Investigating the Association of Hypertension and Alcohol Intake

Abstract

Background

Previous studies have shown an association between hypertension and alcohol consumption. In this study, we investigated the association between systolic blood pressure (a biomarker for hypertension) and alcohol consumption through linear regression.

Methods

Using a sample of 4,802 adults from the National Health and Nutrition Examination Survey, we analyzed the relationship between combined systolic blood pressure reading and alcohol consumption (having 12 or more drinks a year, number of drinks had on an average drinking day, and number of days drinking alcohol throughout the year), adjusting for age, BMI, and gender through multiple linear regression. We also analyzed the change in the effect of alcohol on systolic blood pressure as age increases through interaction terms.

Results

For each one drink increase per day, a subject's log-transformed systolic blood pressure significantly increases by 0.003 (95% CI: 0.001, 0.004, $p < 0.001$) units, adjusting for all other covariates. Also, those who have consumed at least 12 drinks in the past year, were on average found to have a log-transformed systolic blood pressure of 0.06 (95% CI: 0.024, 0.097, $p = 0.001$) units significantly higher than those who did not consume 12 drinks in the past year, adjusting for all other covariates. The interaction term used in our final model between age and 12 or more alcoholic drinks per year showed a slight negative association with Systolic Blood Pressure ($\beta = -0.001$, 95% CI: -0.002 to 0, $p = 0.005$).

Conclusions

We found a positive association between alcohol consumption and log-transformed systolic blood pressure, adjusting for age, BMI, and gender. Due to the missingness in our sample, these results generalize to the population of US adults who consume alcohol. However, with an adjusted R squared of 0.185, there is still a lot of variation in systolic blood pressure that is unexplained by our model. Future research into this association is needed to overcome these limitations.

Introduction

Hypertension is the leading cause of premature death, and cardiovascular disease worldwide.[3] The presence of hypertension is globally rising, especially in low and middle income countries. This rise is partly explained by increases in lifestyle risks, such as obesity, low physical activity, alcohol consumption, and an overall unhealthy diet.[3]

Previous studies have shown an association between hypertension and alcohol consumption. A study which divided alcohol consumption into five categories, (≥1 to <140; ≥140 to <210; ≥210 to <280; ≥280 to <420; and ≥420 g/week), showed an association between elevated systolic blood pressure in all categories of consumption. [4] Another study found alcohol consumption was significantly associated with elevated blood pressure, especially at the highest consumption amount.[2] Past studies have also shown specific confounding variables which should be considered when studying the association between hypertension and alcohol consumption. A study found that high levels of alcohol consumption was associated with an increased risk of hypertension when adjusting for age, education, height, bmi, gender, smoking status, and physical activity.[5] A second study found the highest association between hypertension and

alcohol assumption in those consuming ≥3 drinks per day, when adjusting for sex, body mass index, age, and parental history[1].

Data from a large cross-sectional survey designed to assess the nutritional and health statuses of children and adults in the United States were analyzed. The survey provides baseline information on demographics, physical measurements including average systolic blood pressure, and health and lifestyle variables including alcohol consumption. In this study, we examined the association between alcohol consumption and hypertension.

Methods

Study population

Our study population comes from the 2009-2010 and 2011-2012 sample cycles of the National Health and Nutrition Examination Survey (NHANES) program at the Centers for Disease Control and Prevention (CDC). This survey assesses the health and nutrition of non-institutionalized civilian resident adults and children in the United States through interviews and physical examinations. The original samples were collected to intentionally oversample certain subpopulations such as racial minorities, so our study cohort started with 10,000 individuals resampled from the original NHANES data from these years. We treat this resampled cohort as a simple random sample from the United States population.

Of the 10,000 individuals, 5,197 were removed due to missingness in the combined systolic blood pressure reading (n = 1,449), BMI (n = 43), or alcohol-related variables (n = 3,705). This process also removed all individuals under the age of 18 (n = 1,346), as the alcohol consumption questions were only asked for participants 18 years or older. Additionally, one participant was removed for being an outlier with a claimed 82 average alcoholic drinks in a day where they drink. We believe that this is an error, as 82 drinks is an improbable amount of alcohol to consume in one day. Thus, we end with a sample of 4,802 participants.

After removing subjects under 18 years old from the analysis, we compare the distributions of our variables of interest between our sample and the remaining subjects with missingness (Table 1). While the distributions of combined systolic blood pressure, age, BMI, and gender are similar between the two groups, the alcohol-related variables have different distributions. We believe that people who drink alcohol were more likely to answer all of the alcohol-related survey questions than those who do not, as 92% of our sample answered “yes” to consuming at least 12 drinks of alcohol per year. We analyzed subjects with complete data, so our study population mostly contains US adults who drink alcohol.

Table 1. Characteristics of Subjects with Missing Data and the Sample

Characteristic	Missing N = 2,679[1]	Sample N = 4,802[1]
Combined Systolic Blood Pressure	122 (18)	120 (16)
Age	48 (19)	45 (16)
BMI	29 (7)	28 (6)
Gender		
female	1,561 (58%)	2,234 (47%)
male	1,118 (42%)	2,568 (53%)
12 or more Alcoholic Drinks in 1 Year: Yes	782 (44%)	4,430 (92%)
Average Alcoholic Drinks in 1 Day	4 (8)	3 (3)
Days Drinking Alcohol in 1 Year	11 (49)	90 (107)

[1] Mean (SD); n (%)

Variables

Predictors of Interest

From Lifestyle Variables, we have three alcohol consumption variables which all reported the age of participants 18 or older. Average Alcoholic Drinks in 1 Day (AlcoholDay) is a continuous variable representing the average number of alcoholic drinks participants consumed on one day. Days Drinking Alcohol in 1 Year (AlcoholYear) is a continuous variable representing the average number of alcoholic drinks participants consumed in the past whole year. 12 or more Alcoholic Drinks in 1 Year (Alcohol12PlusYr) is a categorical variable representing Participants who consumed 12 or more alcoholic drinks in any of the years, with "No" serving as the reference group.

For the interaction term, 12+ Drinks x Age: This takes into account whether the effect of 12 or more alcohol consumption on systolic blood pressure changes across different ages, adjusting for other covariates. Drinks in 1 Day x Age: This interaction term takes into account whether the effect of Alcoholic Drinks in 1 Day on systolic blood pressure changes across different ages, adjusting for other covariates. This interaction term takes into account whether the effect of Alcoholic Drinks in 1 year on systolic blood pressure changes across different ages, adjusting for other covariates. Drinks in 1 Year x Age: This interaction term takes into account whether the effect of Alcoholic Drinks in 1 year on systolic blood pressure changes across different ages, adjusting for other covariates.

Outcome

Subjects were analyzed for the relationship between systolic blood pressure (BPSysAve) and alcohol consumption patterns, specifically focusing on daily alcohol consumption

(AlcoholDay) and annual drinking frequency (AlcoholYear). The analysis adjusted for critical demographic and health-related covariates, including age, BMI, and gender.

The primary focus of the study was to evaluate how these factors influence systolic blood pressure while addressing key statistical assumptions through model diagnostics. Analytical adjustments included the exclusion of extreme values (e.g., implausible alcohol consumption of 82 drinks per day), imputation or removal of missing values, and the standardization of continuous covariates like age and BMI. Interaction terms were incorporated to capture age-dependent effects, and a log transformation of BPSysAve was applied to address skewness and heteroscedasticity.

This comprehensive approach ensured the robustness of the analysis while providing a clear framework for understanding the interplay between alcohol consumption and cardiovascular health.

Covariates

In this analysis, covariates played a crucial role in controlling for confounding factors and isolating the effects of alcohol consumption on systolic blood pressure (BPSysAve). The primary covariates considered were age, BMI, and gender, each selected for its well-documented association with blood pressure. Age, treated as a continuous variable, is widely recognized as a critical predictor of systolic blood pressure, with levels typically increasing due to physiological changes such as arterial stiffening. BMI, a standard measure of overall body fat, was included as an important predictor given its established link to cardiovascular strain and hypertension. Gender, included as a binary variable (Male = 1, Female = 0), accounted for the general observation that men tend to have higher blood pressure levels than women, particularly in middle-aged and older populations.

Interaction terms, such as age by alcohol consumption, were incorporated to capture potential age-dependent variations in the relationship between alcohol consumption and blood pressure. Diagnostic checks were performed to assess multicollinearity among the covariates, with Variance Inflation Factors (VIFs) calculated to confirm independence. All covariates were found to have acceptable VIF values, indicating minimal multicollinearity issues.

Statistical Analyses

In this study, we aimed to evaluate the relationships between systolic blood pressure and alcohol consumption patterns, adjusted for demographic and health-related covariates. Alcohol consumption has been widely recognized as a modifiable risk factor for hypertension, but its effects can vary depending on consumption patterns and individual characteristics such as age, BMI, and gender.

To address the association between systolic blood pressure and alcohol consumption, three regression models we have developed:

1) Baseline Model: A linear regression model was constructed to examine the direct effects of Average Alcoholic Drinks in 1 Day, Days Drinking Alcohol in 1 Year, 12 or more Alcoholic Drinks in 1 Year, Age, BMI, and Gender on Systolic Blood Pressure
2) Interaction Model: Interaction terms (e.g., Age x AlcoholDay) were included to explore age-dependent variations in the effects of alcohol consumption.
3) Log-Transformed Model: The dependent variable systolic blood pressure(BPSysAve) was log-transformed to address skewness and heteroscedasticity in the residuals, and the same predictors as the adjusted model were included.

To ensure the validity of the regression models, we used rigorous diagnostic tests to evaluate key assumptions of those linear regressions:

1) Linearity: Partial regression plots were examined to confirm linear relationships between predictors and Systolic Blood Pressure.
2) Independence of Residuals: The Durbin-Watson test was conducted to evaluate autocorrelation in residuals.
3) Normality: Residual histograms, Q-Q plots, and Shapiro-Wilk tests were used to assess the normality assumption.
4) Homoscedasticity: Residuals vs. fitted values plots were reviewed to detect heteroscedasticity.
5) Multicollinearity: Variance Inflation Factors (VIFs) were calculated to assess collinearity among predictors.

To find out and delete the outlier, we provided the plot of influence diagnostics for the outcome(Systolic Blood Pressure), cook's d bar plot and outlier and leverage diagnostics plot. Influence diagnostics were performed to identify and remove outliers. Cook's distance was used to detect observations exerting undue influence on the regression coefficients. A bar plot of Cook's distance values and diagnostic plots for leverage and residuals were generated to visually assess potential outliers and high-leverage points. Observations exceeding the threshold of 4/n for Cook's distance or exhibiting extreme leverage were flagged for further inspection.

To find out the association between alcohol consumption and high blood pressure, we also provided the correlation plot with its analysis. Correlation analysis and plots were employed to explore the relationships between alcohol consumption and demographic factors on systolic blood pressure. This method facilitated the identification of potential multicollinearity, guided the selection of key predictors, and ensured the integrity of the data. By providing both numerical and visual insights, correlation analysis served as a foundational step in building a robust regression model and communicating the data structure effectively.

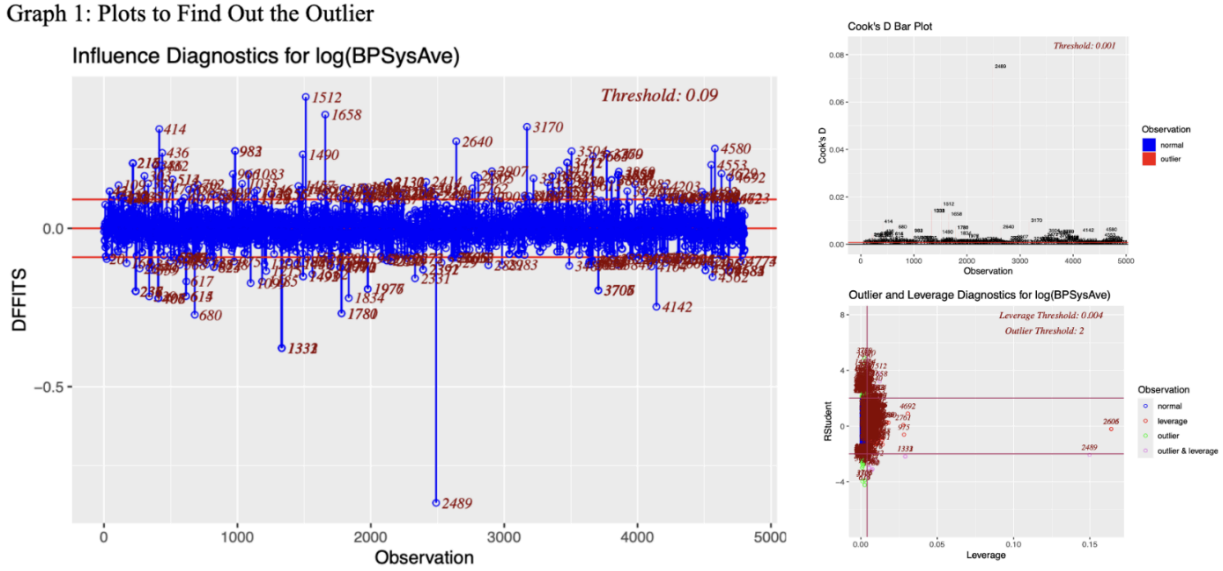
All reported estimation intervals are 95% confidence intervals. A critical level of 0.05 was used to determine statistical significance. All results were performed using R 4.3.1.

Results

Remove the Outlier

Influence diagnostics, including Cook's distance and leverage plots, identified observation 2489 as an influential outlier. Given its potential to unduly impact the model, we excluded this observation at the outset of the analysis to ensure robust and reliable results

Graph 1: Plots to Find Out the Outlier



Regression Model Findings

In the baseline model ($SBP_i = \beta_0 + \beta_1 AD_i + \beta_2 AY_i + \beta_3 12AD_i + \beta_4 Age_i + \beta_5 BMI_i + \beta_6 G_i + \varepsilon_i$), the R^2 value was 0.1868, which is around the 0.2 threshold, indicating that the results reliably reflect the influence of average alcoholic drinks in 1 day on systolic blood pressure while accounting for confounders such as Age, BMI, and Gender. Average alcoholic drinks in 1 day was positively associated with systolic blood pressure (β = 0.354, 95% CI: 0.204 to 0.505), suggesting that increased daily alcohol consumption is linked to higher systolic blood pressure. Since the regression equation was statistically significant ($p < 0.001$), we can conclude that average alcoholic drinks in 1 day is significantly associated with systolic blood pressure after adjusting for other variables (Table 2).

There was no multicollinearity between those variables with all Variance Inflation Factor (VIF) values below 5. Residual diagnostics confirmed linearity, with residuals showing a little bit right-skewed (Table 3). The Durbin-Watson statistic was 1.23, suggesting positive autocorrelation that warrants further model refinement.

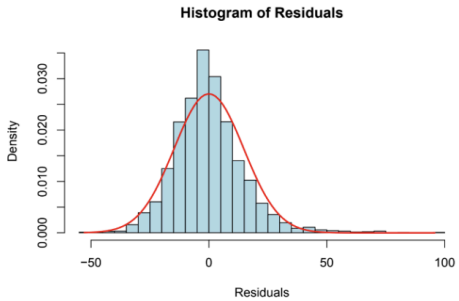
Table 2: Baseline Model Output

Model Results
Linear Regression: BPSysAve[1]

Characteristic[2]	Beta	95% CI	p-value	VIF
Average Alcoholic Drinks in 1 Day	0.354	(0.204, 0.505)	< 0.001	1.14
Days Drinking Alcohol in 1 Year	0.013	(0.009, 0.017)	< 0.001	1.15
12 or more Alcoholic Drinks in 1 Year: Yes	0.142	(-1.485, 1.769)	0.864	1.08
Age	0.362	(0.335, 0.389)	< 0.001	1.15
BMI	0.333	(0.266, 0.399)	< 0.001	1.03
Male Gender	4.341	(3.471, 5.212)	< 0.001	1.08

[1] R^2: 0.1868, Adjusted R^2: 0.1857
[2] Reference Groups: Female, Alcohol12PlusYrNO

Table 3: Normality of Baseline Model



In the interaction model ($SBP_i = \beta_0 + \beta_1 AD_i + \beta_2 AY_i + \beta_3 12AD_i + \beta_4 Age_i + \beta_5 BMI_i + \beta_6 G_i + \beta_7 12AD_i * Age_i + \beta_8 AD_i * Age_i + \beta_9 AY_i * Age_i + \varepsilon_i$), the R^2 value was 0.1927, which is around the 0.2 threshold, indicating that the results reliably reflect the influence of average alcoholic drinks in 1 day on systolic blood pressure. The inclusion of the interaction term Average Alcoholic Drinks in 1 Day:Age (Beta = 0.023, $p < 0.001$) reveals that the effect of daily alcohol

consumption on Systolic Blood Pressure becomes more pronounced with age, while 12 or more Alcoholic Drinks in 1 Year:Age (Beta = -0.138, $p < 0.01$) suggests that heavy drinking has a diminishing impact as age increases. The inclusion of interactions alters the main effect of Average Alcoholic Drinks in 1 Day, which now shows a negative association with Systolic Blood Pressure (Beta = -0.422, $p = 0.024$) (Table 4).

Additionally, high VIFs for interaction terms (up to 19.3) indicate multicollinearity, complicating the interpretation of coefficients and limiting the reliability of this model. Residual diagnostics confirmed linearity, with residuals showing a little bit right-skewed (Table 5). The Durbin-Watson statistic was 1.226, suggesting mild positive autocorrelation that should be considered in further modeling efforts.

Table 4: Interaction Model Output

Model Results

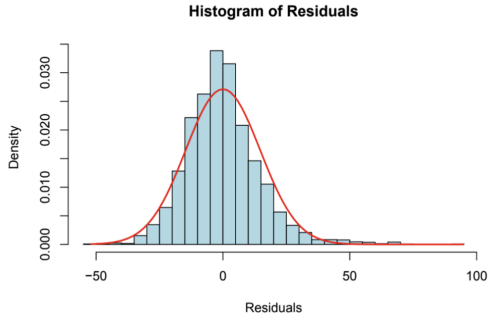
Linear Regression: BPSysAve[1]

Characteristic[2]	Beta	95% CI	p-value	VIF
Average Alcoholic Drinks in 1 Day	-0.422	(-0.788, -0.055)	0.024	6.86
Days Drinking Alcohol in 1 Year	0.025	(0.011, 0.039)	< 0.001	12.42
12 or more Alcoholic Drinks in 1 Year: Yes	6.538	(1.842, 11.234)	0.006	9.09
Age	0.457	(0.371, 0.543)	< 0.001	11.61
BMI	0.325	(0.259, 0.391)	< 0.001	1.04
Male Gender	4.152	(3.28, 5.024)	< 0.001	1.09
Interaction: 12+ Drinks x Age	-0.138	(-0.23, -0.045)	0.004	19.30
Interaction: Drinks in 1 Day x Age	0.023	(0.013, 0.034)	< 0.001	6.56
Interaction: Drinks in 1 Year x Age	0.000	(0, 0)	0.06	14.30

[1] R^2: 0.1927, Adjusted R^2: 0.1912

[2] Reference Groups: Female, Alcohol12PlusYrNO

Table 5: Normality of Interaction Model



Since the histogram shows the models' residuals are a little bit right-skewed and the interaction of Days Drinking alcohol in 1 Year and Age has β and the critical interval of 0 with p-value larger than the significance level 0.05. We delete the covariate of Days Drinking alcohol in 1 Year, only interact 12 or more Alcoholic Drinks in 1 Year with Age and log the outcome Systolic Blood Pressure to get our final model ($SBP_i = \beta_0 + \beta_1 AD_i + \beta_2 12AD_i + \beta_3 Age_i + \beta_4 BMI_i + \beta_5 G_i + \beta_6 12AD_i * Age_i + \varepsilon_i$)

The final model incorporated log-transformation of Systolic Blood Pressure to address skewness and heteroscedasticity, and the R^2 value was 0.186 which is around the 0.2 threshold, indicating that the results reliably reflect the influence of average alcoholic drinks in 1 day on systolic blood pressure. Average alcoholic drinks in 1 day were positively associated with the log-transformed Systolic Blood Pressure (β = 0.003, 95% CI: 0.001 to 0.004, $p < 0.001$), and other covariates also remained significant predictors in the final model. The interaction term between 12 or more alcoholic drinks in 1 year and Age showed a modest negative association with log(Systolic Blood Pressure) (β = -0.001, 95% CI: -0.002 to 0, $p = 0.005$), indicating that the effect of heavy drinking on systolic blood pressure diminishes with increasing age (Table 6).

Multicollinearity diagnostics revealed VIF values below 2 for most predictors, except for the interaction term and age, indicating potential multicollinearity since we did the interaction of the model. Residual diagnostics showed improved normality and stabilized variance (Table 7), confirming the appropriateness of the log-transformation.

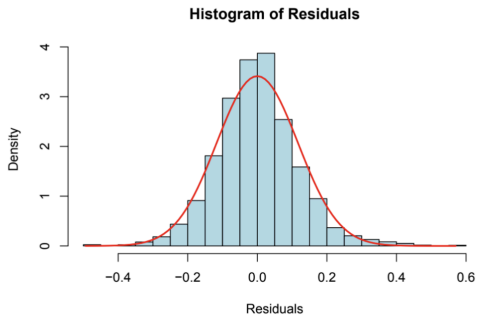
Table 6: Final Model Results

Final Model Results
Linear Regression: log(BPSysAve)[1]

Characteristic[2]	Beta	95% CI	p-value	VIF
Average Alcoholic Drinks in 1 Day	0.003	(0.001, 0.004)	< 0.001	1.15
12 or more Alcoholic Drinks in 1 Year: Yes	0.060	(0.024, 0.097)	0.001	8.59
Age	0.004	(0.003, 0.005)	< 0.001	11.33
BMI	0.003	(0.002, 0.003)	< 0.001	1.01
Male Gender	0.041	(0.034, 0.048)	< 0.001	1.07
Interaction: 12+ Drinks x Age	-0.001	(-0.002, 0)	0.005	17.98

[1] R^2: 0.186, Adjusted R^2: 0.185
[2] Reference Groups: Female, Alcohol12PlusYrNO

Table 7: Histogram of Final Model



From the outcome of our final model's results (Table 6), we can interpret that the significant effects of both average alcoholic drinks in 1 day (β1=0.003) and 12 or more alcoholic drinks in 1 year (β2 = 0.060) suggest that alcohol consumption is positively associated with log of systolic blood pressure. The interaction term (β6) indicates that younger individuals are more adversely affected by heavy drinking compared to older individuals. Other covariates such as age, BMI, and male gender are also important predictors, emphasizing the multifactorial nature of blood pressure regulation.

Correlation Analysis

According to the final model's correlation results (Table 8), we found that there is a strong positive correlation between age and the interaction term (r = 0.61) and 12 or more alcoholic drinks in 1 year and the interaction term (r = 0.73). This is expected because the interaction term directly incorporates age and 12 or more alcoholic drinks in 1 year. There is a moderate positive correlation between average daily alcohol consumption and male gender (r = 0.21), suggesting males tend to drink more on average and males are more likely to report consuming 12 or more alcoholic drinks in a year(r = 0.14). There is a moderate negative correlation between average daily alcohol consumption and age (r = -0.25), suggesting that younger individuals tend to drink more alcohol daily.

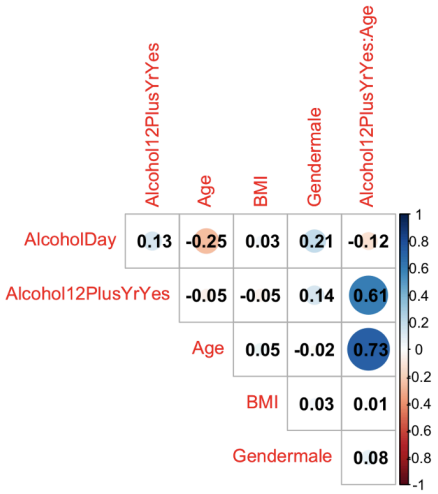


Table 8: Correlation Plot and Results

Conclusions and Discussion

In summary, we found alcohol consumption is significantly associated with increased systolic blood pressure. Using our final model, for each one drink increase per day, a subject's log-transformed systolic blood pressure increases by 0.003 (95% CI: 0.001, 0.004, p < 0.001) units, adjusting for age, BMI, gender, and 12 or more drinks in one year. Also, those who have consumed at least 12 drinks in the past year, were on average found to have a log-transformed systolic blood pressure of 0.06 (95% CI: 0.024, 0.097, p = 0.001) units higher than those who did not consume 12 drinks in the past year, adjusting for age, BMI, gender, and drinks per day. The interaction term used in our final model between age and 12 or more alcoholic drinks per year

showed a slight negative association with Systolic Blood Pressure ($\beta = -0.001$, 95% CI: -0.002 to 0, $p = 0.005$). This association indicates that as an individual gets older, the effect of alcohol consumption on log-transformed systolic blood pressure decreases.

Specific limitations of our study include the following. The NHANES dataset includes extensive missing data within our predictor and confounding variables. After removing the missing data, we compared the subjects with missing data with those included in our final sample. Notable differences occurred within the gender variable, and the 12 or more drinks per year variables. A larger proportion of subjects with missing data are female, but in our sample a larger proportion of subjects are male. Looking at the 12 or more drinks per year variable, the subjects who answered this question are extremely more likely to drink than not. In addition, the adjusted R squared for our final model is quite low 0.185. With a low adjusted R squared, it is reasonable to assume we are missing an important confounding variable, which may be responsible for explaining a larger proportion of the variation in average systolic blood pressure. Also, including interaction terms involving age introduced moderate multicollinearity, complicating precise estimation and interpretation. Additionally, the inclusion of age and BMI did not fully address residual autocorrelation in the models, suggesting that some unmeasured confounders may still influence the results. Furthermore, the binary coding of gender may oversimplify its effects on blood pressure, as hormonal and behavioral factors were not explicitly accounted for in this variable.

In summary, we found a positive association between alcohol consumption and log-transformed systolic blood pressure, in a fully adjusted model. Given the significant limitations of the study, it is not appropriate to generalize our findings to larger populations. Moving forward, alternative approaches could improve the handling of covariates in similar analyses. Multi-level coding or additional interaction terms for gender could provide a more nuanced understanding of its role. Expanding the covariates to include lifestyle factors such as smoking, physical activity, and family history could address residual confounding and provide a more comprehensive model. To address multicollinearity in interaction terms, dimensionality reduction techniques such as Principal Component Analysis (PCA) may also be considered. These adjustments would strengthen future analyses and provide deeper insights into the complex factors influencing blood pressure.

References

1. Gillman, M. W., Cook, N. R., Evans, D. A., Rosner, B., & Hennekens, C. H. (1995). Relationship of Alcohol Intake With Blood Pressure in Young Adults. *AHAIASA Journals*, *25*(5). https://doi.org/https://doi.org/10.1161/01.HYP.25.5.1106
2. Marmot, M. G., Elliott, P., Shipley, M. J., Dyer, A. R., Ueshima, H. U., Beevers, D. G., Stamler, R., Kesteloot, H., Rose, G., & Stamler, J. (1994). Alcohol and Blood Pressure: The Intersalt Study. *BMJ*, *308*(6939), 1263–1267. https://doi.org/10.1136/bmj.308.6939.1263
3. Mills, K. T., Stefanescu, A., & He, J. (2020). The Global Epidemiology of Hypertension. *Nature Reviews Nephrology*, *16*(4), 223–237. https://doi.org/10.1038/s41581-019-0244-2
4. Santana, N. M., Mill, J. G., Velasquez-Melendez, G., Moreira, A. D., Barreto, S. M., Viana, M. C., & Molina, M. del. (2018). Consumption of alcohol and blood pressure: Results of the elsa-brasil study. *PLOS ONE*, *13*(1). https://doi.org/10.1371/journal.pone.0190239
5. Stranges, S., Wu, T., Dorn, J. M., Freudenheim, J. L., Muti, P., Farinaro, E., Russell, M., Nochajski, T. H., & Trevisan, M. (2004). Relationship of alcohol drinking pattern to risk of hypertension. *Hypertension*, *44*(6), 813–819. https://doi.org/10.1161/01.hyp.0000146537.03103.f2

Group Contributions

Sara Elfring - Data Cleaning, Study Population section, Abstract
Jingyi Chen - Regression Models and LINE assumption in R, a part of statistical analyses and results in the report
Shuoyuan Gao - Proposal Presentation, Data Influential point check, Original model LINE assumption, Final model, Predictor(s) of Interest
Alexandra Schmalzel - Literature Review, Introduction, Conclusions and Discussion, Proposal Presentation
Haowen Wu - Outcome, covariates and a part of statistical analyses and results in the report