From Hazard to Health: A Regression Analysis of Life Expectancy Near New Jersey's Superfund Sites

Dipesh Bhattarai[1*†]

[1]Bucknell University, Lewisburg PA.

12/08/2024

Abstract

This study sought to investigate the relationship between Life expectancy (LE) and Superfund sites within New Jersey's 21 counties, integrating socioeconomic and demographic variables through regression analyses. While this research's final model found no direct correlation between Superfund site proximity and LE at the county level, it showcased significant associations between factors, such as median household income, minority income, Black population percentages, and education level. The final model's adjusted R^2 was 0.8895, suggesting that socioeconomic factors are stronger predictors of life expectancy than proximity to hazardous waste sites, though this may be due to the limitations of county level data availability.

1 INTRODUCTION

Average life expectancy (LE) serves as a powerful indicator of the societal well-being of a region, reflecting the infrastructures of the economy, available healthcare, and society as a whole. The growth of LE signifies improvements in social well-being, public health, medical accessibility, and overall health quality. Countries in the global north are considered to have the highest LE, with their robust and continually developing healthcare systems providing a foundation of social well-being. In the United States, there is criticism that even with our advanced healthcare system, there are roadblocks that prevent the promotion of our social well-being. The LE in the US, from 1970 to 2022, has grown from 70.78 to 77.5 years (for both sexes) (CDC, 2024), while other nations, such as Japan, have seen growth from 72.16 to 84.91 years in the same time frame (MacroTrends, 2024). The drastic differences can be attributed to many factors, but one focus is on land that is too hazardous and toxic for human habitation, such that the Environmental Protection Agency (EPA) denotes them as Superfund sites.

The United States' large quantity of Superfund sites can be attributed to the rapid initial industrialization that occurred in the north-east regions. Manufacturing plants and facilities were constructed, where large volumes of synthetic materials were used, and the resulting toxic effluents were not disposed of correctly (Thompson Earth Systems Institute, 2024). The Comprehensive Environmental Response, Compensation, and Liabilities Act (CERCLA) of 1980 provided the EPA lawful authority to denote and clean up these toxic sites. However, even with CERCLA active, there are currently a total of 1,340 sites on the National Priority List (NPL), of which New Jersey (NJ) has 115 sites (EPA, 2024), representing 8.5% of all Superfund sites in the US, and more than any other state. With NJ's quantity of Superfund sites and its status as the fifth smallest state in the US, this paper seeks to investigate whether proximity to a Superfund site impacts residents' average life expectancy within counties.

2 BACKGROUND

The prevalence of Superfund sites may go unnoticed, yet their impacts are profoundly significant. Millions of Americans have lived near these contaminated areas, with an estimated three to four million children residing within one mile of a federally designated site (Landrigan et al., 1999). The hazards associated with these sites encompass a wide range of toxic substances, including heavy metals (e.g. lead, chromium, and arsenic), as well as organic compounds such as benzene and carcinogenic halocarbons like trichlorethylene. Children, especially in their developmental stage (six months to five years) are increasingly exposed and vulnerable, consuming seven times more water, three/four times more food, and two times more air than the average American (Currie et al., 2011; Landrigan et al., 1999). As these toxic substances are known to leach into food and water and dissipate in the air, children's immature immune system and metabolic pathways only further the risk posed by proximity to a Superfund site (Landrigan et al., 1999; NORD, 2024).

Since the 1980 CERCLA, there have been 440 sites cleaned and removed from the NPL, though the cleaning and remediation process can take years or even decades, depending on the site (Morrison, 2023). While over 1,300 sites remain on the NPL nationwide (Figure 1), the Environmental Protection Agency (EPA) has identified between 11,700 and 15,000 sites as hazardous waste locations (Kiaghadi et al., 2021; Landrigan et al., 1999). Within NJ, 134 sites are either on the NPL, part of the NPL, or proposed for the NPL, with over 2,000 additional sites not included on the NPL (Figure 2).

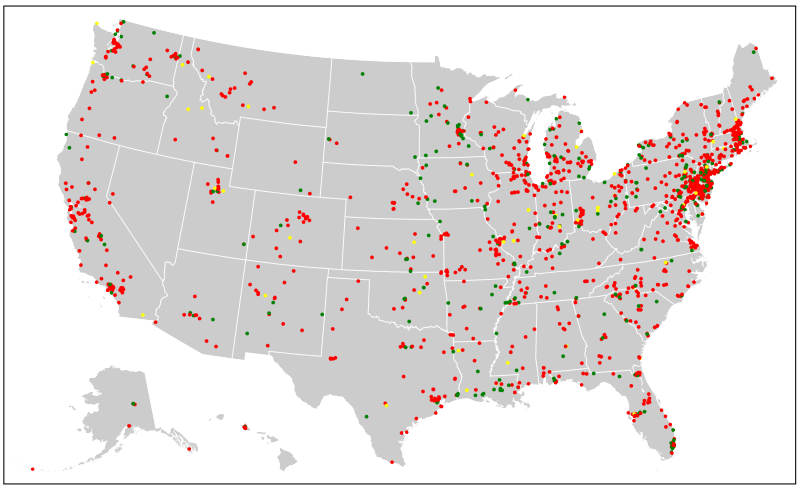

Figure 1: U.S. Superfund Sites: Active (red), Proposed (yellow), and Removed (green).

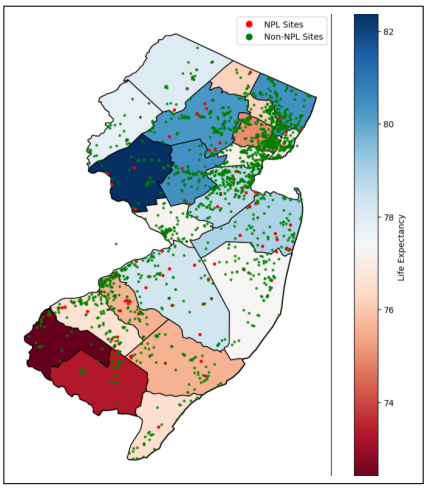

Figure 2: Superfund Sites and Life Expectancy by County in New Jersey

Demographics near Superfund sites are not distributed equally and reveal disparities across the United States. African American, Native American, and Hispanic communities largely comprise the populations near these sites. Children of these residents, as a consequence, face elevated risks of toxic exposure compared to children in other regions of the US or the state. These environmental and social justice concerns are amplified knowing that 60% of Superfund sites in the EPA's jurisdiction are vulnerable to natural hazards, such as flooding, wildfires, and chemical runoff. The environmental fallout from these sites, regardless of whether they're listed on the NPL or not, pose a drastic concern to public health risk due to the documented presence of carcinogenic substances, including heavy metals and hazardous organic compounds (Kiaghadi et al., 2021). One research study found that proximity (< 2000m) to a Superfund site prior to a successful cleanup was associated with a 20 to 25% increase in the risk of congenital (pre-birth) anomalies in children (Currie et al., 2011). As proximity to Superfund sites has shown to be detrimental to human health, research suggests that the presence of a site in an area with median income below than $52,580 can reduce residents' LE by an average of 0.58 years (Kiaghadi et al., 2021). Within our dataset, nine of the 21 NJ counties have an average minority income (Black, Native American, and Hispanic) below this threshold. These socioeconomic statuses suggest that specific populations may face increased health risks due to both their economic status and proximity to Superfund sites. Asian populations were not included as the NIH dataset did not have data available for AAPI populations (DHS, 2022).

3 METHODS

3.1 Statistical Design

This research paper aims to investigate the variables that influence life expectancy among residents in New Jersey's 21 counties. The study will examine socioeconomic and demographic factors, including median family income (all races), median family income (minority), minority racial composition (percentage of Black and Hispanic populations), poverty rates (in percentages), highest education obtained (percentage of population with a bachelor's degree or higher) and other key variables. The goal of this study is to showcase the effects of hazardous sites on residents' overall life expectancy, as LE provides insight into overall health of a region.

3.2 Statistical Analysis

Through statistical analysis, the research will employ use of the Statistical Analysis System (SAS) and Python, with data visualization supported by additional Python libraries such as matplotlib, geopandas, and pandas. A custom dataset will be developed, encompassing NJ's 21 counties alongside ten quantitative variables procured from government and state agencies (e.g. DHS, CDC, SSA, NJ DHS) to produce a regression model to predict LE based on the input variables. The model will follow Equation1 listed below, where β_0 is the regression model constant, β_i is the coefficient for the i^{th} socio-demographic variables, n is the total number of variables in the regression equation and X_i is the socio-demographic variable (Kiaghadi et al., 2021).

$$LE = \beta_0 + \sum_{i=1}^{n} + \beta_i X_i \tag{1}$$

4 RESULTS

Regression models that are developed will use the 10 variables (Table1), alongside the four interaction terms (Table 2) that have been created via the original variable: *Percent_Black, Percent_Poverty, Percent_Unemployed, Percent_Hispanic, Log_Income_Minority, and Percent_BachelorHigher.* These interaction terms examine how poverty compounds with racial demographics and unemployment to create multiplicative effects on life expectancy, while the education-income interaction explores how higher education levels might modify the relationship between minority income and life expectancy outcomes.

Variable	Definition
LE	The Average Life Expectancy
Income_All	Average median family income (all races, both sexes)
Income_Minority	Average median family income (Black, Hispanic, Native)
Percent_Black	Percent of population that is Black
Percent_Hispanic	Percent of population that is Hispanic
Percent_Poverty	Percent of population that is in poverty
Percent_BachelorHigher	Percent of population that has a Bachelor's Degree or higher
Percent_Unemployed	Percent of population that is unemployed
NPL	Number of Superfund sites on the NPL
Not_NPL	Number of Superfund sites not on the NPL

Table 1: Variable Definitions

Interaction Term	Terms
Percent_Black_Poverty	Percent_Black * Percent_Poverty
Percent_Poverty_Unemployed	Percent_Poverty * Percent_Unemployed
Percent_Hispanic_Poverty	Percent_Hispanic * Percent_Poverty
Log_IMBH	Log(Income_Minority) * Percent_BachelorHigher

Table 2: Interaction Terms

4.1 Models

Initially, a full model was developed using all of the ten variables. While the adjusted R^2 value was relatively high (0.8738), there were multiple variables (Table 3) that were deemed statistically insignificant ($p > 0.1$). Two other methods were used for model development: stepwise regression and backward elimination.

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	79.77271	4.43297	18.00	<0.0001
Income_All	61.71771	37.95980	1.63	0.1323
Income_Minority	-123.42793	75.91502	-1.63	0.1323
Percent_Black	-0.04268	0.03231	-1.32	0.2134
Percent_Hispanic	0.00328	0.03888	0.08	0.9343
Percent_Poverty	-0.16247	0.33951	-0.48	0.6416
Percent_BachelorHigher	0.14520	0.07336	1.98	0.0734
Percent_Unemployed	-0.31673	0.48280	-0.66	0.5253
NPL	-0.05173	0.07771	-0.67	0.5193
Not_NPL	0.00498	0.00695	0.72	0.4880

Table 3: Parameter Estimates for Full Life Expectancy Model

Using stepwise regression and backwards elimination, we arrived at a model that had an R^2 of 0.8895. While it has a high R^2 value, its non-usage of crucial variables such as *NPL* and *Not_NPL* suggests a further testing for an accurate model. There was another interaction term that was added: *Inv_PL*), the inverse of the variables *NPL* and *Not_NPL* multiplied together. Its inclusion in another stepwise and backwards elimination method yielded a model that had all significant terms (Table 4), alongside an R^2 of 0.8895.

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	73.55977	0.95265	3612.87924	5962.31	<0.0001
PL*	-35.04791	17.62270	2.39673	3.96	0.0631
Percent_Black	-0.08366	0.02154	9.14212	15.09	0.0012
Log_IMBH	0.03477	0.00442	37.48521	61.86	<0.0001

Table 4: Parameter Estimates from Stepwise Selection

Six other models were developed and their respective parameters (total variables, sig. variables, SSE, RMSE, and adjusted R^2) are transcribed (Table 5). Comparing these models, all have relatively high adjusted R^2 values below 0.90. Their SSE ranges from 7.69 to 10.7, indicating that our models have relatively low unexplained variability in predicting LE, with observations falling close to the fitted values. These small SSE values suggest our models effectively capture the relationships between our predictor variables and life expectancy.

Model	**Variables**	**Sig. Variables**	**SSE**	**RMSE**	**Adj.** R^2
Full Model	10	1	7.69	0.877	0.867
Stepwise	3	3	10.30	0.606	0.8945
Socioeconomic	3	2	10.767	0.796	0.89
Proximity	4	4	17.19	1.03	0.813
Forward w/Interactions	4	4	10.28	0.80	0.89
Mixed Selection w/Validation	4	2	10.112	0.795	0.89
AIC/BIC	5	2	9.51	0.796	0.89
AIC/BIC Modified	3	3	10.26	0.777	0.895

Table 5: Parameter Estimates from Stepwise Selection

Our final model, chosen based on the outlined parameters—maximizing the adjusted R^2 value, accounting for the significance and number of variables, as well as evaluating the SSE and RMSE values–is the Forward Selection model with Interaction Terms (Model 5), described by Equation 2 below.

$$\text{LE} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \tag{2}$$

where:

$\beta_0 = 74.36$ (Intercept)

$\beta_1 = 55.06, X_1 =$ Income_All

$\beta_2 = -110.09, X_2 =$ Income_Minority

$\beta_3 = -0.07, X_3 =$ Percent_Black

$\beta_4 = 0.17, X_4 =$ Percent_BachelorHigher

5 Conclusion

5.1 Findings and Thoughts

The final model (Equation 2) provides insights into life expectancy through socioeconomic and demographic parameters, despite not including Superfund site variables. Our model predicts the life expectancy of NJ residents (in a given county) using median household income, minority median income (Black, Hispanic, and Native residents), percentage of Black population, and percentage of residents with a bachelor's degree or higher.

The absence of Superfund parameters may be attributed to the data collection process, where segmenting NJ by county potentially normalized the localized effects of toxic sites. County wide averages might not capture the relationship between Superfund sites and health outcomes in specific municipalities. Analysis at the census tract level has shown to provide more depth, as it showcases the specific impacts of Superfund sites to general health (Kiaghadi et al., 2021).

5.2 Looking Forward

While our findings did not support the original thesis, Superfund sites remain a significant public health concern. Research has shown that successful Superfund cleanups can reduce congenital anomalies in children by 20-25% (Currie et al., 2011). Despite EPA's continuous monitoring and cleanup efforts, the sheer number of sites means millions of children still live within a mile of these locations (Landrigan et al., 1999).

5.3 Continuing Research

This research project enhanced both statistical and analytical research skills. Future development includes creating an interactive tool for plotting geodata, similar to the visualization of NJ county life expectancy and Superfund site locations.

References

CDC. (2024). *Faststats - life expectancy* (tech. rep.). Center for Disease Control and Prevention.

Currie, J., Greenstone, M., & Moretti, E. (2011). Superfund cleanups and infant health. *American Economic Review*, *101*(3), 435–441. https://doi.org/10.1257/aer.101.3.435

DHS. (2022). National institute of minority health and health disparities [Accessed: 2024-12-08]. https://hdpulse.nimhd.nih.gov/data-portal/social/table?socialtopic=030&socialtopic_options=social_6&demo=00010&demo_options=income_3&race=04&race_options=race_7&sex=0&sex_options=sexboth_1&age=001&age_options=ageall_1&statefips=34&statefips_options=area_states

EPA. (2024). Current npl updates [Accessed: December 7, 2024]. https://www.epa.gov/superfund/current-npl-updates-new-proposed-npl-sites-and-new-npl-sites

Kiaghadi, A., Rifai, H. S., & Dawson, C. N. (2021). The presence of superfund sites as a determinant of life expectancy in the united states. *Nature Communications*, *12*, 1947. https://doi.org/10.1038/s41467-021-22249-2

Landrigan, P. J., Suk, W. A., & Amler, R. W. (1999). Chemical wastes, children's health, and the superfund basic research program. *Environmental Health Perspectives*, *107*(6), 423–427. https://doi.org/10.1289/ehp.99107423

MacroTrends. (2024). Japan life expectancy 1950-2024 [Accessed: December 7, 2024]. https://www.macrotrends.net/global-metrics/countries/JPN/japan/life-expectancy

Morrison, J. (2023). Polluted sites linger under u.s. cleanup program [Accessed: December 8, 2024]. *Chemical & Engineering News*. https://cen.acs.org/articles/95/i14/Polluted-sites-linger-under-US-clean-up-program.html

NORD. (2024). Heavy metal poisoning - lead, mercury, arsenic, and cadmium [Accessed: December 8, 2024]. https://rarediseases.org/rare-diseases/heavy-metal-poisoning/

Thompson Earth Systems Institute. (2024). Tell me about: Superfund sites [Accessed: December 7, 2024]. https://www.floridamuseum.ufl.edu/earth-systems/blog/tell-me-about-superfund-sites/