

County-Level Risk Analysis of Healthcare Access and Racial Disparities in Preventable Hospitalizations

AUTHOR

Arturo Espana

PUBLISHED

July 25, 2025

Introduction

Hospital care leads as the most expensive segment of U.S. healthcare, with \$33.7 billion spent on 3.5 million preventable hospitalizations in 2017 alone. Preventable hospitalizations are measured as patients who are admitted for conditions such as asthma, diabetes, and hypertension, which are among many conditions that potentially could have been resolved in outpatient settings. Beyond their steep financial toll, preventable hospitalizations expose systemic flaws in healthcare, highlighting inequities and suboptimal care delivery.

Early data exploration revealed that counties with larger Black and Hispanic populations experience significantly higher rates of preventable hospitalizations, highlighting entrenched racial disparities. This pattern highlighted the importance of identifying group-specific factors that contribute to inadequate care. While decades of research have documented overall disparities in hospitalization rates, few studies examine how the underlying predictors contrast by race. Given the lack of research, our project investigates how disparities in healthcare access contribute to preventable hospitalizations across racial groups at the county level.

Although discussions of healthcare access often focus on clinician supply and uninsured rates, our analysis indicates that social determinants—such as high school completion and unemployment—are more strongly associated with preventable hospitalizations among counties with higher People of Color (POC) populations. To investigate these dynamics, we employ race-stratified models at the county level, applying random forests and negative binomial regression to assess how healthcare access and socioeconomic factors jointly influence preventable hospitalization rates.

Key Findings:

- Structural inequities, particularly in education and unemployment, exert a greater influence on preventable hospitalizations among POC populations than clinical access alone.
- Rurality and clinician supply paradoxically correlate with higher hospitalization rates in some contexts, suggesting systemic inefficiencies such as fragmented care or over-hospitalization.
- Interventions must be race-conscious: Expanding health infrastructure alone may reduce preventable hospitalizations for White populations, but POC communities require integrated strategies that address upstream socioeconomic barriers.

By mapping these group-specific drivers, our research challenges one-size-fits-all approaches to reducing preventable hospitalizations. These findings can inform more targeted policies and interventions aimed at addressing structural inequities driving preventable hospitalizations in marginalized communities.

Ultimately, acknowledging how disparities in healthcare access, shaped by both socioeconomic status and race, contribute to preventable hospitalization is essential for building a more equitable and effective healthcare system.

Data

Dataset Sources:

Our analysis uses the 2025 County Health Rankings & Roadmaps (CHRR) dataset, a publicly available resource that aggregates U.S. county-level data across a wide range of health indicators, infrastructure metrics, and socioeconomic variables. CHRR compiles data from multiple sources, including the Behavioral Risk Factor Surveillance System (BRFSS), Centers for Medicare & Medicaid Services (CMS), and the American Community Survey (ACS). All variables reflect the most recent available measurements as of 2025.

The unit of analysis is the county, with each observation representing a U.S. county identified by its FIPS code. This allows for geographically granular analysis of disparities in healthcare access and outcomes.

Outcome Variables:

Our primary outcomes are race-stratified preventable hospitalization rates, defined as inpatient admissions for ambulatory care-sensitive conditions that are considered avoidable with timely, high-quality outpatient care. Specifically we use

- Preventable Hospital Stays (White): Number of preventable hospital stays among non-Hispanic White populations
- Preventable Hospital Stays (POC): A composite rate for People of Color, computed as a population-weighted average of subgroup-specific rates for Black, Hispanic, American Indian/Alaska Native, and Asian/Pacific Islander populations.

Each rate is normalized per 10,000 individuals in the relevant racial group to allow for fair comparison across counties with different population sizes.

Predictor Variables:

We selected predictor variables based on three categories.

- Clinical Access Factors: Primary Care Physician providers per 10,000 residents, Mental Health providers per 10,000 residents, Dentist providers per 10,000 residents, Other than Primary Care Providers per 10,000 residents, Mammography Screening rates, Uninsured Rate
- Socioeconomic and Structural Factors: Income Inequality (Gini coefficient), High School Completion rate, Unemployment rate, Percent of population living in rural areas, Percent of population with a disability, Percent not proficient in English, Percent of population with Severe Housing Problems, Broadband access

Data preparation & Transformations:

For the final modeling stage, all variables were cleaned, scaled, and transformed to enhance interpretability and ensure comparability across counties. Continuous predictors were standardized (mean = 0, standard deviation = 1), and skewed variables—such as the unemployment rate—were log-transformed to mitigate the influence of extreme values.

To account for variation in population size, we included log-transformed population offsets (race-specific) in the negative binomial models. These offsets are essential for accurately modeling preventable hospitalization rates using count-based methods.

Because disaggregated data were unavailable for many race-specific predictors, we constructed a composite People of Color (POC) group by computing a population-weighted average of preventable hospitalization rates and mammography screening rates for Black, Hispanic, American Indian/Alaska Native, and Asian/Pacific Islander subgroups. This enabled consistent comparisons with non-Hispanic White populations while preserving interpretability.

For the random forest models, we created a racial makeup classification variable to obtain a more precise depiction of each county's demographic structure. To mark a county as a majority of one race, we required the race variable in that county to reach a threshold of 0.7. To determine a county of two majority races, we required the top two races to add to 0.6, with the larger race proportion being listed first. This resulted in different racial subsets such as White, White-Black, White-Hispanic, Black-White, and Hispanic-White, which allowed for comparison of variable importance among counties with different populations. As White counties outnumbered all minority counties by a large margin, we aggregated Black-White, Hispanic-White, Hispanic, and Black counties into a POC subset, which allowed us to compare POC counties to White counties in terms of variable importance.

To ensure data quality, counties with substantial missingness in any key outcome or predictor variable were excluded. The final analytic dataset consists of complete county-level observations suitable for robust, race-stratified modeling and exploratory analysis.

Methods

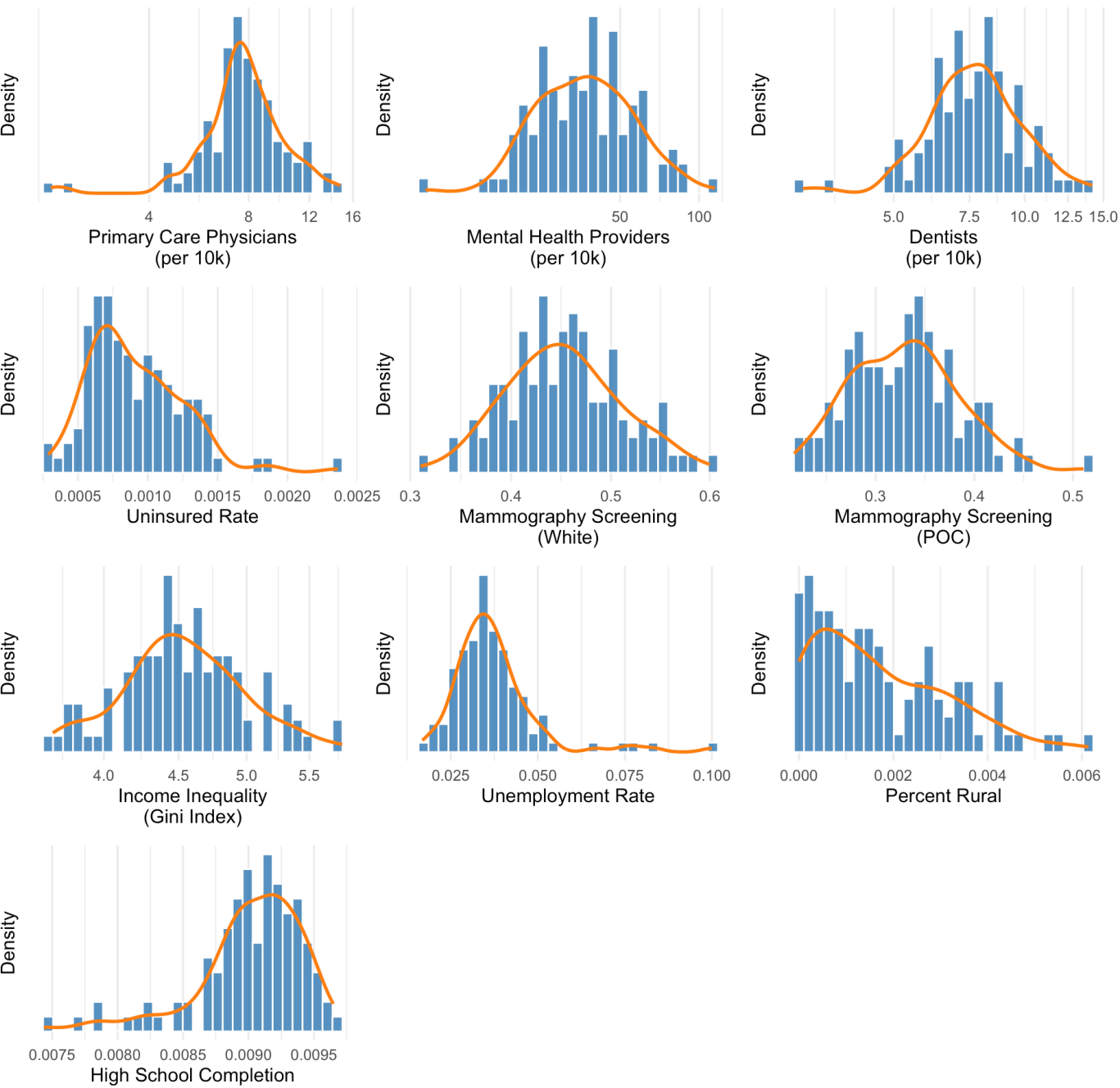
Exploratory Data Analysis:

Before formal modeling, we performed exploratory data analysis (EDA) to understand variable distributions, detect racial disparities in outcomes, and identify geographic and structural patterns relevant to our modeling framework.

Histograms:

We first examined the distributions of clinical and socioeconomic predictors to better understand our variables.

Figure 1. Distribution of Healthcare Access and Socioeconomic Predictors



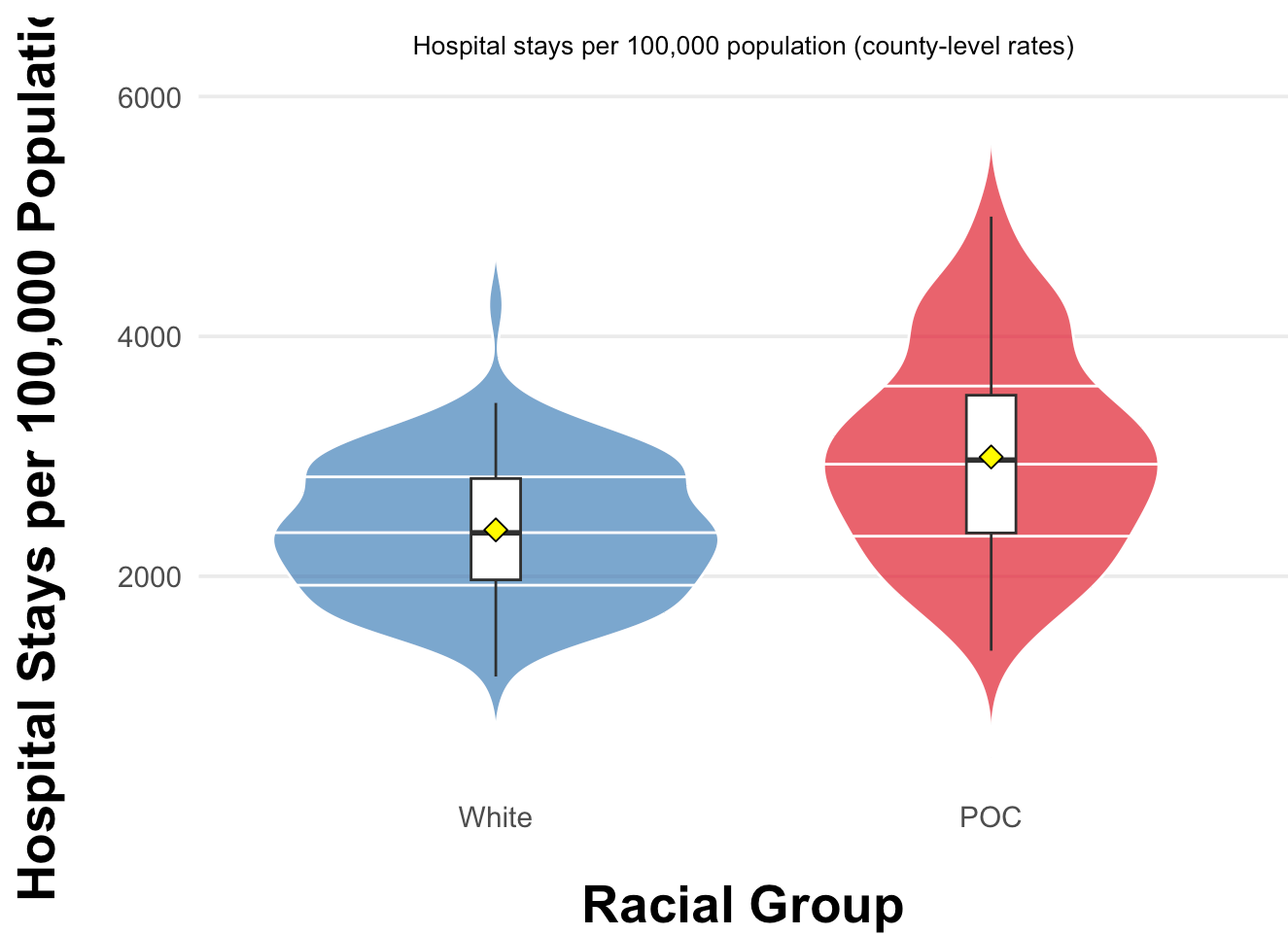
Key Patterns (Figure 1):

- **Uninsured Rate:** The distribution was strongly right-skewed, with a tail of counties exceeding 25% uninsurance—often concentrated in Medicaid non-expansion states.
- **Unemployment:** Similarly skewed, with elevated rates in economically distressed regions.
- **Rurality:** Displayed a bimodal distribution, capturing the divide between urbanized and predominantly rural counties.

To stabilize variance and reduce outlier influence, we applied log-transformations to right-skewed predictors, preserving interpretability and zero values.

Violin Plots:

To assess racial disparities, we visualized the distribution of preventable hospitalizations separately for White and POC populations.



Key Patterns (Figure 2):

- POC counties exhibit consistently higher median rates of preventable hospitalizations compared to White counties.
- The distribution for POC counties also shows greater variability, with a wider range of extreme values, suggesting that many POC communities face systemic barriers to timely, effective outpatient care.

These disparities highlight the importance of modeling outcomes separately by race.

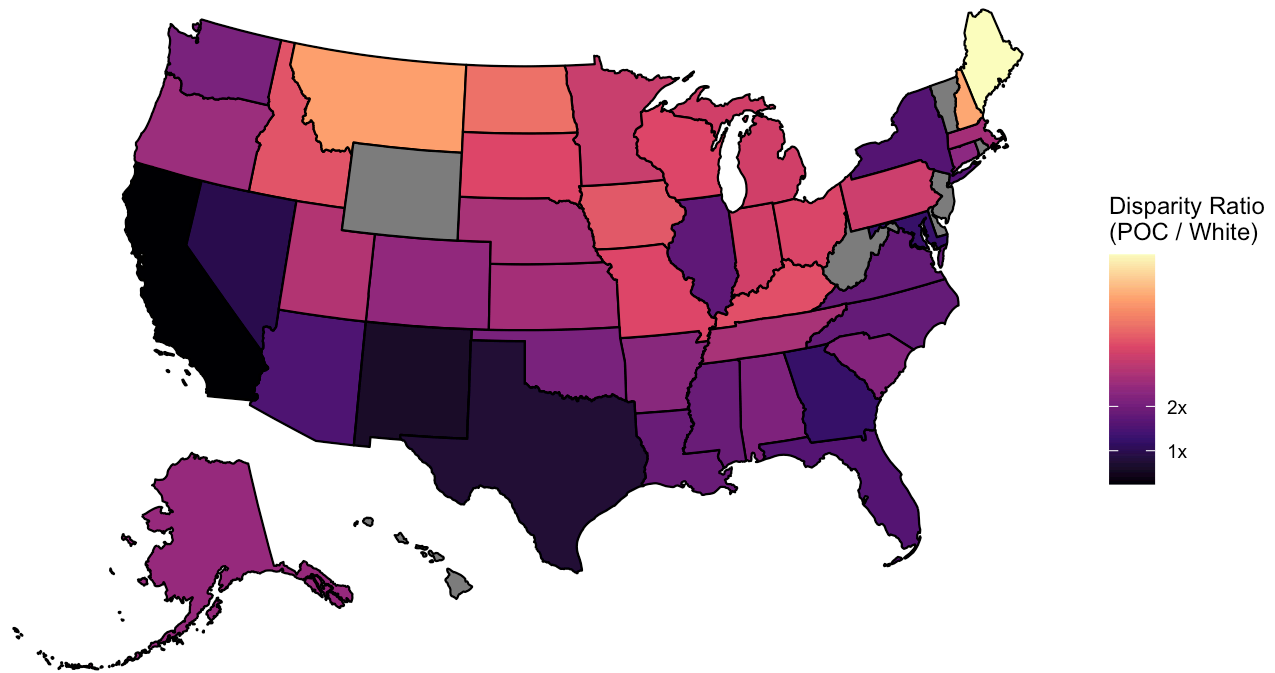
Choropleth Maps:

We aggregated county-level data to compute state-level disparity ratios, defined as:

Disparity Ratio = $\frac{\text{Preventable Hospital Stays}_{\text{POC}}}{\text{Preventable Hospital Stays}_{\text{White}}}$

Figure 4. State-Level Disparities in Preventable Hospital Stays

Ratio of POC to White hospitalization rates



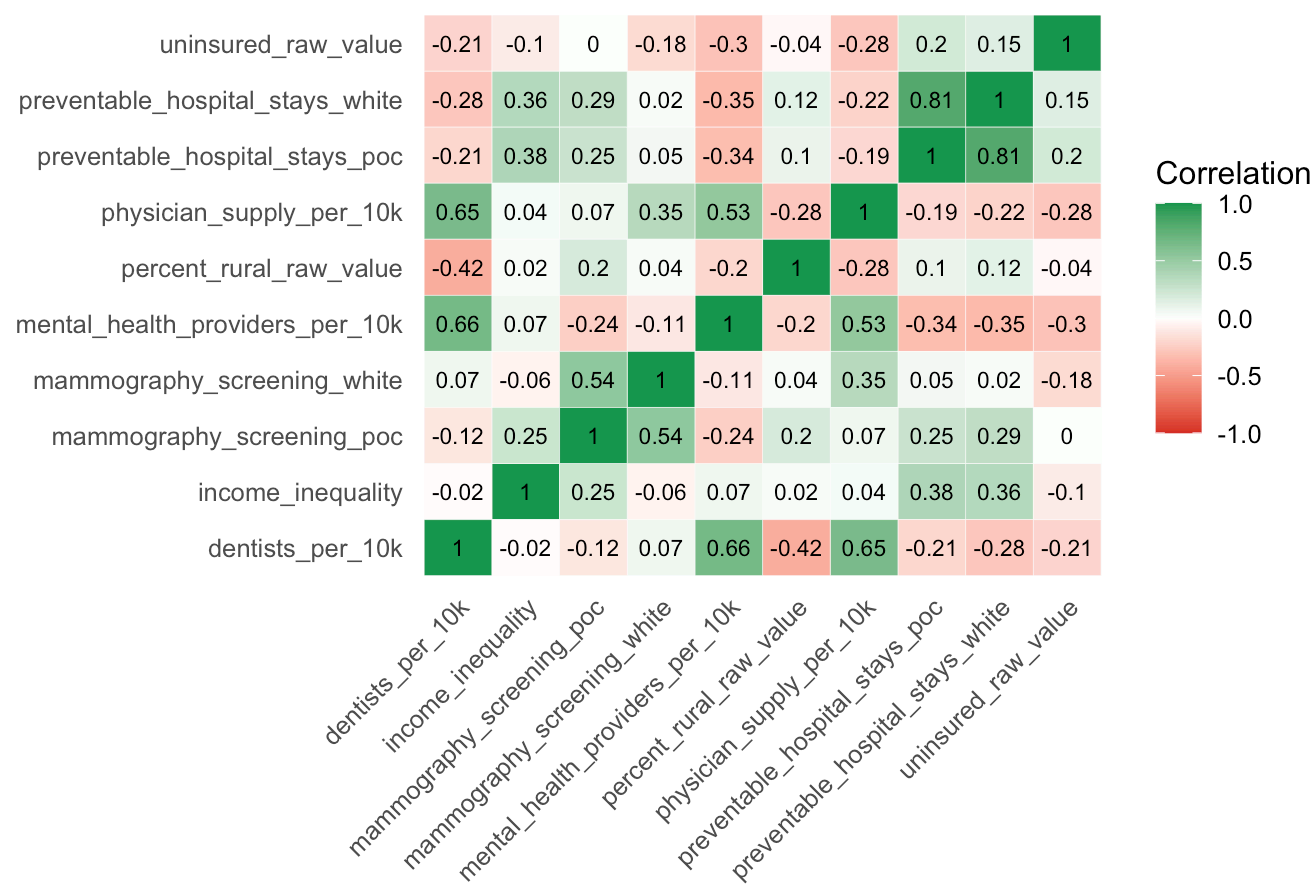
Key Patterns (Figure 3):

- Disparities were most severe in Southern and Midwestern states, where ratios often exceeded 2x.
- Ratios were closer to parity in some Western and Northeastern states, though disparities still persisted.

Correlation Heatmap:

To assess multicollinearity and support variable selection, we computed Pearson correlations among outcomes and key predictors—including clinical access and socioeconomic indicators.

Figure 3. State-Level Disparities in Preventable Hospital Stays



Key Patterns (Figure 4)

- Most predictors showed only weak-to-moderate pairwise correlations, indicating they contribute distinct information and minimizing concerns about overfitting.
- Moderate correlations were observed among provider supply variables (e.g., dentists, physicians, mental health), reflecting shared infrastructure without full redundancy.

These findings justified retaining a broad set of predictors in our race-stratified regression models.

Statistical Modeling Approach:

To quantify the relationship between healthcare access, socioeconomic conditions, and preventable hospitalizations—and how these relationships vary by race—we used two modeling strategies.

1. Random Forests for flexible, nonparametric variable importance analysis
2. Negative Binomial Regression for interpretable inference and effect estimation

Random Forests (Exploratory, Nonparametric Analysis)

Purpose and Justification:

Random Forests were used to identify the most influential predictors of preventable hospitalizations without assuming linearity or specific distributions. This tree-based ensemble method is well-suited for high-dimensional,

multicollinear data and can capture nonlinear relationships and interaction effects that traditional regression models may miss.

While more complex machine learning models (e.g., XGBoost) may offer marginally better predictive accuracy, we prioritized interpretability and transparency. Our goal was not prediction, but insight into the relative importance of predictors—especially as it varies across racial contexts.

Implementation:

We trained separate Random Forest models for majority-White and majority-POC counties. Each forest was constructed using 500 decision trees, with each tree trained on a bootstrapped sample of the data and a random subset of predictors at each split. Final predictions were averaged across all trees.

Two variable importance metrics were extracted:

- %IncMSE: Measures the increase in prediction error when a variable is permuted, reflecting its importance to model accuracy
- IncNodePurity: Measures the cumulative reduction in node impurity from splits involving that variable, indicating contribution to decision-making across the forest

Negative Binomial Regression (Primary Inference Model)

Purpose and Justification:

To estimate the direction and magnitude of predictor effects while accounting for overdispersed count outcomes, we used Negative Binomial (NB) regression. Poisson models assume equal mean and variance—a condition violated in our data, as confirmed by formal overdispersion tests. The NB model relaxes this assumption, offering both statistical rigor and coefficient interpretability.

Stratification Strategy:

Separate NB models were estimated for:

- Non-Hispanic White populations
- Composite People of Color (POC) populations

This approach allows for direct comparisons of how predictors influence outcomes across racial groups.

Model Specification:

(Y_i) : Preventable hospitalizations in county i

(Pop_i) : Race-specific population in county i

(\mathbf{X}_i) : Vector of scaled predictors for county i

Model:

$$\log(E[Y_i]) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \log(Pop_i)$$

Where $\log(\text{Pop}_i)$ is the **offset**, adjusting for population size so the model estimates rates rather than raw counts.

Predictors By Category:

Clinical Access: Primary Care Physician Supply, Dentist supply, Mental health Provider supply, Mammography Screening

Affordability: Uninsured Rate, Income Inequality

Social/Geographic Factors: High school Completion, Unemployment, Rurality

Outcome: Preventable Hospitalizations per 10,000 people in the respective racial group

Model Diagnostics and Robustness Checks

To ensure model validity and reliability, we conducted the following checks:

- Overdispersion Tests: Confirmed Poisson model misspecification; validated the need for Negative Binomial.
- Residual Analysis: Pearson residuals were plotted against fitted values to assess model fit and detect systematic bias.
- Multicollinearity Check: Variance Inflation Factors (VIF) were computed for all predictors; all values were < 5 , indicating acceptable collinearity levels.
- Robust Standard Errors: Heteroskedasticity-consistent (HC0) standard errors were used to ensure reliable inference.

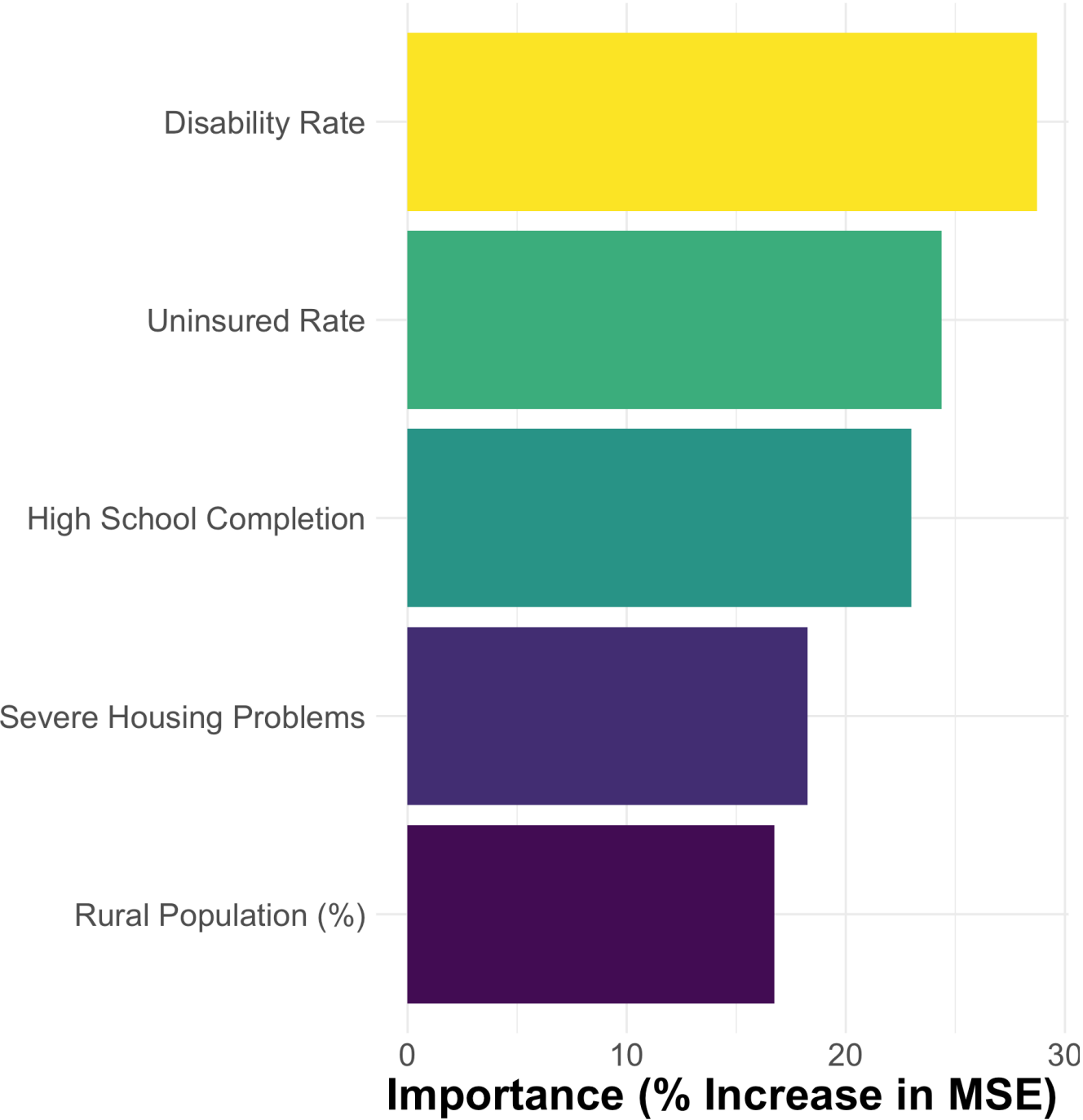
Results

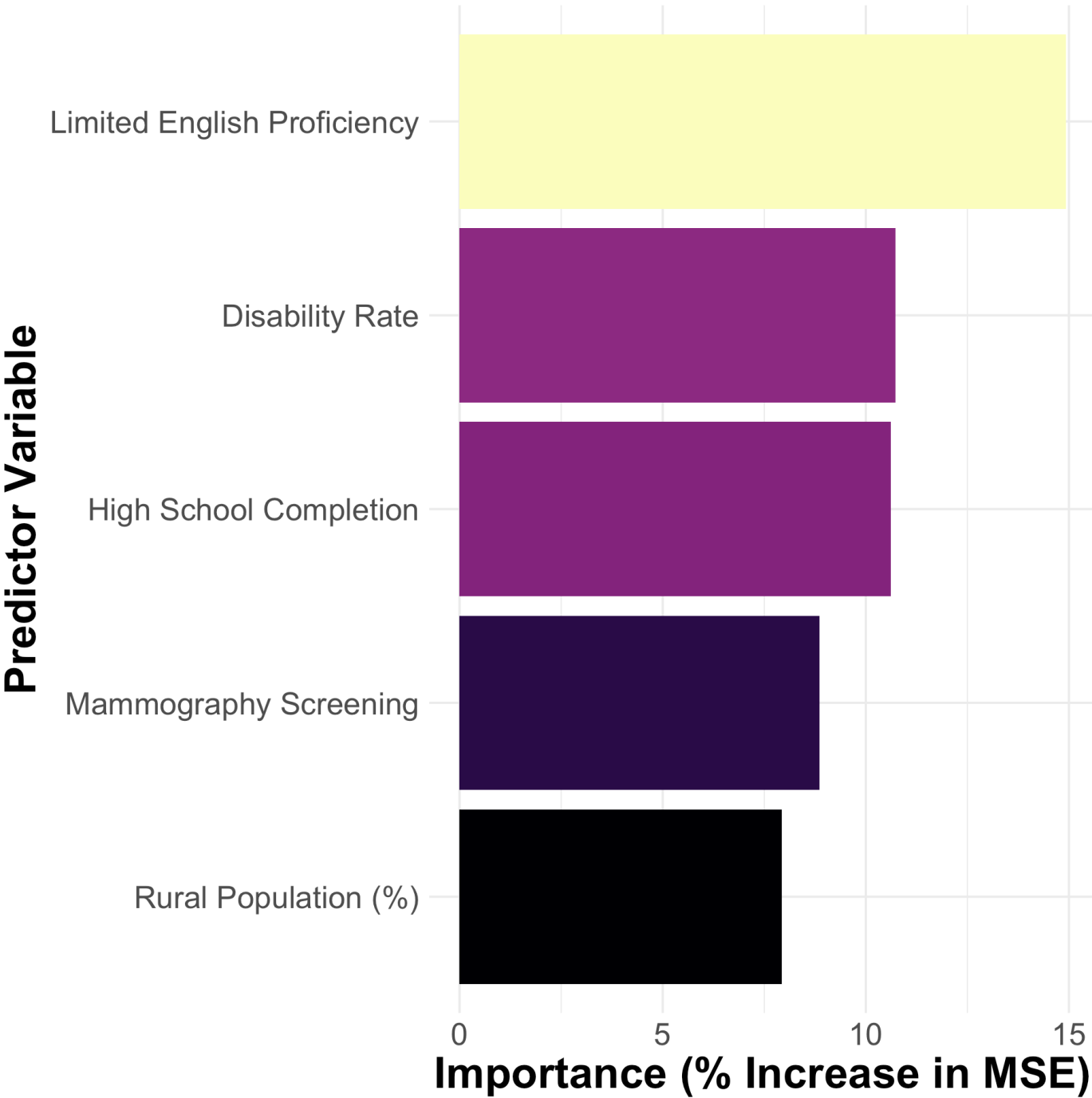
Our analysis revealed clear racial disparities in preventable hospitalizations (PH) and demonstrated that socioeconomic factors consistently rank among the most influential predictors—even more so than traditional clinical access measures. These results were consistent across both exploratory (Random Forest) and inferential (Negative Binomial) approaches.

Variable Importance from Random Forests

To identify the most influential predictors without imposing linear assumptions, we first trained race-stratified Random Forest models. Across both White and POC counties, the top-ranked variables were consistently social determinants of health, especially:

Predictor Variable





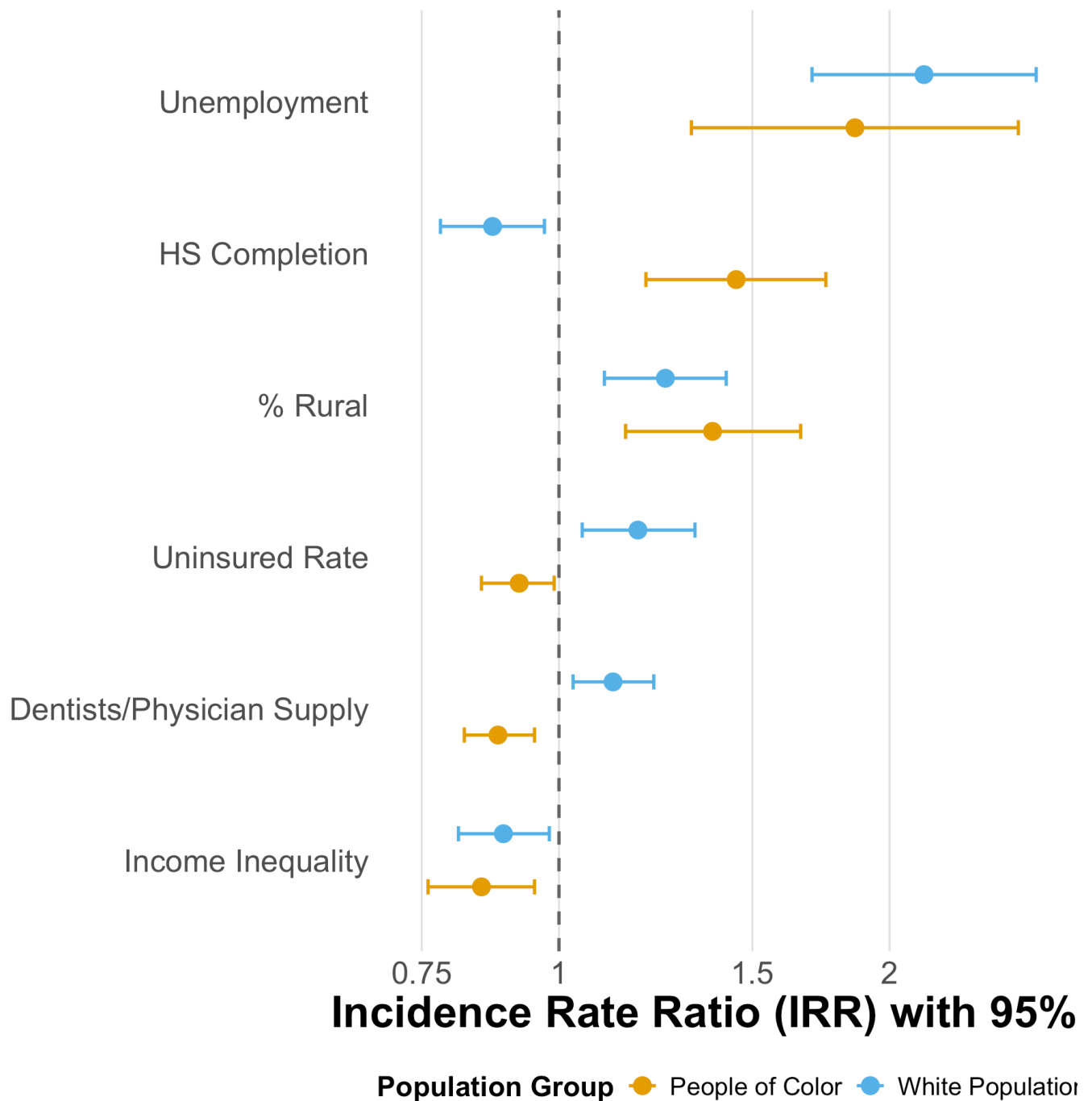
- High school completion
- Uninsured Rate
- Rurality

These findings emphasize that social and economic context—not just healthcare access—shapes hospitalization outcomes. In particular, high school completion ranked as the top variable across both racial groups, suggesting that structural barriers to education are tightly linked to downstream health outcomes.

This pattern reinforces what we observed in our EDA: POC counties face higher median PH rates and more variability than White counties (Figure 2), and these disparities are geographically concentrated in the South and Midwest (Figure 3).

Incidence Rate Ratios from Negative Binomial Regression:

To estimate the magnitude and direction of predictor effects, we fit race-stratified Negative Binomial (NB) regression models. This approach adjusts for population size (via offsets) and accounts for overdispersion in count data. Coefficients were exponentiated to yield Incidence Rate Ratios (IRRs):



Interpretation:

- An IRR > 1 suggests that an increase in the predictor is associated with higher PH rates.
- An IRR < 1 suggests a protective effect—higher values of the predictor are associated with fewer PHs.

Key Takeaways from the NB Models

For White counties, clinical access factors like primary care supply and mammography screening played a more consistent role, suggesting that improving access may reduce preventable stays in these areas.

For POC counties, the dominant drivers were economic and social—notably unemployment, education, and income inequality. A 1 standard deviation increase in unemployment was associated with a 20% increase in preventable hospitalizations (IRR = 1.20).

Recommendations

Our findings reveal that preventable hospitalizations are disproportionately concentrated in counties with high unemployment, income inequality, and lower rates of high school completion—particularly among People of Color (POC). Clinical access factors (e.g., provider supply, insurance coverage) remain important, but their effects vary by race and geography. These insights call for a multifaceted approach that addresses both medical care and the underlying structural determinants of health.

To reduce preventable hospitalizations and close racial disparities, we recommend the following evidence-informed strategies:

1. Expand Access to Primary and Preventive Care

Justification: Our models show that clinical provider supply, while not the top predictor overall, plays a more pronounced role in White counties and remains essential for baseline access.

- Support Federally Qualified Health Centers (FQHCs): Expand funding to deliver primary and preventive care to uninsured and underinsured populations in medically underserved areas.
- Leverage Nurse Practitioners and Allied Health Professionals: Loosen scope-of-practice restrictions to address clinician shortages, especially in rural and high-need regions.
- Invest in Community Health Workers (CHWs): Deploy CHWs to deliver culturally competent outreach, navigation, and chronic care support in marginalized communities.

2. Address Socioeconomic Barriers and Health Literacy

Justification: Across both Random Forests and regression models, unemployment, income inequality, and education emerged as top drivers of preventable hospitalizations—especially for POC.

- Boost Health Insurance Enrollment: Fund navigators and outreach efforts to connect individuals with affordable coverage, particularly in states with large uninsured populations.
- Promote Health Literacy: Improve communication strategies and simplify patient materials to enable better self-management and engagement with care.
- Support Underserved-Area Providers: Expand financial incentives (loan forgiveness, salary support) to recruit and retain clinicians in high-need, low-resource areas.

3. Enhance Chronic Disease Management and Behavioral Health Integration

Justification: Chronic conditions like diabetes and asthma are leading causes of preventable hospitalizations. Mental health access also showed moderate correlation with hospitalization rates.

- Implement Evidence-Based Chronic Disease Programs: Fund interdisciplinary teams and community-based interventions for managing high-risk conditions.
- Integrate Behavioral Health into Primary Care: Address comorbid mental health and substance use issues by embedding behavioral care into routine visits.
- Expand Telemental Health and Mobile Clinics: Use technology and outreach vans to reach patients in areas with geographic, mobility, or stigma-related barriers.

4. Promote Equity Through Culturally Tailored Interventions

Justification: The wide variability in POC hospitalization rates suggests that one-size-fits-all interventions are inadequate. Targeted solutions are essential.

- Train Providers in Cultural Competence: Equip clinicians with tools to engage diverse patients respectfully and effectively.
- Adapt Care to Patient Contexts: Provide interpreters, culturally tailored health materials, and patient navigators who can guide individuals through complex care systems.

5. Improve Care Coordination for High-Risk Populations

Justification: Preventable hospitalizations often stem from fragmented or poorly managed care, especially for older adults and patients with multiple conditions.

- Develop Case-Managed Models for Frail Elders: Coordinate home care, outpatient services, and social support to avoid unnecessary hospital use.
- Expand Medical Homes: Invest in integrated primary care models that emphasize prevention, continuity, and patient-centered care.

Discussion

Our findings confirm the presence of substantial racial disparities in preventable hospitalizations across U.S. counties. These disparities are not solely attributable to clinical access gaps but are closely linked to broader socioeconomic and structural inequities—including education, income inequality, and unemployment. While both Random Forest and Negative Binomial models yielded consistent insights, they offered unique advantages that shaped our interpretation.

Model Interpretation and Strengths:

Random Forests excelled at uncovering nonlinearities and interactions, allowing us to explore complex relationships without assuming a specific functional form. This was especially useful in identifying nuanced predictors in POC-majority counties, where the effects of variables like rurality or income inequality are often compounded. Negative Binomial regression, on the other hand, provided interpretable effect sizes, enabling clear comparisons across predictors and racial groups—essential for policy translation.

Limitations:

Despite the robustness of our modeling framework, several limitations remain:

- **Ecological Data:** Our analysis is based on aggregated, county-level data. While this captures broad geographic trends, it cannot make individual-level inferences or account for within-county disparities.
- **Measurement Constraints:** Racial proportions are derived from census estimates, and preventable hospitalizations are reported by major race categories, potentially obscuring variation within racial subgroups (e.g., Asian, Native American).
- **Data Gaps:** Counties with missing data on key predictors were excluded. This may bias findings toward more well-resourced areas and underrepresent marginalized rural or tribal regions.

Practical Implications and Policy Relevance:

Our study has several practical and policy-relevant takeaways:

- **Equity-Focused Interventions:** Results support tailoring interventions by race and geography. For example, expanding primary care access may reduce hospitalizations in White-majority counties, but in POC-majority counties, addressing education and employment barriers may yield a greater impact.
- **Improved Risk Adjustment:** Policymakers and insurers (e.g., under the ACA) should consider including social determinants of health in risk adjustment formulas to better account for structural disadvantage.
- **Use of Ensemble Methods in Public Health Surveillance:** Random Forests and similar ensemble models could enhance early detection of counties at high risk for preventable hospitalizations, especially when paired with real-time data systems.

Future Directions:

- Test for interaction effects between race and predictors to identify whether certain access barriers are more impactful in specific racial groups.
- Incorporate spatial autocorrelation and clustering methods to detect regional “hotspots” that may benefit from coordinated policy action.
- Integrate ED visits and outpatient care data to build a more complete picture of access and care-seeking behavior.
- Extend to longitudinal designs to examine causal relationships over time and assess the impact of policy changes (e.g., Medicaid expansion).

Conclusion

Preventable hospitalizations are not just a reflection of individual health behaviors or isolated gaps in clinical access—they are a structural outcome shaped by decades of policy, geography, and inequality. Our analysis shows that socioeconomic and educational disadvantage are more strongly associated with preventable hospitalization rates than provider supply, particularly among POC populations.

These insights underscore the urgent need for place-based, race-conscious interventions that address both medical and social determinants of health. Public health agencies, policymakers, and health systems must move beyond one-size-fits-all solutions and embrace tailored strategies grounded in local context and demographic realities.

Future work should deepen this analysis by:

- Building longitudinal models to assess trends and policy impacts
- Exploring hospital overuse in high-access but high-utilization counties
- Applying causal inference frameworks to validate associations
- Advocating for richer, more disaggregated race/ethnicity data collection to illuminate hidden disparities

In conclusion, reducing preventable hospitalizations requires not just better healthcare, but a more equitable society.

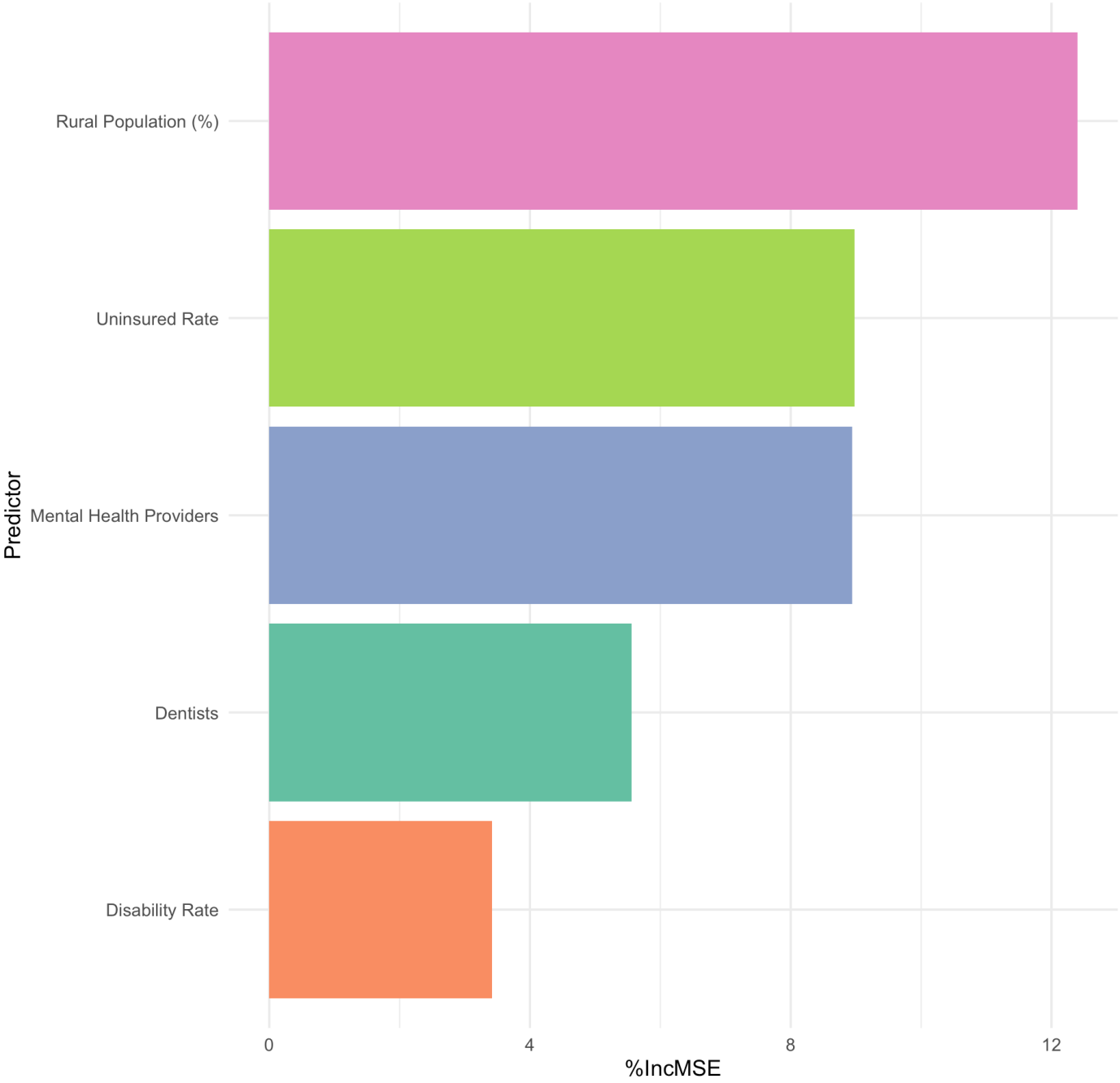
References

University of Wisconsin Population Health Institute. County Health Rankings & Roadmaps 2025

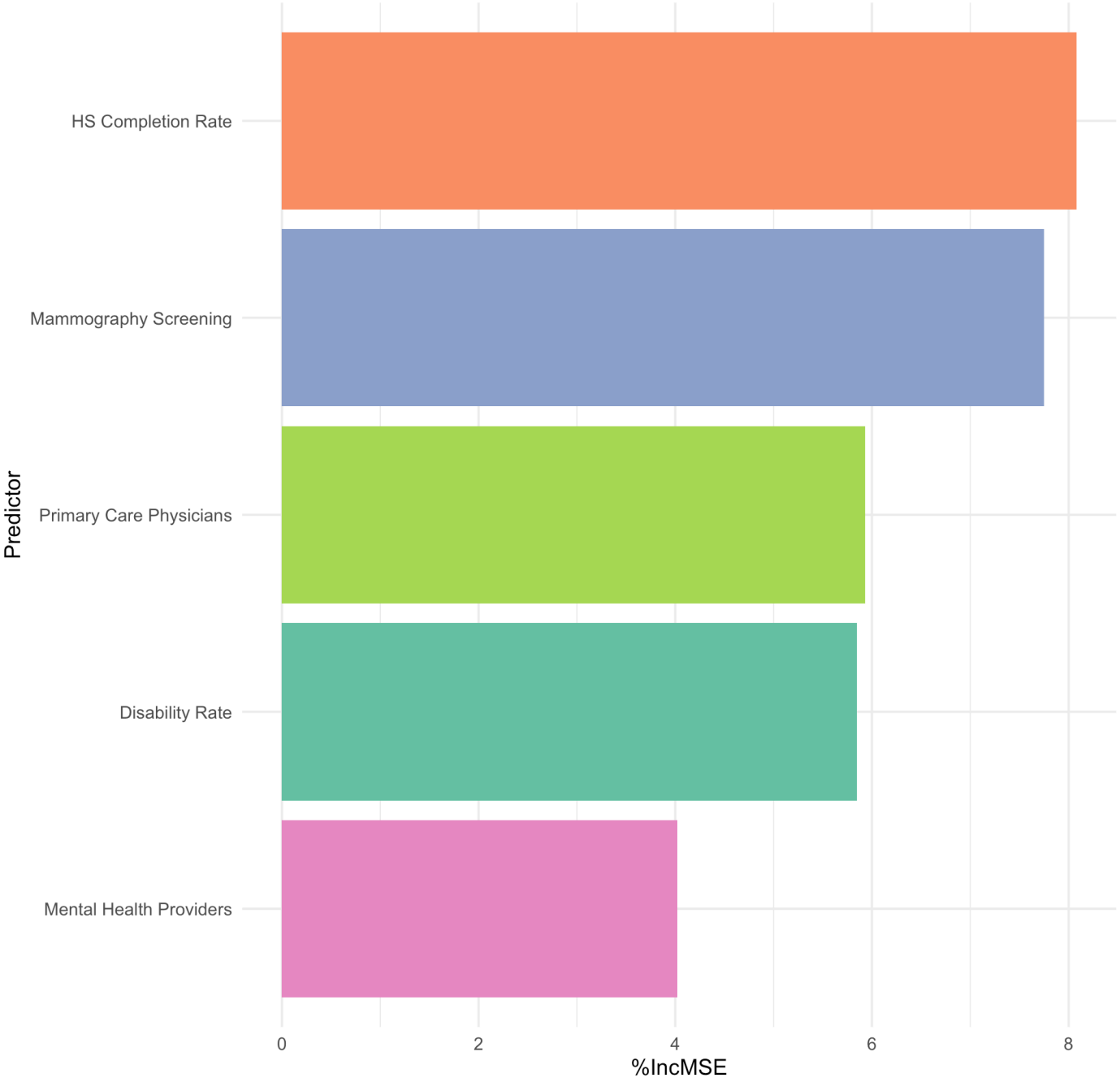
Appendix:

Below are additional visualizations that further support our core findings: socioeconomic determinants—such as high school completion, unemployment, and income inequality—consistently emerge as the most influential predictors of preventable hospitalizations. Model diagnostics are included as well.

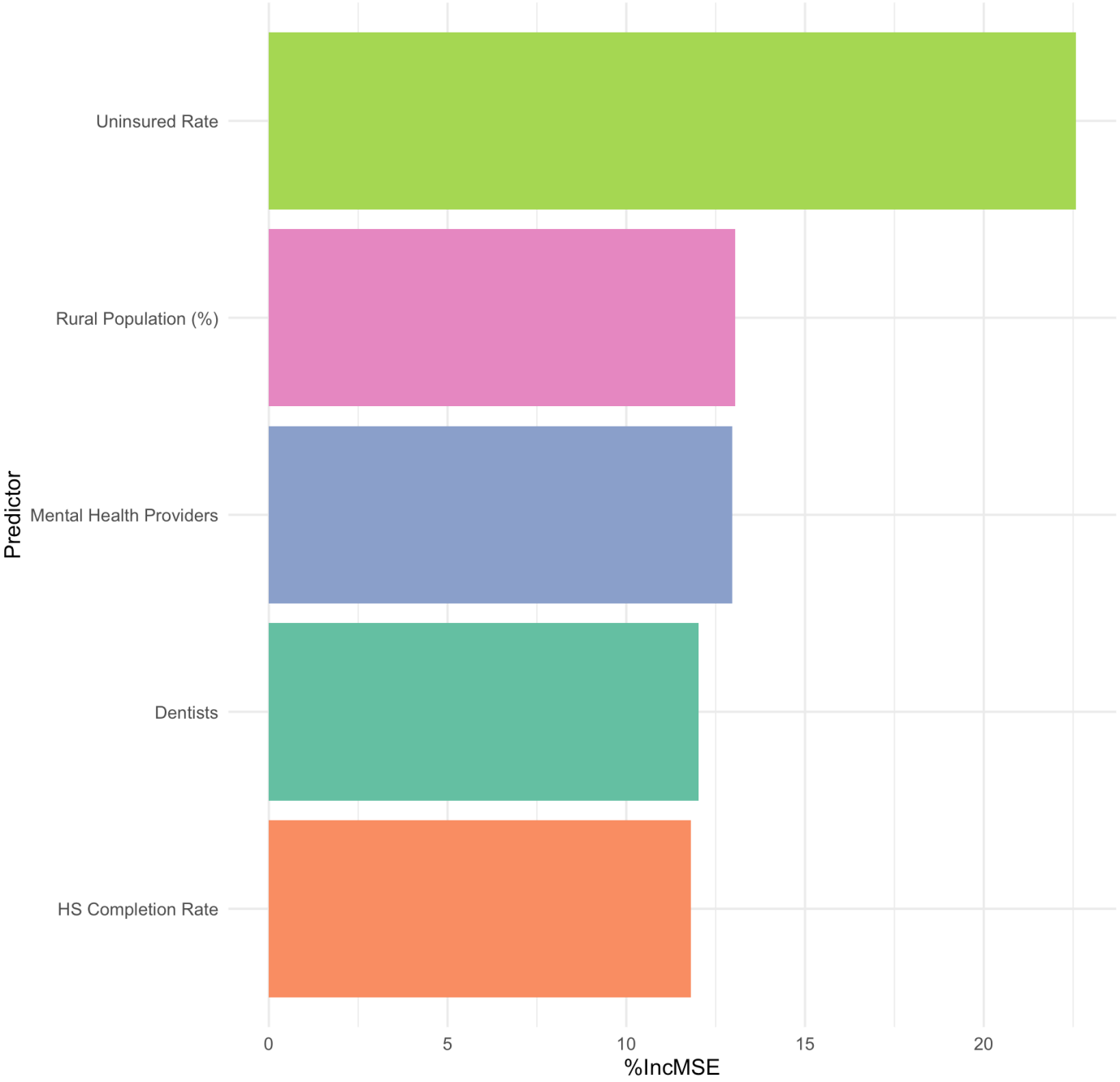
Most Important Variables (Hispanic-White Majority Counties)



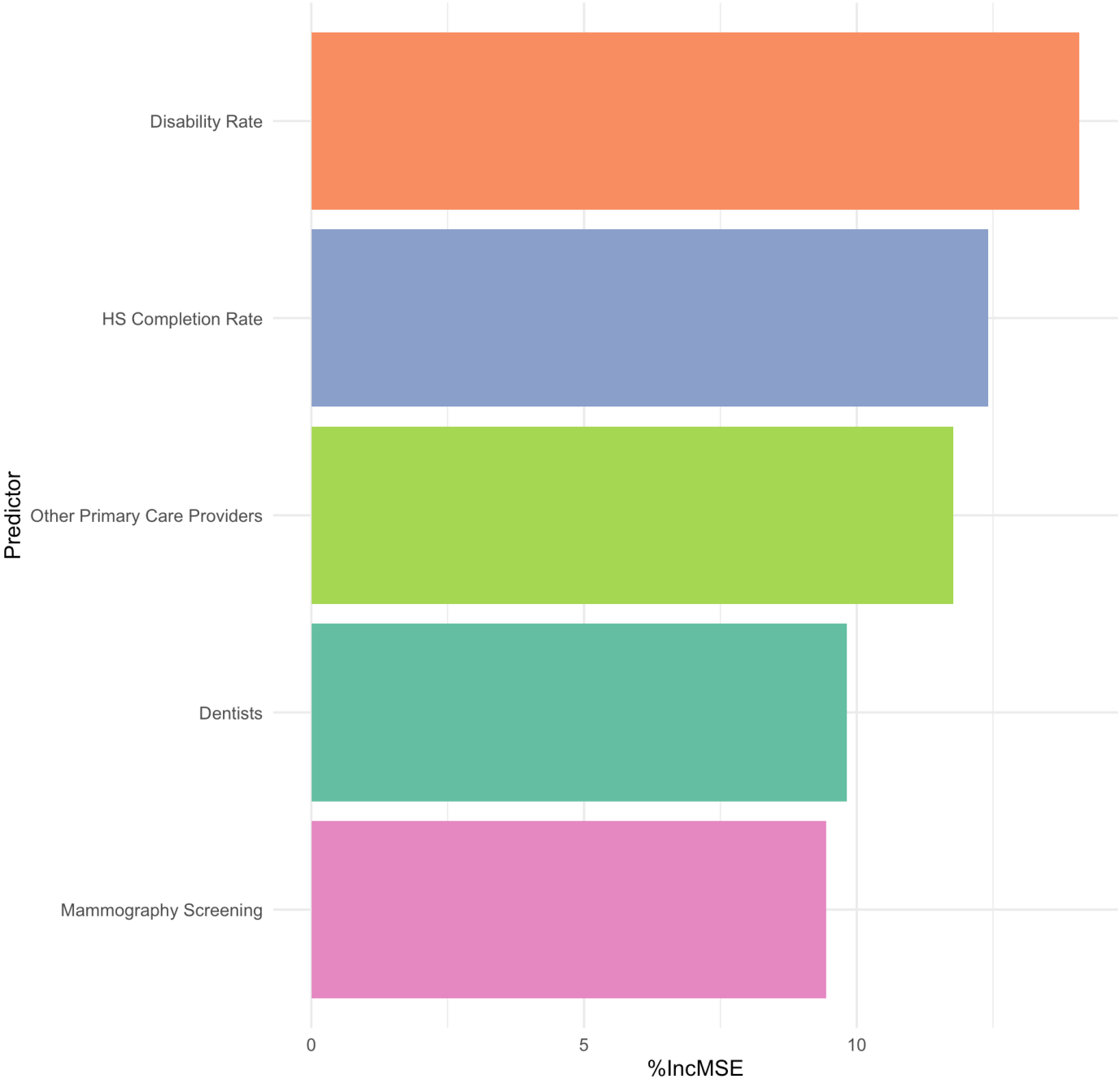
Most Important Variables (Black-White Majority Counties)



Most Important Variables (White-Hispanic Majority Counties)



Most Important Variables (White-Black Majority Counties)



Disability and High School Completion Rates among Top Predictors in White

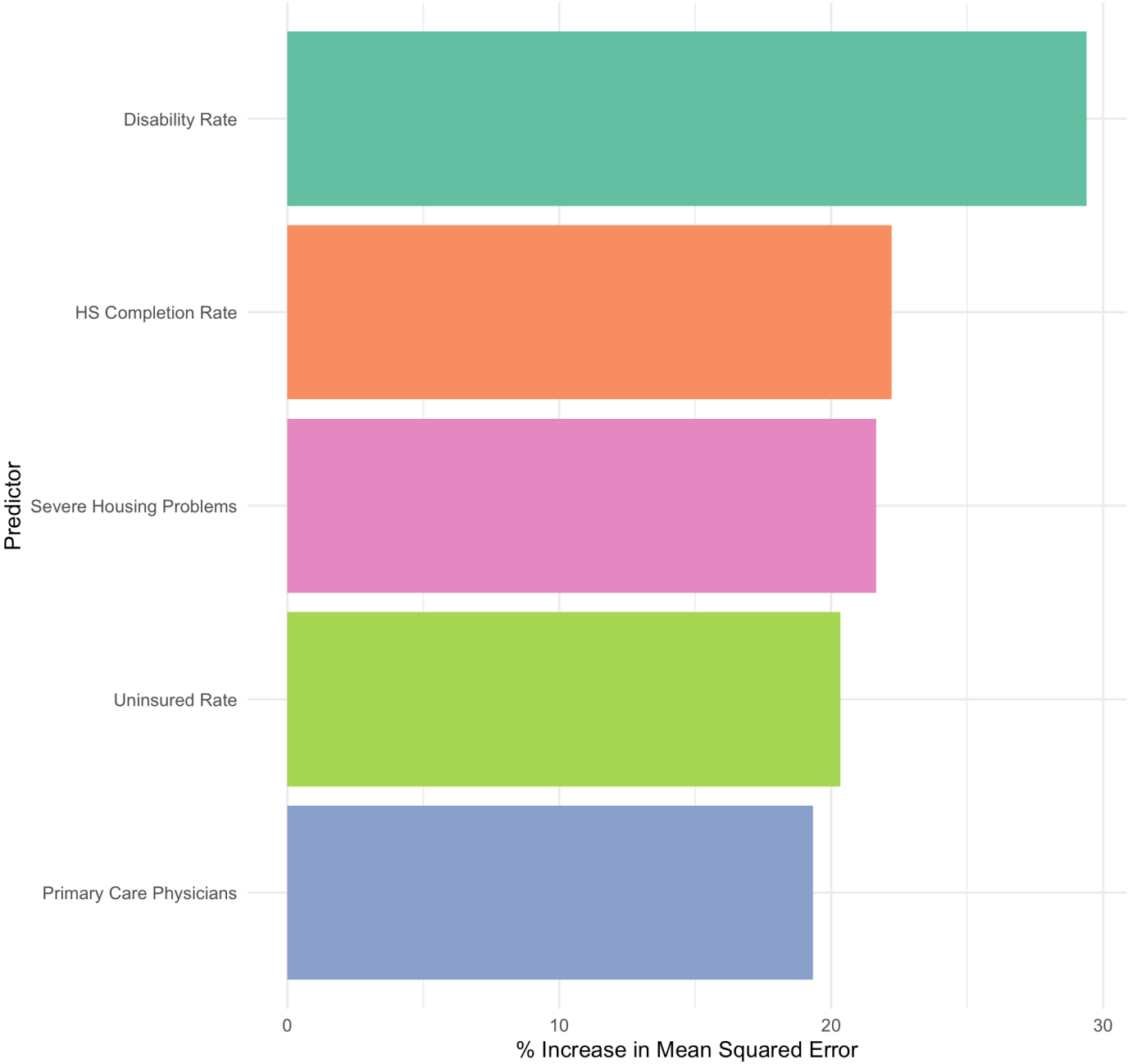
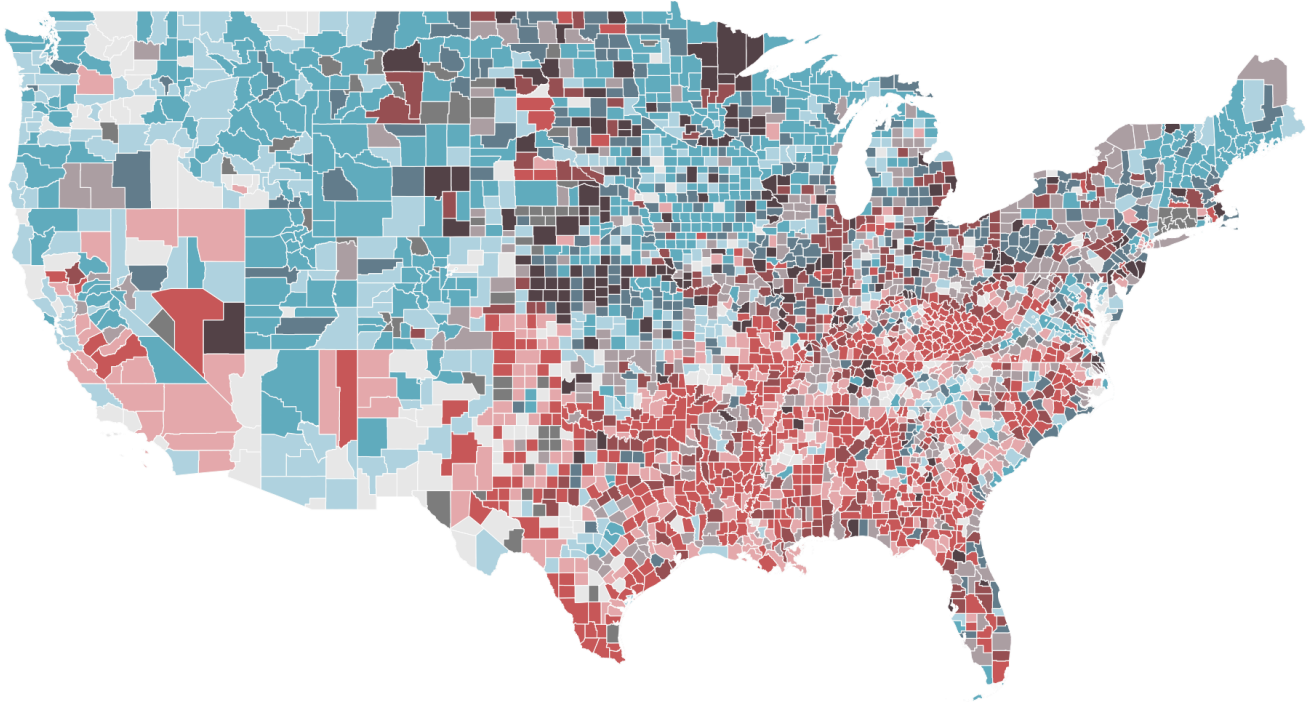
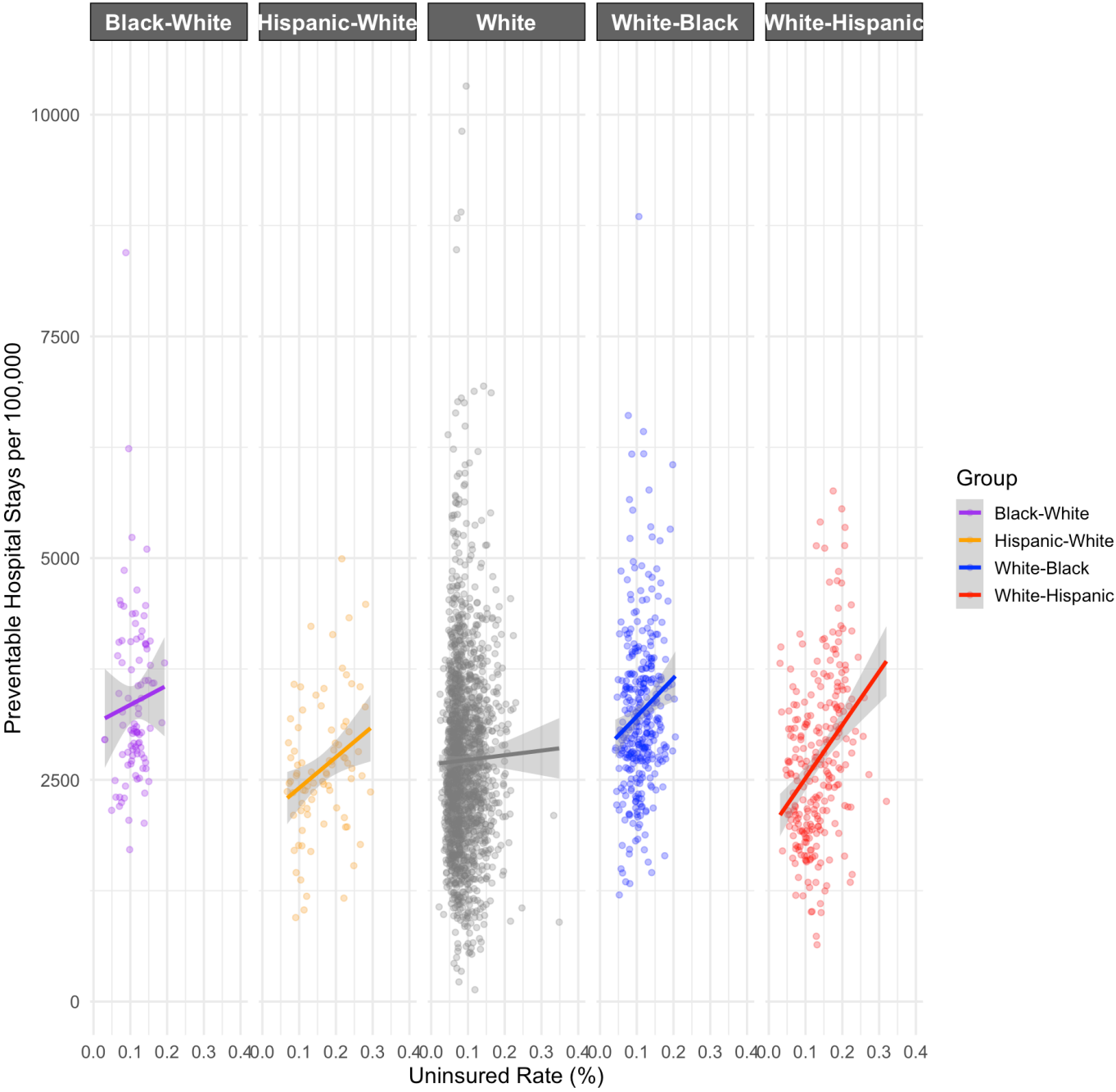


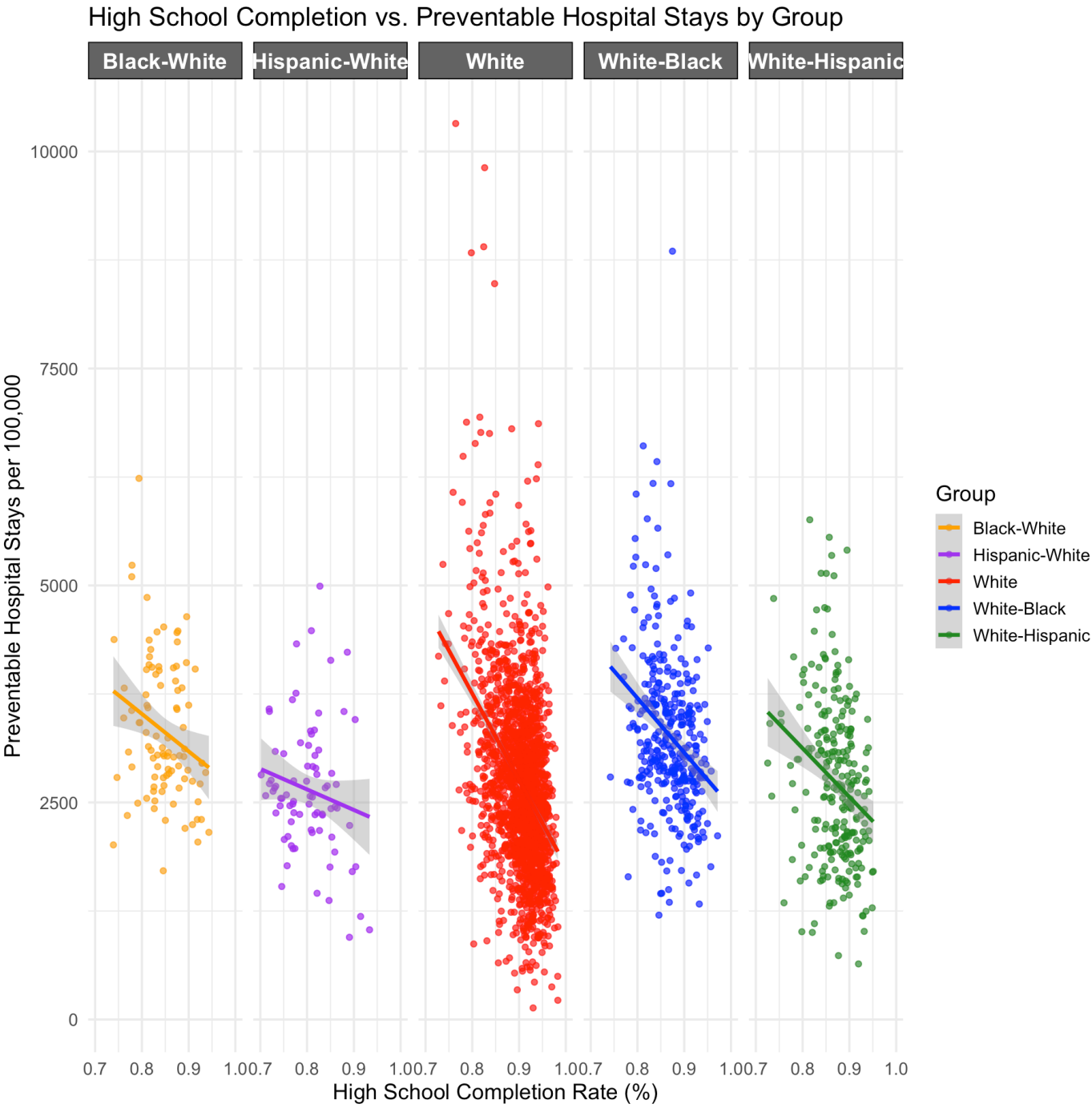
Figure 1: Random Forest variable importance among White Majority Counties

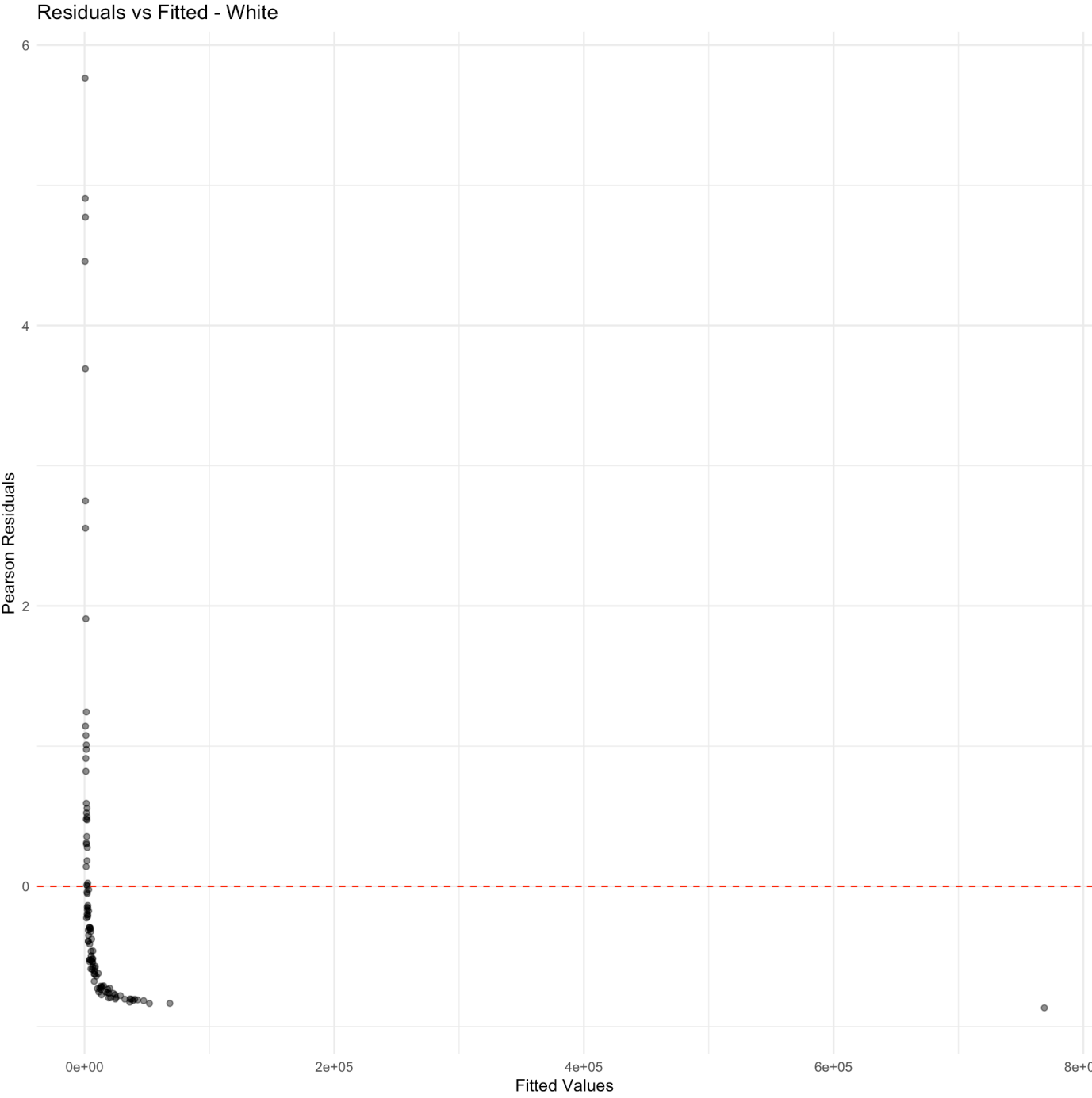
ed Preventable Hospital Stays among counties with lower High School Completion Rates (US C



Uninsured vs. Preventable Hospital Stays by Group







Residuals vs Fitted - POC

