ATTACHMENT 31

Final Report
For

Pennsylvania Department of Health,
Bureau of Epidemiology

Hydraulic Fracturing Epidemiology Research Studies: Childhood Cancer Case-Control Study

Prepared by:
University of Pittsburgh
School of Public Health
Contract number: 4400018535

August 3, 2023

List of Contributors

Faculty Investigators

Evelyn O. Talbott, MPH, DrPH, FAHA, Professor, Department of Epidemiology, Principal Investigator, Childhood Cancer Case-Control Study
Jeanine M. Buchanich, MPH, MEd, PhD, Associate Professor, Department of Biostatistics
Principal Investigator, Retrospective Cohort Studies
Todd M. Bear, PhD, MPH, Assistant Professor, Department of Family Medicine, School of Medicine
James P. Fabisiak, PhD, Associate Professor, Department of Environmental Occupational Health
Jean Tersak MD, MS, Professor, Division of Pediatric Hematology/Oncology, UPMC, Children's Hospital of Pittsburgh
Sally E. Wenzel, MD, Chair, Department of Environmental and Occupational Health
Ada O. Youk, PhD, Associate Professor, Department of Biostatistics
Jian-Min Yuan, MD, PhD, Professor, Department of Epidemiology, Co-leader of the Cancer Epidemiology and Prevention Program, UPMC Hillman Cancer Center

Consultants

Vincent C. Arena, PhD, Associate Professor Emeritus, Department of Biostatistics
Ravi Sharma, PhD, Adjunct Assistant Professor, Department of Epidemiology

Staff

Jennifer F. Fedor, MS, Data Analysis
Abigail Foulds, MA, PhD, Research Coordinator
Allison C. Koller, MS, Project Director
Michael F. Lann, MSIS, Programmer
David Maynard, Administrative Assistant
Becky Meehan, MS, Research Assistant
Andrew Mrkva, MA, Data Manager/Web Developer
Natalie F. Price, MPH, Data Manager
Melanie Stangl, Administrative Assistant
Rachel Taber, MPH, PhD, Recruitment
Renwei Wang, MD, MS, Biostatistician, Senior Scientist, UPMC Hillman Cancer Center

Students

Samantha Bayer, MPH, Department of Epidemiology
Kathleen Gruschow, MPH, Department of Epidemiology
Caroline Hoffman, MPH, Department of Epidemiology
Madelyn Kapfhammer, Department of Epidemiology
Kristen Steffes, MPH, Department of Epidemiology
Fan Wu, MPH, Department of Environmental and Occupational Health

Acknowledgments

A special thank you to our study participants who gave of their time to participate in this study. It is through their efforts that this research was made possible.

The study team wishes to acknowledge and thank our External Advisory Board members for their thoughtful input throughout the study, including at meetings on August 5, 2021, November 10, 2021, January 26, 2022, April 13, 2022, and August 11, 2022.

External Advisory Board Members

David Allard, MS, Pennsylvania Department of Environmental Protection, Bureau of Radiation Protection (retired)
Uni Blake, PhD, American Petroleum Institute
Laura Dagley, Physicians for Social Responsibility (resigned)
Erica Jackson, FracTracker Alliance (resigned)
Ned Ketyer, MD, Physicians for Social Responsibility (resigned)
Jan Maund, Trinity Area School District (retired)
Curtis Schreffler, Deputy Director, Hydrologic Studies, PA Division, United States Geological Survey
Heaven Sensky, Center for Coalfield Justice (resigned)
Logan Spector, PhD, University of Minnesota Medical School (EAB Chair)
Judy Wendt Hess, PhD, Shell Oil
Edward Yorke, Canonsburg Councilman

We wish to thank the following organizations for providing study data. All interpretations are solely the responsibility the investigators.

Pennsylvania Department of Health

Pennsylvania Department of Environmental Protection

We also wish to thank Allison Hydzik, Media Relations, University of Pittsburgh Medical Center, for her invaluable assistance and advice.

Table of Contents

Abbreviations

Abbreviation	Definition
ALL	Acute Lymphocytic Leukemia
ATSDR	Agency for Toxic Substances and Disease Registry
CATI	Computer–Assisted Telephone Interviewing
CHP	Children's Hospital of Pittsburgh
CI	Confidence Intervals
CNS	Central Nervous System
CPDB	Carcinogenic Potency Database
CT	Computed Tomography
EFOT	Ewing Family of Tumor
EPA	US Environmental Protection Agency
GIS	Geographic Information System
HF	Hydraulic Fracturing or fracking
IARC	International Agency for Research on Cancer
ICCC	International Classification of Childhood Cancer
IDW	Inverse Distance Weighting
IRB	Institutional Review Board
NCI	National Cancer Institute
NHL	Non-Hodgkin Lymphoma
NPL	National Priorities List
OR	Odds Ratio
PA	Pennsylvania
PADEP	Pennsylvania Department of Environmental Protection
PADOH	Pennsylvania Department of Health
SEER	Surveillance, Epidemiology, and End Results
SIR	Standard Incidence Ratios
TRI	Toxic Release Inventory
UMTRA	Uranium Mill Tailing Remedial Action
UNGD	Unconventional Natural Gas Development
UV	Ultraviolet
WCS	World Geocoding Service
WHO	World Health Organization

CHILDHOOD CANCER CASE-CONTROL STUDY

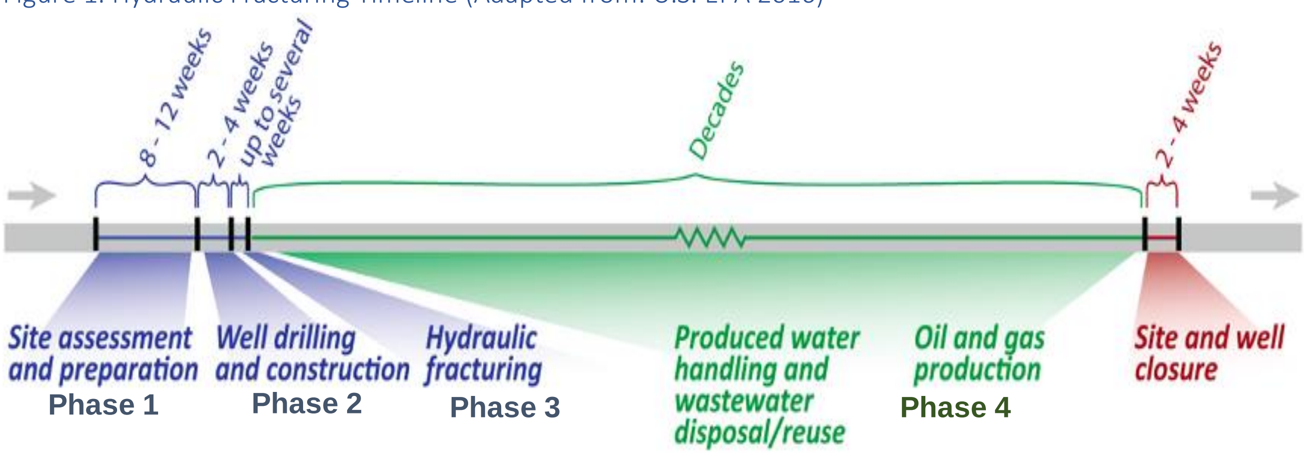
I. Background

Hydraulic fracturing (or fracking) is a type of unconventional natural gas development (UNGD) used to extract natural gas from underground shale rock formations. After obtaining the necessary permits, the first phase of hydraulic fracturing (HF) is well pad preparation. This includes preparing a site for one or more fracturing wells by building access roads and clearing land to build infrastructure. The next phase is drilling in which a borehole is drilled vertically 1 to 2 miles into the ground then turned horizontally into the shale rock (Deziel et al., 2022). Then the steel casing is installed in the borehole and sealed with cement.

Fracturing fluid consists of 90-97% of a base fluid, which is usually water. A fracturing well uses an average of 1.2 million gallons of water. A proppant, usually sand, composes 2-10% of the fracturing fluid. Chemical additives make up less than 2% of the fracturing fluid, though hundreds of chemicals have been reported (Deziel et al., 2022). More information on the chemical additives and their function in fracturing fluid, as well as common constituents reported by the EPA analysis of FracFocus 1.0 (2015) is shown in **Appendix A**. A number of these chemicals include known and suspected endocrine inhibitors and carcinogens (Deziel et al., 2022).

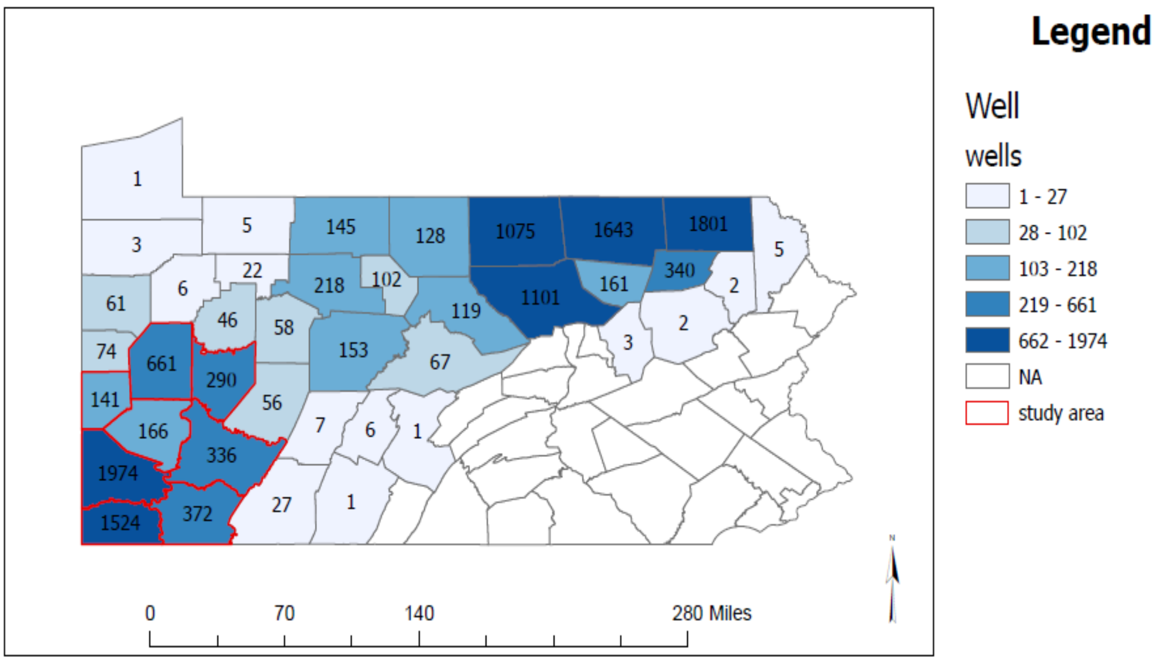
Workers inject this fracturing fluid into the well under high pressure which 'fractures' the rock and releases the natural gas. Once the pressure is released, a mixture of the gas, fracturing fluid, and other compounds found in the rock flow back through the well to the surface. This mixture is often called flowback or produced water. The production phase refers to the separation of the gas from the flowback water, which is then transported through pipelines to a storage facility or processing plant (Deziel et al., 2022). See **Figure 1**.

Figure 1. Hydraulic Fracturing Timeline (Adapted from: U.S. EPA 2016)



The first recorded shale gas well in Pennsylvania was drilled in Erie County in 1860, though modern hydraulic fracturing began in earnest in 2005 in Southwestern Pennsylvania (PA). Currently, Washington County has the largest number of UNGD wells in operation in this region. As of December 2020, there were 12,903 unconventional wells active throughout PA and 5,464 in the 8 county Southwestern PA area. See **Figure 2**. The last county to begin with UNGD drilling was Allegheny County in 2008. The highlighted area on the map includes Allegheny, Armstrong, Beaver, Butler, Fayette, Greene, Washington, and Westmoreland counties, where each had >100 active unconventional oil and natural gas wells in 2020.

Figure 2. Distribution of Wells in Each PA County, with a Total of 12,903 Wells Throughout PA as of December 2020



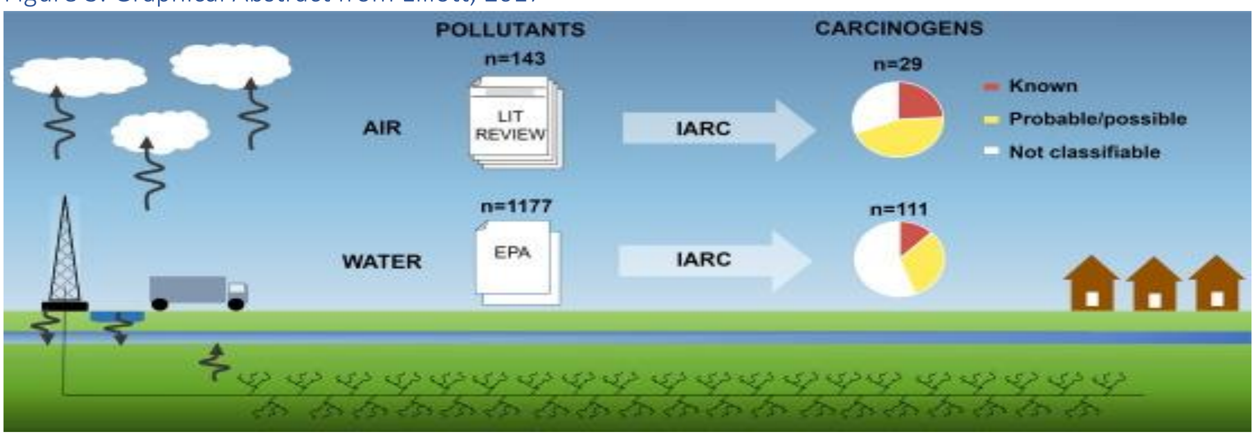
UNGD-related chemicals in the environment

A systematic assessment of carcinogenicity of chemicals in fracturing fluid and flowback water was conducted by Xu et al. (2019). The group assessed 1,173 fracturing fluid-related chemicals identified by the US Environmental Protection Agency (EPA) (Xu et al., 2019). They then linked the fracturing fluid chemical data to the agent classification data from the International Agency for Research on Cancer (IARC) at the World Health Organization (WHO), which was evaluated for human carcinogenic risk. Using IARC's database of 998 chemicals, they found information on 104 fracturing fluid-related chemicals with different evidence in carcinogenicity: 14 were carcinogenic to humans, 7 were probably carcinogenic, and 27 were possibly carcinogenic.

Some of these carcinogenic compounds include 1,3-butadiene, ethanol, ethylene oxide, and formaldehyde, which are found in fracturing fluids; benzo(a)pyrene, beryllium, cadmium, radium-226 and -228 found in flowback; and arsenic, benzene, and chromium (VI) found in both. Additional assessment of the Carcinogenic Potency Database (CPDB) suggested that 66 fracturing fluid-related chemicals are potentially carcinogenic based on rats and mouse models (Xu et al., 2019). Xu et al.'s evaluation suggests that individuals with exposure to certain chemicals in fracturing fluids and wastewater may be at increased risk of cancer, as these chemicals can make their way into ground water and drinking water.

Elliott (2017) also systematically assessed evidence for potential carcinogenicity of both air and water pollutants from hydraulic fracturing exposures but specific to childhood leukemia and lymphoma risk. They likewise evaluated 1,177 chemicals in fracturing fluids and wastewater, finding similar results as those described by Xu et al. They additionally considered 143 UNGD-related air pollutants by review of scientific papers published through 2015 using both PubMed and ProQuest Database, and assessing carcinogenicity evidence of increased risk of leukemia and lymphoma from these chemicals using the IARC monographs. See **Figure 3**.

Figure 3: Graphical Abstract from Elliott, 2017



Of 143 potential air pollutants, 29 (20%) have been evaluated for carcinogenicity by IARC and the remaining 114 (80%) have not been evaluated (Elliot, 2017). Of the 29 air pollutants evaluated, 7 (24%) were carcinogenic to humans, 2 (7%) were considered probably carcinogenic to humans, 11 (38%) were considered possibly carcinogenic to humans, and the remaining 9 (31%) could not be classified with respect to their carcinogenicity. Of the 20 known, probable, or possible carcinogens, there has been supporting evidence for 11 air pollutants that were associated with an increased risk of leukemia or lymphoma. These included 5 known human carcinogens (1,3-butadiene, benzene, ethanol, formaldehyde, diesel engine exhaust), 2 probable human carcinogens (dibenz[*a,h*]anthracene, tetrachloroethylene), and 4 possible human carcinogens (carbon tetrachloroethylene, chrysene, indenol[1,2,3-*cd] pyrene* and styrene).

Risk Factors for Childhood Cancer

Although cancer in children and adolescents is rare, it is the leading cause of death by disease past infancy among children in the United States, according to the National Cancer Institute (NCI, 2021).

In 2021, it was estimated that 15,590 children and adolescents ages 0 to 19 were diagnosed with cancer and 1,780 died of the disease in the United States (Siegel, 2021). Overall, among children and adolescents (ages 0 to 19) in the United States, the most common types of cancer are leukemias, brain and central nervous system (CNS) tumors, and lymphomas (NCI, 2021). These are also the types of cancers found to be associated with various environmental exposures in both adults and children in the literature (NCI, 2021).

Many childhood cancers are caused by genetic mutations that increase cancer risk. Germline alterations (or variants) associated with an increased risk of cancer can be passed down from parents to their offspring, or somatic mutations in cells can occur spontaneously in cells during development (NCI, 2021). About 6-8% of all cancers in children are caused by an inherited pathogenic variant (harmful alteration) in a cancer predisposition gene (Gröbner et al., 2018, Zhang et al., 2015). For example, children with Li-Fraumeni syndrome, Beckwith-Wiedemann syndrome, Fanconi anemia, Noonan syndrome, and von Hippel-Lindau syndrome, have an increased risk of childhood cancer.

Genomic changes that arise during development of one of the germ cells (sperm or egg) which unite to form the zygote that becomes a child can increase the risk of cancer in that child (NCI, 2021). Genomic changes can include broken, missing, rearranged, or extra chromosomes and gene variants. One such alteration is trisomy 21, or the presence of an extra copy of chromosome 21, which causes Down syndrome. Children with Down syndrome are 10 to 20 times more likely to develop leukemia than children without Down syndrome (Ross, 2005). However, only a small proportion of childhood leukemia is linked to Down syndrome (NCI, 2021).

Genetic changes associated with cancer can also occur in different cells of the body after birth, as the body is actively growing and developing during early childhood (Moore et al., 2021). The extent to which these changes react to environmental exposures is unclear. In adults, exposure to cancer-causing substances in the environment, such as cigarette smoke, asbestos, and ultraviolet (UV) radiation from the sun is known to cause genetic changes that can lead to cancer (NCI, 2021). However, environmental causes of childhood cancer have been particularly difficult to identify, this is partly because cancer in children is rare and because it is difficult to determine what children may have been exposed to early in their development (NCI, 2021).

Nevertheless, several environmental exposures, such as ionizing radiation, can lead to the development of leukemia and other cancers in children and adolescents (NCI, 2021). Children and adolescents who were exposed to radiation from the atomic bombs dropped in Japan during the Second World War had an elevated risk of leukemia (Hsu et al., 2013). Also, children exposed to radiation from the Chernobyl nuclear plant accident had an elevated risk for thyroid cancer (Cardis, 2011).

Exposure of parents to ionizing radiation is also a concern in terms of the development of cancer in their future offspring. Exposure to diagnostic medical radiation from computed tomography (CT) scans by children whose mothers had x-rays during pregnancy (that is, children who were exposed before birth) and children exposed after birth has been linked to a slight increase in risk of leukemia and brain tumors, and possibly other cancers (Pearce et al., 2012). However, genomic analysis of children born to people exposed to radiation at Chernobyl indicates that this exposure did not lead to an increase in new genetic changes passed from parent to child (Yeager et al., 2021).

Several other environmental exposures have also been associated with childhood cancer; however, it is difficult to draw firm conclusions because of challenges in studying these exposures. For some types of childhood leukemia (particularly acute lymphoblastic leukemia), researchers have identified associations with paternal tobacco smoking (Liu, 2011, Cao, 2020); exposure to certain pesticides used in and around the home (Bailey et al., 2015) or by parents at their workplaces (Van Maele-Fabry, 2010, Vinson, 2011); use of solvents, organic chemicals found in some household products; and outdoor air pollution (NCI, 2021).

Investigations of childhood brain tumors and leukemia and lymphomas have studied associations with exposures to pesticides in and around the home. A meta-analysis of 277 studies found an increased risk of leukemia and lymphomas in children exposed to indoor residential pesticides. A significant increase in the odds of leukemia was also associated with herbicide exposure. Also observed was a positive but not statistically significant association between childhood home pesticide or herbicide exposure and childhood brain tumors. (Chen et al., 2015). Johnson et al, 2014 reported an association of maternal consumption of cured meats and childhood brain tumors. A recent study (Lombardi et al, 2021) used the California cancer registry to identify childhood cases of brain tumors and linked residence to agricultural pesticide exposure. They noted a significant increased risk of CNS tumors and proximity to residences.

Researchers have also identified factors that may be associated with reduced risk of childhood cancer (NCI, 2021). For example, maternal consumption of folate has been associated with reduced risks of both leukemia and brain tumors in children (Chiavarini, 2018). Also, being breastfed and having been exposed to routine childhood infections are both associated with a lowered risk of developing childhood leukemia (Amitay, 2015).

Previous Hydraulic Fracturing and Childhood Cancer Studies

Three studies have been published that examined a possible association between hydraulic fracturing and the risk of childhood cancer. The study populations and main findings are briefly summarized in **Table 1.** Below are more details for each of these three studies.

Fryzek et al. (2013) were the first to investigate a potential relationship between childhood cancer and hydraulic fracturing in Pennsylvania. The study compared cancer incidence rates at the county level before and after hydraulic fracturing to determine if rates increased. The study did not find a significant increase in the incidence of total cancers or leukemia. It did find a slightly elevated incidence rate for central nervous tumors after drilling began. The ecological study design employed has major limitations due to a lack of individual level data. Further studies were required to draw solid conclusions about the relationship between hydraulic fracturing and childhood cancer.

Table 1: Comparison of Previous HF and Peer-Reviewed Childhood Cancer Studies

	Fryzek et al., 2013	**McKenzie et al., 2017**	**Clark et al. 2022**
Study area	Pennsylvania	Rural Colorado	Pennsylvania
Time period	1990-2009 (stopped data collection 2 years after hydraulic fracturing began - latency issues)	2001-2013	2009-2017
Study population size/design	Standardized incidence rates by county for cases of CNS and leukemia, age 0-20 (N =1,874)	Case-control: aged 0-24, Final sample: 87 ALL, 50 lymphoma and 528 controls diagnosed with non-hematologic cancer sample	Case-control study, N=405 cases of ALL and 2,080 controls
Data source	PA Cancer Registry, US Census Bureau	Colorado Central Cancer Registry	PA Cancer Registry, PA Vital Records (Bureau of Health Statistics and Registries)
Exposure metrics	Compared SIRs before and after drilling using spud dates (date drilling operations begin)	Inverse distance weighted oil and gas well counts within a 16.1 km radius of the residence at time of diagnosis	Inverse distance-squared weighted well counts with buffer sizes 2, 5, and 10 km from birth address for the association between residential proximity to UNGD and ALL in primary exposure and perinatal window
Outcome	Childhood cancer, childhood leukemia, and CNS tumors	ALL and NHL	ALL
Results	1. The observed number of childhood cancers both before and after drilling were as expected (based on SEER cancer incidence rates) 2. **No evidence** that persons living in counties with HF experienced higher childhood cancer rates overall or for childhood leukemia	1. Children aged 0-24 years diagnosed with NHL were **no more likely** to live in areas with active oil and gas development than children diagnosed with non-hematologic cancer 2. Children aged 5-24 years diagnosed with ALL were **more likely** than children diagnosed with non-hematologic cancer to live within 16.1-km of an active oil and gas well	1. Children with at least one UNGD well within 2 km of their birth residence during the primary window had *1.98 (95% CI: 1.06, 3.69) times* the odds of developing ALL in comparison with those with no UNGD wells 2. Children with at least one vs. no UNGD wells within 2 km during the perinatal window had *2.80 (95% CI: 1.11, 7.05) times the odds of developing* ALL

Two case-control studies have been published in the US involving individual data on childhood cancer risk and hydraulic fracturing. The first was conducted between 2001-2013 in Colorado by McKenzie et al. (2017); and the other was conducted between 2009-2017 in Pennsylvania by Clark et al. (2022).

McKenzie et al. (2017) conducted a case-control study in rural Colorado and included participants who were 0-24 years old and diagnosed with cancer between 2001-2013. For each child, they estimated exposure to hydraulic fracturing activity by calculating the distance between the participants' residences and oil and gas wells within a ten-mile radius. Exposure metrics accounted for both the density and proximity of wells to the child. The logistic regression utilized adjusted for age, race, gender, income, and elevation.

Children aged 0-24 with acute lymphoblastic leukemia (ALL) were more likely to live in areas with active wells. For ages 5-24, ALL cases were 4.3 times as likely to be in the highest exposure category. Further adjustment for year of diagnosis increased the association. The study's limitations included the use of non-hematologic cancer cases as a control group, the substantial number of cancer cases that could not be geocoded (28%), and the sole use of residence at cancer diagnosis to calculate exposure, which is not static and can result in misclassification bias.

A more recent case-control study was reported by Clark et al. (2022), which included 405 children aged 2-7 diagnosed with ALL in Pennsylvania between 2009–2017, and 2,080 controls matched on birth year. They calculated a similar exposure metric to the McKenzie study (2017) but used different distance cutoffs to better understand how distance affects exposure levels. They investigated two time-based exposure windows: a primary window (3 months preconception to 1 year prior to diagnosis/index date) and a perinatal window (3 months preconception to birth).

Clark et al. used logistic regression to estimate odds ratios (ORs) and 95% confidence intervals (Cis) for the association between residential proximity to UNGD and ALL in two exposure windows. Children with at least one UNGD well within 2 km (1.2 mile) of their birth residence during the primary window had 1.98 times the odds of developing ALL in comparison with those with no UNGD wells (95% CI: 1.06, 3.69). This result was only based on 7 cases. After adjusting for maternal race and other potential confounders, the OR was no longer statistically significant (OR=1.74, 95% CI: 0.93, 3.27). Similar ORs were produced by models using the water pathway-specific metric.

A major limitation of the Clark et al. study was that a considerable proportion (93-98%) of the study population had no exposure to any UNGD activity within a 10-mile radius. Regulations in metropolitan areas such as Philadelphia and Pittsburgh, or the lack of shale deposits, prohibit hydraulic fracturing activity in sizable portions of Pennsylvania. High proportions of unexposed participants within the study hindered the investigators' ability to identify associations.

In addition to the three peer-reviewed studies, on February 13, 2019, the Pittsburgh-based TV news channel WPXI aired a story regarding a potential cluster of Ewing sarcoma, also sometimes called the Ewing family of tumors (EFOT), a specific type of bone or soft tissue cancer usually occurring in childhood or adolescence. Subsequently, the PA Department of Health received many calls concerning multiple children in the Canon-McMillan School District in Washington County, reporting that they had been diagnosed with EFOT. Several parents came forward to say that their children were also diagnosed with the same disease.

This prompted a cancer incidence survey reported on April 22, 2019 (PADOH, 2019). The PA Department of Health analyzed cancer registry data in three time periods: 1985−1994, 1995−2004 and 2005−2017. These three time periods were used to assess cancer incidence trends over time. This analysis used the mid-time period census population (1990, 2000, and 2010 census data) for age adjustment. Age-standardized SIRs for various childhood cancer types and their 95% CIs for Washington County and Canon-McMillan School District residents were calculated respectively by gender to determine whether the residents experienced a significant excess of cancer incidence compared to the rest of the Pennsylvania population.

Study results for Canon-McMillan School District and incidence of EFOT indicated that there were no cases reported during the first two time periods before hydraulic fracturing. However, there were three cases reported during the 2005-2017 period, which coincided with hydraulic fracturing. The SIRs of Ewing sarcoma estimated based on this small number of cases were considered unstable and difficult to interpret. Overall, total childhood cancer incidence rates were also calculated, and both female and male childhood cancer rates were not appreciably different from the rest of the Commonwealth during any of the three time periods. Moreover, childhood cancer rates in the school district decreased over the last two time periods. The PADOH, however, stated that it would continue to closely monitor EFOT and pediatric cancer incidence in Pennsylvania over the next several years as new data becomes available through the PA cancer registry.

Community concerns persisted, prompting a supplemental analysis reported in March 2020 in addition to advancing other research studies. The present case-control study was initiated by PA Governor Wolf's administration due to concerns about the Ewing sarcoma cluster and a significant rise in hydraulic fracturing and UNGD drilling in western PA since 2005.

Study Aims and Objectives

This study aims to investigate the risk for childhood cancer related to environmental exposures from UNGD hydraulic fracturing in Southwestern Pennsylvania.

Objectives:

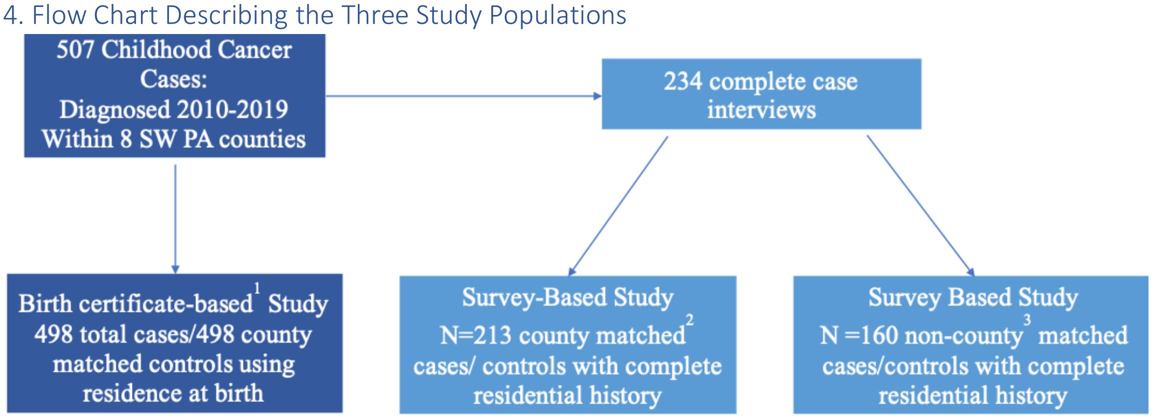
1) We built upon previous studies of exposure to hydraulic fracturing and risk of childhood cancer by conducting a matched case control study using the entire sample of cancer cases identified within the 8-county study area and identifying one randomly selected age, gender, race, and county matched control. Birth records were used to extract information on the mother's and newborn's residence and their characteristics. This birth record-based /cancer registry study enabled comparison with earlier studies conducted by McKenzie (2017) and Clark (2022).

2) An overall UNGD well activity metric was created using each of the individual phases to investigate the childhood cancer risk while controlling for sociodemographic, health history, and behaviors in the year before birth up to the child's cancer diagnosis date.

3) This study also sought to collect more detailed residential histories that can be applied to individual phases and overall UNGD well activity in childhood cancer cases and controls.

Study Design: The study examined three populations derived from the 507 childhood cancer cases diagnosed from 2010-2019 in the eight-county Southwestern Pennsylvania area. The study team completed 234 residency interviews for cases and were able to match 213 of these cases with controls born in the same county, and 160 with controls born in different counties (but still in the eight-county area). Of the total of 507 childhood Cancer Cases, a total of 498 cases were matched to a new group of county-matched controls using only birth certificate data. Nine cases were removed from the full list of cases during data verification.

Figure 4. Flow Chart Describing the Three Study Populations



1. Birth certificate-based means the exposure is based on the mother's residence at birth.
2. County-matched means controls came from the same county as the case.
3. Non-county-matched means controls were chosen at random from the eight-county area.

II. Methods

Study Population

All cases and controls were born in one of the eight counties selected for this study, including Allegheny County (except city of Pittsburgh), Armstrong, Beaver, Butler, Fayette, Greene, Washington, and Westmoreland. Case children were diagnosed with any of four types of malignancies described below and had an address within the defined study area at the time of cancer diagnosis between the years of 2010-2019.

Due to restrictions in hydraulic fracturing within city limits of Pittsburgh, it was necessary to exclude any cases or controls whose parents lived in a zip code located in, or part of, the City of Pittsburgh, as indicated on the birth record or at time of cancer diagnosis. Zip codes excluded from the City of Pittsburgh are shown in **Appendix B.**

Case Inclusion Criteria

All cases of childhood cancer in the present study were identified through the PA Cancer Registry diagnosed from 2010-2019. The cancer types were leukemia, lymphoma, CNS tumors, and malignant bone tumors diagnosed at 0-19 years of age. We extended the age range up to 29 years for malignant bone tumors, including EFOT, to increase sample size due to the rarity of the condition and its later presentation. These specific malignancy types were defined according to the International Classification of Childhood Cancer Recode Third Edition (ICD-O-03/IARC 2017), which is recommended by the NCI Surveillance, Epidemiology, and End Results (SEER) Program. See **Table 2**.

Table 2. Definition of Childhood Cancer Cases for the Case-Control Study in Western PA (International Classification of Childhood Cancer Recode Third Edition, ICD-O-3/IARC 2017)

Cancer type	**ICCC Recode 3rd ICD-O-3/ IARC 2017 morphology codes**	**Behavior codes**	**ICD-O-3 primary site code**
I. Leukemias, Myeloproliferative, and Myelodysplastic Diseases (0-19 years of age)			
1. Precursor cell leukemia	9811-9818, 9837	3	C420, C421, C423, C424, C809
	9835, 9836	3	C000-C809
2.Mature B-cell leukemias	9823	3	C420, C421, C423, C424, C809
	9826, 9832, 9833, 9940	3	C000-C809
3. Mature T-cell and Natural Killer (NK) cell leukemias	9827	3	C420, C421, C423, C424, C809
	9831, 9834, 9948	3	C000-C809
4. Lymphoid leukemia, NOS	9591	3	C420, C421, C423, C424
	9820	3	C000-C809
5. Acute myeloid leukemias	9840, 9861, 9865-9867, 9869-9874, 9891, 9895-9897, 9898, 9910, 9911, 9920, 9930, 9931	3	C000-C809
6. Chronic myeloproliferative diseases	9863, 9875, 9876, 9950, 9960-9964	3	C000-C809
7. Myelodysplastic syndrome and other myeloproliferative diseases	9945, 9946, 9975, 9980, 9982-9987, 9989, 9991, 9992	3	C000-C809
8. Unspecified and other specified leukemias	9800, 9801, 9805-9809, 9860, 9965-9967	3	C000-C809
II. Lymphoma (0-19 years of age)			
1. Precursor cell lymphomas	9727-9729	3	C000-C809
	9811-9818, 9837	3	C000-C419, C422, C440-C779
2. Mature B-cell lymphomas (except Burkitt lymphoma)	9597, 9670, 9671, 9673, 9675, 9678-9680, 9684, 9688-9691, 9695, 9698, 9699, 9712, 9731-9735, 9737, 9738, 9761, 9762, 9764-9766, 9769, 9970, 9971	3	C000-C809
	9823	3	C000-C419, C422, C440-C779
3. Mature T-cell and NK-cell lymphomas	9700-9702, 9705, 9708, 9709, 9714, 9716-9719, 9724-9726, 9767, 9768	3	C000-C809
	9827	3	C000-C419, C422, C440-C779
4. non-Hodgkin lymphomas, NOS	9591	3	C000-C419, C422, C440-C779, C809
	9760	3	C000-C809
5. Burkitt lymphoma	9687	3	C000-C809
6. Miscellaneous lymphoreticular neoplasms	9740-9742, 9750, 9751, 9754-9759	3	C000-C809
7. Unspecified lymphomas	9590, 9596	3	C000-C809

Table 2 Continued. Definition of Childhood Cancer Cases for the Case-Control Study in Western PA (International Classification of Childhood Cancer Recode Third Edition, ICD-O-3/IARC 2017)

Cancer type	ICCC Recode 3rd ICD-O-3 IARC 2017 morphology codes	Behavior codes	ICD-O-3 primary site code
III. CNS and Miscellaneous Intracranial and Intraspinal Neoplasms (0-19 years of age)			
1. Ependymomas and choroid plexus tumor	9383, 9390, 9391-9394, 9396	0-1, 3	C000-C809
2. Astrocytomas	9380	0-1, 3	C723
	9384, 9400-9411, 9420-9424, 9425, 9440-9442	0-1, 3	C000-C809
3. Intracranial and intraspinal embryonal tumors	9470-9478, 9480, 9508	0-1, 3	C000-C809
	9501-9504	0-1, 3	C700-C729
4. Other gliomas	9381, 9382, 9385, 9430, 9431, 9444, 9445, 9450, 9451, 9460	0-1, 3	C000-C809
	9380	0-1, 3	C700-C722, C724-C729, C751, C753
5. Other specified intracranial and intraspinal neoplasms	9840, 9861, 9865-9867, 9869-9874, 9891, 9895-9897, 9898, 9910, 9911, 9920, 9930, 9931	3	C000-C809
	8158, 8290	0-1, 3	C751
6. Unspecified intracranial and intraspinal neoplasms	8000-8005	0-1, 3	C700-C729, C751-C753
IV. Malignant Bone Tumor (0-29 years)			
1. Osteosarcoma	9180–9187, 9191–9195, 9200	3	C400-C419, C760-C768, C809
2. Chondrosarcomas	9210, 9220, 9240	3	C400-C419, C760-C768, C809
	9211-9213, 9221, 9222, 9230, 9241-9243		C000-C809
	9231		C400-C419
3. Ewing tumor and related sarcomas of bone	9260	3	C400-C419, C760-C768, C809
	9365		C000-C809
	9364		C000-C809
4.. Other specified malignant bone tumors	8810, 8811, 8818, 8823, 8830	3	C400-C419
	8812, 9262, 9370-9372, 9270-9275, 9280-9282, 9290, 9300-9302, 9310-9312, 9320-9322, 9330, 9340-9342, 9250, 9261		C000-C809
5. Unspecified malignant bone tumors	8000-8005, 8800, 8801, 8803-8805	3	C400-C419

Exclusion of Ineligible Cases

A total of 593 cancer cases were identified from the PA Cancer Registry between 2010-2019 according to the case eligibility criteria described above. During the data checking and cleaning process, the study team identified the following number of cancer cases were ineligible, and thus were excluded from the final statistical analysis:

- 41 based on the Third Edition ICD-O-3/IARC 2017
- 25 diagnosed within the City of Pittsburgh
- 20 born outside of the eight-county study area.

After these cases were excluded, a total of 507 cancer cases were deemed eligible for the study.

Control Selection

We referenced the birth record registry at PA Bureau of Health Statistics and Registries to select age-, sex- and race-matched controls for either the county-matched or non-county-matched groups. The details of the specific control selection algorithm are provided in Appendix B of this report.

The following steps were followed to obtain a county-matched control:

- A control was selected among children whose mother's residence was recorded on the birth record in the same county as the index case at birth.
- In addition to age, sex, and race, a control without matching on county was selected among children whose mother's residence was within the eight counties of the study area.
- Eligible controls were born within ± 45 days of the index case and were of the same sex and mother's race. For each case, up to 40 county-matched controls and 40 non-county-matched controls were randomly chosen by the PADOH without replacement.
- If the number of eligible controls was fewer than 40 for a given index case, the PA Bureau of Health Statistics and Registries provided information on all eligible controls.
- If a control was matched to multiple cases, a simple random sampling algorithm without replacement was used to determine the matched index case.

We made attempts to locate and update the information of current and past residence history of all cases and 20 of the 40 eligible controls (due to time limitations) through the contact information tracing service Lexis Nexis (described in detail below). Additionally, we used Spokeo, an online tracing service that provides property records, emails, addresses, and phone numbers to confirm residential history and contact information when needed. A unique random number was generated during the control selection process for each of 40 eligible controls per case.

The county-matched control was chosen to help adjust for both urban/rural differences within each county and to assure the greatest similarity of sociodemographic and environmental characteristics to the index cancer case. The non-county match was chosen to limit potential bias from over-matching. The duration of the exposure data collected for the control subject was the same as for the index case, and personal history was obtained up to the index date, which was defined as the date of cancer diagnosis for cases. The same date was applied to matched controls.

Survey

A survey questionnaire was developed based on an ATSDR (Agency for Toxic Substances and Disease Registry) childhood cancer cluster investigation (State of New Jersey Department of Health, 2017) and was modified to include hydraulic fracturing, and industrial and farming activity with an emphasis on residential history. The objective of the survey was to capture the mother's and child's environmental exposure history, residential history, sociodemographic information, health history, and behaviors in the year prior to birth up to the cancer diagnosis date. The survey was then uploaded to a Qualtrics (Provo, UT) software platform. If there were any questions the parent was uncomfortable addressing, they could decline to answer at any time. See **Appendix D**.

As will be described below, the initial response rate from the PADOH recruitment brochure was low (20%) and it was determined that the at least 45 minutes needed to answer the survey questions was negatively affecting the response rate. It became necessary to shorten the questionnaire into a more user-friendly online version, which could be taken at any time. The revised survey included many of the same sections but included fewer questions. See **Table 3**.

Table 3. Main Sections of Case-control Survey

1. Parental background and demographics	5. Maternal reproductive history
2. Residential history, home characteristics, and environmental risk factors for all addresses	6. Maternal medical procedures that occurred during pregnancy with case/control child
3. Occupational and lifestyle histories of the parent(s)	7. Child's medical procedure and infection history
4. Familial cancer history	8. Optional questions regarding household income, interest in future studies, opportunity to share any additional relevant information

The shortened survey is included in **Appendix D**. The longer survey is available upon request.

Overview of Recruitment and Enrollment Process

The Institutional Review Board (IRB)/consent application for this study (protocol number 21020141) was approved by the University of Pittsburgh IRB on March 16, 2021. The PADOH-specific IRB application was approved on June 17, 2021. The University of Pittsburgh applied for and was granted access to protected health information in a data sharing agreement from the PADOH on April 19, and July 7, 2021, respectively. Parents of case and control children, not the children themselves, were asked to participate in the study. The information collected included residence of the mother, and both parents' occupation and health behaviors, including the pregnancy period and early years of the child's life. There was no assent process for children under 18. IRB materials, the timeline of study events, and outreach and recruitment materials are included in **Appendix C**.

PADOH leadership strongly recommended a government-approved third-party tracing agency, LexisNexis, to provide updated and confirmed contact information for recruitment mailings, phone calls, text messages, and emails. The LexisNexis contract was finalized in August 2021, and updated contact information was provided in September 2021, prior to the dissemination of the first round of case recruitment mailings. The initial case dataset was received from the PADOH in September 2021, with the decedent cases received in April 2022.

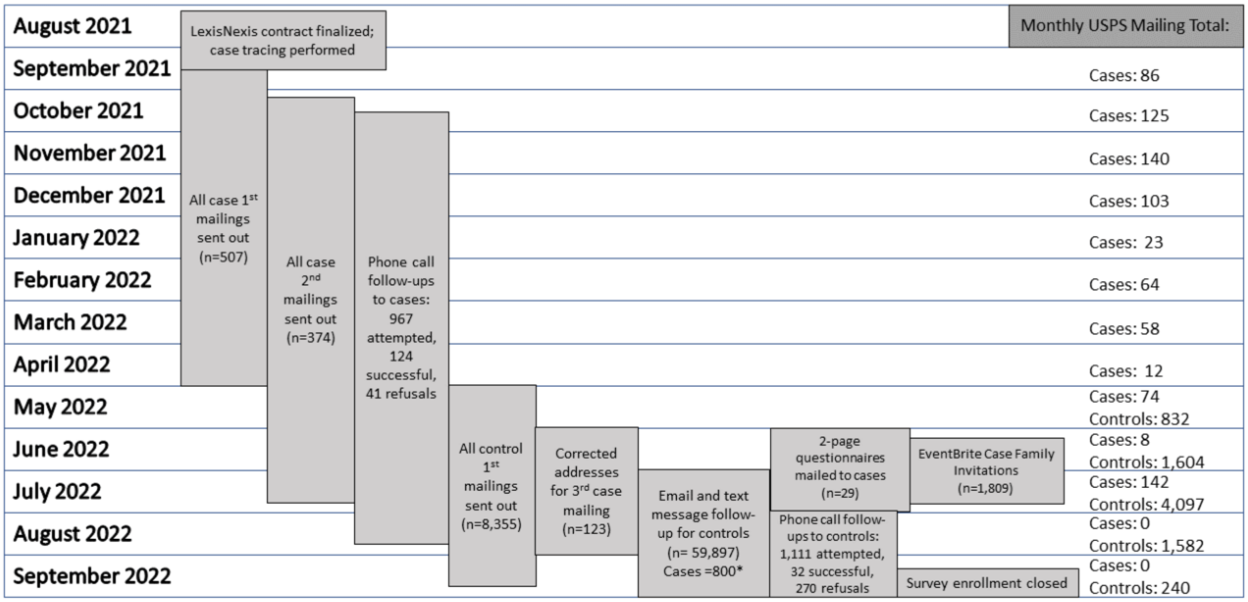
The initial case recruitment protocol, beginning in late September 2021, included a letter from the PADOH Secretary of Health inviting families to schedule a 45–60-minute telephone interview, a brochure explaining the study, and an opt-in/opt-out card with a pre-addressed return envelope. The study team's strategy was to prioritize case recruitment given the need for a sample of controls matched on age, race, gender, and county. Participants who did not respond were sent an additional letter.

Telephone interviewers attempted to contact all parents who opted in using a computer–assisted telephone interviewing (CATI) system to manage sample and call attempts. The CATI system was linked to a Qualtrics-based survey which interviewers used to administer the survey instrument. The PADOH protected-access protocol mandated that only one phone call be made to request participation after receipt of the two recruitment mailings.

Due to concern about the initial low response rate (<20%) after the two letters were sent and follow-up calls were made, the study team initiated a briefer questionnaire that included an online 20–25-minute interview facilitated by co-investigator Dr. Todd Bear and the Population Survey Facility in Pitt School of Medicine in March 2022. In addition, in May 2022 the survey team initiated a shortened two-page residential questionnaire that captured a complete residential history. See **Figure 5** for a timeline of recruitment efforts.

To augment the study response rate and enhance communication with families, the study team solicited support from Dr. Jean Tersak, of UPMC Children's Hospital of Pittsburgh, who provided a letter of support for the study which was subsequently included in all study recruitment mailings. Dr. Tersak was added as a study co-investigator in June 2022.

Figure 5. Timeline of Recruitment Efforts for Cases and Controls



*in follow-up of phone calls

In summer 2022, the study team worked with community nurses and supervisors at state health centers in Washington and Westmoreland counties to facilitate in-person informational sessions at respective health centers in Washington and Greensburg. The goal of these planned sessions was to make the study team available to answer any questions the invited case families may have had regarding the study and their invitation to participate, as well as to facilitate their participation. The study team utilized the email addresses provided by LexisNexis (up to three addresses per parent, a maximum of six addresses per family) to send e-vites to these events, with RSVP capabilities provided through Eventbrite.

The study team sent 1,809 invitations to unique email addresses, of which 415 emails were found to be undeliverable or incorrect; 1,394 were successfully delivered. While 258 recipients clicked the link to the Eventbrite page, no confirmed responses were received for the events. One case family contacted the study team through the publicly available study email address to posit a question about the events, but no families expressed interest in attending the information sessions or completing the online survey. The lack of interest in attending these events was most likely due to remaining COVID school closures and protocols.

Control families were sent an initial mailing between May-September 2022. The study team was permitted to pivot to electronic methods of contact for the second mailing, and emails were sent September 8-22. Priority was given to contacting matched controls of the cases who had already completed an interview. Once a control for each case and each group had participated, and the survey was deemed eligible (completing the residential history at a minimum), no more controls for that case were contacted. Only a few matched controls were contacted at a time to reduce the number of duplicate controls, and to minimize extraneous recruitment outreach efforts.

Control enrollment was closed on September 27, 2022, to allow the study team sufficient time to clean, analyze, and summarize the data. 8,355 initial recruitment letters were mailed to control families between May-September 2022 and 48,298 reminder letters were sent as emails. Telephone interviewers were given case records of anyone who had not responded to previous mail invitations. These individuals were contacted a maximum of five times in seven days. See **Appendix B** for a summary of activities for recruitment of controls.

Incentives

Incentives were provided for all participants who did not refuse payment. The study team used two University of Pittsburgh-approved incentive programs. Initially, the Vincent Card program was used, which involved sending a payment card loaded with a specified amount of money to the participant after the survey. The participant then called the university, reaching a member of the study team who would activate their card. Participants were followed-up if they did not call to activate their card. A new program, called the Tango Card System, was implemented halfway through the recruitment process to simplify the process and to be more conducive to the new online method of completing the survey independently.

The Tango Card system involved the participants entering an email address at the end of the survey. Upon the survey's completion in the Qualtrics software platform, a link was automatically sent to their provided email address, giving the participants access to a site where a variety of gift cards could be selected. Email addresses could not be used multiple times to receive additional payments. Cases were provided $25 compensation, and controls were provided $15. The decrease in incentive for controls was due to the shortening of the survey, which preceded control participation. Case participants who took the shorter survey had their incentives kept at $25 to align with initial communications about the study. 804 participants completed the study, with 731 accepting and receiving paid incentives.

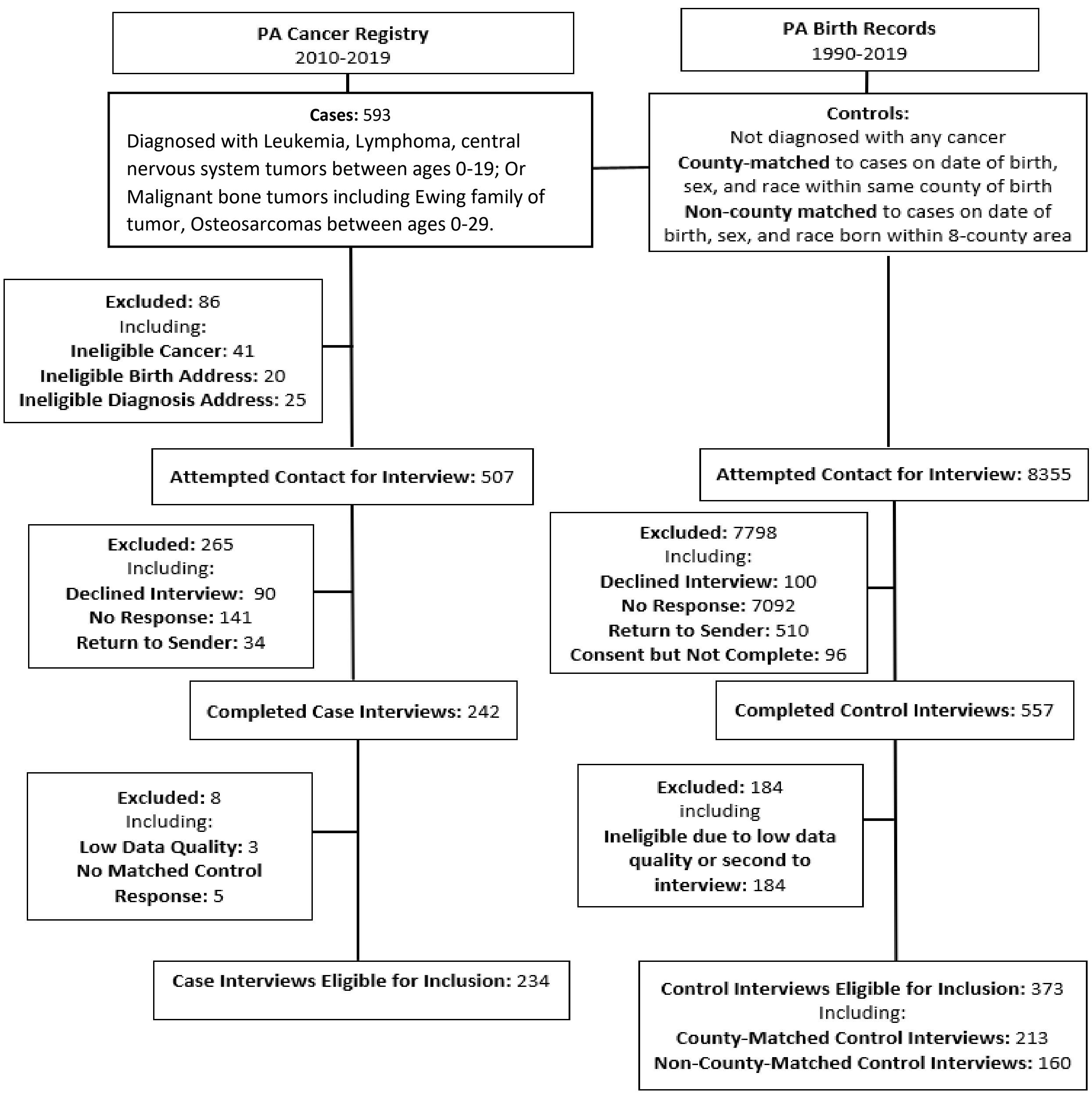
Final Enrollment Numbers

A total of 593 cancer cases were originally identified by the study team. A shift to the use of the ICD-O-3/IARC 2017 coding from an earlier version was recommended by PADOH, leading the study team to reclassify 41 eligible cases to ineligible. Of the 507 remaining eligible cases which the study team attempted to contact, 265 were excluded because 90 refused to participate,141 did not respond to contact attempts, and 34 mailings were "return to sender." An additional 8 cases were excluded from post-data collection; 5 cases were unmatched to a control, and 3 cases were excluded due to low data quality. These exclusions resulted in 234 eligible case interviews.

The research team attempted to contact 8,355 controls, with a priority for interviews with controls whose matched case had already been interviewed. Multiple potential controls for each case were contacted, with the first control who had an eligible response used as the match. 7,798 controls were excluded during recruitment: 7,092 did not respond, 510 were unable to be traced after the letter returned as return to sender, 100 declined interviews, and 96 consented to participate but did not complete the survey. 557 controls were interviewed, but 184 either had low quality data or were second responses for cases who already had a matched control interview completed for that group (county-matched or non-county-matched). 373 controls were included in the analysis. See **Figure 6** for the final enrollment diagram of the case-control study.

Of the 234 eligible case interviews, 147 cases had both county-match and non-county-match controls. A total of 13 cases only had a non-county-matched control and 66 cases only had a county-matched control. After excluding those who refused and the study team was unable to contact, the cooperation rate was 63%.

Figure 6. Enrollment Diagram: Childhood Cancer Case-Control Study



Exposure Measures

UNGD Activity Overview

The primary exposure measure for this study was an inverse distance-weighted index of UNGD activity within 5 miles of parent and child residence. The study team also considered additional buffers: 0.5, 1, and 2 miles. There were four phases of UNGD, including well pad preparation, drilling, hydraulic fracturing, and production, which varied in duration and exposures to potential carcinogens. Therefore, the UNGD activity metric was calculated separately for each of the four phases, for each study subject. Additionally, the study team created an overall activity metric structured the same way as the phase specific metrics, but the duration of activity spanned from the start date of well pad construction until the end of the production phase for each relevant well. Due to the way the phase metrics were structured, the overall activity metric was also equivalent to the sum of the 4-phase metrics. Lastly, the study team calculated well count and inverse distance weighting (IDW) well count to measure the density of and proximity to well sites without integrating duration of exposure. These two metrics were used to align with previous studies.

For wells located in Pennsylvania, data required to calculate the UNGD activity metric were obtained from the Pennsylvania Department of Environmental Protection and the Pennsylvania Department of Conservation and Natural Resources. For wells in Ohio and West Virginia, data were obtained from the Ohio Department of Natural Resources and the West Virginia Department of Environmental Protection, respectively. Due to the difference in the reported data in Ohio and West Virginia (provided annually, rather than daily), the study team was unable to incorporate these data into analyses. Although the analyses focus on residences within the bounds of the eight-county study, the study team had to account for residences located on the geopolitical borders of the study region. To account for this, buffer regions that extended five miles into adjacent counties were included and exposure data within these buffer regions were captured. UNGD phase descriptions are below:

1. **Well pad preparation** – the process of preparing a site where one or more wells were located. It is defined as the period beginning 30 days before the first well on the pad is spudded and ending when the first well is spudded.
2. **Drilling** – the creation of the wellbore. This phase begins on the well's spud date and ends on the drilling completion date; the median for the wells was 104 days.
3. **Hydraulic fracturing** – the process of injecting large volumes of water at high pressure into the wellbore to fracture the shale layer. This period is defined as beginning on the stimulation commencement date and ending on the stimulation completion date. Hydraulic fracturing may be repeated over time for a given well. The median for the wells was 12 days.
4. **Production** – the process of collecting natural gas or oil that—following hydraulic fracturing—travels through the wellbore to the surface. Production durations are variable. A well was defined as being in production for reporting periods when production was indicated and reported production volume was non-zero. The minimum amount of time in the production phase was 30 days (as per how the data were reported). The maximum number of days was 8,769 days. The mean number of days was 2,239 and the median was 2,193 days. An individual well could have had multiple production periods with gaps in which the well was inactive. Calculations include all production period durations but not the gaps between them.

UNGD Exposure Metrics Calculation

Inverse distance weighting (IDW) is a metric used to account for both the proximity and density of wells within a designated buffer distance from a participant's residence. It is a commonly used metric in environmental epidemiological studies. The metric includes a numerator value which is typically 1 but can also take on other quantifying values, such as daily volume of gas production or well depth, adding further information to the metric. The denominator is a measure of distance, typically the distance measured squared. Then these individual fractions are summed across all wells located within a designated buffer distance. See **Figure 7**.

In previous studies, a well was included in the IDW metric if it was both within the designated buffer and there was at least one day of overlap between the well's activity and the participant's study period of interest. This kind of metric did not account for the duration of overlap. For example, two wells that were equidistant from a participant's residence would have made the same contribution to their exposure metric, even if one well was active for one day, whereas the other for one year during the participant's study period. The study team created this metric because it was commonly used in existing literature. To account for duration of exposure, the study team also created an overall activity metric that integrated both the distance and duration of every active well.

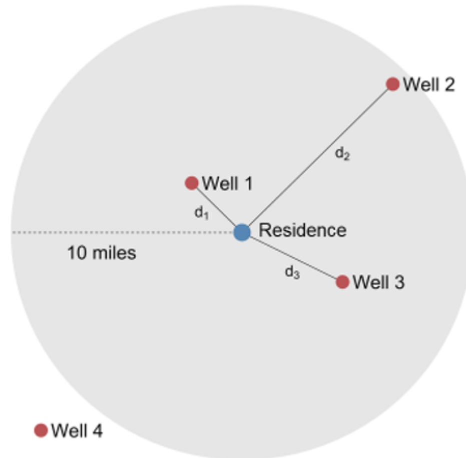
To include a duration element, the numerator for the IDW overall activity metric, as well as the well pad construction, drilling metrics were the sum of days of activity overlap, over the distance squared of each well. This number was summed across all wells within the designated buffer distance. The numerator for IDW hydraulic fracturing and production metrics was well depth in meters and daily average volume of gas production in cubic meters (m^3), respectively, summed over the days of overlap between each respective phase and the participant's study period, then summed across all wells within the designated buffer distance. These two metrics were calculated with additional information to examine how well depth and gas production volume contributed to exposure metric for a given participant.

An IDW overall activity metric and well count metric was calculated as the primary exposure variables. Additionally, 4 IDW metrics corresponding to each phase were calculated as secondary exposure variables. An additional metric of well count (without the use of IDW) was calculated. While examining each phase alone may introduce some issues because many individuals can be exposed to more than one phase simultaneously, the analysis can still contribute to the study's overall conclusions. These 7 metrics were calculated for each residence of the case or control subject. Because each participant could move multiple times during the period of exposure, these metrics were first calculated by residence and then aggregated to create one metric per participant. Further description about how metrics were aggregated provided in the Data Processing section.

Figure 7. Inverse Distance Weighting Example

Inverse distance weighting is a method for calculating exposure to nearby locations of interest, such as UNGD wells.

- The resulting IDW metric not only takes into account the number of wells nearby a residence, but also how close the wells are.
- Wells located close to the residence (like Well #1) contribute more to the IDW metric value, while wells farther away (like Well #2) contribute less.
- Often a buffer distance is used as a boundary, beyond which a well (like Well #4) no longer contributes to the IDW metric value.



Definition of Time Periods

A participant's study period of interest included two time periods. Pregnancy (exposure time window 1, or T1) was defined as conception through date of birth. Date of conception was calculated by subtracting gestational age (in weeks) from the date of birth. Total exposure (exposure time window 2, or T2) was defined as date of birth through the index date, which was date of cancer diagnosis for cases. The same date was applied to controls so the period for both cases and controls was identical.

UNGD activities for a given well had 4 phases as described previously. The duration of each phase was defined in **Table 4**. Each of the data was found, or calculated, using datasets from the Pennsylvania Department of Environmental Protection and the Pennsylvania Department of Conservation and Natural Resources. If a phase for well or well pad overlapped with the case's study exposure time windows T1 and/or T2, all or in part, the overlapping portion of that phase contributed to the calculation of the activity metric for that individual case. See **Tables 5a** and **5b** for the equations of these metrics with an explanation of each term.

Table 4. Definition of UNGD Activity Metric Phase Durations

Metric	Variable name	Definition of Duration
1	Overall Activity	Production period end date minus start date of the well pad preparation variable minus (if applicable) periods of inactivity between production periods
2	IDW Well Count	Numerator was 1 if there were any days overlap between spud date until the most recent production period end date (wells can have multiple production periods), and the participant's exposure period
3	Well Count	Count of 1 if there were any days overlap between spud date until the most recent production period end date (wells can have multiple production periods over time), and the participant's exposure period
4	Well Pad Preparation	Spud date minus 30 days
5	Drilling	Stimulation commencement date minus spud date +1 day
6	Hydraulic Fracturing	Stimulation completion date minus the commencement date + 1 day
7	Production	Production period end date minus production period start date
*Spud date is a fracking industry term meaning the first day of drilling.		

Table 5a. Definition of Primary UNGD Activity Metrics

Metric	Variable Name	Calculation of phase-specific activity metric
1	Overall Activity	$\text{Overall well activity for maternal residence } j = \sum_{i=1}^{n}\sum_{k=1}^{l}\frac{I_A(K)}{d_{ij}^2}$ Where: • *n* was the number of wells within 0.5, 1, 2, or 5 miles of maternal residence *j* • *k* was equal to the date of the beginning of conception and *l* the date of birth (for T1), or *k was* equal to date of birth and *l* the index date (for T2) • $I_A(K)$ was equal to 1 when $d_{ij} \le 0.5$, 1, 2, or 5 miles, respectively, and the overall activity (from well pad construction to the end of production not including any inactive periods of production for a given well) overlapped with the defined exposure time window (T1 or T2), or equal to 0 otherwise • d^2_{ij} was the squared distance (m^2) between well *i* and maternal residence *j*
2	Well Count IDW	$\text{IDW well count for maternal residence } j = \sum_{i=1}^{n}\frac{I_A(K)}{d_{ij}^2}$ Where: • *n* was the number of wells within 0.5, 1, 2, or 5 miles of maternal residence *j* • *k* was equal to the date of the beginning of gestation and *l* the date of birth, or *k* was equal to date of birth and *l* the index date for maternal residence *j* • $I_A(K)$ was equal to 1 when $d_{ij} \le 0.5$, 1, 2, or 5 miles, respectively, and the activity of a well (between spud date and the end date of the last production period) overlapped with the defined exposure time window (T1 or T2), or equal to 0 otherwise • d^2_{ij} was the squared distance (m^2) between well *i* and maternal residence *j*
3	Well Count* *(Results for this metric presented in **Supplement**)	$\text{Well count metric for maternal residence } j = \sum_{i=1}^{n}\sum_{k=1}^{l} I_A(K)$ Where: • *n* was the number of wells within 0.5, 1, 2, or 5 miles of maternal residence *j* • *k* was equal to the date of the beginning of gestation and *l* the date of birth, or *k* was equal to date of birth and *l* the index date for maternal residence *j* • $I_A(K)$ was equal to 1 when $d_{ij} \le 0.5$, 1, 2, or 5 miles, respectively, and the activity of a well (between spud date and the last production period end date) overlapped with the defined exposure time window (T1 or T2), or equal to 0 otherwise

Table 5b. Definition of secondary phase specific UNGD activity metrics

Phase	Phase name	Calculation of phase-specific activity metric
4	Well pad preparation	$$\text{Phase 1 metric for maternal residence j} = \sum_{i=1}^{n}\sum_{k=1}^{l}\frac{I_A(K)}{d_{ij}^2}$$ Where: • *n* was the number of well pads within 0.5, 1, 2, or 5 miles of maternal residence *j* • *k* was equal to the date of the beginning of gestation and *l* the date of birth (T1), or *k* was equal to date of birth and *l* the index date (T2) • $I_A(K)$ was equal to 1 when $d_{ij} \leq$ 0.5, 1, 2 or 5 miles, respectively, and the phase overlapped with the defined exposure time window (T1 or T2), or equal to 0 otherwise • d^2_{ij} was the squared distance (m^2) between well pad *i* and maternal residence *j*
5	Drilling	$$\text{Phase 2 metric for maternal residence j} = \sum_{i=1}^{n}\sum_{k=1}^{l}\frac{I_A(K)}{d_{ij}^2}$$ Where: • *n* was the number of wells within 0.5, 1, 2, 5 miles of maternal residence *j* • *k* was equal to the date of the beginning of gestation and *l* the date of birth (T1), or *k* was equal to date of birth and *l* the index date (T2) • $I_A(K)$ was equal to 1 when $d_{ij} \leq$ 0.5, 1, 2, or 5 miles, respectively, and the phase overlapped with the defined exposure time window (T1 or T2), or equal to 0 otherwise • d^2_{ij} was the squared distance (m^2) between well *i* and maternal residence *j*
6	Hydraulic fracturing	$$\text{Phase 3 metric for maternal residence j} = \sum_{i=1}^{n}\sum_{k=1}^{l}\frac{w_i \times I_A(K)}{d_{ij}^2}$$ Where: • *n* was the number of wells within 0.5, 1, 2, or 5 miles of maternal residence *j* • *k* was equal to the date of the beginning of gestation and *l* the date of birth (T1), or *k* was equal to date of birth and *l* the index date (T2) • w_i was the depth in meters of well *i* • $I_A(K)$ was equal to 1 when $d_{ij} \leq$ 0.5, 1, 2, or 5 miles, respectively, and the phase overlapped with the defined exposure time window (T1 or T2), or equal to 0 otherwise • d^2_{ij} was the squared distance (m^2) between well *i* and maternal residence *j*
7	Production	$$\text{Phase 4 metric for maternal residence j} = \sum_{i=1}^{n}\sum_{k=1}^{l}\frac{v_i \times I_A(K)}{d_{ij}^2}$$ Where: • *n* was the number of wells within 0.5, 1, 2, or 5 miles of maternal residence *j* • *k* was equal to the date of the beginning of gestation and *l* the date of birth (T1), or *k* was equal to date of birth and *l* the index date (T2) • v_i was the daily average produced gas volume (m^3) of well *i* , which was calculated as the reported produced gas volume during the reporting period divided by the number of days the well was actively producing during that reporting period. • $I_A(K)$ was equal to 1 when $d_{ij} \leq$ 0.5, 1, 2, or 5 miles, respectively, and the phase overlapped with the defined exposure time window (T1 or T2), or equal to 0 otherwise • d^2_{ij} was the squared distance (m^2) between well *i* and maternal residence *j*

Calculating IDW Metrics

Addresses were geocoded using ArcMap 10.6 to calculate distances between the wells and residences. Distances were calculated between every residence and well within the study area in MySQL server. Once distances were calculated, data was filtered to include only those that were closer than, or equal to, each respective buffer distance 0.5, 1, 2, and 5 miles. Unexposed individuals were those who had never lived within 5 miles of any UNGD site. Time spent in each residence was truncated for each person to ensure that the dates were within the study periods of interest for each person (T1 – conception to birth, and T2 – birth to the diagnosis/index date). Subsequently, the days that overlapped between time spent in each residence and well activity was calculated. For the hydraulic fracturing and production metrics, the days of overlap were multiplied by well depth and average daily gas volume production, respectively. IDW metrics were built by dividing these numerators by the distance in meters squared for all wells located within each residence's buffer distance. These numbers were then aggregated across all wells for one metric per residence. For those who did not remain consistently within the study area, the study team developed methods to handle lapses in exposure estimation. To aggregate exposure metrics across residences for each case and control, a dataset representing individual participants was used. See **Appendix B** for in-depth descriptions of the geocoding process and methods used to handle incomplete data, as well as calculation methods.

Other UNGD-Related Exposures

Impoundment Ponds

Impoundment ponds store water and other fluids from the hydraulic fracturing process. Using SkyTruth, a nonprofit that uses satellite imagery to identify the locations of possible environmental exposure sites, locations and proximity measures were located and created using the same process described above.

Compressor Stations

Compressor stations are facilities where natural gas is received, repressurized, and sent back out in pipelines. Compressor station data was obtained from the PADEP. Their database was used to identify locations of compressor stations and create inverse distance-weighted proximity measures described above.

Waste Facilities

Waste facilities store waste from the hydraulic fracturing process. Waste facility data was obtained from the PADEP. Their database was used to identify locations of waste facilities and create inverse distance-weighted proximity measures described above.

Other Environmental Exposures

In addition to the UNGD activity metrics, the study team also considered additional sources of environmental exposures in the study area during the study period. These included additional components of oil and gas-related activity (e.g., impoundment ponds, compressor stations, waste disposal facilities), other industrial activities (e.g., toxic release inventory sites), and water source measures. Inverse distance-weighting and other modeling approaches were used, as appropriate, to quantify exposure to these additional sources using the same defined buffer zones.

The study team utilized the following environmental exposures including Uranium Mill Tailing Remedial Action (UMTRA) sites, Toxic Release Inventory (TRI) sites, and Superfund sites. The exposure variables created for UMTRA, TRI, Superfund sites were IDW metrics where the numerator was 1 and denominator was the distance in meters squared summed across each respective site. There was no duration component included. The same buffer distances for UNGD activity metrics were considered. The water source variable was a dichotomous variable with public or private source of water. Below are detailed descriptions of these environmental exposures.

UMTRA Sites

There were four UMTRA sites in the study area. Mill tailings are defined as the sandy waste material from a conventional uranium mill. Milling is the first step in making fuel for nuclear reactors from natural uranium ore. UMTRA sites are areas designated by the US Department of Energy who monitor the clean-up of these mills and prevent further contamination of ground water. The IDW was calculated for the four sites in the study area, as well as the eleven sites outside of Pennsylvania, in case the participants' residential history included areas near those sites.

TRI Sites

Facilities in the United States must report toxic chemical releases to the EPA through the TRI program. For the present analysis, the study team downloaded the 2015 data on all TRI inventory sites for the eight-county study area and all surrounding counties. The year 2015 was chosen as a representative time-point based on the midpoint of the diagnosis time (i.e., 2010 -2019) of cancer cases included in the study. For more information on TRI, visit https://www.epa.gov/toxics-release-inventory-tri-program.

Superfund Sites

Superfund is an environmental remediation program established by the EPA. The program is designed to investigate, and clean-up sites contaminated with hazardous substances and include seven EPA PA sites within the eight-country area, and several sites within the study area.

Other Covariates

In the present analysis, in addition to matching factors on age, sex, race, and county of residence between cases and controls, the following set of variables were considered as potential confounders derived from birth records. These covariates are included in all of the logistic regression models.

1. Maternal age at childbirth
2. Maternal education level (a measure of socioeconomical status)
3. Maternal smoking status (any time during pregnancy) reported at childbirth
4. Gestational age in weeks at birth
5. Birth weight of the study subject

Definition of Exposed and Unexposed

IDW metrics are commonly summarized into levels of exposure for increased ability to meaningfully interpret results. Means and standard deviations (SDs), and medians and inter-quartile values were calculated for each of 7 UNGD activities metric for T1 and T2 time periods for all buffer distances. The distributions of all UNGD activity metrics were used to determine dichotomous exposure or exposure by tertiles or quartiles. Cut points in these variables (between exposed and unexposed or between levels of exposure) are set specifically to increase the contrast.

Few participants in any one level of exposure may yield unstable risk estimates with wide 95% CIs. Beyond this practice, there is currently no agreement in the literature on the best way to summarize IDW variables. The study team chose to display results for several distinct kinds of summary variables where appropriate to see how results may have shifted between options. Four different summary variables were provided for all IDW metrics when there were appropriate numbers of participants within exposure levels as described below:

1. **Dichotomous Exposure** – This variable takes on values of either an exposed or unexposed category. The exposed category was defined for individuals who had any history of residence that was located within 5 miles of any UNGD activity, whereas unexposed category was those who did not have a history of residence within 5 miles of UNGD activity. The unexposed group was used for all analyses for different UNGD-derived metrics described below.
2. **Exposure levels within 5-mile or 2-mile buffer zone** – Exposed individuals were further divided by level of cumulative exposure to UNGD activities over time within the defined buffer zone. The median value among the control group was used to classify individuals into high or low category— tertiles classified individuals into the lowest, middle, and highest-thirds of exposure, and the quartiles classified individuals into the lowest, middle-low, middle-high, and highest-quarters of exposure. In the risk modeling, the unexposed group (defined above) was always used as the reference group.
3. **Proximity measure of UNGD activity** – The proximity measure (i.e., buffer zone) was defined as the shortest distance from a residence to any UNGD activity. Conventional cut-off values [0-0.5], (0.5-1], (1-2] and (2-5] miles were used when appropriate. The reference group consisted of individuals who did not have any wells within 5 miles as defined above. When there were too few subjects in each category, the cut points were set as [0-2], and (2-5]. A square bracket indicates that the value was included within the bound, whereas a parenthesis indicates the value was not included within the bound.
4. **Standardized exposure using phase specific z-score values** – IDW metrics for each phase (well pad construction, drilling, hydraulic fracturing, and production) were calculated and standardized by the standard deviation (i.e. the z-score). The phase-specific z-scores were summed using the following formula: $\sum_{ij}^{k} \frac{x_{ij}-\mu_{.j}}{\sigma_{.j}}$, where i is for subject; j, specific phases of UNGD activities (k=4); x, individual measurement of phase-specific UNGD activity; μ, mean; and σ, standard deviation. The summed z-score was another measure of total UNGD activities per individual exposure. The z-score was unitless and accounted for different values and units of all phase-specific UNGD activities.

Statistical Analysis

Primary Strategy

Descriptive statistics were computed and assessed for all outcome and exposure measures, covariates, and characteristics of the study participants. For continuous variables, mean/standard deviation and median/inter quartile range were used; for categorical variables, frequency/percentiles were used. These variables were estimated for the total population and for the birth record-based and survey-based populations separately and stratified by case-control status and various covariates. Chi-square testing was used to compare differences in percentages for social/demographic and maternal characteristics between groups (e.g., cases vs. controls) when categorical; t-tests were used to evaluate differences in means between groups when continuous. When appropriate, nonparametric tests were used.

The study's main aim was to examine the link between UNGD activity and childhood cancer. As such, logistic regression modeling was used to assess this relationship. To preserve the matched study design, conditional logistic regression modeling was done whenever possible. However, some analyses were performed using an unconditional model including the matching variables as covariates.

Separate conditional logistic regression models were used to estimate ORs and the 95% CIs for all four types of cancer combined (i.e., leukemia, lymphoma, CNS tumors, and bone cancer) comparing exposed with unexposed, as well as comparing various levels of exposure by buffer zone and/or levels of overall UNGD activity. The regression analyses were performed, with and without adjustment for additional covariates. In addition to the primary exposure (UNGD metrics) variable, the multivariable-adjusted models included the following covariates: maternal age at childbirth (continuous), maternal education level (≤ 8th grade, high school, some college, or college degree or higher), maternal smoking status at childbirth (yes/no), gestational age (continuous in weeks), birthweight (continuous in grams), TRI (delineated as non-exposed or exposed within 5 miles), UMTRA (non-exposed or exposed within 5 miles), as well as for Superfund sites (non-exposed or exposed within 5 miles).

Significance testing was performed for individual ORs, as well for evaluation of linear trend for increasing level of UNGD activities using an ordinal variable (i.e., 0 for non-exposed and 1, 2 and 3 for tertiles or 1, 2, 3, 4 for quartiles) with the risk of disease of interest. Similar logistic models were used for the decreasing buffer zone (non-exposed, 2-5 miles, 1-2 miles, 0.5-1.0 miles, and 0-0.5 miles) with the risk of disease of interest. All ORs in this report are shown with 95% CIs for UNGD activities and other exposure variables with adjustment for additional covariates. These models were used to analyze data for all three study populations (two survey-based and one birth record-based).

Although underpowered, regression modeling was done for each of the four individual cancer types. The study team believed it was important to separately examine them due to their different biological characteristics. For EFOT (n=20), unconditional logistic regression modeling was performed separately from other malignant bone tumor cases by including all controls in both survey- and birth record-based studies with adjustment for matching variables (i.e., age at diagnosis, sex, race/ethnicity, and county of residence).

Primary Study Population: Use of the Birth Record Study

The primary study population for analysis was the 498 cancer cases and their county-matched controls. Information on the mothers' and newborns' residence and characteristics from birth certificates was extracted from both cancer registry and birth certificates. For analyses of all malignancies combined, this samples (i.e., 498 cases and 498 matched controls) has sufficient statistical power (>80%) to detect odds ratio of 1.5 and greater assuming 25% UNGD exposure within the control group; when exposure among controls is 20%, there is high power (>90%) to detect odds ratios of 1.75 and greater. Furthermore, this sample had sufficient power to detect odds ratios of 1.75 and greater when exposure among controls is 10%. **(Table 6A).** For analyses of site-specific cancers, power is shown in **Table 6B-D** can detect odds ratios of 2.0 for leukemia and CNS and 2.25 for lymphoma with 80% power within the exposure ranges shown. Power estimates assume a two-sided test with alpha = 0.05, a value of 0.20 for the correlation of exposure status in the matches. Power estimates were calculated using https://sampsize.sourceforge.net/iface/s3.html#ccp).

Table 6: Estimated Power to Detect a Specified Odds Ratio and Probability of Exposure in the Control Sample: (Based on Sample Size Available for Study)

6A. 498 case control pairs

	Odds Ratio				
Probability of exposure in controls	1.5	1.75	2.0	2.25	2.5
0.05	0.326	0.582	0.796	0.922	0.977
0.10	0.543	0.841	0.966	0.996	1.0
0.15	0.684	0.935	0.993	1.0	1.0
0.20	0.772	0.970	0.998	1.0	1.0
0.25	0.826	0.983	0.999	1.0	1.0

6B. Leukemia 157 case control pairs for the Birth Record Study of 498 Cancer Cases

	Odds Ratio				
Probability of exposure in controls	1.5	1.75	2.0	2.25	2.5
0.05	0.129	0.219	0.327	0.447	0.567
0.10	0.207	0.37	0.546	0.705	0.827
0.15	0.272	0.483	0.683	0.832	0.922
0.20	0.323	0.564	0.765	0.893	0.958
0.25	0.363	0.619	0.814	0.924	0.974

6C. Lymphoma 105 case control pairs for Birth Record Study of 498 Cancer Cases

	Odds Ratio				
Probability of exposure in controls	1.5	1.75	2.0	2.25	2.5
0.05	0.0988	0.157	0.228	0.31	0.398
0.10	0.151	0.2599	0.388	0.521	0.646
0.15	0.195	0.342	0.504	0.655	0.778
0.20	0.2299	0.405	0.584	0.736	0.846
0.25	0.2578	0.451	0.637	0.784	0.883

6D. CNS 193 case control pairs for the Birth Record Study of 498 Cancer Cases

	Odds Ratio				
Probability of exposure in controls	1.5	1.75	2.0	2.25	2.5
0.05	0.15	0.261	0.394	0.533	0.664
0.10	0.246	0.441	0.639	0.796	0.899
0.15	0.324	0.569	0.774	0.903	0.965
0.20	0.386	0.655	0.848	0.946	0.984
0.25	0.433	0.712	0.888	0.966	0.991

In contrast and as shown in **Table 6E**, the resulting sample size of the survey 213 cases and 213 matched controls would not provide sufficient power to consider individual cancer specific sites (e.g. leukemia). For all sites combined, however, the resultant sample size *is powered* to detect an odds ratio 2.00 or greater with 80% power. Power estimates assume a two-sided test with alpha = 0.05, a value of 0.20 for the correlation of exposure status in the matches. Please see **Supplementary Tables S3-5** for the overall four malignancies combined risk estimates involving the survey-based population and a few descriptive tables for this second arm of the study.

6E. 213 case control pairs with two-sided test (Survey Sample size) Overall Combined Cancer Risk

	Odds Ratio				
Probability of exposure in controls	1.5	1.75	2.0	2.25	2.5
0.05	0.162	0.285	0.439	0.577	0.71
0.10	0.267	0.479	0.684	0.836	0.927
0.15	0.353	0.612	0.815	0.929	0.98
0.20	0.419	0.699	0.882	0.964	0.991
0.25	0.469	0.755	0.917	0.978	0.996

The decision to use birth residence as the primary location for determining UNGD activity until diagnosis comes into question if the case or control moves during the time from birth until diagnosis. This can lead to misclassification of the exposure and can affect exposure estimates. We carried out a cross tabulation of the county of birth residence for the 498 cases using birth records and the residence county at time of diagnosis using PA Cancer registry. Shown in **Table 7A,** there is high agreement within this study population in that over 85% of cases' parents remained in SW PA counties and the majority also remained within the same county over this period. Likewise shown in **Table 7B** are the results for the controls interviewed for their residential history as part of the survey study. Similarly, the cross tabulation indicates that there is high concordance of residence of controls remaining in the same county of their child's birth and maternal residence.

Table 7A. County of the mother's residence when giving birth, vs. County at diagnosis for the 498 childhood cancer cases

Child's Birth County	Child's Diagnosis County									
	Allegheny *	Armstrong	Beaver	Butler	Fayette	Greene	Washington	Westmoreland	Total	%
Allegheny- **outPGH	188	0	1	8	1	0	6	9	213	88.3
Armstrong	0	13	0	0	0	0	0	3	16	81.3
Beaver	1	1	30	3	0	0	0	0	37	81.1
Butler	0	0	1	55	0	0	0	0	58	94.8
Fayette	2	0	0	0	23	1	2	1	29	79.3
Greene	0	0	0	0	0	9	3	0	12	75.0
Washington	4	0	0	0	0	2	49	0	55	89.1
Westmoreland	7	0	0	0	1	0	1	78	87	89.7
Total	204	14	32	68	25	12	61	91	507	

Table 7B. County of the mother's residence when giving birth vs county at diagnosis for 213 controls

Child's Birth County	Child's Diagnosis County									
	Allegheny	Armstrong	Beaver	Butler	Fayette	Greene	Washington	Westmoreland	Total	%
Allegheny	92	0	1	1	0	0	4	1	99	92.9
Armstrong	0	4	0	0	0	0	0	0	4	100
Beaver	2	0	14	2	0	0	0	0	18	77.8
Butler	2	0	0	16	0	0	0	0	18	88.9
Fayette	0	0	0	0	6	0	0	1	7	85.7
Greene	0	0	0	0	0	6	1	0	7	85.7
Washington	1	0	0	1	0	0	24	0	26	92.3
Westmoreland	0	1	0	0	0	0	0	39	40	97.5
Total	97	5	15	20	6	6	29	41	219*	

*Six controls were excluded due to low data quality or did not meet the resident location requirements

III. Results

Birth Record Sample Characteristics

Table 8 presents the distribution of the 507 childhood cancer cases by primary site for the Birth Record Study. These are newly diagnosed cases excluding relapses and secondary diagnoses. CNS and miscellaneous intracranial and intraspinal neoplasms comprised the largest group, with 38.3% of all cases, followed by leukemias and myeloproliferative diseases accounting for 32.5%, lymphomas (20.7%), and malignant bone tumors including EFOT (8.5%). (See **Supplementary Table S1** for more details).

Table 9 presents the number of total childhood cancer cases for the birth record study by county, year of birth, age group and year of diagnosis (2010-2019). Among the 507 childhood cancer cases eligible for the study, Allegheny County, being the most populous, contributed 204 (40.2%) of these cases followed by Westmoreland, Washington, and Butler counties with 90, 68, and 61 cases, respectively. Fewer cases were included in the 1990-1994 birth cohort as some of children "aged out", (i.e., older than 19 years for the period of cancer diagnosis from 2010-2019). The number of cases by year at diagnosis appears to be evenly distributed from 2010 to 2019. The distribution for the four childhood cancers for ages 0 to 19 years was similar within the total study population, as well as for the two survey populations. They were also similar to the national data recorded by the NCI SEER Program (Cronin et al, 2022).

Table 8 Primary Classes of Childhood Cancer Included in the Birth Record Study (2010-2019)

Primary Cancer Classes	All Cases N (%)
I. Leukemias, myeloproliferative diseases, and myelodysplastic diseases	165 (32.5)
II. Lymphomas and reticuloendothelial neoplasms	105 (20.7)
III. CNS and miscellaneous intracranial and intraspinal neoplasms	194 (38.3)
IV. Malignant bone tumors including EFOT	43 (8.5)†
TOTAL	507 (100)

† Including 20 cases of Ewing tumor and related sarcomas of bone.

Table 9. Characteristics of Childhood Cancer Cases in the Birth Record study, South Western PA 2010-2019

	Total cases (N=507)
	N (%)
Year of Birth	
1990-1994	46 (9.1)
1995-1999	107 (21.1)
2000-2004	115 (22.7)
2005-2009	104 (20.5)
2010-2014	96 (18.9)
2015-2018	39 (7.3)
County of Residence	
Allegheny†	204 (40.2)
Armstrong	14 (2.8)
Beaver	32 (6.3)
Butler	68 (13.4)
Fayette	25 (4.9)
Greene	12 (2.4)
Washington	61 (12.0)
Westmoreland	91 (18.0)
Year of Diagnosis	
2010	60 (11.8)
2011	63 (12.4)
2012	45 (8.9)
2013	52 (10.2)
2014	47 (9.3)
2015	51 (10.1)
2016	52 (10.2)
2017	41 (8.1)
2018	51 (10.1)
2019	45 (8.9)
Age Group at Diagnosis	
0-4	149 (29.4)
5-9	98 (19.3)
10-14	111 (21.9)
15-19	146 (28.8)
20-24‡	2 (0.4)
25-29‡	1 (0.2)

† Excluding the City of Pittsburgh where UNGD is not permitted.
‡ Applicable for malignant bone tumors only.

Table 10. Distributions of Sociodemographic Characteristics of Childhood Cancer Cases Using Birth Record Information in the Birth Record-Based Studies with County-Matched Controls

Sociodemographic Characteristic	Birth Record-Based Study	
	Cases (%)	*Controls (%)*
Total number	498 (100)	498 (100)
Sex at Birth		
Female	216 (43.4)	216 (43.4)
Male	282 (56.6)	282 (56.6)
Maternal Age (years)		
<20	33 (6.6)	25 (5.0)
20-24	79 (15.9)	83 (16.7)
25-29	132 (26.5)	124 (24.9)
30-34	146 (29.3)	160 (32.1)
≥35	108 (21.7)	106 (21.3)
Maternal Race		
White	480 (96.4)	480 (96.4)
Black	12 (2.4)	12 (2.4)
Other	5 (1)	6 (1.2)
Maternal Education [1]		
≤ 8th Grade	2 (0.4)	3 (0.6)
Some High School	36 (7.2)	25 (5)
High School Diploma	145 (29.1)	141 (28.3)
Some College	124 (24.9)	123 (24.7)
College Degree or Higher	186 (37.4)	198 (39.8)
Unknown	5 (1)	8 (1.6)
Number of Prenatal Visits		
0-7	41 (8.2)	48 (9.6)
8-12	241 (48.4)	245 (49.2)
13-16	177 (35.5)	176 (35.3)
≥17	20 (4.0)	17 (3.4)
Unknown	19 (3.8)	12 (2.4)
Birth weight		
≤2500 g	28 (5.4)	23 (4.6)
2501- 4000 g	411 (82.5)	426 (85.5)
>4000 g	60 (12.1)	49 (9.8)
Unknown	28 (5.4)	23 (4.6)
Smoking during pregnancy[2]		
Never	397 (79.7)	408 (81.9)
Ever	92 (18.5)	89 (17.9)
Unknown	9 (1.8)	1 (0.2)
Gestation in weeks		
Mean (±S.D.)	38.7 (1.8)	38.8(1.6)

[1] p value=.08 survey based education >college; p value<.01 for birth record based> college

[2] p value=.28 survey based ever smoked during pregnancy ; p value<.026 for birth record based smoking

Maternal and Birth Characteristics of Birth Record Based Study

Table 10 presents characteristics of cancer cases and their matched controls for the birth-record based study. Childhood cancer cases and their matched controls were 56.6% male, and approximately 96% of the maternal study population reported a race of white. Case mothers reported an educational level of some college (24.9%) or completed college degree or higher (37.4%). The control distribution of education was similar (24.7% and 39.8%, respectively). There was also a similar proportion of cases and county-matched controls with a birth weight between 2501-4000g (82.5% and 85.5%, respectively). The proportion of mothers who reported never smoking during pregnancy was similar for cases and county-matched controls (79.7% and 81.9%, respectively). The birth weight of case infants versus control infants between 2501-4000g was also similar (82.3% and 85.6%, respectively). Similarly, 79.7% of mothers of cases and 82% of mothers of controls reported never having smoked cigarettes during their pregnancy. The average gestational age was 38 weeks for both groups.

Supplementary Table S2 presents the distributions of the eight UNGD activities metrics within a 5-mile radius of the residence among all 498 cancer cases and their 498 county-matched birth certificate controls for the two exposure time windows.

Exposure to UNGD Activity and Risk of Childhood Cancer

The study team analyzed the association between UNGD exposures and risk of four childhood malignancies (lymphoma, leukemia, CNS tumor and malignant bone tumor) combined for all 498 cases and their matched controls based on the information on birth records.

In the birth record-based analyses, the study team presented the results for two exposure time windows separately: T1 was mother's pregnancy period and T2 was from birth to the index date. The index date was the date of malignancy diagnosis for cases and the corresponding date for the matched controls. In addition to matching factors (date of birth, sex, and race), results presented were adjusted for maternal age at childbirth, education level, smoking status at childbirth, as well as gestation age, birthweight, TRI, UMTRA, and superfund site.

Four Malignancy Types Combined

Table 11 presents UNGD activities related to the risk of childhood malignancies. During pregnancy, mothers of 39 (18.3%) cases and of 41 (19.2%) county-matched controls in the survey-based study (213 pairs) reported a history of residence within 5 miles of a UNGD site. In the birth record-based study (498 pairs), the corresponding numbers were 94 (18.9%) cases and 99 (19.9%) controls. Compared with non-exposed group, there was no evidence to support an association between exposure to UNGD activity during mother's pregnancy and risk of malignancy in childhood and adolescence.

In the birth record-based analysis (498 case-control pairs), children diagnosed with any of the four malignancies included in the study were about four times more likely to live in a house within 0.5 miles of a UNGD site than controls (OR=3.94, 95% CI [1.66-9.30], P=0.002). There was a statistically significant linear trend for close-proximity and risk of childhood malignancy (p=0.004) When the subjects were divided into quartiles of overall UNGD activities, increasing levels of these were associated with increased risk of the four childhood malignancies. For example, children diagnosed with any of the four malignancies were more than two times more likely to be in the highest quartile of overall UNGD activities within 2 miles (OR=2.16, 95% CI [1.10-4.25], p=0.026) than their matched controls, and the linear trend for the overall UNGD activities with risk of these malignancies was statistically significant (p for trend=0.032).

Table 11. Overall Unconventional Natural Gas Drilling Activities and Risk of Four Childhood/Adolescent Malignances Combined During Two Exposure Periods in Southwestern PA 2010-2019

Overall UNGD activities by exposure period	**Birth Record-Based Study with County-Matched Controls (498 case-control pairs)**		
	Controls	***Cases***	**OR (95% CI)†**
T1: During Mother's Pregnancy			
Non-exposed	399	404	1.00
Exposed*	99	94	0.82 (0.47-1.41)
By buffer zone			
Non-exposed	399	404	1.00
(2-5] miles	64	63	0.84 (0.48-1.46)
(1-2] miles	24	22	0.72 (0.31-1.67)
(0.5-1] miles	9	7	0.65 (0.19-2.26)
[0-0.5] miles	2	2	0.81 (0.05-14.62)
P trend‡			*0.3817*
By overall UNGD activities within 5 miles			
Non-exposed	399	404	1.00
Lowest (1st) quartile	24	17	0.63 (0.29-1.34)
Low-middle (2nd) quartile	25	22	0.77 (0.37-1.64)
High-middle (3rd) quartile	25	36	1.40 (0.63-3.14)
Highest (4th) quartile	25	19	0.75 (0.31-1.83)
P trend‡			*0.7587*

* Exposed included individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls) (T2); non-exposed otherwise.

† All ORs and their 95% CIs for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and the following variables, including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and Superfund site (no, yes). **Odds ratios and confidence ratios which are bolded are significant at P < .05.**

‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.

Table 11 Continued. Overall Unconventional Natural Gas Drilling Activities and Risk of Four Childhood/Adolescent Malignances Combined During Two Exposure Periods in Southwestern PA 2010-2019

Overall UNGD activities by exposure period	**Birth Record-based Study with County-matched Controls (498 case-control pairs)**		
	Controls	***Cases***	**OR (95% CI)†**
***T2: From Birth to Index Date*§**			
Non-exposed	201	187	1.00
Exposed*	297	311	1.24 (0.87-1.78)
By buffer zone			
Non-exposed	201	187	1.00
(2-5] miles	178	170	1.18 (0.82-1.71)
(1-2] miles	72	77	1.49 (0.89-2.51)
(0.5-1] miles	37	38	1.61 (0.85-3.03)
[0-0.5] miles	10	26	**3.94 (1.66-9.39)**
P trend‡			***P=0.0041***
By overall UNGD activities within 5 miles			
Non-exposed	201	187	1.00
Lowest (1st) quartile	74	86	1.40 (0.91-2.14)
Low-middle (2nd) quartile	74	50	0.76 (0.46-1.25)
High-middle (3rd) quartile	74	88	**1.69 (1.01-2.82)**
Highest (4th) quartile	75	87	1.79 (1.00-3.19)
P trend‡			*0.0975*
By overall UNGD activities within 2 miles**			
Non-exposed	201	187	1.00
Lowest (1st) quartile	29	37	1.74 (0.93-3.27)
Low-middle (2nd) quartile	30	32	1.48 (0.77-2.84)
High-middle (3rd) quartile	30	30	1.41 (0.72-2.77)
Highest (4th) quartile	30	42	**2.16 (1.10-4.25)**
P trend‡			***P=0.0321***

* Exposed included individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls) (T2); non-exposed otherwise.

† All ORs and their 95% CIs for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and the following variables, including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and Superfund site (no, yes). **Odds ratios and confidence ratios which are bolded are significant at P < .05.**

§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.

** The same data for those with UNGD exposure within 2-5 miles of buffer zone were included in this modelling but not presented repeatedly.

Lymphoma

An analysis was carried out on the 105 lymphoma cases and their matched controls using the overall UNGD activity metric with consideration by exposure within five miles versus no exposure within five miles. See **Table 12**. The analysis is shown for both T1 (based on residence during pregnancy till birth) and T2 periods (residency from birth till index date). There is no significant relationship between overall UNGD activity and lymphoma risk for the T1 period. However, for the T2 period involving UNGD activity from birth to date of diagnosis, the point estimate for exposure to UNGD activity was (OR=2.24, 95% CI [0.92-5.47], p=0.076). The data were analyzed by buffer zone, the ORs (95% CIs) of lymphoma for the distance of 2-5, 1-2, 0.5-1, and <0.5 miles from residence to a UNGD site were 2.06 (0.83-5.13), 2.45 (0.77-7.83), 5.05 (1.09-23.39), and 7.71 (1.01-59.00), respectively, compared with non-exposed group (p value for trend=0.015). When the subjects were grouped by the overall UNGD activities over time, the ORs for lymphoma increased with greater levels of UNGD activities within both 5 and 2 miles of buffer zones. For example, the ORs (95% CIs) of lymphoma for children with the first, second, and third tertile of overall UNGD activities limited to two miles of radius surrounding their residences were 2.12 (0.51-8.79), 2.66 (0.66-10.72), and 7.73 (1.63-36.87), respectively, compared with non-exposed individuals (p value for trend=0.020).

When the UNGD activities were summed over the number of standard deviations for each of the four phase-specific UNGD activities, ORs (95% CIs) of lymphoma for children in the first, second, third, and fourth quartile of summed scores were 1.39 (0.44-4.37), 1.89 (0.62-5.80), 4.35 (1.26-15.01), and 5.15 (1.35-19.63), respectively (p values for trend = 0.011), compared with the non-exposed group in the birth record-based analysis.

Table 12. Overall Unconventional Natural Gas Drilling Activities and Risk of Childhood Lymphoma During Two Exposure Periods in Southwestern PA 2010-2019

Overall UNGD activities by exposure period	**Birth Record-based Study with County-matched Controls (105 Lymphoma case-control pairs)**		
	Controls	***Cases***	**OR (95% CI)†**
Period T1: During Mother's Pregnancy			
Non-exposed	89	90	1.00
Exposed*	16	15	0.91 (0.26-3.12)
By buffer zone			
Non-exposed	89	90	1.00
(2-5] miles	10	9	0.96 (0.27-3.48)
(1-2] miles	3	2	0.77 (0.09-6.34)
(0.5-1] miles	1	2	1.82 (0.11-30.83)
[0-0.5] miles	2	2	2.26 (0.06-85.26)
P trend‡			*0.6818*
By overall UNGD activities within 5 miles			
Non-exposed	89	90	1.00
Lowest (1st) quartile	5	1	0.28 (0.03-2.60)
Low-middle (2nd) quartile	5	5	0.82 (0.13-5.06)
High-middle (3rd) quartile	3	6	4.83 (0.4-58.83)
Highest (4th) quartile	3	3	3.59 (0.25-50.69)
P trend‡			*0.4023*

* Exposed included individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls) (T2); non-exposed otherwise.

† All ORs and their 95% CIs for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and the following variables, including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and Superfund site (no, yes). **Odds ratios and confidence ratios which are bolded are significant at P < .05.**

‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.

§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.

Table 12. Continued. Overall Unconventional Natural Gas Drilling Activities and Risk of Childhood Lymphoma During Two Exposure Periods in Southwestern PA 2010-2019

Overall UNGD activities by exposure period	**Birth Record-based Study with County-matched Controls (105 Lymphoma case-control pairs)**		
	Controls	***Cases***	**OR (95% CI)†**
Period T2: From Birth to Index Date§			
Non-exposed	40	32	1.00
Exposed*	65	73	2.24 (0.92-5.47)
By buffer zone			
Non-exposed	40	32	1.00
(2-5] miles	39	39	2.06 (0.83-5.13)
(1-2] miles	17	16	2.45 (0.77-7.83)
(0.5-1] miles	6	12	**5.05 (1.09-23.39)**
[0-0.5] miles	3	6	**7.71 (1.01-59.00)**
P trend‡			***0.0149***
By overall UNGD activities within 5 miles			
Non-exposed	40	32	1.00
Lowest (1st) quartile	13	15	1.74 (0.53-5.77)
Low-middle (2nd) quartile	18	11	1.14 (0.35-3.72)
High-middle (3rd) quartile	15	24	**5.68 (1.58-20.48)**
Highest (4th) quartile	19	23	**3.96 (1.01-15.49)**
P trend‡			***0.0155***
By overall UNGD activities within 2 miles**			
Non-exposed	40	32	1.00
Lowest (1st) tertile	8	7	2.12 (0.51-8.79)
Middle (2nd) tertile	10	12	2.66 (0.66-10.72)
Highest (3rd) tertile	8	15	**7.73 (1.63-36.67)**
P trend‡			***0.0201***

* Exposed included individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls) (T2); non-exposed otherwise.

† All ORs and their 95% (CIs for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and the following variables including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and Superfund site (no, yes). **Odds ratios and confidence ratios which are bolded are significant at P < .05.**

‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.

§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.

** The same data for those with UNGD exposure within 2-5 mile of buffer zone were included in this modelling but not presented repeatedly.

Leukemia

During both the mother's pregnancy and postnatal period, there was no elevated risk of childhood leukemia noted with exposure to any UNGD activities (or overall cumulative activities) or proximity to UNGD sites, in the birth record analysis. In the birth record-based analysis, for the postnatal (T2) period overall, any exposure to UNGD was not associated with the risk of leukemia (OR = 0.79, 95% CI = 0.35-1.79, *P* = 0.574). **See Table 13.**

Table 13. Overall Unconventional Natural Gas Drilling Activities and Risk of Childhood Leukemia During Two Exposure Periods in Southwestern PA 2010-2019

Overall UNGD activities by exposure period	**Birth Record-based Study with County-matched Controls (157 Leukemia case-control pairs)**		
	Controls	***Cases***	**OR (95% CI)†**
Period T1: During Mother's Pregnancy			
Non-exposed	120	122	1.00
Exposed*	37	35	0.73 (0.25-2.10)
By buffer zone			
Non-exposed	120	122	1.00
(2-5] miles	21	25	0.77 (0.27-2.24)
[0-2] miles	16	10	0.27 (0.05-1.36)
P trend‡			*0.1288*
By overall UNGD activities within 5 miles			
Non-exposed	120	122	1.00
Lowest (1st) quartile	8	8	0.89 (0.24-3.27)
Low-middle (2nd) quartile	10	6	0.44 (0.10-1.90)
High-middle (3rd) quartile	9	14	1.12 (0.24-5.25)
Highest (4th) quartile	10	7	0.47 (0.08-2.64)
P trend‡			*0.4337*

* Exposed included individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls) (T2); non-exposed otherwise.
† All ORs and their 95% CIs for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and the following variables, including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and Superfund site (no, yes). **Odds ratios and confidence ratios which are bolded are significant at P < .05.**
‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.
§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.
** The same data for those with UNGD exposure within 2-5 mile of buffer zone were included in this modelling but not presented repeatedly.

Table 13 Continued. Overall Unconventional Natural Gas Drilling Activities and Risk of Childhood Leukemia During Two Exposure Periods in Southwestern PA 2010-2019

Overall UNGD activities by exposure period	**Birth Record-based Study with County-matched Controls (157 Leukemia case-control pairs)**		
	Controls	***Cases***	**OR (95% CI) †**
Period T2: From Birth to Index Date§			
Non-exposed	67	69	1.00
Exposed*	90	88	0.79 (0.35-1.79)
By buffer zone			
Non-exposed	67	69	1.00
(2-5] miles	56	50	0.77 (0.34-1.75)
(1-2] miles	21	20	0.97 (0.28-3.33)
(0.5-1] miles	12	10	0.92 (0.24-3.46)
[0-0.5] miles	1	8	7.69 (0.70-83.91)
P trend‡			*0.3203*
By overall UNGD activities within 5 miles			
Non-exposed	67	69	1.00
Lowest (1st) quartile	25	31	1.16 (0.46-2.90)
Low-middle (2nd) quartile	23	9	0.38 (0.13-1.16)
High-middle (3rd) quartile	26	25	0.98 (0.29-3.27)
Highest (4th) quartile	16	23	1.51 (0.35-6.42)
P trend‡			*0.7676*
By overall UNGD activities within 2 miles**			
Non-exposed	67	69	1.00
Lowest (1st) tertile	14	11	0.62 (0.16-2.4
Middle (2nd) tertile	14	12	0.77 (0.20-2.92)
Highest (3rd) tertile	6	15	3.97 (0.66-23.95)
P trend‡			*0.2648*

* Exposed included individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls) (T2); non-exposed otherwise.

† All ORs and their 95% CIs for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and the following variables, including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and Superfund site (no, yes). **Odds ratios and confidence ratios which are bolded are significant at P < .05.**

‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.

§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.

** The same data for those with UNGD exposure within 2-5 mile of buffer zone were included in this modelling but not presented repeatedly.

Central Nervous System (CNS) Tumor

Similarly, analyses for the risk of CNS tumor from exposure to UNGD during the mother's pregnancy and the period from birth to the index date were conducted separately. There was no association between any measure of UNGD exposure and risk of childhood CNS among the 193 pairs of cases and county-matched controls studied. **See Table 14**. In this birth record-based analysis, any exposure to UNGD within five miles of the mother's residence at birth was not associated with the risk of CNS tumor either during pregnancy or from birth to the index date, (OR = 0.85, 85% CI = 0.35-2.03) and OR = 1.28, 95% CI= 0.74-2.22), respectively. There was one occurrence of a significant increase in risk of CNS tumor in the T2 period from birth to the index date in the lowest tertile of exposure by overall UNGD activities within two miles (OR= 2.79, 95% CI:1.08-7.24).

Table 14. Overall Unconventional Natural Gas Drilling Activities and Risk of Childhood Central Nervous System Tumor During Two Exposure Periods in Southwestern PA 2010-2019

Overall UNGD activities by exposure period	**Birth Record-based Study with County-matched Controls (193 CNS case-control pairs)**		
	Controls	***Cases***	**OR (95% CI)†**
Period T1: During Mother's Pregnancy			
Non-exposed	151	152	1.00
Exposed*	42	41	0.85 (0.35-2.03)
By buffer zone			
Non-exposed	151	152	1.00
(2-5] miles	29	28	0.84 (0.34-2.06)
(1-2] miles	7	8	1.07 (0.26-4.46)
[0-1] miles	6	5	0.68 (0.13-3.59)
P trend‡			*0.7712*
By overall UNGD activities within 5 miles			
Non-exposed	151	152	1.00
Lowest (1st) quartile	9	8	0.77 (0.18-3.30)
Low-middle (2nd) quartile	10	10	0.99 (0.28-3.47)
High-middle (3rd) quartile	11	14	1.09 (0.34-3.53)
Highest (4th) quartile	12	9	0.56 (0.15-2.03)
P trend‡			*0.5827*

* Exposed included individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls) (T2); non-exposed otherwise.

† All ORs and their 95% CIs for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and the following variables, including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and Superfund site (no, yes). **Odds ratios and confidence ratios which are bolded are significant at P < .05.**

‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.

§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.

** The same data for those with UNGD exposure within 2-5 mile of buffer zone were included in this modelling but not presented repeatedly.

Table 14 continued. Overall Unconventional Natural Gas Drilling Activities and Risk of Childhood Central Nervous System Tumor During Two Exposure Periods in Southwestern PA 2010-2019

Overall UNGD activities by exposure period	**Birth Record-based Study with County-matched Controls (193 CNS case-control pairs)**		
	Controls	***Cases***	**OR (95% CI)†**
***Period T2: From Birth to Index Date*§**			
Non-exposed	83	74	1.00
Exposed*	110	119	1.28 (0.74-2.22)
By buffer zone			
Non-exposed	83	74	1.00
(2-5] miles	62	62	1.23 (0.71-2.16)
(1-2] miles	28	30	1.54 (0.69-3.47)
(0.5-1] miles	15	15	1.38 (0.49-3.89)
[0-0.5] miles	5	8	1.96 (0.53-7.26)
P trend‡			*0.2818*
By overall UNGD activities within 5 miles			
Non-exposed	83	74	1.00
Lowest (1st) quartile	29	34	1.32 (0.69-2.50)
Low-middle (2nd) quartile	24	24	1.06 (0.48-2.33)
High-middle (3rd) quartile	24	30	1.55 (0.71-3.35)
Highest (4th) quartile	33	31	1.15 (0.47-2.79)
P trend‡			*0.6205*
By overall UNGD activities within 2 miles**			
Non-exposed	83	74	1.00
Lowest (1st) tertile	13	24	**2.79 (1.08-7.24)**
Middle (2nd) tertile	14	11	0.84 (0.29-2.49)
Highest (3rd) tertile	21	18	1.06 (0.39-2.87)
P trend‡			*0.9850*

* Exposed included individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls) (T2); non-exposed otherwise.

† All ORs and their 95% CIs for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and the following variables, including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and Superfund site (no, yes). **Odds ratios and confidence ratios which are bolded are significant at P < .05.**

‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.

§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.

** The same data for those with UNGD exposure within 2-5 mile of buffer zone were included in this modelling but not presented repeatedly.

Malignant Bone tumors

In the birth record-based study (43 case-control pairs), 3 mothers in the cases and 4 in the controls reported a similar exposure to UNGD activities. No risk of malignant bone tumor was associated with exposure to UNGD activities during mother's pregnancy. See **Table 15**. However, the small sample size of malignant bone tumors provided limited statistical power.

Table 15. Overall Unconventional Natural Gas Drilling Activities and Risk of Childhood/Adolescent Malignant Bone Tumor During Two Exposure Periods in Southwestern PA 2010-2019

Overall UNGD activities by exposure period	**Birth Record-based Study with County-matched Controls (43 case-control pairs)**		
	Controls	***Cases***	**OR (95% CI)†**
T1: During Mother's Pregnancy			
Non-exposed	39	40	1.00
Exposed*	4	3	0.22 (0.01-8.58)
T2: From Birth to Index Date§			
Non-exposed	11	12	1.00
Exposed*	32	31	1.01 (0.25-4.15)
By Buffer zone			
(2-5] miles	21	15	1.02 (0.25-4.12)
[0-2] miles	11	16	3.32 (0.42-26.24)
P trend			*0.2550*
By overall UNGD activities within 5 miles			
Lowest (1st) tertile	11	9	1.20 (0.25-5.85)
Middle (2nd) tertile	12	9	0.63 (0.1-4.03)
Highest (3rd) tertile	9	13	3.52 (0.30-40.73)
P trend‡			*0.5410*

* Exposed included individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls) (T2); non-exposed otherwise.

† All ORs and their 95% CIs for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and the following variables, including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and Superfund site (no, yes). **Odds ratios and confidence ratios which are bolded are significant at P < .05.**

‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.

§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.

Ewing Family of Tumor

In the birth record-based study, Ewings cases, which numbered only 20 in the present study, were compared using unconditional logistic regression to the total sample of 498 controls. This was done to increase the power to assess the relationship of UNGD activities with adjustment by matching variables, age, race, sex and county of birth as well as the other covariates. There were no significant findings from this analysis. **See Table 16.** Additional analysis did not reveal any dose-response relationships for different buffer zones and overall UNGD activities with risk of EFOT (both p values for trend >0.48). To align with previous studies in UNGD and childhood cancer risk in the literature, similar UNGD exposure metrics were created using well counts and IDW well counts. Overall, the associations between these well count measures and risk of childhood malignancies were like those of the newly created UNGD measurements described above. For example, levels of well counts and IDW well counts were associated with higher ORs for lymphoma, CNS tumor, and malignant bone tumor and EFOT. However, none of the point estimates or linear trend tests were statistically significant.

Table 16. Overall Unconventional Natural Gas Drilling Activities and Risk of Childhood/Adolescent Ewing Family of Tumor During Two Exposure Periods in Southwestern PA 2010-2019

Overall UNGD activities by exposure period	Birth Record-based Study with County-matched Controls (20 cases *vs*. 498 controls)		
	Controls	***Cases***	**OR (95% CI)†**
T1: During Mother's Pregnancy			
Non-exposed	399	18	1.00
Exposed*	99	2	0.55 (0.10-2.86)
T2: From Birth to Index Date§			
Non-exposed	201	6	1.00
Exposed*	297	14	1.55 (0.46-5.17)
By Buffer zone			
Non-exposed	201	6	1.00
(2-5] miles	178	9	1.50 (0.43-5.21
[0-2] miles	119	5	1.72 (0.36-8.36)
P trend			*0.4879*
By overall UNGD activities within 5 miles			
Non-exposed	201	6	1.00
Low (below median)	148	8	1.62 (0.46-5.7)
High (above median)	149	6	1.39 (0.32-5.96)
P trend‡			*0.6763*

* Exposed included individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls) (T2); non-exposed otherwise.

† All ORs and their 95% CIs for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and the following variables, including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and Superfund site (no, yes). **Odds ratios and confidence ratios which are bolded are significant at P < .05.**

‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.

§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.

Exposure to Other Environmental Risk Sites and Risk of Childhood Cancer

We examined the association for risk of childhood malignancies with exposures to TRI, UMTRA, and Superfund sites using the case and control mothers' residence for the birth-record study. These analyses were adjusted for age at childbirth, maternal education level, maternal smoking, gestational age, and birth weight. Overall, 86.7% of the children diagnosed with any of the 4 malignancies studied and 84.7% of their matched controls had a birth residence within 5 miles of a TRI site. Compared with non-exposed groups, living close to a TRI site was not associated with an elevated risk of 4 childhood malignancies combined. The malignancy-specific analysis revealed that children with leukemia were no more likely to have lived within 0.5-1 miles of a TRI site, *(***Table 17)**, and no consistent dose-response relationship was observed for proximity and level of exposure to TRI with risk of leukemia (both *Ps* for trend >0.32). No association with elevated risk of other childhood malignancy types including lymphoma, CNS tumor and osteosarcoma was observed for exposure to TRI site. (**Table 17**).

The proportions of children who were exposed to UMTRA and superfund sites within 5 miles of residence from birth to the index date were low. Overall, 8.4-10.6% of children in the study had a history of residence within 5 miles of UMTRA and superfund site. There was no increased risk in children for the four childhood malignancies combined nor for leukemia, lymphoma, and osteosarcoma. However, the risk of childhood CNS Tumors was significantly elevated **OR=2.68 (1.11-6.44) p=.028)** (**Table 18.)**

The proportions of children who were exposed to a Superfund site within five miles of residence from birth to index date was 8.8% for cases and 7.8% for controls. For the overall combined four malignancies, the odds ratio of 1.12 (95% CI: .71-1.76) was not significant. Moreover, leukemia, lymphoma, and osteosarcoma showed no significant results. However, the risk of CNS associated with proximity to a superfund site was OR=2.16 (0.96-4.86), p=.06 after adjustment for all covariates. **(Table 19).**

Table 17. Birth Record Exposure to Inverse-Distanced Weighed (IDW) Toxic Release Inventory (TRI) (US EPA) and Risk of Childhood Malignancies in Western Pennsylvania 2010-2019

Exposure to IDW TRI	Controls	Cases	OR (95% CI)†	*P*	*P* for trend‡
4 Cancer types combined (498 Pairs)					
Non exposed/[5-10] miles	76	66	1 (reference)	-	.5368
[2-5] miles	194	197	1.23 (.81-1.86)	0.3432	-
[1-2] miles	125	132	1.27 (0.8-2.01)	0.3179	-
[.5-1] miles	72	69	1.15 (0.69-1.92)	0.5845	-
[0-.5] miles	31	34	1.31 (0.71-2.42)	0.3909	-
Leukemia (157 pairs)					
Non exposed/[5-10] miles	20	19	1 (reference)	-	0.3228
[2-5] miles	64	61	1.23 (0.55-2.74)	0.6209	-
[1-2] miles	46	43	1.12 (0.48-2.63)	0.7932	-
[.5-1] miles	17	24	1.86 (0.68-5.05)	0.2252	-
[0-.5] miles	10	10	1.61 (0.47-5.55)	0.4535	-
Lymphoma (105 pairs)					
Non exposed/[5-10] miles	16	15	1 (reference)	-	0.3916
[2-5] miles	38	36	1.14 (0.37-3.44)	0.8226	-
[1-2] miles	30	34	1.45 (0.46-4.51)	0.5237	-
[.5-1] miles	17	10	0.59 (0.14-2.51)	0.4749	-
[0-.5] miles	4	10	3.89 (0.71-21.41)	0.1187	-
CNS tumor (193 pairs)					
Non exposed/[5-10] miles	29	29	1 (reference)	-	0.8641
[2-5] miles	82	78	0.99 (0.52-1.91)	0.9844	-
[1-2] miles	40	44	1.16 (0.54-2.46)	0.7096	-
[.5-1] miles	29	31	1.11 (0.51-2.4)	0.8019	-
[0-.5] miles	13	11	0.92 (0.36-2.34)	0.8564	-
Malignant bone tumor (43 pairs)					
Non exposed/[5-10] miles	11	3	1 (reference)	-	0.7340
[2-5] miles	10	22	10.51 (1.47-75.37)	0.0193	-
[0-2] miles	22	18	2.82 (0.52-15.43)	0.2312	-

† Odds ratios (ORs) were adjusted for maternal age at childbirth, maternal education level, maternal smoking status at childbirth, gestation age, and birthweight.

‡ Linear trend test for the exposure variable in ordinal values (1, 2, 3, 4 for quartile) that also included non-exposed.

Table 18. Birth Record Exposure to Inverse-Distance Weighted (IDW) Uranium Mill Tailings Remedial Action (UMTRA) (US DOE) and Risk of Childhood Malignancies in Western Pennsylvania 2010-2019

Exposure to IDW UMTRA	**Controls**	**Cases**	**OR (95% CI)†**	***P***	***P* for trend‡**
4 Cancer types combined (498 Pairs)					
Non exposed/[5-10] miles	456	445	1 (reference)	-	.1884
[0-5] miles	42	53	1.37 (0.86-2.2)	.1884	-
Leukemia (157 pairs)					
Non exposed/[5-10] miles	140	140	1 (reference)	-	.9098
[0-5] miles	17	17	.95 (.37-2.43)	.9098	-
Lymphoma (105 pairs)					
Non exposed/[5-10] miles	95	97	1 (reference)	-	0.5978
[0-5] miles	10	8	0.75 (0.25-2.2)	0.5978	-
CNS tumor (193 pairs)					
Non exposed/[5-10] miles	184	172	1 (reference)	-	**0.0281**
[0-5] miles	9	21	2.68 (1.11-6.44)	**0.0281**	-
Malignant bone tumor (43 pairs)					
Non exposed/[5-10] miles	37	36	1 (reference)	-	0.6164
[0-5] miles	6	7	1.40 (0.38-5.13)	0.6164	-

† Odds ratios (ORs) were adjusted for maternal age at childbirth, maternal education level, maternal smoking status at childbirth, gestation age, and birthweight.
‡ Linear trend test for the exposure variable in ordinal values (1, 2, 3, 4 for quartile) that also included non-exposed.

Table 19. Birth Record Exposure to Inverse-Distance Weighted (IDW) Superfund Site (US EPA) and Risk of Childhood Malignancies in Western Pennsylvania 2010-2019

Exposure to IDW TRI	**Controls**	**Cases**	**OR (95% CI)†**	***P***	***P* for trend‡**
4 Cancer types combined (498 Pairs)					
Non exposed/[5-10] miles	459	454	1 (reference)	-	0.6403
[0-5] miles	39	44	1.12 (0.71-1.76)	0.6403	-
Leukemia (157 pairs)					
Non exposed/[5-10] miles	139	142	1 (reference)	-	0.2679
[0-5] miles	18	15	0.64 (0.29-1.41)	0.2679	-
Lymphoma (105 pairs)					
Non exposed/[5-10] miles	97	99	1 (reference)	-	0.7097
[0-5] miles	8	6	0.82 (0.28-2.4)	0.7097	-
CNS tumor (193 pairs)					
Non exposed/[5-10] miles	182	172	1 (reference)	-	.0545
[0-5] miles	11	21	2.16 (0.96-4.86)	.0612	-
Malignant Bone Tumor (43 pairs)					
Non exposed/[5-10] miles	41	41	1 (reference)	-	0.0612
[0-5] miles	2	2	0.77 (0.1-6.01)	0.8055	-

† Odds ratios (ORs) were adjusted for maternal age at childbirth, maternal education level, maternal smoking status at childbirth, gestation age, and birthweight.
‡ Linear trend test for the exposure variable in ordinal values (1, 2, 3, 4 for quartile) that also included non-exposed.

IV. Discussion

The present study performed three separate analyses derived from 507 cases with childhood cancer newly identified throughout eight counties within Southwestern Pennsylvania between 2010 – 2019, a period of extensive hydraulic fracturing activity. The primary analyses were focused on 498 case-control pairs based on birth certificate data.

The following criteria were used to summarize results:

1. There are no data to suggest/support an increased risk
   a. No statistically significantly elevated odds ratios
   b. Odds ratios at or near 1
   c. Odds ratios below 1 (with or without statistical significance)
2. There are limited data to suggest/support an increased risk
   a. Statistically significantly elevated odds ratios in a low or moderate tertile
   b. Not statistically significant elevated odds ratios in multiple tertiles
3. There are moderate data to suggest/support an increased risk
   a. Statistically significantly elevated odds ratios in multiple low or moderate tertiles
   b. Statistically significantly elevated odds ratios in a high tertile
4. There are strong data to suggest/support an increased risk
   a. Statistically significantly elevated odds ratios in multiple tertiles
   b. Statistically significantly elevated odds ratios that increase across low, moderate, and high tertiles

Table 20. Summary of Results of Association Between UNGD Activities and Childhood Cancer in Southwestern PA 2010-2019

Analysis	**Exposure**	**Four Malignancy Types Combined**	**Lymphoma**	**Leukemia**	**CNS Tumor**	**Malignant Bone Tumor**	**Ewing Family of Tumor**
Birth-record based study with county matched controls (498 pairs)	**Overall UNGD**	**Moderate evidence**	**Moderate evidence**	None	Limited evidence	None	None

Four Childhood Malignancies Combined

In the birth record-based analyses with county-matched controls, there was limited to moderate evidence in support of an association between overall UNGD exposure and the combined four malignancies studied. See **Table 20**. No evidence was observed that exposure to other UNGD-related sites (i.e., compressor station, impoundment pond, and wastewater facility sites) or to other environmental risk sites (i.e., TRI, UMTRA and superfund site) was associated with the risk.

Childhood Lymphoma

This study provided moderate evidence suggesting an association between UNGD activity and childhood lymphoma. Analyses revealed statistically significant elevated ORs in multiple higher levels of overall UNGD activities. ORs for lymphoma increased as residential distances from UNGD sites decreased. These odds also increased as overall UNGD activities within both five miles and two miles of buffer zone increased. respectively. See **Table 12**. Although these positive associations between UNGD activities and risk of lymphoma were stronger in the birth record-based analysis than the survey-based analysis, size of the risk estimates and their direction and magnitude wee similar among the two analyses.

Childhood Leukemia

There was no evidence in support of an association between exposure to UNGD activities and other environmental factors with the risk of childhood leukemia was found in this study. See **Table 13**.

Childhood CNS

Limited data suggesting an association between exposure to overall UNGD activities and risk of childhood CNS was found in this study. See **Table 14**. Analyses revealed a significantly elevated risk of CNS in the lowest tertile of the overall UNGD activities during the primary study period, but no elevated risk estimates were observed for higher exposure levels, nor was there a dose-response relationship.

Malignant Bone Tumor and Ewing Family of Tumor

In this study, no evidence was found to support an association between exposures to UNGD activities and other environmental factors and the risk of malignant bone tumors, including EFOT. Given the small sample size of children with malignant bone tumor, particularly EFOT, additional studies with a larger sample size may be warranted.

Previous Studies

One investigation thus far (McKenzie et al., 2017) considered the association of hydraulic fracturing and the risk of childhood lymphoma and included only non-Hodgkin’s lymphoma (N=50) cases which were matched to other cancer controls without “environmentally mediated” cancers.

Within a ten-mile buffer, the researchers observed no statistically significant associations between density of oil and gas development and NHL in either model, based on trend analysis across categorical IDW well counts adjusted for age, race, gender, socioeconomic status, elevation, and year of diagnosis. Of the 50 cases, 18 were unexposed and 32 were within 8 km or a five-mile buffer with UNGD activity exposure. McKenzie et al. noted odds ratios of 1.5 (95% CI; 0.72, 3.3) in the lowest tertile of exposure, 0.91 (95% CI; 0.37, 2.2) in the medium tertile, and 1.6 (95% CI; 0.77, 3.4) in the highest tertile with the closest buffer. They did, however, note an association of increased risk of Leukemia with UNGD in Colorado in ages 5-24, Acute lymphoblastic leukemia cases were 4.3 times as likely to be in the highest exposure category.

The current study team considered all forms of lymphoma (52 Hodgkin’s, 22 NHL, 5 Burkitt's lymphoma, 25 miscellaneous lymphoreticular neoplasm, and 5 unspecified), and were able to consider multiple buffer distances and individual hydraulic fracturing phases as well as an overall metric that considered birth residence. In contrast, McKenzie et al. used geocoded addresses at time of cancer diagnosis as the only residence.

Lymphoma is more likely to emerge in the presence of infectious stimuli, chemical toxicity, or an immune system that has lost the ability for self-regulation (Skrabek, 2013). There are several studies investigating possible environmental risk factors for lymphoma in children and adults. Some of the environmental risk factors investigated include polychlorinated biphenyls, organophosphate and organochlorine pesticides, benzene, nitrogen dioxide, and in utero exposure to smoking. Many of these chemicals are in the IARC carcinogen list and are also found in hydraulic fracturing fluids (Mcnally, 2006). Future studies with biomarkers for exposure to UNGD activities may clarify the current study's observed association between hydraulic fracturing and risk of lymphoma.

Strengths and Limitations

This study has many strengths. It is only the second population-based study on UNGD activities and childhood cancer risk randomly sampling age, race, and sex matched controls from birth records. The study population was restricted to Western Pennsylvania counties which permitted UNGD activities since 2005. As such, the City of Pittsburgh was excluded due to a ban on hydraulic fracturing. This minimized potential confounding and bias due to other environmental risk factors. The rigid matching criteria (less than 45 days of difference in birth dates between a case and matched control) eliminated potential confounding effect by age. The collection of other environmental exposure data through publicly available sources provided additional information on factors (e.g., TRI, UMTRA, Superfund sites, impoundment ponds, compressor stations, and facilities accepting oil and gas waste), which were adjusted for through multivariable logistic models.

In addition to conventionally used well counts and IDW well counts as exposure variables, the study team was able to create a new metric called "overall activity" in estimates to evaluate cancer risk. The challenge in considering the health effects of individual hydraulic fracturing phases is that they may be occurring simultaneously in the background with other co-located wells. This overall metric accounted for the duration of UNGD activity and IDW components for each phase during the period of exposure studied. Moreover, phases of hydraulic fracturing and other potential environmental covariates including proximity to TRI, UMTRA, and Superfund sites were included in the overall analysis. An additional strength was the application of multiple buffers for proximity of residences within < 0.5, 0.5-1.0, 1-2, and 2-5 miles of these sites, which allowed for the assessment of cancer risk with UNGD proximity. The increased risk of childhood cancer with decreasing residential distance from UNGD sites suggests a probable link between UNGD activities and childhood cancer risk.

This comprehensive analysis also revealed consistent associations for various metrics of UNGD activities, which were highly correlated with each other and the risk of childhood cancer outcomes, further strengthening a probable link between UNGD activities in general and risk of childhood cancer.

This is the first study to include the four most common childhood cancers – leukemia, lymphoma, CNS tumors and malignant bone tumors. The inclusion of multiple cancer types provided a larger sample size for the study and allowed for the assessment of cancer-specific risk with UNGD activities. The strongest association was observed between UNGD activities and risk of childhood lymphoma, which are novel findings and warrant assessment by future studies.

The present study also has some limitations. The chief limitation is using distance as a proxy exposure measurement for UNGD activities. Exposure may be affected by many factors such as the nearby topography and geological formations, weather patterns, and water sources, and the behaviors of individuals residing near UNGD activity. It is possible that using distance as a proxy has resulted in exposure misclassification, which may identify an association where there is not one or vice versa. In addition, although the study team focused much attention on data cleaning and geocoding, the accuracy and completeness of the UNGD activity data used for the calculation of UNGD metrics cannot be certain. In addition, the use of residence from the birth records as a proxy for UNGD exposure from birth until index date to increase sample size also introduces the possibility of misclassification bias. However as shown in previous Table 8, there was an extremely high concordance (85%) with cases' residence at birth compared to their residence at diagnosis remaining in SW PA and an almost 80% of cases remaining in the same county. This adds validity to the use of birth certificates as a proxy for UNGD metrics for this study. Another limitation of the study was the small sample size particularly for Bone Cancer and Ewing Family of Tumor which resulted in large variations in risk estimates and wider confidence intervals.

V. Conclusion

There were no associations between unconventional natural gas development activities and childhood leukemia, brain and bone cancers, including Ewing's family of tumors. Results indicated that children who lived within 1 mile of a well had approximately 5 to 7 times the chance of developing lymphoma, a relatively rare type of cancer, compared to children who lived in a place with no wells within 5 miles. Data suggests that those who lived closer, especially in areas with greater intensity of unconventional natural gas development activities, had the highest risk. There was also a strong dose-response relationship between the overall UNGD activities over the four phases and risk of lymphoma. In addition, the closer the proximity of a residence to an UNGD site, the higher the risk of lymphoma, which further supports a possible link between UNGD activity and risk of childhood lymphoma.

For perspective, the incidence of lymphoma is, on average, 0.0012% in U.S. children under 20 years of age. Our study estimates that rate would be 0.006% to 0.0084% for children living within 1 mile of a well.

No evidence was observed for exposures to other environmental sites (i.e., TRI, UMTRA and Superfund sites), and any childhood cancers.

In this study, no evidence was found to support an association between exposures to UNGD activities and other environmental factors and the risk of leukemia, CNS tumors, and malignant bone tumors, including EFOT. Given the small sample size of malignant bone tumors, due to a very low incidence rate in the population, especially for EFOT, additional studies with a larger sample size are warranted.

References

1. Amitay EL, Keinan-Boker L. Breastfeeding and childhood leukemia incidence: A meta-analysis and systematic review. JAMA Pediatrics 2015; 169(6):e151025.
2. Bailey HD, Infante-Rivard C, Metayer C, et al. Home pesticide exposures and risk of childhood leukemia: Findings from the childhood leukemia international consortium. International Journal of Cancer 2015; 137(11):2644–2663.

3. Batterman, S., Jia, C., & Hatzivasilis, G. (2007). Migration of volatile organic compounds from attached garages to residences: a major exposure source. *Environmental research*, *104*(2), 224-240.
4. Bryan L, Kaye W, Antao V, Mehta P, Muravov O, Horton DK (2016) Preliminary Results of National Amyotrophic Lateral Sclerosis (ALS) Registry Risk Factor Survey Data. PLoS ONE 11(4): e0153683. doi:10.1371/journal.pone.0153683
5. Cao Y, Lu J, Lu J. Paternal smoking before conception and during pregnancy is associated with an increased risk of childhood acute lymphoblastic leukemia: A systematic review and meta-analysis of 17 case–control studies. Journal of Pediatric Hematology/Oncology 2020; 42(1):32–40.
6. Cardis E, Hatch M. The Chernobyl accident--an epidemiological perspective. Clinical Oncology: A Journal of the Royal College of Radiologists 2011; 23(4):251–260.
7. Carter KM, Harper JA, Schmid KW, Kostelnik J. Unconventional natural gas resources in Pennsylvania: The backstory of the modern Marcellus Shale play. *Environ Geosci*. 2011;18(4):217-257. doi:10.1306/eg.09281111008
8. Chen M, Chang CH, Tao L, Lu C. Residential exposure to pesticide during childhood and childhood cancers: A meta-Analysis. Pediatrics 2015; 136(4):719–729.
9. Chiavarini M, Naldini G, Fabiani R. Maternal folate intake and risk of childhood brain and spinal cord tumors: A systematic review and meta-analysis. Neuroepidemiology2018; 51(1–2):82–95.
10. Clark, C. et al. "Unconventional Oil and Gas Development Exposure and Risk of Childhood Acute Lymphoblastic Leukemia: A Case–Control Study in Pennsylvania, 2009–2017", Environmental Health Perspectives, volume 108, no 8. August 2022.
11. Cronin KA, Scott S, Firth AU, et al. Annual Report to the Nation on the Status of Cancer, Part 1: National Cancer Statistics. *Cancer*. October 27, 2022. DOI: 10.1002/cncr.34479.
12. Deziel NC. et al. "Assessing exposure to unconventional oil and gas development: strengths, challenges, and implications for epidemiologic research", Curr Environ Health Rep. 2022;9(3):436-450. doi:10.1007/s40572-022-00358-4
13. Elliott EG, Trinh P, Ma X, Leaderer BP, Ward MH, Deziel NC. Unconventional oil and gas development and risk of childhood leukemia: Assessing the evidence. *Sci Total Environ*. 2017;576:138-147. doi:10.1016/
14. Fryzek J, Pastula S, Jiang X, Garabrant DH. Childhood cancer incidence in Pennsylvania counties in relation to living in counties with hydraulic fracturing sites. *J Occup Environ Med*. 2013;55(7):796-801. doi:10.1097/JOM.0b013e318289ee02
15. Gröbner SN, Worst BC, Weischenfeldt J, et al. The landscape of genomic alterations across childhood cancers. Nature 2018; 555(7696):321–327.
16. Hsu WL, Preston DL, Soda M, et al. The incidence of leukemia, lymphoma and multiple myeloma among atomic bomb survivors: 1950–2001. Radiation Research 2013; 179(3):361–382.
17. Hu, L., Luo, D., Zhou, T., Tao, Y., Feng, J., & Mei, S. (2017). The association between non-Hodgkin lymphoma and organophosphate pesticides exposure: A meta-analysis. *Environmental pollution*, *231*, 319-328.
18. Hurley T, Chippi-Shrestha G, Gheisi A, Hewage K, Sadiq R. Characterizing hydraulic fracturing fluid greenness: application of a hazard-based index approach. *Clean Techn Environ Policy*. 2016;18(3):647-668. doi:10.1007/s10098-015-1054-2
19. Johnson KJ, Cullen J, Barnholtz-Sloan JS, et al. Childhood brain tumor epidemiology: a brain tumor epidemiology consortium review. Cancer Epidemiology, Biomarkers & Prevention 2014; 23(12):2716–2736.

20. Kuo C-L, Duan Y and Grady J (2018) Unconditional or Conditional Logistic Regression Model for Age-Matched Case–Control Data? Front. Public Health 6:57. doi: 10.3389/fpubh.2018.00057
21. Liu R, Zhang L, McHale CM, Hammond SK. Paternal smoking and risk of childhood acute lymphoblastic leukemia: Systematic review and meta-analysis. Journal of Oncology 2011; 2011:854584.
22. Lombardi, C et al. 2021 Residential proximity to pesticide application as a risk factor for childhood central nervous system tumors. Environmental Research vol 197, June 2021
23. McKenzie, L et al. "Childhood hematologic cancer and residential proximity to oil and gas development," PLoS ONE 12(2): e0170423. doi: 10.1371/journal.pone.0170423 February 15, 2017
24. Mcnally, RJQ, Parker L., Environmental factors and childhood acute leukemias and lymphomas, Leukemia & Lymphoma, 2006 47:4, 583-598, DOI: 10.1080/10428190500420973
25. Moore L, Cagan A, Coorens THH, et al. The mutational landscape of human somatic and germline cells. Nature 2021; 597(7876):381–386.
26. National Cancer Institute (NCI), Cancer in Children and Adolescents. National Cancer Institute. https://www.cancer.gov/types/childhood-cancers/child-adolescent-cancers-fact-sheet#r12. Published November 4, 2021. Accessed January 27, 2023.
27. PADOH Bureau of Epidemiology, Division of Community Epidemiology. Ewing's Family of Tumors, Childhood Cancer and Total Cancer Standard Incidence Ratio Results for Washington, Fayette, Greene and Westmoreland Counties in Pennsylvania, Published March 2020. Accessed September 2022.
28. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukemia and brain tumors: a retrospective cohort study. Lancet 2012; 380(9840):499–505.
29. Pennsylvania, Bureau of Epidemiology, Division of Community Epidemiology. April 22, 2019 "Ewing's Family of Tumors, Childhood Cancer, and Radiation-Related Cancer Incidence" Review for Washington County and Canon-McMillan School District I
30. Ross JA, Spector LG, Robison LL, Olshan AF. Epidemiology of leukemia in children with Down syndrome. Pediatric Blood and Cancer 2005; 44(1):8–12.
31. S.K. Alawattegama et al."Well water contamination in a rural community in southwestern Pennsylvania near unconventional shale gas extraction", J Environ Sci Health A Tox Hazard Subst Environ Eng,2015
32. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA: A Cancer Journal for Clinicians 2021; 71(1):7–33.
33. Skrabek, P., Turner, D., & Seftel, M. (2013). Epidemiology of non-Hodgkin lymphoma. *Transfusion and Apheresis Science*, *49*(2), 133-138.
34. State of New Jersey Department of Health. (2017, February 17). *Toms River Township Childhood Cancer Investigation*. NJ Health | Environmental Health. https://www.nj.gov/health/ceohs/environmental-occupational/hazardous-waste-sites/ocean/dovertwp.shtml
35. U.S. Environmental Protection Agency. 2015. Analysis of Hydraulic Fracturing Fluid Data from the FracFocus Chemical Disclosure Registry 1.0. Office of Research and Development, Washington, DC. EPA/601/R-14/003.

36. U.S. EPA. Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-16/236F, 2016.
37. Van Maele-Fabry G, Lantin AC, Hoet P, Lison D. Childhood leukaemia and parental occupational exposure to pesticides: a systematic review and meta-analysis. Cancer Causes & Control 2010; 21(6):787–809.
38. Vinson F, Merhi M, Baldi I, Raynal H, Gamet-Payrastre L. Exposure to pesticides and risk of childhood cancer: a meta-analysis of recent epidemiological studies. Occupational and Environmental Medicine 2011; 68(9):694–702.
39. Wollin K-M, Damm G, Foth H, et al. Critical evaluation of human health risks due to hydraulic fracturing in natural gas and petroleum production. *Arch Toxicol*. 2020;94(4):967-1016. doi:10.1007/s00204-020-02758-7
40. Xu, Xiaohui et al, “A systematic assessment of carcinogenicity of chemicals in hydraulic fracturing fluids and flowback water,” Science direct, Journal of Environmental Pollution, Volume 25 1, August 2019, Pages 128-136.
41. Yeager M, Machiela MJ, Kothiyal P, et al. Lack of transgenerational effects of ionizing radiation exposure from the Chernobyl accident. Science 2021; 372(6543):725–729.
42. Zhang J, Walsh MF, Wu G, et al. Germline mutations in predisposition genes in pediatric cancer. New England Journal of Medicine 2015; 373:2336–2346.

Appendices

Appendix A: Background Reference Materials

Common Hydraulic Fracturing Fluid Constituents (U.S. EPA 2015, Hurley 2015, Wollin 2020)

Additive	Common Chemical Constituents	Function
Acid	Hydrochloric acid	Cleans casing and formation prior to injection; dissolves cement, minerals, and clays to reduce clogging of pore space
Antibacterial agent/biocide	Glutaraldehyde	Controls or eliminates bacterial growth that may reduce well productivity
Breaker	Peroxydisulfuric acid diammonium salt, sodium chloride	Reduces viscosity of gels and foams and promotes recovery of fracturing fluid
Clay controller	Choline Chloride, potassium chloride	Prevents mobilization of formation clays
Corrosion inhibiter	Methanol, propargyl alcohol, isopropanol	Protects steel tubing and other equipment from corrosion
Crosslinker	Ethylene glycol, potassium hydroxide, sodium hydroxide, borate salts	Increases gel viscosity by connecting polymer molecules
Friction reducer	Hydrotreated light petroleum distillates, mineral oil	Minimizes friction when pumping fluids to optimize fluid injection
Gelling agent	Guar gum, hydrotreated light petroleum distillates	Increases fluid viscosity to promote proppant transport and reduce fluid loss
Iron controller	Citric acid	Prevents precipitation of iron compounds
pH control	Carbonic acid, dipotassium salt, potassium hydroxide, sodium hydroxide, acetic acid	Regulates pH of a solution by either inducing a change (pH adjuster) or stabilizing and resisting change (buffer) to achieve desired qualities
Scale controller	Ethylene glycol, methanol	Controls or prevents scale deposits in production conduit or completion system
Solvent	Hydrochloric acid	Controls wettability of contact surfaces or prevents or breaks emulsions
Surfactant	Naphthalene	Decrease fluid surface tension, promote injection, and fluid recovery

Appendix B: Methods Reference Materials

City of Pittsburgh Zip Codes Excluded from the Study Area

Zip code	All or part City of Pittsburgh	Zip code	All or part City of Pittsburgh	Zip code	All or part City of Pittsburgh
15106	Part City	15212	Part City	15224	All City
15120	Part City	15213	All City	15226	Part City
15201	All City	15214	Part City	15227	Part City
15203	All City	15215	Part City	15230	All City
15204	Part City	15216	Part City	15232	All City
15205	Part City	15217	All City	15233	All City
15206	All City	15218	Part City	15234	Part City
15207	All City	15219	All City	15235	Part City
15208	All City	15220	Part City	15240	Part City
15210	Part City	15221	Part City	15260	All City
15211	All City	15222	All City	15282	All City

Summary Activities for Recruitment of Controls

Mode	Number of control mothers and fathers	Number of invitations sent/calls to control mothers and fathers	Number of calls/reminders sent	Total calls/messages sent	Bounced/ spam/ duplicate	Started	Finished	Completion Rate	Response Rate
US Mail	8355	8355					357		4.3%
Email	7062	16198	32096	48294	15235	179	167	93.0%	2.4%
SMS Text	4832	8991	2612	11603	0	394	84	21.0%	1.7%
Phone follow-up	1091	831	280	1111			32		2.9%
Totals	8355	34375	34988	61008	15235	573	640	89.8%	7.7%

The Population Survey Facility (PSF) at the University of Pittsburgh assisted the research team in recruiting matched controls. Following the initial mailing to 8,355 potential controls, the PSF employed a multimode approach for recruiting controls which entailed a combination of email, text message, and follow-up phone calls. Before data cleaning and across all modes the response rate was 7.7%. Contact information was obtained from Lexis-Nexus and consisted of up to 6 emails for each control (i.e., up to 3 emails for both mothers and fathers) and 4 cell phone numbers (i.e., up to 2 for both mothers and fathers). Approximately 61,000 total calls or electronic messages were sent to recruit matched controls, resulting in 640 completed surveys prior to data cleaning.

The IRB Approval Letter

University of Pittsburgh | Institutional Review Board
Office of Research Protections

EXEMPT DETERMINATION

Date:	March 16, 2021
IRB:	STUDY21020141
PI:	Evelyn Talbott
Title:	Heath Effects of Hydraulic Fracturing
Funding:	Name: Pennsylvania Department of Health, Funding Source ID: Contract number: 4400018535
Grant Title:	None

The Institutional Review Board reviewed and determined the above referenced study meets the regulatory requirements for exempt research under 45 CFR 46.104.

Determination Documentation

Determination Date:	3/16/2021
Exempt Category:	(2)(iii) Tests, surveys, interviews, or observation (identifiable); and for which limited IRB review was conducted via expedited review

Determinations:	None
Approved Documents:	• Questionnaire, Category: Data Collection; • Attachment 1 Agency Request for Project 09-04-20_Version_0.01.doc, Category: Sponsor Attachment; • Case Brochure, Category: Recruitment Materials; • CASE PRENOTIFICATION LETTER.docx, Category: Recruitment Materials; • Control Brochure, Category: Recruitment Materials; • CONTROL PRENOTIFICATION LETTER.docx, Category: Recruitment Materials; • Exempt Application Form, Category: IRB Protocol; • Phone Call Script - Scheduling Interview, Category: Recruitment Materials; • Verbal Consent Phone Script, Category: Recruitment Materials

If you have any questions, please contact the University of Pittsburgh IRB Coordinator, Dana DiVirgilio.

Please take a moment to complete our Satisfaction Survey as we appreciate your feedback.

Human Research Protection Office 3500 Fifth Avenue, Suite 106 Pittsburgh, PA 15213 www.hrpo.pitt.edu

Steps for Selection of County-Matched and Non-County-Matched Controls by PADOH Bureau of Health Statistics and Registries

Step 1) Import birth data for all Pennsylvania Cancer Registry (PCR) patients eligible for this study.

Step 2) To prepare for control selection, two fields were created for every patient – "Patient_Bin_1" for resident county-matched controls and "Patient_Bin_2" for those controls not matched to resident county. "Patient_Bin_1" was created by concatenating the mother's Race, the patient's sex per the birth record, and the mother's resident County at time of the patient's birth. "Patient_Bin_2" was created by concatenating the mother's race and the patient's sex per the birth record. The mother's race as reported on the birth record was recoded as the field "Moth_Race_Bin". The following logic was used to recode the mother's race:

Mother's Reported Race ("Moth_Race" via Birth data)	**Recoded Field ("Moth_Race_Bin")**
White	Whi
Black/African-American	Bla
All other entries	Oth

Step 3) To create the pool of potential controls, birth records from 1990-2019 (inclusive) were imported. Due to differences in the layout of these data, three separate data sets were created based on the following years of birth: 1990-2002, 2003-2012, and 2013-2019. Births that did not occur in one of the eight counties of interest for this study were removed from the pool of potential controls. Additionally, certain birth records were removed if, based on the mother's residence zip code, the mother resided in the City of Pittsburgh at the time of the birth. Two bins were created for each potential control: "Control_Bin_1" and "Control_Bin_2". "Control_Bin_1" leveraged the same methodology as described in Step 2 to create the "Patient_Bin_1" field, and "Control_Bin_2" leveraged the same methodology as described in Step 2 to create the "Patient_Bin_2" field.

Step 4) Prior to selecting the controls, all years of birth data were combined into one data set containing the respective bins used as part of the matching criteria, a unique ID for the birth record, and the potential control's date of birth. A random number was also associated with each respective birth record for use later in the selection process. A comprehensive data set was also created for the eligible patients that only included the respective bins used as part of the matching criteria, a unique ID for the birth record, and the patient's date of birth.

Step 5) County-matched controls were identified for all patients in a single Procedure in SAS SQL (Structure Query Language) step. This initial group of record pairings, "Control Group 1", contain patient-control record pairings that were matched on sex, race, and mother's residence county (contained in the "Control_Bin_1" field). Additionally, the matching criteria also included logic to only retain record pairings where the patient's date of birth was within 45 days of the control's date of birth. Controls that matched to multiple patients were isolated, and a single patient-control pairing was selected using simple random sampling (without replacement) via the SAS procedure Proc SurveySelect. Controls identified for "Control Group 1" were sorted by the random number assigned to the respective record during Step 4. A maximum of 40 controls were selected for each patient. Final checks were made to ensure all eligible patients matched to a set of controls, verify there were no duplicate controls represented in the final data set, and determine the final frequency of patient-control pairings.

Step 6) The selection process for "Control Group 2" followed the same logic as described in Step 5 for "Control Group 1", however, controls identified in Step 5 were removed from the pool of eligible birth records prior to the selection process, and the residence county parity requirement was removed from the matching criteria. Sex, race, and date of birth proximity (i.e., controls born within 45 days of the respective patient) were leveraged during the record matching process. The sex and race fields were contained in the "Control_Bin_2" field.

Step 7) The final release files were created for the study group using the controls selected for "Control Group 1" and "Control Group 2".

Dated Summary of Protocol Modifications.

Modification	Summary	Date Approved
Pitt IRB Modification #1	Revision of consent methodology from verbal to written Addition of osteosarcoma and EFOT cases aged 20-29 (previously restricted to 0-19)	September 20, 2021
Pitt IRB Modification #2	Addition of QR code for ease of obtaining (electronic) written consent Revision of LexisNexis contract to allow for phone number and email address tracing Approval of text and email-based recruitment strategies Revision of phone call script for non-response follow-up	February 2, 2022
Pitt IRB Modification #3	Revision of survey mode from 45-60 minutes by phone to 20-25 minutes by phone or online Revision of recruitment flyer to be included in recruitment emails Inclusion of Qualtrics-based online survey link in recruitment emails	February 23, 2022
Pitt IRB Modification #4	Addition of Dr. Jean Tersak as study co-investigator Survey staff personnel updates	May 5, 2022
Pitt IRB Modification #5	Addition of paper-based residential history for eligible case families Addition of Qualtrics-based text message and email recruitment methodology Revision of postcard to indicate survey mode preference	May 16, 2022
Pitt IRB Modification #6	Approval of Dr. Jean Tersak's letter of support for case recruitment materials Approval to host in-person informational sessions for eligible case families at State Health Centers	June 6, 2022
Pitt IRB Modification #7	Revision of Control Incentive to $15; Updated verbiage to reflect shortened survey length (20-25 min)	July 22, 2022
DOH IRB Modification #1	Verbal consent approved for cases and controls (double check)	August 21, 2022

Timeline of Study Activities

Action	Date
DOH Contract Effective Date	September 1, 2020
Study activities commenced by Pitt Study Team (kick-off meeting)	November 20, 2020
Study funding received by Pitt Public Health	December 8, 2020
Initial Pitt IRB Submission	February 23, 2021
Pitt IRB Approval	March 16, 2021
DOH Protected Use Agreement submission	April 19, 2021
Initial DOH IRB submission	June 14, 2021
DOH IRB Approval	June 17, 2021
DOH Protected Use Agreement Approval	July 7, 2021
External Advisory Board Inaugural Meeting	August 5, 2021
Initial case dataset received from DOH (survivors only)	September 2, 2021
Pitt IRB Modification #1 Approval	September 20, 2021
LexisNexis Contract Finalized	September 21, 2021
Case recruitment period commenced	September 28, 2021
Conclusion of 1st quarter of recruitment efforts: n= 71 case interviews	December 31, 2021
Revised case dataset received from DOH includes corrected classification of cancer cases)	January 15, 2022
Pitt IRB Modification #2 Approval	February 2, 2022
Pitt IRB Modification #3 Approval	February 23, 2022
Revised case dataset received from DOH (includes decedents)	February 25, 2022
Conclusion of 2nd quarter recruitment efforts: n= 107 case interviews	March 31, 2022
Complete control dataset received from DOH	April 21, 2022
Pitt IRB Modification #4 Approval	May 5, 2022
Pitt IRB Modification #5 Approval	May 16, 2022
Control recruitment period commenced	May 18, 2022
Pitt IRB Modification #6 Approval	June 6, 2022
Conclusion of 3rd quarter of recruitment: n= 140 case interviews, n=126 control interviews	June 30, 2022
Pitt IRB Modification #7 Approval	July 22, 2022
SMS text message recruitment of control families commenced	September 8, 2022
Email recruitment of control families commenced	September 14, 2022
Electronic recruitment of control families (Emails and Texts) done	September 22, 2022
Conclusion of 4th quarter of recruitment efforts: n= 234 case interviews, n= 640 Controls in	September 27th, 2022
Case/control recruitment period closure	September 27th, 2022
Data cleaning phase commencement	August 2022
Data cleaning phase closure: n= 234 case interviews, n= 373 Control interviews	October 2022
Data analysis phase commencement	September 2022
Data analysis phase closure	October 2022
Report writing phase commencement	October 2022
Report writing phase complete	November 2022
Report 1A submitted to DOH, Report 1B submitted to DOH	11/16 &11/23 2022
Final report submitted to DOH	March 1, 2023

Geocoding Addresses

Addresses of cases and controls were geocoded in ArcMap 10.6, using ArcGIS World Geocoding Service (WCS). All addresses were matched to a set of geocoordinates. WCS included a percentage of accuracy for each match that it found. A decrease in percentage could be due to a typo in the address such as “Street” versus “Avenue” or a misspelling of street name. Sometimes WCS returned a match for a street, but the number provided by the participant was not a currently recognized address along with that street. WCS then identified the centroid of the street. Lastly, it was possible that WCS was not able to find a street with the same name that matched the city and zip code. In that case, WCS defaulted to selecting the centroid of the zip code. In some scenarios, WCS finds multiple potential matches with varying levels in the percentage of accuracy. The analyst can review these other potential matches and evaluate if another one could fit better to the information provided by a participant. If an alternative match was better, the analyst can manually match that set of geocoordinates instead of what was originally selected by WCS. If the other options are less well fitting, the analyst keeps the match the same.

A total of 892, or 78%, of addresses were matched with 100% accuracy, and 257 of the remaining addresses had certainty scores below 100%. However, upon review of these 257, 163 addresses were correctly matched to point addresses. In these instances, typos or inclusions of unit numbers, etc. caused a decrease in the accuracy percentage, but the correct point was identified. Of the remaining addresses with accuracy below 100%, 74 were matched to the centroid of the street and 19 used a zip code centroid where no street could be identified. Only 6 of the centroid addresses were manually rematched with a potential match not originally selected by WCS. In all other cases, the analyst agreed with the choice of geocoordinate selected by WCS. Once the review was done, the geocoding results were exported into a csv file to be uploaded to GCP to the data programmer for exposure metrics calculation. ArcMap was not used to calculate the IDW exposure metrics due to the computing power required to measure distances between all houses and wells.

Aggregating Exposure Metrics Across Residential History

To have a dataset representing individual participants as opposed to houses, exposure metrics were then aggregated across residences for each case and control. Metrics were first calculated by house and by time period as described above. Inverse distance weighted metrics were then summed across houses for all time periods.

Since IDW Well counts cannot appropriately be summed across residences, as this would artificially inflate the counts of individuals who moved often, a different method was used for aggregation. Proportions were calculated for time spent in each individual house as part of the total time period of all residences listed per person. IDW well counts were multiplied by the proportion and then summed to get a time-weighted sum of wells for each person and time period. This potential inflation only occurs with this IDW well count variable but would not occur with the other metrics as they include a duration element. This is how the additional metrics calculated in this study improve upon metrics in the existing literature. For the other environmental exposure variables, the same procedure was used.

Addressing Issues with Incomplete Data

The study team anticipated incomplete data in exposure metrics and well data for the entire exposure period. To address these issues, the following protocol was used:

- For gaps in residency: If residency or well data were missing for some of the exposure period, the metric was based on available data. For each metric computed, a companion variable was calculated indicating the proportion of the time period with available data (variable name: data completeness). For example, the value ranges from 0 to 1 (depending on the proportion of residential history provided), a value of 1 indicates data was provided for the 100% of the participant's time period, while a value of 0.94 indicates data residential history was provided for 94% of the participant's time period. In the complete analysis, only 7 of 213 cases and 7 of 213 controls had less than 100% completion. A sensitivity analysis found that excluding these pairs did not change the results.
- For study participants who relocated to residences outside the eight-county study area: A buffering zone of 5 miles from all borders of the eight-county study area extending into the surrounding counties has been considered when downloading exposure data. Data within the buffering zone or of the adjacent counties that the buffering zone was in were downloaded.
- For study participants who relocated outside of the study area and its buffering area to another hydraulic fracturing county within Pennsylvania: DEP data was used to determine if the participant lived within ten miles of an area with hydraulic fracturing. If the participant lived within an area where hydraulic fracturing occurred, their exposure was considered unknown for that residence, which is accounted for in the data completeness variable described above. Residential histories for study participants who relocated outside of the study area and its buffering area to other states with hydraulic fracturing (West Virginia, Ohio, Texas, etc.) were flagged based on whether a hydraulic fracturing timeline and estimated exposure was able to be shown. If unable to be shown we their exposure was considered to be unknown for that residence, which is accounted for in the data completeness variable described above.
- Residential histories for study participants who relocated outside of the study area and its buffering area to other states without hydraulic fracturing were considered to have no exposure to hydraulic fracturing.
- For missing date information:
    - If the day of the month was missing: the 15th of the month was used
    - If the month was missing: the 7th month and 1st day was used
    - If the end date (move-out date) for a residence was missing: the date 1 day prior to the next listed residence was used
- For missing GIS information which could not be resolved to house number and street name:
    - If data had only street name, GIS coordinates corresponding to the centroid of the street were used
    - If data had only town/city, GIS coordinates corresponding to centroid of town/city used
    - If data had only zip code, GIS coordinates corresponding to centroid of zip code used

Appendix C. Outreach and Subject Recruitment Materials

Letter from the Secretary of Health

COMMONWEALTH OF PENNSYLVANIA
OFFICE OF THE SECRETARY OF HEALTH

The University of Pittsburgh Graduate School of Public Health is collaborating with the Pennsylvania Department of Health in conducting valuable research into the possible environmental risk factors for childhood cancer including exposures related to Hydraulic Fracturing in SW PA. Childhood cancer is the third leading cause of death for those under age nineteen.

To complete this research, the University must compare interview and environmental exposure information between children who have been diagnosed with cancer with data on those who have not. The cancer-free group is referred to as "controls" while those with cancer are referred to as "cases" for this type of study. The University is asking your assistance in this important study.

Parents of children with cancer will be identified through the Pennsylvania Cancer Registry as being diagnosed with this condition between 2010-2019, and parents of control children were identified from a sample of Pennsylvania birth records (by county) which were then selected by birth year and matched by gender and race with a child with cancer.

Participation in this study is entirely voluntary, and if you do not wish to be contacted again, simply return the enclosed card with the "NO" box checked. However, I encourage you to give serious consideration to participating in this valuable research. We need studies such as this one to find the possible causes and risk factors for childhood cancers. Your participation in this study will serve as a small but very personal contribution in helping to find the risk factors for childhood cancer, leading to possible improvements in the lives of others.

If you have any questions about the study, please contact Dr. Evelyn O. Talbott, DrPH, MPH, at 412-624-3074. For any information related to the opt-out option that cannot be answered by the University study team, you may call the Pennsylvania Department of Health at 717-783-2548.

Thank you in advance for considering participation in this important study.

Sincerely,

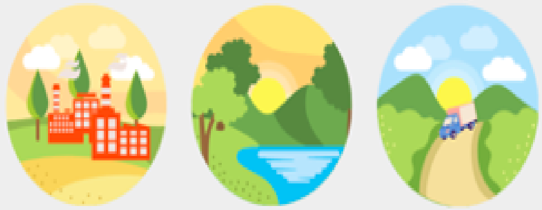
Denise Johnson, MD
Acting Secretary and Physician General
Pennsylvania Department of Health

Case Letter from the Pitt Study Team

University of Pittsburgh

pennsylvania DEPARTMENT OF HEALTH

PA Health and Environment Study



ENROLL ONLINE using this QR code!

June 1, 2022 **STUDYID####**

Dear Ms. and Mr ,

We are asking *the parents of children who were diagnosed with cancer* to participate in the PA Health and Environment Study. The study is a ***one-time online OR telephone survey*** examining possible environmental risk factors of childhood cancer including hydraulic fracturing. This study was initiated by the PA Department of Health in response to community concerns about environmental exposures. A letter from the PA Acting Secretary of Health, and a brochure explaining the study is enclosed.

*We need your help to make this study representative***.** Your residential history may be the key to understanding the environmental determinants of health. After your participation in this one-time 20 minute online OR telephone survey, you will receive a ***$25 payment card*** as compensation for your time.

To enroll or decline participation, you can scan the QR code above or navigate to the link paenv.pitt.edu/enroll, which will take you to an online enrollment and survey form. OR if you prefer, you can return the postcard enclosed here, and we will contact you to take the survey.

If you have any questions email me at eot1@pitt.edu or paenv@pitt.edu. My office phone number is 412-624-3074; and our project office phone number is 412-648-5185. You can read more about the study at paenv.pitt.edu/ccs.html.

Thank you so much for your consideration of this important request.

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

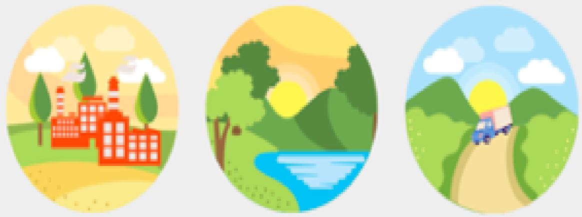
Jian-Min Yuan, MD, PhD
Professor, Department of Epidemiology
UPMC Hillman Cancer Center, University of Pittsburgh
Arnold Palmer Endowed Chair-Cancer Prevention

Control Letter from the Pitt Study Team

University of Pittsburgh

pennsylvania DEPARTMENT OF HEALTH

PA Health and Environment Study



ENROLL ONLINE using this QR code!

July 1, 2022 **STUDYID ######**

Dear Ms. and Mr.,

We are asking *the parents of children who were **NOT** diagnosed with cancer* to participate in the PA Health and Environment Study. The study is a *one-time online survey* examining possible environmental risk factors of childhood cancer including hydraulic fracturing. This study was initiated by the PA Department of Health in response to community concerns about environmental exposures. A letter from the PA Secretary of Health and a brochure explaining the study is enclosed.

We need your help to make this study representative. Your residential history may be the key to understanding the environmental determinants of health. After your participation in this one-time 20-minute online survey, you will receive a ***$15 payment card*** as compensation for your time.

To enroll or decline participation, you can scan the QR code above or navigate to the link paenv.pitt.edu/enroll, which will take you to an online enrollment and survey form.

If you have any questions email me at eot1@pitt.edu or paenv@pitt.edu. Dr. Talbott's office phone number is 412-624-3074; and our project office phone number is 412-648-5185. You can read more about the study at paenv.pitt.edu/ccs.html.

Thank you so much for your consideration of this important request.

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

Jian-Min Yuan, MD, PhD
Professor, Department of Epidemiology
UPMC Hillman Cancer Center, University of Pittsburgh
Arnold Palmer Endowed Chair-Cancer Prevention

Opt-In/Opt-Out Postcard

Please check the statement that represents your decision about participation in this study and sign and date at the bottom:

☐ I **DO** wish to be contacted regarding this study.

If yes, please fill out the contact information below:

Signature: ______________________________ Date: ___/___/20___

Name: ____

Email: ____

Phone Number: ____

Survey preference (check one): Text ___ Online ___ Phone ___

Current Address: ____

☐ I **DO NOT** wish to be contacted regarding this study.

Signature: ______________________________ Date: ___/___/20___

Name: ____

Please return this card in the envelope that has been supplied.

Research ID: _____________

Case Brochure

Why is this research being done?

Childhood cancer is the third leading cause of death in US children, yet there are very few known risk factors.

Pitt Public Health is conducting this study to consider some of the risks that may play a role. These include lifestyle behaviors, residential history, family medical history, workplace and environmental exposures, and other exposures during childhood and early life.

University of Pittsburgh
Graduate School of Public Health
Department of Epidemiology
130 DeSoto Street
Pittsburgh, PA 15261

paenv.pitt.edu

Case Control Study: Childhood Cancer in Southwestern Pennsylvania

Recruiting Parents for an Important Study

Please see why inside!

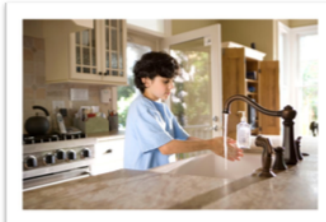
University of Pittsburgh

How did we get your name?

- Information was obtained through the Pennsylvania Cancer Registry as well as from PA birth records (by county) from the Department of Health.

Participation is voluntary!

- If you do choose to participate:
  - This will not impact your access to healthcare or treatment
  - You can withdraw from the study at any time



Who will be asked to participate in this research study?

Parents who have a child:

- Who was diagnosed with Ewing's/ bone cancers at age 0-29 years during 2010 through 2019, or
- Who was diagnosed with Childhood Leukemia, Lymphoma and Central Nervous System tumors at age 0-19 years during 2010 through 2019
- Resided in one of the following Pennsylvania counties:
  - **Allegheny County**
  - **Armstrong County**
  - **Beaver County**
  - **Butler County**
  - **Fayette County**
  - **Greene County**
  - **Washington County**
  - **Westmoreland County**

What will parents be asked to do?

- Complete a one-time, 45-60 minute telephone interview
  - Includes questions about individual, occupational, and environmental exposures
- Your time will be compensated

Other Information

- We will only be speaking with parents
- Any information provided for this research study will be confidential

Contact Information

Evelyn O. Talbott, DrPH, MPH
Principal Investigator

- **Phone:** (412) 648-5185
- **Email:** paenv@pitt.edu
- **Website:** paenv.pitt.edu
- **Project Office Location:**
  University of Pittsburgh
  Graduate School of Public Health
  A545 Public Health Building,
  130 De Soto St
  Pittsburgh, PA 15261

Control Brochure

Why is this research study being done?

Childhood cancer is the third leading cause of death in US children, yet there are very few known risk factors.

Pitt Public Health is conducting this study to consider some of the risks that may play a role. These include lifestyle behaviors, residential history, family medical history, workplace and environmental exposures, and other exposures during childhood and early life.

University of Pittsburgh
Graduate School of Public Health
Department of Epidemiology
130 DeSoto Street
Pittsburgh, PA 15261

paenv.pitt.edu

Case Control Study:
Childhood Cancer in Southwestern Pennsylvania

Recruiting Parents of Children Without Cancer as Controls for an Important Study

Please see why inside!

University of Pittsburgh

How did we get your name?

- Information was obtained from the PA birth records (by county) from the Department of Health.
- The participants in both groups must be matched in the following categories: **Age, Sex, Race, and County**.

What will parents be asked to do?

- Complete a one-time, 45-60 minute telephone interview
  - ◊ Includes questions about individual, occupational, and environmental exposures
- Your time will be compensated

Why are we asking you to participate in this study?

- We are recruiting a control group—families of children without cancer—to compare to families of children with this condition.
- Participation in this study will serve as an important and personal contribution in helping identify risk factors for childhood cancer.

- **Participation in this study is limited to the following counties:**
  - ◊ **Allegheny County**
  - ◊ **Armstrong County**
  - ◊ **Beaver County**
  - ◊ **Butler County**
  - ◊ **Fayette County**
  - ◊ **Greene County**
  - ◊ **Washington County**
  - ◊ **Westmoreland County**

Other Information

- We will only be speaking with parents
- Any information provided for this research study will be kept strictly confidential

Participation is voluntary!

- If you do choose to participate, you can withdraw from the study at any time

Contact Information

Evelyn O. Talbott, DrPH, MPH
Principal Investigator

- **Phone:** (412) 648-5185
- **Email:** paenv@pitt.edu
- **Website:** paenv.pitt.edu
- **Project Office Location:**
  University of Pittsburgh
  Graduate School of Public Health
  A545 Public Health Building,
  130 De Soto St
  Pittsburgh, PA 15261

Recruitment Text Message Scripts

Text Message Enrollment Scripts

Script 1 (briefest, requires no interaction with study team):
Header: **Pitt Public Health Needs Your Help.**
Important study on childhood cancer and hydraulic fracturing in SW PA!
Brief online survey, click here to consent and enroll: paenv.pitt.edu/enroll
$25 dollars for your time.
Reply NO to decline.

Script 2 (brief, requires no interaction with study team):
Header: **Pitt Public Health Needs Your Help.**
Hi (participant)! This is (staff) at Pitt Public Health. We're trying to reach you regarding a childhood cancer case-control survey. If you haven't already, will you consider participating in our brief, online survey? You will receive $25 for your time.
Here is the link to consent and enroll: paenv.pitt.edu/enroll
Reply NO to decline enrollment.

Script 3 (extended, requires interaction with study team):
Header: **Pitt Public Health Needs Your Help.**
Hi, this is Dr. Talbott's study team at the University of Pittsburgh School of Public Health. We are trying to contact Mr./Ms. _____ regarding a childhood health study. Do we have the right person? Reply YES or NO to decline.

No – Thank you, have a nice day.
Yes – We had sent a letter to him and wanted to confirm that he received it. The first letter was sent on date and the second letter was sent on date. Can you confirm that you received these letters? Reply YES or NO.

No, I did not receive the letters.

If you would like information about this study or would like to enroll, you can do so at paenv.pitt.edu/enroll

Yes, I received the letters.

That's great. As you know, we are studying the risk factors for childhood cancer, which we know very little about. Your participation would allow us to make accurate conclusions and help prevent childhood cancer in the future. You would receive $25 for the one-time online survey. Would you be interested in participating? Reply YES or NO to decline.

No, I would not like to participate.

Thanks for your response. Have a nice day.

Yes, I would like to participate.

Thank you! You can enroll online at paenv.pitt.edu/enroll

Recruitment Letter from Dr. Tersak

UPMC | CHILDREN'S HOSPITAL OF PITTSBURGH

Hematology/Oncology
Survivorship

Children's Hospital Drive
4401 Penn Avenue
Pittsburgh, PA 15224
T 412-692-8570
F 412-692-3412
SurvivorConnect@chp.edu

www.chp.edu/survivorship

Dear Mr. and Mrs.:

I am writing to you regarding an important study at the University of Pittsburgh, the "PA Health and Environment Study." I was asked to be involved due to my work as a pediatric oncologist. This study has the potential to help answer critical questions concerning environmental exposures within Southwestern Pennsylvania. A large number of participants from our region will increase the likelihood that we are able to answer the important questions of this study.

As the enclosed brochure describes, Pitt Public Health, in partnership with the Pennsylvania Department of Health, is conducting a case-control study of environmental risk factors and childhood cancer. Studies like this are necessary to evaluate the impact of industrial activities, including hydraulic fracturing ("fracking") on human health, especially on children's overall health and cancer risk.

I am writing to you in support of this state funded study and to encourage you to please consider participating when you are contacted by the Pitt study team. Participation in this study would require you to complete a short survey regarding your residential history, done over the phone or online, and should take approximately 20 minutes. When your answers are aggregated together with more than 1,000 participants like you, we can conduct detailed analysis and learn if the industrial activities are related to childhood cancer. Such knowledge is crucial for the development of strategies to mitigate or even eliminate such environmental risk factors in our community and beyond.

Your participation will be a critical contribution to advancing our understanding of pediatric cancer's environmental origins. I thank you in advance for your consideration to participate. Please reach out to the study team or directly to me if you have any questions about the study.

Sincerest thank you,

Jean M Tersak, MD

Eventbrite Email Invitation

University of Pittsburgh

pennsylvania DEPARTMENT OF HEALTH

PA Health and Environment Study

Hello!

The University of Pittsburgh study team is hosting two informational sessions for parents who are eligible to participate in a paid survey for the PA Health and Environment Study. You can read more about the study at paenv.pitt.edu/ccs.html.

These informational sessions will be held on:

Wednesday, August 10th from 9-11 AM
Westmoreland County's State Health Center,
233 W. Otterman St
Greensburg, PA 15601

and

Friday, August 12th from 1-3 PM
Washington County's State Health Center,
167 N. Main St., Suite 100
Washington, PA 15301

If you would like to attend, we kindly ask that you RSVP online
by Sunday, August 7th with a free ticket
using password **PAENV**
You can RSVP for the Westmoreland session here:
https://www.eventbrite.com/e/393175698097
or the Washington session here:
https://www.eventbrite.com/e/393191595647

If you have any questions email the study team at paenv@pitt.edu or call (412) 648-5185.

Thank you so much for your time and we hope to see you at the informational session.

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

Jian-Min Yuan, MD, PhD
Professor, Department of Epidemiology
UPMC Hillman Cancer Center, University of Pittsburgh
Arnold Palmer Endowed Chair-Cancer Prevention

2-Page Residential Questionnaire

ID Number ________

I consent to participate in the Pennsylvania Health and Environment Study:

Name: _________________________Signature: ____Date___________

Re(child's name)_ _______, please list your residences one year before your child was conceived _________through (date DX)___________

Email: ____--

Address of Residence: Street Address and City/Town	Zip Code	Move-in Date (Month & Year)	Move-out Date (Month & Year)	Was home within 1 mile of at least one major industrial facility? (check one)	Was there any oil/ gas activity or facility nearby? (check one)	Was this home within 1 mile of a farm/agricultural facility? (check one)
1.				Yes___ No__ Unknown__ If yes, describe:	Yes___ No__ Unknown__ If Yes, approx. date you noticed activity ___/___/___ or Don't Know __	Yes_____ No___ Unknown_____ If yes: please describe type (ex: dairy farm, apple orchard, etc.)
2.				Yes___ No__ Unknown__ If yes, describe:	Yes___ No__ Unknown__ If Yes, approx. date you noticed activity ___/___/___ or Don't Know__	Yes_____ No___ Unknown_____ If yes: please describe type
3.				Yes___ No__ Unknown__ If yes, describe:	Yes___ No__ Unknown__ If Yes, approx. date you noticed activity ___/___/___ or Don't Know __	Yes_____ No___ Unknown_____ If yes: please describe type
4.				Yes___ No__ Unknown__ If yes, describe:	Yes___ No__ Unknown__ If Yes, approx. date you noticed activity ___/___/___ or Don't Know __	Yes_____ No___ Unknown_____ If yes: please describe type
5.				Yes___ No__ Unknown__ If yes, describe:	Yes___ No__ Unknown__ If Yes, approx. date you noticed activity ___/___/___ or Don't Know __	Yes_____ No___ Unknown_____ If yes: please describe type

Regarding (child's name)_________, please list your residences from one year before your child was conceived through (date)________

Address of Residence: Street Address and City/Town	Zip Code	Move-in Date (Month & Year)	Move-out Date (Month & Year)	Was home within 1 mile of at least one major industrial facility? (check one)	Was there any oil/ gas activity or facility nearby? (check one)	Was this home within 1 mile of a farm/agricultural facility? (check one)
6.				Yes___ No__ Unknown__ If yes, describe:	Yes___ No__ Unknown__ If Yes, approx. date you noticed activity ___/___/___ or Don't Know __	Yes_____ No___ Unknown_____ If yes: please describe type (ex: dairy farm, apple orchard, etc.)
7.				Yes___ No__ Unknown__ If yes, describe:	Yes___ No__ Unknown__ If Yes, approx. date you noticed activity ___/___/___ or Don't Know __	Yes_____ No___ Unknown_____ If yes: please describe type
8.				Yes___ No__ Unknown__ If yes, describe:	Yes___ No__ Unknown__ If Yes, approx. date you noticed activity ___/___/___ or Don't Know __	Yes_____ No___ Unknown_____ If yes: please describe type
9.				Yes___ No__ Unknown__ If yes, describe:	Yes___ No__ Unknown__ If Yes, approx. date you noticed activity ___/___/___ or Don't Know __	Yes_____ No___ Unknown_____ If yes: please describe type
10.				Yes___ No__ Unknown__ If yes, describe:	Yes___ No__ Unknown__ If Yes, approx. date you noticed activity ___/___/___ or Don't Know __	Yes_____ No___ Unknown_____ If yes: please describe type

Appendix D. Medium-Length Qualtrics Survey (20-25 min)

SWPA Child Cancer - Shortened

Thank you, for participating in our study.

Childhood Cancer is the third leading cause of death among children in the US and yet there are very few known risk factors. This study will examine some risks that may play a role. These include environmental exposures, residential history, and lifestyle behaviors during childhood and early life. You will receive $25 for your time completing the survey. If there are any questions that you are uncomfortable about, you may decline to answer at any time.

Please do not hesitate to contact our project office at 412-648-5185 or email paenv@pitt.edu, if you have any questions.

1. What is your full name?

First Name ____
Last Name ____

2. **What is your child's name? This is your child that was diagnosed with cancer between** the ages of 0-29, in the years of 2010**-2019.**

First Name ____
Last Name ____

3. If you remember your four digit study ID number included in our enrollment materials please enter it here. ______________________________

4. What is your relationship to the child?

a) Biological Mother
b) Biological Father
c) Step Mother
d) Step Father
e) Other ____

5. What is the child's date of birth? ______________________________

6. Confirm your child's gender.

a) Male
b) Female
c) Child is Non-binary/third gender
d) Prefer not to say

7. Would you describe the child as being of Hispanic origin?

a) Yes

b) No
c) Unknown

8. Which of the following terms best describes the child's racial background? Check all that apply.

a) White
b) Black or African American
c) Native American/American Indian or Alaska Native
d) Asian or Pacific Islander
e) Other ______________________________
f) Unknown

9. Now we would like to ask what daycares and schools the child has attended, beginning with their first daycare or school and continuing in order:

Please include ANY address outside the home where the child spent long periods of time during the day.

	Name of School or Daycare	Year Attended - From YEAR to YEAR		School or Daycare Address			
	Name of Daycare or School	Year Start	Year End	Street	City	State	Zip
Daycare / School 1							
Daycare / School 2							
Daycare / School 3							
Daycare / School 4							
Daycare / School 5							
Daycare / School 6							
Daycare / School 7							
Daycare / School 8							

MOTHER'S BACKGROUND

10. What was the highest grade or year of school you / the mother had completed at the time that the child was born?

a) No formal schooling
b) Less than high school
c) 12 years, completed high school or equivalent
d) 1-3 years of college
e) Completed technical college
f) Associates degree
g) 4 years of college or Bachelors degree
h) Advanced degree
i) Don't know

11. What was your / the mother's marital status at the time the child was born?

a) Married or living with partner
b) Separated
c) Divorced
d) Widowed
e) Never married and not living with partner
f) Other ____

FATHER'S BACKGROUND

12. What was the highest grade or year of school you / the father had completed at the time that the child was born?

a) No formal schooling
b) Less than high school
c) 12 years, completed high school or equivalent
d) 1-3 years of college
e) Completed technical college
f) Associates degree
g) 4 years of college or Bachelors degree
h) Advanced degree
i) Don't know

13. What was your / the father's marital status at the time the child was born?

a) Married or living with partner
b) Separated
c) Divorced
d) Widowed
e) Never married and not living with partner
f) Other ____

RESIDENTIAL HISTORY

How many residences did you live in starting from one year before the conception of the child and ending with the date of the child's first cancer diagnosis?

14. How many residences did the biological mother live in starting from one year before the conception of the child and ending with the date of the child's first cancer diagnosis? ______________________________

15. How many residences did you live in starting from one year before the conception of the child and ending with the date of the child's first cancer diagnosis?

	Residences					Approximate Move-IN Date		Approximate Move-OUT Date	
	Street Address	City/Town	State	ZIP	County	Move-IN Month	Move-IN YEAR	Move-OUT Month	Move-OUT YEAR
Address 1 (starting 1 Year BEFORE CONCEPTION)									
Address 2									
Address 3									
Address 4									
Address 5									
Address 6									
Address 7									
Address 8									
Address 9									
Address 10									

Now we are going to ask question about your house at Address 1.

16. What year was this residence built? ______________________________

17. Which PRIMARY FORM of heating fuel do/did you use at this residence? (choose all that apply)

a) Natural Gas
b) Electricity
c) Propane
d) Kerosene
e) Wood
f) Coal
g) Solar
h) Don't know

18. What type of air conditioning did you use at this residence?

a) Central air conditioning
b) Window/wall air conditioning units
c) No air conditioning
d) Other - Please describe ___________________________________
e) Don't know

19. Did you or a family member/other resident operate a business out of this home, such as an auto mechanic shop or hair salon?

a) Yes (Please describe business) ________________________________
b) No

c) Don't know

I am now going to ask you some questions about pesticide, herbicide, and insecticide use for your residence at Address 1.

20. Was this residence ever exterminated for insects and pests so that you had to leave the house for a few hours?

a) Yes
b) No
c) Don't know

Display This Question:

If: Was this residence ever exterminated for insects and pests so that you had to leave the house for... = Yes

21. How often was this residence treated for pests?

a) Once a week
b) Once a month
c) Once every 2-3 months
d) Once a year
e) Don't know
f) Other, please specify ____

22. Was the yard or garden around this residence ever treated with insecticides or herbicides to control insects or weeds?

a) Yes
b) No
c) Don't know

Display This Question:

If: Was the yard or garden around this residence ever treated with insecticides or herbicides to cont... = Yes

23. How often was this yard or garden treated for pests?

a) Once a week
b) Once a month
c) Once every 2-3 months
d) Once a year
e) Don't know
f) Other, please specify ________________________________

24. What was the primary source of water for drinking and cooking at this residence? Please check all that apply:

a) City or township water supply
b) Well
c) Bottled water (for cooking and drinking only, not for showering)
d) Don't know

25. Did you ever have your water tested at this residence?

a) Yes
b) No
c) Don't know

26. Did you ever have this residence tested for radon?

a) Yes
b) No
c) Don't know

27. Did this residence ever require radon remediation?

a) Yes
b) No
c) Don’t know

Display This Question:

If: Did you ever have this residence tested for radon? = Yes

28. If you can recall, what were the approximate levels of radon detected?

____

29. Did this residence have an attached garage?

a) Yes
b) No
c) Don't know

I am now going to ask you some questions about the proximity of Address 1 to some facility types.

30. Was this residence located within 1 mile of a MAJOR INDUSTRIAL FACILITY?
Examples of these are: a factory, agricultural site or farm, power plant, steel mill, cement factory, chemical plant, etc.

a) Yes
b) No
c) Don't know

Display This Question:

If was this residence located within 1 mile of a MAJOR INDUSTRIAL FACILITY? = Yes

31. Were there more than one MAJOR INDUSTRIAL facility within 1 mile of this residence?

a) Yes. If yes, how many? ____
b) No
c) Don't know

Display This Question:

If was this residence located within 1 mile of a MAJOR INDUSTRIAL FACILITY? = Yes

32. If YES, can you describe all of these facilities?

____

33. Was this residence located within 1 mile of any OIL & GAS ACTIVITY or FACILITY

a) Yes
b) No
c) Don't know

Display This Question:

If Loop current: Was this residence located within 1 mile of any OIL & GAS ACTIVITY or FACILITY... = Yes

34. Was there considerable noise at this residence due to OIL & GAS ACTIVITIES?

a) Yes
b) No
c) Don't know

Display This Question:

If Loop current: Was this residence located within 1 mile of any OIL & GAS ACTIVITY or FACILITY... = Yes

35. Did you or any of your household members notice excessive dust generated from the OIL & GAS ACTIVITIES?

a) Yes
b) No
c) Don't know

36. Was this residence located within 1 mile of a FARM or AGRICULTURAL facility?

a) Yes
b) No
c) Don't know

Display This Question:

If Loop current: Was this residence located within 1 mile of a FARM or AGRICULTURAL facility? = Yes

37. Did you or any of your household members notice excessive dust, noise, odors, or other irritants generated from the agricultural activities that impacted your daily quality of life? ____

MOTHER'S OCCUPATIONAL HISTORY

How many jobs did you/the mother have in the period starting one year before the conception of the child and ending 2 years after the child's birth.

38. During the year before you were/the mother was pregnant with the child, did you work outside of the home?

a) Yes
b) No
c) Other ____________________________________

39. How many jobs did you / the Mother have in the period starting one year before the conception of the child and ending 2 years after the child's birth. ______________

Please tell me all of the different jobs you/the mother had outside of the home during this period - from 1 year before conception to 2 years post the birth of the child.

Please give the job title and month and year when you started and stopped working at that job.

40. How many jobs did you/the mother have in the period starting one year before the conception of the child and ending 2 years after the child's birth.

	Job Title	Date Started		Date Stopped	
	Job Title	Month	Year	Month	Year
Job 1					
Job 2					
Job 3					
Job 4					
Job 5					
Job 6					
Job 7					
Job 8					
Job 9					
Job 10					

41. For the first job you listed – as first job title, which of these categories are most similar to your occupational category?

11 = Agriculture, Forestry, Fishing and Hunting ... Refused

42. For the first job you listed - as first job title, which of these occupations are most similar to your occupation?

1 = Accountant, auditor, or bookkeeper... Refused

Display This Question:

If For the first job you listed -- as first job title, which of these occupations a... = 27 = Other (specify):

43. You said "Other" for job title. Please specify: ________________________________

For the first job you listed - - as first job title, - please answer the questions below.

44. Did/do you/the mother work at this job part time or full time?

a) Part time
b) Full Time
c) Don't Know

45. Did you/the mother continue to work at this job while pregnant?

a) Yes
b) No
c) Don't Know

46. If you were / the mother was at this job at the time you gave birth, did you / the mother take maternity leave?

d) Yes
e) No
f) Don't Know

Now I would like to ask you more about the chemicals or substances that you/the mother may have used at work. Some of the names may not sound familiar to you, but please answer as best you can.

47. Did you/the mother work with any of the following materials?

	Did you work with these?			[IF YES} Were you working with them during preconception or pregnancy?	
	Yes	No	Don't Know	Pre-conception	Pregnancy
1. Adhesives or glues, like rubber cement	○	○	○	☐	☐
2. Alcohols, such as methanol or ethanol, formaldehyde	○	○	○	☐	☐
3. Anesthetic gases	○	○	○	☐	☐
4. Automotive fluids, such as antifreeze, brake fluid, degreasers, freon, gasoline	○	○	○	☐	☐
5. Benzene	○	○	○	☐	☐
6. Volatile organics, such as: carbon disulfide, carbon tetrachloride, diesel fumes, ethylene oxide, glycol ethers, styrene, toluene, trichloroethylene (TCE) or trichlorethane (TCA), xylene	○	○	○	☐	☐
7. Metals, such as: chromium, lead, manganese, nickel, metal dust or fules	○	○	○	☐	☐
8. Paint products, such as: oil-based paints, paint strippers, paint thinners, lacquers	○	○	○	☐	☐
9. Pesticides, herbicides, fungicides, or insecticides	○	○	○	☐	☐
10. Pharmaceuticals or drugs	○	○	○	☐	☐
11. Phthalates	○	○	○	☐	☐
12. Vinyl chloride	○	○	○	☐	☐
13. X-ray or radioactive materials	○	○	○	☐	☐
14. Hair dyes	○	○	○	☐	☐
15. Any other? ______	○	○	○	☐	☐

FATHER'S OCCUPATIONAL HISTORY

How many jobs did you / the father have in the period starting one year before the conception of the child and ending 2 years after the child's birth.

Please tell me all of the different jobs you/the father had outside of the home during this period - from 1 year before conception with the child to 2 years after the birth of the child.

48. Please give the job title and month and year when you/ the father started and stopped working at that job.

	Job Title	Date Started		Date Stopped	
	Job Title	Month	Year	Month	Year
Job 1					
Job 2					
Job 3					
Job 4					
Job 5					
Job 6					
Job 7					
Job 8					
Job 9					
Job 10					

49. For the first job you listed – as first job title, which of these categories are most similar to your occupational category?

11 = Agriculture, Forestry, Fishing and Hunting ... Refused

50. For the first job you listed - as first job title, which of these occupations are most similar to your occupation?

1 = Accountant, auditor, or bookkeeper... Refused

Display This Question:

If For the first job you listed -- as first job title, which of these occupations a... = 27 = Other (specify):

51. You said "Other" for job title. Please specify: ________________________________

For the first job you listed - - as first job title, - please answer the questions below.

52. Did/do you/the father work at this job part time or full time?

a) Part time
b) Full Time
c) Don't Know

Now I would like to ask you more about the chemicals or substances that you/the father may have used at work. Some of the names may not sound familiar to you, but please answer as best you can.

53. Did you/the father work with any of the following materials?

	Did you work with these?			[IF YES} Were you working with them during preconception or pregnancy?	
	Yes	No	Don't Know	Pre-conception	Pregnancy
1. Adhesives or glues, like rubber cement	○	○	○	☐	☐
2. Alcohols, such as methanol or ethanol, formaldehyde	○	○	○	☐	☐
3. Anesthetic gases	○	○	○	☐	☐
4. Automotive fluids, such as antifreeze, brake fluid, degreasers, freon, gasoline	○	○	○	☐	☐
5. Benzene	○	○	○	☐	☐
6. Volatile organics, such as: carbon disulfide, carbon tetrachloride, diesel fumes, ethylene oxide, glycol ethers, styrene, toluene, trichloroethylene (TCE) or trichlorethane (TCA), xylene	○	○	○	☐	☐
7. Metals, such as: chromium, lead, manganese, nickel, metal dust or fules	○	○	○	☐	☐
8. Paint products, such as: oil-based paints, paint strippers, paint thinners, lacquers	○	○	○	☐	☐
9. Pesticides, herbicides, fungicides, or insecticides	○	○	○	☐	☐
10. Pharmaceuticals or drugs	○	○	○	☐	☐
11. Phthalates	○	○	○	☐	☐
12. Vinyl chloride	○	○	○	☐	☐
13. X-ray or radioactive materials	○	○	○	☐	☐
14. Hair dyes	○	○	○	☐	☐
15. Any other? ______	○	○	○	☐	☐

MOTHER'S SMOKING HISTORY

54. Have you/ has the mother smoked more than 100 cigarettes in your lifetime?

a) Yes
b) No
c) Don't know

Display this Question:

If Have you/ has the mother smoked more than 100 cigarettes in your lifetime? = Yes

55. How many cigarettes a day did you / the mother usually smoke during the following time periods?

One pack is usually 20 cigarettes.

56. What about e-cigarettes (like vaping) or other tobacco products like a cigar or hookah?

57. During what time periods did you / the mother smoke, vape or use other tobacco products?

	How many cigarettes did you smoke in the [read time period]?	How many of times a day did you vape/use e-Cigs or other types of Tobacco?
	Number of Cigarettes/Day	Number of E-Cigs or Vape or Tobacco/Day
12 months prior to pregnancy		
1st trimester of pregnancy		
2nd trimester of pregnancy		
3rd trimester of pregnancy		
0-24 months after ${q://QID987/ChoiceTextEntryValue/1}'s birth		
After 24 months of ${q://QID987/ChoiceTextEntryValue/1}'s birth until the reference date		

Family Cancer History

Now I would like to ask you some questions about your family's medical history. Please take your time and focus on the blood relatives of the child. Please try to recall whether any of the relatives were ever diagnosed with cancer. Leukemia, brain tumors, lymphomas, and Hodgkin's disease are all types of cancer and should be included.

58. Please record any relatives that have had cancer, and what kinds of cancer they had?

	*Any information about the cancer type, site, etc. should be entered here			If answer yes to cancer
	Yes	No	Don't know	Cancer Type
Mother	○	○	○	
Father	○	○	○	
Maternal Grandmother	○	○	○	
Maternal Grandfather	○	○	○	
Paternal Grandmother	○	○	○	
Paternal Grandfather	○	○	○	
Siblings of CHILD (1)	○	○	○	
Siblings of CHILD (2)	○	○	○	
Siblings of CHILD (3)	○	○	○	

59. During pregnancy, did you/ the mother ever have any of the following medical procedures?

During pregnancy, did you/ the mother ever have any of the following medical procedures?

	Column Options			Column Options					Column Options
	Receive this procedure?			During what time period?					Frequency
	Yes	No	Don't know	1 yr prior to conception	Pregnancy 1st Trimester	Pregnancy 2nd Trimester	Pregnancy 3rd Trimester	Don't know	How many times did this happen?
Diagnostic X-rays	○	○	○	☐	☐	☐	☐	☐	
Radiation therapy	○	○	○	☐	☐	☐	☐	☐	
Dental X-rays - Traditional	○	○	○	☐	☐	☐	☐	☐	
Dental X-rays - Panoramic	○	○	○	☐	☐	☐	☐	☐	

60. Did the child ever have any of the following procedures, prior to their first cancer diagnosis?

	Did your child ever receive this procedure?			If yes, what was the reason?	Frequency of Procedure	Age
	Y	N	Don't know	Age	Frequency of Procedure	Age
Diagnostic X-rays	○	○	○			
Radiation therapy	○	○	○			
Dental X-rays - Traditional	○	○	○			
Dental X-rays - Panoramic	○	○	○			

The following questions focus on your child's medical history before their first cancer diagnosis.

61. Did the child ever have any of the following infections?

	${q://QID987/ChoiceTextEntryValue/1}'s Infections			Age at diagnosis
	Y	N	Don't know	Age at diagnosis
Measles	○	○	○	
Chickenpox	○	○	○	
Shingles	○	○	○	
Cytomegalovirus	○	○	○	
Hepatitis	○	○	○	
Monocleosis	○	○	○	
Herpes virus	○	○	○	

62. At the time the child was born, what was your estimated total household income before taxes?

Please include income such as Medicaid, Social Security, and Unemployment payments.

a) Less than 10 Thousand Dollars per year
b) 10 to 30 Thousand Dollars
c) 30 to 50 Thousand Dollars
d) 50 to 70 Thousand Dollars
e) 70 to 90 Thousand Dollars
f) 90 to 110 Thousand Dollars
g) More than 110 Thousand Dollars
h) Don't know

63. Is there anything else you would like to share with the research team regarding your residence, occupation, exposures, or anything else addressed in this questionnaire that you feel is relevant to this study?

Please describe here: ____

Thank you for completing this questionnaire. Now that you have completed the survey, the research team will be mailing your $25 payment card to the address you provided on your postcard.

We send out the payment cards every Thursday, so you can likely expect to receive it within two weeks of this date. If you don't receive it within 2 weeks, please call the project office at 412-648-5185, and we can investigate.

Upon receipt, you will need to call a project staff member to activate your card. These instructions will be included with the card mailing.

Thank you again for your participation in this research study. Your information could be used to further other studies in this area.

1. Would you be willing to participate in follow-up studies or to give us additional information after the survey has concluded? (not including studies with specimen collections - like blood, saliva, etc.)

a) Yes
b) No
c) Don't know

2. Would you be willing to participate in follow-up studies to give us biosamples after the survey has concluded? Some examples of these may include blood sample, buccal swabs, other specimens.

a) Yes
b) No
c) Don't know

Supplementary Tables

Supplementary Table S1. Distribution of Cases by Fine Categories of Childhood Malignancies in Southwestern PA 2010-2019)

Class (most detailed)	**Frequency**	**Percent**
(a.1) Precursor cell leukemias	112	22.1
(a.2) Mature B-cell leukemias	2	.4
(b) Acute myeloid leukemias	30	5.9
(c) Chronic myeloproliferative diseases	14	2.8
(d) Myelodysplastic syndrome and other myeloproliferative diseases	5	1.0
(e) Unspecified and other specified leukemias	2	.4
(a) Hodgkin lymphomas	52	10.3
(b.1) Precursor cell lymphomas	5	1.0
(b.2) Mature B-cell lymphomas (except Burkitt lymphoma)	12	2.4
(b.3) Mature T-cell and NK-cell lymphomas	5	1.0
(c) Burkitt lymphoma	5	1.0
(d) Miscellaneous lymphoreticular neoplasms	25	4.9
(e) Unspecified lymphomas	1	.2
(a.1) Ependymomas	9	1.8
(a.2) Choroid plexus tumor	5	1.0
(b) Astrocytomas	87	17.2
(c.1) Medulloblastomas	13	2.6
(c.2) PNET	1	.2
(d.1) Oligodendrogliomas	3	.6
(d.2) Mixed and unspecified gliomas	31	6.1
(e.1) Pituitary adenomas and carcinomas	12	2.4
(e.2) Tumors of the sellar region (craniopharyngiomas)	7	1.4
(e.3) Pineal parenchymal tumors	1	.2
(e.4) Neuronal and mixed neuronal-glial tumors	20	3.9
(e.5) Meningiomas	3	.6
(f) Unspecified intracranial and intraspinal neoplasms	2	.4
(a) Osteosarcomas	18	3.6
(b) Chondrosarcomas	2	.4
(c.1) Ewing tumor and Askin tumor of bone	20	3.9
(d.2) Malignant chordomas	2	.4
(d.4) Miscellaneous malignant bone tumors	1	.2

Supplementary Table S2. Distributions of UNGD Activities Metric Within 5 Miles of Buffer Zone among Children with Any of the Four Malignancies and their County-Matched Controls by Different Time Periods of Exposure in the Birth Record-Based Analysis (n=498)

Exposure Metrics within 5 miles*	Group	Time period†	Exposed N‡	Mean	Std Dev	Minimum	Maximum	10th Pctl	25th Pctl	Median	75th Pctl	90th Pctl
Overall UNGD activities	Cases	Pregnancy (T1)	94	3.50E-5	5.8E-5	6.06E-7	4.22E-4	4.71E-6	6.31E-6	12.0E-6	3.30E-5	10.9E-5
		Postnatal (T2)	311	30.2E-5	74.3E-5	7.21E-7	79.5E-4	8.91E-6	24.0E-6	82.0E-6	21.7E-5	65.0E-5
	County-Matched Controls	Pregnancy (T1)	99	3.70E-5	8.40E-5	1.43E-7	7.60E-4	2.73E-6	5.42E-6	10.0E-6	4.5E-5	7.80E-5
		Postnatal (T2)	297	24.3E-5	67.1E-5	8.99E-7	71.6E-4	10.0E-6	28.0E-6	61.0E-6	20.2E-5	54.5E-5
Well pad construction (counts/m²)	Cases	Pregnancy (T1)	48	4.54E-6	5.90E-6	4.32E-7	2.40E-5	6.04E-7	7.91E-7	2.03E-6	5.74E-6	1.60E-5
		Postnatal (T2)	287	39.0E-6	105.0E-6	4.70E-7	125.0E-5	7.71E-7	23.1E-7	7.54E-6	28.0E-6	9.30E-5
	County-Matched Controls	Pregnancy (T1)	50	9.06E-6	22.0E-6	1.28E-7	12.8E-5	5.59E-7	7.50E-7	1.87E-6	6.57E-6	1.8E-5
		Postnatal (T2)	272	26.0E-6	55.0E-6	0.61E-7	43.6E-5	6.41E-7	16.4E-7	6.18E-6	22.0E-6	6.2E-5
Drilling (counts/m²)	Cases	Pregnancy (T1)	60	3.20E-5	5.00E-5	3.36E-8	2.88E-4	8.96E-7	2.81E-6	8.86E-6	4.50E-5	10.0E-5
		Postnatal (T2)	295	22.7E-5	64.1E-5	10.21E-8	74.8E-4	23.3E-7	9.49E-6	49.0E-6	16.2E-5	47.6E-5
	County-Matched Controls	Pregnancy (T1)	62	3.40E-5	7.00E-5	7.69E-8	5.02E-4	3.61E-7	1.58E-6	13.0E-6	3.90E-5	7.00E-5
		Postnatal (T2)	280	18.1E-5	58.7E-5	12.98E-8	65.0E-4	18.5E-7	9.37E-6	37.0E-6	12.5E-5	43.3E-5
Hydraulic fracturing (depth in m/m²)	Cases	Pregnancy (T1)	60	0.019	0.060	3.60E-5	0.445	1.83E-4	7.59E-4	3.82E-3	0.012	0.031
		Postnatal (T2)	283	0.084	0.202	4.90E-5	1.331	9.51E-4	30.9E-4	16.1E-3	0.059	0.197
	County-Matched Controls	Pregnancy (T1)	60	0.016	0.042	6.40E-5	0.309	1.31E-4	9.28E-4	3.57E-3	0.018	0.033
		Postnatal (T2)	268	0.077	0.249	7.00E-5	3.150	9.46E-4	42.2E-4	15.3E-3	0.052	0.201
Production (volume in m³/m²)	Cases	Pregnancy (T1)	88	0.787	4.64	20.0E-5	43.12	2.35E-3	0.013	0.075	0.316	0.813
		Postnatal (T2)	279	2.741	14.85	6.70E-5	190.9	6.93E-3	0.048	0.348	1.347	3.540
	County-Matched Controls	Pregnancy (T1)	88	0.302	0.857	5.58E-6	7.40	1.46E-3	0.011	0.046	0.321	0.725
		Postnatal (T2)	269	2.145	12.30	1.43E-6	154.8	9.59E-3	0.072	0.445	1.225	2.621
Summed Z score§	Cases	Pregnancy (T1)	94	2.251	4.518	-0.476	33.49	-0.075	0.082	0.681	2.249	6.944
		Postnatal (T2)	311	0.817	3.806	-1.001	25.90	-0.942	-0.819	-0.481	0.656	3.091
	County-Matched Controls	Pregnancy (T1)	99	2.569	7.219	-0.565	64.86	-0.270	0.004	0.366	2.920	5.274
		Postnatal (T2)	297	0.463	3.368	-0.999	29.02	-0.923	-0.807	-0.560	0.238	2.178
Well counts	Cases	Pregnancy (T1)	97	27.48	35.82	1.00	154.00	1.00	4.00	9.00	34.00	85.00
		Postnatal (T2)	306	39.26	46.82	1.00	296.00	2.00	7.00	21.50	59.00	103.00
	County-Matched Controls	Pregnancy (T1)	99	22.31	29.05	1.00	117.00	1.00	2.00	10.00	28.00	67.00
		Postnatal (T2)	293	37.97	47.26	1.00	333.00	2.00	6.00	18.00	58.00	101.00
IDW well counts (counts/m²)	Cases	Pregnancy (T1)	97	1.44E-6	2.44E-6	1.68E-8	1.40E-5	4.49E-8	1.08E-7	3.26E-7	1.86E-6	3.98E-6
		Postnatal (T2)	306	3.09E-6	5.74E-6	1.56E-8	4.30E-5	6.40E-8	2.02E-7	8.94E-7	3.38E-6	7.84E-6
	County-Matched Controls	Pregnancy (T1)	99	1.31E-6	2.37E-6	1.56E-8	1.40E-5	2.44E-8	6.76E-8	3.55E-7	1.23E-6	4.29E-6
		Postnatal (T2)	293	2.47E-6	4.70E-6	1.65E-8	4.40E-5	5.04E-8	18.45E-8	6.48E-7	2.81E-6	6.68E-6

* See the formulas for calculation of all metrics in Table 14a.

† The pregnancy period was defined from the conception to birth using the gestation age on the birth records whereas the postnatal period from birth to the index date, which was the date of cancer diagnosis for cases and the corresponding date for the matched controls.

‡ The difference between total N and Exposed N was the number of subjects with non-exposure (not shown).

§ calculated as $\sum_{ij}^{k}\frac{x_{ij}-\mu_{.j}}{\sigma_{.j}}$; where i is for subject; j, specific phases of UNGD activities (=k); x, individual measurement of UNGD activity; μ, mean; and σ, standard deviation.

Supplementary Table S3. Distributions of Sociodemographic Characteristics of Childhood Cancer Cases Using Birth Record Information: 213 County-Matched Case-Control pairs

Sociodemographic Characteristic	Cases (N=213)		County-Matched Controls (N=213)	
	Frequency	*Percent*	*Frequency*	*Percent*
Sex at Birth				
Female	99	46.5	99	46.5
Male	114	53.5	114	53.5
Maternal Age (years)				
<20	7	3.3	7	3.3
20-24	25	11.7	24	11.3
25-29	54	25.4	60	28.2
30-34	74	34.7	81	38.0
≥35	53	24.9	41	19.2
Maternal Race				
White	209	98.1	209	98.1
Black	2	0.9	2	0.9
Other	2	0.9	2	0.9
Maternal Education				
≤ 8^{th} Grade	0	0.0	1	0.5
Some High School	10	4.7	10	4.7
High School Diploma	50	23.5	30	14.1
Some College	43	20.2	45	21.1
College Degree or Higher	108	50.7	127	59.6
Unknown	2	0.9	0	0.0
Number of Prenatal Visits				
0-7	13	6.1	16	7.5
8-12	106	49.8	111	52.1
13-16	77	36.1	77	36.1
≥17	10	4.7	5	2.4
Unknown	7	3.3	4	1.9
Birth weight				
≤2500 g	12	5.6	10	4.7
2501- 4000 g	173	81.2	180	84.5
>4000 g	28	13.2	22	10.3
Unknown	0	0.0	1	0.5
Smoking during pregnancy				
Never	184	86.4	192	90.1
Ever	25	11.7	20	9.4
Unknown	4	1.9	1	0.5
Gestation in weeks				
Mean (S.D.)	38.9(1.66)		38.7(2.02)	

Supplementary Table S4. Descriptives of Residential History Characteristics for Cases and County-Matched Controls

Variable	**Cases (N=213) ***		**County Matched Controls (N=213) ****	
	Frequency	*Percent*	*Frequency*	*Percent*
Pre-1970s Housing				
Ever	71	58.2	102	62.6
Never	51	41.8	61	37.4
Missing/dk	27		8	
Item not presented	64		42	
Residence Exterminated				
Ever	19	15.2	26	17.8
Never	106	84.8	120	82.2
Missing/dk	24		25	
Item not presented	64		42	
Pesticide/Herbicide Used in Yard				
Ever	54	45.0	82	55.4
Never	66	55.0	66	44.6
Missing/dk	29		23	
Item not presented	64		42	
Water Tested				
Ever	26	23.4	29	27.6
Never	85	76.6	76	72.4
Missing/dk	38		46	
Item not presented	64		42	
Radon Tested				
Ever	66	58.4	75	63.0
Never	47	41.6	44	37.0
Missing/dk	36		52	
Item not presented	64		42	
Radon Remediation				
Ever	26	22.2	25	19.5
Never	91	77.8	103	80.5
Missing/dk	32		43	
Item not presented	64		42	

*Out of 213 cases, a total of 149 cases had the opportunity to respond to surveys with a complete survey/residential history, 64 additional participants answered the short residential questionnaire without these items

**Out of 213 county-matched controls, a total of 171 county-matched controls had the opportunity to respond to surveys with a complete residential history, 42 filled out the short residential questionnaire without these items

Supplementary Table S4 Continued. Residential History Characteristics for Cases and County-Matched Controls

Variable	**Cases (N=213) ***		**County-matched Controls (N=213) ****	
	Frequency	*Percent*	*Frequency*	*Percent*
Attached Garage				
Ever	80	62.5	85	49.7
Never	48	37.5	86	50.3
Missing/dk	21		0	
Item not presented	64		42	
Well Water at Home				
Ever	20	14.8	18	10.4
Never	109	85.2	155	89.6
Missing/dk	20			
Item not presented	64		42	
[1]**Perception – Residence within 1 mile of Industrial Facility**				
Ever	36	25.0	46	30.1
Never	108	75.0	107	69.9
Missing/dk	5		18	
[1]**Perception – Residence within 1 mile of Farm**				
Ever	40	27.6	37	25.9
Never	105	72.4	106	74.1
Missing/dk	4		28	
[1]**Perception – Residence within 1 mile of Oil and Gas Industry**				
Ever	23	17.4	23	18.1
Never	109	82.6	104	81.9
Missing/dk	15		44	

*Out of 213 cases, a total of 149 cases had the opportunity to respond to surveys with a complete survey/residential history, 64 additional participants answered the short residential questionnaire only
**Out of 213 county-matched controls, a total of 171 county-matched controls had the opportunity to respond to surveys with a complete residential history, 42 filled out the short residential questionnaire only

1 item presented to all 213 cases and control survey respondents

Supplementary Table S5. Total overall unconventional natural gas drilling (UNGD) activities and risk of four childhood/adolescent 4 malignances combined during two exposure periods in Southwestern Pennsylvania 2010-2019

Overall UNGD activities by exposure period	**Survey-based Study with County-matched Controls (213 case-control pairs)**			**Birth Record-based Study with County-matched Controls (498 case-control pairs)**		
	Controls	***Cases***	**OR (95% CI)†**	***Controls***	***Cases***	**OR (95% CI)†**
T1: During Mother's Pregnancy						
Non-exposed	172	174	1.00	399	404	1.00
Exposed*	41	39	0.76 (0.30-1.89)	99	94	0.82 (0.47-1.41)
By buffer zone						
Non-exposed	172	174	1.00	399	404	1.00
(2-5] miles	26	30	0.80 (0.32-2.03)	64	63	0.84 (0.48-1.46)
(1-2] miles	6	6	0.46 (0.08-2.47)	24	22	0.72 (0.31-1.67)
(0.5-1] miles	9	3	0.16 (0.02-1.08)	9	7	0.65 (0.19-2.26)
[0-0.5] miles				2	2	0.81 (0.05-14.62)
P trend‡			*0.0643*			*0.3817*
By overall UNGD activities within 5 miles						
Non-exposed	172	174	1.00	399	404	1.00
Lowest (1st) quartile	10	14	1.17 (0.37-3.68)	24	17	0.63 (0.29-1.34)
Low-middle (2nd) quartile	10	8	0.51 (0.11-2.36)	25	22	0.77 (0.37-1.64)
High-middle (3rd) quartile	10	12	0.72 (0.20-2.58)	25	36	1.40 (0.63-3.14)
Highest (4th) quartile	11	5	0.26 (0.05-1.29)	25	19	0.75 (0.31-1.83)
P trend‡			*0.1443*			*0.7587*

* Exposed were individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls); non-exposed otherwise.

† All odds ratios (ORs) and their 95% confidence intervals (CIs) for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and following variables including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and superfund site (no, yes).

‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.

§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.

Supplementary Table S5 Continued. Total overall unconventional natural gas drilling (UNGD) activities and risk of four childhood/adolescent 4 malignances combined during two exposure periods in Southwestern Pennsylvania 2010-2019

Overall UNGD activities by exposure period	Survey-based Study with County-matched Controls (213 case-control pairs)			Birth Record-based Study with County-matched Controls (498 case-control pairs)		
	Controls	***Cases***	**OR (95% CI)†**	***Controls***	***Cases***	**OR (95% CI)†**
T2: From Birth to Index Date§						
Non-exposed	84	74	1.00	201	187	1.00
Exposed*	129	139	1.48 (0.88-2.5)	297	311	1.24 (0.87-1.78)
By buffer zone						
Non-exposed	84	74	1.00	201	187	1.00
(2-5] miles	72	75	1.43 (0.83-2.46)	178	170	1.18 (0.82-1.71)
(1-2] miles	24	38	2.09 (0.97-4.49)	72	77	1.49 (0.89-2.51)
(0.5-1] miles	21	14	0.82 (0.32-2.11)	37	38	1.61 (0.85-3.03)
[0-0.5] miles	12	12	1.47 (0.56-3.86)	10	26	**3.94 (1.66-9.39)**
P trend‡			*0.6289*			***0.0041***
By overall UNGD activities within 5 miles						
Non-exposed	84	74	1.00	201	187	1.00
Lowest (1st) quartile	32	48	**2.24 (1.14-4.41)**	74	86	1.40 (0.91-2.14)
Low-middle (2nd) quartile	32	16	0.70 (0.33-1.49)	74	50	0.76 (0.46-1.25)
High-middle (3rd) quartile	32	39	1.55 (0.79-3.04)	74	88	**1.69 (1.01-2.82)**
Highest (4th) quartile	33	36	1.40 (0.61-3.21)	75	87	1.79 (1.00-3.19)
P trend‡			*0.4496*			*0.0975*
By overall UNGD activities within 2 miles**						
Non-exposed	84	74	1.00	201	187	1.00
Lowest (1st) quartile	14	17	1.84 (0.74-4.61)	29	37	1.74 (0.93-3.27)
Low-middle (2nd) quartile	14	23	2.07 (0.84-5.08)	30	32	1.48 (0.77-2.84)
High-middle (3rd) quartile	14	9	0.72 (0.25-2.11)	30	30	1.41 (0.72-2.77)
Highest (4th) quartile	15	15	1.87 (0.66-5.3)	30	42	**2.16 (1.10-4.25)**
P trend‡			*0.4837*			***0.0321***

* Exposed were individuals who lived within 5 miles of any UNGD activity during mother's pregnancy (T1) or from birth to the index date (i.e., date of cancer diagnosis for cases or the same date for matched controls); non-exposed otherwise.

† All odds ratios (ORs) and their 95% confidence intervals (CIs) for different buffer zones or levels of exposures against non-exposed group were derived from unconditional logistic regression models with adjustment for matching factors (age, sex, race, and county of residence) and following variables including maternal age at childbirth (years), maternal education level, maternal smoking status at childbirth (no, yes), gestation age (weeks), birthweight (g), toxics release inventory (TRI) (no, yes), uranium mill tailings remedial action sites {UMTRA} (no, yes), and superfund site (no, yes).

‡ The same unconditional logistic models were used for linear trend test for the exposure variable in ordinal values (1, 2 for high or low) that also included non-exposed individuals (coded as 0) to maintain the case-control matched pairs.

§ The index date was the date of malignancy diagnosis for cases and the same corresponding date for matched controls.

** The same data for those with UNGD exposure within 2-5 mile of buffer zone were included in this modelling but not presented repeatedly.