Assessment of Pediatric Cancer and Its Relationship to Environmental Contaminants: An Ecological Study in Idaho

Naveen Joseph and Alan S. Kolok

1Idaho Water Resources Research Institute, University of Idaho, Moscow, ID, USA

Abstract The primary aim of this study was to determine the degree to which a multivariable principal component model based on several potentially carcinogenic metals and pesticides could explain the county-level pediatric cancer rates across Idaho. We contend that human exposure to environmental contaminants is one of the reasons for increased pediatric cancer incidence in the United States. Although several studies have been conducted to determine the relationship between environmental contaminants and carcinogenesis among children, research gaps exist in developing a meaningful association between them. For this study, pediatric cancer data was provided by the Cancer Data Registry of Idaho, concentrations of metals and metalloids in groundwater were collected from the Idaho Department of Water Resources, and pesticide use data were collected from the United States Geological Survey. Most environmental variables were significantly intercorrelated at an adjusted P-value <0.01 (97 out of 153 comparisons). Hence, a principal component analysis was employed to summarize those variables to a smaller number of components. An environmental burden index (EBI) was constructed using these principal components, which categorized the environmental burden profiles of counties into low, medium, and high. EBI was significantly associated with pediatric cancer incidence (P-value <0.05). The rate ratio of high EBI profile to low EBI profile for pediatric cancer incidence was estimated as 1.196, with lower and upper confidence intervals of 1.061 and 1.348, respectively. A model was also developed in the study using EBI to estimate the county-level pediatric cancer incidence in Idaho (Nash-Sutcliffe Efficiency = 0.97).

Plain Language Summary Previous studies focused on human exposure to environmental contaminants and their potential associations to cancer incidence among children. However, most do not consider how several environmental pollutants may act together, potentially leading to human health outcomes. The study aimed to understand how several environmental contaminants, such as agricultural pesticides, metals, etc., together are associated with cancer incidence among children. The Cancer Data Registry of Idaho provided the data on cancer incidence among children, information on metals was provided by the Idaho Department of Water Resources, and pesticides from the United States Geographic Survey. This study developed a new measure of Environmental Burden for each county in Idaho. The study identified that the counties with high Environmental Burden were more closely associated with cancer incidence among children than counties with low Environmental Burden.

1. Introduction

Several studies have analyzed the geospatial distribution of pediatric cancer globally (Corley et al., 2018; Farazi et al., 2018; Goujon et al., 2018; Lawson & Rotejanaprasert, 2014; Ramis et al., 2015; Thompson et al., 2007; Wheeler, 2007) and its potential associations to several environmental agents (Filippini et al., 2019; Konstantinidis et al., 2020; Little et al., 2018; Wakeford, 2013; Zachek et al., 2015). For instance, geospatial studies on pediatric cancer have been undertaken in the U.S (Siegel et al., 2018), pediatric leukemia in Ohio (Wheeler, 2007), pediatric leukemia and Hodgkin lymphoma in Texas (Thompson et al., 2007), childhood brain cancer in Florida (Lawson & Rotejanaprasert, 2014). On the other hand, studies focused on associations between pediatric cancer and environmental agents do not model how multiple potential cancer initiators may act together (Morett et al., 2017). Rather than focusing on an individual contaminant and its relationship to pediatric cancer, there is a need to shift towards the cumulative assessment of contaminants for addressing the combined exposure to multiple chemicals (Evans et al., 2019; Meek et al., 2011; Morett et al., 2017; NRC, 2009; SCHER, 2013; Sexton, 2015).
While using a multivariable approach, statistical techniques, such as principal component analysis (PCA), may explain the association between the cancer data and multiple environmental agents. For instance, Lobdell et al. (2014) used PCA to analyze the cumulative impact of many environmental agents and identified that PCA is an effective tool to summarize the environmental quality variables to assess the relationship to adverse health outcomes. However, to our knowledge, there are no studies that used PCA to explain the association of environmental agents to pediatric cancer. On the other hand, previous studies have used PCA to study several factors associated with cancer. These include: (a) modeling risk-based biomarkers (Shigemizu et al., 2019), (b) gene expression profiling (D.-T. Chen et al., 2011; Hsu et al., 2014; Pomeroy et al., 2002; Price et al., 2006), (c) reducing data dimensionality when classifying patients with tumors into several sub-groups (Khan et al., 2001), (d) diagnosing tissue samples as normal or neoplastic based on diffuse reflectance spectra (Skala et al., 2007), (e) reducing multiple biomarkers to generate a new optimized biomarker panel for lung cancer detection (Flores-Fernández et al., 2012), and (f) comparing the metabolic profiles from human rectal cancer biopsies and colorectal xenografts (Seierstad et al., 2008). However, while the approach has proven successful in the studies listed above, PCA to model cancer incidence in conjunction with multiple environmental agents has yet to be performed.

The associations between pesticides and pediatric cancer have received considerable attention in the past (M. Chen et al., 2015; Daniels et al., 1997; Hernández & Menéndez, 2016; LaFiura et al., 2007; Leiss & Savitz, 1995; Pogoda & Preston-Martin, 1997; Van Maelle-Fabry et al., 2017; Wigle et al., 2008; Zahm & Ward, 1998). For instance, studies have identified associations between parental, prenatal, and childhood exposures to pesticides and pediatric leukemia in residential and occupational settings (Reynolds et al., 2003; Scélo et al., 2009; Steffen et al., 2004; Vinceti et al., 2012). Supporting this, global-scale studies have explored the association between environmental agents such as pesticides and pediatric cancer and have identified that acute myeloid leukemia was associated with paternal exposure to herbicides and insecticides (Patel et al., 2020). Similarly, maternal exposure to herbicides such as diuron and propanil, along with exposure to insecticides such as phosmet, were found to have a close association with an elevated risk for leukemia (Park et al., 2020). In addition, the Park et al. (2020) study implicated 2,6-dinitroanilines and anilides relative to elevated risk. Furthermore, M. Chen et al. (2015) identified a significant association of insecticides to childhood leukemia and lymphoma and an association between herbicides and childhood leukemia.

Additionally, several studies have identified a significant association between metals and pediatric cancer incidence, including the associations between arsenic and pediatric cancer incidence (Abernathy et al., 2003; Heck et al., 2014; Hong et al., 2014), cadmium and pediatric cancer (Sherief et al., 2015), and chromium and pediatric cancer (Beaumont et al., 2008; Halasova et al., 2009; Yang & Massey, 2019; Zhitkovich, 2011). Relative to drinking water, Drozd et al. (2015) identified significant relationships between childhood thyroid cancer and the nitrates in drinking water. Barrett et al. (1998) also identified a similar relationship between pediatric brain cancer and nitrates. However, there were no supportive studies to confirm the claim.

The study’s objective was to determine the degree to which a multivariable principal component model based on several potentially carcinogenic metals and pesticides explains county-level rates of pediatric cancer in Idaho. Although the pediatric cancer incidence in Idaho has increased at a rate of 0.6% per year from 1975 to 2018 (Johnson et al., 2020), few studies have focused on environmental carcinogens in the state and their association with pediatric cancer incidence. The current study compiled pediatric cancer data from the Cancer Data Registry of Idaho and compared cancer incidence rates to concentrations of metals in groundwater and the mass of specific pesticides used across the state on a county basis.

2. Materials and Methods

2.1. Pediatric Cancer Data

Pediatric cancer data was provided under a data-sharing agreement with the Cancer Data Registry of Idaho for age at diagnosis 0–4 years, 5–9 years, 10–14 years, and 15–19 years for the period 1990–2015 at county resolution. The pediatric cancer incidence by county was estimated using Equation 1, after summing up all the pediatric cancer cases from 1990 to 2015. The county population was calculated by summing the pediatric population using the census data from 1990 to 2015 (Census-Bureau, 2020).

\[
\text{Pediatric cancer incidence} = 100,000 \times \frac{\text{No: of pediatric incidence per county}}{\text{County population}}
\] (1)
The age stratum-specific pediatric cancer incidence was also estimated for ages 0–4, 5–9, 10–14, and 15–19 ages using Equation 1, and its spatial distribution is shown in Figure 1. A visual inspection of the maps suggests that the incidence is highest within the 0–4 age stratum, followed by 15–19, 10–14, and 5–9 age strata, respectively. The pediatric cancer incidence also varies spatially, with counties in the southwestern, southeastern, and northwestern regions of Idaho having higher incidence rates compared to other counties.

2.2. Environmental Data

The environmental datasets considered for use in this study included 198 groundwater quality variables (mg/L) from 3,548 well locations, sampled between 1990 and 2015, as collected by the Idaho Department of Water Resources (IDWR). Data from each wellhead location were averaged across the entire 25-year period. The groundwater well locations are shown in Figure 2a. The groundwater variables were then mapped to county
Table 1
The Thirty Groundwater Variables (mg/L) Chosen for Subsequent Data Analysis After Data Filtering Eliminated Variables That had Considerable Gaps in the Overall Datasets

| Alkalinity (as CaCO₃) | Chromium | Phosphate |
|-----------------------|----------|----------|
| Ammonia               | Copper   | Orthophosphate |
| Arsenic               | Ethylbenzene | Potassium |
| Barium                | Fluoride | Selenium |
| Boron                 | Iron     | Silica |
| Bicarbonate           | Lead     | Silver |
| Bromine               | Magnesium| Sodium |
| Cadmium               | Manganese| Sulphate |
| Calcium               | Nitrate  | Toluene |
| Chloride              | Nitrite  | Zinc |

Note. Data Source: IDWR–Environmental data management system

Metam (Na) and Metam (K) were combined into a single variable metam; likewise, data on the pesticides metochlor and metolachlor-S were combined into a single variable metolachlor. It should be noted that the pesticide data was the usage data by county (kg/acre), while the water quality data was concentration data (mg/L).

2.2.1. Environmental Carcinogen Identification

A total of 30 groundwater contaminant concentrations (mg/L) (Table 1) and an additional 25 pesticide use profiles (kg(county) were considered for inclusion in the PCA (Table 2). Of those variables, eighteen were listed as carcinogens by various sources (e.g., USEPA, IARC, CDC), as shown in Table 3. The other variables were removed from further analysis due to the lack of evidence of carcinogenicity.

2.3. Multivariable Statistical Analysis and Modeling

The relationships among the selected 18 environmental carcinogens were evaluated using the Spearman correlation coefficient, as shown in Table 4. The relationships which are statistically significant at adjusted p-value <0.01 are highlighted with an asterisk. The p-value was adjusted to account for multiple comparisons using the false discovery rate method (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001). Of 153 comparisons, 97 were statistically significant, illustrating a strong intercorrelation between the environmental variables.

Since many environmental carcinogens are intercorrelated, a multivariable regression model with these variables may result in over-parameterization. Hence, PCA was employed in this study to reduce input redundancy (Abdi & Williams, 2010; Wold et al., 1987). PCA is a multivariate statistical technique by which dominant patterns of a data matrix are extracted in the form of a set of loading and score vectors (Wold et al., 1987). This method helps to decrease the dimensionality of the large data set by replacing some correlated variables with a more compact set of orthogonal predictors (e.g., Joseph et al. (2019, 2020)). Each of these new predictors is defined by a linear association of the original variables (I. Jolliffe, 2002). The components are defined so that the first component corresponds to the maximum variance, the second component corresponds to the second maximum variance in the direction orthogonal to the first component, and so on (Renchler, 2002).

The input variables for PCA included pesticides (kg/acre/year) and groundwater variables (mg/L). Since the variables are in different units, the input data set was standardized by subtracting the mean and dividing by standard deviation. This standardized data was used for PCA, where it was consolidated into principal components. The generated principal components resolution. Since the data were not continuous and there were gaps in data acquisition, a data filtering process was adopted to remove the data with significant gaps. For instance, information on cobalt was only available for the years 2001–2005; consequently, it was not included in any further statistical analysis. Among the 198 groundwater quality variables, only 30 had continuous datasets with measurement for each year for at least one county in Idaho. Those 30 datasets, as listed in Table 1, are selected for the next step (Section 2.2.1). This study selected groundwater data as a proxy for the environmental exposures as the groundwater is the major water source supplying drinking water to 95% of Idaho citizens (IDEQ, 2021).

Data regarding pesticide mass (kg/acre/year) were collected from the United States Geographic Survey Pesticide National Synthesis Project (USGS-PNSP) at county resolution for 2017. While the database contains pesticide use estimates for over 500 different pesticides, this study focused on the 25 most applied conventional pesticide active ingredients in the agricultural market sector from 2008 to 2012 (as per Atwood and Paisley-Jones (2017)). Table 2 lists the selected 25 pesticides after data filtering. Figure 2b shows the spatial distribution of the total mass of pesticides per unit acre for each county in Idaho for the year 2017. Data on the pesticides
were then selected in place of the original variables for further analysis. The number of components to be retained was identified using eigenvalue spectra analysis (I. T. Jolliffe, 1986). Eigenvalue spectra is a line plot of eigenvalues of principal components sorted from largest to smallest to determine the number of components retained in the analysis (Guo et al., 2009). The number of components to be retained was identified based on the line plot's slope nearing zero. Further, the fraction of variability explained by each component was estimated. A range of 70%–90% explanation of variability was selected to finalize the number of components (I. Jolliffe, 2002).

The principal components based on the eigenvalue spectra analysis were summarized into a single environmental burden index (EBI). The counties were analyzed for the factor loading of all four components. This was performed by estimating the median of the magnitude of the selected components. The median of the magnitude value was selected as the magnitude indicates the relevance of that variable (Ouyang, 2005), and the median was used as a consolidation measure. This measure of EBI was used to categorize the counties into low, medium, and high EBI profiles. The low EBI profile corresponded to the lower tertile of EBI (≤0.33), the high EBI profile corresponded to the higher tertile (≥0.66), and the medium EBI profile corresponded to the medium tertile (>0.33, <0.66).

2.3.1. Modeling Pediatric Cancer Incidence

Mixed-effects regression with a Poisson distribution and log link was adopted to model the cancer incidence data, as the Pearson Chi-square dispersion statistic was equal to one (Loukas & Kemp, 1986). The modeling was performed using the MATLAB R2019a tool (MATLAB, 2019). The mean response variable was pediatric cancer frequency with person-years as the offset. Age was included as a fixed effect in the model. The fit method selected was Laplace, with the county as intercept. The model coefficients were evaluated using F-statistics and P-value.

The sensitivity of pediatric cancer incidence to EBI profiles was estimated by choosing different cut points for the EBI tertiles. Additionally, EBI profiles classified based on 25th percentile (4 groups) and 20th percentile (5 groups) to evaluate whether the association of EBI to pediatric cancer still holds similar in both cases. In addition

| Chemical | Type and units | References |
|----------|----------------|------------|
| Arsenic  | Groundwater (mg/L) | International Agency for Research on Cancer (IARC, 2012, 2019) |
| Cadmium  | National Cancer Institute (Huff et al., 2007; McElroy et al., 2006) |
| Chromium | The United States Environmental Protection Agency (USEPA, 2004) |
| Lead     | International Agency for Research on Cancer (IARC, 2012, 2019) |
| Nitrate  | United States Occupational Health and Safety Administration (OSHA, 2010; Steenland & Ward, 2014) |
| Nitrite  | United States Occupational Health and Safety Administration (OSHA, 2010; Steenland & Ward, 2014) |
| Silica   | United States Occupational Health and Safety Administration (OSHA, 2010; Steenland & Ward, 2014) |
| 2,4-D    | Pesticide use (kg/county/year) | International Agency for Research on Cancer (IARC, 2012, 2019) |
| Acetochlor | The United States Environmental Protection Agency (USEPA, 2004) |
| Atrazine | California Office of Environmental Health Hazard Assessment (COEHHA, 2012) |
| Chloropicrin | The United States Environmental Protection Agency (USEPA, 2004) |
| Chlorpyrifos | California Office of Environmental Health Hazard Assessment (COEHHA, 2012) |
| Di-chloropropene | Centers for Disease Control and Prevention (CDC, 2006) |
| Glyphosate | International Agency for Research on Cancer (IARC, 2012, 2019) |
| Mancozeb | The United States Environmental Protection Agency (USEPA, 2004) |
| Metalochlor | The United States Environmental Protection Agency (USEPA, 2004) |
| Metam | The United States Environmental Protection Agency (USEPA, 2004) |
| Trifluralin | The United States Environmental Protection Agency (USEPA, 2004) |

Note. The references included in the table provide evidence that these chemicals are potentially carcinogenic.
to the statistical parameters, the rate ratio of pediatric cancer incidence was also estimated, which is defined as
the ratio of the pediatric crude incidence rate of high exposure to low exposure. The final model results were
then evaluated using Nash-Sutcliffe Efficiency (NSE). NSE is a statistical measure to quantify the efficiency of
hydrological models (Nash & Sutcliffe, 1970). It measures the relative magnitude of residual variance compared
to observed data (Moriasi et al., 2007).

3. Results
3.1. Multivariable Analysis Results
Principal component analysis was performed on the data set of environmental carcinogens to avoid input redun-
dancy. A scree plot (eigenvalue spectra), representing eigenvalues sorted from largest to smallest, is shown in
Figure 3a. After the inclusion of PC5, the scree plot flattened, indicating that additional components did not
significantly increase the model’s explanatory power. Figure 3b shows the fraction of variability explained by
each component plotted against the number of components. The first component represents approximately 45%
of the total variability of the original set of variables (eleven pesticides and seven groundwater variables), the first
two components together 66%, the first three components together 79%, and the first four components together
88%. Since the first four components represent the original data set’s variability within the range of 70%–90%
and close to the upper threshold of 90%, these four components were used as an alternative to the original set of
variables.

The most dominant variables within the first principal component (PC1) were glyphosate and atrazine; the most
dominant within the second (PC2) were 2,4-D and metam; the most dominant for the third (PC3) component were
di-chloropropene and chloropicrin; fourth (PC4) were trifluralin and metam. The most dominant components
were pesticide variables, as expressed in kg/acre of application on a county-by-county scale.

| Correlation  | Metam | Glyphosate | 2,4-D | Acetochlor | Atrazine | Chloropicrin | Chlorpyrifos | Di-chloropropene |
|--------------|-------|------------|--------|------------|----------|--------------|--------------|-----------------|
| Metam        | 1.0   | 0.4*       | 0.5*   | 0.6*       | 0.9*     | 0.7*         | 0.9*         |                 |
| Glyphosate   | 1.0   | 0.6*       | 0.6*   | 0.7*       | 0.3      | 0.7*         | 0.2          |                 |
| 2,4-D        | 1.0   | 0.1        | 0.9*   | 0.5*       | 0.8*     | 0.5*         | 0.4*         |                 |
| Acetochlor   | 1.0   | 0.9*       | 0.5*   | 0.6*       | 0.9*     | 0.5*         | 0.5*         |                 |
| Atrazine     | 1.0   | 0.6*       | 1.0    | 0.6*       | 0.9*     | 0.5*         | 1.0          |                 |
| Chloropicrin |       |            |        |            |          | 0.6*         | 0.6*         |                 |
| Chlorpyrifos |       |            |        |            |          |              |              | 0.9*            |
| Di-chloropropene |   |            |        |            |          |              |              |                 |

*The correlations that are statistically significant at P-value <0.01.
3.2. Environmental Burden Index

Figure 4 shows the spatial distribution of the EBI, developed from the consolidated components, categorizing them into low, medium, and high environmental profiles in Idaho. Many of the counties in the southern Snake River region of Idaho and a few counties in the northern region of Idaho have the largest environmental burden. It is important to note that many counties falling within the high environmental burden profile have higher pediatric cancer incidence.

Figure 5 shows the box plot of pediatric cancer incidence corresponding to low, medium, and high EBI profiles. The pediatric cancer crude incidence (per 100,000) for the high EBI profile is the highest (mean = 18.97, median = 17.19), compared to medium EBI profile (mean = 18.77, median = 16.13), and low EBI profile (mean = 14.44, median = 12.22). The rate ratio for EBI high profile to EBI low profile was also estimated as 1.196 with lower and upper confidence intervals of 1.061 and 1.348, respectively. These results indicate the significant increase of pediatric cancer incidence when shifting from a low to a high EBI profile.

3.3. Model Results

A mixed-effects regression model with Poisson distribution was applied to model the pediatric cancer incidence in Idaho. The estimated coefficients of the model with standard error, t-statistic, and P-value are shown in Table 5. A type-3 test for fixed effects was also performed, and the results are shown in Table S1 in the Supporting Information S1. The results indicate that EBI was statistically significant to pediatric cancer incidence in Idaho (P-value < 0.05). The intercept variable, county, was highly significant in the model, which indicates that the inter-county differences remain even after adjusting for the environmental variables. We also conducted sensitivity tests by choosing different cut points for the EBI, as shown in Tables S2 and S3 in the Supporting Information S1. Furthermore, EBI profiles for 25 percentile and 20 percentiles were also estimated as shown in Tables S4 and S5 in the Supporting Information S1, respectively. The results were similar in the model runs, confirming the statistical significance of the EBI profile in all the cases.
Figure 6 shows the modeled and observed pediatric cancer incidence for the ages 0–4, 5–9, 10–14, and 15–19. The modeled results match the observed data with NSE of 0.989, 0.974, 0.961, and 0.92 for ages 0–4, 5–9, 10–14, and 15–19, respectively. The modeled against observed pediatric cancer incidence for all ages is also shown in Figure S1 in the Supporting Information (NSE = 0.97).

4. Discussion

The objective of this study was to determine the degree to which a multivariable principal component model based on several potentially carcinogenic metals and pesticides can be used to explain county-level pediatric cancer in Idaho. To our knowledge, this is the first study that has pediatric cancer incidence data and then related it to water quality parameters and pesticide application rates (kg/county) across the state. Using a multivariable analysis approach, this study consolidated several environmental variables into four consolidated principal components and developed an EBI at the county scale. Based on EBI, the counties in Idaho were classified into low, medium, and high environmental burden profiles. The findings indicated that EBI was significantly associated with pediatric cancer incidence at $P$-value <0.05. The variables predominantly contributing to the environmental burden index were pesticides. A statistical model was also developed which accurately estimated the pediatric cancer incidence in Idaho (NSE = 0.97). The findings from the study indicate that the consolidated analysis of environmental variables provides insights into the geospatial distribution of pediatric cancer incidence in Idaho.

4.1. Use of PCA in Cancer Research

PCA has been used in cancer research to analyze high-dimensional data and extract relevant features for various applications (Abdelrehim et al., 2018; Adiwijaya et al., 2018; D.-T. Chen et al., 2011; De Stefani et al., 2010; Hsu et al., 2014; Price et al., 2006; Silvera et al., 2011). For example, PCA was used in discoveries involving
genomic profiling and personalized medicine (Hunter et al., 2007; Pomeroy et al., 2002; Price et al., 2006; Seierstad et al., 2008). It has also been used by Brown et al. (2013) to analyze the relationships between various adverse childhood experiences, such as physical, emotional, or sexual abuse, family/household member's mental illness, incarceration, and substance abuse, and cancer incidences. Silvera et al. (2011) used PCA to evaluate the impact of various dietary and lifestyle patterns (food habits, tobacco and alcohol usage, occupational history, use of medications, etc.) on cancer incidence and identified that nitrate/nitrite intake is associated with an elevated risk of cancer incidence.

4.2. Importance of Pesticides as Predictors of Pediatric Cancer

While conducted in a more traditional manner, several published studies have demonstrated a link between pesticides and pediatric cancer (M. Chen et al., 2015; Coste et al., 2020; Daniels et al., 1997; Hernández & Menéndez, 2016; Iqbal et al., 2021; Karalexi et al., 2021; LaFuria et al., 2007; Leiss & Savitz, 1995; Pascale & Laborde, 2020; Soldin et al., 2009; Van Maelle-Fabry et al., 2017). In this current study, several pesticides were identified as major contributors to the consolidated principal components. Among these pesticides, metam and chloropicrin are considered high-ranking pesticides relative to cancer risk (Gunier et al., 2001). Supporting this, Reynolds et al. (2005) found that exposure to metam was linked to a two-fold increase in the odds ratio for pediatric leukemia. Evidence regarding a link between the pesticides such as glyphosate (Andreotti et al., 2018; A. J. De Roos et al., 2005; Mink et al., 2012), atrazine (Freeman et al., 2011; Rhoades et al., 2013; Rusiecki et al., 2004), and trifluralin (A. De Roos et al., 2003; Kang et al., 2008) are mixed, or in the case of di-chloropropene much less compelling (Lee et al., 2002) than that for metam and chloropicrin.

This study used pesticide information from USGS-PNSP at county resolution for evaluating its association to pediatric cancer incidence in Idaho. Previous...
studies focusing on pesticides and pediatric cancer collected data through a parental interview (e.g., Gunier et al. (2017)). One of the limitations of survey-based approaches is the lack of statistical power to investigate associations with individual pesticide active ingredients, which was overcome by this study to an extent. Although this study did not focus on specific cancer sites such as leukemia or lymphoma, age at diagnosis was a parameter considered. The embryonal tumors (neuroblastoma, nephroblastoma, and retinoblastoma) are generally diagnosed in very early childhood prior to age 5. Acute lymphoblastic leukemia (ALL) and Wilms tumor diagnoses peaks ages 3–4, and it is lymphomas and germ cell tumors that rise in adolescence (Howell et al., 2007; Withrow et al., 2019). Thus, the stratified analyses by age are a proxy into what cancer types might be most associated with pesticides.

4.3. Comparison of Consolidated Analysis in This Study to Existing Environmental Burden Indices

More to the point of this current research, Scott et al. (2019) developed and used an environmental burden index focusing on socioeconomic and demographic variables including poverty, income, food access, and race. The results of this index were modeled and observed pediatric cancer incidence for ages 0–4, 5–9, 10–14, and 15–19. The axes correspond to the 44 counties in Idaho. Table 5 displays the estimated coefficients of the model for total pediatric cancer incidence with standard error, t-statistic, P-value, and confidence intervals.

Table 5: Estimated Coefficients of the Model for Total Pediatric Cancer Incidence With Standard Error, t-statistic, P-value, and Confidence Intervals

| Name               | Estimate | SE     | t-stat | P-value | Lower C.I. | Upper C.I. |
|--------------------|----------|--------|--------|---------|------------|------------|
| Intercept          | -8.5901  | 0.065776 | -130.6 | 2.01E−172 | -8.7199    | -8.4602    |
| Age Group 5–9      | -0.42718 | 0.065964 | -6.4759 | 9.74E−10 | -0.55739   | -0.29696   |
| Age Group 10–14    | -0.42357 | 0.065367 | -6.4799 | 9.54E−10 | -0.55261   | -0.29454   |
| Age Group 15–19    | 0.07654  | 0.057326 | 1.3352  | 0.1836   | -0.03662   | 0.1897     |
| EBI Profile Medium | 0.1258   | 0.078369 | 1.6053  | 0.11029  | -0.0289    | 0.28051    |
| EBI Profile High   | 0.17925  | 0.061235 | 2.9275  | 0.003886 | 0.05838    | 0.30012    |

Note: SE, Standard error; t-stat, t-statistic; C.I., Confidence interval.
Similarly to the consolidated analysis of environmental variables performed in this study, USEPA (United States Environmental Protection Agency) has developed the Environmental Quality Index (EQI) (Lobdell et al., 2014), and Yale University has developed the Environmental Performance Index (EPI) (Emerson et al., 2012), focusing on several environmental domains. While relevant, these indices are developed at a national level and used datasets at a coarser resolution. For instance, many of the environmental datasets used in these indices are collected at state scale due to the data scarcity in spatial and temporal coverage (Messer et al., 2014). Moreover, the study was conducted for the whole U.S., and many environmental datasets were downscaled or extrapolated using multiple statistical methods, increasing the uncertainty of the modeling approach. To this point, this study has accessed datasets from the USGS-PNSP, Idaho Department of Water Resources, and correlated with pediatric cancer incidence datasets from Cancer Data Registry of Idaho across Idaho county resolution.

It is important to note that most of the variables in the consolidated principal components are pesticide use by county and not the concentration of metals/metalloids in groundwater. The groundwater data may not have been particularly explanatory because the value used was a mean concentration (mg/L) averaged over a 25-year period with no regard paid to the timing of groundwater sampling, both annually and seasonally. Given that the concentration of potentially hazardous chemicals in groundwater may vary seasonally or vary from year to year, it may not be surprising that an average value has low inter-county variability, and hence it had little explanatory power. Due to the lack of a standardized sample regime, averaging the low concentration of groundwater variables in months such as February and high concentration of groundwater variables in months such as July may bias the final concentration reported.

In contrast to the groundwater samples, the kg of pesticides used in each county across Idaho is a fairly robust measurement. While the amount of pesticide used may vary from year to year, the rank order of the counties relative to their pesticide use probably does not vary appreciably from 1 year to the next, as counties, where the landscape is dominated by row crops, will invariably use large amounts of pesticides, while counties dominated by forested areas will consistently use low amounts of pesticide. As an example of this, the Spearman Rank Correlation between the use of metam (the primary pesticide in PC1) in 2013 and 2017 for 24 Idaho counties was extremely high ($R > 0.99$). We did the analysis for 24 counties as metam was applied only on these 24 counties.

### 4.4. Limitations and Future Directions

Recent studies have explored the global associations between all cancer incidence and climate change on a global scale (Nogueira et al., 2020; Piacentini et al., 2018; Vineis et al., 2021), pediatric cancer incidence, and socioeconomic variables in the U.S. (Swensen et al., 1997), or specific cancer incidences such as lung cancer and gender differences (Hanagiri et al., 2007; Zang & Wynder, 1996). These studies are uncovering geospatial relationships between cancer incidence and other variables, but to date, little has been done with respect to the use of water quality criteria or pesticide application as they pertain to large geospatial distributions of cancer. The results of this current study illustrate that such analysis may prove to be fruitful.

This study did not consider the temporal changes in exposure measurements relative to disease onset, including lag time between exposure and disease. Part of the reason we chose pediatric cancers was because of the short latency period between exposure and onset of disease. Moreover, we are using the pesticide mass in kg/acre/county as the input variable, which is an indirect measurement of exposure at county scale. We hypothesize that the greater the use of pesticide in the county, the more probability there would be for human exposure.

One other limitation of the current study is that the differences in pediatric cancer incidence across different races such as Hispanic population, or with differences in parental age at the birth of a child are neglected. This was primarily because of the lack of a dataset available. Moreover, this study did not analyze the pediatric cancer incidence by site, instead focused on the total pediatric cancer incidence. The future scope of the study is to include these parameters after considerable data mining, and to evaluate its association to specific cancer sites such as leukemia.
This study did not analyze cancer incidences by various sites primarily due to the low sample size issue. Idaho is a low population state, with several counties being frontier counties (population density < 6/square miles). The total number of pediatric cancer incidence in Idaho across counties, categorized by age-group itself is very low. The three major pediatric cancer types in Idaho are leukemia, lymphoma, and CNS. For leukemia, 83% of the counties have less than or equal to 3 incidences when categorized by age group. The corresponding percentage for lymphoma is 89% and for CNS is 80%.

Furthermore, the results from this study depend on the spatial unit selected (county) and may vary with analyzing at a different spatial resolution, which is referred to as the modifiable areal unit problem (Openshaw, 1984). Moreover, this study did not perform an exposure-based assessment, rather correlated the mass of pesticides applied in counties of Idaho (kg/county/area/year) and water quality variables (mg/L) to pediatric cancer incidence. While this study overcame the limitations of recent studies such as Scott et al. (2019), by using potential carcinogenic pesticides and groundwater quality variables in the analysis, exposure routes were neglected. The future scope of this work is to overcome these limitations and collect additional temporal datasets on environmental variables and pediatric cancer in Idaho, and to perform exposure-based assessments. This would help comprehensively analyze pediatric cancer distribution in Idaho concerning specific environmental contaminants based on the biological plausibility between exposure and incidence.

Conflicts of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Pediatric cancer incidence data of current work is not open to the public as the health information is confidential as required by state law (Idaho Code §57-1706). Investigators seeking pediatric cancer incidence data in Idaho should directly make a request to the Cancer Data Registry of Idaho at https://www.idcancer.org/index.html. Pesticide data can be accessed from the United States Geographic Survey Pesticide National Synthesis Project (USGS-PNSP) at https://water.usgs.gov/nawqa/pnsp/usage/maps/compound_listing.php. Groundwater quality data in Idaho can be accessed from the Idaho Department of Water Resources at https://idwr.idaho.gov/water-data.

References

Abdelrahim, M. G., Malfouh, E. M., Ewis, A. A., Seedhom, A. E., Alifii, H. M., & Shebl, F. M. (2018). Dietary factors associated with pancreatic cancer risk in Minia, Egypt: Principal component analysis. Asian Pacific Journal of Cancer Prevention: Asian Pacific Journal of Cancer Prevention, 19(2), 449.

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley interdisciplinary reviews: Computational Statistics, 2(4), 433–459. https://doi.org/10.1002/wics.101

Abernathy, C. O., Thomas, D. J., & Calderon, R. L. (2003). Health effects and risk assessment of arsenic. Journal of Nutrition, 133(S5), 1536S–1538S. https://doi.org/10.1093/jn/133.5.1536s

Adiwijaya, W. U., Lisnawati, E., Aditsania, A., & Kusumo, D. S. (2018). Dimensionality reduction using principal component analysis for cancer detection based on microarray data classification. Journal of Computer Science, 14(11), 1521–1530. https://doi.org/10.3844/jcssp.2018.1521.1530

Andretti, G., Koutros, S., Hofmann, J. N., Sandler, D. P., Lubin, J. H., Lynch, C. F., et al. (2018). Gluphosate use and cancer incidence in the agricultural health study. JNCI: Journal of the National Cancer Institute, 110(5), 509–516. https://doi.org/10.1093/jnci/djx233

Atwood, D., & Paisley-Jones, C. (2017). Pesticides industry sales and usage: 2008–2012 market estimates. US Environmental Protection Agency. Barrett, J. H., Parslow, R. C., McKinney, P. A., Law, G. R., & Forman, D. (1998). Nitrate in drinking water and the incidence of gastric, esophageal, and brain cancer in Yorkshire, England. Cancer Causes & Control, 9(2), 153–159. https://doi.org/10.1023/a:1000878126535

Beaumont, J. J., Sedman, R. M., Reynolds, S. D., Sherman, C. D., Li, L.-H., Howd, R. A., et al. (2008). Cancer mortality in a Chinese population exposed to hexavalent chromium in drinking water. Epidemiology, 12–23. https://doi.org/10.1097/ede.0b013e31815ceac4c

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B, 57(1), 289–300. https://doi.org/10.1111/j.2517-6616.1995.tb02031.x

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Annals of Statistics, 1165–1188. https://doi.org/10.1214/aos/1069326564

Brown, M. J., Thacker, L. R., & Cohen, S. A. (2013). Association between adverse childhood experiences and diagnosis of cancer. PLoS One, 8(6), e65524. https://doi.org/10.1371/journal.pone.0065524

CDC. (2006). Centers for Disease Control Prevention, United States Department of Health and Human Services. National Institutes of Health. Treatment of opportunistic infections.
Census-Bureau. (2020). U.S. Population data - 1969 - 2019: County-level population files - 19 age groups. From surveillance, Epidemiology, and End results program. National Cancer Institute.

Chen, D.-T., Hsu, Y.-L., Fulp, W. J., Coppola, D., Haura, E. B., Yeatman, T. J., & Cress, W. D. (2011). Prognostic and predictive value of a malignancy-risk gene signature in early-stage non-small cell lung cancer. Journal of the National Cancer Institute, 103(24), 1859–1870. https://doi.org/10.1093/jnci/djr220

Chen, M., Chang, C.-H., Tao, L., & Lu, C. (2015). Residential exposure to pesticide during childhood and childhood cancers: A meta-analysis. Pediatrics, 136(4), 719–729. https://doi.org/10.1542/peds.2015-0006

COEHHA (2012). The proposition 65 list. California Office of Environmental Health Hazard Assessment. Retrieved from https://oehha.ca.gov/proposition-65/proposition-65-list

Corley, B., Bartelt-Hunt, S., Rogan, E., Coulter, D., Sparks, J., Baccaglini, L., et al. (2018). Using watershed boundaries to map adverse health outcomes: Examples from Nebraska. USA. Environmental Health Insights, 12, 1178630217751906. https://doi.org/10.1177/1178630217751906

Coste, A., Bailey, H. D., Kartal-Kaess, M., Renella, R., Berthet, A., & Spycher, B. D. (2020). Parental occupational exposure to pesticides and risk of childhood cancer in Switzerland: A census-based cohort study. BMC Cancer, 20(1), 1–11. https://doi.org/10.1186/s12885-020-07319-w

Daniels, J. L., Olshan, A. F., & Savitz, D. A. (1997). Pesticides and childhood cancers. Environmental Health Perspectives, 105(Suppl 10), 1068–1077. https://doi.org/10.1289/ehp.971051068

De Roos, A., Zahn, S., Cantor, K., Weisenburger, D., Holmes, F., Burmeister, L., & Blair, A. (2003). Integrative assessment of multiple pesticides as risk factors for non-Hodgkin’s lymphoma among men. Occupational and Environmental Medicine, 60(9), e1. https://doi.org/10.1136/ oem.60.9.e11

De Roos, A. J., Blair, A., Rusiecki, J. A., Hoppin, J. A., Svec, M., Dosemeci, M., et al. (2005). Cancer incidence among glyphosate-exposed pesticide applicators in the Agricultural Health Study. Environmental Health Perspectives, 113(1), 49–54. https://doi.org/10.1289/ehp.7340

De Stefani, E., Ronco, A. L., Deneo-Pellegrini, H., Bogetto, F., Aune, D., Acosta, G., et al. (2010). Dietary patterns and risk of advanced prostate cancer: A principal component analysis in Uruguay. International Journal of Hygiene and Environmental Health, 213(4), 397–404. https://doi.org/10.1016/j.ijheh.2010.05.007

Elgo, J., Aran, A., Rovira, J., Espinosa, J. A., Piñol, M., Mateo, M. A., et al. (2008). Reduction of cigarette smoke-related indoor air pollution in a medical center: A randomized controlled trial. International Journal of Hygiene and Environmental Health, 211(1), 48–53. https://doi.org/10.1016/j.ijheh.2007.03.001

Fulghesu, G., Gugliuzza, F., Costas, C., Nardone, A., D’Aquila, G., & Massara, E. (2018). Occupational exposure to pesticides and risk of cancer: A review of the literature. Expert Systems with Applications, 92, 38–44. https://doi.org/10.1016/j.eswa.2018.03.008

Gunier, R. B., Harnly, M. E., Reynolds, P., Hertz, A., & Von Behren, J. (2001). Agricultural pesticide use in California: Pesticide prioritization, use densities, and population distributions for a childhood cancer study. Environmental Health Perspectives, 109(10), 1071–1078. https://doi.org/10.1289/ehp.011091071

Heck, J. E., Park, A. S., Qiu, J., Cockburn, M., & Ritz, B. (2014). Risk of leukemia in relation to exposure to ambient air toxics in pregnancy and nancy-risk gene signature in early-stage non–small cell lung cancer. Journal of Molecular Sciences, 5(9), e02314. https://doi.org/10.3390/ijms17040461

Huff, J., Lunn, R. M., Waalkes, M. P., Tomatis, L., & Infante, P. F. (2007). Cadmium-induced cancers in animals and in humans. International Journal of Occupational and Environmental Health, 13(2), 202–212. https://doi.org/10.1179/oeh.2007.13.2.202

Hunter, D. J., Kraft, P., Jacobs, K. B., Cox, D. G., Yeager, M., Hankinson, S. E., et al. (2007). A genome-wide association study identifies alleles in FGF22 associated with risk of sporadic postmenopausal breast cancer. Nature Genetics, 39(7), 870–874. https://doi.org/10.1038/ng2075

IARC. (2012). Agents classified by the international agency for research on cancer for monographs (Vol. 1–106). Retrieved from http://monographs. iarc.fr/ENG/Classification/index.php

IARC. (2019). International Agency for Research on Cancer monographs on the identification of carcinogenic hazards to humans. World Health Organization.
Withrow, D. R., de Gonzalez, A. B., Lam, C. J., Warren, K. E., & Shiels, M. S. (2019). Trends in pediatric central nervous system tumor incidence in the United States, 1998–2013. *Cancer Epidemiology and Prevention Biomarkers*, 28(3), 522–530. https://doi.org/10.1158/1055-9965.epi-18-0784

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1–3), 37–52. https://doi.org/10.1016/0169-7439(87)80084-9

Yang, F., & Massey, I. Y. (2019). Exposure routes and health effects of heavy metals on children. *Biometals*, 32(4), 563–573.

Zachek, C. M., Miller, M. D., Hsu, C., Schiffman, J. D., Sallan, S., Metayer, C., & Dahl, G. V. (2015). Children's cancer and environmental exposures: Professional attitudes and practices. *Journal of pediatric hematology/oncology*, 37(7), 491. https://doi.org/10.1097/mph.0000000000000316

Zahn, S. H., & Ward, M. H. (1998). Pesticides and childhood cancer. *Environmental Health Perspectives*, 106(suppl 3), 893–908. https://doi.org/10.2307/3434207

Zang, E. A., & Wynder, E. L. (1996). Differences in lung cancer risk between men and women: Examination of the evidence. *JNCI: Journal of the National Cancer Institute*, 88(3–4), 183–192. https://doi.org/10.1093/jnci/88.3-4.183

Zhukovich, A. (2011). Chromium in drinking water: Sources, metabolism, and cancer risks. *Chemical Research in Toxicology*, 24(10), 1617–1629. https://doi.org/10.1021/tx200251t