## Title

Understanding the Role of Socioeconomic, Health Behavioral, and Genetic Factors in Cardiovascular Disease Risk



## Author

Schell, Robert C

## Publication Date

2023

Peer reviewed|Thesis/dissertation

# Understanding the Role of Socioeconomic, Health Behavioral, and Genetic Factors in Cardiovascular Disease Risk

By

Robert Schell

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Health Policy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor William H. Dow, Co-Chair
Professor Lia C.H. Fernald, Co-Chair
Professor Patrick T. Bradshaw
Professor David H. Rehkopf

Spring 2023

Understanding the Role of Socioeconomic, Health Behavioral, and Genetic Factors in Cardiovascular Disease Risk

Abstract

Understanding the Role of Socioeconomic, Biological, Health Behavioral, and Genetic Factors in Cardiovascular Disease Risk Disparities

by

Robert C. Schell

Doctor of Philosophy in Health Policy

University of California, Berkeley

Professors William H. Dow and Lia C.H. Fernald, Co-Chairs

The age-standardized incidence of cardiovascular disease in high-income countries has fallen precipitously from 844 cases per 100,000 in 1990 to 597 in 2019. However, this reduction in average incidence obscures significant disparities in risk by socioeconomic, genetic, biological, and health behavioral factors. To date, most studies of the risk factors for cardiovascular disease attempt to isolate the effect of a single risk factor. This single-exposure focus belies the dynamic interplay between genetics, socioeconomic status, and traditional health and health behavioral factors that ultimately determine an individual's cardiovascular disease risk. This dissertation explores the ways in which different domains of risk interact to produce heterogeneous, individualized risk of cardiovascular disease.

The second chapter of this dissertation employs Mendelian randomization, an instrumental variable technique that uses random genetic variation as the instrument, to understand the causal effect of adiposity on cardiovascular disease incidence for individuals at different levels of socioeconomic status. Both high adiposity and low socioeconomic status increase cardiovascular disease risk. However, it is unclear whether the risk caused by an increase in adiposity is itself identical for individuals at varying levels of socioeconomic status or if, instead, individuals at lower levels of socioeconomic status face a greater risk from an increase in adiposity than their peers. If such a disparity did exist, it would imply that the risk implications of weight gain may differ by socioeconomic status. In UK Biobank data, we find that differences in risk in these Mendelian randomization models existed only for individuals with versus without a university degree. The differences in cardiovascular disease risk from higher body mass index by educational attainment or income – if they exist - are small in magnitude, though imprecisely estimated. The results for waist-to-hip ratio adjusted for body mass index, a measure of central adiposity, generally showed even smaller differences between socioeconomic groups.

The third chapter of this dissertation explores whether a health behavior related to lower cardiovascular disease risk, physical activity, might ameliorate some of the risk of coronary artery disease caused by an individual's genetic predisposition. Specifically, I use wrist-worn accelerometer data combined with the most powerful polygenic risk score in the

literature to determine how physical activity volume and intensity impact an individual's risk of coronary artery disease at different levels of genetic risk. If individuals at higher risk benefit more from greater physical activity than their lower risk peers, this could imply that more personalized physical activity standards should be considered for at-risk patients. Physical activity volume and intensity each had significant independent associations with incident coronary artery disease, with physical activity intensity demonstrating the strongest association. However, no interactive effect of physical activity and genetic risk on coronary artery disease risk was found.

Future research should focus both on the impact of individual risk factors on cardiovascular disease risk and - particularly - how different risk factors and domains might contribute and interact to determine an individual's overall cardiovascular disease risk. With the proliferation of large-scale cohort datasets such as the UK Biobank, the Million Veteran Program, and All of Us, researchers face an unprecedented opportunity to understand how every aspect of a person's life from their social and built environment to their medical care access, genetic code, and health behaviors interconnect to determine disease risk.

# Table of contents

# List of figures

# List of tables

# Acknowledgments

I would like to start by thanking my dissertation chairs, Professors Will Dow and Lia Fernald. Will, your careful work and considered advice has had a profound impact on the type of researcher I aspire to become. The pre-analysis plan you encouraged me to create to keep me focused (still working on that!) had an immense impact on me. I have been fixated on conducting rigorous research with publicly available code since then and I consider this focus on transparency in everything one does the most fundamentally important aspect of being a good researcher. I admire the way you think through problems in research and have modeled my own process after your own, with your emphasis on understanding the assumptions - and there are always assumptions - behind the analytical decisions we make. Lia, thank you for keeping me grounded when I get too excited about applying a new research methodology I learned. You kept me on track to answer research questions of consequence instead of falling into the trap of answering unimportant questions extremely well. Thank you, too, for the immensely valuable time with you and your research group, who provided me vital feedback without ever being critical.

Thank you to my other dissertation committee members, Professors David Rehkopf and Patrick Bradshaw. David, you played a massive role both in providing the vitally needed subject matter expertise on my methods and my dataset without which this dissertation could not have been written. Thank you also for the many interesting conversations about the existing literature at the intersection of public health and genetics and for your patience when I would occasionally miss our 8 am meeting time. Thank you, Patrick, both for guiding me with your knowledge of survival analysis but also for letting me think through problems on the fly with you. I appreciate the variety of resources you provided me, the time and engagement with the many questions I posed to you, and - perhaps most importantly - that you often served as my north star when I wanted to figure out if a simulation or result seemed plausible.

I would also like to thank my peers in the Health Policy program. We all have such idiosyncratic interests that discussing research can be challenging but you provided a much-needed ear when I wanted to rant and I always appreciated both the time we spent together and the non-competitive atmosphere, which I have on good authority is rarely the case in PhD programs. Thank you particularly to Lakshmi, without whom I would surely not have graduated in four years, and my mental health would have been in a far worse state.

Lastly, thank you to my friends and family and especially my fiancé, Lindsay. You were in the unusual position of having to get me to slow down, take breaks, and do things outside of research and you performed that task with aplomb. We've gone through an incredible amount over the past several years and I know without a doubt I would not be in this position without you. I'm looking forward to supporting each other in our journey together regardless of what comes next.

# Dedication

To Lindsay, Bill, David Allen, Nance, and Mr. Schell, who patiently spent the past three years hearing about this project but had the grace to keep listening.

# 1 Introduction

Cardiovascular disease is the leading cause of death worldwide. While the emergence of treatments such as statins and hypertensive medications has greatly reduced mortality, the risk of developing this disease is largely determined beyond the walls of a doctor's office (1). The American Heart Association advocates focusing on "Life's Simple 7," a set of behaviors including not smoking, maintaining a healthy bodyweight, sufficient physical activity, a high-quality diet, and low levels of cholesterol, blood pressure, and fasting glucose without medication use, to lower the incidence of cardiovascular disease (2). However, adherence to these principles is low among the American public with only 13% of adults meeting 5 of the 7 criteria (2). Optimal physical activity and ideal bodyweight, which serves as a proxy for adiposity, have especially low rates of adherence, with under half of Americans meeting physical activity guidelines and over 70% of Americans above an ideal bodyweight according to body mass index (3)(4). These low rates of adherence become even more concerning when coupled with risk factors such as an individual's socioeconomic status and genetic susceptibility that may further increase the risk of developing cardiovascular disease.

This dissertation explores how adiposity, socioeconomic status, genetics, and physical activity interact to produce cardiovascular disease risk. Chapter two focuses on whether socioeconomic status - and particularly educational attainment - alters the causal effect of adiposity on cardiovascular disease. Presently, we have a limited understanding of how adiposity and low socioeconomic status interact to contribute to an individual's cardiovascular disease risk. Chapter two demonstrates their combined impact via a Mendelian randomization survival analysis, which will demonstrate whether an increase in adiposity causes a similar increase in risk for individuals at varying levels of socioeconomic status. Chapter three focuses on the degree to which physical activity volume and intensity can mitigate a high genetic risk of coronary artery disease. This chapter hypothesizes that if individuals at high genetic risk benefit more from greater physical activity, a targeted intervention approach focusing on these high-risk individuals may be warranted. The rest of this introduction outlines the rationale of chapters two and three.

## 1.1 Overview of Chapter 2

Because a randomized controlled trial is neither ethical nor possible to study the effects of socioeconomic status and adiposity on cardiovascular disease risk, researchers must rely on observational studies, for which endogeneity, or correlation between the exposure and error term, obscures the underlying causal relationship. Endogeneity results from two issues: omitted variable bias (or unmeasured confounding) and reverse causation. Omitted variable bias occurs because of persistent unmeasured differences between people of different levels of adiposity. Reverse causation can occur, for instance, if a cardiovascular event causes disability that results in weight gain. To isolate the causal effect of adiposity on cardiovascular disease risk in different socioeconomic strata, I

employ Mendelian randomization, which can solve this endogeneity problem provided its core assumptions are met.

The basic idea behind Mendelian randomization is that an individual randomly inherits one allele from each parent at every single nucleotide polymorphism, a binary genetic variant where an individual could receive one of two alleles from each parent, which makes the genetic variation random conditional on parental genotype (5). Because they are randomly determined, these genetic variants can, in turn, serve as valid instrumental variables for adiposity subject to the three traditional instrumental variable assumptions: relevance, exchangeability, and the exclusion restriction (denoted in order by the arrows labeled 1,2, and 3 in **Figure 1.1**) (6).

Numerous Mendelian randomization studies have explored the causal link between adiposity and cardiovascular disease. Early studies relied on smaller datasets and a relatively small number of genetic variants that accounted for under 2% of the variation in adiposity. Despite the resultant low statistical power, these studies often found a strong association between an increase in adiposity and an increased risk of coronary heart disease, heart failure, and ischemic heart disease, with mixed results for stroke (7-10). More recent Mendelian randomization studies have explored the association between adiposity and cardiovascular disease and, while the percent of variation in adiposity explained by genetic variants remains largely unchanged, the datasets are far larger, which gives the analyses more statistical power. These better powered studies have reinforced the strength of the association between adiposity and cardiovascular disease, with a standard deviation increase in waist-to-hip ratio associated with a 1.46 increase in the odds of coronary heart disease (11-17). However, all of these studies either control for or do not consider the role of socioeconomic status in producing cardiovascular disease risk.

Socioeconomic status, a multidimensional concept that includes an individual's income, educational attainment, occupational status, environment, and social environment, has long been understood as a "fundamental cause" of health disparities in high-income countries.[32] The view of socioeconomic status as a stationary attribute to abstract away from instead of as a key contributor in the production of health that works in conjunction with traditional risk factors in shaping an individual's health has kept the literature from exploring how socioeconomic status modifies health risks. Traditionally, socioeconomic status and adiposity are both thought of as having their own independent (and stable) average treatment effects and as simple confounders instead of as interacting factors in the production of health. However, socioeconomic status is known to impact both biological risk factors for cardiovascular disease such as hypertension and diabetes and social risk factors such as access to medical care and health knowledge that could interact with adiposity's own effects on cardiovascular disease to further increase risk (30). This study is the first to view socioeconomic status as an effect modifier – not confounder – of the impact of adiposity on cardiovascular disease risk.

**Research Question for Chapter 2:** Does the effect of an increase in adiposity on cardiovascular disease risk vary by socioeconomic status among the adults in the UK Biobank?

**Hypothesis:** People of higher socioeconomic status face a lower increase in cardiovascular disease risk from high adiposity than their lower status peers.

**Key Dependent and Independent Variables and Covariates**

*Incident Cardiovascular Disease Case Definition*

The main outcome of interest in this study is incident cardiovascular disease, for which I included angina, myocardial infarction, ischemic heart disease, atrial fibrillation, heart failure, and stroke with the following ICD-10 codes: I20, I21, I25, I48, I50, I60, I61, I63, I64.

*Socioeconomic Status*

I utilized two different measures of self-reported socioeconomic status in the analysis, which reflect the different dimensions encompassed by the term: educational attainment and household income. I treated educational attainment as a categorical variable separated into less than a high school education compared to at least high school graduate given the strong discontinuity in cardiovascular disease risk in the literature at this point. Household income is discretized to above or below the poverty level, owing to the importance of this threshold in the literature.

*Anthropometry*

I used measured waist-hip-ratio adjusted for body mass index (WHRadjBMI) and body mass index (BMI) to reflect adiposity. While most analyses focus on BMI alone to measure adiposity, it is a controversial and imprecise measure that fails to account for fat free mass. The WHRadjBMI, on the other hand, which measures central adiposity, has a far more consistent association with both adiposity and cardiovascular disease risk than BMI (18). Therefore, I used both WHRadjBMI and BMI as measures of central and total adiposity (18-20).

*Genetic Variants Selected*

I selected the genetic variants associated with an elevated WHRadjBMI based on a recent genome-wide association study (GWAS) performed by the GIANT Consortium. Shungin *et al.* measured WHRadjBMI among 210,088 individuals of European ancestry (21). The authors found 49 independent single nucleotide polymorphisms that reached genome-wide significance of $p<1x10^{-8}$, the standard critical value in GWAS studies that comes from multiple testing correction. For BMI, I focused on the most recent GWAS that excludes the UK Biobank, which identified 97 significant variants (22).

## 1.2 Overview of Chapter 3

Physical activity has a well-established influence on coronary artery disease risk. Professional organizations from the World Health Organization to the American Heart Association advocate for a minimum of 150 minutes of moderate intensity physical activity or 75 minutes of vigorous intensity physical activity for adults aged 18 to 64 to minimize the risk of cardiovascular disease (23-24). However, it is unclear whether heterogeneity in the ideal level of physical activity exists for different people based on their underlying risk. Could a person genetically predisposed toward coronary artery disease meaningfully benefit from engaging in physical activity beyond the blanket standards currently recommended? Chapter 3 focuses on the possibility that genetic risk of coronary artery disease and objectively measured physical activity volume and intensity may interact to alter coronary artery disease risk.

This study provides two vital innovations to address this question. Firstly, until recently physical activity in large cohort studies has been measured exclusively via self-reported questionnaires. This measure of physical activity suffers from the biases typical of self-reporting, including most notably recall and social desirability bias. The UK Biobank, the dataset used for this chapter, tracked physical activity over the course of a week with wrist-worn accelerometers for over 100,000 participants. As a result, for the first time in a large cohort study, objectively measured physical activity is available. While this method has its own limitations, previous studies have shown that accelerometers provide a more accurate picture of actual physical activity undertaken than questionnaire-based assessments. Secondly, our understanding of the genetic variants associated with coronary artery disease has increased substantially over the past ten years with the proliferation of both genome-wide association studies and increasingly sophisticated techniques to create polygenic scores that summarize an individual's overall genetic risk of a disease.

**Research Question for Chapter 3:** Can greater accelerometer-measured physical activity volume or intensity offset a high genetic predisposition to coronary artery disease among the adults in the UK Biobank?

**Hypothesis:** People of higher genetic risk benefit more from an increase in physical activity volume and intensity than their lower risk peers.

**Key Dependent and Independent Variables and Covariates**

*Coronary Artery Disease Case Definition*

The main outcome of interest was incident coronary artery disease. I defined cases based on hospital inpatient episodes, surgeries, and deaths using ICD-10 codes I20 to I25, I46, and R96 to determine CAD for cause of death, ICD-10 codes I20.0, I21-I22, and ICD-9 codes 410 and 4110 for an event in hospital inpatient records, and OPCS-4 codes K40 to K46, K49, K501, K75 and OPCS-3 code 3043 for surgeries.

*Polygenic Score*

I applied the best performing polygenic risk score to date that did not include the UK Biobank to avoid the potential for winner's curse bias. This score consisted of 1,090,048 variants, from which I excluded 332 by utilizing a more stringent minor allele frequency and INFO score requirement than the original authors.

*Physical Activity Volume and Intensity*

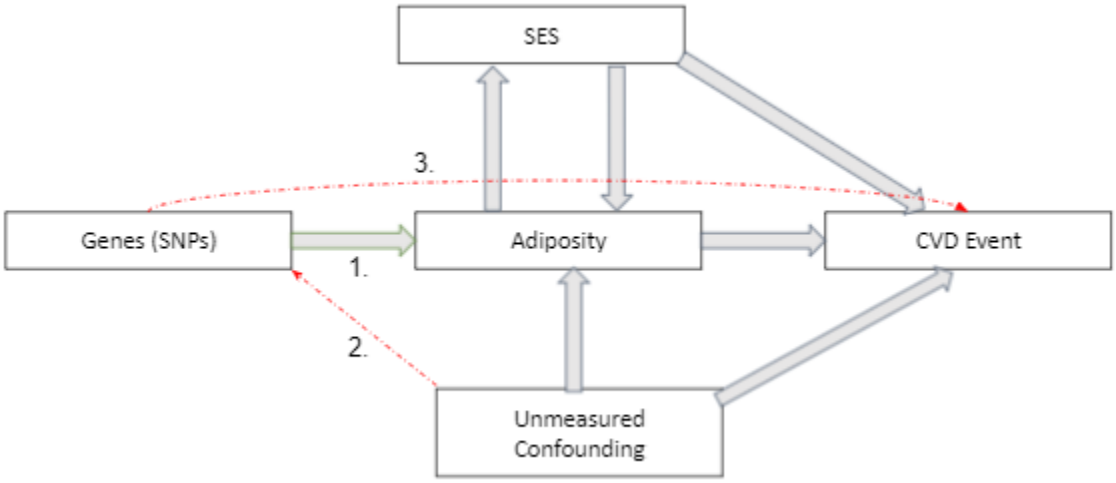
Following the work of Dempsey, *et al*, I converted the raw accelerometer Euclidean norm minus one (ENMO) from the wrist-worn accelerometer data to physical activity energy expenditure to represent physical activity volume (25). I then calculated the percent of this physical activity energy expenditure derived from moderate to vigorous physical activity as the percent of physical activity energy expenditure that took place above 125 milligravities.

## 1.3 Data

The UK Biobank (UKB) is a massive population-based cohort of over 500,000 people aged 40-69 years old at recruitment in 2006 from England, Scotland, and Wales with demographic, genetic, biomarker, detailed health status, socioeconomic status, anthropometric measurement, and disease data (26). This dataset is unparalleled both in its size and in the breadth of variables measured longitudinally since 2006. Participants in the UK Biobank were genotyped using either the UK BiLEVE or the UK Biobank Axiom Array, which each genotyped over 800,000 single-nucleotide polymorphisms (SNPs). Using either the Haplotype Reference Consortium panel or the UK10k and 1000 Genomes phase 3 panels, additional SNPs were imputed, yielding roughly 96 million variants assayed or imputed. The UKB contains detailed demographic, genetic, biomarker, health status, socioeconomic status, anthropometric, and disease data. Almost all of the enrolled subjects (> 96.9%) had valid measures of waist-to-hip ratio, genetic data, and social factors, which makes it an ideal resource to answer the question in chapter two. Between 2013 and 2015, participants with an email address were invited except those in the North West region due to concerns about participant burden. A subsample of 103,712 individuals responded to an email recruiting them to wear a wrist-worn Axivity AX3 triaxial accelerometer continuously for seven days on their dominant wrist and provided data (26). This group of participants constitutes the largest cohort with registered accelerometer data I am aware of and will serve as the study population for chapter three.

## 1.4 Tables and Figures

Figure 1.1: Demonstration of Mendelian randomization analysis and necessary assumptions



# 2 Does educational attainment modify the causal relationship between adiposity and cardiovascular disease? A Mendelian randomization study


## Abstract

A greater risk of cardiovascular disease is associated with low educational attainment and high adiposity. Despite the correlation between low educational attainment and high adiposity, whether educational attainment modifies the risk of CVD caused by high adiposity remains poorly understood. We investigated the effect of adiposity (body mass index [BMI] and waist-to-hip ratio adjusted for BMI [WHRadjBMI]) on incident CVD among individuals with varying education levels, using associational and one-sample Mendelian randomization (MR) survival analyses. Data were collected from 2006 to 2021, and sample sizes were 254,281 (27,511 CVD cases) for BMI and 253,968 (27,458 CVD cases) for WHRadjBMI. In the associational model, a standard deviation (SD) higher BMI was associated with 19.81 (95% CI: 18.55–21.06) additional cases of incident CVD per 10,000 person-years for individuals with a secondary education, versus 32.96 (95% CI: 28.75–37.17) for those without. When university degree served as the education variable, education group differences attenuated, with 18.26 (95% CI: 16.37–20.15) cases from a one SD higher BMI for those with a university degree versus 23.18 [95% CI: 21.56–24.72] for those without. For the MR model, an SD higher BMI resulted in 11.75 (95% CI: −0.84-24.38) and 29.79 (95% CI: 17.20–42.44) additional cases of incident CVD per 10,000 person-years for individuals with versus without a university degree. WHRadjBMI exhibited no effect differences by education. While the associational model showed evidence of educational attainment modifying the relationship between adiposity and incident CVD, it does not modify the association between adiposity and incident CVD in the MR models. This suggests either less education does not cause greater risk of incident CVD from high adiposity, or MR models cannot detect the effect difference. The associational point estimates exist within the MR models' confidence intervals in all BMI analyses, so we cannot rule out the effect sizes in the associational models.


## 2.1 Introduction

One of the great public health achievements ever has been the substantial reduction in cardiovascular disease (CVD) incidence and mortality in high-income countries (27-29). However, these reductions were not equally distributed, and people from lower socioeconomic strata, for example as defined by lower educational attainment, and people with higher levels of adiposity still face a disproportionately high risk of having a CVD event (30-34). A recent meta-analysis of observational studies found that people with a high school education or less faced a 27% to 50% greater risk of a CVD event, while controlling for body mass index (BMI) (34). Just as low educational attainment is related to a substantial increased risk of CVD, so too is high adiposity. A recent Mendelian

randomization (MR) analysis showed that a standard deviation higher waist-to-hip ratio adjusted for BMI (WHRadjBMI) is associated with an odds ratio of 1.46 for coronary heart disease (CHD), which reinforces decades of evidence from observational studies (17,35-36).

These disparities in CVD risk are especially concerning given the high prevalence of obesity, at 28% in the UK, and the fact that over half of UK residents never graduate college and 21% do not have a secondary education (37-39). Low educational attainment and high adiposity also tend to coexist, with one study finding that 22.5% of Europeans with only primary school education had obesity compared to 9.9% of individuals with a university degree (40). Despite their high prevalence and tendency to co-occur, the degree to which low educational attainment modifies the risk of CVD caused by high adiposity remains poorly understood.

MR is an instrumental variable technique that relies on random genetic variation as a natural experiment and, if the core assumptions are met, provides a causal effect robust to reverse causation and confounding bias. The elimination of reverse causation is especially important when considering adiposity as an exposure because of its tendency to blur the association of illness and adiposity in older age (41). While a person's adiposity derives from both genetic and lifestyle factors, MR provides a unique opportunity to isolate the health effects of adiposity from other confounding health behaviors, such as smoking, that could affect CVD risk and adiposity. Recent MR studies on adiposity's effect on CVD improve on the methodological limitations of the earliest studies, but they still largely treat educational attainment as a confounder for which to adjust (7,13-14,42). However, two people with identical levels of adiposity and different levels of education may face different associations between adiposity and CVD risk. Differences in incident CVD risk from adiposity could exist between educational attainment groups because of differences in medical care access and utilization, alcohol consumption patterns, and from higher levels of inflammation, hypertension, and hyperlipidemia (30,31,33,43-47). This study is the first to directly explore how different levels of education modify the effect of adiposity on incident CVD.

We estimate the relationship between adiposity and incident CVD at different levels of education via an MR survival analysis. We hypothesize that an increase in adiposity leads to a greater incidence of CVD among adults without a secondary education in the UK Biobank (UKB) compared to the risk faced by their better educated peers, as shown in the DAG in **Figure 2.1**. We also explore potential effect heterogeneity for individuals with more or less than a university degree as a sensitivity analysis.

## 2.2 Methods

The UKB is a population-based cohort of over 500,000 people from England, Scotland, and Wales aged 40-69 years old at recruitment designed to explore the genetic and environmental determinants of a variety of diseases (48). It began in 2006 and is uniquely suited to perform a One Sample MR survival analysis because of its size and the wide array of data collected longitudinally (26). We included only a subset of the full UKB

dataset based on the criteria outlined in **Figure 2.2**. MR requires us to restrict to unrelated individuals with high-quality genetic data. Because of the possibility of population stratification, or spurious associations which can occur if a disease and a genetic variant are more or less common in a specific ancestry group, we must also restrict to only individuals of white British ancestry. Pooled analysis of ancestry groups would create these spurious associations and there are simply too few non-white subjects in the UK Biobank to produce separate, well-powered analyses in these other ancestry groups. We describe briefly the criteria in **Figure 2.2** below.

First, we dropped any subjects who were not genotyped (14,239 subjects). Because subjects of European ancestry are the largest group by far in the UK Biobank (comprising over 94% of the sample) and population stratification concerns prevent us from analyzing an ancestrally diverse sample, we then dropped all individuals without European ancestry (78,620 subjects). Next, we removed individuals who withdrew consent to continue participating in data collection (96 subjects). We next dropped all individuals related, defined as third degree relatives or closer, to at least one other person in the UK Biobank to avoid creating spurious associations caused by familial effects (131,927 subjects) (26,49).

We excluded individuals whose genetic data failed the standard inclusion quality control procedures created by the MRC Integrative Epidemiology Unit (731 subjects) (50). In short, this excludes individuals who have a mismatch between genetically inferred and reported sex, duplicates, and individuals who are outliers in terms of heterozygosity or missing rates. We excluded individuals with prevalent cardiovascular disease at baseline, defined as those who experienced a CVD event prior to their first adiposity measurement in the UKB (19,259 subjects) and those with missing adiposity measurement (1,278 subjects) or educational attainment status (1,979 subjects). The final sample size of the study was 254,281 participants with 27,511 cases of incident CVD.

### 2.2.1 Incident Cardiovascular Disease Case Definition

The main outcome of interest in this study is time to incident cardiovascular disease (CVD), defined as including angina, myocardial infarction, ischemic heart disease, heart failure, and stroke with the following ICD-10 codes (and OPCS-4 codes for operations): I20 to I25, I44, I50, I60 to I64, I69, K40 to K46, K49, K50, K75 (51,52). If ICD-10 codes were unavailable for an individual, we relied on the corresponding self-reported events. We focused on incident instead of prevalent CVD, so only CVD cases that occurred after the first adiposity measurement for a subject who has not yet experienced CVD qualified as events.

### 2.2.2 Educational Attainment Definition

We treated education as a self-reported categorical variable separated into individuals who did not complete secondary education and those that did. In secondary analyses, we also split education into those with and without a university degree. We chose these

dichotomizations of educational attainment because past studies have shown significant differences in CVD risk between these educational groups (30,31,34).

### 2.2.3 Adiposity Definition

We used first-measured WHRadjBMI, defined as the ratio of the circumference of the waist compared to the hip controlling for BMI and first measured BMI as the two measures of adiposity in this study. This first adiposity measurement occurred at enrollment for all of the individuals in the present study. The benefit of using WHRadjBMI is that it has a more direct correlation with adiposity and measures central adiposity, which serves as an independent predictor of CVD risk beyond total body adiposity (53)(20). We stratified the WHRadjBMI analysis by sex due to the sexually dimorphic nature of the exposure (21). Therefore, the BMI models are presented pooled by biological sex, while the WHRadjBMI models are presented stratified by sex. Both exposures were standardized by subtracting the mean from each individual and dividing by the standard deviation.

### 2.2.4 Selected Genetic Variants

We identified 97 variants significantly associated with BMI that explain 2.7% of its variance from the most recent genome-wide association study (GWAS) that excluded the UKB (22). We selected the 49 variants associated with an elevated WHRadjBMI at genome-wide significance based on a recent GWAS performed by the GIANT Consortium (21). Because genetic variants are often co-inherited and inferring output from genotyping platforms can be difficult, imputation is a common technique used to allow a person's available genetic data to predict any missing genotypes. Imputation enables us to both keep individuals with incomplete genetic data, while also ensuring we can draw accurate inference from this population. The UK Biobank has developed their own imputation methods, detailed in the documentation at https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/impute_ukb_v1.pdf. To assess the quality of the UKB's imputation for our variants, we rely on INFO score, a measure of genetic variants' imputation quality scaled between 0 and 1, with a 1 denoting perfect imputation. Every genetic variant in this analysis had an INFO score higher than 0.95, which signifies the high level of imputation accuracy in this analysis. For individuals with imputed SNPs, we perform hard-call genotyping, which assigns variants to the most likely allele count between 0, 1, and 2 based on an individual's related genetic information.

We next checked for the existence of multi-allelic or palindromic SNPs. In brief, a multi-allelic SNP is one with more than the conventional two alleles, which would make applying any of our estimators impossible. Palindromic SNPs are those with the same letters on the forward and reverse strands, which makes identifying the effect allele challenging, particularly if the effect allele frequency is roughly 0.5. We identified no multi-allelic but two palindromic SNPs:rs1558902 (16:52,361,075) and rs9641123 (7:93,035,668) for BMI. We reran these models without the two above SNPs, which did not affect the results. No such variants existed for WHRadjBMI.

These genetic variants can serve as valid instrumental variables for adiposity subject to three assumptions: relevance, exchangeability, and no horizontal pleiotropy (54). Relevance, which requires that the variant influences adiposity, represents the only empirically verifiable assumption. We verify relevance by determining whether the instruments combine to have a partial F-statistic over 10, a conventional threshold for instrument strength. Exchangeability, or a lack of confounding between the outcome and variant, seems plausible in this scenario given the random assignment of genes at birth, although population stratification could violate it. The last assumption, no horizontal pleiotropy, is the most contentious and likely to fail in practice. This happens when a variant affects CVD risk both through adiposity and some other exposure.

### 2.2.5 Statistical Analyses

The following analyses are stratified by educational attainment, with subjects grouped as those with versus without a secondary education and those with and without a university degree. All analyses consist of an associational model and an MR model. We used Aalen's additive hazards model for both the MR and associational survival analyses. We chose the additive hazard model instead of the Cox model to avoid the issue of non-collapsibility of the hazard ratio as a measure of association (55,56). In a survival model with death as a competing risk, the quantity estimated is a cause-specific hazard difference. The timescale used is time since first measured adiposity. Time to event, the outcome of interest, signifies time from first measurement of adiposity to first incident CVD event. We controlled for baseline age, genotyping array, whether a subject ever smoked (defined as over 100 lifetime cigarettes), the first ten genetic ancestry principal components as a standard way to further control for population stratification confounding, and biological sex in the BMI models (the WHRadjBMI models stratify by sex). We relied on complete case analysis due to low rates of missingness in the UKB with only 3257 individuals missing exposure or outcome values at the final stage of screening.

For the MR analyses we first performed inverse variance weighted (IVW) regression, which is the most efficient MR estimator because it more heavily weights variants with more precise effects on adiposity and, thus, CVD incidence (57). We utilized a fixed effects modeling approach, which assumes one underlying "true" effect of adiposity on CVD incidence in each educational attainment stratum. While the IVW regression provides the most statistical efficiency of any MR design, it is also the most susceptible to bias due to horizontal pleiotropy (58).

We performed sensitivity analyses that vary the assumptions underlying the model to make the causal effect more plausible. The weighted median estimator is an alternative model utilized in many MR analyses that offers less efficiency than the IVW estimator but provides robustness to certain forms of horizontal pleiotropy (58). The weighted median estimator is consistent provided at least half of its weight is placed on variants acting as valid instruments. The causal effect is then the effect of the variant at the 50th percentile of the weights (58). Because the MR estimators do not allow formal

assessments of effect modification, we performed stratified analyses and evaluated confidence intervals and point estimates. While non-overlapping confidence intervals imply statistically significant differences, their overlap does not imply that no difference exists and so we only interpreted confidence intervals in the cases where they diverged or overlapped extensively, thus our conclusions should be conservative (59). All analyses were performed using R 4.1.3. Because individuals with at least (less than) a secondary education and with (without) a university degree overlap, we focused on comparisons between mutually exclusive groups. We pre-specified all analyses and hypotheses on Open Science Framework. This study follows the STROBE-MR guidelines (60).

## 2.3 Results

Out of the 254,281 (253,968) subjects that fit the inclusion criteria for the BMI (WHRadjBMI) analysis, 27,511 (27,458) experienced a CVD incident over 2,970,344 (2,969,741) person-years. Subjects with more education had lower BMIs and incidence of CVD, were less likely to smoke, and were generally younger at baseline than their peers with less education (**Tables 2.1 and 2.2**). The genetic variants displayed relationships of adequate strength according to our pre-analysis plan for every group except for subjects with less than secondary education and males with at least a university degree for WHRadjBMI, as the partial F-statistic for each of the rest exceeded 10. It is important to note that this instrument strength estimate was derived from estimating a first-stage regression with the genetic variants as variables and the confounders, which differs from the models used in the results. **Figures 2.2 to 2.4** present a detailed breakdown of follow-up time by event, Kaplan-Meier curves for overall survival probability, and cumulative incidence by event type.

In the adjusted model in **Table 2.3** and **Figure 2.5**, a standard deviation higher BMI (4.69 kg/m2) results in 22.60 (95% CI: 21.39 to 23.82) excess cases of incident CVD per year among 10,000 individuals in the pooled sample. This effect increases to 32.96 (95% CI: 28.75 to 37.17) for individuals with less than a secondary education compared to only 19.81 (95% CI: 18.55 to 21.06) for individuals with at least a secondary education. Given the over 33% increased rate of incident CVD per standard deviation increase in BMI between individuals in these groups and their non-overlapping confidence intervals, it appears education may act as an effect modifier in the associational models. Individuals with a university degree and those without one also diverge in terms of hazard of CVD, but the effect size difference is far smaller in magnitude. These differences do not appear to exist for WHRadjBMI for either sex, as the educational groups' effect sizes largely overlap.

In the IVW model results, shown in **Figure 2.6** and **Table 2.3**, the differences between educational groups for BMI appear less pronounced. For individuals with less than a secondary education, the relatively small sample size results in imprecise estimates and a confidence interval that reaches below the lowest and above the highest point of any other educational group. The point estimate for individuals with a university degree is less than half of the next closest point estimate and appears to differ meaningfully from individuals with less than a university degree. A standard deviation higher BMI results in

29.79 (95% CI: 17.20 to 42.44) excess cases of incident CVD per year among 10,000 individuals in the sample of individuals with less than a university degree compared to only 11.75 (95% CI: -0.84 to 24.38) for individuals with a university degree. **Figure 2.7** and **Table 2.3** suggest that the weighted median estimator is too imprecise to draw firm conclusions about differences in hazard between educational groups.

The IVW model results for WHRadjBMI differed from our hypothesis. While the results were difficult to distinguish for males, with all groups' confidence intervals overlapping with zero, females experienced a decrease in the hazard of incident CVD from a standard deviation higher WHRadjBMI. Specifically, the model suggests that a standard deviation higher WHRadjBMI results in a decrease in incidence of CVD of 31.12 per 10,000 individuals. Because some of the estimates are in opposite directions, it is important to note that the pooled estimates in this analysis have no straightforward interpretation. Likewise, in the weighted median model females experience an increase in incident CVD from a higher WHRadjBMI only if they have at least a secondary education, which also contradicts our hypothesis of an inverse association between CVD hazard and educational attainment. We repeat these analyses for household income, another dimension of socioeconomic status associated with incident CVD in the literature, in the appendix in **Table 5.1** and **Figures 5.1 to 5.3** (61).

### 2.3.1 Sensitivity Analyses

The simplest method to minimize the potential role of horizontal pleiotropy in this analysis is a "leave one out" (LOO) method, where the model is re-estimated excluding each SNP separately (62). This analysis should demonstrate whether any outlier SNPs have a disproportionate impact on the effect estimate. If each of these iterations provides directionally consistent results, it decreases the likelihood that horizontal pleiotropy plays a large role in driving the existence of a causal effect because the different SNP subsets would likely not suffer from the exact same source of bias.

A more direct way of assessing the causal effect of adiposity on CVD would involve including only SNPs with an established mechanistic link to adiposity that makes their association more plausible (63). Because BMI is determined by a host of mechanisms not directly related to adiposity, we are only able to perform this sensitivity analysis for WHRadjBMI. Out of the original set of 49 SNPs in WHRadjBMI, only 9 are associated with gene expression in either subcutaneous or omental adipose tissue. Because there were only 9 variants associated with WHRadjBMI that had a plausible functional component identified, we tested the strength of this instrument. The combined effect of the genetic variants produced an F-statistic under 10, suggesting the estimator would suffer from weak instrument bias. As a result, we did not perform this sensitivity analysis. As a result, we only present the leave-one-out analyses below.

**Figure 2.8** plots the distribution of point estimates across the iterations of the LOO analyses. The LOO analysis for the effect of WHRadjBMI on incident CVD for females suggests that no one genetic variant appeared to have an outsized influence on the results except in the case of females with less than a secondary education. However, because this

effect was already so imprecisely estimated, it is unclear how much these genetic variants actually impact the observed results.

The LOO analysis for the effect of WHRadjBMI on incident CVD in **Figure 2.9** for males suggests that no one genetic variant appeared to have an outsized influence on the results except in the case of females with less than a secondary education. There were some notable outliers for males with at least a secondary education and a wide range for males with at least a university degree.

The LOO analysis for the effect of BMI on incident CVD in **Figure 2.10**, which produced far more precise estimates than WHRadjBMI, unsurprisingly demonstrates more consistency in results across genetic variants. There are notable outliers for subjects with less than and at least a university degree; however, these were of much smaller magnitude than the outliers in the WHRadjBMI analysis. Because there were only 9 variants associated with WHRadjBMI that had a plausible functional component identified, we tested the strength of this instrument. The combined effect of the genetic variants produced an F-statistic under 10, suggesting the estimator would suffer from weak instrument bias. As a result, we did not perform this sensitivity analysis.

## 2.4 Discussion

The associational models in this analysis suggest differences in hazard of incident CVD from a higher BMI exist between individuals with and without a secondary education, as well as between those with and without a university degree. However, in the IVW models only individuals with and without a university degree still faced significantly different hazards of incident CVD from a higher BMI and the weighted median models were too imprecise to detect differences between educational attainment groups. Overall, we conclude that differences in hazard of incident CVD from higher BMI by educational attainment – if they exist - are small in magnitude. The results for WHRadjBMI were generally even less different between educational attainment groups, although a counterintuitive relationship emerged in the MR models for females where individuals with less than a secondary education experienced a decrease in hazard from increasing WHRadjBMI.

While higher BMI is related to an increased hazard of incident CVD for every educational attainment group in the associational and IVW models, WHRadjBMI increases the hazard of incident CVD for every group in the associational models and then only a subset of educational attainment groups thereafter. WHRadjBMI has been criticized in previous work for introducing collider bias, which could explain both the inconsistent association between WHRadjBMI and incident CVD in the MR models and the negative association found between WHRadjBMI and incident CVD in individuals with less than a secondary education (64).

Unlike much of the MR literature on CVD, we focused on incident CVD which should assuage concerns regarding reverse causation in studies of prevalent CVD and adiposity (65). We employed an additive hazard model instead of the more commonly

used Cox model, which avoids the issue of non-collapsibility and puts the effect measure in a more easily interpreted additive form. This is the first study to our knowledge to consider SES as a potential effect modifier instead of simply as a confounder for the impact of adiposity on CVD, which acknowledges the ways in which education and adiposity could interact in the production of population health. Lastly, we utilized a variety of sensitivity analyses, perform numerous confirmatory analyses, and pre-specified the analysis plan to increase confidence in our results.

As with many MR studies, we face some statistical power limitations, especially for the smaller educational attainment subgroups. The measures of adiposity used, BMI and WHRadjBMI, have also been criticized for not capturing the underlying variable of interest (total adiposity) and producing collider bias, respectively. As with any MR analysis, we cannot ensure the assumptions outside of relevance are met or know the 'best' model in terms of balance between robustness and statistical efficiency. We also treated education as exogenous, as the MR design only applies to adiposity. The UK has universal healthcare and the UKB is an unusually healthy and high-status snapshot of the country (66). While health disparities by educational attainment persist in countries with universal healthcare, our analysis likely understates the differences in CVD incidence between educational attainment groups in countries without universal healthcare and in more socioeconomically diverse samples (67). Additionally, some researchers have attempted to make the UKB more representative of the broader population with sampling weights, which we do not apply in this analysis (68). We restricted the analysis to only White Europeans and control for genetic ancestry principal components to reduce the possibility of population stratification, but social stratification remains a potential concern (69). Past evidence also suggests nonlinear effects of adiposity on CVD risk, with individuals at higher levels of adiposity facing greater risks, but we treated adiposity's impact as linear in this analysis.

Lastly, unique forms of selection bias represent a general limitation of MR studies. Because high adiposity has an association with mortality in young adulthood, the MR analysis here necessarily restricts to individuals who survived long enough to participate (70,71). Therefore, even with valid instruments or methods robust to invalid instruments, the possibility remains for selection bias from loss to follow up correlated with the instruments (70).

In the associational models, individuals with lower educational attainment face a higher hazard of CVD from an increase in adiposity. However, hazard differences between educational attainment groups are only detected in the MR models for individuals with and without a university degree for BMI and not detected at all for WHRadjBMI.

## 2.5 Tables and Figures

Figure 2.1: Directed acyclic graph of adiposity, educational attainment, and cardiovascular disease incidence

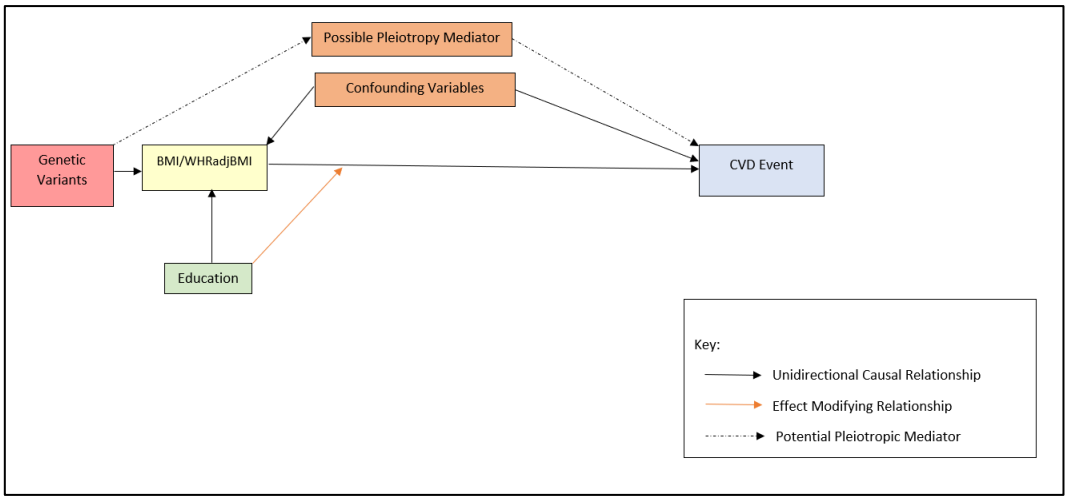


Figure 2.2: Subject exclusion criteria flowchart

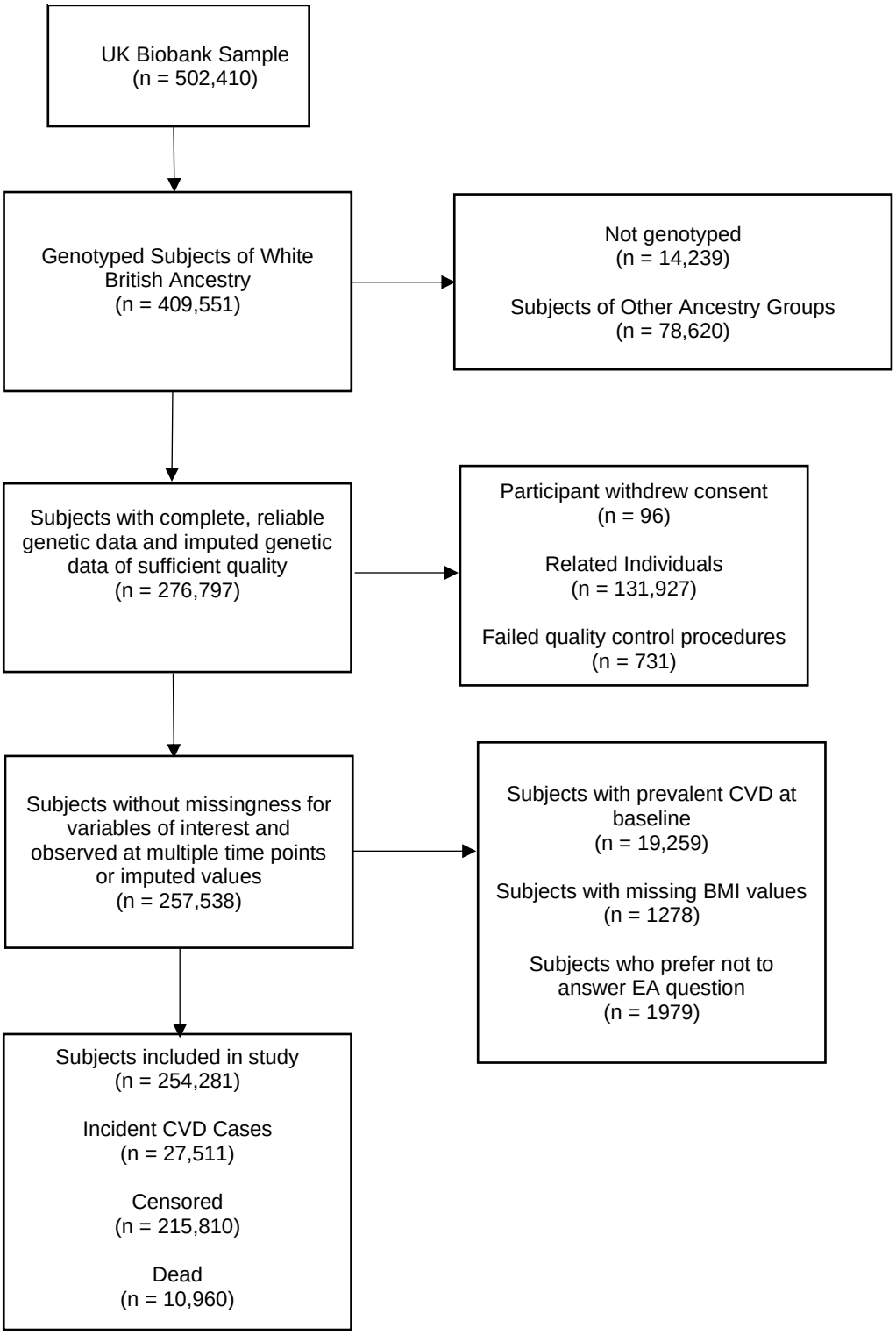


Table 2.1: Characteristics of cohorts with less than and at least secondary education

| | **Less than Secondary Education** | | **At Least Secondary Education** | |
|---|---|---|---|---|
| | **Mean (SD)** | **Range** | **Mean (SD)** | **Range** |
| Age at Baseline | 61.55 (6.28) | 40.31-71.19 | 56.16 (7.97) | 40.02-72.95 |
| BMI at Baseline | 28.17 (4.84) | 14.28-68.41 | 27.05 (4.65) | 12.12-74.68 |
| WHR at Baseline[a] | 0.89 (0.09) | 0.54-1.56 | 0.86 (0.09) | 0.20-1.65 |
| Follow Up Time | 11.23 (3.05) | 0.005-14.45 | 11.78 (2.41) | 0.0027-15.55 |
| CVD Incidence | 0.17 (0.38) | 0-1 | 0.10 (0.30) | 0-1 |
| Ever Smoked | 0.64 (0.48) | 0-1 | 0.59 (0.49) | 0-1 |
| **Biological Sex** | **N (%)** | | **N (%)** | |
| Female | 21,581 (56.06%) | | 117,910 (54.64%) | |
| Male | 16,915 (43.94%) | | 97,875 (45.36%) | |
| **Region** | **N (%)** | | **N (%)** | |
| England | 33,574 (87.21%) | | 190,418 (88.24%) | |
| Scotland | 3276 (8.51%) | | 15,814 (7.33%) | |
| Wales | 1646 (4.28%) | | 9553 (4.43%) | |
| **Total Participants** | 38,496 (100%) | | 215,785 (100%) | |

[a]Note: 253,968 subjects with valid WHR measure (38,484 less than secondary education; 215,484 at least secondary education)

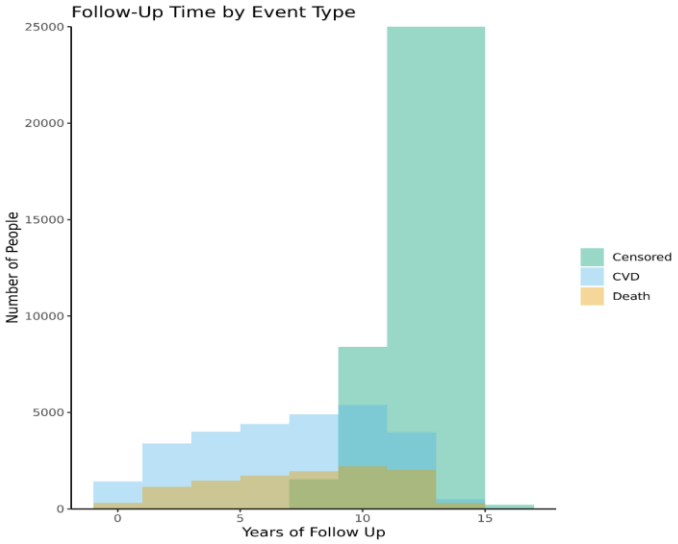
Table 2.2: Characteristics of cohorts with less than university degree and at least university degree

| | Less than University | | At Least University | |
|---|---|---|---|---|
| | **Mean (SD)** | **Range** | **Mean (SD)** | **Range** |
| Age at Baseline | 57.64 (7.94) | 40.20-72.95 | 55.70 (7.90) | 40.02-70.49 |
| BMI at Baseline | 27.63 (4.79) | 12.12-74.68 | 26.43 (4.40) | 13.12-65.23 |
| WHR at Baseline[a] | 0.87 (0.09) | 0.20-1.65 | 0.86 (0.09) | 0.45-1.48 |
| Follow Up Time | 11.60 (2.63) | 0.003-15.55 | 11.88 (2.29) | 0.005-14.46 |
| CVD Incidence | 0.12 (0.33) | 0-1 | 0.08 (0.28) | 0-1 |
| Ever Smoked | 0.61 (0.49) | 0-1 | 0.58 (0.49) | 0-1 |
| **Biological Sex** | **N (%)** | | **N (%)** | |
| Female | 94,368 (56.21%) | | 45,123 (52.23%) | |
| Male | 73,525 (43.79%) | | 41,265 (47.77%) | |
| **Region** | **N (%)** | | **N (%)** | |
| England | 149,201 (88.87%) | | 74,791 (86.58%) | |
| Scotland | 11,244 (6.70%) | | 7846 (9.08%) | |
| Wales | 7448 (4.44%) | | 3751 (4.34%) | |
| **Total Participants** | 167,893 (100%) | | 86,388 (100%) | |

[a]Note: 253,968 patients with valid WHR measure (167,595 less than university degree; 86,373 at least university degree)

Figure 2.3: Follow-up time by event type for body mass index and waist-to-hip ratio adjusted for body mass index

*BMI*




*WHRadjBMI*

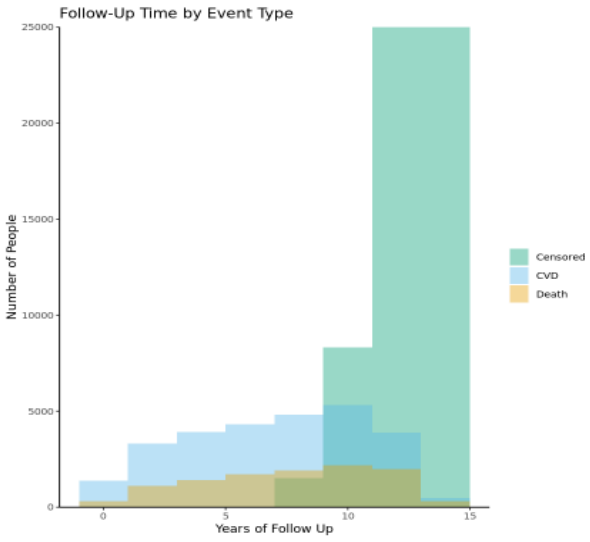
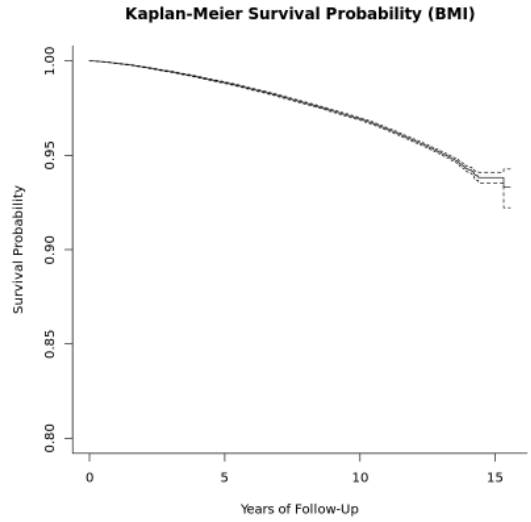


Figure 2.4: Kaplan-Meier survival probability for body mass index and waist-to-hip ratio adjusted for body mass index

*BMI*




*WHRadjBMI*

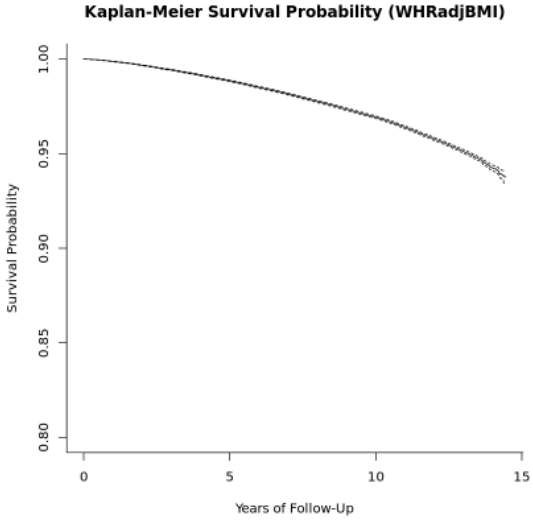
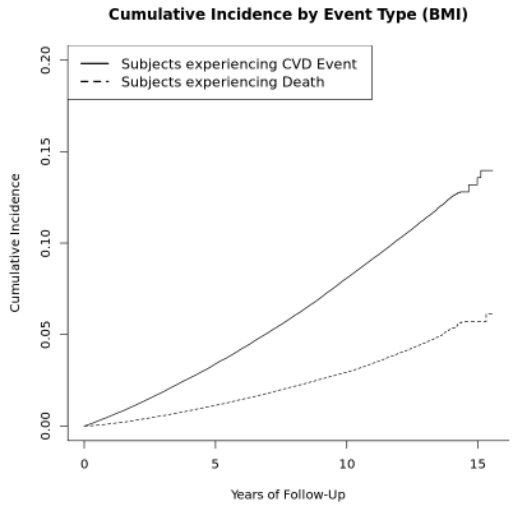


Figure 2.5: Cumulative incidence functions for body mass index and waist-to-hip ratio adjusted for body mass index

*BMI*




*WHRadjBMI*

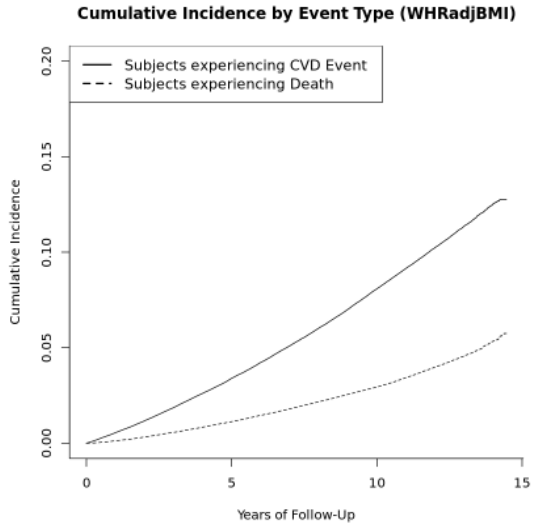
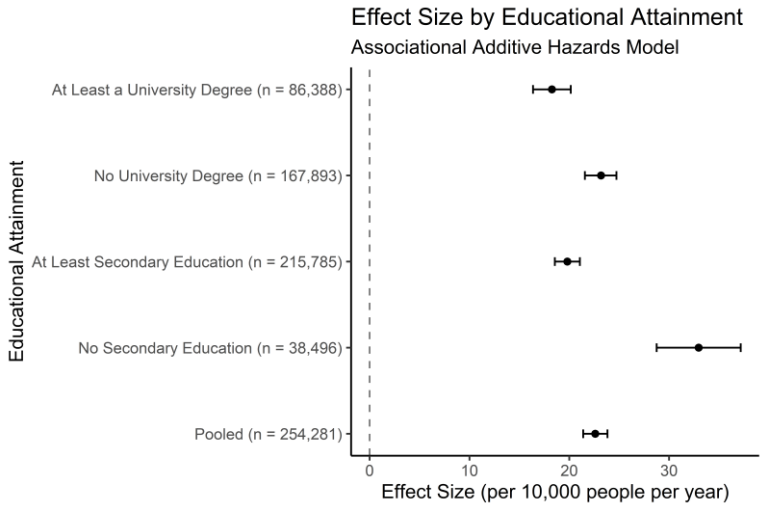


Table 2.3: Association between adiposity and incident cardiovascular disease by educational status and model choice

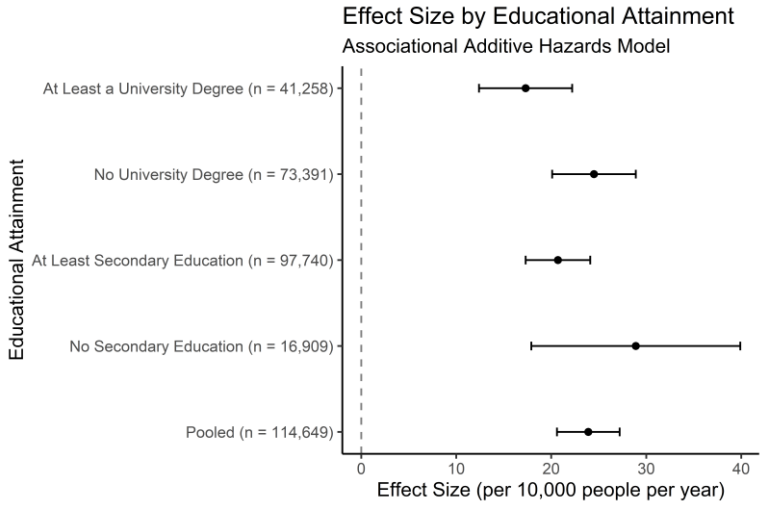
| | | Body Mass Index | | WHRadjBMI (Male) | | WHRadjBMI (Female) | |
|---|---|---|---|---|---|---|---|
| | | **Additive Hazard (CVD incidents per 10,000 person-years)** | **95% Confidence Interval** | **Additive Hazard (CVD incidents per 10,000 person-years)** | **95% Confidence Interval** | **Additive Hazard (CVD incidents per 10,000 person-years)** | **95% Confidence Interval** |
| Less than Secondary Education | Associational | 32.96 | (28.75, 37.17) | 28.90 | (17.90, 39.90) | 8.61 | (3.08, 14.10) |
| | IVW | 28.75 | (-1.89, 59.44) | -28.37 | (-64.16, 7.41) | -31.12 | (-60.38, -1.87) |
| | Weighted Median | 14.28 | (-27.30, 55.85) | -28.67 | (-82.10, 24.75) | -26.67 | (-69.95, 16.61) |
| At Least Secondary Education | Associational | 19.81 | (18.55, 21.06) | 20.70 | (17.30, 24.10) | 5.83 | (4.09, 7.57) |
| | IVW | 23.76 | (14.23, 33.34) | 25.02 | (-14.44, 64.47) | 11.21 | (2.21, 20.20) |
| | Weighted Median | 21.16 | (6.70, 35.62) | 6.26 | (-46.88, 59.41) | 17.39 | (4.04, 30.74) |
| Less than University Degree | Associational | 23.18 | (21.56, 24.72) | 24.50 | (20.10, 28.90) | 7.04 | (4.84, 9.24) |
| | IVW | 29.79 | (17.20, 42.44) | 0.34 | (-56.59, 55.27) | 2.29 | (-8.98, 13.56) |
| | Weighted Median | 19.40 | (2.92, 35.93) | -19.87 | (-88.56, 48.83) | 1.03 | (-16.15, 18.22) |
| At Least University Degree | Associational | 18.26 | (16.37, 20.15) | 17.30 | (12.40, 22.20) | 5.16 | (2.55, 7.77) |
| | IVW | 11.75 | (-0.84, 24.38) | 38.46 | (-14.66, 91.55) | 7.98 | (-5.81, 21.77) |
| | Weighted Median | 12.45 | (-6.86, 31.77) | 34.28 | (-42.43, 111.00) | 8.02 | (-12.35, 28.38) |

*Figure 2.6: Associational relationship between adiposity and incident cardiovascular disease*

*Body mass index*




*Waist-to-hip ratio adjusted for body mass index (Male)*



*Waist-to-hip ratio adjusted for body mass index (Female)*

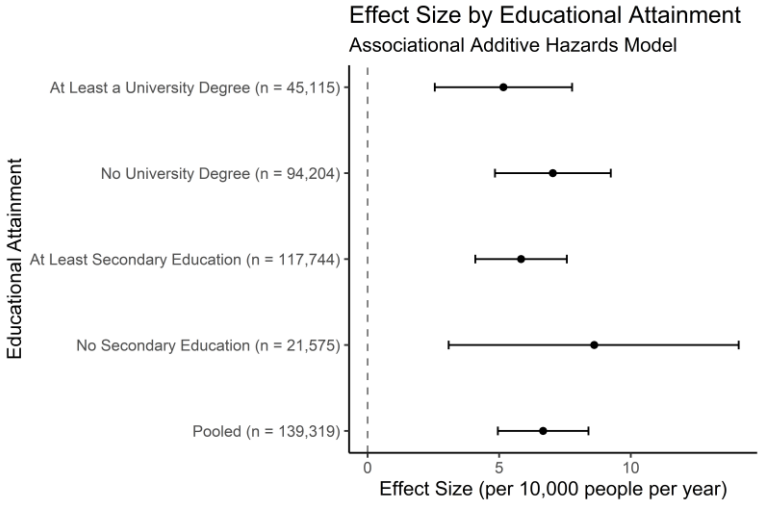
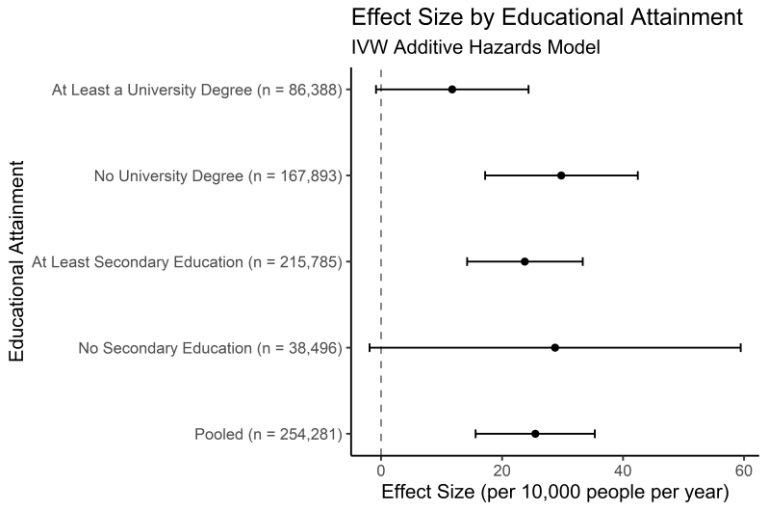
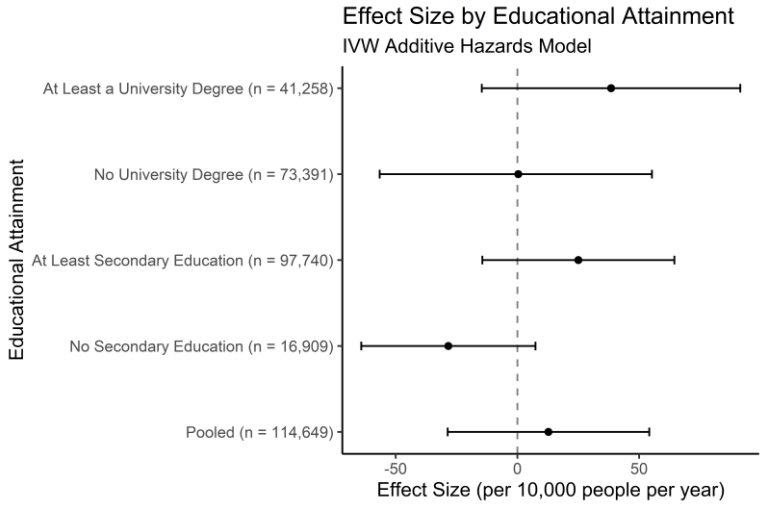


Figure 2.7: Inverse-variance weighted association between adiposity and incident cardiovascular disease

*Body mass index*




*Waist-to-hip ratio adjusted for body mass index (Male)*



*Waist-to-hip ratio adjusted for body mass index (Female)*

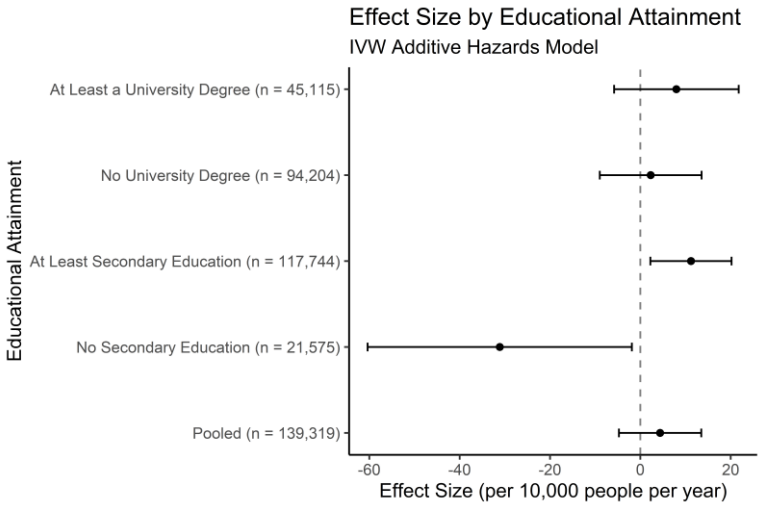
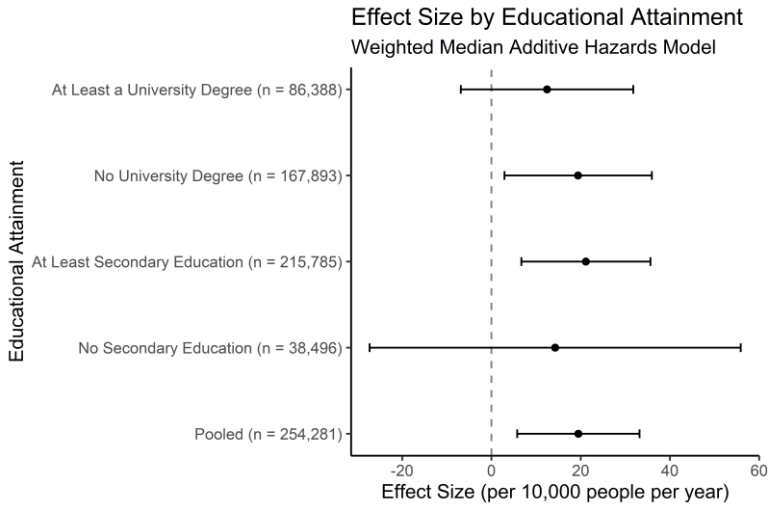
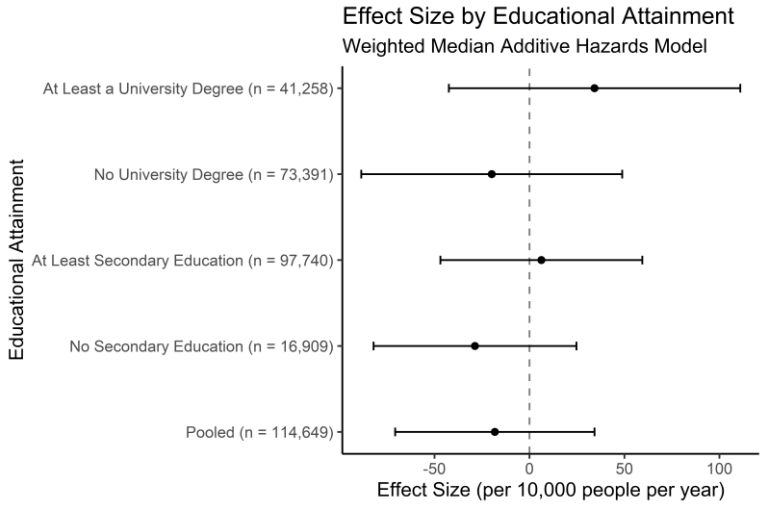


Figure 2.8: Weighted median estimator association between adiposity and incident cardiovascular disease

*Body mass index*




*Waist-to-hip ratio adjusted for body mass index (Male)*



*Waist-to-hip ratio adjusted for body mass index (Female)*

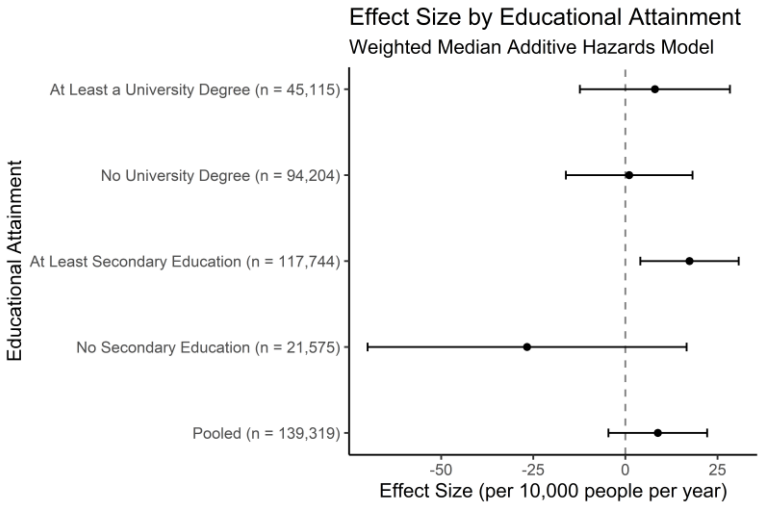
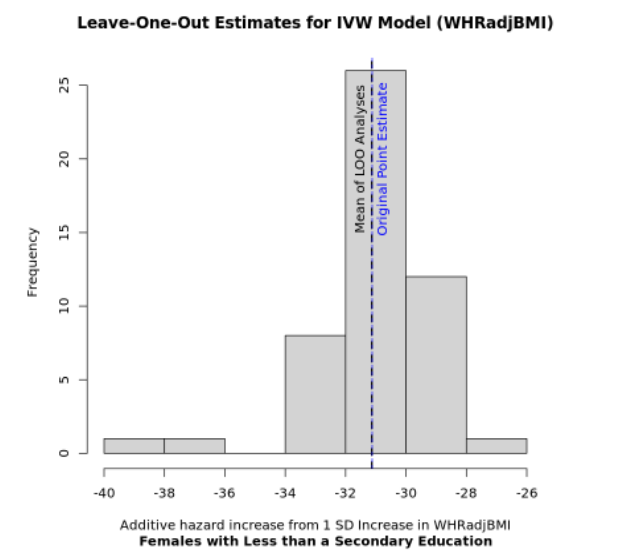
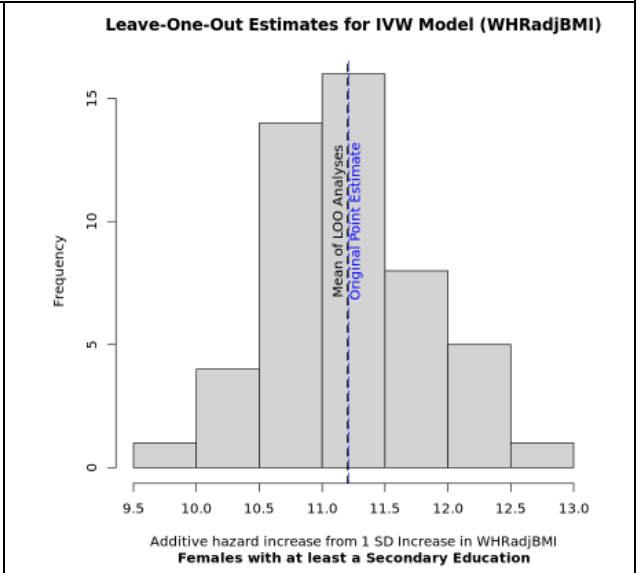
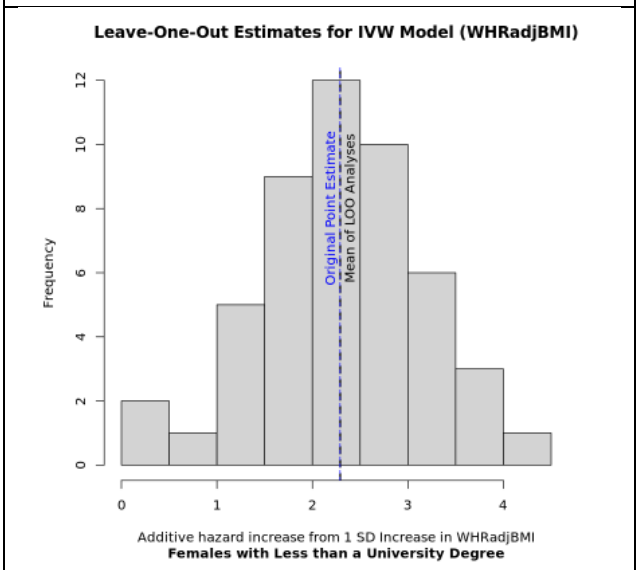


Figure 2.9: Leave-one-out analyses for waist-to-hip ratio adjusted body mass index (Female)



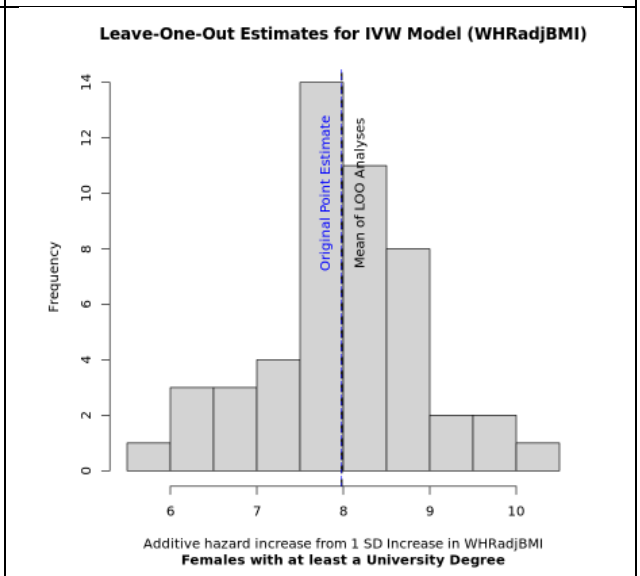



Figure 2.10: Leave-one-out analyses for waist-to-hip ratio adjusted for body mass index (Male)

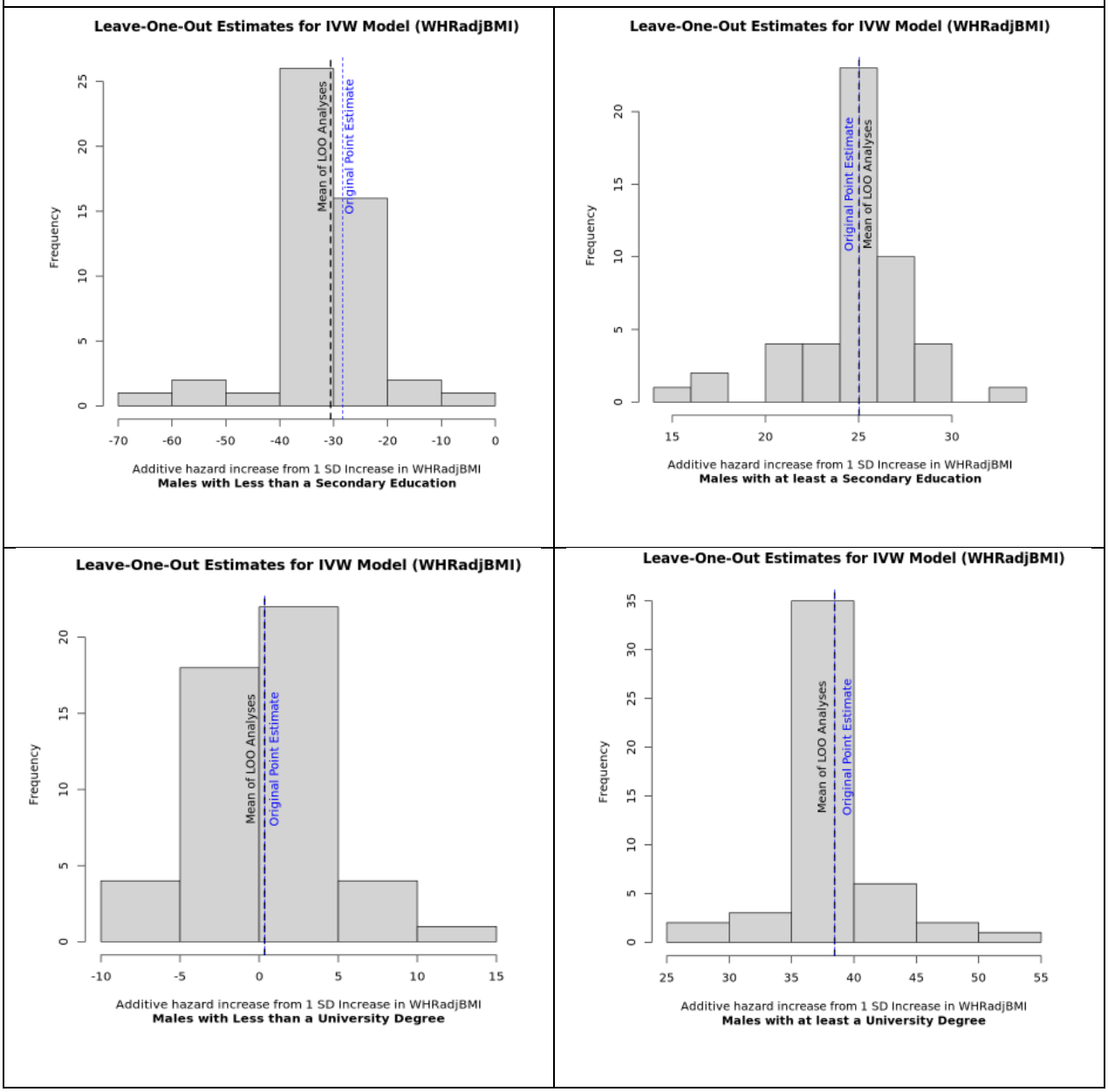
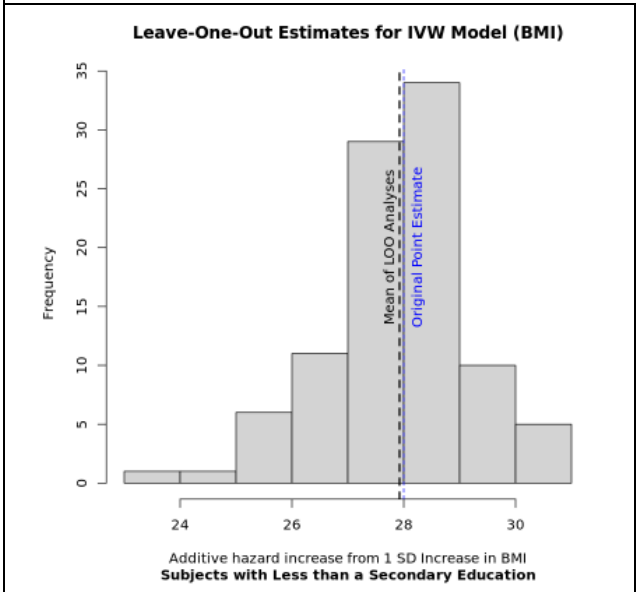
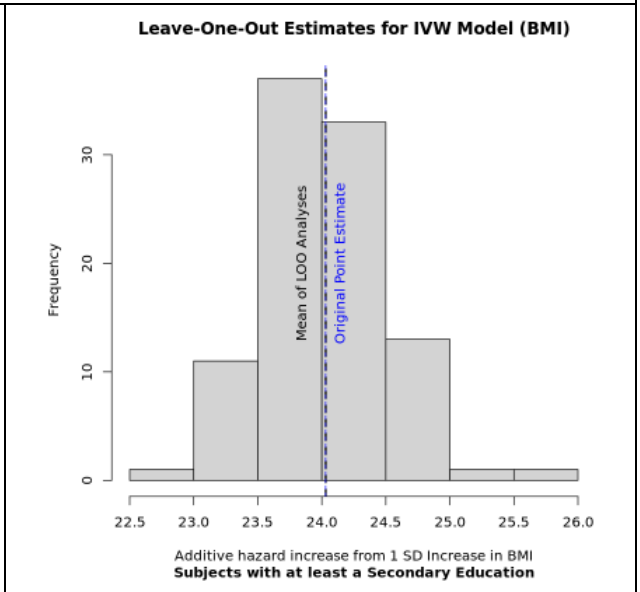
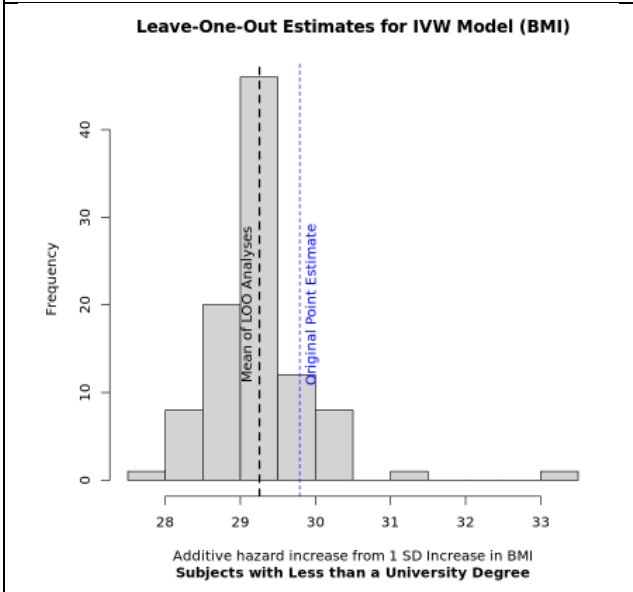
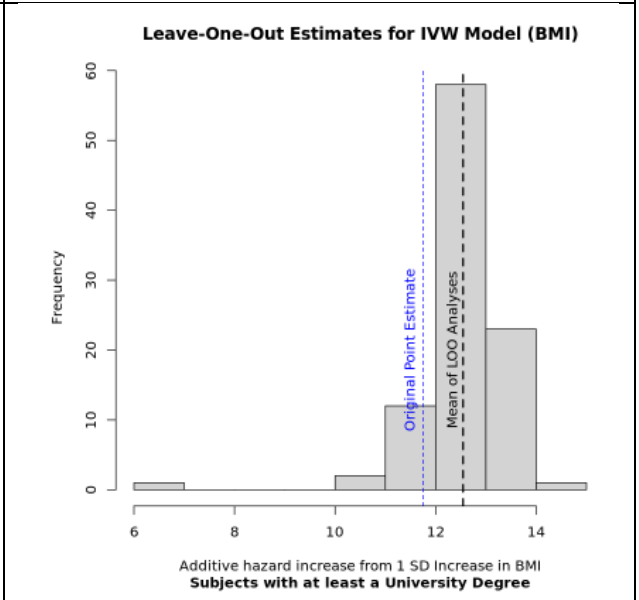


Figure 2.11: Leave-one-out analyses for body mass index



# 3 Joint Association of Genetic Risk and Accelerometer-Measured Physical Activity with Incident Coronary Artery Disease in the UK Biobank Cohort


## Abstract

Previous research demonstrates the joint association of self-reported physical activity and genetics with coronary artery disease. However, an existing research gap is whether accelerometer-measured physical activity volume or intensity can offset genetic predisposition to coronary artery disease. This study explores the independent and joint associations of accelerometer-measured physical activity and genetic predisposition with incident coronary artery disease. The UK Biobank population-based cohort recruited over 500,000 individuals aged 40 to 69 between 2006 and 2010, with 103,712 individuals participating in a weeklong wrist-worn accelerometer study from 2013 to 2015. Individuals of White British ancestry (n = 65,079) meeting the genotyping and accelerometer-based inclusion criteria and with no missing covariates were included in the analytic sample. Incident coronary artery disease based on hospital inpatient records and death register data serves as the outcome of this study. Polygenic risk score and physical activity volume, measured as physical activity energy expenditure, and intensity, measured as percent of physical activity of moderate-to-vigorous intensity, are examined both linearly and at the 20th and 80th percentiles. In the sample of 65,079 individuals, the mean (SD) age was 62.51 (7.76) and 61% were female. During a median follow-up of 6.8 years, 1,382 cases of coronary artery disease developed. At the same genetic risk, physical activity intensity had a hazard ratio (HR) of 1.53 (95% CI: 1.32-1.76) at the 20th compared to 80th percentile versus an HR of 1.35 (95% CI: 1.22-1.50) for physical activity volume. The combination of high genetic risk and low physical activity intensity showed the greatest risk, with an individual at the 80th percentile of genetic risk and 20th percentile of intensity facing an HR of 2.20 (95% CI: 1.45-3.32) compared to an individual at the 20th percentile of genetic risk and 80th percentile of intensity. Physical activity, especially physical activity intensity, is associated with an attenuation of some of the genetic risk of coronary artery disease. This accelerometer-based study provides the clearest evidence to date regarding the joint influence of genetics and physical activity volume and intensity on coronary artery disease.

## 3.1 Introduction

Coronary artery disease (CAD) is a leading cause of death and disability worldwide (72,73). Both physical activity and genetic risk play a crucial role in its development (74,75). Decades of evidence demonstrate the importance of physical activity volume, referring to total physical activity accumulated, and intensity, referring to the proportion of time spent at higher levels of exertion, in reducing the risk of CAD (75-78). However, in recent years, large-scale studies with accelerometer-measured physical activity suggest both that the benefits of physical activity in reducing the risk of CAD may be greater than previously realized and physical activity intensity and volume may each contribute to this risk reduction (25,79-81).

While genetic susceptibility to CAD was established decades ago using twin studies, recent genome-wide association studies have identified millions of variants associated with CAD (72,82-83). Methods of combining these variants have enabled the construction of polygenic risk scores that have improved researchers' ability to understand the genetic risk of developing CAD (84,85).

Several studies have explored the combined impact of genetic susceptibility and self-reported lifestyle factors, including physical activity, on cardiovascular diseases (80,86-90). Genetic risk and physical activity had independent associations with cardiovascular disease and jointly increased overall risk in each study. However, these studies relied on questionnaire-assessed physical activity defined either dichotomously or as quantiles.

This subjective measure of physical activity has several limitations. In doubly labeled water studies, questionnaire-assessed physical activity demonstrated a weaker correlation with PAEE than objective measures (91-92). This method also does not account for incidental physical activity throughout the day. Administering longer questionnaires to provide a more holistic view of an individual's daily physical activity results in higher levels of misclassification (93-94). Even when administered by a trained professional, questionnaire-based techniques suffer from recall and social desirability bias and perform poorly for people of less advantaged sociodemographic backgrounds (95-96). These sources of bias may obscure the associations between physical activity, genetic risk, and incident CAD. Additionally, modeling physical activity dichotomously or in quantiles ignores the continuous relationship between physical activity and CAD risk (75,81). Because these quantiles group physical activity intensity and volume together, these previous efforts could not distinguish their relative importance (80).

This study evaluated the extent to which objective physical activity volume and intensity, measured by a wrist-worn accelerometer and modeled continuously, can offset an individual's genetic susceptibility to incident CAD in the UK Biobank (26). We utilized the best performing polygenic risk score to date, allowing for more precise genetic risk stratification than in previous efforts. Secondarily, we explored whether a gene-environment interaction exists between physical activity volume and intensity and genetic risk.

## 3.2 Methods

### 3.2.1 Accelerometer Cohort

We used the UK Biobank (application # 79654), a population-based cohort of over 500,000 individuals from England, Scotland, and Wales aged 40-69 at recruitment between 2006 and 2010. Follow-up time was censored at March 31st, 2016 in Wales, September 30th, 2021 in England, and July 31st, 2021 in Scotland. This dataset contains information on genetics, health behaviors, socioeconomic status, and health status and is described in detail elsewhere (26). Between 2013 and 2015, participants with an email address were invited except those in the North West region due to concerns about participant burden. A subsample of 103,712 individuals responded to an email recruiting them to wear a wrist-worn Axivity AX3 triaxial accelerometer continuously for seven days on their dominant wrist and provided data. We applied exclusion criteria used previously in this dataset and dropped participants who failed calibration through either insufficient or unreliable data, had implausibly high overall acceleration averages, had wear time under three days, or did not have 24 unique hours of wear in a 24-hour cycle (97-98).

### 3.2.2 Genotyping and Imputation

Participants in the UK Biobank were genotyped using either the UK BiLEVE or the UK Biobank Axiom Array, which each genotyped over 800,000 single-nucleotide polymorphisms (SNPs). Using either the Haplotype Reference Consortium panel or the UK10k and 1000 Genomes phase 3 panels, additional SNPs were imputed, yielding roughly 96 million variants assayed or imputed (99). Following standard genetic quality control criteria in this dataset, we dropped individuals who withdrew consent or were not genotyped, had a mismatch between genetic and reported biological sex, sexual aneuploidy, outliers for missingness or heterozygosity, and we limited the dataset to the maximal set of individuals not related by third degree or closer (50). We also split the dataset by ancestry, with those of White British ancestry as the sample for the analyses. Other ancestry groups contribute too few cases for analysis.

### 3.2.3 Polygenic Score

We applied the most predictive polygenic risk score available for CAD (84). This score was derived by obtaining weights from the largest European-ancestry focused GWAS excluding the UK Biobank; and used PRS-CS, a polygenic risk score prediction method utilizing a Bayesian framework and continuous shrinkage robust to varying genetic architecture. We screened out multi-allelic SNPs, restricted to SNPs with an INFO score greater than 0.6, and restricted minor allele frequency to at least 0.01, yielding 1,087,647 variants included in the score. We then applied the scoring file available on PGS Catalog to recreate the scores derived in the original study (100). We transformed the score into zero mean and unit variance.

### 3.2.4 Physical Activity Measures

Previous researchers processed the raw accelerometer data in the UK Biobank by calibrating to local gravity, filtering out sensor noise and gravity, and detecting and imputing non-wear time data segments to calculate the Euclidean norm minus one (ENMO) (97,101). The average ENMO was summarized as an average proportion of daily time spent at different categories of intensity measured in milligravities (mgs) based on measurements taken every 5 seconds. Following Dempsey *et al.*, we used a formula shown in **Table 3.1** to convert these categorical midpoints of ENMO from dominant wrist-worn accelerometer data into instantaneous physical activity energy expenditure (PAEE) (25). This measure was validated in free-living populations by both doubly labeled water and a combined heart rate monitor and trunk acceleration, the gold and silver standards of physical activity energy expenditure measurement, respectively (91,102-103). PAEE serves as our measure of physical activity volume in kJ/kg/day. In order to calculate physical activity intensity, we categorized physical activity above 125 milligravities as moderate-to-vigorous physical activity (MVPA) and then divided this value by total PAEE and multiplied by 100 to yield the percentage of PAEE from MVPA (percent MVPA) (25,81,91,104).

### 3.2.5 Outcome Definition

We defined CAD based on hospital inpatient episodes, surgeries, and deaths. Specifically, we used ICD-10 codes I20 to I25, I46, and R96 to determine CAD as a cause of death, ICD-10 codes I20.0, I21-I22, and ICD-9 codes 410 and 4110 to denote a CAD event in hospital inpatient records, and OPCS-4 codes K40 to K46, K49, K501, K75 and OPCS-3 code 3043 to denote a CAD-related surgery. We restricted to incident CAD by excluding individuals with an event prior to the start of accelerometer wear. **Figure 3.1** shows the Kaplan-Meier plot for survival in the sample.

### 3.2.6 Covariates

In several waves, participants self-reported information on diet, health behaviors, parental heart disease history, mobility, employment status, and educational attainment pertinent to this analysis. These questionnaires did not occur at the same time as accelerometer wear. To minimize the bias from this discrepancy, we chose the value of the covariates from the most recent wave of self-reported data before accelerometer wear began. Diet consists of several variables, including whether an individual often adds salt to their food, past day consumption of fruits and vegetables, and weekly consumption frequency of oily fish and processed meat. Educational attainment denotes whether a person has a university degree, any other degree, or no degree. Health behaviors include smoking status divided into never, previous, or current and alcohol consumption measured as frequency of consumption per week. Employment status is defined as whether an individual is currently employed, and mobility problems denotes whether an individual

has indicated any issues walking. **Table 3.2** shows how we created these variables from UK Biobank data fields. We controlled for the first 10 genetic principal components, region, biological sex, the Townsend index measuring material deprivation, and season of wear, which as static variables did not depend on the wave selected.

### 3.2.7 Statistical Analyses

We fit a Cox proportional hazards model with age as the timescale to measure the association between physical activity volume and intensity, genetic risk, and incident CAD with time-to-event as the outcome of interest. The model stratified on biological sex, the only covariate violating the proportional hazards assumption based on Schoenfeld residuals. Because the functional form of physical activity volume and intensity's relationship with CAD could be nonlinear, we assessed model fit between the exposures modeled linearly or as a restricted quadratic or cubic spline. The linear model performed best for both physical activity exposures according to BIC. We ran the model with PAEE and polygenic risk score as continuous exposures and an interaction term between these exposures controlling for sex and then the full covariate set. Using this continuous model, hazard ratios and 95% confidence intervals were then calculated for the 20th and 80th percentiles of genetic risk and physical activity volume with the 20th and 80th percentile (lowest), respectively, serving as the reference group. We restricted to the 20th and 80th percentiles of risk instead of the maximum and minimum to avoid interpreting results based on the sparsely populated extremes of the distributions. We ran a model with percent MVPA and polygenic risk score as continuous exposures with an interaction term and controlling for PAEE and adjusting for sex and then the full covariate set and repeated the percentile-based analysis. In sensitivity analyses, we excluded cases occurring within the first year of accelerometer wear to minimize possible reverse causation, explored the impact of measured body mass index, average sleep duration, and cholesterol and blood pressure medication, all potential mediators, as well as manual labor conducted for one's occupation on the results. We relied on complete case analysis but imputed via multivariate imputation by chained equations as a sensitivity analysis.

We explored whether genetic risk and physical activity volume and intensity interact to increase risk of incident CAD by fitting interaction terms between the PA exposures and the polygenic risk score. All analyses were performed using R 4.1.3 (105).

## 3.3 Results

### 3.3.1 Population Characteristics

After screening individuals for valid accelerometer wear data, 96,660 participants remained in the study. We excluded 17,206 participants not meeting the genetic quality control criteria. 1,587 participants had missing covariate data, and 1,980 had prevalent CAD at baseline, which left a final analytic sample of 75,887, among whom 65,079 participants were of White British ancestry as outlined in **Figure 3.2**. Compliance was high, with a median wear time of 6.9 days. **Table 3.3** shows the characteristics of the

participants in our sample. The median follow-up time was 6.8 years with a total of 430,160 cumulative person-years and 1,368 CAD cases. The average age at baseline was 62.5 and participants in this sample were generally higher educated, less likely to smoke, and had lower levels of material deprivation than the larger population in the UK, which coheres with previous research (66). **Table 3.4** shows the percentiles of PAEE, % MVPA, and the polygenic score. Model 1 refers to the fully adjusted model and model 0 refers to the model adjusted for biological sex.

### 3.3.2 Linear Association of Genetic Risk, Physical Activity, and Incident CAD

As **Table 3.5** demonstrates, the hazard ratio for a standard deviation increase in polygenic risk is 1.51 (95% CI: 1.43-1.60) in model 1. The hazard ratio from a standard deviation increase (11.49 kJ/kg/day) in PAEE is 0.83 (95% CI: 0.78-0.88) and for percent MVPA (11.39%) 0.79 (95% CI: 0.73-0.85), which includes PAEE as a confounder, in model 1. **Table 3.6** presents results for model 0.

### 3.3.3 Physical Activity Volume and Genetic Risk Percentile Comparison

**Table 3.7** presents the hazard ratios of participants at different genetic risk and PAEE percentiles. All results are for model 1 and within stratum hazard ratios refer to hazard ratio from a change in one variable at a set value of the other variable. Hazard increases substantially at the highest levels of inactivity, with an individual at the 20th percentile of PAEE (30.07 kJ/kg/day) facing a 35% greater hazard of incident CAD compared to an individual of the same genetic risk at the 80th percentile of PAEE (48.50 kJ/kg/day). Genetic risk has a stronger association as an individual at the 80th percentile of genetic risk within the same PAEE stratum faces a 51% greater hazard of incident CAD than if they were in the 20th percentile of genetic risk. While PAEE and genetic risk each have important independent associations with incident CAD, they combine to create the highest risk of incident CAD. An individual at the 80th percentile of genetic risk and 20th percentile of PAEE faces an 80% greater hazard of incident CAD than the reference group.

### 3.3.4 Physical Activity Intensity and Genetic Risk Percentile Comparison

Controlling for PAEE in model 1, **Table 3.8** shows that percent MVPA has a stronger association with incident CAD than PAEE. An individual at the 20th percentile of percent MVPA (25.97%) faces a 53% greater hazard of incident CAD compared to an individual of the same genetic risk at the 80th percentile (45.46%). The combined association between a participant at the 80th percentile for genetic risk and 20th percentile for percent MVPA results in a 120% higher hazard of incident CAD relative to an individual in the reference group. We explored possible interaction between physical activity

volume and intensity and concluded that no significant interaction exists in this sample. We found no significant interactions between PAEE and genetic risk or percent MVPA and genetic risk, which is similar to what other studies found (80,89).

### 3.3.5 Sensitivity Analyses

We excluded individuals with cases occurring within the first year of follow-up in **Tables 3.9 and 3.10**, reran the analyses with multivariate imputation by chained equations in **Tables 3.11 and 3.12**, and added potential mediators and occupation into the model with results in **Tables 3.13 through 3.15**. None of these choices substantially affected the results.

## 3.4 Discussion

### 3.4.1 Overview of Principal Findings

In this study of 65,079 participants from the UK Biobank, genetic risk was associated with a higher risk of incident CAD regardless of physical activity volume or intensity. Physical activity volume and intensity each had significant independent associations with incident CAD, with physical activity intensity demonstrating the strongest association. While low physical activity volume and intensity increased risk of CAD within a genetic risk stratum, low levels of physical activity volume and intensity were associated with greater risk of incident CAD in the highest genetic risk group. This suggests that physical activity behavior may attenuate some of the high genetic risk of CAD. Specifically, an individual at the 80th percentile of genetic risk and physical activity volume or intensity risk in the percentile analysis faced a 51% or 61% greater hazard of CAD compared to an 80% increase and 120% increase if they also had 20th percentile levels of physical activity volume or intensity, respectively.

### 3.4.2 Comparison with Existing Literature

Because previous studies discretize subjective physical activity, a direct comparison to estimates from the existing literature is not possible. However, the estimates for physical activity's association with cardiovascular diseases in Said, et al. and Tikkanen et al. appear consistent with this study in size and direction of association (80,89). Zaccardi et al., rely on self-reported walking pace as the measure of physical activity and show that this has a large association with CAD, which is also consistent with our stronger results for physical activity intensity (90). Because none of the above studies separate physical activity volume and intensity, we demonstrate that intensity may supersede volume in terms of reducing risk of CAD from a high genetic risk. Our results within genetic risk strata largely agree with existing accelerometer-based studies, although we model physical activity volume and intensity linearly (25,81).

### 3.4.3 Strengths and Limitations

This study is among the first to explore the association of genetic risk and accelerometer-measured physical activity volume and intensity with incident CAD. We use the strongest polygenic risk score and the largest sample of individuals with accelerometer measurements to date. By modeling physical activity continuously and objectively, we avoid the significant misclassification problems from discretizing subjective physical activity (106-107). The exploding commercial popularity of wrist-worn accelerometers has decreased the relevance of current physical activity standards for the population relying on these devices (23,108-109). The current standards do not account for incidental physical activity, or physical activity performed as part of one's normal activities, which means accelerometer-measured physical activity may make users appear more adherent to current guidelines than they are in reality. Studies relying on accelerometer-measured physical activity can help close this gap (108).

This study has several limitations. The UK Biobank sample is disproportionately White and affluent relative to the general population and the sample who responded to take place in the accelerometer study represents further selection bias. However, previous studies have found in terms of physical activity, this cohort appears representative of the general population (110). The covariates used rely on self-reporting and are measured at different times than accelerometer wear. Accelerometer wear occurred over seven days, which makes it cross-sectional, although we validate this against two waves of subjective physical activity in **Figure 3.3**, which found a stronger correlation between more recent subjective physical activity and accelerometer wear. Previous studies have shown reactivity, or a behavioral response to accelerometer wear, may bias measured physical activity volume, although not MVPA (111). More sophisticated machine learning methods can better discriminate between activity types and studies have shown our method of segregating percent MVPA is prone to misclassification (112-113). Wrist-worn accelerometers have limited ability to capture all physical activity, with housework, cycling, and weightlifting especially poorly captured (114-115). Because physical activity is not determined randomly, unmeasured confounding exists. We mitigate this concern by adjusting for related health behavioral factors, socioeconomic status, season of wear, and by performing sensitivity analyses adjusting for potential mediators.

### 3.4.4 Conclusion

High genetic risk and low levels of physical activity volume and intensity were associated with large increases in incident CAD. This study showed physical activity is beneficial regardless of an individual's underlying genetic risk and that genetic risk does not determine an individual's fate regarding CAD (116).

# 3.5 Tables and Figures

Figure 3.1: Kaplan-Meier survival estimates for main sample

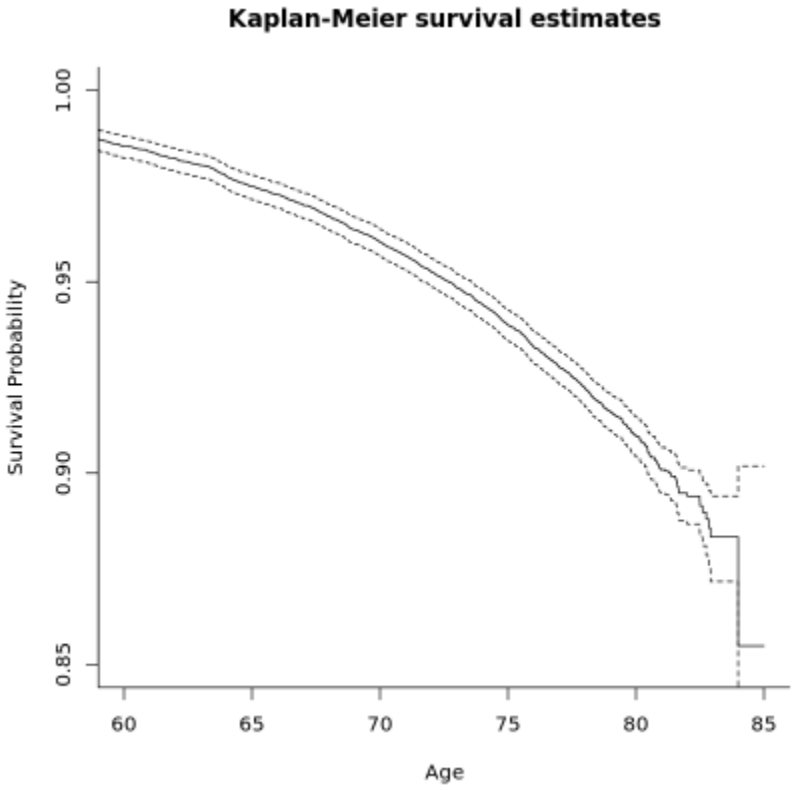



Table 3.1: Equation converting ENMO into PAEE and derivation of percent MVPA

| Physical Activity Exposure | Definition |
|---|---|
| Physical Activity Energy Expenditure (PAEE) | Following the work of Demspey, *et al*. (25), we apply the quadratic equation from White, *et al.*[2] to convert ENMO from wrist-worn accelerometer on dominant hand to PAEE with:<br><br>$PAEE = -10.58 + 1.1176 * (1.5 + 0.8517 * x) + 2.9148 * \sqrt{(1.5 + 0.8517 * x)} - 0.00059277 * (1.5 + 0.8517 * x)^2$<br><br>where x is the midpoint of one of UK Biobank's predefined categories in milligravities. We then convert from J/min/kg into J/kg/day by multiplying by 1.44. We then sum over all of the intervals to get cumulative PAEE. |
| Percent MVPA | We divide PAEE spent above 125 mgs (equivalent to 3 METs) by overall PAEE to get the proportion of PAEE from moderate-to-vigorous physical activity and multiply this by 100 to yield the percent MVPA. |

Table 3.2: Definitions and conversions for covariates in model 1

| Covariate | Final Definition | UK Biobank Definition | Conversion |
|---|---|---|---|
| Smoking Status | Categorical Variable with categories of Current, Never, and Previous following UK Biobank definition. | Categories of Current, Never, and Previous following UK Biobank definition. Derived from Current tobacco smoking and Past tobacco smoking fields. | None |
| Educational Attainment | Categorical variable with University Degree, Other Degree, and No Degree as levels. | Question asked: "which of the following qualifications do you have?" | Converted to University if "College or University Degree" selected, "None" if "None of the above" selected, and "Other" otherwise. |
| Employment Status | Binary variable with employment = 1 and other = 0. | Question asked: "which of the following describes your current situation?" | Converted to binary variable that equals 1 if "in paid employment or self-employed" and 0 if not. |
| Mobility Problems | Binary variable with mobility problems = 1 and no mobility problems = 0. | Question asked: "Please click the ONE box that best describes your health TODAY." | Converted to binary variable that equals 1 if they indicate any issues walking and 0 otherwise. |
| Parental History of Heart Disease | Binary variable with existence of history = 1 and no history = 0. | Question asked: "Has/did your mother ever suffer from?" and "Has/did your father ever suffer from?" | Converted to binary variable that equals 1 if they indicate the mother OR father suffered from heart disease and 0 otherwise. |
| Weekly Processed Consumption | Numeric variable on frequency of processed meat consumption. | Question asked: "How often do you eat processed meats?" and lists options as never, less than once a week, once a week, 2-4 times a week, 5-6 times a week, once or more daily. | Converted to numeric variable with never = 0, less than once a week = 0.5, once a week = 1, 2-4 times a week = 3, 5-6 times a week = 5.5, once or more daily = 7. |
| Fruit & Vegetable Consumption Quartile | Quartiles from 1 to 4 denoting total fruit and vegetable consumption. | Composite of four total questions. Combined cooked and raw vegetable intake "On average how many heaped tablespoons of COOKED(RAW) vegetables would you eat per day?" Combined dried and fresh fruit | Added fruit total and vegetable totals together and then split these totals into quartiles. |

| | | | |
|---|---|---|---|
| | | intake “On average how many pieces of DRIED(FRESH) fruit would you eat per day?” | |
| Weekly Oily Fish Consumption | Numeric variable on frequency of oily fish consumption per week. | Question asked: “How often do you eat oily fish?” and lists options as never, less than once a week, once a week, 2-4 times a week, 5-6 times a week, once or more daily. | Converted to numeric variable with never = 0, less than once a week = 0.5, once a week = 1, 2-4 times a week = 3, 5-6 times a week = 5.5, once or more daily = 7. |
| Weekly Alcohol Consumption | Numeric variable on frequency of alcohol consumption per week. | Question asked: “about how often do you drink alcohol?” and lists daily or almost daily, three or four times a week, once or twice a week, one to three times a month, special occasions only, never. | Converted to numeric variable with never = 0, daily or almost daily = 7, three or four times a week= 3.5, once or twice a week = 1.5, one to three times a month = 0.4, special occasions only = 0.03. |
| Added Salt Intake | Factor variable with four levels never/rarely, sometimes, usually, always. | Question asked: “do you add salt to your food?” with options Never/rarely, sometimes, usually, always. | None |
| Season of Wear | Factor variable coded as Fall, Spring, Winter, or Summer based on date range. | Start time of wear denotes the date they began wearing an accelerometer. | Derived season ranges based on “Start time of wear” variable. |

Figure 3.2: Subject Exclusion Criteria Flowchart

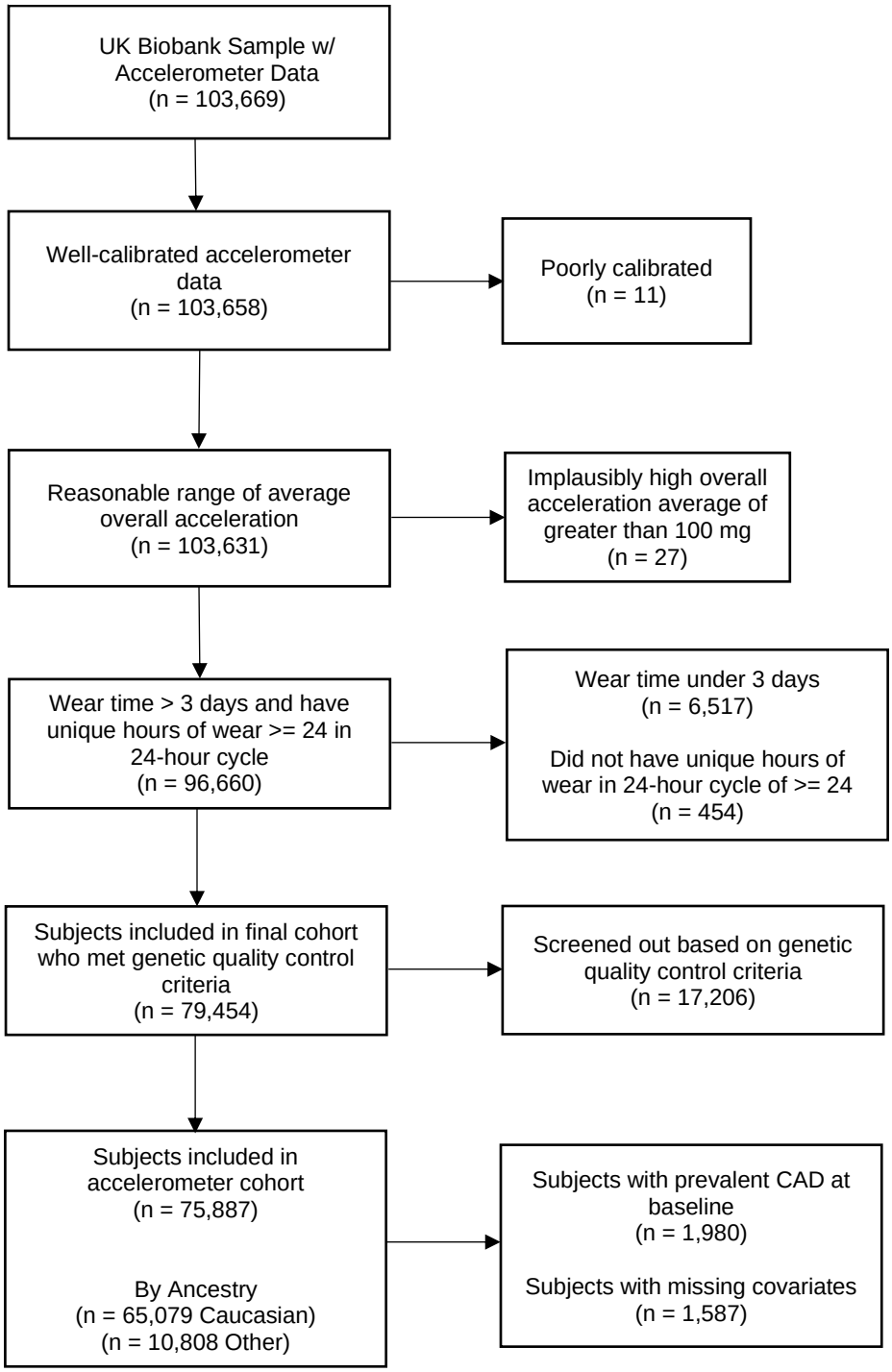


Table 3.3: Baseline characteristics

| Summary Statistics (n = 65,079; Incident CAD = 1368) | | |
|---|---|---|
| *Variable* | | |
| Follow-up Time, median(IQR) | | 6.82 (6.29, 7.36) |
| Physical Activity Energy Expenditure (PAEE), mean(SD) | | 39.56 (11.49) |
| Percent moderate-to-vigorous physical activity (percent MVPA), mean(SD) | | 35.79% (11.39) |
| Standardized Polygenic Risk Score, mean(SD) | | 0 (1.00) |
| Person-Years | | 430,160 |
| Age, mean(SD) | | 62.51 (7.76) |
| **Highest Education Level, n(%)** | | |
| | *University* | 27,779 (42.69%) |
| | *Any Other Qualification* | 32,076 (49.29%) |
| | *No qualification* | 5,224 (8.03%) |
| Townsend Index, mean(SD) | | -1.92 (0.08) |
| Currently Employed, n(%) | | 38,614 (59.33%) |
| Fruit & Vegetable Intake Quartile, mean(SD) | | 2.10 (0.59) |
| Weekly Alcohol Consumption, mean(SD) | | 3.02 (0.58) |
| Weekly Oily Fish Consumption, mean(SD) | | 1.10 (1.00) |
| Female, n(%) | | 36,790 (61.14%) |
| Parental History of Heart Disease, n(%) | | 26,737 (41.08%) |
| **Cigarette Smoking Status, n(%)** | | |
| | *Never* | 37,773 (58.04%) |
| | *Previous* | 23,166 (35.60%) |
| | *Current* | 4,140 (6.36%) |
| **Added Salt Intake, n(%)** | | |
| | *Never* | 39,573 (60.81%) |
| | *Rarely* | 17,085 (26.25%) |
| | *Sometimes* | 6,561 (10.08%) |
| | *Always* | 1,860 (2.86%) |
| **Season Accelerometer Worn, n(%)** | | |
| | *Fall* | 19,329 (29.70%) |
| | *Spring* | 14,810 (22.76%) |
| | *Summer* | 17,086 (26.25%) |
| | *Winter* | 13,854 (21.29%) |

| | | |
|---|---|---|
| **Region, n(%)** | | |
| | *England* | 58,225 (89.47%) |
| | *Scotland* | 4,322 (6.64%) |
| | *Wales* | 2,532 (3.89%) |
| Mobility Limitations, n(%) | | 12,676 (19.48%) |

Table 3.4: Percentiles of PAEE, %MVPA, and PGS

| Percentiles | PAEE | % MVPA | Standardized PGS |
|---|---|---|---|
| 20th Percentile | 30.07 kJ/kg/day | 25.97% | -0.83 |
| 80th Percentile | 48.50 kJ/kg/day | 45.46% | 0.81 |

Table 3.5: Model 1 - controlling for full set of covariates in main analyses (exposures standardized)

| Exposure | Hazard Ratio |
|---|---|
| PAEE | HR = 0.83 (95% CI: 0.78 to 0.88) |
| Standardized PGS (in PAEE eqtn) | HR = 1.51 (95% CI: 1.43 to 1.60) |
| % MVPA | HR = 0.79 (95% CI: 0.73 to 0.85) |
| Standardized PGS (in MVPA eqtn) | HR = 1.50 (95% CI: 1.42 to 1.60) |
| PAEE (in MVPA eqtn) | HR = 0.98 (95% CI: 0.91 to 1.07) |

Table 3.6: Model 0 - controlling for age and biological sex (exposures standardized)

| Exposure | Hazard Ratio |
|---|---|
| Standardized PAEE | HR = 0.80 (95% CI: 0.75 to 0.85) |
| Standardized PGS (in PAEE eqtn) | HR = 1.54 (95% CI: 1.46 to 1.63) |
| Standardized % MVPA | HR = 0.77 (95% CI: 0.71 to 0.83) |
| Standardized PGS (in MVPA eqtn) | HR = 1.54 (95% CI: 1.45 to 1.63) |
| Standardized PAEE (in MVPA eqtn) | HR = 0.97 (95% CI: 0.90 to 1.05) |

Table 3.7: Overview of physical activity volume and genetic susceptibility results

| | 20th Percentile Genetic Risk (-0.83 units) | 80th Percentile Genetic Risk (0.81 units) | |
|---|---|---|---|
| | HR (95% CI) | HR (95% CI) | HR (95% CI) for Genetic Risk within strata of PAEE |
| 80th Percentile PAEE (48.50 kJ/kg/day) | 1.0 (Reference Group) | 1.51 (1.11-2.06) | 1.51 (1.11-2.06) |
| 20th Percentile PAEE (30.07 kJ/kg/day) | 1.35 (1.22-1.50) | 1.80 (1.17-2.79) | 1.85 (1.67-2.05) |
| HR (95% CI) for PAEE within strata of Genetic Risk | 1.35 (1.22-1.50) | 1.27 (1.15-1.41) | |
| Interaction on Multiplicative Scale (Ratio of HRs) | 1.00 (1.00-1.01) | | |

[a] Model adjusted for first 10 genetic principal components, season of wear, salt intake frequency, weekly alcohol intake, weekly oily fish intake, fruit and vegetable consumption quartile, weekly processed meat consumption, parental history of heart disease, mobility problems, employment status, Townsend index, educational attainment, smoking status, region, and biological sex

Table 3.8: Overview of physical activity intensity and genetic susceptibility results

| | 20th Percentile Genetic Risk (-0.83 units) | 80th Percentile Genetic Risk (0.81 units) | |
|---|---|---|---|
| | HR (95% CI) | HR (95% CI) | HR (95% CI) for Genetic Risk within strata of Percent MVPA |
| 80th Percentile Percent MVPA (45.46%) | 1.0 (Reference Group) | 1.61 (1.23-2.11) | 1.61 (1.23-2.11) |
| 20th Percentile Percent MVPA (25.97%) | 1.53 (1.32-1.76) | 2.20 (1.45-3.32) | 1.85 (1.66-2.06) |
| HR (95% CI) for Percent MVPA within strata of Genetic Risk | 1.53 (1.32-1.76) | 1.44 (1.25-1.67) | |
| Interaction on Multiplicative Scale (Ratio of HRs) | 1.40 (0.87-2.28) | | |

[a] Model adjusted for first 10 genetic principal components, season of wear, salt intake frequency, weekly alcohol intake, weekly oily fish intake, fruit and vegetable consumption quartile, weekly processed meat consumption, parental history of heart disease, mobility problems, employment status, Townsend index, educational attainment, smoking status, region, PAEE, and biological sex

Table 3.9: Overview of physical activity volume and genetic susceptibility results (first year excluded)

| | 20th Percentile Genetic Risk (-0.83 units) | 80th Percentile Genetic Risk (0.81 units) | |
|---|---|---|---|
| | HR (95% CI) | HR (95% CI) | HR (95% CI) for Genetic Risk within strata of PAEE |
| 80th Percentile PAEE (48.50 kJ/kg/day) | 1.0 (Reference Group) | 1.46 (1.05-2.04) | 1.46 (1.05-2.04) |
| 20th Percentile PAEE (30.07 kJ/kg/day) | 1.29 (1.16-1.44) | 1.64 (1.03-2.62) | 1.84 (1.65-2.06) |
| HR (95% CI) for PAEE within strata of Genetic Risk | 1.29 (1.16-1.44) | 1.20 (1.08-1.34) | |
| Interaction on Multiplicative Scale (Ratio of HRs) | 1.00 (1.00-1.01) | | |

Table 3.10: Overview of physical activity intensity and genetic susceptibility results (first year excluded)

| | 20th Percentile Genetic Risk (-0.83 units) | 80th Percentile Genetic Risk (0.81 units) | |
|---|---|---|---|
| | HR (95% CI) | HR (95% CI) | HR (95% CI) for Genetic Risk within strata of Percent MVPA |
| 80th Percentile Percent MVPA (45.46%) | 1.0 (Reference Group) | 1.51 (1.13-2.01) | 1.51 (1.13-2.01) |
| 20th Percentile Percent MVPA (25.97%) | 1.53 (1.31-1.79) | 1.97 (1.27-3.08) | 1.83 (1.63-2.05) |
| HR (95% CI) for Percent MVPA within strata of Genetic Risk | 1.53 (1.31-1.79) | 1.97 (1.27-3.08) | |
| Interaction on Multiplicative Scale (Ratio of HRs) | 1.60 (0.95-2.68) | | |

Table 3.11: Overview of physical activity volume and genetic susceptibility results (MICE imputation)

| | 20th Percentile Genetic Risk (-0.83 units) | 80th Percentile Genetic Risk (0.81 units) | |
|---|---|---|---|
| | HR (95% CI) | HR (95% CI) | HR (95% CI) for Genetic Risk within strata of PAEE |
| 80th Percentile PAEE (48.50 kJ/kg/day) | 1.0 (Reference Group) | 1.47 (1.08-2.00) | 1.47 (1.08-2.00) |
| 20th Percentile PAEE (30.07 kJ/kg/day) | 1.36 (1.23-1.51) | 1.74 (1.13-2.68) | 1.85 (1.67-2.04) |
| HR (95% CI) for PAEE within strata of Genetic Risk | 1.36 (1.23-1.51) | 1.27 (1.15-1.41) | |
| Interaction on Multiplicative Scale (Ratio of HRs) | 1.00 (1.00-1.01) | | |

Table 3.12: Overview of physical activity intensity and genetic susceptibility results (MICE imputation)

| | 20th Percentile Genetic Risk (-0.83 units) | 80th Percentile Genetic Risk (0.81 units) | |
|---|---|---|---|
| | HR (95% CI) | HR (95% CI) | HR (95% CI) for Genetic Risk within strata of Percent MVPA |
| 80th Percentile Percent MVPA (45.46%) | 1.0 (Reference Group) | 1.58 (1.21-2.06) | 1.58 (1.21-2.06) |
| 20th Percentile Percent MVPA (25.97%) | 1.53 (1.32-1.76) | 2.12 (1.41-3.20) | 1.84 (1.66-2.05) |
| HR (95% CI) for Percent MVPA within strata of Genetic Risk | 1.53 (1.32-1.76) | 1.44 (1.24-1.66) | |
| Interaction on Multiplicative Scale (Ratio of HRs) | 1.46 (0.91-2.36) | | |

Table 3.13: Add BMI, sleep duration, medication use to model 1 for physical activity volume and genetic susceptibility

| | 20th Percentile Genetic Risk (-0.83 units) | 80th Percentile Genetic Risk (0.81 units) | |
|---|---|---|---|
| | HR (95% CI) | HR (95% CI) | HR (95% CI) for Genetic Risk within strata of PAEE |
| 80th Percentile PAEE (48.50 kJ/kg/day) | 1.0 (Reference Group) | 1.45 (1.06-1.97) | 1.45 (1.06-1.97) |
| 20th Percentile PAEE (30.07 kJ/kg/day) | 1.22 (1.10-1.36) | 1.55 (1.00-2.39) | 1.80 (1.62-1.99) |
| HR (95% CI) for PAEE within strata of Genetic Risk | 1.22 (1.10-1.36) | 1.14 (1.03-1.27) | |
| Interaction on Multiplicative Scale (Ratio of HRs) | 1.00 (1.00-1.01) | | |

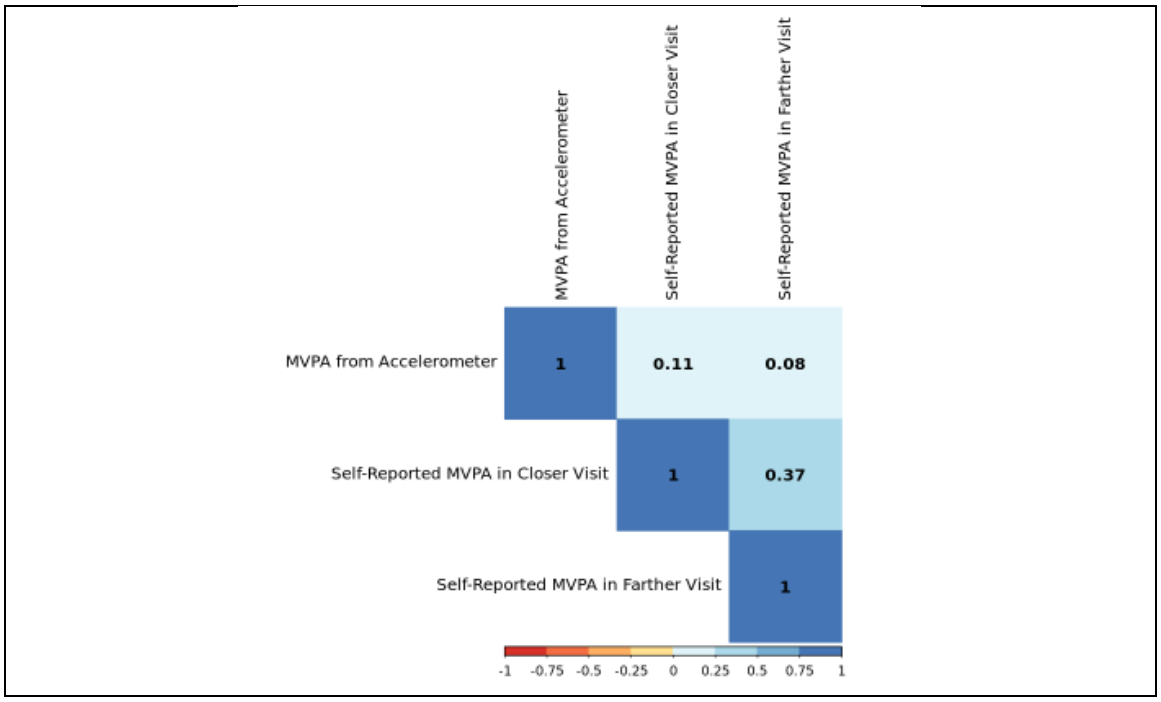
Table 3.14: Add BMI, sleep duration, medication use to model 1 for physical activity intensity and genetic susceptibility

| | 20th Percentile Genetic Risk (-0.83 units) | 80th Percentile Genetic Risk (0.81 units) | |
|---|---|---|---|
| | HR (95% CI) | HR (95% CI) | HR (95% CI) for Genetic Risk within strata of Percent MVPA |
| 80th Percentile Percent MVPA (45.46%) | 1.0 (Reference Group) | 1.57 (1.20-2.05) | 1.57 (1.20-2.05) |
| 20th Percentile Percent MVPA (25.97%) | 1.41 (1.22-1.62) | 1.97 (1.30-2.98) | 1.80 (1.62-2.01) |
| HR (95% CI) for Percent MVPA within strata of Genetic Risk | 1.41 (1.22-1.62) | 1.33 (1.15-1.54) | |
| Interaction on Multiplicative Scale (Ratio of HRs) | 1.41 (0.87-2.28) | | |

Table 3.15: Add BMI, sleep duration, medication use, whether individual is physically active in occupation to model 1 for physical activity volume and genetic susceptibility

| Exposure | Hazard Ratio |
|---|---|
| PAEE | HR = 0.87 (95% CI: 0.82 to 0.93) |
| Standardized PGS (in PAEE eqtn) | HR = 1.50 (95% CI: 1.41 to 1.58) |
| % MVPA | HR = 0.83 (95% CI: 0.76 to 0.90) |
| Standardized PGS (in MVPA eqtn) | HR = 1.49 (95% CI: 1.41 to 1.58) |
| PAEE (in MVPA eqtn) | HR = 1.00 (95% CI: 0.92 to 1.09) |

Figure 3.3: Objective physical activity vs longitudinal subjective physical activity correlation




[a] While correlation between self-reported MVPA and MVPA from accelerometer are low, this correlation changes relatively little between closer or farther visit from accelerometer wear start date and is in line with low correlations even between self-reported MVPA and MVPA from accelerometers measured contemporaneously.[3]

# 4 Conclusion

This dissertation focuses on the ways in which different aspects of an individual's life including their health behaviors, socioeconomic status, and genetics interact to produce cardiovascular disease risk. In chapter two I explore whether socioeconomic status modifies the causal effect of adiposity on incidence of cardiovascular disease. In chapter three, I determine whether objectively measured physical activity volume and intensity could reduce the risk of coronary artery disease among individuals at high genetic risk. In this conclusion, I will briefly discuss the principal findings from these two chapters, the policy and clinical implications of this work and conclude with the future directions for this broader research agenda.

## 4.1 Overview of Principal Findings

Chapter two provides ambiguous evidence regarding the existence of differences in cardiovascular disease risk caused by an increase in adiposity among people of varying socioeconomic backgrounds. For body mass index, clear differences in risk appear to exist in the associational models but these differences only persist between individuals with and without a university degree in the base Mendelian randomization model. While the imprecision of Mendelian randomization is one of its principal drawbacks, the relatively small effect size differences in both the associational and Mendelian randomization models provide convincing evidence that any effect differences that do exist by socioeconomic status are small in magnitude. The measure of central adiposity, waist-to-hip ratio adjusted for body mass index, provided less intuitive results. Among individuals with less than a secondary education, the association was negative, which defies current scientific understanding of the role of central adiposity in cardiovascular disease development. It is likely that collider bias, an issue discussed in the existing literature, and the relatively small number of participants with less than a secondary education are driving this anomalous result.

Chapter three provides evidence that physical activity intensity reduces risk of coronary artery disease to a greater degree than physical activity volume alone and that both can reduce the impact of a high genetic risk of coronary artery disease. However, no evidence of the hypothesized interaction between physical activity and genetic risk was found. This suggests that while physical activity can reduce the genetic risk of coronary artery disease, there are not differential benefits to physical activity between individuals at differing genetic risk levels. While individuals at higher genetic risk may emphasize the need for physical activity owing to their greater susceptibility to the disease, this study does not provide clear evidence of a need for different physical activity standards for this high-risk group. It is important to note that unlike the causal design in chapter two, this chapter relies on a strictly associational model. Other health behaviors such as diet and alcohol consumption imperfectly controlled for in this study likely confound physical activity's association with coronary artery disease and so these conclusions should be interpreted cautiously.

## 4.2 Policy & Clinical Implications

It is reasonable to wonder what, if any, policy implications could be derived from two studies that did not strongly support their original hypotheses. After all, socioeconomic status, genetics, adiposity, and physical activity have well-established influences on cardiovascular disease incidence. The possibility of heterogeneous policy effects would be an important consideration if socioeconomic status impacted the causal effect of adiposity. For instance, the health effects of an intervention to increase educational attainment might include both a decrease in adiposity and making an individual's existing level of adiposity less dangerous. Likewise, if genetics and physical activity had an interactive association with coronary artery disease, this would suggest that personally tailored physical activity recommendations based on an individual's underlying genetic risk be worthwhile. However, studies with null effects are vital precisely because they provide researchers a clearer understanding of when the presence of two risk factors in a single person may not warrant special consideration above and beyond each risk factor's independent effects.

If no evidence of an interaction between genetic risk and physical activity exists, this implies that public health researchers and clinicians should focus on providing accurate and uniform recommendations on physical activity to individuals at different levels of genetic risk all things equal. Likewise, the existence of little - if any - interaction between socioeconomic status and adiposity has important policy implications. When implementing a soda tax, for example, to reduce caloric intake, the primary mechanism for improved health is through reduced adiposity. The benefits of this policy would differ by socioeconomic group primarily due to differences in soda consumption and price sensitivity, which makes its effects easier to project. These studies provide early evidence of the promise of precision health in guiding disease prevention interventions.

## 4.3 Future Directions

Precision medicine has received significant public attention over the past decade. In 2015, President Obama enacted the Precision Medicine Initiative to personalize treatment choices and dosages based on a holistic understanding of a patient, including their genetics, lifestyle, and environment. This initiative was undertaken on the premise that the 'average patient' typified by clinical trials does not provide an accurate understanding of a drug's effects for most patients. However, this focus on precision medicine distracts from the fact that health is largely determined outside of the doctor's office. While precision medicine is vital to ensure patients receive proper medical care, precision health is a concept explicitly focused on utilizing every aspect of a patient's life from their genetic code to their zip code to understand the most effective strategies for disease prevention in everyday life. Fully accounting for all of the risks of disease a person could face through their lifestyle, social status, and genetic makeup is essential to fully quantify individual disease risk and to ultimately prevent these diseases from occurring. By exploring how risks factors from these different aspects of a person's life may interact, this dissertation asks two of the many questions that need to be answered for the successful implementation of precision health.

There are many barriers to this vision of precision health being fully realized for all people. It is imperative that largescale cohort datasets become far more socially and ethnically diverse in the immediate future. Initiatives such as All of Us are a step in the right direction but these datasets and especially the populations most frequently genotyped are too often overwhelmingly white, wealthy, and well-educated. Precision health can never fulfill its promise to improve the entire population's health with the current status quo. If these large datasets are not diversified to include all ancestries, precision health research will instead become just another tool in the perpetuation of health disparities.

Genetics research has a long and sordid history of promoting inequality by class and race and this legacy can still be seen today in both the ancestral homogeneity of genetic datasets and in the field's continued fixation on the genetic predictors of intelligence, educational attainment, and income. These prejudices and lines of inquiry distract from the essential goal of gaining a complete understanding of the genetics behind common diseases among populations in all their diversity.

Lastly, and most importantly, there is a gap between an individual knowing how best to improve their health and actually implementing these changes. Most people know that they get too little exercise or drink too often or eat too much but behavioral change at a population level remains stubbornly out of reach. These new risk prevention tools enabled by precision health must be combined with the arduous "boots on the ground" work that has always been the hallmark of good public health. This dissertation provides only drops in the ocean of discoveries that await for this burgeoning field.

# References


1. Wilson L, Bhatnagar P, Townsend N. Comparing trends in mortality from cardiovascular disease and cancer in the United Kingdom, 1983-2013: Joinpoint regression analysis. Popul Health Metr. 2017;15(1):1-9. doi:10.1186/S12963-017-0141-5/TABLES/4.

2. Sanchez E. Life's simple 7: Vital but not easy. J Am Heart Assoc. 2018;7(11). doi:10.1161/JAHA.118.009324/FORMAT/EPUB.

3. Fryar CD, Carroll MD, Ogden CL. Prevalence of Overweight, Obesity, and Severe Obesity Among Adults Aged 20 and Over: United States, 1960–1962 Through 2015–2016. 2018. https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx. Accessed May 4, 2023.

4. Schiller JS, Bernadel L. Summary health statistics for U.S. adults; National Health Interview Survey, 2002 : data from the National Health Interview Survey. Vital Heal Stat Ser 10 Data from Natl Heal Surv. 2004;10(220). https://stacks.cdc.gov/view/cdc/7008. Accessed May 4, 2023.

5. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Smith GD. Best (but oft-forgotten) practices: The design, analysis, and interpretation of Mendelian randomization studies. Am J Clin Nutr. 2016;103(4):965-978. doi:10.3945/ajcn.115.118216.

6. Tchetgen EJT, Walter S, Vansteelandt S, Martinussen T, Glymour M. Instrumental variable estimation in a survival context. Epidemiology. 2015;26(3):402-410. doi:10.1097/eDe.0000000000000262.

7. Hägg S, Fall T, Ploner A, et al. Adiposity as a cause of cardiovascular disease: A Mendelian randomization study. Int J Epidemiol. 2015;44(2):578-586. doi:10.1093/ije/dyv094.

8. Nordestgaard BG, Palmer TM, Benn M, et al. The effect of elevated body mass index on ischemic heart disease risk: Causal estimates from a mendelian randomisation approach. PLoS Med. 2012;9(5):e1001212. doi:10.1371/journal.pmed.1001212.

9. Fall T, Hägg S, Mägi R, et al. The Role of Adiposity in Cardiometabolic Traits: A Mendelian Randomization Analysis. PLoS Med. 2013;10(6):76. doi:10.1371/journal.pmed.1001474.

10. Riaz H, Khan MS, Siddiqi TJ, et al. Association Between Obesity and Cardiovascular Outcomes: A Systematic Review and Meta-analysis of Mendelian Randomization Studies. JAMA Netw open. 2018;1(7):e183788. doi:10.1001/jamanetworkopen.2018.3788.

11. Lyall DM, Celis-Morales C, Ward J, et al. Association of body mass index with cardiometabolic disease in the UK biobank: A mendelian randomization study. JAMA Cardiol. 2017;2(8):882-889. doi:10.1001/jamacardio.2016.5804.

12. Van Oort S, Beulens JWJ, Van Ballegooijen AJ, Grobbee DE, Larsson SC. Association of Cardiovascular Risk Factors and Lifestyle Behaviors with Hypertension: A Mendelian Randomization Study. Hypertension. 2020:1971-1979. doi:10.1161/HYPERTENSIONAHA.120.15761.

13. Farmer RE, Mathur R, Schmidt AF, et al. Associations between measures of sarcopenic obesity and risk of cardiovascular disease and mortality: A cohort study and mendelian randomization analysis using the UK biobank. J Am Heart Assoc. 2019;8(13). doi:10.1161/JAHA.118.011638.

14. Larsson SC, Bäck M, Rees JMB, Mason AM, Burgess S. Body mass index and body composition in relation to 14 cardiovascular conditions in UK Biobank: A Mendelian randomization study. Eur Heart J. 2020;41(2):221-226. doi:10.1093/eurheartj/ehz388.

15. Carreras-Torres R, Johansson M, Haycock PC, et al. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. BMJ. 2018;361:1767. doi:10.1136/BMJ.K1767.

16. Taylor AE, Richmond RC, Palviainen T, et al. The effect of body mass index on smoking behaviour and nicotine metabolism: a Mendelian randomization study. Hum Mol Genet. 2019;28(8):1322-1330. doi:10.1093/HMG/DDY434.

17. Emdin CA, Khera A V., Natarajan P, et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes and coronary heart disease. JAMA. 2017;317(6):626. doi:10.1001/JAMA.2016.21042.

18. Lu Y, Day FR, Gustafsson S, et al. New loci for body fat percentage reveal link between adiposity and cardiometabolicdisease risk. Nat Commun. 2016;7. doi:10.1038/NCOMMS10495.

19. Pulit SL, Stoneman C, Morris AP, et al. Meta-analysis of genome-wide association studiesfor body fat distribution in 694 649 individuals of European ancestry. Hum Mol Genet. 2019;28(1):166. doi:10.1093/HMG/DDY327.

20. Dale CE, Fatemifar G, Palmer TM, et al. Causal Associations of Adiposity and Body Fat Distribution with Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis. Circulation. 2017;135(24):2373-2388. doi:10.1161/CIRCULATIONAHA.116.026560.

21. D S, TW W, DC C-C, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518(7538):187-196. doi:10.1038/NATURE14132.

22. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197-206. doi:10.1038/nature14177.

23. Bull FC, Al-Ansari SS, Biddle S, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med. 2020;54(24):1451-1462. doi:10.1136/BJSPORTS-2020-102955.

24. American Heart Association Recommendations for Physical Activity in Adults and Kids | American Heart Association. https://www.heart.org/en/healthy-living/fitness/fitness-basics/aha-recs-for-physical-activity-in-adults. Accessed May 4, 2023.

25. Dempsey PC, Rowlands A V, Strain T, et al. Physical activity volume, intensity, and incident cardiovascular disease. Eur Heart J. 2022;43(46):4789-4800. doi:10.1093/EURHEARTJ/EHAC613.

26. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z.

27. Koppaka R. Ten Great Public Health Achievements — United States, 2001–2010. MMWR Morb Mortal Wkly Rep. 2011;60(19):605–9.

28. Tran J, Norton R, Conrad N, Rahimian F, Canoy D, Nazarzadeh M, et al. Patterns and temporal trends of comorbidity among adult patients with incident cardiovascular disease in the UK between 2000 and 2014: A population-based cohort study. PLOS Med. 2018 Mar 1;15(3):e1002513.

29. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012 Dec 15;380(9859):2197–223.

30. Winkleby MA, Jatulis DE, Frank E, Fortmann SP. Socioeconomic Status and Health: How Education, Income, and Occupation Contribute to Risk Factors for Cardiovascular Disease. Am J Public Heal. 1992;82(6):816–20.

31. Loucks EB, Gilman SE, Howe CJ, Kawachi I, Kubzansky LD, Rudd RE, et al. Education and Coronary Heart Disease Risk: Potential Mechanisms Such as Literacy, Perceived Constraints, and Depressive Symptoms. Heal Educ Behav. 2014 Nov 27;42(3):370–9.

32. Cao M, Cui B. Association of Educational Attainment With Adiposity, Type 2 Diabetes, and Coronary Artery Diseases: A Mendelian Randomization Study. Front Public Heal. 2020 Apr 22;8:112.

33. Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. Understanding the consequences of education inequality on cardiovascular disease: mendelian randomisation study. BMJ. 2019 May 22;365:1185.

34. Khaing W, Vallibhakara SA, Attia J, McEvoy M, Thakkinstian A. Effects of education and income on cardiovascular outcomes: A systematic review and meta-analysis. Eur J Prev Cardiol. 2017 Jul 1;24(10):1032–42.

35. Yusuf S, Hawken S, Ôunpuu S, Bautista L, Franzosi MG, Commerford P, et al. Obesity and the risk of myocardial infarction in 27 000 participants from 52 countries: a case-control study. Lancet. 2005 Nov 5;366(9497):1640–9.

36. Manson JE, Colditz GA, Stampfer MJ, Willett WC, Rosner B, Monson RR, et al. A Prospective Study of Obesity and Risk of Coronary Heart Disease in Women. N Engl J Med. 1990 Jan 14;322(13):882–9.

37. NCES. International Educational Attainment. 2020 [cited 2022 Aug 22]; Available from: https://nces.ed.gov/fastfacts/display.asp?id=27.

38. Nomis. Labour Market Profile - Official Census and Labour Market Statistics [Internet]. [cited 2022 Aug 22]. Available from: https://www.nomisweb.co.uk/reports/lmp/gor/2092957698/report.aspx#defs.

39. Baker C. Obesity statistics [Internet]. House of Commons Library. 2022 [cited 2022 Aug 22]. Available from: https://commonslibrary.parliament.uk/research-briefings/sn03336/.

40. Hermann S, Rohrmann S, Linseisen J, May AM, Kunst A, Besson H, et al. The association of education with body mass index and waist circumference in the EPIC-PANACEA study. BMC Public Health. 2011 Mar 17;11(1):1–12.

41. Flegal KM, Graubard BI, Williamson DF, Cooper RS. Reverse Causation and Illness-related Weight Loss in Observational Studies of Body Weight and Mortality. Am J Epidemiol [Internet]. 2011 Jan 1 [cited 2023 Jan 16];173(1):1–9. Available from: https://academic.oup.com/aje/article/173/1/1/128615.

42. Riaz H, Khan MS, Siddiqi TJ, Usman MS, Shah N, Goyal A, et al. Association Between Obesity and Cardiovascular Outcomes: A Systematic Review and Meta-analysis of Mendelian Randomization Studies. JAMA Netw open. 2018 Nov 2;1(7):e183788.

43. Gill D, Zuber V, Dawson J, Pearson-Stuttard J, Carter AR, Sanderson E, et al. Risk factors mediating the effect of body mass index and waist-to-hip ratio on cardiovascular outcomes: Mendelian randomization analysis. Int J Obes (Lond). 2021 Jul 1 ;45(7):1428.

44. Havranek EP, Mujahid MS, Barr DA, Blair I V., Cohen MS, Cruz-Flores S, et al. Social determinants of risk and outcomes for cardiovascular disease: A scientific statement

from the American Heart Association. Vol. 132, Circulation. Lippincott Williams and Wilkins; 2015. p. 873–98.

45. Schultz WM, Kelli HM, Lisko JC, Varghese T, Shen J, Sandesara P, et al. Socioeconomic Status and Cardiovascular Outcomes. Circulation. 2018 ;137(20):2166–78.

46. Beauchamp A, Peeters A, Wolfe R, Turrell G, Harriss LR, Giles GG, et al. Inequalities in cardiovascular disease mortality: the role of behavioural, physiological and social risk factors. J Epidemiol Community Heal. 2010 Jun 1;64(6):542–8.

47. Dégano IR, Marrugat J, Grau M, Salvador-González B, Ramos R, Zamora A, et al. The association between education and cardiovascular disease incidence is mediated by hypertension, diabetes, and body mass index. Sci Rep. 2017 Dec 1;7(1):1–8.

48. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Med [Internet]. 2015 Mar 1 [cited 2023 Jan 16];12(3):e1001779. Available from: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001779.

49. Brumpton B, Sanderson E, Heilbron K, Hartwig FP, Harrison S, Vie GÅ, et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. Nat Commun 2020 111 [Internet]. 2020 Jul 14 [cited 2022 May 6];11(1):1–13. Available from: https://www.nature.com/articles/s41467-020-17117-4.

50. Mitchell R, Hemani G, Dudding T, Corbin L, Harrison S, Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, version 2. [Internet]. 2019. Available from: https://doi.org/10.5523/bris.1ovaau5sxunp2cv8rcy88688v.

51. World Health Organization. International Statistical Classification of Diseases and Related Health Problems. 2nd ed. World Health Organization; 2004. 1–677 p.

52. Connecting for Health (Organization). OPCS classification of interventions and procedures, version 4.5 (April 2009). 4.5. National Health Services, editor. London: TSO; 2009. 1–363 p.

53. Prentice AM, Jebb SA. Beyond body mass index. Obes Rev. 2001 Aug 1;2(3):141–7.

54. Sheehan NA, Didelez V, Burton PR, Tobin MD. Mendelian Randomisation and Causal Inference in Observational Epidemiology. PLOS Med. 2008 Aug;5(8):e177.

55. Cho Y, Rau A, Reiner A, Auer PL. Mendelian randomization analysis with survival outcomes. Genet Epidemiol. 2021 Feb 1;45(1):16–23.

56. Martinussen T, Vansteelandt S. On collapsibility and confounding bias in Cox and Aalen regression models. Lifetime Data Anal. 2013 Jul 1;19(3):279–96.

57. Lee CH, Cook S, Lee JS, Han B. Comparison of Two Meta-Analysis Methods: Inverse-Variance-Weighted Average and Weighted Sum of Z-Scores. Genomics Inform. 2016;14(4):173.

58. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016 May 1;40(4):304–14.

59. Payton ME, Greenstone MH, Schenker N. Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? J Insect Sci [Internet]. 2003 Oct 30;3.

60. Smith GD, Davies NM, Dimou N, Egger M, Gallo V, Golub R, et al. STROBE-MR: Guidelines for strengthening the reporting of Mendelian randomization studies. PeerJ. 2019 Jul 15;e27857v1.

61. Odutayo A, Gill P, Shepherd S, Akingbade A, Hopewell S, Tennankore K, et al. Income disparities in absolute cardiovascular risk and cardiovascular risk factors in the United States, 1999-2014. JAMA Cardiol. 2017 Jul 1;2(7):782–90.

62. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. Hum Mol Genet [Internet]. 2018 Aug 1 [cited 2022 Mar 24];27(R2):R195–208. Available from: https://academic.oup.com/hmg/article/27/R2/R195/4996734.

63. S W, JR M, LD K, ER M, PK C, SC C, et al. Diabetic Phenotypes and Late-Life Dementia Risk: A Mechanism-specific Mendelian Randomization Study. Alzheimer Dis Assoc Disord [Internet]. 2016 [cited 2021 Nov 5];30(1):15–20. Available from: https://pubmed.ncbi.nlm.nih.gov/26650880/.

64. Hartwig FP, Tilling K, Davey Smith G, Lawlor DA, Borges MC. Bias in two-sample Mendelian randomization when using heritable covariable-adjusted summary associations. Int J Epidemiol. 2021 Nov 10;50(5):1639–50.

65. Sattar N, Preiss D. Reverse Causality in Cardiovascular Epidemiological Research: More Common Than Imagined? Circulation. 2017 Jun 13;135(24):2369–72.

66. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol. 2017 Nov 1;186(9):1026–34.

67. Wang JY, Wang CY, Juang SY, Huang KY, Chou P, Chen CW, et al. Low socioeconomic status increases short-term mortality of acute myocardial infarction despite universal health coverage. Int J Cardiol. 2014 Mar 1;172(1):82–7.

68. Van Alten S, Domingue BW, Galama T, Marees AT. Reweighting the UK Biobank to reflect its underlying sampling population substantially reduces pervasive selection bias due to volunteering. 2022.

69. Abdellaoui A, Hugh-Jones D, Yengo L, Kemper KE, Nivard MG, Veul L, et al. Genetic correlates of social stratification in Great Britain. Nat Hum Behav 2019 312. 2019 Oct 21;3(12):1332–42.

70. Swanson SA. A Practical Guide to Selection Bias in Instrumental Variable Analyses. Epidemiology. 2019 May 1;30(3):345–9.

71. Seidell JC. Waist circumference and waist/hip ratio in relation to all-cause mortality, cancer and sleep apnea. Eur J Clin Nutr 2010 641. 2009 Jul 29;64(1):35–41.

72. Kassebaum NJ, Arora M, Barber RM, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1603-1658. doi:10.1016/S0140-6736(16)31460-X.

73. Wang H, Naghavi M, Allen C, et al. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet. 2016;388(10053):1459-1544. doi:10.1016/S0140-6736(16)31012-1.

74. Nikpay M, Goel A, Won H-H, et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121. doi:10.1038/NG.3396.

75. Sattelmair J, Pertman J, Ding EL, Kohl HW, Haskell W, Lee IM. Dose response between physical activity and risk of coronary heart disease: A meta-analysis. Circulation. 2011;124(7):789-795. doi:10.1161/CIRCULATIONAHA.110.010710/FORMAT/EPUB.

75. Sattelmair J, Pertman J, Ding EL, Kohl HW, Haskell W, Lee IM. Dose response between physical activity and risk of coronary heart disease: A meta-analysis. Circulation. 2011;124(7):789-795. doi:10.1161/CIRCULATIONAHA.110.010710/FORMAT/EPUB.

76. Tanasescu M, Leitzmann MF, Rimm EB, Willett WC, Stampfer MJ, Hu FB. Exercise Type and Intensity in Relation to Coronary Heart Disease in Men. JAMA. 2002;288(16):1994-2000. doi:10.1001/JAMA.288.16.1994.

77. Nn JOA, Anson EM, Ich -E Dwards AWR, et al. A Prospective Study of Walking as Compared with Vigorous Exercise in the Prevention of Coronary Heart Disease in Women. https://doi.org/101056/NEJM199908263410904. 1999;341(9):650-658. doi:10.1056/NEJM199908263410904.

78. Hawley JA, Hargreaves M, Joyner MJ, Zierath JR. Integrative Biology of Exercise. Cell. 2014;159(4):738-749. doi:10.1016/J.CELL.2014.10.029.

79. Dempsey PC, Aadland E, Strain T, et al. Physical activity intensity profiles associated with cardiometabolic risk in middle-aged to older men and women. Prev Med (Baltim). 2022;156:106977. doi:10.1016/J.YPMED.2022.106977.

80. Tikkanen E, Gustafsson S, Ingelsson E. Associations of fitness, physical activity, strength, and genetic risk with cardiovascular disease: Longitudinal analyses in the UK biobank study. Circulation. 2018;137(24):2583-2591. doi:10.1161/CIRCULATIONAHA.117.032432/-/DC1.

81. Ramakrishnan R, Doherty A, Smith-Byrne K, et al. Accelerometer measured physical activity and the incidence of cardiovascular disease: Evidence from the UK Biobank cohort study. PLOS Med. 2021;18(1):e1003487. doi:10.1371/JOURNAL.PMED.1003487.

82. Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic Susceptibility to Death from Coronary Heart Disease in a Study of Twins. https://doi.org/101056/NEJM199404143301503. 1994;330(15):1041-1046. doi:10.1056/NEJM199404143301503.

83. Nelson CP, Goel A, Butterworth AS, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet 2017 499. 2017;49(9):1385-1391. doi:10.1038/ng.3913.

84. Tamlander M, Mars N, Pirinen M, Widén E, Ripatti S. Integration of questionnaire-based risk factors improves polygenic risk scores for human coronary heart disease and type 2 diabetes. Commun Biol 2022 51. 2022;5(1):1-13. doi:10.1038/s42003-021-02996-0.

85. Mars N, Kerminen S, Feng Y-CA, et al. Genome-wide risk prediction of common diseases across ancestries in one million people. Cell Genomics. 2022;2(4):100118. doi:10.1016/J.XGEN.2022.100118.

86. Khera A V., Emdin CA, Drake I, et al. Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease. N Engl J Med. 2016;375(24):2349-2358. doi:10.1056/NEJMOA1605086/SUPPL_FILE/NEJMOA1605086_DISCLOSURES.PDF.

87. Rutten-Jacobs LCA, Larsson SC, Malik R, et al. Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306 473 UK Biobank participants. BMJ. 2018;363:4168. doi:10.1136/BMJ.K4168.

88. Wang M, Brage S, Sharp SJ, et al. Associations of genetic susceptibility and healthy lifestyle with incidence of coronary heart disease and stroke in individuals with hypertension. Eur J Prev Cardiol. 2022;29(16):2101-2110. doi:10.1093/EURJPC/ZWAC135.

89. Abdullah Said M, Verweij N, Van Der Harst P. Associations of Combined Genetic and Lifestyle Risks With Incident Cardiovascular Disease and Diabetes in the UK Biobank Study. JAMA Cardiol. 2018;3(8):693-702. doi:10.1001/JAMACARDIO.2018.1717.

90. Zaccardi F, Timmins IR, Goldney J, et al. Self-reported walking pace, polygenic risk scores and risk of coronary artery disease in UK biobank. Nutr Metab Cardiovasc Dis. 2022;32(11):2630-2637. doi:10.1016/J.NUMECD.2022.08.021.

91. White T, Westgate K, Hollidge S, et al. Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. Int J Obes 2019 4311. 2019;43(11):2333-2342. doi:10.1038/s41366-019-0352-x.

92. Maddison R, Ni Mhurchu C, Jiang Y, et al. International physical activity questionnaire (IPAQ) and New Zealand physical activity questionnaire (NZPAQ): A doubly labelled water validation. Int J Behav Nutr Phys Act. 2007;4(1):1-9. doi:10.1186/1479-5868-4-62/TABLES/4.

93. Hagstromer M, Ainsworth BE, Oja P, Sjostrom M. Comparison of a Subjective and an Objective Measure of Physical Activity in a Population Sample. J Phys Act Heal. 2010;7(4):541-550. doi:10.1123/JPAH.7.4.541.

94. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med. 2003;37(3):197-206. doi:10.1136/BJSM.37.3.197.

95. Sabia S, Van Hees VT, Shipley MJ, et al. Association Between Questionnaire- and Accelerometer-Assessed Physical Activity: The Role of Sociodemographic Factors. Am J Epidemiol. 2014;179(6):781-790. doi:10.1093/AJE/KWT330.

96. Adams SA, Matthews CE, Ebbeling CB, et al. The Effect of Social Desirability and Social Approval on Self-Reports of Physical Activity. Am J Epidemiol. 2005;161(4):389-398. doi:10.1093/AJE/KWI054.

97. Doherty A, Jackson D, Hammerla N, et al. Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study. PLoS One. 2017;12(2):e0169649. doi:10.1371/JOURNAL.PONE.0169649.

98. Trost SG, Mciver KL, Pate RR. Conducting Accelerometer-Based Activity Assessments in Field-Based Research. Med Sci Sport Exerc. 2005;37(11):531-543. doi:10.1249/01.mss.0000185657.86065.98.

99. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. bioRxiv. July 2017:166298. doi:10.1101/166298.

100. Birling M-C, Yoshiki A, Adams DJ, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. Nat Genet 2021 534. 2021;53(4):420-425. doi:10.1038/s41588-021-00783-5.

101. Van Hees VT, Fang Z, Langford J, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: An evaluation on four continents. J Appl Physiol. 2014;117(7):738-744. doi:10.1152/JAPPLPHYSIOL.00421.2014/ASSET/IMAGES/LARGE/ZDG0201411640001.JPEG.

102. White T, Westgate K, Wareham NJ, Brage S. Estimation of Physical Activity Energy Expenditure during Free-Living from Wrist Accelerometry in UK Adults. PLoS One. 2016;11(12):e0167472. doi:10.1371/JOURNAL.PONE.0167472.

103. Villars C, Bergouignan A, Dugas J, et al. Validity of combining heart rate and uniaxial acceleration to measure free-living physical activity energy expenditure in young men. J Appl Physiol. 2012;113(11):1763-1771. doi:10.1152/JAPPLPHYSIOL.01413.2011/ASSET/IMAGES/LARGE/ZDG0221203780002.JPEG.

104. Strain T, Wijndaele K, Dempsey PC, et al. Wearable device measured physical activity and future healthrisk. Nat Med. 2020;26(9):1385. doi:10.1038/S41591-020-1012-3.

105. R: The R Project for Statistical Computing. https://www.r-project.org/. Accessed March 23, 2023.

106. Thompson D, Batterham AM, Markovitch D, Dixon NC, Lund AJS, Walhin JP. Confusion and Conflict in Assessing the Physical Activity Status of Middle-Aged Men. PLoS One. 2009;4(2):e4337. doi:10.1371/JOURNAL.PONE.0004337.

107. Steene-Johannessen J, Anderssen SA, Van Der Ploeg HP, et al. Are Self-report Measures Able to Define Individuals as Physically Active or Inactive? Med Sci Sports Exerc. 2016;48(2):235. doi:10.1249/MSS.0000000000000760.

108. Thompson D, Batterham AM, Peacock OJ, Western MJ, Booso R. Feedback from physical activity monitors is not compatible with current recommendations: A recalibration study. Prev Med (Baltim). 2016;91:389-394. doi:10.1016/J.YPMED.2016.06.017.

109. Eckel RH, Jakicic JM, Ard JD, et al. 2013 AHA/ACC guideline on lifestyle management to reduce cardiovascular risk: A report of the American College of cardiology/American Heart Association task force on practice guidelines. Circulation. 2014;129(25 SUPPL. 1):76-99. doi:10.1161/01.CIR.0000437740.48606.D1/-/DC1.

110. Brage S, Lindsay T, Venables M, et al. Descriptive epidemiology of energy expenditure in the UK: findings from the National Diet and Nutrition Survey 2008–15. Int J Epidemiol. 2020;49(3):1007-1021. doi:10.1093/IJE/DYAA005.

111. Baumann S, Groß S, Voigt L, et al. Pitfalls in accelerometer-based measurement of physical activity: The presence of reactivity in an adult population. Scand J Med Sci Sports. 2018;28(3):1056-1063. doi:10.1111/SMS.12977.

112. Walmsley R, Chan S, Smith-Byrne K, et al. Reallocation of time between device-measured movement behaviours and risk of incident cardiovascular disease. Br J Sports Med. 2022;56(18):1008-1017. doi:10.1136/BJSPORTS-2021-104050.

113. Narayanan A, Desai F, Stewart T, Duncan S, MacKay L. Application of Raw Accelerometer Data and Machine-Learning Techniques to Characterize Human Movement Behavior: A Systematic Scoping Review. J Phys Act Heal. 2020;17(3):360-383. doi:10.1123/JPAH.2019-0088.

114. Ham SA, Reis JP, Strath SJ, Dubose KD, Ainsworth BE. Discrepancies between methods of identifying objectively determined physical activity. Med Sci Sports Exerc. 2007;39(1):52-58. doi:10.1249/01.MSS.0000235886.17229.42.

115. Cosoli G, Spinsante S, Scalise L. Wrist-worn and chest-strap wearable devices: Systematic review on accuracy and metrological characteristics. Measurement. 2020;159:107789. doi:10.1016/J.MEASUREMENT.2020.107789.

116. Brown SAN, Jouni H, Marroush TS, Kullo IJ. Effect of Disclosing Genetic Risk for Coronary Heart Disease on Information Seeking and Sharing: The MI-GENES Study (Myocardial Infarction Genes). Circ Cardiovasc Genet. 2017;10(4). doi:10.1161/CIRCGENETICS.116.001613/-/DC1.

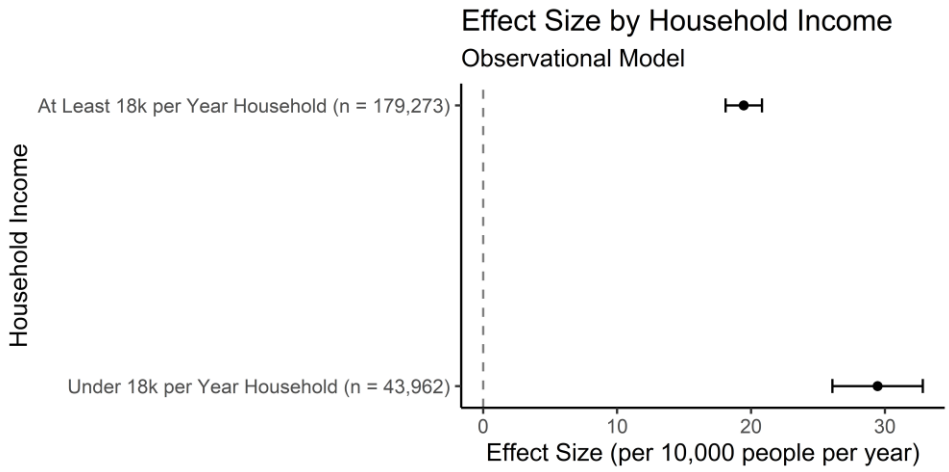
# 5 Supplementary Tables and Figures

Table 5.1: Association between adiposity and incident cardiovascular disease by household income and model choice

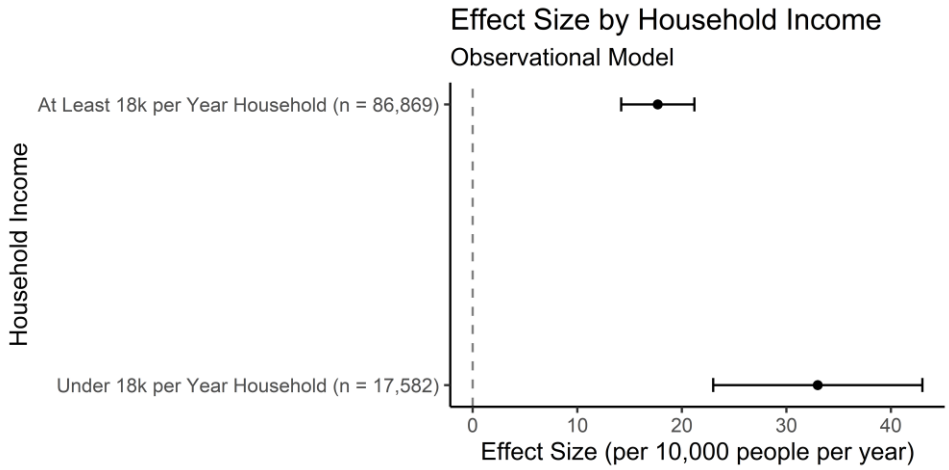
| | | Body Mass Index | | WHRadjBMI (Male) | | WHRadjBMI (Female) | |
|---|---|---|---|---|---|---|---|
| | | **Additive Hazard** | **95% Confidence Interval** | **Additive Hazard** | **95% Confidence Interval** | **Additive Hazard** | **95% Confidence Interval** |
| HH Income Under £18,000 | Associational | 29.45 | (26.08, 32.83) | 33.00 | (23.00, 43.00) | 13.40 | (8.54, 18.30) |
| | IVW | 25.04 | (5.25, 44.84) | -93.80 | (-195.20, 7.60) | 9.61 | (-14.69, 33.92) |
| | Weighted Median | 13.79 | (-18.90, 46.48) | -70.25 | (-216.15, 75.65) | 13.16 | (-22.26, 48.57) |
| HH Income at least £18,000 | Associational | 19.46 | (18.10, 20.82) | 17.70 | (14.20, 21.20) | 3.94 | (2.08, 5.80) |
| | IVW | 25.28 | (15.66, 34.85) | 29.83 | (-11.07, 70.72) | 4.99 | (-4.55, 14.53) |
| | Weighted Median | 27.76 | (12.94, 42.59) | 16.76 | (-37.74, 71.26) | 3.46 | (-10.55, 17.46) |

Figure 5.1: Associational relationship between adiposity and incident cardiovascular disease

*BMI*




*WHRadjBMI (Male)*




*WHRadjBMI (Female)*

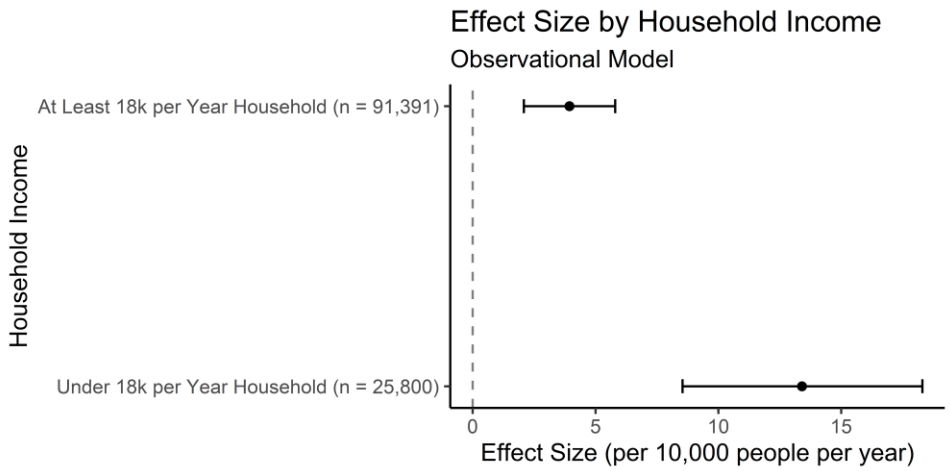
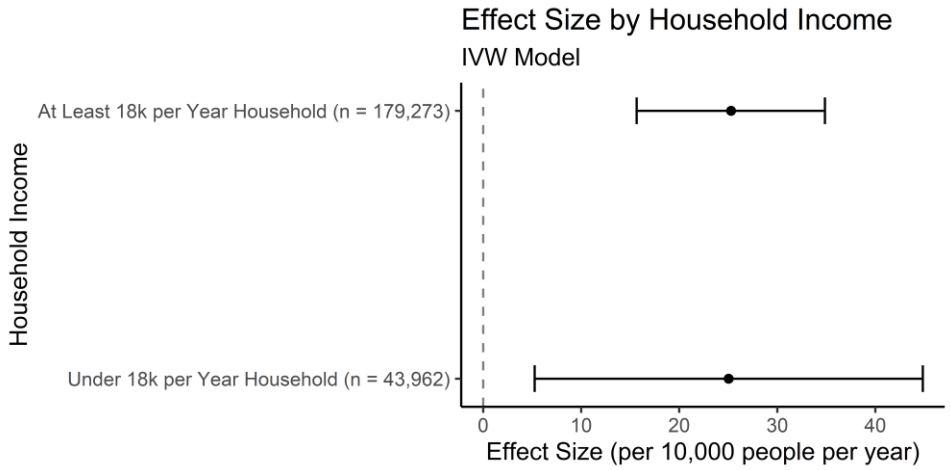
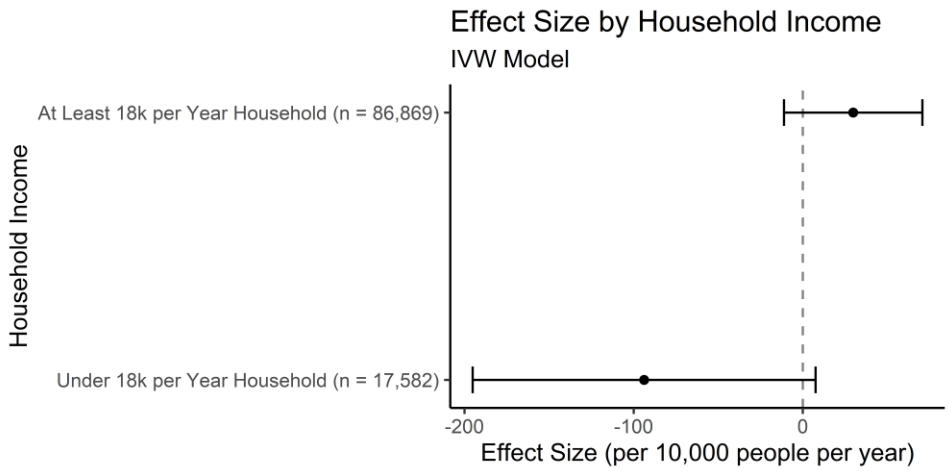



Figure 5.2: IVW association between adiposity and incident cardiovascular disease

*BMI*




*WHRadjBMI (Male)*




*WHRadjBMI (Female)*

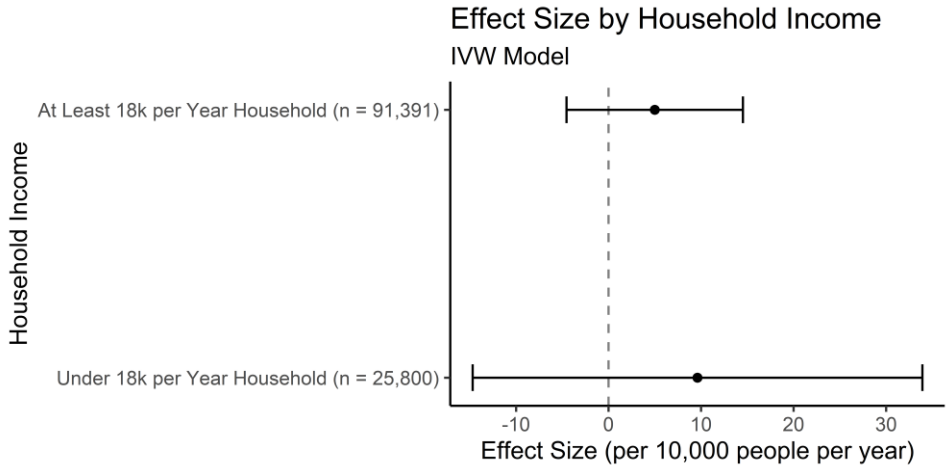
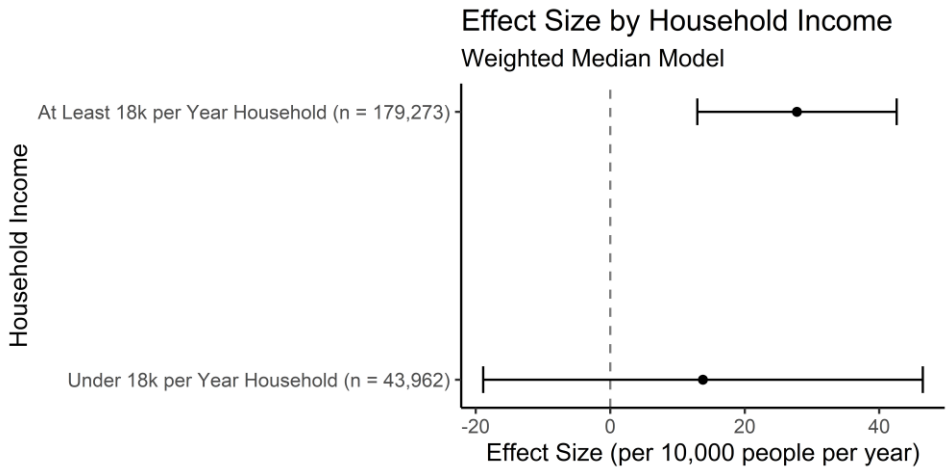
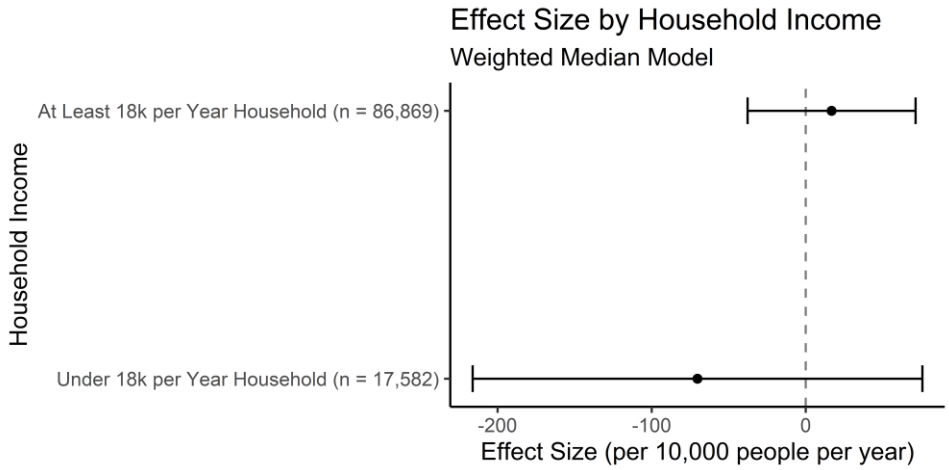


Figure 5.3: Weighted median association between adiposity and incident cardiovascular disease

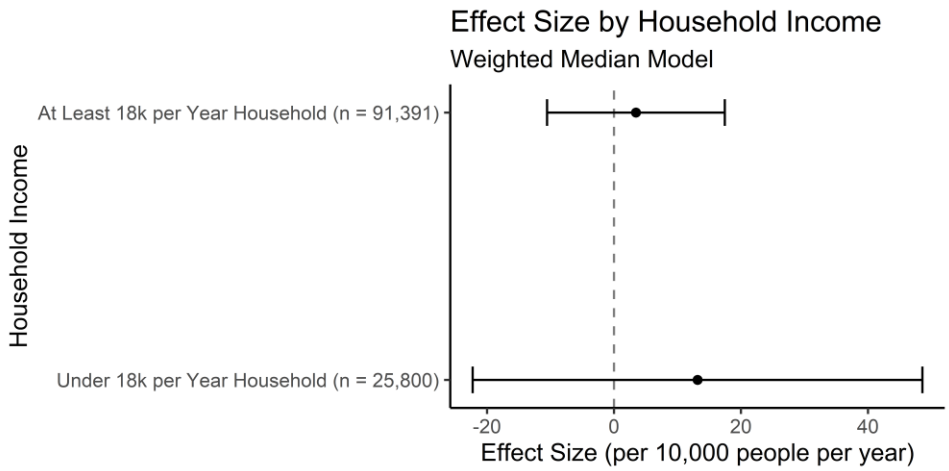
*BMI*




*WHRadjBMI (Male)*




*WHRadjBMI (Female)*


Effect Size by Household Income
Weighted Median Model
At Least 18k per Year Household (n = 91,391)
Under 18k per Year Household (n = 25,800)
Household Income
-20
0
20
40
Effect Size (per 10,000 people per year)