age group (95% compared to 5.0% for ages 12 –to-20 years). These descriptive results are displayed in Table 1, along with the race/ethnicity of the study sample.

For my sample, of the $N = 1,724$ there were 1,706 non-missing observations for abnormal BMI (defined in my study as a BMI > 85^{th} percentile for ages 12-20 and a BMI ≥ $25kg/m^2$ for ages ≥ 21). There was a total of 1.0% missing BMI observations in my dataset. While there is not an established cutoff in the literature regarding the percent of missing data acceptable for valid statistical inferences, Bennett (2001) has suggested that missing data less than 10% for a given variable may present less issues of bias. In general, given that my data was based on a convenience sample, statistical inference was only applicable to my study sample (Kline, 2017). Therefore, because my missing BMI

values were less than 10 % of the total BMI observations for my study sample, I deemed this as insignificant and missing observations for BMI were treated as missing in my moderation analysis in response to RQ2.

Table 1

Descriptive Statistics of Study Sample (N = 1,724, Sex, Age, Race/Ethnicity)

			Frequency	Percentage	Valid percentage
Sex					
Valid		Male	878	50.9	50.9
		Female	846	49.1	49.1
Age*					
Valid	AgeGTE21		1642	95.2	95.2
	AgeGTE12_LE20		82	4.8	4.8
Race/Ethnicity					
Valid	Mexican American		430	24.9	24.9
	Other Hispanic		264	15.3	15.3
	Non-Hispanic White		426	24.7	24.7
	Non-Hispanic Black		353	20.5	20.5
	Other non-Hispanic race including Multiracial		251	14.6	14.6
	Total		1724	100.0	100.0

Note. * = Age at time of NHANES medical exam. GTE = greater than or equal to. LE = less than or equal to. Per the NCHS RDC Disclosure Manual, publication of descriptive statistics on small cells were not permissible. This limited the display of frequency tables on continuous variables of the raw NHANES data. However, categorical variables as above approved by the RDC.

Evaluation of Statistical Assumptions

RQ1 and the Mann-Whitney *U*

Given the two independent samples in my study population, I applied statistical assumptions for the independent samples *t*-test during my pre-analysis. The statistical assumptions for the independent samples *t*-test include: that the dependent variable should be measured on a continuous scale and that the independent variable should

consist of two categorical independent groups; that there is an independence in observations; no significant outliers; that the dependent variable should be approximately normally distributed for each group; and that the variance of the population means are equal between the two groups (or presence of homogeneity of variances). However, the dependent variable was not normally distributed in my study sample. This suggested that I apply the non-parametric Mann-Whitney *U* test as the inferential statistical test to test the hypothesis of RQ1

The Continuous NHANES Survey data collection process clearly indicates that observations per study subject are independent. My dataset included A1c values (my dependent variable) ranging from 5.7% to greater. My independent variable, remote exposure groups were aggregated and categorized as either residing ≤ 1 mile of a HWS or residing > a mile of a HWS (within the county of residency). However, given the complexity of my dataset (which included restricted data) and disclosure risk, I was not able to publish frequency tables of A1c values for my sample. While it was permissible for me to view frequency tables of continuous variables, primarily, no frequency tables on continuous variables of the raw data were permissible for publication outside of the RDC.

As predicted in the assumptions for the study (Chapter 1), it was anticipated that most subjects would not live within less or equal to 1 mile of a HWS in their county of residency. Given the non-normal distribution of the dependent variable (A1c) in this study sample, I applied the Mann-Whitney *U* test to respond to RQ1. Table 2 and Figures 6 and 7 graphically demonstrate the statistical assumptions for RQ1 regarding sample

size and distribution of the dependent variable within the population of both independent groups in my study.

Table 2

Evaluation of Assumptions (RQ1)

Statistics (N = A1c ≥ 5.7%)			
Glycohemoglobin (%)			
GTmile	*n*	Valid	1,676
		Missing	0
LE1mile	*n*	Valid	48
		Missing	0

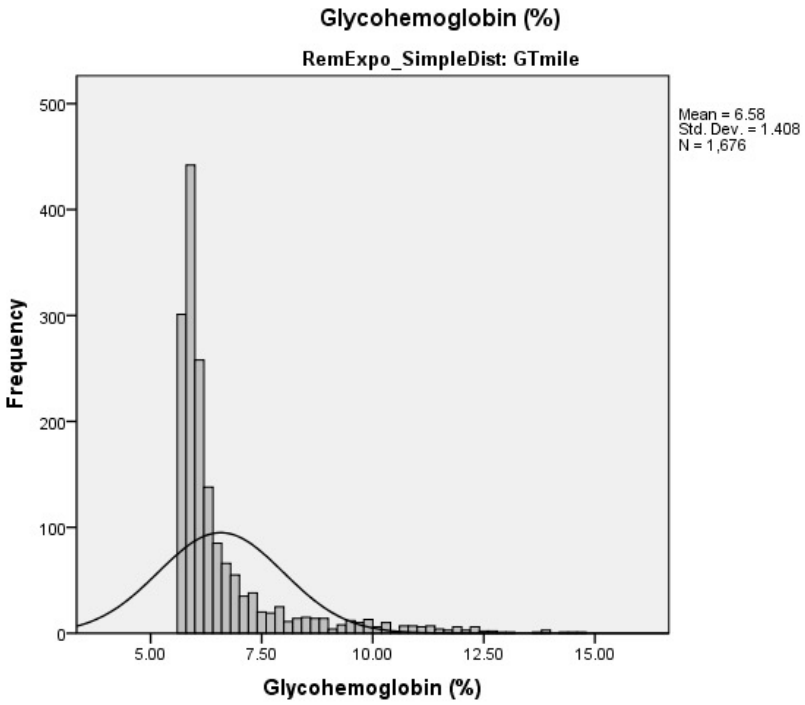

Figure 6. A1c distribution in group residing > 1 mile from a HWS.

As proposed, a skewed distribution was found in this sample with most A1c values nestled between 5.7% to < 10%.

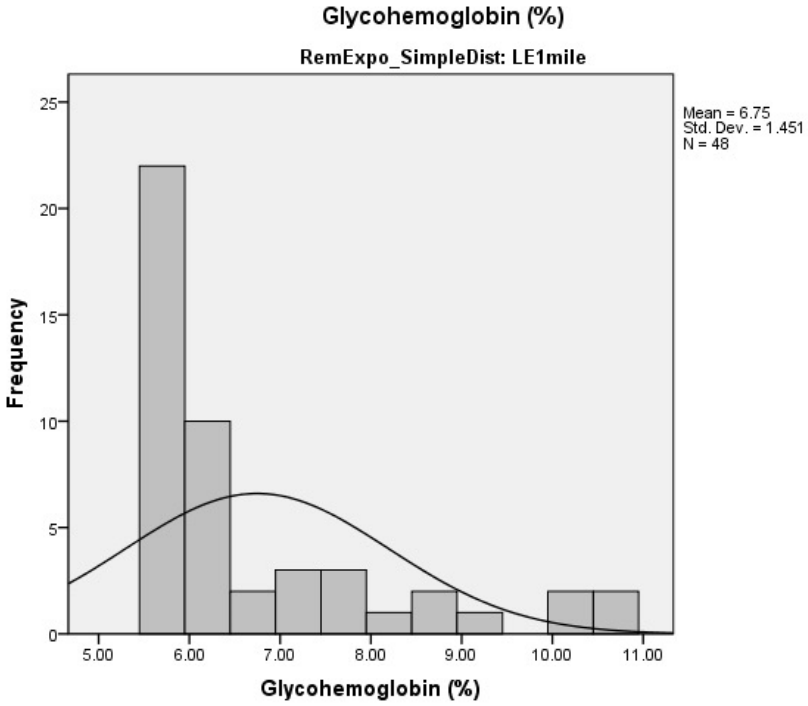

Figure 7. A1c distribution of group residing ≤ 1mile from a HWS.

As proposed, fewer subjects in the study sample resided within a mile of a HWS; of those who did there was a similar range in A1c values (between 5.7% to < 10%) as in the other group.

RQ2 and Binary Logistic Regression Using Moderation Interactions

RQ2 Analysis modifications and rationale. The results of the independent samples *t*-tes*t* (conducted for research question 1) indicated a non-linear relationship between residential proximity to a HWS and abnormal A1c. As a result of this analysis, I applied the Mann-Whitney *U* for RQ1 which indicated no significant mean difference in abnormal A1c between residential groups (i.e. those residing ≤ 1 mile or residing > 1 mile from a HWS within the county of residency). Given these results for RQ1, multiple linear regression moderation analysis could not be applied to RQ2 as originally plan as the assumptions for such inferential analysis were not met. However, given that studies, as discussed in Chapter 3 (see section on Studies Related to the Moderating Variables), describe obesity as a risk factor for T2DM (WHO 2016), and suggest adipose tissue as a storage container for POPs (Merrill et al., 2013), the affect of abnormal BMI on the relationship between the independent and dependent variable remained a rational inquiry for this study. In addition, the literature supports an increased incidence of T2DM with advancing age, disease predisposition among females, and increased disease prevalence among racial/ethnic groups of color (Dabela et al., 2014). Therefore, it remained relevant to this study to assess the effect of the moderators: age, sex, race/ethnicity, and abnormal BMI on the relationship between remote exposure to HWS and abnormal A1c. Thus, I applied binary logistic regression with moderator (covariate) interaction variables to respond to RQ2.

Statistical Assumptions for Binary Logistic Regression

Binary dependent variable. Given the continuous dependent variable abnormal A1c (defined in my study as an A1c of ≥ 5.7%), I created a binary variable using the median for abnormal A1c in the study sample (*Mdn* = 6.0). An A1c value of 6.00% accounted for 53.0% of the study population with greater values accounting for the remaining study population. Therefore, I established a binary dependent variable for abnormal A1c ≥ 5.7%: low abnormal A1c (< 6.10%) coded as "0" in the logistic analysis and high abnormal A1c (≥ 6.10%) coded as "1", which met the assumption for the logistic regression dependent variable.

Binary variables. I evaluated all variables to determine their fit for the logistic regression model. The independent variable (remote exposure or REMEXP) fit the model as residential groups were categorized on two levels: residing ≤ 1 mile or residing > 1 mile from a HWS. Sex, fit the model as it had two levels male and female. Race/ethnicity variables where already identified categorically, thus this variable met the assumptions for logistic regression analysis. However, as with the abnormal A1c dependent variable, the moderators chronological age at time of exam and abnormal BMI required the use of the *Median* from the continuous variable of each to create a binary variable to fit the logistic regression model.

Binary age variable. Using the median age from the continuous age variable for the study sample (AgeYrs ≥ 12), I created a binary age variable (AgeYrs2GTET12) to meet the requirements for logistic regression. Because the sample size when categorized as adolescent or adult presented a very skewed population (i.e., age ≥12 ≤ 20, *n* = 82; age

≥ 21, *n* = 1,642; See Table 1), I was concerned that the small cell size would not be an adequate fit for the logistic regression analysis. Therefore, using the continuous age variable of the study sample, the median age was derived (*Mdn* = 58.0 years) and used to create a binary age variable for logistic regression analysis. Age 58.0 years accounted for 50.5% of the study population with ages greater than 58 accounting for the remaining population. Thus, I created a binary age variable for the study population ≥ 12yrs such that: AgeYrs2 < 58 = low age for the population and AgeYrs2 ≥ 58 accounted for the study population in the high age category. This provided the binary age variable that met the assumptions for logistic regression analysis.

Binary abnormal BMI variable. Similarly, as with the binary abnormal A1c and binary Age variable, I obtained the median from the continuous abnormal BMI variable (*Mdn* = 28.94 kg/m^2). A BMI of 28.94 kg/m^2 accounted for 50% of the study population (adolescents and adults) with frequency ranges from 13.40 kg/m^2 to 72.56 kg/m^2. The frequency output reported 18 missing cases for abnormal BMI in the dataset. However, all cases in the dataset where reviewed and noted to include participants meeting the study age parameters, abnormal A1c threshold, having a remote exposure observation, and having had both an interview and examination as part of the NHANES study. The "18" calculated missing accounted for one percent (1.0%) of the study population. While there is not an established cutoff in the literature regarding the percent of missing data acceptable for valid statistical inferences, Bennett (2001), suggested that missing data less than 10% for a given variable may present less issues of bias. In general, given that my data was based on a convenience sample, statistical inference was only applicable to

my study sample (Kline 2017). Therefore, because my missing abnormal BMI values were less than 10% of the total abnormal BMI observations in the study sample, I viewed this as insignificant and missing observations for abnormal BMI were treated as non-missing in the moderation interaction logistic regression analysis. Given a median abnormal BMI of 28.94 kg/m2, I defined the binary abnormal BMI categorical variable as: AbnormalBMI_Low = 28.94 kg/m2 or less and AbnormalBMI_High = ≥ 28.95 kg/m2, thus establishing an Abnormal BMI variable meeting the assumptions for logistic regression analysis.

Additional abnormal BMI review. Given the growing epidemic of T2DM in youth, while the *Mdn* abnormal BMI above included adolescents and adults in the study population, I separately assessed the median BMI for age percentile. The BMI for youths is a calculated BMI based on the height and weight of youth ages 2 –to 20 years and charted on CDC growth charts according to sex and age which aligns the subject to a BMI for age percentile (see Figures 1 and 2). This calculation and coding of which subjects met the ≥ 85th percentile threshold for my study sample (coded as 1 = BMI ≥ 85th percentile, or 0 threshold not met) was completed by the RDC analyst during the establishment of my RDC dataset with merged public and restricted data (which initially contained 10,942 cases prior to the convenience sampling process as described in Chapter 3). In my study sample the frequency table reported a median abnormal BMI for age percentile of 88.97% (*n* = 81), which translated to a BMI between 26-27 kg/m2 for boys and girls in my study population ages 12 – 20 years. Thus, subjects 12 to 20 years in my study sample were within the Abnormal BMI_Low category for the logistic regression

analysis. The variation of $n = 81$ in the frequency table for establishing the *Mdn* Abnormal BMI for the youth population and $n = 82$ in Table 1 Descriptive Statistics of the Study Sample, representing the youth population in the study, is suspected due to rounding from age in months to age in years in creating the AgeYrs variable.

Finally, logistic regression requires that each observation is independent, that there is linearity of the independent variable and log odds, and a large sample size. Having met all the assumptions for logistic regression, I completed the analysis to respond to RQ2 (See results below under Statistical Analysis).

Statistical Analysis RQ1

Within select counties of New Jersey, Pennsylvania, New York and California, as an aggregate, are there significant differences in the abnormal glucose metabolism measurement (defined in my study as a glycohemoglobin or A1c ≥ 5.7%) between survey respondents ages 12 years and older residing within one mile of a HWS compared to respondents within the same age group residing greater than one mile from a HWS located in their county of residency?

Table 3

RQ1, Distance From HWS, and A1c

	RemExp	*N*	Mean	*SD*	Std. Error Mean
Glycohemoglobin(%)	LE1mile	48	6.746	1.451	.209
	GTmile	1676	6.578	1.408	.034

Note. RemExp = remote exposure. Non-normal distribution of glycohemoglobin (A1c).

Table 4

Mann-Whitney U Test

	Null hypothesis	Test	*p* value	decision
RQ1	The distribution of glycohemoglobin (%) is the same across categories of RemExp_SimpleDist	Independent-samples Mann-Whitney *U* test	.635	Retain the null hypothesis

Note. Asymptomatic significance are displayed. The significance level is .05.

The results of the Mann-Whitney *U* test showed that the mean glycohemoglobin percent (or mean A1c %) between residents ages 12 years and older residing less than or equal to 1 mile from a HWS and residents 12 years and older residing greater than 1 mile of a HWS was not statistically significant at the .05 level of significance ($p = .635$). On average, the hemoglobin A1c between residents residing within 1 mile and greater than 1 mile of a HWS were approximately the same. The null hypothesis which suggested that there was no significant difference in the mean A1c values between the study population residing less than or equal to 1 mile of a HWS and those residing greater than 1 mile of a HWS cannot be rejected.

Statistical Analysis RQ2

For the study subjects ages 12 years and older residing within select counties of New Jersey, Pennsylvania, New York and California, as an aggregate what is the effect of the moderators: abnormal BMI, age, sex and race/ethnicity on the relationship between residential proximity to a HWS within the county of residency (categorized as residing ≤ 1 mile or residing > 1 mile from a HWS) and abnormal A1c% (i.e. A1c ≥ 5.7%)?

Logistic Regression I (Abnormal A1c and All moderators/covariates)

Dependent Variable:

Low abnormal A1c (< 6.10%) coded as 0;
High abnormal A1c (≥ 6.10%) codes as 1.

Categorical Variables in Model

AbnormBMI_cat:

AbnormBMI_Low ≤ 28.94 kg/m^2, coded as 0 (*n* = 871)
AbnormalBMI_High = ≥ 28.95 kg/m^2, coded as 1 (*n* = 853)

AgeYrs2GTET12

Age Low < 58 years coded as 0 (*n* = 839)
Age High ≥ 58 years codes as 1 (*n* = 885)

Sex

Male coded as 0 (*n* = 878)
Female coded as 1 (*n* = 846)

Race/Ethnicity

Mexican American, reference (*n* = 430)
Other Hispanic, coded as 1 (*n* = 264)
Non-Hispanic White coded as 2 (*n* = 426)
Non-Hispanic Black coded as 3 (*n* = 353)
Other non-Hispanic including non-Hispanic MultiRacial coded as 4 (*n*=251)

Table 5

Model Summary Logistic Regression I: Abnormal A1c and All Moderators/Covariates

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	2296.754	.049	.066

Table 6

Hosmer and Lemeshow Test for Logistic Regression I

Step	Chi-square	*df*	*p*
1	8.059	8	.428

Table 7

Variables in the Equation: Logistic Regression I (Abnormal A1c and All Moderators/Covariates)

		B	*SE*	Wald	*df*	*p*-value	Odds Ratio (*OR*)	95% C.I.for *OR* Lower	Upper
Step 1a	AbnormBMI_cat(1 = AbnormBMI_High ≥ 28.95kg/m^2)	.604	.104	33.518	1	.000	1.829	1.491	2.244
	AgeYrs2GTET12(1= Age High ≥ 58yrs)	.764	.102	55.839	1	.000	2.148	1.757	2.624
	Sex (1; Female)	-.145	.101	2.071	1	.150	.865	.711	1.054
	MexicanAmerican			7.372	4	.117			
	Other Hispanic(1)	.050	.161	.097	1	.756	1.051	.767	1.441
	Non-Hisp White (2)	-.156	.143	1.184	1	.277	.856	.647	1.133
	Non-Hisp Black (3)	-.046	.148	.098	1	.755	.955	.714	1.277
	Multi –Racial (4)	.291	.168	2.993	1	.084	1.338	.962	1.860
	Constant	-.750	.133	31.710	1	.000	.473		

The first logistic regression analysis was conducted to investigate if there was a relationship between the covariates/predictor variables (abnormal BMI, age, sex and race/ethnicity) and abnormal A1c%. All of the predictor variables in the model were tested *a priori* to verify there were no violation of the assumption of the linearity of the logit. The Hosmer-Lemeshow goodness-of-fit was not significant ($p > .05$), indicating the model was correctly specified. Additionally, the [-2 log Likelihood = 2296.754] and the [Nagelkerke R squared = .066], supported a good model fit. For this Logistic Regression I model (Tables 5-7), the results indicate that the covariates (sex and race/ethnicity) were not significant ($p > .05$). However, the predictor variables AbnormBMI_High (BMI ≥ 28.95 kg/m^2) and Age_High (≥ 58yrs) were statistically significant ($p = < .05$). Controlling for sex and race/ethnicity, the predictor variables BMI ≥ 28.95 kg/m^2 and Age ≥ 58yrs, were found to contribute to the logistic regression model. The unstandardized Beta weight for the Constant; $B = (-.750)$, $SE = .133$, Wald = 31.710, $p < .05$.

- For AbnormBMI High (BMI ≥ 28.95 kg/m^2), the unstandardized Beta weight for this predictor variable: $B = (.604)$, $SE = .104$, Wald = 33.518, $p < .05$. The estimated odds ratio [OR = [1.829], 95%CI (1.491, 2.244)] favored that study participants with high abnormal BMI were 1.8 times more likely to have an Abnormal A1c compared to the participants with a low abnormal BMI.

- For Age ≥ 58yrs, the unstandardized Beta weight for this predictor variable: $B = (.764)$, $SE = .102$, Wald = 55.839, $p = < .05$. The estimated odds ratio [OR = [2.148], 95%CI (1.757, 2.624)] favored that study participants in the high age category (age ≥ 58yrs) were 2.1 times more likely to have an Abnormal A1c compared to the participants in the low age category.

Logistic Regression II (Covariates and Remote Exposure Moderator Interactions)

Dependent Variable:

Low abnormal A1c (< 6.10%) coded as 0
High abnormal A1c (≥ 6.10%) codes as 1

Categorical Variables in Model. Same as in Logistic Regression I above as well as the following:

SimpDis (RemExposure variable)

GT mile coded as "1" (*n* = 1676)
LE1mile coded as "0" (*n* = 48)

Table 8

Model Summary Logistic Regression II: Covariates and Remote Exposure Moderator Interactions

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	2286.360	.055	.073

Table 9

Hosmer and Lemeshow Test for Logistic Regression II: Covariates and Remote Exposure Moderator Interactions

Step	Chi-square	*df*	*p*-value
1	4.646	8	.795

Table 10

Logistic Regression II (Variables in Equation) Covariates and Remote Exposure Moderator Interactions.

		B	*SE*	Wald	*df*	*p*	Odds Ratio (*OR*)	95% C.I. for *OR*	
								Lower	Upper
Step 1a	AbnormBMI_cat(1= AbnormBMI_High ≥ 28.95kg/m2)	.065	.640	.010	1	.919	1.067	.304	3.743
	AgeYrs2GTET12(1= Age High ≥ 58yrs)	.216	.604	.127	1	.721	1.241	.380	4.055
	Sex(1=Female)	-.617	.654	.890	1	.345	.539	.150	1.945
	MexicanAmerican			3.373	4	.497			
	Other Hispanic(1)	.423	.811	.272	1	.602	1.527	.312	7.481
	Non-Hisp White(2)	1.531	.870	3.096	1	.078	4.621	.840	25.423
	Non-Hisp Black (3)	.666	.624	1.139	1	.286	1.946	.573	6.606
	Multi-Racial (4)	22.185	23015.551	.000	1	.999	4314288637.359	.000	.
	AbnormBMI_cat(1= AbnormBMI_High ≥ 28.95kg/m2) by distLE1mile(1=GTmile)	.565	.647	.764	1	.382	1.760	.495	6.253
	Sex(1=Female) by distLE1mile(1)	.483	.662	.534	1	.465	1.621	.443	5.929

MexicanAmerican * distLE1mile			4.258	4	.372			
Other Non-Hispanic(1) by distLE1mile(1)	-.362	.819	.195	1	.659	.697	.140	3.470
Non-Hisp White(2) by distLE1mile(1=GTmile)	-1.720	.874	3.869	1	.049	.179	.032	.994
Non-Hisp Black (3) by distLE1mile(1)	-.721	.629	1.313	1	.252	.486	.142	1.669
Multi-Racial (4) by distLE1mile(1)	-21.922	23015.551	.000	1	.999	.000	.000	.
AgeYrs2GTET12(1= Age High ≥ 58yrs) by distLE1mile(1)	.568	.612	.861	1	.353	1.764	.532	5.849
Constant	-.764	.134	32.658	1	.000	.466		

In the second logistic regression analysis conducted to respond to RQ2, I aimed to investigate the effect of the interaction (residential proximity or remote exposure to a HWS and abnormal BMI, age, sex, and race/ethnicity) on the outcome (Abnormal A1c ≥ 5.7%). All of the predictor variables in the model were tested *a priori* to verify there were no violation of the assumption of the linearity of the logit. The Hosmer-Lemeshow goodness-of-fit was not significant ($p > .05$), indicating the model was correctly specified.

Additionally, the [-2 log Likelihood = 2286.360] and the [Nagelkerke R squared = .073], supported a good model fit.

For the Logistic Regression II model, the results of the interaction (RemExp*covariate(s) indicate that remote exposure to a HWS (either residing ≤ 1 mile or > 1 mile) coupled with abnormal BMI, age, or sex was not statistically significant (p >.05) towards the outcome of abnormal A1c. However, there was a statistically significant relationship in the interaction of race/ethnicity with residential proximity to a HWS and outcomes of abnormal A1c. The results indicate that for the non-Hispanic White study participants, the unstandardized Beta weight for the interaction of this predictor variable with RemExpo was: B = (-1.720), SE = .874, Wald = 3.869, p = < .05. The estimated odds ratio [OR = [.179], 95% CI (.032, .994)] favored that study participants of non-Hisp White race/ethnicity residing > 1 mile from a HWS had an 82.1% reduced risk of Abnormal A1c compared to participants of the same race residing ≤ 1 mile from a HWS.

Summary

The results indicate that for my study population as an aggregate of residents residing within a county having a HWS in New Jersey, New York, Pennsylvania, and California, there was not a statistically significant mean group difference in abnormal A1c % outcome between residents residing less than or equal to 1 mile of a HWS compared to the study population residing greater than 1 mile of a HWS (Mann Whitney U results Table 4). Thus, the null hypothesis for RQ1 was retained.

In response to the hypothesis for RQ2, the logistic regression analysis I (relationship between the covariates and abnormal A1c, Table 7), and the logistic regression analysis II (interaction of the moderators: abnormal BMI, sex, age, race/ethnicity with remote exposure and outcomes of A1c, Table 10), the null hypothesis is rejected and the alternative hypothesis accepted as the moderator race/ethnicity did have a significant effect on the relationship between residential proximity to a HWS and abnormal A1c outcomes. What also emerged from the logistic regression analysis was the likelihood of abnormal A1c outcomes for study subjects with an abnormally high BMI ($BMI \geq 28.95\ kg/m^2$) or advanced age (≥ 58 years). In Chapter 5, I discuss the findings of my study, the relevance of my results to the study purpose and validity, and propose considerations for future research relevant to the study purpose.

Chapter 5: Discussion, Conclusions, and Recommendations

Introduction

I conducted this ecologically designed, quantitative research study to examine the relationship between residential proximity to HWS (remote exposure) and the abnormal glycohemoglobin (A1c) values in a U.S. population ages 12 years and older. I used archival data from NHANES public use files and restricted NHANES data (residential geocoordinates of respondents) for survey cycles between 2005 and 2012. I used HWS data compiled from the NPL list of Superfund Sites for the states of New York, New Jersey Pennsylvania, and California. Using a convenience sample, I focused on two different residential groups and investigated two questions.

As expressed in Chapter 2, this study was guided by literary evidence indicating the growing global epidemic of T2DM (CDC, 2018; Pinhas-Hamiel & Zeitler, 2005); the increasing concern of POPs and EDCs in association with endocrine related disorders such as diabetes (Bijlsma & Cohen, 2016; Codru et al., 2007; Environmental Work Group, 2011; Gore, 2015; Lee et al., 2014; Taylor et al., 2013); and the changing face of T2DM as diagnosed cases emerge in adolescents (Baker et al., 2015; Dabela et al., 2014;), thus, altering the historical presentation of T2DM as an adult-onset disease. After analysis, I retained the null hypothesis for RQ. I rejected the null hypothesis and accepted the alternative hypothesis for RQ2.

Interpretation of Findings

Key Findings of the Study

Key findings of this investigation include that there was no statistically significant mean group difference in abnormal A1c between residential groups (those residing > 1 mile compared to those residing ≤ 1 mile of a HWS (Table 4). However, study participants with an abnormally high BMI (≥ 28.95 kg/m^2) or advanced age ≥ 58 years had a statistically significant greater risk or probability of having an abnormal A1c outcome. Specifically, study participants classified as being overweight or obese (BMI ≥ 28.95 kg/m^2) were 1.8 times and persons ≥ 58 years of age were 2.1 times more likely to have an A1c ≥ 5.7%, which fits within the parameters of prediabetes and T2DM (Table 7). These study results support current literature, which documents obesity and increased age as risk factors for pre-diabetes and T2DM (https://www.cdc.gov/diabetestv/risk-factors.html; Rodriguez & Campbell, 2017; Kulick et al., 2016).

In addition, I demonstrated that for the study population, residential proximity to a HWS and outcome of abnormal A1c was affected by race/ethnicity with statistical significance. Thus, implicating race/ethnicity as a moderator in the relationship between remote exposure to a HWS and A1c ≥ 5.7%. Participants of non-Hispanic White race/ethnicity residing > 1 mile from a HWS within their counties of residency had an 82.1% reduced risk for an abnormal A1c outcome than non-Hispanic Whites residing ≤ 1 mile of a HWS (Table 10). I assert that this study is the first of its kind to implicate the influence of race/ethnicity in relationship to residential proximity (or remote exposure) to a hazardous waste within the same county of residency and T2DM outcomes. I

anticipated that the further away a subject resided from a HWS, the less might be their risk for abnormal A1c. This finding is supported by other studies that suggest risk of disease outcome is associated with proximity to the HWS. For example, Kouznetsova et al. (2006) reported increased rates of hospitalizations for diabetes among persons residing in ZIP codes containing POP sites in comparison to "clean" sites. In my study, "clean" sites would be represented by remote exposure >1 mile. In another study, Liu and Carpenter (2012) concluded an increased risk of hospitalization rates for acute respiratory infection and chronic obstructive lung disease (COPD) amongst participants living in a ZIP code with a HWS. Theoretically, this aides in supporting the validity of my investigation in that literature with similar environmental exposure studies link proximity to a HWS with increased risk for disease outcome yielding the reverse (further distance) with reduced disease risk.

Given that race/ethnicity was statistically significant as a moderator, hypothetically, if non-Hispanic Whites residing > 1 mile from a HWS had a significant reduction in T2DM risk, I would expect a similar outcome in T2DM risk reduction for Mexican Americans, other Hispanics, non-Hispanic Blacks, and Multi- racial study subjects residing > 1 mile from a within county HWS. The fact that this was not the result, may speak to the influence of other risk factors (i.e., obesity, advancing age, and genetic predisposition) as reported in the literature (see Chapter 2) and reflected in my study results (see Table 7) on outcomes of A1c $\geq$ 5.7%. In addition, because of the ubiquitous nature of POPs, I was unable to control for all pathways of direct or remote

exposures the study subjects might have encountered in their lifetime. This may have also influenced the race/ethnicity moderation effect in this study.

Support for use of NPL Superfund Sites in This Study

By definition, hazardous wastes sites on the NPL list of Superfund Sites (*https://www.epa.gov/superfund*) are areas across the United States containing chemicals with suspected risks for human health. Based on the NPL list at the close of November 2016, I submitted to the RDC an Excel spread sheet containing a list of HWS by county for each state used in my study, along with the HWS geocoordinates. The list included 98 Superfund sites from California, 114 Superfund sites from New Jersey; 95 Superfund sites from Pennsylvania and 86 Superfund sites from New York. Based on the reported number of counties per state in Nov. 2016, the 98 HWS of California represented 58 counties for that state. The 114 New Jersey HWS represented 21counties; the 95 New Jersey HWS represented 67 counties and the 86 New York HWS represented 62 counties.

Due to the risk of disclosure (as discussed in Chapters 1 and 3), the exact number of counties included in my study per state were not made available to me. I predicted that fewer subjects would live near a HWS within their county. However, the fact that at least 48 subjects did reside ≤ 1 mile of a HWS; and the fact that all of the HWS included in this study (based on review of the site narratives – see examples Appendix D) contained some level of pollutants considered harmful to human health (such as volatile organic compounds, vinyl chloride, 1,1,1-trichloroethane (TCA), trichloroethene (TCE), arsenic, etc.) as per the ATSDR 2017 Completed Exposure Pathway Report (*https://www.atsdr.cdc.gov/cep/index.html*); and the fact that HWS on the NPL Superfund

list require continued monitoring, supports the literature discussed in Chapter 2 which extends concerns about the possibility of adverse health effects related to potential human exposure to POPs, even remotely, given that there is still a degree of unknown with respect to the exact concentration of POPs considered safe (X. Liu et al., 2012; Navas-Acien et al., 2008; Sergeev et al., 2011; Tseng et al., 2002).

Retention of the Null Hypothesis

Retention of the null hypothesis for RQ1, supports recommendations as expressed by Liese et al. (2010) to strengthen the precision of distance-based analysis using geocoding and census tract data. Though I used geocoordinates of the HWS and residential geocoordinates of the study subjects, the study was limited by not being able to calculate the miles in distance from the borders of the HWS to the borders of the subject's residence. Given the length or distance in miles of a HWS as reported in the NPL site narrative, I could calculate an estimate of the HWS area and subsequent radial miles (i.e., miles from HWS midpoint to its borders).

However, the distance from the midpoint of the residential geocoordinates to its borders could not be incorporated in this study as part of the distance based analysis. The area or square miles of the subject's residence was not a collected variable in the NHANES data file. It is plausible that if this variable was available on both sides (for the HWS and the residence of the study subjects), this may have enhanced the distance based analysis by creating a border–to-border difference in simple distance miles using geocoordinates. This would respond to the call by Liese and colleagues to strengthen the remote exposure variable and possibly strengthen the moderation effects, especially given

the statistically significant outcomes in the logistic analysis of my study that indicate subjects with abnormally high BMI and subjects with advanced aged are at increased risk for abnormal A1c outcomes, and prior research (as discussed in Chapter 2) citing adipose tissue as a depository for POPs and EDCs.

Limitations of the Study

There were several limitations in this study that restricted my ability to display frequency data or create graphics of data for discussion. The foundation for these limitations were risk of disclosure of the NHANES survey respondents.

Given that public use NHANES data and HWS geocoordinates from the NPL list of Superfund sites were merged with residential latitude and longitude of NHANES survey respondents, the RDC proposal review committee raised concerns about the risks of me being able to identify or suggest a specific geographic location of the NHANES respondents beyond what was already known, which was that the study subjects resided in a county from either New Jersey, New York, Pennsylvania or California that contained a HWS. This limited my residential groups to two large aggregate groups. Regardless of the state of residence, all subjects categorized as residing ≤ 1 mile from a HWS within their county of residency were placed in one residential group and all others of the study sample where categorized as residing > 1 mile of a HWS within their county of residency and grouped together.

The aggregation of subjects into two large groups limited my ability to conduct analysis on mean group differences between counties within the same state or between states as no more than two groups could be created, thus narrowing down the statistical

analysis to the Mann-Whitney *U* (given the non-normal distribution of the study sample). This is significant because, per the RDC analyst, some respondents lived in counties containing more than one HWS. If I had the RDC permission to graphically display in a frequency table or scatter plot the raw scaled A1c values and remote exposure variable between smaller groups (within the same county), this may have added to the discussion on the significance or non-significance of residential proximity to one or more than one HWS and the mean group differences in A1c values ≥ 5.7%.

Given the non-probability sampling methods I applied to gather the study population, statistical inferences could only be made on the study sample. This was a proposed limitation and limits the generalizability of the study results. Due to disclosure risk that might have identified the geographic location of a specific survey respondent, the exact HWS county matched to any subject's county of residence was restricted from viewing and restricted from my NHANES data file containing merged NHANES public use data and the categorical remote exposure data. As a result, I was not able to match specific HWS narratives (from the NPL Superfund list) to the HWS used in my study. This limited my ability to present the specific contaminants of the HWS involved in my study. Having this information may have added to the discussion on POPs and EDCs and possible geographic concentrations of contaminants with respect to remote exposure and A1c outcomes, as well as the strength and challenges of Superfund site clean-ups.

Although I could view (in my prepared RDC data file), the values for original NHANES scaled/continuous variables from the public use data file (for example age and A1c% values) publication of small cell sizes was not permitted due to disclosure risk.

This limited the publication of frequency tables or scatter plots on original NHANES scaled variables (such as the distribution of A1c values by age). And, given that the raw data indicating the difference in geocoordinates between the respondent's residence and the HWS was not disclosed, this limited some graphic representation of the study variables but did not take away from the statistical analysis in response to the research questions. Categorical variable frequency data relevant to the study were reviewed and approved by the RDC analyst for publication as displayed in the results chapter. In addition, this study was based on remote exposure and did not include the investigation of serum concentrations of POPs or EDCs in the study subjects.

Recommendations

While the limitations of the study were significant, they were for the protection of the study subjects. However, the limitations and strengths of this study carve out avenues for further research. This study was strengthened by using geocoordinates to form the remote exposure variable (i.e. residential distance in miles from a HWS within the community of residency). Liese et al. (2010), used geocoding in their study on geographic variation in Type 1 and Type 2 diabetes and concluded that analysis using geocoding and census tract rather than ZIP code level data, was the most unbiased type of distance based analysis. The creation of the remote exposure variable for this study was limited to simple distance calculation based on the difference in latitude and longitude coordinates of the HWS and residential latitude and longitude coordinates of the study subjects. As I proposed, these geocoordinates were considered to be at the center of each location. Future researchers may benefit from the inclusion of the area radius taken from

each location (i.e., HWS and residence of the subject) to create a remote exposure variable that considers the distance between the geographic borders of each site to further fine-tune this distance. The radial distance for this study could only be created with the HWS data as the residential property area of the subjects was not collected by NHANES.

The name of the HWS involved in this study as well as how many HWS were included from each state or county was not available to me due to disclosure risk. This limited my ability to highlight the contaminants contained at a specific site. Due to disclosure risk, I was also not permitted to display frequency tables on continuous A1c values (as afore discussed), nor did I have access to the remote exposure raw data, thus disabling my ability to create a scatter plot of scaled remote exposure geocoordinates against scaled A1c outcomes.

These limitations did not exclude the premise that HWS on the NPL Superfund list are there because of the potential threat the contaminants at these sites pose to human health (*https://www.epa.gov/superfund/superfund-national-priorities-list-npl*) and thus, potential contribution to the T2DM epidemic in the United States. However, the limitations did disable me from conducting a more finite observation of the relationship between the independent variable (remote exposure) and the outcome variable (A1c values $\geq$ 5.7%), given that these variables were categorical in this study and not continuous (for reasons described in Chapter 3). In addition, the study limitations disabled me from comparing observations of POPs and EDCs (based on the description of site contaminants in the HWS narratives) to A1c values in conjunction with residential distance of the study subjects from the HWS. Future researches may benefit from

correlating the number of HWS and contents of HWS with scaled residential geocoordinates and scaled A1c observations to broaden the investigation of a relationship between residential proximity to a HWS and abnormal A1c values ≥ 5.7%. My study also lends itself to future research that includes comparing the mean difference A1c values ≥ 5.7% between remote exposure groups within the same state or between states to identify any potential geographic association or interrelationship between proximity to a HWS and T2DM, thereby furthering investigations grounded in one of the core constructs of the ecosocial theory, i.e., pathways of embodiment [of disease] (Krieger, 2011).

My study was based on remote exposure and did not include the investigation of serum concentrations of POPs or EDCs in the study subjects. This was a proposed limitation. While future research may be enhanced by the application of serum measurements for POPs and EDCs in study subjects, this may be challenging economically and practically for some researchers. Strengthening the ecological study design with more scaled variables, may enhance the opportunity for more robust statistical analysis. The generalizability of results from ecological studies will nearly always be a factor in the discussion of results. However, the feasibility of these studies, particularly with the use of archival data, allows for the recycling of data for good use in the promotion of public health research.

Implications for Positive Social Change

For several decades, incidence of T2DM has been the highest among adult populations. However, the global face of this disease has drastically changed over the last 10 years to include an increased incidence of new cases in persons younger than 20 years

of age. For example, in the Diabetes 2017 Report Card (CDC, 2018), the CDC reported that of the 30.3 million people of the U.S. population diagnosed with diabetes in 2015, 30.2 million (or 12.2% of all U.S. adults) were adults aged 18 or older (Figure 4). The Diabetes 2017 Report also stated an estimated 34% of all U.S. adults in 2015 had prediabetes (defined in my study as an A1c of 5.8 –to 6.4%), and that most were not aware. With respect to children and adolescents, the Diabetes 2017 Report described that more than 5,000 youths aged 10-19 years were newly diagnosed with T2DM between 2011 and 2012, with the highest incidence rates among people of color (Figure 5).

Indeed, my study sample parallels the dynamics of T2DM in the general population and the significance of T2DM as a public health care concern. All subjects in my sample had an A1c value that met the definition of either pre-diabetes or T2DM, and these did include persons younger than 20 years of age (although a small n = 4.8% of the study population). There was a near even distribution of males and females in the study sample and race/ethnicity was well represented in the sample with non-Hispanic Whites representing 24.7% of the population and the remainder persons of color (see Table 1). Given the prevalence of diabetes as reported in the literature, the ubiquitous nature of EDCs, POPs and their effect on endocrine disorders (Gore et al., 2015), public health researchers should be compelled to seek out potential environmental influences that maybe impacting the global diabetes epidemic.

Influenced by the knowledge gap left by Kouznetsova et al. (2006) concerning the potential association between residential proximity to HWS and T2DM as well as the ecosocial theory of disease distribution (Krieger, 2011), the social change significance of

my study was to provide data on A1c outcomes between two different residential groups, those residing ≤ 1 mile from a HWS in their county of residency and those residing > 1 mile from a county based HWS. Using geocoordinates to create the remote exposure variable, the study goals fulfilled were the investigation of mean group differences in A1c values ≥ 5.7% between residential groups and effect of abnormal BMI, age, sex, and race/ethnicity on the relationship between remote exposure or residential proximity to a HWS and pre-diabetes or diabetes risk (i.e., abnormal A1c ≥ 5.7%). Extending the diabetes literature with knowledge on remote exposure to HWS and A1c outcomes in populations 12 years and older, has the probability to enhance the understanding of potential environmental influences on the diabetes disease burden in the United States. Such knowledge, then has the potential to affect social change through diabetes health assessments in clinical practice, public health education about diabetes risk factors, children's environmental health and raise dialogue about social justice and the geographic distribution of HWS.

For this study population, subjects with an abnormally high BMI (≥ 28.95 kg/m^2) were 1.8 times and persons ≥ 58 years of age were 2.1 times more likely to have an A1c ≥ 5.7%. This parallels the current literature indicating abnormal BMI and advancing age as risk factors for pre-diabetes and T2DM. These results foster the continued public health education efforts regarding reduction of T2DM risk and thus, the global disease burden of T2DM.

In addition, non-Hispanic whites in my study population residing > 1 mile of a hazardous was site, demonstrated a statistically significant 82.1% reduced risk for

abnormal A1c compared to participants of the same race residing < 1 mile of a HWS. For this study population, this result implicates race/ethnicity as a moderator in the relationship between remote exposure to HWS and abnormal A1c values reflective of pre-diabetes and T2DM. Further, with non-Hispanic Whites representing 24.7% of this study population, yet with an 82.1% reduced risk for abnormal A1c for those > 1 mile of a HWS, raises the question of 'whose back yard' are HWS in? I propose that these results may challenge professionals involved in social environmental justice, invoking discussions, policies, and practices regarding the distribution of HWS and may further the curiosity of researchers to explore what findings might be discovered in the exploration of HWS proximity and T2DM relevant to social environmental justice given an enhanced remote exposure variable as aforementioned.

Conclusion

The findings of my investigation, extrapolated to the study population, support that for the study population ages 12 years and older, the moderator race/ethnicity had a statistically significant effect on the relationship between residential proximity to a HWS and abnormal A1c%. Specifically, the results indicated that non-Hispanic Whites residing > 1 mile of a HWS, had 82.1% reduced risk for abnormal A1c compared to participants of the same race residing < 1 mile of a HWS. In addition, study participants with an abnormally high BMI classified as being overweight or obese (BMI $\geq$ 28.95 kg/m^2) were 1.8 times and study subjects $\geq$ 58years of age were 2.1 times more likely to have an A1c $\geq$ 5.7%, which fits within the parameters of pre-diabetes or T2DM.

This study supports the significance of the ecosocial theory and its construct of embodiment of disease and core proposition of determining patterns of disease distribution (Kreiger, 2012). The literature is clear on the ubiquitous existence of POPs and EDCs in the environment (Bijlsma & Cohen, 2016) and the threat and effect of these contaminants on human health and disease (Environmental Work Group, 2011; Tyrrell et al., 2013; United Nations Environmental Programme, 2012), as well as the unknown about *safe* environmental limits of POPs (Codru et al., 2007; Stockholm Convention, 2016a).

I proposed that this was the first study of its kind to investigate remote residential exposure to HWS and T2DM inclusive of pre-diabetes and T2DM diagnostic lab values (i.e. A1c $\geq$ 5.7%) in a population 12 years and older. A single study, however, is not conclusive evidence of significant relationships between residential proximity to HWS and T2DM outcomes. This study forges opportunities for additional analysis with an enhanced remote exposure variable, using geocoordinates that consider border distance between residence and the HWS.

References

Anderson, Z., Raaschou-Nielsen, O., Ketzel, M., Jensen, S., Hvidberg, M., Loft, S.,…Sørensen, M. (2012). Diabetes incidence and long-term exposure to air pollution. *Diabetes Care, 35,* 92-98. doi:10.2337/dc11-1155

Andre, A. S., Burgess, J. L., Meza, M. M., Demidenko, E., Waugh, M. G., Hamilton, J. W., & Karagas, M. R. (2006). Arsenic exposure is associated with decreased DNA repair in vitro and in individuals exposed to drinking water arsenic. *Environmental Health Perspectives*, *114*, 1193-1198. doi:10.1289/ehp.9008

Aylward, L. L., Brunet, R. C., Carrier, G., Hays, S. M., Cushing, C. A., Needham, L. L., & ... Mocarelli, P. (2005). Concentration-dependent TCDD elimination kinetics in humans: toxicokinetic modeling for moderately to highly exposed adults from Seveso, Italy, and Vienna, Austria, and impact on dose estimates for the NIOSH cohort. *Journal Of Exposure Analysis And Environmental Epidemiology*, *15*, 51-65. doi:10.1038/sj.jea.7500370

Aylward, L. & Hays, S. (2002). Temporal trends in human TCDD body burden: Decreases over three decades and implications for exposure levels. *Journal of Exposure Analysis and Environmental Epidemiology, 12*, 319-328. doi:10.1038/sj.jea.7500233

Baguley, T. (2009). Standardized or simple effect size: What should be reported*? British Journal of Psychology, 100,* 603-617. doi:10.1348/000712608X377117

Baker, S., Barlow, S., Cochran, W., Fuchs, G., Klish, W., Krebs, N.,… Udall, J. (2005). Overweight children and adolescents: A clinical report of the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition. *Journal of Pediatric Gastroenterology Nutrition, 40,* 533-543. doi:10.1097/01.MPG.0000161147.16590.12

Benachour, N., Moslemi, S., Sipahutar, H., & Seralini, G. (2007). Cytotoxic effects and aromatase inhibition by xenobiotic endocrine disrupters alone and in combination. *Toxicology And Applied Pharmacology*, *222*, 129-140. doi:10.1016/j.taap.2007.03.033

Bennett, D. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health, 25*: 464-469. doi:10.1111/j.1467-842X.2001.tb00294.x

Bertazzi, P., Bernucci, I., Brambilla, G., Consonni, D., & Pesatori, A. (1998). The Seveso studies on early and long-term effects of dioxin exposure: A review. *Environmental Health Perspectives, 106*, 625-633. doi:10.1289/ehp.98106625

Bertazzi, P. A., Consonni, D., Bachetti, S., Rubagotti, M., Baccarelli, A., Zocchetti, C., & Pesatori, A. C. (2001). Health effects of dioxin exposure: A 20-year mortality study. *American Journal of Epidemiology*, *153*, 1031-1044. doi:10.1093/aje/153.11.1031

Bijlsma, N., and Cohen, M. (2016). Environmental chemical assessment in clinical practice: unveiling the elephant in the room. *Australian College of Nutritional & Environmental Medicine (ACNEM) Journal, 35*, 3-17. Retrieved from https://ezp.waldenulibrary.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=117004351&site=ehost-live&scope=site

Boberg, E., Lessner, L., & Carpenter, D. (2011). The role of residence near hazardous waste sites containing benzene in the development of hematologic cancers in Upstate New York. *International Journal of Occupational Medicine and Environmental Health, 24*, 327-338. doi:10.2478/s13382-011-0037-8

Breton, C. V., Mack, W. J., Yao, J., Berhane, K., Amadeus, M., Lurmann, F.,… Avol, E. (2016). Prenatal air pollution exposure and early cardiovascular phenotypes in young adults. *Plos ONE*, *11*, 1-12. doi:10.1371/journal.pone.0150825

Carpenter, D., Ma, J., Lessner, L. (2008). Asthma and infectious respiratory disease in relation to residence near hazardous waste sites. *Annals N.Y. Academy of Science, 1140,* 201-208. doi:10.1196/annals.1454.000

Centers for Disease Control and Prevention [CDC]. (2016a). National Health and Nutrition Examination Survey 1999-2016 Survey
Content Brochure. Retrieved from https://www.cdc.gov/nchs/data/nhanes/survey_content_99_16.pdf

Centers for Disease Control and Prevention. (2016b). NCHS, National Health and Nutrition Examination Survey. Questionnaires, datasets, and related documentation. Retrieved from https://wwwn.cdc.gov/nchs/nhanes/default.aspx

Centers for Disease Control and Prevention [CDC]. (2016c). Key concepts about NHANES survey design. Retrieved from http://www.cdc.gov/nchs/tutorials/nhanes/SurveyDesign/SampleDesign/Info1.htm

CDC/National Center for Health Statistics. (2016d). About the National Health and Nutrition Examination Survey. Retrieved from http://www.cdc.gov/nchs/nhanes/about_nhanes.htm

Centers for Disease Control and Prevention. (2016e). National Center for Health Statistics, NCHS Research Data Center [RDC]. Retrieved from http://www.cdc.gov/rdc

Centers for Disease Control and Prevention. (2018). *Diabetes Report Card 2017*. Atlanta, GA: Centers for Disease Control and Prevention, US Dept of Health and Human Services. Retrieved from www.cdc.gov/diabetes/library/reports/congress.html

Codru, N., Schymura, M., Negoita, S., The Akwesasne Task Force on the Environment, Rej, R., & Carpenter, D. (2007). Diabetes in relation to serum levels of polychlorinated biphenyls and chlorinated pesticides in adult Native Americans. *Environmental Health Perspectives, 115*,1442-1447. doi:10.1289/ehp.10315

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Multiple regression/correlation analysis for the behavioral sciences, 3rd Ed.* Mahway, NJ: Lawrence Erlbaum Associates.

Collaborative on Health and the Environment (2014). CHE Toxicant and Disease Database. Retrieved from https://www.healthandenvironment.org/

Cranmer, M., Louie, S., Kennedy, R., Kern, P., & Fonseca, V. (2000). Exposure to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) is associated with hyperinsulinemia and insulin resistance. *Toxicology Sciences 56*, 431-436. doi:10.1093/toxsci/56.2.431

Creswell, J. (2008). *Educational research: Planning, conducting and evaluating quantitative and qualitative research, 3rd Ed.* Columbus, OH: Pearson.

Dabelea, D., Bell, R.A., D'Agostino, R. B. Jr., Imperatore, G., Johansen, J.M., Linder, B.,… Waitzfelder, B. (2007). Incidence of diabetes in youth in the United States. *Journal of the American Medical Association 297*, 2716-2724. doi:10.1001/jama.297.24.2716

Dabela, D., Mayer-Davis, E., Savdah, S., Imperatore G., Linder, B., Divers, J.,… Liu, L. (2014). Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. *Journal of the American Medical Association, 311,* 1778-1786. doi:10.1001/jama.2014.3201

Eeg-Olofsson, K., Cederholm, J., Nilsson. P.M., Zethelius, B., Svensson, A., Gudbjörnsdóttir, S., & Eliasson, B. (2010). New aspects of HbA1c as a risk factor for cardiovascular disease in type 2 diabetes: an observational study from Swedish National Diabetes Register (NDR). *Journal of Internal Medicine, 268*, 471-482. doi:10.1111/j.1365-2796.2010.02265.x

Environmental Work Group (2011). EPA must finish the job of protecting people from dioxin: dioxin research. Retrieved from http://www.ewg.org/research/epa-must-finish-job-protecting-people-dioxin/dioxin-research

Environmental Protection Agency (2016). Resource Conservation and Recovery Act of 1976, Biennial Report Overview. Retrieved from https://catalog.data.gov/dataset/national-rcra-hazardous-waste-biennial-report-data-files

Færch, K., Højlund, K., Vind, B., Vaag, A., Dalgård, C., Nielsen, F.,... Grandjean, P. (2012). Increased serum concentrations of persistent organic pollutants among prediabetic individuals: potential role of altered substrate oxidation patterns. *Journal Of Clinical Endocrinology & Metabolism*, *97*, E1705-13. doi:10.1210/jc.2012-1342

Fierens, S., Mairesse, H., Heilier, J., Burbure, C., Focant, J., Eppe, G., De Pauw, E., & Bernard, A. (2003). Dioxin/polychlorinated biphenyl body burden, diabetes and endometriosis: findings in a population-based study in Belgium. *Biomarkers, 8,* 529-534. doi:10.1080/1354750032000158420

Friis, R., & Sellers, T. (2004*). Epidemiology for public health practice*. Sudbury, MA: Jones and Bartlett.

Goncharov, A., Bloom, M., Pavuk, M., Birman, I., & Carpenter, D. (2010). Blood pressure and hypertension in relation to levels of serum polychlorinated biphenyls in residents of Anniston, Alabama. *Journal of Hypertension, 28*, 10, 2053-2060. doi:10.1097/HJH.0b013e32833c5f3e

Gore, A. C., Chappell, V. A., Fenton, S. E., Flaws, J. A., Nadal, A., Prins, G. S.,… Zoeller, R. T. (2015). Executive summary to EDC-2: The endocrine society's second scientific statement on EDCs. *Endocrine Reviews 36, 593-602.* doi:10.1210/er.2015-1093.

Hall, J., McLaughlin, M. & Stamm, S. (1975). Coarctation of the aorta in male cousins with similar maternal environmental exposure to insect repellent and insecticides. *Pediatrics, 55*, 425-427. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/?term=Coarctation+of+the+Aorta+in+Male+Cousins+With+Similar+Maternal+Environmental+Exposure+to+Insect+Repellent+and+Insecticides

Harari, R., Julvez, J., Murata, K., Barr, D., Bellinger, D., Debes, F. & Grandjean, P. (2010). Neurobehavioral deficits and increased blood pressure in school-age children prenatally exposed to pesticides. *Environmental Health Perspectives, 118*, 890-896. doi:10.1289/ehp.0901582

Hayes, A. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach, 2nd Ed*. New York, NY: The Guilford Press.

Hazra, A. & Gogtay, N. (2017). Biostatistics series module 10: Brief overview of multivariate methods. *Indian Journal of Dermatology, 62, 358-366*. Retrieved from http://www.e-ijd.org/text.asp?2017/62/4/358/210078

Hermanson, M. H. & Johnson, G.W. (2007). Polychlorinated biphenyls in tree bark near a former manufacturing plant in Anniston, Alabama. *Chemosphere, 68*, 191-198. doi:10.1016/j.chemosphere.2006.11.068

Hill, Austin Bradford (1965). The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine, 58*, 295-300. Retrieved from http://www.edwardtufte.com/tufte/hill

Kline, Theresa J. B. (2017). Sample issues, methodological implications and best practices. *Canadian Journal of Behavioural Science, 49,* 71-77. Retrieved from http://psycnet.apa.org/buy/2017-20844-001

Kouznetsova, M., Huang, X., Ma, J., Lessner, L. & Carpenter, D. (2006). Increased rate of hospitalization for diabetes and residential proximity of hazardous waste sites. *Environmental Health Perspectives 115*, 75-79. doi:10.1289/ehp.9223

Kramarova, E., Kogevinas, M., Anh, C., Cau, H., Dai, L., Stelman, S. & Parkin, D. (1998). Exposure to Agent Orange and occurrence of soft-tissue sarcomas or non Hodgkin lymphomas: an ongoing study in Vietnam. *Environmental Health Perspectives 106*, 671-678. doi:10.1289/ehp.106-1533419

Krieger, N. (1994). Epidemiology and the web of causation: Has anyone seen the spider? *Social Science & Medicine, 39*, 887-903. doi:10.1016/0277-9536(94)90202-X

Krieger, N. (2001). A glossary for social epidemiology. *Journal of Epidemiology And Community Health*, *55*, 693-700. doi:10.1136/jech.55.10.693

Krieger, N. (2012). Methods for the Scientific Study of Discrimination and Health: An Ecosocial Approach. *American Journal of Public Health 102*, 936-945. doi:10.2105/AJPH.2011.300544

Krieger, N (2011). *Epidemiology and the people's health: Theory and context*. New York, NY: Oxford University Press.

Kulick, E., Moon, Y., Cheung, K., Willey, J., Sacco, R., & Elkind, M. (2016). Racial-ethnic disparities in the association between risk factors and diabetes: The Northern Manhattan study. *Preventive Medicine, 83*, 31-36. doi:10.1016/j.ypmed.2015.11.023

Lawson, C., Schnorr, T., Whelan, E., Deddens, J., Dankovic, D., Piacitelli, L.,... Connally, L. (2004). Paternal occupational exposure to 2,3,7,8 tetrachlorodibenzo-p-dioxin and birth outcomes of offspring: birth weight, preterm delivery, and birth defects. *Environmental Health Perspectives*, *112*, 1403-1408. doi:10.1289/ehp.7051

Lee, D., Lee, I., Song, K., Steffes, M., Toscano, W., Baker, B., & Jacobs, D. R., Jr. (2006). A strong dose-response relation between serum concentrations of persistent organic pollutants and diabetes: results from the National Health and Examination Survey 1999-2002. *Diabetes Care, 29,* 1638-1644. doi:10.2337/dc06-0543

Lee, D., Porta, M., Jacobs, D. Jr., & Vandenberg, L. (2014). Chlorinated persistent organic pollutants, obesity, and type 2 diabetes. *Endocrine Reviews, 35*, 557-601. doi:10.1210/er.2013-1084

Lee, D., Steffes, M., Sjödin, A., Jones, R., Needham, L., & Jacobs, D. (2010). Low dose of some persistent organic pollutants predicts type 2 diabetes: A nested case-control study. *Environmental Health Perspectives, 118,* 1235-1242. doi:10.1289/ehp.0901480

Liem, A., Furst, P., & Rappe, C. (2000). Exposure of populations to dioxins and related compounds. *Food Additives and Contaminants, 4,* 241-259. doi:10.1080/026520300283324

Liese, A., Lawson, A., Song, H., Hibbert, J., Porter, D., Nichols, M.,… D'Agostino, R. (2010). Evaluating geographic variation in Type 1 and Type 2 diabetes mellitus incidence in youth in four U.S. regions. *Health Place, 16,* 547-556. doi:10.1016/j.healthplace.2009.12.015

Lind, P., Bavel, B., Salihovic, S. & Lind, L. (2012). Circulating levels of persistent organic pollutants (POPs) and carotid atherosclerosis in the elderly. *Environmental Health Perspectives, 120*, 38-43. doi:10.1289/ehp.1103563

Liu, C., Ying, Z., Harkema, J., Sun, Q. & Sanjay, R. (2013). Epidemiological and experimental links between air pollution and type 2 diabetes. *Toxicologic Pathology, 41,* 361-373. doi:10.1177/0192623312464531

Liu, X., Lessner, L., & Carpenter, D. (2012). Association between residential proximity to fuel-fired power plants and hospitalization rate for respiratory diseases. *Environmental Health Perspectives, 120*, 807-810. doi:10.1289/ehp.1104146

Longnecker, M., & Daniels, J. (2001). Environmental contaminants as etiologic factor of diabetes. *Environmental Health Perspective*, *109*, 871-876. doi:10.2307/3454649

Martuzzi, M., Mitis, F., & Forastiere, F. (2010). Inequalities, inequities environmental justice in waste management and health. *European Journal of Public Health, 20*, 21-26. doi:10.1093/eurpub/ckp216

Méndez-Gómez, J., García-Vargas, G., López-Carrillo, L., Calderón-Aranda, E., Gómez, A., Vera, E.,… Rojas, E. (2008). Genotoxic effects of environmental exposure to arsenic and lead on children in Region Lagunera, Mexico. *Annals of The New York Academy of Sciences, 1140,* 358-367. doi:10.1196/annals.1454.027

Merrill, M. L., Emond, C., Kim, M., Antignac, J., Bizec, B. L., Clement, K.,… Barouki, R. (2013). Toxicological function of adipose tissue: Focus on persistent organic pollutants. *Environmental Health Perspectives, 121,* 162-169. doi:10.1289/ehp.1205485

Mohai, P., Lantz, P., Morenoff, J., House, J., & Mero, R. (2009). Racial and socioeconomic disparities in residential proximity to polluting industrial facilities: Evidence from the Americans' Changing Lives Study. *American Journal of Public Health, 99*, 649-656. doi:10.2105/AJPH.2007.131383

Morris, D., Khunti, K., Achana, F., Srinivasan, B., Gray, L., Davies, M., Webb, D. (2013). Progression rates from HbA1c 6.0-6.4% and other prediabetes definitions to type 2 diabetes: a meta-analysis. *Diabetologia, 56*, 1489-1493. doi:10.1007/s00125-013-2902-4

Nathan, D., Davidson, M., DeFronzo, R., Heine, R., Henry, R., Pratley, R., Zinman, B. (2007). Impaired fasting glucose and impaired glucose tolerance: Implications for care. *Diabetes Care, 30,* 753-759. doi:10.2337/dc07-9920

National Institute of Diabetes and Digestive and Kidney Diseases (2018). Diabetes Test & Diagnosis. National Institutes of Health. Retrieved from https://www.niddk.nih.gov/health-information/diabetes/overview/tests-diagnosis

Navas-Acien, A., Silbergeld, E., Pastor-Barriuso, R., & Guallar, E. (2008). Arsenic exposure and prevalence of type 2 diabetes in US adults. *Journal American Medical Association, 300*, 814-822. doi:10.1001/jama.300.7.814

Nichols, G., McBurnie, M., Paul, L., Potter, J., McCann, S., Mayer, K.,… DeVoe, J. (2016). The high prevalence of diabetes in a large cohort of patients drawn from safety net clinics. *Preventing Chronic Disease, 13,* 1-8. doi:10.5888/pcd13.160056

Oluoch-Otiego, J., Oyoo-Okoth, E., Kiptoo, K., Chemioiwa, E., Ngugi, C., Simiyu, G.,… Opiyo, M.A. (2016). PCBs in fish and their cestode parasites in Lake Victoria. *Environmental Monitoring and Assessment, 188*, 1-11. doi:10.1007/s10661-016 5483-0

Perreault, L., & Faerch, K. (2014). Approaching pre-diabetes. *Journal of Diabetes and Its Complications, 28,* 226-233. doi:10.1016/j.jdiacomp.2013.10.008

Pinhas-Hamiel, O., & Zeitler, P. (2005). The global spread of type 2 diabetes mellitus in children and adolescents. *The Journal of Pediatrics, 146*, 693-700. doi:10.1016/j.jpeds.2004.12.042

Porta M, Bosch de Basea M, Benavides F., Lopez, T., Fernandez, E., Marco, E.,...Puigdomènech, E. (2008). Differences in serum concentrations of organochlorine compounds by occupational social class in pancreatic cancer. *Environmental Research, 108*, 370 –379. doi:10.1016/j.envres.2008.06.010

Portnov, B., Dubnov, J., & Barchana, M. (2007). On ecological fallacy, assessment errors stemming from misguided variable selection, and the effect of aggregation on the outcome of epidemiological study. *Journal of Exposure Science and Environmental Epidemiology, 17,* 106-121. doi:10.1038/sj.jes.7500533

Pratschke, J., Haase, T., Comber, H., Sharp, L., de Camargo Cancela, M., & Johnson, H. (2016). Mechanisms and mediation in survival analysis: towards an integrated analytical framework. *BMC Medical Research Methodology*, *16,* 1-13. doi:10.1186/s12874-016-0130-6

Reinehr, T. (2013). Type 2 diabetes mellitus in children and adolescents. *World Journal of Diabetes, 4,* 270-281. doi:10.4239/wjd.v4.i6.270

Rodriguez, J. & Campbell, K. (2017). Racial and ethnic disparities in prevalence and care of patients with type 2 diabetes. *Clinical Diabetes, 35,* 66-70. doi:10.2337/cd15-0048

Rosenbloom, A., young, R., Joe, J., & Winter, W. (1999). Emerging epidemic of type 2 diabetes in youth. *Diabetes Care*, *22*, 345-354. doi:10.2337/diacare.22.2.345

Santaguida, P., Balion, C., Hunt, D., Morrison, K., Gerstein, H., Raina, P.,… Yazdi, H. (2005). *Diagnosis, prognosis, and treatment of impaired glucose tolerance: Summary*. 2005 Aug. In AHRQ Evidence Report Summaries. Rockville (MD): Agency for Healthcare Research and Quality (US);1998-2005. 128. Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK11923

Schecter, A., Cramer, P., Boggess, K., Stanley, J., Papke, O., Olson, J.,… Schmitz, M. (2001). Intake of dioxins and related compounds from food in the U.S. population. *Journal of Toxicology Environmental Health, 63*, 1-18. doi:10.1080/152873901750128326

SEARCH (2010). A multicenter study of diabetes in children and youth. Centers for Disease Control and Prevention, National Institutes of Health. Retrieved from www.searchfordiabetes.org/index.cfm

Sergeev, A., & Carpenter, D. (2011). Geospatial patterns of hospitalization rates for stroke with comorbid hypertension in relation to environmental sources of persistent organic pollutants: Results from a 12-year population-based study. *Environmental Science and Pollution Research*, *18*, 576-585. doi 10.1007/s11356-010-0399-7

Suarez-Lopez, J., Lee, D., Porta, M., Steffes, M., and Jacobs, D. (2015). Persistent organic pollutants in young adults and changes in glucose related metabolism over a 23-year follow-up. *Environmental Research, 137*, 485-494. doi:10.1016/j.envres.2014.11.001

Shaw, J., Zimmet, P., de Courten, M., Dowse, G., Chitson, P., Gareeboo, H.,… Alberti, K. (1999). Impaired fasting glucose or impaired glucose tolerance: What best predicts future diabetes in Mauritius? *Diabetes Care, 22,* 399-402. doi:10.2337/diacare.22.3.399

Springer, S., Silverstein, J., Copeland, K., Moore, K., Prazar, G., Raymer, T.,… Flinn, S. (2013). Technical Report: Management of type 2 diabetes mellitus in children and adolescents. *Pediatrics*. doi:10,1542/peds.2012-3496

Stockholm Convention (2016a). The 12 initial POPs under the Stockholm Convention. Retrieved from http://chm.pops.int/TheConvention/ThePOPs/ListingofPOPs/tabid/2509/Default.aspx

Stockholm Convention (2016b). The new POPs under the Stockholm Convention. Retrieved from http://chm.pops.int/TheConvention/ThePOPs/TheNewPOPs/tabid/2511/Default.aspx

Stockholm Convention (2016c). What are POPs? Retrieved from http://chm.pops.int/TheConvention/ThePOPs/tabid/673/Default.aspx

Tabacova, S., Balabaeva, L., & Little R. (1997). Maternal exposure to exogenous nitrogen compounds and complications of pregnancy. *Archives of Environmental Health, 52*, 341-347. doi:10.1080/00039899709602209

Taylor, K., Novak, R., Anderson, H., Birnbaum, L., Blystone, C., Devito, M.,… Lind, L. (2013). Evaluation of the association between persistent organic pollutants (POPs) and diabetes in epidemiological studies: A National Toxicology Program workshop review. *Environmental Health Perspectives, 121,* 774-783. doi:10.1289/ehp.1205502

Thayer, K., Heindel, J., Bucher, J., & Gallo, M. (2012). Role of environmental chemicals in diabetes and obesity: A National Toxicology Program workshop report. *Environmental Health Perspectives, 120*, 779-789. doi:10.1289/ehp.1104597

Trochim, W. and Donnelly, J. (2007). *The research methods knowledge base, 3rd edition*. Mason, OH: Thompson Corporation.

Tyrell, J., Melzer, D., Henley, W., Galloway, T., Osborne, N. (2010). Association between socioeconomic status and environmental toxicant concentrations in adults in the USA: NHANES 2001-2010. *Environmental International, 59,* 328-335. Retrieved from https://doi.org/10.1016/j.envint.2013.06.017

Tseng, C., Tseng, C., Chiou, H., Hsueh, Y., Chong, C., Chen, C. (2002). Epidemiologic evidence of the diabetogenic effect of arsenic. *Toxicology Letters, 133*, 69-76. doi:10.1016/S0378-4274(02)00085-1

United Nations Environment Programme (UNEP). (2013). Global Chemicals Outlook. Towards sound management of chemicals. UNEP: Geneva, Switzerland. Retrieved from https://wedocs.unep.org/bitstream/handle/20.500.11822/8455/-Global%20chemicals%20outlook_%20towards%20sound%20management%20of%20chemicals-2013Global%20Chemicals%20Outlook.pdf?amp%3BisAllowed=&sequence=3

U. S. Environmental Protection Agency (2012). EPA's Reanalysis of Key Issues Related to Dioxin Toxicity and Response to NAS Comments, Volume 1. U.S EPA, Washington, DC. Retrieved from www.epa.gov/iris

U.S. Department of Health and Human Services (2016). Agency for Toxic Substances and Disease Registry (ATSDR). Retrieved from https://www.atsdr.cdc.gov

World Health Organization (2016). Global Report on Diabetes. p.11 and 27. World Health Organization, ISBN: 978 92 41565257. Retrieved from www.who.int

World Health Organization (WHO) (2013). Diabetes Fact Sheet No 312. Retrieved from http://www.who.int/mediacentre/factsheets/fs312/en/

WHO (2014). Diabetes Programme. Retrieved from http://www.who.int/diabetes/en/

Appendix A: Process for Creating Simple Distance Calculation and Merging of Hazardous Waste Site and NHANES Public Data

Prepared by RDC Analyst F. McCarty for Researcher Theresa Johnson
May 2018

The hazardous waste site data was provided by the researcher in a single excel file with multiple sheets, one for each state (n=4) included in the study. A file for each state was created by using SAS PROC IMPORT. After each state file was created, the files were converted from long format to wide format resulting in files with a single record for each county within a state. A series of variables, numbered sequentially, was created for each facility in the county.

The necessary geographic variables (state, county, latitude and longitude) were added to the public use data file provided by the researcher using the variable SEQN.

The hazardous waste data file was merged with the public use file containing the geographic variables using state and county.

For each waste site location, represented by a series of variables with a common numeric suffix, the distance (miles) from that location and the NHANES respondent residence was calculated using the GEODIST function in SAS. This function requires four arguments with the option of specifying the units for the distance measure. An example of the call would look like the following:
dist1=geodist(lat, lon, HWSLAT1, HWSLON1, 'M')

In this example, "dist1" would be the calculated distance in miles (M) given the latitude and longitude coordinates (lat, lon) for the residence and the coordinates for the waste site location (HWSLAT1, HWSLON1). This calculation was completed for each facility (1 to n) for each NHANES respondent.

Since each respondent could be associated with more than one waste site based on the state/county merge, the distance used for analysis was determined by using the smallest distance amongst the sites for that respondent. The following SAS code was used to identify the smallest distance and then this distance was used to calculate an indicator for <=1 mile (1) or >1 mile (0):
array values dist1-distN;
smallest = min(of values[*]);

The distance measure (dist1 or smallest) described above could be thought of as a simple distance between the two points.

An additional distance measure was computed by subtracting the radius from the distance measure (dist1) described above. The radius was first calculated using the following SAS code: Radmiles1=sqrt(AreaHWS_Miles21/3.14159). This calculation was completed for each site linked to the respondent.

The following formula was used to compute a new distance variable that was adjusted for the site radius: rrdist1=dist1-RadMiles1. As with previous computations, this computation was applied to all sites for the respondent.

Using the same methods as described above, the smallest distance was identified and then used to compute an additional indicator variable. The following code was used:

```
array values rrdist1-rrdistN;
rrsmallest = min(of values[*]);
```

As above, the rrsmallest distance shown above was used to calculate an indicator for <=1 mile (1) or >1 mile (0). It should be noted that in some cases the rrsmallest value was negative; cases with these values were included in the <=1 mile category.

The final analytic data file that was provided to the researcher included only the <=1 mile indicator variable and a flag variable that was created to indicate that the distance value used was negative (case for rrsmallest variable; note, negative values were not an issue for the simple distance calculation). Based on the approved RDC proposal, the final analytic file could not contain any other geographic information, only <=1 mile indicators.

SAS 9.4 was used to create the analytic data file and then StatTransfer was used to convert the file to SPSS format.

Appendix B: Walden University Dissertation Examples Related to My Study Topic

Cappello, M. A. (2012). *Radon contaminated drinking water from private wells: an environmental health risk assessment* (Doctoral dissertation). Accessed from Dissertations & Theses @ Walden University. (Order No. 3503427).

Childs, D.B. (2016). *Comparison of thyroid disease mortality between urban and rural populations in Southwest Georgia* (Doctoral dissertation). Accessed from Dissertations & Theses @ Walden University. (Order No. 10025736).

Hawk, N. V. (2012). *Risk assessment from agent orange exposure in Vietnam* (Doctoral dissertation). Accessed from Dissertations & Theses @ Walden University. (Order No. 3542468).

Koller, K. R. (2013). *Association between diabetes incidence and metabolic syndrome in western Alaska native people* (Doctoral dissertation). Accessed from Dissertations & Theses @ Walden University. (Order No. 3554414).

Lasker, G. A. (2012). *The association of organochlorine pesticide concentration with migraine headaches, body mass index, gender, and age* (Doctoral dissertation). Accessed from Dissertations & Theses @ Walden University. (Order No. 3521710).

Appendix C: Other University Dissertations Related to My Study Topic

Cook, M. M. (2015). *Endocrine-disrupting compounds: measurement in Tampa Bay, removal from sewage and development of an estrogen receptor model* (Doctoral dissertation, University of South Florida). Accessed from ProQuest Dissertations & Theses Global. (Order No. 3688382).

Hofe, C. R. (2012). *Associations between serum concentrations of polychlorinated biphenyls, serum carotenoids, and the probability of metabolic syndrome in the National Health and Nutrition Examination Survey 2003-2004* (Doctoral dissertation, University of Kentucky). Accessed from ProQuest Dissertations & Theses Global. (Order No.3538063).

Jensen, C. D. (2012). *The three w's of hazardous waste: who, why, and where?* (Doctoral dissertation, West Virginia University). Accessed from ProQuest Dissertations & Theses Global. (Order No. 3530428).

Mutter, E. A. (2014). *Assessment of contaminant concentrations and transport pathways in rural Alaska communities' solid waste and wastewater sites* (Doctoral dissertation, University of Alaska Fairbanks). Accessed from ProQuest Dissertations & Theses Global. (Order No. 3624450).

Tyrrell, J. B. (2013). *Linking environmental toxicant exposure to diabetes susceptibility* (Doctoral dissertation, Wayne State University). Accessed from ProQuest Dissertations & Theses Global. (Order No. 3594725).

Appendix D: Excerpts of NPL Superfund Site Hazardous Waste Site Narratives

AMOC Chemical Co., Oakland, CA

The treatment system operated from January 1997 through July 1998 and extracted approximately 7,000 pounds of VOCs, approximately 40 pounds of which were vinyl chloride. Operation of the system ceased in July 1998, due to community concern over the potential for a release of dioxins from the thermal oxidation unit. On December 5 and 14, 1996, during construction of the treatment system collection trench, the EPA On-Scene Coordinator observed shimmering vapors emanating from the open trench. SUMMA (tm) canister sampling indicated the presence of vinyl chloride; methylene chloride; 1,1,1-trichloroethane (TCA); and trichloroethene (TCE) in the immediate area of the trench. In addition, one SUMMA canister sample collected from in front of a residence adjacent to the site contained TCE.

Unimatic Manufacturing Corp., Fairfield, NJ

Potential Impacts on Surrounding Community/Environment

The former Unimatic building is severely contaminated with PCBs, which have entered indoor air. The Unimatic operation resulted in contaminated soil, ground water and surface water due to decades of wastewater discharge through leaky pipes. The nearest drinking water wells are located less than one-half mile downgradient of the site, and ground water wells within 4 miles of the site provide drinking water to more than 20,000 people. The well water used at the site until 1989 contained high levels of VOCs, and the effluent to the unnamed tributary to Deepavaal Brook was shown to contain petroleum hydrocarbons, oil and high levels of VOCs in the 1980s. Deepavaal Brook flows into the Passaic River and there is a drinking water intake that serves more than 450,000 people located 2.2 miles downstream of Deepavaal Brook.

Boarhead Farms, Bridgeton Township, PA

In 1984, EPA detected elevated levels of 1,1,1-trichloroethane (1,1,1-TCEA), trichloroethylene, and zinc in wells on the site. EPA also detected 1,1,1-TCEA and zinc in nearby residential wells. The 6,000 people living in the area obtain drinking water from public and private wells within 3 miles of the site.

Eighteenmile Creek, Lockport, NY

Sampling events indicate that Eighteenmile Creek sediments are contaminated with a variety of pollutants, including mercury, lead, copper, pesticides/insecticides; PCBs, dioxins, and furans. PCBs are the primary contaminants in sediment samples collected from Eighteenmile Creek. Although the highest PCB concentrations have been reported within the Corridor site, sampling data indicate that contamination extends about 13 miles downstream to Burt Dam.