

https://dspace.org/

Evaluation of causal association and risk stratification of body mass index polygenic risk scores with cardiovascular risk factors in African populations

Mogongoa, Lebogang Francis

2025
University of the Witwatersrand, Johannesburg
https://hdl.handle.net/10539/49757

Mogongoa, Lebogang Francis. (2025). Evaluation of causal association and risk stratification of body mass index polygenic risk scores with cardiovascular risk factors in African populations [Master's dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/49757

Downloaded from a DSpace repository. https://dspace.org/



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Evaluation of causal association and risk stratification of body mass index polygenic risk scores with cardiovascular risk factors in African populations

By

Lebogang Francis Mogongoa
Student Number: 2485509

A research report submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements of the Master of Science (MSc) in Epidemiology and Biostatistics

Supervisors:
Prof Juliana Kagura
Dr Tinashe Chikowore

18 November 2025

DECLARATION

I, Lebogang Francis Mogongoa, declare that this research report is my original unaided work. This report is submitted in partial fulfilment of the requirements for the Master of Science in Epidemiology & Biostatistics degree at the University of the Witwatersrand, Johannesburg, South Africa. This research report has not been submitted before for any degree or examination at the University of the Witwatersrand or any other university.



_____

(Candidate Signature)

I declare this on the 18th of November 2025 in Bloemfontein, South Africa.

DEDICATION

I dedicate the thesis to my late mother, Selemeneng Leah Mogongoa, who was my rock, inspiration, and hero. I also extend my gratitude to my wife, father, siblings, and children for their unwavering support. Without your support, this would not be feasible. Most importantly, I thank the Gracious and Merciful Almighty for providing me with the strength and wisdom to complete the project.

ABSTRACT

Introduction

Cardiovascular diseases (CVDs) are a major cause of morbidity and mortality globally and in sub-Saharan Africa. Body mass index (BMI) is a crucial risk factor for non-communicable diseases, such as CVD. However, the genetic mechanisms by which it contributes to non-communicable diseases are poorly understood. Multivariant recall by genotype (RbG) studies, using polygenic risk scores (PRS) offer an opportunity to unravel mechanistic insights into how BMI causes CVD. However, a major limitation is the poor performance of Europe-designed PRS in the African population. The aim was to compare the causal association and risk stratification of BMI PRSs using the RbG design in Africa's East, South, and West regions.

Methods

This cross-sectional secondary data analysis study utilised the Africa Wits-INDEPTH Partnership for Genomic Research (AWI-Gen) data set to compare seven validated BMI PRS methodologies using the RbG approach to evaluate their utility for causal inferences in carotid intima-medial thickness (CIMT) and blood pressure (BP). The AWI-Gen data set comprised 12,000 participants aged 29 and above from East, West and South regions of Africa. Linear regression was used for predictivity and causal inference analysis. Confounding analyses were performed using the T-test and the Chi-square test. BMI was an outcome variable, and PRSs were exposure variables. Age, sex, education and socioeconomic status were included as confounders. The analyses were performed for the entire study population and per region.

Results

In 10,457 continental Africans, we utilised the RbG approach to evaluate the causal association of BMI with average CIMT and BP. In linear models, BMI was positively associated with CIMT (in AWI-Gen: Beta 0.0022 ($p<0.0001$)) and blood pressure (In AWI-Gen: Beta 0.5065 ($p<0.0001$) for diastolic BP; Beta 0.72667 ($p<0.0001$) for systolic BP) and African

regions. The predictivity of BMI PRSs was low, ranging from 0% for LDPred2_MA to 1.67% for PRSice_MAMA. In the RbG, the PRSs failed to control for confounding of all the covariates.

Conclusion

The BMI PRSs are not useful for stratification and causal inference through RbG due to their low predictivity in continental Africans. There is a need for genome-wide association studies (GWAS) in continental Africa, which could compute more predictive PRS that can be used in RbG approaches for causal inference.

Keywords: Body mass index, Blood pressure, Cardiovascular disease, Carotid intima-media thickness, Recall by genotype

ACKNOWLEDGEMENTS

- Praise be to Jesus, my redeemer. *'"Suppose one of you has a hundred sheep and loses one of them. Doesn't he leave the ninety-nine in the open country and go after the lost sheep until he finds it? And when he finds it, he joyfully puts it on his shoulders and goes home. Then he calls his friends and neighbours together and says, 'Rejoice with me; I have found my lost sheep.' I tell you that in the same way there will be more rejoicing in heaven over one sinner who repents than over ninety-nine righteous persons who do not need to repent.'* Luke 15: 4–7
- Heartfelt gratitute to my supervisors for their guidance and overwhelming support. Without your persistence and belief in me, this project would have failed. I especially applaud you for the sacrifices concerning your family time, I am eternally indebted to you.
- Thanks to the Sydney Brenner Institute for Molecular Bioscience (SBIMB) for granting data access.
- My wife, for your unwavering and humbling support and belief in me.
- My family and friends for their words of encouragement.
- My work supervisor for understanding and granting leave, where possible

TABLE OF CONTENTS

Table of Contents

DECLARATION I
DEDICATION II
ABSTRACT III
ACKNOWLEDGEMENTS V
TABLE OF CONTENTS VI
LIST OF FIGURES VIII
LIST OF TABLES IX
LIST OF ABBREVIATIONS/ NOMENCLATURE X

CHAPTER 1: INTRODUCTION 1

1.1 BACKGROUND 1
1.2 LITERATURE REVIEW 3
1.2.1 Recall-by-Genotype as a Study Method 3
1.2.2 Polygenic risk score computation 4
1.2.3 Cardiovascular risk factors 6
1.3 CAROTID INTIMA-MEDIA THICKNESS AS A SURROGATE MARKER FOR CVD 7
1.4 CONCEPTUAL FRAMEWORK 8
1.5 PROBLEM STATEMENT 9
1.6 JUSTIFICATION 10
1.7 RESEARCH QUESTION 10
1.7.1 Aim 10
1.7.2 Objectives 10
REFERENCES 12

CHAPTER 2: SUBMISSIBLE DRAFT MANUSCRIPT 21

ABSTRACT 22
INTRODUCTION 22
MATERIALS AND METHODS 24
RESULTS 28
DISCUSSION 42
ACKNOWLEDGEMENTS 43
CONFLICT OF INTEREST STATEMENT 43
FUNDING 44
DATA ACCESS 44

ETHICS AND CONSENT 44
AUTHORS' CONTRIBUTION 44
REFERENCES 45
SUPPLEMENTARY INFORMATION 50

CHAPTER 3: EXTENDED METHODS 59

3.1 EXTENDED METHODOLOGY 59
3.1.1 Study setting of the primary study 59
3.1.2 Study design of the secondary study 60
3.1.3 Power calculation 60
REFERENCES 61
APPENDICES 63
Appendix 1: Plagiarism declaration form 63
Appendix 2: Plagiarism/Turnitin cover page 64
Appendix 3: Evidence of article submission 65
Appendix 4: Journal authors' guidelines 66
Appendix 5: Data access permission letter 68
Appendix 6: Ethics clearance certificate 70

LIST OF FIGURES

Figure 1.1: Conceptual framework of the proposed relationship between BMI, hypertension, CIMT, and CVD. 8

Figure 1: BMI association with average CIMT in AWI-Gen, East, South and West African regions. 27

Figure 2: BMI association with diastolic and systolic blood pressure in AWI-Gen, East, South and West African regions. 28

Figure 3: BMI PRS Predictivity summary of the entire AWI-Gen study, East, South and West populations. 29

Figure 4: Comparison of the mean BMI of the bottom and top 10% deciles of the population as stratified PRSs within the AWI-Gen population and per region by ANOVA with Bonferroni correction. 30

Figure 5: Comparison of p-values of confounders (covariates) in BMI PGS organised deciles one and ten based on the RbG study design for the AWI-Gen and regional populations. 32

Figure 6: Forest plot causal association evaluation using BMI PRS deciles and average CIMT. 34

Figure 7: Forest plot causal association evaluation using BMI PRS deciles and diastolic blood pressure. 36

Figure 8: Forest plot causal association evaluation using BMI PRS deciles and systolic blood pressure. 38

Figure 3.1: Map indicating the six sites where participants were recruited for the AWI-Gen study. 56

LIST OF TABLES

Table 1: PRS methods development information of the seven utilised for the analysis 24

Table 2: Population descriptive statistics of the AWI-Gen data set used for analysis 26

Table S1: BMI association with CIMT, diastolic and systolic blood pressure in AWI-Gen and regionally 47

Table S2. Predictivity of BMI PRSs for AWI-Gen and regionally 48

Table S3a. Comparison of decile one (low-risk = bottom 10%) and ten (high-risk = top 10%) after BMI PRS stratification for AWI-Gen 48

Table S3b. Comparison of decile one (low-risk = bottom 10%) and ten (high-risk = top 10%) after BMI PRS stratification for the East region 49

Table S3c. Comparison of decile one (low-risk = bottom 10%) and ten (high-risk = top 10%) after BMI PRS stratification for the South region 49

Table S3d. Comparison of decile one (low-risk = bottom 10%) and ten (high-risk = top 10%) after BMI PRS stratification for the West region 50

Table S4: Confounding analysis for AWI-Gen and regions based on covariates 51

Table S5: RbG BMI PRS causal association with CIMT with covariates in the model 52

Table S6: RbG BMI PRS causal association with diastolic blood pressure with covariates in the model 53

Table S7: RbG BMI PRS causal association with systolic blood pressure with covariates in the model 55

Table 3.1: Power calculation results per study population. 57

LIST OF ABBREVIATIONS/ NOMENCLATURE

AWI-Gen	Africa Wits-INDEPTH partnership for genomic studies
BMI	Body mass index
BP	Blood pressure
CIMT	Carotid intima-media thickness
CVD	Cardiovascular disease
DALY	Disability-adjusted life-years
GWAS	Genome-wide association studies
LD	Linkage disequilibrium
PGS	Polygenic score
PRS	Polygenic risk scores
QC	Quality control
RbG	Recall-by-Genotype
SES	Socio-economic status
SNP	Single-nucleotide polymorphisms

Chapter 1: Introduction

1.1 Background

Cardiovascular disease encompasses conditions that are associated with fatty deposits in arteries and increased risk for blood coagulation, resulting in heart attack and stroke. CVD has been a leading cause of death globally over the past three decades, although there has been a decline in the number of cases (Naghavi et al., 2024). In 2021, CVD accounted for 32% of deaths globally, with 20.5 million deaths (Lindstrom et al., 2022; Naghavi et al., 2024). Of these deaths, heart attack and stroke accounted for 85% (WHO, 2024). Furthermore, approximately 4 out of 5 deaths occurred in low- and middle-income countries (WHO, 2024). The importance of addressing the CVD burden is underpinned by the United Nations Sustainable Development Goal Target 3.4. That requires a one-third reduction in premature death from non-communicable diseases by 2030 (Bennett et al., 2018). This CVD burden can be addressed by handling modifiable risk factors. Among these, obesity is an important risk factor. In 2015, more than two-thirds of 4 million obesity related deaths were caused by CVD (Afshin et al., 2017). Where CVD was the outcome and obesity the exposure.

CVD poses a threat to the well-being of the individual and the public through the cost of medical care. To investigate public health intervention strategies, the mechanistic insights into why obesity is uniquely more prevalent in women than men in Africa (Ramsay et al., 2018) and how it causes cardiovascular diseases must be investigated to enable the design of targeted strategies. One way of investigating causal inferences of multifaceted disorders like obesity on cardiovascular disease is by using multivariate recall-by-genotype study designs based on polygenic risk scores.

Polygenic risk scores (PRS) are the summation of risk for developing a particular trait or disease based on the aggregation of risk-associated alleles for that disease. PRS, a derivative of genome-wide association studies (GWAS), has opened an avenue for investigating the genetic contribution to non-communicable diseases or phenotypes and possible intervention

(Knowles & Ashley, 2018; Martin et al., 2018; McCarthy et al., 2008). One of the applications of PRS is in RbG studies, which involve sampling individuals from the tails of the distribution of the PRS associated with the trait of interest. This results in two groups of people, with high and low genetic susceptibility to the trait of interest and who have an equal distribution of confounding factors since this approach leverages Mendel's principle of independent assortment.

The RbG principle was illustrated by Corbin et al. (2018) where the top and bottom 30% were sampled by BMI (phenotype) and BMI PRS distributions, using the Avon Longitudinal Study of Parents and Children (ALSPAC). For the phenotype sampling, the confounding factors, alcohol, education and income, were significantly different for the people at the 30% tails of the distribution. However, this was contrary in the BMI PRS case, where confounders were observed to be insignificant in the recall strata. However, PRS performed very well in European, Asian, Oceanian and American settings due to the critical mass of GWAS, which is lacking in Africa (Chikowore et al., 2016; Kamiza et al., 2022; Majara et al., 2023). Nevertheless, low PRS predictivity and differential generalizability were observed with African populations across different regions.

The poor performance of PRS in Africans is related to the Eurocentric development of the scores, the variability of genetic frequencies, different linkage disequilibrium (LD) patterns, and environmental exposures. PRS are most accurate when applied to European ancestry, contrary to African ancestry, due to the critical mass of GWAS used to compute these PRSs in this population, leading to Eurocentric study bias (S. M. Purcell et al., 2009). However, there is a case for including the African population because of the genetic diversity observed among the African population (Auton et al., 2015; Campbell & Tishkoff, 2008a; Henn et al., 2012). Moreover, an improved resolution to fine-map causal variants is observed because of shorter blocks of linkage equilibrium among African populations (Campbell & Tishkoff, 2008a).

Several PRS methods have been developed to improve the portability of these scores in African populations. However, the impact of these methods in RbG studies to overcome confounding and their usefulness in causal inferences is yet to be determined. Our study set

out to evaluate the utility of selected PRSs computed through varied approaches in controlling confounding in the RbG study design in continental Africans.

1.2 Literature review

Disease aetiology has been investigated, especially in observational studies, using phenotypic expression, but a move toward the contribution of genetics is now being observed. These phenotypes have been utilised to stratify risk factors, understand disease aetiology and identify possible therapeutic interventions (Franks & Pomares-Millan, 2020). Although this approach is very effective, it is limited by confounding, thus, causality cannot be easily determined. However, genetic studies have created an avenue to study aetiology by stratifying based on genetic markers that are randomly attained at conception, independent of confounding factors. This type of study is summarised under RbG, which could be synonymous with random controlled trials (Purcell et al., 2009), with the advantage of eliminating confounding, reverse causation and therefore enabling causal inference. RbG can be based on a single or multiple gene associations (S. M. Purcell et al., 2009). In this literature review, we will outline RbG as a study design, especially in the realm of multiple gene studies, PRS computation methods, and risk factors for CVDs.

1.2.1 Recall-by-Genotype as a Study Method

Recall-by-Genotype is a design that can exploit genetic traits that have been associated with disease/risk factors identified by GWAS to enable an explanation of mechanisms related to genetic traits and to allow causal inference (Corbin et al., 2018). RbG can be applied to a single variant or multiple variant studies, where the former involves stratifying research participants based on the presence or absence of a genetic trait similar to case-control studies. Contrary to the latter, which will be the focus of this study, where the summation of associated

traits is utilised to compute a PRS. The stratification of research participants is achieved by utilising the subgroups of the extreme phenotypic expression (Corbin et al., 2018a). This approach enables the use of a limited number of participants with exact phenotypic expression while not being affected by confounding. Other advantages include enabling causal inference while maintaining statistical power. Although small participant numbers are utilised, a large pool of GWAS studies is required to select the population of interest, which is not found among African populations (Corbin et al., 2018a; Majara et al., 2023). This is caused by the cost associated with GWAS studies. However, this scenario is slowly changing as the associated cost is slowly decreasing, and more African researchers are involved in these genetic studies. Another limiting factor to the application of PRS to the African population is transferability from one population to the next, which is related to how PRS are calculated among others.

1.2.2 Polygenic risk score computation

Polygenic risk scores have variable predictivity across populations because of the population from which they were constructed, the type of genetic traits, and the method of calculation (Chikowore et al., 2016; Kamiza et al., 2022; Majara et al., 2023). BMI or other complex diseases are illuminated by a myriad of genetic variants that explain a small portion of their trait variation. Therefore, PRS utilises the combined effects of the variants to describe the variants of specific phenotypes. The computational applications of PRS and their underlying genetic traits bear on the interpretation of study results. Hence, evaluating different PRS may explain whether they can still eliminate confounding and be used for causal inference despite their variable and low predictive index.

GWAS provides information on numerous genetic variants contributing to the risk of complex diseases. Using the associated genetic variants, PRS, can calculate genetic disease susceptibility. The calculations are based on genetic variants' allele numbers, weighted by the variant's effect, estimated from GWAS data. PRS computation methods are derived from

GWAS summary statistics of single-discovery or multi-ancestry discovery samples, with LDpred2 (Privé et al., 2021) and PRS-CSx (Ruan et al., 2022) as respective examples.

The PRSs included the following calculation methods: LDpred2, PRSCSx, PRSice, clumping and thresholding. LDpred2 is a modification of LDpred1 (Vilhjálmsson et al., 2015), based on summary statistics and an LD matrix. The modification involved estimating the sparsity p, the SNP heritability h_2, and providing a sparse effects vector. Furthermore, PRS-CSx is an extension of PRS-CS, which integrates multi-ancestry GWAS summary statistics and uses a shared continuous shrinkage before coupling single-nucleotide polymorphism effects across different populations (Ruan et al., 2022). PRS-CS (Ge et al., 2019) utilises a Bayesian polygenic modelling and prediction framework. PRSice employs P-value thresholding (P + T) and linkage disequilibrium (LD) clumping. CT-SLEB is a multi-ancestry PRS method that entails three main steps (Choi & O'Reilly, 2019). Clumping and thresholding to select SNPs, the empirical-Bayes method to estimate the effect sizes for the SNPs, and super-learning to combine PRS under tuning parameters (H. Zhang et al., 2023a).

The seven selected PRSs in this study were multi-ancestral and used different calculation methods as depicted in the preceding paragraph. The method selection was to evaluate if the calculation methods, the number of variants, the source of variant association and the ancestry target of computation would make a difference as these were varied, see Table 1 (submissible article) for further details (Chikowore et al., 2024; Shim et al., 2023; Zhang et al., 2023) . The selected PRS were as follows: PGS003844 (CT_AFR), PGS003848 (CTSLEB_AFR), PGS003845 (LDPred2_AFR), PGS03848 (LDPred2_MA), PGS003847 (PRSCSx _AFR_ltd), PGS003884 (PRSCSx_His/Lat_AFR-EUR), and PGS004902 (PRSCice_MAMA). These PRSs were multi-ancestral and computed for the African population, as indicated by the African 100% in the computation column.

1.2.3 Cardiovascular risk factors

The risk factors for CVD can be summarised into modifiable categories, including behavioural, intermediate/metabolic, and environmental factors, and unmodifiable categories, including sex, age, and family history (genetics) (WHO, 2024; World Heart Federation, 2023). The behavioural risk factors include smoking, physical inactivity, a high-sodium diet, and high alcohol consumption. These behaviours lead to changes in the intermediate risk factors that can be assessed to estimate the risk of CVD development (WHO, 2024). These risk factors include obesity, measured by body mass index, hypertension, diabetes, and hypercholesterolaemia (World Heart Federation, 2023). The environmental risk factor is ambient air pollution. It is also essential not to overlook the underlying determinants of CVDs that play a significant role. These determinants include urbanisation, an ageing population, poverty, stress, hereditary factors, and globalisation (World Heart Federation, 2023).

1.2.3.1 Body mass index

Obesity, computed by body mass index (BMI), is a common global phenomenon that varies among different regions and is associated with mortality and morbidity of different non-communicable diseases. Obesity's importance is highlighted by its inclusion in the World Health Organization's target for global non-communicable diseases (Kontis et al., 2014). Globally in 2015, approximately 101.1 to 115.1 million children and 592.9 to 615.6 million adults, that is an overall prevalence of 5% and 12% among children and adults, respectively (Afshin et al., 2017). Obesity among adults was generally higher in females than in men, contrary to those under 20 years, where a sex difference was not observed. The peak obesity prevalence was 60 to 64 years for females and 50 to 54 years for males. The prevalence of obesity among males and females in East, South, and West Africa is 5-10% and 15-20%, 10-15% and >35%, and 5-10% and 5-20%, respectively (Afshin et al., 2017).

Besides sex, age and regional variation, high BMI is a risk factor for CVD (Singh et al., 2013), diabetes (Singh et al., 2013), kidney disease (Singh et al., 2013), cancers (Lauby-Secretan et al., 2016) and other disorders. In 2015, high BMI was responsible for 4 million deaths and 120

million disability-adjusted life-years (DALY) of any cause globally, which represents 7.1% and 4.9%, respectively(Afshin et al., 2017) . The leading cause of death and DALY related to high BMI was cardiovascular diseases, responsible for 2.7 million deaths and 66.3 million DALY globally. Obese persons accounted for 41% and 34% of BMI-related deaths and DALY, respectively (Afshin et al., 2017).

1.2.3.2 Hypertension

Hypertension is a preventable risk factor for premature death and a major risk factor for CVD globally. In 2019, approximately 10.8 million deaths were attributed to hypertension (Abbafati et al., 2020). Approximately 50% of all heart disease and stroke-related deaths are attributed to hypertension (WHO, 2013). From 1990 to 2015, there has been an increase in the number of people with hypertension (Forouzanfar et al., 2017). There are more than one billion people globally living with hypertension(Zhou et al., 2017), which accounts for more than 20% of CVD (Yusuf et al., 2020). Three-quarters of people with hypertension are found in low and medium-income countries (Zhou, Carrillo-Larco, et al., 2021; Zhou, Perel, et al., 2021). The prevalence of hypertension in women was 48%, and in males was 34%, in sub-Saharan Africa in 2019 (Zhou, Carrillo-Larco, et al., 2021). The major drivers for the increase include age, population growth, and behavioural factors like increased alcohol consumption, suboptimal nutrition and obesity (WHO, 2013).

1.3 Carotid intima-media thickness as a Surrogate Marker for CVD

Atherosclerosis is characterised by thickening of the arterial wall and is measured by a non-invasive procedure, CIMT. CIMT has been proposed as a surrogate marker for subclinical atherosclerosis (Touboul, Grobbee, et al., 2012; Willeit et al., 2020) and has been associated with CVD (Chambless et al., 2000; Elias-Smale et al., 2012; Kitamura et al., 2004; Rosvall et al., 2005). Furthermore, CIMT has been associated with CVD risk prediction in diabetes (Sibal

et al., 2011), chronic kidney disease (Lawal et al., 2019) and intervention evaluation (Willeit et al., 2020). The CIMT prediction risk can be improved by combining it with the carotid plaque formation assessment (Ho, 2016; Nambi et al., 2010) and other risk factors (Elias-Smale et al., 2012). The association with CVD is seen in different races (Kitamura et al., 2004; Pflederer et al., 2016), including Africans (Holland et al., 2009; Mohamed et al., 2022; Villines et al., 2017).

1.4 Conceptual framework

Cardiovascular diseases are precipitated by atherosclerotic plaque formation, leading to blood vessel narrowing and decreased tissue blood flow distal to the plaque (Ajoolabady et al., 2024). The development of CVDs depends on non-modifiable and modifiable risk factors (Lucas-Herald & Delles, 2022a). As depicted in Figure 1.1, obesity is influenced by genetic factors, age, hypertension and sex. In addition, obesity plays a role in increasing CIMT, and it can lead to CVD development. At the same time, hypertension can increase CIMT, which can lead to CVD development. Thus, we are looking into the association between BMI and BP and CIMT, using PRS as predictors.

The genetic contribution to increased CVD risk can be summarised using PRS to assess the impact of genetic background. This is especially important in the studied population, as hypertension is a significant risk factor in people of African descent (Olowoyo et al., 2025). Furthermore, obesity, which affects low- and middle-income countries, is an important factor to consider (WHO, 2024). Hypertension and BMI are directly linked to increased CIMT and can influence each other, thereby increasing the risk of CVD, as shown in Figure 1.1.

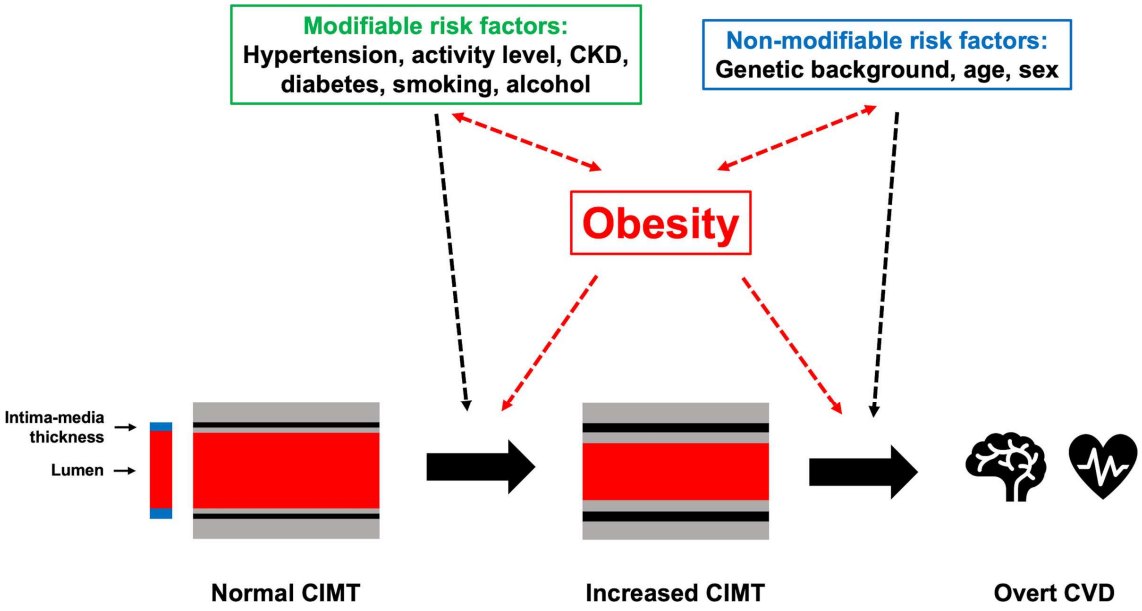


Figure 1.1: Conceptual framework of the proposed relationship between BMI, hypertension, CIMT, and CVD (Lucas-Herald & Delles, 2022b)

1.5 Problem statement

Polygenic risk scores have been noted to perform poorly in people of African descent and to vary in predictivity ability across different African regions. The multi-ancestry PRS prediction has been reported at a 4.5-fold (Martin et al., 2019) and 7.2-fold (0.93% vs 6.72%) lower rate in Africans compared to Europeans (Chikowore et al., 2024). Currently, several approaches are being developed and are promising to improve PRS predictivity. However, the optimum PRS approaches for RbG designs are yet to be clarified in Africa (Ruan et al., 2022). It is unclear how low and variable PRS predictivity impacts the RbG approach's ability to stratify individuals, control for confounding, and determine causal association in the African population.

1.6 Justification

The ability of these PRS to stratify and be implemented in causal inference studies may contribute to shedding mechanistic insights into how BMI causes cardiovascular diseases and thus lead to the discovery of more relevant and applicable interventions to address the CVD burden (WHO, 2024) in Africa. Furthermore, the stratification in Africans will shed light on the genetic contribution, especially due to the genetic diversity encountered (Campbell & Tishkoff, 2008a). Moreover, due to the burden of obesity (Afshin et al., 2017) and hypertension (Zhou, Carrillo-Larco, et al., 2021) on the healthcare system, such effective solutions can decrease morbidity and thus healthcare costs in the Sub-Saharan healthcare system, which is under pressure. The PRS could also be utilised to design genetic evidence-based interventions to advance precision medicine on the African continent.

1.7 Research question

Are PRSs derived from BMI causally associated with cardiovascular disease in East, West, and Southern Africa?

1.7.1 Aim

To compare the causal association and risk, stratification of BMI PRSs surrogate with cardiovascular markers (CIMT and BP) using the RbG design in Africa's East, South, and West regions

1.7.2 Objectives

1. To describe the population based on sex associated with cardiovascular risk markers and potential confounding factors in Africa's East, South, and West regions.
2. To evaluate the risk stratification of seven BMI PRS with cardiovascular markers using RbG design in Africa's East, South and West regions.

3. To evaluate the causal association of seven BMI PRS with cardiovascular markers using RbG design in Africa's East, South and West regions.

References

Abbafati, C., Abbas, K. M., Abbasi-Kangevari, M., Abd-Allah, F., Abdelalim, A., Abdollahi, M., Abdollahpour, I., Abegaz, K. H., Abolhassani, H., Aboyans, V., Abreu, L. G., Abrigo, M. R. M., Abualhasan, A., Abu-Raddad, L. J., Abushouk, A. I., Adabi, M., Adekanmbi, V., Adeoye, A. M., Adetokunboh, O. O., … Murray, C. J. L. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, *396*(10258), 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2

Afshin, A., MH, F., MB, R., P, S., K, E., A, L., L, M., AH, M., M, M.-L., M, N., JS, S., T, V., KH, A., C, A., MB, A., Z, A.-A., A, A., R, A.-R., AT, A., … CJL, M. (2017). Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *The New England Journal of Medicine*, *377*(1), 13–27. https://doi.org/10.1056/NEJMOA1614362

Ajoolabady, A., Pratico, D., Lin, L., Mantzoros, C. S., Bahijri, S., Tuomilehto, J., & Ren, J. (2024). Inflammation in atherosclerosis: pathophysiology and mechanisms. *Cell Death & Disease*, *15*(11), 817. https://doi.org/10.1038/S41419-024-07166-8,

Ali, S. A., Soo, C., Agongo, G., Alberts, M., Amenga-Etego, L., Boua, R. P., Choudhury, A., Crowther, N. J., Depuur, C., Gómez-Olivé, F. X., Guiraud, I., Haregu, T. N., Hazelhurst, S., Kahn, K., Khayeka-Wandabwa, C., Kyobutungi, C., Lombard, Z., Mashinya, F., Micklesfield, L., … Ramsay, M. (2018). Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Global Health Action*, *11*(sup2). https://doi.org/10.1080/16549716.2018.1507133

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., …

Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/NATURE15393

Bennett, J. E., Stevens, G. A., Mathers, C. D., Bonita, R., Rehm, J., Kruk, M. E., Riley, L. M., Dain, K., Kengne, A. P., Chalkidou, K., Beagley, J., Kishore, S. P., Chen, W., Saxena, S., Bettcher, D. W., Grove, J. T., Beaglehole, R., & Ezzati, M. (2018). NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. *Lancet (London, England)*, *392*(10152), 1072–1088. https://doi.org/10.1016/S0140-6736(18)31992-5

Campbell, M. C., & Tishkoff, S. A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics*, *9*, 403–433. https://doi.org/10.1146/ANNUREV.GENOM.9.081307.164258

Cardiovascular diseases (CVDs). (n.d.). Retrieved April 14, 2025, from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

Chambless, L. E., Folsom, A. R., Clegg, L. X., Sharrett, A. R., Shahar, E., Nieto, F. J., Rosamond, W. D., & Evans, G. (2000). Carotid wall thickness is predictive of incident clinical stroke: The Atherosclerosis Risk in Communities (ARIC) study. *American Journal of Epidemiology*, *151*(5), 478–487. https://doi.org/10.1093/OXFORDJOURNALS.AJE.A010233,

Chikowore, T., van Zyl, T., Feskens, E. J. M., & Conradie, K. R. (2016). Predictive utility of a genetic risk score of common variants associated with type 2 diabetes in a black South African population. *Diabetes Research and Clinical Practice*, *122*, 1–8. https://doi.org/10.1016/J.DIABRES.2016.09.019

Choi, S. W., & O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*, *8*(7). https://doi.org/10.1093/GIGASCIENCE/GIZ082,

Corbin, L. J., Tan, V. Y., Hughes, D. A., Wade, K. H., Paul, D. S., Tansey, K. E., Butcher, F., Dudbridge, F., Howson, J. M., Jallow, M. W., John, C., Kingston, N., Lindgren, C. M., O'Donavan, M., O'Rahilly, S., Owen, M. J., Palmer, C. N. A., Pearson, E. R., Scott, R. A.,

… Timpson, N. J. (2018). Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference. *Nature Communications*, *9*(1), 711. https://doi.org/10.1038/S41467-018-03109-Y

Elias-Smale, S. E., Kavousi, M., Verwoert, G. C., Koller, M. T., Steyerberg, E. W., Mattace-Raso, F. U. S., Hofman, A., Hoeks, A. P. G., Reneman, R. S., & Witteman, J. C. M. (2012). Common carotid intima-media thickness in cardiovascular risk stratification of older people: The Rotterdam Study. *European Journal of Preventive Cardiology*, *19*(4), 698–705. https://doi.org/10.1177/1741826711414623,

Forouzanfar, M. H., Liu, P., Roth, G. A., Ng, M., Biryukov, S., Marczak, L., Alexander, L., Estep, K., Abate, K. H., Akinyemiju, T. F., Ali, R., Alvis-Guzman, N., Azzopardi, P., Banerjee, A., Bärnighausen, T., Basu, A., Bekele, T., Bennett, D. A., Biadgilign, S., … Murray, C. J. L. (2017). Global burden of hypertension and systolic blood pressure of at least 110 to 115mmHg, 1990-2015. *JAMA - Journal of the American Medical Association*, *317*(2), 165–182. https://doi.org/10.1001/JAMA.2016.19043,

Franks, P. W., & Pomares-Millan, H. (2020). Next-generation epidemiology: the role of high-resolution molecular phenotyping in diabetes research. *Diabetologia*, *63*(12), 2521. https://doi.org/10.1007/S00125-020-05246-W

Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A., & Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications*, *10*(1). https://doi.org/10.1038/S41467-019-09718-5

Henn, B. M., Cavalli-Sforza, L. L., & Feldman, M. W. (2012). The great human expansion. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(44), 17758–17764. https://doi.org/10.1073/PNAS.1212380109

Ho, S. S. Y. (2016). Current status of carotid ultrasound in atherosclerosis. *Quantitative Imaging in Medicine and Surgery*, *6*(3), 285–296. https://doi.org/10.21037/QIMS.2016.05.03,

Holland, Z., Ntyintyane, L. M., Gill, G. V., & Raal, F. J. (2009). Carotid intima–media thickness is a predictor of coronary artery disease in South African black patients. *Cardiovascular*

Journal of Africa, *20*(4), 237. https://pmc.ncbi.nlm.nih.gov/articles/PMC3721310/

Kamiza, A. B., Toure, S. M., Vujkovic, M., Machipisa, T., Soremekun, O. S., Kintu, C., Corpas, M., Pirie, F., Young, E., Gill, D., Sandhu, M. S., Kaleebu, P., Nyirenda, M., Motala, A. A., Chikowore, T., & Fatumo, S. (2022). Transferability of genetic risk scores in African populations. *Nature Medicine*, *28*(6), 1163–1166. https://doi.org/10.1038/S41591-022-01835-X

Kitamura, A., Iso, H., Imano, H., Ohira, T., Okada, T., Sato, S., Kiyama, M., Tanigawa, T., Yamagishi, K., & Shimamoto, T. (2004). Carotid intima-media thickness and plaque characteristics as a risk factor for stroke in Japanese elderly men. *Stroke*, *35*(12), 2788–2794. https://doi.org/10.1161/01.STR.0000147723.52033.9E,

Knowles, J. W., & Ashley, E. A. (2018). Cardiovascular disease: The rise of the genetic risk score. *PLoS Medicine*, *15*(3). https://doi.org/10.1371/JOURNAL.PMED.1002546

Kontis, V., Mathers, C. D., Rehm, J., Stevens, G. A., Shield, K. D., Bonita, R., Riley, L. M., Poznyak, V., Beaglehole, R., & Ezzati, M. (2014). Contribution of six risk factors to achieving the 25×25 non-communicable disease mortality reduction target: a modelling study. *Lancet (London, England)*, *384*(9941), 427–437. https://doi.org/10.1016/S0140-6736(14)60616-4

Lauby-Secretan, B., Scoccianti, C., Loomis, D., Grosse, Y., Bianchini, F., & Straif, K. (2016). Body Fatness and Cancer--Viewpoint of the IARC Working Group. *The New England Journal of Medicine*, *375*(8), 794–798. https://doi.org/10.1056/NEJMSR1606602

Lawal, O. M., Balogun, M. O., Akintomide, A. O., Ayoola, O. O., Mene-Afejuku, T. O., Ogunlade, O., Okunola, O. O., Lawal, A. O., & Akinsola, A. (2019). Carotid Intima-Media Thickness: A Surrogate Marker for Cardiovascular Disease in Chronic Kidney Disease Patients. *Clinical Medicine Insights: Cardiology*, *13*. https://doi.org/10.1177/1179546819852941,

Lindstrom, M., DeCleene, N. , Dorsey, H. , F. V. , J. C. O. , L. K. E. , M. G. A. , R. C. , S. B. , Varieur Turco, J. , & Roth, G. A. (2022). Global Burden of Cardiovascular Diseases and Risks Collaboration, 1990-2021. *Journal of the American College of Cardiology*, *80*(25),

2372–2425. https://doi.org/10.1016/J.JACC.2022.11.001

Lucas-Herald, A. K., & Delles, C. (2022a). Carotid Intima-Media Thickness Is Associated With Obesity and Hypertension in Young People. *Hypertension*, *79*(6), 1177–1179. https://doi.org/10.1161/HYPERTENSIONAHA.122.19163,

Lucas-Herald, A. K., & Delles, C. (2022b). Carotid Intima-Media Thickness Is Associated With Obesity and Hypertension in Young People. *Hypertension*, *79*(6), 1177–1179. https://doi.org/10.1161/HYPERTENSIONAHA.122.19163/ASSET/2D9A249F-C887-47F4-9A67-2B744CF22AA9/ASSETS/GRAPHIC/HYPERTENSIONAHA.122.19163.FIG01.JPG

Majara, L., Kalungi, A., Koen, N., Tsuo, K., Wang, Y., Gupta, R., Nkambule, L. L., Zar, H., Stein, D. J., Kinyanda, E., Atkinson, E. G., & Martin, A. R. (2023). Low and differential polygenic score generalizability among African populations due largely to genetic diversity. *Human Genetics and Genomics Advances*, *4*(2), 100184. https://doi.org/10.1016/J.XHGG.2023.100184

Martin, A. R., Daly, M. J., Robinson, E. B., Hyman, S. E., & Neale, B. M. (2018). Predicting polygenic risk of psychiatric disorders. *Biological Psychiatry*, *86*(2), 97. https://doi.org/10.1016/J.BIOPSYCH.2018.12.015

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews. Genetics*, *9*(5), 356–369. https://doi.org/10.1038/NRG2344

Mohamed, S. F., Khayeka-Wandabwa, C., Muthuri, S., Ngomi, N. N., Kyobutungi, C., & Haregu, T. N. (2022). Carotid intima media thickness (CIMT) in adults in the AWI-Gen Nairobi site study: Profiles and predictors. *Hipertension y Riesgo Vascular*, *40*(1), 5. https://doi.org/10.1016/J.HIPERT.2022.08.001

Naghavi, M., Ong, K. L., Aali, A., Ababneh, H. S., Abate, Y. H., Abbafati, C., Abbasgholizadeh, R., Abbasian, M., Abbasi-Kangevari, M., Abbastabar, H., Abd ElHafeez, S., Abdelmasseh, M., Abd-Elsalam, S., Abdelwahab, A., Abdollahi, M., Abdollahifar, M. A.,

Abdoun, M., Abdulah, D. M., Abdullahi, A., … Murray, C. J. L. (2024). Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet (London, England)*, *403*(10440), 2100. https://doi.org/10.1016/S0140-6736(24)00367-2

Nambi, V., Chambless, L., Folsom, A. R., He, M., Hu, Y., Mosley, T., Volcik, K., Boerwinkle, E., & Ballantyne, C. M. (2010). Carotid Intima-Media Thickness and Presence or Absence of Plaque Improves Prediction of Coronary Heart Disease Risk. The ARIC (Atherosclerosis Risk In Communities) Study. *Journal of the American College of Cardiology*, *55*(15), 1600–1607. https://doi.org/10.1016/j.jacc.2009.11.075

Pflederer, M. C., Long, C. S., Beaty, B., Havranek, E. P., Mehler, P. S., Keniston, A., & Krantz, M. J. (2016). Longitudinal Changes in Vascular Risk Markers and Mortality Rates among a Latino Population with Hypertension. *Texas Heart Institute Journal*, *43*(2), 131. https://doi.org/10.14503/THIJ-15-5053

Privé, F., Arbel, J., & Vilhjálmsson, B. J. (2021). LDpred2: better, faster, stronger. *Bioinformatics (Oxford, England)*, *36*(22–23), 5424–5431. https://doi.org/10.1093/BIOINFORMATICS/BTAA1029

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., Ruderfer, D. M., McQuillin, A., Morris, D. W., Oğdushlaine, C. T., Corvin, A., Holmans, P. A., Oğdonovan, M. C., MacGregor, S., Gurling, H., Blackwood, D. H. R., Craddock, N. J., Gill, M., Hultman, C. M., … Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(7256), 748–752. https://doi.org/10.1038/NATURE08185

Ramsay, M., Crowther, N. J., Agongo, G., Ali, S. A., Asiki, G., Boua, R. P., Gómez-Olivé, F. X., Kahn, K., Khayeka-Wandabwa, C., Mashinya, F., Micklesfield, L., Mukomana, F., Nonterah, E. A., Soo, C., Sorgho, H., Wade, A. N., Wagner, R. G., Alberts, M., Hazelhurst, S., … Tollman, S. (2018). Regional and sex-specific variation in BMI distribution in four sub-Saharan African countries: The H3Africa AWI-Gen study. *Global Health Action*,

11(sup2). https://doi.org/10.1080/16549716.2018.1556561

Ramsay, M., Crowther, N., Tambo, E., Agongo, G., Baloyi, V., Dikotope, S., Gómez-Olivé, X., Jaff, N., Sorgho, H., Wagner, R., Khayeka-Wandabwa, C., Choudhury, A., Hazelhurst, S., Kahn, K., Lombard, Z., Mukomana, F., Soo, C., Soodyall, H., Wade, A., … Sankoh, O. (2016). H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Global Health, Epidemiology and Genomics*, *1*, e20. https://doi.org/10.1017/GHEG.2016.17

Rosvall, M., Janzon, L., Berglund, G., Engström, G., & Hedblad, B. (2005). Incident coronary events and case fatality in relation to common carotid intima-media thickness. *Journal of Internal Medicine*, *257*(5), 430–437. https://doi.org/10.1111/J.1365-2796.2005.01485.X,

Ruan, Y., Lin, Y. F., Feng, Y. C. A., Chen, C. Y., Lam, M., Guo, Z., Ahn, Y. M., Akiyama, K., Arai, M., Baek, J. H., Chen, W. J., Chung, Y. C., Feng, G., Fujii, K., Glatt, S. J., Ha, K., Hattori, K., Higuchi, T., Hishimoto, A., … Ge, T. (2022). Improving polygenic prediction in ancestrally diverse populations. *Nature Genetics*, *54*(5), 573–580. https://doi.org/10.1038/S41588-022-01054-7

Sibal, L., Agarwal, S. C., & Home, P. D. (2011). Carotid intima-media thickness as a surrogate marker of cardiovascular disease in diabetes. *Diabetes, Metabolic Syndrome and Obesity*, *4*, 23–24. https://doi.org/10.2147/DMSO.S8540,

Singh, G. M., Danaei, G., Farzadfar, F., Stevens, G. A., Woodward, M., Wormser, D., Kaptoge, S., Whitlock, G., Qiao, Q., Lewington, S., Di Angelantonio, E., Vander Hoorn, S., Lawes, C. M. M., Ali, M. K., Mozaffarian, D., Ezzati, M., Jørgensen, T., Borch-Johnson, K., Nissinen, A., … Danesh, J. (2013). The age-specific quantitative effects of metabolic risk factors on cardiovascular diseases and diabetes: a pooled analysis. *PloS One*, *8*(7). https://doi.org/10.1371/JOURNAL.PONE.0065174

Touboul, P. J., Grobbee, D. E., & Den Ruijter, H. (2012). Assessment of subclinical atherosclerosis by carotid intima media thickness: Technical issues. *European Journal of Preventive Cardiology*, *19*(2 Suppl), 18–24. https://doi.org/10.1177/2047487312448990,

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P. R., Bhatia, G., Do, R., Hayeck, T., Won, H. H., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K. H., Holmans, P. A., Lee, P., Bulik-Sullivan, B., … Price, A. L. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American Journal of Human Genetics*, *97*(4), 576–592. https://doi.org/10.1016/J.AJHG.2015.09.001

Villines, T. C., Hsu, L. L., Blackshear, C., Nelson, C. R., & Griswold, M. (2017). Relation of Carotid Intima-Media Thickness to Cardiovascular Events in Black Americans (From the Jackson Heart Study). *American Journal of Cardiology*, *120*(9), 1528–1532. https://doi.org/10.1016/J.AMJCARD.2017.07.046,

Willeit, P., Tschiderer, L., Allara, E., Reuber, K., Seekircher, L., Gao, L., Liao, X., Lonn, E., Gerstein, H. C., Yusuf, S., Brouwers, F. P., Asselbergs, F. W., Van Gilst, W., Anderssen, S. A., Grobbee, D. E., Kastelein, J. J. P., Visseren, F. L. J., Ntaios, G., Hatzitolios, A. I., … Lorenz, M. W. (2020). Carotid Intima-Media Thickness Progression as Surrogate Marker for Cardiovascular Risk: Meta-Analysis of 119 Clinical Trials Involving 100 667 Patients. *Circulation*, *142*(7), 621–642. https://doi.org/10.1161/CIRCULATIONAHA.120.046361

WHO (2023). *A global brief on hypertension: silent killer, global public health crisis: World Health Day 2023*. Retrieved April 25, 2025, from https://www.who.int/publications/i/item/a-global-brief-on-hypertension-silent-killer-global-public-health-crisis-world-health-day-2013

World Heart Federation (2023): *World Heart report 2023: Full Report - World Heart Federation*. Retrieved April 15, 2025, from https://world-heart-federation.org/resource/world-heart-report-2023/

Yusuf, S., Joseph, P., Rangarajan, S., Islam, S., Mente, A., Hystad, P., Brauer, M., Kutty, V. R., Gupta, R., Wielgosz, A., AlHabib, K. F., Dans, A., Lopez-Jaramillo, P., Avezum, A., Lanas, F., Oguz, A., Kruger, I. M., Diaz, R., Yusoff, K., … Dagenais, G. (2020). Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-

income, middle-income, and low-income countries (PURE): a prospective cohort study. *The Lancet*, *395*(10226), 795–808. https://doi.org/10.1016/S0140-6736(19)32008-2

Zhang, H., Zhan, J., Jin, J., Zhang, J., Lu, W., Zhao, R., Ahearn, T. U., Yu, Z., O'Connell, J., Jiang, Y., Chen, T., Okuhara, D., Aslibekyan, S., Auton, A., Babalola, E., Bell, R. K., Bielenberg, J., Bryc, K., Bullis, E., … Chatterjee, N. (2023). A new method for multiancestry polygenic prediction improves performance across diverse populations. *Nature Genetics*, *55*(10), 1757. https://doi.org/10.1038/S41588-023-01501-Z

Zhou, B., Bentham, J., Di Cesare, M., Bixby, H., Danaei, G., Cowan, M. J., Paciorek, C. J., Singh, G., Hajifathalian, K., Bennett, J. E., Taddei, C., Bilano, V., Carrillo-Larco, R. M., Djalalinia, S., Khatibzadeh, S., Lugero, C., Peykari, N., Zhang, W. Z., Lu, Y., … Eggertsen, R. (2017). Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19·1 million participants. *The Lancet*, *389*(10064), 37–55. https://doi.org/10.1016/S0140-6736(16)31919-5

Zhou, B., Carrillo-Larco, R. M., Danaei, G., Riley, L. M., Paciorek, C. J., Stevens, G. A., Gregg, E. W., Bennett, J. E., Solomon, B., Singleton, R. K., Sophiea, M. K., Iurilli, M. L. C., Lhoste, V. P. F., Cowan, M. J., Savin, S., Woodward, M., Balanova, Y., Cifkova, R., Damasceno, A., … Zuñiga Cisneros, J. (2021). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *The Lancet*, *398*(10304), 957–980. https://doi.org/10.1016/S0140-6736(21)01330-1

Zhou, B., Perel, P., Mensah, G. A., & Ezzati, M. (2021). Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nature Reviews Cardiology*, *18*(11), 785–802. https://doi.org/10.1038/S41569-021-00559-8,

Chapter 2: SUBMISSIBLE DRAFT MANUSCRIPT

Utility of polygenic risk scores in evaluating the causal association of body mass index and cardiovascular risk factors in African populations

Short running title: Body mass index polygenic risk scores causal utility in Africans

Lebogang Francis Mogongoa [1,2], Juliana Kagura [1*], Michele Ramsay[3], Tinashe Chikowore [3,4,5*]

[1] Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[2] Department of Health Sciences, Central University of Technology, Free State, Bloemfontein, South Africa

[3] Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[4] Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States of America

[5] Harvard Medical School, Boston, Massachusetts, United States of America

*Joint last authors

Corresponding authors :LF Mogongoa 2485509@students.wits.ac.za and Tinashe Chikowore: tinashe.chikowore@channing.harvard.edu

Abstract

Cardiovascular disease (CVD) is the foremost cause of death worldwide. Epidemiological studies prone to bias have indicated body mass index (BMI) as a key risk factor. Recall by genotype (RbG) studies that use polygenic risk scores have been used to evaluate the causal association of BMI and CVD outcomes in Europeans. However, this approach is yet to be evaluated in continental Africans.

In 10,457 continental Africans, we used the RbG approach to evaluate the causal association of BMI with average carotid intima-media thickness (CIMT) and blood pressure (BP). We also used seven PRS for BMI computed through multiple approaches from the PGS catalog.

In linear models, BMI was positively associated with CIMT (In AWI-Gen: Beta 0.0022 ($p<0.0001$)) and BP (In AWI-Gen: Beta 0.5065 ($p<0.0001$) for diastolic BP, Beta 0.72667 ($p<0.0001$) for systolic BP) and African regions. The predictivity of BMI PRSs was low, ranging from 0 to 1.67%. In the RbG, the PRSs failed to fully control for confounding of the covariates.

The BMI PRSs are not yet useful for causal inference through RbG due to their low predictivity in continental Africans. There is a need for GWAS studies in continental Africa, which can be used to compute more predictive PRS that can be implemented in RbG approaches for causal inference.

Keywords: Body mass index, Blood pressure, Cardiovascular disease, Carotid intima-media thickness, Recall by genotype

Introduction

Cardiovascular diseases are among the leading causes of morbidity and mortality globally and in Sub-Saharan Africa (SSA). In 2021, 205 million deaths occurred, and approximately 80%

of the deaths occurred in low- and middle-income countries (Lindstrom et al., 2022). The health systems in SSA are stretched, and poverty levels are high, further compounding the problem. Therefore, there is an urgent need to address this situation in SSA. Understanding the causal risk factors for CVD will help inform evidence-based medicine interventions to curb this burden.

Epidemiology studies have discovered that risk factors for CVD include high blood pressure, obesity, family history of CVD, inactivity, smoking, diabetes, age, gender, and others (Bays et al., 2021). However, observational studies are biased by confounding. This challenge can be overcome by conducting randomised controlled trials (RCTs), as the randomisation of participants minimises selection bias and enhances causal inference. However, RCTs are not plausible to conduct in certain instances, such as when it is not ethical to expose people to a particular risk factor to determine its causal association with CVD (Saldanha et al., 2022).

Recently, genetic studies have been employed to evaluate causal inferences, using genetic variants that are assigned randomly at birth, based on Mendel's laws of inheritance (Corbin et al., 2018b; Wade et al., 2018). Current approaches include Mendelian randomisation (MR) and recall by genotype (RbG). MR approaches have been extensively evaluated, and the utility of RbG in continental Africans remains to be explored (Corbin et al., 2018b). RbG studies involve sampling individuals from the tails of the normal distribution of the polygenic risk score (PRS) associated with the trait of interest. This results in two groups of people with high and low genetic susceptibility to the trait of interest, and who have an equal distribution of confounding factors since this approach leverages Mendel’s principle of the random assortment. This approach achieves randomisation and control for confounding, like RCT (Corbin et al., 2018b). We set out to evaluate how PRS computed through multiple approaches, leveraging the RbG approach, can be used to evaluate the causal association of BMI with CIMT and BP in continental Africans.

Materials and methods

Study design and population

Our study was a cross-sectional secondary data analysis study in the Africa Wits-INDEPTH partnership for genomic studies (AWI-Gen). AWI-Gen study was a population-based longitudinal study that involved six study sites, which represented East Africa (Nairobi - Kenya), South Africa (Agincourt, Dikgale and Soweto) and West Africa (Nanoro - Burkina Faso and Navrongo - Ghana) (Ali et al., 2018; Ramsay et al., 2018). Phenotypic and genotypic data to assess risk factors and endpoint cardiometabolic disease in East, South and West African regions were undertaken in 10,457, aged 29 years and older (Ramsay et al., 2018).

Genetic data

Genomic DNA was isolated and purified from nucleated buffy coat cells using automated and manual methods as described by Ali et al. (2018). Three methods were used for the quality control of the genomic DNA.

Quality Control:

Genotype data from about 11,000 samples were initially generated on the 2.3 million single-nucleotide polymorphisms (SNPs) H3Africa array (https://chipinfo.h3abionet.org). The data were subjected to QC using the H3ABioNet/H3Agwas QC pipeline (https://github.com/h3abionet/h3agwas/tree/master/qc) (Baichoo et al., 2018; Brandenburg et al., 2022), resulting in a cleaned dataset of 10,903 samples and 1.7 million SNPs.

Genotype Imputation:

Imputation was performed on the cleaned dataset with 1,729,661 SNPs and 10,903 remaining after quality control. To enhance genomic coverage for fine-mapping, imputation was performed using the EAGLE2 (Loh et al., 2016) + PBWT pipeline at the Sanger Imputation Server with the African Genome Resources reference panel (Wang et al., 2024). Post-imputation filtering (removing indels and rare/low-quality SNPs) yielded a final set of 13,976,041 SNPs (MAF> 0.01 and info score > 0.6) (Choudhury et al., 2022).

Final GWAS Cohort:

The final GWAS was conducted on the GRCh37/hg19 assembly, including 10,775 participants who possessed reliable phenotype and genotype data.

Assessment of cardiovascular outcomes (BP and CIMT), exposure variable (BMI) and covariate (SES)

A dual B-mode ultrasound was used to determine CIMT, as described by Ali et al. (2018). The average left and right CIMT was used to compute the average CIMT used in this study. Mannheim Consensus was used for CIMT data interpretation and quality control (Touboul, Hennerici, et al., 2012).

Trained health care providers used a digital sphygmomanometer to measure BP. A standardised method was used, and participants were advised about precautionary measures to ensure accuracy (Ali et al., 2018). Duplicate BP readings were obtained after a five-minute rest period, and an average was reported.

For BMI assessment, a wall-mounted Harpenden digital stadiometer (Holtain, Wales) and digital Physician Large Dial 200 kg capacity scales (Kendon Medical, South Africa) were used to determine the height and weight of participants, respectively (Ali et al., 2018). Precautionary measures were undertaken to ensure accurate measurements, like heels touching the wall, removing heavy clothing and shoes. BMI was calculated by dividing the weight in grams by height in meters squared.

Socioeconomic status (SES) was determined by counting the household appliances, indirectly estimating the income (Ali et al., 2018). Thus, it is expressed as a continuous variable.

Computation of Polygenic Risk Scores

The seven pre-validated BMI PRS utilised for comparison were obtained from the PRS catalog (https://www.pgscatalog.org/search/?q=bmi). These PRS scores were applied to the AWI-Genstudy data using PLINK 2 software (Purcell et al., 2007). Table 1 depicts information related to the selected PRS methodologies and their development.

Table 1: PRS methods development information of the seven utilised for the analysis

Polygenic risk score	Calculation Parameters	Number of variants	Source of variant Association (GWAS)	Ancestry of the target dataset during computation.
PGS003844 (H. Zhang et al., 2023b) (**CT_AFR_87**)	"clumping window size=500kb, clumping r^2=0.1, p-value threshold = c(5E-08,5E-07,...,0.5,1)"	87	African 100%	African American or Afro-Caribbean
PGS003845 (H. Zhang et al., 2023b) (**LDpred2_AFR)**	"causal SNPs portions vary with length 17 that are evened spaced on a logarithmic scale from 1E-04 to 1" *	752 209	African: 100%	African: 100%
PGS003846 (H. Zhang et al., 2023b) (**LDpred2_MA)**	"Linear weight combination of the best single ancestry LDpred2" *	817 198	Multi-ancestry (including European): 100%	African: 100%
PGS003847 (H. Zhang et al., 2023b) (**PRSCSx_AFR_ltd)**	"gamma-gamma prior hyperparameters a and b to default values of 1 and 0.5, respectively. Parameter phi is varied over the default set of values 1E-06, 1E-04), 1E-02), and 1" *	300 864	Multi-ancestry (including European): 100%	African: 100%
PGS003848 (H. Zhang et al., 2023b) (**CTSLEB_AFR)**	"w_b= (50kb, 100kb), clumping r^2=(0.01, 0.05, 0.1, 0.2, 0.5, 0.8), clumping window size = w_b/r2, p-value threshold = c(5E-08,5E-07,...,0.5,1), the super leaner method was set to be glmnet and ridge to obtain linear weight" *	741 732	Multi-ancestry (including European): 100%	African: 100%
PGS003884 (Shim et al., 2023) (**PRScsx_His/Lat_AFR_EUR)**	"a=1, b=0.5, phi=auto"	1 067 771	"Hispanic or Latin American: 44.5% African: 34.7% Multi-ancestry (excluding European): 9.4% Additional Diverse Ancestries: 9.3% Not Reported: 2.1%" *	Greater Middle Eastern: 100%
PGS004902 (Chikowore et al., 2024) (**PRSice_MAMA)**	"LD r^2=0.8, clumping distance=250kb; p value threshold = 0.024" *	322 639	"East Asian: 71.4% African: 14.1% Hispanic or Latin American: 9.9% Multi-ancestry (excluding European): 2.1% Additional Diverse Ancestries: 2.1% Not Reported: 0.5%" *	African: 100%

***** Methodology data in the table is obtained from (https://www.pgscatalog.org/search/?q=bmi)

Statistical Analysis

Descriptive statistics

The outcome variables (CIMT and BP), exposure variable (BMI) and covariates (age, socioeconomic status (SES), education level, and African regions) were stratified by sex. These included comparisons using a t-test for continuous variables (age, BP, BMI, CIMT, SES) by sex, and a chi-square test for categorical variables (education level, region). Statistical analyses were performed using Stata ®18.

Linear models for outcome association and PRS predictivity

Linear regression models were conducted with either CIMT or BP as an outcome variable and BMI as an exposure, while adjusting for covariates (age, education, SES, sex, and region) included in the model. For PRS predictivity evaluation, linear regression models with BMI as an outcome and the covariates (age, sex, education, SES and genetic principal components (PC) 1 to 5) as predictors formed the null model. The five PC were utilised as the addition of PC 6 to 10 did not add value to the observed variation. For the full model, a linear regression for BMI as an outcome and the covariates and each PGS as the predictor. Predictivity of the PRS was calculated as adjusted residual incremental variance = R^2 (Full model) – adjusted R^2 (Null model).

Confounding evaluation using recall by genotype

We assessed whether confounding was accounted for when the selected PRS were utilised in the following RbG approach. The data was divided into deciles based on BMI PRS, and the top and bottom 10% of BMI PRS. This was followed by comparing whether the confounders (covariates) in the RbG strata (bottom 10% of the PRS vs top 10%) differed significantly among the seven BMI PRS approaches. The 10% threshold was selected because the sample size was large, and to ensure the furthest population within the distribution will be included. This was done using the T-test for continuous variables and the Chi-square test for categorical variables.

Causal Inference Analysis

Linear regressions with CIMT and BP as outcome variables, and BMI as exposure variables, while correcting for covariates (PRS, age, SES, education), were used to evaluate the associations of the RbG category of each of the seven PRS. The data was summarised employing Forest plots, using the beta coefficient and confidence intervals. The linear regression assumptions results are presented in the supplementary table (Table S8).

Results

Descriptive statistics

The AWI-Gen data set used for analysis consisted of 10,457 participants (45% males) with 1,767, 4,823 and 3,867 participants from the East, South and West regions, respectively. Age was not statistically different between males and females, as depicted in Table 2. The mean average CIMT, SES, diastolic and systolic BP were statistically significantly higher in males, while females had a higher mean BMI. More males were found in the higher education levels than females.

Table 2: Population descriptive statistics of the AWI-Gen dataset used for analysis

Variables	Men n median (IQR) n [%]	Women N median (IQR) n [%]	p-value
Age	4 713 50.0 (45.0 – 56.0)	5 774 51.0 (45.0 – 56.0)	0.1089
BMI	4 583 21.8 (19.6 – 24.9)	5 574 25.5 (20.9 – 31.9)	0.0001
CIMT (average)	3 555 0.63 (0.56 – 0.72)	4 352 0.62 (0.56 – 0.71)	0.0064
DBP	4 702 78.5 (70.5 – 88.0)	5 739 77.5 (69.5 – 87.0)	0.0003
SBP	4 702 123.5 (111.5 – 138.0)	5 735 121.0 (107.5 – 137.0	0.0001
SES	4 682 10.0 (8.0 – 13.0)	5 570 9.0 (7.0 – 11.0)	0.0001
Education			
No primary	1 519 [32.5%]	2 282 [41.4%]	0.0001
Primary	1 311 [28.0%]	2 072 [37.6%]	
Secondary	1 591 [34.0%]	1 073 [19.5%]	
Tertiary	255 [5.5%]	80 [1.5%]	
Regions			0.0001
East	805 [17.0%]	962 [16.8%]	
South	2 025 [43.0%]	2 798 [48.7%]	
West	1 883 [40.0%]	1 984 [34.5%]	

BMI = body mass index, CIMT = carotid intima-media thickness, DBP = diastolic blood pressure, IQR = 25^{th} – 75^{th} percentile; SBP = systolic blood pressure, SES = socioeconomic status, SD = standard deviation, X = mean.

Association of BMI with CIMT and BP

BMI was positively associated with CIMT as determined by linear regression controlling for age, education, region, SES and sex. The association was observed for the entire AWI-Gen and African regions, as depicted in Figure 1 (Table S1).

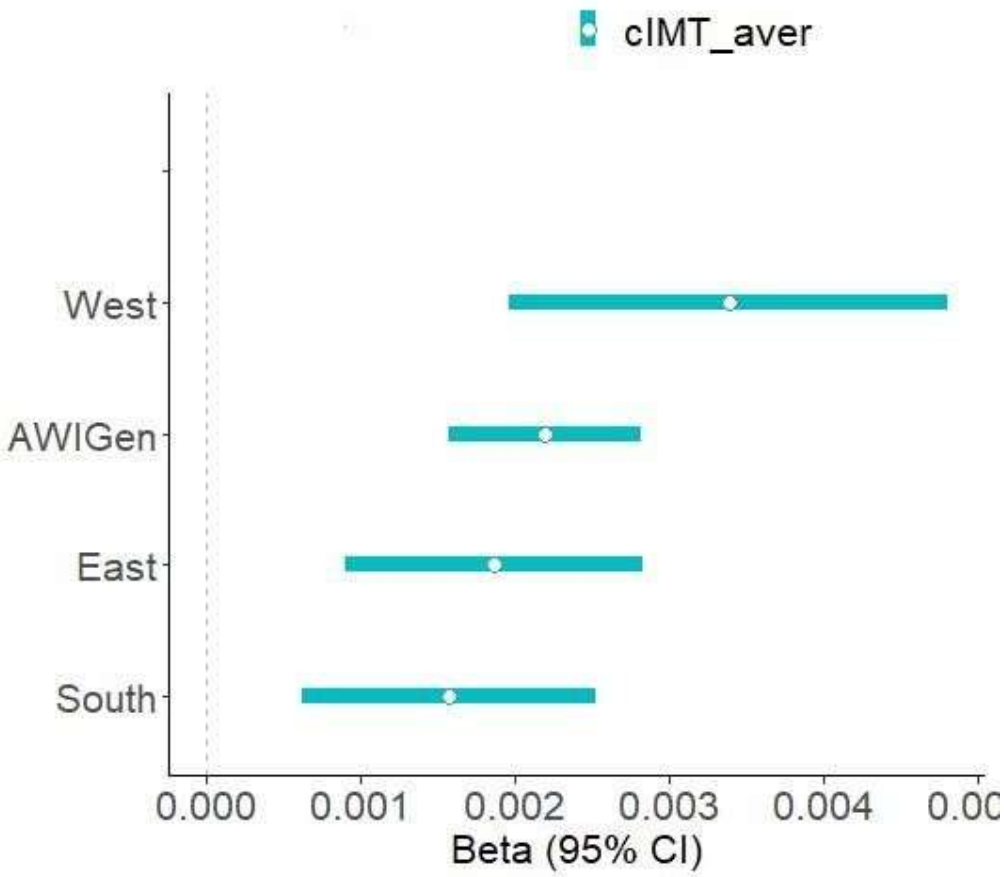


Figure 1: BMI association with average CIMT in AWI-Gen, East, South and West African regions. Linear regression results with CIMT as an outcome variable and BMI as an exposure variable, with covariates (age, education, region, SES, sex) included in the model. CIMT = carotid intima-media thickness. CIMT_aver = average (left and right) carotid intima-media thickness. The x-axis presents the beta coefficient and the 95% confidence interval, y-axis presents total (AWI-Gen) and regional (East, South & West) populations.

Diastolic and systolic blood pressure were positively associated with BMI as determined by linear regression, controlling for age, education, region, SES and sex. The association was for the entire AWI-Gen and regions, as depicted in Figure 2 (Table S1).

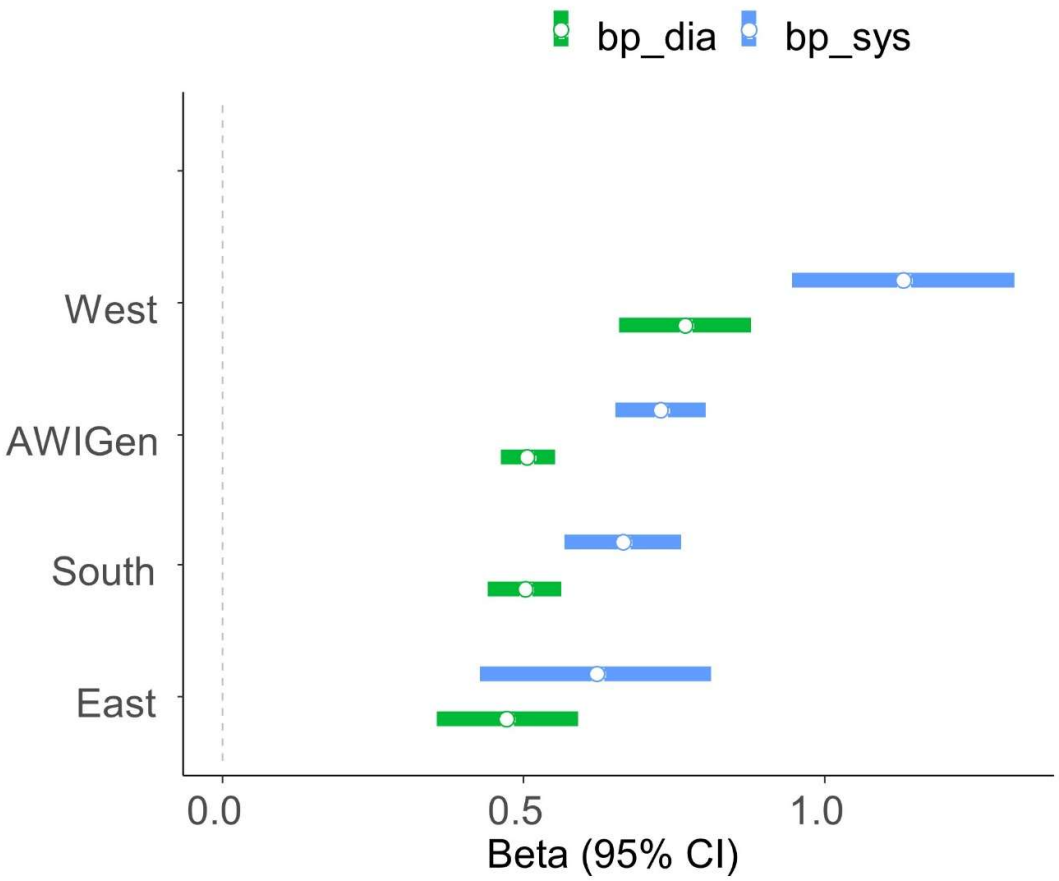


Figure 2: BMI association with diastolic and systolic blood pressure in AWI-Gen, East, South and West African regions. Linear regression results with diastolic or systolic blood pressure as an outcome variable and BMI as an exposure variable, with covariates (age, education, region, SES, sex) included in the model. bp_dia = diastolic blood pressure, bp_sys =systolic blood pressure. The x-axis presents the beta coefficient and the 95% confidence interval, y-axis presents total (AWI-Gen) and regional (East, South & West) populations.

Predictivity and decile analysis results

All PRSs showed transferability except for the BMI PRS developed by Zhang et al. (2023), using LDpred2; full and null models adjusted R^2 were the same, as shown in Figure 3 (Table S2). By total population and regional analysis, all PRSs were statistically significant in their models except CT_AFR_87 and LDpred2_AFR in the East region. PRScsx_His/Lat_Afr_Eur (1.08% in East region) and PRSice_MAMA (1.67% in South and 1.43% in West region) showed the highest predictivity in all analyses. However, PRSice_MAMA predictivity was lower for the Eastern region.

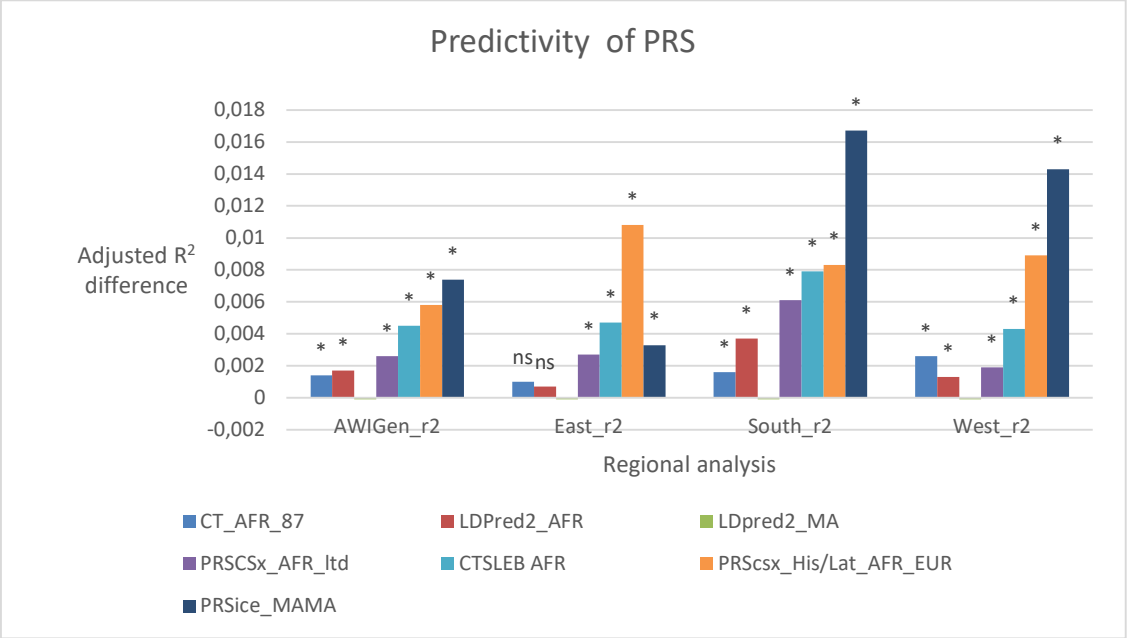


Figure 3: BMI PRS Predictivity summary of the entire AWI-Gen study (n = 9,842), East (n = 1,745), South (n = 4,418) and West (n = 3,719) populations. Predictivity = full – null model adjusted R^2. **Full model** = Body mass index, PRS (CT_AFR_87 or other), age, sex, education, socioeconomic status, principal components 1 to 5; **Null model** = BMI, age, sex, education, socioeconomic status, principal components 1 to 5 (all variables except the PRS). * Indicates the statistical significance of the PRS in the model, while 'ns' indicates not significant.

For all regions and AWI-Gen, PRScsx_His/Lat_Afr_Eur was the most predictive of BMI, accounting for 3.7 kg/m^2 (14%) trait variance, with a large BMI difference between the bottom 10% and top 10%, as depicted in Figure 4 (Table S3a-d). For PRSice_MAMA and CTSLEB_AFR, the significant increase was found in AWI-Gen and the South region. The highest mean difference was observed in AWI-Gen (20% or 5.4 kg/m^2). Furthermore, PRSCSx_AFR_ltd also showed a statistically significant increase in AWI-Gen. However, it should be noted that the other changes were below ten per cent.

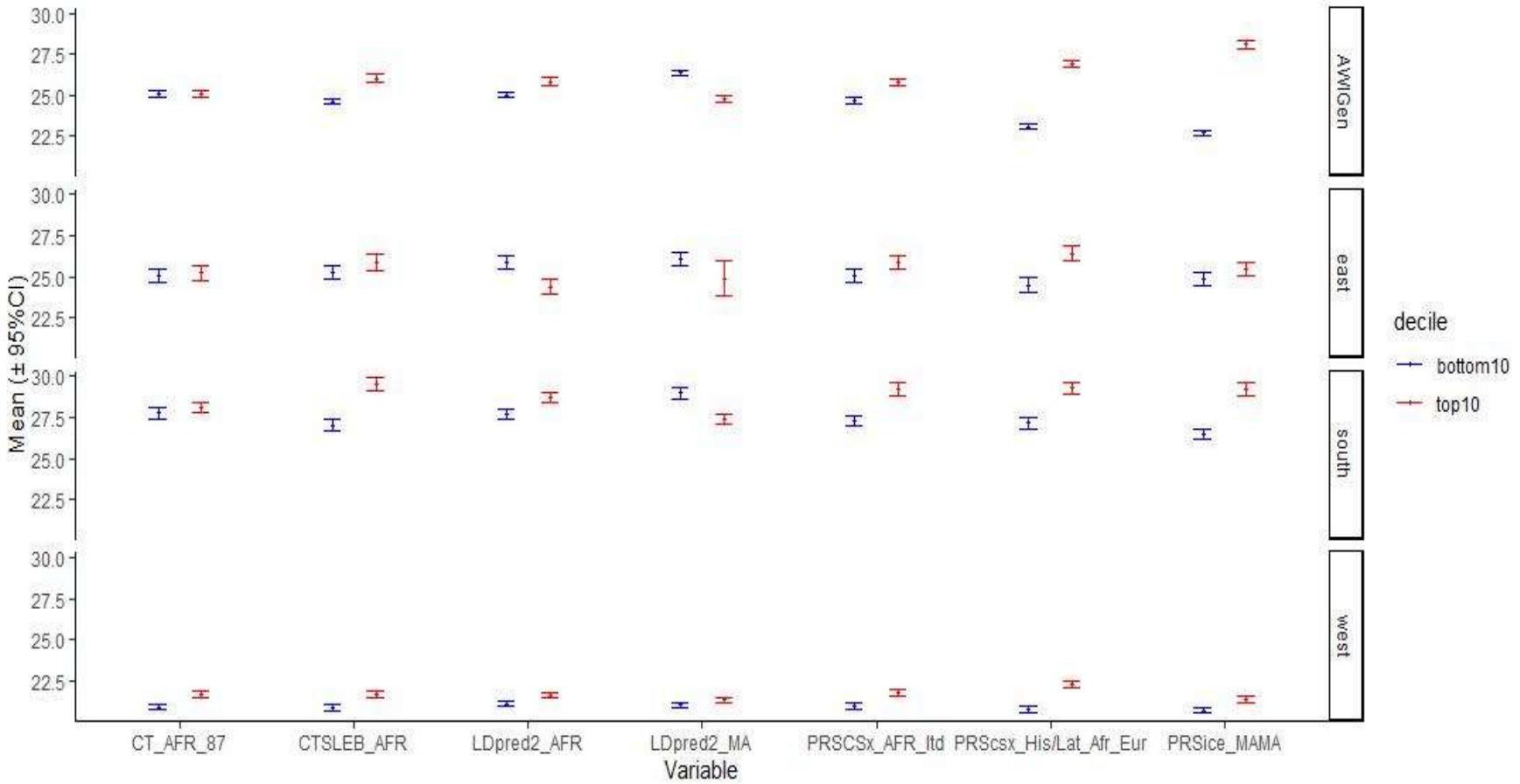


Figure 4: Comparison of the mean BMI of the bottom and top 10% deciles of the population stratified by PRSs within the AWI-Gen population and per region by ANOVA with Bonferroni correction. The y-axis presents the mean, and error bars indicate the 95% confidence interval (CI). The data is grouped by study population, presented on the right side. The x-axis represents the different PRSs used for stratification.

Confounding evaluation

For AWI-Gen, all PRSs showed statistically significant differences among the covariates, as depicted in Figure 5 (Table S4). For the East region, the only PRS that did not have a statistically significant difference among the covariates was PRSCSX_His/Lat_Afr_Eur. A similar pattern was observed for the South and West regions, where CT_AFR_87, CTSLEB_AFR, LDpred2_AFR, and PRSCSX_His/Lat_Afr_Eur did not have statistically significant differences among the covariates. LDpred2_MA had statistically significant differences in all regions, while PRSCSX_His/Lat_Afr_Eur did not have statistically significant differences in the East region. Age and sex were not statistically different for PRSice_MAMA, in the South and West regions. Education and SES showed a statistically significant difference in the South region, while SES was observed in the West region.

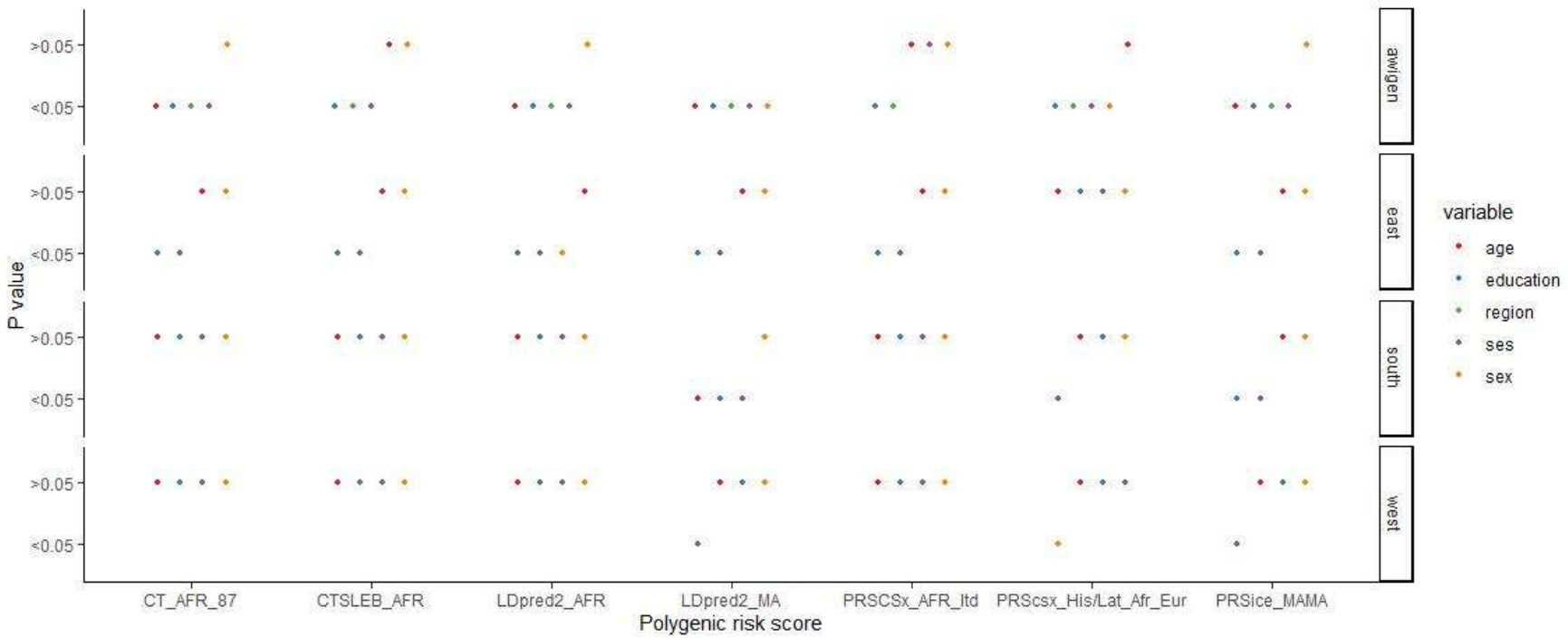


Figure 5: Comparison of p-values of confounders (covariates) in BMI PRS organised deciles one and ten based on the RbG study design for the AWI-Gen and regional populations. The covariates compared the top and bottom 10% based on age and SES, utilising a T-test. Sex, education, and regions were compared using the Chi-square. Region analysis was only performed for AWI-Gen.

A statistically significant positive association of CIMT was found with CTSLEB_AFR in AWI-Gen and CT_AFR_87 in the South population, as shown in Figure 6 (Table S5). However, no PRS controlled for confounding in AWI-Gen (Figure 5 & Table S4). The bottom and top 10% did not differ significantly for CT_AFR_87 for the South region (Figure 4 & Table S3c). Although a statistically negative association of CIMT and LDpred2_MA for both AWI-Gen and the South region was found. It should be noted that LDpred2_MA did not control for confounding (Figure 5 & Table S4), was not predictive (Figure 3 & Table S2), and the mean BMI for the top 10% was lower than the bottom 10% (Figure 4 & Table S3a).

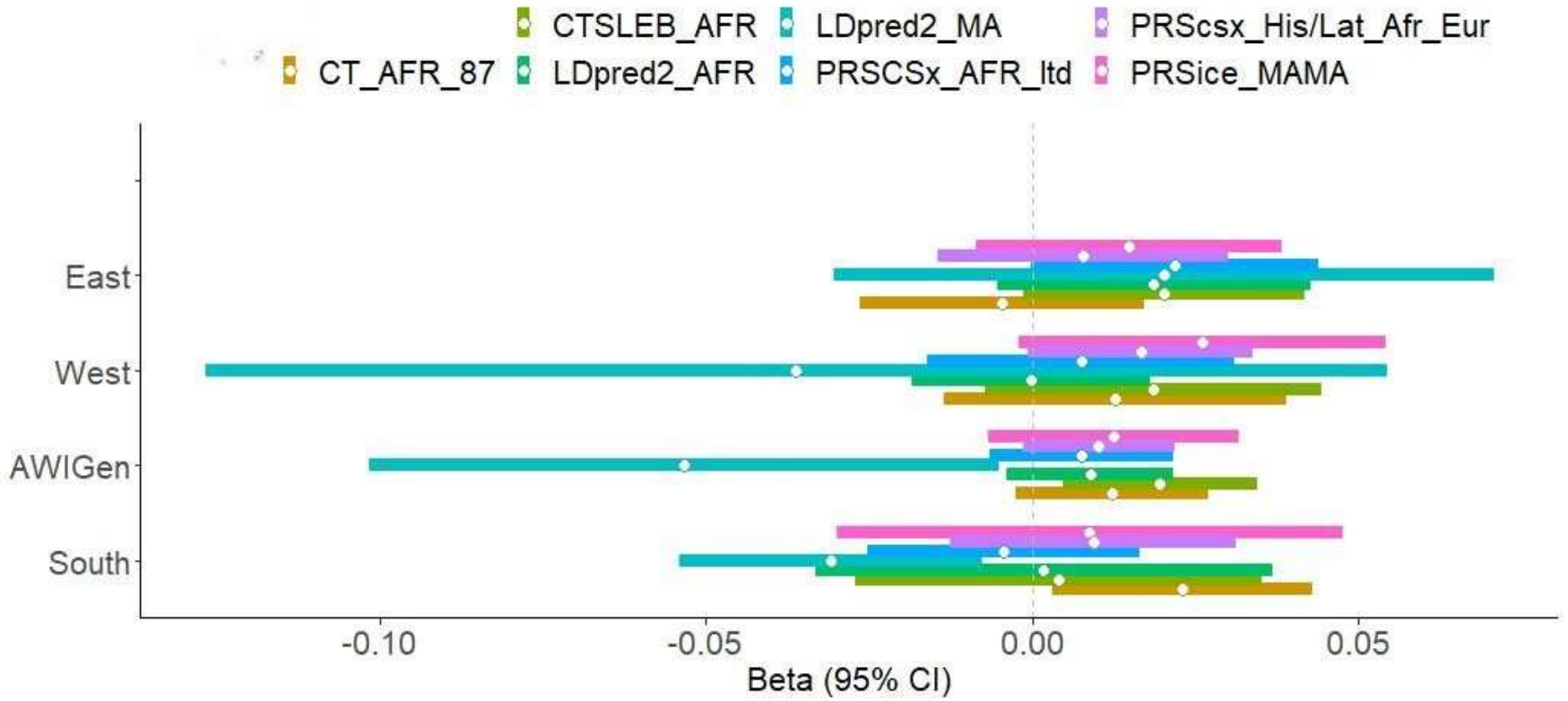


Figure 6: Forest plot causal association evaluation using BMI PRS deciles and average CIMT. The evaluation utilised linear regression with CIMT as the outcome variable and BMI PRS as the exposure variable, correcting for confounding (age, education, region (only AWI-Gen), SES, sex). Two PRSs for both AWI-Gen and the South region showed an association. The results are presented for the AWI-Gen and regional populations (y-axis). The x-axis presents the beta coefficient and the 95% confidence interval.

A statistically significant positive association between BMI PRS top and bottom 10% and diastolic BP for CTSLEB_AFR in AWI-Gen and the East population was found, as presented in Figure 7 (Table S6). However, no PRS controlled for confounding in AWI-Gen (Figure 5 & Table S4). A statistically negative association between diastolic BP and LDpred2_MA was observed for both AWI-Gen and the South region. However, LDpred2_MA did not control for confounding (Figure 5 & Table S4), was not predictive (Figure 3 & Table S2), and the mean BMI for the top 10% was lower than the bottom 10% (Figure 4 & Table S3a, c).

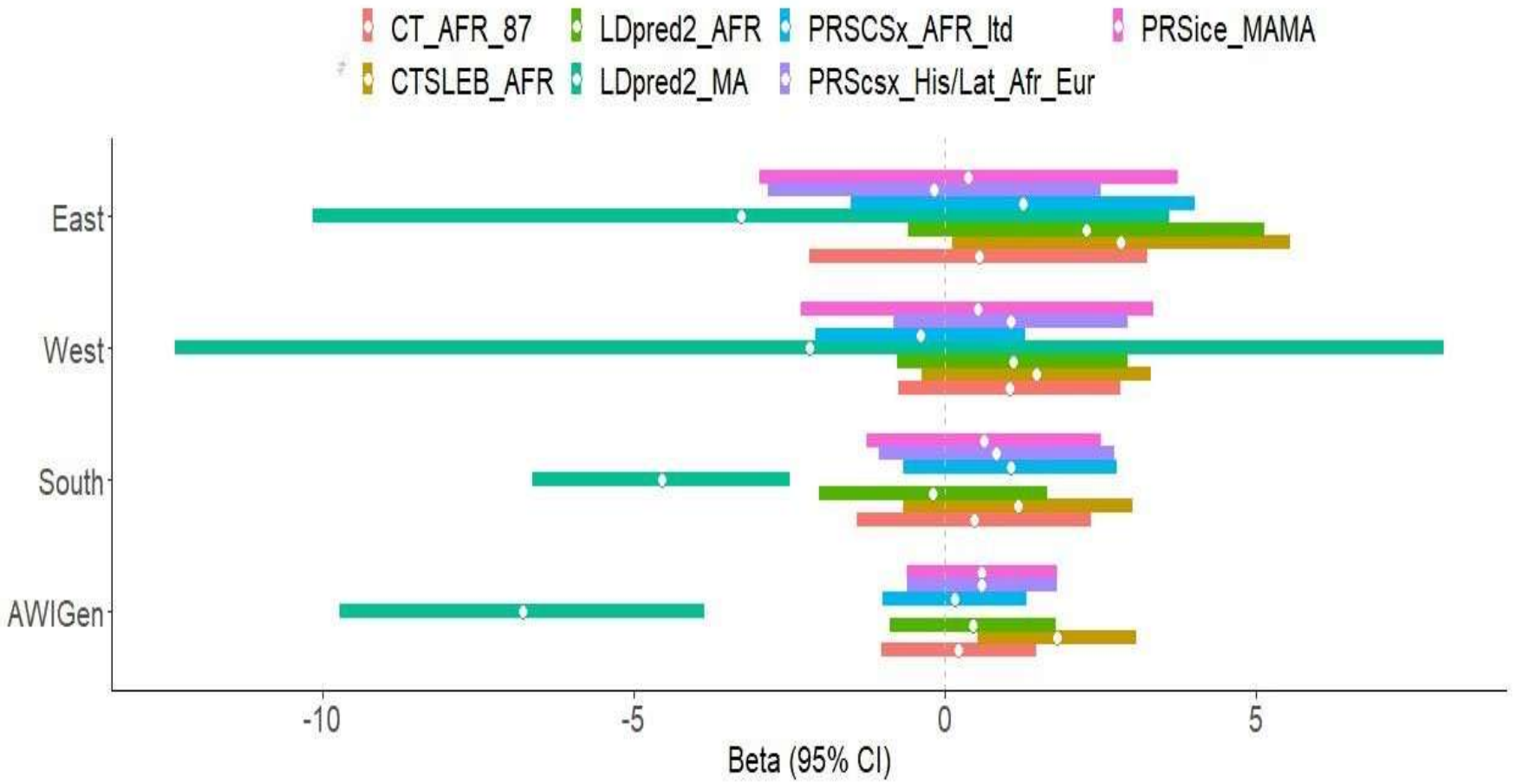


Figure 7: Forest plot causal association evaluation using BMI PRS deciles and diastolic blood pressure. The evaluation utilised linear regression with diastolic blood pressure as the outcome variable and BMI PRS as the exposure variable, correcting for confounding (age, education, region (only AWI-Gen), SES, sex). Two PRSs in AWI-Gen and one in the East and South regions showed associations. The results are presented for the AWI-Gen and regional populations (y-axis). The x-axis presents the beta coefficient and the 95% confidence interval.

A statistically significant positive association was observed with BMI PRS top and bottom 10% and systolic BP for CTSLEB_AFR in AWI-Gen and the East population, plus LDpred2_ARF for East, as presented in Figure 8 (Table S7). However, no PRS controlled for confounding in AWI-Gen (Figure 5 & Table S4). However, LDpred2_AFR was not statistically significant in the East population's full predictive model (Figure 3 & Table S2).

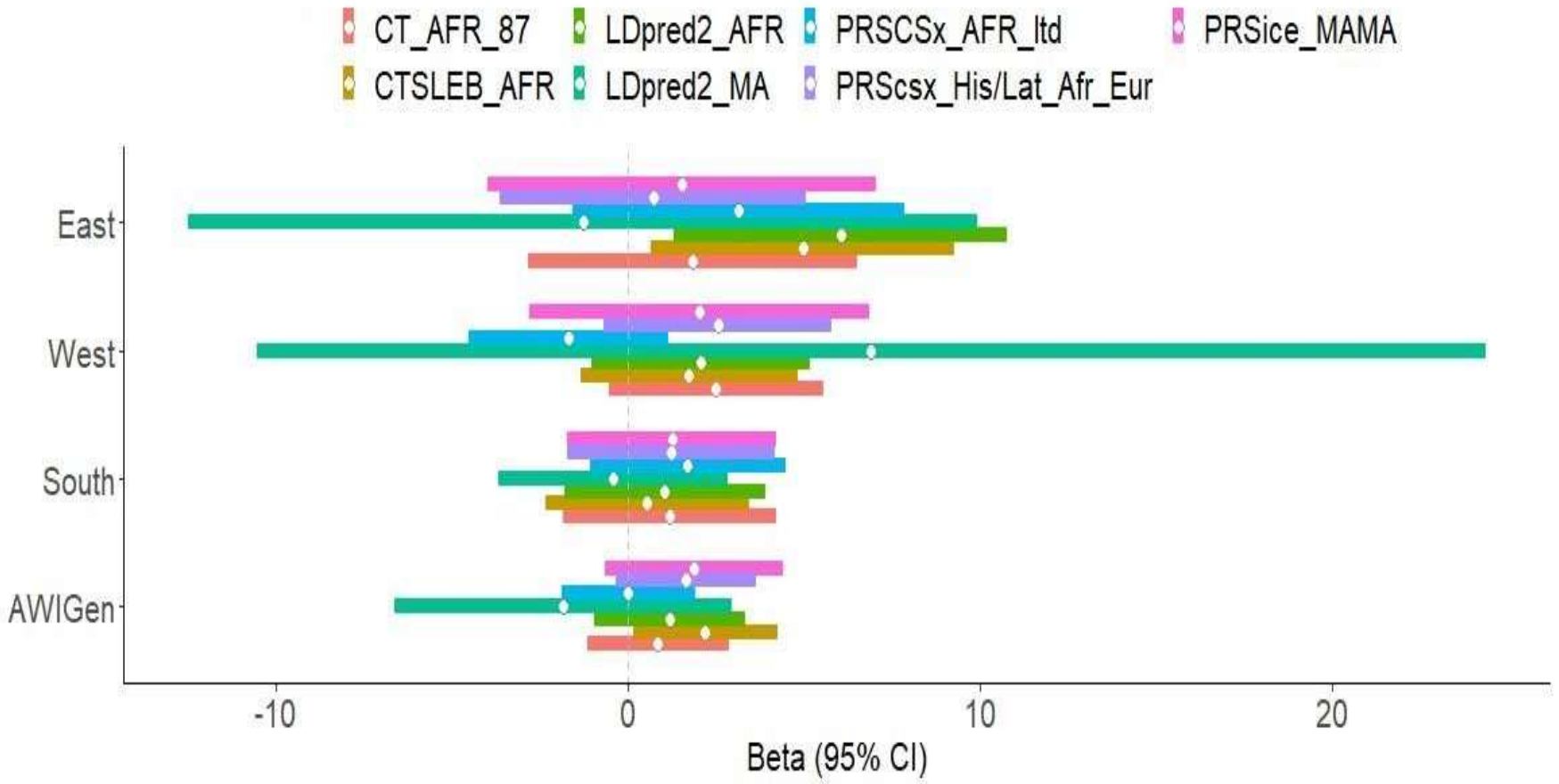


Figure 8: Forest plot causal association evaluation using BMI PRS deciles and systolic blood pressure. The evaluation utilised linear regression with systolic blood pressure as the outcome variable and BMI PRS as the exposure variable, correcting for confounding (age, education, region (only AWI-Gen), SES, sex). Two PRSs in the East region and one in AWI-Gen showed associations. The results are presented for the AWI-Gen and regional populations (y-axis). The x-axis presents the beta coefficient and the 95% confidence interval.

Discussion

The study aimed to evaluate the utility of BMI PRS in the causal association with CIMT and BP in the AWI-Gen data set. Our work showed a positive association of BMI with CIMT and BP. However, polygenic prediction was low in continental African populations compared to European populations due to the limited representation of Africans in the discovery GWAS. Furthermore, most BMI PRSs could control for confounding to enhance causal inference in the RbG strata.

We observed a significant positive association of BMI with CIMT and BP in the combined and AWI-Gen and region-stratified analyses. The association of BMI with CIMT is thought to be related to atherosclerosis, which is a major cause of CVD, among other reasons (Mannami et al., 1997; Takase et al., 2023). Hypertension could lead to atheroma formation (Bots et al., 1997) and be the target that incurs damage because of hypertension (L. Zhang et al., 2019). This association has been observed in multiple ancestries throughout the literature(Agyei et al., 2024; Mannami et al., 1997; Naya et al., 2008; Takase et al., 2023).

Overall, the predictivity of the PRSs was low, however, two multiancestry PRS (PRScsx_His/Lat_Afr_Eur (1.08%) and PRSice_MAMA (1.67%)), as has been reported in other studies(Chikowore et al., 2024; Dikilitas et al., 2020; Fahed et al., 2021; Kullo & Dikilitas, 2020; Majara et al., 2023). PRS are most accurate when applied to European ancestry, contrary to African ancestry, due to the critical mass of GWAS used to compute these PRSs in this population, leading to Eurocentric study bias (Majara et al., 2023). The poor performance of PRS in Africans is also related to the variability of genetic frequencies, different linkage disequilibrium (LD) patterns, and environmental exposures (Campbell & Tishkoff, 2008b). However, gene-environment interaction may play a significant role (Chikowore et al., 2024).

Inconsistent results were found for RbG BMI PRSs association with surrogate markers for CVD: CIMT and BP. The negative association of LDpred2_MA with CIMT and BP in AWI-Gen

and the South region was consistent with the low mean BMI for the top 10%. However, this association can be overlooked since the PRS were not predictive and did not control for confounding. This signifies that RbG was unsuccessful in regions. However, other multivariate RGB studies have performed well in Europeans, contrary to our findings. We suspect that the poor performance of the RgB analysis was impacted by the low predictivity of the PRS.

Our study was limited by the small size of the East African region, the lack of representativeness of the East and West African regions, and the lack of PRS developed using continental African data. These limit the generalisability of the findings to the entire African population.

Conclusion

The study reports an association of BMI with CIMT and BP. However, this association does not translate into a causal association based on the RbG design between the variables. This could be due to the low PRS predictivity. There is a need for larger GWAS studies in continental Africa, which can be used to compute more predictive PRS that can be used in RbG approaches for causal inference. This will enable Africans to benefit from precision medicine and thus assist with the health-associated Sustainable Development Goals in low- and middle-income countries.

Acknowledgements

We acknowledge the AWI-Gen research team, staff, and participants.

Conflict of Interest statement

The authors declare no competing interests.

Funding

“The AWI-Gen collaborative Center is funded by the Office of the Director (OD), the National Institute of Environmental Health Sciences (NIEHS), National Human Genome Research Institute (NHGRI), the Office of AIDS Research (OAR), Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD), and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), of the National Institutes of Health under award number U54HG006938 and its supplements, as part of the H3Africa Consortium. Additionally, funding was granted by the African Partnership for Chronic Disease Research (APCDR) and the Department of Science and Innovation, South Africa, award number DST/CON 0056/2014” (Ramsay et al., 2018).

Data access

“The AWI-Gen dataset is available in the European Genome-phenome Archive (EGA) database (https://ega-archive.org/). The Data and Biospecimen Access Committee of the H3Africa Consortium is responsible for controlling access to this dataset” (Chikowore et al., 2024).

Ethics and consent

The AWI-Gen study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (protocol number M221008).

Authors’ contribution

Conceptualisation: TC, MR, JK, LF; supervision: JK, TC; data analysis: LF, TC; writing original draft: LF; writing, reviewing and editing: LF, TC, JK. MR. All authors read and approved the final manuscript.

References

Agyei, B. A., Wiafe, Y. A., Donkor, A., Anyitey-Kokor, I. C., Yahya, K., & Sarfo, F. S. (2024). The association of blood pressure, body mass index and carotid intima-media thickness as childhood predictors of cardiovascular disease risk: A systematic review and meta-analysis. *Health Science Reports*, *7*(9). https://doi.org/10.1002/HSR2.70033

Ali, S. A., Soo, C., Agongo, G., Alberts, M., Amenga-Etego, L., Boua, R. P., Choudhury, A., Crowther, N. J., Depuur, C., Gómez-Olivé, F. X., Guiraud, I., Haregu, T. N., Hazelhurst, S., Kahn, K., Khayeka-Wandabwa, C., Kyobutungi, C., Lombard, Z., Mashinya, F., Micklesfield, L., … Ramsay, M. (2018). Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Global Health Action*, *11*(sup2). https://doi.org/10.1080/16549716.2018.1507133

Baichoo, S., Souilmi, Y., Panji, S., Botha, G., Meintjes, A., Hazelhurst, S., Bendou, H., Beste, E. de, Mpangase, P. T., Souiai, O., Alghali, M., Yi, L., O'Connor, B. D., Crusoe, M., Armstrong, D., Aron, S., Joubert, F., Ahmed, A. E., Mbiyavanga, M., … Mulder, N. (2018). Developing reproducible bioinformatics analysis workflows for heterogeneous computing environments to support African genomics. *BMC Bioinformatics [Internet]*, *19*(1), 457. https://doi.org/10.1186/s12859-018-2446-1

Bays, H. E., Taub, P. R., Epstein, E., Michos, E. D., Ferraro, R. A., Bailey, A. L., Kelli, H. M., Ferdinand, K. C., Echols, M. R., Weintraub, H., Bostrom, J., Johnson, H. M., Hoppe, K. K., Shapiro, M. D., German, C. A., Virani, S. S., Hussain, A., Ballantyne, C. M., Agha, A. M., & Toth, P. P. (2021). Ten things to know about ten cardiovascular disease risk factors. *American Journal of Preventive Cardiology*, *5*, 100149. https://doi.org/10.1016/J.AJPC.2021.100149

Bots, M. L., Hoes, A. W., Koudstaal, P. J., Hofman, A., & Grobbee, D. E. (1997). Common

carotid intima-media thickness and risk of stroke and myocardial infarction: The Rotterdam Study. *Circulation*, *96*(5), 1432–1437. https://doi.org/10.1161/01.CIR.96.5.1432

Brandenburg, J.T., Clark, L., Botha, G., Panji, S., Baichoo, S., Fields, C., & Hazelhurst, S. (2022). H3AGWAS: a portable workflow for genome wide association studies. *BMC Bioinformatics,* **23**, 1–15

Campbell, M. C., & Tishkoff, S. A. (2008). African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. In *Annual Review of Genomics and Human Genetics* (Vol. 9, pp. 403–433). https://doi.org/10.1146/annurev.genom.9.081307.164258

Chikowore, T., Läll, K., Micklesfield, L. K., Lombard, Z., Goedecke, J. H., Fatumo, S., Norris, S. A., Magi, R., Ramsay, M., Franks, P. W., Pare, G., & Morris, A. P. (2024). Variability of polygenic prediction for body mass index in Africa. *Genome Medicine*, *16*(1), 1–13. https://doi.org/10.1186/S13073-024-01348-X/FIGURES/6

Choudhury, A., Brandenburg, JT., Chikowore, T. *et al.* (2022). Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits. *Nature Communications,* **13**, 1–13 (2022)

Corbin, L. J., Tan, V. Y., Hughes, D. A., Wade, K. H., Paul, D. S., Tansey, K. E., Butcher, F., Dudbridge, F., Howson, J. M., Jallow, M. W., John, C., Kingston, N., Lindgren, C. M., O'Donavan, M., O'Rahilly, S., Owen, M. J., Palmer, C. N. A., Pearson, E. R., Scott, R. A., … Timpson, N. J. (2018). Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03109-y

Dikilitas, O., Schaid, D. J., Kosel, M. L., Carroll, R. J., Chute, C. G., Denny, J. A., Fedotov, A., Feng, Q. P., Hakonarson, H., Jarvik, G. P., Lee, M. T. M., Pacheco, J. A., Rowley, R., Sleiman, P. M., Stein, C. M., Sturm, A. C., Wei, W. Q., Wiesner, G. L., Williams, M. S., … Kullo, I. J. (2020). Predictive Utility of Polygenic Risk Scores for Coronary Heart Disease in Three Major Racial and Ethnic Groups. *American Journal of Human Genetics*, *106*(5),

707–716. https://doi.org/10.1016/J.AJHG.2020.04.002

Fahed, A. C., Aragam, K. G., Hindy, G., Chen, Y. D. I., Chaudhary, K., Dobbyn, A., Krumholz, H. M., Sheu, W. H. H., Rich, S. S., Rotter, J. I., Chowdhury, R., Cho, J., Do, R., Ellinor, P. T., Kathiresan, S., & Khera, A. V. (2021). Transethnic Transferability of a Genome-Wide Polygenic Score for Coronary Artery Disease. *Circulation. Genomic and Precision Medicine*, *14*(1), E003092. https://doi.org/10.1161/CIRCGEN.120.003092

Kullo, I. J., & Dikilitas, O. (2020). Polygenic Risk Scores for Diverse Ancestries: Making Genomic Medicine Equitable ∗. *Journal of the American College of Cardiology*, *76*(6), 715–718. https://doi.org/10.1016/J.JACC.2020.06.028

Lindstrom, M., DeCleene, N. , Dorsey, H. , F. V. , J. C. O. , L. K. E. , M. G. A. , R. C. , S. B. , Varieur Turco, J. , & Roth, G. A. (2022). Global Burden of Cardiovascular Diseases and Risks Collaboration, 1990-2021. *Journal of the American College of Cardiology*, *80*(25), 2372–2425. https://doi.org/10.1016/J.JACC.2022.11.001

Loh, P. R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G. R., Durbin, R., & Price, A. L. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet*, *48*(11), 1443–1448. https://doi.org/10.1038/ng.3679

Majara, L., Kalungi, A., Koen, N., Tsuo, K., Wang, Y., Gupta, R., Nkambule, L. L., Zar, H., Stein, D. J., Kinyanda, E., Atkinson, E. G., & Martin, A. R. (2023). Low and differential polygenic score generalizability among African populations due largely to genetic diversity. *Human Genetics and Genomics Advances*, *4*(2), 100184. https://doi.org/10.1016/J.XHGG.2023.100184

Mannami, T., Konishi, M., Baba, S., Nishi, N., & Terao, A. (1997). Prevalence of asymptomatic carotid atherosclerotic lesions detected by high-resolution ultrasonography and its relation to cardiovascular risk factors in the general population of a Japanese city: The Suita study. *Stroke*, *28*(3), 518–525. https://doi.org/10.1161/01.STR.28.3.518

Naya, T., Hosomi, N., Ohyama, H., Ichihara, S. I., Ban, C. R., Takahashi, T., Taminato, T., Feng, A., Kohno, M., & Koziol, J. A. (2008). Smoking, fasting serum insulin, and obesity

are the predictors of carotid atherosclerosis in relatively young subjects. *Angiology*, *58*(6), 677–684. https://doi.org/10.1177/0003319707303589

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Ramsay, M., Crowther, N. J., Agongo, G., Ali, S. A., Asiki, G., Boua, R. P., Gómez-Olivé, F. X., Kahn, K., Khayeka-Wandabwa, C., Mashinya, F., Micklesfield, L., Mukomana, F., Nonterah, E. A., Soo, C., Sorgho, H., Wade, A. N., Wagner, R. G., Alberts, M., Hazelhurst, S., … Tollman, S. (2018). Regional and sex-specific variation in BMI distribution in four sub-Saharan African countries: The H3Africa AWI-Gen study. *Global Health Action*, *11*(sup2). https://doi.org/10.1080/16549716.2018.1556561

Saldanha, I. J., Skelly, A. C., Ley, K. Vander, Wang, Z., Berliner, E., Bass, E. B., Devine, B., Hammarlund, N., Adam, G. P., Duan-Porter, D., Bañez, L. L., Jain, A., Norris, S. L., Wilt, T. J., Leas, B., Siddique, S. M., Fiordalisi, C. V., Patino-Sutton, C., & Viswanathan, M. (2022). *Strengths and Limitations of RCTs*. https://www.ncbi.nlm.nih.gov/books/NBK584466/

Shim, I., Kuwahara, H., Chen, N. N., Hashem, M. O., AlAbdi, L., Abouelhoda, M., Won, H. H., Natarajan, P., Ellinor, P. T., Khera, A. V., Gao, X., Alkuraya, F. S., & Fahed, A. C. (2023). Clinical utility of polygenic scores for cardiometabolic disease in Arabs. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-41985-1

Takase, M., Nakaya, N., Nakamura, T., Kogure, M., Hatanaka, R., Nakaya, K., Chiba, I., Kanno, I., Nochioka, K., Tsuchiya, N., Hirata, T., Hamanaka, Y., Sugawara, J., Kobayashi, T., Fuse, N., Uruno, A., Kodama, E. N., Kuriyama, S., Tsuji, I., & Hozawa, A. (2023). Carotid Intima Media Thickness and Risk Factor for Atherosclerosis: Tohoku Medical Megabank Community-Based Cohort Study. *Journal of Atherosclerosis and Thrombosis*, *30*(10), 1471. https://doi.org/10.5551/JAT.64039

Touboul, P. J., Hennerici, M. G., Meairs, S., Adams, H., Amarenco, P., Bornstein, N., Csiba,

L., Desvarieux, M., Ebrahim, S., Hernandez Hernandez, R., Jaff, M., Kownator, S., Naqvi, T., Prati, P., Rundek, T., Sitzer, M., Schminke, U., Tardif, J. C., Taylor, A., … Woo, K. S. (2012). Mannheim carotid intima-media thickness and plaque consensus (2004-2006-2011). An update on behalf of the advisory board of the 3rd, 4th and 5th watching the risk symposia, at the 13th, 15th and 20th European Stroke Conferences, Mannheim, Germany, 2004, Brussels, Belgium, 2006, and Hamburg, Germany, 2011. *Cerebrovascular Diseases (Basel, Switzerland)*, *34*(4), 290–296. https://doi.org/10.1159/000343145

Wade, K. H., Chiesa, S. T., Hughes, A. D., Chaturvedi, N., Charakida, M., Rapala, A., Muthurangu, V., Khan, T., Finer, N., Sattar, N., Howe, L. D., Fraser, A., Lawlor, D. A., Davey Smith, G., Deanfield, J. E., & Timpson, N. J. (2018). Assessing the Causal Role of Body Mass Index on Cardiovascular Health in Young Adults: Mendelian Randomization and Recall-by-Genotype Analyses. *Circulation*, *138*(20), 2187–2201. https://doi.org/10.1161/CIRCULATIONAHA.117.033278

Wang, S., Ojewunmi, O. O., Kamiza, A., Ramsay, M., Morris, A. P., Chikowore, T., Fatumo, S., & Asimit, J. L. (2024). Accounting for heterogeneity due to environmental sources in meta-analysis of genome-wide association studies. *Communications Biology*, *7*(1), 1–12. https://doi.org/10.1038/S42003-024-072369

Zhang, H., Zhan, J., Jin, J., Zhang, J., Lu, W., Zhao, R., Ahearn, T. U., Yu, Z., O'Connell, J., Jiang, Y., Chen, T., Okuhara, D., Aslibekyan, S., Auton, A., Babalola, E., Bell, R. K., Bielenberg, J., Bryc, K., Bullis, E., … Chatterjee, N. (2023). A new method for multiancestry polygenic prediction improves performance across diverse populations. *Nature Genetics*, *55*(10), 1757–1768. https://doi.org/10.1038/S41588-023-01501-Z

Zhang, L., Fan, F., Qi, L., Jia, J., Yang, Y., Li, J., & Zhang, Y. (2019). The association between carotid intima-media thickness and new-onset hypertension in a Chinese community-based population. *BMC Cardiovascular Disorders*, *19*(1), 269. https://doi.org/10.1186/S12872-019-1266-1

Supplementary information

Presented in this section are supplementary tables in support of the manuscript figures. They are grouped as those related to BMI association with BP and CIMT, predictivity, comparison of the top and bottom 10% population, confirmation of successful stratification, and RbG PRS BMI causal inference against BP and CIMT. Index for the supplementary tables is as follows:

Table S1: BMI association with CIMT, diastolic and systolic blood pressure in AWI-Gen and regionally. Table S2: Predictivity of BMI PRSs for AWI-Gen and regionally. Table S3a-d: Comparison of decile one (low-risk) and ten (high-risk) after BMI PRS stratification for AWI-Gen(a), East (b), South (c) and West (d). Table S4: Confounding analysis for AWI-Gen and regions based on covariates. Table S5: RbG BMI PRS causal association with CIMT with covariates in the model. Table S6: RbG BMI PRS causal association with diastolic blood pressure with covariates in the model. Table S7: RbG BMI PRS causal association with systolic blood pressure with covariates in the model.

Table S1: BMI association with CIMT, diastolic and systolic blood pressure in AWI-Gen and regionally.

Variable	Region	Beta coefficient	CI_lower	CI_upper	p value
CIMT_aver	AWI-Gen	.0021922	.0015747	.0028097	0.0001
bp_dia	AWI-Gen	.5065628	.4614633	.5516624	0.0001
sys_dia	AWI-Gen	.7269671	.6531048	.8008294	0.0001
CIMT_aver	East	.0018683	.0009055	.002831	0.0001
bp_dia	East	.472005	.3536153	.5903947	0.0001
sys_dia	East	.6192407	.4265655	.8119159	0.0001
CIMT_aver	South	.0015723	.0006217	.0025229	0.001
bp_dia	South	.5007154	.440583	.5608479	0.0001
sys_dia	South	.6651394	.5679276	.7623511	0.0001
CIMT_aver	West	.0033868	.0019659	.0048078	0.0001
bp_dia	West	.7680775	.6592921	.8768629	0.0001
sys_dia	West	1.129727	.9444139	1.31504	0.0001

bp_dia = diastolic blood pressure, bp_sys =systolic blood pressure, CIMT-aver = carotid media-intimal thickness average (Left + Right), CI = confidence interval

Table S2: Predictivity of different BMI PRS in AWI-Gen and regionally

pgs_name	AWI-Gen_r2 (p-value)	East_r2 (p-value)	South_r2 (p-value)	West_r2 (p-value)
CT_AFR_87	0,0014 (0,0001)	0,001 (0,069)	0,0016 (0,001)	0,0026 (0,001)
LDPred2_AFR	0,0017 (0,0001)	0,0007 (0,112)	0,0037 (0,0001)	0,0013 (0,012)
LDpred2_MA	-0,0001 (0,538)	-0,0001 (0,37)	-0,0001 (0,636)	-0,0001 (0.538)
PRSCSx_AFR_ltd	0,0026 (0,0001)	0,0027 (0,007)	0,0061 (0,0001)	0,0019 (0.003)
CTSLEB AFR	0,0045 (0,0001)	0,0047 (0,001)	0,0079 (0,0001)	0,0043 (0,0001
PRScsx_His/Lat_AFR_EUR	0,0058 (0,0001)	0,0108 (0,0001)	0,0083 (0,0001)	0,0089 (0,0001)
PRSice_MAMA	0,0074 (0,0001)	0,0033 (0.001)	0,0167 (0,0001)	0,0143 (0,0001)

R2 represents the difference between the full and the null model.

Table S3a: Comparison of decile one (low-risk = bottom 10%) and ten (high-risk = top 10%)

after BMI PRS stratification for AWI-Gen

PGS_name	decile	CI_lower	mean	CI_upper	p -value
CT_AFR_87	bottom10	24,67	25,06	25,46	1,0000
CT_AFR_87	top10	24,66	25,07	25,47	
LDpred2_AFR	bottom10	24,65	25,03	25,40	0,2700
LDpred2_AFR	top10	25,42	25,84	26,27	
LDpred2_MA	bottom10	25,97	26,35	26,73	0,0001
LDpred2_MA	top10	24,33	24,74	25,15	
PRSCSx_AFR_ltd	bottom10	24,27	24,65	25,02	0,0060
PRSCSx_AFR_ltd	top10	25,33	25,79	26,25	
CTSLEB_AFR	bottom10	24,26	24,62	24,98	0,0001
CTSLEB_AFR	top10	25,56	26,03	26,50	
PRScsx_His/Lat_Afr_Eur	bottom10	22,74	23,09	23,44	0,0001
PRScsx_His/Lat_Afr_Eur	top10	26,46	26,91	27,36	
PGS_nama	bottom10	22,40	22,73	23,05	0,0001
PGS_nama	top10	27,59	28,10	28,61	

CI = confidence interval, se = standard error

Table S3 b: Comparison of decile one (low-risk = bottom 10%) and ten (high-risk = top 10%) after BMI PRS stratification for the East region

PGS_name	decile	CI_lower	mean	CI_upper	p-value
CT_AFR_87	bottom10	24,27	25,06	25,85	
CT_AFR_87	top10	24,34	25,24	26,15	1,0000
LDpred2_AFR	bottom10	25,06	25,89	26,71	
LDpred2_AFR	top10	23,53	24,39	25,26	1,0000
LDpred2_MA	bottom10	25,32	26,11	26,91	
LDpred2_MA	top10	22,87	24,92	26,96	1,0000
PRSCSx_AFR_ltd	bottom10	24,32	25,11	25,90	
PRSCSx_AFR_ltd	top10	25,00	25,87	26,73	1,0000
CTSLEB_AFR	bottom10	24,50	25,26	26,03	
CTSLEB_AFR	top10	24,86	25,85	26,84	1,0000
PRScsx_His/Lat_Afr_Eur	bottom10	23,59	24,47	25,35	
PRScsx_His/Lat_Afr_Eur	top10	25,52	26,42	27,32	0,0540
PGS_nama	bottom10	24,10	24,87	25,64	
PGS_nama	top10	24,65	25,49	26,33	1,0000

CI = confidence interval, se = standard error

Table S3c: Comparison of decile one (low-risk = bottom 10%) and ten (high-risk = top 10%) after BMI PRS stratification for the South region

PGS_name	decile	CI_lower	mean	CI_upper	p-value
CT_AFR_87	bottom10	27,08	27,75	28,42	
CT_AFR_87	top10	27,43	28,09	28,75	1,0000
LDpred2_AFR	bottom10	27,04	27,69	28,35	
LDpred2_AFR	top10	28,05	28,70	29,35	1,0000
LDpred2_MA	bottom10	28,33	29,00	29,67	
LDpred2_MA	top10	26,72	27,37	28,01	0,0420
PRSCSx_AFR_ltd	bottom10	26,70	27,32	27,94	
PRSCSx_AFR_ltd	top10	28,44	29,20	29,95	0,0060
CTSLEB_AFR	bottom10	26,41	27,03	27,65	
CTSLEB_AFR	top10	28,72	29,51	30,29	0,0001
PRScsx_His/Lat_Afr_Eur	bottom10	26,50	27,15	27,80	
PRScsx_His/Lat_Afr_Eur	top10	28,58	29,29	30,01	0,0010
PGS_nama	bottom10	25,83	26,46	27,09	
PGS_nama	top10	28,39	29,18	29,96	0,0001

CI = confidence interval, se = standard error

Table S3d: Comparison of decile one (low-risk = bottom 10%) and ten (high-risk = top 10%) after BMI PRS stratification for the West region

PGS_name	decile	CI_lower	mean	CI_upper	p-value
CT_AFR_87	bottom10	20,59	20,91	21,22	0,1010
CT_AFR_87	top10	21,35	21,69	22,02	
LDpred2_AFR	bottom10	20,78	21,13	21,47	1,0000
LDpred2_AFR	top10	21,27	21,64	22,01	
LDpred2_MA	bottom10	20,72	21,05	21,38	1,0000
LDpred2_MA	top10	20,99	21,35	21,72	
PRSCSx_AFR_ltd	bottom10	20,67	21,00	21,34	0,1020
PRSCSx_AFR_ltd	top10	21,41	21,79	22,17	
CTSLEB_AFR	bottom10	20,57	20,88	21,19	0,0740
CTSLEB_AFR	top10	21,34	21,69	22,04	
PRScsx_His/Lat_Afr_Eur	bottom10	20,43	20,78	21,13	0,0001
PRScsx_His/Lat_Afr_Eur	top10	21,85	22,28	22,71	
PGS_nama	bottom10	20,38	20,71	21,04	0,6610
PGS_nama	top10	20,99	21,42	21,84	

CI = confidence interval, se = standard error

Table S4: Confounding analysis for AWI-Gen and regions based on covariates

		p – value			
pgs_name	**variable**	**AWI-Gen**	**East**	**South**	**West**
CT_AFR_87	region	0,0001			
CT_AFR_87	age	0,0023	0,6235	0,9047	0,5879
CT_AFR_87	education	0,0001	0,048	0,681	0,604
CT_AFR_87	SES	0,0023	0,00001	0,6082	0,3531
CT_AFR_87	sex	0,204	0,075	0,147	0,42
CTSLEB_AFR	region	0,0001			
CTSLEB_AFR	age	0,083	0,2896	0,0812	0,9545
CTSLEB_AFR	education	0,001	0,0001	0,692	0,528
CTSLEB_AFR	SES	0,00001	0,00001	0,5247	0,8156
CTSLEB_AFR	sex	0,819	0,059	0,643	0,215
LDpred2_AFR	region	0,0001			
LDpred2_AFR	age	0,00001	0,4795	0,1363	0,7865
LDpred2_AFR	education	0,0001	0,007	0,655	0,16
LDpred2_AFR	SES	0,00001	0,00001	0,3746	0,5392
LDpred2_AFR	sex	0,378	0,022	0,575	0,954
LDpred2_MA	region	0,0001			
LDpred2_MA	age	0,00001	0,0973	0,00001	0,3631
LDpred2_MA	education	0,0001	0,0001	0,0001	0,719
LDpred2_MA	SES	0,00001	0,00001	0,00001	0,0011
LDpred2_MA	sex	0,034	0,448	0,95	0,185
pgs_nama	region	0,0001			
pgs_nama	age	0,0001	0,7372	0,5995	0,638
pgs_nama	education	0,0001	0,045	0,013	0,457
pgs_nama	SES	0,00001	0,00001	0,025	0,0013
pgs_nama	sex	0,178	0,241	0,507	0,726
PRSCSx_AFR_ltd	region	0,0001			
PRSCSx_AFR_ltd	age	0,347	0,508	0,4034	0,6687
PRSCSx_AFR_ltd	education	0,001	0,01	0,229	0,303
PRSCSx_AFR_ltd	SES	0,569	0,0104	0,2677	0,368
PRSCSx_AFR_ltd	sex	0,855	0,317	0,779	0,191
PRScsx_His/Lat_Afr_Eur	region	0,0001			
PRScsx_His/Lat_Afr_Eur	age	0,969	0,5139	0,5527	0,2071
PRScsx_His/Lat_Afr_Eur	education	0,0001	0,683	0,103	0,242
PRScsx_His/Lat_Afr_Eur	SES	0,00001	0,132	0,0001	0,9473
PRScsx_His/Lat_Afr_Eur	sex	0,006	0,709	0,791	0,017

Age & SES compared by T-test, education, region & sex compared by Chi Square
SES = socioeconomic status

Table S5: RbG BMI PRS causal association with CIMT with covariates in the model

pgs_name	region	Beta coefficient	CI_lower	CI_upper	p value
CT_AFR_87	AWI-Gen	.0122095	-.0025454	.0269645	0.105
CTSLEB_AFR	AWI-Gen	.0196077	.0046645	.0345508	0.010
LDpred2_AFR	AWI-Gen	.0089038	-.0038095	.021617	0.170
LDpred2_MA	AWI-Gen	-.0533731	-.1015438	-.0052024	0.030
PRSice_MAMA	AWI-Gen	.0124047	-.0067755	.0315849	0.205
PRSCSx_AFR_ltd	AWI-Gen	.007623	-.0064064	.0216523	0.287
PRScsx_His/Lat_Afr_Eur	AWI-Gen	.0102051	-.0013985	.0218088	0.085
CT_AFR_87	East	-.0046905	-.0265078	.0171269	0.673
CTSLEB_AFR	East	.0203087	-.0012292	.0418466	0.065
LDpred2_AFR	East	.0186111	-.0054228	.042645	0.129
LDpred2_MA	East	.0202421	-.0303101	.0707944	0.431
PRSice_MAMA	East	.0147661	-.0086897	.038222	0.216
PRSCSx_AFR_ltd	East	.0217795	-.0002577	.0438167	0.053
PRScsx_His/Lat_Afr_Eur	East	.0077722	-.0144088	.0299533	0.491
CT_AFR_87	South	.0229914	.0030802	.0429026	0.024
CTSLEB_AFR	South	.0040441	-.0271068	.0351951	0.799
LDpred2_AFR	South	.0017944	-.0331334	.0367222	0.920
LDpred2_MA	South	-.0308387	-.0541228	-.0075546	0.010
PRSice_MAMA	South	.0087946	-.0299917	.0475809	0.656
PRSCSx_AFR_ltd	South	-.0044092	-.0252768	.0164585	0.678
PRScsx_His/Lat_Afr_Eur	South	.0093554	-.0126103	.0313211	0.403
CT_AFR_87	West	.0126605	-.0135648	.0388859	0.344
CTSLEB_AFR	West	.0185492	-.0071894	.0442878	0.158
LDpred2_AFR	West	-.0-.0002211	-.0185453	.01810310	0.981
LDpred2_MA	West	-.0361815	-.1267359	.054373	0.433
PRSice_MAMA	West	.0260764	-.0020684	.0542211	0.069
PRSCSx_AFR_ltd	West	.0074956	-.0160851	.0310762	0.533
PRScsx_His/Lat_Afr_Eur	West	.0165751	-.0005598	.03371	0.058

Table S6: RbG BMI PRS causal association with diastolic blood pressure with covariates in the model

pgs_name	region	Beta coefficient	CI_lower	CI_upper	p value
CT_AFR_87	AWI-Gen	.2257163	-1.012721	1.464153	0.721
CTSLEB_AFR	AWI-Gen	1.808675	.5299273	3.087423	0.006
LDpred2_AFR	AWI-Gen	.4548259	-.8837041	1.793356	0.505
LDpred2_MA	AWI-Gen	-6.786674	-9.717538	-3.855809	0.000
PRSice_MAMA	AWI-Gen	.592578	-.6139968	1.799153	0.336
PRSCSx_AFR_ltd	AWI-Gen	.1608236	-.9942265	1.315874	0.785
PRScsx_His/Lat/Afr/Eur	AWI-Gen	.592578	-.6139968	1.799153	0.336
CT_AFR_87	East	.5428249	-2.173876	3.259526	0.695
CTSLEB_AFR	East	2.829717	.1175334	5.541901	0.041
LDpred2_AFR	East	2.276413	-.5839255	5.136752	0.118
LDpred2_MA	East	-3.269125	-10.15014	3.611889	0.350
PRSice_MAMA	East	.3834689	-2.972885	3.739823	0.822
PRSCSx_AFR_ltd	East	1.2612	-1.502786	4.025186	0.370
PRScsx_His/Lat_Afr_Eur	East	-.1684864	-2.849269	2.512296	0.902
CT_AFR_87	South	.4712872	-1.408139	2.350713	0.623
CTSLEB_AFR	South	1.17949	-.6678789	3.02686	0.210
LDpred2_AFR	South	-.1866131	-2.015435	1.642209	0.841
LDpred2_MA	South	-4.55532	-6.624131	-2.486509	0.000
PRSice_MAMA	South	.6308554	-1.255945	2.517656	0.512
PRSCSx_AFR_ltd	South	1.053766	-.6619486	2.769481	0.228
PRScsx_His/Lat_Afr_Eur	South	.8317773	-1.055225	2.71878	0.387
CT_AFR_87	West	1.0378	-.7495452	2.825145	0.255
CTSLEB_AFR	West	1.471654	-.3769588	3.320267	0.118
LDpred2_AFR	West	1.094892	-.7598823	2.949666	0.247
LDpred2_MA	West	-2.168199	-12.35984	8.023445	0.676
PRSice_MAMA	West	.5226516	-2.304641	3.349945	0.717
PRSCSx_AFR_ltd	West	-.3954846	-2.078237	1.287268	0.645
PRScsx_His/Lat_Afr_Eur	West	1.061148	-.8119843	2.934281	0.266

Table S7: RbG BMI PRS causal association with systolic blood pressure with covariates in the model

pgs_name	region	Beta Coefficient	CI_lower	CI_upper	p value
CT_AFR_87	AWI-Gen	.8598903	-1.156481	2.876262	0.403
CTSLEB_AFR	AWI-Gen	2.202398	.1560457	4.24875	0.035
LDpred2_AFR	AWI-Gen	1.19726	-.9505706	3.34509	0.274
LDpred2_MA	AWI-Gen	-1.827824	-6.615321	2.959673	0.454
PRSice_MAMA	AWI-Gen	1.894413	-.6396933	4.428519	0.143
PRSCSx_AFR_ltd	AWI-Gen	.0289909	-1.848564	1.906546	0.976
PRScsx_His/Lat_Afr_Eur	AWI-Gen	1.654272	-.3233769	3.631922	0.101
CT_AFR_87	East	1.835247	-2.835031	6.505525	0.440
CTSLEB_AFR	East	4.979644	.668131	9.291157	0.024
LDpred2_AFR	East	6.051396	1.330208	10.77258	0.012
LDpred2_MA	East	-1.271674	-12.45516	9.911814	0.823
PRSice_MAMA	East	1.542638	-3.976188	7.061465	0.583
PRSCSx_AFR_ltd	East	3.151191	-1.546496	7.848877	0.188
PRScsx_His/Lat_Afr_Eur	East	.7336326	-3.609633	5.076898	0.740
CT_AFR_87	South	1.209227	-1.820933	4.239387	0.434
CTSLEB_AFR	South	.5588837	-2.321766	3.439533	0.703
LDpred2_AFR	South	1.057934	-1.798428	3.914296	0.467
LDpred2_MA	South	-.4066584	-3.64679	2.833473	0.805
PRSice_MAMA	South	1.270894	-1.695716	4.237504	0.401
PRSCSx_AFR_ltd	South	1.713192	-1.055552	4.481937	0.225
PRScsx_His/Lat_Afr_Eur	South	1.243496	-1.697358	4.184349	0.407
CT_AFR_87	West	2.506522	-.5349478	5.547992	0.106
CTSLEB_AFR	West	1.752853	-1.328861	4.834568	0.264
LDpred2_AFR	West	2.074288	-1.027606	5.176182	0.190
LDpred2_MA	West	6.919755	-10.53037	24.36988	0.436
PRSice_MAMA	West	2.037065	-2.779092	6.853222	0.407
PRSCSx_AFR_ltd	West	-1.667755	-4.490809	1.1553	0.247
PRScsx_His/Lat_Afr_Eur	West	2.563517	-.6700771	5.797112	0.120

CI = confidence interval

Chapter 3: EXTENDED METHODS

3.1 EXTENDED METHODOLOGY

This chapter entails methods that could not fit into the manuscript. These included the study setting of the primary study and power calculations,

3.1.1 Study setting of the primary study

The primary study, Africa Wits-INDEPTH partnership for genomic studies (AWI-Gen), collected phenotypic and genotypic data to assess risk factors and endpoint cardiovascular and metabolic disease in East, South and West African regions (Ali et al., 2018). Phase 1 of the study was a population-based cross-sectional study that involved six centres: East Africa (Nairobi, Kenya), South Africa (Agincourt, Dikgale and Soweto) and West Africa (Nanoro, Burkina Faso, and Navrongo, Ghana), see Figure 3.1 for further details. Phase 2 involved follow-up after 3 to 4 years; thus, a longitudinal and heterogeneous sample pool was obtained. Participants were recruited; measures like anthropometrics, diet, physical activity, alcohol/tobacco use, socioeconomic status, and fat distribution were performed. In addition, blood was obtained for genotyping, and disease biomarker analysis was undertaken. The study population consisted of approximately 12,000 participants, both genders aged 29 years and older.

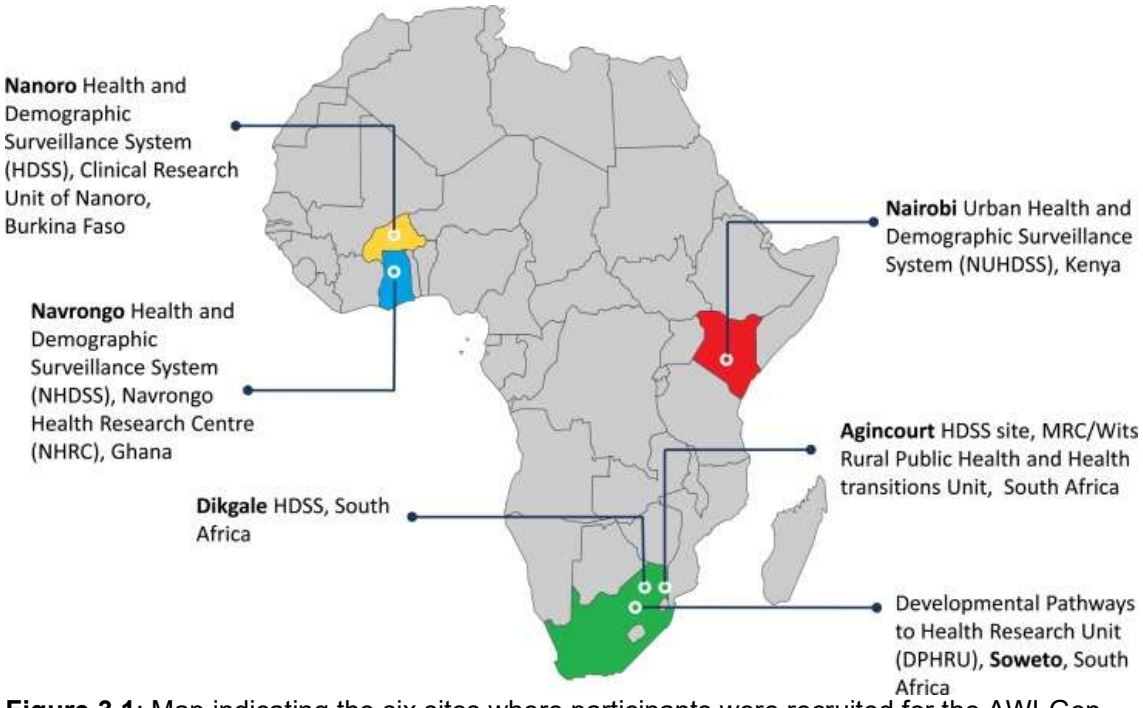


Figure 3.1: Map indicating the six sites where participants were recruited for the AWI-Gen study (Ramsay et al., 2016)

3.1.2 Study design of the secondary study

The cross-sectional study design where a comparison of different PRS methods undertaken utilising data from phase 1 of the AWI-Gen study for analysis. The top and bottom BMI PRS participants from a normal distribution will be sampled and compared. The seven different PRS utilised for comparison will be obtained from the PRS catalogue (https://www.pgscatalog.org/search/?q=bmi) and applied to the AWI-Gen study data. The selected PRS were those that account for multi-ancestral populations.

3.1.3 Power calculation

The power was calculated using the analytical mode of the Recall by Genotype Study Planner app (http://ieushiny.biocompute.org.uk/rbg-app/) (Corbin et al., 2018a). The following inputs were used for calculation: total genotyped population (depended on the study population), GRS percentile included (10), sample size of the percentile (based on study population), alpha level (0.05), R2 for both between PRS and exposure, plus exposure and outcome (0.1). The

genotype population input was 9,842, 1,745, 4,418 and 3,719 for AWI-Gen, East, South and West populations, respectively. The sample size was based on the lowest group during decile computation with 964, 115, 423 and 339 for AWI-Gen, East, South and West populations, respectively. The study power was superior for AWI-Genas; all calculations were above 80%, as depicted in Table 3.1. From the regions, the South and West had good power except for the random recall study, which was low; the South region had slightly better power than the West. The East region showed weak study power as depicted by the percentages, thus, a more representative sample should be obtained.

Table 3.1: Power calculation results per study population.

	AWI-Gen	East	South	West
Minimum recruitment rate (%)	48.97	32.95	47.87	45.59
Power to detect a difference in exposure in 'RbGmv study'	1.00	1.00	1.00	1.00
Power of 'RbGmv study'	1.00	0.47	0.95	0.90
Power of 'random recall study'	0.88	0.19	0.54	0.46
Power of 'total cohort study'	1.00	0.99	1.00	1.00

References

Ali SA, Soo C, Agongo G, Alberts M, Amenga-Etego L, Boua RP, et al. Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. Glob Health Action [Internet]. 2018 Nov 16 [cited 2025 Apr 6];11(sup2). Available from: https://www.tandfonline.com/doi/abs/10.1080/16549716.2018.1507133

Corbin LJ, Tan VY, Hughes DA, Wade KH, Paul DS, Tansey KE, et al. Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference. Nat

Commun [Internet]. 2018 Dec 1 [cited 2025 Apr 23];9(1):711. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5818506/

Ramsay M, Crowther N, Tambo E, Agongo G, Baloyi V, Dikotope S, et al. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. Glob Health Epidemiol Genom [Internet]. 2016 [cited 2025 Apr 25];1:e20. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5732578/

APPENDICES

Appendix 1: Plagiarism declaration form



PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I Lebogang Francis Mogongoa (Student number: 2485509) am a student registered for the degree of

_Master of Science in Epidemiology and Biostatistics in the academic year 2025.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included as an appendix a report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document.

Signature: _ ___ Date: _16 May 2025_____

Appendix 2: Plagiarism/Turnitin cover page

Lebogang Mogongoa 

Prof Juliana Kagura 

MSc updated

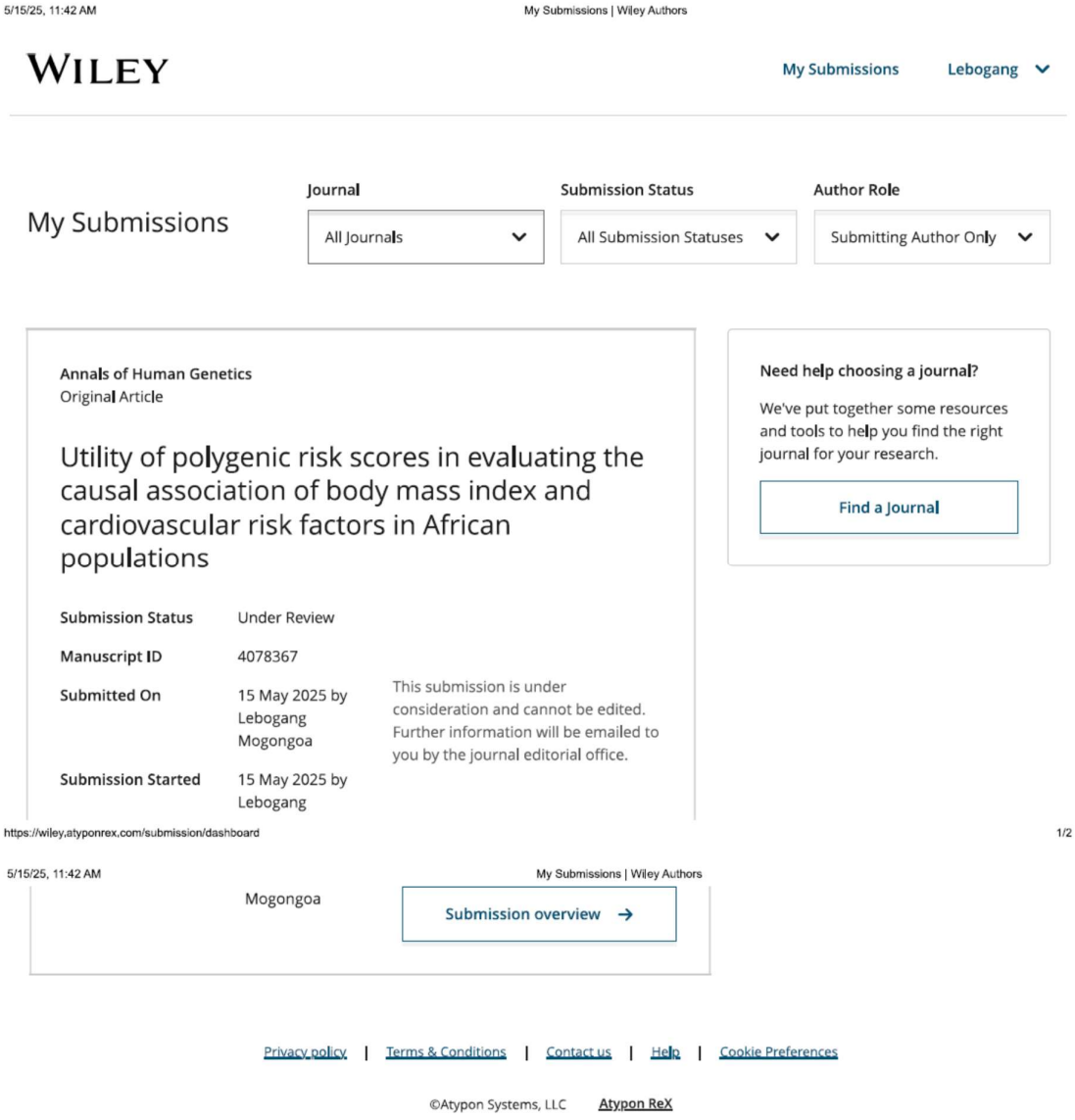
ORIGINALITY REPORT

11%	8%	10%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	wiredspace.wits.ac.za Internet Source	<1%
2	www.frontiersin.org Internet Source	<1%
3	hdl.handle.net Internet Source	<1%
4	Mathenge, Violet Wairimu. "The Association Between Alcohol Consumption and the Metabolic Syndrome and its Component Diseases in a Rural Ghanaian Population", University of the Witwatersrand, Johannesburg (South Africa), 2025 Publication	<1%
5	genomemedicine.biomedcentral.com Internet Source	<1%
6	www.tandfonline.com Internet Source	<1%
7	bmcpublichealth.biomedcentral.com Internet Source	<1%
8	pmc.ncbi.nlm.nih.gov Internet Source	<1%
9	research.edgehill.ac.uk Internet Source	<1%

Appendix 3: Evidence of article submission



Appendix 4: Journal authors' guidelines

Annals of human genetics

Sections

1. Submission
2. Aims and Scope
3. Manuscript Categories and Requirements
4. Preparing Your Submission
5. Editorial Policies and Ethical Considerations
6. Author Licensing
7. Publication Process After Acceptance
8. Post Publication
9. Editorial Office Contact Details

1. SUBMISSION

Thank you for your interest in the *Annals of Human Genetics*. Note that submission implies that the content has not been published or submitted for publication elsewhere except as a brief abstract in the proceedings of a scientific meeting or on a preprint server.

Both new and revised submissions should be made via the Research Exchange submission portal at **https://wiley.atyponrex.com/journal/AHG**. For technical help with the submission system, please review our **FAQs** or contact **submissionhelp@wiley.com**.

4. PREPARING YOUR SUBMISSION

Free Format Submission

Annals of Human Genetics now offers **Free Format submission** for a simplified and streamlined submission process.

Before you submit, you will need:

- Your manuscript: this should be an editable file including text, figures, and tables, or separate files – whichever you prefer. All required sections should be contained in your manuscript, including abstract (which does need to be correctly styled), introduction, methods, results, and conclusions. Figures and tables should have legends. Figures should be uploaded in the highest resolution possible. If the figures are not of sufficiently high quality your manuscript may be delayed. We also encourage you to include your figures within the main document to make it easier for editors and reviewers to read your manuscript. References may be submitted in any style or format, as long as it is consistent throughout the manuscript. Supporting information should be submitted in separate files. If the manuscript, figures or tables are difficult for you to read, they will also be difficult for the editors and reviewers, and the editorial office will send it back to you for revision. Your manuscript may also be sent back to you for revision if the quality of English language is poor.
- An ORCID ID, freely available at https://orcid.org. *(Why is this important? Your article, if accepted and published, will be attached to your ORCID profile. Institutions and funders are increasingly requiring authors to have ORCID IDs. See Section 1 for more details.)*
    - Your ORCID profile should include:
        1. Name (First, last)
        2. Employment (current affiliation and history preferably with a link to an organization identifier such as the Research Organization Registry (ROR) https://ror.org/about/)
        3. Work (list of publications with full citations)
- The title page of the manuscript, including:
    - Your co-author details, including affiliation and email address. *(Why is this important? We need to keep all co-authors informed of the outcome of the peer review process.)*
    - Statements relating to our ethics and integrity policies, which may include any of the following *(Why are these important? We need to uphold rigorous ethical standards for the research we consider for publication):*
        - data availability statement
        - funding statement
        - conflict of interest disclosure

- ethics approval statement
- patient consent statement
- permission to reproduce material from other sources
- clinical trial registration

To submit, login at **https://submission.wiley.com/journal/ahg** and create a new submission. Follow the submission steps as required and submit the manuscript.

Parts of the Manuscript

For manuscripts submitted via the new Wiley Authors submission portal, manuscripts can be uploaded either as a single Microsoft Word document (containing the main text, tables and figures), or with figures and tables provided as separate files. Should your manuscript reach revision stage, figures and tables must be provided as separate files. The main manuscript file should be submitted in Microsoft Word (.doc or .docx) or LaTex (.tex) formats.

Cover Letters and Conflict of Interest statements may be provided as separate files, included in the manuscript, or provided as free text in the submission system. A statement of funding (including grant numbers, if applicable) should be included in the "Acknowledgements" section of your manuscript.

Text File

The text file should include:

- A short informative title containing the major key words. The title should not contain abbreviations (see Wiley's best practice SEO tips);
- A short running title of less than 70 characters;
- The full names (First, Middle and Last) of the authors written in capitals;
- The author's institutional affiliations at which the work was carried out, (footnote for author's present address if different to where the work was carried out);
- Summary and Keywords;
- Main Text;
- Acknowledgements;
- Conflict of Interest statement;
- References;
- Tables (each table complete with title and footnotes);
- Figure Legends;
- Appendices (if relevant).

Appendix 5: Data access permission letter

AWI-Gen Data Access and Use Agreement for internal(SBIMB/Wits) users

This agreement stipulates the terms and conditions of data access and use of the AWI-Gen data (the data) which include:

Title of your project: Evaluation of applicability of multivariant-based recall genotype for causal inference in African populations

Brief description of the project: Body mass index (BMI) is an important risk factor for cardiovascular disease. However, the mechanisms by which it contributes to cardiovascular diseases are poorly understood. Multivariant recall by genotype (RbG) studies offer an opportunity to unravel mechanistic insights into how BMI causes cardiovascular diseases. RbG studies have mostly been explored in European populations. The troublesome aspect is the poor performance of Europe-designed PRS in the African population. Studies are underway to determine how this poor performance of PRS in Africa can be improved but the use of these PRS in causal inferences has not been elucidated. Especially among the African populations who show regional variation of polygenic prediction that is thought to be a result of genetic and environmental diversity in Africa.
The aim of the study is to evaluate the effect of BMI PRS performance on RbG stratification and causal inference in three African regions.
This cross-sectional secondary data analysis study will compare the predictability and utility of four BMI PRS methodologies' ability to deal with confounding and utility for causal inferences. The AWIGen data set, which consists of 12 000 participants from three sub-Saharan regions, will be used for analysis. Data will be summarised using mean, standard deviation, median, interquartile range, percentage, linear regression (utilising RbG study design) and Forest plots. Statistical analysis will be performed on Stata ®17. Ethical approval or waiver will be sort for this study.
The outcome of the study can positively impact the design of intervention strategies to mitigate cardiovascular outcomes among African populations.

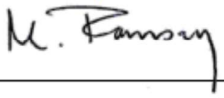
By signing this document, I Lebogang Mogongoa, agree to the following conditions:

1. I will use the Data and analyses only for the agreed purpose
2. I will only use the data for analysis on the Wits cluster and I will not copy the data to another location without prior written consent
3. I will not post the data publicly
4. I will not distribute the data to any third party
5. I will make no attempt to connect the data back to personal identifying information from any participants in the study
6. I agree that I will not publish any analysis or work based on the data without written permission from the AWI-Gen PI
7. I will not write and describe the study outcomes in such a way that will cause damage or distress, or that could be potentially stigmatizing to any data participants
8. **I will appropriately acknowledge all the funders in publications resulting from these data as instructed by the PI**
9. **Before submission I will share the manuscript with PI to ensure the attribution to the grant is accurate (applicable for AWI-Gen consortium members and students)**

10. I agree to provide an update on my progress to the supervisory group as requested
11. Unless otherwise approved, I will delete data and analyses (and in any other form) that may be stored on my personal computer on the occasion that I leave the AWI-Gen study
12. In accordance with the Wits Intellectual Property Policy, I understand that I have no claim to intellectual property emanating from the work I perform on these data

I understand and acknowledge that the data are experimental in nature, and that access to the data is provided without any representations or warranties of any kind in relation to the database or the data.

Should I have any questions or concerns I will contact the AWI-Gen Principal Investigator or project supervisors.

Data Recipient
Name: Lebogang Mogongoa Affiliation: MSc Epidemiology and Biostatic student, School of Public Health, Wits Email: 2485509@students.wits.ac.za Signature:  Date:23 May 2023
Supervisor 1 (if applicable)
Name: Dr Tinashe Chikorowe Affiliation: Sydney Brenner Institute for Molecular Bioscience (SBIMB) Email: tinashe.chikowore1@wits.ac.za Signature:  Date: 23 May 2023
Supervisor 2 (if applicable)
Name: Dr Juliana Kagura Affiliation: School of Public Health, Wits Email: juliana.kagura@wits.ac.za Signature:  Date: 24 May 2023
AWI-Gen Principal Investigator
Name: Michèle Ramsay Signature:  Date: 7 March 2024

Appendix 6: Ethics clearance certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research and Innovation)

TO: Professor M Ramsay, et al
Faculty of Health Sciences
Sydney Brenner Institute for Molecular Bioscience
Medical School
University

E-mail: Michele.Ramsay@wits.ac.za

CC: Supervisor: Not applicable
<>
and <HREC-Medical Research Office@wits.ac.za>

FROM: Mr Iain Burns
Human Research Ethics Committee (Medical)
Tel: 011 717 1252

E-mail: Iain.Burns@wits.ac.za

DATE: 2022/11/24

REF: R14/49

PROTOCOL NO: **M2210108** (This is your ethics application reference number. Please quote it in all enquiries, oral or written, relating to this study.)

PROJECT TITLE: *AWI-Gen Phases 1 and 2: Genomic and environmental risk factors for cardiometabolic disease in Africans*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to Government funding of the University.

MSWorks2000/Iain0007/Clearscan.wps

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R49 Professor M Ramsay, et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M2210108

NAME: (Principal Investigator)	Professor M Ramsay, et al
DEPARTMENT:	Faculty of Health Sciences Sydney Brenner Institute for Molecular Bioscience Medical School University
PROJECT TITLE:	*AWI-Gen Phases 1 and 2: Genomic and environmental risk factors for cardiometabolic disease in Africans*
DATE CONSIDERED:	Ad hoc
DECISION:	Approved unconditionally
CONDITIONS:	Subsumes M17/08/80 and M18/05/35
NOTE:	If contact information regarding student study participants is required, please contact the Registrar's office - <Nicoleen.Potgieter@wits.ac.za>
SUPERVISOR:	Not applicable
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2022/11/24

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **2022/10/01** and therefore reports and re-certification will be due in the month of **2022/10/01** each year. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

Signature of Principal Investigator _____ Date _____