Impact of BMI on Diabetes and Cardiovascular Disease Risk

Emily Chiang, Emma Mo, Shengge Cheng, Junlian Cao

I. INTRODUCTION

According to the World Health Organization (WHO), the number of people with diabetes rose from 108 million in 1980 to 422 million in 2014 (WHO, Diabetes), and cardiovascular diseases (CVDs) remain the leading cause of death globally, claiming an estimated 17.9 million lives each year (WHO, Cardiovascular disease). The rise in chronic conditions like diabetes and heart disease in the United States has become a growing concern for public health, making it crucial to understand the underlying risk factors. One important factor in both diseases is the body mass index (BMI), a key measure in categorizing individuals by weight status.

Using large-scale public health datasets, this report aims to explore the relationship between BMI and the risk of developing diabetes and heart disease. The analysis includes both descriptive and hypothesis testing methods to examine whether higher BMI levels correlate with increased disease prevalence. The research can help develop more targeted public health interventions and encourage lifestyle changes to mitigate these risks.

II. METHODS & MATERIALS

A. Data Set

We used two comprehensive datasets: the Diabetes Health Indicators Dataset and the Heart Disease Health Indicators Dataset, both available publicly on Kaggle. These datasets include critical variables such as BMI, physical activity, smoking habits, and blood pressure, providing a strong foundation to investigate the relationship between BMI and chronic diseases like diabetes and cardiovascular disease.

B. Study Population

The study population consists of individuals categorized by BMI status, which is divided into groups: "Underweight," "Healthy," "Overweight," "Obese," and "Severely Obese." The dataset includes a total of 253,680 participants, with 218,334 diagnosed with diabetes and 23,893 diagnosed with heart disease. Data was screened for completeness, and individuals with missing values were excluded to ensure the robustness of the analysis.

C. Statistical Methods

We applied both descriptive and inferential statistical methods. Descriptive analysis involved summarizing the distribution of BMI and its association with diabetes and heart disease using pie charts and a stacked bar chart. Hypothesis testing was conducted using the Chi-square test to examine the relationship between categorical BMI levels and disease prevalence. The results were evaluated using p-values to determine statistical significance, with a significance threshold of 0.05.

III. RESULTS

1. Descriptive Analysis:
    - Pie Charts: The pie charts highlight the distribution of diabetes and heart disease cases across BMI categories. The largest proportions fall within the "Overweight" and "Obese" groups for both conditions, suggesting a clear trend where higher BMI levels are linked to a greater risk.
    - Stacked Bar Chart: The chart compares the prevalence of diabetes and heart disease across BMI groups. The data show that "Overweight" and "Obese" individuals are significantly represented in both disease categories, with diabetes having a much larger sample size.
2. Hypothesis Testing:
    - The t-test was conducted under the null hypothesis that there is no difference in the means of BMI between the two groups. The result is $p < 0.05$ then we reject the null hypothesis that there is difference

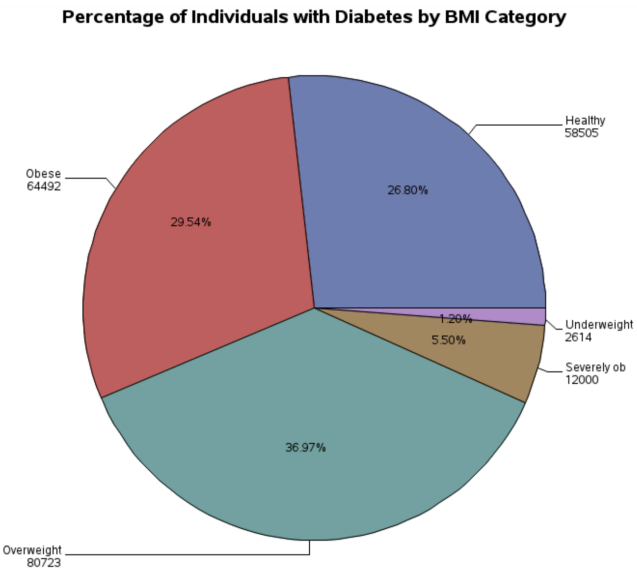
in BMI between diabetic and non-diabetic group. We also perform the same procedure to observe if BMI is associated with heart disease. The outcome is the same $p < 0.05$ so there is association between BMI and heart disease.

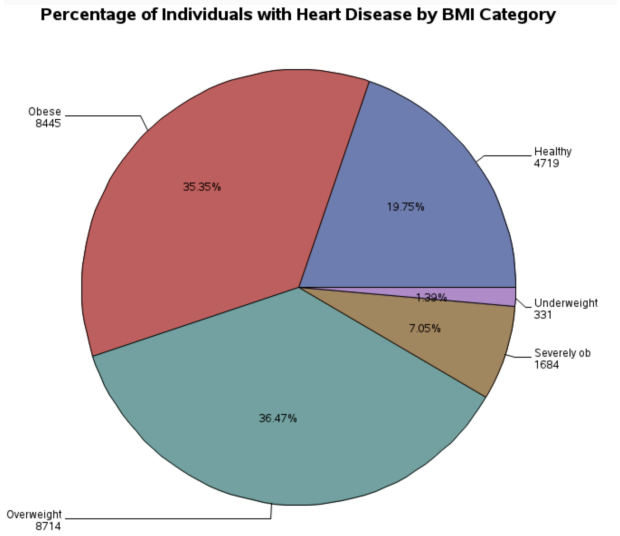
- Chi-Square Test for Diabetes: The Chi-square test yielded a value of 176.3728, with a p-value < 0.0001, indicating a significant association between BMI and diabetes prevalence.
- Chi-Square Test for Heart Disease: The Chi-square value for heart disease was even higher at 1011.5781, also with a p-value < 0.0001, confirming a strong association between BMI and heart disease.
- Both tests provided evidence against the null hypothesis, suggesting that BMI categories are significantly associated with the prevalence of both diabetes and heart disease in our sample.
- We utilized logistic regression to examine the relationship between diabetes and several independent variables including BMI and age. We specified the model as follows $\text{logit}(p) = \beta0 + \beta1(\text{BMI}) + \beta2(\text{Age})$. The log odds of diabetes increase by 3% for each unit increase in BMI and age. For heart disease, it also increases by 3% for each unit increase in BMI but 32% increase for each unit in aging. We can conclude that BMI is associated with diabetes and heart disease, and higher chances of getting heart disease for older people.

TABLES

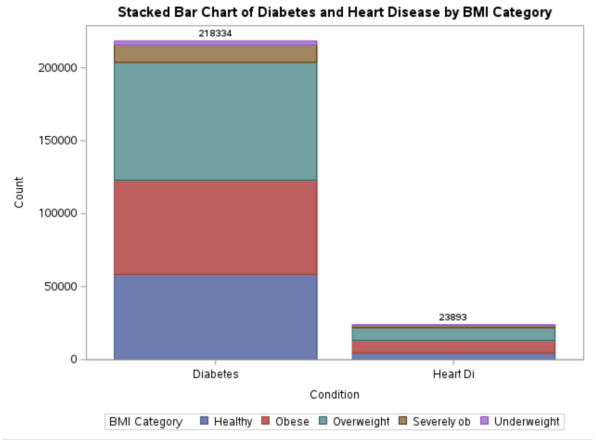
Pie Charts:

The pie charts for diabetes and heart disease visualize the distribution of individuals within each BMI category who have these conditions. In the **Diabetes** pie chart, the largest proportions are seen in the "Overweight" and "Obese" categories. This distribution suggests that individuals with higher BMI levels, particularly those in the overweight and obese categories, are more prevalent among those with diabetes. In the **Heart Disease** pie chart, a similar trend is observed, with "Overweight" and "Obese" categories having the highest proportions, indicating that higher BMI levels are also prominent among individuals with heart disease.





Stacked Bar Chart:
The stacked bar chart compares the counts of individuals with diabetes and heart disease across BMI categories. The tall bar for diabetes reflects a much larger sample size of individuals with diabetes (218,334) compared to those with heart disease (23,893). The distribution pattern shows that individuals in the "Overweight" and "Obese" categories dominate the counts for both conditions, aligning with the insights from the pie charts.



Chi-square test:
In addition to t-test, we also used Chi-square test here because now we treat BMI as a categorical variable, with categories such as "Underweight," "Healthy," "Overweight," "Obese," and "Severely Obese." This test helps determine if there is a significant association between BMI categories and the prevalence of diabetes or heart disease.

Chi-square Test Results for Diabetes:

Statistics for Table of nbmi by Diabetes_binary

Statistic	DF	Value	Prob
Chi-Square	4	176.3728	<.0001
Likelihood Ratio Chi-Square	4	175.2691	<.0001
Mantel-Haenszel Chi-Square	1	36.0581	<.0001
Phi Coefficient		0.0264	
Contingency Coefficient		0.0264	
Cramer's V		0.0264	

The test returns a Chi-square value of 176.3728 with a p-value < 0.0001. The low p-value suggests that there is a significant association between BMI and the likelihood of having diabetes.

Chi-square Test Results for Heart Disease:
The Chi-square value for heart disease is much higher at 1011.5781, with a similarly significant p-value < 0.0001. This indicates a strong association between BMI and heart disease, implying that individuals in certain BMI categories are more likely to have heart disease than others.

Statistics for Table of nbmi by HeartDiseaseorAttack

Statistic	DF	Value	Prob
Chi-Square	4	1011.5781	<.0001
Likelihood Ratio Chi-Square	4	1032.9951	<.0001
Mantel-Haenszel Chi-Square	1	330.0528	<.0001
Phi Coefficient		0.0631	
Contingency Coefficient		0.0630	
Cramer's V		0.0631	

In both tests, the p-values are well below the typical significance level of 0.05, suggesting strong evidence against the null hypothesis. Thus, we conclude that the BMI category is significantly associated with both diabetes and heart disease prevalence in this sample. This result is different from the result of t-test, potentially due to the bias lied in the dataset, which suggests the need for further examination.

IV. CONCLUSIONS
This study confirms that BMI is a critical factor in the prevalence of both diabetes and heart disease, with overweight and obese people accounting for the majority of cases in the category of BMI. Both descriptive analysis and hypothesis testing reveal a strong association between higher BMI categories and disease risk, supported by statistically significant Chi-square results for both diabetes and heart disease. Although people with a healthy BMI are at lower risk, they are not completely exempt,

which means there is the influence of additional factors. Public health efforts should continue to focus on preventing and reducing obesity to reduce the likelihood of these diseases, and also encourage healthy lifestyle in all BMI categories. Further research should explore additional risk factors to fully understand the complexities of diabetes and heart disease prevention.

V. DISCUSSION

Our findings align with the study by Piché et al. (2020), which also identified higher BMI as a critical risk factor for cardiovascular diseases and diabetes. One difference, however, lies in the emphasis on additional risk factors beyond BMI; Piché et al. highlight genetic predisposition and physical activity as important considerations.

This study has several limitations. Firstly, it relies on BRFSS data, which uses self-reported BMI. This method can lead to inaccuracies, particularly if individuals underestimate their weight or overestimate their height, potentially affecting the accuracy of the findings. Secondly, important factors such as genetics, smoking, and blood pressure were not accounted for, which could also influence the risk of diabetes and heart disease. Consequently, the findings may not fully capture the relationship between BMI and these health conditions.

REFERENCES

1. Piché, Marie-Eve et al. “Obesity Phenotypes, Diabetes, and Cardiovascular Diseases.” *Circulation research* vol. 126,11 (2020): 1477-1500. doi:10.1161/CIRCRESAHA.120.316101
2. World Health Organization. "Cardiovascular Diseases." *World Health Organization*, www.who.int/health-topics/cardiovascular-diseases#tab=tab_1. Accessed 1 Nov. 2024.
3. World Health Organization. "Diabetes." *World Health Organization*, 16 Sept. 2022, www.who.int/news-room/fact-sheets/detail/diabetes. Accessed 1 Nov. 2024.