Age and Beyond: Investigating Hypertension Risk Factors

Carol Sanchez Garibay, Irene Na, Francesca Scipioni

April 2024

Introduction

Hypertension occurs when blood pressure consistently exceeds normal levels, typically measured through systolic (during heartbeats) and diastolic (between beats) readings. Diagnosis occurs if systolic readings exceed 140 mmHg or diastolic readings exceed 90 mmHg on multiple occasions. Recent research suggests systolic pressure is a stronger indicator, especially for older adults[1]. Hypertension, associated with serious health complications like heart disease, stroke, and kidney failure, affects an estimated 1.28 billion adults aged 30–79 worldwide[2]. Highlighted in American Heart Association reports, recent medical research underscores systolic blood pressure's pivotal role in hypertension diagnosis, and its connection to age[3]. Pinto (2007)[4] attributes age-related blood pressure increases to arterial changes and stiffness. The Centers for Disease Control and Prevention (CDC) estimates from 2015-2016 reveal hypertension prevalence rises with age: 7.5% among adults aged 18–39, 33.2% among those aged 40–59, and 63.1% among those aged 60 and over[5]. Based on these researches, age's connection with blood pressure dynamics is significant, yet hypertension's complexity also extends to factors including genetics, race, overweight/obesity, sedentary lifestyle, salt consumption, and alcohol and tobacco use[3]. Our research aims to delve deeply into the relationship between age and systolic blood pressure (SBP), while acknowledging the multifactorial nature of hypertension. Our work will center around the following research question:

> **How does age relate to systolic blood pressure, and to what extent does this relationship change when considering additional risk factors?**

Leveraging CDC survey data and contextualizing our analysis within hypertension risk factors, we aim to provide comprehensive insights into blood pressure patterns across diverse populations, informing management strategies that promote public health and reduce cardiovascular disease burden.

Data, Models and Visualizations

This study uses data from the 2017-2018 National Health and Nutrition Examination Survey (NHANES), conducted by the CDC's National Center for Health Statistics (NCHS). The sample includes non-institutionalized U.S. civilians over 20 years old. NHANES over-samples of certain groups to ensure representative statistics. The survey collects person-level demographic, health, and nutrition information through a standardized process.

In our study, we utilize variables from the NHANES survey dataset to represent the key input variable age (X), SBP (Y) and other important hypertension risk factors identified in existing research. The "Age" and "SBP" variables directly reflect the conceptual information of each participant, while the risk factors are measured using closely related data proxies (as seen in below discussion). The NHANES data comprises five main categories: Demographic, Dietary, Examination, Laboratory, and Questionnaire, each containing multiple datasets with specific information. We downloaded relevant datasets and developed functions to clean and merge them into a single dataframe, eliminating missing values and entries with invalid SBP values (i.e., SBP=0), and using the unique identifier $SEQN$ for merging. The final dataset has 4539 entries and 7 features. Subsequently, the dataset was randomly split into exploration and confirmation sets, with a 30% and 70% split respectively. All decisions regarding building intuition, data exploration, and model specification were based on the exploration set.

Our model specifications are closely tied to our research question: whether a relationship exists between age and SBP. We explored this relationship through a series of OLS regressions using SBP as the outcome

[1] Thor Christensen, American Heart Association News. (2021)."Which blood pressure number matters most? The answer might depend on your age."

[2] World Health Organization. (2023). "Hypertension."

[3] American Heart Association. (2023). "Know Your Risk Factors for High Blood Pressure"

[4] Pinto E. Blood pressure and aging. Postgrad Med J. 2007 Feb;83(976):109-14. doi: 10.1136/pgmj.2006.048371. PMID: 17308214; PMCID: PMC2805932.

[5] Fryar C.D. et al. NCHS Data Brief No. 289, October 2017.

variable and age as the primary input variable. We begin with the simplest regression model including only age. Sequentially, we incorporate other established hypertension risk factors (identified in peer-reviewed literature and summarized in the introduction) to assess how those factors affect the "Age" coefficient of the regression. The other risk factors include (data proxies denoted in brackets): Body Mass Index (BMI), physical inactivity (sedentary hours per day), alcohol consumption (drinks per day), and genetics (race, gender). In our models, we did not intentionally leave out any variable to assess incremental impact.

Table 1: Summary Statistics of the Dataset

Variable	Observations	Mean	Std. Dev	Median
Sys. Blood Preassure (mmHg) (Y)	1325	127	21	123
Age (X)	1325	52	17	54
BMI (X)	1325	30	7.2	28
Sedentary hours/day (X)	1325	5.5	3.3	5
Alcoholic drinks/day (X)	1325	1.6	2.1	1

Table 1 presents summary statistics for the key numeric variables. The average SBP is 127 mmHg, with a median of 123 mmHg, indicating a slight right skew, as also observed in *Figure 1A*. The average age is 52, and the mean BMI is 30, suggesting that individuals in the sample are, on average, overweight (BMI > 25) and older than the general population. Alcohol consumption is low, with an average of 1.6 drinks per day. The other panels of **Figure 1** offer insights for model decisions. *Figure 1B* illustrates a visually positive association between Age and SBP, with a concentration of values around the age of 80 (due to all individuals older than 80 years being reported as age 80). While weight is expected to influence SBP, the BMI vs SBP plot in *Figure 1C* does not reveal a clear association. However, after applying a logarithmic transformation to BMI, the data spreads out, revealing a linear pattern (*Figure 1D*). Lastly, *Figure 1E* shows that the African American group exhibits the highest average SBP.

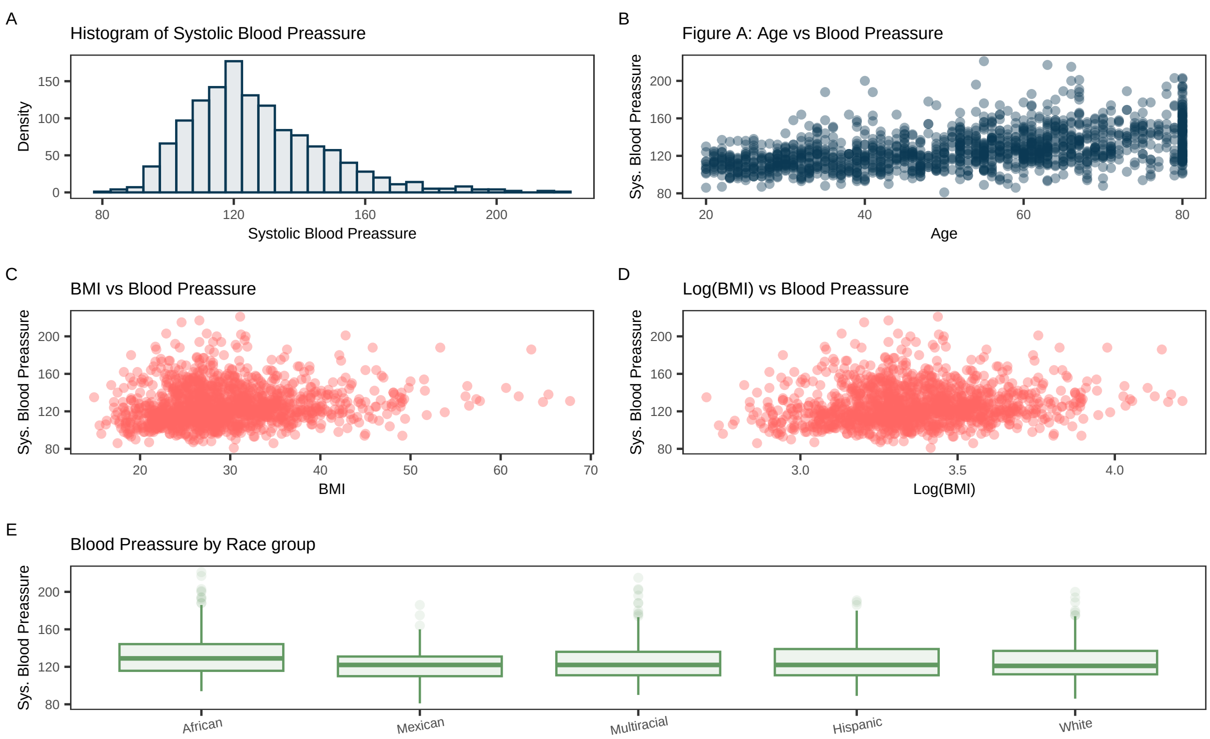

Figure 1: Grid plots of relationships between explanatory variables and blood preassure

Large sample linear regression requires two key assumptions: I.I.D. samples, and the unique BLP exists. Given the strict NHANES sampling process and wide coverage, we considered samples are largely independent and from the same distribution. However, the independence may be challenged given the samples are tilted towards elderly population. We examined and ruled out perfect collinearity and heavy tails (See *Figure 1A*).

Upon examining data and fundamental relationships, we explored two input variable transformations. **Firstly**, we considered using the logarithm of BMI to address BMI's skewed distribution and ensure more consistent residuals' dispersion from a regression of SBP on BMI. **Secondly**, we introduced an interaction term between log(BMI) and sedentary hours to understand how time of physical inactivity directly and indirectly affects SBP through BMI changes. Per peer-reviewed literature, sedentary lifestyle is known to contribute to weight gain, which also influences blood pressure. This allows us to capture both the individual effects of these variables and their combined influence on SBP.

Table 2: Systolic Blood Pressure and Risk Factors: OLS Regression Summary

	Dependent variable:		
	Systolic Blood Pressure (mmHg)		
	(1)	(2)	(3)
Age	0.519*** (0.017)	0.516*** (0.017)	0.541*** (0.018)
log(BMI)		11.420*** (1.371)	12.553*** (2.609)
Sedentary hrs/day			0.940 (1.297)
log(BMI):Sedentary			−0.310 (0.383)
Alcohol drinks/day			0.502** (0.177)
Mexican American			−4.381*** (1.112)
Other/Multi-Racial			−4.599*** (0.955)
Other Hispanic			−5.125*** (1.288)
White			−6.470*** (0.870)
Male			1.666** (0.620)
Constant	100.397*** (0.813)	62.087*** (4.578)	60.250*** (8.835)
Two-model F-test Pr(>F)(vs lhs model)	N/A	1.306e-17 ***	1.314e-14 ***
Observations	3,214	3,214	3,214
Adjusted R^2	0.215	0.232	0.250
F Statistic	878.544***	486.149***	107.909***
Note:	*p<0.05; **p<0.01; ***p<0.001		
	Race: African American is omitted; gender: female is omitted; last F-stat is vs constant		

Results, Limitations and Conclusions

Key model findings are derived from the confirmation set, using robust standard errors, and summarized in **Table 2**. **1.** The coefficient for age is consistently significant at 0.1% significance level across all model specifications and maintains a stable magnitude. This indicates a meaningful linear relationship between SBP and age, suggesting that a one-year increase in age is associated with a approximately 0.5 mmHg increase in SBP, all else equal. **2.** BMI and race all have statistically significant coefficients at 0.1% significance level, supporting their established roles in influencing SBP. **3.** Both alcohol consumption and gender coefficient is significant at 1% significance level. **4.** Neither the direct effect of sedentary hours on SBP nor the indirect effect through BMI changes were statistically significant.

In terms of practical significance, these results addressed the main research question regarding SBP and age, and regarding the additional key risk factors, which are aligned with existing peer-reviewed publications. While the adjusted R^2 is only around 20, indicating a moderate level of variance explained, the significant F-statistic for all models (compared to the restrictive model with only a constant term), confirm that the included input variables help to describe changes in SBP.

We recognized that our descriptive model analysis has limitations. **Population representation and modeling**: The dataset, although collected from a reputable source, does not reflect the US population's age distribution, skewing towards the elderly population and potentially compromising the I.I.D. assumption. **Data availability on other factors**: Public literature highlights various environmental and genetic factors influencing blood pressure. Unfortunately, our dataset lacks comprehensive data on these factors (e.g. salt overconsumption, family history, ect.). **Medical expertise**: due to lack of domain expertise, we relied on variables discussed in literature, potentially overlooking relevant factors.

Despite these limitations, our study provides evidence that age has a significant linear relationship with SBP. Although the coefficient's magnitude is small, the practical significance remains meaningful. This study could be of interest to healthcare and research institutions aiming to develop preventative healthcare strategies and promote healthy lifestyle.