ASSOCIATIONS BETWEEN DRINKING WATER SOURCE WATERSHED AND ADVERSE BIRTH OUTCOMES IN CENTRAL APPALACHIA

Cameron S. Cornwell

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

In

Biological Systems Engineering

Leigh-Anne H. Krometis, Chair

Jonathan A. Czuba

Julia M. Gohlke

Korine N. Kolivras

May 9, 2022

Blacksburg, VA

Keywords: adverse birth outcomes, Appalachia, land cover, watersheds

ASSOCIATIONS BETWEEN DRINKING WATER SOURCE WATERSHED AND ADVERSE BIRTH OUTCOMES IN CENTRAL APPALACHIA

Cameron Cornwell

ABSTRACT

In order to ensure clean drinking water for all, it is crucial to understand potential upland stressors that compromise the quality of source waters treated by local community water systems (CWSs). Contamination associated with specific types of land cover can result in downstream water quality degradation, which may reduce the effectiveness of treatment by CWSs. Surface mining has been hypothesized as a source of drinking water degradation within the Central Appalachian region, which may result in adverse exposures and health disparities. The purpose of this study was to identify potential correlations between land cover and adverse birth outcomes (ABOs) through the application of watershed epidemiology, an emerging environmental health paradigm. Birth records for the Central Appalachian region were acquired from their respective state health departments from 2001 to 2015: each record contained the mother's street address, outcome variables, and covariates. Records were included in later analyses if they fell within an approximated CWS service area. Contributing land cover to each CWS was determined via previously delineated watersheds that relied on CWS intake points. A binomial generalized linear model was used to compare low birth weight (LBW), term low birth rate (tLBW), and preterm birth (PTB) incidence to CWS source watershed land cover, Safe Drinking Water Act (SDWA) violations, CWS size, and covariates related to the birth records. Source watershed mining and SDWA health based (HB) violations were significantly associated with greater risks for preterm birth (PTB) and low birth weight (LBW). Future work should be conducted to explore upstream flow impacts, address missing data in the birth records, and to more accurately represent CWS service areas to better characterize exposure.

ASSOCIATIONS BETWEEN DRINKING WATER SOURCE WATERSHED AND ADVERSE BIRTH OUTCOMES IN CENTRAL APPALACHIA

Cameron Cornwell

GENERAL AUDIENCE ABSTRACT

Millions of individuals throughout the world are sickened by waterborne exposures every year. To ensure clean drinking water long-term, it is crucial to understand how human land cover might change the water quality of source watersheds, as this may impact the effectiveness of water treatment and increase adverse human health exposures. The goal of this effort is to understand whether land cover is linked to downstream adverse birth outcomes (ABOs) in Central Appalachia, a region of the United States previously associated with high disease incidence suspected to be partially linked to environmental exposure. Birth records were acquired for the years of 2001 to 2015 from four (VA, WV, TN, KY) respective state health departments. Each record contained the mother's address, outcome variables, and covariates (e.g., race, ethnicity). Births were located within approximate service areas for 140 surface water dependent community water systems (CWS) within the region. Data from each CWS, including weighted land cover proportions for their source watershed, were merged with the birth records according to approximate service areas. Statistical analysis suggested that higher source watershed levels of mining and urban development were associated with higher risks of preterm birth (PTB) and low birth weight (LBW). The number of health based (HB) violations associated with each CWS was also associated with both of these outcomes. Major limitations of this work include birth record data gaps and the lack of publicly available CWS service areas and/or water consumption rates, which does increase the risk of exposure misclassification.

ACKNOWLEDGMENTS

This study would not be possible without the guidance and support of numerous individuals. I owe a great deal of appreciation to Dr. Leigh-Anne Krometis, my advisor, as she sought me out in one of her undergraduate courses and helped me every step of the way. Her undivided attention, insight, feedback, and encouragement were crucial in the completion of this project. Furthermore, I also owe much gratitude towards my other committee members. Dr. Jon Czuba's expertise within hydrology and related ArcGIS Pro applications was a driving force in much of the methodology and data preparation. Dr. Julia Gohlke provided valuable public health insight allowing for much of the statistical modeling setup and interpretation. Expertise within geography and land cover applications from Dr. Korine Kolivras was also found to be extremely valuable towards the completion of this work.

I am also very grateful to the Krometis research group. Hannah Patton, Charles Sterling, Kathleen Hohweiler, and Sara Price provided unconditional support, feedback, and assistance on this project. I also value the friendships that were made over the last couple of years and look forward to remaining in touch. Additionally, I owe much appreciation to Ethan Smith who provided feedback, data, and guidance throughout this effort.

I owe much gratitude to the National Institute of Health (NIH) for this research as this project wouldn't be possible without their funding. I would also like to thank the Statistical Applications and Innovations Group (SAIG) here at the university for their assistance with the statistical modeling. Lastly, a great deal of appreciation goes out to Molly McKnight who got me started on this work.

Lastly, I would like to thank all my friends and family for the unconditional love and support throughout this work. Pushing to graduate one year after obtaining my bachelor's degree was challenging. However, having such a great support system has made all of this possible and I truly appreciate everyone's contributions regardless of the amount.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. **INTRODUCTION**

Longstanding health disparities within the Central Appalachian region cannot be entirely attributed to sociobehavioral factors, suggesting that environmental exposures (e.g., air and water) may play a significant role in disease (Krometis et al., 2017). Although water is frequently cited as an environmental health concern by Appalachian communities (Levêque & Burns, 2017; McSpirit & Reid, 2011), the role of waterborne exposures is poorly quantified for this region. Though past studies have suggested that community water system (CWS) violations may be linked to adverse health outcomes, these have been criticized for ecological bias (Hendryx & Ahern, 2012). A more recent examination incorporating watersheds found unclear evidence of watersheds as mediators of the association between proximity to surface mining and adverse birth outcomes, emphasizing that this route of exposure is quite complex (Ruktanonchai et al., 2022).

This research aims to address the uncertainty of watersheds as mediators by applying the emerging watershed epidemiology paradigm (Corley et al. 2018) to evaluate the associations between upstream land cover and adverse birth outcomes (ABOs). A child is described to have an ABO if they are less than 2,500 g in weight at birth (low birth weight: LBW), and or a gestational term less than 37 weeks (preterm birth: PTB), or if they weigh less than 2,500 g after a gestational term 37 week or more (term low birth weight: tLBW). Although watershed management is largely absent from the public health literature (Jenkins et al., 2018), past research suggests that land cover can impact downstream rates of Safe Drinking Water Act (SDWA) violations (Smith, 2020). Associations between ABOs and surface water management would provide justification for increased environmental protection and/or infrastructure investment to reduce regional burdens of disease.

2. LITERATURE REVIEW

2.1. Introduction: The Role of the Environment

Two decades ago, mapping and understanding the human genome appeared to offer a primary solution to address human disease; however, increasing research suggests that many diseases are strongly linked to key environmental exposures (Brunekreef, 2013). Reviews of epidemiological studies emphasize that 70 to 90% of chronic diseases originate from non-genetic factors (Pleil, 2012; Rappaport & Smith, 2010) and suggest the utility of genetic typing in addressing disease is often compromised by poor concomitant measures of environmental exposure (Wild, 2005). Given the current assessment by many human geneticists that “genetics loads the gun, but the environment pulls the trigger” (Pleil, 2012), defining and quantifying environmental exposures is critical to improving public health.

Delineation of exposure pathways is essential to the identification of environmental contaminants of concern. The 2004 International Programme on Chemical Safety (IPCS) Risk Assessment Terminology report defines exposure as, “contact between an agent and a target,” where such contact is made on a given exposure surface (oral, respiratory, or dermal) over a defined period (IPCS, 2004). While this definition is easily applied to acute occupational exposures, where and when critical contact occurs with chronic environmental contaminants can be quite difficult to identify. Likewise, the relationship between chronic environmental exposure and disease remains understudied (Wild, 2005). Individual level environmental exposure assessments are often logistically difficult and costly (Brunekreef, 2013); therefore, environmental epidemiologists sometimes rely on combining landscape data with health outcomes to approximate exposure (e.g., Buttling et al., 2021). However, while occupational exposure assessments have predefined boundaries and exposure definitions, environmental exposure assessments can struggle to account for spatial heterogeneity as geospatial boundaries delineating levels of exposure are not generally obviously defined (Corley et al., 2018).

The analysis of 10 million single nucleotide polymorphisms within the human genome will be of limited use without an understanding of the environmental factors that drive genetic expression and the onset and progression of disease (Wild, 2005). In response, an increasing number of researchers advocate adoption of the human exposome framework: the exposome encompasses all environmental exposures and environmental influences with their associated biological responses from the prenatal period onward (Brunekreef, 2013; Garner et al., 2016; Wild, 2005). The exposome comprises everything outside of the individual's genome including the metabolome, microbiome, adductome, and the proteome (Pleil, 2012). This framework specifically considers not only adverse exposures, but also an individual's mental, physical, and social factors to consider potential predictors of disease holistically (Brunekreef, 2013). Of critical note when applying this framework, the exposome is highly variable and constantly changes throughout an individual's lifetime, which is in direct contrast to the individual's genome (Wild, 2005).

2.2. Environmental Health in Central Appalachia

Populations within the Central Appalachian region, which includes portions of the states of Virginia, Kentucky, West Virginia, and Tennessee, continue to struggle to address significant and seemingly persistent health disparities (Behringer & Friedell, 2006; Borak et al., 2012; Gohlke, 2021; Krometis et al., 2017; Mcgarvey et al., 2011; Meit et al., 2019; Woolley et al., 2015). As previous analyses suggest that these disparities are not completely explained by genetic or behavioral (e.g., smoking, diet) factors, these disparities must be examined in connection to the unique land cover and potential environmental exposures of the region in order to improve community health. Appalachia as a region is often popularly associated with mining to produce thermal and metallurgical coal. In recent years, a large uptick in production between 1985 and 2015 resulted in roughly 1,100 square miles of newly mined land (Pericak et al., 2018). This increase in mining is partially attributable to amendments made to the Clean Air Act that resulted in higher demand for low-sulfur, high efficiency coal (Boyles et al., 2017). Mountaintop removal (MTR) is currently the primary method of coal mining in Central Appalachia as it is fast, cheap, and less labor intensive (Boyles et al., 2017).

Although the vast majority of the Central Appalachia footprint is forested or agricultural (Smith, 2020), the impact of MTR on water resources can be significant. The process of MTR requires the withdrawal of up to 1,000 ft of rock and soil via explosives at the surface, and this "overburden" is subsequently deposited in nearby valleys and waterways, which are often headwater streams (Ahern et al., 2011; Boyles et al., 2017; Holzman 2011). Explosion effluent, overburden, and runoff from surface mines can release sediment, metals, polycyclic aromatic hydrocarbons (PAHs), and salts into downstream surface waters and groundwater (Ahern et al., 2011; Boyles et al., 2017; Holzman 2011). Impacts on natural hydrology are significant: between 1985 and 2001, 724 miles of Appalachian streams were permanently destroyed along with four million acres impacted by the surface mining technique of MTR (Ahern et al., 2011). Coal processing, often on-site, also contaminates billions of gallons of water which are then stored in impoundment ponds or injected underground (Ahern et al., 2011; Holzman 2011). It is also important to note that standard surface mining (non-MTR), as well as underground mining, often occur alongside MTR mines, and that these sites are associated with the leaching of metals and salts via runoff and weathering of new, top-layer rock, as well as downstream sedimentation (Cook et al., 2015; Hopkins et al., 2013; Palmer et al., 2010; Sarver and Cox, 2013). There is increasing recent interest in potential links between MTR and non-MTR and adverse health outcomes (Holzman, 2011), which have long been recognized within coal-mining areas (Guidotti 1979).

Given the obvious and well-documented environmental consequences associated with coal mining, particularly MTR (Palmer et al. 2010), several environmental epidemiology studies have examined potential links between health disparities and mining land cover. Both Ahern et al. (2011) and Woolley et al. (2015) reported higher total mortality rates in coal-producing counties as compared to non-producing counties. Woolley et al. (2015) also observed higher rates of respiratory system mortality specifically in coal-mining counties in both WV and VA. Gohlke (2021) evaluated life expectancy, mortality rates, and low birth weight and preterm birth rates in four high-producing coalfield counties (Pike, KY; Boone, WV; Mingo, WV; and Logan, WV) and compared these observations relative to US national rates to other nations around the world. Low birth weight and preterm birth rates were elevated in WV, TN, and KY as compared to VA and the broader U.S. Of note, between 2014-2017, rates of low birth weight and preterm birth were higher in these Appalachian counties (12-14%) than the global average (11%), China

(7%), and even Ethiopia (12%) (Gohlke, 2021). In agreement with these findings, Ahern et al. (2011) reported, in comparison with non-mining counties, a 16% increase in risk of low birth weight infants in high coal production counties and a 14% increase in low birth weight in lower producing coal counties. Overall, although infant mortality and age-adjusted mortality rates are decreasing regionally, the rate of decline is slower when compared to the rest of the U.S. (Gohlke, 2021). It is important to note that many covariates including smoking, maternal education, and other socioeconomic/personal risk factors are often associated with some uncertainty with regards to accurate reporting and therefore could play a larger role than expected (Ahern et al., 2011; Boyles et al., 2017; Gohlke, 2021; Woolley et al., 2015). However, in a recent study Buttling et al. (2021) reported a significant positive association between preterm birth, low birth weight, and term low birth weight and proximity to active surface mining during gestation in Central Appalachia, even when accounting for the impacts of maternal age, parity, educational level, payment method, race, Hispanic origin, tobacco use, and the child’s sex (Buttling et al., 2021).

2.3. Defining Waterborne Exposures

A recent systematic review by Boyles et al. (2017) reported that of 11 studies on mortality in Central Appalachia, the majority reported higher rates associated with coal mining, but cautioned that conclusions were difficult to draw due to potential sources of bias in assigning exposure, which often relies on political boundaries (e.g., county of residence, census blocks). Since this meta-analysis was concluded, more recent examinations of potential connections between mining and birth outcomes have relied on natural boundaries of exposure, including airsheds (Buttling et al., 2021; Ruktanonchai et al., 2022).

Water is frequently cited as a potential vector for pollutants of human health concern from upland sources, including minelands (Hendryx 2015; Hendryx & Ahern 2012), and drinking water quality is often specifically mentioned as an issue of concern in Appalachian communities (Krometis et al. 2017). However, waterborne exposures are poorly defined by political boundaries (e.g., Hendryx & Ahern, 2012), which rarely follow hydrologic pathways. Previous works comparing census tracts to watersheds note that there is generally little to no congruence (Corley et al., 2018). Heterogeneity of waterborne exposure is therefore expected to occur across census tracts due to the lumping of differing geological characteristics within the same political boundary (Kolok et al., 2009). Differential boundaries between watersheds and political boundaries are likely to generate over- or underestimates of exposure, resulting in classification error/exposure bias.

Watersheds are recognized as extremely useful in linking natural (e.g., topography) and anthropogenic (e.g., land cover) characteristics with changes in downstream water quality and quantity (Omernik et al., 2017). Receiving water quality and quantity within a watershed depends heavily on upland geology, topography, human uses, and land cover (Jordan and Benson, 2015). Many anthropogenic land covers including mining, oil and gas extraction, urban development, and agriculture are associated with adverse impacts on water quality and quantity, and these impacts are likely to worsen given projected increases in population and climate change (Jordan and Benson, 2015). Given both the wide range of downstream human uses, as well as the sensitivity of receiving waters to anthropogenic impacts, the watershed provides a potentially useful boundary for transdisciplinary research efforts examining changes in the health of humans, non-human species, and ecosystems (Jenkins et al., 2018).

However, although widely used in evaluations of ecological health and non-human exposure to adverse contaminants, watershed management remains almost absent from public health literature; between 2000 and 2010, only 3.5% of the academic journals on watershed management mention human health (Jenkins et al., 2018). Combining public and ecological health frameworks offers an opportunity to better understand relationships between exposure (human and ecological) and contaminants (chemical or microbial) to simultaneously improve environmental and health management (Angermeier et al., 2021; Jordan & Benson, 2015).

One strategy that aims to connect watershed management and public health is the emerging field of watershed epidemiology, which delineates potential waterborne environmental exposures via hydrologic boundaries (Jenkins et al., 2018). Scaling exposure geographically within watersheds simultaneously encompasses relationships between the environment, water resources, and human health (Jordan & Benson, 2015) and can be used to explore relationships between watershed geography and contaminant distribution (Kolok et al., 2009). Contaminants flow in a defined manner within a watershed, allowing exposure assessment to be more accurately represented than through the use of political boundaries such as counties (Corley et al., 2018). Furthermore, a common level of exposure can be evaluated as any point on a stream reflects the collection of the characteristics upgradient from that stream point (Omernik et al., 2017). This approach offers a substantial reduction in potential exposure bias when compared to political boundaries, i.e., individuals residing in the same watershed miles apart are more likely to share similar waterborne contaminant exposures as compared to individuals living in close proximity yet in separate watersheds (Corley et al., 2018; Jenkins et al., 2018; Kolok et al., 2009). Corley et al. (2018) demonstrated this new technique through the use of hydrologic boundaries to evaluate adverse health outcomes in Nebraska. Incidence rate per watershed was calculated and evaluated to determine that the three health metrics of interest (birth defects, pediatric cancers, and thyroid cancer) all had independent patterns across all watersheds, highlighting the significance of using these boundaries rather than census tracts (Corley et al., 2018). However, it is important to note that like census boundaries, watershed are still evaluating exposure at an approximated level.

Ruktanonchai et al. (2022) recently explicitly examined watershed and airshed exposure pathways to compare proximity to surface mining and adverse birth outcomes in Central Appalachia through a mediation analysis. Although there was a strong relation between proximity to surface mining and adverse birth outcomes, the watershed exposure pathways were not significantly and uniquely associated with the adverse birth outcomes. It is important to recognize that in this analysis, HUC10s were used as a primary watershed unit, i.e., all births within a HUC10 boundary were assigned the same “exposure”. HUCs (Hydrologic Unit Codes) are a “nationwide set of geographical polygons based on drainage subdivisions,” which ultimately are sorted in hierarchical levels (Omernik et al., 2017). Although HUCs have previously been incorporated in human health behavioral research and ecological modeling (Corley et al., 2018), these watersheds are nonspecific to drinking water source watersheds and do not always include the full upstream watershed area (Omernik et al., 2017). Ruktanonchai et al. (2022) therefore posited that the complexities of source water use, treatment, and distribution may render standard uniform size HUC10s less useful in defining exposure.

3. PROJECT MOTIVATION AND OBJECTIVES

Central Appalachia is home to previously identified recalcitrant health disparities that cannot be entirely explained by sociobehavioral factors or issues of health access (Gohlke, 2021; Krometis et al., 2017). Although waterborne exposures are frequently cited as a significant concern of residents, previous examinations linking land cover and health outcomes largely have relied on political boundaries, which can increase the potential for ecological exposure bias. However, water as a medium of exposure is worth investigating as a potential source of concern; as described by (Garner et al., 2016), water and its constituents are a fundamental component of the exposome and play a significant role in human health, survival, hygiene, and indirect and direct exposures via bioaerosols.

Environmental exposures potentially linked or suspected to be linked to adverse birth outcomes include lead, pharmaceuticals, nutrients, herbicides, pesticides, polycyclic aromatic hydrocarbons (PAHs), particulate matter (PM), arsenic, mercury, cadmium, selenium, nickel, copper, sulfur dioxide, nitrous dioxide, carbon monoxide, nitrates, tobacco smoke, and pesticides (Ahern et al., 2011; Corley et al., 2018; Gohlke, 2021). Previous research suggests that mining activity may be associated with ABOs in the Central Appalachian region (Ahern et al. 2011; Buttling et al. 2021), although associations between ABOs and potential waterborne exposure remain unclear (Ruktanonchai et al, 2022).

Therefore, the goal of this effort is to define exposure via hydrologic boundaries to examine the potential correlation between a set of adverse birth outcomes (preterm birth, low birth weight, and term low birth weight) and land cover in Central Appalachia. This effort aims to answer the following question: Does upstream land cover influence rates of adverse birth outcomes in Central Appalachia?

As roughly 55% of the Central Appalachian region is within the southern coalfields, with an approximate cumulative historic surface mining footprint of 2,300 square miles in 2015 (Pericak et al., 2018), this effort will add to previous examinations of the impact of coal mining on human health. It will also consider agricultural land cover, which is second only to forest cover in terms of regional footprint (Smith, 2020). This is potentially of interest beyond the target region as water quality degradation of surface and groundwater attributed to agriculture was the leading cause in the 2000 National Water Quality Inventory (Environmental Protection Agency, 2005) and previous examinations of predominantly agricultural communities have noted potential links between related contaminants and adverse health outcomes (Kolok et al., 2009). This work will improve on earlier work that used HUC10s to define potential waterborne exposures (Ruktanonchai et al, 2022) and instead link exposures to specific water treatment plant source watersheds previously defined by Smith (2020).

4. METHODS

4.1. Defining Watershed Boundaries

The project focuses on the Central Appalachian region of the United States, which includes counties within the four states of KY, TN, VA, and WV (ARC, 2021). Central Appalachia was selected as a target given significant local health disparities (Krometis et al. 2017; Gohlke 2021), current and historical resource extraction (Pericak et al. 2018; Marston and Kolivras, 2020), and past research group experience/availability of health data (Buttling et al., 2021; Ruktanonchai et al., 2022; Smith, 2020).

Drinking water source watersheds were delineated using previous work by Smith (2020). In brief, Smith (2020) delineated watersheds for the 140 surface water reliant, active CWSs in the Central Appalachian region. Watershed boundaries were delineated from a point between 100 to 1,200 m downstream from an intake pump, treatment plant, or center of the community served (>90% within 300 m of treatment plant) following visual inspection of Google Maps. As 13 of the 140 water systems included multiple intake points, 153 source watersheds were delineated; land cover "exposure" from CWSs served by multiple intakes were considered the average of both upstream source watersheds. The smallest watershed was at roughly 0.2 square miles and the largest at approximately 23,700 square miles. It is important to note that a majority of the watersheds overlap one another, as some CWSs were served by the same primary source water (e.g., the Cumberland or Ohio River).

Community water system (CWS) data were obtained from the USEPA's Safe Drinking Water Information System (SDWIS; https://www.epa.gov/enviro/sdwis-search) for KY, TN, VA, and WV from 2001 to 2015. Data from the SDWIS for each system include: system identifier, population served, county served, primary source water type, ownership (private vs. public), address, and phone number. An additional field was added to represent the system's size as classified by the USEPA, i.e., a CWS serving less than 501 people was identified as very small, 501-3,300 as small, 3,301-10,000 as medium, 10,001-100,000 as large, and anything greater than 100,000 as very large (USEPA, 2011).

All project mapping and analyses were conducted in ArcGIS Pro (Version 2.7). County level data were acquired for the 82 Central Appalachian counties from the United States Census Bureau in TIGER/Line ASCII format for the year of 2019. Both the county and watershed data were projected to the North American Datum (NAD) of 1983 and Universal Transverse Mercator (UTM) coordinate system of zone 17 in the northern hemisphere (Figure 4-1), and these settings were used for the remainder of the research.

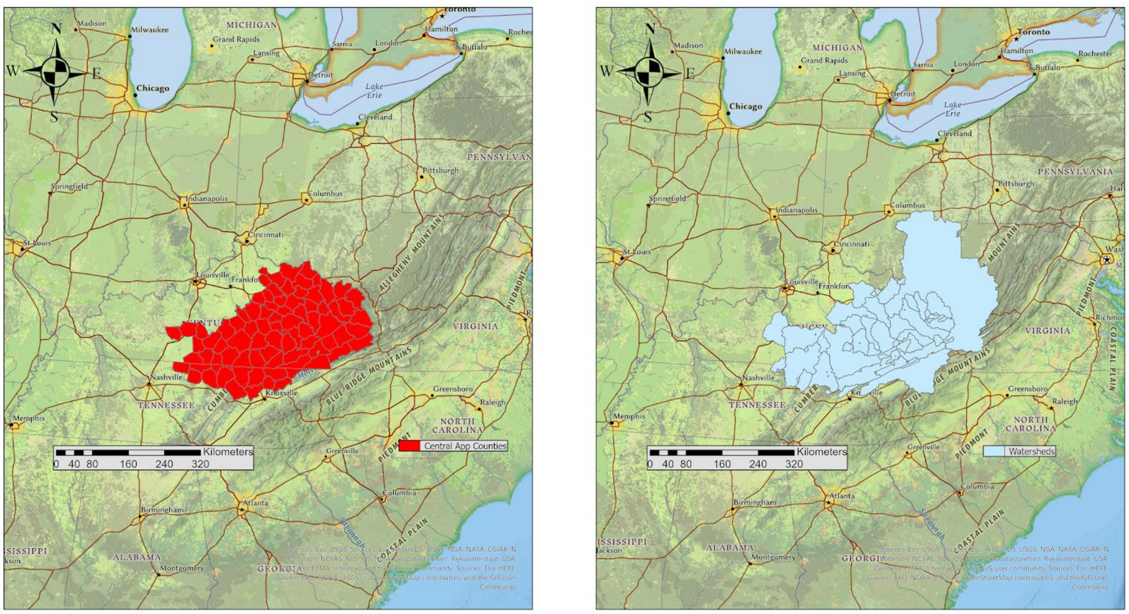

Figure 4-1. Comparison of Central Appalachian A) county boundaries and B) approximated drinking water exposure watersheds

4.2. Birth Record Overview

This work relies on the same dataset of 194,084 birth records used by Buttling et al. (2021) and McKnight et al. (in revision). In brief, the dataset represents a subset of all the birth records in Central Appalachia from 1989 to 2015, excluding those without full addresses, such as those that include only Post Office box addresses (Buttling et al., 2021). The street-level subset contains the mother's address, outcome variables, and covariates, all of which were provided by a respective state health department. Outcome variables include the three adverse birth outcomes (ABOs) of interest: low birth weight (LBW; child born weighing less than 2500 g), preterm birth (PTB; child born at less than 37 weeks gestational age), and term low birth weight (tLBW; child born after 37 weeks gestational age and weighing less than 2500 g). Covariates included mother's age, parity, mother's race, ethnicity (Hispanic vs non-Hispanic), and the child's sex. Due to higher rates of LBW and PTB in plural births regardless of exposure, this dataset represents only singleton births (Buttling et al., 2021; McKnight et al., in revision).

4.3. Defining CWS Service Areas

As noted by Ruktanonchai et al. (2022), drinking water exposure is complex. As some source water watersheds delineated by Smith (2020) were as large as 23,700 mi^2, it is unlikely that the entire population within these watersheds were served by municipal water from the single CWS serving as the outlet point. Unfortunately, CWS service areas (i.e. distribution system extents) are generally not publicly available (Chini & Stillwell, 2017; Marcillo et al., 2021). To approximate the population served by a given CWS, buffers were created around CWS based on system size; larger systems were assumed to support longer trunk lines and larger service areas (Table 4-1 and Figure 4-2). Buffer shapefile layers were used to clip the birth records to assign them to an approximate CWS approximate service area (and associated exposure). To account for births falling within multiple service areas, all birth records were snapped to the nearest CWS. Once snapped, respective CWS data (e.g., land cover, system size, system violations) were spatially joined with each birth, creating a singular data table.

Table 4-1. CWS count and buffer distance by system size

CWS Size	Count	Buffer Distance (km)
Very Small	3	1
Small	42	2
Medium	56	5
Large	52	10

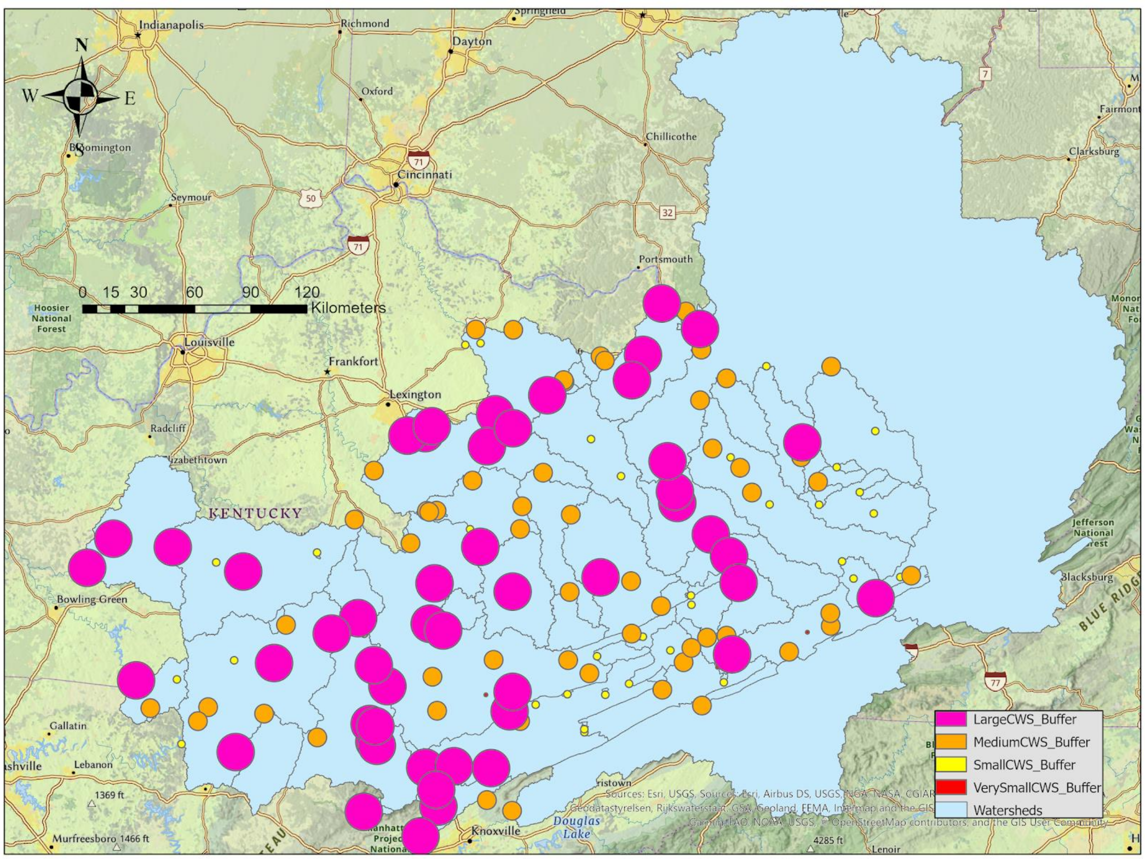

Figure 4-2. Approximate CWS service areas

A total of 109,505 birth records (~61.2%) between 1990-2010 analyzed by Buttling et al 2021 and McKnight et al (in revision) fell within all the approximated CWS service areas illustrated in Figure 4-2. However, CWS violation data are only available since 2001, and 10% of the records did not include maternal Hispanic origin. Birth records with a gestational year prior to 2001 or missing Hispanic origin were therefore removed (Figure 4-3), yielding a final subset of 62,656 birth records for analysis.

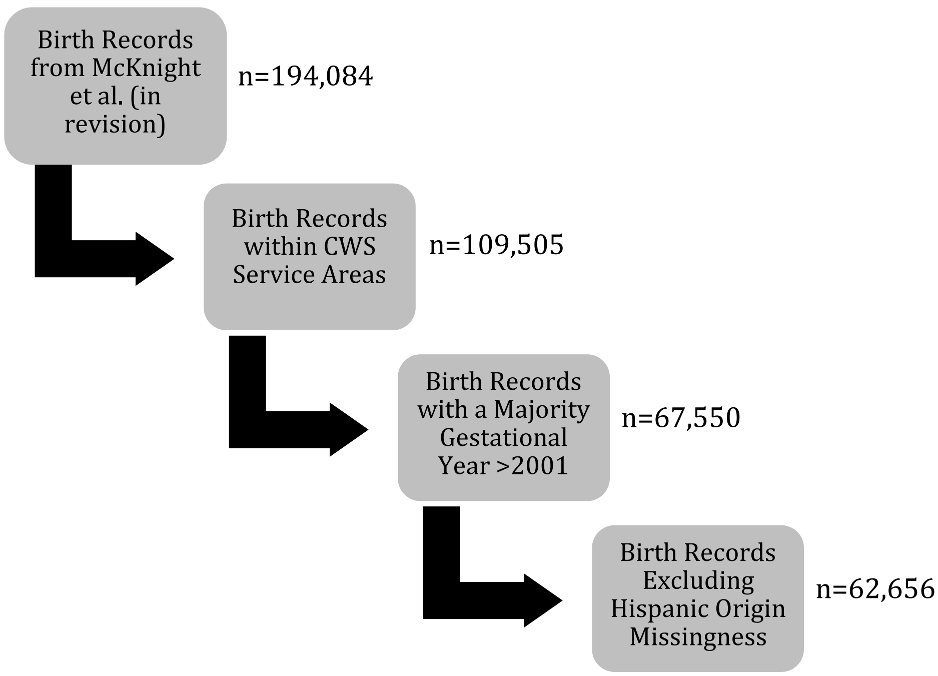

Figure 4-3. Remaining birth records for statistical modeling

4.4. Land Cover Classification

Land cover classification relied on simplified categories developed by Smith (2020) for the years 2001, 2004, 2006, 2008, 2011, 2013, and 2016 at a 30 m resolution. In brief, Smith (2020) coupled data from the national land cover database (NLCD) with surface mining data from Marston and Kolivras (2020), which defined mining inputs by evaluating the normalized difference vegetation index (NDVI) for all pixels in the study region from 1985 through 2015.

The original twenty NLCD land cover classifications within the study region were reduced to an initial six broad categories of interest (Table 4-2). To account for surface mining, any pixel that overlapped with surface mining pixels defined by Marston and Kolivras (2020) were reclassified as surface mining (i.e., a seventh category) for each respective year (Table 4-2).

Table 4-2. Land Cover Reclassification

Reclassified Class	**NLCD Original Class**
Water	Open Water Perennial Ice/Snow Woody Wetlands Emergent Herbaceous Wetlands
Low Developed	Developed, Open Space Developed, Low Intensity
High Developed	Developed, Medium Intensity Developed, High Intensity
Barren	Barren Land (Rock/Sand/Clay)
Vegetation	Deciduous Forest Evergreen Forest Mixed Forest Dwarf Scrub Shrub/Shrub Grassland/Herbaceous Sedge/Herbaceous Lichens Moss
Agriculture	Pasture/Hay Cultivated Crops
Surface Mining	N/A

It is important to note that Smith (2020) combined the 2015 mining extents with the 2016 NLCD land cover data due to the temporal constraints on the mining data. Also, any area defined as mined but less than 9000 m^2 was removed from the surface mining class as this was considered the minimum threshold for an active surface mine. The final reclassified land cover rasters were clipped to the extents of the delineated watersheds. Land cover classification percentages were averaged between 2001 and 2016 incorporating the final seven, reclassified land cover rasters (Table 4-3).

Table 4-3. Average watershed land cover percentages from 2001 to 2016 for Central Appalachian target region

Land Cover	**Mean (%)**	**Min (%)**	**Max (%)**
Water	1.16	1.14	1.17
Low Developed	5.88	5.82	5.95
High Developed	0.75	0.69	0.80
Barren	0.49	0.40	0.55
Vegetation	75.84	75.68	75.96
Agriculture	15.64	15.44	15.96
Surface Mining	0.25	0.16	0.36

Observing Table 4-3, vegetation (~76%) and agriculture (~16%) had the highest average land cover percentages with surface mining having the lowest (~0.25%). It is important to note that there is little variance between the mean, minimum, and maximum for each of the seven reclassified land cover types, suggesting consistency over the study duration at the source watershed scale. Rather than including average land cover percentages per watershed in statistical modeling efforts, distance weighted land cover proportions were incorporated as land closer to a CWS intake point was assumed to have a greater impact on water quality. Proportions were estimated for all 153 previously delineated watersheds using the following equation from Smith (2020):

$$P_l = \sum_l (1 - \frac{d_l}{d_m}) \, / \, \sum_i \left(\frac{d_i}{d_m}\right) * 100\% \tag{4-1}$$

where

P_l = a proportion of a land cover class within a watershed weighted by its proximity to the watershed outlet
d_l = the distance from a cell within the target land cover class to the outlet (m)
d_m = the distance to the outlet from the watershed cell furthest from the outlet (m)
d_i = a cell's distance to the outlet (m)

For CWSs containing multiple intake points, all the distance weighted land cover proportions were averaged for each respective CWS. Once complete, each birth record was assigned specific distance weighted land cover percentages based on their respective CWS.

4.5. Statistical Modeling Approach

Once each individual birth record had been assigned a CWS and its respective distance weighted land cover proportions for each of the seven land cover classes, the data subset was then exported from ArcGIS Pro to RStudio. Potentially predictive variables included maternal age, education level, race, Hispanic origin, tobacco consumption, child's sex, parity, CWS size, weighted land cover percentages, and CWS health based (HB) and monitoring and reporting (MR) violations. Variables were compared with one another to check respective correlation values which recommended removal of vegetation as a land cover variable as the correlation value between vegetation and agriculture was extremely high. Agriculture was retained given its anthropogenic origin and anticipated higher likelihood of negative downstream water quality impacts (Hoorman et al., 2008). Variables that were not an integer or a decimal value were evaluated as factor variables in order to address categorical and or yes/no data types. Following variable classification, a binomial generalized linear model was fit for each of the three outcome variables of interest (LBW, PTB, and tLBW) using the aforementioned predictive variables. A B-spline with four degrees of freedom was utilized to account for secular trends in outcomes over the time period of study within all model runs (found in RStudio spline package). Splines allow for a nonlinear function to account for overall rise then fall of preterm birth and low birth weight observed in the mid-2000s likely due to the rise then fall in caesarean deliveries (Tilstra & Masters, 2020). Complete RStudio code and initial outputs are available in Appendix A. For each model output, odds ratios were calculated for all predictor variables using the following equation:

$$OR = e^{\beta} \tag{4-2}$$

where

OR = Odds Ratio

β = beta value represented by the estimate column in each model output

5. RESULTS

5.1. Preterm Birth

Variables statistically significant for preterm birth (PTB) include maternal age, parity, the child's sex, tobacco use, community water system (CWS) size, land cover, health based (HB) violations, and year spline 2 (Table 5-1). With regards to maternal age, mothers less than 18 years of age were 15.7% ($p < 0.05$) less likely to undergo PTB as compared to mothers between the ages of 18 and 35. In contrast, mothers greater than 35 years of age were 40.1% ($p < 0.001$) more likely to have a PTB compared to mothers between the ages of 18 and 35. Parity, the numerical order of a birth, was found to be statistically significant for PTBs on the second and fourth or greater birth order compared to first order births. Second order births presented lesser odds of a PTB at 6.1% ($p < 0.05$) compared to first order births. However, fourth and larger order births were more likely of a PTB at 30% ($p < 0.001$) compared to first order births. Female infants were 12.1% ($p < 0.001$) less likely to be a PTB compared to infants born as males. Tobacco use was associated with a 20.9% ($p < 0.001$) greater odds of a PTB when compared to non-smoking. With regards to the CWS size, infants born in medium systems were less likely to be preterm compared to large systems at 8.8% ($p < 0.01$). A majority of land cover types displayed a statistical significance for every 1% increase. High intensity development and surface mining provided greater odds of a PTB by 10.5% ($p < 0.01$) and 23.8% ($p < 0.001$) respectively for every 1% increase in land cover. Low intensity development and barren land displayed lesser odds of a PTB at 3.1% ($p < 0.001$) and 8.5% ($p < 0.05$) respectively for every 1% increase in land cover. HB violations were associated with an increase in PTB by 3.7% ($p < 0.001$) for every 1% increase in the variable. Although birth within Tennessee (TN) was significant, it is important to note these births comprised a small total percent within the birth record subset and this finding was associated with a high standard error (SE) of 47.9%.

Table 5-1. Odds ratios of PTB from variables included in the final model

Preterm Birth				
Variable	Estimate	Odds Ratio	Std. Error	Pr (> | Z |)
(Intercept)	-2.453	0.086	0.171	< 2e-16 ***
Mother's Age (Reference: 18-35)				
<18 Years of Age	-0.171	0.843	0.074	0.020 *
>35 Years of Age	0.337	1.401	0.054	4.25e-10 ***
Parity (Reference: 1st Order)				
2nd Order	-0.063	0.939	0.031	0.043 *
3rd Order	0.019	1.019	0.039	0.623
4th Order or Greater	0.262	1.300	0.047	2.29e-08 ***
Mother's Education (Reference: 8th Grade or Less)				
9th-12th Grade	0.094	1.099	0.083	0.256
Post High School Education	-0.041	0.960	0.085	0.632
Mother's Race (Reference: White)				
Black	0.123	1.131	0.107	0.250
Other	-0.129	0.879	0.123	0.292
Mother Hispanic Origin (Reference: Yes)				
No	0.281	1.324	0.147	0.056 ·
Child Sex (Reference: Male)				
Female	-0.129	0.879	0.026	7.86e-07 ***
Tobacco Use (Reference: No)				
Yes	0.189	1.209	0.029	3.13e-11 ***
Size (Reference: Large)				
Medium	-0.092	0.912	0.034	0.007 **
Small	-0.088	0.916	0.090	0.328
High Intensity Development	0.100	1.105	0.031	0.002 **
Low Intensity Development	-0.032	0.969	0.009	4.67e-04 ***
Agriculture	-0.001	0.999	0.001	0.260
Mining	0.214	1.238	0.049	1.34e-05 ***
Barren	-0.089	0.915	0.038	0.020 *
MR Violations	-0.001	0.999	0.004	0.806
HB Violations	0.037	1.037	0.010	1.66e-04 ***
State (Reference: Kentucky)				
Tennessee	1.487	4.424	0.479	0.002 **
Virginia	-0.340	0.712	0.381	0.372
West Virginia	-0.282	0.755	0.179	0.115
bs(Gestational Year, df =4)1	-0.074	0.928	0.115	0.519
bs(Gestational Year, df =4)2	0.447	1.563	0.098	5.05e-06 ***
bs(Gestational Year, df =4)3	-0.067	0.936	0.088	0.452
bs(Gestational Year, df =4)4	0.022	1.022	0.066	0.739

Significant Codes: 0 **\|** ' *** ' 0.001 **\|** ' ** ' 0.01 **\|** ' * ' 0.05 **\|** ' · ' 0.1 **\|** ' ' 1

5.2. Low Birth Weight

Statistically significant variables with regards to low birth weight (LBW) include maternal age, parity, maternal education level, mother's race, child's sex, maternal tobacco use, land cover types, HB violations, and year splines 2, 3, and 4 (Table 5-2). As observed for PTB, as maternal age increased, the likelihood of a LBW child did as well. A significant increase in odds of a LBW infant was found in mothers greater than the age of 35 by 59% ($p < 0.001$) compared to mothers between the ages of 18-35. With regards to parity, second and third order births were less likely of being classified as LBW compared to first order births at 31.3% and 29.1% (both at $p < 0.001$) respectfully. Maternal education at any level post high school education decreased the odds of a LBW infant by 32.3% ($p < 0.001$) compared to mothers receiving an educational level of eighth grade or less. Black mothers were 102.7% ($p < 0.001$) more likely to have a LBW child when compared to white mothers as reference. Female births were 21.6% ($p < 0.001$) more likely to be classified as LBW, which is contrary to findings for PTB. Similar to PTB, tobacco use presented greater odds of LBW by 117.6% ($p < 0.001$) compared to mothers who didn't smoke. Again, corresponding to PTB, high intensity development and surface mining were associated with greater odds of a LBW infant: for every 1% increase in land cover, odds of LBW increased by 9.2% and 14.3% (both at $p < 0.05$), respectively. Low intensity development displayed lesser odds of a LBW infant at 2.2% ($p < 0.05$) for every 1% increase in land cover. HB violations provided a relatively similar increase in likelihood of a LBW child compared to PTB at 2.3% ($p < 0.05$). Again, Tennessee was significant as the home state of birth but was associated with a high standard error (52.6%).

Table 5-2. Odds ratios of LBW from variables included in the final model

Low Birth Weight				
Variable	Estimate	Odds Ratio	Std. Error	Pr (> | Z |)
(Intercept)	-2.875	0.056	0.201	< 2e-16 ***
Mother's Age (Reference: 18-35)				
<18 Years of Age	-0.110	0.896	0.076	0.150
>35 Years of Age	0.464	1.590	0.063	2.68e-13 ***
Parity (Reference: 1st Order)				
2nd Order	-0.375	0.687	0.036	< 2e-16 ***
3rd Order	-0.344	0.709	0.047	5.10e-14 ***
4th Order or Greater	-0.077	0.926	0.054	0.150
Mother's Education (Reference: 8th Grade or Less)				
9th-12th Grade	-0.097	0.907	0.087	0.262
Post High School Education	-0.390	0.677	0.090	1.37e-05 ***
Mother's Race (Reference: White)				
Black	0.706	2.027	0.101	3.04e-12 ***
Other	-0.127	0.881	0.149	0.394
Mother Hispanic Origin (Reference: Yes)				
No	0.226	1.254	0.176	0.199
Child Sex (Reference: Male)				
Female	0.196	1.216	0.030	6.21e-11 ***
Tobacco Use (Reference: No)				
Yes	0.777	2.176	0.032	< 2e-16 ***
Size (Reference: Large)				
Medium	-0.024	0.976	0.039	0.530
Small	0.103	1.108	0.095	0.280
High Intensity Development	0.088	1.092	0.035	0.012 *
Low Intensity Development	-0.022	0.978	0.010	0.031 *
Agriculture	-0.001	0.999	0.001	0.384
Mining	0.134	1.143	0.054	0.013 *
Barren	-0.048	0.953	0.041	0.240
MR Violations	-0.005	0.995	0.005	0.357
HB Violations	0.023	1.023	0.011	0.041 *
State (Reference: Kentucky)				
Tennessee	1.135	3.110	0.526	0.031 *
Virginia	-0.236	0.790	0.388	0.543
West Virginia	0.088	1.092	0.172	0.609
bs(Gestational Year, df =4)1	-0.018	0.982	0.134	0.890
bs(Gestational Year, df =4)2	0.266	1.304	0.113	0.019 *
bs(Gestational Year, df =4)3	0.271	1.311	0.101	0.008 **
bs(Gestational Year, df =4)4	0.252	1.129	0.077	9.82e-04 ***

Significant Codes: 0 | ' *** ' 0.001 | ' ** ' 0.01 | ' * ' 0.05 | ' · ' 0.1 | ' ' 1

5.3. Term Low Birth Weight

Variables that are significant with regards to a child being classified as term low birth weight (tLBW) include maternal age, parity, mother's educational level and race, child's sex, maternal tobacco use, CWS system size and year splines 3 and 4 (Table 5-3). No land cover variables were significant. Results related to maternal age were similar to those observed for PTB and LBW. Mothers over the age of 35 were 61.2% ($p < 0.001$) more likely to have a tLBW infant compared to mothers between the ages of 18 and 35. Compared to first order births, all other birth order categories displayed lesser odds of a tLBW classification with second order births at 33% ($p < 0.001$), third order birth at 29.2% ($p < 0.001$), and fourth or higher order births at 22.1% ($p < 0.01$). Mothers who participated in any level of post high school education, whether complete or not, were 37.2% ($p < 0.001$) less likely to have a tLBW infant compared to mothers who were at most in eighth grade. The likelihood of tLBW was 111% ($p < 0.001$) higher for black mothers as compared to white mothers. A female infant had a 69.9% ($p < 0.001$) greater odds of a tLBW classification compared to a baby classified as male. Tobacco consumption presented a strong odds of tLBW infant at 214.9% ($p < 0.001$) compared to mothers who didn't smoke. Mothers located within the approximated service area of a small CWS were 33.4% ($p < 0.05$) more likely to have a tLBW infant as compared to those served by large systems. Again, Tennessee was significant as the home state of birth but was associated with a high standard error (63.9%).

Table 5-3. Odds ratios of tLBW from variables included in the final model

Term Low Birth Weight				
Variable	Estimate	Odds Ratio	Std. Error	Pr (> | Z |)
(Intercept)	-3.987	0.019	0.311	< 2e-16 ***
Mother's Age (Reference: 18-35)				
<18 Years of Age	-0.098	0.906	0.121	0.417
>35 Years of Age	0.478	1.612	0.103	3.47e-06 ***
Parity (Reference: 1st Order)				
2nd Order	-0.401	0.670	0.058	6.41e-12 ***
3rd Order	-0.345	0.708	0.072	1.86e-06 ***
4th Order or Greater	-0.250	0.779	0.088	0.004 **
Mother's Education (Reference: 8th Grade or Less)				
9th-12th Grade	-0.193	0.824	0.131	0.141
Post High School Education	-0.465	0.628	0.137	6.99e-04 ***
Mother's Race (Reference: White)				
Black	0.747	2.110	0.155	1.46e-06 ***
Other	-0.171	0.843	0.249	0.493
Mother Hispanic Origin (Reference: Yes)				
No	-0.103	0.902	0.270	0.702
Child Sex (Reference: Male)				
Female	0.530	1.699	0.049	< 2e-16 ***
Tobacco Use (Reference: No)				
Yes	1.147	3.149	0.052	< 2e-16 ***
Size (Reference: Large)				
Medium	-0.018	0.982	0.063	0.776
Small	0.288	1.334	0.139	0.038 *
High Intensity Development	0.064	1.066	0.054	0.238
Low Intensity Development	-0.017	0.983	0.016	0.274
Agriculture	-0.001	0.999	0.002	0.600
Mining	-0.021	0.979	0.083	0.801
Barren	0.037	1.037	0.058	0.530
MR Violations	-0.004	0.996	0.008	0.623
HB Violations	-0.034	0.967	0.020	0.088 ·
State (Reference: Kentucky)				
Tennessee	1.472	4.356	0.639	0.021 *
Virginia	-0.317	0.729	0.614	0.606
West Virginia	0.223	1.249	0.253	0.379
bs(Gestational Year, df =4)1	0.211	1.234	0.217	0.332
bs(Gestational Year, df =4)2	0.221	1.248	0.183	0.226
bs(Gestational Year, df =4)3	0.378	1.460	0.164	0.021 *
bs(Gestational Year, df =4)4	0.524	1.690	0.123	2.11e-05 ***

Significant Codes: 0 | ‘ *** ’ 0.001 | ‘ ** ’ 0.01 | ‘ * ’ 0.05 | ‘ · ’ 0.1 | ‘ ’ 1

6. DISCUSSION

6.1. Land Cover

The presence of MTM has been consistently linked with downstream water quality degradation (Palmer et al., 2010), and several past studies have hypothesized that waterborne exposures as a result of this degradation might contribute to adverse health outcomes observed for the Central Appalachian region (Hendryx 2015; Hendryx & Ahern, 2012). In this model, increases in surface mining within a drinking water source watershed was significantly linked with greater chances of both preterm birth (PTB) and low birth weight (LBW), i.e., a 1% increase in surface mining within the drinking water service area watershed was associated with a 23.8% greater chance of PTB and a 14.3% greater chance for LBW when controlling for all covariates (Tables 5-1 and 5-2). These findings are in keeping with past studies that found evidence of links between increases in surface mining and increases in adverse birth outcomes (ABOs) (Ahern et al., 2011; Buttling et al., 2021; Gohlke, 2021; Ruktanonchai et al., 2022). For example, Ahern et al. (2011) demonstrated that both high and moderate-producing coal-mining counties of West Virginia (WV) were associated with greater chances of LBW at very similar levels (16% and 14% greater likelihood, respectively). In assigning exposure based on CWS contributing source watershed, rather than political boundaries, this work continues to expand on work by Buttling et al. (2021) and Ruktanonchai et al. (2022) that aims to reduce exposure bias. Buttling et al. (2021) assigned land cover exposure through proximity and adjusted for rates of ABOs in the same area prior to active mining, and determined increases of 6% for PTB and LBW for mothers per 1% increase in active surface mining within five kilometers while controlling for demographic characteristics, but did not explicitly examine waterborne pathways. Although Ruktanonchai et al. (2022) situated exposure within HUC10 watersheds, results were less clear, perhaps due to the complexity of drinking water exposure. It is worth noting that in the present study, term low birth weight (tLBW) was not significantly related to surface mining, contrasting the 2% increase found in Buttling at al. (2021).

This study uniquely also examined potential links between other land covers in Central Appalachia and ABOs. In addition to surface mining, urbanization was significantly associated with both PTB and LBW. For every 1% increase in high intensity development, the likelihood of PTB and LBW increased by 10.5% and 9.2% respectively (Tables 5-1 and 5-2). Conversely, for every 1% increase in low intensity development, PTB and LBW chances decrease by 3.1% and 2.2% respectively (Tables 5-1 and 5-2). It is not surprising that high intensity development is associated with increases in both PTB and LBW as urbanization is strongly linked to water quality degradation (Wenger et al., 2009). Conversely, the seeming "protective" impact of low impact development may be the result of increasing access to health care as compared to rural areas, though more research would be required to confirm this hypothesis. Barren land, an area with sparse vegetation due to infertile soil, displayed lesser odds of PTB at 8.5% for every 1% increase (Table 5-1). The relationship between barren land and PTB is still largely unclear which can be seen by a lack of previous literature/studies.

Contrasting previous literature and studies, increases in agricultural land cover were not significantly associated with PTB, LBW, or tLBW. For example, Corley et al. (2018) found a strong association between agriculture in Nebraska with birth defects. One potential explanation is that Nebraska has a higher agricultural land cover at 92% of the state (NDA, 2022) compared to 24% of the Central Appalachian region. Furthermore, the type of agriculture can also play a large role, as different pesticides, herbicides, and other chemicals are applied for different types of crops and or pasture.

6.2. Demographic Variables

Relationships between maternal age, smoking, and infant sex were largely in keeping with previous studies. For example, the present findings suggest that likelihood of PTB, LBW, and tLBW rates increase along with maternal age. In this analysis, maternal age was separated into three subcategories as previous studies have shown increased risk of ABOs at younger maternal ages (Buttling et al., 2021; Fuchs et al., 2018; McKnight et al., in revision). In this work, a maternal age greater than 35 was associated with higher odds of PTB (40.1%), LBW (59%), and tLBW (61.2%) when compared to maternal age between the ages of 18 and 35 (Tables 5-1, 5-2, and 5-3). A maternal age less than 18 was associated with reduced risk of PTB (15.7%), LBW (10.4%) and tLBW (9.4%) when compared to mothers between the ages of 18 and 35 (Table 5-1). These findings are consistent with the results in Ahern et al. (2011) where mothers greater than 39 years of age were 44% more likely of LBW and similarly with McKnight et al. (in revision) where a 1% increase in maternal age led to increased chances of PTB at 6%. Tobacco consumption was associated greater odds for PTB (20.9%), LBW (117.6%), and tLBW (214.9%). This is to be expected, as smoking has long been associated with ABOs (Lewandowska et al., 2020; Silverman, 1977; Wouters et al., 1987). Infant sex was significant: female infants had a greater likelihood of LBW (19.5%) and tLBW (53%) compared to male infants (Tables 5-2 and 5-3). Due to males tending to weigh more at birth than females (Dominguez, 2008; McKnight et al., in revision), this was expected and is consistent with McKnight et al (in revision) where female births weighed on average 119.24 g less than males. However, PTB odds decreased among female babies at 12.1% compared to males (Table 5-1) which is again consistent with McKnight et al. (in revision) where female births resulted in a 10% decrease in chances of having a PTB compared to male births. Increases in parity were associated with significant decreases in LBW and tLBW, but increases in PTB (Tables 5-2 and 5-3).

Many previous studies have noted differences in ABOs across maternal race/ethnicity and educational attainment (Ahern et al., 2011; Luo et al., 2006; McKnight et al., in revision). For example, a rurality and population demographics study in Alabama, demonstrated higher rates of PTB and LBW in black and less educated mothers (Kent et al., 2013). In the present study, mothers who participated in any level of post high school education were associated with lesser odds of LBW (32.3%) and tLBW (37.2%) when compared to mothers who were in eighth grade or less (Tables 5-2 and 5-3). These results are congruent with the findings of Ahern et al. (2011) where for every 1% increase in educational years, LBW decreased by 4%. Maternal race was significant for LBW and tLBW: black mothers had elevated risks of 102.7% and 111% respectively compared to white mothers (Tables 5-2 and 5-3). Findings are consistent with McKnight et al. (in revision) where infants with a black mother weighed 203.44 g less than infants with a white mother.

6.3. Community Water System Size and Violations

Relationships between community water system (CWS) size and ABOs varied. For both the LBW and tLBW models, medium systems conveyed lesser odds at 2.4% and 11.8% respectively and small systems displayed larger chances of 10.8% and 33.4% respectively compared to large systems (Table 5-2 and 5-3). Higher rates of ABO would be expected for smaller systems as multiple past studies have indicated that smaller rural CWS incur more SDWA violations than larger more urban systems (Allaire et al., 2018; Marcillo & Krometis, 2019), likely due to management and financial struggles (Fu et al., 2020; Smith, 2020; USEPA, 2011). However, the PTB model counterintuitively reported larger systems having the highest odds of PTB compared to small and medium systems at 8.4% and 8.8% respectively (Table 5-1). This may be the result of more complex interactions between urban land cover, demographics, and healthcare availability, i.e., larger systems are likely in more urban areas with more services.

Health based (HB) violations significantly associated with PTB and LBW: a 1% increase in number of violations was associated with increases in odds of 3.7% and 2.3% (Tables 5-1 and 5-2). This is to be expected, as a HB violation indicates water quality levels in exceedance of maximum contaminant level guidelines; the most common HB violations for this dataset were for exceedances of coliform and disinfection byproducts (Smith, 2020). However, it is worth noting that overall, the most common violations for CWSs in Central Appalachia for the target study period were monitoring and reporting (MR) violations at 94% (Smith, 2020). Of the MR violation totals, 78% fell under the chemicals rule group, a list of regulations encompassing arsenic, lead and copper, inorganic chemicals, nitrate, radionuclides, synthetic organic chemicals (SOCs) and volatile organic chemicals (VOCs) (Smith, 2020). Surprisingly, relationships between MR violations and ABOs were statistically insignificant in all model runs. Previous work by Marcillo and Krometis (2019) and (Fu et al., 2020) suggest that monitoring and reporting violations, while not always necessarily indicative of an immediate health threat, do potentially obscure HB violations and should be of more concern, particularly for small, underfunded, and rural CWS.

7. CONCLUSION

This study examined potential associations between land cover and adverse birth outcomes (ABOs) in Central Appalachia through a watershed epidemiology approach that considered upstream land cover in drinking water source watersheds as a potential adverse exposure. Although no source watershed land cover was predictive of changes in tLBW, it is notable that increases in both high impact development and mining land cover were significantly associated with increased likelihoods of both PTB and LBW. Numbers of CWS health based (HB) violations of the Safe Drinking Water Act were also significantly associated with an increased likelihood of PTB and LBW; however, it is critical to note that births were assigned to CWSs via approximated service areas, which introduces the potential for classification bias. Community water system (CWS) size was significantly associated with PTB and tLBW outcomes, though these relationships appeared complex and sometimes counter-intuitive. Demographic variables significantly predictive of ABOs included maternal age, the mother's educational level, parity, maternal race, the child's sex, and maternal smoking use.

Despite the limitations posed by data availability, the associations between PTB and LBW and upstream land cover are compelling and demand further attention. Both urbanization and surface mining are acknowledged to result in detrimental impacts to downstream water quality, which can present challenges to CWS treatment processes. Given increasing national attention on issues of water equity, particularly in the Central Appalachian region, further direct examination of the impacts of anthropogenic land cover on water treatment and typical consumption rates is warranted to better characterize exposure.

8. FUTURE WORK

Limitations and sources of bias in this study largely stem from data availability, and required several assumptions that are worth noting. Removing all births without complete street addresses, such as those associated with Post Office boxes, reduced the total dataset by nearly half, with a larger number of births removed from rural counties. Covariates associated with births are potentially at risk for inaccuracies, e.g. mothers who smoked during pregnancy may inaccurately report tobacco use based on social perceptions (Ahern et al., 2011). Delineations of CWS service areas are publicly unavailable (McDonald et al., 2022), and approximated service areas are unlikely to truly capture the extents of water service, as infrastructure development depends on both topography and local political and economic pressures. It is likely that this resulted in some exposure misclassification, as some birth records could belong to different CWSs than assigned, or were reliant on private systems. Simply residing in a CWS service area does not necessarily ensure that the residents regularly consume the water provided in their homes; reliance on bottled water in the United States is considerable (Javidi & Pierce, 2018), and may be particularly widespread in Central Appalachia (Levêque & Burns, 2017; McSpirit & Reid, 2011).

To better understand the association between land cover and adverse birth outcomes, future work should be conducted to address the limitations and sources of bias listed above as a first step. Ideally, community-scale efforts to sample home water use and quality would better characterize exposure, although it is uncertain how generalizable these results would be to the region more broadly and it is unlikely that every occupied home would ever be measured. Similarly, individual level surveys of mothers might improve demographic information and better represent rural areas. Publicly available service area extents, as well as information on service interruptions (e.g., pipe breaks) and servicing would improve exposure classification at a broader scale. To better characterize the influence of land cover, upstream flow paths should be individually delineated for each watershed in the study region to better understand the natural flow of these contaminants. Furthermore, in-stream measurements should be taken at the intake points to observe water quality and quantity accurately. Lastly, this effort focused solely on public water systems specifically dependent on surface water in order to explore the use of watersheds to approximate exposure. Therefore it is important to consider surface water and groundwater exchanges along with groundwater pathway drainages as these are much more complicated than surface water flow paths individually. Ideally, groundwater and private systems data should be included in future analyses, although previous examinations suggest that identifying and quantifying contaminant origins in these systems is particularly complex (Hynds et al., 2014).

REFERENCES

Ahern, M., Mullett, M., MacKay, K., & Hamilton, C. (2011). Residence in coal-mining areas and low-birth-weight outcomes. *Maternal and Child Health Journal*, *15*(7), 974–979. https://doi.org/10.1007/s10995-009-0555-1

Allaire, M., Wu, H., & Lall, U. (2018). National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(9), 2078–2083. https://doi.org/10.1073/PNAS.1719805115/SUPPL_FILE/PNAS.201719805SI.PDF

Angermeier, P. L., Krometis, L. A., Stern, M. J., & Hemby, T. L. (2021). Exploring relationships among stream health, human well-being, and demographics in Virginia, USA. *Ecological Indicators*, *121*(July 2020), 107194. https://doi.org/10.1016/j.ecolind.2020.107194

ARC. (2021). *Subregions in Appalachia - Appalachian Regional Commission*. https://www.arc.gov/map/subregions-in-appalachia/

Behringer, B., & Friedell, G. H. (2006). Appalachia: Where place matters in health. *Preventing Chronic Disease*, *3*(4), 4–7.

Boyles, A. L., Blain, R. B., Rochester, J. R., Avanasi, R., Goldhaber, S. B., McComb, S., Holmgren, S. D., Masten, S. A., & Thayer, K. A. (2017). Systematic review of community health impacts of mountaintop removal mining. In *Environment International* (Vol. 107, pp. 163–172). Elsevier Ltd. https://doi.org/10.1016/j.envint.2017.07.002

Brunekreef, B. (2013). Exposure science, the exposome, and public health. In *Environmental and Molecular Mutagenesis* (Vol. 54, Issue 7, pp. 596–598). https://doi.org/10.1002/em.21767

Buttling, L. G., McKnight, M. X., Kolivras, K. N., Ranganathan, S., & Gohlke, J. M. (2021). Maternal proximity to Central Appalachia surface mining and birth outcomes. *Environmental Epidemiology*, *5*(1), e128. https://doi.org/10.1097/ee9.0000000000000128

Chini, C. M., & Stillwell, A. S. (2017). Where Are All the Data? The Case for a Comprehensive Water and Wastewater Utility Database. *Journal of Water Resources Planning and Management*, *143*(3), 01816005. https://doi.org/10.1061/(asce)wr.1943-5452.0000739

Corley, B., Bartelt-Hunt, S., Rogan, E., Coulter, D., Sparks, J., Baccaglini, L., Howell, M., Liaquat, S., Commack, R., & Kolok, A. S. (2018). Using Watershed Boundaries to Map Adverse Health Outcomes: Examples From Nebraska, USA. *Environmental Health Insights*, *12*. https://doi.org/10.1177/1178630217751906

Dominguez, T. P. (2008). Race, racism, and racial sisparities in adverse birth outcomes. *Clin Obstet Gynecol*, *52*(2), 360–370. https://doi.org/10.1097/GRF.0b013e31816f28de

Environmental Protection Agency, U. (2005). *A R GRICULTURAL UNOFF from Protecting Water Quality*. www.fsa.usda.gov/dafp/cepd/default.

Fu, G., Liu, P., & Swallow, S. K. (2020). Effectiveness of Public versus Private Ownership: Violations of the Safe Drinking Water Act (SDWA). *Agricultural and Resource Economics Review*, *49*(2), 291–320. https://doi.org/10.1017/age.2020.4

Fuchs, F., Monet, B., Ducruet, T., Chaillet, N., & Audibert, F. (2018). *Effect of maternal age on the risk of preterm birth: A large cohort study*. https://doi.org/10.1371/journal.pone.0191002

Garner, E., Zhu, N., Strom, L., Edwards, M., & Pruden, A. (2016). A human exposome framework for guiding risk management and holistic assessment of recycled water quality. In *Environmental Science: Water Research and Technology* (Vol. 2, Issue 4, pp. 580–598). Royal Society of Chemistry. https://doi.org/10.1039/c6ew00031b

Gohlke, J. M. (2021). Human Health in Coalfield Communities of Appalachia. In *Appalachia's Coal-Mined Landscapes* (pp. 311–336). Springer International Publishing. https://doi.org/10.1007/978-3-030-57780-3_12

Hendryx, M. (2015). The public health impacts of surface coal mining. *The Extractive Industries and Society*, *2*(4), 820–826. https://doi.org/10.1016/J.EXIS.2015.08.006

Hendryx, M., & Ahern, M. (2012). Reply to Borak et al "mortality disparities in appalachia: Reassessment of major risk factors." In *Journal of Occupational and Environmental Medicine* (Vol. 54, Issue 7, pp. 768–769). https://doi.org/10.1097/JOM.0b013e318254622f

Holzman, D. C. (2011). Mountaintop Removal Mining: Digging Into Community Health Concerns. *Environmental Health Perspectives*, *119*(11), a476. https://doi.org/10.1289/EHP.119-A476

Hoorman, J., Hone, T., Sudman, T., Dirksen, T., Iles, J., & Islam, K. R. (2008). Agricultural impacts on lake and stream water quality in Grand Lake St. Marys, western Ohio. *Water, Air, and Soil Pollution*, *193*(1–4), 309–322. https://doi.org/10.1007/s11270-008-9692-1

Hynds, P. D., Thomas, M. K., & Pintar, K. D. M. (2014). Contamination of Groundwater Systems in the US and Canada by Enteric Pathogens, 1990–2013: A Review and Pooled-Analysis. *PLoS ONE*, *9*(5). https://doi.org/10.1371/JOURNAL.PONE.0093301

IPCS, (International Programme on Chemical safety). (2004). IPCS Risk Assessment Terminology Part 1: IPCS/OECD Key Generic Terms used in Chemical Hazard/Risk Assessment and Part 2: IPCS Glossary of Key Exposure Assessment Terminology. *Geneva: World Health Organization*. http://www.inchem.org/documents/harmproj/harmproj/harmproj1.pdf

Javidi, A., & Pierce, G. (2018). U.S. Households' Perception of Drinking Water as Unsafe and its Consequences: Examining Alternative Choices to the Tap. *Water Resources Research*, *54*(9), 6100–6113. https://doi.org/10.1029/2017WR022186

Jenkins, A., Capon, A., Negin, J., Marais, B., Sorrell, T., Parkes, M., & Horwitz, P. (2018). Watersheds in planetary health research and action. In *The Lancet Planetary Health* (Vol. 2, Issue 12, pp. e510–e511). Elsevier B.V. https://doi.org/10.1016/S2542-5196(18)30228-6

Jordan, S. J., & Benson, W. H. (2015). Sustainable Watersheds: Integrating Ecosystem Services and Public Health. In *Environmental Health Insights* (Vol. 9). SAGE Publications Inc. https://doi.org/10.4137/EHI.S19586

Kent, S. T., McClure, L. A., Zaitchik, B. F., & Gohlke, J. M. (2013). Area-level risk factors for adverse birth outcomes: Trends in urban and rural settings. *BMC Pregnancy and Childbirth*, *13*. https://doi.org/10.1186/1471-2393-13-129

Kolok, A. S., Beseler, C. L., Chen, X.-H., & Shea, P. J. (2009). The Watershed as A Conceptual Framework for the Study of Environmental and Human Health. In *Environmental Health Insights*.

Krometis, L. A., Gohlke, J., Kolivras, K., Satterwhite, E., Marmagas, S. W., & Marr, L. C. (2017). Environmental health disparities in the Central Appalachian region of the United States. *Reviews on Environmental Health*, *32*(3), 253–266. https://doi.org/10.1515/reveh-2017-0012

Levêque, J. G., & Burns, R. C. (2017). Predicting water filter and bottled water use in Appalachia: A community-scale case study. *Journal of Water and Health*, *15*(3), 451–461. https://doi.org/10.2166/wh.2017.219

Lewandowska, M., Więckowska, B., Sztorc, L., & Sajdak, S. (2020). Smoking and smoking cessation in the risk for fetal growth restriction and low birth weight and additive effect of maternal obesity. *Journal of Clinical Medicine*, *9*(11), 1–18. https://doi.org/10.3390/jcm9113504

Luo, Z. C., Wilkins, R., & Kramer, M. S. (2006). Effect of neighbourhood income and maternal education on birth outcomes: A population-based study. *CMAJ*, *174*(10), 1415–1420. https://doi.org/10.1503/CMAJ.051096

Marcillo, C., & Krometis, L. A. (2019). Small Towns, Big Challenges: Does Rurality Influence Safe Drinking Water Act Compliance? *AWWA Water Science*, *1*(1), e1120. https://doi.org/https://doi.org/10.1002/aws2.1120

Marcillo, C., Krometis, L. A., & Krometis, J. (2021). Approximating Community Water System Service Areas to Explore the Demographics of SDWA Compliance in Virginia. *International Journal of Environmental Research and Public Health*, *18*(24). https://doi.org/10.3390/IJERPH182413254

McDonald, Y. J., Anderson, K. M., Caballero, M. D., Ding, K. J., Fisher, D. H., Morkel, C. P., & Hill, E. L. (2022). A systematic review of geospatial representation of United States community water systems. *AWWA Water Science*, *4*(1), e1266. https://doi.org/10.1002/AWS2.1266

Mcgarvey, E. L., Leon-Verdin, M., Killos, L. F., Guterbock, T., & Cohn, W. F. (2011). *Health Disparities Between Appalachian and Non-Appalachian Counties in Virginia USA*. https://doi.org/10.1007/s10900-010-9315-9

McKnight, M. X., Kolivras, K. N., Gohlke, J. M., & Thomas, J. P. (in revision). *Examining The Modifiable Areal Unit Problem: Associations between Surface Mining and Birth Outcomes in Central Appalachia at Multiple Spatial Scales*.

McSpirit, S., & Reid, C. (2011). Residents' perceptions of tap water and decisions to purchase bottled water: A survey analysis from the Appalachian, Big Sandy coal mining region of West Virginia. *Society and Natural Resources*, *24*(5), 511–520. https://doi.org/10.1080/08941920903401432

Meit, M., Heffernan, M., & Tanenbaum, E. (2019). Investigating the impact of the diseases of despair in Appalachia. *Journal of Appalachian Health*, *1*(2), 7–18.

NDA. (2022, February). *Nebraska Agriculture Fact Card*. https://nda.nebraska.gov/facts.pdf

Omernik, J. M., Griffith, G. E., Hughes, R. M., Glover, J. B., & Weber, M. H. (2017). How Misapplication of the Hydrologic Unit Framework Diminishes the Meaning of Watersheds. *Environmental Management*, *60*(1). https://doi.org/10.1007/s00267-017-0854-z

Palmer, M. A., Bernhardt, E. S., Schlesinger, W. H., Eshleman, K. N., Foufoula-Georgiou, E., Hendryx, M. S., Lemly, A. D., Likens, G. E., Loucks, O. L., Power, M. E., White, P. S., & Wilcock, P. R. (2010). Mountaintop mining consequences. *Science*, *327*(5962), 148–149. https://doi.org/10.1126/SCIENCE.1180543/FORMAT/PDF

Pericak, A. A., Thomas, C. J., Kroodsma, D. A., Wasson, M. F., Ross, M. R. V., Clinton, N. E., Campagna, D. J., Franklin, Y., Bernhardt, E. S., & Amos, J. F. (2018). Mapping the yearly extent of surface coal mining in Central Appalachia using Landsat and Google Earth Engine. *PLOS ONE*, *13*(7), e0197758. https://doi.org/10.1371/JOURNAL.PONE.0197758

Pleil, J. D. (2012). Categorizing biomarkers of the human exposome and developing metrics for assessing environmental sustainability. *Journal of Toxicology and Environmental Health - Part B: Critical Reviews*, *15*(4), 264–280. https://doi.org/10.1080/10937404.2012.672148

Rappaport, S. M., & Smith, M. T. (2010). Environment and disease risks. *Science*, *330*(6003), 460–461. https://doi.org/10.1126/science.1192603

Ruktanonchai, C. W., McKnight, M. X., Buttling, L., Kolivras, K., Krometis, L.-A., & Gohlke, J. (2022). Identifying exposure pathways mediating adverse birth outcomes near active surface mines in Central Appalachia. *Environmental Epidemiology*, *6*(3), e208. https://doi.org/10.1097/EE9.0000000000000208

Salerno, F., Viviano, G., & Tartari, G. (2018). Urbanization and climate change impacts on surface water quality: Enhancing the resilience by reducing impervious surfaces. *Water Research*, *144*, 491–502. https://doi.org/10.1016/j.watres.2018.07.058

Silverman. (1977). *This study examines the relationship between maternal smoking and birth weight . The association of these two fac- tors has been extensively investigated (1-. 105*(6), 513–521.

Smith, E. P. (2020). *Land Cover As a Predictor of Safe Drinking Water Act Violations in Central Appalachia*. M.S. Thesis, Virginia Tech, Blacksburg, VA, USA

Tilstra, A.M., Masters, R.K. (2020). Worth the Weight? Recent Trends in Obstetric Practices, Gestational Age, and Birth Weight in the United States. *Demography, 57*, 99-121.

USEPA. (2011). National Characteristics of Drinking Water Systems: Serving 10,000 or Fewer People. *Water*, *July*.

Wenger, S. J., Roy, A. H., Jackson, C. R., Bernhardt, E. S., Carter, T. L., Filoso, S., Gibson, C. A., Hession, W. C., Kaushal, S. S., Mart, E., Meyer, J. L., Palmer, M. A., Paul, M. J., Purcell, A. H., Ramrez, A., Rosemond, A. D., Schofield, K. A., Sudduth, E. B., & Walsh, C. J. (2009). Twenty-six key research questions in urban stream ecology: an assessment of the state of the science. *Https://Doi.Org/10.1899/08-186.1*, *28*(4), 1080–1098. https://doi.org/10.1899/08-186.1

Wild, C. P. (2005). Complementing the genome with an “exposome”: The outstanding challenge of environmental exposure measurement in molecular epidemiology. In *Cancer Epidemiology Biomarkers and Prevention* (Vol. 14, Issue 8, pp. 1847–1850). https://doi.org/10.1158/1055-9965.EPI-05-0456

Woolley, S. M., Meacham, S. L., Balmert, L. C., Talbott, E. O., & Buchanich, J. M. (2015). Comparison of mortality disparities in central appalachian coal- and non-coal-mining counties. *Journal of Occupational and Environmental Medicine*, *57*(6), 687–694. https://doi.org/10.1097/JOM.0000000000000435

Wouters, E. J. M., de Jong, P. A., Cornelissen, P. J. H., Kurver, P. H. J., van Oel, W. C., & van Woensel, C. L. M. (1987). Smoking and low birth weight: absence of influence by carbon monoxide? *European Journal of Obstetrics and Gynecology and Reproductive Biology*, *25*(1), 35–41. https://doi.org/10.1016/0028-2243(87)90090-6

APPENDIX A

A.1. RStudio Code for the Binomial Generalized Linear Model

```
library(glm)
library(splines)
data=read.csv(choose.files())
attach(data)
data$paritycat[Parity == 1] <- "1"
data$paritycat[Parity == 2] <- "2"
data$paritycat[Parity == 3] <- "3"
data$paritycat[Parity >= 4] <- "4"
detach(data)
#Correlation Check Example
#cor(data$LowDev.,data$Vegetation)
#Variable Classification
data$MotherAge=as.factor(data$MotherAge)
data$Mother_Edu=as.factor(data$Mother_Edu)
data$Mother_Rac=as.factor(data$Mother_Rac)
data$Mother_His=as.factor(data$Mother_His)
data$Child_Sex=as.factor(data$Child_Sex)
data$Tobacco=as.factor(data$Tobacco)
data$Low_Birthw=as.factor(data$Low_Birthw)
data$Preterm=as.factor(data$Preterm)
data$Term_Low_B=as.factor(data$Term_Low_B)
data$PWS_ID=as.factor(data$PWS_ID)
data$Size=as.factor(data$Size)
data$State=as.factor(data$State)
#Covariated Models
Model1=glm(formula=Term_Low_B~MotherAge+Parity+Mother_Edu+Mother_Rac+Mother_H
is+Child_Sex+Tobacco+Size+HighDev.+LowDev.+Ag.+Mining.+Barren.+MR+HB+State+bs(
Gestational_Year,df=4),family = "binomial",data = data)
Model2=glm(formula=Low_Birthw~MotherAge+Parity+Mother_Edu+Mother_Rac+Mother_Hi
s+Child_Sex+Tobacco+Size+HighDev.+LowDev.+Ag.+Mining.+Barren.+MR+HB+State+bs(G
estational_Year,df=4),family = "binomial",data = data)
Model3=glm(formula=Preterm~MotherAge+Parity+Mother_Edu+Mother_Rac+Mother_His+Ch
ild_Sex+Tobacco+Size+HighDev.+LowDev.+Ag.+Mining.+Barren.+MR+HB+State+bs(Gestati
onal_Year,df=4),family = "binomial",data = data)
```

A.2. Raw RStudio Code Outputs

```
Call:
glm(formula = Preterm ~ MotherAge + paritycat + Mother_Edu +
    Mother_Rac + Mother_His + Child_Sex + Tobacco + Size + HighDev. +
    LowDev. + Ag. + Mining. + Barren. + MR + HB + State + bs(Gestational_Year,
    df = 4), family = "binomial", data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0788  -0.4971  -0.4573  -0.4213   2.4561

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                   -2.4528235  0.1707899 -14.362  < 2e-16 ***
MotherAge2                    -0.1707566  0.0736162  -2.320 0.020365 *
MotherAge3                     0.3369657  0.0539619   6.245 4.25e-10 ***
paritycat1                    -0.0630849  0.0311549  -2.025 0.042880 *
paritycat2                     0.0190089  0.0386519   0.492 0.622863
paritycat3                     0.2623010  0.0469344   5.589 2.29e-08 ***
Mother_Edu2                    0.0943235  0.0830626   1.136 0.256136
Mother_Edu3                   -0.0406508  0.0847985  -0.479 0.631667
Mother_Rac2                    0.1226783  0.1065634   1.151 0.249641
Mother_Rac3                   -0.1294187  0.1228260  -1.054 0.292032
Mother_His2                    0.2806184  0.1471109   1.908 0.056452 .
Child_Sex2                    -0.1287948  0.0260779  -4.939 7.86e-07 ***
Tobacco2                       0.1894041  0.0285233   6.640 3.13e-11 ***
SizeMedium                    -0.0924785  0.0343039  -2.696 0.007021 **
SizeSmall                     -0.0875420  0.0895530  -0.978 0.328300
HighDev.                       0.0995930  0.0314056   3.171 0.001518 **
LowDev.                       -0.0315701  0.0090226  -3.499 0.000467 ***
Ag.                           -0.0010265  0.0009116  -1.126 0.260155
Mining.                        0.2135317  0.0490447   4.354 1.34e-05 ***
Barren.                       -0.0886731  0.0382534  -2.318 0.020447 *
MR                            -0.0009857  0.0040043  -0.246 0.805556
HB                             0.0365293  0.0097000   3.766 0.000166 ***
StateTN                        1.4871352  0.4787731   3.106 0.001895 **
StateVA                       -0.3397315  0.3806146  -0.893 0.372079
StateWV                       -0.2815005  0.1785331  -1.577 0.114855
bs(Gestational_Year, df = 4)1 -0.0743546  0.1151870  -0.646 0.518595
bs(Gestational_Year, df = 4)2  0.4469134  0.0979475   4.563 5.05e-06 ***
bs(Gestational_Year, df = 4)3 -0.0665003  0.0884800  -0.752 0.452300
bs(Gestational_Year, df = 4)4  0.0220803  0.0663965   0.333 0.739473
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 42468  on 62655  degrees of freedom
Residual deviance: 42133  on 62627  degrees of freedom
AIC: 42191

Number of Fisher Scoring iterations: 5
```

Figure A-1. Model results for the PTB outcome variable

```
Call:
glm(formula = Term_Low_B ~ MotherAge + paritycat + Mother_Edu +
    Mother_Rac + Mother_His + Child_Sex + Tobacco + Size + HighDev. +
    LowDev. + Ag. + Mining. + Barren. + MR + HB + State + bs(Gestational_Year,
    df = 4), family = "binomial", data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7472  -0.2730  -0.2015  -0.1626   3.1978

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                 -3.9872941  0.3105733 -12.838  < 2e-16 ***
MotherAge2                  -0.0982328  0.1210563  -0.811 0.417099
MotherAge3                   0.4775718  0.1029121   4.641 3.47e-06 ***
paritycat1                  -0.4010397  0.0583739  -6.870 6.41e-12 ***
paritycat2                  -0.3452066  0.0724053  -4.768 1.86e-06 ***
paritycat3                  -0.2498003  0.0875256  -2.854 0.004317 **
Mother_Edu2                 -0.1933338  0.1313620  -1.472 0.141085
Mother_Edu3                 -0.4646999  0.1370816  -3.390 0.000699 ***
Mother_Rac2                  0.7466780  0.1550355   4.816 1.46e-06 ***
Mother_Rac3                 -0.1706428  0.2490307  -0.685 0.493200
Mother_His2                 -0.1031501  0.2697413  -0.382 0.702162
Child_Sex2                   0.5301721  0.0491381  10.789  < 2e-16 ***
Tobacco2                     1.1470882  0.0523829  21.898  < 2e-16 ***
SizeMedium                  -0.0177957  0.0625559  -0.284 0.776045
SizeSmall                    0.2880253  0.1388871   2.074 0.038097 *
HighDev.                     0.0639073  0.0541779   1.180 0.238166
LowDev.                     -0.0174403  0.0159432  -1.094 0.273998
Ag.                         -0.0008904  0.0016979  -0.524 0.600017
Mining.                     -0.0209205  0.0829736  -0.252 0.800937
Barren.                      0.0366020  0.0582662   0.628 0.529883
MR                          -0.0040304  0.0081964  -0.492 0.622908
HB                          -0.0339024  0.0198987  -1.704 0.088428 .
StateTN                      1.4715878  0.6394685   2.301 0.021377 *
StateVA                     -0.3167176  0.6140091  -0.516 0.605981
StateWV                      0.2225433  0.2530568   0.879 0.379173
bs(Gestational_Year, df = 4)1  0.2105170  0.2171516   0.969 0.332322
bs(Gestational_Year, df = 4)2  0.2214609  0.1829708   1.210 0.226140
bs(Gestational_Year, df = 4)3  0.3782689  0.1642819   2.303 0.021304 *
bs(Gestational_Year, df = 4)4  0.5244376  0.1233060   4.253 2.11e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 16419  on 62655  degrees of freedom
Residual deviance: 15554  on 62627  degrees of freedom
AIC: 15612

Number of Fisher Scoring iterations: 7
```

Figure A-2. Model results for the tLBW outcome variable

```
Call:
glm(formula = Low_Birthw ~ MotherAge + paritycat + Mother_Edu +
    Mother_Rac + Mother_His + Child_Sex + Tobacco + Size + HighDev. +
    LowDev. + Ag. + Mining. + Barren. + MR + HB + State + bs(Gestational_Year,
    df = 4), family = "binomial", data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9697  -0.4464  -0.3593  -0.3064   2.6924

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)
(Intercept)                  -2.8749225  0.2006411 -14.329  < 2e-16 ***
MotherAge2                   -0.1101185  0.0764855  -1.440 0.149944
MotherAge3                    0.4636150  0.0634263   7.310 2.68e-13 ***
paritycat1                   -0.3750262  0.0362200 -10.354  < 2e-16 ***
paritycat2                   -0.3438412  0.0456664  -7.529 5.10e-14 ***
paritycat3                   -0.0772064  0.0536083  -1.440 0.149812
Mother_Edu2                  -0.0970748  0.0867964  -1.118 0.263388
Mother_Edu3                  -0.3904553  0.0897845  -4.349 1.37e-05 ***
Mother_Rac2                   0.7063646  0.1012592   6.976 3.04e-12 ***
Mother_Rac3                  -0.1268453  0.1488480  -0.852 0.394114
Mother_His2                   0.2262013  0.1762967   1.283 0.199467
Child_Sex2                    0.1958988  0.0299608   6.539 6.21e-11 ***
Tobacco2                      0.7774534  0.0319326  24.347  < 2e-16 ***
SizeMedium                   -0.0244811  0.0389697  -0.628 0.529868
SizeSmall                     0.1025909  0.0948846   1.081 0.279601
HighDev.                      0.0880143  0.0348434   2.526 0.011537 *
LowDev.                      -0.0218135  0.0100879  -2.162 0.030592 *
Ag.                          -0.0009174  0.0010529  -0.871 0.383558
Mining.                       0.1337469  0.0541364   2.471 0.013490 *
Barren.                      -0.0477551  0.0406859  -1.174 0.240496
MR                           -0.0045240  0.0049066  -0.922 0.356516
HB                            0.0231385  0.0113187   2.044 0.040926 *
StateTN                       1.1346322  0.5258350   2.158 0.030946 *
StateVA                      -0.2358613  0.3875201  -0.609 0.542761
StateWV                       0.0876697  0.1715915   0.511 0.609406
bs(Gestational_Year, df = 4)1 -0.0184478  0.1337659  -0.138 0.890311
bs(Gestational_Year, df = 4)2  0.2657533  0.1133451   2.345 0.019046 *
bs(Gestational_Year, df = 4)3  0.2709602  0.1013832   2.673 0.007526 **
bs(Gestational_Year, df = 4)4  0.2522652  0.0765483   3.296 0.000982 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 34627  on 62655  degrees of freedom
Residual deviance: 33478  on 62627  degrees of freedom
AIC: 33536

Number of Fisher Scoring iterations: 5
```

Figure A-3. Model results for the LBW outcome variable