HEALTH DISPARITIES IN ALABAMA'S BLACK BELT: A CROSS-SECTIONAL DECOMPOSITION OF THE COUNTY-LEVEL GAP, 2023

Sanket Kanekar, PhD

Department of Economics, Alabama State University

Montgomery, AL 36101

Corresponding author: Sanket Kanekar, PhD

Department of Economics, Alabama State University

915 S. Jackson Street, Montgomery, AL 36101

Email: skanekar@alasu.edu

Word count: approximately 3,800 (excluding abstract, references, tables, and figures)

Tables: 5. Figures: 4.

Keywords: health disparities; rural health; Alabama Black Belt; social determinants of health; Oaxaca-Blinder decomposition.

Funding: None. Conflicts of interest: None declared. Acknowledgments: None.

ABSTRACT

Purpose. Alabama's Black Belt, an 18-county region with a majority African American population and persistent poverty, has long been identified as one of the most disadvantaged areas in the United States. Existing studies typically focus on a single condition or rely on descriptive comparisons. This study quantifies the breadth of the Black Belt health gap across 14 outcomes and decomposes it into components attributable to socioeconomic status, racial composition, and unobserved factors.

Methods. I analyzed all 67 Alabama counties using a 2023 cross-section that combined the American Community Survey (ACS), CDC PLACES, and County Health Rankings (CHR). I compared 18 Black Belt counties with 49 non-Black Belt counties using ordinary least squares with heteroskedasticity-robust (HC3) standard errors, a threefold Oaxaca-Blinder decomposition (endowments, coefficients, interaction), and causal mediation analysis with bootstrap confidence intervals (2,000 replications) to estimate the share of the gap mediated by socioeconomic conditions and racial composition.

Findings. Black Belt counties exhibited worse outcomes across nearly every indicator: higher diabetes prevalence (17.1% vs. 12.7%), obesity (47.3% vs. 40.0%), hypertension (46.9% vs. 39.7%), and premature mortality (14,837 vs. 12,092 years of life lost per 100,000), and lower life expectancy (71.3 vs. 73.1 years). The Oaxaca-Blinder decomposition attributed 62% to 136% of the raw gap to differences in endowments (poverty, income, education, and racial composition) rather than to differences in returns to those endowments. Causal mediation indicated that 84% to 123% of the total Black Belt effect operated indirectly through socioeconomic and racial mediators (95% bootstrap CIs excluded zero for every outcome), while the average direct effect was statistically indistinguishable from zero in every model.

Conclusions. The Alabama Black Belt health penalty is large, broad, and almost entirely attributable to differences in measurable structural conditions. There is no detectable residual Black Belt effect once poverty, income, education, and racial composition are accounted for. Policies that reduce poverty, expand healthcare coverage, and invest in human capital in the region are likely to close most of the observed gap.

INTRODUCTION

Rural health disparities remain a pressing public health challenge in the United States. Rural communities experience higher rates of chronic disease, shorter life expectancy, and greater barriers to healthcare access than urban and suburban areas.[1,2] A growing body of evidence indicates that these disparities have widened in recent years, particularly for mortality outcomes, and that socioeconomic disadvantage is a primary driver.[3,4] Within the rural South, these challenges are especially acute in regions shaped by the historical legacies of slavery, disinvestment, and structural racism.[5]

Alabama's Black Belt is among the most disadvantaged regions in the nation. The name originally referred to the dark, fertile soil of central Alabama but became synonymous with the region's large African American population, a demographic legacy of the plantation economy and slavery.[6] The Black Belt encompasses 18 counties stretching across the center of the state. These counties are characterized by persistent poverty, population decline, limited healthcare infrastructure, and some of the worst health outcomes in the country.[7,8] In 2020, the region had a 30% higher COVID-19 death rate than non-Black Belt counties in Alabama.[9] County-level analyses have shown that high-poverty rural counties in the South carry a disproportionate burden of cardiovascular and chronic disease mortality.[10]

Despite longstanding recognition of these challenges, the empirical literature on Black Belt health has important gaps. Most existing studies focus on a single condition such as diabetes,[11] hypertension,[12] HIV,[13] or respiratory illness,[14] rather than documenting the breadth of disparities across multiple domains. Few studies have used formal decomposition methods to quantify the share of the Black Belt gap that is attributable to measurable structural conditions versus unobserved residual factors. Long and colleagues[4] showed that socioeconomic variables account for much of the rural mortality penalty at the national level, but this question has not been examined within a specific high-disparity region using modern decomposition tools.

This study addresses these gaps. Using a 2023 cross-section of all 67 Alabama counties, I document the Black Belt health gap across 14 outcomes spanning chronic disease, mortality, behavioral risk, and access to care. I then apply two complementary decomposition strategies, the Oaxaca-Blinder threefold decomposition and causal mediation analysis, to estimate how much of the gap operates through differences in socioeconomic endowments and racial composition versus a residual Black Belt effect that is not explained by those factors. The research questions are: (1) How large is the Black Belt health gap, and how broad is it across outcomes? (2) What share of the gap is attributable to differences in measurable structural conditions? (3) Is there a residual Black Belt effect after accounting for those conditions?

METHODS

Study Design and Setting

I conducted a county-level cross-sectional ecological analysis of all 67 Alabama counties for 2023. The unit of observation is the county. The exposure of interest is membership in the Black Belt region, defined here as the 18-county historical core identified by the Alabama Department of Public Health: Barbour, Bullock, Butler, Choctaw, Crenshaw, Dallas, Greene, Hale, Lowndes, Macon, Marengo, Montgomery, Perry, Pickens, Pike, Russell, Sumter, and Wilcox. The remaining 49 counties form the comparison group.

The single-year cross-sectional design was adopted for two reasons. First, the ACS 5-year estimates available for adjacent years (e.g., 2019–2023 and 2020–2024) share four of five underlying survey years and therefore do not provide independent observations. Second, the CDC

PLACES estimates are revised across release years using updated BRFSS waves and modeling assumptions, which introduces measurement non-comparability that is difficult to disentangle from true temporal change. Restricting attention to the most recent year for which all three data systems publish coterminous estimates (2023) yields a defensible single-year comparison.

Data Sources

Three publicly available data systems were combined.

1. American Community Survey 5-Year Estimates, 2019–2023. Census tract-level ACS estimates were aggregated to the county level using Census API endpoints. Variables included total population, the share of the population that is Black or African American, the share Hispanic, the poverty rate, median household income, the unemployment rate, and the share of adults aged 25 years or older with a bachelor's degree or higher.

2. CDC PLACES Project, 2024 release. PLACES provides modeled, age-adjusted county-level prevalence estimates of self-reported health conditions and behaviors based on multilevel regression with poststratification of Behavioral Risk Factor Surveillance System (BRFSS) data.[15] Outcomes used here include diabetes, obesity, hypertension, heart disease, stroke, COPD, current smoking, physical inactivity, fair or poor self-rated health, frequent mental distress, frequent physical distress, and uninsurance.

3. County Health Rankings and Roadmaps (CHR), 2024 release. CHR aggregates administrative and survey measures from federal sources including the National Center for Health Statistics, the Bureau of Labor Statistics, and the Health Resources and Services Administration.[16] From CHR, I obtained premature death (years of potential life lost before age 75 per 100,000 population), life expectancy at birth, and county uninsurance rates.

Statistical Analysis

Three estimators were applied sequentially. First, I estimated ordinary least squares (OLS) regressions of each health outcome on the Black Belt indicator with progressively richer control sets, specified as follows.

Model 1 (unadjusted):

$$y_i = \alpha + \beta \cdot BB_i + \varepsilon_i \quad (1)$$

Model 2 (adjusted for socioeconomic status):

$$y_i = \alpha + \beta \cdot BB_i + \gamma' SES_i + \varepsilon_i \quad (2)$$

Model 3 (adjusted for socioeconomic status and racial composition):

$$y_i = \alpha + \beta \cdot BB_i + \gamma' SES_i + \delta \cdot Black_i + \varepsilon_i \quad (3)$$

In these specifications, y_i denotes the health outcome in county i; BB_i is an indicator that equals one for Black Belt counties; SES_i is a vector of socioeconomic covariates (poverty rate, median household income, unemployment rate, and share of adults with a bachelor's degree or higher); $Black_i$ is the share of the county population that is Black or African American; and ε_i is a county-specific error term. Heteroskedasticity-consistent standard errors of the HC3 form, appropriate for small samples, were used throughout.[17]

Second, I estimated a threefold Oaxaca-Blinder decomposition.[18,19] Separate OLS regressions were estimated for the Black Belt and non-Black Belt groups. Letting $\bar{X}_m$ and β_m denote the mean covariate vector and coefficient vector for group g, the raw gap is decomposed as in Equation (4).

$$E[y|BB = 1] - E[y|BB = 0] = (\bar{X}_{BB} - \bar{X}_{NBB})'\beta_{NBB} + \bar{X}_{NBB}{}'(\beta_{BB} - \beta_{NBB}) + (\bar{X}_{BB} - \bar{X}_{NBB})'(\beta_{BB} - \beta_{NBB}) \quad (4)$$

The first term on the right-hand side is the endowments component, which captures differences attributable to mean differences in covariates evaluated at the non-Black Belt coefficient vector (the advantaged reference). The second term is the coefficients component, which captures differences in returns to those covariates. The third term, the interaction, reflects the joint contribution of differences in covariates and returns. Covariates entered the decomposition were poverty, median household income, unemployment, share of adults with a bachelor's degree or higher, and share Black.

Third, I conducted causal mediation analysis following the counterfactual framework of Imai, Keele, and Tingley.[20] For each outcome and each mediator block (socioeconomic status alone; socioeconomic status plus race), I estimated the system in Equations (5) and (6).

$$M_k = a_{0,k} + a_k \cdot BB + u_k \quad (5)$$

$$y = c_0 + c' \cdot BB + \sum_{k=1}^{K} b_k M_k + v \quad (6)$$

In linear models, the average causal mediation effect (ACME, or indirect effect) equals the sum across mediators of $a_k \cdot b_k$, the average direct effect (ADE) equals c', and the total effect equals their sum. The proportion of the total effect mediated is ACME divided by the sum of ACME and ADE. Bootstrap 95% confidence intervals were obtained from 2,000 nonparametric resamples.

All analyses were conducted in Python 3.13 using statsmodels 0.14.6.[21] Replication code is available from the corresponding author on reasonable request.

Sensitivity Analyses

To assess robustness of the main results to definitional and sample choices, the fully adjusted OLS specification (Model 3) was re-estimated under five alternative samples. (S1) Montgomery County, the largest Black Belt county and the source of approximately 40% of the region's total population, was excluded. (S2) Crenshaw County, the Black Belt county with the lowest share of Black residents (23.4%) and a poverty rate indistinguishable from the non-Black Belt average, was excluded. (S3) A restrictive core 12-county definition was applied, retaining as Black Belt only the 12 counties with the highest share of Black residents (range 47% to 82%). (S4) A 17-county definition that reclassified Crenshaw as non-Black Belt while retaining it in the sample was applied. (S5) An expansive 23-county definition added the five non-Black Belt counties (Conecuh, Clarke, Monroe, Chambers, and Mobile) whose share of Black residents exceeded 35%.

RESULTS

Sample Characteristics

Table 1 reports group means for 2023. The 18 Black Belt counties had higher poverty rates (24.6% vs. 16.2%), lower median household incomes ($42,896 vs. $58,347), higher unemployment (7.2% vs. 5.2%), lower educational attainment (17.0% vs. 20.4% with a bachelor's degree), and a higher share of Black residents (57.2% vs. 17.6%). Black Belt counties were also smaller in population (mean 30,872 vs. 91,807), reflecting the rural character of the region.

The Black Belt Health Gap

Across the 14 outcomes considered, Black Belt counties consistently fared worse (Table 2, columns under Unadjusted). Diabetes prevalence was 4.4 percentage points higher (17.1% vs. 12.7%, $P < .001$); obesity was 7.3 points higher ($P < .001$); and hypertension was 7.2 points higher ($P < .001$). Premature death (years of potential life lost before age 75) was 22.7% higher in Black Belt counties (14,837 vs. 12,092 per 100,000; $P < .001$), and life expectancy was 1.8 years lower (71.3 vs. 73.1 years; $P < .001$). The pattern holds for behavioral risk factors (smoking +2.6

percentage points; physical inactivity +6.2 percentage points) and self-rated health (fair or poor health +6.3 percentage points).

Adjusted Gap

Table 2 also reports the residual Black Belt coefficient after sequentially adding socioeconomic controls (Model 2) and racial composition (Model 3). In the fully adjusted model, the Black Belt coefficient was statistically indistinguishable from zero for every outcome ($P > .19$ in all 14 cases), including diabetes (−0.11; P = .65), obesity (+0.56; P = .60), hypertension (+0.35; P = .42), premature death (−582.4; P = .44), and life expectancy (+0.12 years; P = .87). Model R^2 values ranged from 0.69 (life expectancy) to 0.96 (hypertension and diabetes), indicating that the controls explain the bulk of the cross-county variation.

Oaxaca-Blinder Decomposition

Table 3 presents the threefold decomposition. The endowments component, which is the share of the gap attributable to differences in poverty, income, education, and racial composition evaluated at non-Black Belt returns, accounted for 62% to 136% of the raw gap across outcomes. The coefficients component (differences in returns to these characteristics) was small and consistently negative across chronic disease outcomes, indicating that returns to socioeconomic and demographic characteristics in Black Belt counties were not systematically different from returns elsewhere in Alabama. The interaction terms were also modest. For three outcomes (frequent mental distress, premature death, and life expectancy), the endowments component exceeded 100% of the gap, implying that if Black Belt counties had non-Black Belt covariate values, they would, on the basis of the estimated returns, have outcomes slightly better than those of non-Black Belt counties.

Sensitivity to Sample and Definition

Table 5 reports the adjusted Black Belt coefficient (Model 3) under five alternative samples. The main null result was robust to dropping Montgomery County (S1: 0 of 14 outcomes statistically significant at the 5% level), dropping Crenshaw County (S2: 0 of 14), and applying the 17-county strict definition (S4: 0 of 14). Two definition changes did produce nominally significant residual coefficients. The 12-county core definition (S3), restricted to the 12 Black Belt counties with the highest share of Black residents, produced positive residuals on seven outcomes at $P < .05$: hypertension (+1.10 percentage points; $P = .029$), heart disease (+0.33; $P = .039$), stroke (+0.25; $P = .031$), COPD (+0.56; $P = .032$), fair or poor self-rated health (+1.64; $P = .013$), frequent physical distress (+0.67; $P = .015$), and uninsurance (+1.19; $P = .012$). The 23-county expansive definition (S5) produced four negative-direction residuals at $P < .05$ (heart disease −0.25; stroke −0.17; smoking −0.81; frequent mental distress −0.44), reflecting the inclusion of urbanized Mobile County (population approximately 413,000), whose health profile differs from that of the rural Black Belt. None of the seven S3 results survived Benjamini-Hochberg false discovery rate control at $q = .05$ across the 14 simultaneous outcome tests (smallest unadjusted $P = .012$; required threshold for the smallest P-value = .0036). The S3 pattern is therefore consistent with a small place-based effect at the demographic extreme of the region rather than with robust evidence of a residual Black Belt penalty. No outcome retained a multiple-testing-adjusted significant residual under any sensitivity sample.

Causal Mediation

Table 4 reports causal mediation results. With socioeconomic and racial mediators, the average causal mediation effect (ACME) accounted for between 84% and 123% of the total Black Belt effect across outcomes. For example, the total effect on diabetes prevalence (4.36 percentage points) was almost entirely mediated (ACME = 4.47 percentage points; 95% CI, 3.31 to 5.56) with a negligible direct effect (−0.11 percentage points). Bootstrap confidence intervals on the ACME excluded zero for every outcome, while bootstrap intervals on the average direct effect (ADE) included zero in every case. With socioeconomic mediators alone (excluding racial composition), proportion-mediated estimates ranged from 57% to 123% across chronic disease outcomes, indicating that racial composition contributes a substantial mediating share for some outcomes (e.g., diabetes 65%, obesity 57%, hypertension 57%) but is offset by socioeconomic status alone for others (heart disease 111%, COPD 123%).

DISCUSSION

This analysis yields three findings. First, the Alabama Black Belt exhibits a large and broad health penalty: across 14 outcomes spanning chronic disease, mortality, behavioral risk, and access to care, Black Belt counties fared worse, in many cases by clinically meaningful margins. Second, nearly all of this penalty is attributable to differences in measurable structural conditions, namely poverty, income, education, employment, and racial composition. The Oaxaca-Blinder endowments component absorbed 62% to 136% of the gap, and causal mediation indicated that 84% to 123% of the total effect operated indirectly through these mediators. Third, no outcome exhibited a statistically detectable residual Black Belt effect once these conditions were accounted for. The 95% bootstrap confidence intervals on the average direct effect included zero in all 14 cases.

These findings sharpen the policy implication. The Black Belt's health disadvantage is not a mystery that requires a region-specific structural explanation beyond what is captured by socioeconomic status and racial composition. The region is unhealthy because it is poor, has lower educational attainment, has high unemployment, and has a population whose racial composition reflects the legacy of historical disinvestment. Policies that act on these channels, including Medicaid expansion, federal nutrition assistance, expansion of the Earned Income Tax Credit, investment in public schooling and healthcare workforce, and broadband and transportation infrastructure, have a strong empirical case as instruments for closing the gap.

The findings should be read alongside four caveats that limit causal interpretation. First, with 67 counties, statistical power to detect a small residual Black Belt effect is limited; absence of a statistically significant direct effect is not equivalent to evidence that no direct effect exists. Second, because the Black Belt indicator and the share of Black residents are highly correlated (Pearson $r = 0.79$), the model that includes both estimates the residual effect of being a Black Belt county net of racial composition. This is the correct quantity for the question asked, but it should not be interpreted as a test of whether place per se has any influence. By construction, including the defining characteristic of the region absorbs much of the variation. The socioeconomic-only mediation specification reported in Table 4 re-estimates the indirect-effect share without the racial-composition control. Third, several health outcomes are derived from CDC PLACES, which produces model-based prevalence estimates from BRFSS rather than direct measurements. The point estimates are subject to model uncertainty that is not propagated into the standard errors reported here, and small-county estimates carry wider underlying confidence bands than large-county estimates. Fourth, 14 outcomes are tested simultaneously in the main analysis and across six samples in the sensitivity analysis (84 tests in total). The reported P-values are unadjusted for multiplicity. Benjamini-Hochberg false discovery rate control at $q = .05$ was applied as a check and produced no significant residual coefficient under any specification or sample, consistent with the unadjusted main result.

The sensitivity analyses sharpen rather than overturn the main finding. The null residual is preserved under every conventional definition variation and survives multiple-testing correction in the most demographically restrictive sample. In the 12-county core definition, which retains only the counties with the highest share of Black residents, seven outcomes exhibit nominally positive residual coefficients (hypertension, heart disease, stroke, COPD, self-rated health, frequent physical distress, and uninsurance), but none survives Benjamini-Hochberg false discovery rate control at $q = .05$. This pattern is consistent with one of two interpretations: a small nonlinear effect of racial composition (returns to majority-Black status differ from those of plurality-Black status), or a place-based mechanism that operates only in counties where structural disadvantage is most concentrated and that is too small to detect with $N = 12$ plus 49 after multiple-testing adjustment. Both interpretations reinforce, rather than challenge, the broader structural account: residual effects emerge, if at all, only where structural conditions are most severe, and they do not survive correction for the multiplicity of outcomes tested. Future work using spatial regression or matched designs in larger samples may help adjudicate between these alternatives.

A further interpretive note concerns cancer. Age-adjusted incidence from the State Cancer Profiles (NCI/CDC) is higher in Black Belt counties for overall cancer (+13.8 per 100,000), prostate cancer (+42 per 100,000), and lung cancer (+19.4 per 100,000), yet self-reported cancer prevalence in PLACES is lower in Black Belt counties. This pattern is consistent with survival bias: higher incidence combined with lower prevalence implies higher case fatality. The discrepancy warrants more focused investigation than is possible within the present study design and is not analyzed further here.

The use of the threefold Oaxaca-Blinder decomposition and the counterfactual mediation framework distinguish this study from prior descriptive comparisons of Black Belt health. Both estimators converge on the same finding: the gap is structural and operates through observable channels, which strengthens confidence in the conclusion. The substantive contribution is not that socioeconomic status matters, which is well established, but rather that, within this specific high-disparity region and using two methodologically distinct decomposition strategies, the residual unexplained component is small and statistically indistinguishable from zero.

CONCLUSION

The Alabama Black Belt's health penalty is large, broad, and structurally explained. Differences in poverty, income, education, employment, and racial composition account for nearly all of the observed gap across 14 outcomes; no outcome shows a residual Black Belt effect that is statistically distinguishable from zero. These results imply that broad-based investments targeting socioeconomic channels, rather than region-specific interventions premised on unobserved place-based mechanisms, are the most direct route to closing the gap. They also caution against interpretive frameworks that treat the Black Belt as a sui generis health environment. The disadvantage is real and severe, but it is not in the soil; it is in the structural conditions that have accumulated on top of it.