Associations Between Hypertension and Demographic Factors

Musembi Nzau: 7877498
Samantha Sénécal: 300214936
Phoenix Bastien: 300199469

Table of contents

Introduction

Hypertension, often symptomless and ominously referred to as the 'silent killer', affects millions of people worldwide, posing a significant public health challenge(1). This common condition, characterized by high blood pressure, can lead to serious health complications, including heart attack/failure, stroke, kidney disease/failure, vision loss, sexual dysfunction, cognitive decline, and reduced life expectancy(1). With its widespread prevalence, understanding the factors contributing to hypertension becomes essential for effective prevention and management(1).

Objectives

This study aims to shed light on the intricate relationships between hypertension and key demographic factors, including gender, age, cholesterol levels, and BMI. By delving into these associations, we aim to provide valuable insights that can inform preventive strategies and personalized healthcare approaches.

We hypothesize that hypertension is conditionally dependent on a person's gender, age, total cholesterol and body mass index (BMI). In the subsequent sections, we will present the data, outline the methods employed in this study, and thoroughly analyze and discuss our findings.

Data

To investigate these relationships, we will utilize a data sample obtained from the National Health and Nutrition Examination Survey (NHANES) 2009-2010 cycle. This program, dating back to the early 1960s, has played a crucial role in assessing the health and nutritional status of a diverse sample of 5,000 people across the United States each year. Combining interviews and physical examinations, NHANES provides a robust dataset for epidemiological studies, health sciences research, and the development of public health policies(2).

Data exploration

We built a data dictionary of the nhanes dataset in order to highlight the different variables that caught our interest:

Table 1: Data dictionary for the nhanes dataset

Variable name	Variable description	Variable type	Accepted values
gender	Gender of participant	Nominal (binary)	Female, male
age	Participant age	numerical	16-80
Marstat	Marital status	nominal	Married, widowed, separated, never married, living together, divorced

tchol	Amount Total cholesterol mg/dL	Numerical (discrete)	90-383
hdl	High density lipoprotein mg/dL	Numerical (discrete)	11-144
sysbp	Systolic blood pressure mm Hg	Numerical (discrete)	90-220
dbp	Diastolic blood pressure mm Hg	numerical (discrete)	40-134
bmi	Body mass index	numerical (continuous)	13.18-65.19
vigwrk	Vigorous work activity	Nominal (binary)	No, yes
modwrk	Moderate work activity	Nominal (binary)	No, yes
vigrecexr	Vigorous recreational exercise	Nominal (binary)	No, yes
modrecexr	Moderate recreational exercise	Nominal (binary)	No, yes

Since we will be employing categorical data analysis, we used clinically established thresholds to categorize the total cholesterol (“tchol”), high-density lipoprotein (“hdl”), systolic blood pressure (“sysbp”) and diastolic blood pressure (“dbp”). We used the blood pressure classification from the National library of Medicine (3) which is:

- Normal blood pressure is defined as a systolic blood pressure of less than 120 mmHg and a diastolic blood pressure of less than 80 mmHg.
- Elevated blood pressure is defined as a systolic blood pressure between 120 and 129 mmHg and a diastolic blood pressure of less than 80 mmHg.
- Stage 1 hypertension is defined as a systolic blood pressure between 130 and 139 mmHg or a diastolic blood pressure between 80 and 89 mmHg.
- Stage 2 hypertension is defined as a systolic blood pressure greater or equal to 140 mmHg and/or a diastolic blood pressure greater or equal to 90 mmHg.

We also further categorized blood pressure by making a binary hypertension variable where people with hypertension stage 1 or stage 2 were classified as having hypertension, and people with normal or elevated blood pressure were categorized as having no hypertension.

For the total cholesterol and high-density lipoprotein classifications, we used the classification from the National library of Medicine (4) which is:

- A total cholesterol level less than 200 mg/dL is desirable.
- A total cholesterol level less than between 200 and 239 mg/dL is borderline high.
- A total cholesterol level greater or equal to 240 mg/dL is high.
- A HDL cholesterol level greater or equal to 60 mg/dL is high.
- A HDL cholesterol level between 40 and 59 mg/dL is medium.
- A HDL cholesterol level less than 40 mg/dL is low.

In addition, we decided to group the ages (Adolescent: 16-17 years old, Adult: 18-64 years old and Older adult: 65+ years old), to get a sense for the sample's demographic. We also decided to merge the vigorous recreational exercise and moderate recreational exercise variables ("vigrecexc" and "modrecexc") into a single category called recreational exercise which has values vigorous/moderate, vigorous, moderate or none. The vigorous work activity and moderate work activity variables ("vigwrk" and "modwrk") were transformed identically. Finally, we decided to make our weight categories using the classification from the National library of Medicine (5) which is:

- Underweight if the body mass index is less than 18.5.
- Normal weight if the body mass index is between 18.5 and 24.9.
- Overweight if the body mass index is between 25 and 29.9.
- Obesity Class 1 if the body mass index is between 30 and 34.9.
- Obesity Class 2 if the body mass index is between 35 and 39.9.
- Extreme Obesity if the body mass index is greater than 40.

Table 2: Data dictionary for our derived dataset

Variable name	Variable description	Variable type	Accepted values
gender	Gender of participant	Nominal (binary)	Female, male
age_category	Age group of participant	Nominal	Adolescent, adult, older adult
age	Participants age	numerical	16-80
marstat	Marital status	Nominal	Married, widowed, separated, never married, living together, divorced
tchol_category	Level of total cholesterol	Nominal	Desirable, borderline high, high
tchol	Amount Total cholesterol mg/dL	Numerical (discrete)	90-383
hdl_category	Level of high-density lipoprotein (good cholesterol)	Nominal	Low, medium, high
bp_category	Level of blood pressure	Nominal	Normal, elevated, high (hypertension s1), high (hypertension s2)
hypertension	If respondent has hypertension	Nominal (binary)	Yes, no
BMI_category	Body mass index	Nominal	Underweight, Normal weight, Overweight, Obesity (class 1), Obesity (class 2), Extreme obesity
bmi	Body mass index	numerical (continuous)	13.18-65.19

Work_activity	Category of work activity	Nominal	None, moderate, vigorous/moderate, vigorous
Recreational_activity	Category of recreational exercise	Nominal	None, moderate, vigorous/moderate, vigorous

The original dataset had missing data, and upon closer examination, we found that the majority of the missing data was related to the marital status question for individuals aged 16-19. To address these missing values, we opted to designate them as 'never married' as it appeared to be the most reasonable solution. As for the remaining missing values in the dataset, we chose to exclude them from our analytical dataset since their quantity was minimal.

Table 3: Sample demographic characteristics stratified by hypertension status

Sample characteristic	Total (n= 5417)	Hypertension (n= 2159) 40%	No hypertension (n= 3258) 60%
Gender			
Male	2703 (50%)	1195 (55%)	1508 (46%)
Female	2714 (50%)	964 (45%)	1750 (54%)
Age			
Adolescent	270 (5%)	23 (1%)	247 (8%)
Adult	4070 (75%)	1445 (67%)	2625 (80%)
Older adult	1077 (20%)	691 (32%)	386 (12%)
Marstat			
Married	2569 (47%)	1157 (54%)	1412 (43%)
Widowed	399 (7%)	252 (12%)	147 (5%)
Divorced	537 (10%)	259 (12%)	278 (9%)
Seperated	160 (3%)	60 (3%)	100 (3%)
Never married	1343 (25%)	289 (13%)	1054 (32%)
Living together	409 (8%)	142 (6%)	267 (8%)
Tchol			
Desirable	3243 (60%)	1109 (51%)	2134 (65%)
Borderline high	1509 (28%)	709 (33%)	800 (25%)
High	665 (12%)	341 (16%)	324 (10%)
Hdl			
Low	1136 (21%)	454 (21%)	682 (21%)
Medium	2745 (51%)	1100 (51%)	1645 (50%)
High	1536 (28%)	605 (28%)	931 (29%)
BMI			
Underweight	100 (2%)	18 (1%)	82 (2%)
Normal weight	1586 (29%)	477 (22%)	1109 (34%)
Overweight	1813 (34%)	752 (35%)	1061 (33%)
Obesity C1	1107 (20%)	518 (24%)	589 (18%)
Obesity C2	497 (9%)	235 (11%)	262 (8%)
Extreme obesity	314 (6%)	159 (7%)	155 (5%)

Work activity			
Vigorous/moderate	756 (14%)	311 (14%)	445 (14%)
Vigorous	235 (4%)	98 (5%)	137 (4%)
Moderate	1144 (21%)	438 (20%)	706 (22%)
None	3282 (61%)	1312 (61%)	1970 (60%)
Recreational activity			
Vigorous/moderate	714 (13%)	196 (9%)	518 (16%)
Vigorous	475 (9%)	133 (6%)	342 (10%)
Moderate	1408 (26%)	592 (28%)	816 (25%)
None	2820 (52%)	1238 (57%)	1582 (49%)

Table 3 displays our sample's demographic characteristics, which we further segmented by hypertension status to gain deeper insights into our population. It indicates that our sample was evenly distributed between males and females. The majority of individuals fell within the adult age group (75%) and were married (47%). Most exhibited desirable cholesterol levels (60%), medium HDL levels (51%), and had normal blood pressure (44%). Additionally, a significant portion of the sample was not involved in either work activities (61%) or recreational activities (52%).

We have made barplots to visualize the different variables of interest by hypertension status.

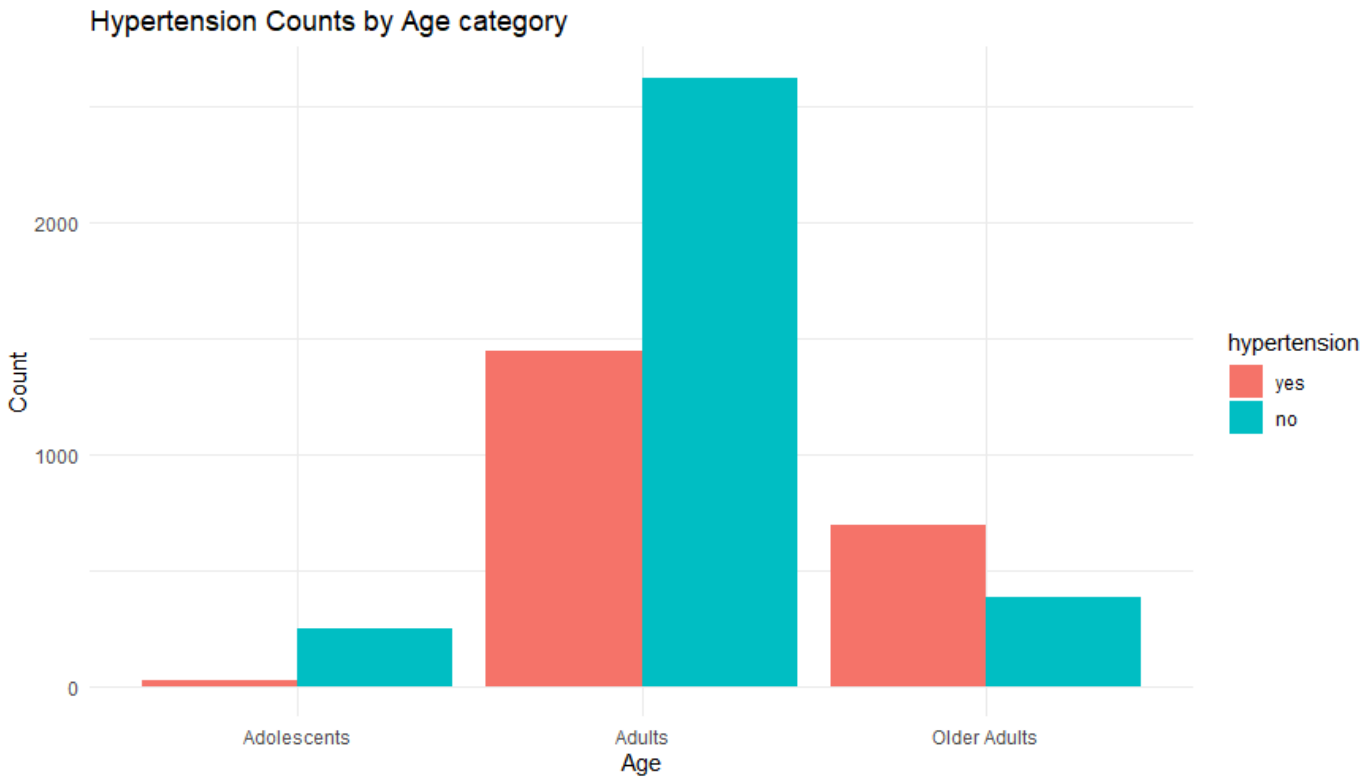

Fig 1: *Hypertension counts by age category*

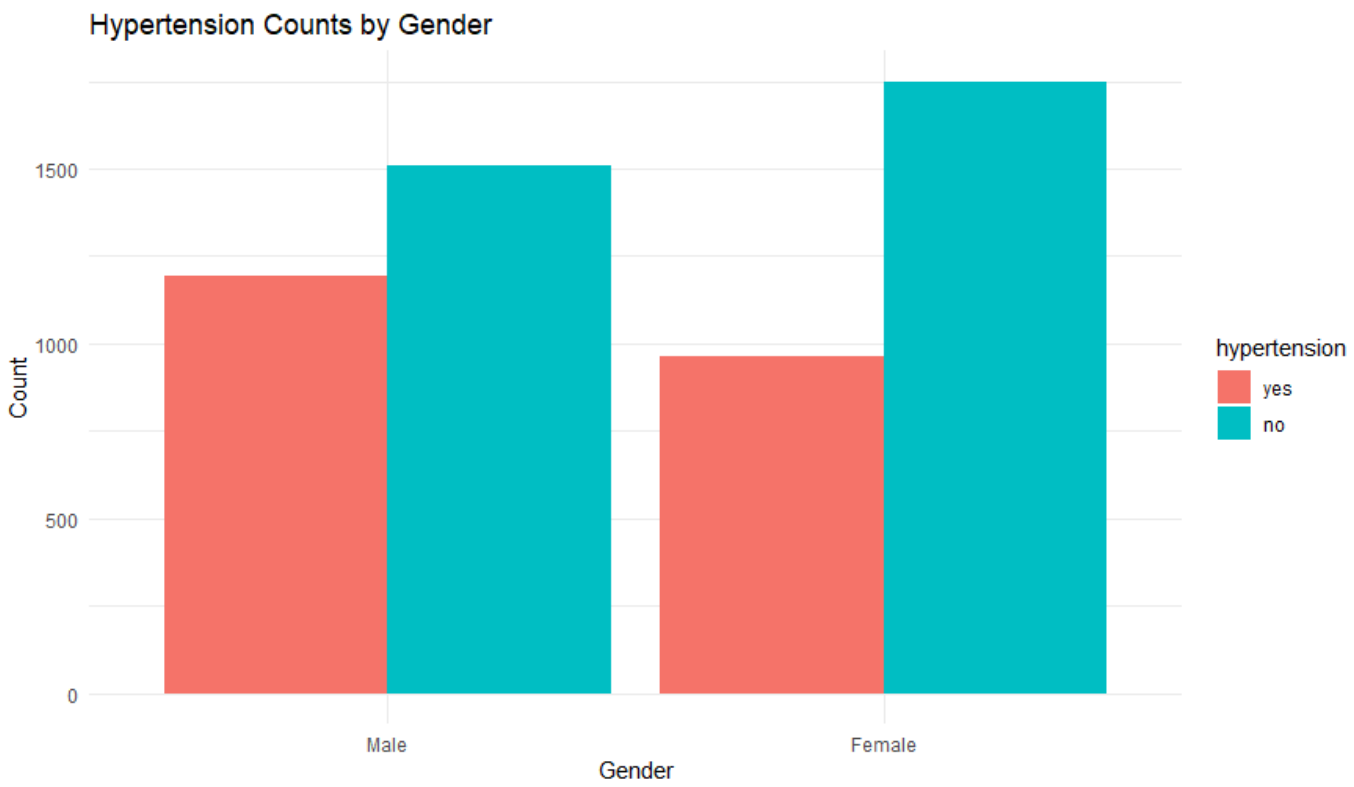

Fig 2: *Hypertension counts by gender*

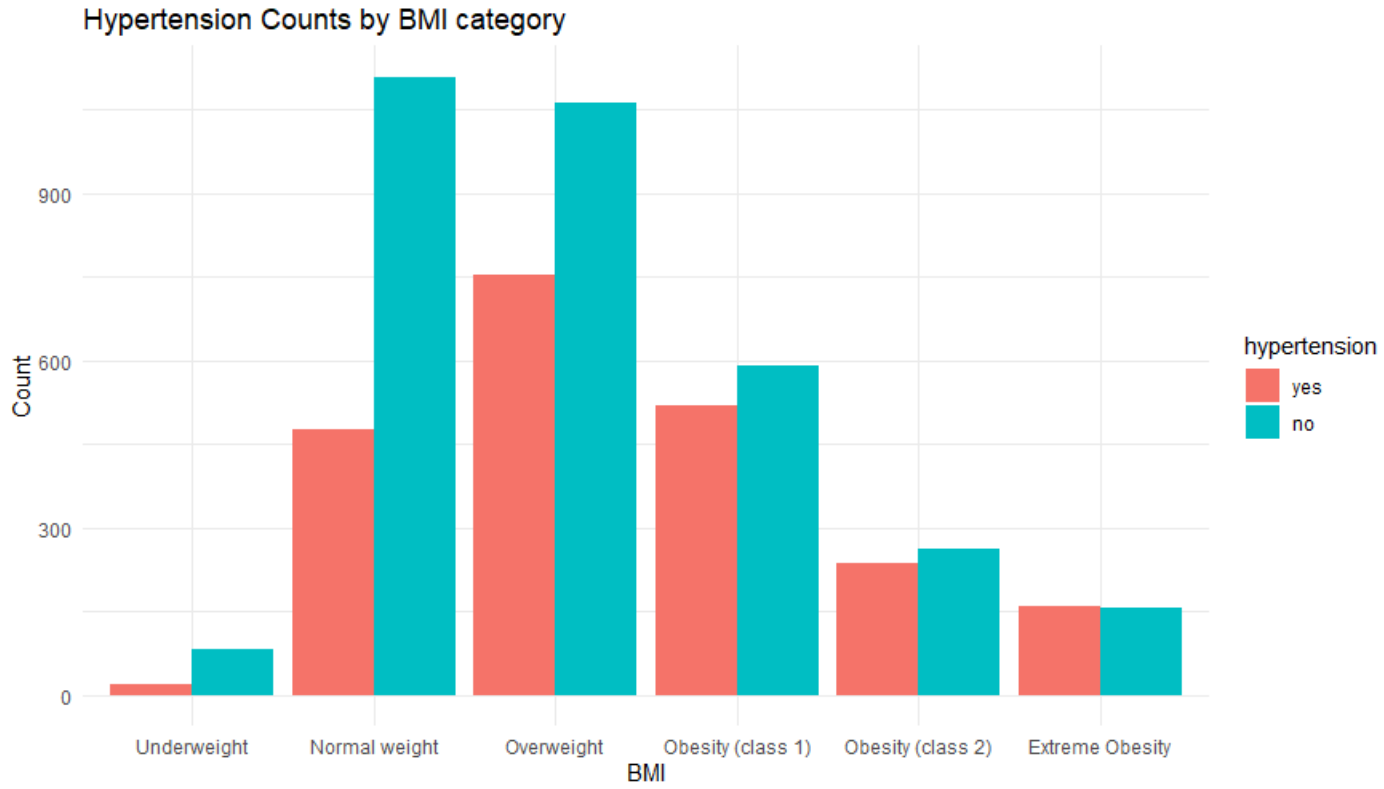

Fig 3: *Hypertension counts by BMI category*

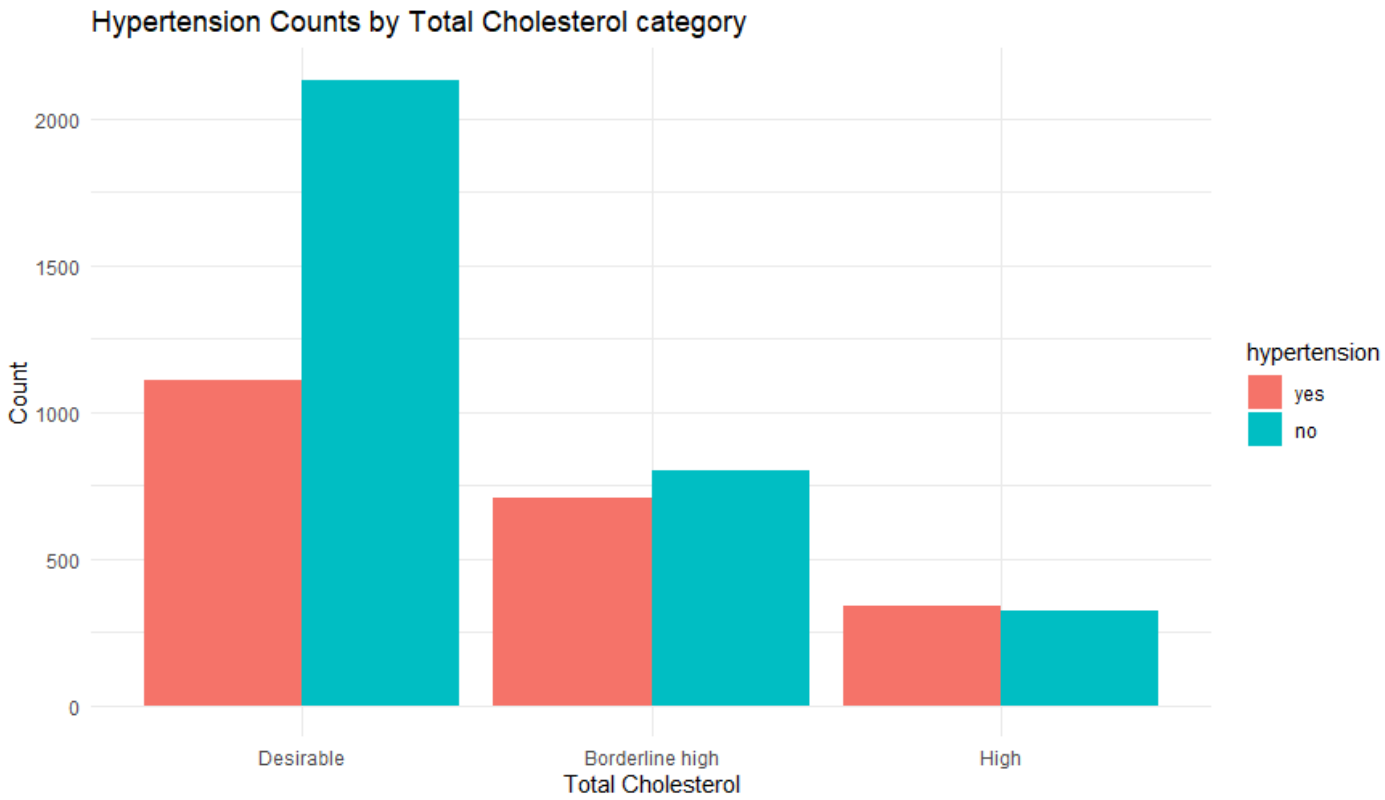

Fig 4: *Hypertension counts by total cholesterol category*

From Figure 1, we can see that there seems to be a relationship between hypertension and age category. The majority of adolescents and adults have no hypertension and the majority of older adults have hypertension. From Figure 2, we can see that hypertension seems to be more likely in males than females. Figure 3 shows a relationship between hypertension and BMI category, hypertension seems to be more common in people with a higher BMI and the majority of extremely obese people have hypertension. Finally, Figure 4 shows the relationship between hypertension and total cholesterol. There seems to be an association again, since the higher the level of cholesterol, the higher the counts for hypertension.

Table 4: *Abbreviations for relevant variables*

Variable Name	Abbreviation
Hypertension	Y
Gender	G
Age	A
BMI	B
Tchol	T

Table 4 consists of each variable their respective abbreviation. This was done to make formulas more concise and legible. Note: Since H is already a symbol used for hypotheses, we use Y as an abbreviation instead of H.

Methods

Logistic regression

To comprehensively examine the relationship between hypertension and demographic factors, including age, gender, total cholesterol, and BMI, we will employ a logistic regression model. By fitting this model, we aim to gain valuable insights into how these variables influence the likelihood of hypertension (identify which predictors (gender, age, tchol, BMI) are statistically significant in predicting the probability of hypertension). Specifically, we anticipate uncovering the magnitude and direction of the associations, allowing us to identify key contributors to hypertension risk.

In constructing our logistic regression model, we initiated the process by fitting the null model (designated as Model 26 in Table 5). We then systematically incorporated individual variables, one by one, along with interaction terms, progressively refining the model (checking the difference in deviance from 1 model to the next) until we identified Model 7 as the most statistically significant. Subsequently, we extended the model by introducing additional interaction terms. However, upon closer examination, we observed that none of the supplementary terms demonstrated sufficient significance (likelihood ratio test between model 7 and models 1-6 were not significant) to justify their retention in the final model.

***Table 5**: Table of deviance and degrees of freedom for various models*

	formulas	df	deviance
1	$Y \sim G + A + B + T + G * A + A * B + G * A * B * T$	5401	6175.7
2	$Y \sim G + A + B + T + G * A + A * B + A * B * T$	5407	6181.0
3	$Y \sim G + A + B + T + G * A + A * B + G * A * T$	5407	6177.0
4	$Y \sim G + A + B + T + G * A + A * B + G * A * B$	5408	6181.1
5	$Y \sim G + A + B + T + G * A + A * B + B * T$	5409	6181.1
6	$Y \sim G + A + B + T + G * A + A * B + A * T$	5409	6182.0
7	$Y \sim G + A + B + T + G * A + A * B$	5410	6182.1
8	$Y \sim G + A + B + T + G * A + G * T$	5410	6213.2
9	$Y \sim G + A + B + T + G * A + G * B$	5410	6213.2
10	$Y \sim G + A + B + T + G * A$	5411	6213.2
11	$Y \sim G + A + B + T$	5412	6263.1

12	$Y \sim A + B + T$	5413	6323.4
13	$Y \sim G + B + T$	5413	6999.2
14	$Y \sim G + A + T$	5413	6328.2
15	$Y \sim G + A + B$	5413	6316.3
16	$Y \sim B + T$	5414	7057.5
17	$Y \sim A + T$	5414	6380.8
18	$Y \sim A + B$	5414	6371.1
19	$Y \sim G + T$	5414	7100.4
20	$Y \sim G + B$	5414	7125.7
21	$Y \sim G + A$	5414	6383.6
22	$Y \sim T$	5415	7150.3
23	$Y \sim B$	5415	7177.3
24	$Y \sim A$	5415	6431
25	$Y \sim G$	5415	7242.3
26	$Y \sim 1$	5416	7285

Let π be the probability that an individual with a specific set of covariate values has hypertension. We fit the following logistic regression model to our data:

$$logit(\pi) = \beta_0 + \beta_1 Gender + \beta_2 Age + \beta_3 BMI + \beta_4 Tchol$$

where we used Age, BMI and Tchol as continuous and with β_0 being the Gender = Male.

We conducted a likelihood ratio test, which assesses the hypothesis of all slope parameters being 0 (no association) as well as a Wald test to examine the presence of an association between the explanatory and response variables. Also, we calculated the odds ratios for hypertension given each explanatory variable to see how these variables affect the probability of developing hypertension (positively or negatively).

Log-linear regression

Log-linear regression models assess the number of occurrences/counts observed for groups of levels of categorical variables. As such, log-linear regression is based on the poisson distribution, where the response variables are logarithmic means of counts per group.

$$ln(\mu_{ijklm}) = hypertensionn_i + Gender_j + Age_k + BMI_l + Tchol_m$$

Such that i,j,k,l and m correspond to each level of their respective categorical variable.

Hypothesis testing

We constructed an alternate hypothesis H_1: there exists a relationship between hypertension (or the lack of) and gender, age, BMI & tchol vs null hypothesis H_0: the status of hypertension is independent of gender, age, BMI and tchol. Afterwards, we conducted an ANOVA test between the following independent model H_0 and saturated model H_1 :

$$H_0 = hypertensionn_i + Gender_j + Age_k + BMI_l + Tchol_m$$

$$H_1 = Y_i + G_j + A_k + B_l + T_m + Y_iG_j + Y_iA_k + Y_iB_l + Y_iT_m$$

We further hypothesized that hypertension is conditionally dependent on each other explanatory variable. For testing conditional independence between hypertension and all other explanatory variables respectively, we conducted an odds ratio test of having hypertension conditioned on our other explanatory variables using the fitted values of our log-linear model. The compared variables are conditionally independent if their odds ratio (OR) equals (or is close to) 1, otherwise they are conditionally dependent. Due to the rounding errors caused by computer imprecision , we chose all values <0.9 and >1.1 to be conditionally dependent.

Logistic regression

The overall model exhibits high significance. The likelihood ratio test yields a statistic of 1103 with 6 degrees of freedom and p-value significantly less than 0.001. Also, we can see from Table 6 that the explanatory variables are all significantly associated with hypertension (p-value <0.001).

***Table 6**: Estimated regression parameters, their standard errors and Wald statistics and P-values*

	Estimate	Std. Error	z value	Pr(>|z|)
genderFemale	1.833	0.193	9.478	<0.001
age	-0.077	0.008	-9.352	<0.001
bmi	-0.115	0.014	-8.029	<0.001
tchol	-0.004	0.001	-5.932	<0.001
genderFemale:age	-0.027	0.004	-7.482	<0.001
age:bmi	0.002	0.000	5.587	<0.001

An interesting observation from Table 6 is that older individuals are less likely to develop hypertension. Surprisingly, individuals with higher BMI and higher total cholesterol are also less likely to develop hypertension. Additionally, females have higher odds of hypertension compared to males. The effect of age on hypertension is less pronounced for females, and the effect of BMI on hypertension increases with age.

***Table 7**: Estimated odds ratios and 95% confidence intervals for developing hypertension*

	Odds ratio	Confidence Interval
genderFemale	6.251	(4.279, 9.131)
age	0.926	(0.911, 0.941)
bmi	0.891	(0.866, 0.917)
tchol	0.996	(0.994, 0.997)
genderFemale:age	0.973	(0.966, 0.980)
age:bmi	1.002	(1.001, 1.002)

Table 7 reveals a substantial gender disparity, indicating that females have more than six times the odds of developing hypertension compared to males. Additionally, the likelihood of hypertension diminishes by 7.4% with each unit increase in age, by 10.9% with each unit increase in BMI, and by 0.4% with each unit increase in total cholesterol. Furthermore, interaction terms highlight a 2.7% reduction in odds for females with each unit increase in age and a 0.2% increase in odds with each unit increase in both age and BMI.

Log-linear regression

Table 8: *regression coefficients for parameters*

coefficient	value	Std error
intercept	-2.91	0.32
Hypertension: no	3.54	0.34
Gender: female	-0.21	0.05
BMI category:Normal weight	3.28	0.24
BMI category:overweight	3.73	0.24
BMI category:Obesity (class 1)	3.36	0.24
BMI category:Obesity (class 2)	2.57	0.24
BMI category:Extreme obesity	2.18	0.24
Tchol category:Borderline high	-0.45	0.05
Tchol category:High	-1.18	0.06
Age group:adults	4.14	0.21
Age group:Older adults	3.40	0.21
Y(no):B(Normal weight)	-0.67	0.27
Y(no):B(Overweight)	-1.17	0.26
Y(no):B(Obesity class 1)	-1.39	0.27
Y(no):B(Obesity class 2)	-1.41	0.28
Y(no):B(Extreme obesity)	-1.54	0.28
Y(no):G(female)	-0.36	0.06
Y(no):T(Borderline high)	-0.53	0.06
Y(no):T(High)	-0.71	0.09
Y(no):A(adults)	-1.78	0.22
Y(no):A(Older adults)	-2.96	0.23

Table 9: *levels of log-linear models*

Intercept	Highest count model
Hypertension: yes	Hypertension: no
Gender: male	Gender: female
Age group: adolescents	Age group: adults
BMI category: underweight	BMI category: normal weight
Tchol: desirable	Tchol: desirable
Expected count: 1	Expected count: 315
Actual count: 2	Actual count: 351

From table 8, which displays all the regression coefficients of our log-linear model, we can derive the reference model (intercept) and the highest expected count model shown in table 9. Since the explanatory variable is the logarithmic mean of the respective count, we summed the largest levels of each explanatory variable and used it as an exponent for e.

From table 9 we can see that the reference model(intercept) is a group of underweight adolescent males with desirable total cholesterol suffering from hypertension, whereas the model for highest count consists of

Normal weight adult females with desirable total cholesterol and no hypertension. The expected count for the reference model was 1, whereas the actual count was 2. This estimation is fairly good although technically, the expected model estimated 50% less counts. However, the aforementioned value is misleading due to the fact that both the expected and actual counts are sparse.

The highest expected count was 315, while the actual count was 351, boasting an accuracy of 91%. Surprisingly, it turns out that the group predicted to have the highest count was also the group that had the highest actual count.

Hypothesis testing

Since the p-value from our ANOVA is <0.05, we reject the null hypothesis and conclude that there exists a relationship between hypertension and gender, age, BMI & tchol.

For the test of conditional independence, we found that hypertension is conditionally dependent on all levels of response variables aside from ($\frac{obesity(class\ 1)}{obesity(class\ 2)}$) and naturally ($\frac{obesity(class2\ )}{obesity(class\ 1)}$) (for information on specific OR values see appendix tables A2-A4). The largest odds ratio for BMI was between extreme obese and underweight people. Extremely obese people were 4.71 times more likely to suffer from hypertension compared to underweight people. For total cholesterol, the largest odds ratio was between people with high tchol, who were twice as likely to have hypertension compared to people with desirable levels. When it comes to age, the largest odds ratio was 19.23 for older adults compared to under age people. Lastly, Males are 1.44 times more likely to have hypertension compared to females.

Summary / Discussion

We used the NHANES dataset to explore the relationship between hypertension and various health metrics within the US population. To do so, we transformed various variables from the NHANES dataset and created our own dataset. We were specifically interested in the

relationship between hypertension and age, body mass index, total cholesterol and gender. We hypothesized that there exists a meaningful association linking these variables and hypertension.

To test our hypothesis, we used logistic regression and log-linear regression. The likelihood ratio test conducted in the logistic regression results demonstrated a significant association between hypertension and age, gender, total cholesterol and age. The presence of these four variables made for a significant model but the inclusion of interactions further enhanced model significance. For instance, including an interaction between gender and age as well as an interaction between age and bmi, in the model, significantly increased its fit. These interactions suggest that there is a joint effect of one variable's impact by the presence of the other.

After our hypothesis testing, we conducted odds ratios for our different explanatory variables from the logistic regression model and found that the odds for developing hypertension in females is significantly higher than the odds in males. Also the odds of developing hypertension decrease by increase in age, bmi and tchol. Finally, the odds for females decrease slightly by unit increase of age and the odds of hypertension decrease as bmi and age increase.

Afterwards we created the following log-linear regression models for our null and alternate hypothesis:

$$H_0 = hypertension_i + Gender_j + Age_k + BMI_l + Tchol_m$$

$$H_1 = Y_i + G_j + A_k + B_l + T_m + Y_iG_j + Y_iA_k + Y_iB_l + Y_iT_m$$

The intercept of our log-linear regression model had an expected count of 2, while the true count was 1, whereas the group with the highest estimated count was 315, whereas the true count for that group was 351. However the group with the highest estimated count was also the group with the actual highest count.

By conducting an ANOVA analysis, we obtained a p-value <0.05, and concluded that there exists a relationship between hypertension and the other variables. To test for conditional independence, we calculated the odds ratios of having hypertension between levels within each response variable. If the odds ratio is 1 or "near" one, then hypertension is conditionally independent from the other response variables. We defined value n to be "near" 1 if 0.90<n<1.10 and found that the only near value was the odds ratio between obesity(class1) and obesity (class 2) and vice versa. We concluded that the hypothesis is conditionally dependent on gender, age, BMI and total cholesterol.

Although the regression models are very significant, it is strange that developing hypertension is less likely the higher their BMI is. This contradicts our prior knowledge on hypertension since we have seen that there is an association between developing hypertension and bmi (see assignment 1). This can be explained by a possible nonlinear relationship between predictors. This would mean that the assumptions for logistic regression (linear relationships between predictors and log-odds of the outcomes) are not

met and the model may not capture these relationships effectively, prompting a need for further investigation and potential consideration of alternative modeling techniques.

References

1. American Heart Association. “What Is High Blood Pressure?” *Www.heart.org*, 25 May 2023,www.heart.org/en/health-topics/high-blood-pressure/the-facts-about-high-blood-pressure/what-is-high-blood-pressure.

2. Centers for Disease Control and Prevention. “NHANES - about the National Health and Nutrition Examination Survey.” Centers for Disease Control and Prevention, 2019, www.cdc.gov/nchs/nhanes/about_nhanes.htm.

3. Blake JM, Evans CV, Webber EM, et al. Screening for Hypertension in Adults: An Updated Systematic Evidence Review for the U.S. Preventive Services Task Force [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2021 Apr. (Evidence Synthesis, No. 197.) Table 1, Blood Pressure Classifications. Available from: https://www.ncbi.nlm.nih.gov/books/NBK570233/table/ch1.tab1/

4. Lee Y, Siddiqui WJ. Cholesterol Levels. [Updated 2023 Jul 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK542294/

5. Dwyer JT, Melanson KJ, Sriprachy-anunt U, et al. Dietary Treatment of Obesity. [Updated 2015 Feb 28]. In: Feingold KR, Anawalt B, Blackman MR, et al., editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, Inc.; 2000-. Table 4, Classification of Weight Status by Body Mass Index (BMI) Available from: https://www.ncbi.nlm.nih.gov/books/NBK278991/table/diet-treatment-obes.table4clas/

Appendix

Table A1: Summary for hypertension by gender and sample population

Weight	hypertension	No hypertension

	male	female	total	male	female	total
Underweight	10	8	18	21	61	82
Normal weight	258	219	477	504	605	1109
Overweight	460	292	752	563	498	1061
Obesity (class 1)	303	215	518	276	313	589
Obesity (class 2)	107	128	235	101	161	262
Extreme obesity	57	102	159	43	112	155

Table A2: Conditional *odds ratios of Tchol categories conditioned on hypertension*

Odds ratios OR($\frac{column}{row}$)	Underweight	Normal weight	Overweight	Obesity (class 1)	Obesity (class 2)	Extreme obesity
underweight	1.00	1.97	3.25	4.04	4.12	4.71
Normal weight	0.51	1.00	1.65	2.05	2.09	2.39
Overweight	0.31	0.61	1.00	1.24	1.27	1.45
Obesity (class 1)	0.25	0.49	0.81	1.00	1.02	1.17
Obesity (class 2)	0.24	0.48	0.79	0.98	1.00	1.14
Extreme obesity	0.21	0.42	0.69	0.86	0.88	1.00

***Table A3*:** Conditional *odds ratios of BMI categories conditioned on hypertension*

Odds ratios OR($\frac{column}{row}$)	desirable	Borderline high	high
desirable	1.00	1.69	2.00
Borderline high	0.59	1.00	1.18
high	0.50	0.85	1.00

***Table A4*:** Conditional *odds ratios of having hypertension between age group*

Odds ratios OR($\frac{column}{row}$)	Under age	adult	Older adult

Under age	1	5.91	19.23
adult	0.17	1	3.25
Older adult	0.05	0.31	1

CODE

```
library(aplore3)
library(dplyr)
library(qwraps2)
library(vtable)
library(epitools)
library(ggplot2)
library(reshape)
library(gridExtra)
library(MASS)

#Building the dataset
################################################################################
#######
data <- nhanes

# Update "marstat" based on age
data$marstat <- ifelse(data$age >= 16 & data$age <= 19,
                       5, as.numeric(data$marstat))
data$marstat <- factor(data$marstat,
                       levels = 1:6,
                       labels = c("Married", "Widowed", "Divorced", "Separated",
                                  "Never Married", "Living Together"))

data <- data %>% na.omit(data)

# tchol category
tchol_cutoffs <- c(0, 200, 240, max(data$tchol) + 1)
tchol_categories <- c("Desirable", "Borderline high", "High")

# add category as new column in data
data <- data %>%
  mutate(tchol_category = cut(tchol, breaks = tchol_cutoffs,
                              labels = tchol_categories, right = FALSE))

# hdl category
hdl_cutoffs <- c(0, 40, 60, max(data$hdl) + 1)
hdl_categories <- c("Low", "Medium", "High")

# add category as new column in data
data <- data %>%
```

```
  mutate(hdl_category = cut(hdl, breaks = hdl_cutoffs,
                            labels = hdl_categories, right = FALSE))

# bp category and add category as new column in data
data <- data %>%
  mutate(bp_category = case_when(
    sysbp < 120 & dbp < 80 ~ "Normal",
    sysbp >= 120 & sysbp <= 129 & dbp < 80 ~ "Elevated",
    (sysbp >= 130 & sysbp <= 139) |
      (dbp >= 80 & dbp <= 89) ~ "High (hypertension S1)",
    sysbp >= 140 ~ "High (hypertension S2)",
    dbp >= 90 ~ "High (hypertension S2)"
  ))

# bmi category and add category as new column in data
data <- data %>%
  mutate(bmi_category = case_when(
    bmi < 18.5 ~ "Underweight",
    bmi >= 18.5 & bmi < 25 ~ "Normal weight",
    bmi >= 25 & bmi < 30 ~ "Overweight",
    bmi >= 30 & bmi < 35 ~ "Obesity (class 1)",
    bmi >= 35 & bmi < 40 ~ "Obesity (class 2)",
    bmi >= 40 ~ "Extreme Obesity"
  ))

# Convert bp_category and bmi_category to factors
data$bp_category <- factor(data$bp_category,
                           levels = c("Normal", "Elevated",
                                      "High (hypertension S1)",
                                      "High (hypertension S2)"))

data$bmi_category <- factor(data$bmi_category,
                            levels = c("Underweight", "Normal weight",
                                       "Overweight", "Obesity (class 1)",
                                       "Obesity (class 2)", "Extreme Obesity"))


# new column "age_category" based on age ranges
data$age_category <- cut(data$age, breaks = c(13, 17, 65, max(data$age) + 1),
                         labels = c("Adolescents", "Adults", "Older Adults"))

# new column "recreational_activity"
data$recreational_activity <-
  ifelse(data$vigrecexr == "Yes" & data$modrecexr == "Yes", "vigorous/moderate",
                                ifelse(data$vigrecexr == "Yes", "vigorous",
                                       ifelse(data$modrecexr == "Yes",
                                              "moderate", "none")))
data$recreational_activity <-
```

```
  factor(data$recreational_activity,
         levels = c("none", "moderate", "vigorous", "vigorous/moderate"))


# new column "work_activity"
data$work_activity <-
  ifelse(data$vigwrk == "Yes" & data$modwrk == "Yes", "vigorous/moderate",
                                ifelse(data$vigwrk == "Yes", "vigorous",
                                       ifelse(data$modwrk == "Yes",
                                              "moderate", "none")))
data$work_activity <-
  factor(data$work_activity,
         levels = c("none", "moderate", "vigorous", "vigorous/moderate"))

#creating dataset for cases of hypertension
data <- data %>%
  mutate(hypertension = case_when(
    bp_category %in% c("Normal", "Elevated") ~ "no",
    bp_category %in% c("High (hypertension S1)",
                       "High (hypertension S2)") ~ "yes",
  ))
data$hypertension <- factor(data$hypertension,
         levels = c("yes", "no"))

#logistic regression
#########################################################################
#######
anova(
  glm(hypertension~gender*age_category*tchol_category*bmi,family=binomial(),
      data=data),
  glm(hypertension~gender*age_category+tchol_category+bmi+age_category*bmi,
      family=binomial(), data=data),test="LRT")

summary(glm(hypertension~gender+age+tchol+bmi+gender*age+age*bmi,family=binomial()
,
            data=data))

model<-glm(hypertension~gender+age+bmi+tchol+gender*age+age*bmi,family=binomial(),
data=data)
mytbl=data.frame(exp(coef(model)),exp(confint.default(model)))[2:7,]
1/mytbl[,c(1,3,2)]
#########################################################################
#########################################################################
#######
#creating loglinear regression model
#######################################
modeldata <- data.frame(hypertension = data$hypertension,
                        bmi_category = data$bmi_category,
```

```
                    gender = data$gender, tchol = data$tchol_category,
                    age = data$age_category)

modeldata


vars = expand.grid(hyp = levels(modeldata$hypertension),
                   bmi = levels(modeldata$bmi_category),
                   gender = levels(modeldata$gender),
                   tchol = levels(modeldata$tchol),
                   age = levels(modeldata$age) )

vars[1,1]
nrow(data)

#finding counts for combinations of levels
counts <- c(1:216)

for (i in 1:216){
  counts[i] <- sum(data$hypertension == vars[i,1] &
                   data$bmi_category == vars[i,2] &
                   data$gender == vars[i,3] &
                   data$tchol_category == vars[i,4] &
                   data$age_category == vars[i,5])
}
cbind(vars,counts)


#dataframe for counts
dat <- data.frame(vars, counts)


logln <- glm(counts ~ hyp*bmi + hyp*gender + hyp*tchol+ hyp*age ,
             family = poisson(), data = dat)
logln2 <-  glm(counts ~ hyp + bmi + gender + tchol + age ,
               family = poisson(), data = dat)

summary(logln)

#conditional independence counts
cnd <- round(fitted(logln),2)
cnd
dat$cond <- cnd
dat


#testing for conditional independence between hypertension & variables
#library(DescTools)
```

```
x <- table(dat$hyp)

OR1 <- tapply(dat$cond,list(dat$hyp,dat$bmi),sum)
OR1


OR2 <- tapply(dat$cond,list(dat$hyp,dat$gender),sum)
OR2

OR3 <- tapply(dat$cond,list(dat$hyp,dat$tchol),sum)
OR3

OR4 <- tapply(dat$cond,list(dat$hyp,dat$age),sum)
OR4

exp(logln2$coefficients)
levels(data$bmi_category)
summary(logln)

anova(logln2, logln, test = "LRT")

summary(logln)$deviance
summary(logln)$df.residual

summary(logln)
#########################################################################
library(ggplot2)

age_plot1 <- ggplot(data, aes(x = age, fill = hypertension)) +
  geom_bar(position = "dodge") +
  labs(title = "Hypertension Counts by Age",
       x = "Age",
       y = "Count") +
  theme_minimal()


age_plot2 <- ggplot(data, aes(x = age_category, fill = hypertension)) +
  geom_bar(position = "dodge") +
  labs(title = "Hypertension Counts by Age category",
       x = "Age",
       y = "Count") +
  theme_minimal()


gender_plot <- ggplot(data, aes(x = gender, fill = hypertension)) +
  geom_bar(position = "dodge") +
  labs(title = "Hypertension Counts by Gender",
       x = "Gender",
```

```
      y = "Count") +
  theme_minimal()


bmi_plot <- ggplot(data, aes(x = bmi_category, fill = hypertension)) +
  geom_bar(position = "dodge") +
  labs(title = "Hypertension Counts by BMI category",
       x = "BMI",
       y = "Count") +
  theme_minimal()


tchol_plot <- ggplot(data, aes(x = tchol_category, fill = hypertension)) +
  geom_bar(position = "dodge") +
  labs(title = "Hypertension Counts by Total Cholesterol category",
       x = "Total Cholesterol",
       y = "Count") +
  theme_minimal()

age_plot1
age_plot2
gender_plot
bmi_plot
tchol_plot

############################################################
# to create table 3
subgrouped_data <- split(data, data$hypertension)
nrow(subgrouped_data$"no")
summary(subgrouped_data$"yes")
summary(subgrouped_data$"no")
```