Original Research Article

Simulation-Based Modeling of Predictors of Hypertension Risk Among Young Adults in the United States

Jihun Jung

Tenafly High School, 19 Columbus Dr, Tenafly, NJ 07670, United States

ABSTRACT

Hypertension risk is an important concern in the area of public health since increased blood pressure elevates the risk of cardiovascular disease, stroke, and other long-term complications. However, it is often under-recognized in young adults. In this study, a simulation-based quantitative modeling approach was taken to examine predictors of hypertension among young adults in the United States. A total of 2,000 individuals aged from 18 to 39 were generated by simulation grounded by prior literature to reflect plausible demographic, behavioral, anthropometric, and metabolic patterns reported in prior epidemiological research in the United States. Variables included age, sex, race/ethnicity, body mass index (BMI), smoking status, alcohol use, physical activity, fasting glucose, total cholesterol, and hypertension status. The simulated sample was summarized by descriptive statistics, and binary logistic regression was used to estimate the independent correlations between aforementioned predictors and hypertension. Hypertension prevalence in the simulated sample was reported to be around 11.5%. Higher age within the young adult range, higher BMI, smoking, increased glucose, male sex, and Black race were positively correlated with greater odds of hypertension. Physical activity indicated a statistically significant protective association with hypertension odds, with participants meeting guidelines for physical activity having lower modeled odds of hypertension. Among the modeled predictors, BMI was reported to be the strongest correlations with hypertension risk. The model indicated acceptable discriminatory performance, and the area under the receiver operating characteristic curve was around 0.77. Overall, the findings in this study suggest that hypertension risk in young adults is shaped by multiple demographic, behavioral, and metabolic factors, emphasizing the value of simulation-based modeling in public health research.

Keywords: Hypertension; young adults; logistic regression; simulation-based modeling; body mass index; cardiometabolic risk; public health

INTRODUCTION

Hypertension is a major public health problem because there is a tendency for persistently increased blood pressure aggravates the risk of heart disease, stroke, kidney damage, and other long-term complications (1, 2). High blood pressure is defined as blood pressure in a level that is consistently at or above 130/80 mm Hg by the Centers for Disease Control and Prevention. In addition, the World Health Organization highlights that

Corresponding author: Jihun Jung, E-mail: jungjihun721@gmail.com.

Accepted June 10, 2026
https://doi.org/10.70251/HYJR2348.43549556

hypertension becomes a serious asymptomatic problem if untreated (1, 2). Because many affected individuals tend not to feel any symptoms, hypertension is often described as a silent condition with a possibility of remaining undetected until serious vascular damage has already occurred (1, 2).

Although hypertension often occurs in middle-aged and older adults, it is also an important health issue in younger populations. According to National U.S. data, hypertension prevalence has been on a rising trend with age but has also become a serious issue among adults aged between 18 and 39 years (3). Recent work also suggests that younger adults are in a situation with poor awareness and control with hypertension. This means that many affected individuals may live with elevated blood pressure for a long time without diagnosis or treatment (4). Because of this phenomenon, early adulthood has become a particularly important period for prevention of hypertension. When high blood pressure begins earlier in life, there is a possibility for it to contribute to a longer cumulative exposure to vascular stress. Therefore, it imposes a higher lifetime risk of cardiovascular disease (2, 4).

Prior research has examined several measurable predictors of hypertension that were specifically relevant to young adults. One of the most outstanding predictors was body mass index (BMI). Elevated BMI tends to be associated with higher blood pressure through mechanisms, including increased vascular resistance, altered neurohormonal regulations, and broader metabolic dysfunction (5). In particular, BMI was reported to be strongly correlated with hypertension prevalence in young adults, while smoking and lower physical activity also showed adverse relationships with blood pressure risk (5). These findings support the idea that hypertension in young adults is influenced by both age and modifiable lifestyle factors.

Behavioral risk has also been emphasized in the literature. For example, smoking is known to contribute to endothelial injury, inflammation, and sympathetic activation that collectively may worsen blood pressure regulation (5, 6). Physical inactivity is another important factor. Physical activity pattern from young adulthood onward is known to be associated with later hypertension risk (7). This suggests that regular activity may protect against blood pressure elevation across the life course. In addition, contributors to hypertension risk include excessive alcohol intake and poor dietary habits as repeatedly cited by public health agencies (1, 2).

Metabolic health is also an important factor in hypertension. Beyond BMI alone, elevated glucose and adverse lipid profiles may show broader cardiometabolic dysfunction that tends to be associated with hypertension development (6). According to a recent study of blood pressure trajectories in young adults, traditional risk factors, including smoking, diabetes, and elevated lipids were reported to be associated with adverse blood pressure patterns over time (7). This suggests that hypertension should be understood as a part of a broader cluster of behavioral and metabolic risk rather than an isolated condition.

Despite this growing literature about hypertension, there is an important gap to note. Many prior studies identified risk factors of hypertension descriptively or examined only a limited number of variables at a time. However, even a fewer studies translated these relationships into an integrated quantitative framework with focus specifically on young adults. A multivariable modeling framework is useful since it estimates the relative contribution of several predictors simultaneously. This allows researchers to compare the strength of demographic, behavioral, and metabolic influences within one analytic structure.

To address this literature gap, this study uses a simulation-based quantitative modeling framework informed by prior epidemiological patterns among young adults in the U.S. By incorporating demographic characteristics, behavioral variables, anthropometric measures, and metabolic indicators into a single quantitative framework, this study seeks to identify which factors most strongly predict hypertension risk. This approach is intended to provide an interpretable modeling analysis that is grounded in established public health evidence (1-7). This study specifically aims to answer the research question about which measurable demographic, behavioral, and metabolic factors best predict hypertension among young adults in the U.S. This study hypothesized that young adults with higher body mass, poorer metabolic health, smoking exposure, and lower physical activity will have a substantially higher probability of hypertension than those with healthier profiles.

METHODS AND MATERIALS

Study Design

This study employed a quantitative modeling approach with a simulation-based dataset that was designed to represent demographic, behavioral, and metabolic characteristics of young adults in the United States. The purpose of the analysis was to assess how hypertension risk was jointly influenced by multiple demographic, behavioral, and metabolic factors within a young adult population. To allow the investigation of relationships among several predictors in a controlled analytic setting with transparency, a simulation framework was chosen to generate dataset.

Simulation-Based Data

As the primary data generation approach, simulation was selected due to the complex nature of publicly available health datasets containing detailed measures of blood pressure, behavioral risk factors, and metabolic biomarkers as they had complex multi-file merging procedures or restricted access conditions. In addition, there was an issue of limited availability of individual-level observations for open analysis due to privacy protections in many health datasets. Epidemiological parameters reported in prior literature were used when constructing a simulation-based dataset, providing a practical alternative for assessing statistical modeling methods while still preserving realistic population characteristics. The dataset was designed to reflect possible demographic distribution and risk factor relationships for the analysis to focus on identifying predictors of hypertension by grouping the simulation parameters in published epidemiological studies and also national health reports that described hypertension risk factors and population characteristics among young adults in the United States (1-3). Based on prior research, factors including elevated body mass, smoking behaviors, and metabolic indicators, such as higher glucose levels, have been consistently shown to be correlated with increased hypertension risk (4-6).

Data Construction

A dataset was generated, comprising a total of 2,000 simulated individuals aged from 18 to 39 years to approximate characteristics commonly reported by young adult populations in the United States. Parameter values used for the distributions of demographic and health variables followed the young adult health patterns reported by previously published epidemiological studies and national health reports (1-3). The simulation was designed to reflect realistic population patterns by the marginal distributions of main variables.

To improve reproducibility, the simulation parameters were specified before analysis. Age was generated as an integer variable in a range from 18 to 39 years. Sex was generated as a binary variable that was coded as 1 for male and 0 for female. Race/ethnicity was generated as a categorical variable with White as the reference group and Black, Hispanic, and Asian/Other as comparison groups. BMI was generated as a continuous variable measured in kg/m^2, and fasting glucose and total cholesterol were generated as continuous variables measured in mg/dL. Smoking status, high alcohol use, and physical activity were generated as binary variables. The simulation was designed to produce a final hypertension prevalence of around 10-15%, consistent with reported hypertension prevalence among young adults in the United States. The generated variables were used to calculate probability of hypertension of each participant through a logistic risk function, and the binary hypertension outcome was then generated probabilistically from that predicted risk.

In the dataset, variables representing demographic characteristic, behavioral risk factors, anthropometric measures, metabolic biomarkers, and a hypertension outcome variable were included. For demographic variables, age, sex, and race/ethnicity were included and categorized as White, Black, Hispanic, and Asian/Other. Behavioral variables included current smoking status, high alcohol use, and physical activity. Current smoking was coded as 1 for current smokers and 0 for non-smokers. High alcohol use was coded as 1 for participants exceeding recommended low-risk alcohol intake levels and 0 otherwise. Physical activity was coded as 1 for participants meeting recommended physical activity guidelines and 0 for participants not meeting those guidelines. In the logistic regression model, physical activity was expected to have a protective association with hypertension risk, while smoking and high alcohol use were expected to increase hypertension risk.

Anthropometric measurements included body mass index (BMI) that was expressed in kilograms per square meter. Fasting glucose levels and total cholesterol levels were measured in milligrams per deciliter and included in metabolic biomarkers. These variables were included particularly because they were identified as important factors of cardiometabolic health and contributors to hypertension risk from prior epidemiological research (4-6).

During the process of simulation, possible relationships between predictors and hypertension risk were applied according to findings from previous epidemiological studies. For example, there was an association between higher BMI values and increased probability of hypertension. Higher levels of physical activity were modeled as protective. In addition, both smoking behavior and elevated metabolic indicators were modeled as contributors to increase hypertension risk. These relationships were applied using probabilistic functions informed by prior epidemiological studies that reported associations between risk factors and hypertension.

Outcome Definition

The main outcome variable was hypertension status, and this was coded as a binary variable (1 = hypertension, 0 = no hypertension). Hypertension probability for each simulated individual was created by using a logistic risk function that applied multiple predictors, including age, BMI, smoking status, sex, physical activity, glucose levels, and race/ethnicity. To produce a hypertension prevalence of around 10-15%, parameter values were specifically selected to be consistent with estimates reported for young adult populations in the United States (3). This allowed the simulated dataset to maintain realistic epidemiological characteristics.

Statistical Analysis

To summarize the characteristics of the simulated population, descriptive statistics were first calculated. Continuous variables, including age, BMI, glucose, and cholesterol were summarized using means and standard deviations. In addition categorical variables, such as sex, race, smoking status, alcohol use, physical activity, and hypertension were summarized using frequencies and percentages.

For the assessment of associations between predictors and hypertension, binary logistic regression was used. The main mathematical model used in the analysis was binary logistic regression, estimating the log-odds of hypertension as a function of demographic, behavioral, and metabolic predictors. The model generated in this study was as follows:

$$log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Age_i + \beta_2 BMI_i + \beta_3 Smoking_i + \beta_4 Alcohol_i + \beta_5 PhysicalActivity_i + \beta_6 Glucose_i + \beta_7 Cholesterol_i + \beta_8 Sex_i + \beta_9 Race_i$$

Where p_i shows the probability of an individual I to have hypertension, β_0 is the intercept, and β_1 through β_9 indicate regression coefficients for the predictors. Since the hypertension status, the dependent variable, is binary, logistic regression was chosen. For the predictor variables, age, BMI, smoking status, alcohol use, physical activity, glucose, cholesterol, sex, and race/ethnicity were used. All predictors were used for the regression model simultaneously to estimate the independent correlation of each factor while controlling for the others.

Using a two-sided significance level of $\alpha = 0.05$, statistical significance was assessed. Regression results were reported as odds ratios (OR) that corresponded with 95% confidence intervals (CI) for the interpretation of the magnitude and direction of associations between predictors and hypertension risk. All simulation and statistical analyses were conducted using Python. Descriptive statistics, logistic regression, confidence intervals, p-values, likelihood ratio testing, and pseudo R^2, and ROC/AUC analysis were calculated using standard Python statistical libraries.

Model performance was assessed using several indicators. Pseudo R^2 and a likelihood ratio test were used to assess overall model fit. With them, the fitted regression model was compared with a null model containing only an intercept. Model discrimination was evaluated using the area under the receiver operating characteristic (ROC) curve. With this curve, the model's ability to correctly distinguish between individuals with and without hypertension was measured. The resulting ROC area under the curve (AUC) was calculated to be around 0.77, and this was the acceptable predictive performance for the multivariable model.

RESULTS

Sample Characteristics

In the simulated dataset, a total of 2,000 young adults in the ages from 18 to 39 were included. The mean age of the sample was around 28 years. This showed a balance distribution across the young adult age range. The mean body mass index (BMI) was around 27kg/m^2. From this, the average participant was placed in the overweight range. The mean fasting glucose was reported to be around 96 mg/dL, and the mean total cholesterol was around 193 mg/dL. The sample was almost evenly distributed by sex, and behavioral variables including smoking, alcohol use, and physical activity were represented at realistic frequencies within the population simulated in this analysis.

The prevalence of hypertension in the dataset was overall around 11.5%. Descriptive comparisons indicated that hypertension cases were more concentrated among participants with higher BMI values, smokers, males, and participants not satisfying physical activity guidelines. These patterns were reported to be consistent with the expected direction of correlation established into the simulation framework.

When grouping participants by BMI category, the percentage of the full simulated sample with hypertension was lowest in the normal-weight group, higher in the overweight group, and highest in the obese group. Since Figure 1 reports the percentage of the full sample rather than within-category prevalence, these values should be interpreted as the distribution of hypertension cases across BMI categories within the total simulated sample. A similar pattern was observed across sex. Particularly, male participants reported higher hypertension prevalence than female participants. Both anthropometric and demographic factors were correlated with variation in hypertension risk within the simulated sample as shown from the descriptive results.

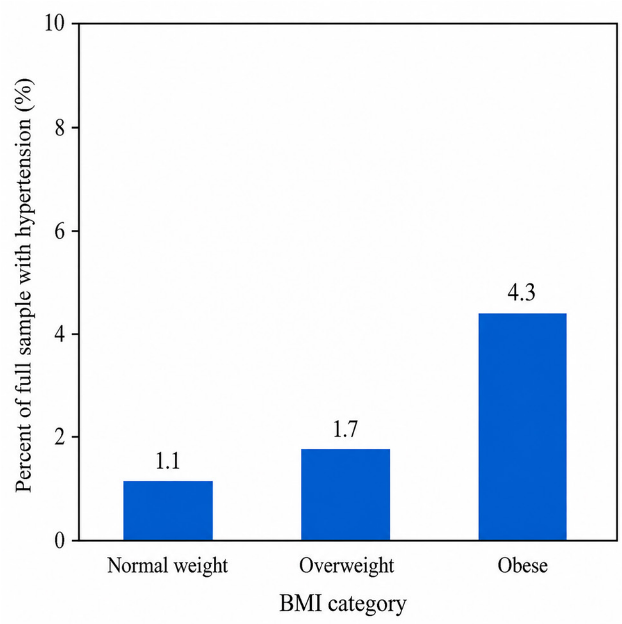

Figure 1. *Percentage of the Full Simulated Sample With Hypertension by BMI Category. This figure indicates the percentage of the total simulated sample with hypertension and belonged to each BMI category. Therefore, the values should be interpreted as each BMI category's contribution to overall hypertension cases in the full sample, rather than as within-category hypertension prevalence. The obese group contributed the largest share of hypertension cases in the simulated sample.*

Logistic Regression Findings

A binary logistic regression model was estimated including age, BMI, smoking status, alcohol use, physical activity, glucose, cholesterol, sex, and race/ethnicity to assess the independent association of each predictor with hypertension. Several predictors were statistically significant after adjusting the other variables in the model. Exact p-values were reported for all predictors in the final logistic regression model. Age, BMI, current smoking, physical activity, fasting glucose, male sex, and Black race were statistically significant predictors of hypertension risk. However, after adjustment, high alcohol use, total cholesterol, Hispanic race/ethnicity, and Asian/other race/ethnicity were not statistically significant predictors. The p-value for high alcohol use was corrected to 0.152 after verification of the statistical output confirming that alcohol use was not statistically significant in the adjusted model. Full adjusted regression results are shown in Table 1.

Table 1. *Adjusted Logistic Regression Results for Predictors of Hypertension Risk. This table reports adjusted odds ratio, 95% confidence intervals, and p-values from the logistic regression model. Values greater than 1 show higher modeled odds of hypertension, while the ones below 1 show lower modeled odds.*

Predictor	Odds Ratio	95% CI	p-value
Age	1.07	1.05-1.10	<0.001
BMI	1.17	1.14-1.21	<0.001
Current smoking	1.70	1.17-2.46	0.005
High alcohol use	1.30	0.91-1.88	0.152
Physical activity	0.71	0.53-0.95	0.022
Fasting glucose	1.03	1.02-1.04	<0.001
Total cholesterol	1.00	1.00-1.01	0.619
Male sex	1.82	1.35-2.45	<0.001
Black race	1.79	1.18-2.71	0.006
Hispanic race/ethnicity	1.33	0.92-1.93	0.130
Asian/other race/ethnicity	0.95	0.60-1.52	0.841

First of all, age was positively correlated with hypertension. The odds ratio for age was calculated to be around 1.07 per year, $p < 0.001$. This showed that older participants within the young adult range had modestly higher odds of hypertension than younger individuals. Even though this study focused only participants in the ages from 18 to 39, this result suggests that hypertension risk gradually increases during early adulthood.

Second, BMI was one of the strongest predictors in the model. To be more specific, even 1-unit increase in BMI was correlated with an odds ratio of around 1.17, $p < 0.001$. This indicated that higher body mass was strongly correlated with higher odds of hypertension. This result was consistent with the descriptive pattern in hypertension prevalence observed across BMI categories.

Third, smoking status was also substantially correlated with hypertension. Current smokers had around 1.70 times the odds of hypertension compared to non-smokers, $p=0.005$. In contrast, physical activity indicated a protective association. Participants meeting physical activity guidelines had an odds ratio of around 0.71, $p=0.022$. This showed lower odds of hypertension relative to less active participants.

Among the metabolic variables, fasting glucose was positively correlated with hypertension, and the odds ratio was reported to be around 1.03 per mg/dL, $p < 0.001$. This shows that higher glucose levels were positively correlated with higher odds of hypertension in the model. Total cholesterol indicated a weaker correlation and was not statistically significant after adjusting for the other predictors, OR = 1.00, $p = 0.619$ (Figure 2).

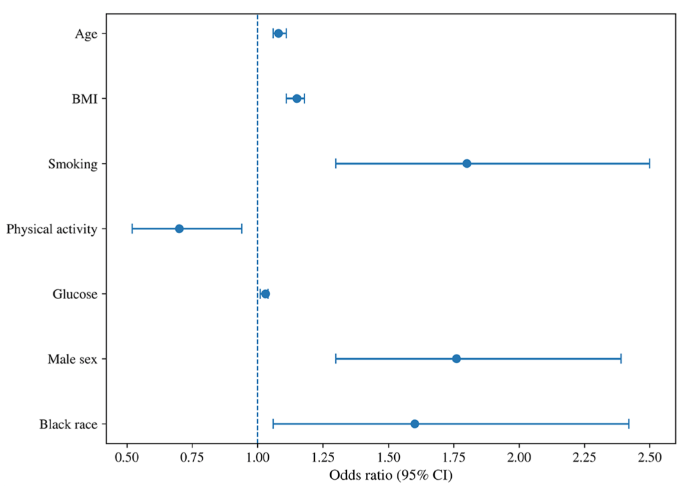

Figure 2. *Adjusted Odds Ratios for Hypertension Predictors. This figure shows summarized logistic regression model using odds ratios, along with 95% confidence intervals. BMI, smoking, glucose, age, male sex, and Black race were positively correlated with hypertension. Physical activity indicated a protective association with an odds ratio below 1.*

Demographic Differences

Several demographic patterns were significant in the adjusted model. Male sex was correlated with higher odds of hypertension, and the odds ratio was reported to be around 1.82 compared with females, $p < 0.001$. In addition, participants categorized as Black reported higher odds of hypertension than the White reference group. The same result was shown after considering behavioral and metabolic variables. The estimated odds ratio for Black participants was around 1.79, $p = 0.006$. Among Hispanic and Asian/Other categories, there were no statistically significant differences compared with the reference group in the final model.

Model Performance

The overall logistic regression model indicated acceptable performance. The likelihood ratio test was statistically significant ($p < 0.001$) when comparing the fitted model with a null model. This indicated that the predictor set improved the model fit. Pseudo of the model suggested moderate explanatory power. Furthermore, the area under the receiver operating characteristic curve (AUC) was around 0.77, and this indicated acceptable discrimination between participants with and without hypertension (Figure 3).

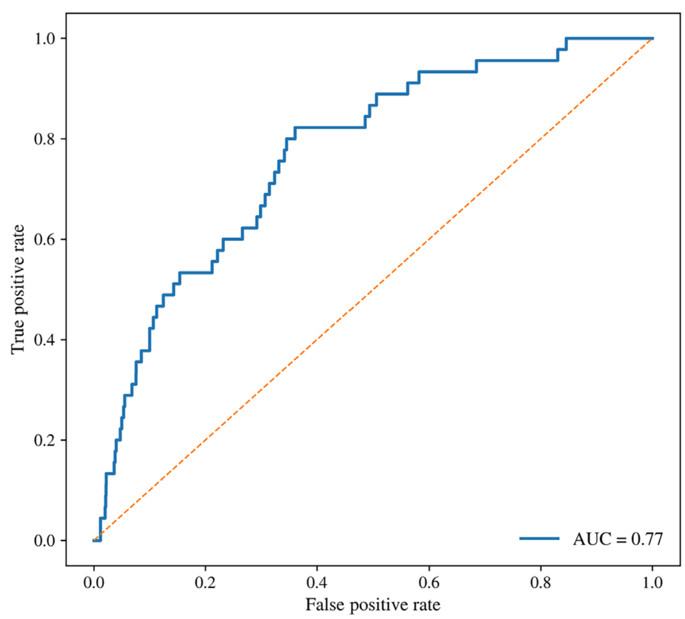

Figure 3. *Receiver Operating Characteristic Curve for the Logistic Regression Model. The ROC curve assesses the ability of the model to distinguish participants with and without hypertension. The area under the curve was around 0.77. This shows acceptable discriminatory performance for the multivariable model.*

DISCUSSION

In this study, a simulation-based quantitative model was used to examine predictors of hypertension among young adults in the United States. The findings demonstrate that a combination of demographic, behavioral, and metabolic factors is influential in hypertension risk in this group of young adults. In the adjusted model, participants with higher BMI, smoking, higher glucose, old age within the range from 18 to 39, male sex, and Black race were correlated with higher odds of hypertension. On the other hand, physical activity was correlated with lower odds. These findings were consistent with prior public health research, demonstrating that blood pressure risk begins before middle age and is shaped by multiple interacting influences rather than a single cause.

Among the predictors used in the analysis, BMI was one of the strongest correlations with hypertension. This finding is important in a sense that excess body weight is a modifiable factor and is substantially associated with prevention strategies in young adults. As shown from the descriptive increase in hypertension prevalence across BMI categories combined with the significant adjusted odds ratio, body weight plays an important role in modeled blood pressure risk. This finding supports prior evidence that obesity is associated with vascular strain, altered metabolic regulation, and greater cardiovascular burden (5).

Smoking was also positively correlated with higher hypertension risk. However, physical activity indicated a protective relationship. These two findings highlight that life style factors in early cardiovascular health are still important. Even in a young adult population, tobacco exposure-related behaviors and exercise appear to substantially shape blood pressure risk. This is especially relevant from a perspective of public health since both decreased smoking and increased activity are realistic targets for intervention.

The positive correlation between glucose and hypertension further demonstrates that cardiometabolic dysfunction may already be present in young adults at risk. Even if cholesterol was not statistically significant in the final model, glucose was still informative. This shows that several metabolic markers may be more closely related to blood pressure than others within this framework. Furthermore, the higher odds were observed among male and Black participants in the analysis. This was consistent with broader patterns of hypertension disparity in the U.S., suggesting that demographic differences were still important even after considering other modeled variables.

Several limitations should be noted in this study. First, the dataset generated in this study was from simulation rather than from observational source. Therefore, the results should be interpreted as estimates based on the model rather than direct population measurements. Second, several potentially relevant predictors, including sodium intake, family history, healthcare access, and socioeconomic status were not included. Despite aforementioned limitations, this study provides an important insight about how a simulation-based model can provide an interpretable and useful framework for assessing hypertension risk.

CONCLUSION

In this study, a simulation-based quantitative framework was used to examine predictors of hypertension among young adults in the United States. The findings indicated that hypertension risk was correlated with multiple factors, including higher body mass index, smoking, increased glucose, older age within the young adult range, male sex, and Black race. Physical activity indicated a protective association with hypertension risk. These results show that a combination of demographic, behavioral, and metabolic influences was influential in hypertension in young adults rather than a single predictor.

This study also indicates the usefulness of logistic regression as a mathematical modeling tool when assessing the relative contribution of multiple risk factors within one framework. Even if the dataset was generated by a simulation and should not be interpreted as direct observational evidence, the modeled relationships were informed by published epidemiological patterns. Overall, the findings in this study emphasize the importance of early prevention efforts with focus on weight management, decreased smoking, increased physical activity, and cardiometabolic health to reduce long-term hypertension risk among young adults. Future studies should validate these simulation-based findings with observational cohort data by including additional predictors, such as sodium intake, family history, socioeconomic status, healthcare access, and medication history to improve the real-world applicability of the model.

CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

REFERENCES

1. Centers for Disease Control and Prevention. About high blood pressure. Atlanta (GA): CDC; 2026 Jan 28. Available from: https://www.cdc.gov/high-blood-pressure/about/index.html. Accessed 2026-01-28.
2. World Health Organization. Hypertension. Geneva: WHO; 2025 Sep 25. Available from: https://www.who.int/news-room/fact-sheets/detail/hypertension. Accessed 2026-01-02.
3. Fryar CD, Ostchega Y, Fakhouri THI. Hypertension prevalence, awareness, treatment, and control among adults age 18 and older: United States, August 2021–August 2023. *NCHS Data Brief*. 2024 Oct; (511): 1-8. Available from: https://www.cdc.gov/nchs/products/databriefs/db511.htm. Accessed 2026-02-21.
4. Tang R, An J, Bellows BK, Moran AE, and colleagues. Trends in hypertension prevalence, awareness, treatment, and control among US young adults, 2003–2023. *Am J Hypertens*. 2025; 38 (8): 551-559. Available from: https://academic.oup.com/ajh/article/38/8/551/8100303. Accessed 2026-01-21. https://doi.org/10.1093/ajh/hpaf044
5. Papathanasiou G, Zerva E, Zacharis I, Papandreou M, Papageorgiou E, Tzima C, *et al*. Association of high blood pressure with body mass index, smoking and physical activity in healthy young adults. *Open Cardiovasc Med J*. 2015; 9: 5-17. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC4378027/. Accessed 2026-01-11. https://doi.org/10.2174/1874192401509010005
6. An J, Fischer H, Ni L, Xia M, Choi SK, Morrissette K, *et al*. Association between young adult characteristics and blood pressure trajectories. *J Am Heart Assoc*. 2024; 13 (8): e033053. Available from: https://www.ahajournals.org/doi/10.1161/JAHA.123.033053. Accessed 2026-01-11. https://doi.org/10.1161/JAHA.123.033053
7. Nagata JM, Vittinghoff E, Pettee Gabriel K, Garber AK, Moran AE, Sidney S, Rana JS, Reis JP, Bibbins-Domingo K. Physical activity and hypertension from young adulthood to middle age. *Am J Prev Med*. 2021; 60 (6): 757-765. Available from: https://pubmed.ncbi.nlm.nih.gov/33867211/. Accessed 2025-12-22. https://doi.org/10.1016/j.amepre.2020.12.018