NC STATE UNIVERSITY

Computational Public Health: Modeling Cancer Inequities and Designing Evidence-Based Policy Interventions

Madison E. Thompson

Department of Computer Science, North Carolina State University

Introduction: Social Determinants of Health

- Social Determinants of Health (SDOH) are the non-medical, cumulative factors that influence individual health outcomes[1]
- Many cancers, including lung, breast, and colorectal, are influenced by modifiable and reversible SDOH risk factors, including chronic stress, healthcare access, and environmental exposures[1]

Figure 1. SDOH as described by the Centers for Disease Control and Prevention.[1]

Introduction: Public Health in North Carolina

- Cancer is the second-leading cause of death in North Carolina, with 151.1 deaths per 100,000 in 2023[2]
- 35% of North Carolinians live in rural NC and have at-risk SDOH metrics compared to their urban peers[3,4]

SDOH Variable	Rural	Urban
Bachelor's Degree	23.0%	40.4%
Low-Income (<$35,000)	27.3%	20.4%
Current Smoking	13.8%	9.1%
Obesity	40.8%	28.6%

Table 1. Select SDOH variables between rural and urban residents of NC.[3]

- SDOH-linked cancers comprise the majority of NC statistics
- Lung (25%), breast (7%), and colorectal (8%) cancer comprise 40% of cancer deaths and were projected to comprise 38.6% of new cancer cases in 2024[5]
- NC has consistently been a national leader in cancer incidence, suggesting targeted policy interventions that consider both rural and urban residents are needed[6]

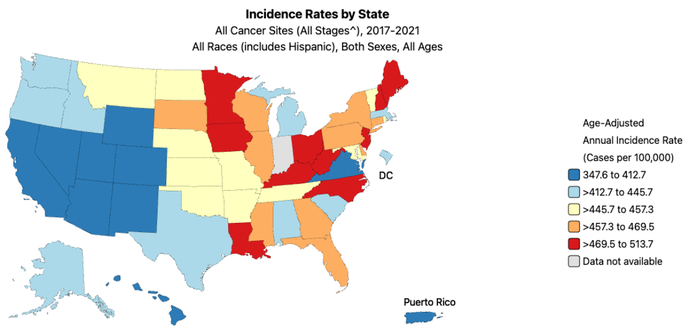

Figure 2. Cancer incidence rates across the US for 2017-2021, showing NC among states with the highest incidence at 475.5 per 100,000.[6]

- This study analyzes SDOH associations with lung, breast, and colorectal cancers across all 100 NC counties, identifies which factors are most closely related to incidence and mortality, and suggests public health policy interventions to address the findings

Methodology: OLS, Ridge, LASSO

Variables

- Age-adjusted incidence and mortality rates per 100,000 individuals as reported by NC State Center for Health Statistics for 2019-2023[7,8]
- County-level SDOH predictors as reported by NC Institute of Medicine for 2021[4]

Models

- Four pooled models were fitted to identify generalizable SDOH associations with cancer outcomes
- OLS and Ridge were fitted to evaluate for multicollinearity after exploratory data analysis, as shown in Figure 3
- Six LASSO models were fitted to determine incidence and mortality outcomes across cancers while reducing some predictors to exactly zero

$$Y_i = \beta_0 + \beta_1(Uninsured\ Adults)_i + \beta_2(PCP)_i + \beta_3(Education)_i + \beta_4(Adult\ Smoking)_i + \beta_5(Poverty)_i + \beta_6(Food\ Insecurity)_i + \beta_7(Transportation)_i + \beta_8(Air\ Pollution)_i + \varepsilon_i$$

Equation 1. Where Y_i represents age-adjusted cancer outcome per 100,000 for observation i, all represented predictors are z-scored, and ε_i is the error term.

$$\text{minimize: } \Sigma(Y_i + \hat{Y}_i)^2 + \lambda\Sigma{\beta_j}^2$$

Equation 2. Where λ is the regularization parameter selected via 10-fold cross-validation to minimize mean squared error (MSE). λ is added to Equation 1 for Ridge regression, which shrinks coefficient estimates toward zero proportionality to account for multicollinearity.

$$\text{minimize: } \Sigma(Y_i + \hat{Y}_i)^2 + \lambda\Sigma|\beta_j|$$

Equation 3. Where Y_i represents age-adjusted cancer outcome per 100,000 for county i, all predictors are z-scored, and λ is the L1 penalty parameter.

Results: Regression Analysis

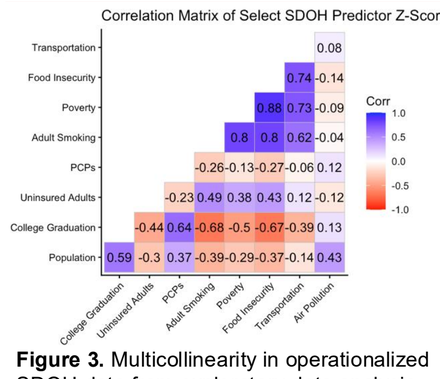

Figure 3. Multicollinearity in operationalized SDOH data from exploratory data analysis.

- Poverty, food insecurity, and transportation show significant multicollinearity
- College graduation is protective against the clustered predictors
- College education and PCP access show moderate collinearity

Model Fit: Pooled Incidence	OLS Estimate	Ridge Estimate
R^2	0.022	0.010
Adjusted R^2	-0.006	-0.018
AIC	3326	2487
Model Fit: Pooled Mortality		
R^2	0.072	0.058
Adjusted R^2	0.039	0.025
AIC	1827	1213

Table 2. OLS and Ridge estimates for for pooled incidence and mortality.

- High AIC values, $R^2 < 0.10$, and adjusted $R^2 < 0.05$ indicate that there is no generalized association between SDOH and cancer outcomes
- This result is expected, as SDOH variables tend to impact specific cancers more than general populations[1,5], however, it raises an interesting conflict with the multicollinearity in Figure 3

Cancer Outcome	R^2	CV-MSE	λ	Variables Retained
Breast Inc.	0.394	37.9	0.358	6
Breast Mor.	0.088	17.5	1.100	1
Lung Inc.	0.489	83.6	0.133	7
Lung Mor.	0.469	31.4	0.092	7
Colorectal Inc.	0.496	31.9	0.142	8
Colorectal Mor.	0.242	8.16	0.327	5

Table 3. LASSO model performance summary, where R^2 is calculated from the CV-MSE, and variables retained shows the number of variables the model found significant. λ is selected to minimize error.

- Lung and colorectal incidence and lung mortality show strong associations between SDOH and cancer outcomes, with 7, 7, and 8 variables retained respectively
- Breast cancer mortality was interestingly unable to be predicted, with λ = 1.100 and only 1 variable retained, despite literature showing a high association with SDOH[9,10]

Predictor	BI	BM	LI	LM	CI	CM
Adult smoking	2.43	-	0.83	2.33	1.10	0.12
Air pollution	10.97	-	1.55	1.68	0.54	-0.05
College graduation	-9.62	-	-8.34	-5.40	-3.83	-0.89
Food insecurity	-	1.06	2.06	-	0.21	0.57
PCPs	3.38	-	0.86	0.86	0.03	-
Poverty	-	-	-	-0.58	0.99	0.12
Uninsured adults	0.35	-	-1.90	-1.53	-0.90	-
Transportation	4.70	-	0.24	-0.65	1.33	-

Table 4. LASSO standardized coefficients for variables retained. A dash indicates that a variable was not retained for a particular outcome.

- College graduation was seen as a consistently protective factor and retained by all meaningful models, aligning with expectations from Figure 3
- Breast cancer incidence showed a strong association with air pollution, indicating that this outcome may be related to built environment more than previously anticipated

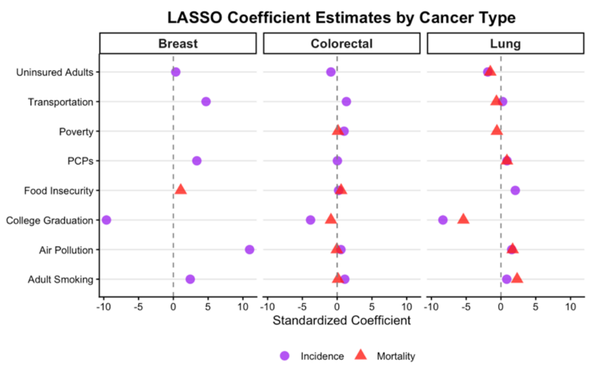

Figure 4. Visual representation of standardized coefficients for retained variables, showing cancer outcomes by type and overlap between incidence and mortality for each.

- Lung incidence and mortality follow the same pattern, indicating a persistent association with SDOH predictors
- Colorectal behaves similarly, with more variables affecting incidence than mortality, indicating a lifecycle shift
- Breast cancer is seemingly unpredictable, with wide ranging influences and low predictive measures, indicating overfitting

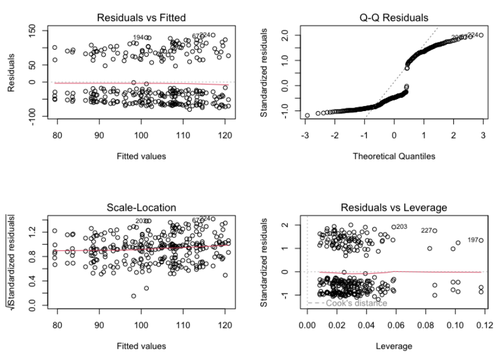

Figure 5. Standard OLS diagnostic plots, indicating violations of linear regression assumptions that justify Ridge regression and LASSO modeling.

The pooled model is weak due to inappropriate aggregation, rather than from data quality issues or outlier influence.

Public Health Policy: Educational Attainment

- The data in Figure 3 and Table 4 demonstrates that college graduation is a protective factor for all cancer outcomes
- Rural populations have significantly lower college graduation rates (23.0%) than their urban peers (40.4%), making them more susceptible to SDOH-associated cancer outcomes
- Public health policies that emphasize improving access to higher education for low-income (Figure 3) and rural (Table 1) residents may be a meaningful intervention for this population
- NC superintendent Maurice "Mo" Green recently began *NC College Connect*, a program that allows public high school students with a GPA of 2.8 or higher to be automatically admitted to 40 public and private NC colleges and universities and all 58 NC community colleges[11]
- This program is an expansion of the College Foundation of North Carolina's *College Application Week* initiative, which allows high schools to apply for college for free during one week per year[12]
- These programs remove red tape from the college application process, allowing rural students to attend college close to home, reducing costs and maintaining family structures
- Programs that fund college for low-income students in NC exist, but not en masse
- Considering the data presented and existing programs, tuition waivers for first-generation college students in NC would be critical to expanding access to higher education to students from rural backgrounds
- Specialized advising for these students would also be monumental, ensuring they can participate in these programs

Public Health Policy: Telehealth Access

- With a population that is 35% rural, NC benefited significantly from Telehealth access made available by federal waivers that were not renewed in the 2025 shutdown[13]
- Governor Stein has passed legislation to support expanded Telehealth in North Carolina, namely the *Rural Health Transformation Program*, which seeks to modernize rural infrastructure, implement AI systems to support overworked rural providers, and teach technological literacy to residents[14]
- Telehealth is also economically beneficial, and has been shown to reduce strain on emergency systems and slow the spread of diseases particularly in rural economies[15]
- Access to Telehealth also increases access to care, resulting in a 15-56% reduction in hospital visits and chronic illness mortality for patients with chronic illnesses[16]
- A state-level Telehealth program to support Stein's five-year initiative, combined with expanded higher education access, has the potential to completely transform public health in rural communities

Limitations

- Breast cancer mortality remains unexplained, indicating that there are confounding variables at play in SDOH associations
- This work explores cancers with known SDOH associations, but does not explore any molecular or genetic data
- The data presented is aggregate from a five-year period and only covers counties in one state

Future Directions

- Studies that combine SDOH associations with genetic data would have significantly more power, as many genes are known to be variably expressed in association with SDOH parameters (epigenetics)
- Data collection from matched counties across the country would add statistical significance to future findings
- Increasing the period analyzed, and setting clear breakpoints for five, ten, and twenty-year lags would more clearly determine whether SDOH maintains association
- A multi-decade study on individuals from at-risk SDOH backgrounds, including evaluation for incidence and mortality, would provide novel case study data
- Modeling to determine whether SDOH associations for incidence and mortality interact significantly differently within cancer types would be novel, particularly if SDOH-associated incidence was considered as a factor in evaluating SDOH-associated mortality

References and Acknowledgements

I would like to thank the US-UK Fulbright Commission for my Summer Institute experience at King's College London, which shaped my interest in health policy inspired me to pursue this work independently. All work expressed is my own, and does not reflect the views, opinions, values, or beliefs of the US-UK Fulbright Commission, King's College London or the Fulbright program.

NATIONAL ACADEMY OF ENGINEERING®

Grand Challenges Scholars Program