UDS Team Pandas: Exploratory Report
Lilah DuBoff, Mu Niu, Samantha/Xiangyu Wang, Su Zhang

Note: We would like to present this report to a stakeholder who has the power to help design and fund initiatives to reduce nation-wide cardiovascular disease risk. For example, this could be a senior or executive member of the CDC or NIH.

I. Executive Summary

Over the past few decades, there has been a concerning increase in the number of younger adults (less than 50 years of age) developing and/or suffering from early warning risk factors of cardiovascular disease (CVD) in the United States (Andersson and Vasan, 2017; Aggarwal et. al, 2023). Cardiovascular disease is the leading cause of death in the United States, costing the healthcare system an estimated $252.2 billion dollars and afflicting millions of people in recent years alone (CDC, 2024). CVD is additionally a preventable disease, as many of the contributing factors are lifestyle choices or habits that can be improved with relatively simple interventions (Martin, 2025; CDC, 2024). Furthermore, there is a lack of data on acquired cardiovascular disease in young adults, and according to Andersson and Vasan, "...estimates are associated with wide confidence intervals owing to the low absolute number of individuals in this population with cardiovascular disease" (Andersson and Vasan, 2017).

This report aims to explore areas to help younger adults offset the risk of CVD, identify at-risk age and ethnic groups, and establish a baseline from which more data can be collected to confirm epidemiological trends in CVD in younger adults. First, we identify prevalent bad habits or lifestyle choices associated with cardiovascular disease patients. Next, we identify an approximation of the distribution of different types of cardiovascular disease and their relationship to demographics - exploring whether factors like age, gender, race, and socioeconomic status have an impact on the development of chronic disease. Finally, we seek to pinpoint when younger patients with cardiovascular disease began to show a noticeable increase in prevalence.

Our preliminary data analysis results were not explicitly consistent with existing literature. Previous research finds that several biomedical conditions and lifestyle habits contribute to the development of CVD - our analysis revealed only weak correlations between CVD and age (0.24), bmi (0.17), and cholesterol (0.22) (Martin, 2025). Historically, men are more at risk for developing CVD than women - our analysis found that gender differences are minimal, with males having a slightly higher prevalence of CVD (50.52%) compared to females (49.67%) (Martin, 2025). Analysis results additionally showed that older individuals are at a higher risk, which is consistent with current findings (Martin, 2025; CDC, 2024).

II. Introduction, Context, and Data

According to the American Heart Association, the absence of cardiovascular disease can be defined by seven guidelines, referred to as Life's Simple Seven (Martin et. al, 2025). For an individual to have good cardiovascular health, they must

> "not clinically manifest cardiovascular disease and also have optimal levels of all 7… components, including not smoking, a healthy diet pattern, sufficient [physical activity], normal body weight, and normal levels of [total cholesterol], [blood pressure], and [fasting plasma glucose] in the absence of medication treatment" (X).

To simplify, The Centers for Disease Control and Prevention (CDC) list that biomedical conditions including high blood pressure, high blood cholesterol, and diabetes are "key risk factors" for heart disease (CDC, 2024). Other preventable conditions that contribute to the development of heart disease include smoking, unhealthy diet, a lack of physical activity, excessive alcohol use, and high BMI (overweight and/or obesity) (CDC, 2024). Andersson and Vasan discuss the data collected in the past few decades that forecasts CVD (both congenital and acquired) among individuals aged 18–45 years seems to be increasing (Andersson and Vasan, 2017). Mitchell explains that medical advancements in the past seventy years have worked to reduce the mortality rate of CVD, but that progress has plateaued (Mitchell, 2023). Medical and public health officials have suggested that this plateau is indicative of a future epidemic of worsened CVD. In particular, among those aged 18-45. There has been little data gathered on CVD in this age group, and even less analyzing the effect by race and ethnicity (Andersson and Vasan, 2017; Mitchell, 2023; CDC, 2024). Therefore, this report incorporates existing literature surrounding risk factors and target demographics, and an original data analysis that together could contribute to developing more effective preventative strategies and resources for at-risk groups.

The data utilized in this report comes from a publicly available file on Kaggle, which offers demographic and biomedical data that is HIPPA compliant (Kaggle, unknown). Variables include patient age, height, weight, gender (for this analysis, this variable has been changed to biological sex, for clarity purposes), systolic blood pressure (SBP), Diastolic Blood Pressure (DBP), cholesterol, blood glucose (BG), smoking habits, alcohol intake, physical activity, and presence or absence of cardiovascular disease. In future reports and analyses, we aim to utilize the ResearchAllOfUs database, which requires ID registration and RCR training[1].

[1] We discovered this database post original analysis, but researched Duke's participation in the organization. We are able to get first tier access by completing an RCR training, and can establish second tier access by communicating with the primary Duke coordinator, Susan Hayden. We anticipate this data being more reliable than Kaggle, as it is operated through the NIH.

III. Lifestyle Choices Associated with Cardiovascular Diseases

Figure 1 displays a correlation matrix of the variables in our dataset and their relationship to CVD. The risk factors most strongly correlated with CVD include age (0.24), bmi (0.17), and cholesterol (0.22), all indicating low-strength relationships.These findings were inconsistent with existing literature, which suggests that SBP/DBP, BG, and smoking should all be correlated as well (Andersson and Vasan, 2017; Mitchell, 2023; CDC, 2024; Martin, 2025).

According to the CDC, key risk factors for cardiovascular disease include smoking, alcohol consumption, and lack of physical exercise (CDC, 2024). From the data analysis, we observe that the prevalence of smoking and alcohol consumption is slightly lower among CVD patients (8.37% and 5.21%, respectively) compared to non-CVD patients (9.25% and 5.54%, respectively). However, a smaller proportion of CVD patients (78.96%) reported engaging in regular physical activity compared to non-CVD patients (81.79%) (Table 1). Both findings suggest that lack of physical exercise is a risk factor for CVD. However, there is a slight discrepancy in alcohol consumption and smoking, where the data shows these behaviors are slightly less prevalent among CVD patients compared to non-CVD patients.

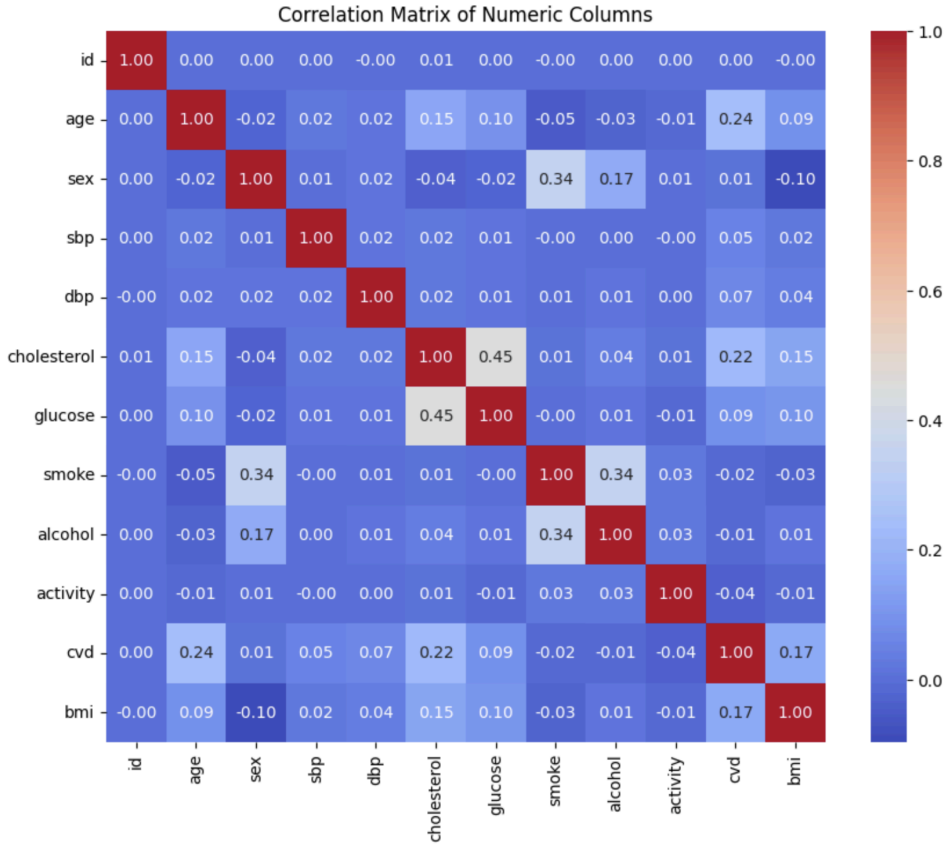

Figure 1: Correlation matrix identifying the strongest risk factors for CVD. Highest correlation values include age (0.24), bmi (0.17), and cholesterol (0.22), all indicating low-strength relationships.

Table 1: Prevalence of bad habits (smoking, alcohol usage, and physical activity) among CVD patients (n=70000), calculated as the proportion of those who take part in these activities out of the entire sample.

Lifestyle Choice/Habit	**CVD Patient**	**Non-CVD Patient**
Smoking	0.0837	0.0925
Alcohol Consumption	0.0521	0.0554
Physical Activity	0.7896	0.8179

IV. Relationship to Demographics

According to the American Heart Association, the data published by governmental agencies for some racial and ethnic groups are considered unreliable because of the small sample size in the studies. This makes finding certain demographics information about cardiovascular disease difficult (Martin, 2025). Therefore, our analysis covers age and sex, while racial and ethnic group and socioeconomic status statistics are supplemented by existing literature.

From the data analysis, we see clear patterns in the distribution of CVD. Older individuals are at a higher risk. Gender differences are minimal, with males having a slightly higher prevalence of CVD (50.52%) compared to females (49.67%). Research suggests that race and ethnicity play a critical role, as Black (Non-Hispanic) individuals experience the highest proportion of CVD-related deaths (22.6%), followed by Asians (18.6%) and Native Hawaiians or Other Pacific Islanders (18.3%), while Hispanic individuals experience the lowest proportion (11.9%)(CDC, 2024). Socioeconomic status is also a significant determinant of CVD outcomes. Individuals with lower socioeconomic status are more likely to experience CVD because of limited access to healthcare and challenges in managing chronic conditions (Schultz et al., 2018).

Table 2: A comparison of the median and mean (sd) age of patients, classified into those with CVD (n=34979) and those without CVD (n=35021).

	CVD Patient	**Non-CVD Patient**
Median Age	55.85	52.07
Mean Age (SD)	54.95 (6.345)	51.73 (6.778)

Table 3: The distribution of biological sex across the entire dataset (total n=70000, males=24470, females=45530)

	Biological Male	**Biological Female**
Prevalence of CVD	0.5052	0.4967

V. When Did Patients Become Younger?

Young patients with cardiovascular disease, particularly those aged 40 or younger, began to show a noticeable increase in prevalence since 2000 (Andersson and Vasan, 2017). A study conducted between 2000 and 2016 found that the proportion of very young individuals experiencing heart attacks has been rising by 2% each year during the study period. Moreover, 1 in 5 heart attack patients in the study group is 40 or younger, with some as young as their 20s and early 30s (American College of Cardiology, 2019).

When looking at the risk factors of CVD identified in Section III, existing research confirms that the rates of certain risk factors have all increased over the past 40 years for younger adults. Andersson and Vasan report several findings: The global age-adjusted prevalence of obesity has risen from 3.2% to 10.8% in men and from 6.4% to 14.9% in women; in the United States, more than 30% of young adults aged 15-29 are classified as being sedentary or physically inactive; In the USA, the prevalence of any type of diabetes increased by 250% in individuals aged 0–44 years and by 220% in individuals aged >44 years between 1980 and 2014. However, they additionally report some positive results: The global prevalence of cigarette smoking in adults aged 15-35 declined globally by ~25% between 1980 and 2012; In high-income countries, blood-pressure levels dropped in individuals aged 5–34 years between 1948 and 1998 (Andersson and Vasan, 2017).

V. Conclusion

The increasing prevalence of CVD risk factors among younger adults in the United States is a growing public health concern, due to the potential for long-term negative impacts. Despite medical advancements that have historically reduced CVD mortality rates, the progress has plateaued, and in some cases, other issues are arising. This analysis, alongside previous literature and research, highlights several critical factors that need addressing. Persistent unhealthy lifestyle habits, obesity, physical inactivity, and diabetes, all have the potential to drastically increase the rate of CVD in this generation of young adults. While our preliminary findings suggest weak correlations between established risk factors and CVD, this further underscores the need for further, more comprehensive data collection and analysis.

As CVD is primarily preventable through lifestyle changes and/or interventions, there is a need for the development of initiatives to encourage healthier habits - resources and projects that promote early screening, improve health education, and address inequalities in access to healthcare. Additional data will also be invaluable - driving research in underrepresented populations and demographics will be crucial to the development of successful prevention strategies. By expanding research efforts, we have the opportunity to have an impact on the reversal of increasing disease rates, therefore reducing the future impact of CVD on public health, and reducing the financial burden placed on our healthcare system by chronic disease.

References

1. Andersson, C., Vasan, R. Epidemiology of cardiovascular disease in young individuals.*Nat Rev Cardiol* 15, 230–240 (2018). https://doi.org/10.1038/nrcardio.2017.154

2. Martin, Seth S., Aaron W. Aday, Norrina B. Allen, Zaid I. Almarzooq, Cheryl A. M. Anderson, Pankaj Arora, Christy L. Avery, et al. n.d. "2025 Heart Disease and Stroke Statistics: A Report of US and Global Data From the American Heart Association," Circulation 0, 0 (0). American Heart Association. doi:10.1161/CIR.0000000000001303.

3. Centers for Disease Control and Prevention. 2024. "Heart Disease Facts." Centers for Disease Control and Prevention. October 24. https://www.cdc.gov/heart-disease/data-research/facts-stats/index.html.

4. Centers for Disease Control and Prevention. 2024. "Preventing Heart Disease." Centers for Disease Control and Prevention. May 24. https://www.cdc.gov/heart-disease/prevention/index.html

5. Andersson, C., & Vasan, R. S. (2017). *Epidemiology of cardiovascular disease in young individuals. Nature Reviews Cardiology, 15(4), 230–240.*doi:10.1038/nrcardio.2017.154 https://www.bidmc.org/about-bidmc/news/2023/03/cardiovascular-risk-factors-rise-among-young-adults

6. Schultz, W. M., Kelli, H. M., Lisko, J. C., Varghese, T., Shen, J., Sandesara, P., Quyyumi, A. A., Taylor, H. A., Gulati, M., Harold, J. G., Mieres, J. H., Ferdinand, K. C., Mensah, G. A., & Sperling, L. S. (2018). Socioeconomic Status and Cardiovascular Outcomes: Challenges and Interventions. *Circulation*, *137*(20), 2166–2178. https://doi.org/10.1161/CIRCULATIONAHA.117.029652