OPTIMIZING BODY MASS INDEX TARGETS USING GENETICS AND BIOMARKERS

OPTIMIZING BODY MASS INDEX TARGETS USING GENETICS AND BIOMARKERS

By IRFAN KHAN, BSc. Kin. (Hon.)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science

McMaster University

McMaster University MASTER OF SCIENCE, Medical Sciences (2021) Hamilton, Ontario

TITLE: Optimizing Body Mass Index Targets Using Genetics And Biomarkers

AUTHOR: Irfan Khan, BSc. Kin. (Hon.) (McMaster University) SUPERVISOR: Dr. Guillaume Paré NUMBER OF PAGES: xiii, 126

Lay Abstract

The World Health Organization (WHO) recommends targeting a body mass index (BMI) between 18.5 - 24.9 kg/m^2 for optimal health. However, this recommendation does not take into account individual differences in genetics or biology. Our project aimed to determine whether the optimal BMI, or the BMI associated with the lowest risk of mortality, varies due to genetic or biological variation. Analyses were conducted across 387,692 individuals. We divided participants into groups according to genetic risk for obesity or clinical biomarker profile. Our results show that the optimal BMI varies according to genetic or biomarker profile. WHO recommendations do not account for this variation, as the optimal BMI can fall under the normal 18.5 - 24.9 kg/m^2 or overweight 25.0 – 29.0 kg/m^2 WHO BMI categories depending on individual genetic or biomarker profile. Thus, there is potential for using genetic and/or biomarker profiles to make more precise BMI recommendations for patients.

Abstract

Introduction/Background: Guidelines from the World Health Organization currently recommend targeting a body mass index (BMI) between 18.5 and 24.9 kg/m^2 based on the lowest risk of mortality observed in epidemiological studies. However, these recommendations are based on population observations and do not take into account potential inter-individual differences. We hypothesized that genetic and non-genetic differences in adiposity, anthropometric, and metabolic measures result in inter-individual variation in the optimal BMI.

Methods: Genetic variants associated with BMI as well as related adiposity, anthropometric, and metabolic phenotypes (e.g. triglyceride (TG)) were combined into polygenic risk scores (PRS), cumulative risk scores derived from the weighted contributions of each variant. 387,692 participants in the UK Biobank were split by quantiles of PRS or clinical biomarkers such as C-reactive protein (CRP), and alanine aminotransferase (ALT). The BMI linked with the lowest risk of all-cause and cause-specific mortality outcomes ("nadir value") was then compared across quantiles ("Cox meta-regression model"). Our results were replicated using the non-linear mendelian randomization (NLMR) model to assess causality.

Results: The nadir value for the BMI–all-cause mortality relationship differed across percentiles of BMI PRS, suggesting inter-individual variation in optimal BMI based on genetics (p = 0.005). There was a difference of 1.90 kg/m^2 in predicted optimal BMI between individuals in the top and bottom 5th BMI PRS percentile.

Individuals having above and below median TG ($p = 1.29\times10^{-4}$), CRP ($p = 7.92 \times 10^{-5}$), and ALT ($p = 2.70 \times 10^{-8}$) levels differed in nadir for this relationship.
There was no difference in the computed nadir between the Cox meta-regression or NLMR models ($p = 0.102$).

Conclusions: The impact of BMI on mortality is heterogenous due to individual genetic and clinical biomarker level differences. Although we cannot confirm that are results are causal, genetics and clinical biomarkers have potential use for making more tailored BMI recommendations for patients.

Acknowledgements

I would like to first give my esteemed gratitude to the entire Genetic and Molecular Epidemiology Laboratory (GMEL) for their kindness, support, and brilliance these past two years. Although we were socially distant due to the COVID-19 pandemic, our love for science allowed us to keep in touch, strengthening the bonds that bind our hearts together. I would like to give special thanks to Michel Kiflen, Sukrit Narula and Dr. Robert Morton for their kind and honest feedback and mentorship. Special regards go to Christina Brinza for her contributions and allowing me the honour to mentor her.

Second, I would like to give my kindest appreciation to my support system, as they helped me emotionally manage the challenging times working towards this thesis. To Dr. Waliul Khan and Salma Khan, my beloved parents; Rushaida Khan, my dear sister, and all my closest friends.

Third, I am truly grateful to my committee members, Dr. Marie Pigeyre and Dr. Katherine Morrison, for their excellent guidance, feedback and encouragement during each committee meeting. I could not have asked for a better committee. I would also like to thank Dr. Russell De Souza for being the external examiner.

Finally, I would like to express my deepest gratitude towards Dr. Guillaume Paré. He is one of the best mentors a student could ever have, not only teaching me how to become a better researcher, but also taking the incredible amount of time and energy to do so, all with great compassion and kindness. I could not have asked for a better influence and mentor than Dr. Paré. I hope to leverage what he has taught me in my future endeavors and beyond.

Table of Contents

LAY ABSTRACT ..... **III**
ABSTRACT ..... **IV**
ACKNOWLEDGEMENTS ..... **VI**
TABLE OF CONTENTS ..... **VII**
LIST OF ABBREVIATIONS AND SYMBOLS ..... **IX**
LIST OF FIGURES ..... **XI**
LIST OF TABLES ..... **XII**
DECLARATION OF ACADEMIC ACHIEVEMENT ..... **XII**
CHAPTER I: INTRODUCTION **1**
1.1 EPIDEMIOLOGY AND BMI 2
1.2 THE GENETICS OF BMI 7
1.3 BIOMARKERS OF BMI 9
1.4 NON-GENETIC COMPONENTS OF BMI 11
1.5 OBJECTIVES 11
1.6 HYPOTHESIS 12
1.7 REFERENCES 12
CHAPTER II: INDIVIDUALIZING BMI TARGETS USING GENETICS AND CLINICAL BIOMARKERS **23**
2.1 INTRODUCTION **24**
2.2 METHODS 25
2.2.1 Study Population *25*
2.2.2 Polygenic Risk Score Calculation *26*
2.2.3 Cox Meta-Regression Model *28*
2.3 RESULTS 30
2.3.1 Optimal BMI across Quantiles of PRS in the UKB Population *30*
2.3.2 Subgroup Analyses *32*
2.3.3 Cause-Specific Analyses *33*
2.3.4 Tertiary Analyses *33*
2.4 DISCUSSION 34
2.4.1 Summary of Results *34*
2.4.2 Clinical Implications *34*
2.4.3 Research Implications *36*
2.4.4 Strengths and Limitations *40*
2.5 CONCLUSION AND SIGNIFICANCE 42
2.6 REFERENCES 42
2.7 FIGURES 52

CHAPTER III: OPTIMIZING BMI RECOMMENDATIONS USING GENETICS AND CLINICAL BIOMARKERS - A MENDELIAN RANDOMIZATION APPROACH 64

3.1 INTRODUCTION 65
3.2 METHODS 69
3.2.1 UK Biobank Study Population 69
3.2.2 Polygenic Risk Score Calculations 69
3.2.3 Non-Linear Mendelian Randomization 69
3.3 RESULTS 73
3.3.1 Primary Analyses 73
3.3.2 Comparison with the Cox Meta-Regression Model 73
3.3.3 Cause-Specific Analyses 73
3.3.4 Secondary Analyses 74
3.3.5 Tertiary Analyses 74
3.3.6 Non-Linear Mendelian Randomization Sensitivity Analysis 74
3.3.7 Calibration Interval Computation 76
3.4 DISCUSSION 76
3.4.1 Summary of Results 76
3.4.2 The causal J-shaped relationship between BMI and all-cause mortality 76
3.4.3 Limitations of the bootstrapping method to generate confidence intervals 79
3.4.4 Calibration interval method for computing confidence intervals 80
3.4.5 Other limitations 81
3.5 CONCLUSION AND SIGNIFICANCE 82
3.6 REFERENCES 83
3.7 FIGURES 92

CHAPTER IV: CONCLUSION 100

4.1 SUMMARY OF MAIN FINDINGS 101
4.2 CLINICAL IMPLICATIONS 101
4.3 RESEARCH IMPLICATIONS 105
4.4 LIMITATIONS AND FUTURE DIRECTIONS 106
4.5 FINAL REMARKS 108
4.6 REFERENCES 109

APPENDICES 117

List of Abbreviations and Symbols

ACM - All-Cause Mortality

ALT - Alanine Aminotransferase

BMI - Body Mass Index

CAD – Canadian Dollar

CRP - C-Reactive Protein

CI - Confidence Interval

CVD - Cardiovascular Disease

GEFOS - GEnetic Factors for OSteoporosis Consortium

GIANT - Genetic Investigation of ANthropometric Traits

GWAS - Genome-Wide Association Study

GWS - Genome-Wide Significance

HR - Hazard Ratio

ICD-10 - International Classification Of Diseases, 10th Revision

IV - Instrumental Variable

Hb1Ac - Hemoglobin A1C

IMD - Index Of Multiple Deprivation

LACE - Local Average Causal Estimate

LD - Linkage Disequilibrium

LDL - Low-Density Lipoprotein

LM - Whole Body Lean Mass

MD - Metabolically Deleterious BMI

MF - Metabolically Favourable BMI

MN - Metabolically Neutral BMI

MR - Mendelian Randomization

MVP - Million Veteran Program (MVP)

NAFLD – Non-Alcoholic Fatty Liver Disease

NLMR - Non-Linear Mendelian Randomization

OR - Odds Ratio

PC - Principal Component

pQTL - Protein Quantitative Trait Loci

PRS - Polygenic Risk Score

RPC - Regional Polygenic Correlation

RCT - Randomized Controlled Trial

SD - Standard Deviation

SNP - Single Nucleotide Polymorphism

TG - Triglyceride

T2D - Type 2 Diabetes

UK - United Kingdom

UKB - UK Biobank

USD - United States Dollar

WHO - World Health Organization

WHR - Waist-To-Hip Ratio

Wt/Ht - Weight-To-Height Ratio

List of Figures

CHAPTER II: INDIVIDUALIZING BMI TARGETS USING GENETICS AND CLINICAL BIOMARKERS

Figure 2.1 - Regional polygenic correlation between BMI and T2D……52

Figure 2.2 - Comparing the optimal BMI across quantiles of BMI PRS……...53

Figure 2.3 - Bootstrapping procedure to calculate confidence intervals for the population optimal BMI…….. 54

Figure 2.4 – Mediation of BMI-mortality relationships by genetics…….56

Figure 2.5 - Mediation of optimal and mean BMI by TG, CRP, and ALT levels……....57

Figure 2.6 - A) The optimal BMI and B) the change in optimal BMI by cause-specific mortality…… 58

Figure 2.7 - Variation in optimal BMI across 20 quantiles of TG, ALT, and CRP……..60

Figure 2.8 - The relationship between WHR and LM, respectively, and all-cause mortality ……62

Figure 2.9 – Variation in optimal LM between quantiles of LM PRS……...62

CHAPTER III: OPTIMIZING BMI RECOMMENDATIONS USING GENETICS - A MENDELIAN RANDOMIZATION APPROACH

Figure 3.1 - The BMI – all-cause mortality relationship using non-linear mendelian randomization (NLMR) in the UKB……..92

Figure 3.2 - Comparing the computed population optimal BMI using the Cox meta-regression and NLMR models……93

Figure 3.3 –Differences between the population optimal BMI against cause-specific mortality using NLMR……. 94

Figure 3.4 – The relationship between BMI and all-cause mortality in the UKB A) smokers only and B) never smokers only using NLMR……96

Figure 3.5 - BMI-all-cause mortality relationships outputted from bootstrapped samples of younger adult subgroup of the UKB using NLMR…….…….. 97

Figure 3.6 - Derivative of the quadratic model fit between mean BMI per quantile and the local average causal estimate for the BMI-all-cause mortality relationship……. 98

Figure 3.7 - Comparison of the population optimal BMI and UKB participants above and below median TG, CRP, and ALT levels using the quadratic model for NLMR……. 99

Figure 3.8 - Comparing Mendelian Randomization and Randomized Control Trial Methodology……118

List of Tables

CHAPTER II: INDIVIDUALIZING BMI TARGETS USING GENETICS AND CLINICAL BIOMARKERS

Table 2.1 – A list of a) GWAS consortia used to derive PRS, b) exposures, c) outcomes, and d) clinical biomarker variables used.....119

Table 2.2 - Baseline characteristics of participants in the UK Biobank (UKB)……..121

Table 2.3 - Association between BMI and major covariates used in analyses……....124

Table 2.4 – Collinearity between ALT, CRP, and TG phenotypes…… 126

Declaration of Academic Achievement

Irfan Khan was the lead author for this project, formulating the purpose of the project, research question, hypothesis, study design, methodology, and collecting, analyzing, and interpreting the data and final results. Dr. Guillaume Paré helped finalize the final purpose, research question, and hypothesis of the study, as well as with data interpretation

and writing the thesis. Michael Chong, Ann Le, Pedrum Mohammadi-Shemirani, Dr. Robert Morton, Christina Brinza, Michel Kiflen, Sukrit Narula, Dr. Loubna Akhabir, Dr. Shihong Mao, Dr. Katherine Morrison, and Dr. Marie Pigeyre helped with data collection, analysis and interpretation, as well as writing the thesis.

Chapter I: Introduction

1.1 Epidemiology and BMI

Body mass index (BMI) is a measurement used to categorize individuals into one of four weight classes: underweight (< 18.5 kg/m^2), normal (18.5-24.9 kg/m^2), overweight (25-29.9 kg/m^2), or obese (>= 30 kg/m^2), based on weight and height[1,2]. BMI is the current metric for determining anthropometric height and weight characteristics in adults[2]. It is commonly interpreted as an index of individual adiposity or "fatness"[2]. Due to the widespread acceptance in defining specific categories of BMI as health issues such as obesity, it has been useful in assessing its association with various outcomes in population-based studies and determining specific public health policies [2]. Since 1980, the global prevalence of people who are overweight or obese has doubled; nearly one-third of the world's population is categorized as overweight or obese, with certain regions experiencing this phenomenon more than others[8].

Causes and consequences of obesity not only have a severe impact on the individual, but also the health care system as a whole. In Canada, from 1994 to 2006, the annual costs of obesity ranged from 1.27 to 11.08 billion CAD[38]. Various interventions are often recommended for obesity and other related chronic conditions (e.g. type 2 diabetes (T2D), cardiovascular disease (CVD))[39,50]. Such recommendations include nutrition therapy, regular physical activity, psychological treatments, pharmacotherapy, and bariatric surgery[39,50]. Many of these interventions try to motivate obese individuals to lose weight, such that they fall into the normal WHO-defined BMI category[39]. Even if effective weight loss interventions were successfully followed through and individuals achieve a BMI in this normal range, there is no guarantee that their BMI is optimal for

their unique body composition and biology[1,2]. For example, variation in waist-to-hip ratio or lean mass can influence what BMI confers optimal health for individuals, as BMI does not take these factors into account[1,2].

The measurement of adiposity has gone through many iterations (e.g. Wt/Ht ratio, $Wt/Ht^{1.6}$) before widespread adoption of the now-familiar BMI weight/height^2 ratio (i.e. "the Quetelet Index", named after its inventor Dr. Quetelet in the 1800s) in the mid-1990s by the World Health Organization (WHO)[2]. Nevertheless, the recommended BMI ranges for optimal health is subject to debate[1,7,9,10]. The 'ideal' weight for health has been evolving over the years: before 1959, the ideal weight was determined based on height class (Wt/Ht ratio); from 1959 to 1983: the relative 'ideal' weight increased as height class increased, varying by age and gender; finally, from 1983 to the mid-1990s, the desirable range was considered to be equivalent to a BMI of < 27 kg/m^2. Ever since the late 1990s, the desirable range of BMI was set to 18.5-25.0 kg/m^2, after it was discovered based on observational studies that this was the range that best protected individuals in a given population from disease and mortality [2,7].

In the mid-2000s, there were an increasing number of epidemiological studies calling into question the established WHO-defined desirable range for BMI[45]. Up until 2006, epidemiological studies, such as the Framingham Heart Study, generally found a positive relationship between obesity and mortality[45]. However, since 2006, several studies have been showing inverse relationships between BMI and all-cause mortality[45]. In 2018, Bhaskaran et al. showed that in 3.6 million British participants, a J-shaped relationship between BMI and all-cause mortality (ACM) was found, with the lowest risk

of mortality at a BMI of 25-26 kg/m^2, outside of the recommended normal WHO BMI recommendation[6,7]. A J-shaped relationship is the non-linear relationship between an independent and dependent variable[43]. Although used interchangeably in some studies, the main difference between a J- and a U-shaped curve is that J-shapes do not have as strong of an inverse relationship between the independent and dependent variable at the lower extreme compared to a U-shaped curve[43]. J-shaped curves have been found between alcohol, blood pressure, caloric intake, and BMI and all-cause mortality[46]. Most statistical methods assume that a linear association underlies a given set of data: if the data deviates from linearity, it could lead to bias in the final estimation[43]. Model misspecification is a significant issue for both epidemiological and genetic studies, as conclusions may be drawn from studies that fit linear models to data not necessarily following a linear relationship[1,6,34]. The relationship between BMI and all-cause mortality is considered to be an example of data that does not necessarily fit a linear model[6]. In recent epidemiological and genetic studies, the relationship between BMI and all-cause mortality was able to be modelled using non-linear modeling techniques, such as restricted cubic spline modeling, which transforms the independent variable (e.g. BMI) being examined in order to fit a non-linear model to the data with the smoothest possible shape[44]. J-shaped relationships were also found between BMI and specific causes of mortality, such as cardiovascular or respiratory disease, corroborating with select epidemiological studies in the past that determined overweight individuals to have the lowest mortality risk in their respective samples[2,7]. Although discovery of J-shaped relationships and optimal thresholds for health between a given exposure and outcome is

important from a public health perspective, the way J-shaped relationships are communicated to the public holds more precedence[46]. For example, Bhaskaran et al. provided the sample size of participants in each WHO-defined BMI category as part of their 3.6 million UK participant cohort: 112,077 individuals were in the underweight category, 1,793,989 individuals were in the normal BMI category, 1,151,359 were in the overweight category, and 575,249 were in the obese category[6]. A significantly greater proportion of the population have a BMI to the right of the optimal BMI of 25.0 kg/m^2 compared to the left[6]. From a public health perspective, the J-shape curve is best contextualized through sample size differences across BMI categories, detailing the issue that is most epidemiologically important to address: in this case, targeting reduction of BMI as a priority [43]. Nevertheless, while this information may help guide public health policymakers on what would be the most pressing epidemiological issue to manage (i.e. reducing BMI in the population), there has not been any literature suggesting that such interventions are guaranteed to be successful in the long-term[35,50,52,53,54,55]. For example, reduced calorie intake, pharmacological treatments such as semaglutide or orlistat, or bariatric surgery have been shown to effectively decrease weight, yet long-term maintenance of weight loss can vary significantly among individuals (e.g. due to poor treatment adherence, obesogenic environments that promote increased food intake and decreased physical activity, and interactions between interventions and patient genetics), rendering interventions to be effective in some but not others in a given population[35,50,52,53,54,55]. J-shaped relationships could also show that not all interventions are effective for everyone: while it is true from epidemiological studies that there is no

safe level of tobacco or trans-fat consumption against mortality, studies have reported a safe level of alcohol consumption or BMI against mortality[46]. Lowering alcohol consumption or reducing BMI may not be the best intervention against mortality for everyone if their respective relationships follow a J-shape curve with mortality and there is an "optimum" BMI or level of alcohol consumption that provides the lowest mortality risk[46]. Public health policymakers also need to be careful when recommending interventions that take into account both extremes of the J-shaped relationship with mortality for a given risk factor[46]. Caloric intake, which strongly impacts BMI, has a J-shaped relationship with all-cause mortality, like BMI[46]. With caloric intake, the left side of the J-shaped curve represents deaths due to malnutrition while the right-hand side represents deaths due to obesity[46]. If done ineffectively, interventions to resolve either malnutrition or obesity could result in more harm than good[46]. For instance, there had been feeding regimens in Chile set in place to address the malnutrition epidemic in the country[46]. Instead, the feeding regimen resulted in a net increase in overweight and obesity in children that were from a higher socioeconomic status background[46]. As such, policymakers should make sure interventions are moderate in their approach, addressing both sides of the J-shaped relationship and optimizing the level of positive health outcomes.

Nevertheless, many of these epidemiological studies are prone to biases such as reverse causality or confounding variables (e.g. smoking), making it difficult to infer causal relationships between BMI and various different outcomes[12]. This further contributes to the lack of consensus in the optimal BMI associated with the lowest

mortality/disease risk. While many have suggested stronger alternatives to BMI in predicting disease risk, such as lean mass (LM) or waist-to-hip ratio (WHR), very few of these studies emphasized individual genomic influence in the context of BMI[1].

1.2 The genetics of BMI

Obesity is a heritable trait[3]. ~40-70% of the variation in BMI is due to genetic variation[3]. The large genetic variation in BMI may be a significant reason for why interventions for obesity may be more effective for some individuals compared to others[50,51,53,55]. Common genetic variation accounts for >20% of variation in BMI[3]. In 2015, Locke et al. determined that 97 loci were significantly associated with BMI at genome-wide significance (GWS), accounting for ~2.7% of variation of BMI[3]. In 2018, Yengo et al. included additional data from the UK Biobank (UKB), a European cohort, and found that 941 independent genetic variants or single nucleotide polymorphisms (SNPs) significantly associated with BMI[31]. These SNPs accounted for ~6.0% of the variation in BMI. Nevertheless, individually, genetic variants at genome-wide significance contribute very little to phenotypic variance and most of the genetic variation in BMI remains to be determined.

Despite the low explanatory power of SNPs on BMI variation, SNPs associated with BMI are often used in Mendelian randomization (MR) studies to establish causal relationships between BMI and a given outcome[32]. MR is able to establish causal relationships due to its application of Mendel's Law of Independent Assortment, which describes how alleles for a given phenotype are randomized at conception[32]. Such

randomization of alleles is akin to how individuals are randomized in a randomized control trial (RCT)[32]. Like RCT, randomization reduces the amount of influence confounding or reverse causality has on alleles[32]. Thus, alleles for a given genotype (often in the form of single nucleotide polymorphisms (SNPs)) can be used to make causal inferences between an exposure and an outcome[32]. Sun et al. determined that SNPs associated with BMI are positively associated with all-cause mortality through MR[1]. However, their findings were obtained through conventional MR methods that only assess the linear causal relationship between variables and are unable to account for any non-linear relationships. The relationship between BMI and all-cause mortality, as observational studies have shown, is considered non-linear[1,7]. In 2017, Staley and Burgess developed a non-linear MR (NLMR) method to assess the causal non-linear relationship between two variables[34]. Sun et al. would later use this method to determine whether the relationship between BMI and all-cause mortality is causal[1]. In their study, they showed that the J-shaped relationship between BMI and all-cause mortality was causal, with the lowest risk of all-cause mortality at a BMI range of ~22-25 kg/m^2, with factors such as smoking or sex influencing the BMI nadir[1]. Specific causes of mortality such as cardiovascular disease or cancer did not have a J-shaped relationship with BMI, which was in contrast with previous epidemiological literature[1,7]. While results from MR studies validate many of the results seen in previous observational studies, they are unable to reveal the optimal BMI at the individual level, given that these causal associations represent population averages. Furthermore, no study has shown whether individual

genetics for either BMI or related factors related to BMI can be used to personalize BMI for optimize health (e.g. optimal protection from mortality).

There have been limited studies that have investigated the beneficial and adverse effects of BMI. Non-obese "metabolically unhealthy" individuals (e.g. those hyperglycemic and/or dyslipidemic) have worse cardiovascular disease (CVD)-related outcomes compared to obese "metabolically healthy" individuals (e.g. those normoglycemic and/or having normal lipid levels). The heterogenous nature of BMI could complicate clinical recommendations for BMI, given that global BMI recommendations from the WHO may not fully reflect metabolic health. Genetic variants that increase BMI had a heterogenous effect on risk for poor metabolic health, with certain variants increasing risk and others decreasing it[4,7]. Therefore, there may be benefit in calculating each individual's 'optimal' BMI through genetic risk scores as it may help account for the genetic heterogeneity of BMI.

1.3 Biomarkers of BMI

Although BMI is an easy measurement for general adiposity, studies have shown it might not be the most reliable way of quantifying adiposity and consequently, a risk factor of disease or mortality[7]. Non-genetic factors such as clinical biomarker levels are associated with BMI and can be used as a more robust proxy for adiposity[7]. For example, triglyceride (TG), the main constituent of body fat in humans, C-reactive protein (CRP), a major marker of inflammation, and alanine aminotransferase (ALT), a marker of liver

function, are associated with BMI[40,41,42]. While it has been assumed that these associations are due to the effect BMI has on these biomarkers, rather than these biomarkers having an effect on BMI, recent evidence has suggested that this might not always be the case: more complex, bi-directional relationships could be present [41,42]. For example, while the majority of epidemiological studies have shown that ALT is positively associated with BMI, results from MR show that ALT is negatively associated with BMI, which is contrary to the consensus regarding the BMI-ALT relationship[42]. More importantly, MR indicated that ALT causally impacts BMI, which may not be biologically plausible.[42] Previous literature has shown that obesity is a major risk factor for non-alcoholic fatty liver disease (NAFLD)[48,49]. Excess adiposity in obese individuals leads to adipose tissue dysfunction and insulin resistance, which subsequently leads to lipolysis or the breakdown of adipose tissue into free fatty acids[49]. Free fatty acids and leptin levels increase, and adiponectin decreases, leading to intrahepatic accumulation of fat[49]. Excess intrahepatic fat leads to mitochondrial dysfunction and oxidative stress, which leads to liver inflammation[49]. Prolonged inflammation can result in the scarring of liver tissue, or fibrogenesis, eventually leading to impaired liver tissue regeneration[49]. Ultimately, impaired regeneration leads to liver cell death and dysfunction[49]. Higher levels of ALT are a marker of liver dysfunction[48,49]. Thus, results from this MR study may have been due to BMI and ALT sharing a common genetic architecture – future studies may need to use gene expression analyses such as cis-protein quantitative trait loci (pQTL) MR to confirm whether ALT truly has a causal impact on BMI through its genetic pathways[42,47]. Otherwise, the consensus that BMI and ALT are positively

correlated, with high ALT as a marker of high BMI, still holds true despite the findings from this MR study. MR analyses looking at BMI-TG and BMI-CRP results were largely consistent with previous epidemiological literature: BMI influences TG and CRP but the reverse is not true[40,41,42].

1.4 Non-genetic components of BMI

Other non-genetic components associated with BMI include diet, physical activity, and sleep. A healthy dietary pattern, such as one rich in fruits and vegetables, was negatively associated with BMI, compared to an unhealthy diet, or a diet rich in meat and fat[35]. Lower exercise capacity was found in one study to be a significant risk factor for all-cause mortality, theorizing that exercise may influence BMI to impact all-cause mortality risk, though this has not been confirmed[36]. The association between BMI and sleep duration can depend on age: a negative and U-shaped association was found in young adults (18-29 years old) and middle-aged adults (30-64 years old) respectively[37].

Nevertheless, no study to date has examined the influence of these non-genetic factors on individual optimal BMI using epidemiological or genetic tools.

1.5 Objectives

The objective of this thesis is to determine whether variation in genetic and non-genetic factors associated with BMI causes variation in the optimal BMI. Our first objective will be to determine whether the optimal BMI varies due to genetic and clinical biomarker variation using Cox meta-regression. Our second objective is to determine

whether optimal BMI variation due to genetic and clinical biomarker variation is causal using non-linear mendelian randomization (NLMR).

1.6 Hypothesis

For objective 1, genetic and clinical biomarker level differences in adiposity, anthropometric, and metabolic measures will result in inter-individual variation in the optimal BMI. For objective 2, the NLMR-determined optimal BMI will not be different from the Cox meta-regression-determined optimum. For all objectives, the NLMR-determined optimal BMI and Cox meta-regression-determined optimum will be distinct from "normal" WHO BMI recommendations (i.e. 18.5-25 kg/m^2).

1.7 References

1. Sun YQ, Burgess S, Staley JR, *et al*. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non- linear mendelian randomisation analyses. *BMJ* 2019*;* **364**: 1-10. DOI: 10.1136/bmj.l1042.

2. Nuttall FQ. Body Mass Index: Obesity, BMI, and Health: A Critical Review. *Nutr Today* 2015; **50**: 117-128. DOI: 10.1097/NT.0000000000000092.

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

4. Paré G, Mao S, Deng WQ. (2019). A robust method to estimate regional polygenic correlation under misspecified linkage disequilibrium structure. *Genet Epidemiol* 2019; **42**: 636-647. DOI: 10.1002/gepi.22149.

5. Marees AT, de Kluiver H, Stringer S, *et al* (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res* 2018; **27**: 1-10. DOI: 10.1002/mpr.1608.

6. Bhaskaran K, Dos-Santos-Silva I, Leon DA, *et al*. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3.6 million adults in the UK. *Lancet Diabetes Endocrinol* 2018; **6**: 944-953. DOI: 10.1016/S2213-8587(18)30288-2.

7. Pigeyre M, Sjaarda J, Mao S, *et al*. Identification of novel causal blood biomarkers linking metabolically favorable adiposity with type 2 diabetes risk. *Diabetes Care* 2019; **42**: 1800-1808. DOI: 10.2337/dc18-2444.

8. Rothman KJ. BMI-related errors in the measurement of obesity. *Int J Obes (Lond)* 2008; **32**:1-4. doi: 10.1038/ijo.2008.87.

9. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med* 2020; **12:** 1-11. https://doi.org/10.1186/s13073-020-00742-5.

10. Wolfe BM, Kvach E, Eckel RH. Treatment of Obesity: Weight Loss and Bariatric Surgery. Circ Res. 2016;**118**:1844-1855. doi:10.1161/CIRCRESAHA.116.307591

11. Echouffo-Tcheugui JB, *et al*. Natural history of obesity subphenotypes: dynamic changes over two decades and prognosis in the Framingham Heart Study. *J Clin Endocrinol Metab* 2019;**104**:738–752. DOI: 10.1210/jc.2018-01321.

12. Henneberg M, Ulijaszek SJ. Body frame dimensions are related to obesity and fatness: Lean trunk size, skinfolds, and body mass index. *Am J Hum Biol*. 2010;**22**:83-91. doi: 10.1002/ajhb.20957.

13. Hu G, Tuomilehto J, Silventoinen K, *et al*. The effects of physical activity and body mass index on cardiovascular, cancer and all-cause mortality among 47 212 middle-aged Finnish men and women. *Int J Obes (Lond).* 2005;**29**: 894-902. doi: 10.1038/sj.ijo.0802870.

14. Mbatchou J, Barnard L, Backman J, *et al*. Computationally efficient whole genome regression for quantitative and binary traits. *bioRxiv* 2020; Published online January 1. DOI:10.1101/2020.06.19.162354.

15. Seidelmann SB, Claggett B, Cheng S, *et al*. Dietary carbohydrate intake and mortality: a prospective cohort study and meta-analysis. *Lancet Public Health.* 2018;**3**:419-428. doi: 10.1016/S2468-2667(18)30135-X.

16. Scrucca L, Santucci A, Aversa F. Competing risk analysis using R: an easy guide for clinicians. *Bone Marrow Transplant*, 2007; **40**, 381–387. DOI: 10.1038/sj.bmt.1705727.

17. Higgins JP, Thompson SG, Deeks JJ, *et al*. Measuring inconsistency in meta-analyses. BMJ. 2003;**327**:557-560. doi:10.1136/bmj.327.7414.557

18. Bhaskaran K, Douglas I, Forbes H, *et al*. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet* 2014;**384**:755-65. DOI: 10.1016/S0140-6736(14)60892-8.

19. Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. Stat Med.

1997;**16**:753-68. doi: 10.1002/(sici)1097-0258(19970415)16:7<753::aid-sim494>3.0.co;2-g.

20. Kautzky-Willer A, Harreiter J, Pacini G. Sex and Gender Differences in Risk, Pathophysiology and Complications of Type 2 Diabetes Mellitus. *Endocr Rev* 2016;**37**:278-316. DOI:10.1210/er.2015-1137

21. Censin JC, Peters SAE, Bovijn J, *et al*. Causal relationships between obesity and the leading causes of death in women and men. *PLoS Genet* 2019; **15**:e1008405. DOI: 10.1371/journal.pgen.1008405.

22. Khera AV, Chaffin M, Wade KH, *et al*. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell* 2019;**177**:587-596. DOI:10.1016/j.cell.2019.03.028.

23. Richardson TG, Harrison S, Hemani G, *et al*. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *Elife* 2019;**8**:e43657. DOI: 10.7554/eLife.43657.

24. Howe LJ, Dudbridge F, Schmidt AF, *et al*. Polygenic risk scores for coronary artery disease and subsequent event risk amongst established cases. *Human Molecular Genetics* 2020; **29**, 1388–1395. DOI: 10.1093/hmg/ddaa052.

25. Manning AK, Hivert MF, Scott RA, *et al*. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genetics* 2012;**44**:659-669. DOI: 10.1038/ng.2274.

26. Udler MS, McCarthy MI, Florez JC, *et al*. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocr Rev* 2019;**40**:1500-1520. DOI: 10.1210/er.2019-00088

27. Shamai L, Lurix E, Shen M, *et al*. Association of body mass index and lipid profiles: evaluation of a broad spectrum of body mass index patients including the morbidly obese. *Obes Surg* 2011;**21**:42-7. doi: 10.1007/s11695-010-0170-7.

28. Morin S, Leslie WD; Manitoba Bone Density Program. High bone mineral density is associated with high body mass index. *Osteoporos Int* 2009;**20**:1267-1271. DOI: 10.1007/s00198-008-0797-6.

29. Clifton EA, Day FR, De Lucia Rolfe E, *et al*. Associations between body mass index-related genetic variants and adult body composition: The Fenland cohort study. *Int J Obes (Lond)* 2017;**41**:613-619. DOI:10.1038/ijo.2017.11.

30. Sahoo K, Sahoo B, Choudhury AK, *et al*. Childhood obesity: causes and consequences. *J Family Med Prim Care* 2015;**4**:187-192. DOI:10.4103/2249-4863.154628.

31. Yengo L, Sidorenko J, Kemper KE, *et al*. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. *Hum Mol Genet*. 2018;**27**(20):3641-3649. DOI:10.1093/hmg/ddy271

32. Bennett, D.A., & Holmes, M.V. Mendelian randomisation in cardiovascular research: an introduction for clinicians. *Heart.* 2017;**103**(18): 1400-1407. doi: 10.1136/heartjnl-2016-310605

33. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet*. 2018;**27**(R2):R195-R208. doi:10.1093/hmg/ddy163

34. Staley JR & Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genet Epidemiol.* 2017; **41**(4): 341-352. doi: 10.1002/gepi.22041

35. Mu M, Xu LF, Hu D, *et al*. Dietary Patterns and Overweight/Obesity: A Review Article. *Iran J Public Health.* 2017;**46**(7):869-876. doi: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5563867/pdf/IJPH-46-869.pdf. Published July 2017. Accessed April 2021.

36. McAuley PA, Blaha MJ, Keteyian SJ, *et al*. Fitness, Fatness, and Mortality: The FIT (Henry Ford Exercise Testing) Project. *Am J Med*. 2016;**129**(9):960-965. doi:10.1016/j.amjmed.2016.04.007.

37. Grandner MA, Schopfer EA, Sands-Lincoln M, *et al*. Relationship between sleep duration and body mass index depends on age. Obesity (Silver Spring). 2015;**23**(12):2491-2498. doi:10.1002/oby.21247.

38. Tran BX, Nair AV, Kuhle S, Ohinmaa A, Veugelers PJ. Cost analyses of obesity in Canada: scope, quality, and implications. *Cost Eff Resour Alloc*. 2013;**11**:3. Published 2013 Feb 8. doi:10.1186/1478-7547-11-3

39. Musich S, MacLeod S, Bhattarai GR, *et al*. The Impact of Obesity on Health Care Utilization and Expenditures in a Medicare Supplement Population. *Gerontol Geriatr Med*. 2016;**2**:2333721415622004. Published 2016 Jan 19. doi:10.1177/2333721415622004

40. C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. *Int J Obes (Lond)* 2011; **35**, 300-308. DOI:10.1038/ijo.2010.137

41. Shamai L, Lurix E, Shen M, *et al*. Association of body mass index and lipid profiles: evaluation of a broad spectrum of body mass index patients including the morbidly obese. *Obes Surg* 2011;**21**:42-7. doi: 10.1007/s11695-010-0170-7.

42. Liu, J, Au Yeung, SL, Kwok, MK, *et al.* The effect of liver enzymes on adiposity: a Mendelian randomization study. *Sci Rep* **9,** 16792 (2019). https://doi.org/10.1038/s41598-019-52489-8

43. Lemmens P. U-Shaped Curve. In: Encyclopedia of Research Design. Thousand Oaks, CA: SAGE Publications, Inc; 2012: 1587- 1588.

44. Gauthier J, Wu QV & Gooley TA. Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians. *Bone Marrow Transplant*; 2020; **55**, 675–680. DOI: 10.1038/s41409-019-0679-x

45. Ades PA, Savage PD. The obesity paradox: perception vs knowledge. Mayo Clin Proc. 2010;**85**(2):112-114. doi:10.4065/mcp.2009.0777

46. Chokshi DA, El-Sayed AM, Stine NW. J-Shaped Curves and Public Health. *JAMA.* 2015 Oct 6;314(13):1339-40. doi: 10.1001/jama.2015.9566. PMID: 26390229

47. Sun BB, Maranville JC, Peters JE, *et al*. Genomic atlas of the human plasma proteome. Nature. 2018;**558**:73-79. doi:10.1038/s41586-018-0175-2

48. Bayard M, Holt J, Boroughs E. Nonalcoholic fatty liver disease. *Am Fam Physician*. 2006;**73**(11):1961-8. Retrieved from: https://www.aafp.org/afp/2006/0601/p1961.html

49. Polyzos SA, Kountouras J, Mantzoros CS. Obesity and nonalcoholic fatty liver disease: From pathophysiology to therapeutics. *Metabolism.* 2019;**92**:82-97. doi: 10.1016/j.metabol.2018.11.014. Epub 2018 Nov 29. PMID: 30502373.

50. Wharton S, Lau DCW, Vallis, M, *et al*. Obesity in adults: a clinical practice guideline. *CMAJ*. 2020; **192**: E875-E891. doi: 10.1503/cmaj.191707

51. Lamiquiz-Moneo I, Mateo-Gallego R, Bea, AM, *et al*. Genetic predictors of weight loss in overweight and obese subjects. *Sci Rep*. 2019; **9**, 1-9. doi: 10.1038/s41598-019-47283-5.

52. Wilding JPH, Batterham RL, Calanna S, *et al*. Once-Weekly Semaglutide in Adults with Overweight or Obesity. *N Engl J Med*. *2021*;**384**:989-1002. doi:10.1056/NEJMoa2032183.

53. Hall KD, Kahan S. Maintenance of Lost Weight and Long-Term Management of Obesity. *Med Clin North Am.* 2018;**102**:183-197. doi:10.1016/j.mcna.2017.08.012

54. Torgerson JS, Hauptman J, Boldrin MN, *et al*. XENical in the prevention of diabetes in obese subjects (XENDOS) study: a randomized study of orlistat as an adjunct to lifestyle changes for the prevention of type 2 diabetes in obese patients. *Diabetes Care.* 2004;**27**:155-61. doi: 10.2337/diacare.27.1.155.

55. Reddon H, Gerstein HC, Engert JC, *et al*. Physical activity and genetic predisposition to obesity in a multiethnic longitudinal study. *Sci Rep.* 2016 Jan 4;**6**:1-10. doi: 10.1038/srep18672.

Chapter II: Individualizing BMI Targets Using Genetics and Clinical Biomarkers

2.1 Introduction

In the 2018 study by Bhaskaran et al., all-cause mortality, cancer, cardiovascular disease (CVD), and respiratory diseases had a J-shaped relationship with BMI, with the lowest risk in the 21-26 kg/m^2 range[6]. The J-shaped relationship also varied by age: individuals younger than 70 had an optimal BMI of 23 kg/m^2 while individuals aged 70 and older had an optimal BMI of 25 kg/m^2, even after adjusting for multiple confounding variables, such as age, sex, alcohol use, or socioeconomic status [6]. Smoking status had a modest confounding influence on this relationship: when the population was restricted to those who never smoked, the J-shaped curve was slightly less prominent[6]. With regards to a sex effect, the association of BMI with higher risk of mortality was much stronger in males compared to females[6]. Their study is generally consistent with other studies investigating the relationship between BMI and mortality, despite varying cohort composition, methods of defining BMI between studies, and number and type of covariates adjusted for across analyses[3,6]. Similar to Bhaskaran et al., J-shaped relationships between BMI and mortality were found, with the optimal BMI ranging from 20-25 kg/m^2, and the most significant variation in the relationship occurring between smokers and non-smokers[6]. Nevertheless, these studies assess population trends without consideration for inter-individual variation in genetic and non-genetic factors such as clinical biomarkers, physical activity, diet, or sleep duration.

Single nucleotide polymorphisms (SNPs) associated with a certain phenotype can be combined by effect size into one aggregated risk score or polygenic risk score (PRS)[22]. PRS are quantitative measures for genetic predisposition to a trait[22]. Aggregating SNPs

addresses the key limitation of modest SNP effects on a given phenotype in most genetic studies[32]. In the context of BMI, PRS can be used to measure individual genetic predisposition for high BMI. Indeed, Khera et al. built a PRS for BMI and found that mean BMI was 30.0 kg/m^2 for individuals within the top 5th percentile of the PRS and 25.2 kg/m^2 for individuals in the bottom 5th percentile[22]. This confirms that higher individual genetic risk correlates with higher BMI in these individuals. However, to the best of our knowledge, no study has tested whether genetic variation in BMI (or non-genetic variation in other adiposity, anthropometric, or metabolic factors) influences the relationship between BMI and mortality.

The objective of this chapter is to determine whether the optimal BMI varies based on genetic and non-genetic variation. Our hypothesis is that the optimal BMI will vary based on inter-individual variation of these factors.

2.2 Methods

2.2.1 Study Population

The UK Biobank (UKB) is a prospective cohort of >500,000 individuals, aged 40-69[9]. Comprehensive genotypic and phenotypic data were collected from 2006-2010[9]. The latest UKB dataset was issued on October 17th, 2020, and included 388,115 unrelated British individuals. At the extreme ends of the BMI spectrum (< 15 kg/m^2 or > 50 kg/m^2), statistical power to detect effect sizes with a reasonable degree of confidence significantly decreases

due to small sample size[1,7,11]. Thus, 423 individuals outside of the 15-50 kg/m^2 range were excluded from subsequent analyses. The final population sample size for analysis was 387,692. Outcomes were defined based on the International Classification of Diseases, 10th revision (ICD-10) codes as described elsewhere (Table 2.1c)[6].

Analyses were divided into primary and secondary analyses: the former is defined by analyses conducted on 387,692 individuals while the latter refers to subgroup and clinical characteristic analyses. In the primary analyses, all-cause mortality, as well as cancer, CVD, respiratory disease, and other disease (i.e., death due to diseases/conditions other than cancer, CVD, or respiratory disease) mortality were the main outcomes. In secondary analyses, all-cause mortality was the only outcome analyzed. Clinical characteristic analyses including waist-to hip ratio (WHR), hemoglobin A1C (Hb1Ac), triglyceride (TG), low-density lipoprotein (LDL), high-density lipoprotein (HDL), C-reactive protein (CRP), cholesterol, and alanine aminotransferase (ALT) were used to assess non-genetic variation in the optimal BMI. Subgroups were defined based on PRS, smoking status, diabetes status, age, sex, dietary factors, physical activity levels (excess metabolic equivalent of task (MET)-h/week), and sleep duration[2,6].

2.2.2 Polygenic Risk Score Calculation

PRS are quantitative measures for genetic disposition to a certain trait derived from the weighted effects of each genetic variant or SNP on the phenotype[22]. SNPs associated with each metabolic phenotype obtained from genome-wide association study (GWAS) consortium data were then incorporated into PRS (see Table 2.1a for GWAS sources used

to derive the PRS)[5,14]. LASSOSUM was used to generate PRS, which is a method that uses a non-Bayesian penalized regression to generate PRS as described elsewhere[14]. UKB participants were excluded from any selected GWAS data used for PRS, in order to avoid circularity.

Although GWAS focus on the association of SNPs with a given phenotype (e.g. if SNPs have a positive, negative, or neutral association with BMI), they lack the ability to account for heterogeneity in effects associated SNPs have on correlated traits (e.g. Type 2 Diabetes (T2D))[4] .GWAS are thus unable to determine whether variants positively associated with BMI ("BMI increasing variants") have a protective effect on a correlated phenotype, like T2D[4]. Regional polygenic correlation (RPC) was used to account for SNP effect heterogeneity through the genetic correlation between BMI or T2D[4,7]. BMI variants extracted from GWAS data from the GIANT consortium were divided into three polygenically determined categories based on their influence on T2D risk: 1) metabolically favourable (MF or BMI-increasing variants associated with decreased T2D risk), 2) metabolically deleterious (MD or BMI-increasing variants associated with increased T2D risk), and 3) metabolically neutral (MN or BMI-increasing variants with no association with T2D risk) (Figure 2.1) [4,7].

A PRS was created for whole genome and polygenically determined BMI, as well as TG, LM, and WHR.

2.2.3 Cox Meta-Regression Model

Non-linear cox proportional hazard models were used to determine the association between BMI and mortality outcomes. Deaths were censored if participants had a previous diagnosis of CVD, cancer, or respiratory disease[6,36]. If baseline disease has an impact on baseline BMI measurements, this may introduce bias due to reverse causality; thus, censoring deaths in this manner accounts for this bias. In a separate sensitivity analysis, deaths were additionally censored for participants diagnosed with CVD, cancer, or respiratory disease within 2 years of the baseline BMI measurement. This was to take into account that disease progression could take time and still result in reverse causation[6]. Our primary model, the minimally adjusted model, included age, sex, and the first 10 genetic principal components (measure for genetic ancestry). Our secondary model, the fully adjusted model, included smoking, UKB assessment centre, alcohol consumption, index of multiple deprivation (metric for socioeconomic status (SES)), and diabetes status[6]. An unadjusted model was also used as a sensitivity analysis. Covariates were selected based on published evidence and strong association with BMI or mortality (Table 2.3)[6].

UKB participants were stratified into quantiles according to each PRS. Twenty quantiles were used in the primary analysis while 10 quantiles were used in the secondary analyses. Non-linear cox regression analyses between BMI and each mortality outcome were conducted within each quantile. In non-linear cox regression, a non-linear model in the form of a restricted cubic spline was fit to the data. The BMI associated with the lowest risk of mortality was determined for each quantile. A linear regression analysis

was then conducted to determine if the BMI-mortality nadir values varied across quantiles (Figure 2.2, using the BMI PRS as the example). The linear regression of results obtained from Cox regression is where the term "Cox meta-regression," is derived from. In separate sensitivity analyses, primary and secondary analyses were repeated with BMI adjusted for each respective PRS used to divide the sample into quantiles. Additionally, the population optimal BMI for primary and secondary analyses was determined by repeating the above analysis without dividing the population/subgroup by quantiles (Figure 2.3).

Confidence intervals for the population optimal BMI were computed using a bootstrapping procedure (Figure 2.3A-C). Participants in the UKB population were re-sampled with replacement 10,000 times. The optimal BMI was computed using the Cox meta-regression model for each re-sample. 95% confidence intervals were computed from the bootstrap distribution generated from these 10,000 optimal BMI values.

Heterogeneity between subgroups was assessed using a fixed effects meta-analysis model[17,19]. All statistical analyses were conducted using R (3.6.0 - 3.6.3). Statistical significance was set at two-tailed p values < 0.05. Bonferroni correction for multiple hypothesis testing was set at i) 0.05/7 = 0.007 for our primary analyses, where p is divided by the number of PRS analyzed; ii) 0.05/4 = 0.0125 for our cause-specific mortality analyses, where p was divided by the number of cause-specific mortality outcomes; and at iii) 0.05/60 = 0.0008 for our subgroup analyses, where p was divided by the number of subgroups analyzed.

We completed tertiary analyses to supplement the primary and secondary analyses. We determined whether WHR and LM followed a J-shaped relationship with mortality outcomes and if so, whether the optimal BMI, if any, varied due to WHR or LM PRS variation respectively using the methods mentioned above. The minimally adjusted model was used, along with models adjusted for height and height^2 to account for the strong correlation between lean mass and height as well as non-linear effects of height[6,44,45].

For the subgroup analyses, the median value for each continuous variable was determined in order to dichotomize UKB participants into two groups based on continuous variables such as triglyceride level or sleep duration. Participants were categorized as having levels above or below the median value of a given continuous variable and then split into two groups based on their categorization.

2.3 Results

2.3.1 Optimal BMI across Quantiles of PRS in the UKB Population

The baseline characteristics of study participants are shown in Table 2.2. The population optimal BMI in the UKB was 25.0 kg/m^2 [95% CI 24.6-25.5 kg/m^2] and 23.7 kg/m^2 [23.5 – 23.9 kg/m^2] using the minimally adjusted and unadjusted models, respectively. We tested polygenically-determined BMI, as well as TG, LM, and WHR PRS for their influence on optimal BMI. The whole genome BMI PRS was significantly associated with optimal BMI

variation across quantiles after adjustment for multiple hypothesis testing (0.10 kg/m^2 per quantile [95% CI 0.04-0.17], $P = 0.005$ ($p < 0.007$ for significance); Figure 2.4). Results remained consistent after censoring participants diagnosed with cancer, CVD, or respiratory disease within 2 years of the baseline BMI measurement (0.11 kg/m^2 per quantile [95% CI 0.04-0.17], $P = 0.003$). Thus, there would be a difference of 1.90 kg/m^2 in the optimal BMI between individuals in the bottom and top 5th BMI PRS percentile. Comparing results with the mean BMI, there is a difference of 3.90 kg/m^2 in mean BMI between bottom and top 5th BMI PRS percentile. Altogether, higher genetic risk for BMI is associated with both increased mean and optimal BMI, but the increase in optimal BMI is not as pronounced. A person with an optimal BMI of 23.5 kg/m^2 in the bottom 5th BMI PRS percentile is predicted to have an optimal BMI of 25.4 kg/m^2 in the top 5th BMI PRS percentile. In other words, an individual in the bottom 5th BMI PRS percentile has an ideal BMI target within the normal WHO-defined BMI range; however, an individual in the top 5th BMI PRS percentile would have an optimal BMI in the overweight range (Figure 2.4).

Using the minimally adjusted model, the difference in BMI nadir between quantiles of the TG PRS was nominally significant (0.10 kg/m^2 per quantile [95% CI 0.03-0.17], $p = 0.01$) (Figure 2.4). Results with the TG PRS remained nominally significant with the fully adjusted model (0.14 kg/m^2 per quantile [95% CI 0.04-0.23], $p = 0.009$ ($p < 0.008$ for significance)) (Figure 2.4). No other PRS was associated with mortality nadir.

All analyses looking at the change in optimal BMI per 1-unit increase in a given PRS quantile using the minimally adjusted model were either nominally significant or not significant, regardless of PRS used.

2.3.2 Subgroup Analyses

The analyses above were repeated for the following UKB subgroups: i) never smokers and ever smokers; ii) diabetics and non-diabetics; iii) younger adults (<= 65 years old) and older adults (> 65 years old); iv) males and females; v) participants above and below median v.1) mean and/or percent carbohydrate, fat, protein, sugar, and energy intake, v.2) physical activity (MET-hr/week), v.3) sleep duration (hrs/day), v.4) WHR, TG, LDL, HDL, Hb1Ac, CRP, cholesterol, and ALT clinical biomarkers, and v.5) whole genome BMI, metabolically favourable BMI, metabolically deleterious BMI, metabolically neutral BMI, LM, WHR, and TG PRS. These subgroups were not tested in the primary analyses and we only used the minimally adjusted model for testing (Figure 2.5). There were significant differences between individuals above and below median TG, C-reactive protein (CRP), and alanine aminotransferase (ALT) in population optimal BMI (TG: p heterogeneity: 1.29×10^{-4}; CRP: p heterogeneity = 7.92×10^{-5}; ALT: p heterogeneity = 2.70×10^{-8}). The differences in population optimal BMI mirrored the differences in mean BMI between these subgroups (Figure 2.5; TG: p heterogeneity: 0, CRP: p heterogeneity = 0, ALT: p heterogeneity = 0). Optimal BMI variation was also present between percentiles of the TG, CRP, and ALT traits (Figure 2.7; TG: p heterogeneity: 2.00×10^{-4}, CRP: p heterogeneity = 1.26×10^{-6}, ALT: p heterogeneity = 5.00×10^{-4}).

Independent of genetic variation, there were nominally significant differences in population optimal BMI between diabetics and non-diabetics and individuals above and below median sleep duration; TG, metabolically neutral BMI, and BMI PRS (not shown;

diabetes diagnosis: p heterogeneity: 0.0009, sleep duration: p heterogeneity = 0.009, TG PRS: p heterogeneity = 0.002, metabolically neutral BMI PRS: p heterogeneity = 0.004, and BMI PRS: p heterogeneity = 0.03).

2.3.3 Cause-Specific Analyses

The above analyses were repeated, focusing only on cancer, CVD, respiratory disease, and other disease (i.e. death due to all other diseases) mortality using the minimally adjusted model (Figure 2.6, shown using the BMI PRS). There were no significant differences between mortality subtypes in population optimal BMI or PRS effect per quantile (population: p = 0.19 , PRS effect per quantile: p = 0.82).

2.3.4 Tertiary Analyses

WHR did not follow a J-shaped relationship with all-cause mortality (Figure 2.8a). Thus, there was no optimal WHR with which to compare PRS percentiles. LM followed a W-shaped relationship (i.e., two inflection points) with all-cause mortality, with the lowest risk of mortality at 43.4 kg (Figure 2.8b). Even after adjusting for height and height^2, the W-shape stayed consistent, although the negative relationship between LM and all-cause mortality on the left-hand side of the curve was slightly attenuated (Figure 2.8e and f). The W-shape was explained by a sex-specific bimodal distribution in LM, given that the nadir at 63.2 kg approximately corresponded with the nadir for LM and all-cause

mortality in the males only subgroup at 59.1 kg (Figure 2.8c) and the nadir at 43.4 kg approximately corresponded with the nadir for LM and all-cause mortality in the females only subgroup at 41.5 kg (Figure 2.8d). However, the optimal LM did not vary with LM PRS variation (-0.47 kg/m^2 per quantile [95% CI -1.30-0.35], p = 0.24, Figure 2.9).

2.4 Discussion

2.4.1 Summary of Results

The results demonstrate that the BMI that confers the lowest risk for mortality, varies for each individual. Specifically, individual genetic variation and circulating triglycerides, C-reactive protein, and alanine aminotransferase modify the BMI-mortality association. In conclusion, the WHO definition of 'optimal' BMI may be misleading as it does not account for inter-individual variation in genetic or clinical biomarker profile.

2.4.2 Clinical Implications

First, independent of genetic or clinical factors, UKB participants had an optimal BMI of ~25.0 kg/m^2, which is 0.1 - 6.5 kg/m^2 beyond the normal WHO BMI recommendations. Our findings are consistent with previous studies and further undermines current recommendations of BMI: certain individuals may have an optimal BMI well into the WHO overweight BMI range through a combination of population (i.e. UKB) and genetic effects[6]. Higher polygenically determined BMI was associated with increased mean BMI and to a lesser degree, increased optimal BMI. Although reducing one's BMI from the mean to the optimal BMI in the top 5th percentile means the difference between a

"normal" or "overweight" category, it is unclear whether there is any benefit in reducing one's BMI to the optimum value when the mean BMI value is only 2 kg/m^2 higher. Further studies should assess the relative benefit of reducing BMI to the genetically-determined optimum against various disease and mortality outcomes at the individual level.

Second, there is variation in optimum BMI within populations, linked with genetic and clinical biomarker variation. Prior studies have challenged the idea that the normal 18.5 – 24.9 kg/m^2 BMI range is always associated with the lowest risk of death after demonstrating J-shaped relationships between BMI and mortality outcomes[1,6]. According to our primary results, if a clinician were to set a 22 kg/m^2 BMI target for a patient in the highest BMI PRS quantile, where the target BMI is actually 26.3 kg/m^2, current recommendations may not be fitted for the best health interests of this patient, even if 22 kg/m^2 appears to be healthy according to "normal" WHO guidelines. Similarly, it would be misguided to set BMI targets based on the population optimum BMI for these individuals, which from our primary analyses, is 25.0 kg/m^2. From our results, variation in circulating TG, CRP, and ALT levels may mean the difference between the optimal BMI value being classified as "normal" or "overweight" according to the WHO, although we cannot conclude that this variation in the optimal BMI is causal. Thus, it may be misguided to make BMI recommendations strictly based on the WHO guidelines or the population optimum BMI for individuals, as their optimal BMI may be above or below these thresholds due to genetic or clinical biomarker variation.

2.4.3 Research Implications

Although we see variation in the optimal BMI due to genetic variation, this may be a result of individuals having a higher BMI due to higher genetic predisposition for BMI. Khera et al. found a positive association between mean BMI and BMI PRS decile[22]. Our study found the same association between mean BMI and PRS quantile, which followed the same direction as the association between optimal BMI and PRS quantile. Nevertheless, the distinction between the predicted optimal and mean BMI difference between individuals in the top and bottom 5th BMI PRS percentile is large enough to conclude that changes in optimal BMI do not simply reflect changes in mean BMI. It is unknown whether the higher optimal BMI for those with high genetic predisposition is beneficial (e.g., metabolically favourable BMI or subcutaneous fat) or detrimental (e.g., metabolically deleterious BMI or visceral/abdominal fat). Finally, while we saw significant variation in the population optimal BMI due to variation in circulating triglyceride (TG) levels, variation in the optimal BMI due to variation in polygenically determined TG was nominally significant. Since these results are not causal, it is still unclear how much of TG-mediated optimal BMI variation is due to genetics. PRS only uses estimates derived from SNPs without taking into account complex interactions of these SNPs with the environment, whereas the traits these SNPs are associated with encompass variability that SNPs cannot capture[36]. The PRS for TG is not capturing enough of the variation in the optimal BMI as the TG trait itself, reflecting the underpowered nature of TG PRS and PRS in general[9]. The TG trait is better powered at capturing optimal BMI variation as it is less prone to misspecification, unlike the PRS,

where genetic variants may be associated with traits other than TG via pleiotropic effects. Future studies would need to conduct larger and more powered GWAS of TG to determine whether or not there is a greater percentage of variation in TG explained by genetics that is not able to be captured using relatively smaller sample sizes[37]. A greater percentage of variation in TG explained by genetics will help improve predictive power of the PRS, though not to the extent the original TG trait has in predictive power[9].

Of note is that while primary analysis results were significant using the minimally adjusted model, they were no longer significant using the fully adjusted and unadjusted model. The relationship between BMI PRS quantile and optimal BMI per quantile becomes flatter with the fully adjusted and unadjusted model, indicating that certain confounders, either unadjusted or adjusted for, have a greater influence on optimal BMI variation than genetic variation. It may also mean that if perfectly adjusted for, the relationship could become completely neutral, where the optimal BMI does not differ with BMI PRS variation. The difference between the unadjusted and minimally adjusted models is that age, sex, and genetic ancestry were adjusted for in the latter, meaning either one or more of these covariates is introducing bias that underestimates the true association between BMI PRS quantile and optimal BMI per quantile. Such covariates are known as negative confounders[38]. However, as all three of these covariates are strongly associated with the BMI PRS and we cannot confirm causality using the Cox meta-regression model, it is difficult to determine which one is responsible for the negative confounding[6,7]. In the fully adjusted model, smoking, alcohol, diabetes status, socioeconomic status, and UKB assessment centre were also included along with the

covariates included in the minimally adjusted model. Results from our fully adjusted model were not significant, meaning one or more of the factors controlled for in the fully adjusted model is a significant confounder of the PRS quantile and optimal BMI per quantile. Similarly to the minimally adjusted model, as all of these covariates are strongly associated with the BMI PRS and causality cannot be confirmed, it is difficult to determine which one is responsible for the confounding[6,7]. Future studies are warranted to disentangle the effects of these covariates using well-powered causal inference methods. Nevertheless, WHO BMI recommendations are not adjusted for specific covariates[2,43]. This creates complications when determining which model is clinically significant, regardless of statistical significance. Although WHO BMI recommendations are not adjusted for covariates, the studies that were used to derive these recommendations may have been heterogeneous in covariates adjustment[2,43]. The unadjusted model may reflect the closest to what might be seen in a regular patient population, as confounders, such as age, sex, or ancestry, are not accounted for when making clinical BMI recommendations[6,7]. However, results from the unadjusted model are also statistically unreliable as a result of not controlling for key confounders of BMI[6,7]. As i) most genetic studies control for sex, age, and genetic ancestry and ii) BMI has been historically shown to significantly vary by sex, age, and ancestry, the minimally adjusted model may be the most balanced model while retaining a reasonable level of external validity for clinical populations[1,2,6,7].

Our results are consistent with previous literature showing that TG, CRP, and ALT are associated with BMI[12,27,30]. While possible mechanisms for how these

biomarkers related to BMI have been described previously, it is generally accepted that these biomarkers are actually a marker of higher BMI rather than true influences on BMI[12,27,30]. From our results, the differences in mean BMI follow the same trend as the differences in optimal BMI. As BMI is highly correlated with TG, CRP and ALT, it is unclear whether the increase in optimal BMI in those with higher circulating levels of these biomarkers is due to an increase in mean BMI. Although previous genetic and epidemiological literature are inconsistent with regards to the directionality of BMI-TG and BMI-ALT relationships, the BMI-CRP relationship was found to follow a causal, positive association in one study[12,27,30,34]. As CRP is strongly associated with ALT and TG (Appendix Table 2.4), it may be that these biomarkers are mere markers of higher BMI. Thus, the difference in BMI nadir between groups above and below median values of these biomarkers may or not be reflective of the difference in mean BMI between individuals above and below these median values.

Although WHR and LM did not follow a J-shaped relationship with all-cause mortality in the whole UKB population, there was a J-shaped relationship between LM and all-cause mortality in both males and females, although causality cannot be concluded from these results. WHR and LM have been shown to be stronger than BMI in predicting disease and mortality risk, both phenotypically and through genetics[2,33]. As such, consideration of using genetically-predicted WHR and LM in conjunction with BMI recommendations may optimize overall adiposity/weight recommendations for patients, potentially addressing much of the limitations of BMI as an indicator for ideal weight.

2.4.4 Strengths and Limitations

Our study had notable limitations. First, while efforts were made to reduce reverse causality or residual confounding, we still cannot be sure whether our results were completely free from these biases[6]. Factors that BMI or genetic variants associated with BMI are sensitive to, such as lung function or eating behaviour, are significantly associated with mortality and may alter the BMI-mortality relationship and consequently, optimal BMI[6,19,39,40]. However, the current WHO BMI recommendations were also determined from observational studies[2,6]. We hypothesize that since the BMI-mortality relationship has been previously validated through Mendelian randomization analyses, our results may still be relatively robust to these confounders[1]. Nevertheless, the source population (e.g., different quantiles of PRS, different subgroups above and below median biomarker values, etc.) used to compute these associations could be the reason for the persistent variation in optimal BMI[1]. Second, while we used a genetically homogeneous British Caucasian population with minimal genetic ancestral confounding, our results cannot be easily generalized to other populations[7,9,41,42]. This is because the BMI and consequently, optimal BMI may vary greatly due to differences in genetic ancestry, which varies across different ethnic groups[7,9,41,42]. The predictive power of PRS also varies with genetic ancestry: if the GWAS used to derive the PRS was conducted in a given discovery cohort, the closer the population being analyzed is to this cohort in genetic ancestry, the better the PRS predictivity[9]. The predictive power of the PRS also decreases if frequencies for alleles that impact the trait varies between populations[42]. Additionally, linkage disequilibrium (LD) structures, or the non-random links between alleles of SNPs

at different loci on the same chromosome, may also differ between the different populations, decreasing PRS predictive power in populations that have a different LD structure from the one present in the population used to derive the PRS[4,5,42]. Hence, using PRS to determine the optimal BMI is subject to genetic ancestral influence and this may have impacted our results with certain PRS derived from discovery samples not matching closely to the UKB[9]. Finally, it is not known what type of adiposity the metabolically beneficial, deleterious, or neutral BMI encapsulates[7]. Future studies are needed to determine whether the genes associated with these respective scores are related to a specific type of fat, such as subcutaneous or visceral fat[7]. Identifying the type of fat presenting greater protection against all-cause mortality for those at the higher end of these PRS could help guide recommendations or treatment plans[9]. Additionally, both the whole genome and polygenically determined BMI PRS could also explain a significant amount of the variation in key non-fat mass components (e.g., lean mass, water mass). Knowing how much variation in non-fat mass components is attributed to these PRS might also be useful in making precise BMI recommendations. Thus, future studies should seek to determine which genes linked with these PRS associate with a given non-fat mass component.

Strengths of our study include the large sample size of the UKB, the opportunity for extensive phenotyping (i.e., the detailed and comprehensive analysis of phenotypes such as BMI), innovative use of genetic tools in combination with a meta-regression framework, and the power of the polygenic risk scores used.

2.5 Conclusion and Significance

The impact of BMI on mortality varies due to individual genetic difference[1,7,8]. Our findings provide support for a more individualized approach to promoting health, especially for those that may have a higher risk for death at a BMI falling under the WHO-defined normal category. Our study demonstrates that the current WHO recommendations may not be universal in achieving optimal BMI, which has potential important implications for patient counselling and selection of patients for obesity-related interventions such as bariatric surgery[10]. The use of PRS in general for clinical practice, despite its strong potential clinical utility for screening diseases such as CVD or breast cancer, is currently unestablished[9]. We hope our research provides further evidence for support of clinical use of PRS to optimize patient health, well-being, and quality of life. Such advances in precision medicine could have large implications for public health recommendations for BMI: just as the definition of BMI has constantly changed over time due to evidence-based medicine, we hope our findings can provide the catalyst to bring change once again.

2.6 References

1. Sun YQ, Burgess S, Staley JR, *et al*. Body mass index and all cause mortality in HUNT and UK Biobank studies: linear and non- linear mendelian randomisation analyses. *BMJ* 2019*;* **364**: 1-10. DOI: 10.1136/bmj.l1042.

2. Nuttall FQ. Body Mass Index: Obesity, BMI, and Health: A Critical Review. *Nutr Today* 2015; **50**: 117-128. DOI: 10.1097/NT.0000000000000092.

3. Aune D, Sen A, Prasad M, *et al.* BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with 3.74 million deaths among 30.3 million participants. *BMJ* 2016; **353**:1-17. doi:10.1136/bmj.i2156

4. Paré G, Mao S, Deng WQ. (2019). A robust method to estimate regional polygenic correlation under misspecified linkage disequilibrium structure. *Genet Epidemiol* 2019; **42**: 636-647. DOI: 10.1002/gepi.22149.

5. Marees AT, de Kluiver H, Stringer S, *et al* (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res* 2018; **27**: 1-10. DOI: 10.1002/mpr.1608.

6. Bhaskaran K, Dos-Santos-Silva I, Leon DA, *et al*. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3.6 million adults in the UK. *Lancet Diabetes Endocrinol* 2018; **6**: 944-953. DOI: 10.1016/S2213-8587(18)30288-2.

7. Pigeyre M, Sjaarda J, Mao S, *et al*. Identification of novel causal blood biomarkers linking metabolically favorable adiposity with type 2 diabetes risk. *Diabetes Care* 2019; **42**: 1800-1808. DOI: 10.2337/dc18-2444.

8. Emdin CA, Khera AV, Natarajan P, *et al*. Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA,* 2017; **317**: 626-634. DOI: 10.1001/jama.2016.21042.

9. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med* 2020; **12:** 1-11. https://doi.org/10.1186/s13073-020-00742-5.

10. Schorr M, Dichtel, LE, Gerweck, AV, *et al.* Sex differences in body composition and association with cardiometabolic risk. *Biol Sex Differ* 2018; **9,** 1-10. DOI: 10.1186/s13293-018-0189-3.

11. Echouffo-Tcheugui JB, *et al*. Natural history of obesity subphenotypes: dynamic changes over two decades and prognosis in the Framingham Heart Study. *J Clin Endocrinol Metab* 2019;**104**:738–752. DOI: 10.1210/jc.2018-01321.

12. Timpson NJ, Nordestgaard BG, Harbord RM, et al. C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal

Mendelian randomization. *Int J Obes (Lond)* 2011; **35**, 300-308. DOI:10.1038/ijo.2010.137

13. Tam V, Patel N, Turcotte M. *et al.* Benefits and limitations of genome-wide association studies. *Nat Rev Genet* 2019; **20**, 467–484. DOI: 10.1038/s41576-019-0127-1.

14. Mbatchou J, Barnard L, Backman J, *et al*. Computationally efficient whole genome regression for quantitative and binary traits. *bioRxiv* 2020; Published online January 1. DOI:10.1101/2020.06.19.162354.

15. Newcombe PJ, Nelson CP, Samani NJ, *et al*. A flexible and parallelizable approach to genome-wide polygenic risk scores. *Genetic Epidemiology* 2019;**43**:730-741. DOI: 10.1002/gepi.22245.

16. Scrucca L, Santucci A, Aversa F. Competing risk analysis using R: an easy guide for clinicians. *Bone Marrow Transplant*, 2007; **40**, 381–387. DOI: 10.1038/sj.bmt.1705727.

17. Klempfner R, Erez A, Sagit BZ, *et al*. Elevated Triglyceride Level Is Independently Associated With Increased All-Cause Mortality in Patients With Established Coronary Heart Disease: Twenty-Two-Year Follow-Up of

the Bezafibrate Infarction Prevention Study and Registry. *Circ Cardiovasc Qual Outcomes,* 2016;**9**:100-108. DOI: 10.1161/CIRCOUTCOMES.115.002104.

18. Bhaskaran K, Douglas I, Forbes H, *et al*. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet* 2014;**384**:755-65. DOI: 10.1016/S0140-6736(14)60892-8.

19. Mafort, TT, Rufino R, Costa CH, *et al.* Obesity: systemic and pulmonary complications, biochemical abnormalities, and impairment of lung function. *Multidiscip Respir Med* 2016; **11;** 1-11. DOI: 10.1186/s40248-016-0066-z.

20. Kautzky-Willer A, Harreiter J, Pacini G. Sex and Gender Differences in Risk, Pathophysiology and Complications of Type 2 Diabetes Mellitus. *Endocr Rev* 2016;**37**:278-316. DOI:10.1210/er.2015-1137

21. Censin JC, Peters SAE, Bovijn J, *et al*. Causal relationships between obesity and the leading causes of death in women and men. *PLoS Genet* 2019; **15**:e1008405. DOI: 10.1371/journal.pgen.1008405.

22. Khera AV, Chaffin M, Wade KH, *et al*. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell* 2019;**177**:587-596. DOI:10.1016/j.cell.2019.03.028.

23. Richardson TG, Harrison S, Hemani G, *et al*. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *Elife* 2019;**8**:e43657. DOI: 10.7554/eLife.43657.

24. Howe LJ, Dudbridge F, Schmidt AF, *et al*. Polygenic risk scores for coronary artery disease and subsequent event risk amongst established cases. *Human Molecular Genetics* 2020; **29**, 1388–1395. DOI: 10.1093/hmg/ddaa052.

25. Manning AK, Hivert MF, Scott RA, *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genetics* 2012;**44**:659-669. DOI: 10.1038/ng.2274.

26. Udler MS, McCarthy MI, Florez JC, *et al*. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocr Rev* 2019;**40**:1500-1520. DOI: 10.1210/er.2019-00088

27. Shamai L, Lurix E, Shen M, *et al*. Association of body mass index and lipid profiles: evaluation of a broad spectrum of body mass index patients including the morbidly obese. *Obes Surg* 2011;**21**:42-7. doi: 10.1007/s11695-010-0170-7.

28. Morin S, Leslie WD; Manitoba Bone Density Program. High bone mineral density is associated with high body mass index. *Osteoporos Int* 2009;**20**:1267-1271. DOI: 10.1007/s00198-008-0797-6.

29. Clifton EA, Day FR, De Lucia Rolfe E, et al. Associations between body mass index-related genetic variants and adult body composition: The Fenland cohort study. *Int J Obes (Lond)* 2017;**41**:613-619. DOI:10.1038/ijo.2017.11.

30. Liu, J., Au Yeung, S.L., Kwok, M.K. *et al.* The effect of liver enzymes on adiposity: a Mendelian randomization study. *Sci Rep 2019;* **9**, 1-10. https://doi.org/10.1038/s41598-019-52489-8

31. Banack, H.R., Kaufman, J.S. From bad to worse: collider stratification amplifies confounding bias in the “obesity paradox”. *Eur J Epidemiol* 2015; **30**, 1111–1114. https://doi.org/10.1007/s10654-015-0069-7

32. Echouffo-Tcheugui JB, Ali MK, Griffin SJ, *et al.* Screening for type 2 diabetes and dysglycemia. *Epidemiol Rev.* 2011;**33**:63-87. doi: 10.1093/epirev/mxq020.

33. Emdin CA, Khera AV, Natarajan P, *et al*. Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA*. 2017;**317**:626-634. doi:10.1001/jama.2016.21042

34. Holmes MV, Lange LA, Palmer T, *et al*. Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis. *Am J Hum Genet.* 2014;**94**:198-208. doi:10.1016/j.ajhg.2013.12.014

35. Ryan DH, Yockey SR. Weight Loss and Improvement in Comorbidity: Differences at 5%, 10%, 15%, and Over. Curr Obes Rep. 2017;6(2):187-194. doi:10.1007/s13679-017-0262-y

36. Choi, SW, Mak, TSH. & O'Reilly, PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* 2020; **15,** 2759–2772. 10.1038/s41596-020-0353-1

37. Hong EP, Park JW. Sample size and statistical power calculation in genetic association studies. *Genomics Inform*. 2012;**10**:117-122. doi:10.5808/GI.2012.10.2.117

38. Choi AL, Cordier S, Weihe P, *et al*. Negative confounding in the evaluation of toxicity: the case of methylmercury in fish and seafood. *Crit Rev Toxicol.* 2008;**38**:877-893. doi:10.1080/10408440802273164

39. Wardle J. Eating behaviour and obesity. *Obes Rev*. 2007**;8**:73-75. doi:10.1111/j.1467-789X.2007.00322.x

40. Qi L, Cho YA. Gene-environment interaction and obesity. *Nutr Rev.* 2008;**66**:684-694. doi:10.1111/j.1753-4887.2008.00128.x

41. Hoffmann TJ, Choquet H, Yin J, *et al*. A Large Multiethnic Genome-Wide Association Study of Adult Body Mass Index Identifies Novel Loci. *Genetics.* 2018;210(2):499-515. doi:10.1534/genetics.118.301479

42. Giri A, Edwards TL, Hartmann KE, *et al*. African genetic ancestry interacts with body mass index to modify risk for uterine fibroids. *PLoS Genet*. 2017;**13**:1-22. doi:10.1371/journal.pgen.1006871

43. Flegal KM, Carroll MD, *et al*. Overweight and obesity in the United States: prevalence and trends, 1960-1994. *Int J Obes Relat Metab Disord*. 1998;**22**:39-47. doi: 10.1038/sj.ijo.0800541.

44. Karasik D, Zillikens MC, Hsu YH, et al. Disentangling the genetics of lean mass. *Am J Clin Nutr.* 2019;**109**:276-287. doi:10.1093/ajcn/nqy272

45. Pei YF, Hu WZ, Yang XL, *et al*. Two functional variants at 6p21.1 were associated with lean mass. *Skelet Muscle.* 2019;**9**:28. doi:10.1186/s13395-019-0212-3

2.7 Figures

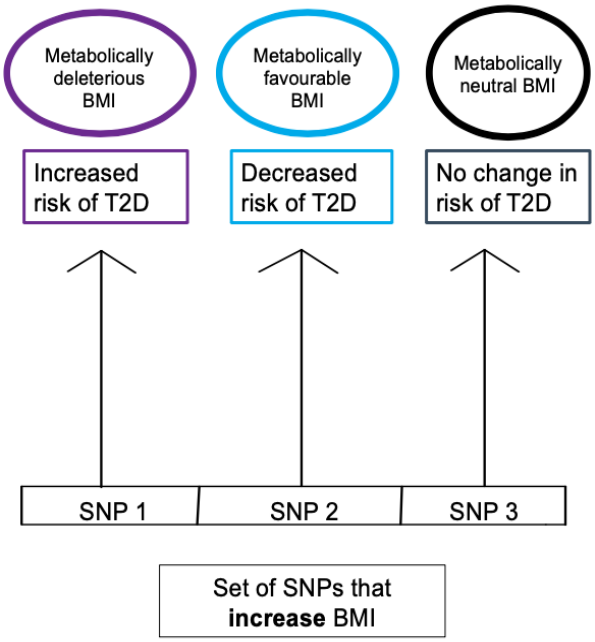


Figure 2.1 - Regional polygenic correlation between BMI and T2D.

3 different SNPs, each having an effect allele that increases BMI, are shown to be correlated with T2D in three ways: increased risk (metabolically deleterious BMI), decreased risk (metabolically favourable BMI), or no risk (metabolically neutral BMI) of T2D. BMI = body mass index, T2D = type 2 diabetes.

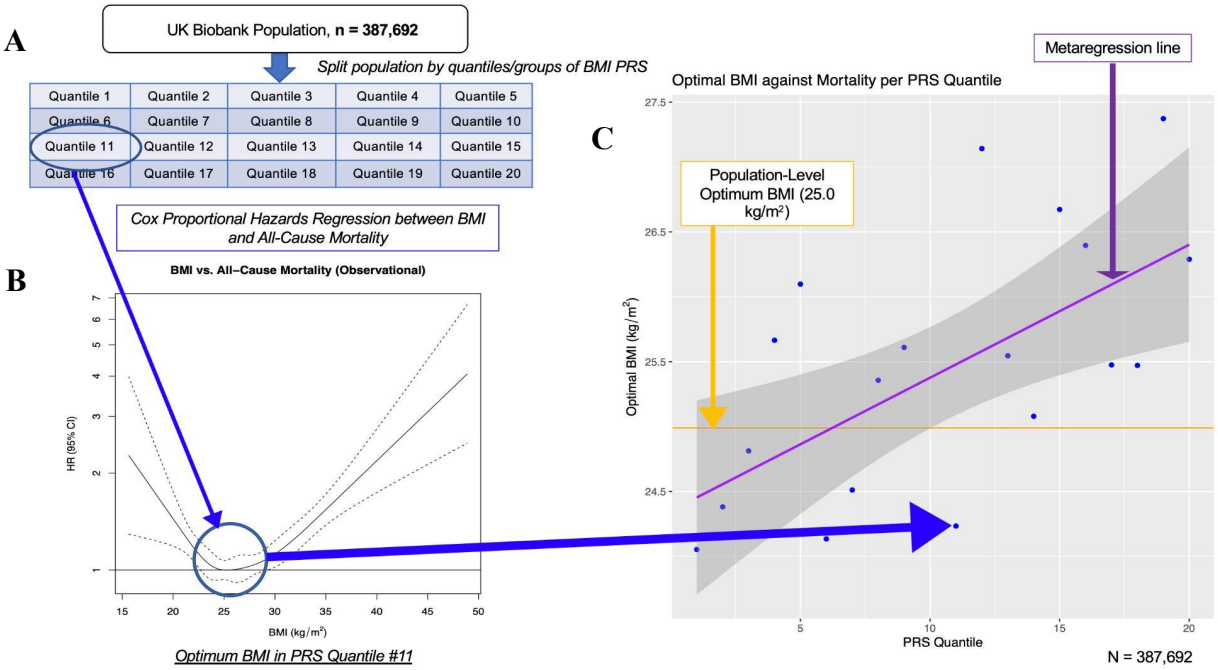


Figure 2.2 - Comparing the optimal BMI across quantiles of BMI PRS.

A) Partition of UKB participants into quantiles of BMI PRS, B) Cox proportional hazards regression model used to compute the BMI-all-cause mortality relationship and extract the optimal BMI for a given quantile, C) Comparing the optimal BMI between PRS quantiles using linear regression ("Cox meta-regression"). Results are also compared to the population optimal BMI computed on the entire UKB (orange line).

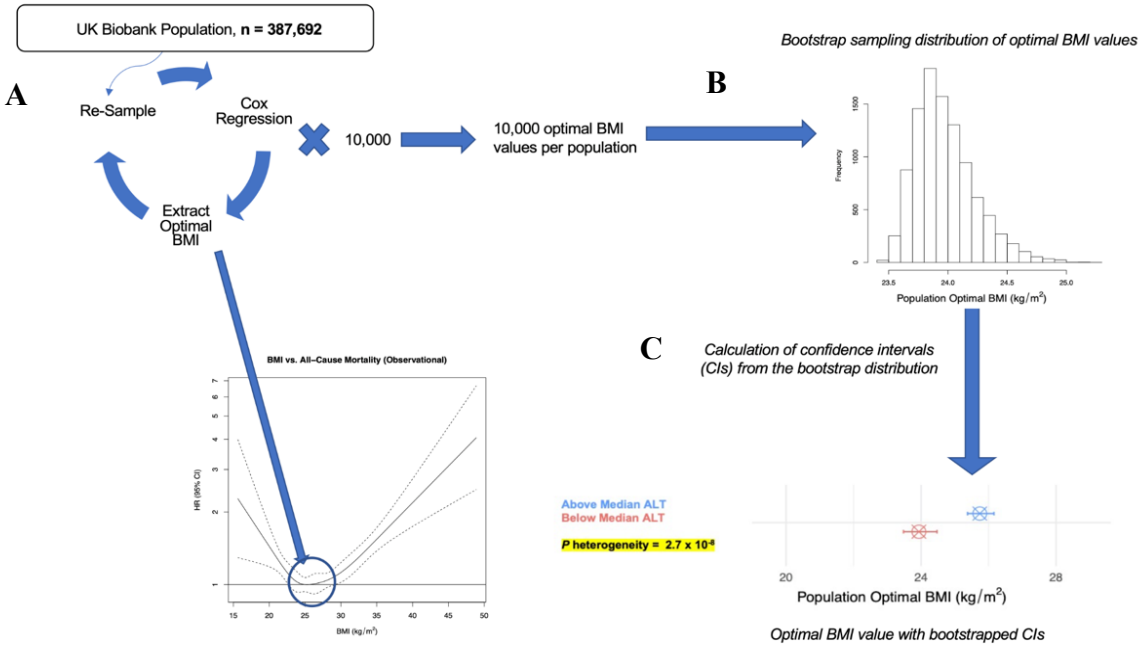
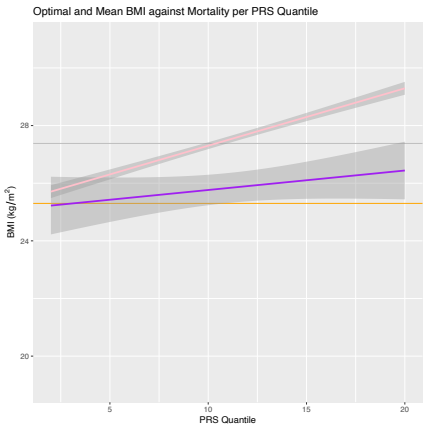
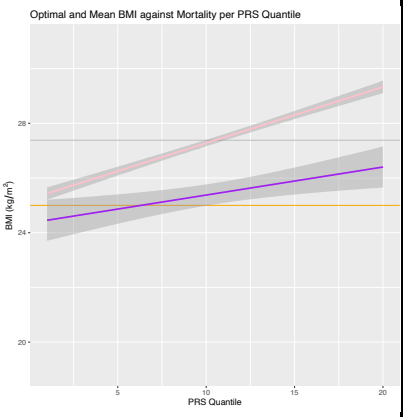
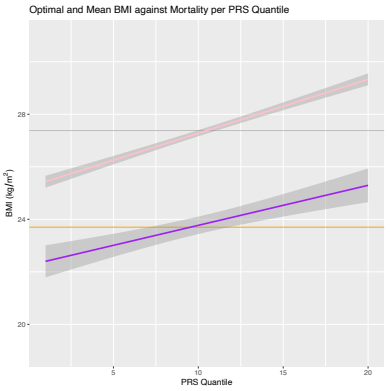


Figure 2.3 - Bootstrapping procedure to calculate confidence intervals for the population optimal BMI.

A) Re-sampling of the UKB population with replacement, followed by computing the optimal BMI using the Cox meta-regression model per re-sample. Our sample was resampled 10,000 times. B) Bootstrap distribution of the 10,000 optimal BMI values generated in A). The frequency for which a nadir value falls under a given population optimal BMI value is shown on the y-axis. C) Calculation of 95% confidence intervals from the bootstrap distribution, shown for the population optimal BMI for individuals above and below median ALT in the UKB. The *p* value for the heterogeneity test used to

compare significant differences in optimal BMI across subgroups is highlighted in yellow. The heterogeneity test uses a fixed-effects meta-analysis model. BMI = body mass index, UKB = UK Biobank. ALT = alanine aminotransferase.

Primary Analyses		
Cox Regression Analysis: All-Cause Mortality and BMI, fully adjusted model	Cox Regression Analysis: All-Cause Mortality and BMI, minimally adjusted model	Cox Regression Analysis: All-Cause Mortality and BMI, unadjusted model
		
PRS: Whole Genome BMI Beta [95% CI]: 0.07 [-0.03-0.16] p = 0.15 Population Optimum BMI [95% CI]: 25.3 kg/m^2 [24.6 – 26.1 kg/m^2] Population Mean BMI [95% CI]: 27.4 kg/m^2 [27.3 – 27.5 kg/m^2]	PRS: Whole Genome BMI Beta [95% CI]: 0.10 [0.04-0.17] p = 0.005 Population Optimum BMI [95% CI]: 25.0 kg/m^2 [95% CI 24.6-25.5 kg/m^2]	PRS: Whole Genome BMI Beta [95% CI]: 0.06 [-0.17– 0.30] p = 0.54 Population Optimum BMI [95% CI]: 23.7 kg/m^2 [23.5 – 23.9 kg/m^2]

	Population Mean BMI [95% CI]: 27.4 kg/m^2 [27.3 – 27.5 kg/m^2]	Population Mean BMI [95% CI]: 27.4 kg/m^2 [27.3 – 27.5 kg/m^2]

Figure 2.4 - Mediation of BMI-mortality relationships by genetics.

Results are shown for both primary and secondary analyses. β values (95% CI) indicate the effect of a 1-unit increase in PRS quantile on BMI nadir value (kg/m^2). Bonferroni significance is considered at $p < 0.007$. BMI = body mass index, PRS = polygenic risk score. The purple line represents the meta-regression line, the orange line represents the population optimum BMI, and the grey line represents the population mean BMI. The pink line represents the regression line of the association between PRS quantile number and the mean BMI per quantile ("average BMI regression line"). Shaded areas represent 95% CI for the meta-regression and average BMI regression lines. For the fully adjusted model, the first quantile was considered an outlier and was subsequently removed.

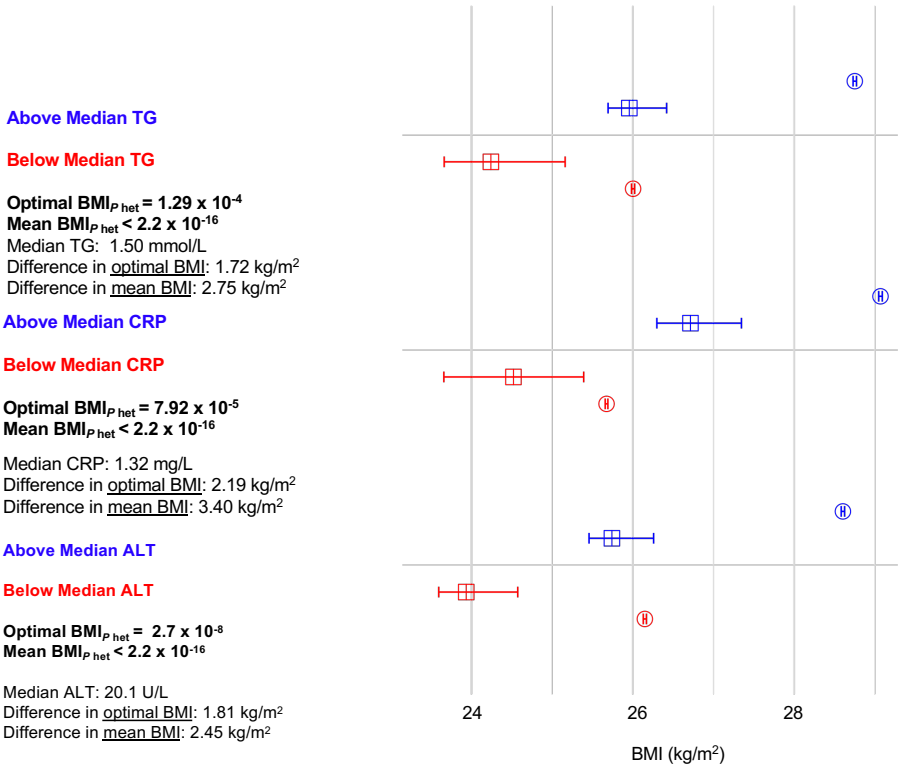


Figure 2.5 - Mediation of optimal and mean BMI by TG, CRP, and ALT levels.

The population optimal BMI is represented by squares and the mean BMI is represented by circles. BMI values are coloured according to the colour of the subgroup name. Results from the minimally adjusted model are shown. Bonferroni significance is considered at $p < 0.0008$. TG = triglyceride, CRP = C-reactive protein, ALT = alanine aminotransferase, BMI = body mass index, Optimal BMI_{Phet} = fixed effects heterogeneity test results comparing the population optimal BMI (kg/m^2) between subgroups, Mean BMI_{Phet} = fixed effects heterogeneity test results comparing the mean BMI (kg/m^2) between subgroups. Error bars represent 95% CI obtained from bootstrapping analyses.

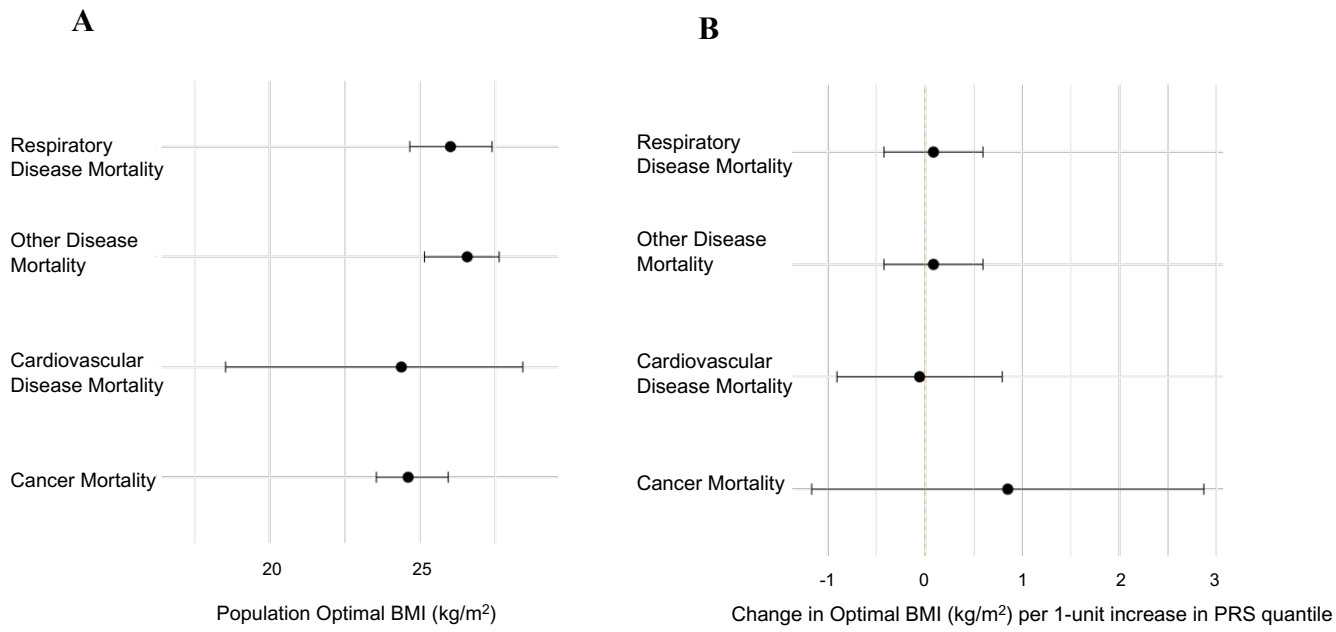


Figure 2.6 - A) The optimal BMI and B) the change in optimal BMI by cause-specific mortality.

Results from the minimally adjusted model are shown. Bonferroni significance is considered at $p < 0.0125$. CVD = cardiovascular disease, other disease= mortality due to diseases/conditions other than cancer, cardiovascular disease, or respiratory disease. BMI = body mass index, PRS = polygenic risk score. Error bars represent 95% CI obtained from bootstrapping analyses in 2.6 A) and the meta-regression line in 2.6 B). For 2.6 B), β values (95% CI) indicate the effect of a 1-unit increase in PRS quantile on BMI nadir value (kg/m^2).

Percentiles of Clinical Biomarker Analyses

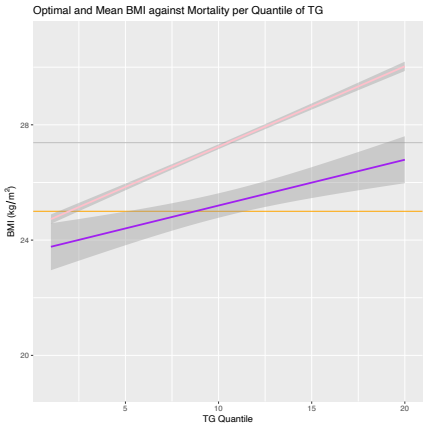
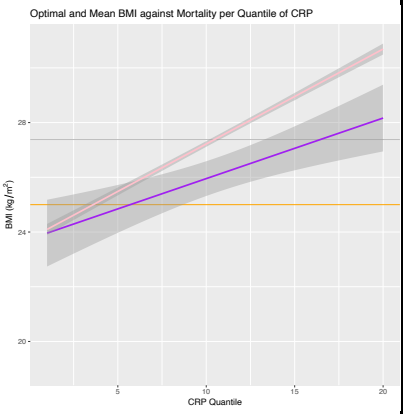
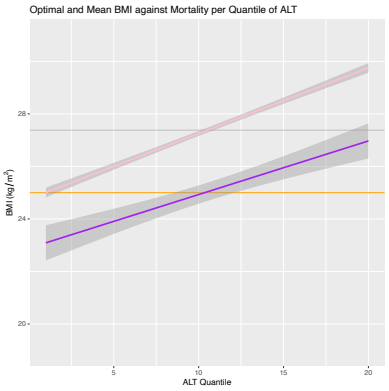
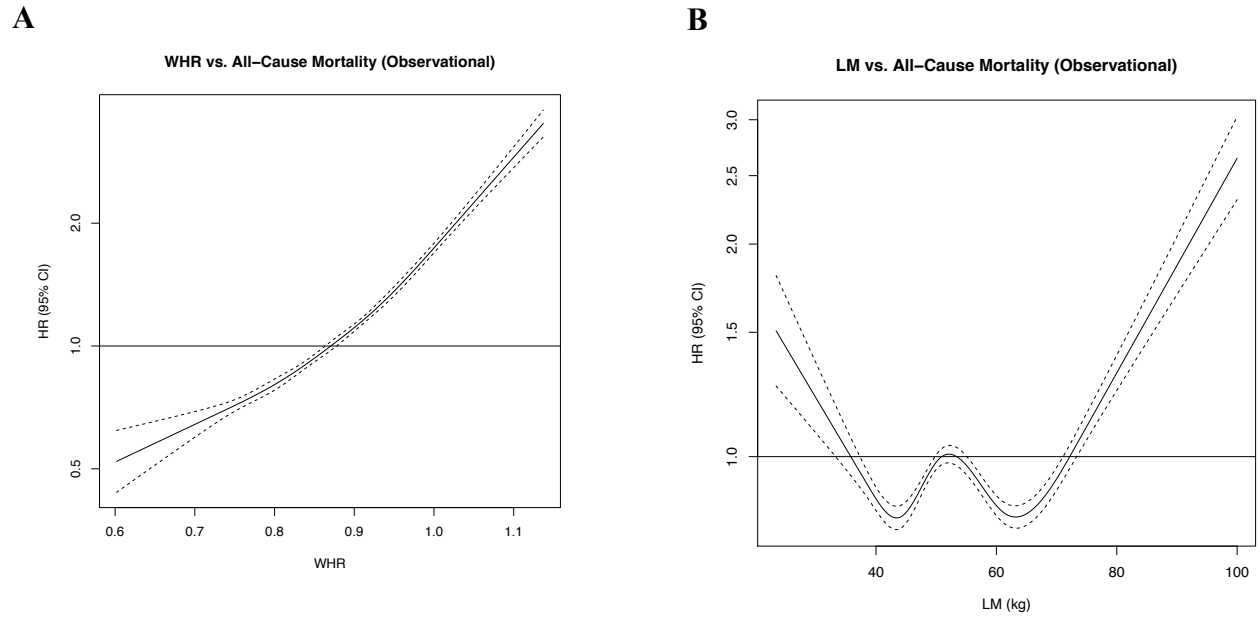
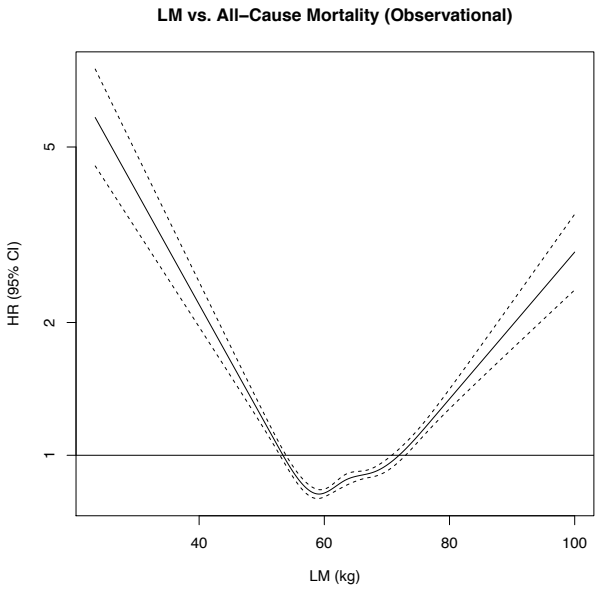
Biomarker: TG	Biomarker: CRP	Biomarker: ALT
		
Beta [95% CI]: 0.16 [0.09-0.23]	Beta [95% CI]: 0.20 [0.14-0.26]	Beta [95% CI]: 0.22 [0.11 – 0.33]
$p = 2.00 \times 10^{-4}$	$p = 1.26 \times 10^{-6}$	$p = 5.00 \times 10^{-4}$
Population Optimum BMI [95% CI]: 25.0 kg/m^2 [95% CI 24.6-25.5 kg/m^2]	Population Optimum BMI [95% CI]: 25.0 kg/m^2 [95% CI 24.6-25.5 kg/m^2]	Population Optimum BMI [95% CI]: 25.0 kg/m^2 [95% CI 24.6-25.5 kg/m^2]
Population Mean BMI [95% CI]: 27.4 kg/m^2 [27.3 – 27.5 kg/m^2]	Population Mean BMI [95% CI]: 27.4 kg/m^2 [27.3 – 27.5 kg/m^2]	Population Mean BMI [95% CI]: 27.4 kg/m^2 [27.3 – 27.5 kg/m^2]

Figure 2.7 – Variation in optimal BMI across 20 quantiles of TG, ALT, and CRP. β values (95% CI) indicate the effect of a 1-unit increase in a clinical biomarker quantile on BMI nadir value (kg/m^2). The minimally adjusted model was used. Bonferroni

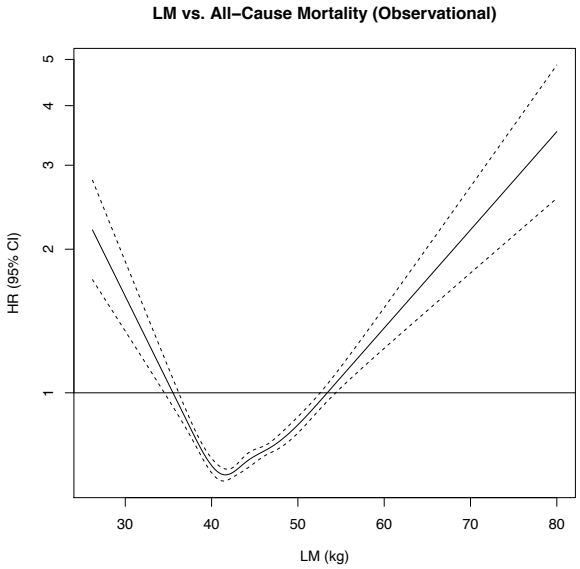
significance is considered at $p < 0.0008$. BMI = body mass index, TG = triglyceride, CRP = C-reactive protein, ALT = alanine aminotransferase, PRS = polygenic risk score. The purple line represents the meta-regression line, the orange line represents the population optimum BMI, and the grey line represents the population mean BMI. The pink line represents the regression line of the association between clinical biomarker quantile number and the mean BMI per quantile ("average BMI regression line"). Shaded areas represent 95% CI for the meta-regression and average BMI regression lines.



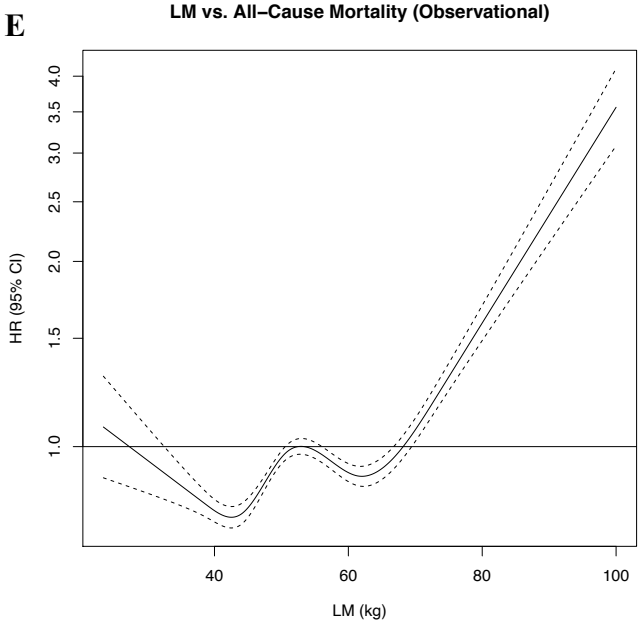
C


LM vs. All−Cause Mortality (Observational)
HR (95% CI)
1
2
5
40
60
80
100
LM (kg)

D


LM vs. All−Cause Mortality (Observational)
HR (95% CI)
1
2
3
4
5
30
40
50
60
70
80
LM (kg)

E


LM vs. All−Cause Mortality (Observational)
HR (95% CI)
1.0
1.5
2.0
2.5
3.0
3.5
4.0
40
60
80
100
LM (kg)

F)

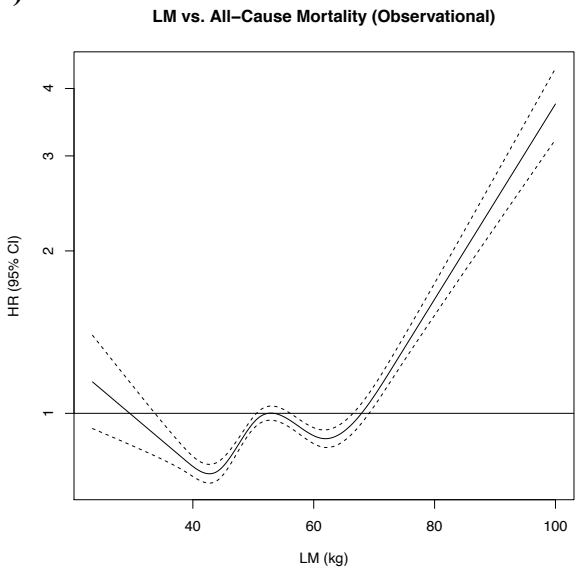
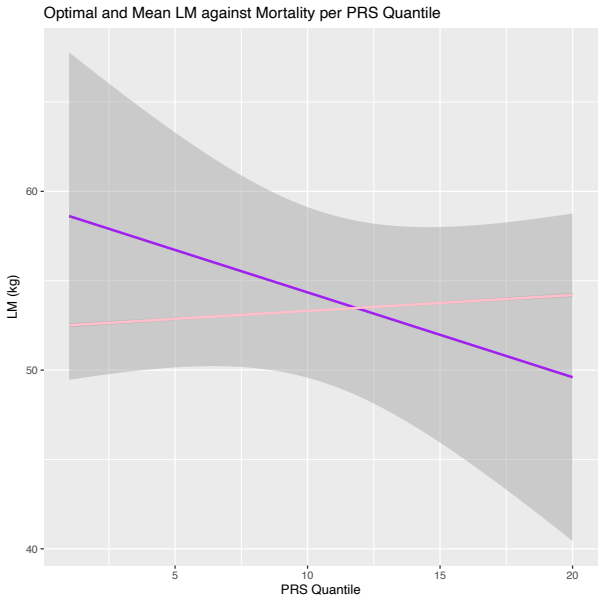

LM vs. All−Cause Mortality (Observational)
HR (95% CI)
1
2
3
4
40
60
80
100
LM (kg)

Figure 2.8 –The relationship between a) WHR and b) LM, respectively, and all-cause mortality. The relationship between LM and all-cause mortality in c) males and d) females is also shown. WHR = Waist-to-hip ratio, LM = lean mass, HR = hazard ratio for all-cause mortality. The minimally adjusted model was used. Figure 2.8e) and f) shows the W-shaped relationship for LM adjusted for height and height^2 respectively. For Figure 2.8 a), WHR values that were considered outliers were removed, reducing the number of deaths and controls in the overall sample to 22,957 and 364,097 respectively.



Beta [95% CI]: -0.47 [-1.30-0.35]

p = 0.24

Figure 2.9 – Variation in optimal LM between quantiles of LM PRS. β values (95% CI) indicate the effect of a 1-unit increase in PRS quantile on LM nadir value (kg). The minimally adjusted model was used. Significance is considered at $p < 0.05$. LM = whole body lean mass, PRS = polygenic risk score. The purple line represents the meta-

regression line. The pink line represents the regression line of the association between PRS quantile number and the mean LM per quantile ("average LM regression line"). Shaded areas represent 95% CI for the meta-regression and average LM regression lines.

Chapter III: Optimizing BMI Recommendations Using Genetics and Clinical Biomarkers - A Mendelian Randomization Approach

3.1 Introduction

Well-powered epidemiological studies are able to confer strong insights into key relationships in nature. Several large epidemiological studies found that BMI and all-cause mortality had a J-shaped relationship, with the optimal BMI, or the lowest risk of mortality, being between 20-25 kg/m^2 [3,6]. The J-shaped relationship holds true with various specific causes of deaths, as well as after stratifying the population by age, sex, and smoking status[3,6]. Nevertheless, epidemiological studies are prone to biases like residual confounding, which includes confounders such as smoking in the case of BMI and all-cause mortality relationships, or reverse causality, creating difficulty in inferring causality[3,6].

Genetic variants comprising our individual genotypes are randomly inherited from our parents[32]. Alleles for a given genotype, in the form of a single nucleotide polymorphism (SNP), are randomized at the time of gamete formation[32]. Mendel's second law of independent assortment explains this process as genetic variants are inherited independent of other characteristics or potential confounding influences[32]. Mendelian randomization (MR) is based on this law: if SNPs either change the level or mirror the biological effects of the phenotype that is linked with the outcome, then the variants should be causally associated with the outcome, at least to the extent predicted by their impact on the phenotype[32]. In this sense, MR is "nature's" equivalent to a

randomized control trial (RCT) (Appendix Figure 3.1) [32,38]. Unlike RCT, where individuals are randomized and split into an exposed group receiving the intervention and control group receiving no intervention or placebo, genotypes are randomized in MR, with the exposure group having the risk allele for the given phenotype (i.e. increased BMI) and the control group having the alternate or null allele (i.e. no effect on BMI)[32,38]. For both MR and RCT, this randomization allows for confounders to be equivalent between the two groups[32]. When the outcome is examined between the two groups using either method, a potentially causal or unconfounded conclusion can be reached, provided that certain assumptions are met for either method[32,39].

In order for MR estimates to be considered valid, three primary assumptions must be met[32]. The genetic variants must be associated with the exposure, there can be no unmeasured confounders of the association between the variants and the outcome, and the variants must affect the outcome only through the outcome of interest[32]. A secondary assumption that must be met is that random mating occurs in the population being analyzed, though this is not often discussed in the literature as a primary assumption for MR[12,21]. While the first assumption can be confirmed by assessing the strength of the association between the genetic variants and the exposure, the latter two assumptions are difficult to prove due to horizontal pleiotropy, or when a single genetic variant influences the exposure and outcome through independent pathways[33]. While there are various methods to detect and account for pleiotropy, it cannot be completely avoided and is considered a key limitation in MR studies[33]. Nevertheless, MR studies allow researchers

to assess the relationship between multiple variables with as little confounding as possible, thus being able to make causal inferences with a high degree of confidence, despite potential pleiotropy[33].

Traditional MR analyses have shown a positive relationship between BMI and all-cause mortality that varies significantly across different WHO-defined BMI categories. Overweight and obese individuals have a 5% and 9% higher risk of mortality respectively while underweight and low normal weight participants have a 34% and 14% lower risk respectively[1]. While this analysis suggests that higher BMI is associated with higher mortality risk, traditional MR methods do not take into account the J-shaped nature of the BMI-all-cause mortality relationship found in previous epidemiological studies[1]. Traditional MR methods are thus insufficient to draw conclusions about non-linear relationships.

In 2017, Staley and Burgess developed a modelling method for MR that takes into account non-linear relationships between a given exposure and outcome[1]. In one study, this model successfully recreated the J-shaped relationship between BMI and all-cause mortality previously seen in epidemiological research, even after exclusion of deaths due to cancer and cardiovascular disease[1]. Thus, the J-shaped relationship between BMI and all-cause mortality was determined to be causal[1]. However, a J-shaped relationship was only found in the smokers UKB subgroup, not the never smokers subgroup, where a positive relationship was found instead[1]. Smoking is still a significant confounder in this J-shaped relationship, despite using NLMR[1]. Overall, this study simultaneously shows the

strengths and limitations of using NLMR for BMI-all-cause mortality relationships[1]. While it was able to replicate prior epidemiological analyses, it did not completely account for all possible confounding or potential pleiotropic influences, like smoking[1]. Nevertheless, even with the confounding influence of smoking, the J-shaped curve is likely the true relationship between BMI and all-cause mortality as large, well-powered genetic and epidemiological studies produced consistent results, with the optimum BMI within a range of 22-25 kg/m^2 [1]. However, Sun et al did not compare the optimal BMI computed from NLMR with the optimal BMI computed using traditional epidemiological methods (i.e. Cox regression)[1,6]. Individual variation in the optimum BMI computed using NLMR due to factors strongly linked with BMI, such as sleep, physical activity, or diet, was also not explored[6]. Metabolically healthy and unhealthy BMI may also influence the BMI-mortality relationship computed by NLMR, which could make a difference in terms of recommendations for potential weight loss interventions targeting the optimal BMI[11].

The first objective of this chapter is to determine the optimal BMI using the NLMR model. The second objective is to compare the optimal BMI computed from NLMR with the optimal BMI computed using the Cox meta-regression model. The third objective is to determine whether the NLMR-derived optimal BMI varies due to individual genetic and non-genetic variation.

3.2 Methods

3.2.1 UK Biobank Study Population

Details describing the UK Biobank study population can be found in Section 2.2.1.

3.2.2 Polygenic Risk Score Calculations

Details describing polygenic risk score calculations can be found in Section 2.2.2.

3.2.3 Non-Linear Mendelian Randomization

The results from traditional Mendelian randomization (MR) analyses are calculated via the Wald ratio: $\beta = \frac{Outcome \sim PRS}{Exposure \sim PRS}$[19]. The numerator is the estimated association of our genetic variants, in the form of polygenic risk scores (PRS), with the outcome while the denominator is the estimated association of the PRS with the exposure[19]. The Wald ratio is the quantified causal estimate between an exposure and an outcome. However, we used non-linear Mendelian randomization (NLMR) to investigate the causal effects of our risk factors (e.g. BMI) on our outcomes (e.g. all-cause mortality or ACM).

Unlike traditional MR, NLMR uses non-linear modelling to capture causal relationships that are non-linear in nature[1, 11]. Suppose BMI is the exposure and we are using the BMI PRS as our genetic variants for NLMR. First, the UKB population is stratified by quantiles of BMI[11]. However, the population cannot be stratified into quantiles of the regular BMI phenotype, as such an approach may induce an association between the BMI PRS and outcome even if it was not present prior to stratification[11].

Instead, the population was stratified into centiles of the BMI residual variation, or residuals of BMI after regressing on the BMI PRS (genetic-free BMI or the non-genetic component of BMI).[1,11]. Second, within each genetic-free BMI centile, the causal association between BMI and the outcome (e.g. ACM), or local average causal estimate (LACE), was estimated via the following formula: $LACE = \frac{Outcome \sim BMI\ PRS}{Exposure \sim BMI\ PRS}$[11]. The numerator is the estimated association of the BMI PRS with the outcome (e.g. ACM) while the denominator is the estimated association of the BMI PRS with the exposure (e.g. BMI) in the UKB[11,20]. Third, the causal relationship between the exposure and the outcome was determined from the LACEs computed per centile[11]. Two distinct methods can be used to accomplish this: the fractional polynomial and the piecewise linear model[11]. The fractional polynomial method uses a smoothing spline to fit non-linear relationships between the exposure and the outcome[11]. The method first uses a meta-regression analysis where the LACE estimates, itself a product of regression, is being regressed against the mean BMI value within each genetic-free BMI stratum[11]. The line or curve of best-fit for this meta-regression is the best-fitting fractional polynomial function for the relationship[11]. The meta-regression between LACEs and mean BMI per quantile represents the derivative of the true exposure-outcome relationship, or the instantaneous slope of the tangent line to the true non-linear model fitting the BMI- all-cause mortality relationship[11,45]. Thus, the integral or anti-derivative of this best-fitting fractional polynomial function, or the true non-linear relationship between BMI and all-cause mortality derived from its derivative function, was then taken, using a smoothing

spline, producing the true exposure-outcome relationship (e.g. Figure 3.1)[11,46]. The piecewise linear method is generated by having each genetic-free BMI stratum contribute to a line segment whose tangent or derivative is the LACE estimate for that stratum[11]. The function is constrained to be continuous, with each line segment starting from where the previous ended[11]. One method is preferred over the other depending on prior belief or evidence of the nature of the true exposure-outcome relationship[11]. If the relationship is believed to be smooth, then the fractional polynomial should be used; otherwise, the piecewise linear should be used[11]. Since a large number of epidemiological studies show a smooth non-linear relationship between BMI and many mortality outcomes, only the fractional polynomial NLMR was used for these analyses[1,10,21]. The exposure-outcome relationship is considered significantly non-linear if the fractional polynomial non-linearity *p*-value is <0.007 (the Bonferroni adjustment for *p* values for primary analyses in Chapter II is used here and is described in further detail in 2.2.3)[1]. If significant, we conclude that a non-linear model fits the data better than a linear model[1].

The optimal BMI, or BMI associated with the lowest risk of a given mortality outcome, was extracted from the true exposure-outcome relationship computed by NLMR. Confidence intervals for the optimal BMI were computed using the bootstrapping procedure described in Figure 2.3.

The covariate models used for NLMR analyses were the same as the ones described in Chapter II, except the fully adjusted model, which included smoking and UKB assessment centre as additional covariates to the minimally adjusted model only.

The only exposures that were used for NLMR analyses were BMI and its metabolically favourable, deleterious, and neutral BMI subcategories. Thus, the PRS associated with these phenotypes were the only ones utilized in NLMR analyses. The outcomes and subgroups used in our primary, secondary, and tertiary analyses, as well as our other methods for statistical analyses, were the same as the ones used in Chapter II.

For our primary analyses, centiles were used to partition the UKB into genetic-free BMI quantiles while 50 quantiles were used for our secondary analyses. For our tertiary analyses, centiles were used when partitioning the UKB into genetic-free exposure quantiles while 50 quantiles were used for our sex-specific analyses.

A modification to the standard NLMR method was used as an alternative method for modeling the non-linear relationship between BMI and mortality outcomes across our primary and secondary analyses. Instead of a fractional polynomial function being fit to the mean BMI per quantile and LACE relationship, a quadratic model was fitted as it may provide more stability and power, while retaining a simple framework. The quadratic model was identical to the original NLMR model, but instead of fitting the best-fitting fractional polynomial function to the LACE-mean BMI per quantile relationship, a linear regression model was fit, since this would be the derivative of the quadratic model (Figure 3.6). The BMI at which the linear regression line crosses the 0 LACE point would be the optimal BMI, as this represents the inflection point between the lower and upper extreme of the J-shaped curve (Figure 3.6). Once we have extracted the optimal BMI, we computed the 95% calibration interval, which is the inverse of the confidence interval, as

we are trying to find the variability in the independent variable, rather than the dependent variable.[40]

3.3 Results

3.3.1 Primary Analyses

Consistent with previous literature, a significant J-shaped relationship between BMI and all-cause mortality was found using NLMR using the minimally adjusted model and BMI PRS (fractional polynomial non-linearity *p*-value = 0.0002, Figure 3.1). A significant relationship was also found using the fully adjusted model, but not the unadjusted model (fully adjusted model: $p = 0.0006$; unadjusted: 0.653). The J-shaped relationship was nominally significant using the metabolically neutral (MN) BMI PRS, fully adjusted for covariates ($p = 0.04$).

3.3.2 Comparison with the Cox Meta-Regression Model

When comparing the optimal BMI computed from the Cox meta-regression model in Chapter II and the one computed using NLMR, there were no significant differences between the two BMI values (*p* heterogeneity = 0.102, Figure 3.2).

3.3.3 Cause-Specific Analyses

Our above analyses were repeated, looking only at cancer, cardiovascular disease, respiratory disease, and other disease (i.e. all other diseases) mortality using the minimally

adjusted model and the BMI PRS. There were no significant differences between mortality subtypes in population optimal BMI using NLMR (*p* heterogeneity = 0.38, Figure 3.3).

3.3.4 Secondary Analyses

Using the minimally adjusted model and BMI PRS, the difference in the population optimal BMI between smokers and non-smokers was nominally significant (*p* heterogeneity = 0.003). There was a significant J-shaped relationship between BMI and all-cause mortality in smokers only and a positive linear relationship in non-smokers only (fractional polynomial non-linearity *p* value in smokers only: 0.0005; fractional polynomial non-linearity *p* value in non-smokers only: 0.129, Figure 3.4a-b).

3.3.5 Tertiary Analyses

The W-shaped and J-shaped relationship between LM and all-cause mortality found in our whole UKB population and sex-specific subgroup analyses respectively using Cox meta-regression were not found using NLMR, whether height or height^2 was adjusted for or not (fractional polynomial non-linearity $p > 0.05$ across all analyses).

3.3.6 Non-Linear Mendelian Randomization Sensitivity Analysis

After constructing bootstrapped confidence intervals for the optimal BMI extracted from the BMI-all-cause mortality relationship using the BMI PRS, there were certain

bootstrapped samples where the NLMR analysis did not output a J-shaped relationship between BMI and all-cause mortality (Figure 2.3). Rather, a positive linear relationship was outputted, meaning no optimal BMI would be extracted other than the lowest BMI value across the relationship, which was often ~15 kg/m^2, in the WHO-defined underweight BMI category. This phenomenon occurred in both our primary and secondary analyses.

Our bootstrapped samples often had different mortality:control ratios (i.e. the ratio of the number of deaths to the number of people alive in the sample) from each other within a given UKB subgroup. For example, bootstrap samples #2 and #6 within the younger adults (age <=65) UKB subgroup had a mortality:control ratio of 15,167: 313,292 and 15,287: 313,172 respectively. Nevertheless, this small fluctuation of 120 deaths between the two samples meant the difference between a significant J-shaped curve in sample #2 with a nadir of 22.8 kg/m^2 and a positive linear relationship in sample #6 with the lowest BMI being 15.1 kg/m^2 (Figure 3.5a-b).

For certain subgroups, like those below median WHR and above median LDL levels, the mean of the bootstrap distribution used to compute confidence intervals was far from the original nadir estimate. For example, those below the median WHR ratio in the UKB had a BMI nadir value of 15.0 kg/m^2, which falls in the underweight WHO BMI category, with the mean of the bootstrap distribution being 26.4 kg/m^2, which falls under the overweight WHO BMI category. For those above median LDL levels, the nadir value was 15.0 kg/m^2 while the mean of the bootstrap distribution was 20.2 kg/m^2, falling in the normal BMI range.

3.3.7 Calibration Interval Computation

Results using the quadratic model to fit the mean BMI per quantile and the local average causal estimate for the BMI-all-cause mortality relationship are shown in Figure 3.6. The differences in optimal BMI computed using the quadratic model for NLMR between individuals above and below median TG, CRP, and ALT are shown in Figure 3.7. The lower confidence interval bound for the optimal BMI computed in UKB participants below median TG and ALT levels was negative (Figure 3.7).

3.4 Discussion

3.4.1 Summary of Results

There was no significant difference between Cox meta-regression and NLMR models in computed population optimal BMI. The results found using NLMR were consistent with previous literature, but additionally, we discovered that metabolically neutral BMI may play a role in driving the J-shaped relationship between BMI and all-cause mortality. However, as our results were found to be nominally significant, we cannot draw further conclusions and thus, future studies need to explore the link between metabolically neutral BMI and the J-shaped BMI-mortality relationship.

3.4.2 The causal J-shaped relationship between BMI and all-cause mortality

The J-shaped relationship between BMI and all-cause mortality is causal, given that the NLMR-derived BMI-all-cause mortality relationship is analogous in shape to the

epidemiologically-derived curve, consistent with previous studies[1,6]. One explanation for why a J-shaped relationship exists between BMI and all-cause mortality may be that for certain cause-specific mortality outcomes, there is a monotonic increase in mortality risk with increasing BMI, while for other mortality outcomes, there is a monotonic decrease in risk with increasing BMI[1,6,42]. This may lead to genetically-increased BMI increasing mortality risk for the former, but decreasing mortality risk for the latter, thus shifting the optimal BMI value to the right. The epidemiological and genetic literature conflict in terms of the directionality of certain mortality outcomes with BMI. For example, epidemiological studies have found that many cardiovascular disease and cancer mortality outcomes had J-shaped relationships with BMI, while one NLMR study by Sun et al found that there was no relationship between cancer mortality and BMI and a positive association between BMI and cardiovascular disease mortality. It should be noted that Sun et al did not assess the relationship between BMI and different cardiovascular or cancer mortality outcomes, unlike other epidemiological studies. In our analyses, we were unable to assess cause-specific mortality outcomes beyond cancer, cardiovascular disease, or respiratory disease as there were less than 1000 deaths for all other mortality outcomes, which is insufficient for both epidemiological and NLMR analyses. Given both analyses require the population to be split into multiple quantiles, the sample size per quantile would be too small and underpowered to detect any meaningful association. Thus, it remains unclear which mortality outcomes are contributing to either end of the J-shaped relationship between BMI and all-cause mortality. Future studies should address this question in larger, better powered populations.

Not only could mortality outcomes modulate the J-shaped relationship, but also body composition measures, such as lean mass, which BMI does not account for[1,6,42]. Several studies have speculated that the lower end of the J-shaped BMI-all-cause mortality relationship may be due to lean mass loss or cachexia, which may be related to a number of different factors, such as age-related sarcopenia or smoking. No study has been done to identify whether specific body compositional change drives the upper or lower extreme of the J-shaped curve. However, studies have demonstrated an inverse relationship between lean mass and mortality, and one study showed that after adjusting for muscle mass, the optimal BMI shifts to the left towards the normal WHO-defined BMI category[44]. Thus, the loss of lean mass explaining the lower extreme of the J-shaped curve is supported by the epidemiological literature. However, our analyses showed that while lean mass and the BMI PRS were strongly correlated, a positive relationship was seen between the two, which is in contrast to previous literature. BMI, fat mass, and lean mass are all highly correlated with each other, both epidemiologically and at the genetic level[42]. Genetic variants for BMI could affect both fat and non-fat mass (i.e. lean mass) components[42]. Thus, even with Mendelian randomization, because it is not clear whether certain BMI genetic variants exclusively impact lean mass or fat mass, it can be difficult to assess the genetic effect of individual body composition components on all-cause mortality[42]. All of the clinical biomarkers we analyzed had a significant positive association with lean mass except total cholesterol, LDL, and HDL. Although HDL has been shown to be inversely related to BMI in epidemiological literature, total cholesterol, LDL, and HDL are highly correlated at the epidemiological and genetic level, and thus,

their specific impact on lean mass or BMI, and by extension, all-cause mortality cannot be delineated[27]. Furthermore, the epidemiologically determined W-shaped and J-shaped relationship between LM and all-cause mortality found in the whole UKB population and our sex-specific subgroups respectively using Cox meta-regression were not causal according to results found using NLMR. Therefore, the specific body composition component that is responsible for driving the lower end of the J-shape curve remains undetermined. Future studies should aim to either use or develop specific genetic techniques to delineate the genetic effect of lean mass, fat mass, cholesterol, LDL, and HDL on their effect on all-cause mortality. Knowing which body composition drives both the lower and upper end of the J-shaped curve can have profound public health implications: BMI recommendations can be tailored specifically to account for low muscle mass or high fat mass.

3.4.3 Limitations of the bootstrapping method to generate confidence intervals

Since our results show that the optimal BMI computed with the Cox meta-regression model and NLMR are not significantly different, it may appear that the optimal BMI of ~22-25 kg/m^2 is causal, but the limitations behind our methodology cast doubt on how confident we are in making this conclusion. After constructing bootstrapped confidence intervals for the optimal BMI extracted from the BMI-all-cause mortality relationship using the BMI PRS, there were certain bootstrapped samples where the NLMR analysis did not output a non-linear relationship between BMI and all-cause mortality. Rather, a positive linear relationship was outputted, with the optimal BMI in the underweight WHO

BMI category. Our sensitivity analyses show that the NLMR model is hypersensitive to small fluctuations in the number of deaths, casting doubt on the stability of the model.

For many subgroups, the mean of the bootstrap distribution used to compute confidence intervals was often quite far from the original nadir estimate. Large differences between the two values indicates that the original BMI nadir is biased, again casting doubt on the stability of NLMR. It also creates great difficulty in making comparisons between subgroups, due to the lack of power. Therefore, any variation in the NLMR-derived optimal BMI due to genetic or clinical biomarker variation cannot be reliably concluded.

The sample size per quantile may be the source of bias present in our NLMR analyses: we are ultimately limited by the size of the overall UKB cohort as the same issues occurred separating the UKB into 50 or 100 quantiles. As such, future studies should replicate this analysis in larger sample sizes to determine whether the NLMR improves in stability with increasing sample size, and by extension, number of deaths per quantile. This can facilitate better comparison between subgroups if confidence intervals are more precise as a result of having greater power.

3.4.4 Calibration interval method for computing confidence intervals

The NLMR model fits a smooth spline curve based on the integral computed from the LACE versus mean BMI per quantile meta-regression analysis, which uses a fractional polynomial model for the meta-regression[34]. However, due to the aforementioned instability of the fractional polynomial model in fitting the data across multiple bootstrap

samples with varying number of deaths and controls, we believed a quadratic model may provide more stability and power. After assessing the BMI-all-cause mortality relationship within the UKB and select subgroups using this method, we determined that the results were even more unstable than the results found using the initial bootstrapping method. The lower confidence interval bound for the optimal BMI computed in UKB participants below median TG levels was negative, which is not possible as BMI cannot be negative (Figure 3.7). Overall, we determined that, despite its limitations, the initial method of using bootstrapping to compute confidence intervals for the population optimal BMI is the most ideal.

3.4.5 Other limitations

While NLMR studies are less susceptible to residual confounding and reverse causality influences, our genetic instruments may still be subject to pleiotropy[1]. While vertical pleiotropy allows for a reliable causal inference as it involves genetic variants associated with an exposure (e.g. BMI) to impact a trait (e.g. systolic blood pressure) immediately downstream of the outcome to influence the outcome (e.g. coronary heart disease), horizontal pleiotropy, or the scenario where genetic variants influence the exposure and outcome through independent pathways, is a major concern in MR studies[32,33]. It is especially problematic with PRS: while PRS have the potential for improving genetic risk prediction as it includes multiple genetic variants associated with the trait to improve predictive power, pleiotropic influences can reduce the validity of PRS for MR[33]. Depending on the *p* value threshold used to select SNPs for the PRS, which impacts the

number of SNPs included in the final PRS, a mixture of true and false positive SNP-exposure associations may arise and this could introduce strong horizontal pleiotropic effects when the PRS is used in analyses[33]. Due to horizontal pleiotropy, it can be very difficult to disentangle the effects of BMI from other traits strongly associated with BMI, like TG, lean mass, or WHR, on mortality outcomes. Previous studies have raised concerns regarding the pleiotropic nature of BMI genetic variants, where variants can either be associated with fat mass or non-fat mass components[42]. However, prior NLMR studies have confirmed very little evidence of pleiotropy within BMI genetic variants, so pleiotropy is unlikely to have had a significant impact on our results[1]. Since MR estimates lifetime trends within individuals in a population as opposed to individual points in time, we cannot conclude that the trends we see in our results vary with time[1]. Finally, the other limitations mentioned at the end of chapter II also apply here.

3.5 Conclusion and Significance

Consistent with previous studies, a causal non-linear relationship between BMI and all-cause mortality was found in the whole UKB sample. It is difficult to make conclusions based on the results obtained from our other NLMR analyses due to the limitations of the model. Future studies should use larger and better powered studies to definitively confirm this and allow for more stable NLMR performance with respect to generating confidence and calibration intervals from bootstrapping. If a causal basis for our epidemiological analyses in Chapter II were to be confirmed, it would increase the level of confidence in

having individual genetic and clinical biomarker variation considered alongside traditional BMI recommendations. Such a discovery would lead to more precise recommendations, which could positively impact health care outcomes related to obesity and malnutrition, reducing the overall burden and cost to the health care system and increasing the quality of life for many patients.

3.6 References

1. Sun YQ, Burgess S, Staley JR*, et al*. Body mass index and all-cause mortality in HUNT and UK Biobank studies: linear and non- linear mendelian randomisation analyses. *BMJ* 2019; **364**: 1-10. DOI: 10.1136/bmj.l1042.

2. Nuttall FQ. Body Mass Index: Obesity, BMI, and Health: A Critical Review. *Nutr Today* 2015; **50**: 117-128. DOI: 10.1097/NT.0000000000000092.

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

4. Paré G, Mao S, Deng WQ. (2019). A robust method to estimate regional polygenic correlation under misspecified linkage disequilibrium structure. *Genet Epidemiol* 2019; **42**: 636-647. DOI: 10.1002/gepi.22149.

5. Marees AT, de Kluiver H, Stringer S, et al (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res* 2018; **27**: 1-10. DOI: 10.1002/mpr.1608.

6. Bhaskaran K, Dos-Santos-Silva I, Leon DA, *et al.* Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3.6 million adults in the UK. *Lancet Diabetes Endocrinol* 2018; **6**: 944-953. DOI: 10.1016/S2213-8587(18)30288-2.

7. Pigeyre M, Sjaarda J, Mao S, *et al.* Identification of novel causal blood biomarkers linking metabolically favorable adiposity with type 2 diabetes risk. *Diabetes Care* 2019; **42**: 1800-1808. DOI: 10.2337/dc18-2444.

8. Rothman KJ. BMI-related errors in the measurement of obesity. *Int J Obes (Lond)* 2008; **32**:1-4. doi: 10.1038/ijo.2008.87.

9. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med* 2020; **12**: 1-11. https://doi.org/10.1186/s13073-020-00742-5.

10. Wolfe BM, Kvach E, Eckel RH. Treatment of Obesity: Weight Loss and Bariatric Surgery. *Circ Res*. 2016;**118**:1844-1855. doi:10.1161/CIRCRESAHA.116.307591

11. Echouffo-Tcheugui JB, *et al.* Natural history of obesity subphenotypes: dynamic changes over two decades and prognosis in the Framingham Heart Study. *J Clin Endocrinol Metab* 2019;**104**:738–752. DOI: 10.1210/jc.2018-01321.

12. Henneberg M, Ulijaszek SJ. Body frame dimensions are related to obesity and fatness: Lean trunk size, skinfolds, and body mass index. *Am J Hum Biol*. 2010;**22**:83-91. doi: 10.1002/ajhb.20957.

13. Hu G, Tuomilehto J, Silventoinen K, *et al.* The effects of physical activity and body mass index on cardiovascular, cancer and all-cause mortality among 47 212 middle-aged Finnish men and women. *Int J Obes (Lond).* 2005;**29**: 894-902. doi: 10.1038/sj.ijo.0802870.

14. Mbatchou J, Barnard L, Backman J, *et al.* Computationally efficient whole genome regression for quantitative and binary traits. *bioRxiv* 2020; Published online January 1. DOI:10.1101/2020.06.19.162354.

15. Seidelmann SB, Claggett B, Cheng S, *et al.* Dietary carbohydrate intake and mortality: a prospective cohort study and meta-analysis. *Lancet Public Health*. 2018;**3**:419-428. doi: 10.1016/S2468-2667(18)30135-X.

16. Scrucca L, Santucci A, Aversa F. Competing risk analysis using R: an easy guide for clinicians. *Bone Marrow Transplant*, 2007; **40**, 381–387. DOI: 10.1038/sj.bmt.1705727.

17. Higgins JP, Thompson SG, Deeks JJ, *et al.* Measuring inconsistency in meta-analyses. *BMJ*. 2003;**327**:557-560. doi:10.1136/bmj.327.7414.557

18. Bhaskaran K, Douglas I, Forbes H*, et al.* Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet* 2014;**384**:755-65. DOI: 10.1016/S0140-6736(14)60892-8.

19. Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Stat Med.* 1997;**16**:753-68. doi: 10.1002/(sici)1097-0258(19970415)16:7<753::aid-sim494>3.0.co;2-g.

20. Kautzky-Willer A, Harreiter J, Pacini G. Sex and Gender Differences in Risk, Pathophysiology and Complications of Type 2 Diabetes Mellitus. *Endocr Rev* 2016;**37**:278-316. DOI:10.1210/er.2015-1137

21. Censin JC, Peters SAE, Bovijn J, *et al.* Causal relationships between obesity and the leading causes of death in women and men. *PLoS Genet* 2019; **15**:e1008405. DOI: 10.1371/journal.pgen.1008405.

22. Khera AV, Chaffin M, Wade KH, *et al.* Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell* 2019;**177**:587-596. DOI:10.1016/j.cell.2019.03.028.

23. Richardson TG, Harrison S, Hemani G, *et al.* An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *Elife* 2019;**8**:e43657. DOI: 10.7554/eLife.43657.

24. Howe LJ, Dudbridge F, Schmidt AF, *et al.* Polygenic risk scores for coronary artery disease and subsequent event risk amongst established cases. *Human Molecular Genetics* 2020; **29**, 1388–1395. DOI: 10.1093/hmg/ddaa052.

25. Manning AK, Hivert MF, Scott RA, *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genetics* 2012;**44**:659-669. DOI: 10.1038/ng.2274.

26. Udler MS, McCarthy MI, Florez JC, *et al.* Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocr Rev* 2019;**40**:1500-1520. DOI: 10.1210/er.2019-00088

27. Shamai L, Lurix E, Shen M, *et al.* Association of body mass index and lipid profiles: evaluation of a broad spectrum of body mass index patients including the morbidly obese. *Obes Surg* 2011;**21**:42-7. doi: 10.1007/s11695-010-0170-7.

28. Morin S, Leslie WD; Manitoba Bone Density Program. High bone mineral density is associated with high body mass index. *Osteoporos Int* 2009;**20**:1267-1271. DOI: 10.1007/s00198-008-0797-6.

29. Clifton EA, Day FR, De Lucia Rolfe E, *et al.* Associations between body mass index-related genetic variants and adult body composition: *The Fenland cohort study.* Int J Obes (Lond) 2017;**41**:613-619. DOI:10.1038/ijo.2017.11.

30. Sahoo K, Sahoo B, Choudhury AK, *et al.* Childhood obesity: causes and consequences. J *Family Med Prim Care* 2015;**4**:187-192. DOI:10.4103/2249-4863.154628.

31. Yengo L, Sidorenko J, Kemper KE, *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet.* 2018;**27**:3641-3649. DOI:10.1093/hmg/ddy271

32. Bennett, D.A., & Holmes, M.V. Mendelian randomisation in cardiovascular research: an introduction for clinicians. *Heart.* 2017;**103**: 1400-1407. doi: 10.1136/heartjnl-2016-310605

33. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet.* 2018;**27**:R195-R208. doi:10.1093/hmg/ddy163

34. Staley JR & Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genet Epidemiol.* 2017; **41**: 341-352. doi: 10.1002/gepi.22041

35. Mu M, Xu LF, Hu D, *et al.* Dietary Patterns and Overweight/Obesity: A Review Article. *Iran J Public Health.* 2017;**46**:869-876. doi: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5563867/pdf/IJPH-46-869.pdf. Published July 2017. Accessed April 2021.

36. McAuley PA, Blaha MJ, Keteyian SJ, *et al.* Fitness, Fatness, and Mortality: The FIT (Henry Ford Exercise Testing) Project. *Am J Med.* 2016;**129**:960-965. doi:10.1016/j.amjmed.2016.04.007.

37. Grandner MA, Schopfer EA, Sands-Lincoln M, *et al.* Relationship between sleep duration and body mass index depends on age. *Obesity (Silver Spring).* 2015;**23**:2491-2498. doi:10.1002/oby.21247.

38. Pingault J-B, Cecil CAM, Murray J ,*et al.* Causal Inference in Psychopathology: A Systematic Review of Mendelian Randomisation Studies Aiming to Identify Environmental Risk Factors for Psychopathology. *Psychopathology Review*. 2017:4-25. doi:10.5127/pr.038115

39. Alexander Krauss. Why all randomised controlled trials produce biased results, *Annals of Medicine*. 2018;**50**:312-322, DOI: 10.1080/07853890.2018.1453233

40. Greenwell, Brandon & Kabban, Christine. investr: An R Package for Inverse Estimation. *R Journal*. 2014;**6**, 90-100. doi: 10.32614/RJ-2014-009.

41. Perperoglou A, Sauerbrei W, Abrahamowicz M, *et al*. A review of spline function procedures in R. *BMC Med Res Methodol.* 2019;**19**:46. Published 2019 Mar 6. doi:10.1186/s12874-019-0666-3.

42. Bradbury KE, Cairns BJ. Understanding the relation between BMI and mortality. *BMJ* 2019; **364** :l1219. doi:10.1136/bmj.l1219

43. Chokshi DA, El-Sayed AM, Stine NW. J-Shaped Curves and Public Health. *JAMA.* 2015 Oct 6;**314**(13):1339-40. doi: 10.1001/jama.2015.9566. PMID: 26390229

44. Abramowitz MK, Hall CB, Amodu A, *et al*. Muscle mass, BMI, and mortality among adults in the United States: A population-based cohort study [published correction appears in PLoS One]. *PLoS One* 2018 ;**13**:e0198318].. 2018;**13**:e0194697. doi:10.1371/journal.pone.0194697

45. The Definition of Differentiation. World d(Math) Web. Updated October 14, 1999. Accessed May 26, 2021. https://web.mit.edu/wwmath/calculus/differentiation/definition.html

46. Antiderivatives. Mathematics – LibreTexts™. Updated November 10th, 2020. Accessed May 26, 2021. https://math.libretexts.org/@go/page/4469.

3.7 Figures

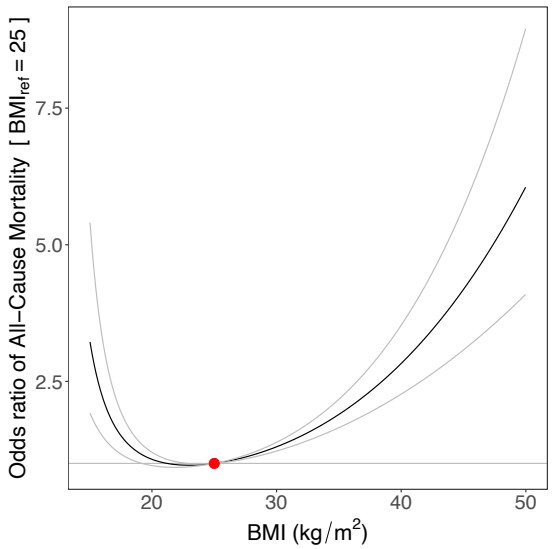


Figure 3.1 - The BMI – all-cause mortality relationship using non-linear mendelian randomization (NLMR) in the UKB.

The minimally adjusted model was used, with age, sex, and the first 10 principal components as covariates. The BMI PRS was used. Grey lines indicate 95% confidence intervals. Odds ratios for all-cause mortality are computed from determining the odds of all-cause mortality across the BMI distribution compared to the odds of all-cause mortality at the reference BMI of 25 kg/m^2, where the OR at this reference point is 1. Fractional polynomial non-linearity *p* value = 0.0002. BMI = body mass index, UKB = UK Biobank, PRS = polygenic risk score, BMIref = the reference BMI of 25 kg/m^2.

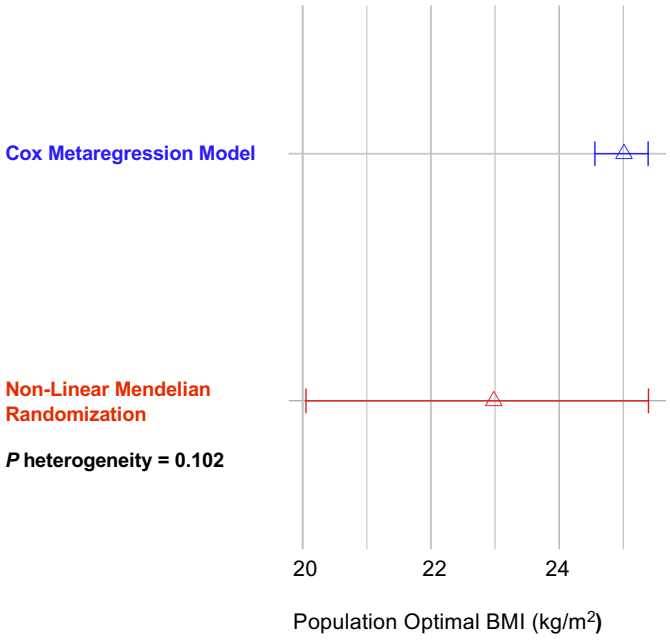


Figure 3.2 - Comparing the computed population optimal BMI using the Cox meta-regression and NLMR models.

P heterogeneity = 0.102. Heterogeneity was assessed using a fixed effects meta-analysis model. The minimally adjusted model was used, with age, sex, and the first 10 principal components as covariates. The BMI PRS was used. BMI = body mass index, PRS = polygenic risk score. Error bars represent 95% confidence intervals generated from bootstrapping analyses.

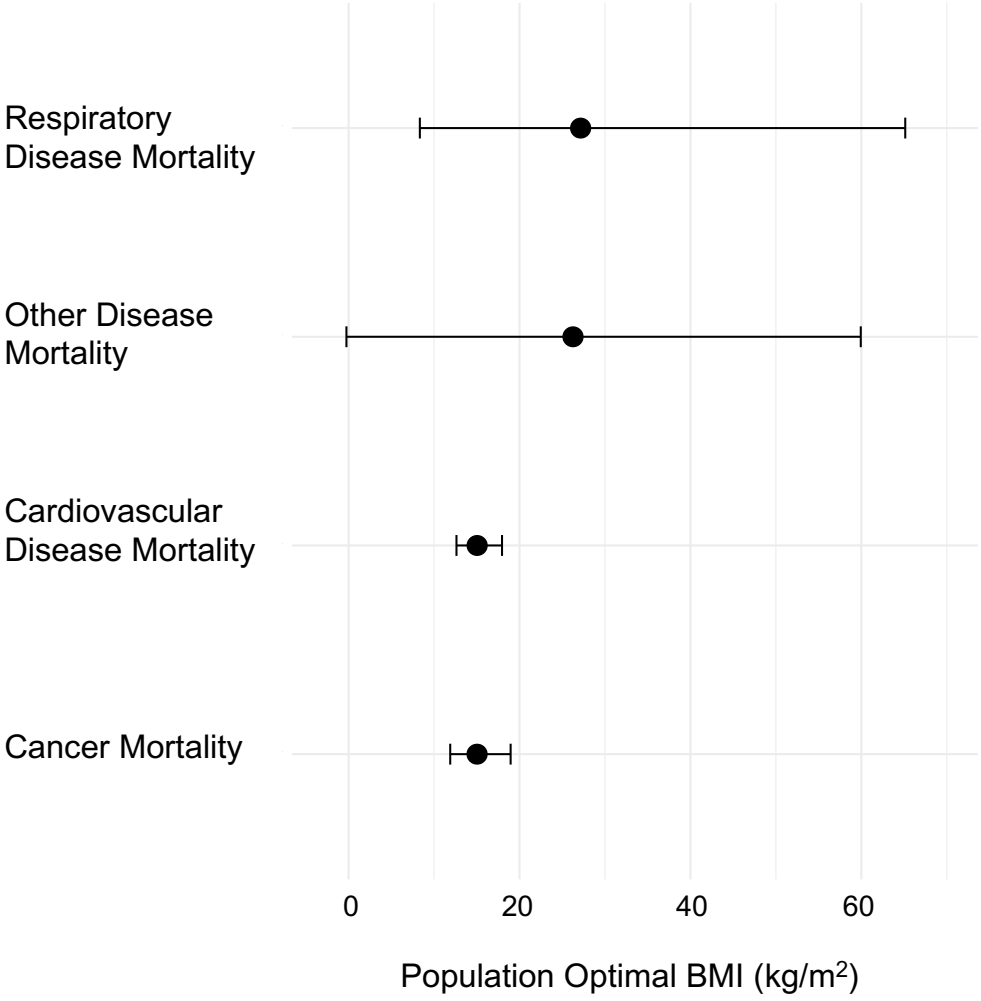


Figure 3.3 – Differences between the population optimal BMI against cause-specific mortality using NLMR.

The minimally adjusted model was used, with age, sex, and the first 10 principal components as covariates. The BMI PRS was used. *p* heterogeneity = 0.38. Other mortality refers to deaths other than from cancer, cardiovascular disease, or respiratory disease. Heterogeneity was assessed using a fixed effects meta-analysis model. Error bars

represent 95% confidence intervals generated from bootstrapping analyses. BMI = body mass index, CVD = cardiovascular disease, PRS = polygenic risk score.

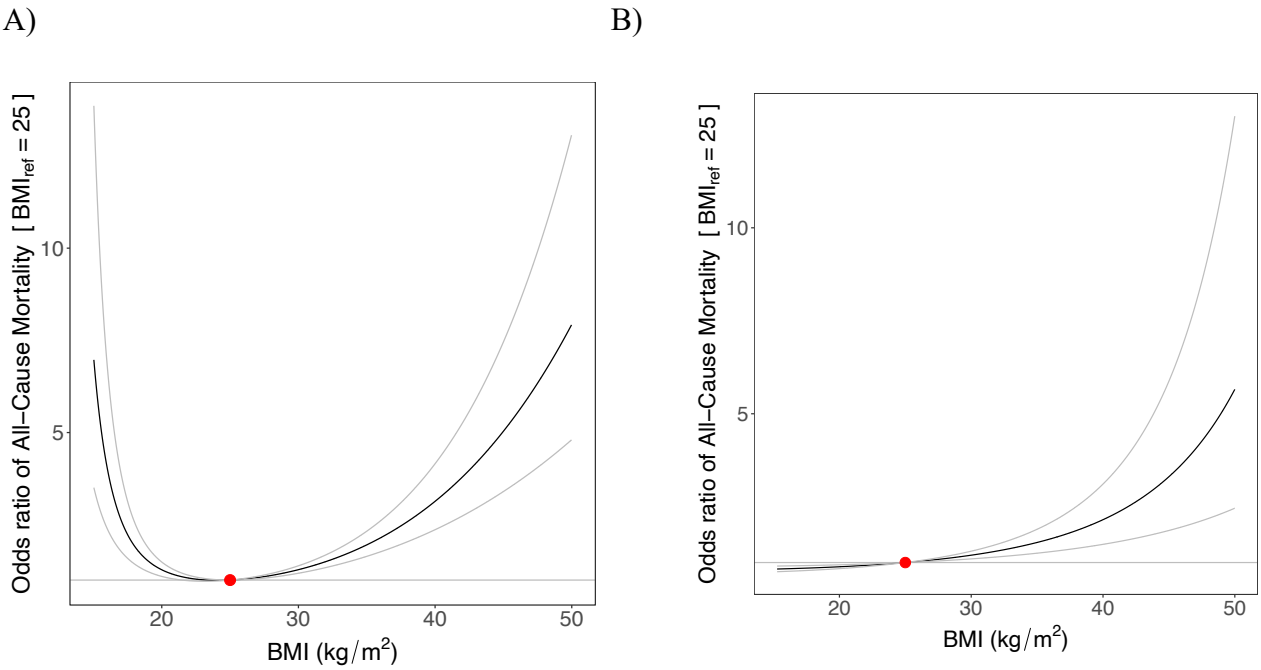


Figure 3.4 – The relationship between BMI and all-cause mortality in the UKB A) smokers only and B) never smokers only using NLMR.

The minimally adjusted model was used, with age, sex, and the first 10 principal components as covariates. The BMI PRS was used. Grey lines indicate 95% confidence intervals. Odds ratios for all-cause mortality are computed from determining the odds of all-cause mortality across the BMI distribution compared to the odds of all-cause mortality at the reference BMI of 25 kg/m^2, where the OR at this reference point is 1. Fractional polynomial non-linearity *p* value for smokers only = 0.0005; for non-smokers only = 0.129. BMI = body mass index, UKB = UK Biobank, BMIref = the reference BMI of 25 kg/m^2.

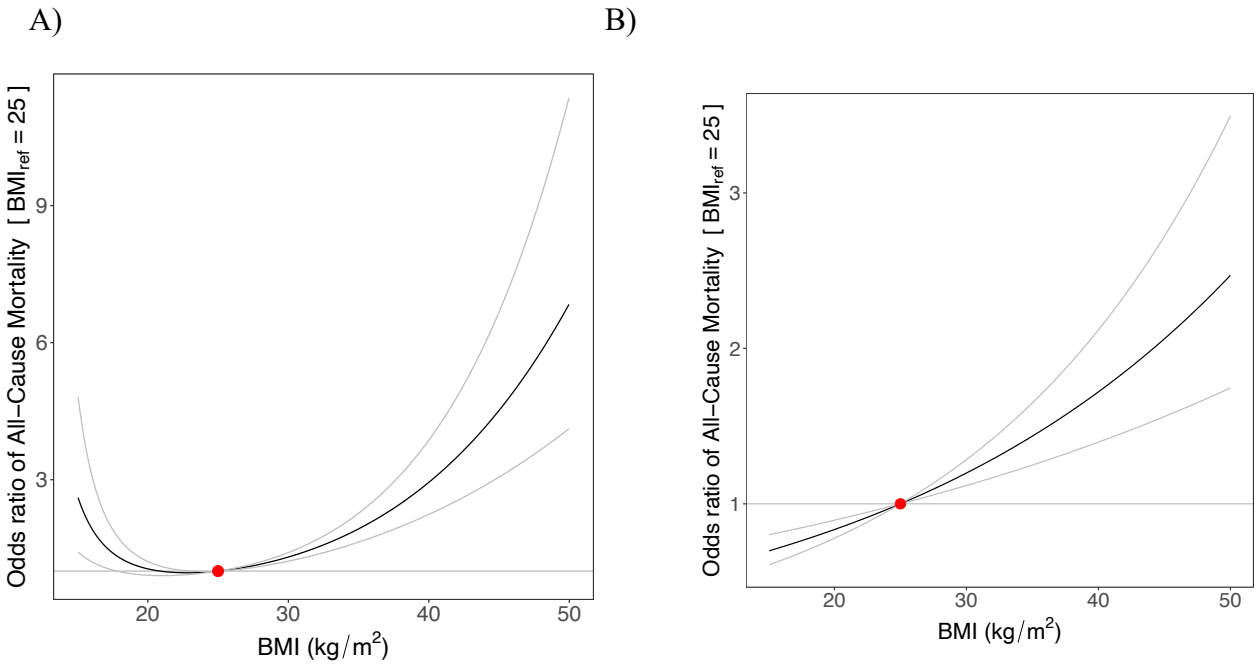


Figure 3.5 - BMI-all-cause mortality relationships outputted from bootstrapped samples A) #2 and B) #6 of the younger adult (age <= 65) subgroup of the UKB using NLMR. The minimally adjusted model was used, with sex and the first 10 principal components as covariates. The BMI PRS was used. Grey lines indicate 95% confidence intervals. Odds ratios for all-cause mortality are computed from determining the odds of all-cause mortality across the BMI distribution compared to the odds of all-cause mortality at the reference BMI of 25 kg/m^2, where the OR at this reference point is 1. Fractional polynomial non-linearity *p* value for sample #2 = 0.001; for sample #6 = 1. BMI = body mass index, UKB = UK Biobank, BMIref = the reference BMI of 25 kg/m^2.

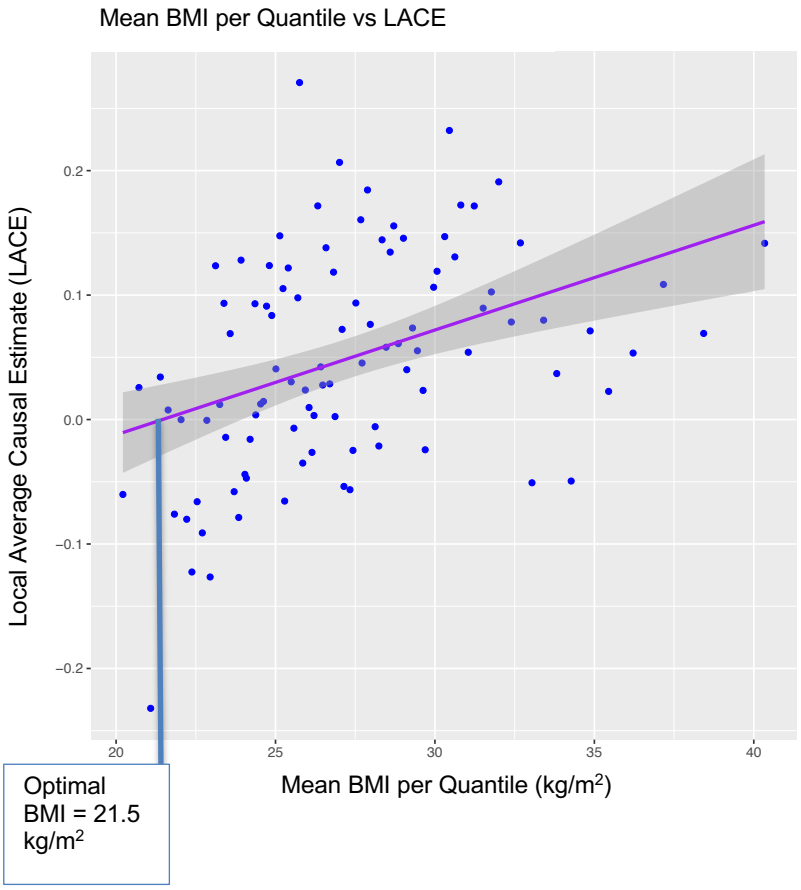


Figure 3.6 - Derivative of the quadratic model fit between mean BMI per quantile and the local average causal estimate for the BMI-all-cause mortality relationship.

The derivative is the linear model fit between the independent and dependent variable (purple line). The BMI associated with the point where the linear model predicts a LACE of 0 is where the optimal BMI resides, as the LACE of 0 indicates the point of inflection between the decreasing and the increasing component of the J-shaped curve (dark blue line). LACE = local average causal estimate, BMI = body mass index.

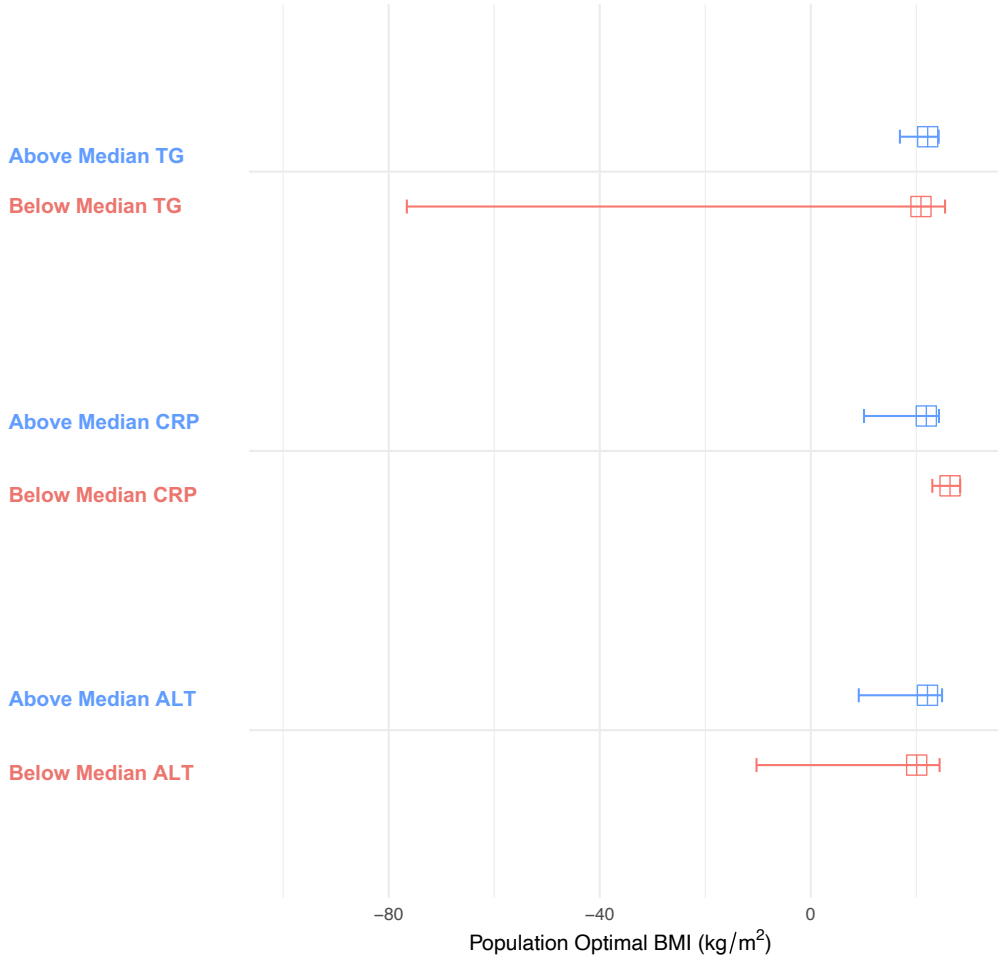


Figure 3.7 - Comparison of the population optimal BMI and UKB participants above and below median TG, CRP, and ALT levels using the quadratic model for NLMR. Results from the minimally adjusted model are shown. Bonferroni significance is considered at $p < 0.0008$. TG = triglyceride, CRP = C-reactive protein, ALT = alanine aminotransferase, BMI = body mass index. Error bars represent 95% CI obtained from bootstrapping analyses. Note that because of the unreliable confidence intervals, no heterogeneity tests were completed.

Chapter IV: Conclusion

4.1 Summary of Main Findings

Our results show that the optimal BMI for the BMI–all-cause mortality relationship varied due to genetic variation in BMI (0.10 kg/m^2 per quantile [95% CI 0.04-0.17], P = 0.005; Figure 2.4). There is a difference of 1.90 kg/m^2 in predicted optimal BMI between individuals in the top and bottom 5th BMI PRS percentile. Variation in circulating TG, CRP, and ALT levels was associated with variation in the optimal BMI for the BMI-all-cause mortality relationship (above/below median biomarker level analyses: TG - p heterogeneity: 1.29×10^{-4}; CRP - p heterogeneity =7.92×10^{-5}; ALT- p heterogeneity = 2.70×10^{-8}; percentile of clinical biomarker analyses: TG - p heterogeneity: 2.00×10^{-4}; CRP - p heterogeneity =1.26×10^{-6}; ALT- p heterogeneity = 5.00×10^{-4}). The population optimal BMI computed using the Cox regression model does not significantly differ from the population optimal BMI computed using the non-linear mendelian randomization (NLMR) model (p heterogeneity = 0.102). Metabolically neutral BMI appeared to modestly contribute to optimal BMI variation using both the Cox meta-regression and NLMR model, although the results were nominally significant and thus warrant further investigation.

4.2 Clinical Implications

BMI recommendations made by the WHO were determined by previous epidemiological studies that reached a consensus as to which BMI range conferred the lowest mortality risk[2,6]. However, recent epidemiological studies have shown that the lowest risk of mortality may be around the 25-26 kg/m^2 range, which falls under the WHO-defined

overweight category of BMI[6]. Such studies already gave rise to the notion that WHO categories may be inadequate to capture the ideal BMI range in the population. Our study demonstrated further evidence of what the WHO BMI recommendations fail to capture: individual genetic and clinical biomarker level variation. This is significant because it raises the question on whether or not genotyping and/or biochemical profiling can be used to make more precise BMI recommendations for individuals.

While our study further highlights the flaws with the WHO BMI recommendations, it cannot itself serve as the basis for change in clinical BMI recommendations. First, the results we have obtained from the Cox meta-regression are observational, meaning they are subject to residual confounding or reverse causality[1,6]. We thought that replicating our results using NLMR, which models the non-linear causal relationship between BMI and all-cause mortality, could help overcome these biases[1]. While our NLMR analyses were consistent with previous literature, overall, NLMR analyses were not consistent with our Cox meta-regression model (i.e. epidemiological) analyses, leaving the question of whether or not our results using Cox meta-regression were truly unconfounded.

The fact that the population optimal BMI computed in the UKB did not differ between the Cox meta-regression and NLMR model may have suggested that the computed optimal BMI was causal or unconfounded. However, as described in Chapter III, the NLMR is highly sensitive to small changes in sample size, calling into question its stability to produce reliable confidence intervals for an accurate comparison between the two nadir values. Furthermore, the quadratic model for the NLMR offered significantly

less stability than the default fractional polynomial model. This is because for our quadratic model, we made the strong assumption that the underlying model was still J-shaped. Since the quadratic model is more simplistic in design compared to the fractional polynomial model, if this assumption were correct, this model would have allowed for a more precise and powered analysis. However, given the instability of the quadratic model, it is likely that this assumption was not met. Thus, this model, however more powered, lacks the validity found in the original fractional polynomial method. Overall, we cannot conclude that our results obtained from Cox meta-regression are causal because of these limitations. Further studies should use more larger and more well powered studies when using causal frameworks like NLMR to replicate our analyses, as sample size, particularly in terms of number of deaths per sample, was most likely not large enough to maintain stability in computing bootstrapped estimates of the optimal BMI. If such a study confirmed that our results were causal, this would make for a stronger case for using genetics and clinical biomarkers when making BMI recommendations in the future.

The use of PRS in general for clinical practice, despite its strong potential clinical utility for screening diseases such as CVD or breast cancer, is currently unestablished and controversial[31]. In several cohorts, clinical use of PRS is most useful for individuals with a higher predisposition to a disease/condition, providing potential therapeutic advantage for those in the early stages of the disease/condition in particular: treatment options or understanding of the disease progression can thus be guided through PRS[9]. The strongest example of this can be seen with CVD: for example, when examining the use of statin, a

cholesterol-lowering medication to manage CVD, in reducing the risk of the first coronary event, several studies demonstrated that the reduction in relative risk for the first coronary event using statin is higher for those with a higher genetic risk for CVD, which is consistent with previous epidemological literature highlighting the stronger beneficial effect of statins in those with diabetes, hypertension, or high C-reactive protein (CRP) levels[9]. In the United Kingdom (UK), mammogram screening for breast cancer is offered to women over the age of 47, where the mean 10-year breast cancer risk is approximately 2.6%[9]. One study by Mavaddat et al. showed that 20% of women with the highest genetic risk for breast cancer according to their PRS reached the 2.6% mean 10-year breast cancer risk before the age of 40, while 20% of women with the lowest PRS for breast cancer never reached this risk[9]. This study demonstrates that PRS for breast cancer can be used to profile a subgroup of women at high-risk for breast cancer that would benefit from mammogram screening for breast cancer at an earlier age than what the current UK public health guidelines provide[9]. In fact, there are available commercial breast cancer risk tests that use PRS: riskScore™ and AmbryScore[9]. There is potential for use of PRS in conjunction with pre-existing WHO BMI recommendations for reasons analogous to breast cancer risk: genetics can capture information that the WHO BMI recommendations cannot and thus, could be useful in making more accurate recommendations. However, to our knowledge, the results obtained from the current study is the only existing source of evidence to support using PRS for this purpose. More research is needed before PRS can be hoped to be considered as an adjunct to current clinical BMI recommendations.

4.3 Research Implications

There was variation in the optimal BMI due to variation in triglyceride (TG), C-reactive protein (CRP), and alanine aminotransferase (ALT) using the Cox meta-regression model. TG, CRP, and ALT have shown to be positively associated with BMI in previous literature[12,27,30]. However, no consensus has been reached in the literature regarding the directionality of certain relationships due to inconsistency between epidemiological and MR studies[12,27,30]. Our results found using the Cox meta-regression model were not replicated using the NLMR model as we were significantly underpowered for our NLMR analyses; thus, we cannot make further conclusions about the biological significance behind the results nor confirm or deny the inconsistency between epidemiological and MR studies. Higher levels of TG, CRP, and ALT may be reflective of individuals having higher BMI due to the strong correlation between these biomarkers and BMI; thus, when we select individuals above median concentrations of these biomarkers, the optimal BMI may be higher because individuals have a higher BMI to begin with[12,27,30]. Whether this higher nadir value is causal remains unanswered. Our results should be replicated using NLMR in larger studies with greater mortality:control ratios for stronger power. If it is true that clinical biomarkers play a significant causal role in optimal BMI variation, then biochemical profiling may be also considered as an adjunct to existing clinical BMI recommendations.

4.4 Limitations and Future Directions

Although the UKB is a large, genetically homogenous cohort with opportunities for expansive phenotyping, the sample may not be large enough for NLMR analyses to have power, considering the model relies on the sample size per quantile, and by extension, the number of deaths per quantile[34]. The issue surrounding limited sample size was manifested within the limitations of both the quadratic and fractional polynomial-based models for NLMR when producing bootstrapped confidence intervals, making it harder to conclude causality in our epidemiological analyses. Future studies should replicate our findings in larger, more powered studies, especially with larger mortality:control ratios to address these limitations. While there were many efforts made to reduce confounders, because all of our results are observational, we cannot be sure whether our studies are still subject to significant residual confounding, horizontal pleiotropic effects, or environmental factors strongly associated with either BMI or SNPs associated with BMI, such as lung function or eating behaviour, especially when we were not powered enough to confirm causality using NLMR. However, since the BMI-mortality relationship has been validated through mendelian randomization analyses in previous studies, our results may still be relatively robust to these confounders[1,19,37,38]. Future studies should explore how other potential confounders of BMI may influence the optimal BMI to confirm this, especially potential confounders currently unavailable in the UKB. While we used a large genetically homogeneous British Caucasian population with minimal genetic ancestral confounding, our results cannot be easily generalized to other populations for which the optimal BMI varies greatly due to genetic ancestry. This limitation also applies for our

PRS[9]. Future studies should replicate our analyses in multi-ethnic studies to prove generalizability of our findings in non-British Caucasian populations. Studies have not determined what type of adiposity the metabolically beneficial, deleterious, or neutral BMI encapsulates[9]. Future studies need to determine whether the genes associated with metabolically favourable, deleterious, or neutral BMI PRS are related to a specific type of fat, such as subcutaneous or visceral fat, as this could provide more information on which type of fat presents greater protection against mortality for those at the higher end of these PRS or how metabolically neutral BMI is contributing to the genetic variation in BMI, which could further help guide recommendations or treatment plans[9]. There are also several practical limitations in using PRS in the clinical setting. Using PRS in a framework that consists of integrating genotype and whole-genome sequence data into clinical records, computing an individual PRS for the patient and interpreting the score against an appropriate genome reference population is not yet justified in the literature and offers a challenge in terms of practical implementation[9]. More importantly, such a reference population may not be available as a point of comparison, leaving PRS computed for individuals in certain populations to be left uninterpreted and thus, not useful[9]. The usefulness of PRS for clinical screening in those with non-European ancestry and those that are genetically admixed (i.e. having genetic ancestry from two or more previously isolated populations, thus having ancestors from multiple sources) is also not clear from the current literature, limiting applicability to populations outside of European descent, which can further weaken the case of bringing PRS into clinical practice for BMI recommendations, as this is true with most obesity genetic studies[1,33].

4.5 Final Remarks

We hope our research provides further evidence for support of clinical use of PRS to optimize patient health, well-being, and quality of life. Although more research needs to be conducted, such advances in precision medicine could have large implications for public health recommendations for BMI. Obesity has been linked with increased hospitalization and usage of medical services[35]. In 2016, one study showed that annualized health care expenditures were $1,496 USD higher for those who were obese compared to those in the normal weight category[35]. Much of the excessive health care utilization and expenditures associated with obesity were explicated by chronic condition/illness and poor health, either as a cause or consequence of obesity[35]. Obesity is considered a preventable condition[35]. BMI is also used to determine eligibility for obesity treatments, such as bariatric surgery[10]. Helping individuals set a BMI target that works best for them based on their unique characteristics can optimize their health and prevention of the negative health care outcomes associated with obesity, both in terms of the medical cost for the individual and economic cost to the health care system. Positive consequences of achieving an ideal BMI potentially include improved cardiovascular, metabolic, and mental health[2]. Using genetic and clinical biomarkers to adjust BMI recommendations could also set precedence for re-consideration of the existing BMI cut-off points for specific obesity interventions, like bariatric surgery[10]. As BMI may not be the best indicator of ideal weight given its limitations, WHR or LM, which have been

shown to be better at predicting risk for disease and/or mortality compared to BMI, may be considered for use alongside BMI recommendations to address much of what BMI cannot account for, both through the traits themselves and through genetics. The sex-specific J-shaped relationship between LM and all-cause mortality indicates that there might be an optimal amount of lean mass to obtain and that this optimum differs between males and females. Although these results are not confirmed to be causal, they nevertheless still show how other body composition measurements could be potentially useful in informing clinical BMI recommendations. Just as the definition of BMI has constantly changed over time due to evidence-based medicine, we hope our findings can provide the catalyst to bring change once again.

4.6 References

1. Sun YQ, Burgess S, Staley JR, *et al*. Body mass index and all-cause mortality in HUNT and UK Biobank studies: linear and non- linear mendelian randomisation analyses. *BMJ* 2019*;* **364**: 1-10. DOI: 10.1136/bmj.l1042.

2. Nuttall FQ. Body Mass Index: Obesity, BMI, and Health: A Critical Review. *Nutr Today* 2015; **50**: 117-128. DOI: 10.1097/NT.0000000000000092.

3. Aune D, Sen A, Prasad M, *et al.* BMI and all cause mortality: systematic review and non-linear dose-response meta-analysis of 230 cohort studies with

3.74 million deaths among 30.3 million participants. *BMJ* 2016; **353**:1-17. doi:10.1136/bmj.i2156

4. Paré G, Mao S, Deng WQ. (2019). A robust method to estimate regional polygenic correlation under misspecified linkage disequilibrium structure. *Genet Epidemiol* 2019; **42**: 636-647. DOI: 10.1002/gepi.22149.

5. Marees AT, de Kluiver H, Stringer S, *et al* (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res* 2018; **27**: 1-10. DOI: 10.1002/mpr.1608.

6. Bhaskaran K, Dos-Santos-Silva I, Leon DA, *et al*. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3.6 million adults in the UK. *Lancet Diabetes Endocrinol* 2018; **6**: 944-953. DOI: 10.1016/S2213-8587(18)30288-2.

7. Pigeyre M, Sjaarda J, Mao S, *et al*. Identification of novel causal blood biomarkers linking metabolically favorable adiposity with type 2 diabetes risk. *Diabetes Care* 2019; **42**: 1800-1808. DOI: 10.2337/dc18-2444.

8. Emdin CA, Khera AV, Natarajan P, *et al*. Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *JAMA,* 2017; **317**: 626-634. DOI: 10.1001/jama.2016.21042.

9. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med* 2020; **12:** 1-11. https://doi.org/10.1186/s13073-020-00742-5.

10. Schorr M, Dichtel, LE, Gerweck, AV, *et al.* Sex differences in body composition and association with cardiometabolic risk. *Biol Sex Differ* 2018; **9,** 1-10. DOI: 10.1186/s13293-018-0189-3.

11. Echouffo-Tcheugui JB, *et al*. Natural history of obesity subphenotypes: dynamic changes over two decades and prognosis in the Framingham Heart Study. *J Clin Endocrinol Metab* 2019;**104**:738–752. DOI: 10.1210/jc.2018-01321.

12. Timpson NJ, Nordestgaard BG, Harbord RM, *et al.* C-reactive protein levels and body mass index: elucidating direction of causation through reciprocal Mendelian randomization. *Int J Obes (Lond)* 2011; **35**, 300-308. DOI:10.1038/ijo.2010.137

13. Tam V, Patel N, Turcotte M. *et al.* Benefits and limitations of genome-wide association studies. *Nat Rev Genet* 2019; **20**, 467–484. DOI: 10.1038/s41576-019-0127-1.

14. Mbatchou J, Barnard L, Backman J, *et al*. Computationally efficient whole genome regression for quantitative and binary traits. *bioRxiv* 2020; Published online January 1. DOI:10.1101/2020.06.19.162354.

15. Newcombe PJ, Nelson CP, Samani NJ, *et al*. A flexible and parallelizable approach to genome-wide polygenic risk scores. *Genetic Epidemiology* 2019;**43**:730-741. DOI: 10.1002/gepi.22245.

16. Scrucca L, Santucci A, Aversa F. Competing risk analysis using R: an easy guide for clinicians. *Bone Marrow Transplant*, 2007; **40**, 381–387. DOI: 10.1038/sj.bmt.1705727.

17. Klempfner R, Erez A, Sagit BZ, *et al*. Elevated Triglyceride Level Is Independently Associated With Increased All-Cause Mortality in Patients With Established Coronary Heart Disease: Twenty-Two-Year Follow-Up of the Bezafibrate Infarction Prevention Study and Registry. *Circ Cardiovasc Qual Outcomes,* 2016;**9**:100-108. DOI: 10.1161/CIRCOUTCOMES.115.002104.

18. Bhaskaran K, Douglas I, Forbes H, *et al*. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet* 2014;**384**:755-65. DOI: 10.1016/S0140-6736(14)60892-8.

19. Mafort, TT, Rufino R, Costa CH, *et al.* Obesity: systemic and pulmonary complications, biochemical abnormalities, and impairment of lung function. *Multidiscip Respir Med* 2016; **11;** 1-11. DOI: 10.1186/s40248-016-0066-z.

20. Kautzky-Willer A, Harreiter J, Pacini G. Sex and Gender Differences in Risk, Pathophysiology and Complications of Type 2 Diabetes Mellitus. *Endocr Rev* 2016;**37**:278-316. DOI:10.1210/er.2015-1137

21. Censin JC, Peters SAE, Bovijn J, *et al*. Causal relationships between obesity and the leading causes of death in women and men. *PLoS Genet* 2019; **15**:e1008405. DOI: 10.1371/journal.pgen.1008405.

22. Khera AV, Chaffin M, Wade KH, *et al*. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell* 2019;**177**:587-596. DOI:10.1016/j.cell.2019.03.028.

23. Richardson TG, Harrison S, Hemani G, *et al*. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *Elife* 2019;**8**:e43657. DOI: 10.7554/eLife.43657.

24. Howe LJ, Dudbridge F, Schmidt AF, *et al*. Polygenic risk scores for coronary artery disease and subsequent event risk amongst established cases. *Human Molecular Genetics* 2020; **29**, 1388–1395. DOI: 10.1093/hmg/ddaa052.

25. Manning AK, Hivert MF, Scott RA, *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genetics* 2012;**44**:659-669. DOI: 10.1038/ng.2274.

26. Udler MS, McCarthy MI, Florez JC, *et al*. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocr Rev* 2019;**40**:1500-1520. DOI: 10.1210/er.2019-00088

27. Shamai L, Lurix E, Shen M, *et al*. Association of body mass index and lipid profiles: evaluation of a broad spectrum of body mass index patients including the morbidly obese. *Obes Surg* 2011;**21**:42-7. doi: 10.1007/s11695-010-0170-7.

28. Morin S, Leslie WD; Manitoba Bone Density Program. High bone mineral density is associated with high body mass index. *Osteoporos Int* 2009;**20**:1267-1271. DOI: 10.1007/s00198-008-0797-6.

29. Clifton EA, Day FR, De Lucia Rolfe E, *et al.* Associations between body mass index-related genetic variants and adult body composition: The Fenland cohort study. *Int J Obes (Lond)* 2017;**41**:613-619. DOI:10.1038/ijo.2017.11.

30. Liu, J., Au Yeung, S.L., Kwok, M.K. *et al.* The effect of liver enzymes on adiposity: a Mendelian randomization study. *Sci Rep* **9,** 16792 (2019). https://doi.org/10.1038/s41598-019-52489-8

31. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med* 2020; **12:** 1-11. https://doi.org/10.1186/s13073-020-00742-5.

32. Pinal-Fernandez I, Casal-Dominguez M, Mammen AL. Statins: pros and cons. *Med Clin (Barc*). 2018;**150**:398-402. doi:10.1016/j.medcli.2017.11.030

33. Korunes KL, Goldberg A. Human genetic admixture. *PLoS Genet*. 2021 Mar 11;**1**7:e1009374. doi: 10.1371/journal.pgen.1009374. PMID: 33705374; PMCID: PMC7951803.

34. Robinson PN. Deep phenotyping for precision medicine. *Hum Mutat.* 2012 May;**33**:777-80. doi: 10.1002/humu.22080.

35. Musich S, MacLeod S, Bhattarai GR, *et al.* The Impact of Obesity on Health Care Utilization and Expenditures in a Medicare Supplement Population. Gerontol Geriatr Med. 2016;2:2333721415622004. Published 2016 Jan 19. doi:10.1177/2333721415622004

36. Tran BX, Nair AV, Kuhle S, Ohinmaa A, Veugelers PJ. Cost analyses of obesity in Canada: scope, quality, and implications. *Cost Eff Resour Alloc*. 2013;**11**:3. Published 2013 Feb 8. doi:10.1186/1478-7547-11-3

37. Wardle J. Eating behaviour and obesity. *Obes Rev*. 2007;**8**:73-75. doi:10.1111/j.1467-789X.2007.00322.x

38. Qi L, Cho YA. Gene-environment interaction and obesity. *Nutr Rev.* 2008;**66**:684-694. doi:10.1111/j.1753-4887.2008.00128.x

Appendices

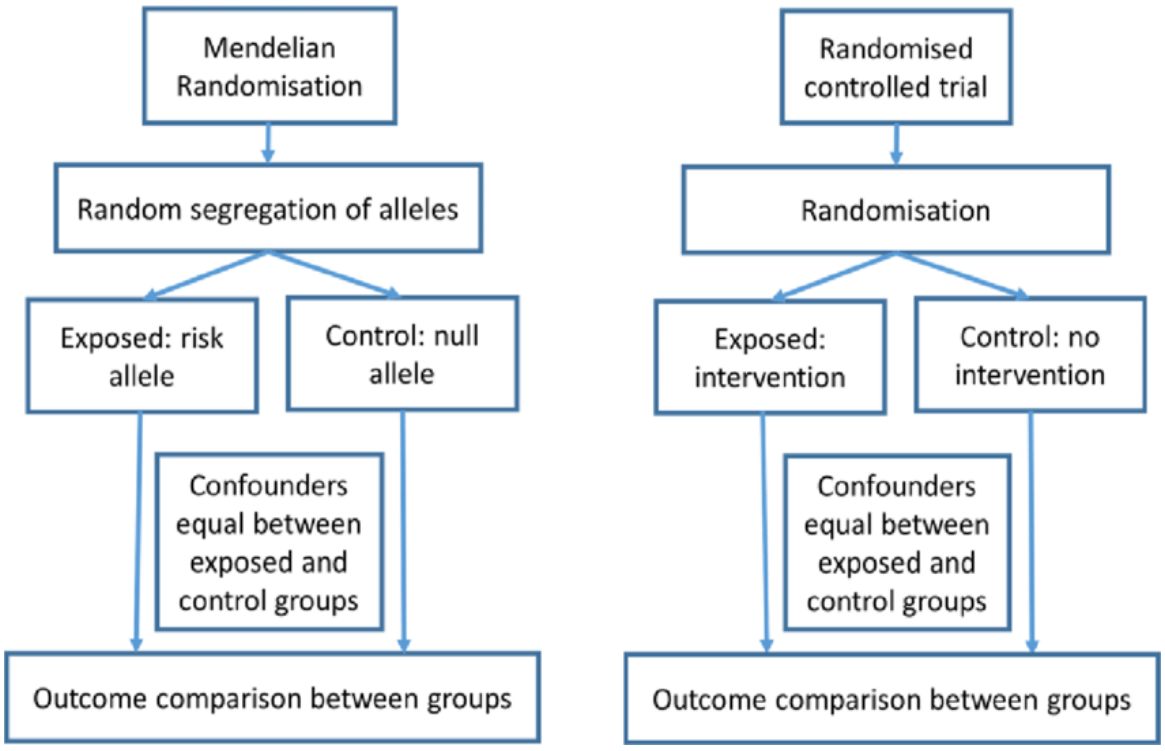


Figure 3.8 - Comparing Mendelian Randomization and Randomized Control Trial Methodology.

Table 2.1 – A list of a) GWAS consortia used to derive PRS, b) exposures, c) outcomes, and d) clinical biomarker variables used

a)

Phenotype	**GWAS Consortia Name**
Body mass index (BMI, including metabolically favourable, deleterious, and neutral BMI)	Genetic Investigation of ANthropometric Traits (GIANT)[1]
Triglyceride (TG)	Million Veteran Program (MVP)[5]
Whole body lean mass (LM)	GEnetic Factors for OSteoporosis Consortium (GEFOS)[6]
Waist-to-hip ratio (WHR)	Genetic Investigation of ANthropometric Traits (GIANT)[1]

b)

Phenotype	**UKB Field ID/ICD-10 Code**
BMI	21001-0.0
Sex	31-0.0
Age	21022-0.0

LM	23101-0.0
Smoking Status	20116-0.0
Diabetes	2443-0.0 and ICD-10 Code E1[0-4]
Alcohol	1558-0.0
Index of Multiple Deprivation (IMD)	IMD data from England: 26410-0.0 IMD data from Scotland: 26427-0.0 IMD data from Wales: 26426-0.0
UKB Assessment Centre	54-0.0
Height	50-0.0

c)

Mortality Outcome	**UKB Field ID/ICD-10 Code**
All-Cause Mortality	40000-0.0
Cardiovascular Mortality	ICD-10 Code: I
Cancer Mortality	ICD-10 Code: C
Respiratory Disease Mortality	ICD-10 Codes: J00-09, J10-19, J20-22, J23-29, J3-9

Other Mortality*	All ICD-10 Codes except I, C, and J00-09, J10-19, J20-22, J23-29, J3-9

d)

Clinical Biomarker	**UKB Field ID/ICD-10 Code**
Triglycerides (TG)	30870-0.0
Low-density lipoprotein (LDL)	30780-0.0
High-density lipoprotein (HDL)	30760-0.0
Hemoglobin 1Ac (Hb1Ac)	30750-0.0
C-reactive protein (CRP)	30710-0.0
Cholesterol	30690-0.0
Alanine Aminotransferase	30620-0.0
WHR	48-0.0 and 49-0.0

Table 2.2 - Baseline characteristics of participants in the UK Biobank (UKB)

Baseline characteristics	All UKB participants	All deaths	Controls
Number of participants	387,692	23,094	364,598
Percentage (%) of men	177,346 (45.7)	13,732 (59.5)	163,614(44.9)
Mean (SD) age at baseline (years)	56.9 (8.0)	61.8(6.3)	56.6(8.0)
Mean (SD) body mass index (kg/m^2)	27.4 (4.6)	28.1(5.1)	27.3(4.6)
Mean (SD) whole body lean mass (kg)	53.4 (11.5)	55.3(11.6)	53.2(11.5)
Mean (SD) whole body fat mass (kg)	24.8 (9.3)	25.6(10.0)	24.8(9.3)
Mean (SD) waist-to-hip ratio	0.87 (0.1)	0.91(0.1)	0.87(0.1)
Percentage (%) of ever smokers	175,504 (45.3)	14,229 (61.6)	161,275 (44.2)

Percentage (%) of people with diabetes	26,162 (6.7)	3,960 (17.1)	22,022 (6.1)
Demographics of alcohol consumption	*Daily or almost daily*: 82,221(21.2) *Three or four times a week*: 93,938 (24.2) *Once or twice a week*: 102,714 (26.5) *One to three times a month*: 43,078 (11.1)	*Daily or almost daily*: 5,473 (23.7) *Three or four times a week*: 4,621 (20.0) *Once or twice a week*: 5,450 (23.6) *One to three times a month*: 2,204 (9.5) *Special occasions only*: 2,983 (12.9)	*Daily or almost daily*: 76,748(21.1) *Three or four times a week*: 89,317 (24.5) *Once or twice a week*: 97,264 (26.7) *One to three times a month*: 40,874 (11.2) *Special occasions only*: 37,743 (10.4) *Never*: 22,652 (6.21)

	Special occasions only: 40,726(10.5) *Never*: 25,015 (6.5)	*Never*: 2,363 (10.2)	
Mean (SD) Index of Multiple Deprivation**	16.5 (13.6)	19.5(15.8)	16.5(13.6)

Table 2.3 - Association between BMI and major covariates used in analyses.

For binary covariates, the odds ratio (OR) was reported; for continuous covariates, the beta regression coefficient (β) was reported. Significance was set at $p < 0.05$. PC = principal component, IMD = index of multiple deprivation, UKB = UK Biobank, BMI = body mass index.

Covariate ~ BMI Association	**β/OR**	***P***	**Adjusted R-squared**
Age ~ BMI	0.08	$< 2 \times 10^{-16}$	0.002378
Sex ~ BMI	1.04	$< 2 \times 10^{-16}$	N/A

PC1 ~ BMI	3.13×10^{-3}	1.99×10^{-8}	7.87×10^{-5}
PC2 ~ BMI	-0.001	0.0227	1.08×10^{-5}
PC3 ~ BMI	-0.002	4.8×10^{-5}	4.8×10^{-5}
PC4 ~ BMI	0.018	$< 2 \times 10^{-16}$	0.0008194
PC5 ~ BMI	0.041	$< 2 \times 10^{-16}$	0.0008179
PC6 ~ BMI	-0.001	0.00215	2.17×10^{-5}
PC7 ~ BMI	0.0005	0.434	-9.999×10^{-7}
PC8 ~ BMI	-0.001	0.0926	4.715×10^{-6}
PC9 ~ BMI	-0.009	1.74×10^{-8}	7.935×10^{-5}
PC10 ~ BMI	0.003	0.001207	2.445×10^{-5}
IMD ~ BMI	0.370	$< 2 \times 10^{-16}$	0.01604
Smoking Status ~ BMI	1.03	$< 2 \times 10^{-16}$	N/A
Diabetes ~ BMI	1.16	$< 2 \times 10^{-16}$	N/A
UKB Assessment Centre ~ BMI	1.05	0.000161	N/A

Table 2.4 – Collinearity between alanine aminotransferase, C-reactive protein, and triglyceride phenotypes. CRP = C-reactive protein, ALT = alanine aminotransferase, TG = triglyceride, β = beta regression coefficient.

Covariate ~ BMI **Association**	**β**	***P***
CRP ~ ALT	0.02	$< 2 \times 10^{-16}$
CRP ~ TG	0.32	$< 2 \times 10^{-16}$
ALT ~ TG	3.63	$< 2 \times 10^{-16}$