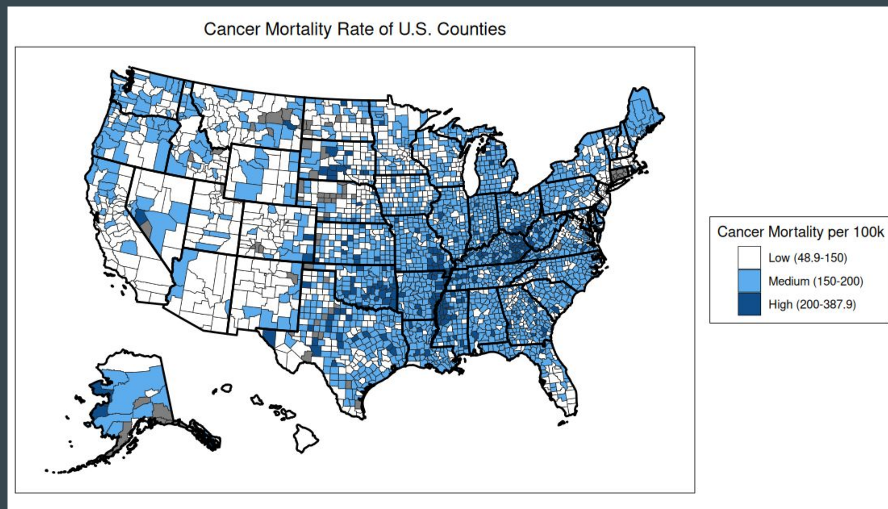
Poverty Level Relationship with Cancer Mortality in U.S. Counties



DATASCI 203 - Fall 2025
Elle Luo, Jeff Houser, Minh Doan, Vamsi Katari
12/11/2025

Our analysis focuses on cancer mortality rates within the United States. This graph here shows cancer mortality rates per 100,000 on a county level. This is simply calculated by counting the number of people who die from cancer in a given year within each county and accounting for the population of the county to establish the rate per 100,000 people. The darker blue colors on this graph represent counties with the highest rates while the white color represents counties with the lowest rates. Gray counties are counties without information regarding cancer mortality. This may be absent due to either the nature of how data is collected in different areas within the United States or due to counties with small populations to protect individual identities. You can see increases in the cancer rates in specific areas with Appalachia and the south exhibiting some of the highest rates.

Introduction

Motivation

- Cancer mortality shows persistent structural disparities.
- Poverty influences access to prevention, early detection, and treatment.
- Insurance coverage and racial composition shape additional inequities.

Research questions:

- What is the relationship between county poverty and mortality?
- How does insurance composition alter this relationship?
- How do racial and ethnic shares relate to mortality outcomes?

Motivation
• The central issue is unequal cancer outcomes across US counties. These disparities consistently align with socioeconomic disadvantage.
• Poverty shapes access to prevention and treatment, which means poorer counties face systematically higher risks long before diagnosis occurs.
• Insurance and racial composition compound these barriers, creating layered inequities that the analysis must account for.
Research questions
• First, we quantify the raw relationship between poverty levels and cancer mortality.
• Second, we assess how much of that relationship is explained or modified by differences in insurance coverage.
• Third, we evaluate whether racial and ethnic composition continues to predict mortality after economic and insurance controls.

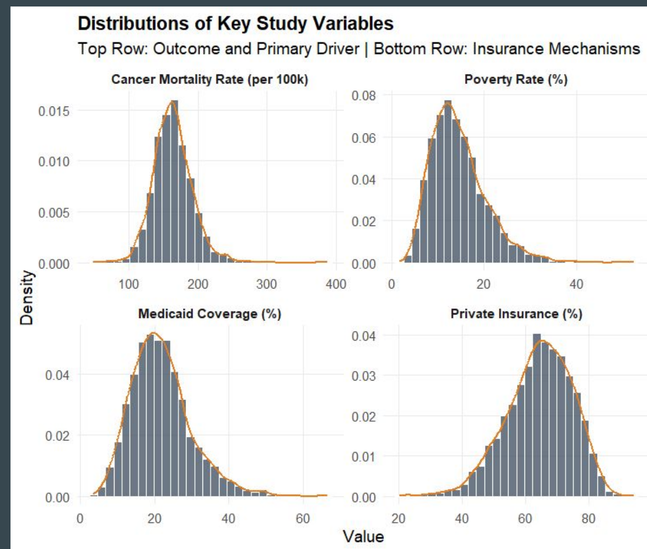
Data Sources & Key Variables

Data Sources:

- SEER*Stat (NIH)
- ACS 5-Year Estimates (Census.gov)

Key Variables:

- Dependent (Y Concept): Age-adjusted cancer mortality per 100k.
- Primary driver (X Concept): Poverty rate (% below federal poverty line).
- Other key variables: Medicaid %, private insurance %, race/ethnicity shares.



For data sources, we use two federal datasets. The mortality data come from SEER*Stat, which standardizes cancer death records across all counties, so the numbers are consistent and comparable. The socioeconomic measures come from the ACS 5-Year Estimates, which give stable county-level statistics.

For key variables, our outcome is age-adjusted cancer deaths per 100,000 people. Age adjustment matters because counties have very different age structures. Our main explanatory variable is the poverty rate, defined as the share of residents living below the federal poverty line. We also bring in Medicaid coverage, private insurance coverage, and race and ethnicity shares to capture differences in healthcare access and demographic structure.

Before diving into the specific variables, we looked at the histograms to check the overall 'shape' of our data. Specifically, we wanted to ensure there were no 'heavy tails' or extreme outliers

1. The cancer mortality rate: the cancer mortality rate follows a bell-shaped distribution centered around about 160 deaths per 100k, which supports running linear models.
2. Poverty and Medicaid: Poverty and Medicaid both show strong right-skew, meaning most counties are in the middle, but some have much higher disadvantage. These are the most vulnerable counties
3. Private insurance: Private insurance shows the opposite pattern, with most counties having relatively high coverage and a long tail of counties with very low coverage. This mirrors the negative relationship we expect: as poverty and

1. Medicaid rise, private insurance tends to fall.

While the socioeconomic variables (like poverty) are skewed, which is expected, they don't show the kind of extreme irregularity that would break a linear model.

Model Assumptions

IID:

- Geographical clustering of data impact on independence
- Counties can vary in size and demographics so they are not an identically distributed population
- Consequence of potential misrepresentation of standard errors

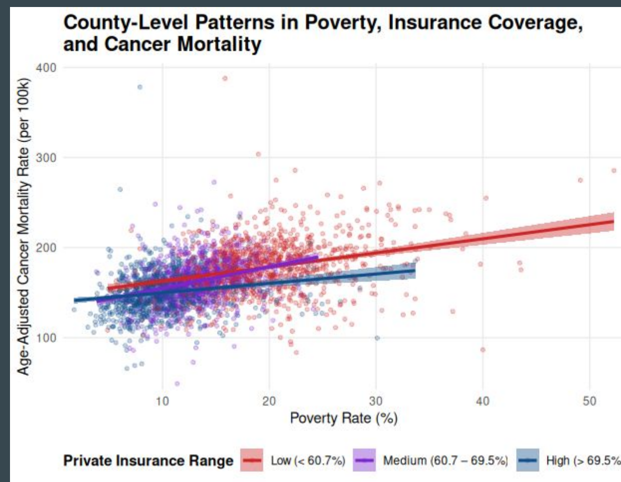
Existence and Uniqueness of BLP:

- Independent variables used in analysis were percentages that were finite without heavy tails- suggesting finite covariance with both the dependent variable and other independent variables.
- None of the variables were dropped in the analysis suggesting uniqueness of the BLP although some variables (racial percentage) were interrelated and had some degree of collinearity

Our dataset is sufficiently large that the assumptions required for our analysis are those of a large sample. Namely we needed to evaluate IID and the existence and uniqueness of the BLP. As seen in our initial slide that showed the cancer rate in different counties, there is obviously some geographic clustering of the data suggesting some of the counties could be thought of as groups of counties in our analysis suggesting a potential violation of independence. Counties as a unit of measure can also vary in size and demographic composition so they may not necessarily represent a draw from an identical population. The consequences of these potential violations are that it would possibly impact the standard errors in our analysis and make our data appear more precise than it necessarily is. This type of data has to be represented on the county level due to the sensitivity of patient information and also this format of the data is fitting with our analysis since our interest is improving health outcomes for different communities which is mediated by county-level data. The assumption of existence of the BLP is satisfied by our data since the independent and dependent variables in the analysis represent finite values. The independent variables are percentages within each county while the dependent variable is a rate per 100,000. The variables also do not exhibit heavy tails as demonstrated by Minh's data on the previous slide. The BLP is unique since none of the variables were dropped when performing the analysis. Although some variables such as racial percentages within each county were interrelated and had some degree of collinearity

Relationship of Poverty and Cancer Mortality

- There is an obvious correlation between an increased poverty rate within each county and an increased cancer mortality rate
- The percentage of people with private health insurance within each county is inversely related to a higher cancer mortality rate and higher poverty rate

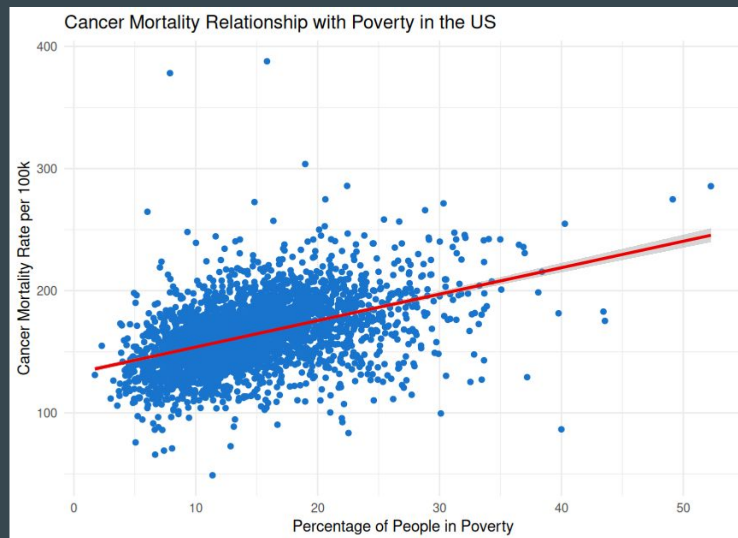


You can see from this graph here showing the poverty rate within each county on the x-axis and the cancer mortality rate on the y axis that a higher rate of poverty is associated with a higher cancer mortality rate. The colors of the individual points represent the amount of people with private insurance within each county. Lower percentages are represented by a red color while higher percentages are in blue. You can see there is a relationship between the amount of private health insurance people have within a county and both the poverty rate and cancer mortality rate with lower private insurance percentages corresponding to worse outcomes for poverty and cancer.

Model 1

- Examines the percentage of people living in poverty in a county, and how it relates to the cancer mortality.

- Shows a strong and positive association between poverty and cancer mortality. The coefficient for poverty in this model is 2.165 suggesting for every 10% increase in poverty within a county, the cancer mortality rate per 100k increases by roughly 21.

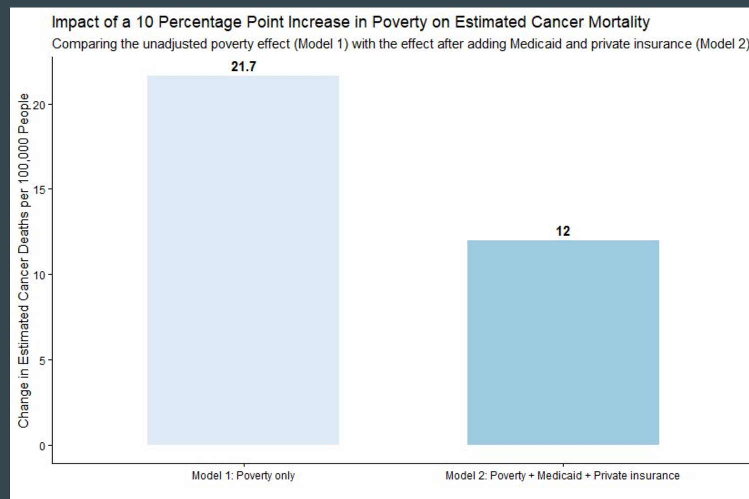


Model 2

Model 2 focusing only on poverty, Medicaid coverage, and private insurance coverage:

10-point increase in poverty associated:

- ~21.7 additional cancer deaths per 100k.
- Controlling for Medicaid reduces to ~12 deaths per 100k.



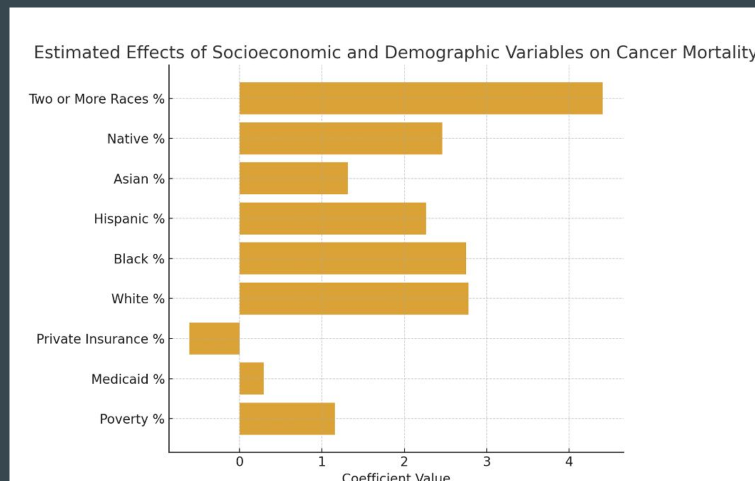
Model 2 focuses on how poverty relates to cancer mortality once we account for Medicaid and private insurance coverage.

The plot shows how the association between poverty and cancer mortality changes once we account for insurance coverage. The first bar is the raw relationship: a ten-point increase in poverty is linked to about 21.7 additional deaths per 100,000 people. When we add insurance to the model, that effect drops to about 12.

This tells us that a large share of the poverty–mortality link is tied to differences in insurance coverage across counties. But even when insurance is held constant, poverty is still associated with substantially higher mortality, which points to other structural factors beyond coverage alone.

Model 3

- Poverty remains a strong predictor of cancer mortality even after adjusting for insurance and demographics.
- Private insurance is associated with lower mortality, while Medicaid enrollment is associated with higher mortality.
- Two or More Races population shows the strongest association with higher cancer mortality.



The Model 3 coefficient plot shows that cancer mortality is shaped by a combination of socioeconomic, insurance, and demographic factors. Poverty remains a strong predictor even after adjusting for other variables, indicating that economically disadvantaged counties consistently face higher mortality burdens. Insurance patterns also align with expectations: higher Medicaid enrollment is associated with higher mortality, while greater private insurance coverage is linked to lower mortality, reflecting differences in access and quality of care. Racial and ethnic composition adds substantial explanatory power, with counties having larger Black, Hispanic, Native, and multiracial populations showing notably higher mortality rates. Taken together, these results highlight that cancer mortality disparities are multidimensional and reflect the intertwined effects of economic hardship, insurance access, and structural inequities.

Results

Strong link between poverty and cancer mortality

- A 10-point increase in poverty is associated with ~21 additional cancer deaths per 100,000.
- Socioeconomic disadvantage remains a consistent predictor across all models.

Insurance coverage modifies the relationship

- Adding Medicaid, private insurance, and uninsured rates reduces but does not eliminate the poverty mortality association.
- Counties with higher uninsured or Medicaid reliance show worse outcomes.

Racial and ethnic composition remains significant

- Even after controlling for poverty and insurance, counties with larger multiracial and marginalized populations exhibit higher mortality rates.

Overall pattern

- Cancer mortality disparities reflect overlapping socioeconomic, insurance-related, and demographic inequalities.

Our results show a clear and consistent pattern. Poverty is strongly associated with higher cancer mortality about 21 additional deaths per 100,000 for every 10 point increase in poverty. When we add insurance variables, the effect of poverty decreases but remains significant, showing that insurance helps explain some, but not all, of the disparities.

We also find that racial and ethnic composition continues to matter. Even after controlling for poverty and insurance, counties with larger multiracial or marginalized populations still experience higher mortality rates.

Overall, these results show that cancer mortality disparities are shaped by a combination of economic disadvantage, insurance coverage gaps, and demographic inequities.