Investigating the incidence and potential risk factors for hypertension-related mortalities in adults in the United States.

Shaun van den Hurk

**Authors**

* First Author: Shaun van den Hurk

**Author affiliations**

1. College of Veterinary Medicine, University of Georgia, Athens, GA, USA.

Corresponding author: shaun.vandenhurk@uga.edu

Disclaimer: The opinions expressed in this article are the author’s own and don’t reflect their employer.

# 1. Summary/Abstract

Cardiovascular disease (CVD) remains a leading cause of mortality in the United States, with hypertension recognized as a major modifiable contributor. This study investigates patterns and predictors of hypertension-related CVD mortality among U.S. adults aged 35 and older using county-level data from 2000 to 2019. We examined temporal, demographic, and geographic disparities to assess both descriptive trends and predictive capacity.

Exploratory analyses revealed persistent disparities in mortality across race, sex, and age groups. Black and American Indian/Alaska Native adults consistently exhibited the highest mortality rates, especially among those aged 65 and older. Temporal models showed diverging trends across subgroups, with steep declines in Black adults and rising mortality among males and American Indian/Alaska Native populations. Spatial clustering was modest but statistically significant, with localized high-mortality areas concentrated in the Southeast.

Predictive modeling using Random Forest and LASSO regression demonstrated that demographic and temporal variables explained a substantial proportion of county-level mortality variation. The Random Forest model achieved an R² of 0.748, with age group emerging as the most influential predictor, followed by race/ethnicity. These findings underscore the enduring impact of demographic and geographic factors on cardiovascular health outcomes and support the use of predictive models to inform targeted public health interventions.

# 2. Introduction

## 2.1 General Background Information

Lifestyle diseases such as cardiovascular disease (CVD) represent some of the most significant causes of death globally, both in low-income regions and high-income countries such as the United States of America (USA). Cardiovascular diseases are a broad umbrella term including different clinical manifestations such as hypertension, stroke, and ischemic heart disease. Cardiovascular diseases were cited as the leading cause of death globally in 2022, in the USA they were responsible for one in every five deaths nationally (1–3). Hypertension is commonly identified as a significant risk factor for multiple cardiovascular conditions, and one that is modifiable, making it an ideal target for public health interventions (4,5). However, not all demographics appear to be equally affected by hypertension and CVD, with differences in rates of CVD and hypertension in people from different races, sexes, age groups, geographic location, income levels, and other groups in previous studies (5,6).

A deeper understanding of these factors helps us to address the multifactorial nature of CVD and also helps guide effective public health interventions. Additionally, it is important to understand if there have been any differences in these rates of CVD and hypertension over time, and if there are particular geographic hotspots. This is essential as it allows public health officials to measure the impact of previous intervention strategies and to ensure that resources strategically target areas of concern.

## 2.2 Description of data and data source

The dataset was collected from the CDC website and the information was sourced from the National Vital Statistics System (7). The dataset contains rates and trends in the hypertension-related cardiovascular disease (CVD) death rates. The study period is from 2000 to 2019, with two main intervals displayed in the data 2000-2009, and 2010-2019. The data is further distributed by county, age group (35-64 years and 65 years and older), race/ethnicity, and sex.

## 2.3 Questions/Hypotheses to be addressed

This data analysis aimed to evaluate the rates of hypertension-related cardiovascular disease mortality in adults in the United States based on county-level data and to identify trends that might be associated with demographic and geographic factors. Furthermore, we aim to identify factors that can be used in predictive models to estimate mortality trends.

#### 2.3.0.1 Study Aims:

- Aim 1: Assess national temporal trends in hypertension-related CVD mortality, stratified by age group.

- Aim 2: Identify spatial clustering of county-level mortality.

- Aim 3: Evaluate disparities in mortality trends by race/ethnicity and sex, stratified by age group.

- Aim 4: Build predictive models using demographic and geographic features to estimate mortality trends.

#### 2.3.0.2 Hypotheses:

- Mortality trends vary significantly over time and across age groups.

- County-level mortality exhibits spatial clustering, reflecting geographic disparities.

- Temporal trends differ across race and sex, with widening inequities over time.

- Predictive models can identify key demographic and spatial features associated with elevated mortality risk.

# 3. Methods

## 3.1 Data import and cleaning

We obtained the dataset from the CDC’s online portal, sourced from the National Vital Statistics System (7). The raw CSV file was approximately 300MB in size and contained 1,103,872 rows and 24 columns. Initial preprocessing was conducted in R to remove redundant variables, reduce file size, and prepare the dataset for analysis. This was all outlined in the “processingcode” document (and README file), which will not run automatically because the Excel (csv) file will not be available due to size concerns with uploading documents (available on request). Full preprocessing steps are documented in the repository’s processingfile-v1.qmd and processingcode.R scripts.

Non-informative or repetitive columns (e.g., columns with a single repeated value or metadata such as DataSource, Topic, Geographic\_Level) were removed. Columns describing demographic stratification were renamed for clarity: Stratification1 was renamed to age\_group, Stratification2 to race\_ethnicity, and Stratification3 to sex. The primary outcome variable, Data\_Value, was renamed mortalities.

All rows with missing values in any of the key variables (mortality rate, year, age group, race/ethnicity, sex, and geographic coordinates) were removed. The resulting cleaned dataset contained 572,572 rows and 14 variables. Summary statistics and exploratory inspection (e.g. using the skimr package) confirmed the data’s readiness for analysis. The dataset included both numeric variables (e.g. mortality rates, year, coordinates) and categorical variables (e.g. sex, race/ethnicity, age group, county).

Age groups were categorized into two broad strata: 35–64 years and 65 years and older. While this limits deeper evaluation, it facilitates comparisons across the two broad groups and supports stratified modeling approaches. The cleaned dataset was saved in .rds format for improved performance and integration with downstream analysis scripts.

## 3.2 Statistical Analysis

Statistical analysis was structured according to the study’s four primary aims and was conducted using R (version 4.3.1). All code is available in the project repository within the statistical-analysis-v2-slim.qmd script. Analyses incorporated regression modeling, spatial statistics, and supervised machine learning to assess temporal trends, demographic disparities, spatial clustering, and predictive performance.

To evaluate national temporal trends in hypertension-related cardiovascular disease (CVD) mortality from 2000 to 2019 (Aim 1), we began by fitting a linear regression model to estimate the overall change in mortality over time. We then used generalized linear models (GLMs) to assess stratified time trends by race/ethnicity, sex, and age group. Interaction terms such as Year × race/ethnicity and Year × age group were included to evaluate whether trends varied between subgroups. We estimated annual mortality slopes for each demographic category using the emtrends() function from the emmeans package and conducted pairwise comparisons using the contrast() function to formally test for differences between groups.

For spatial analysis (Aim 2), we examined county-level clustering of mortality rates using spatial statistics. County geometries were created by converting longitude and latitude coordinates into spatial features using the sf package. Spatial weights were computed using 5-nearest-neighbor relationships. Global spatial autocorrelation was assessed using Moran’s I statistic. Local clustering patterns were evaluated using Local Indicators of Spatial Association (LISA) via the spdep package. We visualized local spatial relationships using standardized mortality z-scores, classifying counties as high–high, low–low, or spatial outliers depending on both their mortality values and those of neighboring counties.

To evaluate demographic disparities in temporal trends (Aim 3), we extended the GLMs used in Aim 1 by including three-way interaction terms between year, race/ethnicity, and sex, and stratified analyses by age group. Estimated marginal trends were obtained separately for adults aged 35–64 and those aged 65 and older. Annual slope estimates and their confidence intervals were extracted and compared across race–sex combinations. Pairwise slope comparisons were again conducted using emmeans, allowing us to quantify and visualize disparities across key demographic intersections.

In the final aim (Aim 4), we developed predictive models to estimate county-level mortality rates using demographic and temporal features. Two modeling approaches were implemented: a random forest model using the ranger engine and a LASSO regression model using the glmnet engine. The modeling dataset included year, age group, race/ethnicity, and sex as predictors. Categorical variables were dummy-encoded using the recipes package, and numeric predictors were normalized; zero-variance predictors were removed. Data were split into training (80%) and test (20%) sets, stratified by mortality outcome. Within the training set, 10-fold cross-validation was used to tune model hyperparameters. For random forests, tuning was performed over a regular grid of mtry and min\_n values. LASSO regression was tuned using a log-scaled penalty grid to identify optimal shrinkage parameters. Final model performance was evaluated on the test set using RMSE (root mean squared error), MAE (mean absolute error), and R² (coefficient of determination). Confidence intervals were computed for cross-validation metrics but not for the final test set performance due to computational constraints. Model interpretability was further assessed using variable importance plots for the random forest model and non-zero coefficient selection in LASSO.

This multi-step analytical framework enabled us to assess trends, identify disparities, detect geographic clustering, and test predictive capacity using a consistent set of demographic and temporal variables across all analytic aims.

# 4. Results

## 4.1 Descriptive and Exploratory Analysis

We began by examining overall patterns in hypertension-related cardiovascular disease (CVD) mortality among U.S. adults aged 35 and older from 2000 to 2019. Summary statistics revealed large disparities in mortality rates across racial/ethnic and age groups. For adults aged 35–64, Black individuals had the highest mean county-level mortality rate (92.2 per 100,000), followed by American Indian/Alaska Native (46.2), and White (32.9). In the 65+ age group, Black adults again had the highest mean mortality (662.6), with American Indian/Alaska Native (451.9) and White (398.3) also experiencing elevated rates. To further characterize these disparities, we summarized the mean and median mortality rates by race/ethnicity across all counties and years. Black adults (all ages) had the highest mean mortality (305 per 100,000), while White adults (all ages) had the highest median mortality (185 per 100,000), reflecting differing distributions and the influence of extreme values (see Supplementary Table S1). These differences are visualized in boxplots by race/ethnicity and sex ([Figure 1](#fig-mortality-rate-by-race) and [Figure 2](#fig-mortality-rate-by-sex)).

Boxplots confirmed these disparities, showing right-skewed distributions with particularly high upper ranges for Black and American Indian/Alaska Native adults. Males consistently had higher mortality distributions than females across all groups. While some groups showed declines in mortality over time, others experienced stable or rising trends, supporting the need for formal modeling to assess temporal and demographic patterns

Temporal visualizations indicated divergent patterns between subgroups in [Figure 3](#fig-glm-interaction). For example, while some populations experienced declines in mortality rates, others showed upward or stable trends. These descriptive findings supported formal modeling to assess temporal and demographic patterns.

|  |
| --- |
| Figure 1: County-Level Mortality Rates by Race/Ethnicity (2000–2019). Boxplot shows distribution across counties and years. |

|  |
| --- |
| Figure 2: County-Level Mortality Rates by Sex (2000–2019). Boxplot shows distribution across counties and years. |

## 4.2 Basic statistical analysis

### 4.2.1 Temporal Trends in Mortality

At the national level, a linear regression of overall mortality against year showed a small but statistically significant increase in mortality from 2000 to 2019 (slope = 1.40, p < 0.001). However, stratified models revealed considerable variation between groups which may otherwise be masked by this overall increase.

To better capture these differences, we fit a generalized linear model (GLM) including an interaction between year and race/ethnicity. The model revealed substantial divergence in mortality trends across groups. Black adults had the highest predicted mortality in 2000 but experienced a steep decline over time (slope = –7.06, p < 0.001), while mortality rates increased among American Indian/Alaska Native adults. White adults showed a modest rise in mortality over the twenty year period. These patterns and contrasts were not evident in a simpler main-effects-only model, which is presented in Supplementary Figure S1 for comparison.

[Figure 3](#fig-glm-interaction) shows observed county-level mean mortality rates overlaid with GLM-predicted trends by group.

|  |
| --- |
| Figure 3: Observed and model-predicted mortality rates by race/ethnicity (2000–2019), based on a generalized linear model with interaction between year and race/ethnicity. Dots represent annual observed means; lines show GLM predictions adjusting for sex. |

Building on the race/ethnicity model, we next assessed differences in mortality trends by age group using an interaction between year and age. This analysis revealed that adults aged 65 and older experienced a significantly steeper increase in mortality over time than those aged 35–64 (interaction slope = +0.22, p < 0.001). Estimated marginal trends from the model further supported this finding: the slope for older adults was 1.26 deaths per 100,000 per year (95% CI: 1.18–1.34), compared to 1.04 (95% CI: 0.96–1.12) among younger adults, a difference that was statistically significant (p = 0.0002). These results are visualized in Figure [Figure 4](#fig-slope-agegroup), which presents the estimated annual mortality increase for each age group with 95% confidence intervals.

|  |
| --- |
| Figure 4: Estimated annual increase in hypertension-related mortality by age group. Slope values represent the yearly increase in predicted mortality rates per 100,000, based on a GLM with interaction between year and age group. Error bars reflect 95% confidence intervals. |

A corresponding figure from the GLM without interaction terms is available in the Supplement (Supplementary Figure S1). It assumes parallel trends across groups and does not capture the observed divergence over time.

### 4.2.2 Geographic Distribution and Clustering

[Figure 5](#fig-spatial-all), [Figure 6](#fig-spatial-black), and [Figure 7](#fig-spatial-asian) show spatial variation in hypertension-related mortality across all U.S. counties, both overall and for selected racial/ethnic groups. Elevated mortality was most prominent among Black individuals, particularly in counties across the Southeast and Midwest. Asian and Pacific Islander populations displayed lower and more geographically diffuse mortality patterns. Additional subgroup maps are included in the Supplement (Figures S2 - S6), and the combined spatial overview plot (Supplementary Figure S3) provides a side-by-side comparison of geographic patterns by group.

|  |
| --- |
| Figure 5: Spatial distribution of hypertension-related mortality – All races. |

|  |
| --- |
| Figure 6: Spatial distribution of hypertension-related mortality – Black population. |

|  |
| --- |
| Figure 7: Spatial distribution of hypertension-related mortality – Asian and Pacific Islander population. |

Global spatial clustering of hypertension-related CVD mortality was modest but statistically significant at the national level, with Moran’s I = 0.0435 (p < 0.001). Similar low levels of spatial autocorrelation were observed in state-level analyses: Georgia (I = 0.0297), Mississippi (0.0289), Texas (0.0168), and Kentucky (0.0227), all p < 0.001. To further explore local patterns, we conducted Local Indicators of Spatial Association (LISA) analyses. Supplementary Figure S7 displays the LISA cluster map for all U.S. counties, revealing several high–high clusters concentrated in the Southeastern United States, particularly in regions with high Black mortality. Supplementary Figures S8 and S9 provide state-level LISA maps for Georgia and Kentucky, which further illustrate the localized nature of spatial clustering. Although some statistically significant clusters were observed, the majority of counties did not exhibit strong local spatial dependence. These findings suggest that while certain mortality hotspots exist, broader structural and demographic factors likely account for the majority of geographic variation, rather than tightly bounded spatial processes.

### 4.2.3 Demographic Disparities in Temporal Trends

We assessed demographic disparities in mortality trends by analyzing race/ethnicity and sex within each age group using GLMs with Year × Race × Sex interaction terms. Among adults aged 35–64, Black individuals experienced the steepest annual decline in mortality (–3.68 per 100,000), while American Indian/Alaska Native individuals showed the greatest increases (+3.38). Similar patterns were observed in the 65+ group, though the absolute changes were larger. Overall, males experienced steeper annual increases in mortality compared to females. Pairwise slope comparisons revealed the greatest disparity between Overall Male and Black Overall groups (+7.00 per year), followed by contrasts involving American Indian/Alaska Native, White, and Asian/Pacific Islander populations. These trends are summarized in [Table 2](#tbl-slope-race-sex) and visualized in [Figure 8](#fig-slope-35-64) and [Figure 9](#fig-slope-65-plus). These results indicate persistent and widening disparities in hypertension-related mortality trends, particularly among historically marginalized populations, underscoring the need for targeted public health strategies.

|  |
| --- |
| Table 1: Estimated annual increase in hypertension-related mortality by age group. Values are based on a generalized linear model with an interaction between year and age group. Slopes represent predicted yearly change in mortality rates per 100,000. |

|  |
| --- |
| Table 2: Estimated annual mortality trends by race/ethnicity and sex. Slopes were derived from a generalized linear model with Year × Race × Sex interaction terms and represent predicted yearly change in mortality per 100,000. Confidence intervals are shown in the final column. |

Additional visualizations from Aim 3.4 reinforce these findings ([Figure 8](#fig-slope-35-64) and [Figure 9](#fig-slope-65-plus)):

|  |
| --- |
| Figure 8: Estimated annual change in hypertension-related mortality for adults aged 35–64, by race and sex. Values represent GLM-estimated slopes with 95% confidence intervals. ‘Overall Male’ and ‘Overall Female’ refer to aggregate trends across all racial/ethnic groups within the respective sex category. |

|  |
| --- |
| Figure 9: Estimated annual change in hypertension-related mortality for adults aged 65 and older, by race and sex. Values represent GLM-estimated slopes with 95% confidence intervals. ‘Overall Male’ and ‘Overall Female’ refer to aggregate trends across all racial/ethnic groups within the respective sex category. |

### 4.2.4 Predictive Modeling of Mortality Rates

To evaluate the predictive value of demographic and temporal factors for hypertension-related CVD mortality at the county level, we developed two supervised machine learning models: a Random Forest and a LASSO regression model. Each model was trained on 80% of the data and evaluated on a 20% held-out test set, with 10-fold cross-validation used to optimize hyperparameters and minimize overfitting.

The Random Forest model achieved the highest predictive accuracy. The optimal model used all four predictors—year, age group, race/ethnicity, and sex—with a minimum node size of 15 and mtry = 4. On the test set, it produced an RMSE of 115.3, MAE of 66.9, and R² of 0.748, explaining approximately 75% of the variance in county-level mortality (see [Table 3](#tbl-model-performance)). Variable importance scores, calculated using permutation-based methods, highlighted age group (particularly 65+) as the most influential predictor, followed by race/ethnicity. In contrast, sex and year had lower relative importance ([Figure 11](#fig-rf-vip)). The predicted versus observed plot ([Figure 10](#fig-rf-pred)) showed strong alignment, although slight underprediction was observed in counties with extreme mortality rates, indicating potential underfitting at the distribution tails.

|  |
| --- |
| Table 3: Test set performance metrics for Random Forest, LASSO regression, and a null model. RMSE = root mean squared error, MAE = mean absolute error, and R² = coefficient of determination. The null model predicts the mean mortality rate and serves as a baseline for comparison. |

|  |
| --- |
| Figure 10: Predicted versus observed mortality rates for the Random Forest model on the test set. Dashed line represents perfect prediction. Test set R² = 0.748. |

|  |
| --- |
| Figure 11: Variable importance from the Random Forest model. Importance was calculated using permutation-based methods. Results are shown for the top 20 predictors. |

The LASSO regression model served as a more interpretable, though slightly less accurate, alternative. On the test set, it achieved an RMSE of 119.5, MAE of 70.7, and R² of 0.726 (see [Table 3](#tbl-model-performance)). The model applied coefficient shrinkage and automatic variable selection using a penalty parameter (λ = 1e–4), which was selected via cross-validation to minimize RMSE. This regularization excluded several dummy-coded predictors, improving interpretability by focusing on the most informative features. However, LASSO’s linear structure and shrinkage approach limited its flexibility in capturing non-linear interactions. The predicted versus observed plot ([Figure 12](#fig-lasso-pred)) displayed clear horizontal banding patterns, a result of many counties sharing identical categorical predictor combinations. These banding effects were more pronounced than in the Random Forest model, which better captured complex interactions through hierarchical tree splits. Supplementary results confirmed strong generalization in LASSO model performance across folds, with a mean RMSE of 119.63 (95% CI: 117.24–122.02) and R² of 0.73 (95% CI: 0.72–0.73) across 10-fold CV resamples (see Supplementary Table S4).

|  |
| --- |
| Figure 12: Predicted versus observed mortality rates for the LASSO regression model on the test set. Dashed line represents perfect prediction. Horizontal banding reflects repeated predictions due to dummy-encoded categorical variables. |

To contextualize model performance, we also evaluated a null model that predicted the mean training set mortality rate for all test observations. As expected, this baseline yielded substantially worse performance: RMSE = 228.65, MAE = 192.28, and an undefined R² due to zero variance in predictions (see Table 3). In contrast, both the Random Forest and LASSO models performed significantly better, confirming their ability to capture meaningful structure in the data. These results underscore that even with a limited set of demographic and temporal features, machine learning models can explain a substantial portion of the variation in county-level mortality.

To assess generalization, we compared model performance across the training and test sets. As shown in Supplementary Table S2, both models yielded nearly identical metrics across splits. For example, the Random Forest achieved RMSE = 115.31 (train) vs. 115.30 (test), and LASSO showed similarly consistent results. This close alignment suggests minimal overfitting. Cross-validation results further supported model stability. The Random Forest model demonstrated mean CV RMSE values ranging from 175.01 to 177.53, with R² values consistently around 0.72–0.73 across hyperparameter settings (see Supplementary Table S3). The LASSO model showed equally strong consistency, with a mean RMSE of 119.63 (95% CI: 117.24–122.02) and R² of 0.73 (95% CI: 0.72–0.73) across the 10 CV folds (Supplementary Table S4). These findings reinforce that both models are well-calibrated and generalize effectively, capturing robust mortality patterns without overfitting.

Despite their differences, both models demonstrated strong performance in capturing meaningful mortality patterns. The Random Forest offered greater predictive accuracy, while the LASSO model provided clearer interpretability through variable selection and shrinkage. Together, these results highlight that demographic and temporal variables, particularly age, race/ethnicity, and year, are robust predictors of county-level hypertension-related mortality. This supports our hypothesis that predictive models can effectively identify demographic patterns associated with elevated mortality risk. As such, these approaches represent promising tools for population-level risk estimation and public health planning.

\_

# 5. Discussion

## 5.1 Summary and Interpretation

We examined patterns and predictors of hypertension-related cardiovascular disease (CVD) mortality among U.S. adults aged 35 and older between 2000 and 2019. Through exploratory visualizations, regression modeling, spatial clustering, and machine learning, we identified consistent disparities in mortality across time, geography, and demographic groups. These findings highlight the persistent burden of hypertension-related CVD and point to critical areas for targeted intervention.

Nationally, overall mortality rates remained relatively stable across the study period. However, stratified analyses revealed substantial variation between subgroups. Black individuals consistently experienced the highest mortality rates, while American Indian/Alaska Native populations showed a notable upward trend in recent years. Adults aged 65 and older exhibited steeper annual increases than younger adults. These patterns align with prior surveillance studies (3,5) and underscore the importance of modeling subgroup-specific trends. Aggregated models with shared slopes risk masking meaningful divergences—such as the sharp declines among Black adults versus rising mortality in American Indian/Alaska Native populations.

Spatial analysis revealed modest but statistically significant clustering of mortality at national and state levels. Global Moran’s I was low (I = 0.0435), and LISA analyses identified limited high–high clusters, primarily in the Southeast. These findings suggest that broad demographic and structural factors, rather than localized spatial effects, may better explain mortality distribution—echoing prior work on the role of social determinants in health disparities (5,6).

Demographic subgroup analyses showed diverging mortality trends. Among adults aged 35–64, Black individuals experienced the largest declines, while mortality increased sharply among American Indian/Alaska Native adults and males. In the 65+ group, mortality rose across nearly all categories. These results reinforce the importance of considering intersecting dimensions of inequality, such as race, sex, and age, in surveillance and intervention planning. They also support the need for demographic-specific public health responses to reduce disparities in chronic disease outcomes (4,5).These stratified insights are further strengthened by the predictive modeling results, which confirmed age, race/ethnicity, and year as the most influential predictors of county-level mortality.

In the predictive modeling component, we assessed whether demographic and temporal variables could accurately predict county-level mortality. The Random Forest model performed best, explaining approximately 75% of the variation in mortality rates (R² = 0.748), with an RMSE of 115.3. Race/ethnicity and year were the most influential predictors. The LASSO regression model, while slightly less accurate (R² = 0.726), offered greater interpretability by selecting a reduced subset of predictors and shrinking non-influential coefficients to zero. These results suggest that relatively simple demographic and temporal features can provide strong predictive power, and that different modeling approaches offer complementary strengths. The Random Forest’s flexibility allowed for more accurate predictions, while LASSO highlighted core factors driving variability.

Notably, both models showed horizontal banding in predicted vs. observed plots—an artifact of dummy-encoded categorical variables and repeated predictor combinations across counties. This highlights a broader methodological challenge in population-level modeling using grouped categorical features. Although machine learning approaches can improve predictive performance, careful attention must be paid to encoding strategies and interpretability.

These findings contribute to a growing body of evidence emphasizing the importance of demographic stratification and predictive modeling in public health research. Importantly, our results reinforce that even well-known risk factors—such as age, sex, race/ethnicity, and time—continue to explain a large proportion of the variance in hypertension-related CVD mortality and may be sufficient for initial risk stratification at the population level.

## 5.2 Strengths and Limitations

This study benefits from the use of a comprehensive, nationwide dataset sourced from the CDC’s National Vital Statistics System, covering nearly two decades of hypertension-related cardiovascular disease mortality records. The integration of traditional statistical models and machine learning approaches allowed us to describe mortality patterns and assess their predictability using demographic and temporal inputs. However, several limitations should be noted. First, the analysis relied on county-level aggregated data, which may obscure within-county heterogeneity and limits the ability to draw individual-level conclusions. Second, although we used age-standardized mortality rates, key contextual variables such as socioeconomic status, healthcare access, or comorbidities were not included, potentially introducing residual confounding. Lastly, while our models achieved high predictive accuracy for county-level outcomes, they were not intended to establish causal relationships or support individual-level risk prediction.

## 5.3 Implications and Future Directions

Our findings highlight persistent and inequitable patterns in hypertension-related CVD mortality that should inform public health surveillance, policy development, and resource allocation strategies. While geographic clustering was modest, identifying high-burden areas remains important for targeting interventions. The predictive models developed here demonstrate that even simple demographic and temporal variables can capture a substantial proportion of county-level variation, making them useful for settings where more granular data are unavailable. These models can support proactive health planning and serve as screening tools when integrated with local or regional surveillance systems.

Future research should incorporate additional contextual predictors—such as healthcare access, environmental risk factors, and social determinants of health—to improve model precision and equity relevance. Efforts to enhance interpretability and generalizability through multi-level modeling, causal frameworks, or longitudinal datasets may also help inform upstream interventions to reduce CVD disparities.

## 5.4 Conclusions

This study presents a comprehensive analysis of hypertension-related cardiovascular disease (CVD) mortality among U.S. adults aged 35 and older from 2000 to 2019, leveraging national county-level data. While national mortality trends remained relatively stable, we identified significant disparities across race, age, sex, and geography. Black and American Indian/Alaska Native populations consistently bore higher mortality burdens, with the most pronounced increases observed among older adults and males. Spatial analysis revealed modest but statistically significant clustering, particularly in the Southeastern U.S., with broader demographic and structural factors driving most of the observed variation.

Temporal modeling showed that subgroup-specific mortality trends diverged considerably over time, suggesting that recent public health gains have not been shared equally. Our predictive modeling confirmed the strong influence of demographic variables—especially age group, which emerged as the most important predictor, followed by race/ethnicity and year. These variables alone explained a substantial portion of the variance in county-level mortality rates. While Random Forest models achieved higher predictive accuracy, LASSO models enhanced interpretability, illustrating complementary strengths for public health modeling.

Collectively, these findings reinforce the need for targeted, equity-focused public health strategies. Future work should incorporate structural and social determinants of health, including healthcare access and policy context, to better understand and address the root causes of persistent CVD disparities. Robust, interpretable models like those presented here can support evidence-based decision-making, guide resource prioritization, and ultimately contribute to more equitable cardiovascular outcomes.

# 6. References

1. Mills KT, Stefanescu A, He J. The global epidemiology of hypertension. *Nature Reviews Nephrology* [electronic article]. 2020;16(4):223–237. (<https://www.nature.com/articles/s41581-019-0244-2>). (Accessed April 28, 2025)

2. Virani SS, Alonso A, Benjamin EJ, et al. Heart Disease and Stroke Statistics—2020 Update: A Report From the American Heart Association. *Circulation* [electronic article]. 2020;141(9). (<https://www.ahajournals.org/doi/10.1161/CIR.0000000000000757>). (Accessed April 28, 2025)

3. Vaughan AS, Coronado F, Casper M, et al. County‐Level Trends in Hypertension‐Related Cardiovascular Disease Mortality—United States, 2000 to 2019. *Journal of the American Heart Association: Cardiovascular and Cerebrovascular Disease* [electronic article]. 2022;11(7):e024785. (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9075476/>). (Accessed April 28, 2025)

4. Whelton PK, Carey RM, Aronow WS, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Hypertension* [electronic article]. 2018;71(6). (<https://www.ahajournals.org/doi/10.1161/HYP.0000000000000065>). (Accessed April 28, 2025)

5. Mahmood SS, Levy D, Vasan RS, et al. The Framingham Heart Study and the epidemiology of cardiovascular disease: A historical perspective. *The Lancet* [electronic article]. 2014;383(9921):999–1008. (<https://linkinghub.elsevier.com/retrieve/pii/S0140673613617523>). (Accessed April 28, 2025)

6. Yusuf S, Joseph P, Rangarajan S, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): A prospective cohort study. *The Lancet* [electronic article]. 2020;395(10226):795–808. (<https://linkinghub.elsevier.com/retrieve/pii/S0140673619320082>). (Accessed April 28, 2025)

7. Rates and Trends in Hypertension-related Cardiovascular Disease Mortality Among US Adults (35+) by County, Age Group, Race/Ethnicity, and Sex – 2000-2019 Data Centers for Disease Control and Prevention. (<https://data.cdc.gov/Heart-Disease-Stroke-Prevention/Rates-and-Trends-in-Hypertension-related-Cardiovas/uc9k-vc2j/about_data>). (Accessed April 28, 2025)