A Comparison of Chronic Disease Prevalence using *All of Us* (*AoU*), National Health Interview Surveillance (NHIS) and National Health and Nutrition Examination Survey (NHANES) Datasets in the Years 2022-2023

Henri Alejandro Coronado-Volta

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Public Health

University of Washington

2025

Committee:

Stephen J. Mooney

Jimmy Phuong

Vignesh Subbian

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

A Comparison of Chronic Disease Prevalence using *All of Us* (*AOU*), National Health Interview Surveillance (NHIS) and National Health and Nutrition Examination Survey (NHANES) Datasets in the Years 2022-2023

Henri Alejandro Coronado-Volta

Chair of the Supervisory Committee:

Stephen J. Mooney

Department of Epidemiology

Background: Chronic diseases remain a primary public health problem in the United States, with disparities enduring across a wide range of racial and ethnic populations. This study

assesses the generalizability of the *All of Us (AoU) Research Program* by analyzing the prevalence of chronic conditions with those from two national surveillance systems – National Health Interview Survey (NHIS) and National Health And Nutrition Examination Survey (NHANES).

Methods: A cross sectional analysis was performed using data from *AoU*, NHIS and NHANES compiled from years 2022 and 2023. Six chronic conditions were analyzed: hypertension, stroke, prediabetes, diabetes, myocardial infarction, cancer, asthma. Prevalence estimates and ratios were computed by overall and stratified by race/ethnicity with a focus on differences of sampling, demographics and collection approaches.

Results: *AoU* reported higher or similar estimates for chronic conditions relative to NHIS and NHANES especially for stroke, prediabetes, diabetes, myocardial infarction and asthma. In contrast, hypertension and cancer prevalence was lower in *AoU*. Racial and ethnic disparities were apparent across these datasets, with underrepresented populations for Black and AIAN populations experiencing a higher burden for many conditions.

Conclusions: These findings indicate that *AoU*, despite convenience sample methods, provide comparable and occasionally higher chronic disease prevalence than traditional surveillance systems. These results draw attention to the need for methodological considerations in generalizing findings from *AoU* to the broader US population. The study emphasizes the value of inclusive and representative data to better understand and address health disparities.

Keywords: Prevalence, Chronic Conditions, Latino population, Generalizability

Introduction

Chronic diseases or chronic health conditions, including hypertension, diabetes, cancer, and asthma, are important public health challenges in the United States. Understanding the prevalences of these conditions and how prevalence differs across population subgroups is crucial for monitoring the success of prevention and treatment strategies.

In the United States, national survey-based surveillance systems, including the National Health and Nutrition Examination Survey (NHANES) and the National Health Interview Survey (NHIS) have long been used to monitor trends in disease prevalence and health-related practices[1,2]. NHIS and NHANES are national surveillance systems that track disease prevalence and health behaviors in the United States. NHIS and NHANES recruit 100,000 and 5,000 to 10,000 people yearly, respectively, using population-based random sampling techniques that allow for generalization to the overall US population[1,2]. In contrast, the *All of Us (AoU) Research Program* is a national longitudinal research cohort with a recruitment goal to achieve 1 million participants. As of May 2025, the *AoU Research Program* has recruited approximately 800,000 convenience-sampled US individuals with cohort data consisting of survey data, electronic health records (EHR) data, biometric measurements, and genomic data[3]. Little is known about how generalizable findings from the *AoU Research Program* are to the general US population. Benavidez and colleagues reported differences in prevalence of hypertension, diabetes, cancer, and asthma across geographic and sociodemographic factors. Findings showed that chronic disease prevalence was higher in the southeastern US and with greater socioeconomic disadvantage[4]. Nevertheless, these prior studies did not report differences by race/ethnicity.

The current study compares prevalence estimates of seven chronic health conditions (i.e., hypertension, hypertension medication, stroke, prediabetes, diabetes, myocardial infarction, cancer, and asthma) derived from the *AoU Research Program* with weighted population-based estimates from NHIS and NHANES.

This comparison will provide insights into consistency of findings and differences due to sampling methods, data collection techniques, and population coverage. If *AoU* approximates disease prevalence in established national surveillance systems, it increases confidence that *AoU* results can generalize, or that results from population-based studies in *AoU* are not compromised by the sampling scheme. On the other hand, all surveys data may be biased by low response and differentially higher response in healthier individuals[5]. The findings will help us understand to what extent *AoU* approximates disease prevalence compared to established national surveillance systems.

My study adds to this research by comparing the different prevalence ratios in each racial and ethnic category. In particular, my research provides insights on the generalizability of *AoU* to other populations and lays the groundwork for future research into interventions to promote chronic disease prevention.

Methods

Objective

A cross sectional study design was used to compare exploratory descriptive statistics of seven chronic health conditions derived from NHIS, NHANES, and *AoU* for the years 2022 and 2023. We compared the reported rates and prevalence ratios (PR) of the seven chronic conditions across racial and ethnic categories.

Prevalence for the NHIS is defined as the weighted proportion of respondents who report a diagnosis of a condition at the time of the interview[1]. The prevalence in the NHANES dataset is the weighted proportion of individuals who meet self-reported criteria for a condition[2]. The *AoU* proportion is defined as the unweighted proportion of participants with a diagnosis and self-reported condition[3]. All estimates are based on individuals aged 18 years or older during the years 2022 or 2023. Those who are taking hypertension medication are assumed to have hypertension. The phrasing of the questions for the conditions prediabetes[6] and hypertension medication[7] make it challenging to individually analyze these data and come up with a conclusion that is impacted solely these conditions and the American Indian / Alaskan Native (AIAN) racial group[8] is also not widely sampled in NHANES dataset. A cumulative analysis was used to analyze the *AoU* dataset. In this analysis the *AoU* utilizes only the self-report and was not confirmed by the EHR. For recurring conditions (eg. cancer) the *AoU's* focus is current presence of the condition while NHIS and NHANES describe it as ever having [condition]. Consequently, it could be a reason why cancer is lower in the *AoU* population.

The NHIS dataset asked the question "Have you EVER been told by a doctor or other health professional that you had…Hypertension, also called high blood pressure". This framing was similar across all the health conditions. This format aligns with the NHANES which asked "Have you (subject participant) ever been told by a doctor or other health profession that (you/s/he) had hypertension, also called high blood pressure. This continues with the other health conditions. In contrast, when the *AoU* asked this question, it was framed "Including yourself, who in your family has had high blood pressure (hypertension)? Select all that apply". The survey would give the responses: mother, father, sibling, daughter, son, grandparent. The participants can select multiple people for each question. These questions are framed the same way as the other health conditions.

We hypothesize that prevalence values will be higher in the *AoU* dataset compared to the NHIS and NHANES datasets due to differences in sampling and recruitment methods deployed (convenience vs population based). That is, because *AoU* recruited at clinics and community settings, respondents may be

more likely to have a chronic condition than those who were recruited using population-based sampling methods.

Data sources

NHIS is a nationally representative program from the National Center for Health Statistics and the Center for Disease Control and Prevention. Participants are recruited for the NHIS national survey using random sampling which are selected based on addresses. The NHIS collects data through personal interviews and uses stratified, multistage probability sampling designs. It collects data on health conditions, behaviors, healthcare access and utilization, functioning limitations, mental health and risk factors. Surveys are administered through an in-person computer-assisted personal interviewing conducted by the U.S. Census Bureau. NHIS has a robust national repository of survey data, and EHR data from a diverse cohort of 87,500 people across roughly 35,000 households; the survey is administered yearly[1].

NHANES is a program used to assess the nutritional status from individuals throughout the United States. This program recruits participants through probability sampling through random selection by primary sampling units across the United States to generalize to a broader population. The NHANES dataset uses multistage probability sampling and survey components to collect data regarding health behaviors, health conditions, dietary and nutrition information, environmental exposures, physical exams including vital signs, laboratory tests, and physical measurements. It collects data from 5,000 people each year. The NHANES is a robust national repository of survey data, and EHR data from a diverse cohort of around 5,000 people per year[2].

As part of the mission for the *AoU Research Program*, the dataset aspires to be diverse and include a wide range of underrepresented and underserved individuals, in terms of racial-ethnic minorities, rurality, lower income among other demographic groups. While *AoU* is a nation wide cohort, it has specific recruitment hubs in Tennessee, Massachusettes, California, Arizona and Washington. Prospective participants are recruited from healthcare providers, hospitals, community organizations and other locations. *AoU* is a

national repository of survey data, biomedical samples, and linked EHR data from a diverse cohort of 800,000 adults (with the goal of enrolling 1 million in total). The *AoU* collects information on health conditions, preexisting conditions, genetic data, environmental exposures, lifestyle behaviors, health-related behaviors, healthcare access and usage, follow up and longitudinal data. It collects data from surveys, the EHR, biological sampling and wearable devices[3].

Comparision of Data Types and Research Strategies

NHIS is an in-person survey that monitors population-level trends to allow for focused interventions while NHANES provides self-reported and physical measurements to understand prevalence and deficiencies of risk factors. NHIS concentrates on collecting data through interviews while NHANES a door to door survey is focused on collecting physical examinations and lab tests which provides a broader scope of health status of nutrition and chronic diseases[2].

During the COVID pandemic, the NHIS continued to collect data, although the recruiters interviewed individuals by phone in order to accommodate COVID regulations introduced in the United States. In contrast to the NHIS, the NHANES suspended all interviews and recruitment which accounts for the year and a half gap in the data collection (2019-2020). Survey collection resumed in in 2021[9,10].

The goal of the *AoU Research Program* is to establish a diverse and representative sample of the United State population. Recruitment takes place using convenience sampling primary at clinic and community-based locations; individuals can also directly enroll through the *AoU* website. Participants are able to complete surveys, provide access to EHR records and sync Fitbit data. They can also visit partner sites for measurements and laboratory samples. The *AoU Research Program* keeps participants updated on new opportunities to participate and further the research of the study[3].

Analysis Plan

This study investigates the prevalence of six chronic conditions and health outcomes – hypertension, hypertension medication, stroke, prediabetes, diabetes, myocardial infarction, cancer, and asthma - across

three major health datasets: NHIS, NHANES, *AoU Research Program*. This analysis concentrates on the comparison across five racial-ethnic groups: Hispanic, White, Black, Asian, AIAN.

Cohort Construction

The analytic cohort includes participants aged 18 years and older from the NHIS, NHANES, and *AoU* datasets. Inclusion criteria are similar, when possible, across all datasets for compatibility.

The NHANES did not have data for hypertension medication or prediabetes and were not used in the analysis portion.

The races 'other single and multiple races', 'refused', 'not ascertained' and 'don't know' in the NHIS; 'other race – including multi-racial', 'missing' in the NHANES; 'I prefer not to answer', 'Middle Eastern or North African', 'More than one population', 'Native Hawaiian or Other Pacific Islander', 'Not Indicated', 'None of these', and 'PMI: Skip' were omitted to support exchangeability. The race American Indian/Alaska Native and American Indian/Alaska Native Other was combined due to small sample size in the NHIS. The NHANES did not have the American Indian/Alaska Native race was not included in the analysis.

Only data collected in the years 2022 and 2023 are used.

For the NHIS and NHANES this ascertainment of these conditions was based on self-report, clinical diagnoses and exams and laboratory measurements. The *AoU* identification of condition based on EHR and survey data, aligned with definitions in the NHIS and NHANES[1,2,3].

Condition Definitions

Each condition is defined using a standardized consistent criterion across all datasets.

Hypertension: self-reported diagnosis; NHANES includes systolic/diastolic blood pressure >= 140/90 mmHg[11,12].

Stroke, Myocardial Infarction, Diabetes, Cancer, and Asthma: Identified from self-report or EHR codes in *AoU*.

Prediabetes: Defined by self-report or HbA1c between 5.7%-6.4% in NHANES[13]

Prevalence Estimation

Prevalence is estimated for all conditions within racial and ethnic groups for all datasets. NHIS and NHANES estimates are weighted using design weights, sampling units and stratification for nationally representative data. *AoU* estimates are unweighted proportions.

Comparative Analysis

To evaluate consistency and the generalizability of *AoU* estimates the absolute differences and prevalence ratios are calculated to compare *AoU* with the NHIS and NHANES for each condition within each demographic group. Forest plots and bar charts were used for the visualization of the PRs and prevalences respectively to compare *AoU* against the NHIS and NHANES benchmarks.

Results

Participant Characteristics

Out of 57,173 participants in NHIS, 1,784,471 participants in NHANES and 633,547 participants in the (Table 1), females comprised more than half of the NHIS and NHANES datasets, and almost two-thirds of the *AoU* cohort. One half of NHIS respondents were 55 and older in comparison to 40% of NHANES and 52% of *AoU* respondents (Table 1). About two thirds of the NHIS and NHANES population identify as White with 56. 5% in *AoU* (Table 1). More than 20% of individuals in NHANES identified as Hispanic, compared with 12.7% and 17.8% in NHIS and *AoU,* respectively.

Overview

In general, *AoU* has a higher or equal prevalence of stroke, prediabetes, diabetes, myocardial infarction and asthma than NHIS and NHANES (Table 2). NHIS and NHANES have higher prevalence of hypertension and cancer than *AoU* (Table 2).

In comparisons of prevalence ratios across the three datasets, Hispanic participants had lower prevalence than White participants for hypertension and stroke. Black participants had a higher prevalence for hypertension medication in NHIS and *AoU*. The prevalence ratios are all greater for the underrepresented groups for prediabetes, except for Hispanic participants in the AoU dataset. The prevalence ratios for diabetes are greater in all three datasets for Black participants. The trends are less than White participants for MI except for Black and AIAN participants in the *AoU.* For cancer the prevalence ratios are all lower in the underrepresented population than White participants. Asthma is lower for Hispanic in the NHIS and *AoU*. Black participants having a higher asthma prevalence NHIS and NHANES.

Hypertension

Hypertension prevalence was 24% in NHANES, 32% NHIS and 27% in *AoU* (Table 2). Compared to White participants in each respective dataset, Hispanic participant prevalence was lower in NHIS (PR 0.73; CI [0.67,0.81]) and *AoU* (PR 0.42; CI [0.41,0.43]) but similar in NHANES (PR 0.66 [0.58,0.74]). Black participants had a higher prevalence in NHIS and NHANES than in *AoU*. For Asian and AIAN participants, there was a higher prevalence in NHIS than *AoU*. The prevalence ratio compared to White participants was higher in the Black participants from NHIS. A similar pattern observed in NHANES with the highest burden of disease for Asian participants. While those individuals in the *AoU* had a greater prevalence in White participants than any other racial-ethnic category.

For the overall population-based prevalence estimates the NHIS was the highest followed by *AoU* and NHANES. The Hispanic prevalence is higher in the NHIS than NHANES and *AoU*. White participants

show a higher prevalence in NHIS than NHANES and *AoU*. For Black participants the estimates are higher than in the NHANES and *AoU*. Asian participant prevalence estimates are higher in the NHANES than NHIS and *AoU*. The AIAN participants are higher in NHIS than *AoU*.

Hypertension Medication

The estimates from participants diagnosed with hypertension are asked if those individuals are taking hypertension medication for both NHIS and *AoU*. The prevalence of those on hypertension medication is 25% in NHIS. From the population that has been diagnosed with hypertension more than four fifths in the *AoU* take medication for hypertension. The prevalence ratio comparing hypertension in non-Hispanic White participants to Hispanic participants was lower in NHIS than *AoU.* Black population prevalence was higher in NHIS than *AoU*.

The overall trends for those individuals who were diagnosed with hypertension that are taking medication, had a threefold increase in the *AoU* compared to NHIS. The Hispanic population prevalence is four times the amount in the *AoU* than the NHIS. In the White population the prevalence is a little over double. For the Black participants, the estimate in NHIS is a third of *AoU*. For Asian participants, the prevalence ratio is four times as much in the *AoU* than the NHIS. In AIAN participants, the prevalence ratio is four times as much in the *AoU*.

Stroke

History of stroke prevalence was 3% in the NHANES and NHIS and 7% in *AoU*. Compared to the White population in this study, Hispanic participant prevalence was lower in NHIS (PR 0.63; CI [0.43,0.92]) and NHANES (PR 0.71 [CI 0.47,0.95]) but similar in *AoU* (PR 1.06; CI [0.97,1.17]). In both the NHIS (PR 1.17; [CI 0.87,1.57]) and NHANES (PR 1.08; CI [0.68,1.48]) Black participant stroke prevalence was equivalent than White participants but higher in *AoU* (PR 1.89; CI [1.76, 2.03]). For the Asian participants stroke prevalence, there was a lower prevalence in NHIS and *AoU* than in NHANES. For the AIAN population stroke prevalence was higher in *AoU* (PR 2.27 [0.38,2.80]).

In terms of the history of stroke, the NHIS and NHANES are similar. The *AoU* prevalence ratio is only slightly higher for the Hispanic population. For White participants, the prevalence ratio estimates are the same for NIHS and NHANES and twofold increase in *AoU*. The Black participants have similar estimates for NHIS and NHANES and tenfold increase in the *AoU*. The estimates for Asian participants remained similar across all three datasets. There was a tenfold increase for the prevalence in the AIAN population.

Prediabetes

Prediabetes prevalence is 16% in NHIS and 51% in *AoU*. Compared to White participants, Hispanic and Asian participants had higher prevalence of prediabetes in NHIS (PR 2.21; CI [1.06,1.39], PR 1.34 [0.86,2.08]) but not in *AoU* (PR 0.91; CI [0.89,0.92], PR 1.00 [0.93,1.06]). Black participants had a higher prevalence in both datasets. For AIAN participants, the prediabetes estimates did not differ from White participants.

Prediabetes trends demonstrate a two to threefold increase from the NHIS to the *AoU*. This pattern continues to hold constant in the race and ethnicities. In Hispanic participants, the prevalence ratio estimate has a twofold increase in the *AoU* compared to the NHIS. The White population is triple the amount in *AoU* compared to the NHIS. For Black participants, the prevalence ratio is triple the amount in *AoU* than the NHIS. The Asian participants have three-fold more in the *AoU* than NHIS. The AIAN participants experience a threefold increase in *AoU* compared to NHIS.

Diabetes

Diabetes prevalence was 9% in NHANES, 8% in NHIS and 18% in *AoU*. Compared to the White population, Hispanic participants had higher prevalence of diabetes in NHIS (PR 1.18; CI [1.00,1.40]) but not in NHANES (PR 0.95; CI [0.78,1.13]) or *AoU* (PR 1.02; CI [0.99,1.05]). Diabetes prevalence among the Black population was higher than among the White population in in all three datasets. The Asian

participants diabetes prevalence was higher compared to White participants in NHIS but lower in NHANES and *AoU*. The AIAN population diabetes prevalence was higher in NHIS than *AoU*.

Diabetes population estimate trends are similar in both the NHIS and NHANES. The prevalence ratio estimate is a twofold increase in the *AoU* compared to the NHANES. The trend is only slightly higher for the *AoU* compared to the NHIS. In the Hispanic population the prevalence is only slightly higher in the *AoU* compared to the NHIS and double that of the NHANES. For the White participants the NHIS and NHANES estimates are similar with a twofold increase compared to the *AoU*. The Black population has a higher prevalence in NHIS compared to NHANES. The prevalence ratio estimate is slightly higher in *AoU* compared to NHIS. It is four times that of the NHANES prevalence ratio estimate. In terms of the Asian population the prevalence is similar across the three datasets. For the AIAN participants the estimate is only slightly higher in the *AoU* compared to NHIS.

Myocardial Infarction

The history of prevalence of MI was 3% in NHIS, 2% in NHANES and 7% in *AoU*. Compared to White participants, Hispanic population had lower prevalence of MI across all datasets (PR 0.40; CI [0.27,0.60], PR 0.43 CI [0.27,0.58], PR 0.91 [0.84,0.99]). The Black population MI prevalence was higher than White population MI prevalence in *AoU* but not in NHIS or NHANES. The Asian participants MI prevalence was lower than White participants MI prevalence in all three datasets. The AIAN population MI prevalence was higher in *AoU* but not in NHIS.

The prevalence estimates for the history of MI are the same for NHIS and NHANES and slightly higher in the *AoU*. The prevalence ratio estimates for the Hispanic population are the same in NHIS and NHANES and marginally higher for the *AoU*. The White population shows slightly higher estimates for *AoU* than NHIS and NHANES (2-4%). For the Black participants the prevalence ratio estimates are the same in NHIS and NHANES and higher in *AoU* (10%). In the Asian population the prevalence ratio is no

different for NHIS and NHANES and only slightly higher for *AoU*. AIAN participants are higher in the *AoU* than NHIS.

Cancer

The history of cancer was 10% in NHIS, 9% in NHANES and for those that had cancer currently was 8% in *AoU*. Compared to White participants the cancer prevalence for the Hispanic population was lower in all the datasets. The ratios for all three datasets show the prevalence ratios less than Whites (PR 0.40; CI [0.27,0.60], PR 0.43; CI [0.27,0.58], PR 0.91; CI [0.84,0.99]). In Asian participants, cancer prevalence is less than one for the three datasets but slightly higher in *AoU* (PR 0.72; CI [0.59,0.87]) than the NHIS (PR 0.40; CI [0.20,0.81]) and NHANES (PR 0.40; CI [0.15,0.64]).

The prevalence trends of cancer are similar in the overall estimates across the three datasets. In the Hispanic population the prevalence is similar. The prevalence ratio has a twofold increase in the *AoU* compared to the NHANES. The White population experiences a higher burden in the NHIS compared to NHANES and *AoU*. The prevalence ratio has a tenfold increase in NHIS compared to the NHANES. The estimates are similar in the NHIS and *AoU*. The Black participants have similar estimates in the NHIS and *AoU*. There is a slightly higher prevalence ratio estimate in NHANES compared to the other two datasets. For Asian participants, the estimates are similar across all three datasets. This similarity holds true for the prevalence ratio estimates of the AIAN population for NHIS and *AoU*.

Asthma

Asthma prevalence was 15% in NHIS datasets, 15% in NHANES, 37% in *AoU*. Compared to White participants, the Hispanic population had lower prevalence of asthma across all datasets (PR range 0.83-0.94). The prevalence of asthma was slightly higher for Black participants in NHIS (PR 1.19; CI [1.02,1.38]) and NHANES but no different in the *AoU*. The Asian participant asthma prevalence was slightly higher in *AoU* but no different in the NHIS and NHANES. The AIAN had a higher rate of asthma compared to Whites in the NHIS than the *AoU.*

For the Hispanic population the prevalence ratio estimate for asthma is the same for NHIS and NHANES and higher in the *AoU*. The prevalence for White participants is slightly higher in NHANES than NHIS and higher in *AoU*. For Black participants it is higher in *AoU* than both NHIS and NHANES. The Asian prevalence ratio is higher in *AoU* than NHIS and NHANES. The estimates for AIAN is slighly higher in *AoU* than NHIS.

Discussion

For this study, we compared the prevalence of chronic conditions across two population-based national surveillance systems (NHIS and NHANES) and the *AoU* dataset, which is a national cohort of participants that are selected through convenience sampling. We hypothesized that prevalence would be higher in the *AoU* dataset compared to the NHIS and NHANES datasets. Consistent with our hypothesis, we found that prevalence of stroke, prediabetes, diabetes, myocardial infarction, and asthma were higher in the *AoU* data, compared to NHIS and NHANES. Contrary to our hypothesis, we found that the prevalence of hypertension and cancer were lower in the *AoU* data, compared to NHIS and NHANES. These findings can inform future interpretations of *AoU* data.

Apart from the difference in sampling methods, recruitment strategies also differed across the datasets. The *AoU* recruitment was done in community and clinic settings while NHIS and NHANES used population-based recruitment, contacting individuals in their homes. Our comparision of prevalence ratios for participants of various races and ethnicities showed that for many of the conditions there was a higher prevalence in underrepresented and underserved populations, especially Black and AIAN populations that carried most of the burden of these conditions.

Table 1 and table 2 findings can be explained by the sampling methods and demographics of the populations. NHIS and NHANES use population-based and household sampling. The *AoU* uses convenience sampling predominately in healthcare settings. *AoU* recruits participants from academic institutions, VA medical centers and community health centers nationwide[3]. Convenience sampling may overrepresent people that are more likely to have chronic health conditions than the general population. Prevalence of hypertension is lower in the Latino population than in other races (AIAN, Asian, Black or African American and White) in the *AoU* than in the NHIS and NHANES datasets[14,15]. Thus, we propose that other chronic disease conditions would be similarly lower in the Latino population in the *AoU* due to the differences in sampling methods from NHIS and NHANES datasets.

The prevalence of hypertension and cancer were lower in the *AoU* compared to NHIS and NHANES. This might be explained by people with cancer being treated in oncology specific facilities, where *AoU* recruitment generally did not occur. The lower prevalence of hypertension may be partially explained by the higher proportion of females in the *AoU* dataset (63.1% vs 54.3% in NHIS and 51.1% in NHANES). Hypertension has a lower prevalence in females (44.6%) compared to males (50.8%)[16].

Age plays a big role in all these conditions. The proportion of participants 65 and older varied across the studies. About one third of NHIS (32.5%) and AoU (33.1%) participants were 65 years or older compared to 21.8% of NHANES participants. Nationally, 85% of individuals 65 years or older and 62.7% of individuals aged 50-64 have at least one chronic condition[17]. A future area of research would be to age adjust these chronic disease estimates. Because the proportion of individuals 65+ is higher in the *AoU*, may have a greater rate of disease. This is in part due to older people having more chronic conditions.

NHIS and NHANES utilize population-based weights which means they survey a limited number (5000-10000) of individuals and apply weights to represent the broader population. Wang and colleagues used raked weights (adjust sample weights in a contingency table to match desired marginal totals) to adjust

the prevalence estimates in *AoU* to approximate the general population. When they used raking, the prevalence of the conditions went down. Diabetes prevalence went from 11.67 to 10.99. The prevalence estimate for hypertension is 32.16 and the weighted prevalence is 29.20[18]. This raking could be a promising approach to comparing the results to the general population.

Prior Research

Prior research studies have used NHIS as a nationally representative population based sample. These data for the six conditions are hypertension (37.6%)[19], hypertension medication (69.0%)[20], stroke (2.9%)[21], prediabetes (17.8)[13], diabetes (15.8%)[22], MI (3.0%)[23], cancer (12.7%)[24], asthma (8.7%)[25]. These NHIS data pair well with similiar estimates from *AoU,* hence *AoU* can be generalized to the overall population.

The *AoU* further succeeded in recruiting participants across racial and ethnic subgroups. The *AoU* estimates include White (56.5%), Hispanic (17.8%), Black (15.8%), Asian (3.54%), AIAN (1.42%) populations. This compares to census data showing the represention of White (75.3%), Hispanic (19.5%), Black (13.7%), Asian (6.4%), AIAN (1.3%) populations[26].

Consistent with these national surveillance system *AoU* showed a higher burden of diabetes, asthma and hypertension. Although the overall prevalence of diabetes is 15.9%, the burden remains high in racial and ethnic subgroups. Diabetes is prevalent in 22.2% for Hispanics, 21.4% among Blacks, 12.2% among Whites, 19.3% among Asians[27]. Disparities are apparent in other chronic conditions with hypertension and asthma having particularly high rates in underrepresented and underserved populations. The healthcare gaps betweeen resources accessible to different groups underscore the importance for targeted strategies to reduce these disparities and address the social determinants of health that impact these communities. This illustrates the importance of combatting these disparities in racial ethnic subgroups.

Study Findings

These findings corroborate existing prior research that is used to validate my prevalence estimates for this study. Over the past two decades, the prevalence of many of these conditions has increased due to aging populations, lifestyle changes (e.g., increase sedentary behavior), environmental factors (e.g., increasing

global temperature, percent of racial demographics age breakdown concentration) and healthcare access and utilization[28]. This will be increasingly important to monitor chronic conditions over time. Since *AoU* collected biological samples, it will allow for robust studies on risk factors and etiology. Our findings can inform interpretation and applicability of results from these studies.

There are three limitations to this analysis: non-responsiveness to the survey questions (also referred to as non-participation bias), the categorization of race ethnicity and underreporting of disease categories. The nonresponsiveness in the NHIS is (61.1%)[1], NHANES (51.9%)[29], *AoU* (30.7)[30]. The non-responsiveness and how many individuals decline could impact the percentage point difference and prevalence ratios which would provide higher or lower estimates for these analyses[31].

The race and ethnic category is an inherent limitation as a person's identity could change over time and the "other" category might not encompass aspects of multiracial identities. Finally, the prevalences of conditions would allow underreporting for those who do not attend doctor's visits. The AIAN population was small in *AoU* dataset, leading to imprecise prevalence ratio estimates.

Conclusion

This analysis highlights both the consistencies and divergences in chronic disease prevalence estimates across three major U.S. data sources: NHIS, NHANES, and *AoU*. While all three datasets consistently show higher burdens of conditions such as hypertension, diabetes, and asthma among Black and Hispanic populations, notable variations in absolute prevalence estimates emerge—driven largely by differences in data collection methods, sample design, and condition ascertainment.

The population-based sampling frames of NHIS and NHANES offer strong generalizability, while the *AoU* cohort provides rich, longitudinal data with enhanced representation of historically underrepresented groups. However, *AoU*'s convenience sampling and reliance on a mix of self-report, EHR, and survey data can introduce biases, especially in conditions requiring clinical confirmation.

Together, these datasets offer complementary strengths. Cross-cohort comparisons, when interpreted with careful attention to methodology, can deepen our understanding of health disparities and inform targeted public health interventions.

Acknowledgements

I would like to express my deepest gratitude to Dr. Steve Mooney, my committee chair, for his invaluable guidance, support, and encouragement throughout the course of this research. I am also sincerely thankful to my committee members, Dr. Jim Phuong and Dr. Vignesh Subbian, for their insightful feedback and thoughtful mentorship.

A special thank you to Dr. Gloria Coronado, my aunt, whose dedication to public health continues to inspire me and James Nusrala my uncle, for his unwavering support and belief in my work. Thank you for answering the phone, I owe you more than a dinner.

Most importantly, I thank my mother, Nora Coronado, whose unconditional love, strength, and sacrifices made this journey possible. Her constant support has been the foundation of my achievements.

This thesis reflects the guidance, mentorship, and support I have been fortunate to receive. Thank you all.

References

1. Zablotsky B, Lessem SE, Gindi RM, Maitland AK, Dahlhamer JM, Blumberg SJ. Overview of the 2019 National Health Interview Survey Questionnaire Redesign. *Am J Public Health*. 2023;113(4):408-415. doi:10.2105/AJPH.2022.307197

2. Tissues NRC (US) C on NM of H. The National Health and Nutrition Examination Survey. In: *Monitoring Human Tissues for Toxic Substances*. National Academies Press (US); 1991. Accessed May 19, 2025. https://www.ncbi.nlm.nih.gov/books/NBK234172/

3. The "All of Us" Research Program. *N Engl J Med*. 2019;381(7):668-676. doi:10.1056/NEJMsr1809937

4. Benavidez GA, Zahnd WE, Hung P, Eberth JM. Chronic Disease Prevalence in the US: Sociodemographic and Geographic Variations by Zip Code Tabulation Area. *Prev Chronic Dis*. 2024;21:E14. doi:10.5888/pcd21.230267

5. Maitland A, Lin A, Cantor D, et al. A Nonresponse Bias Analysis of the Health Information National Trends Survey (HINTS). *J Health Commun*. 2017;22(7):545-553. doi:10.1080/10810730.2017.1324539

6. Mosslemi M, Park HL, McLaren CE, Wong ND. A treatment-based algorithm for identification of diabetes type in the National Health and Nutrition Examination Survey. *Cardiovasc Endocrinol Metab*. 2020;9(1):9-16. doi:10.1097/XCE.0000000000000189

7. Derington CG, King JB, Herrick JS, et al. Trends in antihypertensive medication monotherapy and combination use among US adults, NHANES 2005–2016. *Hypertens Dallas Tex 1979*. 2020;75(4):973-981. doi:10.1161/HYPERTENSIONAHA.119.14360

8. Johnson PJ, Blewett LA, Davern M. Disparities in Public Use Data Availability for Race, Ethnic, and Immigrant Groups. *Med Care*. 2010;48(12):1122-1127. doi:10.1097/MLR.0b013e3181ef984e

9. Blumberg SJ, Parker JD, Moyer BC. National Health Interview Survey, COVID-19, and Online Data Collection Platforms: Adaptations, Tradeoffs, and New Directions. *Am J Public Health*. 2021;111(12):2167-2175. doi:10.2105/AJPH.2021.306516

10. Paulose-Ram R, Graber JE, Woodwell D, Ahluwalia N. The National Health and Nutrition Examination Survey (NHANES), 2021–2022: Adapting Data Collection in a COVID-19 Environment. *Am J Public Health*. 2021;111(12):2149-2156. doi:10.2105/AJPH.2021.306517

11. Bays HE, Chapman ,Richard H., Fox ,Kathleen M., and Grandy S. Comparison of self-reported survey (SHIELD) versus NHANES data in estimating prevalence of dyslipidemia. *Curr Med Res Opin*. 2008;24(4):1179-1186. doi:10.1185/030079908X280527

12. Guo JW, Ning H, Lloyd-Jones DM. Cardiovascular Health Status in US Adults With Chronic Diseases: National Health and Nutrition Examination Survey (NHANES), 2013–2018. *J Am Heart Assoc Cardiovasc Cerebrovasc Dis*. 2024;14(1):e034388. doi:10.1161/JAHA.124.034388

13. Bullard KM, Saydah SH, Imperatore G, et al. Secular Changes in U.S. Prediabetes Prevalence Defined by Hemoglobin A1c and Fasting Plasma Glucose. *Diabetes Care*. 2013;36(8):2286-2293. doi:10.2337/dc12-2563

14. Mapes BM, Foster CS, Kusnoor SV, et al. Diversity and inclusion for the All of Us research program: A scoping review. *PloS One*. 2020;15(7):e0234962. doi:10.1371/journal.pone.0234962

15. Chandler PD, Clark CR, Zhou G, et al. Hypertension prevalence in the All of Us Research Program among groups traditionally underrepresented in medical research. *Sci Rep*. 2021;11(1):12849. doi:10.1038/s41598-021-92143-w

16. Fryar CD, Kit B, Carroll MD, Afful J. Hypertension Prevalence, Awareness, Treatment, and Control Among Adults Age 18 and Older: United States, August 2021-August 2023. *NCHS Data Brief*. 2024;(511).

17. Fong JH. Disability incidence and functional decline among older adults with major chronic diseases. *BMC Geriatr*. 2019;19:323. doi:10.1186/s12877-019-1348-z

18. Wang VHC, Lei J, Shi T, Pagán JA. Weighting the United States All of Us Research Program data to known population estimates using raking. *Prev Med Rep*. 2024;43:102795. doi:10.1016/j.pmedr.2024.102795

19. Agho AV, Jiwo HL, Oni TB, et al. Trends in Hypertension Diagnosis and Self-Reported Cases: A Retrospective Analysis of National Health Interview Survey (NHIS) Database. *Cureus*. 17(3):e80655. doi:10.7759/cureus.80655

20. Chang TE, Ritchey MD, Park S, et al. National Rates of Nonadherence to Antihypertensive Medications Among Insured Adults With Hypertension, 2015. *Hypertension*. 2019;74(6):1324-1332. doi:10.1161/HYPERTENSIONAHA.119.13616

21. Imoisili OE. Prevalence of Stroke — Behavioral Risk Factor Surveillance System, United States, 2011–2022. *MMWR Morb Mortal Wkly Rep*. 2024;73. doi:10.15585/mmwr.mm7320a1

22. Gwira JA, Fryar CD, Gu Q. Prevalence of Total, Diagnosed, and Undiagnosed Diabetes in Adults: United States, August 2021-August 2023. *NCHS Data Brief*. 2024;(516):CS354814. doi:10.15620/cdc/165794

23. Ojha N, Dhamoon AS. Myocardial Infarction. In: *StatPearls*. StatPearls Publishing; 2025. Accessed January 26, 2025. http://www.ncbi.nlm.nih.gov/books/NBK537076/

24. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7-33. doi:10.3322/caac.21708

25. Swed S, Sawaf B, Al-Obeidat F, et al. Asthma prevalence among United States population insights from NHANES data analysis. *Sci Rep*. 2024;14:8059. doi:10.1038/s41598-024-58429-5

26. U.S. Census Bureau QuickFacts: United States. Accessed June 2, 2025. https://www.census.gov/quickfacts/fact/table/US/PST045224#PST045224

27. Zhu Y, Sidell MA, Arterburn D, et al. Racial/Ethnic Disparities in the Prevalence of Diabetes and Prediabetes by BMI: Patient Outcomes Research To Advance Learning (PORTAL) Multisite Cohort of Adults in the U.S. *Diabetes Care*. 2019;42(12):2211-2219. doi:10.2337/dc19-0532

28. Ansah JP, Chiu CT. Projecting the chronic disease burden among the adult population in the United States using a multi-state population model. *Front Public Health*. 2023;10:1082183. doi:10.3389/fpubh.2022.1082183

29. Fakhouri TH, Martin CB, Chen TC, et al. An Investigation of Nonresponse Bias and Survey Location Variability in the 2017−2018 National Health and Nutrition Examination Survey. Accessed June 1, 2025. https://stacks.cdc.gov/view/cdc/156535

30. Schulkey CE, Litwin TR, Ellsworth G, et al. Design and Implementation of the All of Us Research Program COVID-19 Participant Experience (COPE) Survey. *Am J Epidemiol*. 2023;192(6):972-986. doi:10.1093/aje/kwad035

31. Davern M. Nonresponse Rates are a Problematic Indicator of Nonresponse Bias in Survey Research. *Health Serv Res*. 2013;48(3):905-912. doi:10.1111/1475-6773.12070

Appendix

Tables

Table 1.

Table 1: The Demographic Characteristics for the NHIS, NHANES and *AoU* Datasets			
	NHIS (N=1,784,471)	NHANES (N=303,667,876)	*AoU* (N = 633,547)
Sex			
Female	968,774 (54.3%)	155,130,208 (51.1%0	395,987 (63.14%)
Male	815,284 (45.7%)	148,537,668 (48.9%)	2,31,210 (36.86%)
Age			
18-24	111,677 (6.3%)	26,615,238 (11.2%)	15,015 (2.37%)
25-34	259,416 (14.5%)	42,176,199 (17.8%)	85,448 (13.49%)
35-44	275,413 (15.4%)	41,136,699 (17.3%)	102,929 (16.25%)
45-54	244,222 (13.7%)	34,147,526 (14.4%)	97,823 (15.44%)
55-64	313,722 (17.6%)	41,124,441 (17.3%)	122,632 (19.36%)
65+	580,021 (32.5%)	51,969,905 (21.9%)	209,700 (33.10%)
Race			
White	1,241,020 (69.5%)	184,910,840 (65.3%)	357,658 (56.45%)
Hispanic	225,930 (12.7%)	61,808,108 (20.4%)	112,751 (17.80%)
Black	179,877 (10.1%)	36,494,046 (12.0%)	99,788 (15.75%)
Asian	89,811 (5.0%)	20,454883 (6.7%)	22,400 (3.54%)
AIAN	27,802 (1.6%)	--	8,973 (1.42%)

Table 2.

Table 2: Rates and Prevalence Ratios of Chronic Conditions in the NHIS, NHANES and *All of Us*						
	NHIS	PR/95%CI*	NHANES	PR/95%CI*	*AoU*	PR/95%CI*
Hypertension						
Overall	0.32		0.24		0.27	
White	0.34	Ref (1.00)	0.18	Ref (1.00)	0.17	Ref (1.00)
Hispanic	0.25	0.73 (0.67,0.81)	0.16	0.66 (0.58,0.74)	0.07	0.42 (0.41,0.43)
Black	0.40	1.21 (1.11,1.32)	0.26	1.21 (1.09,1.34)	0.11	0.66 (0.65,0.67)
Asian	0.25	0.75 (0.65,0.87)	0.31	0.71 (0.57,0.86)	0.07	0.41 (0.39,0.43)
AIAN	0.36	0.92 (0.66,1.28)	--	--	0.07	0.40 (0.37,0.44)
Hypertension Medication						
Overall	0.25		--	--	0.87	
White	0.27	Ref.	--	--	0.87	Ref.
Hispanic	0.18	0.68 (0.60,0.76)	--	--	0.83	0.96 (0.95,0.97)
Black	0.33	1.22 (1.10,1.35)	--	--	0.91	1.05 (1.05,1.06)
Asian	0.20	0.75 (0.64,0.89)	--	--	0.82	0.94 (0.92,0.97)
AIAN	0.26	0.84 (0.53,1.34)	--	--	0.83	0.96 (0.93,1.00)
Stroke						
Overall	0.03		0.03		0.07	
White	0.03	Ref (1.00)	0.03	Ref (1.00)	0.06	Ref (1.00)
Hispanic	0.02	0.63 (0.43,0.92)	0.01	0.71 (0.47,0.95)	0.06	1.06 (0.97,1.17)
Black	0.04	1.17 (0.87,1.57)	0.03	1.08 (0.68,1.48)	0.13	1.89 (1.76,2.03)
Asian	0.02	0.49 (0.27,0.91)	0.03	0.60 (0.27,0.92)	0.03	0.49 (0.38,0.62)
AIAN	0.06	1.04 (0.41,2.63)	--	--	0.16	2.27 (1.84,2.80)
Prediabetes						
Overall	0.16		--	--	0.51	
White	0.14	Ref (1.00)	--	--	0.71	Ref (1.00)
Hispanic	0.17	1.21 (1.06,1.39)	--	--	0.64	0.91 (0.89,0.92)
Black	0.20	1.41 (1.23,1.60)	--	--	0.80	1.13 (1.11,1.15)
Asian	0.18	1.31 (1.07,1.60)	--	--	0.65	1.01 (0.97,1.05)
AIAN	0.20	1.34 (0.86,2.08)	--	--	0.71	1.00 (0.93,1.06)
Diabetes						
Overall	0.10		0.09		0.18	
White	0.09	Ref (1.00)	0.08	Ref (1.00)	0.17	Ref (1.00)
Hispanic	0.11	1.18 (1.00,1.40)	0.08	0.95 (0.78,1.13)	0.17	1.02 (0.99,1.05)
Black	0.13	1.31 (1.20,1.66)	0.08	1.55 (1.26, 1.85)	0.25	1.47 (1.43,1.51)

Table 2 cont.

Asian	0.10	1.17 (0.90,1.52)	0.12	0.73 (0.46,0.99)	0.11	0.67 (0.62,0.72)
AIAN	0.13	1.47 (0.92,2.34)	--	--	0.20	1.17 (1.06,1.28)
Myocardial Infarction						
Overall	0.03		0.03		0.07	
White	0.04	Ref (1.00)	0.02	Ref (1.00)	0.07	Ref (1.00)
Hispanic	0.02	0.40 (0.27,0.60)	0.01	0.43 (0.27,0.58)	0.05	0.91 (0.84,0.99)
Black	0.03	0.68 (0.49,0.94)	0.03	0.70 (0.41,0.99)	0.10	1.54 (1.44,1.67)
Asian	0.02	0.40 (0.20,0.81)	0.02	0.40 (0.15,0.64)	0.04	0.72 (0.59,0.87)
AIAN	0.05	1.11 (0.32,3.75)	--	--	0.12	1.79 (1.46,2.19)
Cancer						
Overall	0.10		0.09		0.08	
White	0.13	Ref (1.00)	0.03	Ref (1.00)	0.08	Ref (1.00)
Hispanic	0.04	0.28 (0.22,0.35)	0.03	0.26 (0.19,0.33)	0.06	0.72 (0.70,0.74)
Black	0.05	0.40 (0.31,0.50)	0.12	0.43 (0.33,0.53)	0.07	0.81 (0.79,0.83)
Asian	0.03	0.26 (0.17,0.42)	0.05	0.21 (0.12,0.31)	0.06	0.79 (0.74,0.84)
AIAN	0.08	0.47 (0.13,1.66)	--	--	0.07	0.85 (0.78,0.92)
Asthma						
Overall	0.15		0.15		0.37	
White	0.15	Ref (1.00)	0.18	Ref (1.00)	0.38	Ref (1.00)
Hispanic	0.13	0.83 (0.71,0.97)	0.13	0.94 (0.82,1.06)	0.38	0.89 (0.87,0.90)
Black	0.18	1.19 (1.02,1.38)	0.17	1.16 (1.00,1.33)	0.36	0.96 (0.94,0.98)
Asian	0.10	0.65 (0.51,0.84)	0.19	0.59 (0.42,0.77)	0.37	0.97 (0.92,1.01)
AIAN	0.22	1.56 (0.92,2.64)	--	--	0.36	0.96 (0.89,1.02)

*Prevalence Ratios and 95% Confidence Intervals

Figures

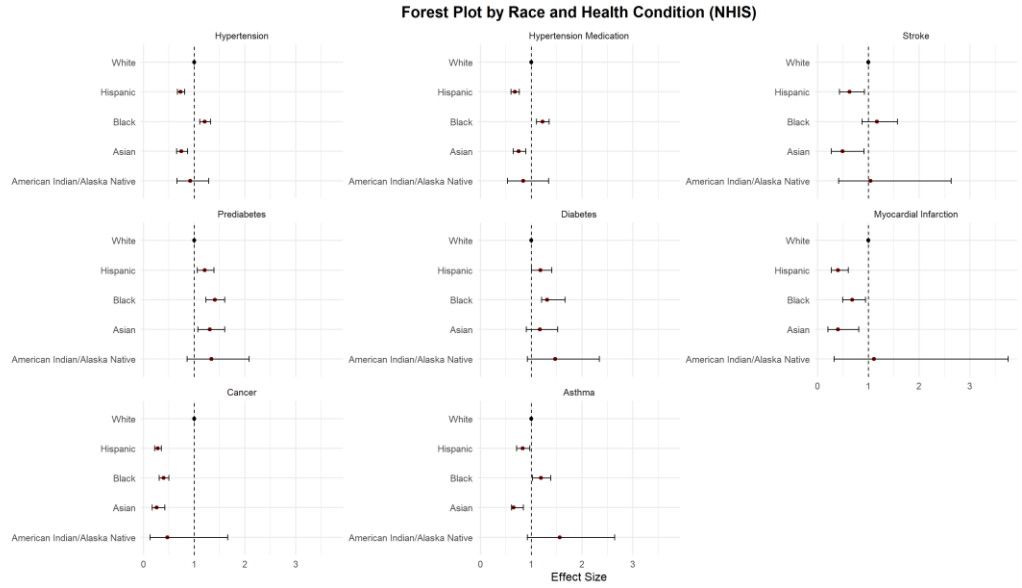

Figure 1. This shows prevalence ratios in the NHIS disaggregated conditions to better illustrate within each race. The red dots are the point estimates and the bar shows the confidence for each prevalence ratio point estimate.

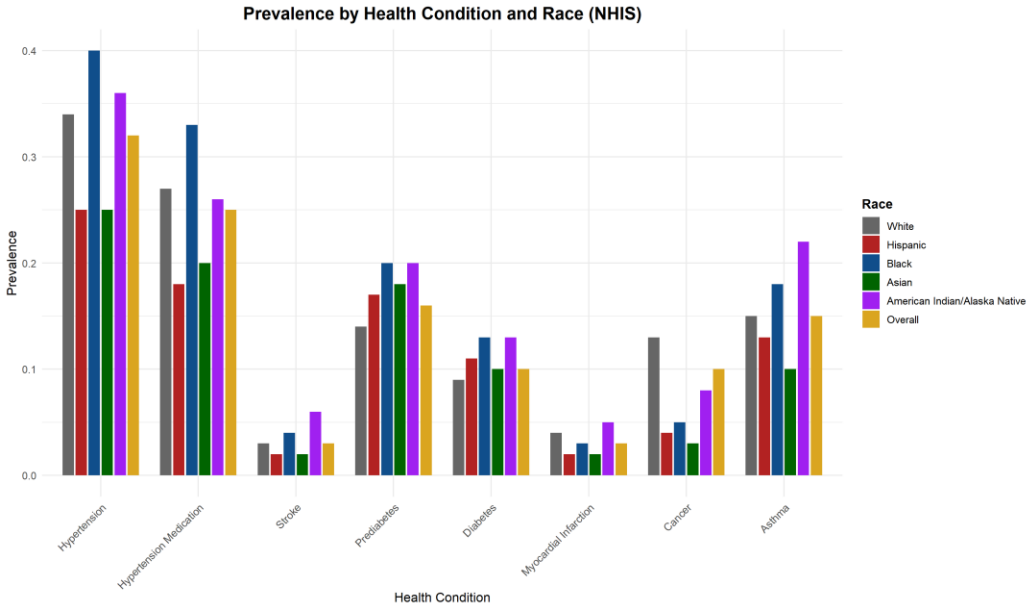

Figure 2. This shows the prevalences within each race stratified by condition in the NHIS dataset.

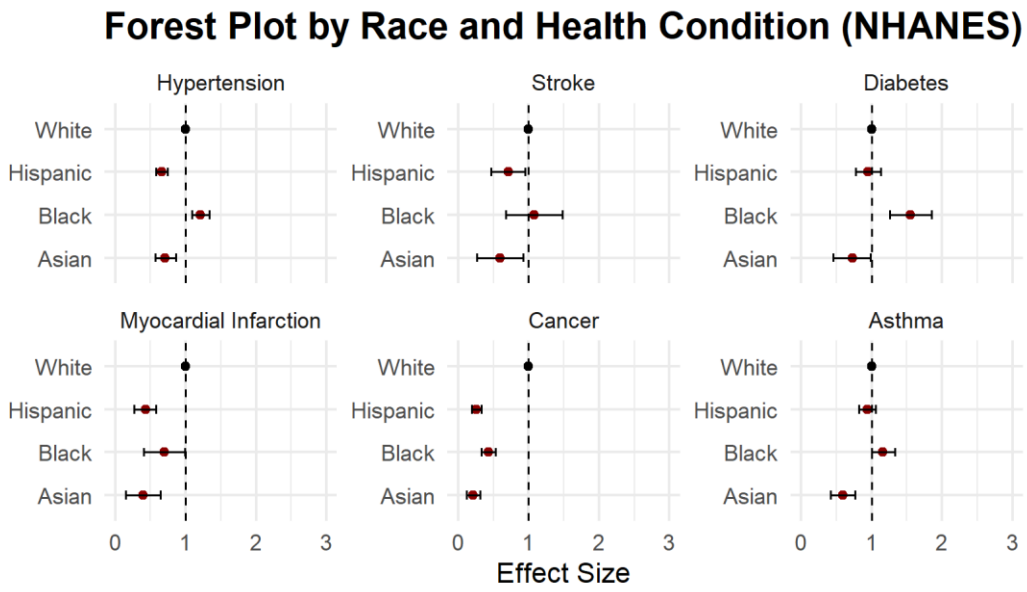

Figure 3. This shows prevalence ratios in the NHANES disaggregated conditions to better illustrate within each race. The red dots are the point estimates and the bar shows the confidence for each prevalence ratio point estimate.

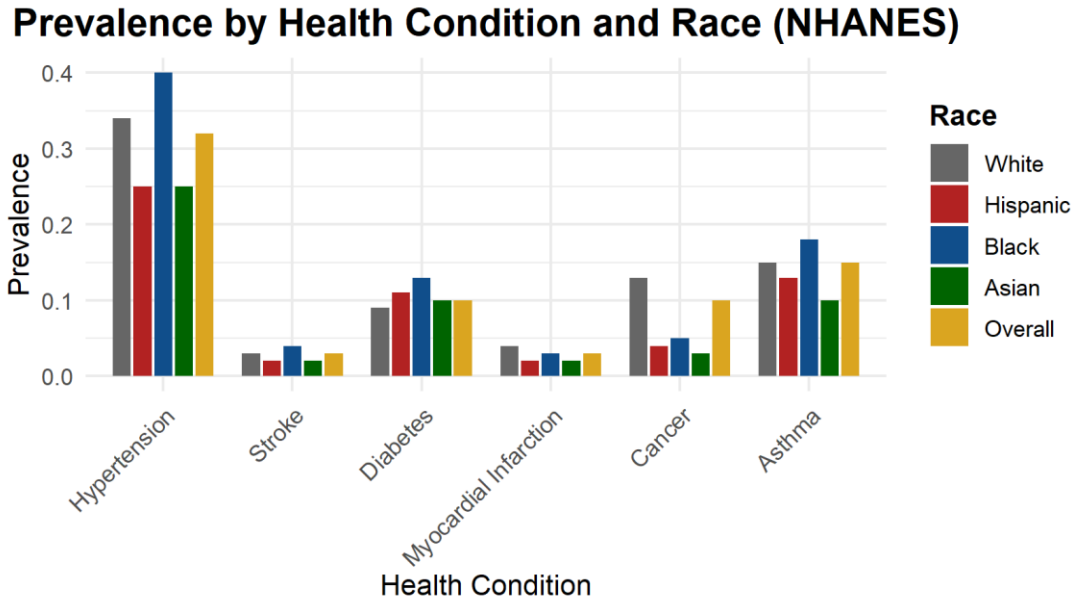

Figure 4. This shows the prevalences within each race stratified by condition in the NHANES dataset.

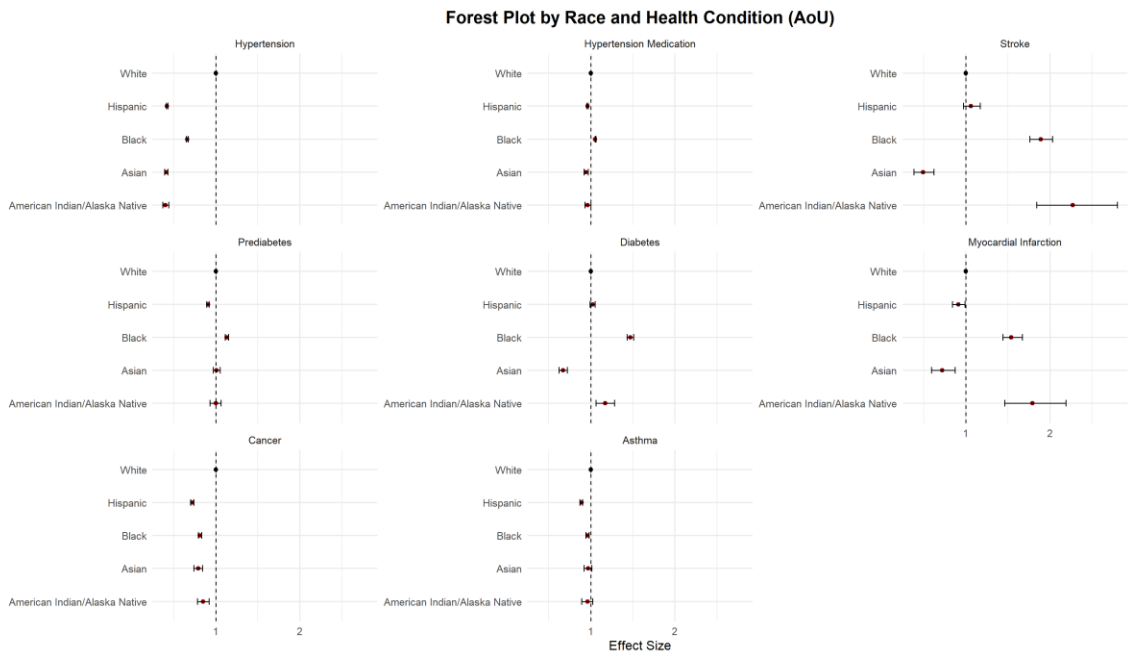

Figure 5. This shows prevalence ratios in the *AoU* disaggregated conditions to better illustrate within each race. The red dots are the point estimates and the bar shows the confidence for each prevalence ratio point estimate.

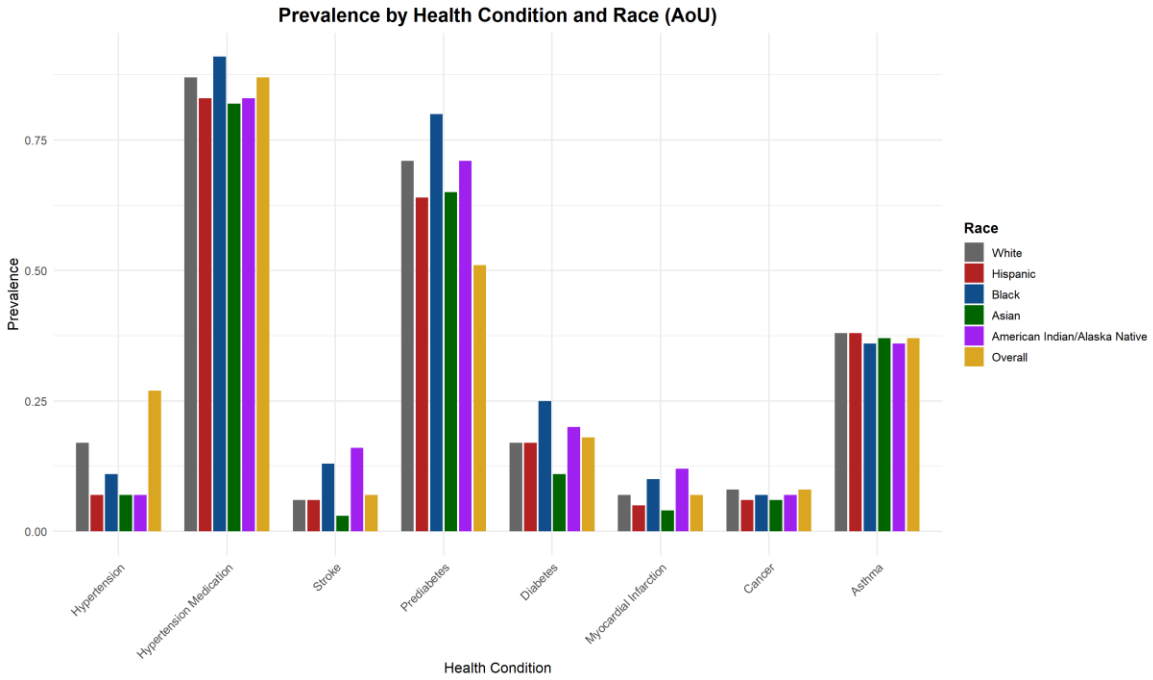

Figure 6. This shows the prevalences within each race stratified by condition in the *AoU* dataset.

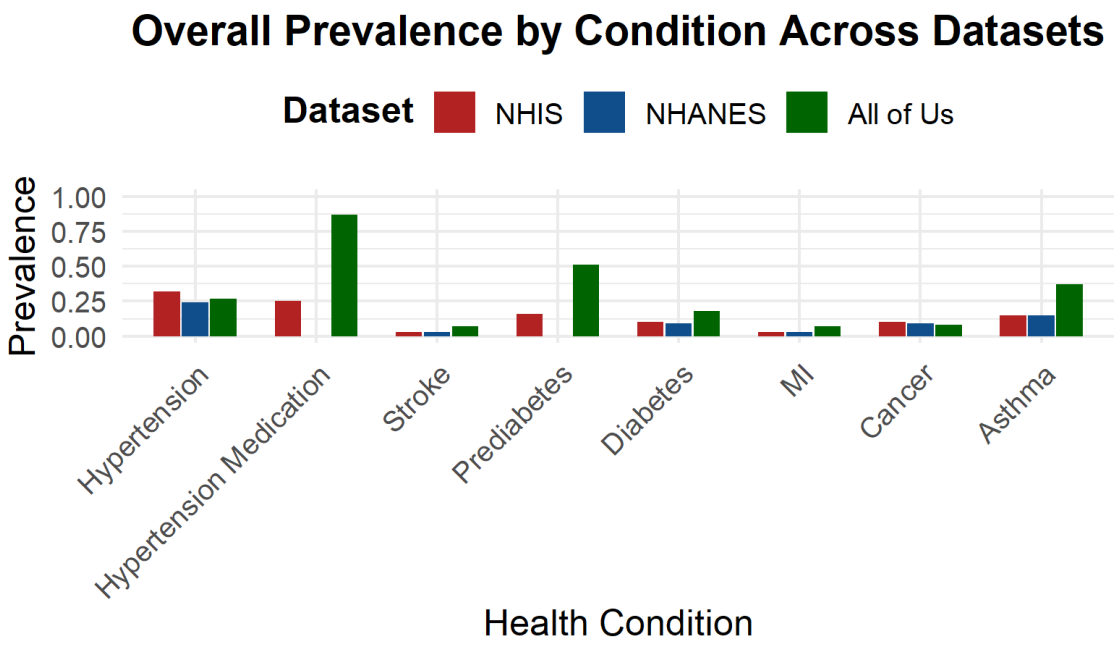

Figure 7. This shows the Overall Prevalence of each chronic condition stratified by each dataset.