Original Research Paper

Analysis of Socio-Demographic, Pollution, And Hazard Risk Factors Affecting Life Expectancy In West Virginia: A Multilevel Regression

Sagar Sharma Timilsina[1], Tilak Bhusal[1], Ashish Nimbarte (Ph.D.)[1], Avishek Choudhury (Ph.D.) [1*]

[1] *Industrial Management and Systems Engineering, West Virginia University, Morgantown, USA.*

Article history
Received: 2 December 2024
Revised: 6 January 2025
Accepted: 8 January 2025

**Corresponding Author*: Avishek Choudhury, West Virginia University, Morgantown, USA; Email: avishek.chudhury@mail.wvu.edu

Abstract: West Virginia, a rural Appalachian state has the second lowest life expectancy at birth in the USA. Multiple factors like demographic, socio-economic, pollution, hazard proximities, etc., are worsening the situation. We have developed multiple multilevel linear regression models with counties as the clustering variable to study the relation of different factors with life expectancy. Data was extracted from the EPA's EJScreen and P2 EJ facility mapping tool datasets. Data covering 1,639 census blocks across all 55 counties of the West Virginia state is considered. We found the model with Hazard Proximity and Exposure Risk variables had the best fit and it suggested that traffic proximities, residential lead exposure, toxic releases to discharge water, proximity to underground storage tanks, proximities to RMP facilities had a significant positive relation to lower life expectancies in West Virginia. Although not a significant parameter, proximity to Superfund sites was also positively related to lower life expectancies in West Virginia. Therefore, state- and federal-level policymakers should consider these hazard proximities and exposure risk factors while making policies related to life expectancies.

Keywords: Life expectancy; Hazard proximity; Exposure risk; Pollution; Socio-economic; Demographic.

Introduction

Global life expectancy has been on a steady upward trajectory, rising from 45.51 years in 1950 to 73.3 years in 2024, with minor periods of decline (UN Department of Economic and Social Affairs, 2024). The World Health Organization (WHO) defines life expectancy at birth as "the average number of years that a newborn could expect to live if exposed to the sex- and age-specific death rates prevailing at the time of their birth, for a specific year, in a given country, territory, or geographic area" (World Health Organization, 2024).

In the United States, the average life expectancy at birth is 77 years, surpassing the global average (Tejada-Vera et al., 2022). However, there are notable disparities among states, with Hawaii reporting the highest life expectancy at 80.7 years and Mississippi the lowest at 71.9 years. These differences are particularly pronounced in Appalachia, where life expectancy disparities compared to the rest of the U.S. have increased over time and are projected to grow further (Singh et al., 2017). Among the 13 states in Appalachia, West Virginia stands out as the only entirely rural state and the one entirely situated within the Appalachian region. With a life expectancy of 72.8 years, West Virginia ranks second lowest in the nation (Tejada-Vera et al., 2022). While literature exists on life expectancy trends across the United States, focused studies on West Virginia, which lies at the core of Appalachia, remain limited. Several socioeconomic and environmental factors can potentially influence the life expectancy of a given region. Poverty has been repeatedly associated with reduced life expectancy (Singh & Lee, 2020; Tafran et al., 2020), and West Virginia had the third-highest poverty rate in the U.S. in 2022, with 17.9% of its population living below the poverty line (DePietro, 2023).

Furthermore, research has established a positive relationship between cancer mortality and carcinogenic discharges in West Virginia, highlighting the potential role of toxic chemical exposure in reducing life expectancy (Ahern et al., 2011). Environmental factors, including pollutants from industrial, agricultural, transportation, and mining operations, also pose significant public health threats in the state (Garry et al., 2002; Hendryx & Ahern, 2008; Khuder et al., 2007; Pope III et al., 2002; Thornton et al., 2002). Additionally, socio-economic disparities exacerbate the health impacts of these environmental burdens (Northridge et al., 2003; Schulz & Northridge, 2004). Despite these findings, there is a lack of comprehensive studies that evaluate the combined effects of demographic, socio-economic, environmental, and hazard proximity factors on life expectancy in West Virginia.

This study aims to fill this gap by examining the multifaceted determinants of low life expectancy in West Virginia.

Materials and Methods

Data Sources

This study utilized two primary datasets from the United States Environmental Protection Agency (EPA): EJScreen and P2 EJ Facility Mapping Data (United States Environmental Protection Agency, 2023a; United States Environmental Protection Agency, 2023b).

The EJScreen dataset provides comprehensive demographic, ecological, and socio-economic information for census block groups across the United States. Key variables include total population, percentage of people of color, percentage of the population with low income, annual average levels of particulate matter 2.5 (PM2.5) in the air, proximity to hazardous waste, and other environmental and health-related metrics.

The P2 EJ Facility Mapping dataset focuses on industrial establishments likely contributing to local pollution levels. It incorporates environment-related data specific to these facilities, leveraging information from multiple sources, including the Toxics Release Inventory, Greenhouse Gas Reporting Program, Water Discharge Monitoring Reports, Hazardous Waste Shipment Manifests, and the Emissions Inventory System.

For this study, data from 1,639 census blocks across all 55 counties in West Virginia were analyzed.

Description and data processing

Table 1 shows the variables, abbreviations, definitions, and data sources. The dataset was initially organized at the census block level. Data processing methods were employed to aggregate variables accordingly to analyze county-level distributions. For continuous variables such as population, PM25, ozone levels, diesel particulate matter, respiratory risks, and others, the average was calculated across census blocks within each county. Similarly, averages were computed for county-level variables like TRI, RCRA, DMR, GHG, and NEI. Percentage-based variables, such as the percentage of people of color, limited English-speaking populations, low-income populations, and others, required additional processing. The raw values for each census block group were divided by their corresponding percentages, and these totals were then summed across blocks to derive county-level aggregates. Using these totals, percentages for each county were calculated. Raw values were not available in the dataset for the LOW_LIFE variable, which represents low life expectancy. Instead, averages of the existing data points at the census block level were used to calculate county-level values. The LOW_LIFE variable served as the outcome variable for this study. It is an inverse measure of life expectancy, with higher values indicating areas where life expectancy at birth is lower than national norms.

The analysis employed a multilevel modeling approach, using census blocks as the base level (Level 1) and counties as the higher level (Level 2). Independent variables were categorized accordingly. Covariates were grouped into three categories to understand the effects of different types of variables on LOW_LIFE. The first category, Demographic and Socio-Economic Variables, included predictors such as POP, CLR, LING, L_INC, and EDU. The second category, Pollution and Health Risk Variables, consisted of factors related to pollutants and health risks, such as PM2.5, PMDSL, RSEI_AIR, and NEI. The third category, Hazard Proximity, and Exposure

Risk Variables included proximity-based measures such as PTRAF, PHWF, and UST, reflecting environmental exposure risks.

Table 1. Variables used in the study

Variable	Abbreviation	Definition	Source
Outcome variable			
LIFEEXPPCT	LOW_LIFE	It is the inverse of normalized life expectancy calculated as (1- (life expectancy/ national max life expectancy). The area with lower life expectancy has higher values and vice versa	EJScreen
Grouping variable			
CNTY_NAME	CNTY	It represents the name of the counties in which the block groups lie.	EJScreen
Demographic and Socio-economic variables			
ACSTOTPOP	POP	It is the total population of the individual block group.	EJScreen
PEOPCOLORPCT	CLR	It is the percentage of all the people who identify as a race other than non-Hispanic white.	EJScreen
LINGISOPCT	LING	It is the percentage of households where no one aged 14 or above can speak English.	EJScreen
UNDER5PCT	U_5	It is the percentage of the population whose age is below 5 years	EJScreen
OVER64PCT	O_64	It is the percentage of the population whose age is above 65 years	EJScreen
LOWINCPCT	L_INC	It is the percentage of the population with income below or equal to two times the poverty level.	EJScreen
UNEMPPCT	UNEMP	It represents the percentage of the population who were unemployed.	EJScreen
LESHSPCT	EDU	It represents the percentage of the population of 25 years or older with an education status of less than high school.	EJScreen
Pollution and health Risk Variables			
PM25	PM25	It is the annual average PM 2.5 (μg/m³) concentration in air.	EJScreen
OZONE	O3	It is the annual average of the top ten daily maximum 8-hour ozone concentrations (in parts per billion) in air.	EJScreen
DSLPM	PMDSL	It is an estimated value of diesel particulate matter (μg/m³) concentration level in air.	EJScreen
RESP	RESP	It is the respiratory hazard index that measures the combined risk from the respiratory system affecting air toxics.	EJScreen
RSEI_AIR	RSEI_AIR	It is the toxicity-weighted concentrations in the air of TRI-listed chemicals modeled by the Risk-Screening Environmental Indicators (RSEI) model of EPA. It helps to quantify the human health impacts of toxic chemicals released into the air. The higher the value, the higher the potential for health impacts.	EJScreen
TRI	TRI	It is the average TRI chemical (onsite and offsite) release in lbs of a facility in a county. It is calculated by dividing the sum of TRI releases reported by the facilities by the total number of Facilities in a county.	EJ Facility Mapping tool
RCRA	RCRA	It is the average amount of hazardous waste measured in tons, transported by a facility in a county	EJ Facility Mapping tool
GHG	GHG	It is the average amount of carbon dioxide equivalent greenhouse gas direct emission from a facility measured in metric tons in a county.	EJ Facility Mapping tool
NEI	NEI	It is the average annual emissions of criteria air pollutants in tons from a facility in a county.	EJ Facility Mapping tool
DMR	DMR	It is the average amount of pollutants discharged into water bodies from a facility in a county. It is measured in kilograms.	EJ Facility Mapping tool
Hazard Proximity and Exposure Risk variables			
PTRAF	PTRAF	It represents the proximity to high traffic volume. It is based on the average annual daily traffic count divided by the distance from the centroid of the census block.	EJScreen
PRE1960PCT	LEAD	It is the percentage of occupied housing units built before 1960. As the paint used in those houses contained lead, it represents the potential lead exposure.	EJScreen
PNPL	PSF	It indicates the proximity to the superfund site.	EJScreen
PRMP	PRMP	It represents the proximity to the RMP facility.	EJScreen
PTSDF	PHWF	It represents the proximity to the hazardous waste facilities.	EJScreen
PWDIS	PWDIS	It represents the potential human health impacts from the toxic chemicals in the discharge water.	EJScreen
UST	UST	It represents the risk of being affected by leaked underground storage tanks.	EJScreen

Data integration

The census block-level data obtained from the EJScreen dataset and the EJ Facility mapping tool were combined based on the county name. Base-level variables were used as they were provided in the EJScreen dataset. For the level 2 predictors, individual data available for the facilities in the EJ facility mapping tool were summed and then averaged based on the number of facilities in that specific county. Then, the calculated average value for a county was assigned to all the block groups present in that particular county for which the average value had been calculated.

Statistical analysis

A correlation analysis was conducted to generate a correlation matrix heatmap in the JASP tool (JASP Team, 2024). It was conducted to find out Pearson's correlation coefficient value of the variables. Studying the results, the variables with very high correlation values (greater than 0.75) were discarded. GHG, NEI, and TRI had very high correlation coefficient values, so only the TRI variable was used, and others were discarded from the analysis. The impact of different variables on life expectancy is studied by taking the LOW-LIFE variable as the outcome.

A simple linear regression model was formulated using the GAMLj package in the JAMOVI (The jamovi project, 2024) software tool. All the predictor variables were used as covariates, and the LOW_LIFE variable was used as the dependent variable. For the model thus formed, its model fit data was studied. The linear regression option available in the analyses tab of the Jamvoi tool was used to develop the model.

Then, a null model (Model 0) was formulated. In the null model, only the intercept and the cluster variable were used, and no predictors were used. It helps to understand how much proportion of the total variance is due to the grouping structure. Multilevel modeling is necessary if the intraclass correlation coefficient (ICC) is significant. Then, multiple hierarchical models were developed to study the individual models for each group of covariates. This was followed by formulating a mixed model whereby all the covariates were used. The linear mixed model option inside linear models in the analyses tab of the JAMOVI tool was used to formulate multilevel models.

First, the measure type of the variables was set. The CNTY variable was set to "nominal" type, whereas other variables were set to "continuous" type. For the models, CNTY is set as the cluster variable. The restricted maximum likelihood (REML) option was chosen to estimate the variance components, as it gives less biased variance than the maximum likelihood option (Maestrini et al., 2024). Following the general practice, 95% was taken as the confidence level for the model parameters estimation. The scale of the data points for different variables had large variability. Therefore, to ensure that the variables with values of a larger scale do not generate a bias in the model, all the covariates and the outcome variables were scaled using z-scores. This helped to create uniformity in the scales of the variables.

The GAMLj statistical package used in the study uses the Sattherwaite approximation method to estimate the degrees of freedom and the Wald method to calculate the confidence intervals. The standard errors of the parameter estimates are used in the Wald method to construct the confidence intervals. The LOW_LIFE was set as the dependent variable. The null model is named Model 0, and the model with demographic and socioeconomic variables as covariates is Model 1. Similarly, the models with pollution and health risk variables as covariates and those with hazard proximity and exposure risk variables are Model 2 and Model 3, respectively. In model 4, all three categories of variables are used as covariates to formulate a single mixed model.

Obtained model residual plots showed that the extremities had skewed the residual plot, making it non-normal. Therefore, 136 data points were identified as outliers using the cluster boxplot in JAMOVI software, which were deleted from the data set, and normality was reassessed using the Shapiro-Wilk test (p-value > 0.05) and Q-Q Plot. The covariates' variance inflation factor (VIF) was measured for the collinearity check, and residual-predicted scatterplots were used to check for homoscedasticity. The model with the best fit was then selected by comparing the model fit statistics of all models. AIC was used to compare the models' fit following the general practice. The model with the lowest AIC value was chosen as the final model owing to the best-fit statistics.

Table 2. Data description (n=1503)

Descriptive	Mean	Standard Error	Median	Standard Deviation	Minimum	Maximum
LOW_LIFE	0.22	0.00	0.22	0.03	0.13	0.31
POP	1,103.47	13.16	1,019.00	510.14	0.00	3,478.00
CLR	0.08	0.00	0.05	0.11	0.00	0.93
LING	0.00	0.00	0.00	0.01	0.00	0.14
U_5	0.05	0.00	0.04	0.04	0.00	0.41
O_64	0.21	0.00	0.20	0.10	0.00	0.77
L_INC	0.38	0.00	0.36	0.19	0.00	1.00
UNEMP	0.07	0.00	0.04	0.09	0.00	1.00
EDU	0.12	0.00	0.10	0.10	0.00	0.66
PM25	7.47	0.02	7.41	0.70	5.77	9.03
O3	57.15	0.03	57.40	1.25	54.27	59.65
PMDSL	0.12	0.00	0.10	0.06	0.04	0.35
RESP	0.30	0.00	0.30	0.13	0.20	2.00
RSEI_AIR	5,292.64	544.98	637.00	21,127.99	0.00	330,995.86
TRI	8,252.63	447.15	1,439.48	17,335.41	0.00	103,905.08
RCRA	43.59	2.36	5.57	91.60	0.00	579.89
DMR	974,336.18	150,815.05	1,220.98	5,850,000.00	0.00	74,400,000.00
PTRAF	52.88	2.46	16.38	95.02	0.00	1,367.11
LEAD	0.35	0.01	0.31	0.24	0.00	1.00
PSF	0.09	0.01	0.04	0.23	0.01	3.15
PRMP	0.34	0.02	0.11	0.66	0.01	6.75
PHWF	0.53	0.03	0.10	1.17	0.01	12.91
PWDIS	3.04	0.85	0.01	32.96	0.00	1,113.48
UST	1.92	0.09	0.37	3.66	0.00	32.74

Results

Data description

Table 2 presents the descriptive statistics of the data. The outcome variable, LOW_LIFE, had a mean of 0.22 with a standard deviation of 0.03, indicating relatively low variation across the census blocks. The median value 0.22 closely matched the mean, with minimum and maximum values ranging from 0.13 to 0.31. For demographic and socio-economic variables, the average population (POP) per block group was 1,103.47, with a standard deviation of 510.14 and a wide range from 0 to 3,478. The percentage of people of color (CLR) had a mean of 0.08, but its distribution was highly skewed, as indicated by a median of 0.05 and a maximum of 0.93. The percentage of limited English-speaking households (LING) was minimal, with a mean close to 0 and a maximum of 0.14. The proportions of children under five U_5) and individuals over 64 years (O_64) were 0.05 and 0.21 on average, respectively, with minimal variation. Socioeconomic indicators, such as the proportion of low-income residents (L_INC) and the unemployment rate (UNEMP), had means of 0.38 and 0.07, respectively, while the educational attainment rate (EDU) averaged 0.12. For environmental variables, the annual average PM25 showed low variability, with a mean of 7.47 and a standard deviation of 0.70. Ozone levels (O3) had a mean of 57.15 and a narrow range from 54.27 to 59.65. Diesel particulate matter levels (PMDSL) averaged 0.12, while respiratory risk scores (RESP) had a mean of 0.30, but some extreme values reached 2.00. The RSEI air toxicity score (RSEI_AIR) showed significant variability, with a mean of 5,292.64 and a maximum value of 330,995.86. Among hazard proximity and exposure risk variables, the toxic release inventory (TRI) averaged 8,252.63, but variability was high, with a standard deviation of 17,335.41 and a maximum of

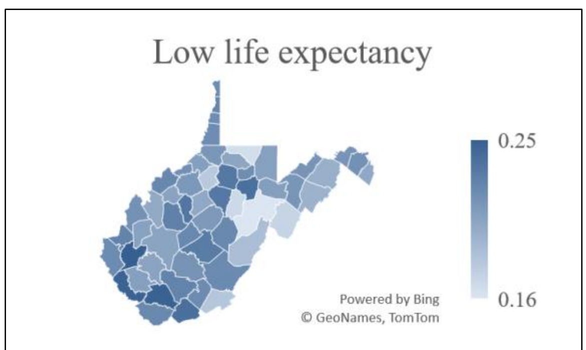

Figure 1. Low-life expectancy distribution.

103,905.08. Hazardous waste facility proximities (RCRA) and wastewater discharge levels (DMR) also exhibited substantial ranges, with means of 43.59 and 974,336.18, respectively, and extreme maximum values. Proximity to traffic (PTRAF) had a mean of 52.88 but ranged widely up to 1,367.11. Lead exposure risk (LEAD) had a mean of 0.35, and other hazard proximity measures like PSF, PRMP, PHWF, and PWDIS showed variation, with mean values of 0.09, 0.34, 0.53, and 3.04, respectively. Proximity to underground storage tanks (UST) averaged 1.92 but reached a maximum of 32.74.

Fig. 1 shows that the southern part of West Virginia has comparatively lower life expectancy, i.e., a higher value of the Low Life expectancy variable (indicated by darker shade).

The population distribution in Fig. 2 shows that Kanawha County has a comparatively higher population concentration than other counties and a very dark shade compared to other parts of the state. The distribution of people of color and limited English-speaking households seems the same throughout the state, with a slightly higher concentration in a few counties. The distribution of low-income, unemployment, and education below high school has a similar pattern.

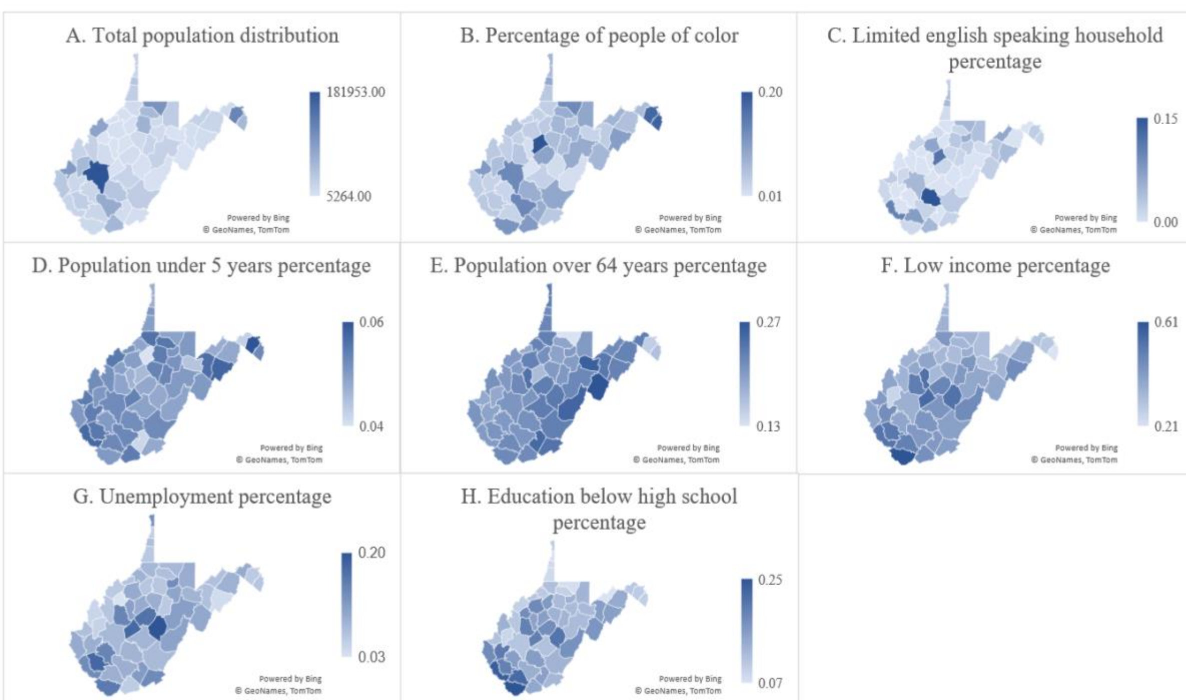

Figure 2. Demographic and socio-economic variables distribution.

The PM 2.5 concentration distribution in Fig. 3 shows a relatively higher concentration in the northwestern part of West Virginia and lower in the eastern parts. The distribution of the air toxic respiratory hazard index and diesel particulate matter concentration also shows a similar pattern to PM 2.5. However, the ozone concentration is lower in the central and northeastern parts of the state than in other parts. The distribution of other considered pollution and hazard risk variables is higher in some counties but almost identical in different parts.

The distribution in Fig. 4 shows that the proximities to all the considered entities, traffic, Superfund sites, RMP facilities, hazardous waste management facilities, water pollutants in discharge water, and underground storage tanks are high in a few counties and similar in other counties. However, the residential lead exposure is higher throughout the state due to the higher number of old houses.

From the correlation analysis, the correlation heatmap in Fig 5 is generated. In the correlation matrix heatmap, purple represents a positive correlation between the variables, and brown represents a negative correlation. The darker the shade, the higher the correlation value. A maximum correlation coefficient of 0.877 was found between the variables NEI and GHG, and the second highest value of 0.864 was between TRI and NEI. So, only the TRI value was used, and the GHG variable and NEI were discarded for the sections of the analysis.

Modal Estimates

The fit data, fixed estimates, and random components for all the models formulated are presented in Table 3. For the simple linear regression model, the adjusted R-squared value is 0.181. This means that the model can justify 18.1% of the variability. The ICC value of the null model is 0.174, with a significant LR test value of 136, which means that clustering has improved the model. So, using multilevel models for the analysis makes sense. The fit statistics of all the models formulated are in Table 2. Model 3 has the lowest value of AIC (i.e., 3956), so it is the best model among all the models formulated.

Figure 3. Pollution and health risk variables distribution.

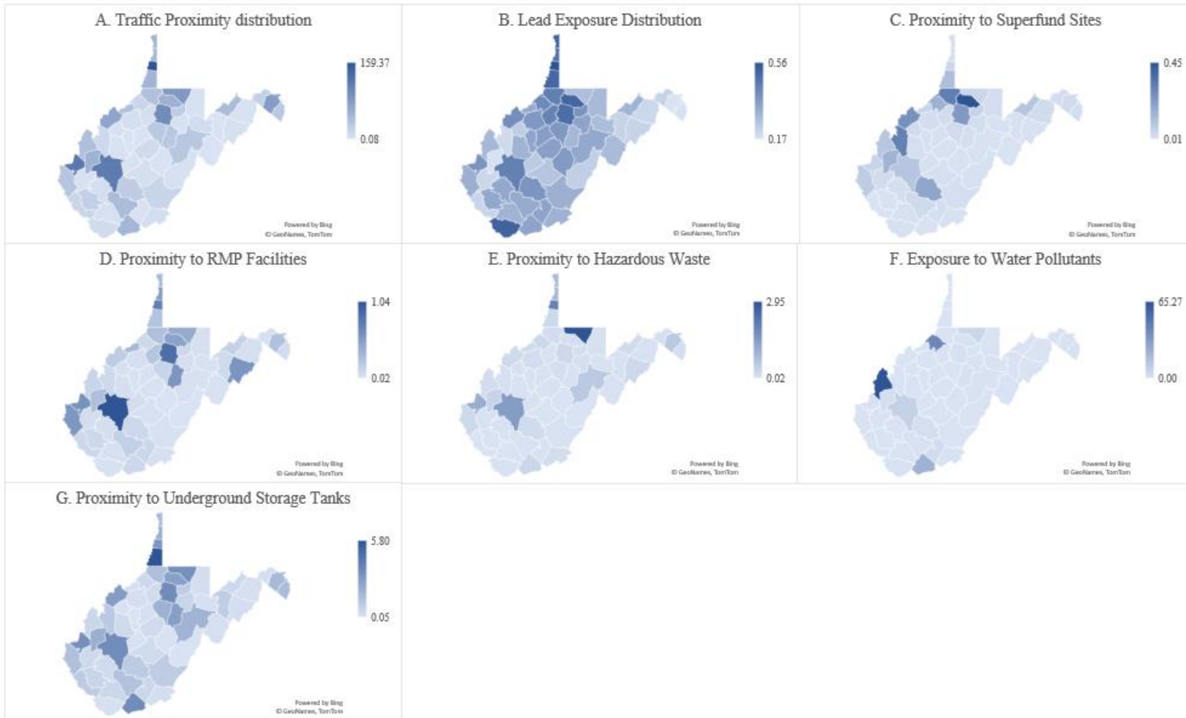

Figure 4. Hazard proximity and exposure risks variables distribution.

Fixed Effects (Measures of Association)

The nearness to traffic volume (β = 0.0647; 95% CI 0.0051 to 0.1243), the exposure to lead by staying in the houses built before 1960 (β = 0.1583; 95% CI 0.10134 to 0.21521), the proximity to an RMP facility (β = 0.1154; 95% CI 0.05027 to 0.18051), Nearness to hazardous waste facilities (β = -0.0773; 95% CI [-0.14532 to -0.00928), the wastewater discharge indicator quantifying the relative risk of exposure to pollutants in downstream water bodies (β = 0.0496; 95% CI 0.00391 to 0.09528), and the indicator quantifies the relative risk of being affected by a leaked underground storage tank (β = 0.164; 95% CI 0.10401 to 0.2239), are the significant predictors of the Low life expectancy in the Model 3. The UST variable has the highest value of the fixed effect, implying that it has the strongest influence over low life expectancy. People living in areas with a higher value of the UST indicator, i.e., areas with a higher risk of being affected by a leaked underground storage tank, have lower life expectancy. Similarly, the variable LEAD has the second most substantial effect. So, there is lower life expectancy in the areas with a higher chance of residential lead exposure due to a higher number of buildings built before 1960. Furthermore, PRMP, PTRAF, and PWDIS also have positive fixed effects, suggesting that life expectancy is lower in areas where these variables have higher values. However, the variable PHWF has a negative impact, which seems counterintuitive. It suggests that life expectancy is higher in areas with greater proximity to hazardous waste facilities.

In addition, the proximity to a superfund site (β = 0.0285; 95% CI -0.02224 to 0.07915) is not a statistically significant predictor in model 3. So, this indicator does not significantly affect the life expectancy of people in a bock group.

Random Effects (Measures of Variation)

The Intraclass correlation coefficient estimated for the null model found that 17.4% of the variance in the LOW_LIFE is due to county-level factors. Similarly, for Model 1, Model 2, Model 3, and Model 4, the variance attributed to the clustering effects are 17.9%, 23.8%, 21.2%, and 21.2%, respectively. Also, the significant values of the LR Test in the random components show that the models' fit statistics improved significantly for all with multilevel modeling.

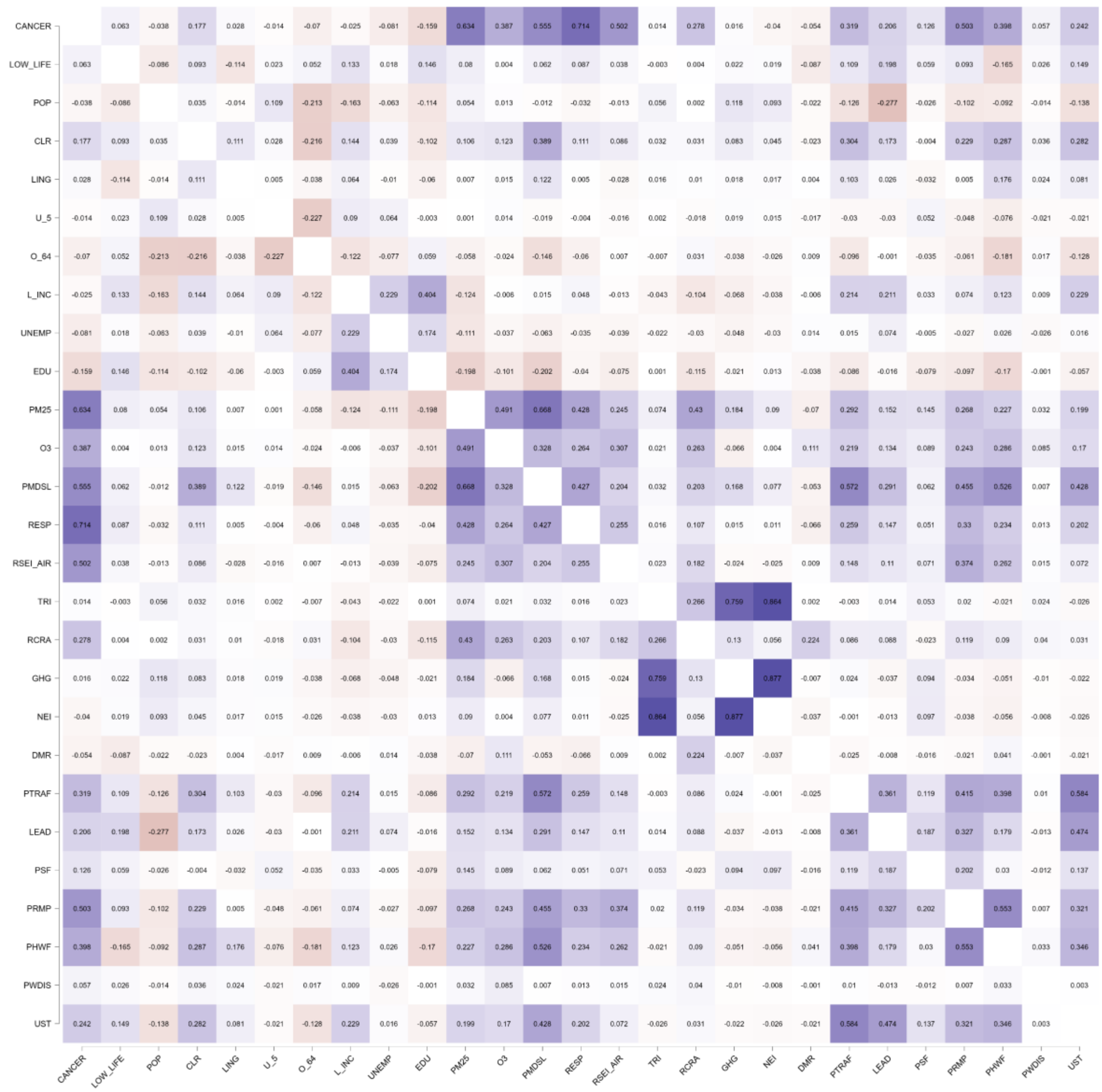

Figure 5. Correlation heatmap of all the parameters. The darker shade of purple color shows higher value of positive correlation between the parameters and the darker shade of brown color shows the higher value of negative correlation. Lower the shade lower is the respective correlation.

Table 3. Model Parameter Estimates

Variables (1,486)	Model 0 [95% CI]	Model 1 [95% CI]	Model 2 [95% CI]	Model 3 [95% CI]	Model 4 [95% CI]
Fixed effects					
(Intercept)	-0.00988 [-0.138, 0.118]	-0.0231 [-0.1463, 0.1001]	0.078 [-0.0739, 0.2299]	0.0378 [-0.09557, 0.17122]	0.04336 [-0.09214, 0.17887]
POP		-0.1001*** [-0.1493, -0.0508]			-0.04297 [-0.09199, 0.00605]
CLR		0.1118*** [0.0615, 0.1621]			0.04985 [-0.0015, 0.1012]
LING		--0.0356 [-0.0813, 0.01]			-0.04029 [-0.0845, 0.00391]
U_5		-0.0173 [-0.0639, 0.0294]			-0.00523 [-0.05049, 0.04003]
O_64		-0.0264 [-0.0762, 0.0234]			-0.00921 [-0.05735, 0.03892]
L_INC		0.214*** [0.1598, 0.2682]			0.15023***[0.09585, 0.20461]
UNEMP		-0.0161 [-0.0639, 0.0316]			-0.01159 [-0.05785, 0.03466]
EDU		0.0824** [0.029, 0.1358]			0.10584** [0.0537, 0.15799]
PM25			-0.1751* [-0.3341, -0.0161]		-0.01991 [-0.16929, 0.12947]
O3			-0.0206 [-0.1502, 0.1089]		-0.00661 [-0.12621, 0.11298]
PMDSL			0.3302*** [0.2481, 0.4123]		0.12628* [0.02486, 0.22769]
RESP			0.1136*** [0.0529, 0.1744]		0.08905** [0.03116, 0.14695]
RSEI_AIR			0.0391 [-0.0122, 0.0904]		0.03802 [-0.01246, 0.08851]
TRI			-0.0257 [-0.167, 0.1155]		-0.03552 [-0.16082, 0.08979]
RCRA			0.0159 [-0.1178, 0.1497]		0.01245 [-0.10731,0.13221]
DMR			-0.0132 [-0.109, 0.0827]		-0.00175 [-0.0877,0.08419]
PTRAF				0.0647* [0.0051, 0.1243]	0.00413 [-0.05767, 0.06593]
LEAD				0.1583*** [0.10134, 0.21521]	0.10573** [0.04774, 0.16373]
PSF				0.0285 [-0.02224, 0.07915]	0.03021 [-0.01967, 0.08008]
PRMP				0.1154*** [0.05027, 0.18051]	0.0775* [0.01216, 0.14285]
PHWF				-0.0773** [-0.14532, -0.00928]	-0.11679*** [-0.18616, -0.04743]
PWDIS				0.0496* [0.00391, 0.09528]	0.0471* [0.00269, 0.09152]
UST				0.164*** [0.10401, 0.2239]	0.13299*** [0.07407, 0.19192]
Random Effects					
ICC	0.174	0.179	0.238	0.212	0.212
LR Test	136***	133***	193***	160***	142***
Model Fitness					
R-squared	0.174	0.259	0.301	0.305	0.348
Log-Likelihood	-2067	-2006	-2033	-1968	-1953
AIC	4140	4034	4088	3956	3958
BIC	4156	4092	4146	4009	4096
Number of Clusters	55	55	55	55	55

Note: p < 0.001 ***, p<0.01, p<0.05***

Discussion

Our study evaluated different factors affecting life expectancy in West Virginia. The key findings show that proximity to traffic, RMP facilities, wastewater discharge, Underground storage tanks, and the risk of Lead Exposure positively correlates with low life expectancy. In addition, this study found that proximity to hazardous waste management facilities negatively correlates with low life expectancy. Also, although not significant, proximity to superfund sites was found to have positively contributed to low life expectancy.

Our study showed that the population living in areas with greater proximity to traffic have lower life expectancy than those with lower magnitude of traffic proximity. This finding is consistent with the common theories as well as past studies showing adverse effects of higher proximity to traffic on human health (Brender et al., 2011; Hoffmann et al., 2009; Sørensen et al., 2011) like higher risk of coronary heart disease (Gan et al., 2010), neurologic disease (Yuchi et al., 2020), greater exposure to pollutant emissions, higher risk of accidents, and many more.

In addition, this study showed that populations with a higher risk of residential lead exposure have lower life expectancy. This finding aligns with the previous studies (Ana et al., 2007; Boskabady et al., 2018; Patočka & Kuča, 2016) indicating that prolonged lead exposure leads to chronic health issues like increased blood pressure (Ana et al., 2007), gastrointestinal effects, anemia (Prüss-Üstün et al., 2004), chronic kidney disease (CKD) (Song et al., 2024), and many other health issues. As found in our results, these factors can be the reasons for the lower expectancy caused by lead exposure.

Next, earlier studies have revealed that the areas near RMP facilities have lower property values (Guignet et al., 2023). Because of this, people of lower income status are likely to live near RMP facilities, which in turn is inversely related to life expectancy (Rogot et al., 1992; G. Singh & Lee, 2020). This supports our finding of a direct relationship between proximity to RMP facilities and low life expectancy.

Different contaminants and toxins in polluted water can enter the food cycle and cause serious health hazards following their use in farm irrigation, human consumption of aquatic life, and groundwater use for drinking (Jiang et al., 2022; Wato & Amare, 2020). Thus, the greater the amount of toxins and pollutants in the downstream water, the greater the impact on human health. These health-related hazards can be acting as the cause for the lower life expectancy, which corroborates with the finding from our study that the Wastewater discharge indicator (indicator presents the relative risk from the pollutants in the downstream water bodies (U.S. Environmental Protection Agency (EPA), 2023)) is positively related to the low life expectancy.

Moreover, underground storage Tanks (USTs) are used for storing energy supply reserves, waste containments, and toxic substances, owing to their large capacity and minimum floor space (Ooi et al., 2019). Also, leaking USTs are considered a major source of groundwater contamination (Nadim et al., 2000). Earlier studies have found that groundwater contamination is related to human health hazards (Jiang et al., 2022; Srivastav & Ranjan, 2020). This supports the positive relation between the predictor UST and the low life expectancy obtained in our study.

Though insignificant, our study has found that proximity to superfund sites is positively related to low life expectancy, which aligns with previous studies' findings. The small number of Superfund sites (11 active Superfund sites (USEPA, 2024)) in the West Virginia State may be the reason behind the insignificance of this variable in the model. Studies have found uncontrolled hazardous waste sites, such as superfund sites, associated with environmental and public health concerns (Johnson Barry L. & DeRosa Christopher, 1997). Several hazardous substances, including VOCs, Arsenic, Cadmium, and Polychlorinated biphenyls (Johnson, 1995), are found in these sites. The health impacts from these dangerous substances may be the reasons behind lower life expectancy for the populations living near these sites.

However, our study finds that proximity to hazardous waste management facilities is negatively related to Low life expectancy. This counterintuitive result contradicts previous findings (Domingo et al., 2020; Saxena & Jotshi, 1996), where studies have found that proximity to hazardous waste management facilities has negative health impacts. The reason behind this finding remains unexplored, and it is a limitation of this study, which calls for further future work.

The findings from our study suggest that for improvement in the life expectancy in West Virginia state, the government needs to increase their focus on parameters like reduction of traffic pollution, providing special health care policies for the population in the proximity of RMP facilities, and superfund facilities. Several laws, regulations, and investments from the state as well as the federal government, like The Clean Water Act (CWA)(Walsh & Ward, 2022), The Resource Conservation and Recovery Act (RCRA) under the Solid Waste Disposal Act, etc., are related to the aforementioned life expectancy affecting factors. CWA is in action to protect water bodies against different kinds of pollution, and RCRA works to protect against leaking underground storage tanks. The recent bipartisan infrastructure law by the federal government has committed much money to improve traffic-related infrastructures (U.S. Department of Transportation, 2021). This improvement in transportation infrastructures can contribute to reducing some of the negative harms, such as reduction in noise and pollutant emissions from traffic proximity. Similarly, the RMP program under the Clean Air Act (Kleindorfer et al., 2004) and the Superfund program under the Comprehensive Environmental Response Compensation and Liability Act (CERCLA) (Johnson, 1995; Johnson Barry L. & DeRosa Christopher, 1997) are the programs launched by the federal government to protect populations from chemical accidents and uncontrolled hazardous wastes. In addition, the US Department of Housing and Urban Development had a Lead Hazard Reduction Grant Program (U.S. Department of Housing and Urban Development, 2024) for fiscal year 2024 that assists projects involving abatement, repair, or rehabilitation of eligible privately owned entities to maximize the number of children below the age of 6 protected from lead poisoning. Though these acts and several others are acting to ensure healthier lives and longer life expectancy, there is still room for improvement. Providing medical benefits for the communities living in the proximities of RMP and superfund facilities, further stringent rules and regulations regarding noise and other pollution related to vehicles, and increasing the outreach of the programs intended for reducing residential lead poisoning programs can be some of the ways forward. Existing laws must be more vigorously implemented, and plans that motivate and incentivize facilities to switch toward less hazardous material usage in their processes must be developed. The state legislature should consider the factors evaluated in this study while formulating new plans and policies.

Limitations

The major limitation of our study is the lack of sufficient publicly available data. The data related to different factors that can impact life expectancy at the block group level, like the impact of drug addiction, access to healthcare services, and many others that could have been incorporated into the model, is lacking. Furthermore, our model showed a negative relation between the proximity to hazardous waste management facilities and the low life expectancy, indicating that the population living near the hazardous waste management facilities has a higher life expectancy. This is counterintuitive and contradicts the findings of previous studies. The reason behind this may be the lack of consideration of other parameters in the model. This calls for further research.

Conclusion

Traffic proximity, proximity to RMP facilities, residential Lead Exposure risk, and risk for toxic chemicals in discharge water significantly impact low life expectancy. These findings showed that to reduce the disparities indicated by Low life expectancy, the government and policymakers of West Virginia state should focus on increasing the outreach of existing programs and formulating new programs, such as providing more health care benefits to populations near RMP facilities, higher water toxicity, and superfund facilities.

Acknowledgments

Dr. Ashish Nimbarte and Dr. Avishek Choudhury were funded by the US EPA Office of Prevention, Pesticides, & Toxic Substances (Award # 84070601). The contents of this publication do not reflect the views of the EPA, nor does mention of trade names constitute endorsement.

Literature Cited

Ahern, M. M., Hendryx, M., Conley, J., Fedorko, E., Ducatman, A., & Zullig, K. J. (2011). The association between mountaintop mining and birth defects among live births in central Appalachia, 1996–2003. *Environmental Research*, *111*(6), 838–846. https://doi.org/https://doi.org/10.1016/j.envres.2011.05.019

Ana, N.-A., Eliseo, G., K, S. E., & J, R. S. (2007). Lead Exposure and Cardiovascular Disease—A Systematic Review. *Environmental Health Perspectives*, *115*(3), 472–482. https://doi.org/10.1289/ehp.9785

Boskabady, M., Marefati, N., Farkhondeh, T., Shakeri, F., Farshbaf, A., & Boskabady, M. H. (2018). The effect of environmental lead exposure on human health and the contribution of inflammatory mechanisms, a review. *Environment International*, *120*, 404–420. https://doi.org/https://doi.org/10.1016/j.envint.2018.08.013

Brender, J. D., Maantay, J. A., & Chakraborty, J. (2011). Residential Proximity to Environmental Hazards and Adverse Health Outcomes. *American Journal of Public Health*, 101(S1), S37–S52. https://doi.org/10.2105/AJPH.2011.300183

DePietro Andrew. (2023, November 9). 2023 Poverty Rate By State, Per The Latest Census Data. Forbes. https://www.forbes.com/sites/andrewdepietro/2023/11/09/2023-poverty-rate-by-state-per-the-latest-census-data/

Domingo, J. L., Marquès, M., Mari, M., & Schuhmacher, M. (2020). Adverse health effects for populations living near waste incinerators with special attention to hazardous waste incinerators. A review of the scientific literature. Environmental Research, 187, 109631. https://doi.org/https://doi.org/10.1016/j.envres.2020.109631

Gan, W. Q., Tamburic, L., Davies, H. W., Demers, P. A., Koehoorn, M., & Brauer, M. (2010). Changes in Residential Proximity to Road Traffic and the Risk of Death From Coronary Heart Disease. Epidemiology, 21(5). https://journals.lww.com/epidem/fulltext/2010/09000/changes_in_residential_proximity_to_road_traffic.18.aspx

Garry, V. F., Harkins, M., Lyubimov, A., Erickson, L., & Long, L. (2002). REPRODUCTIVE OUTCOMES IN THE WOMEN OF THE RED RIVER VALLEY OF THE NORTH. I. THE SPOUSES OF PESTICIDE APPLICATORS: PREGNANCY LOSS, AGE AT MENARCHE, AND EXPOSURES TO PESTICIDES. Journal of Toxicology and Environmental Health, Part A, 65(11), 769–786. https://doi.org/10.1080/00984100290071333

Guignet, D., Jenkins, R. R., Belke, J., & Mason, H. (2023). The property value impacts of industrial chemical accidents. Journal of Environmental Economics and Management, 120, 102839. https://doi.org/https://doi.org/10.1016/j.jeem.2023.102839

Hendryx, M., & Ahern, M. M. (2008). Relations Between Health Indicators and Residential Proximity to Coal Mining in West Virginia. American Journal of Public Health, 98(4), 669–671. https://doi.org/10.2105/AJPH.2007.113472

Hoffmann, B., Moebus, S., Dragano, N., Möhlenkamp, S., Memmesheimer, M., Erbel, R., & Jöckel, K.-H. (2009). Residential traffic exposure and coronary heart disease: results from the Heinz Nixdorf Recall Study. Biomarkers, 14(sup1), 74–78. https://doi.org/10.1080/13547500902965096

JASP Team. (2024). JASP (Version 0.18.3.0). https://jasp-stats.org/

Jiang, X., Zhang, H., Wang, X., Zhang, X., & Ding, K. (2022). Comprehensive Analysis of the Association between Human Diseases and Water Pollutants. International Journal of Environmental Research and Public Health, 19(24). https://doi.org/10.3390/ijerph192416475

Johnson, B. L. (1995). Nature, extent, and impact of superfund hazardous waste sites. Chemosphere, 31(1), 2415–2428. https://doi.org/https://doi.org/10.1016/0045-6535(95)00112-L

Johnson Barry L., & DeRosa Christopher. (1997). The Toxicologic Hazard of Superfund Hazardous-Waste Sites. 12(4), 235–252. https://doi.org/doi:10.1515/REVEH.1997.12.4.235

Khuder, S., Milz, S. A., Bisesi, M., Vincent, R., McNulty, W., & Czajkowski, K. (2007). Health Survey of Residents Living Near Farm Fields Permitted to Receive Biosolids. Archives of Environmental & Occupational Health, 62(1), 5–11. https://doi.org/10.3200/AEOH.62.1.5-11

Kleindorfer, P. R., Elliott, M. R., Wang, Y., & Lowe, R. A. (2004). Drivers of accident preparedness and safety: evidence from the RMP Rule. Journal of Hazardous Materials, 115(1), 9–16. https://doi.org/https://doi.org/10.1016/j.jhazmat.2004.06.016

Maestrini, L., Hui, F. K. C., & Welsh, A. H. (2024). Restricted maximum likelihood estimation in generalized linear mixed models. ArXiv Preprint ArXiv:2402.12719.

Nadim, F., Zack, P., Hoag, G. E., Liu, S., & Carley, R. J. (2000). Non-Uniform Regulations of Underground Storage Tanks in the United States: Calls for a National-Scale Revision. Spill Science & Technology Bulletin, 6(5), 341–348. https://doi.org/https://doi.org/10.1016/S1353-2561(01)00051-2

Northridge, M. E., Stover, G. N., Rosenthal, J. E., & Sherard, D. (2003). Environmental Equity and Health: Understanding Complexity and Moving Forward. American Journal of Public Health, 93(2), 209–214. https://doi.org/10.2105/AJPH.93.2.209

Ooi, C. S., Keng Ngui, W., Hui, K. H., Lim, H., & Leong, M. S. (2019). Review of Underground Storage Tank Condition Monitoring Techniques; Review of Underground Storage Tank Condition Monitoring Techniques. https://doi.org/10.1051/matecconf/20192

Patočka, J., & Kuča, K. (2016). LEAD EXPOSURE AND ENVIRONMENTAL HEALTH. Military Medical Science Letters, 85(4), 147–163. https://doi.org/10.31482/mmsl.2016.027

Pope III, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution. JAMA, 287(9), 1132–1141. https://doi.org/10.1001/jama.287.9.1132

Prüss-Üstün, A., Fewtrell, L., Landrigan, P. J., Luis Ayuso-Mateos, J., Ezzati, M., Lopez, A. D., Rodgers, A., & Murray, C. J. (2004). World Health Organization Report Part Title: Lead exposure Report Part Title: Comparative Quantification of Health Risks Report Subtitle: Global and Regional Burden of Disease Attributable to Selected Major Risk Factors Report Editor(s). In World Health Organization.

Roffia, P., Bucciol, A., & Hashlamoun, S. (2023). Determinants of life expectancy at birth: a longitudinal study on OECD countries. International Journal of Health Economics and Management, 23(2), 189–212. https://doi.org/10.1007/s10754-022-09338-5

Rogot, E., Paul Sorlie, M. D., & Johnson, N. J. (1992). Life Expectancy by Employment Status, Income, and Education in the National Longitudinal Mortality Study.

Saxena, S. C., & Jotshi, C. K. (1996). Management and combustion of hazardous wastes. Progress in Energy and Combustion Science, 22(5), 401–425. https://doi.org/https://doi.org/10.1016/S0360-1285(96)00007-X

Schulz, A., & Northridge, M. E. (2004). Social Determinants of Health: Implications for Environmental Health Promotion. Health Education & Behavior, 31(4), 455–471. https://doi.org/10.1177/1090198104265598

Singh, G. K., Kogan, M. D., & Slifkin, R. T. (2017). Widening Disparities In Infant Mortality And Life Expectancy Between Appalachia And The Rest Of The United States, 1990–2013. Health Affairs, 36(8), 1423–1432. https://doi.org/10.1377/hlthaff.2016.1571

Singh, G., & Lee, H. (2020). Marked Disparities in Life Expectancy by Education, Poverty Level, Occupation, and Housing Tenure in the United States, 1997-2014. International Journal of Maternal and Child Health and AIDS (IJMA), 10(1), 7–18. https://doi.org/10.21106/ijma.402

Sørensen, M., Hvidberg, M., Andersen, Z., Nordsborg, R., Lillelund, K., Jakobsen, J., Tjønneland, A., Overvad, K., & Raaschou-Nielsen, O. (2011). Road Traffic Noise and Stroke: A Prospective Cohort Study. Epidemiology, 22(1). https://journals.lww.com/epidem/fulltext/2011/01001/road_traffic_noise_and_stroke__a_prospective.127.aspx

Srivastav, A. L., & Ranjan, M. (2020). Chapter 1 - Inorganic water pollutants. In P. Devi, P. Singh, & S. K. Kansal (Eds.), Inorganic Pollutants in Water (pp. 1–15). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-818965-8.00001-9

Tafran, K., Tumin, M., & Osman, A. F. (2020). Poverty, Income, and Unemployment as Determinants of Life Expectancy: Empirical Evidence from Panel Data of Thirteen Malaysian States. Iran J Public Health, 49(2), 294–303. https://pmc.ncbi.nlm.nih.gov/articles/PMC7231709/

Tejada-Vera B, Salant B, & Arias E. (2022). U.S. Life Expectancy by State and Sex for 2020. National Center for Health Statistics. https://www.cdc.gov/nchs/data-visualization/state-life-expectancy/index_2020.htm

The jamovi project. (2024). jamovi (Version 2.5.6). https://www.jamovi.org/

Thornton, J. W., McCally, M., & Houlihan, J. (2002). Biomonitoring of Industrial Pollutants: Health and Policy Implications of the Chemical Body Burden. Public Health Reports®, 117(4), 315–323. https://doi.org/10.1093/phr/117.4.315

UN Department of Economic and social affairs. (2024). World Population Prospects. United Nations. https://population.un.org/wpp/Graphs/DemographicProfiles/Line/900

United States Environmental Protection Agency. (2023a, June). EJScreen. https://www.epa.gov/ejscreen/download-ejscreen-data

United States Environmental Protection Agency. (2023b, June). Pollution Prevention (P2) Environmental Justice (EJ) Facility Mapping Tool. https://awsedap.epa.gov/public/extensions/P2_EJ/P2_EJ.html

U.S. Department of Housing and Urban Development. (2024). FY24 Lead Hazard Reduction Grant Program. U.S. Department of Housing and Urban Development. https://www.hud.gov/program_offices/cfo/gmomgmt/grantsinfo/fundingopps/lead_hazard_reduction

U.S. Department of Transportation. (2021, November). The Bipartisan Infrastructure Law Will Deliver for West Virginia. Office of Public Affairs, U.S. Department of Transportation. www.transportation.gov/newsroom

U.S. Environmental Protection Agency (EPA). (2023). EJScreen Technical Documentation. www.epa.gov/EJScreen

USEPA. (2024, July). National Priorities List (NPL) Sites - by State. Environment Protection Agency. https://www.epa.gov/superfund/national-priorities-list-npl-sites-state#WV

Walsh, R., & Ward, A. S. (2022). An overview of the evolving jurisdictional scope of the U.S. Clean Water Act for hydrologists. WIREs Water, 9(5), e1603. https://doi.org/https://doi.org/10.1002/wat2.1603

Wato, T., & Amare, M. (2020). The Agricultural Water Pollution and Its Minimization Strategies – A Review. Journal of Resources Development and Management, 64, 10–22. https://doi.org/10.7176/jrdm/64-02

World Health Organization. (2024). Life expectancy at birth (years). World Health Organization. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/life-expectancy-at-birth-(years)

Yuchi, W., Sbihi, H., Davies, H., Tamburic, L., & Brauer, M. (2020). Road proximity, air pollution, noise, green space and neurologic disease incidence: A population-based cohort study. Environmental Health: A Global Access Science Source, 19(1). https://doi.org/10.1186/s12940-020-0565-4