The development and value assessment of an integrated cardiovascular disease risk score in an African setting

Michelle Kamp

Supervisors:

Prof. Michèle Ramsay

Prof. Cathryn Lewis

Dr Oliver Pain

A Thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy.

May 2024

NATIONAL HEALTH LABORATORY SERVICE

WITS UNIVERSITY

SDIMB SYDNEY BRENNER INSTITUTE FOR MOLECULAR BIOSCIENCE

DECLARATION

I, Michelle Kamp, declare that this Thesis is my own, unaided work, unless otherwise specified in the text. This thesis is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg, South Africa. It has not been submitted before for any degree or examination at this or any other university.

Michelle Kamp

Signed on the 18th of May 2024

To all those who have supported me on this journey, especially my husband and son, and in memory of my dearest friend Zannie and my father, Francois.

PRESENTATIONS ARISING FROM THIS THESIS

SCIENTIFIC SEMINARS

1. Kamp, M. *Clinicians' Perceptions Towards Precision Medicine Tools for CVD Risk Stratification in South Africa*. Division of Human Genetics' Seminar series, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, 31 August 2022 (virtual).
2. Kamp, M. *The development of an integrated risk score for Cardiovascular diseases in African populations.* Social, Genetic & Developmental Psychiatry Centre (SGDP) Seminar series, SGDP, King's College London, London, UK, 27 April 2023 (in-person).
3. Kamp, M. *The development of an integrated risk score for Cardiovascular diseases in African* populations. Medical and Molecular Genetics Seminar series, King's College London, London, UK, 21 June 2023 (in-person).

CONFERENCE PROCEEDINGS

1. Kamp, M., Pain, O., Lewis, C.M., and Ramsay, M. *Study design: The development and value assessment of a precision medicine-based Cardiovascular disease risk score.* Poster presentation at the London Genetic Network Annual Meeting, London, UK (virtual), 03 December 2021.
2. Kamp, M., Pain, O., May, A., Lewis, C.M., and Ramsay, M. *Clinicians' Perceptions Towards Precision Medicine Tools for CVD Risk Stratification in South Africa*. Poster presentation at the European Society of Human Genetic Conference, Vienna, Austria (in-person), 14 June 2022.
3. Kamp, M., Pain, O., May, A., Lewis, C.M., and Ramsay, M. *Clinicians' Perceptions Towards Precision Medicine Tools for CVD Risk Stratification in South Africa*. Poster presentation at the Annual Wits Faculty of Health Sciences Research Day, Johannesburg, South Africa (in-person), 15 September 2022.
4. Kamp, M., Pain, O., Lewis, C.M., and Ramsay, M. *The development and value assessment of a precision medicine-based Cardiovascular disease risk score in a South African setting*. Oral presentation at the 14th International Congress of

Human Genetics (ICHG), Cape Town, South Africa (in-person), 23-26 February 2023.

5. Kamp, M., Pain, O., May, A., Lewis, C.M., and Ramsay, M. *Clinicians' Perceptions Towards Precision Medicine Tools for CVD Risk Stratification in South Africa.* Poster presentation at the 7th Global Genomic Medicine Conference (G2MC), Geneva, Switzerland (in-person), 2-4 October 2023.
6. Kamp, M., Pain, O., Lewis, C.M., and Ramsay, M. The development and value assessment of a precision medicine-based Cardiovascular disease risk score in a South African setting. Oral presentation at the 7th Global Genomic Medicine Conference (G2MC), , Geneva, Switzerland (in-person), 2-4 October 2023.
7. Kamp, M., Achilonu, O., Kisiangani, I., Nderitu, DM., Mpangase, PT., Tadesse, G.A., Adetunji, K., Iddi, S., Speakman, S., Hazelhurst, S., Asiki, G. and Ramsay, M. *Multimorbidity in African populations: a scoping review.* Poster presentation at the 3rd Data Science for Health Discovery and Innovation in Africa (DS-I Africa) Consortium meeting, Kigali, Rwanda (in-person), 3-9 November 2023.

PUBLICATIONS ARISING FROM THIS THESIS

The research findings from this PhD study were written up as four separate manuscripts and submitted to international peer-reviewed scientific journals. Three manuscripts have been accepted and published by the respective journals, and one manuscript is currently under review. The publications are listed below, and the student's contribution to each publication follows. The signed co-author agreement for the use of each publication within this thesis is in Appendix A.

Published papers:

1. **Kamp, M**., Krause, A., and Ramsay, M., 2021. Has translational genomics come of age in Africa?, *Human Molecular Genetics.* 30(R2): R164–R173, https://doi.org/10.1093/hmg/ddab180.
2. **Kamp M.**, Pain O, May A, Lewis CM, Ramsay M. 2022. Clinicians' Perceptions towards Precision Medicine Tools for Cardiovascular Disease Risk Stratification in South Africa. *Journal of Personalized Medicine*. 12(9):1360. https://doi.org/10.3390/jpm12091360
3. **Kamp M**., Achilonu, O., Kisiangani, I., Nderitu, DM., Mpangase, PT., Tadesse, G.A., Adetunji, K., Iddi, S., Speakman, S., Hazelhurst, S., Asiki, G. and Ramsay, M. as members of the MADIVA Research Hub and the DS-I Africa Consortium. 2023. Multimorbidity in African ancestry populations: a scoping review. *BMJ Global Health.* 8:e013509. http://dx.doi.org/10.1136/bmjgh-2023-013509

Papers under review:

4. **Kamp, M**., Pain, O., Lewis C.M., and Ramsay, M., 2023. Combining ancestry-aligned polygenic risk scores with conventional risk factors improves cardiovascular disease risk stratification in African populations. *Genome Medicine.*

PhD student's contribution to each publication:

In each publication, in partnership with supervisors, the student was responsible for the study design and conceptualisation of each sub-study, as well as project coordination, including data collection (where appropriate) and data management. The student performed data cleaning, statistical analyses, interpretation of results, and the

writing and revision of each manuscript. Co-authors provided methodological advice, coding support, statistical assistance, and critically reviewed the manuscripts.

GRANTS AND AWARDS

- **2020**: Recipient of a PhD fellowship for PhD fees and stipend through the from the NIH AWI-Gen grant held by Prof. Michèle Ramsay.
- **2020**: Recipient of the Postgraduate Merit Award for PhD fees and stipend through the University of the Witwatersrand.
- **2021**: Recipient of the Postgraduate Merit Award for PhD fees and stipend through the University of the Witwatersrand.
- **2022**: Recipient of the Postgraduate Merit Award for PhD fees and stipend through the University of the Witwatersrand.
- **2022**: Recipient of a University of the Witwatersrand Faculty Research Committee (FRC) Individual Research Grant (R3 000).
- **2022 and 2023**: Recipient of the Phillip Vallentine Tobias Educational Bursary through the University of the Witwatersrand.
- **2020**: Recipient of a PhD fellowship for PhD fees and stipend through the NIH MADIVA Grant held by Prof. Scott Hazelhurst.
- **2023**: Recipient of the NIH Travel Fellowship to attend the International Congress of Human Genetics (ICHG) 2023.
- **2023**: Recipient of the DS-I Africa Travel Fellowship to attend the 3rd Data Science for Health Discovery and Innovation in Africa (DS-I Africa) Consortium meeting 2023.

ABSTRACT

Cardiovascular diseases (CVD) are a significant health threat in Africa and are a leading cause of mortality and morbidity in the region. Risk stratification is the preferred approach to disease prevention but is challenging to apply due to scarce data and current scores not being validated in African populations. CVD risk is influenced by genetic and environmental factors, with the latter mostly considered in risk calculators. Precision medicine (PM) aims to tailor medical diagnostics and therapeutics by leveraging the genetics, environment, and lifestyle of individual patients. Polygenic scores (PGS) can quantify an individual's genetic burden for CVD and improve the predictive value of risk tools when included with conventional risk factors. PGS for European ancestry populations are reaching the point where they may be useful in clinical care. However, transferability of European-derived PGS to African populations is limited and may hamper application in clinical practice.

The overarching aim of the study was to develop and evaluate an integrated CVD risk score incorporating both genetic and non-genetic factors for continental African populations. This comprised of developing ancestry-aligned PGS for various cardiometabolic traits, and then deriving, evaluating, and comparing the predictive utility of genetic, non-genetic, and integrated (genetic + non-genetic) CVD risk prediction models using elastic net regression with nested 10-fold cross validation. Furthermore, we aimed to contextualise the potential utility of such PM-based tools in Africa through an assessment of the current landscape of translational genomics on the continent and by gauging clinician's perceptions on implementing the derived CVD-risk stratification tool in South Africa's public health setting.

This study demonstrated the potential of genetic information to enhance disease risk stratification among continental African populations. Results revealed PGS provide both independent and complementary information in predicting dyslipidaemia, hypertension, and obesity - integrated scores improved prediction by at least 2.5% compared to models consisting of non-genetic factors alone. The study also highlights the challenges limiting the advancement of PM in Africa – 1. the paucity of genetic and phenotypic data within continental African populations limits the development and validation of robust and clinically useful models. This challenge is exacerbated by

Euro-centric genomic and computational tools and echoes the call for more inclusive methodological approaches. 2. Despite positive perceptions of PM, there is an urgent need to address complex structural and operational barriers, including insufficient funding, lack of political will, healthcare infrastructure deficits, and training and education gaps.

ACKNOWLEDGEMENTS

Many people have supported me on my PhD journey, and whom without, I would not have reached this milestone. My heartfelt thanks and appreciation go out to each and every one of you.

- My supervisors – Professor Michèle Ramsay, Professor Cathryn Lewis, and Dr Oliver Pain. Your mentorship, insights, and unwavering support have been invaluable. Your combined knowledge and wisdom have greatly enriched my journey and inspired me to continue in academia. Thank you for your time, patience, and willingness to walk this path with me.
- My family and friends – I am so grateful for your continuous support and encouragement throughout the highs and lows of my PhD. Thank you to my amazing husband and son for believing in me. Your patience and support mean the world to me.
- My colleagues at the SBIMB, University of the Witwatersrand, and at the SGDP, King's College London, who were always willing to lend a helping hand and provide encouragement and motivation when needed. A special thank you to David Twesigomwe, Andrew May and Helena Davies for the many laughs and wonderful discussions about the joy and complexities of genetics.
- The projects office at the SBIMB, specifically Jocelyne Gayenga, Kerry Glover and Furahini Tluway, who have assisted with all the administration associated with the larger AWI-Gen and MADIVA projects, my fellowship, travel, and tuition. Your support and seamless organisation has allowed me to focus on my studies.
- Research participants – Without the individuals who participated in APHRC and AWI-Gen, this research would not have been possible. I want to thank you for your time and willingness to share your data. Thank you to everyone involved in the management of these datasets. A heartfelt appreciation goes to the clinicians who completed the survey about implementing precision medicine in South Africa despite their significant time constraints and the challenges posed by COVID-19.
- My funders – The National Institute of Health (NIH), specifically through the AWI-Gen (U54HG006938 and U24HG009780) and MADIVA (U54TW012077) Grants, and the National Research Foundation (NRF) South African Research Chair

Initiative (SARChI) for funding my research and tuition. Thank you to the Philip Tobias Educational Trust for their financial support while completing my degree.

STATEMENT OF AUTHORSHIP

The work presented in this thesis was conducted and written by Michelle Kamp, except for data collection, as noted below:

- Individual-level data from the AWI-Gen study was used in Section 2: Chapter 2. Other researchers performed data cleaning and preliminary quality control of the genetic data linked to the AWI-Gen project without my involvement.

- Summary statistics from the African Partnership for Chronic Disease Research, specifically the meta-analysis described by Gurdasani et al. (2019), and which includes the Uganda Genome Resource (UGR), the Africa America Diabetes Mellitus Study (AADM), the Durban Case-Control study (DCC), and the Durban Diabetes Study (DDS). These association statistics were used to derive the polygenic risk score and subsequent integrated risk score described in Section 2: Chapter 2. The data collection, GWAS, and associated analysis was performed without my involvement.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AADM	Africa America Diabetes Mellitus Study
AFR	African
AfSHG	African Society for Human Genetics
AIDS	Acquired immune deficiency syndrome
ASSIGN	Assessing Cardiovascular Risk Using SIGN Guidelines
AWI-Gen	Africa Wits-INDEPTH partnership for Genomic Studies
BMI	Body mass index
BP	Blood pressure
CAD	Coronary Artery Disease
CI	Confidence interval
COVID-19	Coronavirus disease 2019
CVDs	Cardiovascular Diseases
DALY	Disability-adjusted life year
DBP	Diastolic blood pressure
DCC	Durban Case Control
DDS	Durban Diabetes Study
DLD	Dyslipidaemia
DNA	Deoxyribonucleic acid
EAS	East Asian
EUR	European
GBD	Global Burden of
GPC	General Population Cohort
GRCh37	Genome Reference Consortium Human Build 37

GRCh38	Genome Reference Consortium Human Build 38
GWAS	Genome Wide Association Studies
HA	Heart attack
H3Africa	Human Heredity & Health in Africa
h^2	Heritability
h^2_{SNP}	SNP-based heritability
HDL	High-density lipoprotein
HIV	Human immunodeficiency virus
HTN	Hypertension
INDEPTH	International Network for the Demographic Evaluation of Populations and Their Health
L1	LASSO regularisation
L2	Ridge regularisation
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LMICs	Low- and middle-income countries
MAF	Minor allele frequency
MCS	Mean confidence score
MKS	Mean knowledge score
MM	Multimorbidity
MPS	Mean perception score
MR	Mendelian randomisation
MRC	Medical Research Council
N, n	Absolute number
NCDs	Non-communicable diseases

PCE	Pooled cohort equations (PCE) for estimation of atherosclerotic cardiovascular disease
PC	Principal component
PCA	Principal component analysis
PGS	Polygenic score
PM	Precision medicine
PPC	Projected principal components
PPH	Precision Public Health
PRS	Polygenic risk score
pT+ clump	LD clumping plus p-value thresholding approach
P value	Probability value
REDCap	Research Electronic Data Capture
RMSE	Root mean square error
QC	Quality control
QRISK	10-year risk estimate for CVD in women and men developed in the UK
SBP	Systolic blood pressure
SD	Standard deviation
SE	Standard Error
SNP	Single nucleotide polymorphism
T2D	Type 2 diabetes
TB	Tuberculosis
TC	Total cholesterol
TG	Triglycerides
UGR	Uganda Genome Resource
UKBB	United Kingdom Biobank

UVRI	Uganda Virus Research Institute
WGS	Whole genome sequencing
WHO	World Health Organization
Wits	University of the Witwatersrand

Units and symbols

The International System of Units (SI) has been used throughout this thesis.

%	Percentage
>	Greater than
<	Smaller than
≥	Greater than or equal to
±	Plus/minus
α	Alpha
β	Beta
$	United States of America Dollar
cm	centimetre
g	gram
hr	hour
kg	kilogram
kg/m2	kilogram per square metre
km2	square kilometre
mmol/l	millimoles per litre
mmHg	millimetre of mercury
R^2	Model fit
ZAR	South African Rand

PREFACE

My journey to a PhD in Human Genetics has been non-linear through diverse corporate and academic roles. A growing curiosity and respect for the rigorous nature of research drew me back to academia. My interest in genomics, especially polygenic risk scores, is due to the promise of genomic medicine and its potential health impact for all. I am particularly interested in health equity and have aimed to consider the implementation of genomic advances where resources are scarce. I am inspired by trailblazing women scientists, such as Prof. Michèle Ramsay, who champions African genomics on a global stage, and hope to contribute to this in some way.

This PhD thesis is presented in the 'integrated' format', which includes four manuscripts. Three manuscripts have been accepted for publication in the respective journals, and one is currently under review. The thesis is divided into three sections that are further divided into additional chapters as outlined below:

- **Section 1: Background and Study Context**

  This section describes the literature around the topic, the aims and objectives of the study, and the approach taken to conduct the research.
  - **Chapter 1**: Introduction
  - **Chapter 2**: Methodological approach

- **Section 2: Publications**

  This section consists of four publications conducted during this PhD study.
  - **Chapter 1**: Has translational genomics come of age in Africa?
  - **Chapter 2**: Combining ancestry-aligned polygenic risk scores with conventional risk factors improves cardiovascular disease risk stratification in African populations.
  - **Chapter 3**: Clinicians' perceptions towards Precision Medicine tools for Cardiovascular disease risk stratification in South Africa.
  - **Chapter 4**: Multimorbidity in African ancestry populations: a scoping review.

  Section 3: Integrated Discussion

This section consists of an integrated discussion that consolidates the findings in all four manuscripts and ends with an overall conclusion.

- **Chapter 1**: Publication summaries
- **Chapter 2**: Key themes emanating from this thesis
- **Chapter 3**: Strengths, limitations, and future directions
- **Chapter 4**: Overarching conclusion

.

SECTION 1: BACKGROUND AND STUDY CONTEXT

Section Overview

This chapter provides the background to the study and key methodological approaches used.

Chapter 1 aims to provide an overview of the literature across key project domains, specifically, precision medicine in Africa, cardiovascular diseases - and the potential of precision medicine-driven care such as enhanced risk stratification through polygenic risk scores, and the clinical translation of such approaches in an African setting (**Figure 1**). In addition, given that cardiovascular diseases rarely occur in isolation in Africa due to the substantial and overlapping chronic infectious disease burden, this chapter introduces the concept of multimorbidity and suggests that effective precision medicine will also require the consideration of concomitant diseases and their impact on diagnosis, prognosis, and treatment. Finally, the motivation behind the study, the overarching aim, and the specific objectives are described.

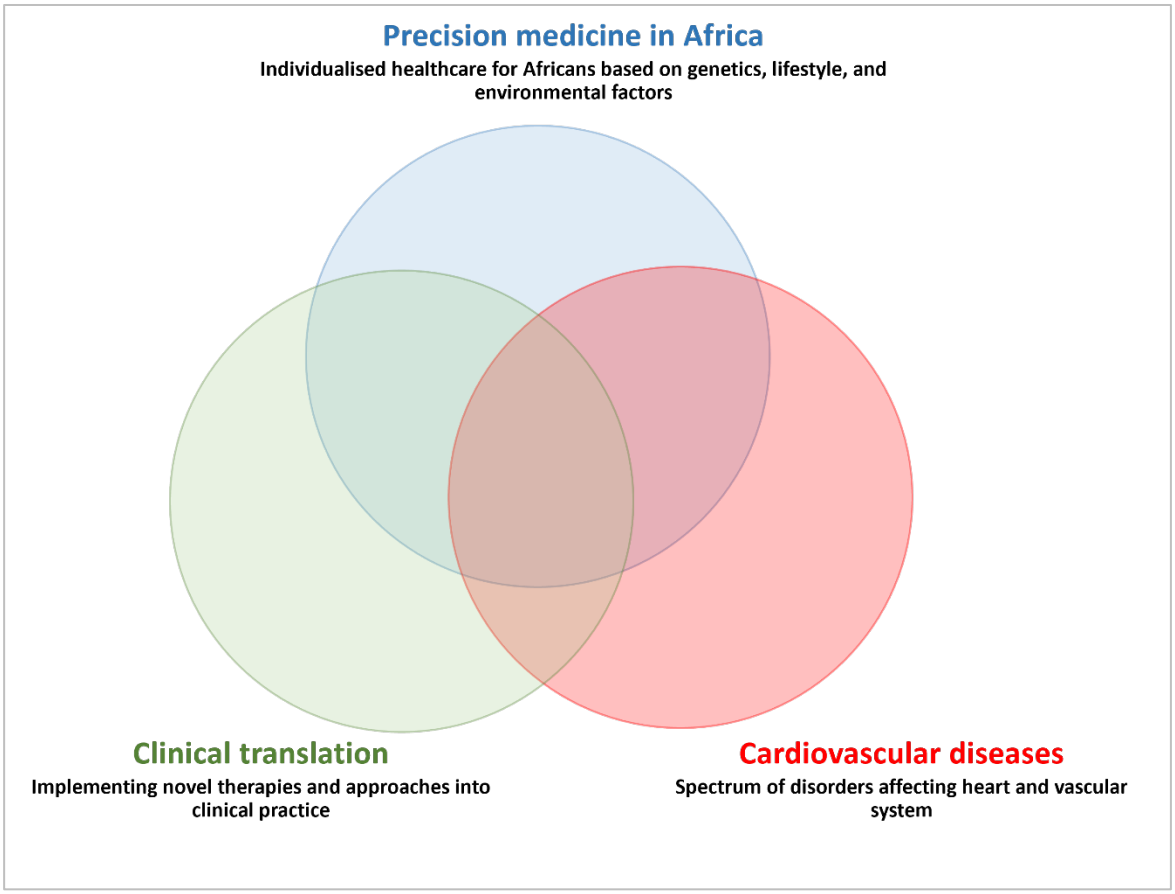


Figure 1: The three key domains of interest in this body of work - Precision medicine, Cardiovascular diseases, and Clinical translation

Chapter 2 describes the study setting, sample, and methodological approaches used to achieve each objective. This chapter aims to provide details on specific methodology and statistical analyses described in the publications (Section Two). Given their complexity, this chapter will focus on polygenic and integrated score development methodology and statistical approaches. Although described in more detail, there will be overlap with the methods described in the draft manuscript (Section 2: Publication 2). The methods used in the clinical utility assessment and multimorbidity scoping review are sufficiently described in their respective publications (Section 2: Publication 3 and 4).

Chapter 1: INTRODUCTION

1.1. Precision Medicine

Precision medicine (PM) is a healthcare model that challenges the prevailing 'one-size-fits-all' care and treatment approach. PM aims to tailor medical diagnostics and therapeutics by leveraging the genetics, environment, and lifestyle of individual patients. It aims to improve clinical decision-making by distinguishing patients more likely to benefit from a specific treatment from those who might not respond or suffer side effects. As a result, healthcare delivery efficiency could be increased; encompassing improved patient treatments, better resource allocation and cost management, and a reduction in the number of return visits (1).

The tools by which PM seeks to achieve its goals are '*omics'* - genomics, epigenomics, transcriptomics, proteomics and metabolomics - large-scale biological information which seek to understand the structure and function of biological organisms, artificial intelligence, machine learning (ML), environmental, social, and behavioural factors and their integration with preventive and public health (2).

Advances in genomics, including large-scale biobanks, advanced genome-based technologies, and powerful computational tools that provide insight into biological mechanisms and drug targets (3) have contributed to PM approaches being increasingly adopted. One such example is Canrisk, the extension of the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA)(4). The updated algorithm includes genetic factors to systematically identify individuals at highest and lowest risk of breast cancer and is routinely used in the UK and other European settings (5). In addition to lifestyle (e.g. alcohol intake), hormonal (e.g. oral contraceptive use), reproductive (e.g. age at menarche), anthropometric (e.g. height and body mass index (BMI)) and mammographic density risk factors, CanRisk now incorporates an assessment of genetic risk components, specifically through the effects of truncating variants in *BRCA1, BRCA2, PALB2, CHEK2*, and *ATM* genes; a polygenic risk score (*described in further detail in Section 1.4.2*) consisting of 313 single-nucleotide polymorphisms (SNPs), and a residual polygenic component accounting for other genetic/familial effects (6).

Risk prediction is central to PM and clinical decision-making, providing the ability to stratify a population into distinct risk categories to optimise the risk–benefit balance of public health or clinical interventions (7), and is a key approach which can be applied in the identification and treatment of patients with cardiovascular diseases (CVD).

Subsequent sections detail the current status of CVD in Africa, the prevention and treatment strategies employed, and the advancements in CVD genetics-associated methodologies for the development of PM-based tools for CVD in African populations.

1.2. Cardiovascular diseases

Cardiovascular diseases (CVDs) are a set of heterogeneous diseases with different but overlapping aetiologies affecting the heart and blood vessels, including coronary artery disease (CAD), stroke, cardiomyopathy, and congenital heart disease. Most CVD fatalities, among men and women, arise from CAD and stroke (8). Atherosclerosis – the accumulation of fatty deposits in the arteries, is a primary cause. Common risk factors include older age, female sex, sedentary behaviour, and intermediary conditions, such as diabetes, hypercholesterolaemia, hypertension, and obesity (9).

CVDs, the leading non-communicable diseases worldwide, accounted for an estimated 17.8 million deaths in 2017, which is expected to reach 22.2 million deaths in 2030 (10). Contrary to historical perceptions, CVDs are no longer limited to affluent nations, with over 75% of deaths occurring in low- and middle-income regions, such as Africa (11), due to population-wide demographic, social, and economic shifts (12), and the associated rise in diabetes and hypertension. South Africans are at particularly high risk as they have some of the highest prevalence of obesity, hypertension, undiagnosed diabetes mellitus, smoking and high cholesterol on the continent (13). These risk factors for CVD also constitute diseases in their own right; hypertension, dyslipidaemia, obesity, and abnormal glucose levels are disorders that require medical intervention. These precursor conditions accelerate pathologies in the vascular, cerebral, and myocardial systems. Over time, these pathologies can evolve into serious complications, including strokes, myocardial infarctions, and kidney diseases (14).

CVDs have diverse symptoms, age of onset, and prevalence characteristics due to their complex and multifactorial aetiology, stemming from a combination of genetic

susceptibility, environmental exposures, and lifestyle factors (15). Key CVD risk factors are split into modifiable (smoking, alcohol, high blood pressure, obesity, poor diet, and physical inactivity) and non-modifiable factors (age, sex, and genetics) (16).

Inadequate preventative screening, delayed diagnosis, and limited access to effective and equitable healthcare services are exacerbating the CVD epidemic in Africa, which in turn, impedes social development. Premature deaths during prime working years limit economic contributions and the potential to uplift households (17). In 2008, South Africa lost 117,963 potentially productive years of life (PPYLL) due to CVDs among the population aged 35–64 years, and this is projected to reach 199,964 in 2050. Although years lost in South Africa are not as high as other middle income countries given population size (Brazil: 741,065; China: 2,247,802; India: 11,920,506; Russia: 1,362,436), the country will see the second greatest increase in PPYLL between the two time periods, 70% (Brazil: 59%; China: 17%; India: 108%; Russia: -19%), and PPYLL are far higher than in Portugal (10,052 in 2050; -16% between 2008 and 2050), the country with lowest CVD death rates in Europe (14). CVDs place a heavy financial burden on households and the economy, with out-of-pocket expenditure common (9,17), and CVDs account for a significant proportion of national health expenditure. Recent estimates of the latter are not readily available, but approximately 25% of South African healthcare expenditure was devoted to the direct treatment of CVD in 1991 (18).

CVD prevention and control requires a two-pronged approach involving both population-wide and individual interventions (17). Public measures include the taxation of foods containing high levels of fat, sugar, or salt, i.e., legislating the reduction of these ingredients in purchased foodstuffs, and implementing tobacco and alcohol control initiatives. Personal interventions focus on health screening, medication, and lifestyle modifications, but these need to be targeted at those with a high CVD risk (17). While South Africa has adopted population-wide approaches - it was the first country to introduce salt reduction legislation (19) and passed the sugary beverage taxation (20) - personalised interventions based on CVD risk stratification remain challenging to implement as data on the genetic and environmental risk factors of CVD in African populations is scarce.

1.2.1. Cardiovascular disease risk stratification

Globally, assessing CVD risk is complex and requires an integrated approach involving multiple risk factors to predict the likelihood of experiencing a CVD event within a specified timeframe (21). Almost all CVD guidelines recommend risk-scoring strategies for clinical decision-making, cost-effective planning of primary prevention interventions, and reduction of absolute risk through behaviour change and medications, such as statins (22). The practice is well-established in most high-income countries and endorsed by the WHO for low- and middle-income countries (21).

Numerous CVD risk prediction models have been developed, including the Framingham 10-Year Risk of General Cardiovascular Disease score, The Pooled Cohort Equation for Atherosclerotic Cardiovascular Disease (PCE), and QRISK3, (Appendix D) (16) to estimates the 10-year risk of developing CVD. The Framingham score and its variations are most widely used and endorsed by various international and local committees, including those in South Africa (e.g., South African Heart Association, Lipid and Atherosclerosis Society of Southern Africa, Southern African Hypertension Society and National Department of Health) (23–25).

The Framingham algorithm was developed based on data from the Framingham Heart Study, a long-term, ongoing cardiovascular cohort study of residents of the town of Framingham, Massachusetts, USA. The score integrates age, gender (meaning biological sex in this context), smoking status, blood pressure, and cholesterol levels, utilising sex-specific Cox proportional-hazards regressions to gauge CVD risk (see Appendix E for sex-specific weightings). It applies to men and women with no prior history of CVD (26). For resource-limited settings, where blood lipid assessments are costly, the Framingham model has been successfully modified to the 'Framingham's non-laboratory test' by substituting cholesterol with body mass index (BMI) (27).

The Framingham 10-Year Risk of General Cardiovascular Disease score formula is as follows:

$$\hat{p} = 1 - S_0(t)^{\exp\left(\sum_{i=1}^{p} ßiXi - \sum_{i=1}^{p} ß\overline{\iota X}i\right)}$$

Where $S_0(t)$ represents baseline survival at follow-up time t, βi denotes the estimated regression coefficient, Xi signifies the log-transformed value of the i^{th} risk factor (if continuous), $\overline{Xi}$ corresponds to the respective mean, and p represents the number of risk factors. The factors by which each risk factor is multiplied (by sex) are detailed in Appendix D.

The model's authors acknowledge several limitations to the model, including its development using a North American-based data set, and primarily with European-ancestry participants. Cultural and lifestyle risk factors differ worldwide and those identified in this US-based dataset may not apply across diverse global populations Also, other risk factors, such as abdominal obesity, indications of insulin resistance, triglycerides, and a strong family history of premature CVD that may be more predictive of CVD in both US-based populations and others are excluded (26).

The Pooled Cohort Equation for Atherosclerotic Cardiovascular Disease (PCE), developed in 2013 but not widely implemented in Africa, aims to address Framingham's shortcomings in estimating risk among ethnic minority groups (22). The PCE model was developed by the American College of Cardiology/American Heart Association (ACC/AHA) using pooled data from several large, diverse, community-based cohorts in the United States, and thus more ethnically diverse. This model includes the same risk factors as the Framingham Risk Score, albeit with different risk factor weightings to reflect the different populations and methodologies used in model development, but also considers ethnicity, and has separate equations for European and African–American groups (28). Similarly, QRISK3 is a risk calculation tool developed in the UK to assess the risk of CVD among those living in the UK. It includes risk factors like age, cholesterol levels, and blood pressure, as well as additional factors such as ethnicity, Body Mass Index (BMI), family history of CVD, socioeconomic status, and medical conditions like rheumatoid arthritis, chronic kidney disease, and diabetes (29).

While the applicability of Framingham, QRISK3, and PCE models to sub-Saharan Africa remains largely unexplored - meaning that they have not been validated nor calibrated for African populations, a study examined the concordance between Framingham's laboratory, Framingham's non-laboratory, and PCE equations among Ghanaian migrant and resident populations without evident CVD and identified

significant differences. Results indicated that CVD risk prediction using the same algorithm differed between migrant and resident populations, that there was a lack of concordance in CVD predictions by different risk algorithms, and despite being adapted for resource-poor settings, the interchangeability of Framingham's laboratory and non-laboratory algorithms within these populations was limited (22). Similarly, another study assessing CVD risk across different African regional populations noted discordance in predictions by six different risk algorithms (30)

A systematic review of risk prediction models revealed that there are approximately 363 different models, with most relying on similar risk factors. However, most are of limited usefulness due to their lack of external validation (28). The same review showed that less than 1% of these models include genetic data in determining CVD risk despite growing evidence that individual genetic variation influences predisposition to complex diseases (28).

Inadequate risk-based management may lead to overtreatment or undertreatment, hampering the effective prevention of CVD and access to appropriate care. Thus, it is imperative to explore a robust and validated CVD risk model that encompasses clinically relevant risk factors, is tailored to local populations, and incorporates individual genetic variability (15,28).

1.2.2. Cardiovascular disease genetics

Heritability (h^2) quantifies the relative contribution of genetic susceptibility to disease predisposition in a population. Defined as the proportion of phenotypic variation in a population that can be explained by genetic variation, heritability is governed by the underlying genetic architecture of the disease of interest (15).

The genetic architecture of disease encompasses the number and frequency of genetic variants influencing disease risk, the magnitude of their influence on disease predisposition, and the prevalence of the disease in the general population (15). Common chronic diseases rarely follow a single-gene Mendelian inheritance pattern. In fact, in many chronic diseases, only a small fraction of individuals with the disease phenotype (approximately 1–10%) result from rare (minor allele frequency (MAF) <0.5%) high-risk pathogenic genetic variants (15). Examples include monogenic forms of obesity (leptin deficiency) and a dominant predisposition to breast and ovarian

cancer (*BRCA1* and *BRCA2* pathogenic variants). The heritability of common adult-onset diseases is conferred by numerous common (MAF >5%) and low-frequency (MAF between 0.5% and 5%) genetic variants that individually contribute small effects (7,31).

Genetic factors contributing to CVD risk and associated phenotypes range from rare high penetrance variants responsible for inherited cardiovascular conditions (e.g., familial hypercholesterolemia) to common low-penetrance single nucleotide polymorphisms (SNPs) that modulate risk of CVD in those with common forms of disease. A SNP is the simultaneous occurrence, at the same locus, of two or more distinct alleles where the frequency of the least common allele is greater than or equal to 1%. There are several genetic tests that inform diagnosis or guide treatment. These tests are largely focused on few variants with high effect. For example, a genetic testing panel including genes such as *KCNQ1, KCNH2*, and *SCN5A* is used to diagnose Long QT Syndrome (32). Also, screening for mutations in the *TITIN* gene aids in confirming a diagnosis of dilated cardiomyopathy (33).

Family- and twin studies estimate the twin heritability of CAD at between 40 and 60% (34,35) and the SNP heritability of stroke between 30 and 40% (36), indicating a substantial genetic predisposition to the diseases. However, given the complex nature of these diseases, identifying robust genetic associations remains an ongoing pursuit. Nevertheless, through the concurrent establishment of large collaborative consortia, such as CARDIoGRAM (Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis) (37) and the Coronary Artery Disease (C4D) Genetics consortium (38), novel technologies, and approaches, such as genome-wide association studies (GWAS), and improved statistical approaches, many loci have been identified, and have provided invaluable insights into biological processes contributing to disease aetiology.

1.2.2.1. Genome-wide association studies (GWAS)

GWAS have contributed to successfully identifying genetic variants associated with CVD phenotypes. GWAS are large-scale, hypothesis-free genetic analyses that explore the relationships between sites of common genome sequence variation and disease predisposition on a genome-wide scale (39). A GWAS identifies differences in the frequency of SNPs across the genome between cases, such as those with coronary

artery disease (CAD), and control groups. In 2007, the first GWAS focusing on CAD identified a disease-associated locus on chromosome 9p21 *(CDKN2A, CDKN2B).* This locus has been replicated across multiple cohorts and is considered one of the strongest genetic risk factors for CAD. Saying that, the effect size of this variant is still low compared to traditional risk factors and cannot be used alone for prediction. A recent GWAS consisting of 181,522 cases among 1,165,690 participants identified 241 independent associations at 198 loci, including *PCSK9*, *APOE*, and *LDLR*, which are related to lipid metabolism, inflammation, blood pressure regulation, and vascular function (40).

Genetic association studies for type 2 diabetes mellitus (41,42) and CAD (43,44) indicate that common variants with small effect sizes explain most of the disease heritability. Little to no evidence of low-frequency variants with moderate or larger effects (odds ratio >1.5) were uncovered despite the studies being well-powered to detect such associations (15). Despite this, all the identified loci together still only explain a small fraction of disease heritability (CAD SNP-based h^2 = ~ 10%, stroke h^2 = ~ 10%, diabetes h^2 = ~ 10%). One reason for this is that common diseases are highly polygenic, and genetic variants with small effect sizes act cumulatively to increase risk. Simultaneously investigating thousands of SNPs, which individually do not reach statistical significance in a GWAS, could account for a greater proportion of the heritability (45). Polygenic risk scoring, described in more detail below, tests the predictive power of such an ensemble of markers.

1.2.2.2. Polygenic risk scores (PGS)

Polygenic risk scores (PRSs), or polygenic scores (PGS), are calculated at the level of the individual and therefore are ascribed potential utility for ‘precision’ medicine. PGS summarise genome-wide disease-associated genotype data into a single value estimate of an individual's total genetic risk (or burden) for a disease over multiple susceptibility variants. They are an attractive biomarker, as they are easy to calculate and store, stable across an individual's lifetime, and provide a ‘baseline risk’ on which external influences act to modulate disease and disease susceptibility (15,46).

PGS are calculated as the summation of the number of risk alleles an individual carries, weighted by corresponding genotype effect size estimates derived from summary statistic GWAS data. Analyses can be characterised by the two input data sets that

they require: (1) base GWAS data, specifically, the summary statistics (e.g., for each associated SNP, the effect size (betas for continuous traits or odds ratios for dichotomous (i.e. case-control) studies), significance of the association of genotype-phenotype associations (p-values), and (2) target data: the genotypes and phenotype(s) of individuals in the target sample. Critical considerations when deriving PGS is which SNPs to select for the score, and what, if any, shrinkage to apply to the GWAS effect size estimates (47).

The Clumping and Thresholding (C+T) technique is the most common method for computing PGS. The clumping step gathers SNPs that are in linkage disequilibrium (LD) within a specific genomic window, prioritising the most significantly associated SNP for PGS construction. LD refers to when alleles at different loci are associated with each other more often than chance. This non-random association, typically due to their proximity on a chromosome, results in alleles at two loci being inherited together (48). LD across a region implies that association will be detected at several SNPs and does not imply independent genetic contributions to the disease. Adjustment for LD is essential to avoid double-counting an association at a specific locus.

Thresholding selects SNPs based on their association p-values. Any SNPs with p-values below a fixed threshold, which is more liberal than genome-wide significance, are included in the PGS calculation, and those above the threshold are excluded. This methodological choice was influenced by the observation that only a fraction of SNP-based heritability is captured by SNPs achieving conventional GWAS significance ($p < 5x10^{-8}$). As such, thresholding is used to identify potentially relevant SNPs with smaller effect sizes that might be overlooked in typical GWAS. To find the best association with the phenotype, a range of thresholds is typically evaluated. While C+T offers an easy-to-understand approach, it may not leverage all the available information (47).

Additional methods dealing with LD include Bayesian inference and penalised regression and have shown enhanced PGS prediction (49,50). Unlike (C+T) which accounts for LD through assessment of SNP-SNP correlations, the Bayesian inference and penalised regression usually shrinks, or downweighs, the effect sizes of individual SNPs before their inclusion into a PRS, taking the local strength of LD into account (51). Simulation studies by Pain et al. (2021) (52) evaluated the predictive performance

of eight PGS methods, including p-value thresholding and clumping (pT+ clump), SBLUP, lassosum, LDpred1, LDpred2, PRScs, DBSLMM and SBayesR in the UK Biobank and the Twins Early Development Study (TEDS); specifically. LDpred2, lassosum, and PRScs were highlighted as top performers when using 10-fold cross-validation, surpassing pT+clump by 16-18% in terms of correlating observed and predicted outcomes. Elastic net models that integrated multiple polygenic scores from a range of parameters consistently enhanced predictions (52).

Genome-wide PGS for many traits have substantially greater predictive power when they include SNPs associated at a higher p-value than using a small number of genome-wide significant SNPs (47). PGS for CAD, atrial fibrillation, type 2 diabetes, inflammatory bowel disease, and breast cancer identified 8.0, 6.1, 3.5, 3.2, and 1.5% of individuals in the study population, respectively, at greater than threefold increased risk. For CAD, this 8% of the population is 20-fold higher than the carrier frequency of rare monogenic mutations, which confer comparable risk. These results suggest that PGS can identify people at elevated risk similar to that conferred by rare variants (53). However, the utility of PGS is limited and more predictive models would include other aspects of the disease causality, including environmental factors, which may not only be important determinants of risk but could also be potentially modifiable through changes in lifestyle or other interventions (7). Using CAD as an example, PGS have shown better prediction of CAD outcomes than any single traditional risk factor alone (54).

1.2.2.3. Integrated risk scores

An integrated risk score combines the assessment of conventional risk factors (e.g. age, blood pressure, cholesterol levels, smoking status) with PGS, which are derived from genetic information, to provide a more comprehensive evaluation of an individual's risk for developing a certain disease or condition. Several studies assessing the predictive performance of integrated risk scores have shown the value of combining different risk factors (54–56). Tada et al. (2015) (56) showed that adding a PGS to a model based on established risk factors for CHD, including family history, had the effect of improving risk classification and discrimination. Inouye, et al. (2018) (54) also noted that integrated scores inclusive of PGS and conventional risk factors,

such as smoking and BMI, substantially increased prediction over PGS and conventional risk factors alone (54).

More recently, Riveros-McKay et al. (2022) (55) developed a novel PGS for CAD using data from the UK Biobank. The PGS was combined with existing CVD risk stratification tools (PCE, QRISK3) to form an integrated risk tool (IRT). The CAD PRS exhibited greater prediction power for CAD occurrences than other available PGSs and had minimal correlation with both PCE and QRISK3. The IRT, when combined with PCE, outperformed in predictive accuracy. When specifically looking at how the IRT classifies individual risk, at high, average, or low risk, the IRT identified 10.4% of CAD cases labelled as low risk by PCE, as high risk , whereas only 4.4% were misclassified by the IRT in contrast to the PCE. The net enhancement in classification by the IRT was 5.9%. Further, when participants were segmented by age and sex, the IRT's superiority was evident across all categories, particularly in men aged 40-54 years, with a 15.4% improvement. Similar favourable results were observed when integrating with QRISK3 and when considering a broader spectrum of cardiovascular diseases (55). In a subsequent paper, this integrated score's improved prediction was validated across ancestries, notably European, African, and South Asian populations. However, wider variations in results were observed for groups with smaller sample sizes (57). Importantly, while these findings suggest transferability across ancestries, the African ancestry individuals in this study are primarily African Americans, Afro-Caribbeans, or Black Africans living in the UK. Using such populations as proxies for continental Africans has limitations (58). Consequently, the score's performance in African ancestries residing in Africa still remains unknown.

1.2.2.4. Transferability of PGS

Although risk loci are often relevant across populations, allele frequencies and LD patterns differ [28]. The majority, close to 90% (88.6%) of PGS have been developed from GWASs consisting primarily of participants from European populations (59). Euro-centric PGSs pose major ethical and scientific challenges surrounding their clinical implementation and relevance in non-European populations. Martin et al. (2019) (60) examined prediction accuracy, defined as the correlation between true and genetically predicted phenotypes, for 17 anthropometric and blood-panel traits in UK Biobank participants, and observed a decline in genetic prediction accuracy across global

populations using European-derived PGS (**Figure 2**). Specifically, accuracy decreased by approximately 1.6 times for both Hispanic/Latino Americans and South Asians (SAS), 2.0 times for East Asians (EAS), and 4.5 times for Africans (AFR). Authors concluded that in their current form, although they provide invaluable insight, European-derived PGS have the potential to exacerbate health disparities and afford greater healthcare benefits to those of European descent (60).

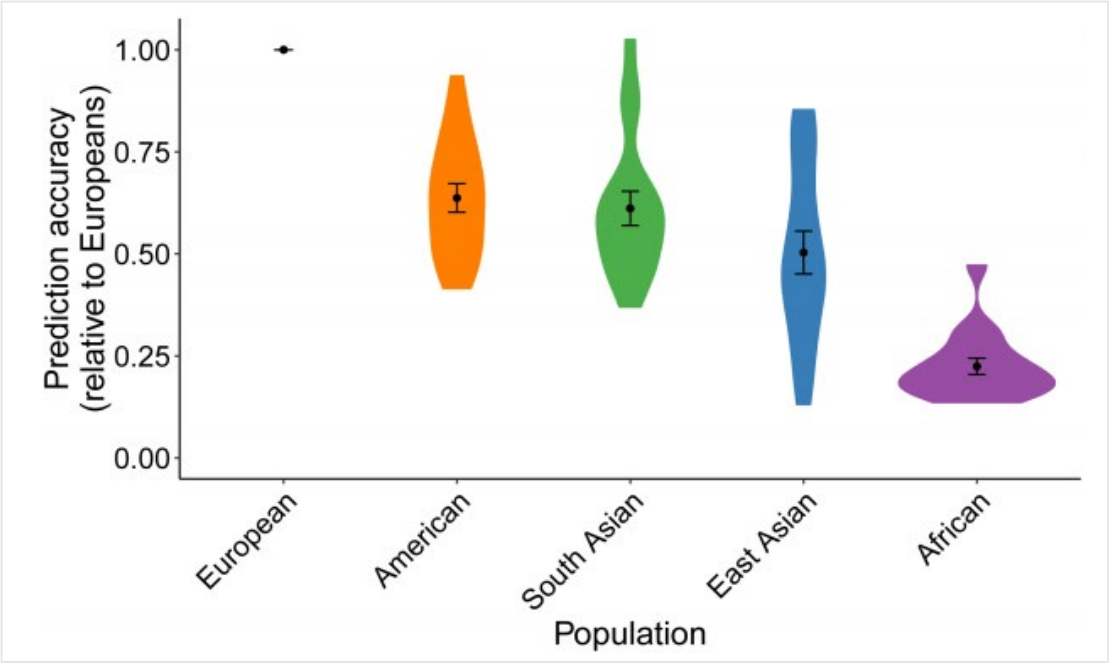


Figure 2: Prediction accuracy relative to European-ancestry individuals across 17 quantitative traits and 5 continental populations in the UK biobank (Martin et al., 2019).

Violin plots show distributions of relative prediction accuracies. Points show mean values, and error bars show standard error of the mean (s.e.m) values.

A recent review of PGS and their use in disease risk prediction in Africa by Fatumo and colleagues (2023) (59) highlight that using large-scale African ancestry cohorts as discovery sets in the development of PGS may generate more generalisable scores. To realise the full and equitable potential of PGS and ensure health disparities are not exacerbated, greater diversity in genetic studies and improved methods must be prioritised. Several novel multi-ancestry methods have been developed, including PRS-CSx (61) and BridgePRS (62), that consider population specific LD structure and coefficient weightings. However, predictive utility remains relatively low in African ancestry populations compared to other ancestry groupings (EUR, EAS, and SAS),

and importantly African representation is largely amongst African diaspora, including African Americans, suggesting prediction may be even lower for continental Africans. Nonetheless, the most appropriate method for capturing polygenic risk in African populations remains unclear. To address some of these challenges, this study aims to develop ancestry-aligned PGS using GWAS summary statistics and target data from continental Africans.

1.3. Clinical translation

As discussed, current clinical risk prediction relies primarily on basic demographic, lifestyle, and clinical characteristics with limited inclusion of underlying genetic predisposition. The clinical application of PGSs relates to the assessment of an individual's score in relation to the population. For example, individuals with the highest disease risk based on PGS values, might be offered regular screening, recommended lifestyle modifications or prescribed therapeutic interventions (7). The potential value of using PGS has been observed in the UK for breast cancer, where PGS is used to stratify women according to breast cancer risk and explore the implications for initiating screening at different ages (63,64). The use of PGS for cardiovascular diseases has yet to be translated widely into the clinical setting, but a recent clinical trial conducted in the UK evaluated the real-world feasibility, acceptability, and impact of an integrated risk tool, inclusive of PGS, for a cardiovascular disease (CVD IRT). This tool combines the standard QRISK2 risk algorithm with a PGS. The trial noted that among healthcare professionals and participants within the trial, CVD IRT implementation was feasible and well accepted. Importantly, the results of the CVD IRT were associated with planned changes in prevention strategies (65).

However, there are substantial barriers to the development and implementation of PGS, particularly in low- to middle-income settings. An integrated CVD risk score approach requires substantial investment in infrastructure and genetic and scientific capacity, including access to large-scale genotyping facilities and extensive training of medical professionals. Importantly, education for clinicians and the public will be necessary to increase genetic literacy and overcome potential negative public attitudes towards genetic testing (46). The attitudes and perceptions of patients, healthcare providers, government, and other stakeholders towards PM and risk scores remains

an understudied area in South Africa, and other resource-constrained settings, and should be explored.

1.3.1. A need to consider multimorbidity

It is critical to recognise that CVDs often do not occur in isolation, especially within Africa, where there is a significant burden of chronic infectious diseases, such as human immuno-deficiency virus (HIV) and tuberculosis (TB) (66). Multimorbidity (MM), the co-occurrence of two or more chronic conditions in an individual, is a growing global challenge (67,68) and is associated with increased premature mortality (68), lowered quality of life (69), and intensified health services utilisation [7]. Consequently, precision medicine requires a paradigm shift not only in the individual treatment approach and the data used but also towards a more comprehensive understanding of an individual's health. The interplay between multiple chronic conditions, on both a genetic and environmental level, and their respective treatment must be considered.

Epidemiological studies indicate that the co-occurrence of certain disease clusters such as CVD with T2D, and depression with both—are more common than others and thus these conditions may be interconnected. Depression has been shown to significantly increase the risk of CVD morbidity and mortality by 80 to 90% (70), and the likelihood of developing T2D by 60% (71). Also, about 40% of individuals with CAD exhibit symptoms of depression (72). The underlying causes of these associations are complex and may include shared lifestyle factors like physical activity, diet, and BMI (73–75) but several meta–analyses have shown that the CAD-depression and the T2D-depression relationship were similar before and after adjusting for major sociodemographic and lifestyle factors (76,77). These persistent associations suggest a potential genetic predisposition. Supporting this, twin and family studies revealed moderate correlations between depression and CAD (42%) (78) and depression and T2D (25%) (79).

A recent study investigated the genetic overlap between depression, CAD, and T2D, and identified a latent genetic factor structure for MM across all three diseases. Bivariate genetic correlations suggested a shared genetic architecture between all three disorders, albeit with the strongest correlation between CAD and T2D. Eleven SNPs and 18 genes were linked to MM, with most SNPs related to CAD and T2D, and a smaller number to depression. Associated genes are mainly active in immune and

cytokine pathways, which are involved in the regulation of immune and inflammatory responses–both of which have been implicated in CAD, T2D and depression (80).

However, this research and most other MM studies have largely been confined to high-income, EUR-ancestry populations, limiting the relevance of these findings globally. Studies that do include African-ancestry populations, primarily African Americans, suggest a higher risk of MM and poorer health outcomes amongst Africans, but African Americans are known to be poorly representative of continental African populations. Furthermore, these are often based on limited sample sizes and a few diseases or comorbidities, leaving a gap in understanding for African-ancestry groups. To improve health outcomes and perform PM effectively in an African setting an understanding of MM prevalence, patterns, and associated factors specific to African-ancestry populations is required.

1.4. Summary of the literature

The field of PM challenges the traditional 'one-size-fits-all' approach to healthcare by tailoring medical diagnostics and treatments based on individual genetics, environment, and lifestyle. Due to the increasing burden of CVD in South Africa and the associated burgeoning social and economic costs, identifying the most appropriate and cost-effective CVD prevention and treatment strategies should be prioritised.

Advancements in genomic technologies and related statistical approaches have the potential to provide information to guide tailored clinical interventions. PGS are one such approach and aim to understand an individual's underlying genetic risk for complex diseases such as CVD. Most PGS have been developed using data from European ancestry individuals and show limited transferability to African populations, potentially further exacerbating existing health disparities. Additionally, CVD is influenced by modifiable environmental risk factors and, hence, when computing CVD risk, both genetic and environmental factors must be accounted for. An integrated CVD risk score developed using data from Africans, which includes both genetic and environmental risks, could lead to improved risk stratification and, thus, the development of more appropriate prevention and treatment strategies for African patients.

However, the clinical utility of an integrated CVD risk score model in a South Africa setting may be limited, given the country's resource constraints. The risk score's value must be explored to identify the key considerations required for effective and feasible use in the South African clinical setting.

1.5. Gaps in the literature and motivation for this study

CVD remain a major health challenge in Africa. Individual risk is underpinned by both genetic predisposition and exposure to environmental risk factors. With advancements in genomics, such as PGS, understanding the role of genetics in CVD risk, and subsequent prevention and treatment of CVD is compelling. However, several gaps and challenges remain:

Integrated Risk Scores: While it is acknowledged that integrating PGS with established risk factors can enhance CVD prediction, a comprehensive risk score that includes both genetic and environmental risks from continental African is yet to be developed and evaluated.

Lack of transferability: Most PGS that assess genetic risk for diseases, including CVD, have been developed from GWAS using European ancestry populations, reducing their performance in other populations. Consequently, the applicability and accuracy of these scores are limited in African-ancestry populations and call for investigation into the predictive utility of ancestry-aligned scores.

Precision Medicine's feasibility: PM promises individualised healthcare solutions based on genetics, environment, and lifestyle. However, the clinical translation of PGS, and other PM tools, in resource-constrained settings, such as South Africa, faces infrastructural, educational, and perception-related barriers. Understanding the current genetic services landscape, and the perceptions and attitudes of PM-based tools amongst different stakeholders is critical to facilitating translation.

CVD and Multimorbidity: Given the substantial and overlapping chronic infectious disease burden in Africa, considering concomitant diseases when developing PM-tools for CVD is important.

Motivated by these significant research and health-impact gaps, this study aims to improve CVD risk stratification in Africa through the development of a CVD risk score

tailored for continental African populations. Furthermore, it aims to understand the feasibility of such genomic-based approaches in an African setting.

1.6. Aim and Objectives

The overarching aim of this study is to develop and critically assess the value of a PM-based risk score that assesses both genetic and conventional risk for CVD in an African setting. A major part of this work will be to determine if genetic factors improve CVD risk prediction over non-genetic factors alone. The second part of the study aims to assess the potential utility of PM-based approaches in resource constrained settings, such as South Africa.

The objectives of this study are:

1. To develop ancestry-aligned polygenic risk scores for cardiometabolic traits by leveraging existing continental African GWAS summary statistics and applying it to the AWI-Gen study as a target population.
2. To integrate derived PGS with conventional non-genetic risk factors to form an integrated risk score.
3. To derive, evaluate, and compare the predictive utility of genetic, non-genetic and integrated CVD risk models in the AWI-Gen population.
4. To understand the current state of genetics and genomics on the African continent, specifically the current approach to clinical genetics, including laboratory and healthcare personnel resources.
5. To determine clinician perceptions, readiness, and willingness to adopt PM-based CVD risk stratification in South Africa's public health setting.
6. To contextualise the CVD burden among the multiple health burdens in Africa.

Chapter 2: METHODS

The polygenic risk score analysis and prediction modelling component of the study used two existing datasets: the publicly available GWAS summary statistics from the African Partnership for Chronic Disease Research (APCDR) and, following a data agreement, the individual-level genetic and phenotypic data from the Africa-Wits INDEPTH Partnership for Genomic Research (AWI-Gen).

2.1. Study datasets

2.1.1. Africa Wits-INDEPTH partnership for Genomic Studies

This study is part of the Africa Wits-INDEPTH (University of the Witwatersrand, Johannesburg (Wits), and the International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH) (81) partnership for Genomic Studies (AWI-Gen) project, supported by the NIH as a Collaborative Centre within the Human Heredity and Health in Africa (H3Africa) Consortium. AWI-Gen's primary objective is to investigate the genetic and environmental contributors to cardiometabolic diseases in African populations (82).

AWI-Gen is a population-based cross-sectional study of 12,045 unrelated African participants (men and women), most of whom are aged between 40-60 years and reside in one of six sites (rural and urban) across sub-Saharan Africa (83) (**Figure 3**). The sites are located in East Africa - Nairobi, Kenya (84); West Africa - Nanoro, Burkina Faso (85) and Navrongo, Ghana (86); and in South Africa - Agincourt (87), Dikgale (88) and Soweto (89). The study excludes closely related individuals, pregnant women, and recent immigrants (<10 years) into the communities. There were no selection criteria based on body composition, infection or disease history (82).

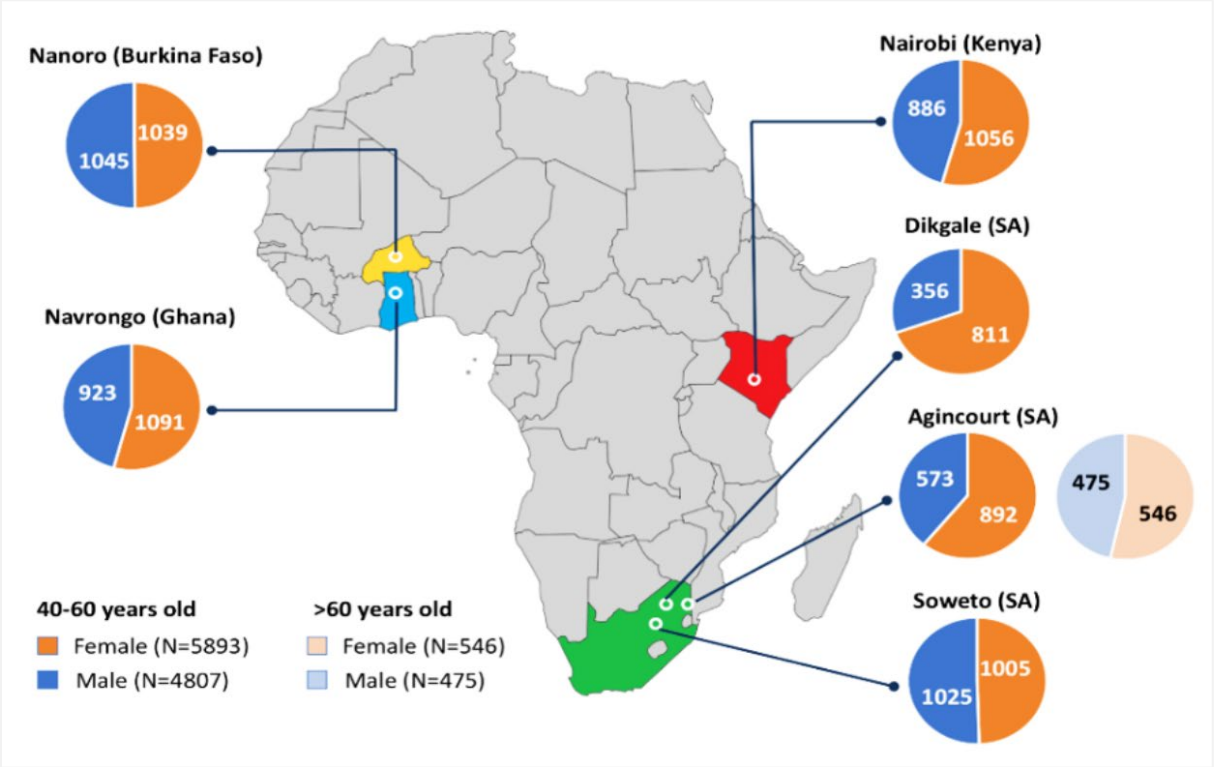


Figure 3: AWI-Gen research project sites, by sex and age split
https://www.wits.ac.za/research/sbimb/research/

Phenotype data

Data collection involved detailed paper-based questionnaires and biomarker assessments to collect demographic, general health and infection history variables as well as information on non-modifiable risk factors (age, sex, genetic predisposition), lifestyle factors (diet, exercise), and biomarkers for disease (BMI, fat distribution, blood pressure, lipid levels) (82).

For this study, variables associated with CVDs and used in conventional CVD risk calculators were included in our models as non-genetic factors (26,29,90). Variables selected included age, sex, current smoker and alcohol consumer status, sleep (hours/night) and moderate and vigorous physical activity (minutes/week) as well as juice (number per week) and sugar drinks (number per week). Weekly consumption of bread (slices per week), fruit (servings per week), and vegetables (servings per week) were derived by multiplying the individual's number of servings per day by the number of times a week each respective food group is consumed. Not all selected variables were available in the Soweto sample; thus, these samples were excluded

from prediction modelling analyses. For the remaining participants, individuals with more than 5% of data missing amongst these selected variables were removed.

Genotype data

Genotyping was performed on approximately 11, 000 individuals using the H3Africa genotyping array (Illumina), enriched for common African variants, and included approximately 2.3 million SNPs, processed by Illumina's FastTrack Sequencing Service. The genotype dataset underwent pre-imputation quality control that included the removal of individuals with missing SNP calling rate greater than 0.05, SNPs with a genotype missingness greater than 0.05, MAF less than 0.01 and Hardy-Weinberg equilibrium (HWE) p-value less than 0.0001. Non-autosomal and mitochondrial SNPs, and ambiguous SNPs that did not match the GRCh37 references alleles or strands were also removed. Samples that were potential duplicates (PIHAT > 0.9), had a missing SNP genotyping rate greater than 0.05, and reported vs. genetic sex inconsistencies were excluded.

Population stratification was assessed using a principal component (PC) analysis based on an LD-pruned subset of SNPs using the smartPCA program implemented in EIGENSTRAT. Imputation was performed using the African Genome Resources reference panel (EAGLE2 + PBWT pipeline) at the Sanger Imputation Server (https://imputation.sanger.ac.uk/). The cleaned dataset, comprising 1,729,661 SNPs from 10,903 participants, was imputed using the Sanger Imputation Server with the African Genome Resources as a reference. EAGLE2 was selected for pre-phasing, and the default PBWT algorithm was used for imputation. Post-imputation QC involved removing indels, rare SNPs (MAF ≤ 0.01), and poorly imputed SNPs (Info score ≤ 0.6), resulting in a final dataset containing 10,603 participants and 13.98M SNPs (91)

2.1.2. African Partnership for Chronic Disease Research

The African Partnership for Chronic Disease Research (APCDR) is a multi-trait pan-African GWAS involving 14,126 individuals. This GWAS meta-analysis combines association statistics from four distinct cohorts: the Uganda Genome Resource (UGR) (n = 6,407), the Durban Diabetes Study (DDS) (n = 1,165), the Diabetes Case-control

study (DCC) (n = 1,542), and the African American Diabetes Mellitus (AADM) study (n = 5,231).

The UGR includes 6,407 individuals (4,429 with genotype, and 1,978 with sequence data) from 9 ethno-linguistic groups from the General Population Cohort (GPC), Uganda (92). The GPC is a population-based cohort study initiated in 1989 by the UK Medical Research Council (MRC) in collaboration with the Uganda Virus Research Institute (UVRI) to study the prevalence and incidence of HIV infection and associated factors. A questionnaire collected health and lifestyle data, and blood samples and biophysical measurements were taken (92).

The UGR genetic data comprises of data from 4,778 individuals with genotypes for ~2.2 million SNPs from the Uganda Genome-wide Association Study (UGWAS) resource, and sequence data on up to 1,978 individuals, including 41.5 M SNPs and 4.5M indels (Uganda 2000 Genomes [UG2G]). To maximise opportunities for genomic discovery, the UGR authors meta-analysed GWAS data from the remaining three study populations (DDS, DCC, and AADM). Genotypes were imputed with combined reference panel developed by authors (93) using the Han-Eskin random-effects meta-analytic approach implemented in METASOFT.

The DDS is a population-based cross-sectional study conducted between November 2013 and December 2014 in KwaZulu-Natal (South Africa). Recruited participants included those over 18 years of age, not pregnant, and residing in urban black African communities (94). A survey was used to collect health, lifestyle, and socioeconomic data. Standardised biophysical measurements, biomarkers for non-communicable and infectious diseases, and genetic data, were collected (94).

In contrast, the DCC targeted individuals with type 2 diabetes to examine the epidemiology and genomics of type 2 diabetes and related cardiometabolic traits in a South African population. Participants older than 40 years, residing in Kwa Zulu Natal and diagnosed with type 2 diabetes according to the WHO criteria were recruited from a tertiary hospital in Durban, South Africa. The design was for a case-control study, though only case data were available by the study's end in 2013.

AADM is a genetic epidemiology study of type 2 diabetes and related traits in African populations, with participants drawn from various study centres in Nigeria (Ibadan,

Lagos, and Enugu), Ghana (Accra and Kumasi), and Kenya (Eldoret). Demographic information was collected using standardized questionnaires. Anthropometric, medical history, and clinical measurements were obtained by trained study staff during a clinic visit.

For the GWAS meta-analysis, 34 traits were assessed, spanning anthropometric, haematological, blood pressure, glycaemic, liver function, and lipid measurements. Of these, this study focused on 14 cardiometabolic traits given their alignment with the AWI-Gen project. These traits include measurements related to anthropometry (Body mass index (BMI); height, hip circumference, waist circumference, weight, Waist-Hip Ratio (WHR)), blood pressure (diastolic blood pressure (DBP), systolic blood pressure (SBP)), lipid measurements (Total Cholesterol (TC), Low-density lipoprotein (LDL), Triglycerides (TG), High-density lipoprotein (HDL)). **Table *1*** lists these traits, with those in bold and italicised assessed in this study. Summary statistics were downloaded from the NHGRI-EBI GWAS Catalog (95) on 01 Jun 2021 for studies GCST009042 to GCST009060.

Table 1: Phenotypic traits analysed in the GPC

Traits	Traits Grouping	Unit	Sample no.	Mean	SD
TC	***Lipids***	***mmol/L***	***4778***	***3.6***	***0.98***
LDL	***Lipids***	***mmol/L***	***4778***	***2.1***	***0.76***
TG	***Lipids***	***mmol/L***	***4778***	***1.2***	***0.61***
HDL	***Lipids***	***mmol/L***	***4778***	***1***	***0.4***
ALT	Liver function	U/L	4778	20.1	12.7
AST	Liver function	U/L	4778	28.1	24.6
Albumin	***Liver function***	***g/L***	***4778***	***41.4***	***4.1***
ALP	Liver function	U/L	4778	133.5	103.4
GGT	Liver function	U/L	4778	29.5	73
Bilirubin	***Liver function***	***µmol/L***	***4778***	***9.9***	***8.5***
DBP	***Blood pressure***	***mm/Hg***	***4773***	***74***	***10.5***
SBP	***Blood pressure***	***mm/Hg***	***4770***	***122.8***	***17***
HbA1c	Glycaemic control	NGSP%	4772	3.4	0.69
Height	***Anthropometric indices***	***cm***	***4732***	***157.3***	***9***
Weight	***Anthropometric indices***	***kg***	***4640***	***52.9***	***11.3***
BMI	***Anthropometric indices***	***kg/m2***	***4618***	***21.2***	***3.8***
Waist circum	***Anthropometric indices***	***cm***	***4617***	***75.2***	***9.2***

Hip circum	***Anthropometric indices***	***cm***	***4616***	***89.5***	***9.8***
WHR	***Anthropometric indices***	***cm***	***4615***	***0.8***	***0.05***
WBC count	Blood haematological factors	x10^9/l	1479	5.2	1.5
RBC count	Blood haematological factors	x10^9/l	1479	4.7	0.61
MCV	Blood haematological factors	fl	1479	85.9	7.8
MCH	Blood haematological factors	pg/cell	1479	28.9	2.9
MCHC	Blood haematological factors	g/L	1479	33.6	1.2
RDW	Blood haematological factors	%	1479	13.1	1.4
PCV	Blood haematological factors	l/l	1479	40.3	4.6
Haemoglobin	Blood haematological factors	g/L	1479	13.6	1.6
MPV	Blood haematological factors	fL	1478	8.7	0.8
PLT	Blood haematological factors	x10^9/l	1478	219	79.1
Lymphocytes	Blood haematological factors	x10^9/l	1423	48.6	9.9
Monocytes	Blood haematological factors	x10^9/l	1423	5.7	1.9
Basophils	Blood haematological factors	x10^9/l	1423	0.9	1
Neutrophils	Blood haematological factors	x10^9/l	1415	38	10.5
Eosinophils	Blood haematological factors	x10^9/l	1415	6.7	6.5

List of the phenotypic traits analysed in the GPC. Bold and italicised traits were the traits assessed in this study. TC, Total Cholesterol; LDL, Low-density lipoprotein; TG, Triglycerides; HDL, High-density lipoprotein; DBP, diastolic blood pressure; SBP, systolic blood pressure; BMI, Body mass index; WHR, Waist-Hip Ratio.

2.1.3. Disease outcomes

Outcome variables in this study comprised 13 cardiometabolic traits and seven CVD-associated outcomes. Cardiometabolic traits included BMI, height, weight, hip circumference, waist circumference, and W-H ratio, DBP, SBP, LDL-C, HDL-C, TC, TG, and albumin and bilirubin blood serum levels. CVD-associated disease outcomes assessed were CVD, diabetes mellitus, dyslipidaemia, heart attack, hypertension, obesity, and stroke.

Disease outcome definitions

CVD: defined as present if the participant reported having had a heart attack, stroke, or transient ischaemic attack. Participants previously diagnosed with congestive heart failure or angina were also classified as having CVD. No data were available for angina or heart attack prevalence for the Soweto men, so CVD prevalence was calculated using the remaining data.

Diabetes: defined using the WHO criteria, which are the presence of a previous diagnosis of diabetes by a healthcare professional or fasting blood glucose≥7 mmol/L or random glucose≥11.1 mmol/L, or on diabetes medication at the time of recruitment.

Dyslipidaemia: defined as the presence of one of the following: total cholesterol (TC)≥5.0 mmol/L, or high-density lipoprotein cholesterol (HDL-C)<1.0 mmol/L for men and <1.3 mmol/L for women, or low-density lipoprotein cholesterol (LDL-C) ≥3. 0 mmol/L, or triglycerides (TG)≥1.7 mmol/L. The Randox Daytona Plus (Randox Laboratories, UK) autoanalyser was used to analyse HDL-C, TG and TC on fasting venous blood samples. LDL-C was calculated using the Friedewald equation.

Heart attack: defined by yes response to self-reported heart attack status. Participant asked if a healthcare professional has diagnosed the participant with a heart attack in the past.

Hypertension : defined as systolic blood pressure≥140 mm Hg and/or diastolic blood pressure≥90 mm Hg, in line with the seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure, or if the participant was taking hypertension medication.

Obesity: defined as having a Body Mass Index (BMI) of 30 kg/m² or higher, in accordance with the guidelines set by the World Health Organization. BMI is calculated by dividing a person's weight in kilograms by the square of their height in meters.

Stroke: defined by yes response to self-reported stroke status. Participant asked if a healthcare professional has diagnosed the participant with a stroke in the past.

2.2. Polygenic risk score derivation

2.2.1. Quality control of datasets

2.2.1.1. QC of Base data: GWAS summary statistics

GWAS summary statistics were identified for phenotypes that were also measured in AWI-Gen. The identified summary statistics underwent a series of quality control (QC) procedures, including the extraction of HapMap3 variants, and the removal of ambiguous variants, or where variants had missing data. Variants were flipped to

match the 1000 Genomes Phase 3 (1KG) reference, following which variants with a MAF >0.01 in the African subset of 1KG (1KG AFR), a MAF >0.01 in the GWAS sample, and an INFO >0.6 were retained. GWAS summary statistics variants were removed if they (1) had a discordant MAF (>0.2) between the reference and GWAS sample, (2) had reported p-values outside the range of 0 to 1, (3) were duplicates, or (4) had a sample size >3 SD from the median sample size.

2.2.1.2. QC of Target data: Ancestry classification

Individuals in AWI-Gen were assigned to the five super populations present in the 1000 Genomes phase 3 (1KG) reference sample (1000 Genomes Project Consortium, 2015), namely EUR, EAS, SAS, AFR, and Admixed American. Super population membership was predicted using a 1KG reference trained elastic net model consisting of the first six reference-projected genetic principal components. Principal components were defined in the 1KG reference using variants present in the unimputed AWI-Gen genotype data with a minor allele frequency >0.05, missingness <0.02 and Hardy-Weinberg p-value >1×10-6. LD pruning for independent variants was performed in PLINK (Chang et al., 2015) after removal of long-range LD regions (Price et al., 2008), using a window size of 1000, step size of 5, and r^2 threshold of 0.2.

A multinomial elastic net model predicting super population membership in the 1KG reference was derived in using the glmnet R package (96), with model performance assessed using 5-fold cross validation. The reference-derived principal components were then projected into AWI-Gen, and the reference-derived elastic net model was used to predict super-population membership in AWI-Gen. AWI-Gen participants were assigned to a super population if the predicted probability was >0.5. As expected, all participants were assigned to the AFR superpopulation.

2.2.2. Score Construction

Typically, a PGS follows the form $\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \ldots + \beta_n X_n$, where β_k represents the effect size attributed to each allele for a given cardiometabolic trait associated with SNP k. X_k is the number of effect alleles at SNP k, and n is the total number of SNPs in the PGS. To derive the PGS for each trait, we used (1) publicly available GWAS summary statistics from APCDR; extracting the disease-associated

variants, the effect allele, the estimated *β*-coefficient for the effect allele, and the p-value of each genetic variant, and (2) LD between genetic variants from the African 1KG reference panel (661 Africans) (97), this is important for reproducibility and generalisability (52).

After preparation of GWAS summary statistics, PGS were derived using the p-value thresholding and clumping (pT+ clump) approach, a PGS method commonly applied across diverse phenotypes and ancestries. The pT + clump method using LD-based clumping to retain only the single most significant variant within each locus. Within this study, LD-clumping parameters were an r^2 threshold of 0.1 and window of 250 kb, i.e. variants are removed if they are within 250kb and have $r^2 \geq 0.1$ with a more significant variant The African subset of the 1000 Genomes (1KG AFR) was used to estimate LD. Ten p-value thresholds were used to select variants: 1×10^{-8}, 1×10^{-6}, 1×10^{-4}, 1×10^{-2}, 0.1, 0.2, 0.3, 0.4, 0.5 and 1. Polygenic scores were then calculated in AWI-GEN, imputing missing variants using the 1KG AFR allele frequency. 1,104,4026 HapMap3 variants were present. Polygenic scores were standardised (scaled and centred) based on the mean and SD of PGS in the reference sample (1KG AFR). The score calculations were implemented using the GenoPredPipe pipeline (https://github.com/opain/GenoPred) and using PLINK v1.9.

2.2.3. Association testing

Following PGS development, regression analysis was used to assess the within- and cross-trait prediction utility of each PGS, and with the seven disease outcomes of interest, while accounting for confounders such as age, sex, and the first eight principal components (98). Similarly, regression analyses were run to assess the association between conventional non-genetic risk factors, including age, smoking status, alcohol consumption, diet-related variables, sleep, and physical activity, with selected disease outcomes. Given the differences by sex of these traits within African populations, the analyses were stratified by sex.

The proportion of variance for a trait explained by the PGS and non-genetic factors was computed as the R^2. For PGS association testing, R^2 was obtained from a full model including both PGS and covariates (PCs, sex, age, and age-squared) minus the R^2 obtained from a model including covariates alone. R^2 was not adjusted to the liability threshold model due to limited disease prevalence estimates available across Africa

and the substantial variation in prevalence noted across cohort sites. For multiple testing, results were corrected for the number of PGS tested for each outcome (i.e., applying a p-value threshold of 0.05/14). We did not correct for the number of PGS p-value thresholds as PGS calculated at each threshold are correlated, and a Bonferroni correction would be overly conservative.

Prediction accuracy was evaluated as the Pearson correlation between the observed and predicted outcome values. Correlation was used as the main test statistic as it is applicable for both binary and continuous outcomes and standard errors are easily computed as:

$$SE_r = \frac{1 - r^2}{\sqrt{n - 2}}$$

Where SE_r is the standard error of the Pearson correlation, *r* is the Pearson correlation, and n is the sample size.

2.2.4. Ancestry predictor derivation

Genetically inferred ancestry was included in prediction models to account for population stratification and potentially improve prediction. To reduce overfitting, ancestry was determined by fitting the data to the 1KG Phase 3 projected principal components (PPCs) of population structure. This is in contrast to within-sample PCs which are traditionally used as covariates in regression models. PPCs are used to reduce potential overfitting. Calculation of PPCs is described in Section 2.2.3.

2.3. Integrated risk score and prediction modelling

2.3.1. Elastic net regression

When modelling the PGS and integrated risk scores, logistic regression was used for predicting binary outcomes, and linear regression was used for predicting continuous outcomes. For genetic and integrated models, for each cardiometabolic trait, rather than selecting the single best-performing PGS (based on max R^2), all PGS demonstrating significance after multiple testing corrections were retained for subsequent predictive modelling analyses.

As models contained more than one predictor (i.e. the PGS for each p-value threshold and several non-genetic factors), an elastic net model was applied to address

overfitting and multicollinearity (99). Overfitting occurs when a regression model becomes too complex, capturing noise and random fluctuations in the data rather than the true underlying relationships, and thus may cause an inflation of association estimations. Multicollinearity occurs when two or more independent variables in a regression model are highly correlated with each other; making it challenging to distinguish the individual effects of each predictor on the dependent variable leading to unstable and unreliable coefficient estimates.

Elastic net is a hybrid regularisation method that combines both L1 (Least Absolute Shrinkage and Selection Operator (LASSO)) and L2 (Ridge) penalties to achieve feature selection and shrinkage simultaneously. The key difference between the L1 and L2 methods are the penalty term. Lasso introduces a penalty based on the magnitude of the coefficients and shrinks the co-efficient of less important features to zero, thus removing some features altogether, whilst ridge takes the square and subsequently distributes the impact of correlated features across multiple coefficients, with coefficients never reaching zero. The loss function of Elastic Net Regression can be represented as follows:

$$\hat{\beta}^{elastic} = arg\ min_{\beta} \left(MSE + \lambda \left[\alpha \|\beta\|_1 + \frac{1}{2}(1 - \alpha)\|\beta\|_2^2 \right] \right)$$

$\hat{\beta}^{elastic}$ represents the estimated coefficients (or parameters) of the regression model that the Elastic Net is trying to find. These coefficients define how each predictor (independent variable) influences the dependent variable. *arg min* is used to denote the value of *β* that minimises the expression. MSE (Mean Squared Error) is a measure of model fit and is calculated as the average of the squares of the differences between the predicted and actual values. β are the regression coefficients (weights) for each predictor. λ (lambda) is the non-negative hyperparameter that controls the overall strength of the regularisation. A larger λ applies more regularisation (more penalty), leading to smaller and potentially sparser coefficients. α (alpha) is a hyperparameter between 0 and 1 that balances the L1 and L2 penalties. An α close to 1 makes the penalty similar to LASSO (favouring sparsity), while an α close to 0 makes it more like Ridge regression (favouring smaller, but not zero, coefficients). $\|\beta\|_1$ is the Lasso penalty term and encourages model sparsity by forcing some coefficients to exactly

zero. $\|\beta\|_2^2$ is the L2 penalty that encourages small, non-zero coefficients by penalising large ones.

As the optimal hyperparameter values, α and λ, are unique to each model and unknown prior to modelling analyses, cross-validation techniques are employed to find optimal values for α and λ.

2.3.2. Nested cross-validation

Cross-Validation (CV) is an out of sampling technique that is used to evaluate the performance of a model and select the best hyperparameters. The technique helps reduce bias and overfitting in model evaluation.

Nested CV reduces the potential for overfitting by using a series of train/validation/test set splits. In this study, and as previously described (100), the data was repeatedly partitioned into training, validation, and testing sets, and consisted of 5 outer folds with a 80-20 data split (80% training, 20% testing) to provide an unbiased estimate of the predictive utility of the model, and 10 inner folds (90% training, 10% testing) for hyperparameter tuning . Different hyperparameter settings are evaluated within this inner loop. The setting that performs best on the validation fold is chosen. This process is repeated for each combination of training-validation split within the inner loop. For each outer loop iteration, the model with the best hyperparameters (found in the inner loop) is trained on the entire training set and tested on the outer test set. The performance scores from the outer loop are averaged to provide an overall performance metric (**Figure 4**). The proportion of variance explained by a model was computed as R^2. Hyperparameters were determined using the 'caret' R package, which optimises the RMSE for continuous outcomes and accuracy for binary outcomes.

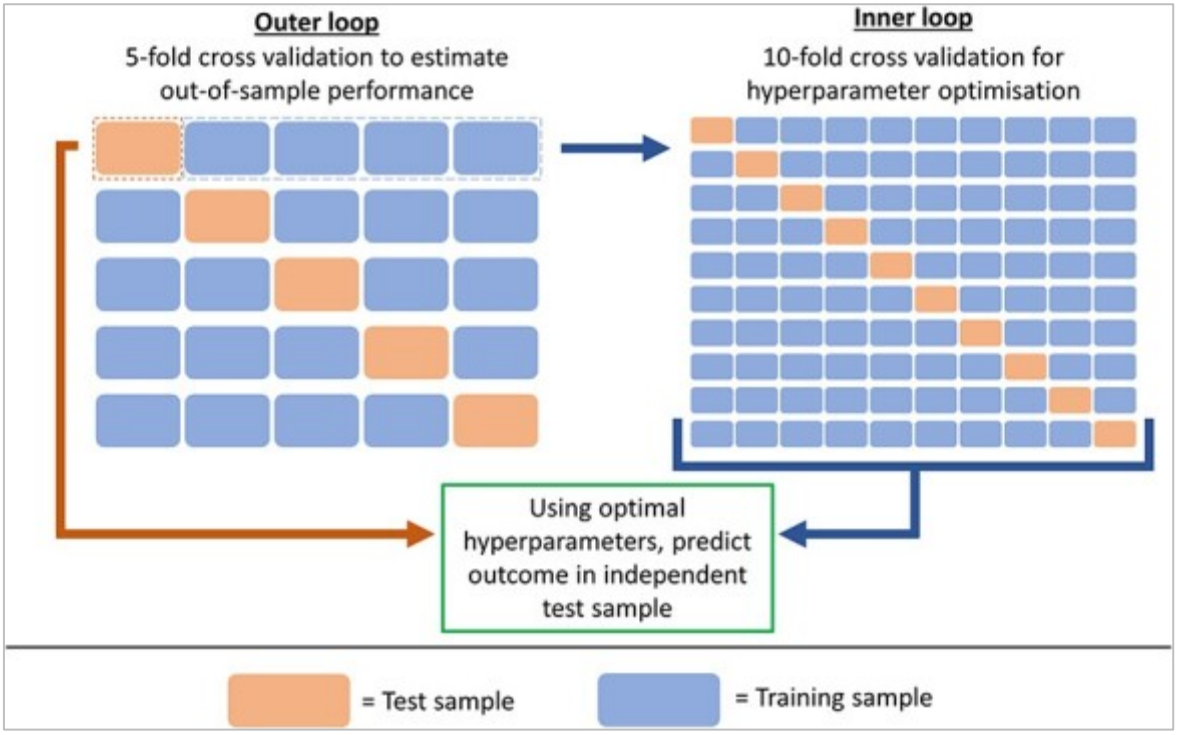


Figure 4: Schematic illustration of nested cross-validation

Schematic representation of nested cross-validation. The sample is split into 5 outer folds, in each outer fold the data is split into 5 parts - four parts are used as a training sample for hyperparameter tuning, and the resulting model is then used to predict the outcome in the remaining part (test sample). This process is repeated until predictions are available for all parts of the sample. Hyperparameter optimisation is carried out within the inner loop, which consists of 10-fold cross validation. Figure used with permission from Oliver Pain (100).

The correlation between observed and predicted values of each model were compared using William's test (also known as the Hotelling-Williams test) (101) as implemented by the 'psych' R package's 'paired.r' function, with the correlation between model predictions of each method specified to account for their non-independence. A two-sided test was used when calculating p-values.

Non-genetic models included ten factors selected based on data availability in AWI-Gen and their known association with CVDs. Integrated models included genetic (PGS and PPCs) and non-genetic models. No data were available for diet-related variables for the Soweto study site for men and women, so this site was excluded from prediction modelling analyses.

All analyses were performed using PLINK v1.9 (https://www.cog-genomics.org/plink/1.9/) (102), and R version 3.4.4 (http://www.r-project.org/) (103) unless specified otherwise.

SECTION 2: PUBLICATIONS

1

Section Overview

This section comprises the four publications emanating from this doctoral study. Each chapter corresponds to a publication. The numbering for tables, figures, and references has been updated to a continuous sequence to maintain a coherent flow and facilitate reading. Also, the spelling has been updated to follow British convention. The original versions of each publication can be found in Appendix F-H.

The four chapters and their associated publication titles are below, and their alignment to the study's key domains are illustrated in **Figure 5**:

Publication 1: Has translational genomics come of age in Africa?

Publication 2: Ancestry-aligned polygenic risk scores combined with conventional risk factors improve cardiovascular disease risk stratification in African populations.

Publication 3: Clinicians' Perceptions towards Precision Medicine Tools for Cardiovascular Disease Risk Stratification in South Africa.

Publication 4: Multimorbidity in African ancestry populations: a scoping review.

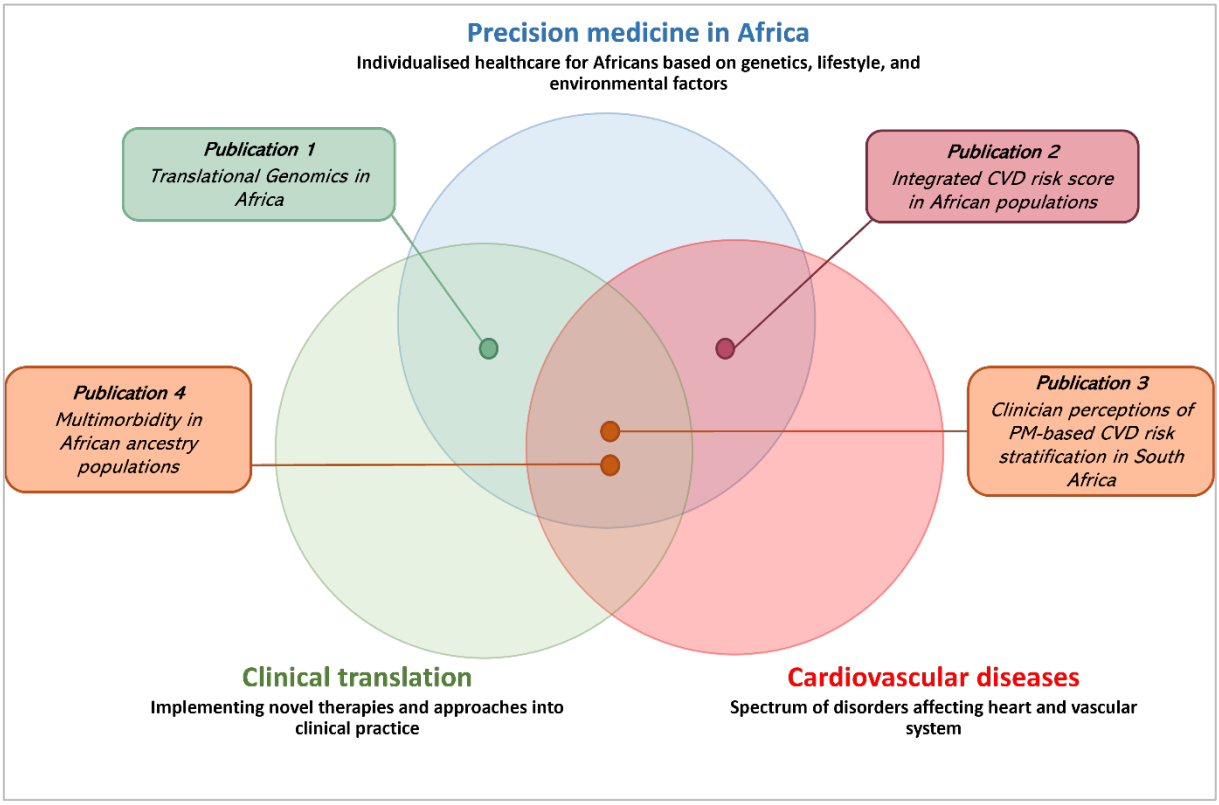


Figure 5: The alignment of publications emanating from this thesis in relation to three key domains of interest (Precision medicine, Clinical translation, and Cardiovascular diseases)

Chapter1: Publication 1

Has translational genomics come of age in Africa?

This chapter describes an invited review (Appendix F) that was performed in order to gain an understanding of the current translational genomics landscape in Africa and what factors would likely be required to accelerate the adoption of genomic medicine on the continent. This paper aims to understand the potential for precision public health, a concept similar to precision medicine in that aims to provide tailored healthcare solutions, except rather than focused on the individual, it aims to provide the right intervention to the right population at the right time based on extensive population-specific data (104), in Africa. Understanding the translational genomics landscape is crucial to ascertaining whether clinical genomic interventions, further research, and potential policy changes are not only warranted but feasible. The work described in this chapter was peer-reviewed and published in the journal Human Molecular Genetics and is an exact copy of the following journal publication:

Kamp, M., Krause, A., and Ramsay, M., 2021. Has translational genomics come of age in Africa?, *Human Molecular Genetics.* 30(R2): R164–R173, https://doi.org/10.1093/hmg/ddab180.

Author contributions: Conceptualisation: MR, AK, and MK. Interviews and synthesis: MK. Writing – Original draft preparation – MK. Writing – review and editing: MR, AK, and MK. Visualisation: MK. Supervision: MR, AK.

ABSTRACT

The rapid increase in genomics research in Africa and the growing promise of precision public health (PPH) begs the question of whether African genomics has come of age and is being translated into improved healthcare for Africans. An assessment of the continent's readiness suggests that genetic service delivery remains limited and extremely fragile. The paucity of data on mutation profiles for monogenic disorders and lack of large genome-wide association cohorts for complex traits in African populations is a significant barrier, coupled with extreme genetic variation across different regions and ethnic groups. Data from many different populations are essential to developing appropriate genetic services. Of the proposed genetic service delivery models currently used in Africa—Uncharacterised, Limited, Disease-focused, Emerging and Established—the first three best describe the situation in most African countries. Implementation is fraught with difficulties related to the scarcity of an appropriately skilled medical genetic workforce, limited infrastructure and processes, insufficient health funding and lack of political support, and overstretched health systems. There is a strong nucleus of determined and optimistic clinicians and scientists with a clear vision, and there is a hope for innovative solutions and technological leapfrogging. However, a multi-dimensional approach with active interventions to stimulate genomic research, clinical genetics and overarching healthcare systems is needed to reduce genetic service inequalities and accelerate PPH on the continent. Human and infrastructure capacity development, dedicated funding, political will and supporting legislation, and public education and awareness, are critical elements for success. Africa-relevant genomic and related health economics research remains imperative with an overarching need to translate knowledge into improved healthcare. Given the limited data and genetic services across most of Africa, the continent has not yet come of 'genomics' age.

1. INTRODUCTION

The past decade has seen significant growth in genomic research in Africa, in large part as a result of the activities of the African Society of Human Genetics (AfSHG) (105–107) and the Human Heredity and Health in Africa Consortium (H3Africa) (108–110). However, the question remains whether the genomic revolution and African

genome research have translated into direct benefits to individuals and families affected by genetic disorders.

Access to genetic technologies and genomic services remain sparse in Africa for several reasons. These include a lack of financial resources, poor national health services and infrastructure and the presence of more compelling health priorities, such as infectious diseases like tuberculosis, malaria, HIV/AIDS (109,111) and now COVID-19. Furthermore, few countries in Africa have formally recognised genetic services supported by clinically trained geneticists and dedicated diagnostic laboratory infrastructure.

The field of genomics holds great promise for health care and medical research, as the identification of the genetic determinants of diseases or phenotypes bring about the potential for significant improvements in diagnosis, prevention, and treatment of disease. Genomics is particularly attractive for Africa, where new technologies and products from genomics research offer the potential for a precision medicine approach to help mitigate the heavy burden of infectious and non-communicable diseases, as well as monogenic disorders (91,112). The availability of large genomic datasets could feed into precision public health (PPH) initiatives to improve population health by translating these new genomics and digital technologies, to guide public health practice through the generation of population-wide tailored interventions and policies.

From a clinical genetics' perspective, interventions could initially focus on the identification and molecular characterisation of monogenic disorders in specific African regions, such that their frequencies and molecular profiles in different areas are well characterised and defined using genomic technologies. Limited studies on sickle cell anaemia (113), fragile X syndrome (114–117), hearing loss (118–122), Huntington's disease (123,124), familial cancer syndromes (125,126) and others, have been performed and show the value of such studies, particularly highlighting the diversity in Africa. In addition, factors that modulate disease severity should be studied (as reviewed in (125–128)). Simultaneously, a profile of common variation in these populations would be established, providing a critical resource for interpreting, and characterising the potential functional impact of variants and their disease associations.

Thus, a two-pronged strategy, one firmly rooted in fundamental principles and the other leveraging state-of-the-art technologies, could propel clinical genetics to new levels in Africa. The first would involve collecting basic data on the prevalence and phenotype of monogenic diseases. The second would be to use next-generation sequencing approaches to establish population architecture and variation and subsequently identify causal mutations and their geographical distribution. In parallel, the development of the computational skills required for the analysis and interpretation of these variants is needed, as is digital innovation in data capture, storage and integration (129,130).

Worldwide, genetic testing is becoming increasingly complex as it inches closer towards the era of precision medicine, where genomic medicine becomes integrated into routine healthcare. However, despite these exciting developments, the fear is that existing health care inequalities are exacerbated as technologies and discoveries emanating from genomics research are not translatable across populations and remain accessible only to those who can afford them (60,131–133). Africa has an urgent need to develop innovative approaches and to provide timeous and effective genetic services to patients in resource-constrained environments.

While acknowledging substantial knowledge, capacity, and sustainability gaps, it is clear that the effective implementation of PPH has the potential to translate into more effective patient care. This review highlights the current state of genetics and genomics on the continent, describing how genetic services are currently delivered and how genomic and digital technologies underpin these services. The goal is to identify key factors needed to ensure that clinical genetic and laboratory services have broad benefit in Africa in the era of precision medicine.

2. GENETIC SERVICE DELIVERY MODELS

In order to gauge the readiness for genomic medicine, and subsequent potential of PPH in Africa, an assessment of the current genetic services offered is required. This will, in turn, assist in identifying the factors necessary to drive a PPH agenda on the continent. Although there is limited published information about the service delivery models in use (apart from some from South Africa (134,135), Egypt (136), Morocco (137), Tunisia (138,139), Nigeria (140) and Rwanda (141) preliminary exploration

suggests the presence of varying strategies from one country to the next. This variability is a direct consequence of the heterogeneity in healthcare contexts, and in the case of North Africa, a response to the consequences of high consanguinity levels (128,136,137,139).

Informal discussions with individuals identified by the authors as being well informed about the status of genetic service delivery in their respective countries explored the following areas: (1) the availability of a clinical genetics service in the country; (2) the level of infrastructure and local capacity available for service delivery and (3) the evaluation of the service within its given context.

These discussions revealed that genetic service delivery remains very limited and extremely fragile in Africa. Genetic services are fraught with difficulties relating to the scarcity of an appropriately skilled dedicated medical genetic workforce, limited formalisation of structures, insufficient health funding and political support, and overstretched health systems contending with competing health priorities (134,137,140,142). Difficulties are exacerbated when cultural and religious aspects that may influence or limit engagement with genetic services are also considered (139,143–145). These have not been explored at any significant level. Furthermore, even where services do exist, access to clinical genetic services remains poor, with services often restricted to major urban centres leaving a large proportion of the population of Africa underserved (142,146).

The current practices across African countries have been evaluated by describing the prevailing service delivery models. We propose that there are five genetic service delivery models currently employed in Africa, namely: Uncharacterised, Limited, Disease focused, Emerging and Established, each differentiated by (i) specialist access (accessibility to clinical geneticists and genetic counselling vs. primary care physicians/other medical specialists, such as haematologists and paediatricians); (ii) availability of laboratory testing to confirm diagnoses and (iii) nature of laboratory testing (where and by whom testing is performed and the type of technology used). The models, described in **Figure 6** reflect successive points on a continuum from no service provision (or unknown) to an established comprehensive service.

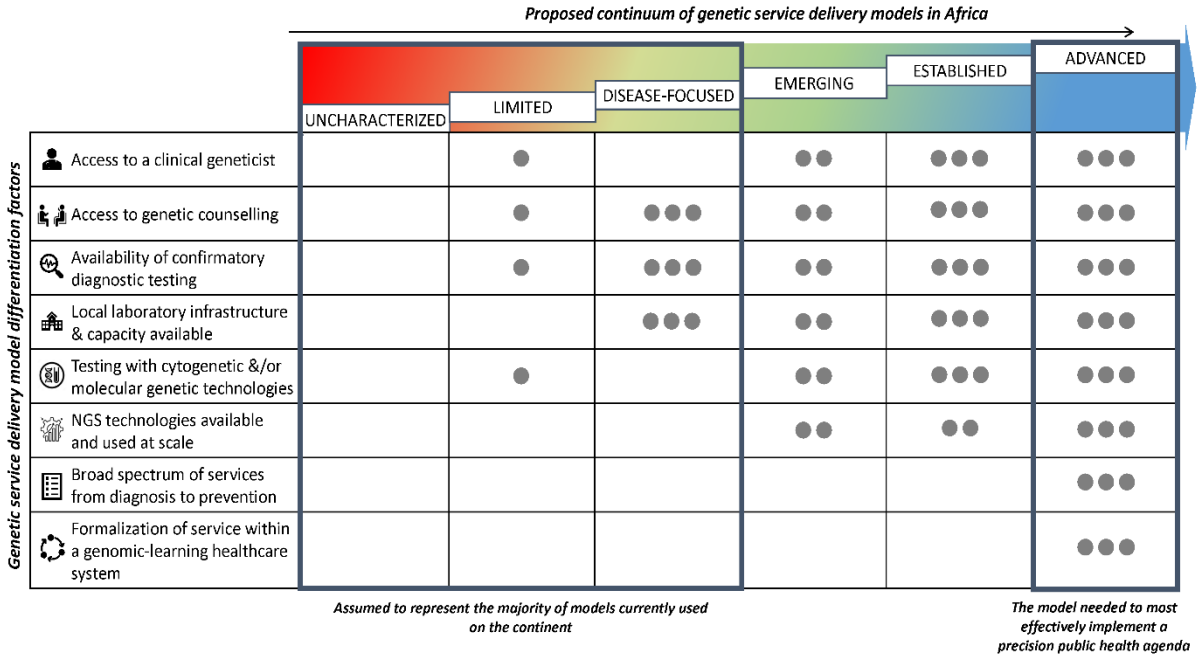


Figure 6: The proposed continuum of genetic service delivery models in Africa

We propose that five genetic service delivery models are currently utilised in Africa. 1. Uncharacterised**:** No service offered or state of service unknown. Limited clinical genetics service may be offered on a non-systematic basis. Individuals involved typically rely on international research collaborations. Few or no resources available for in-country laboratory testing. 2. Limited: Genetic services are provided by non-genetic specialists. Clinical diagnoses are not regularly confirmed with laboratory testing. Genetic testing, when requested, is conducted overseas, often leveraging research collaborations and networks. Services are often limited to centres where interested specialists are located—uneven distribution across city/province/ region/country. 3. Disease-focused: Directed genetic services provided by non-genetic specialists in response to high health burden. Clinical diagnoses confirmed with basic laboratory *testing. Testing* done within the country by locally trained individuals. 4. Emerging: Small number of clinical geneticists provide genetic services. Some clinical diagnoses confirmed by genetic testing. Cytogenetics and limited molecular genetics technologies are available. Testing is conducted both locally and/or internationally depending on the test complexity and the technology required or available. 5. Established: Several clinical geneticists provide formalised genetic services and do so in collaboration with other specialists. Clinical diagnoses are regularly confirmed by laboratory testing. Cytogenetics and molecular genetics technologies are available for testing for an array of common diseases. The majority of testing is done within the country by locally trained capacity.

Uncharacterised, Limited, and Disease-focused represent the majority of models currently operational on the continent. The '*Advanced*' model is deemed necessary to implement a PPH agenda effectively. However, our belief is that this model is not yet operational in Africa. The Advanced model has formalised services, and all

components are present to provide access appropriate for given populations. Also, services are offered beyond just the diagnosis of monogenic conditions and include broad disease prevention strategies (screening, availability of prenatal diagnosis, pre-clinical diagnosis, and genetic risk stratification). In addition, clinical services need to be operating within genomic-learning healthcare systems.

The models we have assessed and described relate to the public sector only. Across most regions private healthcare is available but, generally serves a small and wealthy segment of the population. In general, private care does afford additional testing, for instance, newborn screening programs, and often uses overseas testing facilities when legislation permits this. In most settings, the public sector serves the majority of the population and requires local genetic services capacity and supportive formal structures. In most cases, even in established settings, legislation and policy development relating to service provision is awaiting approval or is not yet developed.

Across all models, services are geared towards confirming diagnosis of common monogenic disorders. Disease prevention strategies, including population screening, risk stratification programs and services associated with complex diseases, are essentially non-existent. Furthermore, testing is focused on affected probands, and subsequent cascade and prenatal diagnosis testing are rare, particularly in limited and disease-focused service settings, due to the lack of financing for such testing or the fear or perceived fear of discrimination amongst family members (144).

In countries with limited or disease-focused models (e.g., focused on sickle cell disease in Central, West and East Africa), genetic services do not extend beyond the common disease of focus and are associated with non-genetic medical specialities, mostly haematology, paediatrics, and pathology. In emerging and established settings, clinical geneticists are active in the system but often at very low numbers, and genetic counselling capacity is severely inadequate (137,147), with access to a trained and certified genetic counsellor only available in South Africa (134,148) or through clinicians (136–138,140,141). In almost all settings, there is a ‘key person(s)’ risk, whereby one or a few individuals are at the core, driving the provision of clinical genetics services, and without them, the system would likely not be sustained.

Furthermore, the retention of medical genetics skills remains challenging, with several countries unable to stem the 'brain drain' (134). This loss is entrenched by a lack of political will, employment and career progression opportunities and limited government financing of the discipline and healthcare system as a whole (91,135). Formal structures and posts are limited or non-existent in most African countries.

Additional barriers impeding genetic services and the more widespread adoption of genomics in Africa are highlighted in **Table 2** and can be classified according to limited opportunities across four dimensions: awareness, accessibility, affordability, and applicability.

Table 2: Barriers impeding genetic services in Africa

Factor	Detail
Awareness	- Patients and healthcare providers' limited knowledge about the availability and benefit of genetic testing restricts testing uptake and limits the understanding of the true burden of genetic disease (149–151). - Lack of awareness amongst the general public contributes to discrimination and social exclusion of affected individuals (144,145). - Funding gaps and a lack of appropriate legislative and regulatory frameworks have resulted from the lack of knowledge and awareness amongst policymakers and key decision-makers (146,152).
Accessibility	- Genetic services are not widely available [54,65,66], and significant clinical genetics and counselling workforce shortages exist [62,67,70]. Limited access to adequate healthcare services, especially in rural areas, exacerbates accessibility challenges as the referral system is compromised [58]. - Scarce laboratory infrastructure and poor financial support for genetic testing and genomic technologies limit the opportunity to broaden genetic services and benefit from translational genomic innovations, particularly those associated with disease prevention and patient management and care [60,66,72,73]. - African healthcare systems lack the multidisciplinary medical specialty resources required for effective genetic disorder patient management and care [57].
Affordability	- Financial barriers are a common hindrance to patients' utilisation of genetic services. National health insurance or subsidies may lower burden but not all tests are covered by national insurance (including cascade testing). - Despite costing decreases in sequencing and associated genomic technologies, they remain out of reach for most African laboratories, especially when coupled with the additional costs, time delays and complex logistics associated with the importation of reagents, and servicing and maintenance of laboratory equipment (39,44). - Limited financial support from national governments for laboratory testing and clinical posts results in a vulnerable system that has a heavy reliance on grant and academic institutional funding to subsidise funding shortfalls.

Applicability	- Africa's burden of genetic disorders remains largely unknown (140,153), and reliable epidemiological data is not available (154). This limited understanding of burden of genetic disease restricts development and subsequent prioritisation of genetic service delivery (146,153). - Limited transferability of mutation profiles for monogenic disorders and genetic associations for complex traits across populations impedes the uptake of genomic interventions into clinical practice. This is exacerbated by already low levels of translation from genomic research. - Need for a culturally competent service that considers and respects the culture, values, and traditions of the individuals and communities accessing genetic services and contributing to genomic research (111,131,132). - Improving the uptake of genomic medicine practice innovations and subsequent quality of clinical care requires locally relevant genetic tests and programs (60,133,155). - Genomic innovations must be successfully translated, scientifically validated and their clinical utility demonstrated in each context. This requires a supportive and robust genomics research community (146,155).

Despite the significant challenges, the African genetics community is generally optimistic about the future of genetic service delivery and the role genomics could play in improving healthcare on the continent (91,155). Africa's clinicians pride themselves on providing a clinical service to their patients and doing so at a level that is on par with international peers despite their resource-constrained environments. However, only a small sector of the population has access to these services (153). The novel discoveries and active research communities that continue to contribute to the global genetics knowledge base despite a challenging operating environment are also a source of optimism.

Moving towards PPH will require new and practical strategies to increase accessibility to genetics and genomics services and to reduce regional or social inequalities. Discussions with the African clinical genetics' community revealed that several critical factors are needed to enhance genetic services on the continent; namely, capacity development (human resources and infrastructure), dedicated funding and support, political will and supporting legislation, public education and awareness, and Africa-relevant genomic and related health economics research, with an overarching need to translate knowledge into improved precision health care.

3. ACCELERATING THE PPH AGENDA FOR AFRICA

Clinical genetics services do not operate in isolation but are nested within a larger healthcare system, which is increasingly translating knowledge gained from genomic research into mainstream health care (2,152,156). On the African continent, there is a need for active intervention in genomic research and clinical genetics, as well as in the overarching healthcare systems within which they function, to reduce genetic service inequalities and accelerate PPH (153,156).

Health systems strengthening is considered fundamental to the successful implementation of genomic medicine in Africa as a functional and well-resourced healthcare setting will provide the platform necessary to elevate Africa's genetic services (109). Factors identified as necessary for health system strengthening include capacity building, increased funding for health, and more political support and commitment (140,157). These factors, along with patient education and awareness, need to be directed specifically to improving clinical services and genomic and clinical research on the continent (109,153,155). Additional accelerators unique to the research and services domain will also be required to advance PPH in Africa. These above-mentioned cross-cutting factors will be described further in the following sections and are illustrated in **Figure 7**.

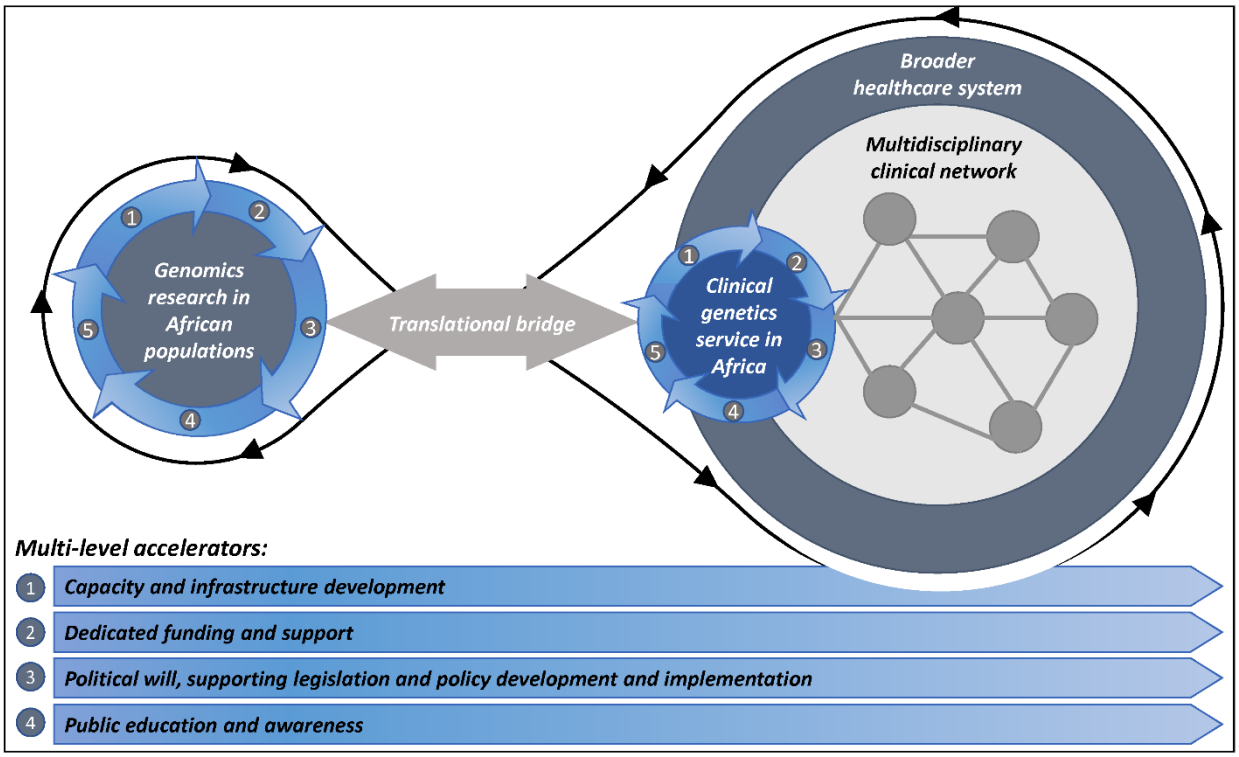


Figure 7: Accelerating the PPH agenda in Africa requires a multi-dimensional approach.

In addition to the multi-dimensional accelerators (capacity and infrastructure, dedicated funding and support, political will and supporting legislation), accelerators are required at the genomic research level (Population genetics and variant profiles, Monogenic disease mutation profiles, Complex disease genomics, Infectious diseases and host response, Pharmacogenomics and Databases & data sharing) and at the clinical genetics level (Patient registries, Prevalence data on genetic disorders, Broad open access, Mutation databases and Healthcare integration). Research discoveries should translate into clinical practice that is beneficial to a broader multidisciplinary clinical network. A virtuous cycle between the research environment and healthcare system should exist where genomic learning is active and continuous, and each domain self-improves after successive advances.

3.1.1. Capacity and infrastructure

African countries vary significantly in the availability of genetics and genomics clinical training programs and infrastructure (146,150). Genomics research and clinical genetics capacity building, both on a human resources and infrastructure level, is urgently required to develop the critical mass of genomics and healthcare expertise needed for translational genomics and precision medicine (2,3,39,48).

Few African countries provide formalised training in clinical genetics and even fewer for genetic counselling (134,135,147). Most clinical geneticists in central and west Africa have been trained at international academic institutions, mainly in Europe. However, the tide is changing, and more African talent is being developed within Africa. Most local and continental genomics training initiatives, such as the H3Africa Consortium (www.h3africa.org) and the African Academy of Sciences (www.aasciences.africa), have promoted and accelerated training opportunities across the continent and contribute to the growing genetics knowledge and exposure amongst the medical fraternity. The African Genomic Medicine Training Initiative (AGMT) (150) and the West African Genetic Medicine Centre (WAGMC) (https://wagmc.org) promise additional training opportunities, providing genomic medicine training for nurses and targeted education on sickle cell anaemia and cancers. These initiatives have not, to date, been involved in training genetic health care professionals, who can deliver dedicated broad high-level genetic counselling and medical genetics services.

Despite this upward trajectory, significant challenges remain. Few efforts train specifically for the healthcare sector or for patient care, but rather for a research setting. In addition, language has been identified as a significant barrier to the uptake of cross-continental training initiatives (158), and the ability of Africa to retain scientific leaders who are capable of developing and maintaining sustainable research programs remains difficult (108).

Clinical and genomics infrastructure shortages further impede the continent's adoption of genomic medicine (107,146,155,159). Robust clinical infrastructure is required for adequate and equitable access to service and to collect, process and store clinical samples destined for diagnostic genomic analysis (160). Additionally, the infrastructure required for analysis and subsequent genomic medicine implementation is expensive and requires considerable expertise and a sophisticated operating environment (155,161). As resources for infrastructure projects remain scarce in Africa (159), developing innovative approaches that leverage existing infrastructure, promote the sharing of facilities at the national and regional level or prioritise subcontracting services to allow efforts to be concentrated on developing services are paramount (109,155). Furthermore, access is but the first step; proper integration of high-

throughput technologies on affordable technical platforms and based on an African knowledge-base is needed if Africans are to reap the health benefits (162).

On a more positive note, Africa's genomics capacity is not restricted to human genetics, and this provides an opportunity for leveraging shared infrastructure, and genetics and bioinformatics expertise. There is significant capacity in infectious disease genomics with institutions such as the African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), the Africa Centre for Disease Control and Prevention (Africa CDC), Nigerian CDC, the Kwa-Zulu-Natal Research Innovation and Sequencing Platform (KRISP) and the National Institute for Communicable Diseases in South Africa. These initiatives spearheaded the response to COVID-19 on the continent (163–165). Furthermore, there are pockets of excellence for agricultural genomics, such as the Agricultural Research Centre (ARC) in South Africa (www.arc.agric.za) and the African Orphan Crops Consortium (http://africanorphancrops.org).

The use of digital tools could provide cost-effective solutions that broaden the reach and efficiency of genomic medicine. Distance-based training initiatives (166) and telegenetics platforms (167) enable increased access to training, testing and counselling resources by reducing traditional cost and time barriers (168). Such approaches would need to be tailored for Africa to assess efficacy and uptake.

While all these initiatives would accelerate the growth of genomics in Africa, dedicated healthcare professionals with broad genetics and genomics training will be required to translate this knowledge into appropriate clinical services. Furthermore, while shared equipment facilities are helpful, human genetics diagnostic laboratories have specific requirements to ensure results are returned to individual patients with high accuracy and appropriate counselling. Interpretation of individual genomic data for feedback to patients needs to be ring-fenced as a distinct requirement.

3.1.2. Dedicated funding and support

Most countries in Africa do not have formal medical specialties in clinical genetics and therefore lack dedicated funding for this medical discipline. South Africa and several North African countries, including Egypt, Morocco and Tunisia, are the exceptions, but

still have challenges in providing broad and equitable services throughout their respective countries (135–138). Genetic services are often considered non-essential in resource-constrained health services and are then only offered where dedicated and informed medical practitioners seek out or develop diagnostic tests and offer patient and family counselling. This hampers access to diagnosis, treatment and prevention strategies, as well as opportunities for family planning across the genetic disease spectrum (153).

There is a need to create new funding opportunities from within the continent, including philanthropic, government and industry funding, that allow for African genomics questions, priorities and intended outcomes to be framed locally, nationally and regionally (159). A variety of co-operative arrangements, between the government and private sector, are becoming increasingly popular in health care due to their ability to mobilise resources, finance and reduce service delivery inefficiencies (169). A venture capital start-up company in Nigeria, 54gene, is developing such a partnership (https://54gene.com) and some genetic researchers in Africa have argued that public–private partnerships are necessary for expansion of genomic medicine on the continent (146).

3.1.3. Political will and supporting legislation

The movement toward integrating genomics into routine medical practice is gaining momentum with country-level, government-funded initiatives underway in at least 14 countries internationally (170). Such large-scale national efforts drive the adoption of genomic medicine through financial support (152), increasing genetic literacy, fostering national and regional collaborations (109), and limiting unnecessary duplication of effort and infrastructure (170). However, no such initiatives exist in any African country as yet (170). The potential for national and regional collaboration across Africa is large.

Some African governments are beginning to recognise the value of local genomic data and have supported research sequencing initiatives (171–173). Recently a call was made to sequence three million African genomes (3MAG) to capture the extent of genetic variation across the continent, which would not only improve global health and research but significantly reduce envisioned health disparities (174). Although such large scale public health genomic projects may not be feasible for now, African

governments must still plan for genomic medicine and prioritise legislation, frameworks and policies that facilitate adequate and equitable genetic service provision and genomics research (108,152,159), with specific goals to translate knowledge gained into improved clinical services.

3.1.4. Public education and awareness

Genomic literacy and awareness are critical in managing genetic diseases, including diagnosis, prevention of complications and therapy (138,144). In Africa, public knowledge of genetics and genomics is still very poor, and few initiatives aim to improve genomic literacy amongst the public despite the potential health benefits (138,150,151,175) and indirect motivation for government buy-in and genomics medicine investment (146). Africa is characterised by great cultural and language diversity, thereby requiring a multidisciplinary approach to improving public and community genomics literacy and engagement (176). The low literacy rates in some Sub-Saharan African countries add a further complication (137,151).

The catalytic role of patient advocacy groups as a driver for change should not be underestimated (127,139,143). For example, the Albinism Society of South Africa (ASSA), which was started in 1994, is very active in providing information on the condition, as well as support services for people with albinism and advocating for their rights (particularly for adequate health care) (177). Similar groups are organised in Kenya, Tanzania, Cameroon and elsewhere (148).

3.1.5. Genomic and public health economics research relevant to Africa

There is unanimous agreement that understanding population genetic variation as it links to health is the gateway to genomic medicine in Africa (91,109). For PPH strategies to be applicable to Africans, African scientists must direct the research agenda and conduct high-quality genomics research that aims to understand health problems relevant to African populations (108,159).

There are almost no large-scale population-based studies in Africa. African participants in GWAS studies contribute <3% of the world genetic data (178), despite harbouring the greatest genetic diversity. This unequal weighting has not only biased our

understanding of genetic disease but has also reduced the potential transferability of translational genomic research. There are also missed opportunities to recognise the value of African data in fine-mapping causal variants given the high diversity and low linkage disequilibrium in African populations (as reviewed in (146)).

Furthermore, there are extremely limited data on mutation profiles for monogenic diseases in Africa, with minimal understanding of the genetic epidemiology. Genomic technologies offer the potential to increase this knowledge rapidly and systematically. Key areas of focus where such interventions offer possibilities for rapid acceleration include baseline characterisation of population variant profiles, monogenic disease mutation profiles, infectious diseases and host responses, pharmacogenomics, and precision therapy for complex diseases. Broad open access databases and data sharing are critical components to ensure accessibility of the data across the continent, for the benefit of individuals on the continent but also for the many individuals distributed across the world who have their origins in Africa.

Concurrent clinical service accelerators are required. Apart from ensuring adequate capacity and training, the establishment of patient registries, biorepositories, electronic health records and mutation databases with broad access are essential for translational genomic medicine. This translation must encompass the key areas of analytical validity, clinical validity, clinically utility and ethical soundness.

Although the economic factors in the establishment of genomics services and their impact on healthcare in low and middle-income countries have not been assessed, it is reasonable to assume that they may have similar value to those demonstrated elsewhere (179–183). On an individual level, families value causal diagnoses to minimise the financial and emotional effects of the diagnostic odyssey. Further value arises from directed management strategies that improve patient health and well-being while expanding evidence-based research to reduce evidentiary uncertainty. The value at a continent-wide level needs to be explored further (184).

4. CONCLUSION

The need for globalisation in research and the recognition of the critical role of African genetic diversity in unravelling the aetiology of disease states (91,109,185) is a useful starting point to justify the development of clinical genetic services in Africa.

Informal discussions with colleagues in different African countries revealed several distinct approaches to genetics service delivery but practicing clinical geneticists could not be identified in most African countries. In some countries, there were no ring-fenced, structured genetic services. Many relied heavily on academic and research environments and funding to support sporadic clinical genetic services. Only a handful had formal training for clinical geneticists and genetic counsellors with dedicated government-supported services.

Service could be expedited by fostering intra- and inter-African institution collaborations to develop and expand the necessary infrastructure and capacity required to promote genomics-based health interventions in Africa. However, there is a need to create new funding opportunities from within the continent, including philanthropic, government and industry funding, for African researchers so that more research questions, priorities and intended outcomes can be framed locally, nationally, and regionally.

Genetic services in Africa need to be culturally competent by considering and respecting the culture, values and traditions of the individuals and communities accessing the genetic services. Much research is needed to explore potential barriers and develop innovative approaches to genetic education and language and the mechanisms best suited to prevent and treat genetic disorders in Africa. Intensive community engagement is required to obtain informed opinions and use these to develop appropriate services. Genetic services need to be based on scientific validity obtained from studies in Africans and should have clinical utility in low-resourced settings to ensure that they lead to health benefits. Overall, this will only be possible with strengthening of underlying general health systems. Considering the limited and fledgling genetic services across most of Africa, the continent has perhaps not yet come of ‘genomics’ age but has a strong nucleus of determined and optimistic clinicians and scientists with a clear vision. There is hope for innovative solutions and technological leapfrogging, but in the absence of political will, and investment in translation through strong health care systems, the people of the continent will need to be patient.

Chapter 2: Publication 2

Ancestry-aligned polygenic risk scores combined with conventional risk factors improve prediction of cardiometabolic outcomes in African populations

This chapter describes the research conducted to develop an integrated risk score for cardiovascular diseases in African populations. This research was performed in order to improve cardiovascular disease risk prediction in African populations. Specifically, the assessment included deriving an integrated risk score that includes genetic and non-genetic factors and compares the predictive utility of population-specific polygenic scores. Given the limited transferability of current polygenic risk scores across populations, developing a score suitable for Africans is crucial to limiting the potential health disparities that could arise from precision-medicine approaches, such as polygenic risk scores. The work described in this chapter is under review at the peer-reviewed journal, *Genome Medicine*:

Kamp, M., Pain, O., Lewis C.M., and Ramsay, M., 2023. Ancestry-aligned polygenic risk scores combined with conventional risk factors improve cardiovascular disease risk stratification in African populations – under review at *Genome Medicine.*

Author contributions: Conceptualisation: MR, CL, OP and MK; Methodology: MR, CL, OP and MK; Formal analysis: MK; Writing – Original draft preparation: MK. Writing – review and editing: MR, CL,OP and MK. Visualisation: MK. Supervision: MR, CL, and OP; Project administration: MK; Resources: MR.

ABSTRACT

Background – Cardiovascular diseases (CVD) are a major health concern in Africa. Identifying and treating high-risk individuals will reduce adverse health outcomes. Current CVD risk calculators are largely unvalidated in African populations and overlook genetic factors. Polygenic scores (PGS) can provide a measure of an individual's genetic susceptibility to CVD and enhance risk prediction beyond traditional risk factors. However, European ancestry bias limits their applicability in genetically diverse populations. To address this, we developed models that integrate genetic and non-genetic data to assess the risk of developing cardiometabolic disease associated outcomes in African populations.

Methods - Summary statistics from a genome-wide association meta-analysis (n=14,126) in African populations were used to derive novel genome-wide PGS for 14 cardiometabolic traits in an independent African target sample (AWI-Gen, n=10,603). Using regression analyses, relationships between each PGS and corresponding cardiometabolic trait, and seven CVD outcomes (CVD, heart attack, stroke, diabetes mellitus, dyslipidaemia, hypertension, and obesity) were assessed. Predictive utility of the genetic data was determined by evaluating elastic net models containing multiple PGS and reference-projected principal components of ancestry. An integrated risk prediction model incorporating genetic and conventional risk factors was developed. To enhance generalisability, nested cross-validation was used when deriving elastic net models.

Results - Our African-specific PGS displayed significant but variable within- and cross-trait prediction (max.R2 = 6.8%, p=1.86 × 10−173). The following PGS were significantly associated with dyslipidaemia: total cholesterol (logOR = 0.210, SE = 0.022, p = 2.18 $x10^{-21}$), and low-density lipoprotein (logOR = -0.141, SE = 0.022, p = 1.30 $x10^{-20}$; the PGS for systolic blood pressure was associated with hypertension (. logOR = 0.150, SE = 0.045, p = 8.34 $x10^{-4}$); and and multiple PGS, specifically body mass index (logOR = 0.131, SE = 0.031, p = 2.22 $x10^{-5}$), hip circumference (logOR = 0.122, SE = 0.029, p = 2.28 $x10^{-5}$), waist circumference (logOR = 0.013, SE = 0.098, p = 8.13 $x10^{-4}$) and weight (logOR = 0.103, SE = 0.029, p = 4.89 $x10^{-5}$) were associated with obesity . Elastic net models combining PGS and reference-projected principal

components significantly improved prediction over PGS alone. Models including genetic and non-genetic risk factors were more predictive than non-genetic models alone (dyslipidaemia: R^2 increase= 2.6%, p=4.45x10^{-12}; hypertension: R^2 increase= 2.6%, p= 2.37x10^{-13}; obesity: R^2 increase = 5.5%, 1.33x10^{-34}).

Conclusion – In African populations, risk prediction models of CVD and associated cardiometabolic traits can be improved by incorporating PGS aligned to the target population and accounting for projected principal components of ancestry. Combining PGS with conventional risk factors further enhances prediction over traditional models based on non-genetic data. The generalisability of international predictive models for CVD to African populations will be enhanced by including data from target populations.

1. INTRODUCTION

Cardiovascular diseases (CVD) are the leading cause of mortality and morbidity worldwide. In 2019, approximately 17.9 million deaths were attributed to CVD, 75% of which occurred in low- and middle-income regions, including Africa (17). Early identification of high-risk patients and timely initiation of appropriate treatment are crucial in mitigating adverse health outcomes associated with CVD. Clinical guidelines recommend using ten-year risk calculators, such as Framingham's and QRISK3, to identify individuals at increased risk of developing CVD (29,186). However, the application of these calculators in Africa is challenging due to limited longitudinal data and limited validation in African populations (22,30). Moreover, most calculators do not consider genetic risk factors, which have been shown to contribute to CVD development (53).

Polygenic scores (PGS), calculated as the weighted sum of trait-associated alleles, can provide a single value estimate of an individual's genetic liability for disease. PGS can potentially improve risk stratification, the systematic process of categorising individuals based on their likelihood of developing a disease, and better identify those at higher risk for developing disease (15,187). Furthermore, including PGS alongside conventional risk factors further improves CVD-risk prediction accuracy (54,55). However, more than 85% of genome-wide association studies (GWAS) have been performed in populations of European ancestry, substantially reducing their predictability in other ancestry groups (59,60,133). This limited portability is likely due

to differences in population allele frequencies (i.e. the variation in the prevalence of specific alleles among different populations), linkage disequilibrium (LD), population-specific causal variants or effects (i.e. the variation in the prevalence of specific alleles among different populations), and potential variations in gene-gene and gene-environment interactions [as reviewed by 8 and 9].

To address these challenges, there is active investment in increasing the representation of diverse populations in GWAS and developing innovative methodological and computational approaches to data analysis (133,188). Research indicates that PGS perform optimally within ancestrally matched populations, including continental African populations due to the increased genetic diversity, low LD, and high population substructure in African populations. Whilst scores that work well across all populations are desired, we believe that developing scores that consider the unique epidemiological characteristics and genetic diversity of African populations is necessary and could inform trans-ancestry method. By integrating population-specific genetic risk factors, we can enhance the accuracy and precision of risk assessment, ultimately improving patient stratification and optimising the allocation of limited healthcare resources (54,57,189). This study aimed to develop and assess an integrated risk score model, considering both genetic and non-genetic factors, for CVD and associated cardiometabolic traits in populations resident in Africa.

2. METHODS

2.1. Study design and overview

The design of this study is illustrated in ***Figure 8***. The objective of the present study was to determine and evaluate the predictive performance of an integrated risk score (IRS), which encompasses genetic and conventional non-genetic factors, for CVD and related cardiometabolic outcomes in African populations. The genetic factors in the IRS consist of PGS and genetic ancestry, as inferred by projected principal components (PC) of ancestry. The non-genetic component included conventional sociodemographic and behavioural variables such as age, sex, smoking status, alcohol consumption, diet-related factors, physical activity, and nightly sleep duration.

Summary statistics from the African Partnership for Chronic Disease Research (APCDR) GWAS meta-analysis for cardiometabolic traits were used to derive PGS. The APCDR cohort consists of 14,126 individuals from different African regions (93). Scores were trained on data from the Africa Wits INDEPTH Partnership for Genomic Research (AWI-Gen) (82) cohort comprising 10,603 individuals from four African countries. Study participants from AWI-Gen resided in Burkina Faso, Ghana, Kenya, and South Africa, while those from ACPDR were from Ghana, Kenya, Nigeria, South Africa, and Uganda. Despite a similar regional mix, there was no recorded sample overlap between the target cohort and the individuals analysed in the GWAS from which summary statistics were obtained.

First, we constructed distinct PGS for each of fourteen cardiometabolic traits, all continuous phenotypes, using the p-value thresholding and LD clumping approach (pT+clump) employing the GenoPred pipeline (https://github.com/opain/GenoPred/tree/master/GenoPredPipe). All approaches utilised in the construction of the PGS, including the quality control of both the base and target data, the calculation of scores at multiple p-value thresholds, and the adjustment for linkage disequilibrium (LD), are standardised and calculated according to proceures outlined by Choi et al. (2020). Scores were derived for the following traits: six anthropometric indices (body mass index (BMI), height, weight, hip circumference, waist circumference, and waist-to-hip ratio (W-H ratio)); two blood pressure measurements (diastolic blood pressure (DBP) and systolic blood pressure (SBP)); four lipid traits (low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein (HDL-C), total cholesterol (TC) and triglycerides (TG)); and two liver function measures (albumin and bilirubin blood serum levels).

Subsequently, the associations of the PGS and non-genetic factors with thirteen of the cardiometabolic traits (excluding bilirubin, which was not measured in AWI-Gen) and seven CVD-associated outcomes, namely CVD, diabetes mellitus, dyslipidaemia, heart attack, hypertension, obesity, and stroke, were tested. Next, to derive and evaluate the predictive utility of genetic, non-genetic, and IRS models, elastic net regression with nested cross-validation (NCV) was used.

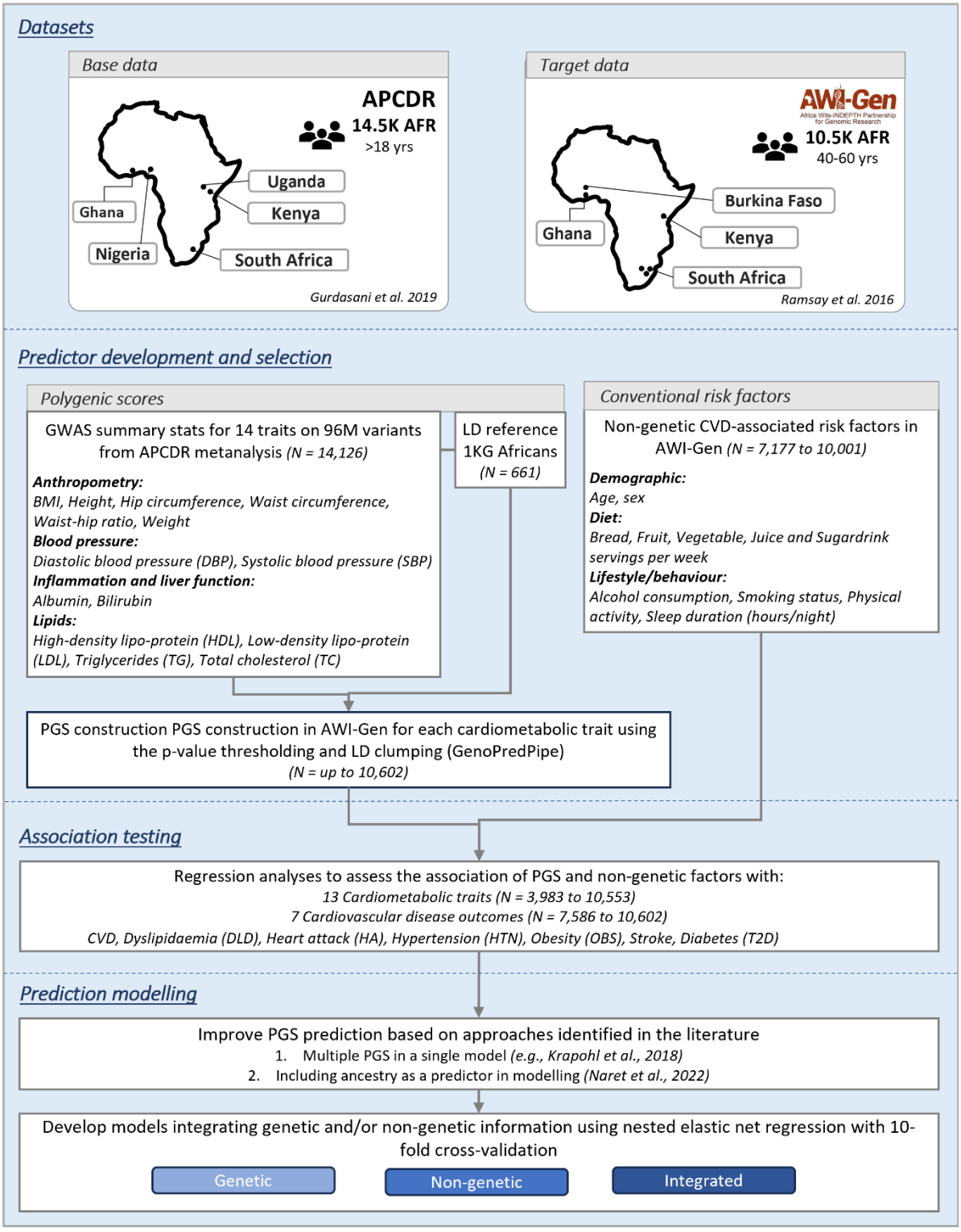


Figure 8: Study design and overview

This study employed two primary datasets, the base and target datasets. Base: The African Partnership for Chronic Disease Research (APCDR) dataset, encompassing 14,126 participants from South Africa, Kenya, Uganda, Ghana, and Nigeria, and the target was the Africa Wits INDEPTH Partnership for Genomic Research (AWI-Gen) dataset which included 10,602 participants from Burkina Faso, Ghana, Kenya, and South Africa. Polygenic risk scores for 14 cardiometabolic traits were derived using the p-value thresholding method combined

with linkage disequilibrium (LD) clumping. The African subset of the 1000 Genomes Project (1KG) served as the reference panel. Predictive modelling evaluated the efficacy of genetic, non-genetic, and integrated models in forecasting disease outcomes. The map included in the figure visually represents the approximate geographical distribution of the cohorts, with the position circles indicating the location. Key acronyms: APCDR - African Partnership for Chronic Disease Research; AWI-Gen - Africa Wits INDEPTH Partnership for Genomic Research; CVD - Cardiovascular Disease; GWAS - Genome-Wide Association Study; LD - Linkage Disequilibrium; PGS - Polygenic Score

2.2. Data sources

2.2.1. Base data set: APCDR Meta-analysis

The APCDR (African Partnership for Chronic Disease Research) is a genome-wide association meta-analysis of association statistics that encompasses association statistics derived from four African cohorts (93). The cohorts included the Uganda Genome Resource (UGR) (n = 6,188), the Africa-America Diabetes Mellitus (AADM) (n = 5,231), the Durban Diabetes Study (DDS) (n = 1,165), and the Durban Case Control (DCC) (n = 1,542) (93,94,190). In brief, the meta-analysis, as conducted by Gurdasani et al. (93), investigated 34 cardiometabolic traits in up to 14,126 individuals aged 18 years and older residing in Ghana, Kenya, Nigeria, South Africa, and Uganda. The APCDR data is publicly available and includes imputed dosage data for all individuals and ~96 million variants. These data were generated using METASOFT (191) with a composite reference panel developed by authors (93). Summary statistics were downloaded from the NHGRI-EBI GWAS Catalog (95) on 01 Jun 2021 for studies GCST009042 to GCST009060 (details of studies are provided in Publication supplementary materials S1) (93).

2.2.2. Target data set: AWI-Gen

AWI-Gen is a cross-sectional cohort study undertaken across four sub-Saharan African countries: Burkina Faso, Ghana, Kenya, and South Africa (82). This study's primary objective is to explore genetic and environmental factors associated with cardiometabolic diseases in Africans. It is part of the Human Heredity and Health in Africa Consortium (H3Africa). From 2012 to 2016, approximately 12,000 participants, primarily between the ages of 40 and 60 years, were enrolled across six study centres, and individual-level genetic, health-related, and phenotypic data relating to lifestyle was collected. Baseline data was used in this study. The study sites are: from South

Africa, the MRC/Wits Agincourt Health and Demographic Surveillance System Site (HDSS) (referred to as Agincourt), the Dikgale HDSS of the University of Limpopo, and the Soweto centre which is coordinated by the South African Medical Research Council/Wits Developmental Pathways for Health Research Unit (DPHRU); in Kenya, the African Population and Health Research Center HDSS in Nairobi; in Ghana, the Navrongo HDSS in the Navrongo Health Research Centre; and in Burkina Faso, the Nanoro HDSS hosted by the Institut de Recherche en Sciences de la Santé Clinical Research Unit (82,83).

Genetic data

Approximately 11,000 individuals were genotyped on the 2.3 M SNP H3Africa array at Illumina® FastTrackTM Microarray services (192). Genotype calling was performed using the Illumina pipeline. Quality control (QC) was performed as described previously, but in summary, pre-imputation QC was performed using the H3ABioNet/H3Agwas pipeline (https://github.com/h3abionet/h3agwas) and variants with a minor allele frequency (MAF) <0.01, missingness >0.05 or Hardy-Weinberg equilibrium p-value <1×10−3 were removed. Additionally, SNPs from the X and Y chromosomes, mitochondrial SNPs, and SNPs that did not match the GRCh37 reference alleles were removed. Samples that were potential duplicates (PIHAT > 0.9), had a missing SNP genotyping rate greater than 0.05, and reported vs. genetic sex inconsistencies were excluded. Population stratification was assessed using principal component (PC) analysis based on an LD-pruned subset of SNPs using the smartPCA program implemented in EIGENSTRAT. Imputation was performed using the African Genome Resources reference panel (EAGLE2 + PBWT pipeline) at the Sanger Imputation Server (https://imputation.sanger.ac.uk/). Post-imputation QC involved removing indels, rare SNPs (MAF ≤ 0.01), and poorly imputed SNPs (Info score ≤ 0.6), resulting in a final dataset containing 10,603 participants and 13.98 M SNPs.

Phenotypic data

In addition to demographic, general health and infection history variables, the AWI-Gen questionnaire provided information on diet, smoking status, alcohol use, physical activity, and sleep. Variables associated with CVDs and used in conventional CVD risk calculators were included in our models and referred to as non-genetic factors throughout our analyses (26,29,90). The variables selected included age, sex, current

smoking status, alcohol consumption status, sleep (hours/night), moderate and vigorous physical activity (minutes/week), juice (number per week) and sugar drinks (number per week).

Current smoking status was obtained from “Yes”, “No” responses to the following question, “Do you currently smoke tobacco?”. Similiarly, alcohol consumption status was determined from “Yes”, “No” responses to the following question “Are you a current alcohol consumer?” Those who prefered not to answer or did not know, were excluded. not inlcuded. The Global Physical Activity Questionnaire (GPAQ) was used to obtain self-reported physical activity. Total moderate-vigorous physical activity (MVPA) in minutes per week was calculated from the accumulation of occupation, travel-related and leisure time physical activity. Sitting time (minutes/week) is used as a proxy for sedentary behaviour (Ref).

Weekly consumption of bread (slices per week), fruit (servings per week), and vegetables (servings per week) was calculated by multiplying the individual's number of servings per day by the number of times a week each respective food group was consumed. Not all the selected variables were available in the Soweto sample; thus, these samples were excluded from the prediction modelling analyses. For the remaining participants, individuals with more than 5% of the data missing among these selected variables were removed. Additional details on the variables and their construction can be found in Publication supplementary materials S2.

2.2.3. Disease outcomes

The outcome variables in this study included 13 cardiometabolic traits and seven CVD-associated outcomes. Cardiometabolic traits included BMI, height, weight, hip circumference, waist circumference, and W-H ratio, DBP, SBP, LDL-C, HDL-C, TC, TG, and albumin and bilirubin blood serum levels. CVD-associated disease outcomes assessed were CVD, diabetes mellitus (T2D), dyslipidaemia (DLD), heart attack (HA), hypertension (HTN), obesity (OBS), and stroke. CVD was defined as present if the participant reported having had a heart attack, stroke, or transient ischaemic attack. Participants previously diagnosed with congestive heart failure or angina were also classified as having CVD. Transient ischaemic attack, congestive heart failure or angina outcomes are not included as single disease endpoints due to small sample

size. Further information regarding the outcome definitions can be found in Publication supplementary materials S3. For disease traits, all cases were included. The maximum sample sizes for each phenotype in the prediction modelling analyses are shown in ***Table 3***.

Table 3: Sample sizes for each disease outcome (e.g., Characteristics of the AWI-Gen cohort for the CVD traits and associated outcomes)

Phenotype	Abbrev	Total sample size	NA	No. cases	No. controls
Cardiovascular disease	**CVD**	**7,586**	**1,517 (20%)**	**219 (2.9%)**	**5,850 (77.1%)**
Dyslipidaemia	**DLD**	**10,602**	**2,121 (20%)**	**5,680 (53.6%)**	**2,801 (26.4%)**
Heart attack	**HA**	**8,598**	**1,719 (20%)**	**46 (0.5%)**	**6,833 (79.5%)**
Hypertension	**HTN**	**10,602**	**2,121 (20%)**	**3,183 (30%)**	**5,298 (50%)**
Stroke	**Stroke**	**9,737**	**1,947 (20%)**	**113 (1.2%)**	**7,677 (78.8%)**
Obesity	**OBS**	**10,602**	**2,121 (20%)**	**1,775 (16.7%)**	**6,706 (63.3%)**
Type 2 Diabetes	**T2D**	**10,537**	**2,109 (20%)**	**568 (5.4%)**	**7,861(74.6%)**

NA refers to participants who reported that they did not know their disease status, and those from the Soweto site – these participants were excluded from modelling analysis given the high level of missingness of non-genetic risk factors. CVD includes heart attack, stroke, or transient ischaemic attack, and participants previously diagnosed with congestive heart failure or angina. Limited case numbers for transient ischaemic attack, congestive heart failure or angina restricted their use as disease endpoinnts themselves. Some individuals experienced multiple CVD outcomes and subsequently a summation of individual outcomes does not equate to the total number of CVD cases reported

2.3. Polygenic scoring

2.3.1. Quality control of datasets

2.3.1.1. QC of Base data: GWAS summary statistics

GWAS summary statistics of traits for inclusion in the study were selected due to their relevance to cardiometabolic disease, presence in the AWI-Gen cohort, as well as sharing a similar ancestry to the target AWI-Gen population. The identified summary statistics underwent a series of quality control (QC) procedures, including the extraction of HapMap3 variants, and the removal of ambiguous variants, or where variants had missing data. Variants were flipped to match the 1000 Genomes Phase 3 (1KG) reference, and then variants were retained if the MAF >0.01 in the African subset of 1KG (1KG AFR), the MAF >0.01 in the GWAS sample, and the INFO >0.6. GWAS summary statistics variants and samples were removed if they (1) had a discordant MAF (>0.2) between the reference and GWAS sample, (2) had reported p-values outside the range of 0 to 1, (3) were duplicates, or (4) had a sample size >3 SD from the median sample size.

2.3.1.2. QC of Target data: Ancestry classification

Individuals in AWI-Gen were assigned to the five super populations present in the 1000 Genomes phase 3 (1KG) reference sample (97), namely European, East Asian, South Asian, African, and Admixed American. Super population membership was predicted using a 1KG reference trained elastic net model consisting of the first six reference-projected genetic principal components (PPCs). Principal components were defined in the 1KG reference using HapMap3 SNPs in common between the 1KG and AWIGEN data with a minor allele frequency >0.05, missingness <0.02 and Hardy-Weinberg p-value >1×10-6. LD pruning for independent variants was performed in PLINK (193) after removal of long-range LD regions (194), using a window size of 1000, step size of 5, and r^2 threshold of 0.2.

A multinomial elastic net model, created using the glmnet R package (96), predicted super population membership in the 1KG reference with 5-fold cross validation. This model, along with reference-derived principal components, was applied to AWI-Gen

for similar predictions. Participants with a predicted probability over 0.5 were assigned to a super population, with all being assigned to the AFR superpopulation as expected.

2.3.2. Score Construction

Typically, a PGS follows the form $\beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \ldots + \beta_n X_n$, where β_k represents the effect size attributed to each allele for a given cardiometabolic trait associated with SNP k. X_k is the number of effect alleles at SNP k, and n is the total number of SNPs in the PGS. To derive the PGS for each trait, we used (1) publicly available GWAS summary statistics described in Gurdasani et al. (2019); extracting the disease-associated variants, the effect allele, the estimated β-coefficient for the effect allele, and the p-value of each genetic variant, and (2) linkage disequilibrium (LD) between genetic variants from the African 1KG LD reference panel (661 Africans) (97). Scores were derived using the pT+clump approach. The pT + clump method is a robust approach that enhances the accuracy and relevance of PGS by selecting the most informative genetic variants while reducing redundancy due to LD. We used default LD-based clumping parameters (r^2 = 0.1, window =250kb) to retain only the single most significant variant within each locus, as overly aggressive LD thresholds can detrimentally affect the predictive power PGS (47). The African subset of the 1000 Genomes (1KG AFR) was used to estimate LD. Ten p-value thresholds were considered (1×10^{-8}, 1×10^{-6}, 1×10^{-4}, 1×10^{-2}, 0.1, 0.2, 0.3, 0.4, 0.5 and 1). Polygenic scores were then calculated in AWI-Gen participants, imputing missing variants using the 1KG AFR allele frequency. In the AWI-Gen sample, 1,104,4026 HapMap3 variants were present. Polygenic scores were standardised (scaled and centred) based on the mean and SD of PGS in the 1KG AFR reference sample. The score calculations were performed using PLINK v1.9 as implemented in the GenoPred pipeline (https://github.com/opain/GenoPred).

2.3.3. Association testing

Following PGS development, regression analysis was used to assess the within- and cross-trait predictive utility of each PGS, and with the seven disease outcomes of interest, while accounting for confounders such as age, sex, and the first eight with-in sample principal components to avoid PGS associations being confounded by population structure (98). Similarly, regression analyses were run to assess the

association between conventional risk factors, including age, smoking status, alcohol consumption, diet-related variables, sleep, and physical activity, with selected disease outcomes. Given the differences by sex of these traits within African populations, the analyses were stratified by sex (13,195,196). The predictive utility of scores was assessed as the proportion of variance explained by the PGS or non-genetic factors, and computed as the R^2. For PGS association testing, R^2 was obtained from a full model including both PGS and covariates (PCs, sex, age, and age-squared) minus the R^2 obtained from a model including covariates alone. R^2 was not adjusted to the liability threshold model due to limited disease prevalence estimates available across Africa and the substantial variation in prevalence noted across cohort sites. For multiple testing, results were corrected for the number of PGS tested for each outcome (i.e., applying a p-value threshold of 0.05/14). We did not correct for the number of p-value thresholds as they are correlated, and a Bonferroni correction would be overly conservative. Prediction accuracy was evaluated as the Pearson correlation between the observed and predicted outcome values. Correlation was used as the main test statistic as it is applicable for both binary and continuous outcomes and standard errors are easily computed.

2.3.4. Ancestry predictor derivation

In addition to PGS, reference-projected genetic principal components (PPCs) were included in prediction models to enhance prediction. Genetic principal components capture major axes of genetic variation, which primarily represent differences in genetic ancestry and can be used to enhance prediction over PGS alone. To prevent overfitting, the principal component SNP-weights should be derived independently of the target sample. Therefore, we used the PPCs described in Section 2.3.1.2, where the first 6 genetic principal components were derived from the 1KG reference, and then projected these principal components into the AWI-Gen target sample.

2.4. Integrated risk score and prediction modelling

Elastic net regression with nested cross-validation (NCV) was used to develop and evaluate the predictive utility of three risk prediction models; specifically genetic, non-genetic, and integrated:

1. **Genetic** -

a. MultiPGS - assessed the predictive utility of utilising multiple PGS compared to single-trait PGS.
b. MultiPGS + Ancestry - assessed the predictive utility of utilising multiple PGS with genetically inferred ancestry, specifically projected principal components.

2. **Non-genetic** - assessed the predictive utility of selected conventional risk factors.
3. **Integrated** - assessed the predictive utility when combining all genetic (MultiPGS and ancestry) and non-genetic predictors.

Elastic net balances feature selection and regularisation to reduce over-fitting and address collinearity among predictors. It combines the properties of both ridge and lasso regression, where similar to ridge regression, elastic net applies a penalty to model coefficients which shrinks them towards zero, thus reducing the impact of less important predictors. And similar to lasso regression, elastic net performs variable selection by setting some coefficients to zero. By balancing the weight of ridge and lasso penalties, this regularisation removes the need for manual selection of predictors and selects and weights the most predictive variables appropriately, reducing redundancy and enhancing model interpretability (99). NCV repeatedly partitioned the dataset into training, validation, and testing sets, and consisted of 5 outer folds with a 90-10 data split (90% training, 10% testing) to provide an unbiased estimate of the predictive utility of the model, and 10 inner folds (80% training, 20% testing) for hyperparameter tuning. The proportion of variance explained by a model was computed as R^2. Hyperparameters were determined using the 'caret' R package, which optimises the RMSE for continuous outcomes and accuracy for binary outcomes.

The predictive performance of the models were assessed by the correlation between observed and predicted values of each model, and the comparative performance of the models were compared using William's test (also known as the Hotelling–William's test) as implemented by the 'psych' R package's 'paired. r' function. The code used to prepare data and conduct analyses is available on the GenoPredPipe GitHub page (see Data and Code Availability).

For the MultiPGS, genetic and integrated models, for each cardiometabolic trait, rather than selecting the single best-performing PGS (based on max R^2), all PGS

demonstrating significance after multiple testing corrections were retained for subsequent predictive modelling analyses were retained for subsequent predictive modelling analyses and elastic net regression was utilised to simultaneously select and weight predictors (52,197). Genetically inferred ancestry was included in prediction models to account for population stratification and potentially improve prediction (198). To reduce overfitting, ancestry was determined by fitting data to the 1KG Phase 3 projected principal components (PPCs) of population structure and not to AWI-Gen sample PCs.

Non-genetic models included ten factors selected based on data availability in AWI-Gen and their known association with CVDs. Integrated models included genetic (PGS and PPCs) and non-genetic models. No data were available for diet-related variables for the Soweto study site for men and women, so this site was excluded from prediction modelling analyses.

2.5. Statistical analysis

All analyses were performed using PLINK v1.9 (https://www.cog-genomics.org/plink/1.9/) (102), and R version 3.4.4 (http://www.r-project.org/) (103) unless specified otherwise. Data are presented as percentages (%) or mean ± SD. Associations between non-genetic factors and PGS and CVD-associated outcomes were assessed by logistic regression with the adjustment of PCs, sex, age, and age squared as described previously (98).

2.6. Ethics statement

This study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand (Wits)(protocol number M210355) as a sub-study of the AWI-Gen project (protocol number M170880).

3. RESULTS

Fourteen PGS were derived, and their within- and cross-trait associations assessed using regression analyses. Elastic net regression with nested 10-fold cross-validation was used to determine the predictive utility of models (genetic, non-genetic, and

integrated), and the performance for each PGS and model was assessed using the correlation between observed and fitted values.

3.1. Derivation, validation, and association testing of PGS

All PGS, except those for albumin and waist-hip ratio, had at least one significant association after correcting for multiple testing (**Figure 9**). There was extensive variability in variance explained across phenotypes, with the variance explained ranging (R^2) from 0.068 (p = 1.86 x 10^{-173}) for LDL to 0.004 (p = 4.70 x 10^{-20}) for height. The most significant associations were found amongst lipid traits (HDL, LDL, TC, and TG). Given the genetic correlation across traits, especially amongst anthropometric and lipid traits, significant cross-prediction was also noted (max R^2 = 0.068, p = 6.23 x 10^{-175} for TC PGS predicting LDL). Table S4 and S5 lists all significant within and cross-trait predictions after correcting for multiple testing.

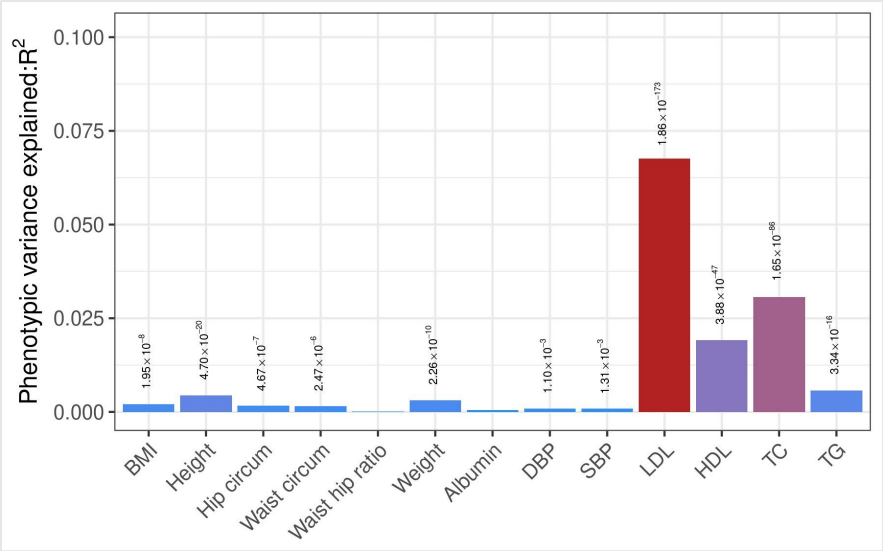


Figure 9: Within-trait prediction of derived PGS.

Variance explained (R-squared) by the PGS for 13 cardiometabolic traits in the AWI-Gen target sample across P-value thresholds. Figure only shows the results for the PGS with the highest variance explained. The values above the bars are P-values indicating whether the variance explained is significantly different from zero. Within-trait predictive utility, described as phenotypic variance explained (R2), of 13 of the derived PGS. *DBP – diastolic blood pressure, SBP – systolic blood pressure, LDL – low-density lipoprotein, HDL – high-density lipoprotein, TC – total cholesterol, TG – triglycerides*.

CVD-associated disease outcomes of DLD, OBS and HTN were significantly associated with the PGS (**Figure 10**). Four PGS (BMI, hip circumference, weight, and waist circumference) were associated with increased risk of OBS, with the largest increase in risk linked to the PGS for BMI (max. logOR = 0.131, SE = 0.031, p = 2.22 $x10^{-5}$). Similarly, three PGS (TC, LDL and HDL) were associated with DLD, with the greatest increase in risk linked to the PGS for HDL (max. logOR = 0.210, SE 0.022, p = 2.18 $x10^{-21}$). For HTN, only the PGS for SBP was associated (max. logOR = 0.150, SE 0.045, p = 8.34 $x10^{-4}$). No significant associations were found for CVD, HA, T2D and stroke disease outcomes (All PGS significant associations provided in Publication supplementary materials S4, S5, and S6). Distribution assessments (mean, standard deviation, interquartile range) of derived PGS across AWI-Gen sites were too small to accurately contrast effect sizes between populations.

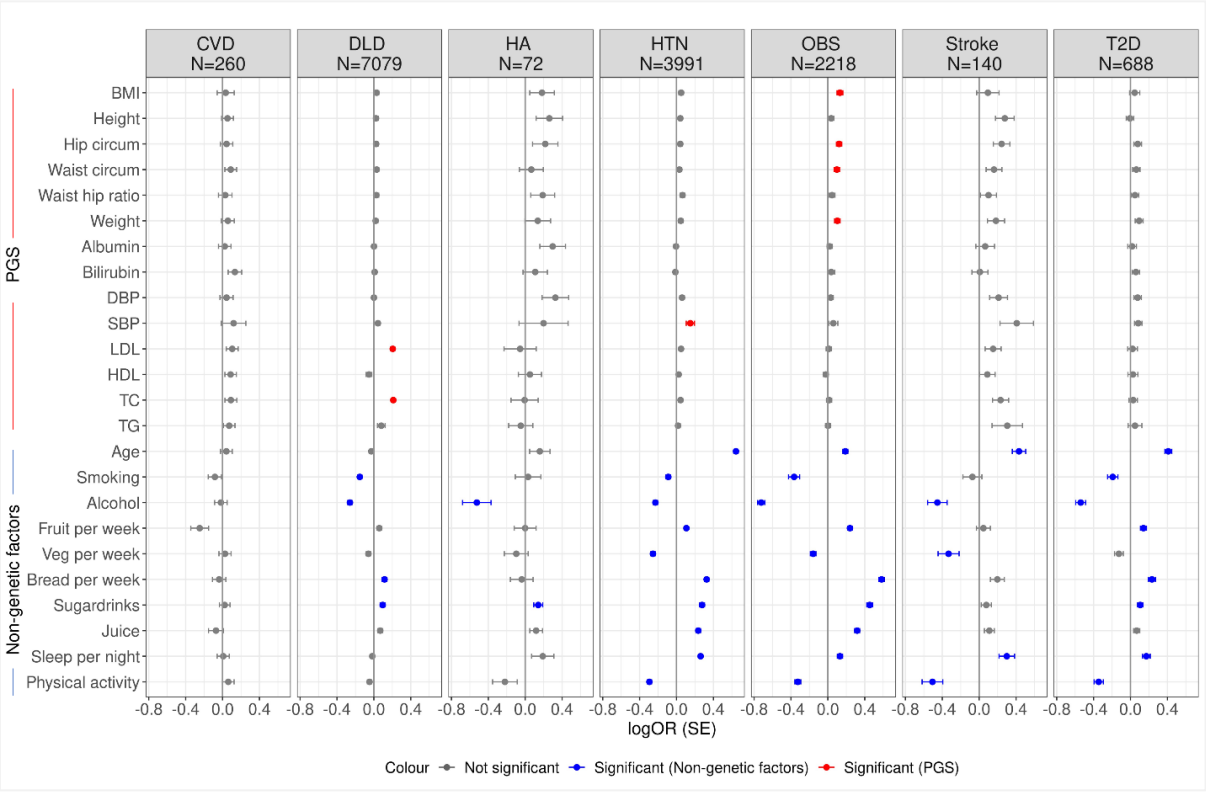


Figure 10: Association analysis of PGS and non-genetic factors with seven CVD-associated outcomes.

Logistic regression analyses assessing the association of each of the 14 derived PGS and 10 non-genetic factors with the seven selected CVD-disease outcomes. Significant associations with PGS are coloured in red, significant associations with non-genetic factors are coloured in blue. Maximum case numbers are indicated by N=. CVD – Cardiovascular disease, DLD – Dyslipidaemia, HA – Heart attack, HTN –

Hypertension, OBS – Obesity, T2D – Type 2 diabetes. DBP – diastolic blood pressure, SBP – systolic blood pressure, LDL – low-density lipoprotein, HDL – high-density lipoprotein, TC – total cholesterol, TG –triglycerides.

3.2. Non-genetic factor association with CVD disease outcomes

Factors previously identified as associated with CVD were selected from the AWI-Gen study: age, sex, smoking status, alcohol consumption, various diet factors, physical activity, and sleep (hours per night). The dietary variables included weekly consumption of fruit, vegetables, bread, fruit juice and sugar drinks. Using regression analysis, and accounting for confounders such as age, sex and principal components, the relationship between each non-genetic factor and disease outcome was assessed. These non-genetic factors accounted for little to no variance explained in CVD and HA (**Figure 10** and Publication supplementary materials S7). In contrast, all factors were associated with HTN (max. logOR = 0.64, SE = 0.023, p = 9.75 $x10^{-173}$) and OBS (max. logOR = 0. 58, SE = 0.029, p = 1.23 $x10^{-92}$). Age was the most significant predictor of HTN ($logOR_{AGE}$ = 0.64, SE = 0.023, p = 9.75 $x10^{-173}$), T2D ($logOR_{AGE}$ = 0.41, SE = 0.037, p = 5.32 $x10^{-28}$), and stroke ($logOR_{AGE}$ = 0.43, SE = 0.073, p = 3.05 $x10^{-9}$); whilst bread servings per week accounted for greatest increased odds in OBS ($logOR_{BREAD}$ = 0.58, SE = 0.029, p = 1.23 $x10^{-92}$) and DLD ($logOR_{BREAD}$ = 0.12, SE = 0.025, p = 5.42 $x10^{-06}$). Alcohol consumption was associated with reduced odds in all diseases except CVD. This effect was most pronounced in OBS ($logOR_{ALC}$ = -0.72, SE = 0.039, p = 9.17 $x10^{-77}$), T2D ($logOR_{ALC}$ = -0.54, SE = 0.054, p = 6.98 $x10^{-24}$), and HA ($logOR_{ALC}$ = -0.53, SE = 0.154, p = 6.73 $x10^{-04}$) (**Figure 5**).

3.3. Prediction modelling for CVD disease outcomes

3.3.1. Genetic prediction models

Models consisting of all PGS combined (MultiPGS) were significantly associated with four outcomes: DLD (p = 1.81 x 10^{-37}), HTN (p = 1.84 x 10^{-162}), OBS (1.69 x 10^{-4}), and T2D (p = 9.43 $x10^{-17}$) (***Figure 11*** orange bar). The greatest improvement was noted in HTN, where phenotypic variance explained more than doubled from 2.8% to 6.7%, followed by OBS (4.0% to 7.5%) and DLD (1.2% to 1.5%). Prediction in the T2D

MultiPGS model remained low, increasing from 0.3% to 0.7%. The MultiPGS model for CVD and HA did not significantly improve prediction (Publication supplementary materials S8).

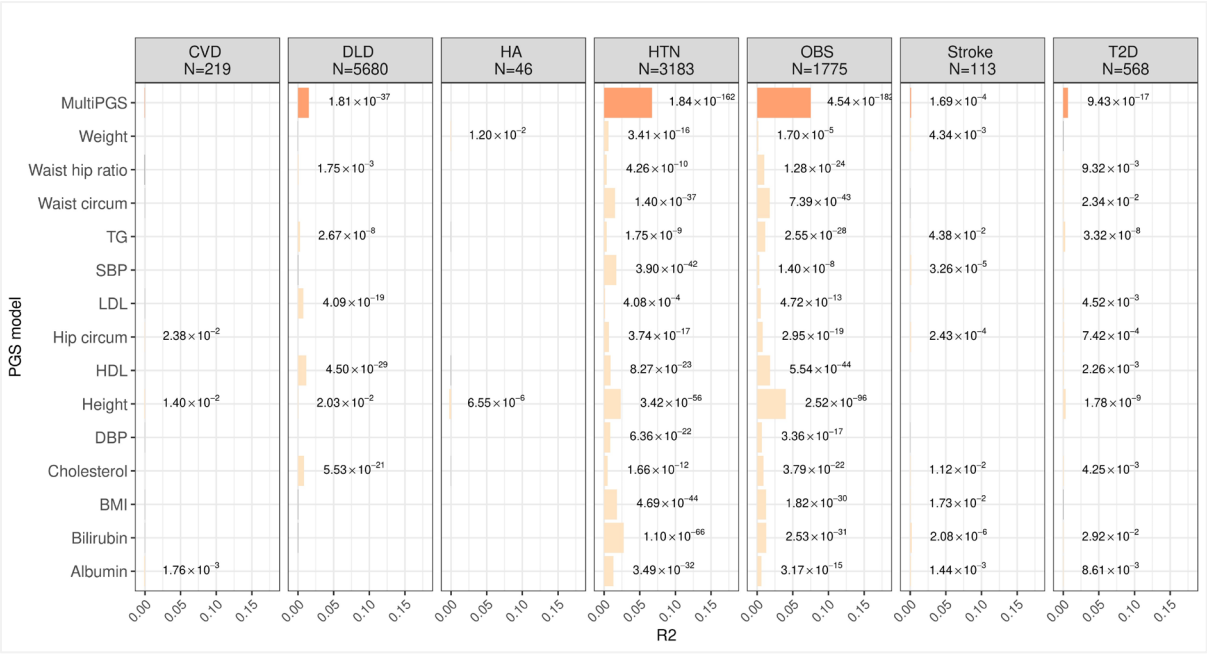


Figure 11: Predictive utility of the single best PGS and MultiPGS models for seven CVD-associated outcomes.

The predictive utility of a single PGS was compared with that inclusive of all cardiometabolic trait PGS (MultiPGS) across seven selected CVD-disease outcomes. Single trait PGS are shown in light orange, and MultiPGS shown in dark orange. Significant associations have p-values displayed. Case numbers for each phenotype are indicated by N=. *CVD – Cardiovascular disease, DLD – Dyslipidaemia, HA – Heart attack, HTN – Hypertension, OBS – Obesity, T2D – Type 2 diabetes. DBP – diastolic blood pressure, SBP – systolic blood pressure, LDL – low-density lipo-protein, HDL – high-density lipo-protein, TC – total cholesterol, TG –triglycerides.*

To further improve prediction of PGS, we assessed whether accounting for population stratification through the incorporation of ancestry could improve performance. As per **Figure 12**, including an ancestry predictor improved prediction, by approximately 2.5% for both HTN (R^2_{PGS} = 6.5%, R^2_{ANS} = 8.9%, $R^2_{PGS+ANS}$ = 9%) and OBS (R^2_{PGS} = 7.1%, R^2_{ANS} = 7.5%, $R^2_{PGS+ANS}$ = 10%), but had little to no effect for DLD, stroke, and T2D with improvements less than 0.1% for each. For DLD, the PGS accounted for greater variance explained and including ancestry reduced prediction performance by 0.1% (R^2_{PGS} = 1.7%, R^2_{ANS} = 0.1%, $R^2_{PGS+ANS}$ = 1.6%). Similarly, to models containing

multiple PGS models, CVD and HA PGS + Ancestry models showed no significant prediction (Publication supplementary materials S9).

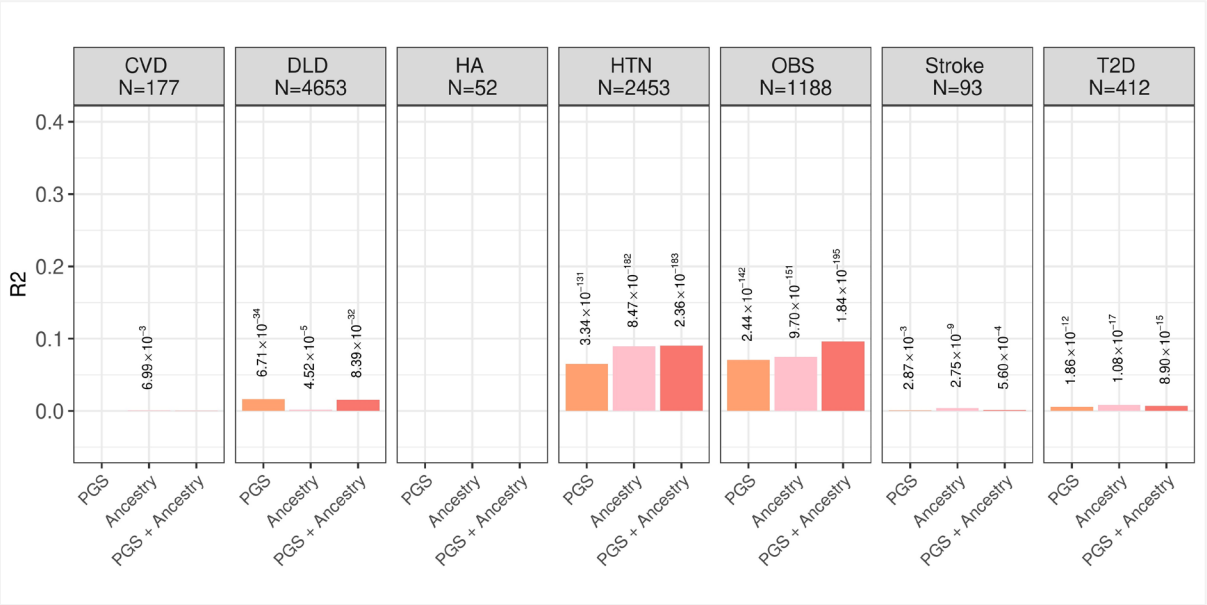


Figure 12: The predictive utility of genetic models (including PGS and projected principal components) for CVD-associated disease outcomes.

Comparing the predictive utility of models inclusive of PGS (orange), projected ancestry (pink), and PGS + Projected Ancestry (red) across seven selected CVD-disease outcomes. Significant associations have p-values displayed. *CVD – Cardiovascular disease, DLD – Dyslipidaemia, HA – Heart attack, HTN – Hypertension, OBS – Obesity, T2D – Type 2 diabetes. DBP – diastolic blood pressure, SBP – systolic blood pressure, LDL – low-density lipoprotein, HDL – high-density lipoprotein, TC – total cholesterol, TG – triglycerides.*

3.4. Integrated prediction models

Lastly, we assessed whether including genetic (PGS and PPCs) and non-genetic factors in an IRS model could improve the predictive performance of models for selected disease outcomes in African populations. The integrated model consisted of 143 predictors and a maximum of 8,057 individuals. As per **Figure 13**, when we examined the non-genetic and genetic contribution to total variation, the non-genetic factors (in blue) consistently accounted for greater variance explained than genetic factors, the greatest prediction from non-genetic factors seen in OBS (R^2_{Nongen} = 19%) and the lowest in Stroke (R^2_{Nongen} = 0.3%). Within DLD, HTN and OBS, non-genetic risk factors explained 7.4% to 9% of the total variation in fully adjusted models depending on the outcome. The predictive utility of genetic models (in red) was variable

across outcomes, with the highest R^2 noted in HTN (R^2_{Gen} = 8%) and OBS (R^2_{Gen} = 9%). For CVD, HA and Stroke, genetic models yielded no significant phenotypic variance explained. In contrast, the IRS models (in green) showed significant prediction in all outcomes except CVD. Models including genetic predictors minimally reduced the prediction of Stroke, HA, and T2D (by 0.5% or less) but increased the prediction of OBS, HTN, and DLD. OBS prediction increased by 5.5% to reach 21% (R^2_{Gen} = 9.3%, R^2_{Nongen} = 15.3%, R^2_{IRS} = 20.9%), HTN by 2.6% to reach 14% (R^2_{Gen} = 8.3%, R^2_{Nongen} = 11.2%, R^2_{IRS} = 13.8%), and DLD by 2.6% to 15% (R^2_{Gen} = 2.9%, R^2_{Nongen} = 11.5%, R^2_{IRS} = 14.1%) (Publication supplementary materials S10A). A pairwise comparison of scores was performed for each model to show the difference in correlation within and between models for outcomes, with p-values for significant differences calculated using the William's test results (Publication supplementary materials S10B and S10C). The integrated models were significantly different to those including non-genetic factors alone for the same traits, suggesting genetic information provides independent and complementary information to non-genetic risk factors for risk prediction (DLD: $p = 4.45 \times 10^{-12}$; HTN: $p = 2.37 \times 10^{-13}$; OBS: $p = 1.33 \times 10^{-34}$). A pairwise comparison of scores was performed for each model to show the difference in correlation within and between models for outcomes, with p-values for significant differences calculated using the William's test results (Publication supplementary materials S10B and S10C).

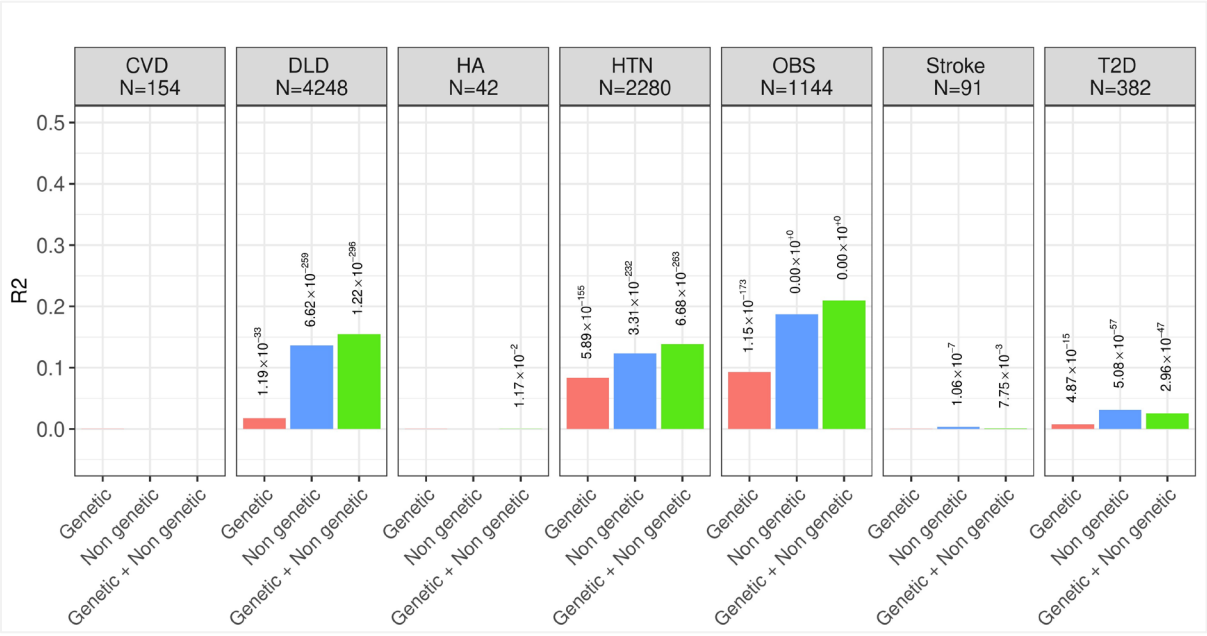


Figure 13: The predictive utility of integrated models (including genetic and non-genetic factors) for four CVD-associated disease outcomes.

Prediction modelling analyses assessing the predictive performance of integrated models across seven selected CVD-disease outcomes. Predictive performance, measured as R2, of Genetic models (PGS + Ancestry), Non-genetic models, and Integrated models, including Genetic and Non-genetic predictors, and denoted as 'Genetic + Non genetic'. Genetic models are shown in red, Non-genetic in blue and Integrated in green. Significant associations have p-values displayed. 0.00 $x10^{+00}$ indicates a significance value of ≤1 $x10^{-300}$. *CVD – Cardiovascular disease, DLD – Dyslipidaemia, HA – Heart attack, HTN – Hypertension, OBS – Obesity, T2D – Type 2 diabetes. DBP – diastolic blood pressure, SBP – systolic blood pressure, LDL – low-density lipoprotein, HDL – high-density lipoprotein, TC – total cholesterol, TG – triglycerides*.

4. DISCUSSION

This study developed and evaluated the predictive utility of 14 cardiometabolic trait ancestry-aligned PGS for seven CVD-associated outcomes in continental African populations. We investigated whether modelling these PGS into a MultiPGS improved prediction and if integrating this MultiPGS with established non-genetic risk factors had greater predictive utility beyond non-genetic factors alone. To date, AWI-Gen provides the largest sample size for the prediction of CVD and associated cardiometabolic traits in continental Africa across genetic and traditional risk factors. The study supports the use of genomic information for enhanced CVD and associated disease outcome risk stratification in African populations.

PGS of cardiometabolic biomarkers showed significant yet variable prediction within and across traits in continental African populations. As previously observed in other populations, lipid measurements had the highest variance explained, likely due to higher reported heritability compared to obesity/anthropometric, and blood pressure traits (199). Although cardiometabolic trait PGS predict several CVD-associated outcomes (Dyslipidaemia, Hypertension and Obesity) in African populations, their predictive utility is relatively modest and typically explain less phenotypic variance than conventional non-genetic risk factors (189). This is in part due to limited GWAS sample sizes within continental populations, although previous research has shown the substantial value of using ancestry-aligned GWAS compared to those derived from genetically distant populations (98,200).

Nonetheless, biomarker PGS provide a potential preliminary step in PM-based risk stratification in Africa - particularly considering the limited availability of disease-specific cohorts in Africa and potential to leverage more readily accessible and routinely collected biomarker data for identification of at-risk individuals. As disease-specific variant data becomes available in African populations, future research should investigate whether models including PGS of both disease outcome and biomarker/trait PGS enhance predictive accuracy and clinical application as seen in other studies (201).

Our research also revealed that combining the PGS of multiple related traits into a single model - MultiPGS - substantially increased the predictive performance for dyslipidaemia, hypertension, obesity, stroke and diabetes compared with single-score predictor models (52,197,201). This improvement is partly due to genetic correlation among traits and leveraging the increased statistical power among discovery GWAS (variance explained and sample size), of either genetically proximal or distantly related traits. Sinnott-Armstrong et al. noted improved predictive accuracy of aggregated PGS – trait PGS plus 35 cardiometabolic PGS - across multiple disease outcomes in European ancestry populations, albeit with limited generalisability to other populations (201).

By including predictors capturing genetic ancestry through projected principal components, we achieved enhanced risk prediction across our seven phenotypes to varying degrees. Naret et al., who investigated the inclusion of ancestry in prediction models using a different approach, observed similar improvements, and hypothesised that traits that show a greater gain in prediction are likely those that are influenced by more population-specific alleles (198). This underscores the importance of accounting for the substantial genetic diversity and regional variation in Africa, since African populations are markedly heterogeneous and exhibit greater genetic diversity than any other ancestry group (171,185). Consequently, a prediction score developed in one African region may not be universally applicable across all African populations (98,202). This regional variation necessitates the development of region-specific models or the inclusion of a wider array of African genetic data in global models to reflect the genetic diversity and population structure of Africans more accurately. For example, in the 1KG reference data, continental representation is limited to Gambian individuals from the Gambia, Esan and Yoruba in Nigeria, Luhya in Kenya, and Mende

in Sierra Leone (97). This approach inadequately captures genetic diversity among Africans and thus may skew prediction models, highlighting the need for broader genetic data representation.

Integrating genetic factors with non-genetic risk factors showed statistically significant improved in predicting dyslipidaemia, hypertension, and obesity. We show that genetic factors provide independent but complementary information in risk prediction models, a finding supported by previous research assessing integrated models for CAD and CVD in primarily European populations (54,55,203).

This research demonstrates the potential of using genetic information, such as PGS, to improve CVD risk calculators in African populations. It also highlights the need for additional research investigating the generalisability of models across diverse African populations and other ancestries, noting that previous work showed that PGSs built from African Americans increased prediction in sub-Saharan Africans compared to using a European GWAS, but not consistently across African populations (202). As data from continental African cohorts grow, the opportunities for validation and refinement of these scores is also expected to increase. However, robust assessment of the clinical, financial, and system benefits these scores provide is crucial to gauge their true translational value (64). Translation will require an understanding of the sensitivity and specificity of scores in different African populations, whilst considering, cost, accessibility, and acceptance within the health ecosystem. In clinical contexts, particularly for PGS, distinguishing between relative and absolute risk is important. Relative risk assesses the connection between genetic traits and disease, contrasting disease incidence in individuals with specific genetic markers against those without. Absolute risk, incorporating factors like age and population disease prevalence, however, reflects the actual probability of disease development. While PGS may indicate increased relative risk, this does not necessarily equate to a high absolute risk. Tools to estimate absolute risk from relative risk (204) require accurate disease prevalence and other data, and their applicability remains limited in the African context due to the paucity of such data. In addition, considerations for resource-constrained settings are essential for the successful integration of genetic approaches into routine practice (205,206).

This study is unique for its use of data solely from continental Africa, a region often underrepresented in genetic research. However, although the study made use of the largest dataset of continental African populations, sample size is orders of magnitude smaller than many European and multi-ancestry studies (55,57,207). Some limitations must be noted. Firstly, the APCDR dataset meta-analysis includes two population-based cohorts and two-disease-based cohorts, specifically diabetes, which may potentially result to an overestimation of genetic factors associated with diabetes and related conditions in the GWAS results. Secondly, while the pT + clump method has been used given its robustness in calculating PGS in different ancestries, it would be valuable for future research to compare and evaluate performance of multiple PGS construction methods in African ancestry populations. Thirdly, is important to acknowledge the extensive diversity within the continent (185), which means that despite both the base and target datasets having representation from multiple African regions, and a similar regional mix, our base and target data are not necessarily ancestrally aligned. The strength of this approach lies in its focus on African-specific genetic profiles, addressing a critical gap in current genomic research which often generalises findings from non-African populations, and contributes to a more tailored understanding of CVD prediction and prevention for African populations. However, the extensive variability means that a model developed for one African region may not generalise across the continent. Also, despite using nested 10-fold CV to reduce overfitting, validation in an appropriate continental African dataset was not possible given the scarcity of such cohorts. Also, the relatively small GWAS sample sizes and imputation efficiencies in African GWAS studies may affect the precision of the estimated impact of individual variants on disease risk. Similarly, case numbers were insufficient to examine CVD and associated disease outcomes such as stroke and heart attack. Lastly, previous studies have shown the added value of including PGS in clinical risk scores, such as Framingham Risk Score and QRISK (57,208). AWI-Gen data does not yet have sufficient 10-year longitudinal data available to undertake such performance comparisons.

Future research should employ large, diverse multi-ancestry cohorts, once these become available, to overcome sample size limitations, reduce overfitting, and enhance generalisability. It is important to note existing multi-ancestry cohorts often disproportionately represent African ancestry through African American participants. To

more accurately capture genetic diversity and enhance research applicability, it is essential to include a broader representation of African ancestry individuals from regions across Africa. In addition, moving to models that account for gene-gene and gene-environment interactions will further advance our understanding in this field.

5. CONCLUSION

The integrated risk score derived in this study demonstrates the value of including genetic and non-genetic risk information for improving CVD risk prediction in African populations. This approach could provide more accurate and personalised risk assessment, tailoring prevention and treatment strategies more effectively. The inclusion of genetic information improves prediction performance over and above traditional non-genetic factors. In African populations, ancestry accounts for a substantial proportion of variance in CVD prediction models, and modelling this variance through principal components, suggests a promising direction for model refinement. However, improving these models will require research across diverse African populations and the use of appropriately ancestry-aligned cohorts. Improving access to extensive African datasets is crucial for refining CVD prediction models and necessary to effectively address health disparities both on the African continent and among global African-ancestry populations.

Chapter 3: Publication 3

Clinicians' Perceptions Towards Precision Medicine Tools for CVD Risk Stratification in South Africa

This chapter describes original research that was performed in order to explore the clinical utility of precision medicine-based approaches, specifically an integrated risk score for cardiovascular disease, in a resource-constrained setting such as South Africa. One critical element of clinical utility is user adoption, and as clinicians would be the primary implementers and at the interface between patient and genomic centres, understanding their perspectives is paramount to successful implementation. This research aimed to gain an understanding of clinicians' perceptions of precision medicine-based medicine and their willingness to integrate polygenic risk scores into their clinical practice. The work described in this chapter was published in the peer-reviewed journal, Journal of Personalized Medicine (Appendix G):

Kamp M, Pain O, May A, Lewis CM, Ramsay M. 2022. Clinicians' Perceptions towards Precision Medicine Tools for Cardiovascular Disease Risk Stratification in South Africa. *Journal of Personalized Medicine*. 12(9):1360. https://doi.org/10.3390/jpm12091360

Author Contributions: Conceptualisation: MK, MR; Methodology: MK, MR; Formal analysis, MK, OP, and AM; Investigation: MK; Resources: MR; Data curation: MK; Writing—original draft preparation: MK; writing—review and editing: MK, MR, CL, OP, AM; Visualization: MK, OP; Supervision: MR, CL, OP; Project administration, MK, Funding acquisition: MR.

ABSTRACT

Objectives: Cardiovascular diseases (CVDs) are a leading cause of mortality and morbidity in South Africa. Risk stratification is the preferred approach to disease prevention, but identifying patients at high risk for CVD remains challenging. Assessing genetic risk could improve stratification and inform a clinically relevant precision medicine (PM) approach. Clinicians are critical to PM adoption; thus, this study explores practising clinicians' perceptions of PM-based CVD risk stratification in South Africa's public health setting.

Design: Practising clinicians (n = 109) at four teaching hospitals in Johannesburg, South Africa, completed an electronic self-administered survey. The effect of demographic and professional characteristics on PM-based CVD risk stratification perceptions was assessed.

Results: Fewer than 25% of respondents used clinical genetic testing, and 14% had formal genetics training. 78% had a low mean knowledge score, with higher scores associated with genetic training ($p < 0.0005$) and research involvement ($p < 0.05$). Despite limited knowledge and resources, 84% perceived PM approaches positively. 57% felt confident in applying the PM-based approach, with those already undertaking CVD risk stratification more confident ($p < 0.001$). High cost and limited access to genetics services are key barriers.

Conclusion: Integrating genetic information into established clinical tools will likely increase confidence in using PM approaches. Addressing the genetics training gap and investment into the country's genomics capacity is needed to advance PM in South Africa.

1. INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of mortality globally (209). South Africans are at risk as they have a high and increasing prevalence of obesity, hypertension and undiagnosed diabetes (13). In 2019, CVDs were the leading cause of mortality (15.85% of deaths) and the third leading cause of morbidity (7.08% of DALYs) in South Africa (210). This growing epidemic is exacerbated by the lack of preventative screening, early diagnosis, and inadequate access to healthcare services.

The burden associated with CVDs emphasises the need for intensified efforts to identify high-risk patients and improve prevention, diagnosis, and management.

Globally, and within South Africa, a risk-based prevention strategy is the most widely accepted approach to disease prevention, with high-risk individuals offered preventative medication and encouraged to adopt healthier lifestyles (25,211). Clinical practice guidelines advocate risk calculators to estimate the 10-year risk of disease in order to identify and target high-risk individuals (25,29,90,212). Risk stratification is calculated from multiple demographic and clinical factors, such as age, sex, smoking status, and blood lipid levels. Despite widespread use, it is estimated that these calculators fail to identify up to 40% of high-risk patients (208,213), an underestimation likely greater in Africa, where scores are yet to be validated in local populations (30). Moreover, although genetics is a known risk factor for CVD (43,53), current clinical tools do not assess genetic risk directly.

The summation of genetic variants into a polygenic risk score (PRS), provide an estimate of genetic risk for disease based on the aggregate effect of many common variants, as originally identified in case-control GWAS studies, and is able to stratify individuals according to their genetic risk (46,53). Integrating PRSs with traditional risk factors can improve the prediction of CVD and associated traits (54,55,208,214). Such integrated tools, or precision medicine (PM) approaches, where individual genetic variation, lifestyle, and health history are leveraged to tailor prevention and treatment strategies, have the potential to improve patient care and reduce the burden of CVDs on healthcare systems (91,205,206).

Despite the anticipated benefits of PM, significant barriers to integrating genomics into clinical practice remain (91,108,109,205,206). In addition to the scientific, educational, ethical, legal, and social barriers, the type of testing and physician-based characteristics may also play a role (215). Given that clinicians are the interface between healthcare systems and patients, understanding their perceptions and willingness to utilise PM-based approaches is critical to ensuring patients reap the benefits of PM in a timely and cost-effective manner, especially where healthcare systems are fragile and under-resourced (215,216).

Little is known about healthcare providers' perceptions of PM tools in Africa (175,217). Most research has explored medical professionals' views in higher and middle-income countries with robust healthcare settings (215,218–226). This study aimed to address this gap by exploring medical doctors' perceptions of a PM-based tool for CVD risk prediction in South Africa. We report our assessment of the findings and discuss clinicians' needs and the broader barriers that need to be addressed to introduce PM-based approaches in the South African public health setting.

2. MATERIALS AND METHODS

2.1. Study design and participants

Study participants were drawn from clinicians primarily affiliated with four public academic hospitals in Johannesburg, South Africa (Charlotte Maxeke Johannesburg Academic Hospital, Chris Hani Baragwanath Academic Hospital, Helen Joseph Academic Hospital, and Rahima Moosa Mother and Child Hospital). Four hospitals were selected to ensure an adequately sized and representative sample. Clinicians at all levels (intern, medical officer, registrar, and specialist) across all medical specialisations were eligible for inclusion. Clinicians were excluded if they had not completed their medical training and/or were not affiliated with the selected hospitals.

The study was approved by the University of the Witwatersrand Human Research Ethics Committee (Medical), Johannesburg, South Africa (Approval #M210355).

2.2. Study questionnaire

To assess clinician experiences and attitudes towards PM-based tools for CVD risk stratification, a 107-item electronic survey was developed. The survey was based on similar validated surveys (224–226), with adaptations to the local context, and was developed in consultation with a multidisciplinary group of academics and clinicians. Four specialists, two clinicians, and two geneticists assessed the survey face validity and ensured question clarity. The survey instrument was tested amongst five clinicians for content, design, and readability, and revised accordingly based on feedback.

The survey was organised into eight sections (**Table 4** and Publication supplementary materials S11) and administered in identical order (i.e., no section/item randomisation).

Items consisted of close-ended questions formatted using a five-point Likert-type rating scale. Items were scored in ascending order for all participants, with higher scores on rating scales implying increased self-perceived knowledge, perception, and confidence of PM-based stratification tools. A higher rating indicated a greater degree of benefit or concern associated with the specific item in the case of benefits and concerns, and an increased level of impact indicated that the barrier would be a greater hindrance to tool implementation. 'I prefer not to answer' was available for all questions, and free text options were available for respondents wishing to provide additional clarification.

Table 4: Survey organisation and associated scales used.

Section	Theme	No. of items	Scale
1	Demographic and professional information	10	Not applicable
2	Exposure: CVD risk stratification, genetics training and testing	21	Not applicable
3	Knowledge: Genetics and PM, and Educational video	15	Understanding: None, Little, Some, Moderate, Expert; Agreement: True, False, I do not know
4	Perception toward PM-based CVD risk stratification	5	Agreement: Strongly disagree, Disagree, Neutral, Agree, Strongly agree
5	Benefits and concerns of PM-based CVD risk stratification	19	Benefit/concern: Not, A little, Somewhat, Moderately, Very
6	Confidence in applying PM-based CVD risk stratification	8	Confidence: None, A little, Somewhat, Moderately, Very
7	Expectations in applying PM-based CVD risk stratification	9	Role: None, Supporting, Primary, Not sure Adoption likelihood: Very unlikely, Unlikely, No difference, Likely, Very likely
8	Barriers to implementing PM-based CVD risk stratification	15	Impact: None, Little, Some, Moderate, Strong

Clinicians working in South Africa have diverse educational backgrounds, with education and training from multiple institutions within South Africa and abroad. To ensure that all respondents had a similar baseline understanding of PM and related concepts, a compulsory 15-min educational video was included in the questionnaire (available on request). To determine whether the video had successfully relayed the

pertinent topics, Section 3 included nine true or false questions. A score of 78% (7 out of 9) and above was interpreted as successfully understanding the topics.

2.3. Data collection and analyses

Invitation to participate was extended via email to clinicians over a six-month period, from 10 September 2021 to 31 March 2022. Consenting respondents completed the survey, which was hosted online (REDCap). Due to low initial uptake, researchers requested permission to attend departmental clinical academic meetings, where the study aim and survey could be better outlined prior to inviting participation. Of the 15 departments approached, nine accepted (Internal Medicine, Cardiology, Clinical Genetics, Clinical Pathology, Emergency Medicine, Paediatrics and Child Health, and Surgery). Study data were collected and managed using REDCap (Research Electronic Data Capture) (version 12.5.4, Vanderbilt University, Nashville, TN, USA) tools hosted at the University of the Witwatersrand (227,228).

Self-perceived knowledge, perception and confidence (Sections 3, 4 and 6; **Table *2***) were calculated as continuous scores by converting responses to a numerical scale and summing the participants' responses, e.g., Knowledge: level of understanding scale converted to numeric scale (No understanding = 0, Little understanding = 1, Some understanding = 2, moderate understanding = 3, Expert understanding = 4) and the total score calculated by summing scores for each question (maximum score = 24). Mean knowledge score (MKS) was derived by dividing the participants' total score by 6 (number of questions). The mean perception (MPS) and confidence (MCS) scores were calculated similarly, with a maximum obtainable score of 4. However, MPS was calculated over five items. Regarding the nine items for assessing comprehension following the educational video, one point was given if the correct answer was chosen, and a score of zero was given if the wrong answer or 'I do not know' was chosen.

For descriptive comparisons, participants were dichotomised into high/low categories for the MKS and MCS scores using a breakpoint mean score of 2 (with low being ≤ 2, high > 2). Similarly, respondents were dichotomised into negative/positive categories for MPS scores at a breakpoint mean score of 2. As categorising a continuous variable causes a loss of appreciable information, inferential analysis was restricted to mean

scores. Similarly, age and number of years of clinical experience were dichotomised by the median.

Differences in MKS, MPS, and MCS distributions between groups were assessed with parametric (*t*-test and ANOVA) or non-parametric tests (Mann–Whitney or Kruskal–Wallis), where appropriate, followed by multivariate regression analysis (see **Table *5*** for study variables). Practice level and medical specialty were removed from the model due to collinearity with other predictors (Publication supplementary materials S12). All analyses were done using RStudio [38] (version 1.1.456), the psych (version 1.8.4; (229)) and ggplot2 (version 3.0.0; (230)) packages. Study responses were presented as percentages (95% confidence intervals; CI) and means (standard deviation; SD). Statistical significance was accepted at a *p*-value < 0.05.

Table 5: Categorical and continuous predictor variables used for analysis in this study.

Variable	Description	Type (range)	Categories
Outcome variables			
Mean knowledge score (MKS)	Average self-reported knowledge relating to genetics and PM	Continuous (0-4)	Not applicable
Mean perception score (MPS)	Average self-reported perception relating to value of PM-based CVD risk stratification tool	Continuous (0-4)	Not applicable
Mean confidence score (MCS)	Average self-reported confidence relating to implementation of PM-based CVD risk stratification tool	Continuous (0-4)	Not applicable
Predictor variable			
Sex	Reported sex	Categorical	Female Male
Medical group	Practitioner type based on self-reported practice level and years of experience	Categorical	Consultant Trainee
Clinical experience	Number of years conducting clinical duties	Categorical	< 8 years ≥ 8 years
Postgraduate qualifications	Have a postgraduate qualification in addition to a medical degree	Categorical	Yes No
Involvement in research	Are involved in medical research activities	Categorical	Yes No
Genetics or PM training	Have training specifically relating to genetics and/or PM	Categorical	Yes No
CVD stratification in clinical practice	Conducts CVD screening in their clinical practice	Categorical	Yes Sometimes No

**CVD – Cardiovascular disease, PM – Precision Medicine.*

3. RESULTS

3.1. Participant characteristics and exposure to genetics and cardiovascular disease screening

A total of 114 participants completed the survey. Five respondents were excluded as the respondents were non-patient-facing, resulting in a final sample of 109 participants. An accurate response rate could not be calculated, as the exact number of clinicians who received the questionnaire is unknown. Power calculations using Raosoft® showed a sample size of at least one hundred was required for an 8% margin of error and 95% confidence interval, assuming a response distribution of 50%. **Table *6*** presents respondent characteristics and exposure to study-related clinical variables. One-quarter of respondents were interns, designated 'trainees' in subsequent analyses, with a maximum of two years clinical experience and no medical specialisation. Over three-quarters were consultant clinicians. 25% of these practiced within internal medicine disciplines, such as Cardiology and Endocrinology, and 70% within non-internal medicine disciplines, including Paediatrics and Clinical Genetics. The mean (SD) age of the study population was 37.1 years (12.5), with a mean of 11.8 years (12.1) of clinical experience. Age and clinical experience were highly correlated (r = 0.97, SE = 0.03). Consequently, inferential analysis was restricted to clinical experience as to avoid multicollinearity.

Table 6: Respondents' demographics, professional characteristics, and study-related clinical exposures (n = 109).

Variable	Detail	N (%)
Hospital affiliation	Chris Hani Baragwanath Academic Hospital	23 (21.1)
	Charlotte Maxeke Johannesburg Academic Hospital	36 (33.0)
	Helen Joseph Hospital	13 (11.9)
	Rahima Moosa Mother and Child Hospital	32 (29.4)
	I prefer not to answer	5 (4.6)
Medical specialty	Internal medicine	21 (19.3)
	Non-internal medicine	57 (52.3)
	Not yet specialised (trainee)	27 (24.8)
	I prefer not to answer	4 (3.7)
Sex	Female	67 (61.5)
	Male	42 (38.5)

Age	< 33 years	57 (51.4)
	≥ 33 years	51 (46.8)
	I prefer not to answer	1 (0.9)
Medical group	Consultant	82 (75.2)
	Trainee	27 (24.8)
Clinical experience	< 8 years	55 (50.5)
	≥ 8 years	50 (45.9)
Postgraduate qualifications	Yes	66 (60.6)
	No	41 (37.6)
	I prefer not to answer	2 (1.8)
Involvement in medical research	Yes	59 (54.1)
	No	50 (45.9)
CVD stratification in clinical practice	Yes	41 (37.6)
	Sometimes	31 (28.4)
	No	36 (33.0)
	I prefer not to answer	1 (0.9)
CVD stratification using a risk score (n = 72)	Yes	25 (34.7)
	Sometimes	19 (26.4)
	No	28 (38.9)
Genetics training	Yes	17 (15.6)
	No	91 (83.5)
	I prefer not to answer	1 (0.9)
Genetics testing in clinical practice	Yes	31 (28.4)
	No	78 (71.6)
Conditions genetics testing discussed* (n = 31)	Monogenic disorders	30 (96.8)
	Cancers	16 (51.6)
	CVD	6 (19.4)
	Nutrigenomics	1 (3.2)
	Pharmacogenomics	2 (6.5)
Genetic testing frequency (n = 31)	1 patient every 2 to 3 months	13 (41.9)
	1 patient per month	2 (6.5)
	2 to 5 patients per month	6 (19.4)
	6 to 10 patients per month	2 (6.5)
	11 to 20 patients per month	3 (9.7)
	20+ patients per month	3 (9.7)
	I prefer not to answer	2 (6.5)

** Each respondent could select multiple options.*

Two-thirds of respondents undertake CVD risk stratification in their clinical practice. Sixty per cent of those use a stratification tool to determine risk, with the Framingham Coronary Heart Disease Risk Score being the most used (80%). 58% undertake CVD screening in response to risk factors, including family history and abnormal laboratory results. Clinicians are confident in the tool, with the majority sharing results with

patients (86%) and using them to guide treatment (96%). Most of those who do not undertake screening (68%) indicate that screening is not standard practice in their patient population, e.g., paediatrics and clinical genetics.

Overall exposure to genetics and PM was low. Fifteen per cent of respondents indicated they had formal genetics training, and approximately one-quarter had performed genetic testing in their clinical practice, although testing is infrequent (**Table 6**).

3.2. Genetics and precision medicine knowledge

Participants rated their level of understanding of six genetics and PM-related topics. The mean knowledge scores per individual across all six items were low (mean (SD) = 1.57 (0.64); max score = 4) (**Figure 14**) resulting in 80% (n = 108; 95% CI 70.5–86.5) of respondents categorised as having 'low knowledge'. A single individual indicated they had 'extensive knowledge' across all six themes, and no respondents answered all six items as 'expert.'

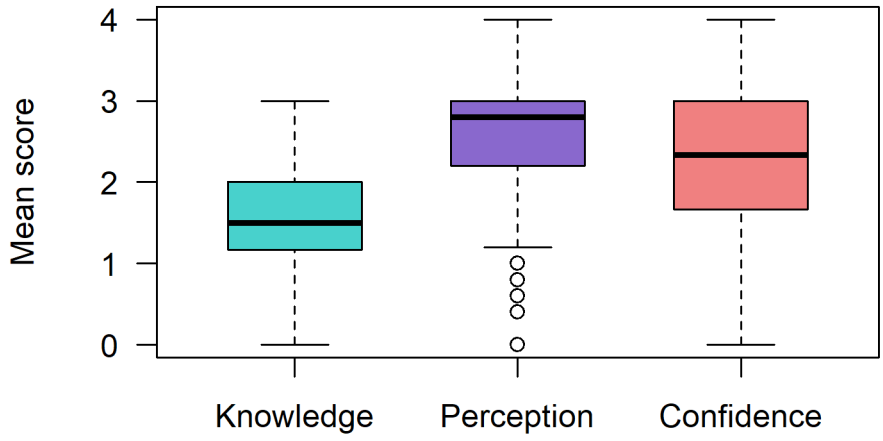


Figure 14: Respondents' overall mean knowledge, perception, and confidence scores.

Mean scores calculated per individual across items relating to knowledge, perception, and confidence.

Variability in understanding shifted towards little or no knowledge for newer genetic concepts, with approximately two-thirds of individuals indicating little or no understanding of Genome-Wide Association Study (GWAS) and PRS ***(Figure 15A)***. Regarding the assessment of knowledge following the introductory video, 97% of

respondents obtained a score of 78% or higher, suggesting the educational lecture successfully relayed the study's pertinent points and assisted in providing context to survey questions.

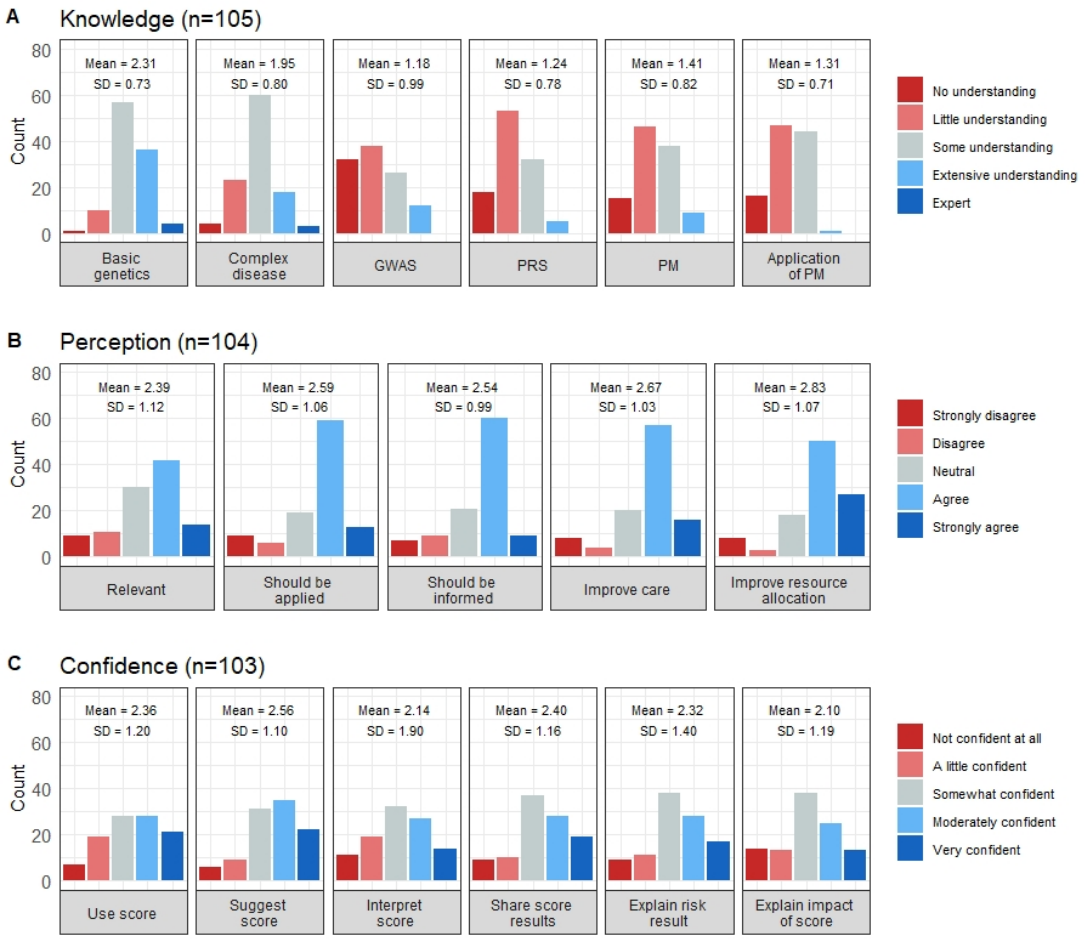


Figure 15: Respondents' knowledge, perception and confidence of items relating to PM-based CVD risk stratification.

The distribution of the clinicians' responses regarding (A) Genetics and PM knowledge, (B) Perceptions towards PM-based CVD risk stratification and (C) Confidence in applying PM-based CVD risk stratification in clinical practice. *GWAS—Genome-wide association studies, PRS—Polygenic risk scores, PM—Precision Medicine.*

3.3. Perceptions toward precision medicine-based CVD risk stratification

Overall, clinicians had a positive perception of a PM-based tool for CVD risk stratification (mean (SD) = 2.60 (0.90); max score = 4). Three-quarters of respondents had a positive mean perception score when categorised as positive or negative (*n* =

104; 76%, 95% CI 67.0–83.7) (**Figure 14**). Despite just over half of respondents believing such a tool would be relevant in their practice, at least two-thirds believed it should be applied in clinical practice (69%) and that it would improve care and prevention of CVD (70%) and national resource allocation (73%) (**Figure 15*B***).

3.4. Confidence in applying a precision medicine-based CVD risk stratification tool in their practice settings

Confidence (self-perceived) scores showed considerable variation (mean (SD) = 2.31 (1.04); max score = 4). Approximately half of respondents (55%) considered themselves confident across all confidence items ***(Figure 14)***. Otherwise, confidence depended on the action required within the clinical pathway. Clinicians were most confident in suggesting and using the risk score, whereas 29% had little to no confidence in interpreting the risk score and explaining any potential adverse insurance policy impacts that may arise from risk information (27% little to no confidence) (***Figure 15C***). Forty per cent of respondents indicated they would be moderately or very comfortable with adapting prevention and treatment strategies based on score results. However, a greater proportion of respondents were comfortable with adapting their treatment strategies (22%) than their prevention approaches (13%).

3.5. Factors influencing knowledge, perceptions, and confidence

We assessed how respondents' characteristics influenced their knowledge, perceptions, and self-confidence toward PM-based CVD risk stratification (***Table 7***). Medical specialty, sex, and clinical experience did not significantly influence knowledge levels. Medical research involvement had a small effect on knowledge, with those involved in research having increased self-perceptions of knowledge [1.71 (0.61) vs. 1.38 (0.61); $p < 0.01$; effect size r = 0.26]. Previous genetics training had a large effect on knowledge, with those having had training having significantly higher mean knowledge scores [2.26 (0.62) vs. 1.42 (0.54); $p < 5 \times 10{-}5$; effect size R = 0.76]. Similarly, medical research involvement influenced overall self-confidence in applying PM-based risk stratification [2.57 (0.96) vs. 2.01 (1.06); $p < 0.05$; effect size R = 0.26]. Self-confidence scores were also influenced, with a moderate effect, by whether

respondents currently undertook CVD risk stratification in their clinical practice or not [2.65 (1.10) vs 1.93 (0.96); $p < 0.05$; effect size $\eta^2 = 0.07$]. Conducting CVD screening on a regular or infrequent basis resulted in a significantly higher mean perception score, and with a large effect [Yes: 2.89 (0.63), Sometimes: 2.65 (0.88) vs No: 2.11 (1.03); $p < 0.0005$; effect size $\eta^2 = 0.13$].

Table 7: Predictors' influence on respondents' mean (SD) knowledge, confidence, and perception scores. P-value is assessing differences in mean scores across variable subgroups.

Variable	Mean (SD) knowledge score (n = 107)	*p*-value	Mean (SD) confidence score (n = 104)	*p*-value	Mean (SD) perception score (n = 106)	*p*-value
Medical specialty						
Internal medicine	1.67 (0.67)		2.70 (0.94)		2.95 (0.64)	
Non-int. medicine	1.47 (0.62)	*0.310*	2.12 (1.09)	*0.098*	2.40 (0.98)	*0.052*
Not yet specialised	1.66 (0.62)		2.30 (0.92)		2.58 (0.85)	
Sex						
Female	1.59 (0.67)	*0.682*	2.41 (0.98)	*0.404*	2.45 (1.11)	*0.632*
Male	1.54 (0.61)		2.26 (1.08)		2.64 (0.74)	
Medical group						
Clinician	1.52 (0.64)	*0.318*	2.26 (1.08)	*0.933*	2.54 (0.93)	0.969
Trainee	1.66 (0.62)		2.30 (0.92)		2.58 (0.85)	
Practice level						
Specialist	1.61 (0.66)		2.25 (1.27)		2.48 (0.90)	
Registrar	1.41 (0.65)	*0.470*	2.32 (0.81)	*0.897*	2.71 (1.05)	*0.526*
Medical officer	1.48 (0.53)		2.56 (0.86)		2.60 (0.82)	
Intern	1.66 (0.62)		2.30 (0.92)		2.58 (0.85)	
Clinical experience						
< 8 years	1.59 (0.60)	*0.580*	2.32 (0.84)	*0.651*	2.59 (0.86)	*0.945*
≥ 8 years	1.53 (0.67)		2.31 (1.23)		2.55 (0.95)	
Postgrad. qualifications						
Yes	1.62 (0.68)	*0.195*	2.32 (1.17)	*0.490*	2.62 (0.890)	*0.317*
No	1.45 (0.53)		2.27 (0.78)		2.48 (0.93)	
Medical research involvement						
Yes	1.71 (0.61)	*0.007***	2.57 (0.96)	*0.007* **	2.69 (0.87)	*0.249*
No	1.38 (0.61)		2.01 (1.06)		2.43 (0.92)	
Genetics training						
Yes	2.26 (0.62)	*3.518 x*	2.46 (0.86)	*0.720*	2.96 (0.61)	*0.146*
No	1.42 (0.54)	*10^{-5}* ***	2.29 (1.17)		2.50 (0.93)	
CVD screening						

Yes	1.75 (0.59)		2.65 (1.10)		2.89 (0.63)	
No	1.46 (0.63)	*0.074*	1.93 (0.96)	*0.010 **	2.11 (1.03)	*4.162 x 10^{-4} ****
Sometimes	1.45 (0.65)		2.33 (0.94)		2.65 (0.88)	

**Non-int – Non-internal Postgrad.—Postgraduate, CVD—cardiovascular diseases; p-value thresholds: *** = p ≤ 0.001, ** = p ≤ 0.01, * = p ≤ 0.05.*

3.6. Multivariable analysis of the factors affecting knowledge, perceptions, and confidence

When adjusting for study predictors, medical research involvement and genetics training remained positively associated with mean knowledge score, with the model explaining 30% of the overall variance. Similarly, for confidence and perception, medical research involvement (confidence only) and CVD screening status remained associated. However, the influence of CVD screening increased, and the model explained 14% and 13%, respectively. The model revealed the medical group to be associated with the mean knowledge and mean confidence score when adjusting for other predictors (Publication supplementary materials S13).

3.7. Benefits and concerns of a PM-based CVD risk stratification

Figure 16 presents the distribution of the clinicians' responses regarding perceived benefits (A) and concerns (B) of a PM-based CVD risk stratification in the South African public setting. Respondents believed the main benefits associated with PM-based CVD risk stratification was the potential availability of a CVD risk score tailored to African populations (very beneficial and moderately beneficial *n* = 91; 86%; 95% CI: 77.4–91.6), followed by the adaption of prevention and treatment strategies (very beneficial and moderately beneficial (*n* = 89; 84%: 95% CI: 75.3–90.1). The most concerning issue for clinicians was the potential adverse impact that score results may have on a patient's life and insurance policies (very concerned and moderately concerned *n* = 77; 73%; 95% CI: 63.0–80.6), and the lack of transferability of current PM risk that scores have across populations (very concerned and moderately concerned *n* = 56; 53%; 95% CI: 42.9–62.5).

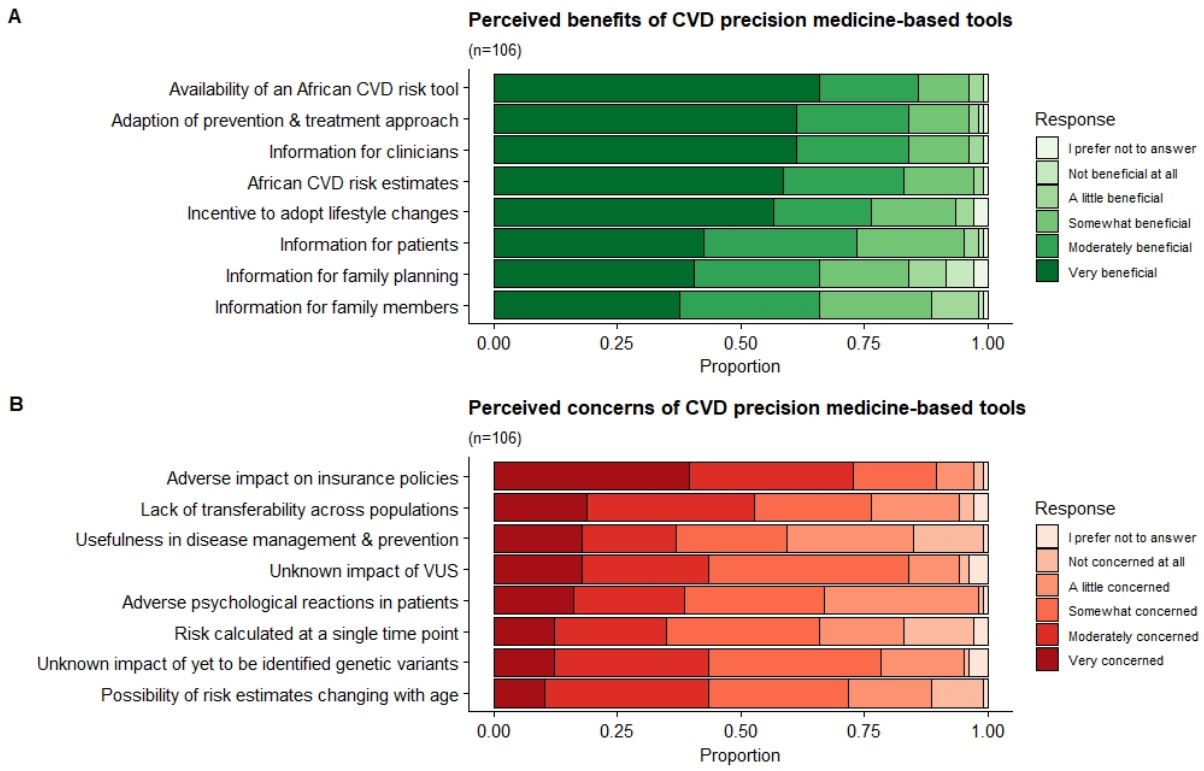


Figure 16: Perceived benefits and concerns of a PM-based CVD risk stratification.

The distribution of the clinicians' responses relating to the perceived benefits (A) and concerns (B) of a PM-based CVD risk stratification in the South African public setting.

3.8. Clinician's expectations: time horizon, funding, and expected role

Assessing clinicians' expectations of PM-based tools for CVD risk stratification in South Africa's public revealed that clinicians anticipate a clinically relevant tool for African populations to be available in the longer term. Over one third (36%) expect the tool within the next six to ten years, and another third in more than ten years (39%). Half of the respondents feel that funding would be the responsibility of multiple players, but at least a third (35%) believe that national and provincial governments should be the primary funder. Despite the expected wait and funding complexities, 85% of respondents indicate that they are likely or very likely to alter their CVD screening approach if a tool becomes available (Publication supplementary materials S14).

When asking clinicians what they believe their expected role will be in the implementation of PM-based CVD risk stratification, approximately two-thirds identify

themselves as having the primary role in traditional patient-facing activities (referring the patient for testing (73%), sharing results with the patient (62%), and using results to guide prevention and treatment approaches (70%). In contrast, at least half of the respondents recognise that interpreting (55%) and explaining results to patients (54%) will require additional expertise and thus a multidisciplinary team (Publication supplementary materials S14).

3.9. Perceived barriers to the implementation of pm-based CVD risk stratification

According to respondents, the top implementation barriers facing PM-based CVD risk stratification in South Africa's public health setting are the cost (strong impact = 89%), and the lack of genetics services (strong impact = 85%) and associated shortage in genetics personnel (strong impact = 83%). Other barriers included the lack of a funder (strong impact = 81%), the absence of clinical guidelines (strong impact = 66%) and limited training in PM and related concepts (62%) (**Figure *17***).

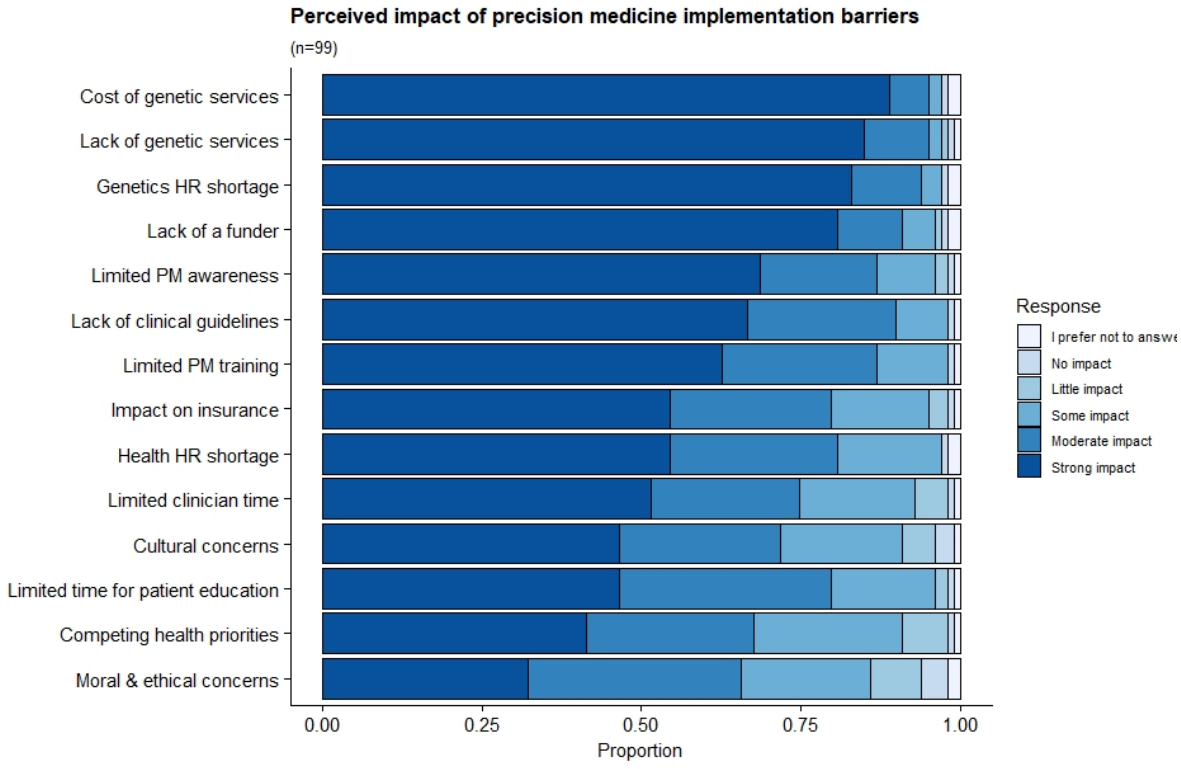


Figure 17: Perceived impact of PM-based CVD risk stratification implementation barriers in the South African public setting.

The distribution of responses regarding the barriers facing the implementation of PM-based CVD risk stratification in the South African public setting.

4. DISCUSSION

As little is known about healthcare providers' perceptions of PM tools in Africa, this study aimed to explore medical doctors' perceptions of a PM-based tool for CVD risk prediction in South Africa. Results revealed clinical exposure to, and knowledge of genetics is limited in the South African public health setting, but most perceive PM approaches positively. Confidence in applying PM-based CVD risk stratification is variable, with those involved in research currently undertaking CVD risk stratification being more confident.

4.1. Positive perceptions of PM-based tools despite knowledge gaps and resource constraints

Our study identified low rates of genomic testing, coupled with low self-perceived knowledge of genetics and PM-related concepts among clinicians and trainees practising in South Africa's public health setting. Despite limited understanding and resource challenges, over two-thirds of respondents expressed positive perceptions of PM-based stratification tools. Such findings are not dissimilar to existing literature, which reveals low self-perceived knowledge but positive perceptions amongst healthcare providers across medical specialisations, geographies and economic regions (175,215,217,221,223–226,231–235). Assessment across five Implementing GeNomics In pracTicE (IGNITE) sites in the USA involving differing precision medicine applications revealed most clinicians believed genetic testing to be clinically useful. However, only a third thought they were adequately trained to provide care for genetically "high-risk" patients (215).

Although the value of integrated CVD risk scores was recognised, the perceived relevance of PM-based risk stratification to individual practice was variable and linked to varying practice standards across medical specialties, highlighting the need for clinician inclusion in the development of appropriate PM tools. The availability of risk estimates tailored to African populations was considered to be of particular value to

providers, but the current limited transferability of PM-based approaches across ancestries (60) was also a key concern.

4.2. Addressing the genetics education gap is paramount to successful adoption

Given that exposure to genetic testing and PM approaches is likely to increase with decreasing costs and increasing efforts to understand African genetic variation, these findings support prior calls for increased training and educational resources (215,221,234,236,237). Furthermore, the limited genetic literacy amongst South African patients further necessitates additional support from healthcare providers to benefit from genetics and PM.

Lower self-perceived knowledge amongst consultants compared to trainees suggests that exposure to PM and related concepts in undergraduate training may have shifted in recent years. Alternatively, inexperienced clinicians may not be aware of the complexities of testing in a clinical setting as yet. Although different, mean knowledge scores for both groups were low, and highlight training requirements at undergraduate and practice levels. Genetics curricula can go unnoticed by clinicians as it is often integrated into the content of a course (238). Simply increasing the volume and duration of genetic curricula is unlikely to be helpful, as advances in genomics do not occur in isolation and compete with other compelling learning demands (236). Specialised courses and qualifications, although increasingly available, require substantial commitments and have relatively low uptake amongst non-genetic specialists (238). Genomics education incorporated into healthcare providers' usual work activities, including departmental presentations and clinical meetings, has been identified as a preferred training strategy (239,240). The most appropriate training approach is debated. However, there is consensus that education should address common misconceptions related to genetics, emphasise the clinical applications of genomics while not negating the complexity of the associated basic science, and ensure that clinicians can correctly utilise genetic-based risk estimations in their clinical practice (236,238,241).

Practicing clinicians have indicated a preference for online continued medical education, conferences, peer-reviewed literature, and in-clinic training (219,240,242).

Historically, most continued medical education has been passive and didactic, with limited impact on changing practice. The most impactful approaches are those that are interactive, use a variety of instructional techniques, and focus on outcomes considered important by healthcare professionals. The appropriateness of such activities in South Africa will need to be further investigated, prioritising culturally sensitive approaches that consider language differences and delivery format challenges.

4.3. Integrated models compatible with current clinical practices are needed

In this study, confidence in utilising the approach was variable and was directly influenced by previous CVD risk stratification exposure. Literature suggests that clinicians lack confidence in using and implementing new genomic tests and precision medicine approaches, and that this low confidence is linked to limited knowledge of genetics and related topics (219,221,224,235,239,243). However, increased confidence has been associated with increased exposure to testing, or when the application is clearly defined and given within the context of a specific disease (215,244). For example, oncologists already using genetic testing in clinical practice reported the highest confidence in using multi-marker tumour panel results to guide patient care (244). Increased confidence amongst those already using CVD stratification algorithms suggests that successfully integrating genetic risk assessment into existing clinical frameworks requires strategies compatible with current practice.

The areas clinicians lacked the most confidence in were related to the interpretation of score results and explaining potential impacts on insurance policies, which may reflect the lack of understanding of the genetic components of the score and the legislative framework. Respondents indicated a strong preference for a multidisciplinary model approach to testing, particularly interpreting score results and explaining such to patients. Multidisciplinary teams require specialised genetics skillsets, such as clinical geneticists and genetic counsellors, which may not be feasible in the context of a skills shortage in South Africa. Increasing clinicians' comfort with using genomics in routine care may reduce the reliance on genetic specialists and improve the feasibility of genomic medicine over the long term. Additionally, the development of guidelines for

PM-based scoring may help improve understanding in these areas and would support testing practices.

4.4. Funding shortfalls and skills shortages hinder PM adoption in resource-scarce settings

Similar to previous research and other settings, surveyed clinicians identified the affordability and accessibility of genetic services and genetics skills shortage as major impediments to PM implementation in South Africa's public health settings (91,152,217,221,234,235). These challenges are not unique to PM and successful innovative models in funding and delivering care, especially within limited resource settings, should be explored.

Additional barriers raised from studies in well-resourced settings include those related to organisational structures (management, organisational culture and/or decentralised care), lack of clinical guidelines and the poor coordination of tests relative to treatment decisions (224). Nephrologists in Australia highlighted barriers relating to their specific practice setting, e.g., leadership endorsement and goal setting (235), and medical oncologists in Canada identified patient genomic literacy and the current clinical utility of genomics as key challenges (226). Although not identified as primary barriers in this South African study, these challenges are likely to be critical in the South African public setting as system-level barriers are addressed. Stakeholders will need to consider these challenges when developing a PM implementation roadmap, and leverage lessons learned as resource challenges are overcome and healthcare systems stabilise.

This study was novel in that it included a compulsory educational video before responding to the questionnaire. The purpose was to provide context and some exposure to PM across the responder group. However, several limitations to the study design merit discussion. Over and above the limited sample size and potential responder biases associated with online convenience sampling strategies (245), the study makes use of a newly developed instrument which has not yet been extensively tested for reliability and validity. There may have been issues with biased response sets (e.g., "faking good") that would be difficult to detect in the absence of reverse-

scored items (246). Reassuringly, however, our results are intuitive and align with literature trends, diminishing concerns around the psychometric rigor of the instrument.

Lastly, for convenience, the study survey was administered to clinical practitioners based in large, relatively well-funded urban medical centres. Consequently, respondents in this study are unlikely to be representative of health professionals across South Africa. Understanding primary care providers' perceptions of and willingness to utilise PM-based tools in the primary care setting should be explored. Primary care provision is the foundation of care in South Africa, serving most communities, especially in rural areas, and where CVD risk stratification could have the greatest benefit (247–249). Successful PM utilisation requires multi-stakeholder co-operation, and thus the views of patients, insurers, policy making bodies and governments, should also be explored in follow up studies.

This study sheds light on an under-researched area of South African healthcare and contributes to a greater understanding of what will be required to successfully implement PM in South Africa, and potentially other resource-constrained environments.

5. CONCLUSION

Practicing clinicians have a positive perception of PM-based CVD risk stratification despite limited genomic knowledge, and they have a desire to adopt such tools in their practices. Limiting adoption barriers requires developing tools that leverage existing clinical approaches. The advancement of PM in South Africa requires an active effort to address the gap in genetic training whilst simultaneously investing in the country's genomics capacity, including both skill force and infrastructure.

Chapter 4: Publication 4

Multimorbidity in African populations: a scoping review

This chapter describes a review that was performed in order to gain an understanding of the current prevalence of multimorbidity in African populations, including continental Africans and African diaspora. Given the multiple disease burdens in Africa, we believe it is crucial to look beyond a single disease in isolation and understand that multiple chronic conditions may be affecting a patient, and thus their treatment and management, as well as risk for other disease. Conducting this scoping review allowed for us to gain an understanding of current literature of multimorbidity research conducted in African ancestry populations and identify research gaps that could be addressed in future work. The work described in this chapter is under review in the peer-reviewed journal, BMJ Global Health (Appendix H):

Kamp M., Achilonu, O., Kisiangani, I., Nderitu, DM., Mpangase, PT., Tadesse, G.A., Adetunji, K., Iddi, S., Speakman, S., Hazelhurst, S., Asiki, G. and Ramsay, M. 2023. Multimorbidity in African ancestry populations: a systematic scoping review. *BMJ Global Health.* 8:e013509

Author contributions: MR, GA, SH and MK conceived the study and participated in the design of the study. MK, OA, IK, DMN, GAT, KA and SI undertook the literature review process. MK, OA and PTM undertook data analysis and visualisation. MK drafted the manuscript with input from OA. MR is acting as guarantor.

ABSTRACT

Objectives: Multimorbidity (MM) is a growing concern linked to poor outcomes and higher healthcare costs. While most MM research targets European ancestry populations, the prevalence and patterns in African ancestry groups remain underexplored. This study aimed to identify and summarise the available literature on MM in populations with African ancestry, on the continent, and in the diaspora.

Design:- A scoping review was conducted in five databases (PubMed, Web of Science, Scopus, Science Direct and JSTOR) in July 2022. Studies were selected based on predefined criteria, with data extraction focusing on methodology and findings. Descriptive statistics summarised the data, and a narrative synthesis highlighted key themes.

Results: Of the 232 publications on MM in African-ancestry groups from 2010 to June 2022—113 examined continental African populations, 100 the diaspora and 19 both. Findings revealed diverse MM patterns within and beyond continental Africa. Cardiovascular and metabolic diseases are predominant in both groups (80% continental and 70% diaspora). Infectious diseases featured more in continental studies (58% continental and 16% diaspora). Although many papers did not specifically address these features, as in previous studies, older age, being women and having a lower socioeconomic status were associated with a higher prevalence of MM, with important exceptions. Research gaps identified included limited data on African-ancestry individuals, inadequate representation, under-represented disease groups, non-standardised methodologies, the need for innovative data strategies, and insufficient translational research.

Conclusion: The growing global MM prevalence is mirrored in African-ancestry populations. Recognising the unique contexts of African-ancestry populations is essential when addressing the burden of MM. This review emphasises the need for additional research to guide and enhance healthcare approaches for African-ancestry populations, regardless of their geographic location.

1. INTRODUCTION

Multimorbidity (MM), the co-occurrence of two or more chronic conditions in an individual, is a growing global challenge (67,68). MM is associated with increased premature mortality (68), lowered quality of life (69), diminished mental health (250,251), increased polypharmacy risk (252), and intensified health services utilisation and associated costs, particularly in resource-poor settings (253).

There are, however, different interpretations of the MM definition and a lack of global consensus. Some use the above-described definition but highlight that there is no index condition or dominant disease in the affected individual (254,255). This is consistent with the WHO's definition emphasising the chronic and long-term nature of conditions and the absence of hierarchy or clustering of diseases (256). Comorbidity, on the other hand, refers to additional conditions in patients with a primary condition of interest. The distinction between MM and comorbidity is important but not always clear, despite its critical importance for accurate phenotyping, research analyses and in healthcare settings. Additionally, there is variability in the nature and number of chronic diseases included in MM studies and the measurement methods used to assess MM across studies, as highlighted by a recent Delphi study (257).

Improved life expectancy resulting from global health efforts has led to a rise in MM due to ageing populations and shifting lifestyle risk factors, including obesity and physical inactivity (67,258–260). MM encompasses various combinations of chronic conditions, with concordant MM sharing origins and treatment requirements, and discordant MM involving unrelated or differently managed conditions (261). As reviewed by Roomaney et al. (2022), in low- and middle-income countries (LMICs), MM is further complicated by overlapping infectious disease burdens, poverty-related environmental stressors, limited social infrastructure and under-resourced healthcare settings (262). It is estimated that at least one-third of adults residing in LMICs have multiple chronic conditions (263), yet the epidemiology of MM in these regions remains relatively unknown.

Despite the substantial growth in MM studies, most have been conducted in high-income countries or among those of European-ancestry, limiting the generalisability of findings (262,264). Although studies involving diverse ancestries, particularly

European and African American (AA) populations, indicate increased MM risk and adverse health outcomes among individuals of African-ancestry (265,266), the analyses are primarily focused on AAs and often based on small sample sizes. Moreover, most studies examine single disease outcomes or comorbidities with an index disease, leaving the true burden of MM, its epidemiological characteristics, and healthcare strategies in African-ancestry populations largely unknown (267).

This scoping review is timely as the global health community is advocating for diversity and inclusion in research studies, a viewpoint strongly echoed by researchers, healthcare professionals, policymakers, and pharma. It is well recognised that African and African-ancestry populations are under-represented and understudied (265) and this review and critical analysis of the published studies are, therefore, important to raise awareness. When global studies report continent-wide prevalence, the projections for African populations are often based on a few small studies on selected communities. They do not reflect the heterogeneity across Africa and in diaspora populations, providing false impressions that could lead to disastrous public health interventions in communities where the data and findings do not apply.

In addition to the limited representation of diverse populations, MM studies face several methodological challenges, including the lack of international consensus on the definition of MM and the frequent use of non-standardised phenotype and laboratory-based measurement methods. These inconsistencies hinder research, comparative assessments and the translation of findings into guidelines and interventions(268).

This scoping review seeks to identify, evaluate, and summarise existing evidence on MM prevalence, patterns, associated factors, and research gaps specific to African-ancestry populations. It provides insight into similar and divergent trends among continental and diaspora African-ancestry individuals, in order to inform the most appropriate and relevant future research priorities.

The following research questions were addressed to guide the scoping review: (1) How many studies have been done on this topic, and what are the geographical locations of the studied populations? (2) Which definitions of MM and what study designs are commonly used in MM studies that include African-ancestry populations?

(3) What are the common MM disease clusters among African-ancestry populations living on the continent and in the diaspora? (4) How does MM prevalence differ across sociodemographic characteristics and are these patterns consistent across continental and diasporic African populations?

2. METHODS

To assess MM within African-ancestry populations, a scoping review methodology was employed rather than a systematic review or meta-analytic approach as a scoping review is more appropriate when aiming to assess an emerging field of literature and identify research gaps (269). This review began with the establishment of a research team consisting of members with expertise in epidemiology, demography, clinical medicine, genetics, and data science research in Africa. Members were from several African countries (Kenya, Nigeria, South Africa, Ghana, and Uganda) and advised on the broad study aim and protocol, including search terms, databases, and thematic synthesis. A review protocol for this study is not available.

2.1. Search strategy and database search

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for scoping reviews (270). The initial search was implemented in July 2022 and included a comprehensive literature search by one researcher (MK) in consultation with a librarian and field experts. Five electronic databases (PubMed, Scopus, Science Direct, Web of Science and JSTOR) were searched to identify relevant publications on MM in individuals of African-ancestry. The search strategy incorporated a combination of MeSH (Medical Subject Headings) terms and keywords, including ‘multimorbidity,’ ‘comorbidity’ and ‘African population’ and their synonyms. The final search terms are listed in **Table *8*** and detailed in Publication supplementary materials S15. Search terms relating to MM were derived from previous systematic reviews, and search themes 1 and 2 were combined using the Boolean operator ‘AND’. The search was limited to the title and abstract fields only, as a trial search found searching full text reduced the ability to detect relevant papers. Similarly, for this reason, a list of African countries was not included in the search strategy. Although the search terms were consistent across all databases, the search query was tailored to the specific requirements of each database.

The publication time frame was limited to 1 January 2010–31 June 2022, as research on MM has increased significantly since 2010, with approximately 80% of publications published after this date (264).

Table 8: Search term strategy

	Search themes	MeSH terms
1	**Multimorbidity**	"comorbidity" OR "co-morbidity" OR "multimorbidity" OR "multi-morbidity" OR "multiple chronic conditions" OR "chronic conditions" OR "chronic disease"
2	**African populations**	"Africa" OR "African" OR "African population" OR "African ancestry" OR "African-ancestry"

2.2. Citation management and study selection

After conducting database searches, the search output citations were downloaded in BibTeX format and uploaded to an electronic screening and data management website, Rayyan (271). The deduplication function was employed to remove duplicates. For the first level of screening, only the title and abstracts of unique entries were reviewed to avoid wasting resources on articles that did not meet the minimum inclusion criteria. Entries were screened independently and blindly by MK and three other reviewers (IK, DMN and GAT), and studies deemed irrelevant were discarded. A third group (GA and MR) assisted with conflict resolution.

2.3. Eligibility criteria

The inclusion criteria were studies reporting epidemiological data on MM or comorbidity in individuals of African-ancestry aged 18 years and above from continental Africa and studies including populations of African-ancestry in other global regions (e.g., Caribbean, Europe, North America). Studies written in French were included to mitigate missing data from Francophone Africa.

The inclusion criteria were studies reporting epidemiological data on MM or comorbidity in individuals of African-ancestry aged 18 years and above from continental Africa and studies including populations of African-ancestry in other global regions (e.g., Caribbean, Europe, North America). Studies written in French were included to mitigate missing data from Francophone Africa.

The definition of MM used in this review was based on the Academy of Medical Sciences definition—the coexistence of two or more chronic conditions, which included conditions in the following categories: physical non-communicable disease (NCD) of long duration, such as cardiovascular disease or cancer, a mental health condition of long duration, such as a mood disorder or dementia, or an infectious disease of long duration, such as HIV or hepatitis (267). The term 'comorbidity' was included, as it has been used interchangeably with MM in the past, although it is now recognised as a distinct concept (268). As per the Academy of Medical Sciences definition, acute conditions (<6 months) were not included in our definition. We use the term diaspora to specifically refer to populations of African-ancestry who no longer reside on the African continent, contrasting them to continental African populations residing in Africa. **Table 9** provides a detailed summary of our inclusion and exclusion criteria.

Table 9: Study inclusion/exclusion criteria for the scoping review

	Inclusion	Exclusion
Methodology	- MM defined as a combination of recognised diseases/conditions (e.g., self-report or International Classification of Disease 9th revision (ICD-9) or 10th revision (ICD-10) codes)	- MM defined as a combination of symptoms or pre-disease conditions, not defined as ICD-9 or ICD-10 diseases (e.g., predisease, frailty, disability, and quality of life) - Transitions or trajectories within a single disease or from one condition into another (e.g., cancer progression)
Study design	- Cross-sectional studies - Longitudinal quantitative studies, including retrospective and prospective cohort studies - Qualitative studies	- Systematic reviews - Meta-analyses - Case studies - Expert opinion/committee reports
Population	- Adult humans (18+ years) - African continent - African-ancestry diaspora	- Infants, children, or adolescents (<18 years) - Animal research
Publication date	- 01 Jan 2010 - 31 June 2022	- Prior to 01 Jan 2010 - Post 31 June 2022
Publication type	- Peer-reviewed journal articles - Clinical trials	- Case reports - Reviews - Editorials - Letters
Language	- English - French	- Other languages

Studies were excluded when both initial reviewers indicated that they did not meet the eligibility criteria. When their responses were discordant, a third reviewer assessed the abstract, and in some cases, reviewed the full publication before deciding whether the publication met inclusion criteria.

2.4. Data extraction and analyses

All citations deemed relevant after title and abstract screening underwent full-text review. Four reviewers (IK, MK, DMN and GAT) independently assessed the full texts (more than one person evaluated some citations), and study characteristics were captured electronically using a template to extract relevant data (Appendix F *Supplementary materials* **Table S1**). Data on MM definition, the approach to measuring MM, the conditions included, and the characteristics of the population under investigation were extracted. Studies were categorised by geographical location, participant ancestry and study design. The chronic conditions assessed to define MM were collected and organised into categories based on the body system affected (257). Five reviewers independently extracted data and updated the extraction spreadsheet (OA, IK, MK, DM, and GAT). For the extracted variables (Publication supplementary materials S16), the mean and standard deviation (SD), the absolute number, or the percentage was recorded as appropriate.

To effectively analyse and interpret the extensive range of diseases evaluated, a categorisation approach was employed. Diseases were grouped into nine broader clusters (**Table 10)**. These clusters were derived using a body system approach with nuances to reflect the major global mortality and morbidity disease burdens over the last decade (272), and through consultation with clinicians.

Table 10: Disease clusters and the associated definition and diseases

Cluster	Cluster detail	Example conditions
Cancer	Diseases in which abnormal cells divide uncontrollably and can invade nearby tissues.	Breast, lung, colon, cervical, and prostate cancers, Kaposi Sarcoma, etc.
Cardiovascular	Chronic diseases that affect the heart and blood vessels.	Arrhythmias, atrial fibrillation, cardiomyopathy, coronary artery disease, heart failure, hypertension,

		peripheral artery disease, stroke, etc.
Infectious	Conditions caused by persistent infections; including bacterial, viral, or parasitic infections.	Chronic viral hepatitis (hepatitis D, E, and G), hepatitis B, hepatitis C, human immunodeficiency virus (HIV), human papillomavirus (HPV), long-COVID, tuberculosis, etc.
Metabolic	Chronic diseases that affect the body's metabolic system.	Dyslipidaemia, fatty liver disease, metabolic syndrome, obesity, type 2 diabetes mellitus, etc.
Musculoskeletal	Chronic conditions that affect the bones, joints, and muscles.	Chronic back pain, osteoarthritis, osteoporosis, rheumatoid arthritis, back syndrome with radiating pain, back syndrome with non-radiating pain, etc.
Neurological	Conditions that affect the brain and nervous system.	Alzheimer's Disease, dementia, epilepsy, multiple sclerosis, Parkinson's disease, autistic spectrum disorder, etc.
Psychiatric	Disorders that affect mood, thinking and behaviour.	Attention deficit hyperactivity disorder (ADHD), bipolar, depression, generalised anxiety disorder (GAD), substance use disorders, post-traumatic stress disorder (PTSD), schizophrenia, etc.
Renal	Chronic diseases that affect the kidneys.	Chronic kidney disease, Renal failure, End-stage renal disease, etc.
Respiratory	Chronic diseases that affect the lungs.	Asthma, Bronchiectasis, Chronic obstructive pulmonary disease (COPD), Cystic fibrosis, Sleep apnoea, Pneumonia, etc.

2.5. Data summary and synthesis

The data extraction spreadsheet was imported into Microsoft Excel (273) for validation and coding. The data were then exported into R statistical Computing Tool (version 4.3) (103) for analysis. Descriptive statistics were calculated to summarise the data, with frequencies and percentages utilised to describe nominal data. An UpSet plot (274) for the two groups of African-ancestry populations was constructed to visualise the disease clusters. A thematic synthesis was used to identify key themes emanating from the included articles.

2.6. Patient and public involvement

Patients and the public were not involved in this study. Clinicians known by the authors and working in MM across Africa were invited to comment.

2.7. Ethics statements

This is a review of already published material; therefore, no ethics approval needed.

3. RESULTS

Below we describe the results of our article search, the analyses of the included papers, and the themes identified through content analyses.

3.1.1. Search results

Figure 18 provides an overview of the search results. The initial search yielded 2397 articles, and after removing duplicates (366), 2031 unique records remained. Out of these, 1678 articles were excluded after screening. During the title and abstract screening phase, reviewers had disagreements on the exclusion/inclusion of 15% of the articles, which were resolved through discussion. A total of 353 abstracts were evaluated for eligibility, with 47 articles excluded due to misclassification (n=23) and the exclusion of papers focusing on MM in COVID-19 severity (n=24). 307 full-text articles were assessed, and an additional 74 exclusions were made due to misclassification (n=30) or the full text being unavailable (n=44). The BibTex citation file of analysed articles is available on GitHub (https://github.com/MADIVA-DSI/MM-in-African-ancestry-pops.git).

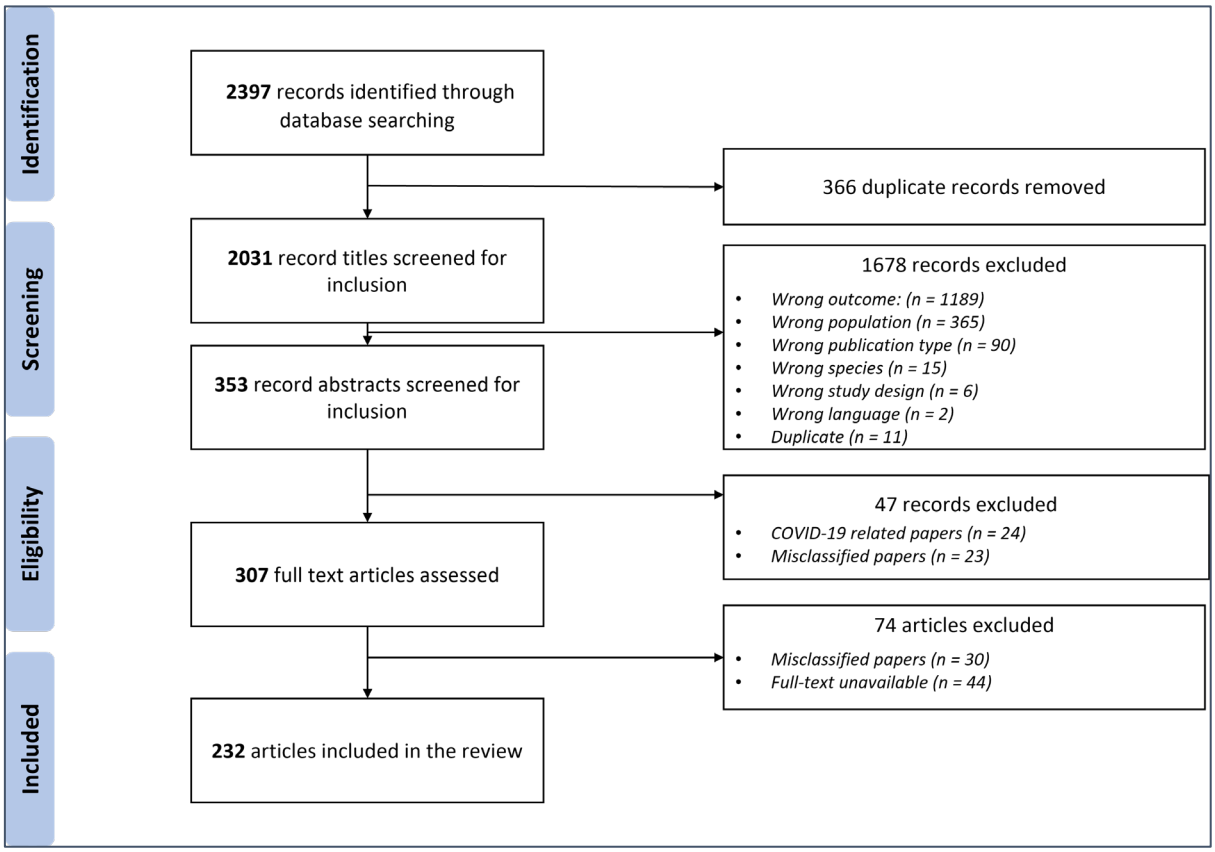


Figure 18: PRISMA-ScR flow diagram of the study selection process.

Misclassified papers refer to those that on further analysis did not meet the inclusion criteria, for example, inappropriate outcome (did not meet the criteria for MM), Incorrect publication type (not original research). *MM, multimorbidity; PRISMA-ScR, Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews*.

While multiple criteria led to paper exclusion, only one reason was assigned to each study. The most common exclusion reason was 'inappropriate outcome,' indicating the research question did not address MM (as defined by the Academy of Medical Sciences) in African-ancestry populations, the focus of this review. These studies often examined comorbidity or broader chronic disease prevalence, prevention, and treatment strategies.

3.1.2. Study characteristics

The number of publications increased over the past decade from 5 in 2010 to 11 in 2022 (**Figure 19**). While there was a decline in papers addressing MM in African-ancestry populations from 2020 to 2022, the upward trend persisted when including papers related to MM and COVID-19. During this period, we identified 23 publications

associating MM as a risk factor for severe COVID-19 in African populations (2020: 8 studies, 2021: 10 studies, 2022: 5 studies).

Geographical location of African-ancestry populations studied

Figure 19 shows the number of publications per year by African ancestry population, distinguishing between those from the diaspora and the continent. The data show there were slightly more publications among continental populations (49%, n=113) than diaspora (43%, n=100), with 8% conducted in multiple countries involving both populations (n=19).

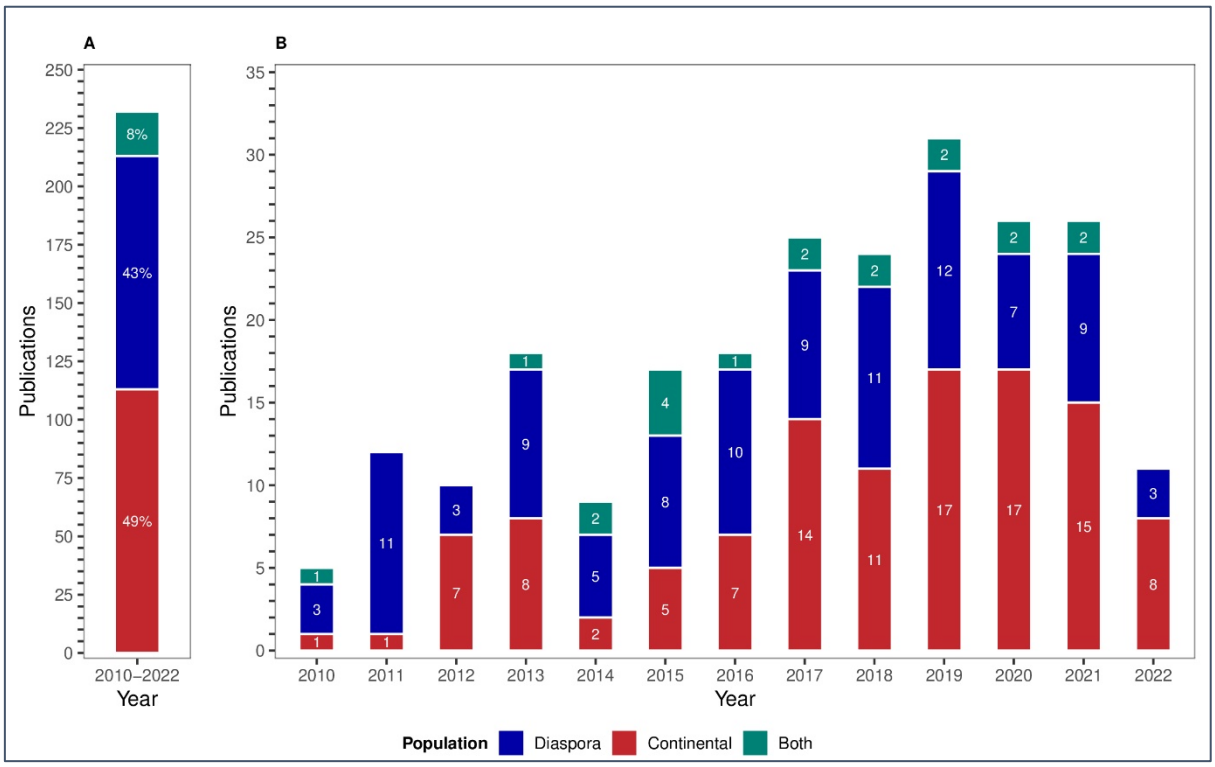


Figure 19: Number of publications by African-ancestry population geographic origin.

(A) Proportion of publications by population group over the full assessed timeframe (January 2010–June 2022) (n=232). (B) Proportion of papers by African-ancestry population origin per year.

Figure 20 illustrates the global distribution of studies investigating MM in African-ancestry populations. **Figure 20A** presents the overall coverage of papers, while **Figure 20B** displays the proportion of papers originating from different regions. Among the 113 papers focused on continental populations, the majority (55%) were conducted

in Southern Africa (n=61), followed by East Africa (22%, n=24) and West Africa (15%, n=17). Less than 5% of studies included multiple African regions (n=3).

Diaspora studies were concentrated in North America (95%, n=94), specifically among AA, who were often part of diverse ancestry studies examining the influence of race on MM-related health outcomes. In Europe, diaspora studies typically involved recently migrated individuals (born in Africa or first-generation migrants).

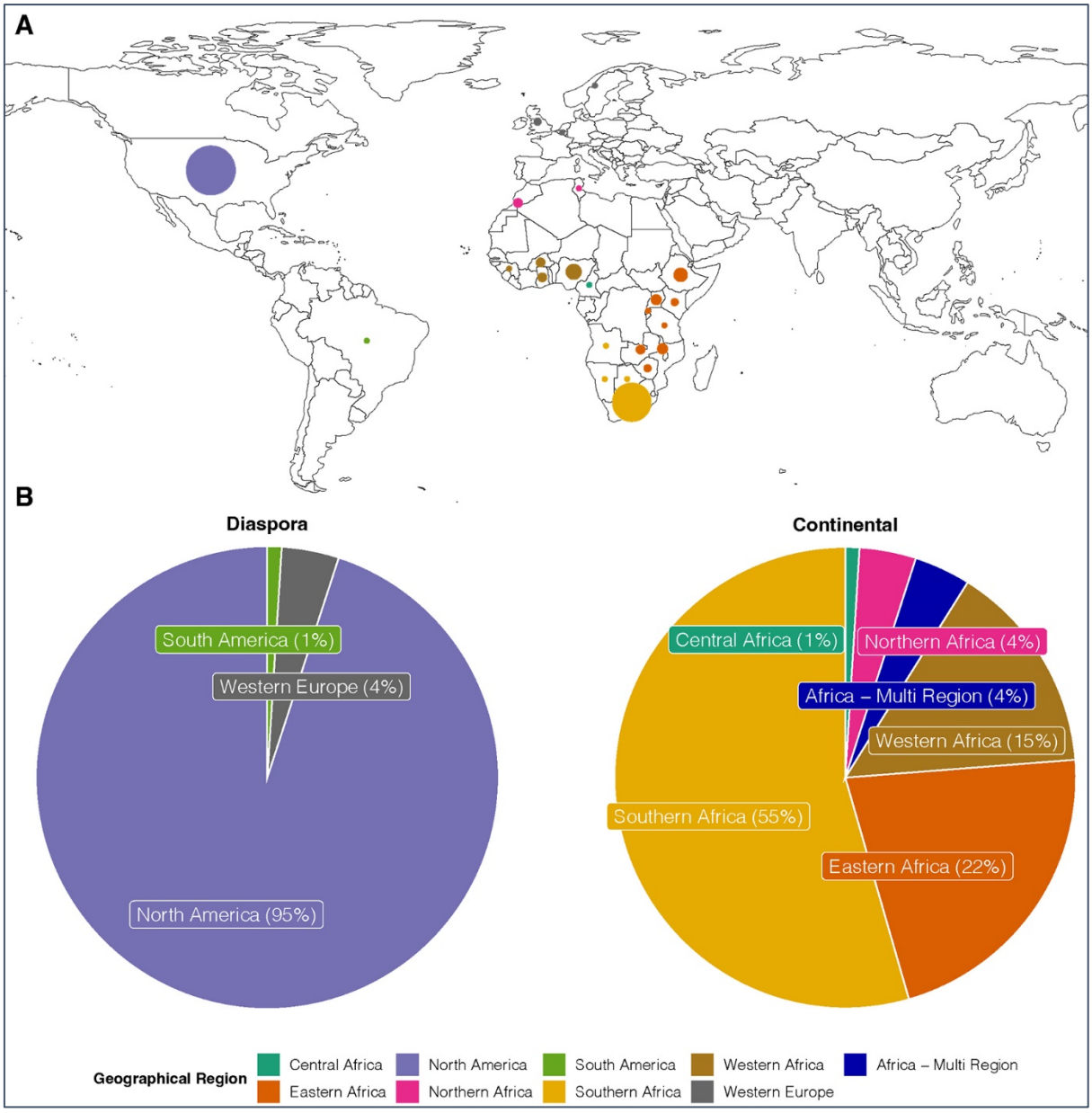


Figure 20: Global distribution of studies investigating Multimorbidity in African populations.

(A) Global location of participants from 213 multimorbidity studies in populations of African-ancestry. Studies including representation from both continental and diaspora

populations (n=19) are not shown. **(B)** Distribution of papers in diaspora (n=100) and continental (n=113) African-ancestry populations by geographical region (n=213). Regional distribution is based on the five regions of Africa described by the United Nations Statistics Division (Publication supplementary materials S17). 'Africa' includes countries from multiple regions.

In continental and diaspora groups, epidemiological studies focused on determining disease clusters and prevalence in specific populations. The most common study designs were cross-sectional (67%, n=155) and longitudinal (31%, n=73). Incidence measurement of MM was primarily observed in diaspora studies. Around two-thirds of the studies were conducted in hospital settings, including primary care and clinical settings, while 36% involved the general population. In the diaspora, secondary analysis of large cohorts or databases was prominent (64% of diaspora studies), while continental studies often utilised prospective hospital-based cohorts (32%). Continental studies had smaller sample sizes (median (range)=990 (14 to 417 786)) compared with diaspora studies (median (range)=2022 (20 to 3 739 528)).

Studies examined various risk factors for MM, including sociodemographic variables like age, sex, and socioeconomic status (SES). Research conducted in North America was particularly interested in comparing MM between what they identified as race/ethnic groups. Some studies explored the impact of MM on mortality, morbidity, hospital resource utilisation, polypharmacy, adverse drug reactions and care integration.

Disease and multimorbidity ascertainment

We considered a study to have a clear definition of MM when it was stated in the objectives, methodology, or discussion of the study. Our literature search found that most studies lacked clear MM definitions and showed considerable variability in disease inclusion. The terms MM and comorbidity were often used interchangeably, particularly in earlier studies.

Studies commonly employed disease counts to identify MM, but diseases included varied widely from 2 to over 200. Many studies used an author-developed disease list or, in some cases, a tool used to determine MM, such as the Charlson Comorbidity Index (275,276). These lists were available in 105 continental studies and 93 diaspora studies. In studies employing electronic health records or chart review, patient disease

was identified mostly by ICD9 or ICD10 codes, based on clinical diagnosis. In smaller cohort studies, primarily in Africa, disease was determined through self-reporting or proxy variables such as medication use, or when clinical biomarkers were outside normal ranges. Although more cost-effective and feasible in resource-constrained research environments, the latter, particularly self-reporting, can introduce biases like recall errors and misclassification, potentially skewing the accuracy and prevalence of diagnoses.

Figure 21 illustrates the frequency of MM combinations in continental and diaspora studies. The most studied cluster in continental studies was cardiovascular-infectious-metabolic diseases (n=16). In the diaspora, psychiatric conditions (represented by more than one condition but grouped for simplicity) were most frequent (n=11), followed by cardiovascular-metabolic (n=10) and metabolic-psychiatric (n=8) clusters. Cardiovascular (e.g., hypertension) and metabolic diseases (e.g., diabetes) were the most widely included in both groups. In continental populations, 80% of studies assessed cardiovascular (n=92) or metabolic (n=90) diseases, while in the diaspora, at least 70% focused on these conditions (cardiovascular (73%, n=72); or metabolic (72%, n=71)) (see **Figure 21**).

In contrast, the diaspora and continental studies differ greatly in their assessment of infectious diseases (predominantly HIV, Tuberculosis (TB) and malaria). Infectious diseases were more prevalent in continental studies (58%) compared with the diaspora (16%), whereas renal diseases, cancer, neurological and musculoskeletal disorders were under-represented in both groups.

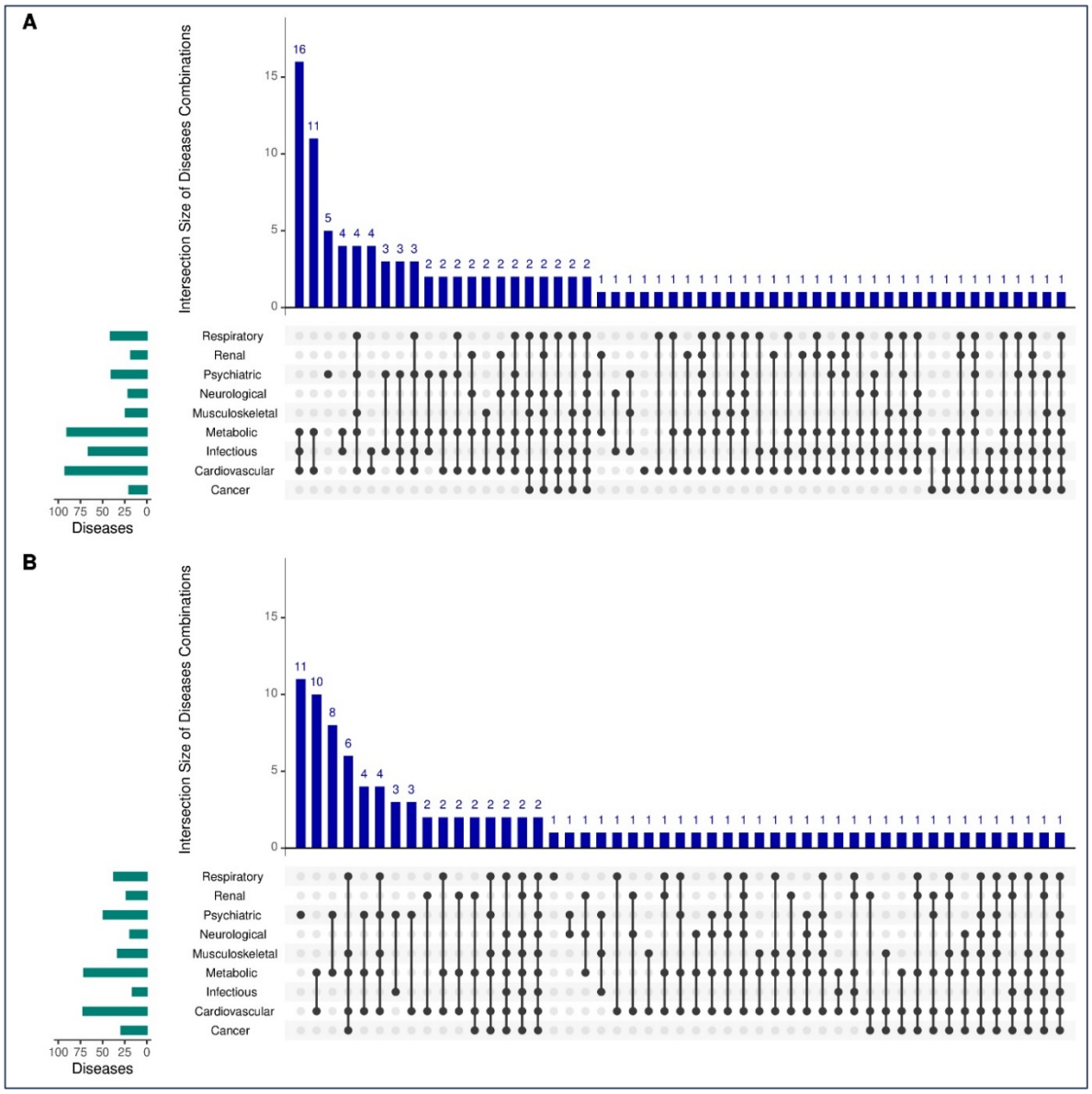


Figure 21: Distribution of disease cluster combinations (Cancer, Cardiovascular, Infectious, Metabolic, Musculoskeletal, Neurological, Psychiatric, Renal and Respiratory) across 213 studies stratified by population.

A. continental (113 studies) and **B.** diaspora African-ancestry populations (100 studies). upSet plots generated using RStudio.

Sociodemographic characteristics such as age, sex, household income and SES were examined in relation to MM prevalence and cumulative disease counts in studies involving continental and diaspora populations. Older individuals, particularly those aged 60 years and above, were found to have a higher risk of MM in both continental and diaspora populations. However, among continental populations, MM was also observed in relatively younger adults, approximately 45 years and below (277–280).

Differences in MM by sex were identified, with women often showing a higher prevalence of MM, particularly in cardiovascular-metabolic clusters (27,280–285). Low SES and income were generally associated with MM, but there were some population-specific differences. In the diaspora, lower SES was linked to higher MM prevalence and increased disease count (286–288) while in continental studies, the impact of SES was variable. Lower SES was associated with higher MM (289–294); but differences were observed. For example, studies in South Africa (295,296), Burkina Faso (285), and Ghana (297) found that higher income was associated with concordant cardiometabolic MM, while lower incomes were associated with infectious disease-related MM (290,292). In both populations, some studies revealed no effect of SES (277,298–300). Ancestry effects in diaspora studies varied depending on recent migrations to Europe or AA in North America. In the USA, AA ethnicity consistently posed a risk factor for MM across most MM clusters (11,265,301–304). One study assessing MM risk amongst men in a racially balanced and economically homogenous cohort in the USA found no difference in MM risk between AA and whites when in the same social environment (305). Conversely, studies on recent migrations indicated lower MM prevalence among African-ancestry migrant populations (306).

Studies also identified significant geographic variations in MM prevalence and patterns (66,284). Using a population-based survey in South Africa, Akindele et al. (2021) observed distinct provincial differences, with lower rates of multiple chronic lifestyle diseases among residents of Limpopo (284), and Wong et al. (2021) (66) demonstrated the intricate geospatial complexities of MM disease clusters within a rural community in Kwa-Zulu Natal.

The impact of behavioural risk factors, such as smoking, alcohol consumption and physical activity, on MM and treatment outcomes was assessed in some studies. When assessed, physical activity was noted as being largely negatively associated with MM, particularly within cardiometabolic MM, due to its impact on cardiometabolic risk factors, such as obesity and waist circumference (307). Alcohol and substance use were largely controlled for or considered as disorders themselves in most studies and therefore their effect on MM was not well described (277,290). However, assessment of studies examining alcohol use and smoking behaviour showed conflicting effects. For example, smoking and alcohol consumption was positively associated with multimorbid patients suffering from post-traumatic stress disorder in South Africa (308)

but negatively associated with multimorbid women newly diagnosed with breast cancer; however, these bivariate associations did not persist in multinomial models (309).

3.1.3. Thematic analysis

An in-depth review of included papers was used to extract recurring themes from the data. These themes were further reviewed and refined in consultation with the authorship team. While the quality of individual studies was not assessed in this review, caution was exercised in comparing the studies due to their diverse approaches in exploring MM in different settings. The themes below should not be viewed in isolation from one another, as they overlap somewhat and need to be integrated when considering solutions. First, we describe the limitations in terms of non-standardised terminology and data collection, a feature common across many disciplines, but we then highlight the heterogeneity among the combinations of chronic conditions that have been included in MM, but that in both African and diaspora studies the cardiovascular and metabolic clusters are most commonly studied.

A. Heterogeneity and inconsistency in terminology and methodological approaches

Similar to previous systematic reviews (262,278,310), our findings highlight substantial heterogeneity in the terminology and methodological approaches used in the identified studies. This variability is not surprising given the relatively recent emergence of MM as a concept in healthcare research and the ongoing debates surrounding its definition and measurement (257). Most studies used cross-sectional designs to estimate MM prevalence, while fewer employed longitudinal designs, possibly due to convenient and cost-effective participant recruitment. Continental African studies, with smaller sample sizes and reliance on self-reported data, may lead to biased prevalence and outcome estimates.

MM can be measured by counting the number of morbidities or using an MM index that considers both the number and severity of diseases, but the latter may be limited by data requirements (310). Variability in MM measurement is influenced by national health priorities and data availability. Furthermore, debates around disease clusters and groupings persist, with some advocating conditions on the same spectrum or

targeting the same organ system collectively rather than separately (311), particularly for well-managed conditions like hypertension (312).

B. Dominance of Cardiovascular-Metabolic disease clusters

Distinct regional variation in disease clusters was observed. Continental populations exhibited a co-occurrence of cardiometabolic clusters and chronic infectious diseases, while the diaspora showed a dominance of psychiatric conditions, often combined with cardiometabolic diseases. This aligns with previous research highlighting depression and cardiometabolic disorders as commonly observed MM clusters (262,313–315). Nonetheless, it is important to acknowledge that the available data suggest that the prevalence of condition clusters is strongly influenced by the context and population in which the research is conducted. Thus, when electronic health records were available, or when utilising large international cohorts, the number of diseases assessed to determine MM were much greater than smaller hospital or community-based cross-sectional cohorts.

In LMICs, such as those in Africa, chronic infectious illnesses like HIV/AIDS and TB are more prevalent. This is partially attributable to the longer life expectancies made possible by antiretroviral therapy, and the subsequent increase in susceptibility to NCDs at older ages [21]. Additionally, the cardiometabolic-infectious cluster may be exacerbated by the adverse effects of antiretroviral therapies on cardiovascular health (316–318).

The dominance of specific clusters limits the understanding of the prevalence of concordant and discordant conditions and their associations with other diseases and health outcomes. However, one continental study found similar prevalence rates for concordant and discordant MM (approximately 12% each) (285), suggesting their comparable significance. Another study examining the association between physical MM and depression in six LMICs, including Ghana and South Africa, demonstrated that physical MM significantly increases the odds of depression, with a dose-dependent relationship as the number of diseases increases (319).

C. Diverse effects of age, sex, and socio-economic status across populations

Factors associated with MM risk in African-ancestry populations, such as age, sex, education, SES, and comorbidities, were examined. Similar to studies conducted in high-income countries, older age, female sex, and lower SES were identified as risk factors for MM. However, nuances were observed, highlighting the importance of context-specific research.

Older age consistently showed a higher prevalence of MM in both populations; however, in continental populations, infectious disease burden contributed to younger adults being diagnosed with MM (66,277,280). Women of all ages were at increased risk of MM, potentially influenced by biological differences or social factors such as living and working environments, care-seeking behaviour and income inequalities (267). The relationship between sex and MM varies depending on the specific diseases considered; as suggested in previous research. We noted men to be more at risk when evaluating psychiatric disorders such as substance abuse disorders. Low SES, typically measured by education, income, occupation, or a composite of these, was identified as a risk factor for MM, with some disease-group intricacies. Also, SES may influence other factors such as sedentary lifestyles, health awareness and access to healthcare (292,298). Furthermore, disentangling social context and ethnicity in MM risk among diaspora populations remains challenging.

Diaspora studies often include race as a covariate to compare MM prevalence among different racial groups. However, variations in study methodology and terminology make it challenging to synthesise and interpret interethnic differences in MM risk. It remains unclear whether genuine disparities exist beyond differences in diagnosis or survival rates. USA-based diaspora studies suggested significant differences in MM prevalence and health outcomes across ethnicities, with AAs often having a higher prevalence. For example, Mochari-Greenberger et al., (2015) reported higher readmission rates among hypertensive Hispanic and AA patients with comorbid diabetes than Whites (320). Clements et al. (2019) noted AA had increased odds of most multiple chronic condition combinations compared to Whites, as well as and an increased risk of mortality even when adjusting for disease combinations (321). However, psychiatric disorders, such as depression and anxiety, were noted as being more prevalent in European-ancestry populations (265,322), likely due to under-reporting or underdiagnosing, particularly among men who may be more reticent to

seek help from mental health services, compared with non-AA. A recent migration study in the Netherlands revealed that ethnic inequalities in MM persisted when adjusting for the lower SES of ethnic minority groups (323).

D. Integrating care to address complex health outcomes associated with Multimorbidity

The literature consistently showed that MM is associated with higher levels of health resource utilisation (e.g., medications, primary care, emergency services) and adverse health outcomes (e.g., mortality, hospitalisations, and reduced quality of life), irrespective of disease clusters. Several studies, including those by Aparasu et al. (2005), Bhagavathula et al. (2021)(294), and Simakoloyi et al. (2022) support these findings. Although patients with MM require frequent healthcare access, social determinants such as SES, geographic location, and healthcare system factors can hinder access. Perry et al. (2016) (326) described these factors as significant barriers to quality patient-centred care for older multimorbid AA men, despite their higher MM burden.

Multiple studies have highlighted the limitations of traditional healthcare systems that focus on single diseases and have recommended a shift towards an integrated care model (ICM) where practitioners from various specialties collaborate in managing patients with multiple conditions (66,327–329). Wong et al. (2021)(66), emphasised the need for integrated management after observing high rates of uncontrolled hypertension (57.5%), diabetes (70.4%) and untreated TB in a rural South African population despite successful HIV management—78% of their cohort had undetectable HIV-1 RNA viral loads. However, implementing ICM poses challenges due to the increased complexity and time required for treating unrelated conditions. Limited evidence exists on the effective management of multiple conditions, particularly in resource-constrained settings like Africa. Nonetheless, Khabala et al. (2018) noted that ICM could improve the flexibility of care delivery and reduce clinician workload among patients with mixed chronic conditions in Kenya, and, in South Africa, ICM reduced HIV stigma in health facilities due to the non-segregation of patients (330), and with consolidated guidelines, patient education materials, and information systems, this could improve treatment and care for MM patients (331,332). Implementing ICM in South Africa is estimated to have minimal additional costs,

approximately USD1.06 (SD:0.33) per patient visit, over the current mean cost of USD4.94 (SD:0.70) (333).

Insufficient data exist to quantify the burden of non-communicable diseases (NCDs) in health facilities in sub-Saharan Africa and their readiness to manage the growing NCD epidemic. Some studies, mainly in South Africa, have investigated the possibility of expanding HIV clinics to address cardiovascular diseases like hypertension and diabetes (334). Despite the potential of integrated care to improve patient care and decrease individual and healthcare system costs, a significant gap in clinic and hospital staff training remains a barrier to effective implementation (335–338).

E. Challenges in statistical analysis approaches

Studies primarily employed univariate descriptive analysis, while logistic regression was the preferred method for assessing the relationship between MM and exposure variables or MM's impact on an outcome. While simple modelling techniques offer interpretability advantages (339), there is room for additional analyses through advanced machine learning (ML) and robust computational approaches. These approaches could reduce dependence on hand crafting by domain experts and scalability for the growing size of data and the number of variables (339,340). Few studies reported model goodness-of-fit assessment or variable selection approaches to refine multivariate models. Using the best subset of variables that contribute to the outcome of interest can lead to more useful and informative models.

Despite the advantages of imputing missing data and available software for implementation, few studies addressed missingness. There was limited assessment of the spatial and temporal distribution of coexisting diseases, which could provide valuable information on MM hotspots and associated factors. Such studies can also offer insights into the future MM population profile and the likely change in MM prevalence.

4. DISCUSSION

This scoping review identified 232 relevant articles and used thematic analysis to address the aim of the review; to identify, evaluate and summarise the available published evidence on MM in African-ancestry populations and identify research gaps

and future recommendations. The themes are often interdependent and should not be viewed in isolation.

4.1.1. Overview of included articles

The review revealed significant heterogeneity in MM terminology, methodology and disease measurement. While cardiovascular-metabolic disease clusters were dominant, regional variations were observed. Risk factors for MM in African-ancestry populations were similar to those found in European populations and high-income countries, including older age, female sex, and lower SES [99,100], with important exceptions. Therefore, the nuanced differences, highlighting the need for context-specific research and the identification of major gaps, need to be considered in the interpretation of MM in these populations.

4.1.2. Research gaps and recommendations for future research

Based on the included literature, we identified six research gaps and below we describe how they could be mitigated. These interconnected areas describe crucial interventions that would advance evidence-based research on MM and thereby potentially enhance patient care and health outcomes.

1. Paucity of data globally (sample size and complexity of data):

Limited data availability and a lack of large-scale epidemiological studies on MM, particularly in African-ancestry populations, hinder our understanding of MM prevalence and patterns in diverse populations. Future studies should increase sample size and undertake longitudinal population-based studies to better understand MM's longitudinal trajectory. Additionally, including genetic data in these studies can offer valuable insights into MM causation and the interplay between genetic and environmental factors. Gathering comprehensive data will improve our understanding of MM and facilitate more targeted interventions.

2. Lack of representation (many African countries and ethnic groups have no published data)

African-ancestry populations are under-represented in the MM literature, both within and outside of Africa. Furthermore, other socioeconomic and health factors impact MM

prevalence in groups and marginalised communities. More research is needed in African-ancestry individuals in Central and South America as well as recent and growing migrant populations in Europe and elsewhere, who may have unique MM experiences. The Research on Obesity and Diabetes among African Migrants (RODAM) study highlights the influence of migration on MM risk factors in people originating from Ghana and emphasises the importance of understanding the sociodemographic, cultural, and healthcare contexts of African-ancestry migrants for a comprehensive understanding of MM (341).

3. Underrepresented disease areas (health priority mismatch)

Upon assessing the Global Burden of Disease (GBD) morbidity estimates both globally and within sub-Saharan Africa (SSA), a mismatch between health priorities and the diseases investigated is evident (GBD morbidity estimates can be found in Publication supplementary materials S18). Cardiometabolic and infectious diseases are the leading disease burdens in SAA (272), which aligns with the focus of continental studies. This alignment is also true for cardiovascular diseases globally. However, mental health conditions among adults younger than 50 years and cancer among individuals older than 50 also account for a significant health burden in SSA, yet these clusters, particularly cancer, appear to be rarely assessed in continental MM studies. Additionally, as life expectancy rises in Africa, musculoskeletal and neurological diseases associated with the elderly will become more prevalent and should be investigated.

Moreover, there is a paucity of evidence on how clusters of conditions develop and change over time, making it challenging to predict how disease burdens might change within an individual's life, and to identify the best interventions and care approaches. Key factors to consider when assessing long-term health in SSA include not only infectious and NCD but also the combined effects of childhood malnutrition, and high rates of maternal HIV infection (66).

4. Lack of standardised study designs and data collection protocols

The absence of standardised definitions and measurements and uncertainties about the applicability and validity of MM assessment tools in African-ancestry populations must be addressed. Moving forward, it is crucial to establish standardised criteria and

methods to ensure consistent and accurate identification of MM across populations and regions while recognising the importance of context-specific assessments. This should be applied both to study communities that remain unexplored and to all the different disease clusters.

5. Need to leverage additional data analysis approaches

In addition to traditional statistical approaches to epidemiological research, recent studies beyond African populations have utilised emerging machine learning (ML) techniques to understand the complex nature of MM (342,343). These ML techniques can be categorised into pairwise, probabilistic, factorisation, and temporal methods (342). Pairwise methods are relatively simple but limited in capturing complex interactions and multiple co-occurring diseases common in aging populations. Probabilistic methods, such as Hidden Markov Models, provide an overall view of disease relationships. Factorisation methods, like non-negative matrix factorisation, offer interpretability but may struggle with encoding suppressiveness among diseases. Advanced ML techniques, including deep learning, show promise in unravelling the complexity of MM (342). These techniques can infer co-occurring diseases, assess data trustworthiness, identify patterns, analyse longitudinal aspects, integrate multiple data sources, and provide interpretable insights. Future studies should also consider employing ML and prediction modelling approaches, building on successful collaborations between domain experts and ML researchers.

6. Invest in translational research

To comprehensively understand the interplay between MM, the healthcare system and the cultural/religious context, diverse perspectives must be considered. The gaps described above, therefore, need to be addressed in the specific milieu of the study communities and adapted to suit the context, while striving to generate interoperable data fit for use in comparative studies. Despite it being crucial to explore patients' experiences, caregivers' challenges and clinicians' perspectives regarding MM in African-ancestry populations, there is currently limited research in these domains. Additionally, cultural, and religious beliefs will likely influence perceptions and behaviours related to MM. Assessing the financial and social costs of MM is also necessary. This knowledge can inform policy shifts, treatment approaches and clinical guidelines, leading to improved diagnosis, risk prediction, and treatment strategies.

4.1.3. Strengths and limitations of this scoping review

This scoping review is the first to summarise the literature on MM prevalence and patterns in African-ancestry populations worldwide. Dividing studies into continental and diaspora African-ancestry populations revealed important differences. By involving a multidisciplinary team of epidemiologists, demographers, clinicians, geneticists, and data scientists, with representation from multiple African countries, we optimised the inclusion of relevant studies and ensured cultural sensitivity during the writing process. However, limitations such as limiting our search strategy to title and abstract only and assessing open access texts or those available through institutional logins may have resulted in some literature being excluded. The literature did not lend itself to a systematic review or meta-analysis and, therefore, we did not perform quality assessment on the publications using standardised tools. Nevertheless, our results emphasise the necessity for additional research on MM in African-ancestry populations and the development of context-specific guidelines for prevention and management across different geographic locations.

5. CONCLUSION

In reviewing the literature on MM in populations with African-ancestry, we observed distinct disease clusters and varying effects of social determinants of health on MM. We provide strong evidence that Black Americans (also referred to as AAs), the most widely studied African-ancestry group, are not a good proxy for all Africans. Furthermore, continental African regions and populations are highly diverse in their health profiles, exposures, and behaviours. To address MM effectively in African-ancestry populations, we identified six research gaps: limited data availability, inadequate representation, under-represented disease clusters, absence of standardised study designs and data collection protocols, the need for innovative data analysis approaches and the need for more translational research. Additionally, considering the complexities observed among populations, it is crucial to account for location, resources and environmental context when developing interventions to mitigate the burden of MM, especially in African-ancestry populations.

SECTION 3: INTEGRATED DISCUSSION AND CONCLUSION

Section overview

The aim of this thesis was to leverage continental African genomic data to better understand the value of including genomic data in CVD risk prediction models in African populations and to evaluate the potential clinical utility of such precision medicine approaches in Africa. This chapter presents a discussion of the conclusions and challenges that emerge from this body of work. First, I summarise the four publications presented in Section 2: Chapters 1 to 4, and then discuss the main themes that they raise, specifically:

1. Including genetic factors enhances CVD risk prediction models in African populations;
2. Current data availability and Euro-centric computational approaches limit the potential of precision medicine in Africa; and
3. Advancing precision medicine in Africa requires context-sensitive, cross-disciplinary collaboration and multistakeholder partnerships.

Second, I outline some key limitations relating to genomic research in African populations. Finally, I discuss potential directions for future research, including multi-ancestry analysis and validation in external datasets, and end with an overall conclusion that describes the key take-home messages.

Chapter 1: PUBLICATION SUMMARIES

1.1. Paper 1: Has translational genomics come of age in Africa?

This paper critically evaluates the progress of genomic research in Africa and its implications for healthcare. It examines whether advancements in African genomics have effectively translated into improved healthcare for Africans. Bearing in mind the complexities and context-specific challenges within Africa, five distinct genetic service models were identified across the continent. Despite the variation in challenges and the extent of those challenges in different locations, the review underscores persistent and common obstacles that genomics encounters in Africa. Obstacles include limited infrastructure, funding gaps, inadequate numbers of skilled geneticists, and the strain of multiple disease burdens on healthcare systems. Despite these difficulties, optimism prevailed among dedicated clinicians and researchers who envision innovative solutions and technological leaps that could bridge the genetic services gap. We advocate for a comprehensive approach involving capacity building, dedicated funding, political support, legislation, public education, and targeted research. We stress the need for cross-disciplinary collaboration to accelerate genomic integration into the clinical setting, narrow the genetic services divide, and realise the full potential of precision healthcare in Africa.

1.2 Paper 2: Ancestry-aligned polygenic risk scores combined with conventional risk factors improve cardiovascular disease risk stratification in African populations.

Cardiovascular diseases (CVD) are the leading health burden in Africa and identifying high-risk individuals using standard risk stratification approaches is challenging. This study aimed to improve CVD risk stratification by developing and critically evaluating an integrated CVD risk score designed for African populations by combining traditional clinical risk factors with genetic risk factors. Using continental African discovery (APCDR, n=14,126) and target (AWI-Gen, n=10,603) datasets, this study developed and evaluated the predictive utility of 14 cardiometabolic trait ancestry-aligned PGS for seven CVD-associated outcomes (CVD, heart attack, stroke, type 2 diabetes, dyslipidaemia, hypertension, and obesity) in continental African populations. We investigated whether modelling these PGS into a MultiPGS improved prediction and if

integrating this MultiPGS with established non-genetic risk factors had greater predictive utility beyond non-genetic factors alone. Results showed variable prediction power of PGS within and across cardiometabolic traits, with significant associations found particularly with dyslipidaemia, hypertension, and obesity. Integrating genetic data with traditional risk factors offered better prediction than non-genetic models alone, significantly enhancing CVD risk prediction in African populations. This study suggests that including population-specific genetic data and conventional risk factors can improve the accuracy of CVD risk prediction models for African populations.

1.3: Paper 3: Clinicians' Perceptions towards Precision Medicine Tools for Cardiovascular Disease Risk Stratification in South Africa.

This study delves into the perceptions of South African clinicians regarding the adoption of precision medicine (PM) for CVD risk stratification. It addresses a gap in the understanding of healthcare workers' attitudes towards PM tools in Africa. Surveying 109 clinicians from Johannesburg's major teaching hospitals, the research examines clinicians' understanding and knowledge of genetics, their perceptions of PM-based CVD risk stratification, and their confidence in applying these approaches. The study also explores barriers to incorporating PM into clinical practice in South Africa. Despite low familiarity with genetics and PM concepts, clinicians' expressed optimism about PM's potential, echoing findings from other regions - the majority (84%) of clinicians hold positive views on PM approaches. However, confidence in adopting these approaches varied, particularly among those engaged in CVD risk assessment, suggesting that integrating genetic risk assessment within existing clinical frameworks could enhance confidence and facilitate PM adoption. Key barriers identified included high costs and limited access to genetics services. The study also underscored the need to address the genetics education gap, suggesting improvements in genetics curricula and clinical training. The findings emphasised that overcoming these barriers and fostering clinician confidence is crucial for successful precision medicine integration in South Africa's public health system.

1.4 Paper 4: Multimorbidity in African ancestry populations: a scoping review

The study conducted a comprehensive review of multimorbidity (MM) in African-ancestry populations, on the continent and in the diaspora. It analysed 232

publications, (113 studies), the diaspora (100 studies), and both settings (19 studies) between 2010 and June 2022. Results revealed diverse MM patterns within and beyond Africa. While both settings frequently assessed cardiovascular and metabolic diseases, infectious diseases are more prevalent in continental studies. Risk factors for MM in African-ancestry populations mirror those in European and high-income countries, including older age, female sex, and lower socioeconomic status. The study underlines the unique MM patterns in these populations and emphasises the need for more representative and context-specific research. Six research gaps were identified:

1. Limited data availability,
2. Underrepresentation of populations,
3. Neglected disease areas,
4. Lack of standardised study designs,
5. Innovative data analysis approaches, and
6. Insufficient translational research.

We emphasised the need for inclusive, nuanced research reflecting the heterogeneity of African-ancestry populations and cautioned against generalising findings from African American studies to all African-ancestry groups. The study advocated for more context-specific research methodologies and policies to address the growing MM burden in these populations.

Chapter 2: KEY THEMES IDENTIFIED FROM THIS RESEARCH

Three key themes emanated from this body of work, namely:

1. Genetic information provides independent and complementary information in CVD risk prediction models
2. Data and computational methods shortfalls hinder the translation of precision medicine in Africa
3. Advancing Precision Medicine in Africa requires context-sensitive, cross-disciplinary collaboration and multistakeholder partnerships

These themes are contextualised within the key domains of this research, *Precision Medicine in Africa, Cardiovascular Disease, and Clinical Translation* (**Figure 22**). These themes, and their intersections, will be addressed in the integrated discussion.

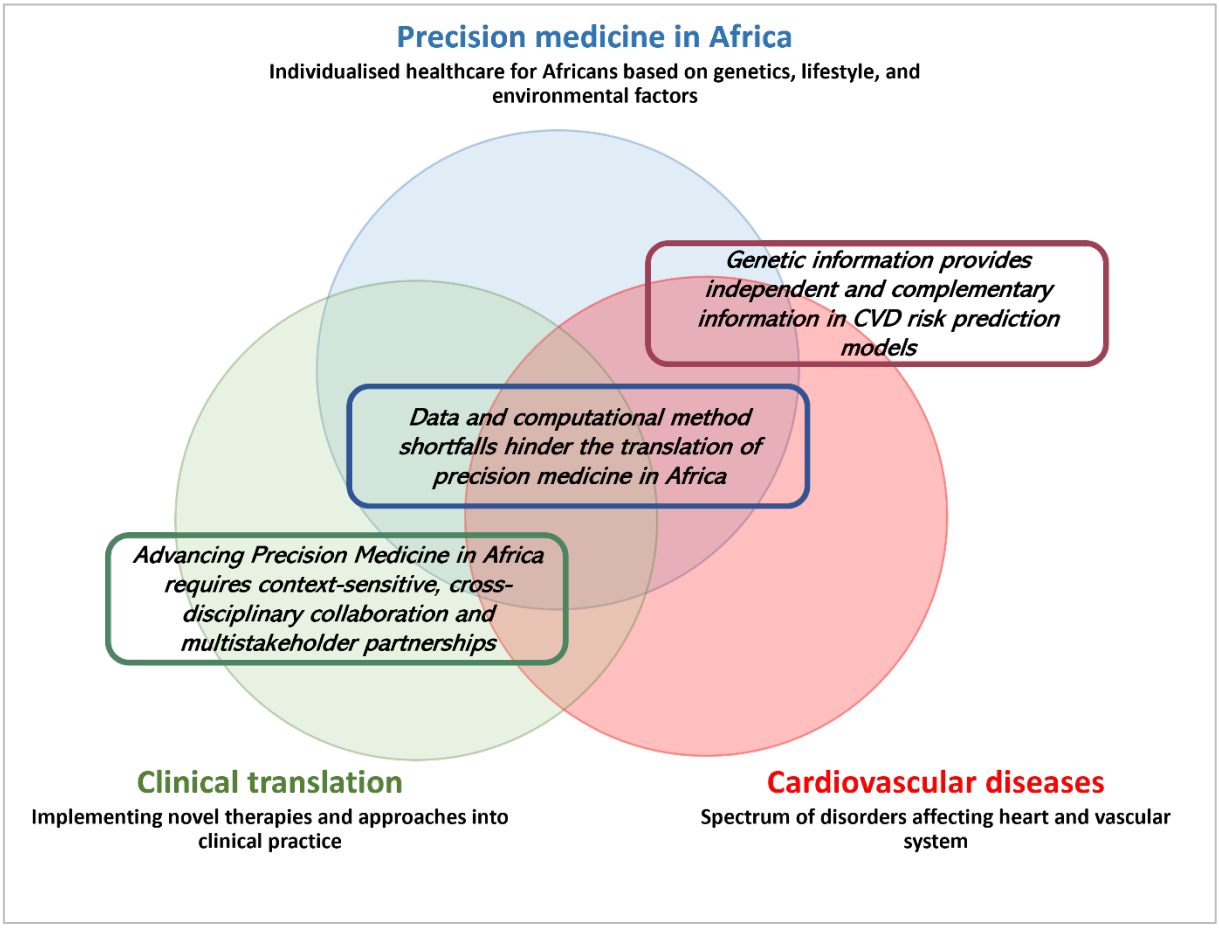


Figure 22: The key themes emanating from this thesis in relation to three key domains of interest.

1. Genetic information provides independent and complementary information in CVD risk prediction models

In our study, we developed novel ancestry-aligned integrated risk scores for various CVD-associated outcomes (Publication 2) and evaluated their predictive utility in the African context. Although a limited sample size hindered the development of a risk score for CVD, we successfully developed predictive models for obesity, hypertension, dyslipidaemia, and diabetes. However, the variance explained for diabetes was small. Given that these factors, along with smoking, have been seen to contribute to over 90% of coronary heart disease cases (344), accurately assessing an individual's risk for these phenotypes may be crucial in preventing cardiovascular morbidities and mortality, especially in high-risk groups yet to experience a cardiovascular event (14).

Our integrated risk score, inclusive of genetic information, through PGS and ancestry, with traditional non-genetic factors demonstrated that incorporating genetic factors alongside conventional non-genetic factors improves prediction over non-genetic factors alone. Risk prediction (measured as R^2) increased by at least 2.5% across disease outcomes, and analysis revealed models to be statistically different from each other, illustrating that genetic information enhances prediction, and provides independent and complementary information in risk assessment (Publication 2). These results have several clinical implications.

First, our study shows that integrated models enhance the prediction of CVD-associated outcomes. Current South African guidelines recommend statin prescriptions for individuals with a 10-year atherosclerotic CVD risk of ≥15% and LDL-C levels above 2.5 mmol/l (25). This risk assessment is based solely on clinical risk factors. Our findings suggest that including genetic information can lead to more accurate predictions. Consequently, some patients, who under current guidelines, would not meet this threshold, may exceed the threshold when their PGS are incorporated, allowing earlier risk mitigation. Conversely, PGS insights might also lower some patients' risk below the threshold, thereby delaying screening and reducing over-medication. However, recommending to withhold screening for low-risk patients would likely face resistance, at least until clinicians and health systems have greater confidence in the technology and there is sufficient data on the clinical use of PGS upon which to formulate guidelines (345).

Second, we showed that PGS of routinely collected clinical measurements, such as blood pressure and cholesterol levels, can be predictive of cardiometabolic outcomes associated with CVD. This finding is particularly useful in the African context as large-scale disease-specific cohorts needed for GWAS and post-GWAS are not readily available. By using routinely collected standard clinical measurements, prediction models can be investigated and the use of available data is optimised.

Third, PGS can be simultaneously calculated at any point for various common diseases, enabling the prediction of lifetime risk early in life and facilitating earlier, more targeted prevention (15). Commonly used clinical risk estimators for coronary disease were optimised for use in middle-aged adult populations in historical cohort studies and consequently underperform in younger populations or individuals of non-European ancestries (346). Genetic information has a key role to play in risk prediction when clinical variables do not yet show risk for disease, particularly at younger and middle ages. Among asymptomatic younger adults, a high CAD PRS was shown to prompt more intensive earlier efforts for lifestyle modification and potentially earlier statin initiation (347). The increasing affordability of genotyping furthers the potential of PGS being adopted at scale, although it is still prohibitively expensive in many LMIC settings and research is lacking on which SNP arrays would most accurately capture relevant variants in different African populations (345).

Lastly, our work moved beyond assessing the transferability of risk models derived in European-ancestry populations to continental African populations but rather developed ancestry-aligned scores using both base and target data from continental African populations. This approach confirms that genetic information has value in CVD risk prediction in African ancestry populations and highlights the effectiveness of scores specifically tailored to African populations residing in Africa.

While our results are promising, further investigation into the scores and their clinical application is necessary. The American Heart Association (AHA), recently released a scientific statement on the use of PGS for CVD and suggested three broad criteria be considered by health care systems before CVD PGS implementation: (1) efficacy, (2) harm, and (3) logistics (348).

Addressing the first consideration, efficacy, i.e. predictivity of the PGS in the target

population, we note that the improvement provided by our risk scores is modest. This is not necessarily surprising as PGS typically explain only a small fraction of trait or disease variance. For complex diseases like CVD, the discriminative ability of PGS is compromised by environmental factors and the imperfect measurement of the full genetic signal associated with the trait or disease - PGS do not consider rare and structural variants nor potential gene–gene interactions (51,345). The value of PGS is correlated to trait heritability and are most effective where heritability is significantly high. For example, if the heritability of a disease is 2%, indicating a minor genetic basis for the disease, PGS would be less useful (345). Translating PGS into clinical tools requires transforming relative risks across the PGS continuum into absolute disease risks, as having a very 'high' PGS, say the 99th percentile, has almost no meaning if that PGS captures only 1% of variance in disease risk (46). This is even more true if the prevalence of the disease in the population is low (204).

Defining the clinical benefit of the modest improvements gained from PGS when applied broadly to diagnostic or treatment regimens is challenging and remains a key debate amongst field experts (51,349). Few studies have sought to quantify the health and economic benefits of cardiometabolic PGS (350,351). Sun et al. (2021) (350) conducted modelling analyses to estimate the number of lives potentially saved in the UK by including PGS and revealed that their use could prevent more than 7% of CVD events than conventional risk factors alone. They also noted that a targeted prevention strategy would be fifteen times more efficient than a general population-wide approach, i.e. assessment only among people at intermediate (i.e., 5% to <10%) 10-year CVD risk compared to screening the entire UK primary care population aged 40–75 years irrespective of risk (350). A Finnish study found that conducting coronary heart disease (CHD) PGS testing is cost-effective, especially when implemented at scale, reducing the average expected costs for patients aged 45 and older by €2.54 over a 10-year periodperiod (351). The authors highlight these savings at scale and show that amongst 100 000 Finnish patients aged 45 plus, approximately 2190 patients would be eligible for PGS testing, and thus total costs associated with CHD would be decreased by €254,000

It is also essential to assess the suitability and effectiveness of current interventions for genetically defined risk groups in managing CVD (64). Given the significant

heterogeneity of healthcare systems across Africa, it is essential to evaluate the feasibility of guided treatment and care approaches within specific contexts (205,206). This evaluation must consider whether awareness of risk, in the absence of adequate clinical options, might inadvertently cause harm (64).

In the African context, addressing potential harms as per the AHA's statement, requires careful consideration of the applicability of PGS across diverse ancestries. Specifically, the lack of transferability of current largely European-derived PGS models to African populations, and their subsequent potential to exacerbate global health inequalities (57,60,200,202). This consideration closely aligns with our second theme; *Data and computational methods shortfalls hinder the translation of precision medicine in Africa*.

2. ***Data and computational methods shortfalls hinder the translation of precision medicine in Africa***.

This theme focuses on the research gaps and the urgent need for context-specific methodologies in studying health burdens among African-ancestry populations. African populations have been historically underrepresented in global health and genetic research (59). Publication 1 and Publication 4 stress the importance of addressing this gap with research that is tailored to the specific health profiles, genetic backgrounds, and environmental contexts of these populations, and call for more inclusive and nuanced research methodologies, emphasising the need for more localised and representative studies.

The need for locally relevant research stems from the region's diverse demographic history and genetics. Unlike European-centric studies, the assumption of homogeneous population structures is unsuitable for Africa, where significant population mixtures and movements have occurred over thousands of years. Africa is a genetically diverse continent representing 54 countries with over 3000 ethnic groups (352). However, much of its genomic data is attributed to Africans living in the diaspora, which is not an adequate representation of the diverse genome that span the continent, and they are often admixed populations where individuals have varying degrees of African ancestry (58,353). Furthermore, Africans live in different environments that can affect health outcomes through gene-environment interactions. Despite our study addressing this challenge and deriving ancestry-aligned scores (Publication 2), the

predictive utility of PGS are not substantially higher to scores developed for similar outcomes using European GWAS (98,200,354).

Eurocentric bias in genomic research has had widespread consequences on the development and utility of genomic tools and resources (e.g. databases) resulting in challenges that need to be addressed (355). Imputation panels are vastly Eurocentric, limiting representation of the greater haplotypic diversity present in Africans from deeper recombination history (171,172,185). The most widely available African sequencing resources have biased representation towards African Americans and West Africans (97), leaving a large proportion of African diversity undocumented (58,355). The lack of extensive, accessible reference panels in Africa hinders data analysis and access. Efforts like the African Genome Variation Project (171) strive to fill this gap but still lag behind in terms of capturing diversity as resources such as the 1000 Genomes Project for more homogeneous populations.

These issues are exacerbated when assessing individuals who have an admixed multi-ancestry genetic background. In these cases, the individual does not fall within any single ancestry group and therefore their ‘true’ score is unlikely to be accurate if one resorted to an averaging of their two or more single-ancestry PGS. Currently, there is no optimised solution for developing PGS for admixed populations and multiple efforts are underway to address this imbalance. These limitations highlight the need for more data from multiple different ancestry groups and the development of methods to address admixture so that PGS can benefit all people (345).

The lack of accurate PGS in African ancestry individuals may cause barrier to achieve precise risk stratification which is critical for precision medicine. Coordinated efforts such as the Polygenic RIsk MEthods in Diverse Populations (PRIMED) (188) aims to deliver new methods for risk prediction in diverse ancestry and specifically a pan-Africa initiative—CARdiometabolic Disorders IN African-ancestry PopuLations (CARDINAL) (356) project which aims to test the performance of PGS in African individuals with phenotype and genotype data available from H3Africa projects.

As data accumulates and analytical methods are refined, the accuracy and reliability of scores will improve, possibly reclassifying patients' risk levels and altering screening and treatment recommendations (51,345). Although individuals with genome-wide

data already recorded should benefit from these improvements without needing to provide another sample, these updates pose significant challenges for hospital systems. Questions remain around who and how the approval, adoption, and communication of updated versions will occur. Adopting updated scores is a complex process and will require a standardised system and clinical guidelines to facilitate the roll out of updated scores (345). As the field is still emerging, there is notable heterogeneity in the application and reporting of risk scores, which hinders the clinical translation of PGS (349). Initiatives such as the Polygenic Risk Score Reporting Standards (PRS-RS) (349) and the publicly accessible Polygenic Score Catalog (357) aim to aid standardisation, transparency, and reproducibility of PGS.

Risk communication is key for reducing harm. Effort must be made to reduce misinterpretation or misapplication of PGS information and risk score results due to clinician knowledge or patient understanding (51). It is also necessary to support patients and their families with understanding polygenic risk and what this might mean for their future health as well as the interventions that they may or may not choose to protect against the risk (345). There are also implications for family members and these need to be addressed. Effective risk communication strategies are needed to make this information accessible and comprehensible, ensuring patients are well-informed without being overwhelmed (358). Excessive details could discourage patients from undergoing such screenings, impeding the future improvement of PGS and their ability to identify high-risk individuals (345,358). Current evidence show that the provision of genetic information alone may improve health-related behaviours, but results are inconsistent (359,360). Widén and colleagues (2022) (359) showed more than two-fifths of study participants at high risk for ASCVD took action to reduce their risk during follow-up, and that both the information on clinical and the genomic risk factors independently contributed to this behavioural change (359). Still, understanding the possible harm of providing genetic information (such as increased anxiety), especially at younger ages, needs to be evaluated.

Closely tied to the AHA's suggested consideration around logistics is the third theme emanating from this work, *"Advancing Precision Medicine in Africa requires context-sensitive, cross-disciplinary collaboration and multistakeholder partnerships"*.

3. ***Advancing Precision Medicine in Africa requires context-sensitive, cross-disciplinary collaboration and multistakeholder partnerships***.

This theme, highlighted across all publications, delves into the multifaceted challenges faced in health and genomic research in Africa and translating discoveries into improved healthcare for Africans. These challenges are rooted in infrastructural inadequacies, such as limited electricity and poor road access, and extend to funding gaps and a scarcity of skilled geneticists and healthcare professionals (205).

Although funding and workforce shortages are shared across emerging and established settings, the scale and impact of these issues differ markedly. In established settings, substantial funding and political support are available for genomic research and clinical translation, as evidenced by national initiatives such as Genomics England in the UK (361) and the All of Us program in the US (362). Also, in these settings, although there is a greater access to specialised skills (bioinformaticians, medical scientists, genetic counsellors, clinical geneticists, etc.) substantially more are needed to scale services. In contrast, less mature settings have limited clinical and genetic research skills to establish basic genetic services, let alone provide them at a population level. Additionally, the strain of high and multiple disease burdens further complicates the landscape (277).

Publication 1 and Publication 3's assessment of barriers to genetic service delivery and the advancement of precision medicine in Africa emphasise the need for context sensitive, locally relevant research with cross-disciplinary collaborations. Furthermore, they highlight that to enable the adoption and scaling of translational genomics across the continent, this research should be supported by substantial investment in upskilling personnel in the field of genomics and investing in relevant infrastructure (Publication 1 and Publication 3).

Conducting genetic research and the subsequent translation of findings remains challenging in Africa with numerous infrastructural, funding, and ethical challenges, as well as methodological and analytical complexities. Genetic services and broader healthcare infrastructure varies substantially across the continent but remains largely limited and highly fragmented, with most services restricted to large cities and/or regional hubs (Publication 1). In addition to being few, laboratory capabilities in the

region are further constrained by the need to import costly reagents, often facing long delivery periods. Furthermore, data and knowledge on developing appropriate genetic diagnostic tests for specific populations in Africa remains limited. The scarcity of biobanks complicates sample storage and processing. While initiatives like H3ABionet have advanced bioinformatics infrastructure, issues with internet connectivity and inadequate computing resources continue to pose significant barriers. Furthermore, Publication 4 showcased the need for understanding health landscapes and associated disease patterns at the continental, regional and local level, and that utilising insights gained from African American populations and other African diaspora populations are not necessarily applicable to Africans residing on the African continent.

To gather support and promote political will conducting cost-effectiveness studies in Africa is crucial but challenging due to extensive data requirements, such as using national electronic health records, which are mostly unavailable across Africa (205,206). Innovative approaches assessing key cost components - one-off costs of genotyping (and the associated infrastructure), algorithm development and ongoing costs such as laboratory and bioinformatics staff - should be explored (348).

Moreover, although clinicians are positive about genetics and precision medicine, and open to incorporating such in CVD risk assessment (Publication 1 and Publication 3) in South Africa, the awareness of genetics among healthcare providers and patients is limited, potentially hindering putative benefits, and exacerbating harms associated with PGS. This finding is not unique to South Africa, and seen in across several African countries (175,363). Human resource issues, such as high turnover and brain drain, exacerbate these difficulties, highlighting an urgent need for stronger support systems for genetic capacity building in the region (345).

Despite these hurdles, there is a sense of optimism and a push for innovative solutions (Publication 1). African investigators need to be meaningfully included, as greater inclusivity and enhanced research capacity afford enormous opportunities to accelerate genomic discoveries that translate more effectively to all populations. Proponents of international collaborations argue that working with high income countries will eventually ensure equity, justice, and benefit to Africans, with capacity building for genomic research providing immediate benefit for African institutions (364).

Genetic services must align with the diverse cultural values and traditions of African communities, and engaging with communities to shape these services is essential.

In short, clinical implementation of precision medicine-based approaches, such as the integrated CVD risk score, in Africa, will require an active effort to address the gap in genetic training whilst simultaneously investing in the genomics capacity, including both skill force and infrastructure (205,365).

Regarding CVDs, and reduced morbidity and mortality on the continent, precision medicine approaches such as the integrated risk score, still remain only part of the solution. The majority of CVD events will likely still occur among lower-risk categories (below intervention thresholds), and therefore to address CVD's effectively a two-pronged prevention strategy will still be the most effective (366). Thus, CVD prevention will include population-based public health strategies aimed at lowering CVD risk, such as increasing physical activity, improving nutrition, and reducing smoking, as well as utilising enhanced individual-based screening approaches (17).

Chapter 3: STRENGTH, LIMITATIONS, AND FUTURE DIRECTIONS

Our study is unique in that it is the first study to derive PGS for cardiometabolic traits using both a continental African base and target dataset. This study moved away from assessing the transferability of European or a multi-ancestry derived scores in continental African populations, but rather assessed the potential of ancestry-aligned scores tools to serve continental African populations. Nonetheless, despite the study's novelty, there are limitations that should be discussed.

Three key limitations were noted across study objectives: 1. Paucity of data, 2.Eurocentric genomic tools and computational approaches, and 3. Validation challenges and generalisability of findings. How they relate to two of the studies is outlined below.

Integrated CVD risk model

Within the integrated modelling component of the work, there was a scarcity of phenotypic and genomic data relating to CVD among continental Africans, and when available, cohorts were often cross-sectional and small. Small sample sizes of cross-sectional studies lead to few diagnosed cases and subsequent limited power to detect robust associations. Single-risk assessments, rather than longitudinal data, resulted in PGS and models being derived from prevalent cases. Although AWI-Gen is a longitudinal cohort, current AWI-Gen data spans only five years and is not yet available. The unavailability of longitudinal data limited longer-term analysis and comparative assessments with existing 10-year CVD risk tools. The study also relied on self-reported data, which could lead to recall bias and misclassification.

Also, although the integrated CVD risk score modelling made use of two datasets (APCDR and AWI-Gen) - both cohorts are in-depth, genetically informative cohort studies - and nested cross validation was used, there are limitations to the study, most of which are shared by other PGS studies. Our study was restricted to participants aged 40 to 60 in AWI-Gen and is, therefore, limited by the characteristics of this cohort and by the lack of additional external datasets for validation. Furthermore, the cohort, comprising individuals from select African countries (Burkina Faso, Ghana, Kenya, and South Africa), may not represent the genetic and environmental heterogeneity of wider African and other super populations (European, Southeast Asian, etc.).

Additionally, similar to previous studies, we applied our PGS uniformly to all individuals, regardless of their age, sex, or other relevant factors. Studies have shown that PGS prediction may vary by sex and age (55), and we note that constructing group-specific PGSs may further improve predictive power if there were sufficient data for training.

Currently, the greatest hurdle for PGS implementation in the clinic is the relative lack of validation studies, with very few PGS studies including prospective cohorts to validate the models developed. The need for validation is two-fold: 1. PGS validation is required to ensure observed model patterns are accurate to real world data and can therefore accurately predict an individual's risk of developing a disease, and 2. validation of these models is critical to their acceptance by accrediting bodies, insurance companies, hospital systems, clinicians, and patients.

A further complication, and discussed in Key Themes: Theme 2, Eurocentric computational tools, including sequencing resources, LD and imputation reference panels, do not adequately represent the diversity present among African-ancestry populations (171,172,185); and this is further exacerbated by the bias of African sequencing resources towards African Americans and West Africans (97). Another constraint of the performance of PGS in African populations is the reliance on the HapMap3 variants SNP list by many PGS scoring methods, typically for gains in computational efficiency. Whilst this may not lead to substantial reduced performance in European populations, research has shown that the HapMap3 variants can be worse at capturing variation within African populations (200). Further research expanding the SNP lists used in polygenic scoring to better capture this variance should be explored

Given the limited sample sizes of continental African cohorts and the development of improved multi-ancestry computational approaches such as BridgePRS and PRS-CSx (61,62,200,346), it would be of interest to meta-analyse GWAS summary statistics from diverse ancestries. Specifically, including the AWI-Gen and APCDR cohorts to boost representation from African ancestry cohorts, which traditionally relies on largely African Americans. Predictive utility may be enhanced due the large sample sizes of European ancestry-derived contributing to a boost in power. An alternate approach may also involve including European-ancestry derived PGS for cardiometabolic traits as predictors in the elastic net models.

Clinician perception survey

The clinician perceptions work sheds light on an under-researched area of South African healthcare and contributes to a greater understanding of what will be required to successfully implement PM in South Africa, and potentially other resource-constrained environments. The study was novel in that it included a compulsory educational video before responding to the questionnaire to provide context and some exposure to precision medicine. The decision to include a video was two-fold, 1. to ensure a baseline understanding amongst participants; and 2. to aid recruitment; each described below:

1. Given that our target population included clinicians from a variety of medical specialties, with varying experience, and graduated from different institutions, the video aimed to ensure an awareness of baseline understanding of key concepts and definitions necessary for completing the questionnaire.
2. Recruiting clinicians during the COVID-19 pandemic was particularly challenging. Recruitment through mailing list survey invitations was extremely low. To address this, and with ethical approval, we connected with clinicians during their weekly compulsory academic meetings as this approach ensured that completing the survey did not encroach on their clinical time. Presenting at these meetings required incorporating a training element such as the video.

However, we are aware that the use of the video as well the use of online convenience sampling strategies there is potential for recruitment and responder bias which may have skewed results (245). Ways in which the influence of the video could be assessed include conducting

1. A control group comparison, a group of participants that do not watch the video. Responses could then be compared to those who viewed the video and any potential differences that may be attributed to the video content identified.

2. A pre- and post-video assessment could provide quantitative data on the effectiveness of the video in conveying the key messages and any potential bias it may introduce.
3. Focus groups or interviews could also be used to gather feedback into how the video's information was perceived and how it may have influenced participants' thinking and response strategies.
4. Statistical techniques, such as covariance analysis, could also be used to analyse questionnaire responses and identify inconsistencies that may point to the video's influence.

Closely related to this is that although the survey was built upon previously validated tools, the specific instrument used in this study was not evaluated for reliability and validity. Moreover, although the questionnaire was developed with substantial input from local clinical and field experts, was piloted with individuals similar to the target population, and the results obtained are intuitively sensible and align with literature trends, further evaluation of questionnaire validity should be explored. Additional methods that could be explored in future research include conducting interviews with a subset of the target population to ensure accurate question interpretation. Also, using reliability testing methods, such as test-retest reliability and Cronbach's Alpha, to ensure questionnaire consistency and dependability.

Assessment of the clinician perceptions of the integrated risk score was restricted to large, relatively well-funded urban medical centres in Johannesburg. Consequently, respondents in this study are unlikely to be representative of health professionals across South Africa, and other African countries.

Furthermore, restricting responses to specialists at tertiary hospitals resulted in no input from primary care physicians. Primary care provision is the foundation of care in South Africa, serving most communities, and where CVD risk stratification could have the greatest benefit [59,60,61]. Primary care physicians are best positioned to implement preventative strategies and inform and educate patients about lifestyle modification and new and innovative technologies. However, given that disparities in genetics training and resource availability can significantly impact clinicians' perceptions (218,219,221), as primary care practitioners may have less exposure to specialised training and research responsibilities as compared to those within tertiary

institutions, and limited access to advanced diagnostic tools and specialised knowledge, their comfort and familiarity with new approaches may be negatively impacted. Nonetheless, additional studies should be undertaken to understand the perceptions of these physicians as they may have a greater understanding of the practical challenges involved in implementing precision medicine on a larger scale in South Africa. Perceptions from other stakeholders including patients, policy makers, payors, and governments are also critical.

Chapter 4: OVERARCHING CONCLUSION

This research shows the value of integrating genetic and non-genetic factors to enhance CVD risk prediction in African populations. Ancestry-aligned integrated risk scores for CVD-associated diseases that include polygenic risk scores of cardiometabolic biomarkers outperform scores containing non-genetic factors alone. Ancestry is a substantial contributor to variance in CVD prediction models, and its effective modelling suggests a promising direction for model refinement. These integrated scores provide an opportunity for earlier, more accurate, and personalised risk assessment, leading to more effective prevention and treatment strategies among populations of African ancestry.

Complementing these findings, this body of work reveals that implementing precision medicine in Africa is a possibility but faces substantial challenges. The research underscores a broader narrative of the need for a shift in health and genetic research on the continent. This shift recognises that simply applying health and genomic insights derived from European and African diaspora populations may exacerbate health disparities and are insufficient proxies for continental Africans due to vast variations in genetics, exposures, behaviours, social determinants of health, and healthcare settings.

The integration of genomic research into healthcare requires a more inclusive, context-sensitive, and locally informed approach that acknowledges and leverages the unique genetic diversity, healthcare challenges, and socio-cultural contexts of African populations. Improving access to large, longitudinal continental African datasets is crucial for refining CVD prediction models, and precision medicine approaches for other diseases. There is also a need to consider the multiple health burdens in Africa and recognise that individuals may be afflicted with multiple conditions thus multiple and inter-connected treatment strategies might be needed.

Additional challenges include addressing the currently limited genetic services capacity and bridging gaps in genomics knowledge and awareness among clinicians and patients. As Africa develops its genetic capacity, with the simultaneous increase in African cohorts and decrease in genetic test costs, future CVD risk tools and precision

medicine in general would be enhanced and have the potential to reduce the heavy burden of CVDs across the continent.

REFERENCES

1. Mandl KD, Kohane IS. Escaping the EHR trap--the future of health IT. N Engl J Med. 2012 Jun;366(24):2240–2.

2. Zeggini E, Gloyn AL, Barton AC, Wain L V. Translational genomics and precision medicine: Moving from the lab to the clinic. Science (80-) [Internet]. 2019 Sep 27 [cited 2021 May 3];365(6460):1409–13. Available from: http://science.sciencemag.org/

3. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015 Feb;372(9):793–5.

4. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med 2019 218 [Internet]. 2019 Jan 15 [cited 2024 Apr 24];21(8):1708–18. Available from: https://www.nature.com/articles/s41436-018-0406-9

5. Lakeman IMM, Rodríguez-Girondo M, Lee A, Ruiter R, Stricker BH, Wijnant SRA, et al. Validation of the BOADICEA model and a 313-variant polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort. Genet Med [Internet]. 2020 Nov 1 [cited 2024 May 13];22(11):1803–11. Available from: https://pubmed.ncbi.nlm.nih.gov/32624571/

6. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med Off J Am Coll Med Genet. 2019 Aug;21(8):1708–18.

7. Chaterjee N, Shi J, Garcia-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet [Internet]. 2017;17(7):392–406. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6021129/pdf/nihms971746.pdf

8. Mendis S, Puska P, Norrving B. Global atlas on cardiovascular disease prevention and control. World Heal Organ. 2011;2–14.

9. Yeates K, Lohfeld L, Sleeth J, Morales F, Rajkotia T, Ogedegbe O. A global perspective on cardiovascular disease in vulnerable populations. Can J Cardiol. 2015;31(9):1081–93.

10. WHO. Global Status Report On Noncommunicable Diseases 2014 [Internet]. Geneva, Switzerland; 2014. Available from: http://apps.who.int/iris/%0Abitstream/10665/148114/1/97892_eng.pdf.

11. Roth GA, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1736–88.

12. Sliwa K, Acquah L, Gersh BJ, Mocumbi AO. Impact of socioeconomic status, ethnicity, and urbanization on risk factor profiles of cardiovascular disease in Africa. Circulation. 2016;133(12):1199–208.

13. Gómez-Olivé FX, Ali SA, Made F, Kyobutungi C, Nonterah E, Micklesfield L, et al. Regional and Sex Differences in the Prevalence and Awareness of Hypertension: An H3Africa AWI-Gen Study Across 6 Sites in Sub-Saharan Africa. Glob Heart. 2017;12(2):81–90.

14. Sivadasanpillai H, Leeder S, M H, Jeemon P, Dorairaj P. A race against time: The Challenge of Cardiovascular Diseases in Developing Economies. 2015.

15. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet [Internet]. 2018;19(9):581–90. Available from: http://dx.doi.org/10.1038/s41576-018-0018-x

16. Moorthie S, Babb de Villiers C, Brigden T, Gaynor L, Hall A, Johnson E, et al. Polygenic scores , risk and cardiovascular disease. Cambridge: PHG Foundation; 2019.

17. WHO. Cardiovascular diseases (CVDs) [Internet]. 2017 [cited 2020 Jan 28]. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

18. Pestana JAX, Steyn K, Leiman A, Hartzenberg GM. The direct and indirect costs of cardiovascular disease in South Africa in 1991. South African Med J. 1996;86(6):679–84.

19. Webster J, Crickmore C, Charlton K, Steyn K, Wentzel-Viljoen E, Naidoo P. South Africa's salt reduction strategy: Are we on track, and what lies ahead? South African Med J. 2017;107(1):20–1.

20. Stacey N, Tugendhaft A, Hofman K. Sugary beverage taxation in South Africa: Household expenditure, demand system elasticities, and policy implications. Prev Med (Baltim) [Internet]. 2017;105:S26–31. Available from: https://doi.org/10.1016/j.ypmed.2017.05.026

21. Collins DRJ, Tompson AC, Onakpoya IJ, Roberts N, Ward AM, Heneghan CJ. Global cardiovascular risk assessment in the primary prevention of cardiovascular disease in adults: Systematic review of systematic reviews. BMJ Open. 2017;7(3).

22. Boateng D, Agyemang C, Beune E, Meeks K, Smeeth L, Schulze MB, et al. Cardiovascular disease risk prediction in sub-Saharan African populations — Comparative analysis of risk algorithms in the RODAM study. Int J Cardiol [Internet]. 2018;254:310–5. Available from: https://doi.org/10.1016/j.ijcard.2017.11.082

23. Peer N, Lombard C, Steyn K, Gaziano T, Levitt N. Comparability of total cardiovascular disease risk estimates using laboratory and non-laboratory based assessments in urban-dwelling South Africans: The CRIBSA study.

South African Med J. 2014;104(10):691–6.

24. Seedat YK, Rayner BL, Veriava Y. South African hypertension practice guideline 2014. Cardiovasc J Afr. 2014;25(6):288–94.

25. Klug EQ, Raal FJ, Marais AD, Taskinen MR, Dalby AJ, Schamroth C, et al. South african dyslipidaemia guideline consensus statement. South African Med J. 2012 Mar;102(3):178–88.

26. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM. General cardiovascular risk profile for use in primary care: the Framingham heart study. Circulation. 2008;117(6):743–753.

27. Gaziano TA, Pandya A, Steyn K, Levitt N, Mollentze W, Joubert G, et al. Comparative assessment of absolute cardiovascular disease risk characterization from non-laboratory-based risk assessment in South African populations. BMC Med [Internet]. 2013;11(1):1. Available from: BMC Medicine

28. Damen JAAG, Hooft L, Schuit E, Debray TPA, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: Systematic review. BMJ. 2016;353.

29. Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. BMJ [Internet]. 2017 May 23;357:j2099. Available from: http://www.bmj.com/content/357/bmj.j2099.abstract

30. Wagner RG, Crowther NJ, Micklesfield LK, Boua PR, Nonterah EA, Mashinya F, et al. Estimating the burden of cardiovascular risk in community dwellers over 40 years old in South Africa, Kenya, Burkina Faso and Ghana. BMJ Glob Heal [Internet]. 2021 Jan 21 [cited 2022 May 10];6(1). Available from: https://pubmed.ncbi.nlm.nih.gov/33479017/

31. Katsanis N. The continuum of causality in human genetic disorders. Genome Biol. 2016;17:233.

32. Kapa S, Tester DJ, Salisbury BA, Harris-Kerr C, Pungliya MS, Alders M, et al. Genetic Testing for Long QT Syndrome - Distinguishing Pathogenic Mutations from Benign Variants. Circulation [Internet]. 2009 Nov 11 [cited 2024 Apr 26];120(18):1752. Available from: /pmc/articles/PMC3025752/

33. Jansweijer JA, Nieuwhof K, Russo F, Hoorntje ET, Jongbloed JDH, Lekanne Deprez RH, et al. Truncating titin mutations are associated with a mild and treatable form of dilated cardiomyopathy. Eur J Heart Fail [Internet]. 2017 Apr 1 [cited 2024 Apr 26];19(4):512–21. Available from: https://pubmed.ncbi.nlm.nih.gov/27813223/

34. Drobni ZD, Kolossvary M, Karady J, Jermendy AL, Tarnoki AD, Tarnoki DL, et al. Heritability of Coronary Artery Disease: Insights From a Classical Twin Study. Circ Cardiovasc Imaging [Internet]. 2022 Mar 1;15(3):e013348. Available from: https://doi.org/10.1161/CIRCIMAGING.121.013348

35. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. J Intern Med. 2002 Sep;252(3):247–54.

36. Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NLM, Jackson C, et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. Stroke. 2012 Dec;43(12):3161–7.

37. Preuss M, König IR, Thompson JR, Erdmann J, Absher D, Assimes TL, et al. Design of the Coronary ARtery DIsease Genome-Wide Replication And Meta-Analysis (CARDIoGRAM) Study: A Genome-wide association meta-analysis involving more than 22 000 cases and 60 000 controls. Circ Cardiovasc Genet. 2010 Oct;3(5):475–83.

38. Peden JF, Hopewell JC, Saleheen D, Chambers JC, Hager J, Soranzo N, et al. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet. 2011 Mar;43(4):339–46.

39. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356–69.

40. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. Nat Genet. 2022 Dec;54(12):1803–15.

41. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. Nature. 2016;536:41–7.

42. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet [Internet]. 2022;54(5):560–72. Available from: https://doi.org/10.1038/s41588-022-01058-3

43. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet. 2015;47(10):1121–30.

44. Tcheandjieu C, Zhu X, Hilliard AT, Clarke SL, Napolioni V, Ma S, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. Nat Med [Internet]. 2022;28(8):1679–92. Available from: https://doi.org/10.1038/s41591-022-01891-3

45. Barton NH, Etheridge AM, Véber A. The infinitesimal model: Definition, derivation, and implications. Theor Popul Biol. 2017 Dec;118:50–73.

46. Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine. Genome Med. 2017;9(1):9–11.

47. Choi SW, Shin T, Mak H, Reilly PFO. A guide to performing Polygenic Risk Score analyses. 2018;2:11–3.

48. Slatkin M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nat Rev Genet [Internet]. 2008;9(6):477–85. Available from: https://doi.org/10.1038/nrg2361

49. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. Am J Hum Genet [Internet]. 2015;97(4):576–92. Available from: https://www.sciencedirect.com/science/article/pii/S0002929715003651

50. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience [Internet]. 2019 Jul 1;8(7):giz082. Available from: https://doi.org/10.1093/gigascience/giz082

51. Koch S, Schmidtke J, Krawczak M, Caliebe A. Clinical utility of polygenic risk scores: a critical 2023 appraisal. J Community Genet [Internet]. 2023;14(5):471–87. Available from: https://doi.org/10.1007/s12687-023-00645-z

52. Pain O, Glanville KP, Hagenaars SP, Selzam S, Fürtjes AE, Gaspar HA, et al. Evaluation of polygenic prediction methodology within a reference-standardized framework. PLOS Genet [Internet]. 2021 May 4;17(5):e1009021. Available from: https://doi.org/10.1371/journal.pgen.1009021

53. Khera A V., Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet [Internet]. 2018;50(9):1219–24. Available from: http://dx.doi.org/10.1038/s41588-018-0183-z

54. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol [Internet]. 2018 Oct 16 [cited 2022 May 11];72(16):1883–93. Available from: https://pubmed.ncbi.nlm.nih.gov/30309464/

55. Riveros-Mckay F, Weale ME, Moore R, Selzam S, Krapohl E, Sivley RM, et al. Integrated Polygenic Tool Substantially Enhances Coronary Artery Disease Prediction. Circ Genomic Precis Med [Internet]. 2021 [cited 2022 May 11];14:192–200. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCGEN.120.003304

56. Tada H, Melander O, Louie JZ, Catanese JJ, Rowland CM, Devlin JJ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. 2015;1–7.

57. Weale ME, Riveros-Mckay F, Selzam S, Seth P, Moore R, Tarran WA, et al. Validation of an Integrated Risk Tool, Including Polygenic Risk Score, for Atherosclerotic Cardiovascular Disease in Multiple Ethnicities and Ancestries. Am J Cardiol. 2021 Jun 1;148:157–64.

58. Fatumo S, Choudhury A. African American genomes don’t capture Africa’s genetic diversity. Vol. 617, Nature. England; 2023. p. 35.

59. Fatumo S, Sathan D, Samtal C, Isewon I, Tamuhla T, Soremekun C, et al. Polygenic risk scores for disease risk prediction in Africa: current challenges and future directions. Genome Med [Internet]. 2023;15(1):87. Available from: https://doi.org/10.1186/s13073-023-01245-9

60. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet [Internet]. 2019;51(4):584–91. Available from: http://dx.doi.org/10.1038/s41588-019-0379-x

61. Ruan Y, Lin Y-F, Feng Y-CA, Chen C-Y, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. Nat Genet [Internet]. 2022;54(5):573–80. Available from: https://doi.org/10.1038/s41588-022-01054-7

62. Hoggart C, Choi SW, García-González J, Souaiaia T, Preuss M, O’Reilly P. BridgePRS : A powerful trans-ancestry Polygenic Risk Score method. bioRxiv : the preprint server for biology. United States; 2023.

63. Mavaddat N, Pharoah PDP, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of Breast Cancer Risk Based on Profiling With Common Genetic Variants. JNCI J Natl Cancer Inst [Internet]. 2015 May 1;107(5):djv036. Available from: https://doi.org/10.1093/jnci/djv036

64. Lewis ACF, Green RC, Vassy JL. Polygenic risk scores in the clinic: Translating risk into action. HGG Adv. 2021 Oct;2(4):100047.

65. Fuat A, Adlen E, Monane M, Coll R, Groves S, Little E, et al. A polygenic risk score added to a QRISK®2 cardiovascular disease risk calculator demonstrated robust clinical acceptance and clinical utility in the primary care setting. Eur J Prev Cardiol [Internet]. 2024 Apr 18 [cited 2024 Apr 24];31(6):716–22. Available from: https://dx.doi.org/10.1093/eurjpc/zwae004

66. Wong EB, Olivier S, Gunda R, Koole O, Surujdeen A, Gareta D, et al. Convergence of infectious and non-communicable disease epidemics in rural South Africa: a cross-sectional, population-based multimorbidity study. Lancet Glob Heal [Internet]. 2021 Jul 1;9(7):e967–76. Available from: https://doi.org/10.1016/S2214-109X(21)00176-5

67. Navickas R, Petric V-K, Feigl AB, Seychell M. Multimorbidity: What do we know? What should we do? J comorbidity [Internet]. 2016 Jan [cited 2023 May 10];6(1):4–11. Available from: https://pubmed.ncbi.nlm.nih.gov/29090166/

68. Nunes BP, Flores TR, Mielke GI, Thumé E, Facchini LA. Multimorbidity and mortality in older adults: A systematic review and meta-analysis. Arch Gerontol Geriatr [Internet]. 2016 Nov 1 [cited 2022 Oct 13];67:130–8. Available from: https://pubmed.ncbi.nlm.nih.gov/27500661/

69. Fortin M, Lapointe L, Hudon C, Vanasse A, Ntetu AL, Maltais D. Multimorbidity and quality of life in primary care: a systematic review. Health Qual Life Outcomes [Internet]. 2004 Sep 20 [cited 2022 Oct 13];2. Available from: https://pubmed.ncbi.nlm.nih.gov/15380021/

70. Nicholson A, Kuper H, Hemingway H. Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. Eur Heart J. 2006 Dec;27(23):2763–74.

71. Yu M, Zhang X, Lu F, Fang L. Depression and Risk for Diabetes: A Meta-Analysis. Can J diabetes. 2015 Aug;39(4):266–72.

72. Dickens C. Depression in people with coronary heart disease: prognostic significance and mechanisms. Curr Cardiol Rep. 2015 Oct;17(10):83.

73. Cahill LE, Pan A, Chiuve SE, Sun Q, Willett WC, Hu FB, et al. Fried-food consumption and risk of type 2 diabetes and coronary artery disease: a prospective study in 2 cohorts of US women and men. Am J Clin Nutr. 2014 Aug;100(2):667–75.

74. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports. 2015 Dec;25 Suppl 3:1–72.

75. Roberts RE, Deleger S, Strawbridge WJ, Kaplan GA. Prospective association between obesity and depression: evidence from the Alameda County Study. Int J Obes Relat Metab Disord J Int Assoc Study Obes. 2003 Apr;27(4):514–21.

76. Gan Y, Gong Y, Tong X, Sun H, Cong Y, Dong X, et al. Depression and the risk of coronary heart disease: a meta-analysis of prospective cohort studies. BMC Psychiatry. 2014 Dec;14:371.

77. Vancampfort D, Correll CU, Wampers M, Sienaert P, Mitchell AJ, De Herdt A, et al. Metabolic syndrome and metabolic abnormalities in patients with major depressive disorder: a meta-analysis of prevalences and moderating variables. Psychol Med. 2014 Jul;44(10):2017–28.

78. Scherrer JF, Xian H, Bucholz KK, Eisen SA, Lyons MJ, Goldberg J, et al. A twin study of depression symptoms, hypertension, and heart disease in middle-aged men. Psychosom Med. 2003;65(4):548–57.

79. Kan C, Pedersen NL, Christensen K, Bornstein SR, Licinio J, MacCabe JH, et al. Genetic overlap between type 2 diabetes and depression in Swedish and Danish twin registries. Mol Psychiatry. 2016 Jul;21(7):903–9.

80. Baltramonaityte V, Pingault J-B, Cecil CAM, Choudhary P, Järvelin M-R, Penninx BWJH, et al. A multivariate genome-wide association study of psycho-cardiometabolic multimorbidity. PLoS Genet. 2023 Jun;19(6):e1010508.

81. Sankoh O, Byass P. The INDEPTH network: Filling vital gaps in global epidemiology. Int J Epidemiol. 2012;41(3):579–88.

82. 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 Heal Epidemiol Genomics. 2016;1.

83. Ali SA, Soo C, Agongo G, Alberts M, Amenga-etego L, Boua, Romuald P. Choudhury A, 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;11(2). Available from: https://doi.org/10.1080/16549716.2018.1507133

84. Beguy D, Elung'ata P, Mberu B, Oduor C, Wamukoya M, Nganyi B, et al. Health & Demographic Surveillance System Profile: The Nairobi Urban Health and Demographic Surveillance System (NUHDSS). Int J Epidemiol. 2015;44(2):462–71.

85. Derra K, Rouamba E, Kazienga A, Ouedraogo S, Tahita MC, Sorgho H, et al. Profile: Nanoro health and demographic surveillance system. Int J Epidemiol. 2012;41(5):1293–301.

86. Oduro AR, Wak G, Azongo D, Debpuur C, Wontuo P, Kondayire F, et al. Profile of the Navrongo health and demographic surveillance system. Int J Epidemiol. 2012;41(4):968–76.

87. Kahn K, Collinson MA, Xavier Gómez-olivé F, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt health and socio-demographic surveillance system. Int J Epidemiol. 2012;41(4):988–1001.

88. Alberts M, Dikotope SA, Choma SR, Masemola ML, Modjadji SEP, Mashinya F, et al. Health & Demographic Surveillance System Profile: The Dikgale Health and Demographic Surveillance System. Int J Epidemiol. 2015;44(5):1565–71.

89. Richter L, Norris S, Pettifor J, Yach D, Cameron N. Europe PMC Funders Group Cohort Profile : Mandela ' s children : The 1990 birth to twenty study in South Africa. Int J Epidemiol. 2007;36(3):504–11.

90. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation [Internet]. 2019 Sep 10 [cited 2022 May 10];140(11):e596–646. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000000678

91. Pereira L, Mutesa L, Tindana P, Ramsay M. African genetic diversity and adaptation inform a precision medicine agenda. Nat Rev Genet [Internet]. 2021 May 1 [cited 2021 May 3];22(5):284–306. Available from: http://www.ncbi.nlm.nih.gov/pubmed/33432191

92. Asiki G, Murphy G, Nakiyingi-Miiro J, Seeley J, Nsubuga RN, Karabarinde A, et al. The general population cohort in rural south-western Uganda: a platform for communicable and non-communicable disease studies. Int J Epidemiol. 2013 Feb;42(1):129–41.

93. Gurdasani D, Carstensen T, Fatumo S, Chen G, Franklin CS, Prado-Martinez J, et al. Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. Cell. 2019;179(4).

94. Hird TR, Young EH, Pirie FJ, Riha J, Esterhuizen TM, O'Leary B, et al. Study profile: the Durban Diabetes Study (DDS): a platform for chronic disease research. Glob Heal Epidemiol genomics. 2016;1:e2.

95. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023 Jan;51(D1):D977–85.

96. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw [Internet]. 2010 Feb 2;33(1 SE-Articles):1–22. Available from: https://www.jstatsoft.org/index.php/jss/article/view/v033i01

97. 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526(7571):68.

98. Choudhury A, Brandenburg J-T, Chikowore T, Sengupta D, Boua PR, Crowther NJ, et al. Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits. Nat Commun. 2022 May;13(1):2578.

99. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. J R Stat Soc Ser B Stat Methodol [Internet]. 2005 Apr 1;67(2):301–20. Available from: https://doi.org/10.1111/j.1467-9868.2005.00503.x

100. Pain O, Glanville KP, Hagenaars S, Selzam S, Fürtjes A, Coleman JRI, et al. Imputed gene expression risk scores: a functionally informed component of polygenic risk. Hum Mol Genet. 2021 May;30(8):727–38.

101. Steiger JH. Tests for comparing elements of a correlation matrix. Psychol Bull. 1980;87(2):245–251.

102. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

103. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.r-project.org/

104. Naumova EN. Precision public health: is it all about the data? J Public Health Policy [Internet]. 2022 Dec 1 [cited 2024 Apr 24];43(4):481–6. Available from: https://link.springer.com/article/10.1057/s41271-022-00367-5

105. Ndiaye Diallo R, Gadji M, Hennig BJ, Guèye M V., Gaye A, Diop JPD, et al. Strengthening human genetics research in Africa: report of the 9th meeting of the African Society of Human Genetics in Dakar in May 2016. Glob Heal Epidemiol Genom [Internet]. 2017 [cited 2021 May 3];2:1–4. Available from: https://pubmed.ncbi.nlm.nih.gov/29868221/

106. El-Kamah GY, Mohamed AM, Gad YZ, Abdelhak S, Hennig BJ, Ramesar RS, et al. Developing a road map to spread genomic knowledge in Africa: 10th Conference of the African Society of Human Genetics, Cairo, Egypt. Am J Trop Med Hyg [Internet]. 2020 [cited 2021 May 3];102(4):719–23. Available from: https://pubmed.ncbi.nlm.nih.gov/32124726/

107. Musanabaganwa C, Mihigo B, Tumusime R, Uwanyirigira M, Da Rocha J, Hayat M, et al. Building skills and resources for genomics, epigenetics, and bioinformatics research for Africa: Report of the Joint 11th Conference of the African Society of Human Genetics and 12th H3 Africa consortium, 2018. Am J Trop Med Hyg [Internet]. 2020 Jun 1 [cited 2021 May 3];102(6):1417–24. Available from: https://pubmed.ncbi.nlm.nih.gov/32207403/

108. Rotimi C, Abayomi A, Abimiku A, Adabayeri VM, Adebamowo C, Adebiyi E, et al. Research capacity. Enabling the genomic revolution in Africa. Science (80-) [Internet]. 2014 [cited 2021 May 3];344:1346–8. Available from: https://pubmed.ncbi.nlm.nih.gov/24948725/

109. Tekola-Ayele F, Rotimi CN. Translational Genomics in Low- and Middle-Income Countries: Opportunities and Challenges. Public Health Genomics [Internet]. 2015 Jul 25 [cited 2021 May 3];18(4):242–7. Available from: https://www.karger.com/Article/FullText/433518

110. Mulder N, Abimiku A, Adebamowo SN, de Vries J, Matimba A, Olowoyo P, et al. H3Africa: Current perspectives. Pharmgenomics Pers Med [Internet]. 2018 Apr 10 [cited 2021 May 3];11:59–66. Available from: https://pubmed.ncbi.nlm.nih.gov/29692621/

111. Wonkam A, de Vries J. Returning incidental findings in African genomics research. Nat Genet. 2020 Jan 1;52(1):17–20.

112. Mitropoulos K, Al Jaibeji H, Forero DA, Laissue P, Wonkam A, Lopez-Correa C, et al. Success stories in genomic medicine from resource-limited countries. Hum Genomics [Internet]. 2015 Dec 18 [cited 2021 May 3];9(11):1–7. Available from: http://humgenomics.biomedcentral.com/articles/10.1186/s40246-015-0033-3

113. Williams TN, Mwangi TW, Wambua S, Peto TEA, Weatherall DJ, Gupta S, et al. Negative epistasis between the malaria-protective effects of α+-thalassemia and the sickle cell trait. Nat Genet [Internet]. 2005 Nov [cited 2021 May 3];37(11):1253–7. Available from: https://pubmed.ncbi.nlm.nih.gov/16227994/

114. Goldman A, Jenkins T, Krause A. Molecular evidence that Fragile X syndrome occurs in the South African black population. J Med Genet. 1998;35(10):878–9.

115. Essop FB, Krause A. Diagnostic, carrier and prenatal genetic testing for Fragile X syndrome and other FMR-1-related disorders in Johannesburg, South Africa: A 20-year review. S Afr Med J [Internet]. 2013 Oct 11 [cited 2021 May 3];103(12):994–8. Available from: http://www.samj.org.za/index.php/samj/article/view/7144

116. Lubala TK, Lumaka A, Mbuyi-Musanzayi S, Kayembe T, Shongo MYP, Mukuku O, et al. Fragile X syndrome with mosaic size mutation in a Bantu patient from Central Africa. Clin Dysmorphol. 2018;27(2):66–9.

117. Kengne Kamga K, Nguefack S, Minka K, Wonkam Tingang E, Esterhuizen A, Nchangwi Munung S, et al. Cascade testing for Fragile X syndrome in a rural setting in Cameroon (Sub-Saharan Africa). Genes (Basel) [Internet]. 2020 Jan 28 [cited 2021 May 3];11(2):136. Available from: https://www.mdpi.com/2073-4425/11/2/136

118. Kabahuma RI, Ouyang X, Du LL, Yan D, Hutchin T, Ramsay M, et al. Absence of GJB2 gene mutations, the GJB6 deletion (GJB6-D13S1830) and four common mitochondrial mutations in nonsyndromic genetic hearing loss in a South African population. Int J Pediatr Otorhinolaryngol [Internet]. 2011 May [cited 2021 May 3];75(5):611–7. Available from: https://pubmed.ncbi.nlm.nih.gov/21392827/

119. Wonkam A, Bosch J, Noubiap JJ, Lebeko K, Makubalo N, Dandara C. No evidence for clinical utility in investigating the Connexin genes GJB2, GJB6 and GJA1 in non-syndromic hearing loss in black Africans. S Afr Med J [Internet]. 2015 Nov 26 [cited 2021 May 3];105(1):23–6. Available from: http://dx.doi.org/10.1093/eurpub/ckr176]

120. Lebeko K, Sloan-Heggen CM, Noubiap JJN, Dandara C, Kolbe DL, Ephraim SS, et al. Targeted genomic enrichment and massively parallel sequencing identifies novel nonsyndromic hearing impairment pathogenic variants in Cameroonian families. Clin Genet [Internet]. 2016 Sep 1 [cited 2021 May 3];90(3):288–90. Available from: https://pubmed.ncbi.nlm.nih.gov/27246798/

121. Adadey SM, Manyisa N, Mnika K, de Kock C, Nembaware V, Quaye O, et al. GJB2 and GJB6 Mutations in Non-Syndromic Childhood Hearing Impairment in Ghana. Front Genet [Internet]. 2019 Sep 18 [cited 2021 May 3];10:1–8. Available from: https://www.frontiersin.org/article/10.3389/fgene.2019.00841/full

122. Arab S Ben, Masmoudi S, Beltaief N, Hachicha S, Ayadi H. Consanguinity and Endogamy in Northern Tunisia and its Impact on Non-syndromic Deafness. 2004 [cited 2021 Jun 21]; Available from: www.interscience.wiley.com

123. Baine FK, Krause A, Greenberg LJ. The Frequency of Huntington Disease and Huntington Disease-Like 2 in the South African Population. Neuroepidemiology [Internet]. 2016 Apr 1 [cited 2021 May 3];46(3):198–202. Available from: https://pubmed.ncbi.nlm.nih.gov/26882115/

124. Krause A, Mitchell C, Essop F, Tager S, Temlett J, Stevanin G, et al.

Junctophilin 3 (JPH3) expansion mutations causing Huntington disease like 2 (HDL2) are common in South African patients with African ancestry and a Huntington disease phenotype. Am J Med Genet B Neuropsychiatr Genet [Internet]. 2015 Oct 1 [cited 2021 May 3];168(7):573–85. Available from: https://pubmed.ncbi.nlm.nih.gov/26079385/

125. Rebbeck TR, Friebel TM, Friedman E, Hamann U, Huo D, Kwong A, et al. Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations. Hum Mutat [Internet]. 2018 May 1 [cited 2021 May 3];39(5):593–620. Available from: https://pubmed.ncbi.nlm.nih.gov/29446198/

126. Zheng Y, Walsh T, Gulsuner S, Casadei S, Lee MK, Ogundiran TO, et al. Inherited breast cancer in Nigerian women. J Clin Oncol [Internet]. 2018 Oct 1 [cited 2021 May 3];36(28):2820–5. Available from: https://pubmed.ncbi.nlm.nih.gov/30130155/

127. Krause A, Seymour H, Ramsay M. Common and founder mutations for monogenic traits in sub-saharan African populations. Annu Rev Genomics Hum Genet. 2018;19:149–75.

128. Romdhane L, Mezzi N, Hamdi Y, El-Kamah G, Barakat A, Abdelhak S. Consanguinity and Inbreeding in Health and Disease in North African Populations. Annu Rev Genomics Hum Genet [Internet]. 2019 Aug 31 [cited 2021 May 3];20(1):155–79. Available from: https://www.annualreviews.org/doi/10.1146/annurev-genom-083118-014954

129. Sankoh O, Ijsselmuiden C. Sharing research data to improve public health: A perspective from the global south. Lancet [Internet]. 2011 [cited 2021 May 3];378(9789):401–2. Available from: https://pubmed.ncbi.nlm.nih.gov/21803205/

130. Popejoy AB. Diversity in precision medicine and pharmacogenetics: Methodological and conceptual considerations for broadening participation. Pharmgenomics Pers Med [Internet]. 2019 [cited 2021 May 3];12:257–71. Available from: https://pubmed.ncbi.nlm.nih.gov/31686892/

131. Kraft SA, Cho MK, Gillespie K, Halley M, Varsava N, Ormond KE, et al. Beyond Consent: Building Trusting Relationships With Diverse Populations in Precision Medicine Research. Am J Bioeth [Internet]. 2018 Apr 3 [cited 2021 May 11];18(4):3–20. Available from: https://pubmed.ncbi.nlm.nih.gov/29621457/

132. Sabatello M, Callier S, Garrison NA, Cohn EG. Trust, Precision Medicine Research, and Equitable Participation of Underserved Populations. Am J Bioeth [Internet]. 2018 Apr 3 [cited 2021 May 3];18(4):34–6. Available from: /pmc/articles/PMC5890957/

133. Duncan L, Shen H, Gelaye B, Meijsen J, Ressler K, Feldman M, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun [Internet]. 2019 Dec 1 [cited 2021 May 3];10(1):1–9. Available from: https://doi.org/10.1038/s41467-019-11112-0

134. Kromberg JG, Sizer EB, Christianson AL. Genetic services and testing in South Africa. J Community Genet. 2013 Jul;4(3):413–23.

135. Kromberg JGR, Krause A. Human genetics in Johannesburg, South Africa: Past, present and future. S Afr Med J [Internet]. 2013 Oct 11 [cited 2021 May 3];103(12):957–61. Available from: http://www.samj.org.za/index.php/samj/article/view/7220

136. Temtamy SA, Hussen DF. Genetics and Genomic Medicine in Egypt: steady pace. Mol Genet Genomic Med [Internet]. 2017 Jan 1 [cited 2021 May 3];5(1):8–14. Available from: /pmc/articles/PMC5241207/

137. Belhassan K, Ouldim K, Sefiani AA. Genetics and Genomic Medicine in Morocco: The present hope can make the future bright. Mol Genet Genomic Med [Internet]. 2016 Nov 1 [cited 2021 Jun 21];4(6):588–98. Available from: /pmc/articles/PMC5118203/

138. Elloumi-Zghal H, Chaabouni Bouhamed H. Genetics and Genomic Medicine in Tunisia. Mol Genet Genomic Med. 2018;6(2):134–59.

139. Chaabouni-Bouhamed H. Country Report Tunisia: Communities and Community Genetics. Community Genet [Internet]. 2008;11:313–23. Available from: www.karger.com

140. Adeyemo AA, Amodu OK, Ekure EE, Omotade OO. Medical Genetics and Genomic Medicine in Nigeria. Mol Genet Genomic Med [Internet]. 2018 May 1 [cited 2021 Jun 21];6(3):314–21. Available from: https://commons.wikimedia.org/wiki/

141. Uwineza A, Mutesa L. Medical Genetics and Genomic Medicine in Rwanda. Mol Genet Genomic Med [Internet]. 2015 Nov 1 [cited 2021 Jun 21];3(6):486–9. Available from: /pmc/articles/PMC4694135/

142. Malherbe HL, Christianson AL, Woods D, Aldous C. The case for the genetic nurse in South Africa. J Community Genet [Internet]. 2017 Apr 1 [cited 2021 May 3];8(2):65–73. Available from: /pmc/articles/PMC5386919/

143. Kromberg J, Jenkins T. Cultural Influences on the Perception of Genetic Disorders in the Black Population of Southern Africa. In: Clarke A PE, editor. Culture, Kinship and Genes [Internet]. London: Macmillan Press; 1997 [cited 2021 May 3]. p. 147–57. Available from: https://link.springer.com/chapter/10.1007/978-1-349-25882-6_11

144. Wonkam A, Njamnshi AK, Angwafo FF. Knowledge and attitudes concerning medical genetics amongst physicians and medical students in Cameroon (sub-Saharan Africa). Genet Med. 2006 Jun;8(6):331–8.

145. Zhong A, Darren B, Loiseau B, He LQB, Chang T, Hill J, et al. Ethical, social, and cultural issues related to clinical genetic testing and counseling in low- and middle-income countries: a systematic review. Genet Med [Internet]. 2018 [cited 2021 May 3];0:1–11. Available from:

https://pubmed.ncbi.nlm.nih.gov/30072741/

146. Munung NS, Mayosi BM, de Vries J. Genomics research in Africa and its impact on global health: insights from African researchers. Glob Heal Epidemiol Genom [Internet]. 2018 [cited 2021 May 3];3:1–8. Available from: /pmc/articles/PMC6152488/

147. Abacan MA, Alsubaie L, Barlow-Stewart K, Caanen B, Cordier C, Courtney E, et al. The Global State of the Genetic Counseling Profession. Eur J Hum Genet [Internet]. 2019 Feb 1 [cited 2021 May 3];27(2):183–97. Available from: https://pubmed.ncbi.nlm.nih.gov/30291341/

148. Kromberg J. Genetic counseling and albinism. In: Kromberg J, Manga P, editors. Albinism in Africa Historical, geographical, medical, genetic and psychosocial aspects. London: Elsevier Academic Press; 2018. p. 203–33.

149. Gillham A, Greyling B, Wessels TM, Mbele B, Schwyzer R, Krause A, et al. Uptake of Genetic Counseling, Knowledge of Bleeding risks and Psychosocial Impact in a South African Cohort of Female Relatives of People with Hemophilia. J Genet Couns [Internet]. 2015 Dec 1 [cited 2021 May 3];24(6):978–86. Available from: https://pubmed.ncbi.nlm.nih.gov/25828422/

150. Nembaware V, Mulder N. The African Genomic Medicine Training Initiative (AGMT): Showcasing a Community and Framework Driven Genomic Medicine Training for Nurses in Africa. Front Genet [Internet]. 2019 Dec 20 [cited 2021 May 3];10:1–14. Available from: https://www.frontiersin.org/article/10.3389/fgene.2019.01209/full

151. Mboowa G, Sserwadda I, Aruhomukama D. Genomics and bioinformatics capacity in Africa: no continent is left behind. Genome [Internet]. 2021 Jan 12 [cited 2021 May 3];64:1–11. Available from: https://pubmed.ncbi.nlm.nih.gov/33433259/

152. Manolio TA, Abramowicz M, Al-Mulla F, Anderson W, Balling R, Berger AC, et al. Global implementation of genomic medicine: We are not alone. Sci Transl Med [Internet]. 2015 Jun 3 [cited 2021 May 11];7(290):1–7. Available from: https://pubmed.ncbi.nlm.nih.gov/26041702/

153. Alwan A, Modell B. Recommendations for introducing genetics services in developing countries. Nat Rev Genet [Internet]. 2003 Jan 1 [cited 2021 May 3];4(1):61–8. Available from: https://pubmed.ncbi.nlm.nih.gov/12509754/

154. Bornstein MH, Hendricks C. Screening for developmental disabilities in developing countries. Soc Sci Med. 2013 Nov;97:307–15.

155. AESA. A Framework for the Implementation of Genomic Medicine for Public Health in Africa [Internet]. Nairobi; 2020 [cited 2021 May 3]. Available from: www.aasciences.africa

156. Green ED, Gunter C, Biesecker LG, Di Francesco V, Easter CL, Feingold EA, et al. Strategic vision for improving human health at The Forefront of

Genomics. Nature [Internet]. 2020 Oct 29 [cited 2021 May 3];586(7831):683–92. Available from: https://doi.org/10.1038/s41586-020-2817-4

157. Oleribe OE, Momoh J, Uzochukwu BS, Mbofana F, Adebiyi A, Barbera T, et al. Identifying key challenges facing healthcare systems in Africa and potential solutions. Int J Gen Med [Internet]. 2019 [cited 2021 May 3];12:395–403. Available from: https://www.dovepress.com/identifying-key-challenges-facing-healthcare-systems-in-africa-and-pot-peer-reviewed-article-IJGM

158. Adebamowo SN, Francis V, Tambo E, Diallo SH, Landouré G, Nembaware V, et al. Implementation of genomics research in Africa: challenges and recommendations. Glob Health Action [Internet]. 2018 Jan 1 [cited 2021 May 3];11(1):1–6. Available from: https://doi.org/10.1080/16549716.2017.1419033

159. Wonkam A, Mayosi BM. Genomic medicine in Africa: Promise, problems and prospects. Genome Med. 2014 Feb 24;6(2):1–3.

160. Christianson A, Zimmern R, Kristoffersson U, Schmidtke J, Kent A, Raouf R, et al. Health needs assessment for medical genetic services for congenital disorders in middle- and low-income nations. J Community Genet [Internet]. 2013 Jul [cited 2021 May 3];4(3):297–308. Available from: /pmc/articles/PMC3739852/

161. Mulder NJ, Adebiyi E, Adebiyi M, Adeyemi S, Ahmed A, Ahmed R, et al. Development of Bioinformatics Infrastructure for Genomics Research. Glob Heart [Internet]. 2017 Jun 1 [cited 2021 May 3];12(2):91–8. Available from: https://pubmed.ncbi.nlm.nih.gov/28302555/

162. Knoppers BM, Sénécal K, Borry P, Avard D. Whole-genome sequencing in newborn screening programs. Sci Transl Med [Internet]. 2014 Mar 26 [cited 2021 May 3];6(229):229cm2. Available from: https://pubmed.ncbi.nlm.nih.gov/24670681/

163. Tessema SK, Inzaule SC, Christoffels A, Kebede Y, de Oliveira T, Ouma AEO, et al. Accelerating genomics-based surveillance for COVID-19 response in Africa. Lancet Microbe [Internet]. 2020 Oct [cited 2021 May 3];1(6):e227–8. Available from: https://pubmed.ncbi.nlm.nih.gov/32838350/

164. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. Nat Med [Internet]. 2021 Apr 1 [cited 2021 May 3];27(4):622–5. Available from: https://doi.org/10.1038/s41591-021-01285-x

165. Tegally H, Wilkinson E, Lessells RJ, Giandhari J, Pillay S, Msomi N, et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. Nat Med [Internet]. 2021 Mar 1 [cited 2021 May 3];27(3):440–6. Available from: https://doi.org/10.1038/s41591-021-01255-3

166. Gurwitz KT, Aron S, Panji S, Maslamoney S, Fernandes PL, Judge DP, et al. Designing a course model for distance-based online bioinformatics training in Africa: The H3ABioNet experience. Ouellette F, editor. PLOS Comput Biol

[Internet]. 2017 [cited 2021 May 3];13(10):1–11. Available from: https://dx.plos.org/10.1371/journal.pcbi.1005715

167. Bombard Y, Hayeems RZ. How digital tools can advance quality and equity in genomic medicine. Nat Rev Genet [Internet]. 2020 Sep 1 [cited 2021 May 3];21(9):505–6. Available from: www.genomicsadviser.com

168. Cohen SA, Huziak RC, Gustafson S, Grubs RE. Analysis of Advantages, Limitations, and Barriers of Genetic Counseling Service Delivery Models. J Genet Couns [Internet]. 2016 Oct 1 [cited 2021 May 12];25(5):1010–8. Available from: https://onlinelibrary.wiley.com/doi/full/10.1007/s10897-016-9932-2

169. Meier F, Schöffski O, Schmidtke J. Public-private partnership as a solution for integrating genetic services into health care of countries with low and middle incomes. J Community Genet [Internet]. 2013 Jul [cited 2021 May 3];4(3):309–20. Available from: /pmc/articles/PMC3739850/

170. Stark Z, Dolman L, Manolio TA, Ozenberger B, Hill SL, Caulfied MJ, et al. Integrating Genomics into Healthcare: A Global Responsibility. Am J Hum Genet [Internet]. 2019 Jan 3 [cited 2021 May 3];104(1):13–20. Available from: https://doi.org/10.1016/j.ajhg.2018.11.014.

171. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. Nature. 2015 Jan 15;517(7534):327–32.

172. Choudhury A, Ramsay M, Hazelhurst S, Aron S, Bardien S, Botha G, et al. Whole-genome sequencing for an enhanced understanding of genetic variation among South Africans. Nat Commun [Internet]. 2017 Dec 1 [cited 2021 May 3];8(1):1–12. Available from: http://www.simplemappr.net/

173. Wohlers I, Künstner A, Munz M, Olbrich M, Fähnrich A, Calonga-Solís V, et al. An integrated personal and population-based Egyptian genome reference. Nat Commun [Internet]. 2020 Dec 1 [cited 2021 May 3];11(1):1–10. Available from: https://doi.org/10.1038/s41467-020-17964-1

174. Wonkam A. Sequence three million genomes across Africa. Nature [Internet]. 2021 Feb 1 [cited 2021 May 3];590(7845):209–11. Available from: https://pubmed.ncbi.nlm.nih.gov/33568829/

175. Muzoriana N, Gavi S, Nembaware V, Dhoro M, Matimba A. Knowledge, attitude, and perceptions of pharmacists and pharmacy students towards pharmacogenomics in Zimbabwe. Pharmacy [Internet]. 2017 Jun 30 [cited 2021 May 3];5(4):36–46. Available from: https://pubmed.ncbi.nlm.nih.gov/28970448/

176. Solomon G, Greenberg J, Futter M, Vivian L, Penn C. Understanding of genetic inheritance among Xhosa-speaking caretakers of children with hemophilia. J Genet Couns [Internet]. 2012 Oct [cited 2021 May 3];21(5):726–40. Available from: https://pubmed.ncbi.nlm.nih.gov/22407306/

177. Mazibuko NG, Kromberg J. A personal perspective: living with albinism. In: Kromberg, J. Manga P, editor. Albinism in Africa Historical, geographic, medical genetic and psychosocial aspects. London: Elsevier Academic Press.; 2018. p. 295–307.

178. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016 Oct 12;538(7624):161–4.

179. Dragojlovic N, Elliott AM, Adam S, van Karnebeek C, Lehman A, Mwenifumbo JC, et al. The cost and diagnostic yield of exome sequencing for children with suspected genetic disorders: a benchmarking study. Genet Med [Internet]. 2018 Sep 1 [cited 2021 May 12];20(9):1013–21. Available from: https://www.nature.com/articles/gim2017226

180. Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikin JE, et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. Sci Transl Med [Internet]. 2014 Dec 3 [cited 2021 May 12];6(265):1–14. Available from: https://stm.sciencemag.org/content/6/265/265ra168

181. Tsiplova K, Zur RM, Marshall CR, Stavropoulos DJ, Pereira SL, Merico D, et al. A microcosting and cost-consequence analysis of clinical genomic testing strategies in autism spectrum disorder. Genet Med [Internet]. 2017 Nov 1 [cited 2021 May 12];19(11):1268–75. Available from: https://www.nature.com/articles/gim201747

182. Stark Z, Schofield D, Alam K, Wilson W, Mupfeki N, Macciocca I, et al. Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement. Genet Med [Internet]. 2017 Aug 1 [cited 2021 May 12];19(8):867–74. Available from: http://www.xe.com

183. Chrystoja CC, Diamandis EP. Whole genome sequencing as a diagnostic test: Challenges and opportunities. Clin Chem [Internet]. 2014 May 1 [cited 2021 May 12];60(5):724–33. Available from: https://academic.oup.com/clinchem/article/60/5/724/5621606

184. Pollard S, Weymann D, Dunne J, Mayanloo F, Buckell J, Buchanan J, et al. Toward the diagnosis of rare childhood genetic diseases: what do parents value most? Eur J Hum Genet [Internet]. 2021 Apr 26 [cited 2021 May 12];1–11. Available from: http://www.nature.com/articles/s41431-021-00882-1

185. Choudhury A, Aron S, Botigué LR, Sengupta D, Botha G, Bensellak T, et al. High-depth African genomes inform human migration and health. Nature [Internet]. 2020 Oct 1 [cited 2021 May 11];586(7831):741–8. Available from: https://doi.org/10.1038/s41586-020-2859-7

186. Goff DCJ, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014 Jun;129(25 Suppl 2):S49-73.

187. Kullo IJ, Lewis CM, Inouye M, Martin AR, Ripatti S, Chatterjee N. Polygenic scores in biomedical research. Nat Rev Genet. 2022 Sep;23(9):524–32.

188. Kachuri L, Chatterjee N, Hirbo J, Schaid DJ, Martin I, Kullo IJ, et al. Principles and methods for transferring polygenic risk scores across global populations. Nat Rev Genet. 2023 Aug;

189. Aragam KG, Dobbyn A, Judy R, Chaffin M, Chaudhary K, Hindy G, et al. Limitations of Contemporary Guidelines for Managing Patients at High Genetic Risk of Coronary Artery Disease. J Am Coll Cardiol. 2020 Jun;75(22):2769–80.

190. Rotimi CN, Chen G, Adeyemo AA, Furbert-Harris P, Guass D, Zhou J, et al. A Genome-Wide Search for Type 2 Diabetes Susceptibility Genes in West Africans : The Africa America Diabetes Mellitus (AADM) Study. Diabetes [Internet]. 2004 Mar 1;53(3):838–41. Available from: https://doi.org/10.2337/diabetes.53.3.838

191. Han B, Eskin E. Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. Am J Hum Genet [Internet]. 2011;88(5):586–98. Available from: https://www.sciencedirect.com/science/article/pii/S0002929711001558

192. Illumina. San Diego, USA;

193. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience [Internet]. 2015 Dec 1;4(1):s13742-015-0047–8. Available from: https://doi.org/10.1186/s13742-015-0047-8

194. Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna K V, et al. Long-range LD can confound genome scans in admixed populations. Vol. 83, American journal of human genetics. United States; 2008. p. 132–9.

195. Ramsay M, Crowther NJ, Agongo G, Ali SA, Asiki G, Boua RP, et al. Regional and sex-specific variation in BMI distribution in four sub-Saharan African countries: The H3Africa AWI-Gen study. Glob Health Action. 2018;11(sup2):1556561.

196. George JA, Brandenburg JT, Fabian J, Crowther NJ, Agongo G, Alberts M, et al. Kidney damage and associated risk factors in rural and urban sub-Saharan Africa (AWI-Gen): a cross-sectional population study. Lancet Glob Heal [Internet]. 2019;7(12):e1632–43. Available from: http://dx.doi.org/10.1016/S2214-109X(19)30443-7

197. Krapohl E, Patel H, Newhouse S, Curtis CJ, von Stumm S, Dale PS, et al. Multi-polygenic score approach to trait prediction. Mol Psychiatry. 2018 May;23(5):1368–74.

198. Naret O, Kutalik Z, Hodel F, Xu ZM, Marques-Vidal P, Fellay J. Improving polygenic prediction with genetically inferred ancestry. Hum Genet Genomics Adv. 2022;3(3).

199. Ekoru K, Adeyemo AA, Chen G, Doumatey AP, Zhou J, Bentley AR, et al. Genetic risk scores for cardiometabolic traits in sub-Saharan African populations. Int J Epidemiol. 2021 Aug;50(4):1283–96.

200. Privé F, Aschard H, Carmi S, Folkersen L, Hoggart C, O'Reilly PF, et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. Am J Hum Genet. 2022 Jan;109(1):12–23.

201. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, Benner C, Aguirre M, et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. Nat Genet. 2021 Feb;53(2):185–94.

202. Kamiza AB, Toure SM, Vujkovic M, Machipisa T, Soremekun OS, Kintu C, et al. Transferability of genetic risk scores in African populations. Nat Med [Internet]. 2022;28(6):1163–6. Available from: https://doi.org/10.1038/s41591-022-01835-x

203. Elliott J, Bodinier B, Bond TA, Marc Chadeau-Hyam, Evangelos Evangelou KGMM, Dehghan A, Muller DC, et al. Predictive Accuracy of a Polygenic Risk Score–Enhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. JAMA. 2020;323(7):636–45.

204. Pain O, Gillett AC, Austin JC, Folkersen L, Lewis CM. A tool for translating polygenic scores onto the absolute scale using summary statistics. Eur J Hum Genet [Internet]. 2022;30(3):339–48. Available from: https://doi.org/10.1038/s41431-021-01028-z

205. Kamp M, Krause A, Ramsay M. Has translational genomics come of age in Africa? Hum Mol Genet [Internet]. 2021 Oct 1 [cited 2021 Oct 7];30(R2):R164–73. Available from: https://academic.oup.com/hmg/article/30/R2/R164/6316672

206. Chikowore T, Kamiza AB, Oduaran OH, Machipisa T, Fatumo S. Non-communicable diseases pandemic and precision medicine: Is Africa ready? EBioMedicine. 2021 Mar 1;65:103260.

207. Chen C-Y, Han J, Hunter DJ, Kraft P, Price AL. Explicit Modeling of Ancestry Improves Polygenic Risk Scores and BLUP Prediction. Genet Epidemiol. 2015 Sep;39(6):427–38.

208. Tamlander M, Mars N, Pirinen M, Widén E, Ripatti S. Integration of questionnaire-based risk factors improves polygenic risk scores for human coronary heart disease and type 2 diabetes. Commun Biol 2022 51 [Internet]. 2022 Feb 23 [cited 2022 May 10];5(1):1–13. Available from: https://www.nature.com/articles/s42003-021-02996-0

209. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update From the GBD 2019 Study. J Am Coll Cardiol. 2020 Dec 22;76(25):2982–3021.

210. Abbafati C, Abbas KM, Abbasi-Kangevari M, Abd-Allah F, Abdelalim A, Abdollahi M, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet [Internet]. 2020 Oct 17 [cited 2022 May 10];396(10258):1223–49. Available from: http://www.thelancet.com/article/S0140673620307522/fulltext

211. Califf RM, Armstrong PW, Carver JR, D RB, Strauss WE. Task Force 5. Stratification of Patients Into High, Medium Risk Subgroups for Purposes of Risk Factor Management General Perspectives on Risk and Effect on Intervention. JACC. 1007;27(5):964–1047.

212. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practiceThe Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J [Internet]. 2016 Aug 1 [cited 2022 May 10];37(29):2315–81. Available from: https://academic.oup.com/eurheartj/article/37/29/2315/1748952

213. Lindbohm J V., Sipilä PN, Mars NJ, Pentti J, Ahmadi-Abhari S, Brunner EJ, et al. 5-year versus risk-category-specific screening intervals for cardiovascular disease prevention: a cohort study. Lancet Public Heal. 2019 Apr 1;4(4):e189–99.

214. Kullo IJ, Jouni H, Austin EE, Brown SA, Kruisselbrink TM, Isseh IN, et al. Incorporating a genetic risk score into coronary heart disease risk estimates: Effect on low-density lipoprotein cholesterol levels (the MI-GENES Clinical Trial). Circulation [Internet]. 2016 Mar 22 [cited 2022 May 11];133(12):1181–8. Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.115.020109

215. Obeng AO, Fei K, Levy KD, Elsey AR, Pollin TI, Ramirez AH, et al. Physician-Reported Benefits and Barriers to Clinical Implementation of Genomic Medicine: A Multi-Site IGNITE-Network Survey. J Pers Med 2018, Vol 8, Page 24 [Internet]. 2018 Jul 24 [cited 2022 May 11];8(3):24. Available from: https://www.mdpi.com/2075-4426/8/3/24/htm

216. Hartzler A, McCarty CA, Rasmussen L V., Williams MS, Brilliant M, Bowton EA, et al. Stakeholder engagement: a key component of integrating genomic information into electronic health records. Genet Med [Internet]. 2013 Oct [cited 2022 May 13];15(10):792–801. Available from: https://pubmed.ncbi.nlm.nih.gov/24030437/

217. Abdela OA, Bhagavathula AS, Gebreyohannes EA, Tegegn HG. Ethiopian health care professionals’ knowledge, attitude, and interests toward pharmacogenomics. Pharmgenomics Pers Med [Internet]. 2017 Dec 5 [cited 2022 Jan 12];10:279–85. Available from: https://www.dovepress.com/ethiopian-health-care-professionals-knowledge-

attitude-and-interests-t-peer-reviewed-fulltext-article-PGPM

218. Haga S, Burke W, Ginsburg G, Mills R, Agans R. Primary care physicians' knowledge of and experience with pharmacogenetic testing. Clin Genet. 2012 Oct;82(4):388–94.

219. Carroll JC, Morrison S, Miller FA, Wilson BJ, Permaul JA, Allanson J. Anticipating the primary care role in genomic medicine: expectations of genetics health professionals. J Community Genet [Internet]. 2021 Oct 1 [cited 2022 May 9];12(4):559–68. Available from: https://link.springer.com/article/10.1007/s12687-021-00544-1

220. Chambers C, Axell-House D, Mills G, Bittner-Fagan H, Rosenthal M, Johnson M, et al. Primary Care Physicians' Experience and Confidence with Genetic Testing and Perceived Barriers to Genomic Medicine. J Fam Med [Internet]. 2015 [cited 2022 Jan 11];2(2). Available from: https://austinpublishinggroup.com/family-medicine/fulltext/jfm-v2-id1024.php

221. Yu MWC, Fung JLF, Ng APP, Li Z, Lan W, Chung CCY, et al. Preparing genomic revolution: Attitudes, clinical practice, and training needs in delivering genetic counseling in primary care in Hong Kong and Shenzhen, China. Mol Genet Genomic Med [Internet]. 2021 Jul 1 [cited 2022 Jan 11];9(7). Available from: /pmc/articles/PMC8372068/

222. Rubanovich CK, Cheung C, Mandel J, Bloss CS. Physician preparedness for big genomic data: a review of genomic medicine education initiatives in the United States. Hum Mol Genet. 2018 Aug 1;27(R2):R250-258.

223. Addis A, Trotta F, Tafuri G, De Fiore L. Information needs on precision medicine: a survey of Italian health care professionals. Ann Ist Super Sanita [Internet]. 2018 [cited 2022 Jan 12];54(4):316–23. Available from: https://pubmed.ncbi.nlm.nih.gov/30575568/

224. Albassam A, Alshammari S, Ouda G, Koshy S, Awad A. Knowledge, perceptions and confidence of physicians and pharmacists towards pharmacogenetics practice in Kuwait. PLoS One [Internet]. 2018 Sep 1 [cited 2022 Jan 12];13(9):e0203033. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0203033

225. Selkirk CG, Weissman SM, Anderson A, Hulick PJ. Physicians' preparedness for integration of genomic and pharmacogenetic testing into practice within a major healthcare system. Genet Test Mol Biomarkers [Internet]. 2013 Mar 1 [cited 2022 May 11];17(3):219–25. Available from: www.liebertpub.com

226. Chow-White P, Ha D, Laskin J. Knowledge, attitudes, and values among physicians working with clinical genomics: A survey of medical oncologists. Hum Resour Health [Internet]. 2017 Jun 27 [cited 2022 Jan 11];15(1):1–9. Available from: https://human-resources-health.biomedcentral.com/articles/10.1186/s12960-017-0218-z

227. Harris P., R Taylor, BL Minor, V Elliott, M Fernandez, L O'Neal, L McLeod, G

Delacqua, F Delacqua, J Kirby, SN Duda TRedc consortium. The REDCap consortium: Building an international community of software partners. J Biomed Inf. 2019;

228. PA Harris, R Taylor, R Thielke, J Payne, N Gonzalez JC. Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inf. 2009;42(2):377–81.

229. Revelle WR. psych: Procedures for Personality and Psychological Research. Software. 2017.

230. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag, editor. New York; 2016. Available from: https://ggplot2.tidyverse.org.

231. Carroll JC, Allanson J, Morrison S, Miller FA, Wilson BJ, Permaul JA, et al. Informing integration of genomic medicine into primary care: an assessment of current practice, attitudes and desired resources. Front Genet. 2019 Nov 21;10:1189.

232. Alharbi AA, Shaqran TM, Eltobgy A, Albalawi AR, Alnawmasi WS. Physicians' Perspective on Diabetes Mellitus Management within the Context of Personalized Medicine Era in Tabuk Governorate, Saudi Arabia. Open Access Maced. J Med Sci [Internet]. 2019;7(10):1706. Available from: https://doi.org/10.3889/oamjms.2019.322

233. Albitar L, Alchamat GA. Pharmacogenetics: Knowledge assessment amongst Syrian pharmacists and physicians. BMC Health Serv Res [Internet]. 2021 Dec 1 [cited 2022 Jan 12];21(1):1–8. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-021-07040-9

234. Spiech KM, Tripathy PR, Woodcock AM, Sheth NA, Collins KS, Kannegolla K, et al. Implementation of a Renal Precision Medicine Program: Clinician Attitudes and Acceptance. Life 2020, Vol 10, Page 32 [Internet]. 2020 Mar 26 [cited 2022 May 12];10(4):32. Available from: https://www.mdpi.com/2075-1729/10/4/32/htm

235. Jayasinghe K, Quinlan C, Mallett AJ, Kerr PG, McClaren B, Nisselle A, et al. Attitudes and Practices of Australian Nephrologists Toward Implementation of Clinical Genomics. Kidney Int Reports. 2021 Feb 1;6(2):272–83.

236. Doll B, De Castro MJ, Fries MH, Pock AR, Seibert D, Yang W. Precision Medicine—A Demand Signal for Genomics Education. Mil Med [Internet]. 2021 Dec 30 [cited 2022 May 13];187(Supplement_1):40–6. Available from: https://academic.oup.com/milmed/article/187/Supplement_1/40/6489944

237. Crellin E, McClaren B, Nisselle A, Best S, Gaff C, Metcalfe S. Preparing Medical Specialists to Practice Genomic Medicine: Education an Essential Part of a Broader Strategy. Front Genet. 2019 Sep 11;10.

238. Cornel MC. Evidence-Based Genetic Education of Non-Genetic-Expert Physicians: Experiences Over Three Decades in Amsterdam. Front Genet [Internet]. 2019 [cited 2022 May 23];10(JUL). Available from: /pmc/articles/PMC6687771/

239. Nisselle A, King EA, Mcclaren B, Janinski M, Metcalfe S, Gaff C. Measuring physician practice, preparedness and preferences for genomic medicine: a national survey. BMJ Open [Internet]. 2021 Jul 1 [cited 2022 Jun 6];11(7):e044408. Available from: https://bmjopen.bmj.com/content/11/7/e044408

240. Laplace B, Calvet B, Lacroix A, Mouchabac S, Picard N, Girard M, et al. Acceptability of pharmacogenetic testing among french psychiatrists, a national survey. J Pers Med [Internet]. 2021 May 21 [cited 2022 May 13];11(6):446. Available from: https://www.mdpi.com/2075-4426/11/6/446/htm

241. Wilcox RL, Adem P V., Afshinnekoo E, Atkinson JB, Burke LW, Cheung H, et al. The Undergraduate Training in Genomics (UTRIG) Initiative: Early & active training for physicians in the genomic medicine era. Per Med. 2018;15(3):199–208.

242. Reed EK, Johansen Taber KA, Ingram Nissen T, Schott S, Dowling LO, O'Leary JC, et al. What works in genomics education: outcomes of an evidenced-based instructional model for community-based physicians. Genet Med. 2016 Jul 1;18(7):737–45.

243. Johnson LM, Valdez JM, Quinn EA, Sykes AD, McGee RB, Nuccio R, et al. Integrating next-generation sequencing into pediatric oncology practice: An assessment of physician confidence and understanding of clinical genomics. Cancer. 2017 Jun 15;123(12):2352–9.

244. de Moor JS, Gray SW, Mitchell SA, Klabunde CN, Freedman AN. Oncologist Confidence in Genomic Testing and Implications for Using Multimarker Tumor Panel Tests in Practice. JCO Precis Oncol [Internet]. 2020 Jun 11 [cited 2022 Jun 6];4(4):620–31. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32923869

245. Fricker RD. Sampling methods for web and e-mail surveys. In: The SAGE handbook of online research methods. London: SAGE Publications Ltd; 2008.

246. Tourangeau, R., Rips, L. and Rasinski K. The psychology of survey response. Cambridge, UK: Cambridge University Press; 2009.

247. Ojji DB, Lamont K, Ojji OI, Egenti BN, Sliwa K. Primary care in the prevention, treatment and control of cardiovascular disease in sub-Saharan Africa. Cardiovasc J Afr [Internet]. 2017 Jul 1 [cited 2022 May 23];28(4):251. Available from: /pmc/articles/PMC5642029/

248. Beaglehole R, Epping-Jordan JA, Patel V, Chopra M, Ebrahim S, Kidd M, et al. Improving the prevention and management of chronic disease in low-income and middle-income countries: a priority for primary health care. Lancet. 2008

Sep 13;372(9642):940–9.

249. Mash R, Almeida M, Wong WCW, Kumar R, von Pressentin KB. The roles and training of primary care doctors: China, India, Brazil and South Africa. Hum Resour Health [Internet]. 2015 Dec 4 [cited 2022 May 23];13(1):1–9. Available from: https://human-resources-health.biomedcentral.com/articles/10.1186/s12960-015-0090-7

250. Cheng GJ, Wagner AL, O'Shea BQ, Joseph CA, Finlay JM, Kobayashi LC. Multimorbidity and Mental Health Trajectories Among Middle-Aged and Older U.S. Adults During the COVID-19 Pandemic: Longitudinal Findings From the COVID-19 Coping Study. Innov aging [Internet]. 2022 Jul 1 [cited 2022 Oct 13];6(5):igac047. Available from: http://www.ncbi.nlm.nih.gov/pubmed/36035631

251. Fortin M, Bravo G, Hudon C, Lapointe L, Dubois MF, Almirall J. Psychological distress and multimorbidity in primary care. Ann Fam Med [Internet]. 2006 Sep [cited 2022 Oct 13];4(5):417–22. Available from: https://pubmed.ncbi.nlm.nih.gov/17003141/

252. Calderón-Larrañaga A, Poblador-Plou B, González-Rubio F, Gimeno-Feliu LA, Abad-Díez JM, Prados-Torres A. Multimorbidity, polypharmacy, referrals, and adverse drug events: are we doing things well? Br J Gen Pract [Internet]. 2012 Dec [cited 2022 Oct 13];62(605):e821. Available from: /pmc/articles/PMC3505415/

253. Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. Arch Intern Med [Internet]. 2002 Nov 15 [cited 2022 Oct 13];162(20):2269–76. Available from: https://pubmed.ncbi.nlm.nih.gov/12418941/

254. Salisbury C, Johnson L, Purdy S, Valderas JM, Montgomery AA. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract [Internet]. 2011 Jan [cited 2023 May 10];61(582). Available from: https://pubmed.ncbi.nlm.nih.gov/21401985/

255. Boyd CM, Fortin M. Future of Multimorbidity Research: How Should Understanding of Multimorbidity Inform Health System Design? Public Heal Rev 2010 322 [Internet]. 2010 Dec 10 [cited 2023 May 10];32(2):451–74. Available from: https://publichealthreviews.biomedcentral.com/articles/10.1007/BF03391611

256. World Health Organization. Multimorbidity: Technical Series on Safer Primary Care: Geneva, Switzerland; 2016.

257. Ho ISS, Azcoaga-Lorenzo A, Akbari A, Davies J, Khunti K, Kadam UT, et al. Measuring multimorbidity in research: Delphi consensus study. BMJ Med [Internet]. 2022 Jul 1 [cited 2022 Oct 13];1(1):e000247. Available from: https://bmjmedicine.bmj.com/content/1/1/e000247

258. Skou ST, Mair FS, Fortin M, Guthrie B, Nunes BP, Miranda JJ, et al.

Multimorbidity. Nat Rev Dis Prim. 2022 Dec 1;8(1).

259. Hajat C, Stein E. The global burden of multiple chronic conditions: A narrative review. Prev Med Reports. 2018 Dec 1;12:284–93.

260. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, et al. Aging with multimorbidity: a systematic review of the literature. Ageing Res Rev [Internet]. 2011 Sep [cited 2022 Oct 13];10(4):430–9. Available from: https://pubmed.ncbi.nlm.nih.gov/21402176/

261. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. Lancet [Internet]. 2012 Jul 7 [cited 2023 May 10];380(9836):37–43. Available from: http://www.thelancet.com/article/S0140673612602402/fulltext

262. Roomaney RA, Van Wyk B, Wyk VP Van. Decolonising multimorbidity? research gaps in low and middle-income countries. Pan Afr Med J [Internet]. 2022 Jan 1 [cited 2023 May 10];41(140):41. Available from: /pmc/articles/PMC9034556/

263. Nguyen H, Manolova G, Daskalopoulou C, Vitoratou S, Prince M, Prina AM. Prevalence of multimorbidity in community settings: A systematic review and meta-analysis of observational studies. J Comorbidity. 2019 Jan 1;9:2235042X1987093.

264. Xu X, Mishra GD, Jones M, Level XX, Health P, Herston B. Mapping the global research landscape and knowledge gaps on multimorbidity: a bibliometric study. 2017 [cited 2022 Oct 12];7(1):10414. Available from: http://cluster.cis.drexel.edu/~cchen/

265. Alshakhs M, Jackson B, Ikponmwosa D, Reynolds R, Madlock-Brown C. Multimorbidity patterns across race/ethnicity as stratified by age and obesity. Sci Reports 2022 121 [Internet]. 2022 Jun 11 [cited 2023 May 10];12(1):1–16. Available from: https://www.nature.com/articles/s41598-022-13733-w

266. Kalgotra P, Sharda R, Croff JM. Examining multimorbidity differences across racial groups: a network analysis of electronic medical records. Sci Rep [Internet]. 2020 Dec 1 [cited 2023 May 10];10(1). Available from: /pmc/articles/PMC7419498/

267. Academy of Medical Sciences. Multimorbidity: a priority for global health research. London, UK; 2018.

268. Johnston MC, Crilly M, Black C, Prescott GJ, Mercer SW. Defining and measuring multimorbidity: a systematic review of systematic reviews. Eur J Public Health [Internet]. 2019 Feb 1 [cited 2022 Oct 12];29(1):182–9. Available from: https://pubmed.ncbi.nlm.nih.gov/29878097/

269. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc

[Internet]. 2015 Sep 1 [cited 2023 May 10];13(3):141–6. Available from: https://pubmed.ncbi.nlm.nih.gov/26134548/

270. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med [Internet]. 2018 Sep 4;169(7):467–73. Available from: https://doi.org/10.7326/M18-0850

271. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev [Internet]. 2016 Dec 5 [cited 2022 Oct 12];5(1):1–10. Available from: https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-016-0384-4

272. Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization [Internet]. 2020. Seattle, Washington: University of Washington; [cited 2023 May 4]. Available from: https://www.healthdata.org/data-visualization/gbd-compare

273. Microsoft Corporation. Microsoft Excel. Redmond, WA; 2018.

274. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of Intersecting Sets. IEEE Trans Vis Comput Graph. 2014 Dec 31;20:1983–92.

275. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis [Internet]. 1987 [cited 2023 May 10];40(5):373–83. Available from: https://pubmed.ncbi.nlm.nih.gov/3558716/

276. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. J Clin Epidemiol. 2004 Dec;57(12):1288–94.

277. Chang AY, Gómez-Olivé FX, Payne C, Rohr JK, Manne-Goehler J, Wade AN, et al. Chronic multimorbidity among older adults in rural South Africa. BMJ Glob Heal [Internet]. 2019 Aug 1 [cited 2023 May 10];4(4):e001386. Available from: https://gh.bmj.com/content/4/4/e001386

278. Sharman M, Bachmann M. Prevalence and health effects of communicable and non-communicable disease comorbidity in rural KwaZulu-Natal, South Africa. Trop Med Int Health [Internet]. 2019 Oct 1 [cited 2023 May 10];24(10):1198–207. Available from: https://pubmed.ncbi.nlm.nih.gov/31389103/

279. Hopkins KL, Hlongwane KE, Otwombe K, Dietrich J, Cheyip M, Olivier J, et al. The substantial burden of non-communicable diseases and HIV-comorbidity amongst adults: Screening results from an integrated HIV testing services clinic for adults in Soweto, South Africa. EClinicalMedicine [Internet]. 2021 Aug 1 [cited 2023 May 10];38. Available from: https://pubmed.ncbi.nlm.nih.gov/34308316/

280. Wade AN, Payne CF, Berkman L, Chang A, Gómez-Olivé FX, Kabudula C, et al. Multimorbidity and mortality in an older, rural black South African population cohort with high prevalence of HIV findings from the HAALSI Study. BMJ Open [Internet]. 2021 Sep 1;11(9):e047777. Available from: http://bmjopen.bmj.com/content/11/9/e047777.abstract

281. Boake M, Mash R. Diabetes in the Western Cape, South Africa: A secondary analysis of the diabetes cascade database 2015 – 2020. Prim Care Diabetes. 2022 Aug 1;16(4):555–61.

282. Ntiyani N, Letamo G, Keetile M. Prevalence of and factors associated with hypertension, diabetes, stroke and heart attack multimorbidity in Botswana: Evidence from STEPS 2014 survey. PLoS One [Internet]. 2022 Mar 1 [cited 2023 May 10];17(3). Available from: https://pubmed.ncbi.nlm.nih.gov/35324986/

283. Mohamed SF, Haregu TN, Uthman OA, Khayeka-Wandabwa C, Muthuri SK, Asiki G, et al. Multimorbidity from Chronic Conditions among Adults in Urban Slums: The AWI-Gen Nairobi Site Study Findings. Glob Heart [Internet]. 2021 [cited 2023 May 10];16(1). Available from: https://pubmed.ncbi.nlm.nih.gov/33598386/

284. Mukadas Akindele O, Ushotanefe U. Multimorbidity of chronic diseases of lifestyle among South African adults. PAMJ [Internet]. 2021;38(332). Available from: https://www.panafrican-med-journal.com/content/article/38/332/full

285. Odland ML, Payne C, Witham MD, Siedner MJ, Bärnighausen T, Bountogo M, et al. Epidemiology of multimorbidity in conditions of extreme poverty: a population-based study of older adults in rural Burkina Faso. BMJ Glob Heal [Internet]. 2020 Mar 1;5(3):e002096. Available from: http://gh.bmj.com/content/5/3/e002096.abstract

286. Cook JA, Burke-Miller JK, Steigman PJ, Schwartz RM, Hessol NA, Milam J, et al. Prevalence, Comorbidity, and Correlates of Psychiatric and Substance Use Disorders and Associations with HIV Risk Behaviors in a Multisite Cohort of Women Living with HIV. AIDS Behav [Internet]. 2018 Oct 1 [cited 2023 May 10];22(10):3141–54. Available from: https://pubmed.ncbi.nlm.nih.gov/29460130/

287. Vennu V, Abdulrahman TA, Alenazi AM, Bindawas SM. Associations between social determinants and the presence of chronic diseases: Data from the osteoarthritis Initiative. BMC Public Health [Internet]. 2020 Aug 31 [cited 2023 May 10];20(1):1–7. Available from: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-020-09451-5

288. Diaz E, Poblador-Pou B, Gimeno-Feliu LA, Calderón-Larrañaga A, Kumar BN, Prados-Torres A. Multimorbidity and Its Patterns according to Immigrant Origin. A Nationwide Register-Based Study in Norway. PLoS One [Internet]. 2015 Dec 1 [cited 2023 May 10];10(12):e0145233. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0145233

289. Mthoko NFN, Pazvakawambwa L, Leonhardt M, Lien L. Physical comorbidity among patients attending mental health services at Windhoek Central Hospital, Namibia. Pan Afr Med J [Internet]. 2022 [cited 2023 May 10];41. Available from: https://pubmed.ncbi.nlm.nih.gov/35734337/

290. Peltzer K. Tuberculosis non-communicable disease comorbidity and multimorbidity in public primary care patients in South Africa. African J Prim Heal care Fam Med [Internet]. 2018 Apr 11 [cited 2023 May 10];10(1). Available from: https://pubmed.ncbi.nlm.nih.gov/29781683/

291. Weimann A, Dai D, Oni T. A cross-sectional and spatial analysis of the prevalence of multimorbidity and its association with socioeconomic disadvantage in South Africa: A comparison between 2008 and 2012. Soc Sci Med [Internet]. 2016 Aug 1 [cited 2023 May 10];163:144–56. Available from: https://pubmed.ncbi.nlm.nih.gov/27423295/

292. Ataguba JEO. Inequalities in multimorbidity in South Africa. Int J Equity Health [Internet]. 2013 [cited 2023 May 10];12(1). Available from: https://pubmed.ncbi.nlm.nih.gov/23962076/

293. Chikezie UE, Otakpor AN, Kuteyi OB, James BO. Depression among people living with human immunodeficiency virus infection/acquired immunodeficiency syndrome in Benin City, Nigeria: a comparative study. Niger J Clin Pract [Internet]. 2013 Apr [cited 2023 May 10];16(2):238–42. Available from: https://pubmed.ncbi.nlm.nih.gov/23563469/

294. Bhagavathula AS, Seid MA, Adane A, Gebreyohannes EA, Brkic J, Fialová D. Prevalence and Determinants of Multimorbidity, Polypharmacy, and Potentially Inappropriate Medication Use in the Older Outpatients: Findings from EuroAgeism H2020 ESR7 Project in Ethiopia. Pharmaceuticals [Internet]. 2021 [cited 2023 May 10];14(9). Available from: /pmc/articles/PMC8468438/

295. Mutyambizi C, Chola L, Groot W, Pavlova M, Labadarios D, Hongoro C. The extent and determinants of diabetes and cardiovascular disease comorbidity in South Africa - results from the South African National Health and Nutrition Examination Survey (SANHANES-1). BMC Public Health [Internet]. 2017 Sep 26 [cited 2023 May 10];17(1). Available from: https://pubmed.ncbi.nlm.nih.gov/28950847/

296. Alaba O, Chola L. The social determinants of multimorbidity in South Africa. Int J Equity Health [Internet]. 2013 [cited 2023 May 10];12(1). Available from: https://pubmed.ncbi.nlm.nih.gov/23962055/

297. Kunna R, San Sebastian M, Stewart Williams J. Measurement and decomposition of socioeconomic inequality in single and multimorbidity in older adults in China and Ghana: Results from the WHO study on global AGEing and adult health (SAGE). Int J Equity Health [Internet]. 2017 May 15 [cited 2023 May 10];16(1):1–17. Available from: https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-017-0578-y

298. Nimako BA, Baiden F, Sackey SO, Binka F. Multimorbidity of chronic diseases

among adult patients presenting to an inner-city clinic in Ghana. Global Health [Internet]. 2013 Nov 26 [cited 2023 May 10];9(1):61. Available from: /pmc/articles/PMC3892007/

299. Zelnick LR, Katz R, Young BA, Correa A, Kestenbaum BR, de Boer IH, et al. Echocardiographic Measures and Estimated GFR Decline Among African Americans: The Jackson Heart Study. Am J Kidney Dis [Internet]. 2017 Aug 1 [cited 2023 May 10];70(2):199–206. Available from: https://pubmed.ncbi.nlm.nih.gov/28143672/

300. Gabriel A, Zare H, Jones W, Yang M, Ibe CA, Cao Y, et al. Evaluating Depressive Symptoms Among Low-Socioeconomic-Status African American Women Aged 40 to 75 Years With Uncontrolled Hypertension: A Secondary Analysis of a Randomized Clinical Trial. JAMA Psychiatry [Internet]. 2021 Apr 1 [cited 2023 May 10];78(4):1. Available from: /pmc/articles/PMC7876618/

301. Dibato JE, Montvida O, Zaccardi F, Sargeant JA, Davies MJ, Khunti K, et al. Association of Cardiometabolic Multimorbidity and Depression With Cardiovascular Events in Early-Onset Adult Type 2 Diabetes: A Multiethnic Study in the U.S. Diabetes Care [Internet]. 2021 Jan 1 [cited 2023 May 10];44(1):231–9. Available from: https://diabetesjournals.org/care/article/44/1/231/33105/Association-of-Cardiometabolic-Multimorbidity-and

302. Erving CL, Frazier C. The Association between Multiple Chronic Conditions and Depressive Symptoms: Intersectional Distinctions by Race, Nativity, and Gender. J Health Soc Behav [Internet]. 2021 Dec 1 [cited 2023 May 10];62(4):599–617. Available from: https://pubmed.ncbi.nlm.nih.gov/34590498/

303. Yang J, Ju X, Liu F, Asan O, Church T, Smith J. Prediction for the Risk of Multiple Chronic Conditions among Working Population in the United States with Machine Learning Models. IEEE Open J Eng Med Biol. 2021;2:291–8.

304. Heckman BD, Merrill JC, Anderson T. Race, psychiatric comorbidity, and headache characteristics in patients in headache subspecialty treatment clinics. Ethn Health [Internet]. 2013 Feb 1 [cited 2023 May 10];18(1):34–52. Available from: https://pubmed.ncbi.nlm.nih.gov/22541025/

305. Thorpe RJJ, Bell CN, Kennedy-Hendricks A, Harvey J, Smolen JR, Bowie J V, et al. Disentangling race and social context in understanding disparities in chronic conditions among men. J Urban Health. 2015 Feb;92(1):83–92.

306. Marzà-Florensa A, Boateng D, Agyemang C, Beune E, Meeks KAC, Bahendeka S, et al. Multimorbidity Among Migrant and Non-Migrant Ghanaians: The RODAM Study. Int J Public Health [Internet]. 2021 Dec 31 [cited 2023 May 10];66. Available from: https://pubmed.ncbi.nlm.nih.gov/35035346/

307. Romano E, Ma R, Vancampfort D, Firth J, Felez-Nobrega M, Haro JM, et al. Multimorbidity and obesity in older adults from six low- and middle-income countries. Prev Med (Baltim) [Internet]. 2021;153:106816. Available from:

https://www.sciencedirect.com/science/article/pii/S0091743521003856

308. Pengpid S, Peltzer K. Mental morbidity and its associations with socio-behavioural factors and chronic conditions in rural middle- and older-aged adults in South Africa. J Psychol Africa (south Sahara, Caribbean, Afro-Latin Am. 2020;30(3):257–63.

309. Ayeni OA, Norris SA, Joffe M, Cubasch H, Nietz S, Buccimazza I, et al. The multimorbidity profile of South African women newly diagnosed with breast cancer. Int J cancer. 2020 Jul;147(2):361–74.

310. Stirland LE, González-Saavedra L, Mullin DS, Ritchie CW, Muniz-Terrera G, Russ TC. Measuring multimorbidity beyond counting diseases: systematic review of community and population studies and guide to index choice. BMJ [Internet]. 2020 Feb 1 [cited 2023 May 10];368. Available from: https://pubmed.ncbi.nlm.nih.gov/32071114/

311. Fortin M, Stewart M, Poitras ME, Almirall J, Maddocks H. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. Ann Fam Med [Internet]. 2012 [cited 2023 May 10];10(2):142–51. Available from: https://pubmed.ncbi.nlm.nih.gov/22412006/

312. Lee ES, Koh HL, Ho EQ-Y, Teo SH, Wong FY, Ryan BL, et al. Systematic review on the instruments used for measuring the association of the level of multimorbidity and clinically important outcomes. BMJ Open [Internet]. 2021 May 1;11(5):e041219. Available from: http://bmjopen.bmj.com/content/11/5/e041219.abstract

313. Sinnige J, Braspenning J, Schellevis F, Stirbu-Wagner I, Westert G, Korevaar J. The prevalence of disease clusters in older adults with multiple chronic diseases--a systematic literature review. PLoS One [Internet]. 2013 Nov 11 [cited 2023 May 10];8(11). Available from: https://pubmed.ncbi.nlm.nih.gov/24244534/

314. Violan C, Foguet-Boreu Q, Flores-Mateo G, Salisbury C, Blom J, Freitag M, et al. Prevalence, Determinants and Patterns of Multimorbidity in Primary Care: A Systematic Review of Observational Studies. PLoS One [Internet]. 2014 Jul 21 [cited 2023 May 10];9(7):e102149. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0102149

315. Prados-Torres A, Calderón-Larrañaga A, Hancco-Saavedra J, Poblador-Plou B, Van Den Akker M. Multimorbidity patterns: a systematic review. J Clin Epidemiol [Internet]. 2014 Mar [cited 2023 May 10];67(3):254–66. Available from: https://pubmed.ncbi.nlm.nih.gov/24472295/

316. Gottlieb S. Antiretroviral therapy increases risk of heart attack. BMJ Br Med J [Internet]. 2003 Nov 11 [cited 2023 May 10];327(7425):1186. Available from: /pmc/articles/PMC1126872/

317. Jarwani B. Cardiovascular Disease and Antiretroviral Therapy. J Glob Infect Dis [Internet]. 2019 Jul 1 [cited 2023 May 10];11(3):91. Available from:

/pmc/articles/PMC6733197/

318. Dubé MP, Lipshultz SE, Fichtenbaum CJ, Greenberg R, Schecter AD, Fisher SD. Effects of HIV Infection and Antiretroviral Therapy on the Heart and Vasculature. Circulation [Internet]. 2008 Jul 8 [cited 2023 May 10];118(2). Available from: https://www.ahajournals.org/doi/abs/10.1161/circulationaha.107.189625

319. Smith L, Shin J Il, Butler L, Barnett Y, Oh H, Jacob L, et al. Physical multimorbidity and depression: A mediation analysis of influential factors among 34,129 adults aged ≥50 years from low- and middle-income countries. Depress Anxiety [Internet]. 2022 May 1;39(5):376–86. Available from: https://doi.org/10.1002/da.23250

320. Mochari-Greenberger H, Mosca L. Differential Outcomes by Race and Ethnicity in Patients with Coronary Heart Disease: A Contemporary Review. Curr Cardiovasc Risk Rep [Internet]. 2015 May 1 [cited 2023 May 10];9(5). Available from: /pmc/articles/PMC4405256/

321. Clements J, West B, Yaker Z, Lauinger B, McCullers D, Haubert J, et al. Disparities in Diabetes-related Multiple Chronic Conditions and Mortality: The Influence of Race. Diabetes Res Clin Pract. 2019 Dec 14;159:107984.

322. Breslau J, Aguilar-Gaxiola S, Kendler KS, Su M, Williams D, Kessler RC. Specifying race-ethnic differences in risk for psychiatric disorder in a USA national sample. Psychol Med [Internet]. 2006 [cited 2023 May 10];36(1):57–68. Available from: https://pubmed.ncbi.nlm.nih.gov/16202191/

323. Verest WJGM, Galenkamp H, Spek B, Snijder MB, Stronks K, van Valkengoed IGM. Do ethnic inequalities in multimorbidity reflect ethnic differences in socioeconomic status? The HELIUS study. Eur J Public Health [Internet]. 2019 Aug 1;29(4):687–93. Available from: https://doi.org/10.1093/eurpub/ckz012

324. Aparasu RR, Mort JR, Brandt H. Polypharmacy trends in office visits by the elderly in the United States, 1990 and 2000. Res Social Adm Pharm [Internet]. 2005 Sep [cited 2023 May 10];1(3):446–59. Available from: https://pubmed.ncbi.nlm.nih.gov/17138489/

325. Simakoloyi N, Erasmus E, van Hoving DJ. The characteristics of geriatric patients managed within the resuscitation unit of a district-level emergency centre in Cape Town. African J Emerg Med [Internet]. 2022 Mar 1 [cited 2023 May 10];12(1):39. Available from: /pmc/articles/PMC8761600/

326. Perry RG, Mitchell JA, Hawkins J, Johnson-Lawrence V. The Role of Age and Multimorbidity in Shaping Older African American Men's Experiences with Patient–Provider Communication. Geriatrics [Internet]. 2018 Dec 1 [cited 2023 May 10];3(4). Available from: /pmc/articles/PMC6371103/

327. Myers B, Joska JA, Lund C, Levitt NS, Butler CC, Naledi T, et al. Patient preferences for the integration of mental health counseling and chronic disease care in South Africa. Patient Prefer Adherence [Internet]. 2018 [cited 2023 May

10];12:1797. Available from: /pmc/articles/PMC6154740/

328. Segafredo G, Kapur A, Robbiati C, Joseph N, de Sousa JR, Putoto G, et al. Integrating TB and non-communicable diseases services: Pilot experience of screening for diabetes and hypertension in patients with Tuberculosis in Luanda, Angola. PLoS One [Internet]. 2019 Jul 1 [cited 2023 May 10];14(7). Available from: https://pubmed.ncbi.nlm.nih.gov/31276500/

329. Khabala KB, Edwards JK, Baruani B, Sirengo M, Musembi P, Kosgei RJ, et al. Medication Adherence Clubs: a potential solution to managing large numbers of stable patients with multiple chronic diseases in informal settlements. Trop Med Int Health [Internet]. 2015 Oct 1 [cited 2023 May 10];20(10):1265–70. Available from: https://pubmed.ncbi.nlm.nih.gov/25962952/

330. Ameh S, D'Ambruoso L, Gómez-Olivé FX, Kahn K, Tollman SM, Klipstein-Grobusch K. Paradox of HIV stigma in an integrated chronic disease care in rural South Africa: Viewpoints of service users and providers. PLoS One [Internet]. 2020 Jul 31;15(7):e0236270. Available from: https://doi.org/10.1371/journal.pone.0236270

331. Matima R, Murphy K, Levitt NS, BeLue R, Oni T. A qualitative study on the experiences and perspectives of public sector patients in Cape Town in managing the workload of demands of HIV and type 2 diabetes multimorbidity. PLoS One [Internet]. 2018 Mar 1 [cited 2023 May 10];13(3). Available from: https://pubmed.ncbi.nlm.nih.gov/29538415/

332. Godongwana M, De Wet-Billings N, Milovanovic M. The comorbidity of HIV, hypertension and diabetes: a qualitative study exploring the challenges faced by healthcare providers and patients in selected urban and rural health facilities where the ICDM model is implemented in South Africa. BMC Health Serv Res [Internet]. 2021;21(1):647. Available from: https://doi.org/10.1186/s12913-021-06670-3

333. Lebina L, Kawonga M, Oni T, Kim HY, Alaba OA. The cost and cost implications of implementing the integrated chronic disease management model in South Africa. PLoS One [Internet]. 2020 Jun 1 [cited 2023 May 10];15(6). Available from: /pmc/articles/PMC7319351/

334. Mayosi BM, Lawn JE, Van Niekerk A, Bradshaw D, Abdool Karim SS, Coovadia HM. Health in South Africa: changes and challenges since 2009. Lancet (London, England) [Internet]. 2012 [cited 2023 May 10];380(9858):2029–43. Available from: https://pubmed.ncbi.nlm.nih.gov/23201214/

335. Adjei KK, Kikuchi K, Owusu-Agyei S, Enuameh Y, Shibanuma A, Ansah EK, et al. Women's overall satisfaction with health facility delivery services in Ghana: A mixed-methods study. Trop Med Health [Internet]. 2019 Jul 5 [cited 2023 May 10];47(1):1–9. Available from: https://tropmedhealth.biomedcentral.com/articles/10.1186/s41182-019-0172-7

336. Lebina L, Oni T, Alaba OA, Kawonga M. A mixed methods approach to

exploring the moderating factors of implementation fidelity of the integrated chronic disease management model in South Africa. BMC Health Serv Res [Internet]. 2020 Jul 6 [cited 2023 May 10];20(1):1–12. Available from: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-020-05455-4

337. Peck CC, Goulet JP, Lobbezoo F, Schiffman EL, Alstergren P, Anderson GC, et al. Expanding the taxonomy of the diagnostic criteria for temporomandibular disorders. J Oral Rehabil [Internet]. 2014 Jan [cited 2023 May 10];41(1):2–23. Available from: https://pubmed.ncbi.nlm.nih.gov/24443898/

338. Peck R, Mghamba J, Vanobberghen F, Kavishe B, Rugarabamu V, Smeeth L, et al. Preparedness of Tanzanian health facilities for outpatient primary care of hypertension and diabetes: a cross-sectional survey. Lancet Glob Heal [Internet]. 2014 [cited 2023 May 10];2(5):e285. Available from: /pmc/articles/PMC4013553/

339. Ley C, Martin RK, Pareek A, Groll A, Seil R, Tischer T. Machine learning and conventional statistics: making sense of the differences. Knee Surgery, Sport Traumatol Arthrosc [Internet]. 2022 Mar 1 [cited 2023 May 10];30(3):753–7. Available from: https://link.springer.com/article/10.1007/s00167-022-06896-6

340. Cruz JA, Wishart DS. Applications of Machine Learning in Cancer Prediction and Prognosis. Cancer Inform [Internet]. 2006 [cited 2023 May 10];2:59. Available from: /pmc/articles/PMC2675494/

341. Asogwa OA, Boateng D, Marzà-Florensa A, Peters S, Levitt N, Van Olmen J, et al. Multimorbidity of non-communicable diseases in low-income and middle-income countries: a systematic review and meta-analysis. BMJ Open [Internet]. 2022 Jan 1 [cited 2023 May 10];12(1):e049133. Available from: https://bmjopen.bmj.com/content/12/1/e049133

342. Hassaine A, Salimi-Khorshidi G, Canoy D, Rahimi K. Untangling the complexity of multimorbidity with machine learning. Mech Ageing Dev. 2020 Sep 1;190:111325.

343. Uddin S, Wang S, Lu H, Khan A, Hajati F, Khushi M. Comorbidity and multimorbidity prediction of major chronic diseases using machine learning and network analytics. Expert Syst Appl. 2022 Nov 1;205.

344. Khot UN, Khot MB, Bajzer CT, Sapp SK, Ohman EM, Brener SJ, et al. Prevalence of Conventional Risk Factors in Patients With Coronary Heart Disease. JAMA [Internet]. 2003 Aug 20;290(7):898–904. Available from: https://doi.org/10.1001/jama.290.7.898

345. Slunecka JL, van der Zee MD, Beck JJ, Johnson BN, Finnicum CT, Pool R, et al. Implementation and implications for polygenic risk scores in healthcare. Hum Genomics [Internet]. 2021;15(1):46. Available from: https://doi.org/10.1186/s40246-021-00339-y

346. Patel AP, Wang M, Ruan Y, Koyama S, Clarke SL, Yang X, et al. A multi-

ancestry polygenic risk score improves risk prediction for coronary artery disease. Nat Med. 2023 Jul;29(7):1793–803.

347. Pradeep N. Polygenic Risk Scoring for Coronary Heart Disease. J Am Coll Cardiol [Internet]. 2018 Oct 16;72(16):1894–7. Available from: https://doi.org/10.1016/j.jacc.2018.08.1041

348. O'Sullivan JW, Raghavan S, Marquez-Luna C, Luzum JA, Damrauer SM, Ashley EA, et al. Polygenic Risk Scores for Cardiovascular Disease: A Scientific Statement From the American Heart Association. Circulation [Internet]. 2022 Aug 23;146(8):e93–118. Available from: https://doi.org/10.1161/CIR.0000000000001077

349. Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, et al. Improving reporting standards for polygenic scores in risk prediction studies. Nature [Internet]. 2021;591(7849):211–9. Available from: https://doi.org/10.1038/s41586-021-03243-6

350. Sun L, Pennells L, Kaptoge S, Nelson CP, Ritchie SC, Abraham G, et al. Polygenic risk scores in cardiovascular risk prediction: A cohort study and modelling analyses. PLOS Med [Internet]. 2021 Jan 14;18(1):e1003498. Available from: https://doi.org/10.1371/journal.pmed.1003498

351. Hynninen Y, Linna M, Vilkkumaa E. Value of genetic testing in the prevention of coronary heart disease events. PLoS One [Internet]. 2019 Jan 15;14(1):e0210010. Available from: https://doi.org/10.1371/journal.pone.0210010

352. Omotoso OE, Teibo JO, Atiba FA, Oladimeji T, Adebesin AO, Babalghith AO. Bridging the genomic data gap in Africa: implications for global disease burdens. Global Health. 2022 Dec;18(1):103.

353. Ramsay M. African genomic data sharing and the struggle for equitable benefit. Patterns [Internet]. 2022;3(1):100412. Available from: https://www.sciencedirect.com/science/article/pii/S2666389921002865

354. Khera A V, Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell. 2019 Apr;177(3):587-596.e9.

355. Martin AR, Teferra S, Möller M, Hoal EG, Daly MJ. The critical needs and challenges for genetic architecture studies in Africa. Curr Opin Genet Dev [Internet]. 2018;53:113–20. Available from: https://www.sciencedirect.com/science/article/pii/S0959437X18300558

356. Adebamowo CA, Adeyemo A, Ashaye A, Akpa OM, Chikowore T, Choudhury A, et al. Polygenic risk scores for CARDINAL study. Nat Genet. 2022 May;54(5):527–30.

357. Lambert SA, Gil L, Jupp S, Ritchie SC, Xu Y, Buniello A, et al. The Polygenic Score Catalog as an open database for reproducibility and systematic

evaluation. Nat Genet [Internet]. 2021;53(4):420–5. Available from: https://doi.org/10.1038/s41588-021-00783-5

358. Rakow T, Heard C, Newell B. Meeting Three Challenges in Risk Communication. Policy Insights from Behav Brain Sci. 2015 Oct 1;2:147–56.

359. Widén E, Junna N, Ruotsalainen S, Surakka I, Mars N, Ripatti P, et al. How Communicating Polygenic and Clinical Risk for Atherosclerotic Cardiovascular Disease Impacts Health Behavior: an Observational Follow-up Study. Circ Genomic Precis Med [Internet]. 2022 Apr 1;15(2):e003459. Available from: https://doi.org/10.1161/CIRCGEN.121.003459

360. Grant RW, O'Brien KE, Waxler JL, Vassy JL, Delahanty LM, Bissett LG, et al. Personalized genetic risk counseling to motivate diabetes prevention: a randomized trial. Diabetes Care. 2013 Jan;36(1):13–9.

361. Turnbull C. Introducing whole-genome sequencing into routine cancer care: The Genomics England 100 000 Genomes Project. Ann Oncol [Internet]. 2018 Apr 1 [cited 2024 May 14];29(4):784–7. Available from: http://www.annalsofoncology.org/article/S0923753419454926/fulltext

362. The All of Us Research Program Investigators. The "All of Us" Research Program. N Engl J Med [Internet]. 2019 Aug 15 [cited 2024 May 14];381(7):668–76. Available from: https://www.nejm.org/doi/full/10.1056/NEJMsr1809937

363. Musanabaganwa C, Ruton H, Ruhangaza D, Nsabimana N, Kayitare E, Muvunyi TZ, et al. An Assessment of the Knowledge and Perceptions of Precision Medicine (PM) in the Rwandan Healthcare Setting. Vol. 13, Journal of Personalized Medicine. 2023.

364. Munung NS, Mayosi BM, de Vries J. Equity in international health research collaborations in Africa: Perceptions and expectations of African researchers. PLoS One [Internet]. 2017 Oct 16;12(10):e0186237. Available from: https://doi.org/10.1371/journal.pone.0186237

365. Kamp M, Pain O, May A, Lewis CM, Ramsay M. Clinicians' Perceptions towards Precision Medicine Tools for Cardiovascular Disease Risk Stratification in South Africa. J Pers Med. 2022 Aug;12(9).

366. WHO. ssssss [Internet]. Cardiovascular Disease. 2018 [cited 2020 Feb 12]. p. 1. Available from: https://annals.org/aim/fullarticle/718212

APPENDICES

APPENDIX A: AUTHORS' AGREEMENT

Declaration: Student's contribution to articles and agreement of coauthors

I, Michelle Kamp, student number 0603250a, declare that this Thesis is my own work and that I contributed adequately towards research findings published in the articles stated below which are included in my Thesis.

Signature of student M. Kamp **Date** 05 Jan 2024

Name of primary supervisor Prof Michele Ramsay

Signature of primary supervisor M. Ramsay **Date** 16 January 2024

Agreement by co-authors: By signing this declaration, the co-authors listed below agree to the use of the article by the student as part of his Thesis.

Article 2

Title: Ancestry-aligned polygenic risk scores combined with conventional risk factors improve cardiovascular disease risk stratification in African populations

Journal name, year, volume and page numbers: *submitted to Genome Medicine*

Authors	Name	Signature	Date
1st author	Michelle Kamp	M. Kamp	05 Jan 2024
2nd author	Oliver Pain		07 Jan 2024
3th author	Cathryn M Lewis		05 Jan 2024
4th author	Michèle Ramsay	M. Ramsay	16 Jan 2024

Comments by primary supervisor:

The PhD candidate developed the project, did the analyses and wrote the first draft of the paper as a major part of her PhD. She was guided by the co-authors who supervised the work and gave inputs which she addressed.

APPENDIX B: INFORMATION AND CONSENT SHEETS

1. AWI-Gen Information Sheet

AWI-Gen H3Africa

Genomic and Environmental Risk Factors for Cardiometabolic Disease in Africans

Trait Association Study Component

INFORMATION SHEET

Consenting around the Wits-INDEPTH Partnership project funded through the Human Heredity and Health in Africa Initiative

Collection of blood samples for DNA extraction, storage and genomic studies related to studies to traits and disease risk and disease, and urine collection for metabolic studies

Dear Potential Participant,

We are a group of scientists from the University of the Witwatersrand who are working in partnership with other scientists across four countries, Ghana, Burkina Faso, Kenya and South Africa, to understand the risk for developing obesity, diabetes, high blood pressure and other diseases. The main aim of our study is to understand how inherited differences (traits that we get from our parents) influence our health. In order to clearly understand this we first need to understand the differences between people in specific regions of Africa. We call this population structure. Populations are different because of their history (for example, migration and mixing between populations) but also because of their environment (if an inherited change gives the person a health advantage, then people with that change will be more likely to survive and pass the change on to their children).

We would like to invite you to take part in our study. In order to take part, you need to agree to provide a sample of blood, and a urine sample and to have various measurements done on your body and to complete a list of questions related to your health and your life and your environment. You also need to give us permission to use the genetic material (DNA) we get out of your blood for the specific studies to understand population history, your exposure to your environment, your health, your body shape and your risk for developing disease. To do this we break open the cells in the blood to release the DNA (this stands for deoxyribonucleic acid). The DNA contains information to program the development of your body and when something goes wrong with the DNA, this can lead to specific diseases.

Once we have collected the blood sample (27 ml in 5 tubes - the equivalent of less than two tablespoons) we will give it a code and this code will be used to identify your sample during

the study. The information that links your name to your code will be safely stored and will not be used in the study.

Taking a blood sample from a vein in your arm may cause a little discomfort and a little pain. This procedure will be done by a qualified nurse, phlebotomist or doctor.

Once the DNA has been taken out of your blood it will be stored in a safe place at the local Research Institution and at the Sydney Brenner Institute for Molecular Bioscience at the University of the Witwatersrand. A small part of the DNA will also be stored in an H3Africa Biobank. There will be strict control about who may use the samples.

From the studies we will be able to reconstruct some of the past history of the group of people who have donated samples to the study. It will tell us if at some time in the past different populations have mixed and also if there was some factor which led to the selection of specific changes in the DNA.

If at some future date we realise there are other studies that we would like to carry out on your DNA samples, such tests would only be performed if permission is given to us by the Human Research Ethics Committee of the University of the Witwatersrand on your behalf. As stated above, no one will know your identity as your sample will be identified by a code. This study may go one for many years (over 5 years) in order to fully understand the connection between DNA changes and the way you look and your possible health.

Your Participation is voluntary, and if you wish not to participate there is no penalty or loss of benefits in any way. You may discontinue participation in this study at any time without penalty or loss of benefits and your DNA sample will be destroyed and your data removed from the database.

Measurements

Your blood pressure, weight, height, waist and hip circumference will be measured. This will take about 10 minutes.

Ultra sound Scans

An ultra sound scan to measure the internal fat in your body will be done. An ultrasound scan uses high frequency sound waves to take pictures of the body. This is not an invasive procedure, will cause you no pain and is perfectly safe. The actual procedure should not take more than 10 minutes

cIMT(Carotid Intima Media Thickness)

This measurement will be done by a trained technician. It is almost the same as the ultrasound procedure above. You will lie on a flat bed and the technician will gently pass a small hand held instrument up and down your neck. This small instrument will measure if the arteries leading to your brain are healthy. This is not an invasive procedure and the actual procedure should not take more than 20 minutes.

Questionnaires

We will ask you to complete one questionnaire. We will fill in the answers for you. You can skip any question that makes you uncomfortable. Answering these questionnaires should take approximately twenty minutes.

HIV rapid test

The rapid HIV test is a **completely voluntary**, antibody test which will enable you to know your HIV status; positive or negative. If you decide to have this test, you will be given pre-test counseling by a fully trained registered counsellor and asked to fill in a separate consent form. This test is always strictly confidential and can only happen if you agree. There is no way in which anyone can link your HIV status to you name as all results in this study are coded with a number. No-one including your doctor, family, or work colleagues will be told about this test without your permission. The advantage of a rapid test is that you do not have to return to get your test result. Results will be available when your check out, after all the other tests, measurements and questionnaires are completed.

You will have your finger pricked with a sterile needle and the drop of blood will be tested on specific HIV testing kits to check for HIV antibodies. Test results will be given to you in private when you check out today by a registered trained counsellor. If the report states negative it means that there are no antibodies to HIV. The window period will be explained. If the report states positive, it means that you are HIV positive and that there are antibodies to HIV. You will be given a letter to refer you to a clinic specialising in HIV treatment and you will be given a second test at the clinic to confirm this result. Sometimes we cannot clearly tell if the results are negative or positive. We will then refer you to a clinic specialising in HIV treatment and you will be given a second test at the clinic to confirm this result.

Urine sample

You will be requested to donate a small urine sample (20ml) as instructed by the field worker in a clean toilet facility. This will be used to test if your kidneys are working well.

Phenotype and DNA sample data management

The scientists who do the research will generate information (data) which will be placed in databases. Some of the information in the databases will be shared with other scientists, possibly around the world. These scientists will not have access to your name and therefore the information will not be linked back to you. It is possible that we may send small amounts of the DNA out of the country where tests will be done that we cannot easily do in South Africa, but that would give us valuable information for our studies. All findings from this study will be reported for the group and so no other persons will know the results for the single individual.

Benefits

The discoveries that come from the studies on your DNA will not be of direct benefit to you and will not be communicated back to you. The discoveries may lead to information that will help us in the future to diagnose disease, understand who is most likely to get ill and how

different people behave when they are given medication. The DNA belongs to you and we are the keepers of the DNA. You may withdraw your sample at any time.

Risks

Since the DNA of every person is different, it is possible that if someone tested your blood and compared it to the data in the databases, they could conclude that the two samples come from the same person.

Should you require further information or wish to withdraw from the study, you may contact us at:

Project coordinator:

Professor Michele Ramsay
Division of Human Genetics
NHLS
PO Box 1038
Johannesburg
2000
Tel: 011 489 9214
Michele.ramsay@nhls.ac.za
(Will also provide project co-ordinator details)

Ethics approval certificate number:
Institution:
Date issued:

2. AWI-Gen Informed Consent

H3A Study ID:.....

AWI-Gen H3Africa

Genomic and Environmental Risk Factors for Cardiometabolic Disease in Africans

INFORMED CONSENT – Trait Association Study

Collection of blood samples for DNA extraction, storage and genomic studies, and urine collection for metabolic studies

Name: _____

Date of birth: d d / m m / y y y y Place of birth: _____

The information around the blood sample taken from me and the DNA that will be extracted from the blood is clear and the purpose of consent is for me to inform the study what they can or cannot do with these samples.

I understand that all procedure/tests on the stored blood and DNA samples will be approved by the Human Research Ethics Committee of the University of the Witwatersrand.	Yes ☐	No ☐
I am in agreement that my DNA may be stored and used for the purposes described above.	Yes ☐	No ☐
I am in agreement that the data generated from my DNA may be made available as stated above.	Yes ☐	No ☐
I am in agreement that the information I have supplied in the list of questions and the information from the tests and measurements taken from me may be used as stated above.	Yes ☐	No ☐
I agree that a small bit of my DNA may be sent out of the country if the research cannot easily be done in South Africa.	Yes ☐	No ☐
I agree that an portion of my DNA may be stored in a repository (laboratory) according to the stipulation of the H3Africa initiative and that some data may be stored in a database as stipulated and that these may be shared according to the processes and procedures of the H3Africa initiative by using my study code or another code that de-identifies my sample and data.	Yes ☐	No ☐
I understand that every time a new study is done on my DNA, permission will be obtained from the ethics committee for the study to make sure that it is used only for the purposes stated above. (check if this is appropriate in your country)	Yes ☐	No ☐
I understand that I will not benefit directly from the research done on my DNA.	Yes ☐	No ☐
I understand that I may withdraw from the study at any time.	Yes ☐	No ☐

RESEARCH ASSISTANT	Printed Name		
	Signature/Mark or Thumbprint	Date d d / m m / y y y y	Time h h : m m
PARTICIPANT	Printed Name		
	Signature/Mark or Thumbprint	Date d d / m m / y y y y	Time h h : m m
WITNESS (if applicable)	Printed Name		
	Signature/Mark or Thumbprint	Date d d / m m / y y y y	Time h h : m m

3. Clinician survey Invitation Letter

Good day,

RE: CPD accredited survey: Utility of precision medicine approaches in South Africa

My name is Michelle Kamp, and I am a PhD student of Prof Michèle Ramsay in the Division of Human Genetics and the Sydney Brenner Institute of Molecular Bioscience (SBIMB), Faculty of Health Sciences, Wits University.

Our research study, The development and value assessment of a precision medicine-based cardiovascular disease risk score in a South African setting, aims to develop and critically assess the value of a cardiovascular disease risk stratification score that assesses both the genetic and behavioural risk factors for cardiovascular disease in a South African setting.

As physicians are critical to the effective implementation of precision medicine, we invite you to participate in a survey assessing Physicians' perceptions of incorporating a precision medicine-based cardiovascular disease risk stratification approach in South Africa. The survey will take approximately 30 minutes to complete and can be found via the link below:

Survey link: https://redcap.link/precision_medicine_south_africa

A 15-minute educational video linked to the survey will provide an introduction to precision medicine, risk stratification, and polygenic risk scores and discusses the clinical utility of precision medicine-based approaches.

Given that this activity contributes to your further education and provides an opportunity to critically assess novel treatment approaches that are relevant to current practice, 1 Continued Professional Development (CPD) point will be awarded for your participation.

The insights generated from the questionnaire responses will not be of direct benefit to you and will not be communicated back to you. However, the results may lead to important information that may help the future effective implementation of cardiovascular disease risk stratification programmes or other genomic-based interventions in South Africa. The results of this survey will be published in a peer-review journal.

You may withdraw at any time, up to the point of submitting your answers, or not answer any question if you do not want to. The survey completed by you will remain confidential. All identifying information, and the information you give will be held securely and not be disclosed to anyone else.

Please note that the survey is restricted to physicians, and to those whose primary affiliation is at one of the following four academic hospitals:

- Chris Hani Baragwanath Hospital,
- Charlotte Maxeke Hospital,
- Helen Joseph Hospital,
- Rahima Moosa Mother and Child Hospital

Should you require further information or wish to withdraw from the study, you may contact us at:

Project Principal Investigator:

Michelle Kamp
SBIMB
9 Jubilee Road
Parktown, Johannesburg
Tel: 011 717 6635
Email: michelleukamp@gmail.com

Project Supervisor

Prof. Michèle Ramsay
SBIMB
9 Jubilee Road
Parktown, Johannesburg
Tel: 011 717 6635
Email: Michele.ramsay@wits.ac.za

This study was approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg in June 2021, Ethics approval certificate number: M210355

If you have any concerns or complaints regarding the ethical procedures of this study, please contact the Committee Chairperson, Professor Clement Penny, telephone +27 (0) 11 717 2301, e-mail Clement.Penny@wits.ac.za. The telephone numbers for the Committee secretariat are +27 (0) 11 717 2700/1234 and the e-mail addresses are Zanele.Ndlovu@wits.ac.za and Rhulani.Mukansi@wits.ac.za

Yours sincerely,

Ms Michelle Kamp

4. Clinician survey Information Sheet

ELECTRONIC STUDY INFORMATION DOCUMENT

Study title: The development and value assessment of a precision medicine-based cardiovascular disease risk score in a South African setting

Dear Potential Participant,

My name is Michelle Kamp, and I am a PhD student in the Division of Human Genetics and the Sydney Brenner Institute of Molecular Bioscience (SBIMB), Faculty of Health Sciences, University of the Witwatersrand in Johannesburg, South Africa. I would like to invite you to participate in our research study.

This Information sheet will provide you with details about the informed consent that is required if you are willing to participate in this study. It also describes how we will collect, use and store the information you will provide if you agree to be part of the study.

This study is nested within a larger study, the Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen), which aims to study the genetic and environmental factors that contribute to cardiometabolic disease in African populations.

The PhD study specifically aims develop and critically assess the value of a cardiovascular disease risk stratification score that assesses both the genetic and behavioural risk factors for cardiovascular disease in a South African setting. A key component in assessing the value of the risk stratification score is understanding its potential clinical utility.

The score's clinical utility will be explored quantitatively using an electronic survey administered to physicians currently posted at one of the following four public academic hospitals in Johannesburg: Chris Hani Baragwanath Hospital, Charlotte Maxeke Hospital, Helen Joseph Hospital, and the Rahima Moosa Mother and Child Hospital.

The purpose of the questionnaire is two-fold. Firstly, to better understand clinicians' current perceptions of a CVD risk stratification score that assess both genetic and environmental risk, and secondly, to identify and the potential barriers to the adoption of such a tool in the South African public healthcare setting.

Your Study Code

If you agree to participate in this study, you will be given a unique Study Code. This Study Code will be used to label all the information you provide us and to ensure that this information remains anonymous. It is important to understand that the database used to store the information that you provide us will not include your name, identity number of date of birth or other personally identifiable information, and therefore the information cannot be linked back to you.

Questionnaire

A weblink will be sent to you via email requesting you to complete an online questionnaire. You can skip any question that makes you uncomfortable. The questionnaire may be broken up into several short sessions and should take approximately 45 minutes in total to complete.

Where will your information (data) be used and stored?

The questionnaire responses (data) will be placed and stored in a database located at the University of the Witwatersrand. The information in the database will not be shared with scientists not related to this specific study.

Your data will be password protected to prevent unauthorized access and will not contain any personal information that you have provided us that could allow you to be identified, such as your name, identity number, or date of birth.

Your right to withdraw from the study

Your participation is completely voluntary, and if you wish not to participate then there is no penalty or loss of benefits in any way.

Benefits

As the electronic questionnaire will include educational materials, specifically a video providing information on genetics, precision medicine, polygenic risk scores, disease risk stratification and complex diseases, the completion of the questionnaire will contribute one Continued Professional Development (CPD) point, please provide your HPCSA number in the survey where directed.

The insights generated from the questionnaire responses will not be of direct benefit to you and will not be communicated back to you. However, the results may lead to important information that may help the future effective implementation of cardiovascular disease risk stratification programmes or other genomic-based interventions in the South African public health setting.

Risks

There are no risks associated with this study.

What will happen when the study is over?

We plan to publish the results in a scientific journal. As the data will be anonymized and the results aggregated, your name will not be used in any publication of the results.

Thank you for reading this Study Information Sheet.

Should you require further information or wish to withdraw from the study, you may contact us at:

Project Principal Investigator:
Michelle Kamp
SBIMB
9 Jubilee Road
Parktown, Johannesburg
Tel: 011 717 6635
Email: michelleukamp@gmail.com

Project Supervisor
Prof. Michèle Ramsay
SBIMB
9 Jubilee Road
Parktown, Johannesburg
Tel: 011 717 6635
Email: Michele.ramsay@wits.ac.za

Contact details of HREC administrator and chair:
This study has been approved by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand, Johannesburg ("Committee"). A principal function of this Committee is to safeguard the rights and dignity of all human subjects who agree to participate in a research project and the integrity of the research.

If you have any concern over the way the study is being conducted, please contact the Chairperson of this Committee who is Professor Clement Penny, who may be contacted on telephone number +27 (0) 11 717 2301, or by e-mail on Clement.Penny@wits.ac.za. The telephone numbers for the Committee secretariat are +27 (0) 11 717 2700/1234 and the e-mail addresses are Zanele.Ndlovu@wits.ac.za and Rhulani.Mukansi@wits.ac.za

Ethics approval certificate number: M210355
Institution: University of the Witwatersrand, Johannesburg
Date issued: 10/06/2021

5. Clinician survey Informed Consent

CONFIRMATION OF CONSENT

Please confirm that your participation in this study has been explained to you (via the participant information sheet) and that you understand the importance for you to consent to the following items:

	Yes	No
I understand that all procedures have been approved by the Human Research Ethics Committee of the University of the Witwatersrand and that my data will have a unique Study Code to protect my identity. No personal information such as name, date of birth, or identity number will be linked to my code.	○	○
I agree to complete the questionnaire. I understand that I am not obliged to answer any question or undergo any procedure that makes me feel uncomfortable.	○	○
I understand that the information collected from me will be stored in secure databases and used as described in the information sheet.	○	○
I understand that I will not benefit directly from the research done on the information (data) I provide, but that I am eligible for HPCSA CPD credit.	○	○

APPENDIX C: ETHICS CLEARANCE CERTIFICATES

1. PhD study Ethics Clearance Certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Mrs Michelle Kamp

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M210355

NAME: (Principal Investigator)	Mrs Michelle Kamp
DEPARTMENT:	Sydney Brenner Institute for Molecular Bioscience
PROJECT TITLE:	The development and value assessment of a precision medicine-based cardiovascular disease risk score in a South African setting
DATE CONSIDERED:	26/03/2021
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Prof Michelle Ramsay and Prof Cathryn Lewis
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	10/06/2021

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. 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 resubmit the application to the Committee. **I agree to submit a yearly progress report.** The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in **March** and will therefore be due in the month of **March** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

M. Kamp — Principal Investigator Signature

30/03/2021 — Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

2. AWI-Gen study Ethics Clearance Certificate

AWI-Gen Phase 1 (2012)

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Professor Michele Ramsay

CLEARANCE CERTIFICATE	**M121029**
PROJECT	WITS-INDEPTH Partnership: Genomic and Environment Risk Factors for Cardiometabolic Disease in Africans
INVESTIGATORS	Professor Michele Ramsay.
DEPARTMENT	School of Pathology/Division of Human Genetics
DATE CONSIDERED	26/10/2012
DECISION OF THE COMMITTEE*	Approved unconditionally

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 21/11/2012 **CHAIRPERSON** .....
(Professor PE Cleaton-Jones)

*Guidelines for written 'informed consent' attached where applicable
cc: Supervisor :

DECLARATION OF INVESTIGATOR(S)
To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University.
I/~~We~~ fully understand the conditions under which I am/~~we~~ are authorized to carry out the abovementioned research and I/~~we~~ guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/~~we~~ undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**
PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES...

M. Ramsay 11 April 2013

3. AWI-Gen Phase 1 renewal (2017)

R14/49 Prof Michele Ramsay

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M170880

NAME: (Principal Investigator)	Prof Michele Ramsay
DEPARTMENT:	School of Pathology/Division of Human Genetics National Health Laboratory Services
PROJECT TITLE:	Wits: INDEPTH Partnership - Genomic and Enviromental Risk Factors for Cardiometabolic Disease in African Populations
DATE CONSIDERED:	03/10/2014(Initial Approval 24/08/2017)
DECISION:	Approved unconditionally
CONDITIONS:	Renewal for 5 Years Valid for the Period 01 September 2017 - 30 September 2022 (Previously M121029)
SUPERVISOR:	
APPROVED BY:	Professor CB. Penny Co-Chairperson, HREC (Medical)
DATE OF APPROVAL:	12/09/2017

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 10004,10th floor, Senate House/3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand. I/We fully understand the conditions under which I am/we are authorised 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 resubmit to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. in this case, the study was initially review September and will therefore be due in the month of September each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

M. Ramsay
Principal Investigator Signature

Date 18 September 2017

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

4. AWI-Gen Phase 1 and Phase 2 combined renewal (2022)

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).

M. Ramsay
Signature of Principal Investigator

28 November 2022
Date

1

2

6. Clinician survey Ethics Clearance Certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Mrs Michelle Kamp

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M210355

NAME: (Principal Investigator)	Mrs Michelle Kamp
DEPARTMENT:	Sydney Brenner Institute for Molecular Bioscience
PROJECT TITLE:	The development and value assessment of a precision medicine-based cardiovascular disease risk score in a South African setting
DATE CONSIDERED:	26/03/2021
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Prof Michelle Ramsay and Prof Cathryn Lewis
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	10/06/2021

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. 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 resubmit the application to the Committee. I **agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in **March** and will therefore be due in the month of **March** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

M. Kamp — Principal Investigator Signature

30/03/2021 — Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

7. Clinician survey Ethics Amendment

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG	HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

2021/10/29

Ms M Kamp
Faculty of Health Sciences
Sydney Brenner Institute for Molecular Bioscience
Medical School
University

Sent by e-mail to: michelleukamp@gmail.com

Dear Ms Kamp

Re: **Protocol Ref No:** M210355
Protocol Title: *The development and value assessment of a precision medicine-based cardiovascular disease risk score in a South African setting*
Principal Investigator: Ms M Kamp

Thank you for your letter of 2021/10/19.

I confirm that we have noted and approve of your proposal to widen your participant base in the manner described.

Thank you for keeping us informed.

Yours Sincerely

J. Burns

.....
Mr I Burns
For the Human Research Ethics Committee (Medical)

CB Penny

.....
Dr CB Penny, Chairperson, Human Research Ethics Committee (Medical)

APPENDIX D: CVD RISK CALCULATORS

Most guidelines rely on the use of a prediction model for the management and prevention of CVD. However, there is no single tool recommended for use. Available risk tools differ in their end points and the risk factors they consider in the risk prediction.

Table 1: Established risk calculators for Cardiovascular Disease

Model	Prediction	Risk factors	Outcome	Population	Guidelines
ASSIGN	10-year risk estimate for CVD	Age, sex, TC, SBP, HDL, smoking, diabetes status, family history of myocardial infarction, area-based deprivation index	Fatal and non-fatal CVD events	European (UK)	Scottish Intercollegiate Guidelines Network http://www.assign-score.com/
Framingham	Risk estimate for CVD (or component of CVD) over a fixed time or lifetime	Age, sex, TC, SBP, HDL, smoking, diabetes status	Fatal and non-fatal CVD events: CHD, myocardial infarction & stroke	European (Australia, USA, South Africa)	Australian Guidelines for the management of absolute CVD risk ACA/AHA Guidelines on the Primary Prevention of CVD https://www.framinghamheartstudy.org/fhs-about/
Pooled Cohort Risk Equations (PCE)	10-year absolute risk estimate for Atherosclerotic Cardiovascular Disease	Age, sex, ethnicity, TC, HDL, SBP, smoking, diabetes status, use of antihypertensive medication	Non-fatal myocardial infarction, non-fatal stroke & fatal CVD event	European (USA) African-American	ACA/AHA Guideline on the Primary Prevention of CVD http://tools.acc.org/ASCVD-Risk-Estimator-Plus/#!/calculate/estimate/
QRISK3	10-year risk estimate for CVD in women and men	QRISK2 + Chronic kidney disease (stage 3), SBP, migraine, corticosteroids, systemic lupus erythematosus, atypical antipsychotics, severe mental	Fatal CVD	European (UK)	NICE Guideline: CVD risk assessment & the modification of blood lipids for the primary & secondary prevention of CVD

		illness, erectile dysfunction			PHE NHS Health Check: best practice guidance https://qrisk.org/
Reynolds	10-year risk estimate for CVD (targeted to people without diabetes or previous CVD)	Age, smoking, SBP, TC, HDL, hsCRP, family history of myocardial infarction (<60), HgBA1c (women only)	Acute myocardial infarction, CVD death (ischemic) or revascularization treatment (coronary),	European (USA)	2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the ACC/AHA Task Force on Clinical Practice Guidelines http://www.reynoldsriskscore.org/
Systematic Coronary Risk Evaluation (SCORE) algorithm	10-year risk estimate for CVD mortality	Age (used as a temporal variable), TC, TC/HDL ratio, SBP, smoking	Fatal CHD, CVD, and non-CVD (congestive heart failure, peripheral vascular disease)	European (European)	ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure http://www.heartscore.org/en_GB

ACA – American College of Cardiology, AHA – American Heart Association, ESC – European Society of Cardiology, HDL – High density lipoprotein, NHS – National Health Service (UK), PHE – Public Health England SBP – Systolic blood pressure, TC – Total cholesterol

APPENDIX E: FRAMINGHAM RISK FACTOR WEIGHTINGS

The factors of the Framingham risk stratification model for CVD utilises seven risk factors. The factors by which risk factor is multiplied (by sex) is in Table 5.

Table 2: Factor associated with each CVD risk factor (by sex) for the Framingham 10 Year Risk of General Cardiovascular Disease

Factor	Female	Male
Age	2.32888	3.06117
Total cholesterol	1.20904	1.12370
HDL	-0.70833	-0.93263
Non-treated SBP	2.76157	1.93303
Treated SBP	2.82263	1.99881
Smoker	0.052873	0.65451
Diabetes	0.69154	0.57367

APPENDIX F: PUBLICATION SUPPLEMENTARY MATERIALS

Publication 1: no supplementary materials

Publication 2: supplementary materials

S1. GWAS Catalog study assession numbers

Study assession number	Reported trait	Sample	Ancestry
GCST009042	Total cholesterol levels	14 126	Sub-Saharan African
GCST009043	LDL cholesterol levels	14 126	Sub-Saharan African
GCST009044	HDL cholesterol levels	14 126	Sub-Saharan African
GCST009045	Triglyceride levels	14 126	Sub-Saharan African
GCST009048	Serum albumin levels	14 126	Sub-Saharan African
GCST009051	Bilirubin levels	14 126	Sub-Saharan African
GCST009052	Diastolic blood pressure	14 126	Sub-Saharan African
GCST009053	Systolic blood pressure	14 126	Sub-Saharan African
GCST009055	Height	14 126	Sub-Saharan African
GCST009056	Weight	14 126	Sub-Saharan African
GCST009057	Body mass index	14 126	Sub-Saharan African
GCST009058	Waist circumference	14 126	Sub-Saharan African
GCST009059	Hip circumference	14 126	Sub-Saharan African
GCST009060	Waist-hip ratio	14 126	Sub-Saharan African

S2. Non genetic factor definitions

Dietary factors that represent key behaviours related to lifestyle were explored using a series of questions guided by the WHO Steps Instrument about the consumption of fruit and vegetables, vendor-supplied meals, bread, and sugar-sweetened drinks including fruit juice.

Current smoking status was obtained from *"Yes", "No"* responses to the following question, "Do you currently smoke tobacco?". Similarly, alcohol consumption status was determined from *"Yes", "No"* responses to the following question "Are you a current alcohol consumer?" Those who preferred not to answer or did not know, were excluded.

The Global Physical Activity Questionnaire (GPAQ) was used to obtain self-reported physical activity. Total moderate-vigorous physical activity (MVPA) in minutes per week was calculated from the accumulation of occupation, travel-related and leisure time physical activity. Sitting time (minutes/week) is used as a proxy for sedentary behaviour (Ref)..

The maximum observations for each non-genetic factor per site is shown in

	Agincourt (N = 2486)	**Dikgale (N = 1399)**	**Nairobi (N = 2003)**	**Nanoro (N = 2097)**	**Navrongo (N = 2016)**	**Total (N = 10001)**
Age	2486	1399	2003	2097	2016	10001
Sex	2486	1399	2003	2097	2016	10001
Smoking	2486	1399	2002	2096	2014	9997
Alcohol	2486	1397	2001	2095	2012	9991
Sugardrinks	2465	1398	2003	2090	2015	9971
Juice	2470	1397	2002	2078	2015	9962
Sleep per night	2386	1396	1997	2089	1978	9846
Physical activity	2469	1393	2000	2097	1995	9954
Fruit per week	2467	1398	2003	2090	2016	9974
Veg per week	2374	1397	2003	1627	2016	9417
Bread per week	2266	1397	2003	745	2015	8426

S3. Disease outcome definitions

CVD was defined as present if the participant reported having had a heart attack, stroke, or transient ischaemic attack. Participants previously diagnosed with congestive heart failure or angina were also classified as having CVD. No data were available for angina or heart attack prevalence for the Soweto men, so CVD prevalence was calculated using the remaining data.

Diabetes was defined using the WHO criteria, which are the presence of a previous diagnosis of diabetes by a healthcare professional or fasting blood glucose≥7 mmol/L or random glucose≥11.1 mmol/L, or on diabetes medication at the time of recruitment.

Dyslipidaemia was defined as the presence of one of the following: total cholesterol (TC)≥5.0 mmol/L, or high-density lipoprotein cholesterol (HDL-C)<1.0 mmol/L for men and <1.3 mmol/L for women, or low-density lipoprotein cholesterol (LDL-C) ≥3. 0 mmol/L, or triglycerides (TG)≥1.7 mmol/L. The Randox Daytona Plus (Randox Laboratories, UK) autoanalyser was used to analyse HDL-C, TG and TC on fasting venous blood samples. LDL-C was calculated using the Friedewald equation.

Hypertension was defined as systolic blood pressure≥140 mm Hg and/or diastolic blood pressure≥90 mm Hg, in line with the seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure, or if the participant was taking hypertension medication.

Obesity was defined as having a Body Mass Index (BMI) of 30 kg/m² or higher, in accordance with the guidelines set by the World Health Organization. BMI is calculated by dividing a person's weight in kilograms by the square of their height in meters.

1 S4. Association testing: Significant within trait associations of derived PGS

Phenotype	PGS	pT	R2	R2 %	BETA	SE	P
BMI	BMI	$1x10^{-04}$	0.0007	0.07%	0.027	0.008	$1.25X10^{-03}$
BMI	BMI	0.01	0.0021	0.21%	0.052	0.009	$1.95X10^{-08}$
BMI	BMI	0.1	0.0017	0.17%	0.055	0.011	$4.51X10^{-07}$
BMI	BMI	0.2	0.0017	0.17%	0.057	0.011	$5.29X10^{-07}$
BMI	BMI	0.3	0.0019	0.19%	0.062	0.012	$6.49X10^{-08}$
BMI	BMI	0.4	0.0018	0.18%	0.061	0.012	$1.45X10^{-07}$
BMI	BMI	0.5	0.0020	0.20%	0.064	0.012	$4.91X10^{-08}$
BMI	BMI	1	0.0020	0.20%	0.065	0.012	$3.55X10^{-08}$
DBP	DBP	0.1	0.0008	0.08%	0.031	0.010	$2.74X10^{-03}$
DBP	DBP	0.2	0.0008	0.08%	0.031	0.010	$2.84X10^{-03}$
DBP	DBP	0.3	0.0008	0.08%	0.033	0.011	$1.90X10^{-03}$
DBP	DBP	0.4	0.0009	0.09%	0.034	0.011	$1.10X10^{-03}$
DBP	DBP	0.5	0.0009	0.09%	0.034	0.010	$1.28X10^{-03}$
DBP	DBP	1	0.0008	0.08%	0.032	0.011	$2.35X10^{-03}$
HDL	HDL	$1x10^{-08}$	0.0182	1.82%	0.136	0.010	$5.91X10^{-45}$
HDL	HDL	$1x10^{-06}$	0.0191	1.91%	0.139	0.010	$3.88X10^{-47}$
HDL	HDL	$1x10^{-04}$	0.0101	1.01%	0.104	0.010	$1.16X10^{-25}$
HDL	HDL	0.01	0.0018	0.18%	0.047	0.011	$1.09X10^{-05}$
Height	Height	$1x10^{-06}$	0.0007	0.07%	0.027	0.007	$2.29X10^{-04}$
Height	Height	$1 x10^{-04}$	0.0006	0.06%	0.025	0.007	$5.53X10^{-04}$
Height	Height	0.01	0.0030	0.30%	0.058	0.008	$3.50X10^{-14}$
Height	Height	0.1	0.0044	0.44%	0.077	0.008	$5.51X10^{-20}$
Height	Height	0.2	0.0044	0.44%	0.079	0.009	$4.70X10^{-20}$
Height	Height	0.3	0.0042	0.42%	0.079	0.009	$2.54X10^{-19}$
Height	Height	0.4	0.0043	0.43%	0.080	0.009	$6.90X10^{-20}$
Height	Height	0.5	0.0043	0.43%	0.081	0.009	$8.03X10^{-20}$
Height	Height	1	0.0043	0.43%	0.080	0.009	$9.96X10^{-20}$
Hip circum	Hip circum	0.01	0.0014	0.14%	0.039	0.008	$1.94X10^{-06}$
Hip circum	Hip circum	0.1	0.0014	0.14%	0.039	0.008	$3.36X10^{-06}$
Hip circum	Hip circum	0.2	0.0014	0.14%	0.039	0.009	$3.98X10^{-06}$
Hip circum	Hip circum	0.3	0.0015	0.15%	0.041	0.009	$1.71X10^{-06}$
Hip circum	Hip circum	0.4	0.0016	0.16%	0.043	0.009	$4.67X10^{-07}$
Hip circum	Hip circum	0.5	0.0016	0.16%	0.043	0.009	$5.28X10^{-07}$
Hip circum	Hip circum	1	0.0015	0.15%	0.042	0.009	$1.31X10^{-06}$
LDL	LDL	$1x10^{-08}$	0.0677	6.77%	0.261	0.009	$1.86X10^{-173}$
LDL	LDL	$1x10^{-06}$	0.0653	6.53%	0.256	0.009	$3.86X10^{-167}$
LDL	LDL	$1x10^{-04}$	0.0511	5.11%	0.227	0.009	$4.20X10^{-130}$
LDL	LDL	0.01	0.0099	0.99%	0.107	0.010	$5.77X10^{-26}$
LDL	LDL	0.1	0.0036	0.36%	0.075	0.012	$2.55X10^{-10}$
LDL	LDL	0.2	0.0029	0.29%	0.071	0.012	$1.09X10^{-08}$
LDL	LDL	0.3	0.0021	0.21%	0.061	0.013	$1.20X10^{-06}$
LDL	LDL	0.4	0.0020	0.20%	0.061	0.013	$1.87X10^{-06}$
LDL	LDL	0.5	0.0019	0.19%	0.059	0.013	$3.92X10^{-06}$

LDL	LDL	1	0.0020	0.20%	0.060	0.013	$2.87X10^{-06}$
SBP	SBP	0.3	0.0007	0.07%	0.055	0.018	$3.07X10^{-03}$
SBP	SBP	0.4	0.0008	0.08%	0.059	0.019	$1.77X10^{-03}$
SBP	SBP	0.5	0.0007	0.07%	0.057	0.019	$2.62X10^{-03}$
SBP	SBP	1	0.0008	0.08%	0.061	0.019	$1.31X10^{-03}$
TC	TC	$1x10^{-08}$	0.0284	2.84%	0.170	0.009	$5.38X10^{-80}$
TC	TC	$1x10^{-06}$	0.0307	3.07%	0.178	0.009	$1.65X10^{-86}$
TC	TC	$1x10^{-04}$	0.0196	1.96%	0.141	0.009	$1.07X10^{-55}$
TC	TC	0.01	0.0034	0.34%	0.060	0.009	$5.52X10^{-11}$
TC	TC	0.1	0.0014	0.14%	0.041	0.010	$3.88X10^{-05}$
TC	TC	0.2	0.0011	0.11%	0.039	0.010	$1.96X10^{-04}$
TC	TC	0.3	0.0011	0.11%	0.039	0.010	$1.64X10^{-04}$
TC	TC	0.4	0.0011	0.11%	0.039	0.011	$2.45X10^{-04}$
TC	TC	0.5	0.0012	0.12%	0.040	0.011	$1.20X10^{-04}$
TC	TC	1	0.0011	0.11%	0.039	0.010	$1.81X10^{-04}$
TG	TG	$1x10^{-08}$	0.0044	0.44%	0.067	0.009	$9.83X10^{-13}$
TG	TG	$1x10^{-06}$	0.0057	0.57%	0.076	0.009	$3.34X10^{-16}$
TG	TG	$1x10^{-04}$	0.0011	0.11%	0.036	0.010	$2.69X10^{-04}$
TG	TG	0.1	0.0008	0.08%	0.079	0.026	$2.61X10^{-03}$
Waist circ	Waist circ	0.01	0.0016	0.16%	0.042	0.009	$2.47X10^{-06}$
Waist circ	Waist circ	0.1	0.0013	0.13%	0.036	0.009	$2.32X10^{-05}$
Waist circ	Waist circ	0.2	0.0013	0.13%	0.037	0.009	$1.76X10^{-05}$
Waist circ	Waist circ	0.3	0.0013	0.13%	0.037	0.009	$1.96X10^{-05}$
Waist circ	Waist circ	0.4	0.0013	0.13%	0.036	0.009	$2.92X10^{-05}$
Waist circ	Waist circ	0.5	0.0013	0.13%	0.037	0.009	$2.12X10^{-05}$
Waist circ	Waist circ	1	0.0012	0.12%	0.036	0.009	$3.35X10^{-05}$
Weight	Weight	0.01	0.0018	0.18%	0.043	0.009	$1.40X10^{-06}$
Weight	Weight	0.1	0.0022	0.22%	0.051	0.009	$6.59X10^{-08}$
Weight	Weight	0.2	0.0028	0.28%	0.058	0.010	$1.33X10^{-09}$
Weight	Weight	0.3	0.0026	0.26%	0.056	0.010	$4.02X10^{-09}$
Weight	Weight	0.4	0.0028	0.28%	0.058	0.010	$1.24X10^{-09}$
Weight	Weight	0.5	0.0030	0.30%	0.060	0.010	$4.61X10^{-10}$
Weight	Weight	1	0.0031	0.31%	0.061	0.010	$2.26X10^{-10}$

*BMI – Body Mass Index, DBP – diastolic blood pressure, HDL – high-density lipoprotein, Hip circum – Hip circumference, LDL – low-density lipoprotein , SBP – systolic blood pressure, TC – total cholesterol, TG – triglycerides, Waist circum – Waist circumference; ; p-*value t*hresholds: '***' = p ≤ 0.0001, '**' = p ≤ 0.001, '*' = p ≤ 0.01*

S5. Association testing: Significant cross-trait associations of derived PGS

Phenotype	PGS	pT	R2	R2 %	BETA	SE	P	N
BMI	HDL	0.01	0.0007	0.07%	-0.029	0.009	$1.24X10^{-03**}$	10546
BMI	Hip circum	0.01	0.0013	0.13%	0.037	0.008	$8.82X10^{-06***}$	10546
BMI	Hip circum	0.1	0.0016	0.16%	0.042	0.009	$1.32X10^{-06***}$	10546
BMI	Hip circum	0.2	0.0015	0.15%	0.042	0.009	$1.53X10^{-06***}$	10546
BMI	Hip circum	0.3	0.0016	0.16%	0.043	0.009	$7.38X10^{-07***}$	10546
BMI	Hip circum	0.4	0.0018	0.18%	0.046	0.009	$1.46X10^{-07***}$	10546
BMI	Hip circum	0.5	0.0018	0.18%	0.046	0.009	$1.37X10^{-07***}$	10546
BMI	Hip circum	1	0.0018	0.18%	0.045	0.009	$2.74X10^{-07***}$	10546
BMI	Waist circum	$1x10^{-04}$	0.0014	0.14%	0.037	0.008	$4.50X10^{-06***}$	10546
BMI	Waist circum	0.01	0.0017	0.17%	0.043	0.009	$5.47X10^{-07***}$	10546
BMI	Waist circum	0.1	0.0012	0.12%	0.034	0.008	$2.68X10^{-05***}$	10546
BMI	Waist circum	0.2	0.0014	0.14%	0.038	0.008	$3.07X10^{-06***}$	10546
BMI	Waist circum	0.3	0.0014	0.14%	0.037	0.008	$5.93X10^{-06***}$	10546
BMI	Waist circum	0.4	0.0013	0.13%	0.037	0.008	$6.85X10^{-06***}$	10546
BMI	Waist circum	0.5	0.0014	0.14%	0.038	0.008	$4.68X10^{-06***}$	10546
BMI	Waist circum	1	0.0014	0.14%	0.038	0.008	$4.84X10^{-06***}$	10546
BMI	Weight	0.01	0.0010	0.10%	0.033	0.008	$1.01X10^{-04***}$	10546
BMI	Weight	0.1	0.0015	0.15%	0.041	0.009	$2.60X10^{-06***}$	10546
BMI	Weight	0.2	0.0018	0.18%	0.047	0.009	$1.73X10^{-07***}$	10546
BMI	Weight	0.3	0.0017	0.17%	0.045	0.009	$3.92X10^{-07***}$	10546
BMI	Weight	0.4	0.0018	0.18%	0.047	0.009	$1.90X10^{-07***}$	10546
BMI	Weight	0.5	0.0019	0.19%	0.048	0.009	$8.25X10^{-08***}$	10546
BMI	Weight	1	0.0020	0.20%	0.049	0.009	$5.38X10^{-08**}$	10546
DBP	SBP	0.4	0.0008	0.08%	0.058	0.019	$2.41X10^{-03**}$	10540
DBP	SBP	0.5	0.0008	0.08%	0.058	0.019	$2.94X10^{-03**}$	10540
DBP	SBP	1	0.0009	0.09%	0.062	0.019	$1.40X10^{-03**}$	10540
DBP	WH ratio	0.01	0.0011	0.11%	0.039	0.011	$2.43X10^{-04***}$	10540
HDL	LDL	$1x10^{-08}$	0.0009	0.09%	-0.030	0.010	$2.18X10^{-03**}$	10463
Height	Weight	0.01	0.0005	0.05%	0.023	0.007	$1.80X10^{-03**}$	10549
Height	Weight	0.2	0.0005	0.05%	0.024	0.008	$2.92X10^{-03**}$	10549
Height	Weight	0.3	0.0004	0.04%	0.023	0.008	$3.51X10^{-03**}$	10549
Height	Weight	0.4	0.0005	0.05%	0.024	0.008	$2.56X10^{-03**}$	10549
Height	Weight	0.5	0.0005	0.05%	0.024	0.008	$2.39X10^{-03**}$	10549
Height	Weight	1	0.0005	0.05%	0.024	0.008	$2.08X10^{-03**}$	10549
Hip circum	BMI	$1x10^{-04}$	0.0008	0.08%	0.029	0.008	$3.47X10^{-04***}$	10157
Hip circum	BMI	0.01	0.0017	0.17%	0.047	0.009	$2.09X10^{-07***}$	10157
Hip circum	BMI	0.1	0.0012	0.12%	0.047	0.011	$1.04X10^{-05***}$	10157
Hip circum	BMI	0.2	0.0013	0.13%	0.050	0.011	$8.19X10^{-06***}$	10157
Hip circum	BMI	0.3	0.0014	0.14%	0.053	0.011	$2.32X10^{-06***}$	10157
Hip circum	BMI	0.4	0.0014	0.14%	0.053	0.011	$2.88X10^{-06***}$	10157
Hip circum	BMI	0.5	0.0015	0.15%	0.057	0.012	$8.81X10^{-07***}$	10157
Hip circum	BMI	1	0.0015	0.15%	0.056	0.012	$1.01X10^{-06***}$	10157
Hip circum	HDL	$1x10^{-04}$	0.0007	0.07%	-0.027	0.008	$9.90X10^{-04***}$	10157

Hip circum	HDL	0.01	0.0007	0.07%	-0.029	0.009	$9.14X10^{-04***}$	10157
Hip circum	HDL	0.4	0.0005	0.05%	-0.032	0.011	$3.33X10^{-03**}$	10157
Hip circum	SBP	$1x10^{-04}$	0.0006	0.06%	0.024	0.008	$2.47X10^{-03**}$	10157
Hip circum	Waist circum	$1x10^{-04}$	0.0015	0.15%	0.039	0.008	$9.25X10^{-07***}$	10157
Hip circum	Waist circum	0.01	0.0018	0.18%	0.044	0.008	$1.23X10^{-07***}$	10157
Hip circum	Waist circum	0.1	0.0014	0.14%	0.037	0.008	$3.87X10^{-06***}$	10157
Hip circum	Waist circum	0.2	0.0015	0.15%	0.039	0.008	$1.31X10^{-06***}$	10157
Hip circum	Waist circum	0.3	0.0014	0.14%	0.039	0.008	$1.81X10^{-06***}$	10157
Hip circum	Waist circum	0.4	0.0015	0.15%	0.039	0.008	$1.26X10^{-06***}$	10157
Hip circum	Waist circum	0.5	0.0015	0.15%	0.039	0.008	$1.24X10^{-06***}$	10157
Hip circum	Waist circum	1	0.0015	0.15%	0.039	0.008	$1.37X10^{-06***}$	10157
Hip circum	Weight	0.01	0.0010	0.10%	0.033	0.008	$6.93X10^{-05***}$	10157
Hip circum	Weight	0.1	0.0013	0.13%	0.040	0.009	$4.18X10^{-06***}$	10157
Hip circum	Weight	0.2	0.0018	0.18%	0.047	0.009	$9.37X10^{-08***}$	10157
Hip circum	Weight	0.3	0.0018	0.18%	0.046	0.009	$1.19X10^{-07***}$	10157
Hip circum	Weight	0.4	0.0018	0.18%	0.047	0.009	$8.38X10^{-08***}$	10157
Hip circum	Weight	0.5	0.0019	0.19%	0.048	0.009	$4.79X10^{-08***}$	10157
Hip circum	Weight	1	0.0019	0.19%	0.048	0.009	$3.02X10^{-08***}$	10157
LDL	TC	$1x10^{-08}$	0.0682	6.82%	0.262	0.009	$6.23X10^{-175***}$	9496
LDL	TC	$1x10^{-06}$	0.0678	6.78%	0.263	0.009	$6.84X10^{-174***}$	9496
LDL	TC	$1x10^{-04}$	0.0332	3.32%	0.183	0.009	$3.02X10^{-84***}$	9496
LDL	TC	0.01	0.0052	0.52%	0.073	0.010	$2.52X10^{-14***}$	9496
LDL	TC	0.1	0.0018	0.18%	0.048	0.011	$6.14X10^{-06***}$	9496
LDL	TC	0.2	0.0014	0.14%	0.044	0.011	$6.67X10^{-05***}$	9496
LDL	TC	0.3	0.0012	0.12%	0.041	0.011	$2.25X10^{-04***}$	9496
LDL	TC	0.4	0.0012	0.12%	0.041	0.011	$2.16X10^{-04***}$	9496
LDL	TC	0.5	0.0011	0.11%	0.040	0.011	$3.65X10^{-04***}$	9496
LDL	TC	1	0.0010	0.10%	0.038	0.011	$7.04X10^{-04***}$	9496
LDL	HDL	$1x10^{-08}$	0.0135	1.35%	-0.117	0.009	$7.91X10^{-35***}$	9496
LDL	HDL	$1x10^{-06}$	0.0117	1.17%	-0.109	0.009	$1.75X10^{-30***}$	9496
LDL	HDL	$1x10^{-04}$	0.0046	0.46%	-0.070	0.010	$7.92X10^{-13***}$	9496
LDL	Weight	1	0.0008	0.08%	-0.031	0.010	$3.48X10^{-03**}$	9496
SBP	WH ratio	0.01	0.0008	0.08%	0.033	0.010	$1.43X10^{-03**}$	10538
TC	LDL	$1x10^{-08}$	0.0273	2.73%	0.167	0.009	$6.11X10^{-77***}$	10476
TC	LDL	$1x10^{-06}$	0.0273	2.73%	0.167	0.009	$6.73X10^{-77***}$	10476
TC	LDL	$1x10^{-04}$	0.0224	2.24%	0.150	0.009	$2.02X10^{-63***}$	10476
TC	LDL	0.01	0.0055	0.55%	0.079	0.010	$1.60X10^{-16***}$	10476
TC	LDL	0.1	0.0017	0.17%	0.052	0.011	$4.27X10^{-06***}$	10476
TC	LDL	0.2	0.0014	0.14%	0.049	0.012	$2.88X10^{-05***}$	10476
TC	LDL	0.3	0.0012	0.12%	0.046	0.012	$1.09X10^{-04***}$	10476
TC	LDL	0.4	0.0013	0.13%	0.049	0.012	$4.70X10^{-05***}$	10476
TC	LDL	0.5	0.0013	0.13%	0.048	0.012	$7.88X10^{-05***}$	10476
TC	LDL	1	0.0012	0.12%	0.047	0.012	$8.73X10^{-05***}$	10476
TG	HDL	$1x10^{-08}$	0.0008	0.08%	-0.028	0.009	$2.65X10^{-03**}$	10454
TG	HDL	0.1	0.0010	0.10%	-0.039	0.012	$8.13X10^{-04***}$	10454
TG	HDL	0.2	0.0010	0.10%	-0.042	0.012	$8.53X10^{-04***}$	10454

TG	HDL	0.3	0.0010	0.10%	-0.043	0.013	$7.93X10^{-04***}$	10454
TG	HDL	0.4	0.0008	0.08%	-0.040	0.013	$1.69X10^{-03**}$	10454
TG	HDL	0.5	0.0008	0.08%	-0.039	0.013	$2.06X10^{-03**}$	10454
TG	HDL	1	0.0009	0.09%	-0.041	0.013	$1.25X10^{-03**}$	10454
Waist circum	BMI	0.01	0.0011	0.11%	0.038	0.010	$6.61X10^{-05***}$	10152
Waist circum	BMI	0.1	0.0008	0.08%	0.039	0.011	$7.15X10^{-04***}$	10152
Waist circum	BMI	0.2	0.0010	0.10%	0.044	0.012	$2.30X10^{-04***}$	10152
Waist circum	BMI	0.3	0.0013	0.13%	0.051	0.012	$2.43X10^{-05***}$	10152
Waist circum	BMI	0.4	0.0012	0.12%	0.049	0.012	$5.28X10^{-05***}$	10152
Waist circum	BMI	0.5	0.0013	0.13%	0.052	0.012	$2.68X10^{-05***}$	10152
Waist circum	BMI	1	0.0012	0.12%	0.051	0.012	$3.80X10^{-05***}$	10152
Waist circum	Hip circum	0.01	0.0009	0.09%	0.031	0.009	$3.51X10^{-04***}$	10152
Waist circum	Hip circum	0.1	0.0010	0.10%	0.033	0.009	$2.07X10^{-04***}$	10152
Waist circum	Hip circum	0.2	0.0010	0.10%	0.033	0.009	$2.56X10^{-04***}$	10152
Waist circum	Hip circum	0.3	0.0010	0.10%	0.033	0.009	$2.60X10^{-04***}$	10152
Waist circum	Hip circum	0.4	0.0011	0.11%	0.035	0.009	$1.34X10^{-04***}$	10152
Waist circum	Hip circum	0.5	0.0010	0.10%	0.034	0.009	$2.09X10^{-04***}$	10152
Waist circum	Hip circum	1	0.0009	0.09%	0.033	0.009	$4.02X10^{-04***}$	10152
Waist circum	Weight	0.01	0.0008	0.08%	0.029	0.009	$9.48X10^{-04***}$	10152
Waist circum	Weight	0.1	0.0011	0.11%	0.035	0.009	$1.19X10^{-04***}$	10152
Waist circum	Weight	0.2	0.0015	0.15%	0.042	0.009	$7.15X10^{-06***}$	10152
Waist circum	Weight	0.3	0.0014	0.14%	0.041	0.009	$1.01X10^{-05***}$	10152
Waist circum	Weight	0.4	0.0015	0.15%	0.042	0.009	$6.52X10^{-06***}$	10152
Waist circum	Weight	0.5	0.0015	0.15%	0.043	0.009	$4.53X10^{-06***}$	10152
Waist circum	Weight	1	0.0016	0.16%	0.044	0.009	$2.55X10^{-06***}$	10152
Weight	BMI	$1x10^{-04}$	0.0008	0.08%	0.029	0.009	$1.28X10^{-03**}$	10553
Weight	BMI	0.01	0.0023	0.23%	0.054	0.010	$3.57X10^{-08***}$	10553
Weight	BMI	0.1	0.0019	0.19%	0.059	0.012	$5.68X10^{-07***}$	10553
Weight	BMI	0.2	0.0017	0.17%	0.058	0.012	$1.62X10^{-06***}$	10553
Weight	BMI	0.3	0.0020	0.20%	0.064	0.012	$2.32X10^{-07***}$	10553
Weight	BMI	0.4	0.0019	0.19%	0.063	0.013	$4.24X10^{-07***}$	10553
Weight	BMI	0.5	0.0021	0.21%	0.066	0.013	$1.66X10^{-07***}$	10553
Weight	BMI	1	0.0021	0.21%	0.066	0.013	$1.95X10^{-07***}$	10553
Weight	HDL	$1x10^{-04}$	0.0007	0.07%	-0.028	0.009	$2.08X10^{-03**}$	10553
Weight	HDL	0.01	0.0009	0.09%	-0.033	0.010	$7.74X10^{-04***}$	10553
Weight	HDL	0.4	0.0007	0.07%	-0.036	0.012	$2.50X10^{-03**}$	10553
Weight	HDL	0.5	0.0007	0.07%	-0.036	0.012	$2.67X10^{-03**}$	10553
Weight	HDL	1	0.0007	0.07%	-0.037	0.012	$2.01X10^{-03**}$	10553
Weight	Hip circum	0.01	0.0020	0.20%	0.046	0.009	$2.90X10^{-07***}$	10553
Weight	Hip circum	0.1	0.0023	0.23%	0.051	0.009	$3.52X10^{-08***}$	10553
Weight	Hip circum	0.2	0.0023	0.23%	0.051	0.009	$5.37X10^{-08***}$	10553
Weight	Hip circum	0.3	0.0024	0.24%	0.052	0.009	$2.10X10^{-08***}$	10553
Weight	Hip circum	0.4	0.0025	0.25%	0.054	0.009	$8.13X10^{-09***}$	10553
Weight	Hip circum	0.5	0.0025	0.25%	0.054	0.009	$1.13X10^{-08***}$	10553
Weight	Hip circum	1	0.0023	0.23%	0.052	0.009	$3.32X10^{-08***}$	10553
Weight	Waist circum	$1x10^{-04}$	0.0012	0.12%	0.034	0.009	$8.36X10^{-05***}$	10553

Weight	Waist circum	0.01	0.0024	0.24%	0.052	0.009	$1.53X10^{-08***}$	10553
Weight	Waist circum	0.1	0.0018	0.18%	0.043	0.009	$1.09X10^{-06***}$	10553
Weight	Waist circum	0.2	0.0021	0.21%	0.047	0.009	$1.15X10^{-07***}$	10553
Weight	Waist circum	0.3	0.0020	0.20%	0.045	0.009	$3.40X10^{-07***}$	10553
Weight	Waist circum	0.4	0.0019	0.19%	0.044	0.009	$6.26X10^{-07***}$	10553
Weight	Waist circum	0.5	0.0020	0.20%	0.045	0.009	$3.52X10^{-07***}$	10553
Weight	Waist circum	1	0.0020	0.20%	0.045	0.009	$4.13X10^{-07***}$	10553

*BMI – Body Mass Index, DBP – diastolic blood pressure, HDL – high-density lipoprotein, Hip circum – Hip circumference, LDL – low-density lipoprotein , SBP – systolic blood pressure, TC – total cholesterol, TG – triglycerides, Waist circum – Waist circumference; ; p-*value t*hresholds: '***' = p ≤ 0.0001, '**' = p ≤ 0.001, '*' = p ≤ 0.01*

S6. Association testing: Significant associations of derived PGS with CVD-associated disease outcomes

Pheno	PGS	pT	OR	Low_CI	Up_CI	logOR	SE	P	N_cont	N_case	N
OBS	BMI	$1x10^{-04}$	1.092	-1.678	0.004	0.088	0.029	$2.38X10^{-03**}$	8384	2218	10602
OBS	BMI	0.01	1.140	-1.685	0.041	0.131	0.031	$2.22X10^{-05***}$	8384	2218	10602
OBS	BMI	0.1	1.130	-1.691	0.018	0.122	0.036	$6.57X10^{-04**}$	8384	2218	10602
OBS	BMI	0.2	1.132	-1.690	0.017	0.124	0.037	$7.54X10^{-04**}$	8384	2218	10602
OBS	BMI	0.3	1.136	-1.692	0.019	0.127	0.037	$6.40X10^{-04**}$	8384	2218	10602
OBS	BMI	0.4	1.128	-1.691	0.010	0.120	0.038	$1.49X10^{-03**}$	8384	2218	10602
OBS	BMI	0.5	1.136	-1.690	0.016	0.127	0.038	$8.38X10^{-04**}$	8384	2218	10602
OBS	BMI	1	1.137	-1.691	0.017	0.128	0.038	$7.69X10^{-04**}$	8384	2218	10602
DLD	TC	$1x10^{-08}$	1.223	1.416	0.137	0.201	0.022	$5.03X10^{-20***}$	3523	7079	10602
DLD	TC	$1x10^{-06}$	1.234	1.414	0.146	0.210	0.022	$2.18X10^{-21***}$	3523	7079	10602
DLD	TC	$1x10^{-04}$	1.126	1.426	0.055	0.119	0.022	$6.92X10^{-08***}$	3523	7079	10602
DLD	HDL	$1x10^{-08}$	0.781	1.468	-0.312	-0.247	0.022	$8.02X10^{-29***}$	3523	7079	10602
DLD	HDL	$1x10^{-06}$	0.784	1.477	-0.308	-0.243	0.022	$4.10X10^{-28***}$	3523	7079	10602
DLD	HDL	$1x10^{-04}$	0.834	1.469	-0.248	-0.182	0.023	$1.04X10^{-15***}$	3523	7079	10602
DLD	HDL	0.01	0.899	1.433	-0.177	-0.106	0.024	$1.45X10^{-05***}$	3523	7079	10602
OBS	Hip circum	0.1	1.130	-1.679	0.038	0.122	0.029	$2.28X10^{-05***}$	8384	2218	10602
OBS	Hip circum	0.2	1.127	-1.685	0.035	0.119	0.029	$3.77X10^{-05***}$	8384	2218	10602
OBS	Hip circum	0.3	1.125	-1.685	0.034	0.118	0.029	$4.31X10^{-05***}$	8384	2218	10602
OBS	Hip circum	0.4	1.125	-1.685	0.034	0.118	0.029	$4.21X10^{-05***}$	8384	2218	10602
OBS	Hip circum	0.5	1.129	-1.685	0.036	0.121	0.029	$3.13X10^{-05***}$	8384	2218	10602
OBS	Hip circum	1	1.122	-1.685	0.030	0.115	0.029	$7.92X10^{-05***}$	8384	2218	10602
DLD	LDL	$1x10^{-08}$	1.223	1.413	0.138	0.202	0.022	$4.77X10^{-20***}$	3523	7079	10602
DLD	LDL	$1x10^{-06}$	1.227	1.421	0.141	0.205	0.022	$1.30X10^{-20***}$	3523	7079	10602
DLD	LDL	$1x10^{-04}$	1.193	1.431	0.112	0.177	0.022	$1.07X10^{-15***}$	3523	7079	10602
HTN	SBP	0.2	1.135	-3.717	0.002	0.126	0.043	$3.07X10^{-03**}$	6611	3991	10602
HTN	SBP	0.3	1.159	-3.719	0.019	0.147	0.044	$8.37X10^{-04**}$	6611	3991	10602
HTN	SBP	0.4	1.161	-3.720	0.019	0.150	0.045	$8.34X10^{-04**}$	6611	3991	10602
HTN	SBP	0.5	1.151	-3.719	0.008	0.140	0.045	$1.93X10^{-03**}$	6611	3991	10602
HTN	SBP	1	1.154	-3.720	0.011	0.143	0.045	$1.57X10^{-03**}$	6611	3991	10602
OBS	Waist circum	0.01	1.103	-1.681	0.013	0.098	0.029	$8.13X10^{-04**}$	8384	2218	10602
OBS	Weight	0.1	1.089	-1.687	0.001	0.085	0.029	$3.32X10^{-03**}$	8384	2218	10602
OBS	Weight	0.2	1.108	-1.685	0.017	0.103	0.029	$4.89X10^{-04**}$	8384	2218	10602
OBS	Weight	0.3	1.105	-1.683	0.014	0.100	0.029	$6.83X10^{-04**}$	8384	2218	10602
OBS	Weight	0.4	1.108	-1.682	0.017	0.103	0.030	$4.95X10^{-04**}$	8384	2218	10602
OBS	Weight	0.5	1.103	-1.681	0.012	0.098	0.030	$8.72X10^{-04**}$	8384	2218	10602
OBS	Weight	1	1.106	-1.681	0.015	0.101	0.030	$6.38X10^{-04**}$	8384	2218	10602

BMI – Body Mass Index, DLD – Dyslipidaemia, HDL – high-density lipoprotein, Hip circum – Hip circumference, HTN – Hypertension, OBS – Obesity, SBP – systolic blood pressure, TC – total cholesterol, Waist circum – Waist circumference; ; p-value t*hresholds: '***' = p ≤ 0.0001, '**' = p ≤ 0.001, '*' = p ≤ 0.01*

S7. Association testing: Significant associations between non-genetic factors and CVD-associated disease outcomes

Pheno	Non-gen factor	OR	Low_CI	Up_CI	logOR	SE	P	N_cont	N_case	N
CVD	Age	1.042	-3.666	-0.138	0.041	0.064	$5.18x10^{-01}$	6931	246	7177
	Smoking	0.920	-3.702	-0.285	-0.083	0.072	$2.49x10^{-01}$	6931	246	7177
	Alcohol	0.982	-3.691	-0.209	-0.018	0.068	$7.88x10^{-01}$	6931	246	7177
	Sugar drinks	1.023	-3.663	-0.139	0.023	0.058	$6.90x10^{-01}$	6931	246	7177
	Juice	0.931	-3.658	-0.298	-0.071	0.081	$3.79x10^{-01}$	6931	246	7177
	Sleep/night	1.007	-3.665	-0.177	0.007	0.066	$9.13x10^{-01}$	6931	246	7177
	Physical activity	1.064	-3.668	-0.115	0.062	0.063	$3.27x10^{-01}$	6931	246	7177
	Fruit /week	0.780	-3.693	-0.519	-0.248	0.097	$1.03x10^{-02}$	6931	246	7177
	Veg/week	1.027	-3.681	-0.157	0.027	0.065	$6.84x10^{-01}$	6931	246	7177
	Bread/week	0.964	-3.647	-0.241	-0.037	0.072	$6.09x10^{-01}$	6931	246	7177
DLD	Age	0.972	1.358	-0.091	-0.028	0.022	$2.11x10^{-01}$	3301	6700	10001
	Smoking	0.858	1.304	-0.215	-0.153	0.022	$6.08x10^{-12***}$	3301	6700	10001
	Alcohol	0.771	1.308	-0.324	-0.260	0.023	$5.48x10^{-30***}$	3301	6700	10001
	Sugar drinks	1.100	1.360	0.020	0.095	0.027	$3.77x10^{-04***}$	3301	6700	10001
	Juice	1.071	1.356	-0.004	0.069	0.026	$7.65X10^{-03**}$	3301	6700	10001
	Sleep/night	0.983	1.355	-0.081	-0.017	0.023	$4.40X10^{-01}$	3301	6700	10001
	Physical activity	0.954	1.359	-0.110	-0.047	0.022	$3.46X10^{-02}$	3301	6700	10001
	Fruit /week	1.060	1.356	-0.008	0.059	0.024	$1.32X10^{-02}$	3301	6700	10001
	Veg/week	0.943	1.344	-0.124	-0.059	0.023	$1.02X10^{-02}$	3301	6700	10001
	Bread/week	1.123	1.337	0.044	0.116	0.025	$5.42X10^{-06***}$	3301	6700	10001
HA	Age	1.172	-5.141	-0.153	0.158	0.111	$1.53X10^{-01}$	9454	72	9526
	Smoking	1.032	-5.134	-0.359	0.031	0.139	$8.21X10^{-01}$	9454	72	9526
	Alcohol	0.592	-5.415	-0.958	-0.525	0.154	$6.73X10^{-04}$	9454	72	9526
	Sugar drinks	1.152	-5.138	0.005	0.141	0.049	$3.73X10^{-03**}$	9454	72	9526
	Juice	1.126	-5.145	-0.076	0.119	0.069	$8.67X10^{-02}$	9454	72	9526

	Sleep/night	1.210	-5.169	-0.149	0.190	0.121	$1.15X10^{-01}$	9454	72	9526
	Physical activity	0.803	-5.184	-0.595	-0.220	0.134	$9.98X10^{-02}$	9454	72	9526
	Fruit /week	0.999	-5.122	-0.334	-0.001	0.119	$9.95X10^{-01}$	9454	72	9526
	Veg/week	0.908	-5.061	-0.458	-0.097	0.129	$4.52X10^{-01}$	9454	72	9526
	Bread/week	0.963	-4.979	-0.384	-0.038	0.123	$7.60X10^{-01}$	9454	72	9526
HTN	Age	1.905	-0.631	0.580	0.644	0.023	$9.75X10^{-173***}$	6543	3458	10001
	Smoking	0.914	-0.625	-0.157	-0.089	0.024	$2.04X10^{-04***}$	6543	3458	10001
	Alcohol	0.796	-0.655	-0.292	-0.229	0.022	$2.55X10^{-24***}$	6543	3458	10001
	Sugar drinks	1.319	-0.594	0.203	0.277	0.026	$4.63X10^{-26***}$	6543	3458	10001
	Juice	1.266	-0.608	0.169	0.236	0.024	$5.08X10^{-23***}$	6543	3458	10001
	Sleep/night	1.298	-0.631	0.200	0.261	0.022	$8.35X10^{-33***}$	6543	3458	10001
	Physical activity	0.745	-0.620	-0.358	-0.294	0.023	$7.62X10^{-38***}$	6543	3458	10001
	Fruit /week	1.113	-0.594	0.049	0.107	0.021	$2.65X10^{-07***}$	6543	3458	10001
	Veg/week	0.775	-0.550	-0.323	-0.255	0.024	$1.63X10^{-26***}$	6543	3458	10001
	Bread/week	1.386	-0.458	0.262	0.327	0.023	$8.63X10^{-46***}$	6543	3458	10001
OBS	Age	1.208	-1.128	0.114	0.189	0.027	$2.13X10^{-12***}$	8296	1705	10001
	Smoking	0.694	-1.268	-0.534	-0.366	0.060	$1.01X10^{-09***}$	8296	1705	10001
	Alcohol	0.486	-1.422	-0.832	-0.722	0.039	$9.17X10^{-77***}$	8296	1705	10001
	Sugar drinks	1.574	-1.134	0.366	0.454	0.031	$1.96X10^{-48***}$	8296	1705	10001
	Juice	1.374	-1.151	0.243	0.318	0.027	$4.94X10^{-33***}$	8296	1705	10001
	Sleep/night	1.140	-1.128	0.051	0.131	0.029	$4.40X10^{-06***}$	8296	1705	10001
	Physical activity	0.722	-1.163	-0.415	-0.326	0.032	$1.86X10^{-24***}$	8296	1705	10001
	Fruit /week	1.271	-1.134	0.171	0.240	0.025	$1.40X10^{-22***}$	8296	1705	10001
	Veg/week	0.853	-1.058	-0.245	-0.159	0.031	$1.93X10^{-07***}$	8296	1705	10001
	Bread/week	1.790	-0.980	0.502	0.582	0.029	$1.23X10^{-92***}$	8296	1705	10001
Stroke	Age	1.540	-4.585	0.227	0.432	0.073	$3.05X10^{-09***}$	9843	140	9983
	Smoking	0.929	-4.510	-0.363	-0.073	0.103	$4.76X10^{-01}$	9843	140	9983

	Alcohol	0.636	-4.690	-0.749	-0.453	0.105	1.74X10^{-05***}	9843	140	9983
	Sugar drinks	1.080	-4.470	-0.077	0.077	0.055	1.59X10^{-01}	9843	140	9983
	Juice	1.117	-4.484	-0.039	0.111	0.053	3.81X10^{-02}	9843	140	9983
	Sleep/night	1.349	-4.576	0.063	0.299	0.084	3.74X10^{-04***}	9843	140	9983
	Physical activity	0.603	-4.638	-0.822	-0.506	0.112	6.59X10^{-06***}	9843	140	9983
	Fruit /week	1.047	-4.468	-0.166	0.046	0.075	5.43X10^{-01}	9843	140	9983
	Veg/week	0.718	-4.505	-0.652	-0.332	0.114	3.62X10^{-03}	9843	140	9983
	Bread/week	1.218	-4.435	-0.014	0.197	0.075	8.87X10^{-03}	9843	140	9983
T2D	Age	1.507	-2.890	0.305	0.410	0.037	5.32X10^{-28***}	9330	600	9930
	Smoking	0.825	-2.883	-0.349	-0.192	0.056	5.58X10^{-04***}	9330	600	9930
	Alcohol	0.583	-3.051	-0.690	-0.540	0.054	6.98X10^{-24***}	9330	600	9930
	Sugar drinks	1.111	-2.808	0.023	0.105	0.029	3.05X10^{-04***}	9330	600	9930
	Juice	1.070	-2.809	-0.027	0.067	0.034	4.49X10^{-02}	9330	600	9930
	Sleep/night	1.187	-2.821	0.053	0.172	0.042	4.68X10^{-05***}	9330	600	9930
	Physical activity	0.709	-2.875	-0.485	-0.344	0.050	6.27X10^{-12}	9330	600	9930
	Fruit /week	1.151	-2.816	0.049	0.141	0.033	1.64X10^{-05***}	9330	600	9930
	Veg/week	0.882	-2.788	-0.258	-0.126	0.047	7.84X10^{-03**}	9330	600	9930
	Bread/week	1.262	-2.739	0.126	0.233	0.038	1.05X10^{-09***}	9330	600	9930

*CVD – Cardiovascular disease, DLD – dyslipidaemia, HA – Heart attack, HTN – Hypertension, OBS – Obesity, T2D – Diabetes mellitus; p-value thresholds: '***' = p ≤ 0.0001, '**' = p ≤ 0.001, '*' = p ≤ 0.01*

S8. Prediction Modelling: MultiPGS model

A. Model predictive utility assessment

Pheno	Model	R	R2	R2%	SE	P	Ncase	Ncont
CVD	Albumin	-0.036	0.001	0.1%	0.011	$1.76X10^{-03**}$	219	5850
	Bilirubin	0.019	0.000	0.0%	0.011	$1.00X10^{-01}$	219	5850
	BMI	-0.020	0.000	0.0%	0.011	$8.02X10^{-02}$	219	5850
	Cholesterol	0.006	0.000	0.0%	0.011	$6.00X10^{-01}$	219	5850
	DBP	-0.011	0.000	0.0%	0.011	$3.38X10^{-01}$	219	5850
	Height	-0.028	0.001	0.1%	0.011	$1.40X10^{-02}$	219	5850
	HDL	-0.005	0.000	0.0%	0.011	$6.33X10^{-01}$	219	5850
	Hip circum	-0.026	0.001	0.1%	0.011	$2.38X10^{-02}$	219	5850
	LDL	-0.020	0.000	0.0%	0.011	$7.89X10^{-02}$	219	5850
	SBP	-0.007	0.000	0.0%	0.011	$5.42X10^{-01}$	219	5850
	TG	0.001	0.000	0.0%	0.011	$9.46X10^{-01}$	219	5850
	Waist circum	-0.015	0.000	0.0%	0.011	$1.93X10^{-01}$	219	5850
	Waist hip ratio	-0.021	0.000	0.0%	0.011	$6.93X10^{-02}$	219	5850
	Weight	0.004	0.000	0.0%	0.011	$7.04X10^{-01}$	219	5850
	MultiPGS	-0.021	0.000	0.0%	0.011	$7.11X10^{-02}$	219	5850
DLD	Albumin	-0.006	0.000	0.0%	0.010	$5.31X10^{-01}$	5680	2801
	Bilirubin	0.018	0.000	0.0%	0.010	$6.57X10^{-02}$	5680	2801
	BMI	0.016	0.000	0.0%	0.010	$1.01X10^{-01}$	5680	2801
	Cholesterol	0.091	0.008	0.8%	0.010	$5.53X10^{-21***}$	5680	2801
	DBP	0.008	0.000	0.0%	0.010	$3.93X10^{-01}$	5680	2801
	Height	0.023	0.001	0.1%	0.010	$2.03X10^{-02}$	5680	2801
	HDL	0.108	0.012	1.2%	0.010	$4.50X10^{-29***}$	5680	2801
	Hip circum	0.002	0.000	0.0%	0.010	$8.62X10^{-01}$	5680	2801
	LDL	0.087	0.008	0.8%	0.010	$4.09X10^{-19***}$	5680	2801
	SBP	0.010	0.000	0.0%	0.010	$3.28X10^{-01}$	5680	2801
	TG	0.054	0.003	0.3%	0.010	$2.67X10^{-08***}$	5680	2801
	Waist circum	-0.006	0.000	0.0%	0.010	$5.50X10^{-01}$	5680	2801
	Waist hip ratio	0.030	0.001	0.1%	0.010	$1.75X10^{-03***}$	5680	2801
	Weight	-0.016	0.000	0.0%	0.010	$9.16X10^{-02}$	5680	2801
	MultiPGS	0.124	0.015	1.5%	0.010	$1.81X10^{-37***}$	5680	2801
HA	Albumin	0.008	0.000	0.0%	0.011	$4.75X10^{-01}$	46	6833
	Bilirubin	0.014	0.000	0.0%	0.011	$1.85X10^{-01}$	46	6833
	BMI	-0.001	0.000	0.0%	0.011	$8.95X10^{-01}$	46	6833
	Cholesterol	-0.009	0.000	0.0%	0.011	$3.94X10^{-01}$	46	6833
	DBP	0.008	0.000	0.0%	0.011	$4.80X10^{-01}$	46	6833
	Height	-0.049	0.002	0.2%	0.011	$6.55X10^{-06***}$	46	6833
	HDL	-0.021	0.000	0.0%	0.011	$5.36X10^{-02}$	46	6833
	Hip circum	0.001	0.000	0.0%	0.011	$9.15X10^{-01}$	46	6833
	LDL	0.015	0.000	0.0%	0.011	$1.77X10^{-01}$	46	6833

	SBP	0.004	0.000	0.0%	0.011	$7.39X10^{-01}$	46	6833
	TG	-0.018	0.000	0.0%	0.011	$9.14X10^{-02}$	46	6833
	Waist circum	0.001	0.000	0.0%	0.011	$8.98X10^{-01}$	46	6833
	Waist hip ratio	-0.004	0.000	0.0%	0.011	$7.32X10^{-01}$	46	6833
	Weight	-0.027	0.001	0.1%	0.011	$1.20X10^{-02}$	46	6833
	MultiPGS	0.003	0.000	0.0%	0.011	$7.72X10^{-01}$	46	6833
HTN	Albumin	0.114	0.013	1.3%	0.010	$3.49X10^{-32***}$	3183	5298
	Bilirubin	0.166	0.028	2.8%	0.010	$1.10X10^{-66***}$	3183	5298
	BMI	0.135	0.018	1.8%	0.010	$4.69X10^{-44***}$	3183	5298
	Cholesterol	0.068	0.005	0.5%	0.010	$1.66X10^{-12***}$	3183	5298
	DBP	0.093	0.009	0.9%	0.010	$6.36X10^{-22***}$	3183	5298
	Height	0.153	0.023	2.3%	0.010	$3.42X10^{-56***}$	3183	5298
	HDL	0.095	0.009	0.9%	0.010	$8.27X10^{-23***}$	3183	5298
	Hip circum	0.082	0.007	0.7%	0.010	$3.74X10^{-17***}$	3183	5298
	LDL	0.034	0.001	0.1%	0.010	$4.08X10^{-04**}$	3183	5298
	SBP	0.132	0.017	1.7%	0.010	$3.90X10^{-42***}$	3183	5298
	TG	0.058	0.003	0.3%	0.010	$1.75X10^{-09***}$	3183	5298
	Waist circum	0.124	0.015	1.5%	0.010	$1.40X10^{-37***}$	3183	5298
	Waist hip ratio	0.061	0.004	0.4%	0.010	$4.26X10^{-10***}$	3183	5298
	Weight	0.079	0.006	0.6%	0.010	$3.41X10^{-16***}$	3183	5298
	MultiPGS	0.259	0.067	6.7%	0.009	$1.84X10^{-162***}$	3183	5298
OBS	Albumin	0.076	0.006	0.6%	0.010	$3.17X10^{-15***}$	1775	6706
	Bilirubin	0.113	0.013	1.3%	0.010	$2.53X10^{-31***}$	1775	6706
	BMI	0.111	0.012	1.2%	0.010	$1.82X10^{-30***}$	1775	6706
	Cholesterol	0.094	0.009	0.9%	0.010	$3.79X10^{-22***}$	1775	6706
	DBP	0.082	0.007	0.7%	0.010	$3.36X10^{-17***}$	1775	6706
	Height	0.200	0.040	4.0%	0.010	$2.52X10^{-96***}$	1775	6706
	HDL	0.134	0.018	1.8%	0.010	$5.54X10^{-44***}$	1775	6706
	Hip circum	0.087	0.008	0.8%	0.010	$2.95X10^{-19***}$	1775	6706
	LDL	0.070	0.005	0.5%	0.010	$4.72X10^{-13***}$	1775	6706
	SBP	0.055	0.003	0.3%	0.010	$1.40X10^{-08***}$	1775	6706
	TG	0.107	0.011	1.1%	0.010	$2.55X10^{-28***}$	1775	6706
	Waist circum	0.133	0.018	1.8%	0.010	$7.39X10^{-43***}$	1775	6706
	Waist hip ratio	0.099	0.010	1.0%	0.010	$1.28X10^{-24***}$	1775	6706
	Weight	0.042	0.002	0.2%	0.010	$1.70X10^{-05***}$	1775	6706
	MultiPGS	0.274	0.075	7.5%	0.009	$4.54X10^{-182***}$	1775	6706
T2D	Albumin	0.026	0.001	0.1%	0.010	$8.61X10^{-03**}$	568	7861
	Bilirubin	0.021	0.000	0.0%	0.010	$2.92X10^{-02**}$	568	7861
	BMI	0.011	0.000	0.0%	0.010	$2.69X10^{-01}$	568	7861
	Cholesterol	0.028	0.001	0.1%	0.010	$4.25X10^{-03**}$	568	7861
	DBP	0.002	0.000	0.0%	0.010	$8.26X10^{-01}$	568	7861
	Height	0.059	0.003	0.3%	0.010	$1.78X10^{-09***}$	568	7861
	HDL	0.030	0.001	0.1%	0.010	$2.26X10^{-03**}$	568	7861
	Hip circum	0.033	0.001	0.1%	0.010	$7.42X10^{-04**}$	568	7861

	LDL	0.028	0.001	0.1%	0.010	$4.52X10^{-03**}$	568	7861
	SBP	0.005	0.000	0.0%	0.010	$5.87X10^{-01}$	568	7861
	TG	0.054	0.003	0.3%	0.010	$3.32X10^{-08***}$	568	7861
	Waist circum	0.022	0.000	0.0%	0.010	$2.34X10^{-02***}$	568	7861
	Waist hip ratio	0.025	0.001	0.1%	0.010	$9.32X10^{-03**}$	568	7861
	Weight	0.010	0.000	0.0%	0.010	$3.00X10^{-01}$	568	7861
	MultiPGS	0.081	0.007	0.7%	0.010	$9.43X10^{-17***}$	568	7861

*CVD – Cardiovascular disease, DLD – dyslipidaemia, HA – Heart attack, HTN – Hypertension, OBS – Obesity, T2D – Diabetes mellitus; p-value thresholds: '***' = p ≤ 0.0001, '**' = p ≤ 0.001, '*' = p ≤ 0.05*

S9. Prediction Modelling: Genetic models; MultiPGS + Ancestry

A. Model predictive utility assessment

Phenotype	Model	R2	R2%	SE	P	Ncase (%)	Ncont (%)
CVD	MultiPGS	0.000	0.0	0.013	$1.74x10^{-01}$	177 (3.5)	4889 (96.5)
	Ancestry	0.001	0.1	0.013	$6.99\ x10^{-03**}$		
	PGS + Ancestry	0.000	0.0	0.013	$8.33\ x10^{-02}$		
DLD	MultiPGS	0.017	1.7	0.011	$6.71\ x10^{-34***}$	4653 (66.0)	2394 (44.0)
	Ancestry	0.002	0.2	0.011	$4.52\ x10^{-5***}$		
	PGS + Ancestry	0.016	1.6	0.011	$8.39\ x10^{-32***}$		
HA	MultiPGS	0.000	0.0	0.011	$7.93\ x10^{-01}$	52 (0.8)	6776 (99.2)
	Ancestry	0.000	0.0	0.011	$2.98\ x10^{-01}$		
	PGS + Ancestry	0.000	0.0	0.011	$4.60\ x10^{-01}$		
HTN	MultiPGS	0.065	6.5	0.010	$3.34\ x10^{-131***}$	2453 (34.8)	4594 (67.3)
	Ancestry	0.090	9.0	0.010	$8.47\ x10^{-182***}$		
	PGS + Ancestry	0.090	9.0	0.010	$2.36\ x10^{-183***}$		
OBS	MultiPGS	0.071	7.1	0.010	$2.44\ x10^{-142***}$	1188 (16.0)	5859 (84.0)
	Ancestry	0.075	7.5	0.010	$9.70\ x10^{-151***}$		
	PGS + Ancestry	0.096	9.6	0.010	$1.84\ x10^{-195***}$		
Stroke	MultiPGS	0.001	0.1	0.011	$2.87\ x10^{-3**}$	93 (1.3)	6945 (98.7)
	Ancestry	0.004	0.4	0.011	$2.75\ x10^{-09***}$		
	PGS + Ancestry	0.001	0.1	0.011	$5.60\ x10^{-04***}$		
T2D	MultiPGS	0.006	0.6	0.011	$1.86\ x10^{-12***}$	412 (5.9)	6583 (94.1)
	Ancestry	0.008	0.8	0.011	$1.08\ x10^{-17***}$		
	PGS + Ancestry	0.007	0.7	0.011	$8.90\ x10^{-15***}$		

*CVD – Cardiovascular disease, DLD – Dyslipidaemia, HA – Heart attack, HTN – Hypertension, PGS – Polygenic risk score, OBS – Obesity, T2D – Diabetes mellitus; p-value thresholds: '***' = p ≤ 0.0001, '**' = p ≤ 0.001, '*' = p ≤ 0.01*

B. Model performance comparison

Pheno	Model 1	Model 2	Model_1_R	Model_2_R	R_diff	R_diff_pval
CVD	MultiPGS	MultiPGS	-0.017	-0.017	0.000	1
	MultiPGS	Ancestry	-0.017	-0.034	0.017	$3.21x10^{-01}$
	MultiPGS	PGS + Ancestry	-0.017	-0.022	0.005	$4.19x10^{-02*}$
	Ancestry	MultiPGS	-0.034	-0.017	-0.017	$3.21x10^{-01}$
	Ancestry	Ancestry	-0.034	-0.034	0.000	1
	Ancestry	PGS + Ancestry	-0.034	-0.022	-0.012	$4.49x10^{-01}$
	PGS + Ancestry	MultiPGS	-0.022	-0.017	-0.005	$4.19x10^{-02*}$
	PGS + Ancestry	Ancestry	-0.022	-0.034	0.012	$4.49x10^{-01}$
	PGS + Ancestry	PGS + Ancestry	-0.022	-0.022	0.000	1
DLD	MultiPGS	MultiPGS	0.129	0.129	0.000	1
	MultiPGS	Ancestry	0.129	0.043	0.085	$7.24x10^{-12***}$
	MultiPGS	PGS + Ancestry	0.129	0.125	0.004	$6.43x10^{-02*}$
	Ancestry	MultiPGS	0.043	0.129	-0.085	$7.24x10^{-12***}$

	Ancestry	Ancestry	0.043	0.043	0.000	1
	Ancestry	PGS + Ancestry	0.043	0.125	-0.081	$1.78x10^{-11***}$
	PGS + Ancestry	MultiPGS	0.125	0.129	-0.004	$6.43x10^{-02*}$
	PGS + Ancestry	Ancestry	0.125	0.043	0.081	$1.78x10^{-11***}$
	PGS + Ancestry	PGS + Ancestry	0.125	0.125	0.000	1
HA	MultiPGS	MultiPGS	0.003	0.003	0.000	1
	MultiPGS	Ancestry	0.003	0.011	-0.008	$5.45x10^{-01}$
	MultiPGS	PGS + Ancestry	0.003	0.008	-0.005	$1.80x10^{-01}$
	Ancestry	MultiPGS	0.011	0.003	0.008	$5.45x10^{-01}$
	Ancestry	Ancestry	0.011	0.011	0.000	1
	Ancestry	PGS + Ancestry	0.011	0.008	0.003	$7.96x10^{-01}$
	PGS + Ancestry	MultiPGS	0.008	0.003	0.005	$1.80x10^{-01}$
	PGS + Ancestry	Ancestry	0.008	0.011	-0.003	$7.96x10^{-01}$
	PGS + Ancestry	PGS + Ancestry	0.008	0.008	0.000	1
HTN	MultiPGS	MultiPGS	0.255	0.255	0.000	1
	MultiPGS	Ancestry	0.255	0.299	-0.044	$5.04x10^{-09}$
	MultiPGS	PGS + Ancestry	0.255	0.301	-0.045	$3.36x10^{-16***}$
	Ancestry	MultiPGS	0.299	0.255	0.044	$5.04x10^{-09}$
	Ancestry	Ancestry	0.299	0.299	0.000	1
	Ancestry	PGS + Ancestry	0.299	0.301	-0.001	$7.50x10^{-01}$
	PGS + Ancestry	MultiPGS	0.301	0.255	0.045	$3.36x10^{-16***}$
	PGS + Ancestry	Ancestry	0.301	0.299	0.001	$7.50x10^{-01}$
	PGS + Ancestry	PGS + Ancestry	0.301	0.301	0.000	1
Stroke	MultiPGS	MultiPGS	0.032	0.032	0.000	1
	MultiPGS	Ancestry	0.032	0.063	-0.032	$7.17x10^{-03**}$
	MultiPGS	PGS + Ancestry	0.032	0.037	-0.005	$1.26x10^{-01}$
	Ancestry	MultiPGS	0.063	0.032	0.032	$7.17x10^{-03**}$
	Ancestry	Ancestry	0.063	0.063	0.000	1
	Ancestry	PGS + Ancestry	0.063	0.037	0.027	$1.00x10^{-02*}$
	PGS + Ancestry	MultiPGS	0.037	0.032	0.005	$1.26x10^{-01}$
	PGS + Ancestry	Ancestry	0.037	0.063	-0.027	$1.00x10^{-02*}$
	PGS + Ancestry	PGS + Ancestry	0.037	0.037	0.000	1
T2D	MultiPGS	MultiPGS	0.075	0.075	0.000	1
	MultiPGS	Ancestry	0.075	0.091	-0.016	$1.09x10^{-01}$
	MultiPGS	PGS + Ancestry	0.075	0.083	-0.008	$2.41x10^{-02*}$
	Ancestry	MultiPGS	0.091	0.075	0.016	$1.09x10^{-01}$
	Ancestry	Ancestry	0.091	0.091	0.000	1
	Ancestry	PGS + Ancestry	0.091	0.083	0.009	$2.82x10^{-01}$
	PGS + Ancestry	MultiPGS	0.083	0.075	0.008	$2.41x10^{-02*}$
	PGS + Ancestry	Ancestry	0.083	0.091	-0.009	$2.82x10^{-01}$
	PGS + Ancestry	PGS + Ancestry	0.083	0.083	0.000	1
Obesity	MultiPGS	MultiPGS	0.266	0.266	0.000	1
	MultiPGS	Ancestry	0.266	0.273	-0.008	$4.18\ x10^{-01}$
	MultiPGS	PGS + Ancestry	0.266	0.310	-0.044	$7.52\ x10^{-13***}$
	Ancestry	MultiPGS	0.273	0.266	0.008	$4.18\ x10^{-01}$
	Ancestry	Ancestry	0.273	0.273	0.000	1

	Ancestry	PGS + Ancestry	0.273	0.310	-0.037	$8.71x10^{-14***}$
	PGS + Ancestry	MultiPGS	0.310	0.266	0.044	$7.52x10^{-13***}$
	PGS + Ancestry	Ancestry	0.310	0.273	0.037	$8.71x10^{-14***}$
	PGS + Ancestry	PGS + Ancestry	0.310	0.310	0.000	1

*CVD – Cardiovascular disease, DLD – Dyslipidaemia, HA – Heart attack, HTN – Hypertension, PGS – Polygenic risk score, OBS – Obesity, T2D – Diabetes mellitus; p-value thresholds: '***' = p ≤ 0.0001, '**' = p ≤ 0.001, '*' = p ≤ 0.01*

S10. Prediction Modelling: Integrated models; Genetic (PGS + Ancestry) and Non-genetic

A. Model predictive utility assessment

Phenotype	Model	R2	R2%	SE	P	Ncase (%)	Ncont (%)
CVD	Gen	0.000	0.0	0.013	2.52×10^{-01}	154 (3.3)	4498 (96.7)
	Non gen	0.000	0.0	0.013	6.13×10^{-01}		
	Gen + Non gen	0.000	0.0	0.013	1.03×10^{-01}		
DLD	Gen	0.018	1.8	0.011	$5.79 \times 10^{-34***}$	4248 (65.9)	2202 (34.1)
	Non gen	0.139	13.9	0.010	$1.52 \times 10^{-264***}$		
	Gen + Non gen	0.154	15.4	0.010	$5.89 \times 10^{-295***}$		
HA	Genetic	0.000	0.0	0.011	9.92×10^{-01}	42 (0.67)	6196 (99.3)
	Non gen	0.001	0.1	0.011	$1.98 \times 10^{-02*}$		
	Gen + Non gen	0.000	0.0	0.011	5.90×10^{-01}		
HTN	Genetic	0.082	8.2	0.011	$7.45 \times 10^{-153***}$	2280 (35.3)	4170 (64.7)
	Non gen	0.112	11.2	0.010	$1.05 \times 10^{-210***}$		
	Gen + Non gen	0.138	13.8	0.010	$5.11 \times 10^{-263***}$		
OBS	Genetic	0.093	9.3	0.011	$7.39 \times 10^{-174***}$	1144 (17.7)	5306 (82.3)
	Non gen	0.154	15.4	0.010	$4.15 \times 10^{-294***}$		
	Gen + Non gen	0.209	20.9	0.010	$<1.0 \times 10^{-300***}$		
Stroke	Genetic	0.001	0.1	0.011	$2.19 \times 10^{-02*}$	91 (1.4)	6355 (98.6)
	Non gen	0.003	0.3	0.011	$6.25 \times 10^{-07**}$		
	Gen + Non gen	0.001	0.1	0.011	$1.08 \times 10^{-03**}$		
T2D	Genetic	0.008	0.8	0.011	$1.65 \times 10^{-15***}$	382 (5.9)	6019 (93.9%)
	Non gen	0.028	2.8	0.011	$1.9 \times 10^{-51***}$		
	Gen + Non gen	0.023	2.3	0.011	$7.83 \times 10^{-43***}$		

*CVD – Cardiovascular disease, DLD – Dyslipidaemia, Gen – Genetic, Gen + Non-gen – Genetic + Non-genetic, HA – Heart attack, HTN – Hypertension, Non-gen – Non-genetic, OBS – Obesity, T2D – Diabetes mellitus; p-value thresholds: '***' = p ≤ 0.0001, '**' = p ≤ 0.001, '*' = p ≤ 0.01*

B. Model performance comparison

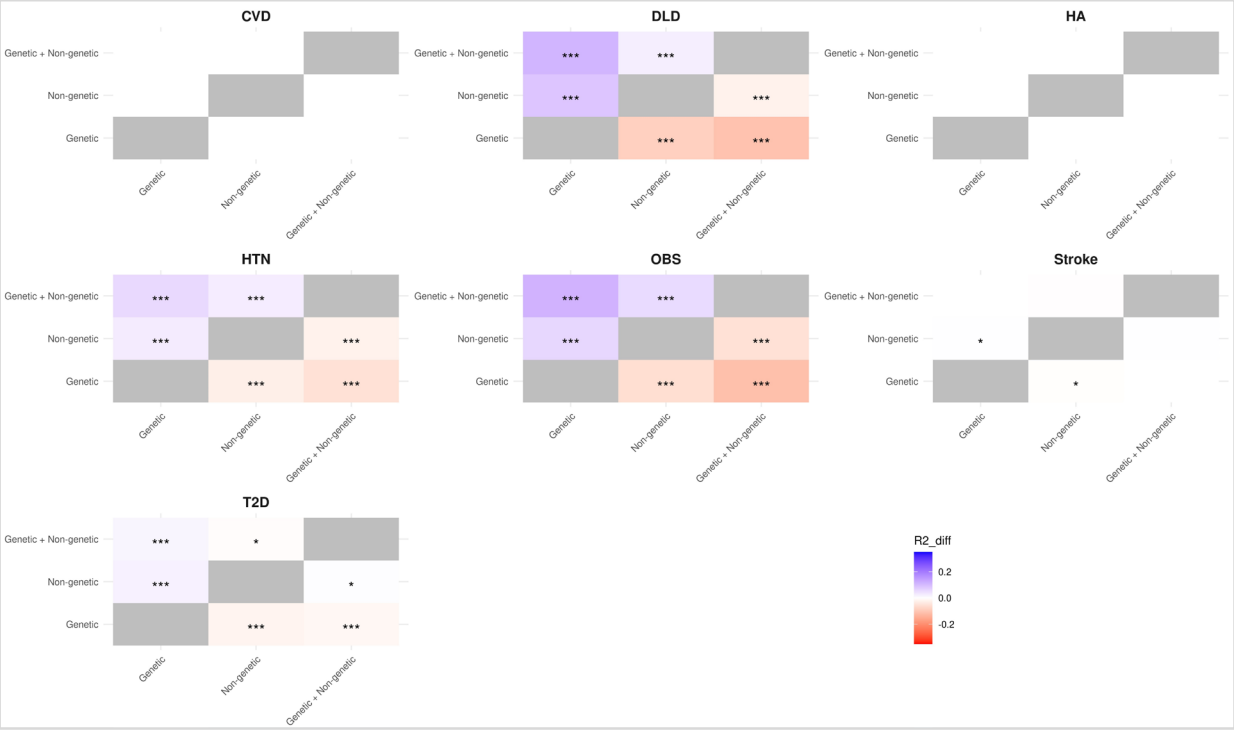
Pheno	Model_1	Model_2	Model_1_R	Model_2_R	R_diff	R_diff_pval
CVD	Gen	Gen	-0.015	-0.015	0.000	1
	Gen	Non-gen	-0.015	0.007	-0.022	2.36×10^{-01}
	Gen	Gen + Non-gen	-0.015	-0.021	0.006	4.58×10^{-01}
	Non-gen	Gen	0.007	-0.015	0.022	2.36×10^{-01}
	Non-gen	Non-gen	0.007	0.007	0.000	1
	Non-gen	Gen + Non-gen	0.007	-0.021	0.028	6.25×10^{-02}
	Gen + Non-gen	Gen	-0.021	-0.015	-0.006	4.58×10^{-01}
	Gen + Non-gen	Non-gen	-0.021	0.007	-0.028	6.25×10^{-02}
	Gen + Non-gen	Gen + Non-gen	-0.021	-0.021	0.000	1
DLD	Gen	Gen	0.172	0.172	0.000	1
	Gen	Non-gen	0.172	0.339	-0.167	$8.89 \times 10^{-32***}$
	Gen	Gen + Non-gen	0.172	0.375	-0.203	$2.01 \times 10^{-84***}$

	Non-gen	Gen	0.339	0.172	0.167	$8.89x10^{-32***}$
	Non-gen	Non-gen	0.339	0.339	0.000	1
	Non-gen	Gen + Non-gen	0.339	0.375	-0.036	$4.45x10^{-12***}$
	Gen + Non-gen	Gen	0.375	0.172	0.203	$2.01x10^{-84***}$
	Gen + Non-gen	Non-gen	0.375	0.339	0.036	$4.45x10^{-12***}$
	Gen + Non-gen	Gen + Non-gen	0.375	0.375	0.000	1
HA	Gen	Gen	0.000	0.000	0.000	1
	Gen	Non-gen	0.000	0.026	-0.026	$8.71x10^{-02}$
	Gen	Gen + Non-gen	0.000	0.006	-0.006	$4.17x10^{-01}$
	Non-gen	Gen	0.026	0.000	0.026	$8.71x10^{-02}$
	Non-gen	Non-gen	0.026	0.026	0.000	1
	Non-gen	Gen + Non-gen	0.026	0.006	0.020	$8.79x10^{-02}$
	Gen + Non-gen	Gen	0.006	0.000	0.006	$4.17x10^{-01}$
	Gen + Non-gen	Non-gen	0.006	0.026	-0.020	$8.79x10^{-02}$
	Gen + Non-gen	Gen + Non-gen	0.006	0.006	0.000	1
HTN	Gen	Gen	0.287	0.287	0.000	1
	Gen	Non-gen	0.287	0.335	-0.048	$9.55x10^{-06***}$
	Gen	Gen + Non-gen	0.287	0.372	-0.085	$1.30x10^{-34***}$
	Non-gen	Gen	0.335	0.287	0.048	$9.55x10^{-06***}$
	Non-gen	Non-gen	0.335	0.335	0.000	1
	Non-gen	Gen + Non-gen	0.335	0.372	-0.037	$2.37x10^{-13***}$
	Gen + Non-gen	Gen	0.372	0.287	0.085	$1.30x10^{-34***}$
	Gen + Non-gen	Non-gen	0.372	0.335	0.037	$2.37x10^{-13***}$
	Gen + Non-gen	Gen + Non-gen	0.372	0.372	0.000	1
Stroke	Gen	Gen	0.026	0.026	0.000	1
	Gen	Non-gen	0.026	0.055	-0.030	$3.08x10^{-02*}$
	Gen	Gen + Non-gen	0.026	0.036	-0.011	$5.93x10^{-02}$
	Non-gen	Gen	0.055	0.026	0.030	$3.08x10^{-02*}$
	Non-gen	Non-gen	0.055	0.055	0.000	1
	Non-gen	Gen + Non-gen	0.055	0.036	0.019	$7.32x10^{-02}$
	Gen + Non-gen	Gen	0.036	0.026	0.011	$5.93x10^{-02}$
	Gen + Non-gen	Non-gen	0.036	0.055	-0.019	$7.32x10^{-02}$
	Gen + Non-gen	Gen + Non-gen	0.036	0.036	0.000	1
T2D	Gen	Gen	0.089	0.089	0.000	1
	Gen	Non-gen	0.089	0.167	-0.079	$9.51x10^{-10***}$
	Gen	Gen + Non-gen	0.089	0.152	-0.064	$5.32x10^{-13***}$
	Non-gen	Gen	0.167	0.089	0.079	$9.51x10^{-10***}$
	Non-gen	Non-gen	0.167	0.167	0.000	1
	Non-gen	Gen + Non-gen	0.167	0.152	0.015	$2.25x10^{-02*}$
	Gen + Non-gen	Gen	0.152	0.089	0.064	$5.32x10^{-13***}$
	Gen + Non-gen	Non-gen	0.152	0.167	-0.015	$2.25x10^{-02*}$
	Gen + Non-gen	Gen + Non-gen	0.152	0.152	0.000	1
Obesity	Gen	Gen	0.306	0.306	0.000	1
	Gen	Non-gen	0.306	0.392	-0.086	$1.73x10^{-13***}$
	Gen	Gen + Non-gen	0.306	0.457	-0.152	$4.85x10^{-79***}$
	Non-gen	Gen	0.392	0.306	0.086	$1.73x10^{-13***}$

	Non-gen	Non-gen	0.392	0.392	0.000	1
	Non-gen	Gen + Non-gen	0.392	0.457	-0.065	$1.33x10^{-34***}$
	Gen + Non-gen	Gen	0.457	0.306	0.152	$4.85x10^{-79***}$
	Gen + Non-gen	Non-gen	0.457	0.392	0.065	$1.33x10^{-34***}$
	Gen + Non-gen	Gen + Non-gen	0.457	0.457	0.000	1

*CVD – Cardiovascular disease, DLD – Dyslipidaemia, Gen – Genetic, Gen + Non-gen – Genetic + Non-genetic, HA – Heart attack, HTN – Hypertension, Non-gen – Non-genetic, OBS – Obesity, T2D – Diabetes mellitus; p-value thresholds: '***' = p ≤ 0.0001, '**' = p ≤ 0.001, '*' = p ≤ 0.01*

C. Integrated model performance comparison matrix



*CVD – Cardiovascular disease, DLD – Dyslipidaemia, Gen – Genetic, Gen + Non-gen – Genetic + Non-genetic, HA – Heart attack, HTN – Hypertension, Non-gen – Non-genetic, OBS – Obesity, T2D – Diabetes mellitus; p-value thresholds: '***' = p ≤ 0.0001, '**' = p ≤ 0.001, '*' = p ≤ 0.01*

Publication 3: supplementary materials

S11. Clinician Survey

PARTICIPANT DEMOGRAPHICS AND OTHER CHARACTERISTICS

---

1. Please select your primary hospital affiliation:

Chris Hani Baragwanath Academic Hospital / Charlotte Maxeke Johannesburg Academic Hospital / Helen Joseph Hospital / Rahima Moosa Mother and Child Hospital / I prefer not to answer

2. What is your medical speciality?

Internal Medicine / Obstetrics and Gynaecology / Orthopaedics / Paediatrics / Surgery / Other / I prefer not to answer

3. Within Internal Medicine, what is your sub-speciality?

Cardiology / Endocrinology / Haematology/Oncology / Infectious Diseases / Nephrology / Neurology / Pulmonology / Rheumatology / Gastroenterology / Dermatology /I prefer not to answer

If other, please specify.

4. What is your current level of practice?

Specialist / Registrar / Medical officer / Community service officer / Intern / Other / I prefer not to answer

If other, please specify.

5. How many years of experience do you have in clinical practice (in years)?

Start from the year after you graduated, including internship and community service, if you prefer not to answer, please put 000

6. At which Institution did you study towards your medical degree?

University of Cape Town / University of the Free State / University of KwaZulu-Natal / University of Limpopo / Sefako Makgatho Health Sciences University (previously known as MEDUNSA) / Stellenbosch University / University of Pretoria / University of the Witwatersrand / Walter Sisulu University / Other / I prefer not to answer

If other, please specify.

7. Do you have any additional qualifications (postgraduate study) in addition to your medical degree?

Yes / No / I prefer not to answer)

8. Are you actively involved in medical research?

Yes / No / I prefer not to answer

9. What is your sex?

Female / Male / I prefer not to answer

10. What is your age (in years)?

If you prefer not to answer, please type 000

CVD SCREENING EXPOSURE

---

1. Do you currently assess patient CVD risk in your clinical practice?

Yes, No / Sometimes / I prefer not to answer

2. When do you assess patient CVD risk?

Initial consultation of all new patients / At every consultation / When patients present with risk factors / At the discretion of the attending physician / Other / I prefer not to answer

If other, please specify when you assess CVD risk.

3. Do you use a CVD risk stratification tool to assess patient risk for CVD?

Yes / No / Sometimes / I prefer not to answer

4. Which tool do you most often use to assess CVD risk?

Framingham risk score / Globorisk / Pooled Cohort Risk Equation (PCE) / QRISK / Systematic Coronary Risk Evaluation (SCORE) algorithm / WHO-CVD / Other / I prefer not to answer

If other, please specify which tool you use to assess CVD risk

5. How confident do you feel in USING the CVD risk score assessment tool?

Not confident at all / A little confident / Somewhat confident / Moderately confident / Very confident / I prefer not to answer

6. How confident do you feel in INTERPRETING the results of the CVD risk score assessment tool?

Not confident at all / A little confident / Somewhat confident / Moderately confident / Very confident / I prefer not to answer

7. Do you share the results of the risk the CVD risk score assessment with the patient?

Yes / No / Sometimes / I prefer not to answer

8. Does the outcome of the risk score assessment guide?

Yes / No / Sometimes / I prefer not to answer

9. Why do you not assess patient CVD risk?

Patients present with competing health issues / Patients referred from another doctor / Patients already considered high or very high risk for CVD / Insufficient time to assess CVD risk / Lack of funding for testing / Not standard practice / Other / I prefer not to answer

If other, please elaborate why you do not assess CVD risk.

GENETICS AND PRECISION MEDICINE EXPOSURE

---

1. Have you completed genetic or precision medicine related training or education?

Yes / No / I prefer not to answer

2. Please specify which training or education you have completed?*

A genetics module as part of my undergraduate medical training / A precision medicine module as part of my undergraduate medical training / A genetics short course after graduating / A precision medicine short course after graduating / A postgraduate genetics qualification / A postgraduate precision medicine qualification / Other / I prefer not to answer

**Selecting multiple options is possible*

If other, please specify.

3. Have you discussed genetic testing/precision-based medicine with a patient?

Yes / No / Sometimes / I prefer not to answer

4. For which diseases have you discussed genetic testing/precision-based medicine?*

Monogenic disorders (cystic fibrosis, familial hypercholesterolaemia, etc.) / Cancers (Breast, prostate, etc.) / Nutrigenomics / Pharmacogenomics / Cardiovascular diseases /Other / I prefer not to answer

**Selecting multiple options is possible*

If other, please specify.

5. With how many patients do you discuss genetic testing/precision-based medicine with on a monthly basis?

Less than 1 patient per month / 1 patient every 2 to 3 months / 1 patient / 2 to 5 patients / 6 to 10 patients / 11 to 20 patients / More than 20 patients / I prefer not to answer

Level of agreement *Yes / No / I prefer not to answer*

8. Have you OFFERED a patient genetic testing/precision-based medicine to determine their disease risk?
9. Have you ORDERED genetic/precision-medicine testing to determine patient disease risk?
10. Have you REFERRED a patient for genetic/precision medicine testing to determine their disease risk?
11. Have you RETURNED the results of genetic/precision medicine testing to a patient?

12. Have you COUNSELLED patients on the results of their genetic/precision medicine testing?
13. Have you APPLIED RESULTS from genetic tests to the treatment and management (drug choice, dosing, monitoring, etc.) of a patient?
14. How regularly do you apply genetic testing results to the treatment and management of your patient?

Never / Rarely / Sometimes / Often / Always I prefer not to answer)

GENETICS AND PRECISION MEDICINE KNOWLEDGE

Please select the corresponding box that BEST fits your level of agreement with each statement.

Level of understanding *No understanding / Little understanding / Some understanding / Extensive understanding / Expert / I prefer not to answer*

1. Basic genetic principles (Mendelian inheritance, penetrance, somatic vs. germline mutation, etc.).
2. Complex disease genetics (Coronary artery disease, diabetes, etc.
3. Genome Wide Association Studies (GWAS)
4. Polygenic risk scores (PRS)
5. Precision medicine
6. When and how to incorporate precision medicine into practice

GENETICS AND PRECISION MEDICINE KNOWLEDGE

Please select the corresponding box that BEST fits your level of agreement with each statement.

Level of agreement *True / False / Do not know / I prefer not to answer*

1. A complex disease is caused by the interaction of multiple gene variants and environmental factors
2. Modifiable risk factors include smoking, alcohol intake and physical activity levels.
3. Examples of complex diseases include cancer and heart disease.
4. Risk stratification aims to stratify patient populations into high-, medium- and low-risk for improved care and better resource allocation.
5. Precision medicine aims to use an individual's genetic profile and other relevant clinical information to guide decisions regarding disease prevention, diagnosis, and treatment.
6. Precision medicine is a medical model that proposes a 'one-drug-fits-all' model.

7. Polygenic risk is a cumulative measure of genetic risk using multiple genetic variants
8. Polygenic risk scores are transferable across all ethnicities
9. Polygenic risk score results provide insight into a patient's total risk for disease.

PERCEPTIONS TOWARDS A PRECISION MEDICINE-BASED CVD RISK STRATIFICATION APPROACH IN SOUTH AFRICA

In relation to the precision medicine-based CVD risk stratification tool proposed in the video, please select the corresponding box that BEST fits your level of agreement with each statement:

Level of agreement *Strongly disagree / Disagree / Neutral / Agree / Strongly agree / I prefer not to answer*

1. Precision medicine-based CVD risk stratification is relevant to my current clinical practice.
2. Precision medicine-based CVD risk stratification should be applied to my clinical practice.
3. I should be able to provide information on the appropriate use of precision medicine-based CVD risk stratification.
4. Precision medicine-based CVD risk stratification will improve my ability to more effectively prevent and treat CVD.
5. Precision medicine-based CVD risk stratification will improve national resource allocation (expenditure, etc.) for CVD disease prevention and management.

CONCERNS

In relation to the precision medicine-based CVD risk stratification tool proposed, please select the corresponding box that BEST fits your level of concern with each statement:

Level of concern *Not concerned at all / A little concerned / Somewhat concerned / Moderately concerned / Very concerned / I prefer not to answer*

1. The potential for adverse psychological reactions in patients receiving CVD risk information.
2. The impact of genetic variants with unknown significance on the current polygenic risk components.
3. The impact of yet to be identified genetic variants on the outcome of the polygenic risk component of the CVD risk score.
4. The variability of the CVD risk score prediction across populations.
5. The score providing a risk estimate is calculated at a single time point.
6. The possibility of risk estimates changing with age.

7. The potential adverse impacts on life or other insurance policies for patients receiving CVD risk information.
8. The test results' usefulness in disease management and prevention.
9. Do you believe there are additional concerns about the risk score that have not been mentioned above?

No / Yes / I prefer not to answer

If yes, please elaborate.

BENEFITS

In relation to the precision medicine-based CVD risk stratification tool proposed in the video, please select the corresponding box that BEST fits the level of benefit with each statement:

Level of benefit *Not beneficial at all / A little beneficial / Somewhat beneficial / Moderately beneficial / Very beneficial / I prefer not to answer*

10. Access to CVD screening approaches tailored to African populations.
11. The provision of population-specific CVD risk estimates.
12. The provision of genetic and environmental CVD risk information to clinicians.
13. The provision of genetic and environmental CVD risk information to patients.
14. The provision of information for family planning.
15. The provision of information for other family members.
16. Adaption of prevention and treatment strategies based on a patient's CVD risk score.
17. Greater incentive for patients to adopt prevention options including lifestyle changes.
18. Do you believe there are additional benefits about the risk score that have not been mentioned above?

No / Yes / I prefer not to answer

If yes, please elaborate.

19. In relation to the precision medicine-based CVD risk stratification tool proposed, do you believe the expected health benefits outweigh the expected negative consequences to justify undertaking the risk assessment?

Strongly disagree / Disagree / Neutral / agree / Strongly agree / I prefer not to answer

CONFIDENCE IN APPLYING A PRECISION MEDICINE-BASED CVD RISK STRATIFICATION APPROACH

If the precision medicine-based CVD risk stratification tool proposed in the video was available, please select the corresponding box that BEST fits your level of confidence with each statement:

Level of confidence *Not confident at all / A little confident / Somewhat confident / Moderately confident / Very confident / I prefer not to answer*

1. SUGGESTING the CVD risk score assessment for a patient.
2. INTERPRETING the results obtained from the CVD risk score test.
3. SHARING a patient's CVD risk score test results with them.
4. EXPLAINING a PATIENT'S RISK based on their CVD risk score test results.
5. EXPLAINING the potential ADVERSE IMPACTS on life or other insurance policies for patients receiving CVD risk information.

Level of comfort *Not comfortable at all / Somewhat comfortable / Moderately comfortable / Very comfortable / I prefer not to answer*

6. If the precision medicine-based CVD risk stratification tool proposed was available, what is your level of comfort with adapting your PREVENTION STRATEGIES on the patient's CVD risk score results?
7. If the precision medicine-based CVD risk stratification tool proposed was available, what is your level of comfort with adapting your TREATMENT STRATEGIES based on the patient's CVD risk score results?

EXPECTATIONS IN APPLYING A PRECISION MEDICINE-BASED CVD RISK STRATIFICATION APPROACH

If the precision medicine-based CVD risk score is implemented on a population level to guide CVD prevention and screening programs, please select the corresponding box that BEST fits with the role clinicians will have?

Envisioned role *No role / Supporting role / Primary role / Unsure / I prefer not to answer*

1. CHOOSING the test for your patient
2. REFERRING the patient to a genetic specialist or counsellor
3. EXPLAINING the test to patients
4. INTERPRETING test results
5. SHARING results with patients
6. GUIDING treatment approach
7. If available, what is the likelihood that the precision medicine-based CVD risk score will influence your patient care in the future?

No difference / Unlikely / Very unlikely Likely / Very likely / I prefer not to answer

8. If the precision medicine-based CVD risk score is developed, when do you believe such a program could be available in South Africa's public clinical setting?

Less than 1 year / 1 to 2 years / 2 to 5 years / 6 to 10 years / More than 10 years / Not implemented / Other / I prefer not to answer

If other, please elaborate.

9. If the precision medicine-based CVD risk score is implemented nationally to guide prevention and screening programs, who do you believe will need to fund the testing?

National/provincial government / Private medical scheme / Patients / All of the above / Other / I prefer not to answer

If other, please elaborate

BARRIERS TO THE IMPLEMENTATION OF A PRECISION MEDICINE-BASED CVD RISK STRATIFICATION APPROACH IN PUBLIC PRACTICE

Please select the corresponding box that BEST fits the level of impact each item may have on the ability to incorporate the precision medicine-based CVD risk stratification score into public practice.

Level of impact *No impact / Little impact / Some impact / Moderate impact / Strong impact / I prefer not to answer*

1. Shortage of healthcare personnel
2. Shortage of genetic service capacity (specialists, counsellors, scientists, etc.)
3. Limited access to genetic services
4. Limited awareness of precision medicine
5. Lack of precision medicine training or education
6. Lack of clinical guidelines on precision-medicine based practice
7. Competing health interests
8. Physician time constraints
9. Time required for patient education
10. Cost of genomic testing
11. Lack of a willing funder
12. Questions about insurance coverage of the testing
13. Patient's concerns about moral/ethical implications
14. Cultural perceptions of genomic testing
15. Do you believe there are additional barriers to incorporating the score than those not mentioned above?

No / Yes / I prefer not to answer

If other, please specify.

S12. Correlation matrix of study predictor variables for (A) Mean knowledge score (n=107), (B) Mean perception score (n=105), and (C) Mean confidence score (n=103)

A. Mean knowledge score with all study predictors

	MKS	Medical specialty	Sex	Medical group	Practice level	Clinical experience (years)	Postgraduate qualifications	Medical research involvement	Genetics training	CVD screening
MKS		0.05	0.06	0.1	-0.05	-0.08	0.16	0.23	0.49	0.16
Medical specialty	0.05		-0.11	0.83	-0.74	-0.59	-0.42	-0.51	0.04	-0.29
Sex	0.06	-0.11		-0.1	0.14	0.15	-0.02	0.02	0.08	-0.01
Medical group	0.1	0.83	-0.1		-0.84	0.57	-0.42	0.5	-0.01	-0.03
Practice level	-0.05	-0.74	0.14	-0.84		0.82	0.48	0.43	0.07	0
Clinical experience (years)	-0.08	-0.59	0.15	0.57	0.82		0.43	0.22	-0.07	0
Postgraduate qualifications	0.16	-0.42	-0.02	-0.42	0.48	0.43		0.37	0.03	0.05
Medical research involvement	0.23	-0.51	0.02	0.5	0.43	0.22	0.37		0.19	0
Genetics training	0.49	0.04	0.08	-0.01	0.07	-0.07	0.03	0.19		0.07
CVD screening	0.16	-0.29	-0.01	-0.03	0	0	0.05	0	0.07	

B. Mean perception score with all study predictors

	MPS	Medical specialty	Sex	Medical group	Practice level	Clinical experience (years)	Postgraduate qualifications	Medical research involvement	Genetics training	CVD screening
MPS		-0.13	-0.11	0	-0.02	-0.01	0.05	0.16	0.18	0.37
Medical specialty	-0.13		-0.11	0.83	-0.75	-0.59	-0.42	-0.52	0.04	-0.29
Sex	-0.11	-0.11		-0.1	0.14	0.14	-0.02	0.02	0.08	-0.01
Medical group	0	0.83	-0.1		-0.85	-0.58	-0.42	-0.51	-0.01	-0.03
Practice level	-0.02	-0.75	0.14	-0.85		0.82	0.49	0.43	0.07	0
Clinical experience (years)	-0.01	-0.59	0.14	-0.58	0.82		0.44	0.22	-0.08	0
Postgraduate qualifications	0.05	-0.42	-0.02	-0.42	0.49	0.44		0.38	0.04	0.05
Medical research involvement	0.16	-0.52	0.02	-0.51	0.43	0.22	0.38		0.19	0
Genetics training	0.18	0.04	0.08	-0.01	0.07	-0.08	0.04	0.19		0.07
CVD screening	0.37	-0.29	-0.01	-0.03	0	0	0.05	0	0.07	

C. Mean confidence score with all study predictors

	MCS	Medical specialty	Sex	Medical group	Practice level	Clinical experience (years)	Postgraduate qualifications	Medical research involvement	Genetics training	CVD screening
MCS		-0.11	0.08	0	-0.05	-0.02	0.04	0.26	0.06	0.27
Medical specialty	-0.11		-0.11	0.83	-0.75	-0.59	-0.43	-0.52	0.04	-0.3
Sex	0.08	-0.11		-0.09	0.13	0.12	0.02	0.04	0.09	0.02
Medical group	0	0.83	-0.09		-0.85	-0.57	-0.44	-0.53	-0.02	-0.04
Practice level	-0.05	-0.75	0.13	-0.85		0.82	0.51	0.45	0.08	0.01
Clinical experience (years)	-0.02	-0.59	0.12	-0.57	0.82		0.48	0.22	-0.08	0
Postgraduate qualifications	0.04	-0.43	0.02	-0.44	0.51	0.48		0.36	0.03	0.02
Medical research involvement	0.26	-0.52	0.04	-0.53	0.45	0.22	0.36		0.18	-0.03
Genetics training	0.06	0.04	0.09	-0.02	0.08	-0.08	0.03	0.18		0.06
CVD screening	0.27	-0.3	0.02	-0.04	0.01	0	0.02	-0.03	0.06	

S13: Multivariate analysis of respondents' mean (SD) knowledge, confidence, and perception scores (n =104)

Factor	Effect group	Reference group	Model coefficients *(95% CI)*	Standard error	Standardized coefficients (β)	*p* value
Mean knowledge score *(F-stat: 6.555, p-value: 8.715 x 10^{-7}, adj R-squared: 0.3014)*						
Medical group	Trainee	Clinician	0.849 *(0.15 – 0.84)*	0.176	0.326	*5.75 x 10^{-3} ****
Gender	Male	Female	0.083 *(-0.14 - 0.30)*	0.111	0.062	*0.459*
Clinical experience	≥ 8 years	< 8 years	0.021 *(-0.25 - 0.29)*	0.136	0.016	*0.879*
Postgraduate qualification	Yes	No	0.242 *(-0.02 - 0.50)*	0.130	0.183	*0.065*
Medical research involvement	Yes	No	0.290 *(0.03-0.55)*	0.132	0.223	*0.030 **
CVD screening	Yes Sometimes	No	0.203 *(0.05 - 0.45)* 0.074 *(-0.20 - 0.35)*	0.127 0.137	0.152 0.052	*0.114* *0.592*
Genetics training	Yes	No	0.758 *(0.46-1.06)*	0.151	0.434	*2.48 x 10^{-6} ****
Mean perception score *(F-stat: 2.962, p-value: 5.391 x 10^{-3}, adj R-squared: 0.1357)*						
Medical group	Trainee	Clinician	0.283 (-0.28 – 0.84)	0.281	0.133	*0.317*
Gender	Male	Female	-0.216 (-0.57 – 0.14)	0.179	-0.115	*0.227*
Clinical experience	≥ 8 years	< 8 years	0.088 (-0.36 - 0.53)	0.224	0.048	*0.694*
Postgraduate qualification	Yes	No	0.020 (-0.40 - 0.44)	0.209	0.011	*0.923*
Medical research involvement	Yes	No	0.371 (-0.05 - 0.79)	0.213	0.201	*0.085*
CVD screening	Yes Sometimes	No	0. 746 (0.34 - 1.15) 0.533 (-0.09 - 0.97)	0.205 0.222	0.396 0.263	*4.49 x 10^{-4} **** *0.018 **
Genetics training	Yes	No	0.349 (-0.14 - 0.84)	0.247	0.139	*0.161*
Mean confidence score *(F-stat: 2.797, p-value: 8. 941 x 10^{-3}, adj R-squared: 0.1279)*						
Medical group	Trainee	Clinician	0.658 (0.00 – 1.32)	0.331	0.271	*0.050 **

Gender	Male	Female	-0.094 (-0.31 - 0.50)	0.183	0.044	*0.1932*
Clinical experience	≥ 8 years	< 8 years	0.034 (-0.54 - 0.48)	0.257	-0.016	*0.8935*
Postgraduate qualification	Yes	No	-0.010 (0.050 - 0.48)	0.247	-0.004	*0.969*
Medical research involvement	Yes	No	0.928 (0.43 - 1.43)	0.252	0.443	*3.91 x 10^{-4} ****
CVD screening	Yes Sometimes	No	0.635 (0.17 - 1.10) 0.369 (-0.14 - 0.88)	0.236 0.258	0.298 0.162	*8.46 x 10^{-3} **** *0.155*
Genetics training	Yes	No	-0.058 (-0.61 - 0.50)	0.279	-0.020	*0.837*

1

2

S14: Clinician's responses relating to their expectations of a PM-based CVD risk stratification tool in the South African public health setting)

Variable	Variable options	Nn	%
Expected role (N =104)			
Choose the CVD risk score	Primary	59	56.7
	Supporting	37	35.6
	Unsure	5	4.8
	No role	2	1.9
	I prefer not to answer	1	1.0
Refer a patient for the CVD score	Primary	75	72.1
	Supporting	23	22.1
	Unsure	3	2.9
	No role	2	1.9
	I prefer not to answer	1	1.0
Explain CVD score results	Primary	47	45.2
	Supporting	51	49
	Unsure	3	2.9
	No role	2	1.9
	I prefer not to answer	1	1.0
Interpret CVD score results	Primary	46	44.2
	Supporting	45	43.3
	Unsure	5	4.8
	No role	7	6.7
	I prefer not to answer	1	1
Share CVD score results	Primary	64	61.5
	Supporting	35	33.7
	Unsure	2	1.9
	No role	4	3.8
	I prefer not to answer	1	1.0
Use the CVD score results to guide care approach	Primary	72	69.2
	Supporting	24	23.1
	Unsure	3	2.9
	None	4	3.8
	I prefer not to answer	1	1.0
Likely behaviour shift (N = 100)			
If available, likelihood CVD score will influence patient care	Very likely	35	35.0
	Likely	54	54.0
	No difference	4	4.0
	Unlikely	4	4.0
	Very unlikely	3	3.0
Funder (N =100)			
Funder	National/provincial governments	36	36.0
	Private medical scheme	6	6.0
	Patients	4	4.0
	All of the above (Governments, Private medical schemes, and patients)	52	52.0
	Other	2	2.0
Time horizon (N = 101)			
Time horizon	Less than 1 year	1	1.0
	1 to 2 years	4	4.0
	2 to 5 years	21	20.8
	6 to 10 years	39	38.6
	More than 10 years	36	25.6

Publication 4: supplementary materials

S15: Search strategy

Pubmed: Search: (comorbidity[Title/Abstract] OR comorbidity [Title/Abstract] OR multimorbidity[Title/Abstract] OR multimorbidity[Title/Abstract] OR "multiple chronic conditions"[Title/Abstract] OR "chronic conditions"[Title/Abstract] OR "chronic disease"[Title/Abstract] OR multipathology[Title/Abstract] OR polypathology[Title/Abstract]) AND (Africa OR African OR "African population" OR "African ancestry" OR "African-ancestry") Filters: Clinical Study, Comparative Study, Multicenter Study, Observational Study, Twin Study, Validation Study, Adult: 19+ years, from 2000/1/1 - 2022/6/30

S16: Variables captured during data capture

No.	Study Characteristic	Characteristic variables (where applicable)
1	Author(s)	
2	Author affiliation	
3	Publication year	
4	Study title	
5	Geographic region	Central Africa, East Africa, North America, etc.
6	Geographic location	Botswana, Malawi, Nigeria, UK, USA., etc.
7	African ancestry population	Continental, Diaspora, Both
8	Aim (and objective)	
9	Study design	Cross-sectional, Experimental, Longitudinal, etc.
10	Type of research	Primary, secondary, Both
11	Methodological approach	
12	Cohort/Database	
13	Cohort type	Hospital-based, Population-based, Clinical database, etc.
14	Data collection dates	
15	Data collected	Socio-demographic, health history, medication, etc.
16	Statistical analysis technique	Regression analyses, machine learning, etc.
17	Sample size	
18	Ancestry distribution	European, African, Hispanic, etc.
19	Age	
20	Mean age (SD)	
21	Sex distribution	Male, Female, Unknown
22	HIV distribution	HIV +, HIV -, NA
23	Urbanicity	Urban, Rural, Both
24	MM definition	
25	MM measure	Disease counts, weighted indices, etc.
26	MM prevalence	
27	No. of conditions assessed	
28	Diseases assessed	
29	Disease clusters assessed	
30	Prevalence of each disease	
31	Outcomes	
32	Exposure/mediators	
33	Key findings	

1 S17: United Nations African subregions

African region	Included countries
Northern	Algeria, Ceuta, Egypt, Libya, Melilla, Morocco, Sudan, Tunisia
Eastern	Burundi, Comoros, Djibouti, Eritrea, Ethiopia, French Southern Territories, Kenya, Madagascar, Malawi, Mauritius, Mayotte, Mozambique, Reunion, Rwanda, Seychelles, Somalia, Somaliland, South Sudan, Tanzania, Uganda, Zambia, Zimbabwe
Central	Angola, Cameroon, Central African Republic, Chad, Congo, Democratic Republic of the Congo, Republic of the Equatorial Guinea, Gabon, São Tomé and Príncipe
Southern	Botswana, Eswatini (Swaziland), Lesotho, Namibia, South Africa
Western	Benin, Burkina Faso, Cabo Verde (Cape Verde), Côte d'Ivoire (Ivory Coast), The Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Saint Helena, Ascension and Tristan da Cunha, Senegal, Sierra Leone, Togo

2

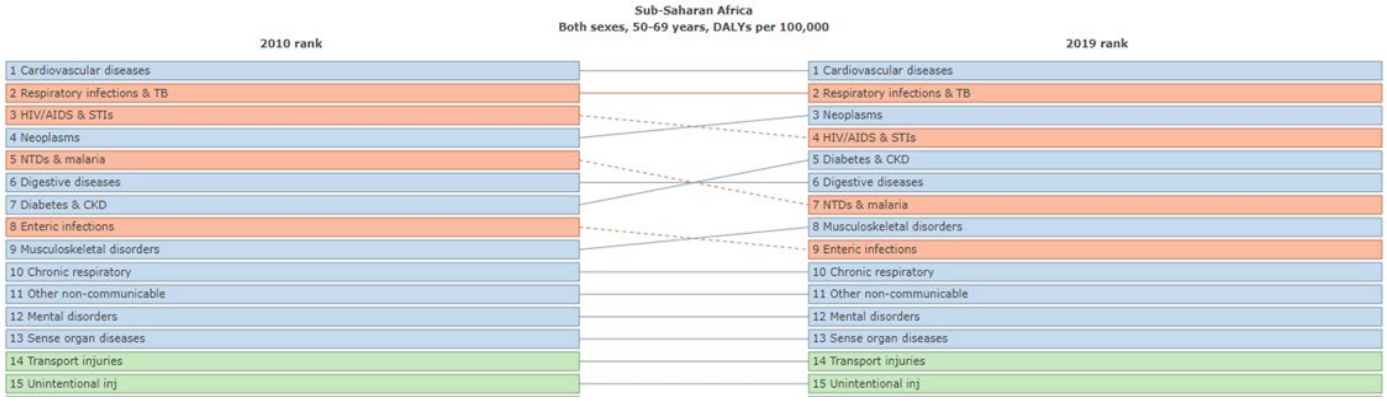
3

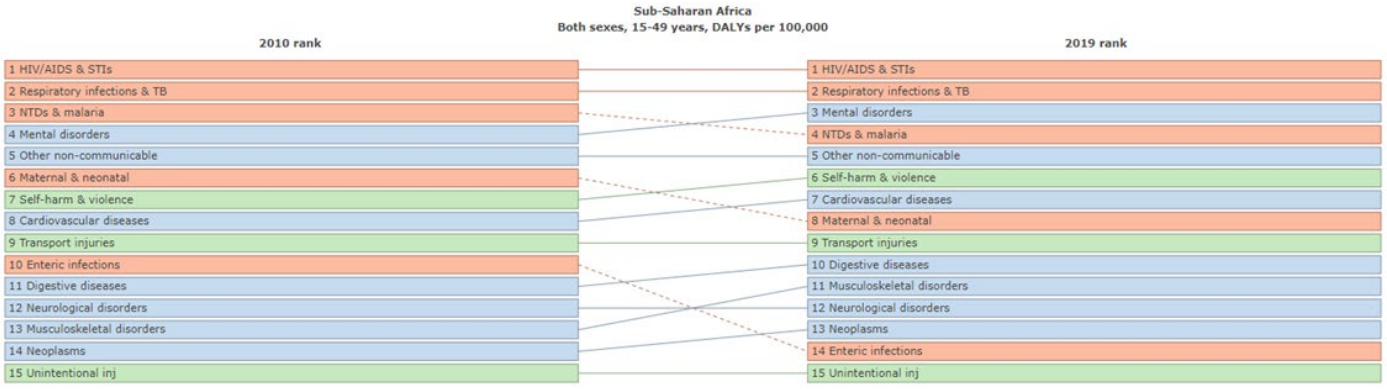
S18: IHME Mortality and morbidity trends

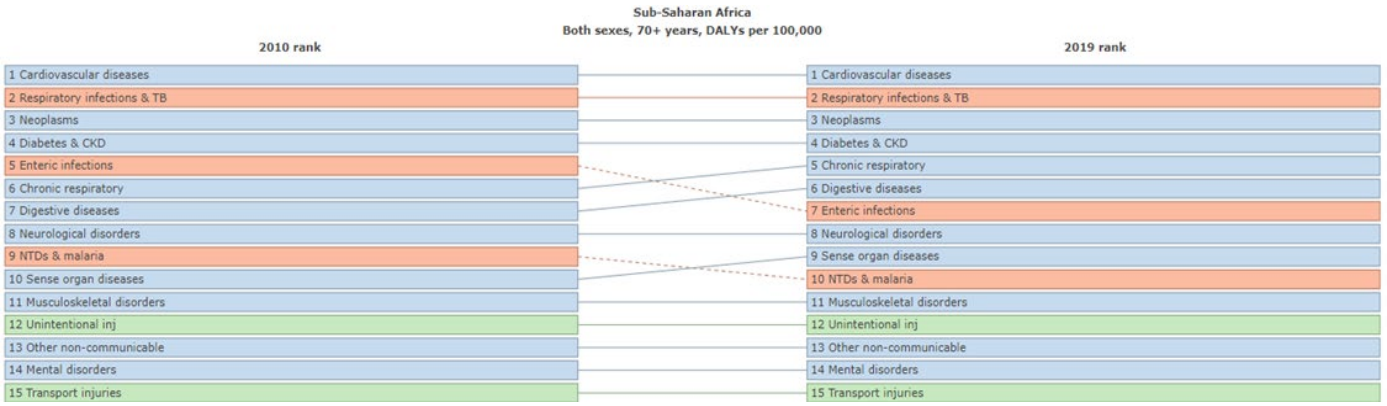
Top 15 disease groups responsible for morbidity, measured as Disability-adjusted life years (DALYs) in A. Sub Saharan Africa and B. Globally, between 2010 and 2019

Data obtained from the Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualisation. Seattle, WA: IHME, University of Washington, 2020.

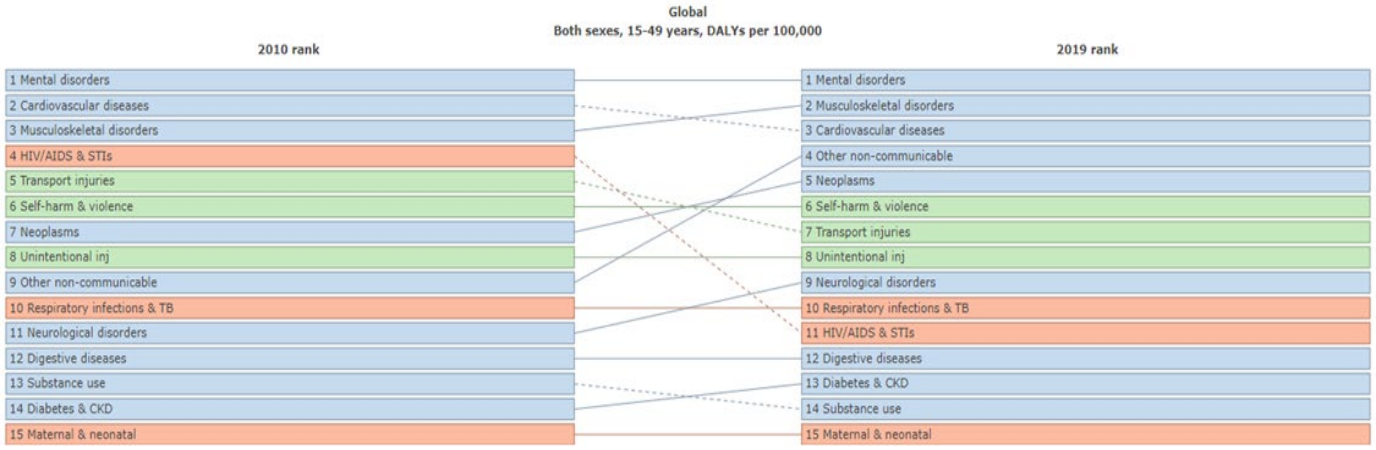
A: Diseases contributing to DALYs in Sub-Saharan Africa between 2010 to 2019 by age.



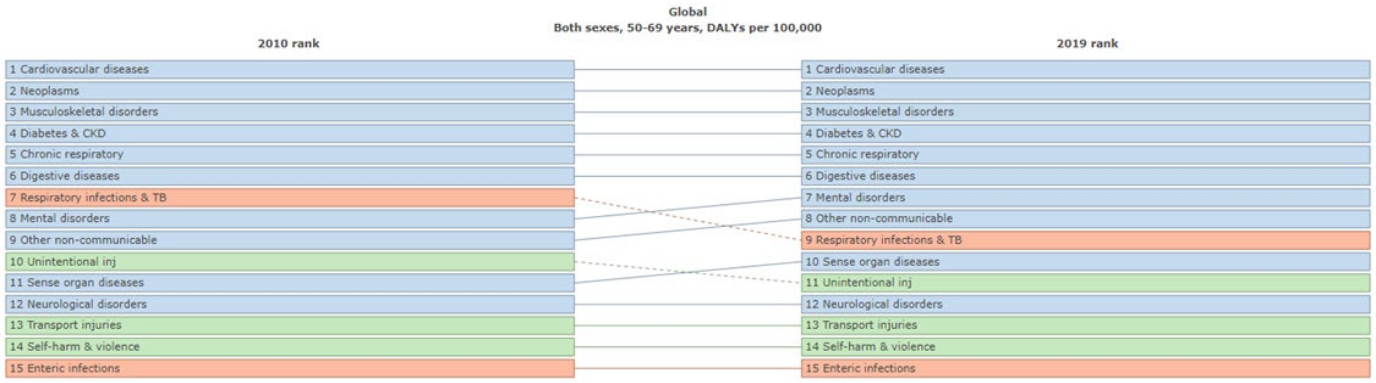




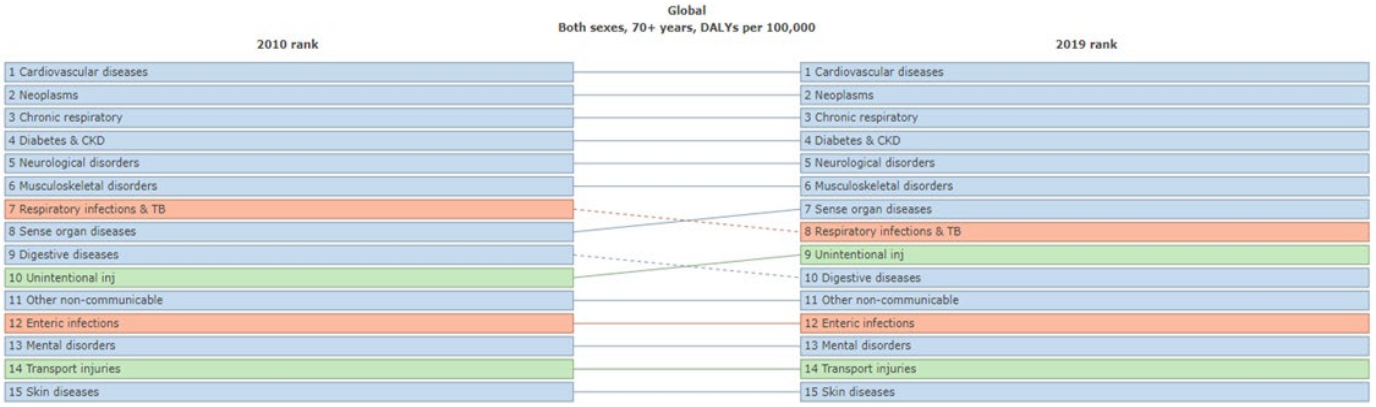
B: Diseases contributing to DALYs globally between 2010 to 2019 by age.



2



3



4

5

6

7

8

9

APPENDIX G: ORIGINAL PUBLICATION 1

OXFORD

Human Molecular Genetics, 2021, Vol. 00, No. 00 1–10

https://doi.org/10.1093/hmg/ddab180
Advance Access Publication Date: 8 July 2021
Invited Review Article

INVITED REVIEW ARTICLE

Has translational genomics come of age in Africa?

Michelle Kamp[1,2], Amanda Krause[1] and Michele Ramsay[1,2,*]

[1]Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, The University of the Witwatersrand, Johannesburg 2193, South Africa and [2]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa

*To whom correspondence should be addressed at: Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa. Tel: (+27) 717 6635; Email: Michele.ramsay@wits.ac.za

Abstract

The rapid increase in genomics research in Africa and the growing promise of precision public health (PPH) begs the question of whether African genomics has come of age and is being translated into improved healthcare for Africans. An assessment of the continent's readiness suggests that genetic service delivery remains limited and extremely fragile. The paucity of data on mutation profiles for monogenic disorders and lack of large genome-wide association cohorts for complex traits in African populations is a significant barrier, coupled with extreme genetic variation across different regions and ethnic groups. Data from many different populations are essential to developing appropriate genetic services. Of the proposed genetic service delivery models currently used in Africa—Uncharacterized, Limited, Disease-focused, Emerging and Established—the first three best describe the situation in most African countries. Implementation is fraught with difficulties related to the scarcity of an appropriately skilled medical genetic workforce, limited infrastructure and processes, insufficient health funding and lack of political support, and overstretched health systems. There is a strong nucleus of determined and optimistic clinicians and scientists with a clear vision, and there is a hope for innovative solutions and technological leapfrogging. However, a multi-dimensional approach with active interventions to stimulate genomic research, clinical genetics and overarching healthcare systems is needed to reduce genetic service inequalities and accelerate PPH on the continent. Human and infrastructure capacity development, dedicated funding, political will and supporting legislation, and public education and awareness, are critical elements for success. Africa-relevant genomic and related health economics research remains imperative with an overarching need to translate knowledge into improved healthcare. Given the limited data and genetic services across most of Africa, the continent has not yet come of 'genomics' age.

Introduction

The past decade has seen significant growth in genomics research in Africa, in large part as a result of the activities of the African Society of Human Genetics (AfSHG) (1–3) and the Human Heredity and Health in Africa Consortium (H3Africa) (4–6). However, the question remains whether the genomic revolution and African genome research have translated into direct benefits to individuals and families affected by genetic disorders.

Access to genetic technologies and genomic services remain sparse in Africa for several reasons. These include a lack of financial resources, poor national health services and infrastructure and the presence of more compelling health priorities, such as infectious diseases like tuberculosis, malaria, HIV/AIDS (5,7) and now COVID-19. Furthermore, few countries in Africa have formally recognized genetic services supported by clinically trained geneticists and dedicated diagnostic laboratory infrastructure.

Received: May 20, 2021. Revised: June 27, 2021. Accepted: June 29, 2021

The field of genomics holds great promise for health care and medical research, as the identification of the genetic determinants of diseases or phenotypes bring about the potential for significant improvements in diagnosis, prevention and treatment of disease. Genomics is particularly attractive for Africa, where new technologies and products from genomics research offer the potential for a precision medicine approach to help mitigate the heavy burden of infectious and non-communicable diseases, as well as monogenic disorders (8,9). The availability of large genomic datasets could feed into precision public health (PPH) initiatives to improve population health by translating these new genomics and digital technologies, to guide public health practice through the generation of population-wide tailored interventions and policies.

From a clinical genetics perspective, interventions could initially focus on the identification and molecular characterization of monogenic disorders in specific African regions, such that their frequencies and molecular profiles in different areas are well characterized and defined using genomic technologies. Limited studies on sickle cell anaemia (10), fragile X syndrome (11–14), hearing loss (15–19), Huntington's disease (20,21), familial cancer syndromes (22,23) and others, have been performed and show the value of such studies, particularly highlighting the diversity in Africa. In addition, factors that modulate disease severity should be studied (as reviewed in (22–25)). Simultaneously, a profile of common variation in these populations would be established, providing a critical resource for interpreting and characterizing the potential functional impact of variants and their disease associations.

Thus, a two-pronged strategy, one firmly rooted in fundamental principles and the other leveraging state-of-the-art technologies, could propel clinical genetics to new levels in Africa. The first would involve collecting basic data on the prevalence and phenotype of monogenic diseases. The second would be to use next-generation sequencing approaches to establish population architecture and variation and subsequently identify causal mutations and their geographical distribution. In parallel, the development of the computational skills required for the analysis and interpretation of these variants is needed, as is digital innovation in data capture, storage and integration (26,27).

Worldwide, genetic testing is becoming increasingly complex as it inches closer towards the era of precision medicine, where genomic medicine becomes integrated into routine healthcare. However, despite these exciting developments, the fear is that existing health care inequalities are exacerbated as technologies and discoveries emanating from genomics research are not translatable across populations and remain accessible only to those who can afford them (28–31). Africa has an urgent need to develop innovative approaches and to provide timeous and effective genetic services to patients in resource-constrained environments.

While acknowledging substantial knowledge, capacity and sustainability gaps, it is clear that the effective implementation of PPH has the potential to translate into more effective patient care. This review highlights the current state of genetics and genomics on the continent, describing how genetic services are currently delivered and how genomic and digital technologies underpin these services. The goal is to identify key factors needed to ensure that clinical genetic and laboratory services have broad benefit in Africa in the era of precision medicine.

Genetic Service Delivery Models

In order to gauge the readiness for genomic medicine in Africa, an assessment of the current genetic services offered is required. This will, in turn, assist in identifying the factors necessary to drive a PPH agenda on the continent. Although there is limited published information about the service delivery models in use (apart from some from South Africa (32,33), Egypt (34), Morocco (35), Tunisia (36,37), Nigeria (38) and Rwanda (39)) preliminary exploration suggests the presence of varying strategies from one country to the next. This variability is a direct consequence of the heterogeneity in healthcare contexts, and in the case of North Africa, a response to the consequences of high consanguinity levels (25,34,35,37).

Informal discussions with individuals identified by the authors as being well informed about the status of genetic service delivery in their respective countries explored the following areas: (1) the availability of a clinical genetics service in the country; (2) the level of infrastructure and local capacity available for service delivery and (3) the evaluation of the service within its given context.

These discussions revealed that genetic service delivery remains very limited and extremely fragile in Africa. Genetic services are fraught with difficulties relating to the scarcity of an appropriately skilled dedicated medical genetic workforce, limited formalization of structures, insufficient health funding and political support, and overstretched health systems contending with competing health priorities (32,35,38,40). Difficulties are exacerbated when cultural and religious aspects that may influence or limit engagement with genetic services are also considered (37,41–43). These have not been explored at any significant level. Furthermore, even where services do exist, access to clinical genetic services remains poor, with services often restricted to major urban centres leaving a large proportion of the population of Africa underserved (40,44).

The current practices across African countries have been evaluated by describing the prevailing service delivery models. We propose that there are five genetic service delivery models currently employed in Africa, namely: *Uncharacterized*, *Limited*, *Disease focused*, *Emerging* and *Established*, each differentiated by (i) specialist access (accessibility to clinical geneticists and genetic counselling vs. primary care physicians/other medical specialists, such as haematologists and paediatricians); (ii) availability of laboratory testing to confirm diagnoses and (iii) nature of laboratory testing (where and by whom testing is performed and the type of technology used). The models, described in Figure 1, reflect successive points on a continuum from no service provision (or unknown) to an established comprehensive service.

Uncharacterized, Limited and Disease-focused represent the majority of models currently operational on the continent. The '*Advanced*' model is deemed necessary to implement a PPH agenda effectively. However, our belief is that this model is not yet operational in Africa. The Advanced model has formalized services, and all components are present to provide access appropriate for given populations. Also, services are offered beyond just the diagnosis of monogenic conditions and include broad disease prevention strategies (screening, availability of prenatal diagnosis, pre-clinical diagnosis and genetic risk stratification). In addition, clinical services need to be operating within genomic-learning healthcare systems.

The models we have assessed and described relate to the public sector only. Across most regions private healthcare is available but, generaly serves a small and wealthy segment of the population. In general, private care does afford additional testing, for instance, newborn screening programs, and often uses overseas testing facilities when legislation permits this. In most settings, the public sector serves the majority of the population and requires local genetic services capacity and

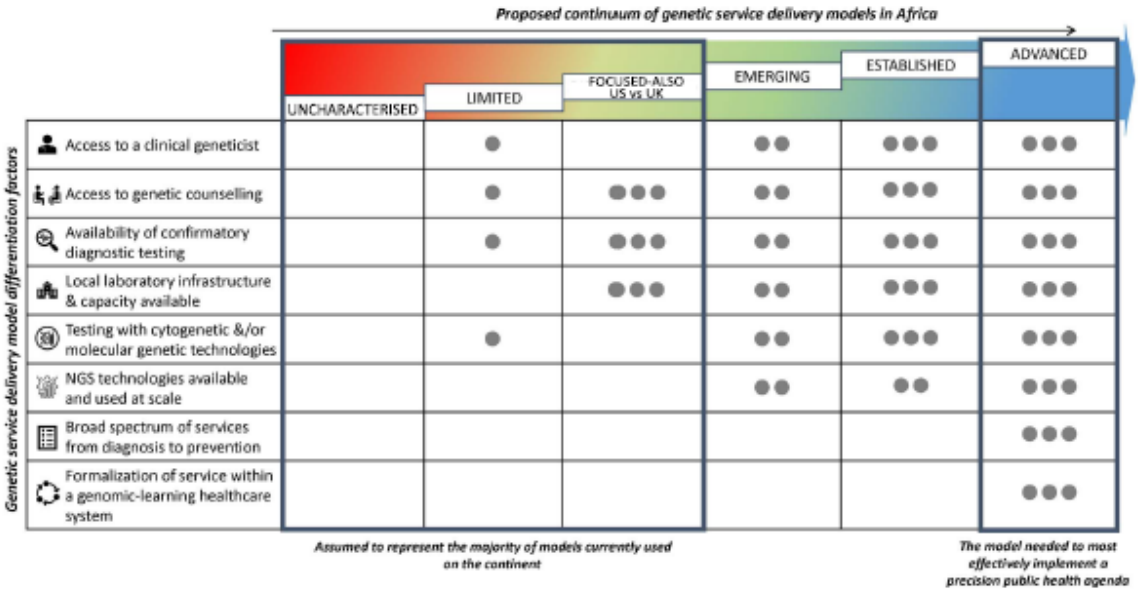


Figure 1. The proposed continuum of genetic service delivery models in Africa: We propose that five genetic service delivery models are currently utilised in Africa. 1. Uncharacterised—No service offered or state of service unknown. Limited clinical genetics service may be offered on a non-systematic basis. Individuals involved typically rely on international research collaborations. Few or no resources available for in-country laboratory testing. 2. Limited—Genetic services are provided by non-genetic specialists. Clinical diagnoses are not regularly confirmed with laboratory testing. Genetic testing, when requested, is conducted overseas, often leveraging research collaborations and networks. Services are often limited to centres where interested specialists are located—uneven distribution across city/province/region/country. 3. Disease focused—Directed genetic services provided by non-genetic specialists in response to high health burden. Clinical diagnoses confirmed with basic laboratory testing. Testing done within the country by locally trained individuals. 4. Emerging—Small number of clinical geneticists provide genetic services. Some clinical diagnoses confirmed by genetic testing. Cytogenetics and limited molecular genetics technologies are available. Testing is conducted both locally and/or internationally depending on the test complexity and the technology required or available. 5. Established—Several clinical geneticists provide formalized genetic services and do so in collaboration with other specialists. Clinical diagnoses are regularly confirmed by laboratory testing. Cytogenetics and molecular genetics technologies are available for testing for an array of common diseases. The majority of testing is done within the country by locally trained capacity.

supportive formal structures. In most cases, even in established settings, legislation and policy development relating to service provision is awaiting approval or is not yet developed.

Across all models, services are geared towards confirming diagnosis of common monogenic disorders. Disease prevention strategies, including population screening, risk stratification programs and services associated with complex diseases, are essentially non-existent. Furthermore, testing is focused on affected probands, and subsequent cascade and prenatal diagnosis testing are rare, particularly in limited and disease-focused service settings, due to the lack of financing for such testing or the fear or perceived fear of discrimination amongst family members (42).

In countries with limited or disease-focused models (e.g. focused on sickle cell disease in Central, West and East Africa), genetic services do not extend beyond the common disease of focus and are associated with non-genetic medical specialties, mostly haematology, paediatrics and pathology. In emerging and established settings, clinical geneticists are active in the system but often at very low numbers and genetic counselling capacity is severely inadequate (35,45) with access to a trained and certified genetic counsellor only available in South Africa (32,46) or through clinicians (34–36,38,39). In almost all settings there is a 'key person(s)' risk, whereby one or a few individuals are at the core, driving the provision of clinical genetics services and, without them, the system would likely not be sustained. Furthermore, the retention of medical genetics skills remains challenging, with several countries unable to stem the 'brain drain' (32). This loss is entrenched by a lack of political will, employment and career progression opportunities and limited government financing of the discipline and healthcare system as a whole (9,33). Formal structures and posts are limited or non-existent in most African countries.

Additional barriers impeding genetic services and the more widespread adoption of genomics in Africa are highlighted in Table 1 and can be classified according to limited opportunities across four dimensions: awareness, accessibility, affordability and applicability.

Despite the significant challenges, the African genetics community is generally optimistic about the future of genetic service delivery and the role genomics could play in improving healthcare on the continent (9,54). Africa's clinicians pride themselves on providing a clinical service to their patients and doing so at a level that is on par with international peers despite their resource-constrained environments. However, only a small sector of the population has access to these services (52). The novel discoveries and active research communities that continue to contribute to the global genetics knowledge-base despite a challenging operating environment are also a source of optimism.

Moving towards PPH will require new and practical strategies to increase accessibility to genetics and genomics services and

Table 1. Barriers impeding genetic services in Africa

Awareness
- **Patients and healthcare providers'** limited knowledge about the availability and benefit of genetic testing restricts testing uptake and limits the understanding of the true burden of genetic disease (47–49).
- Lack of awareness amongst the **general public** contributes to discrimination and social exclusion of affected individuals (42,43).
- Funding gaps and a lack of appropriate legislative and regulatory frameworks have resulted from the lack of knowledge and awareness amongst **policymakers and key decision makers** (44,50).

Accessibility
- Genetic services are not widely available (32,43,44), and significant clinical genetics and counselling **workforce shortages** exist (40,45,48). Limited access to adequate healthcare services, especially in rural areas, exacerbates accessibility challenges as the referral system is compromised (36).
- **Scarce laboratory infrastructure** and **poor financial support** for genetic testing and genomic technologies limit the opportunity to broaden genetic services and benefit from translational genomic innovations, particularly those associated with disease prevention and patient management and care (38,44,50,51).
- African healthcare systems **lack the multidisciplinary medical specialty resources** required for effective genetic disorder patient management and care (35).

Affordability
- **Financial barriers** are a common hindrance to patients' utilization of genetic services. National health insurance or subsidies may lower burden but not all tests are covered by national insurance (including cascade testing).
- Despite costing decreases in sequencing and associated genomic technologies, they remain out of reach for most African laboratories, especially when coupled with the **additional costs, time delays and complex logistics** associated with the importation of reagents, and servicing and maintenance of laboratory equipment (39,44).
- Limited financial support from national governments for laboratory testing and clinical posts results in a vulnerable system that has a heavy **reliance on grant and academic institutional** funding to subsidize funding shortfalls.

Applicability
- Africa's burden of genetic disorders remains largely unknown (38,52) and reliable epidemiological data is not available (53). This limited understanding of **burden of genetic disease** restricts development and subsequent prioritization of genetic service delivery (44,52).
- Limited **transferability of mutation profiles** for monogenic disorders and genetic associations for complex traits across populations impedes the uptake of genomic interventions into clinical practice. This is exacerbated by already low levels of translation from genomic research.
- Need for a **culturally competent service** that considers and respects the culture, values, and traditions of the individuals and communities accessing genetic services and contributing to genomic research (7,28,29).
- Improving the uptake of genomic medicine practice innovations and subsequent quality of clinical care requires **locally relevant** genetic tests and programs (30,31,54).
- Genomic innovations must be successfully translated, scientifically validated and their clinical utility demonstrated in each context. This requires a supportive and robust genomics research community (44,54).

to reduce regional or social inequalities. Discussions with the African clinical genetics community revealed that several critical factors are needed to enhance genetic services on the continent; namely, capacity development (human resources and infrastructure), dedicated funding and support, political will and supporting legislation, public education and awareness, and Africa-relevant genomic and related health economics research, with an overarching need to translate knowledge into improved precision health care.

Accelerating the PPH Agenda for Africa

Clinical genetics services do not operate in isolation but are nested within a larger healthcare system, which is increasingly translating knowledge gained from genomic research into mainstream health care (50,55,56). On the African continent, there is a need for active intervention in genomic research and clinical genetics, as well as in the overarching healthcare systems within which they function, to reduce genetic service inequalities and accelerate PPH (52,55).

Health systems strengthening is considered fundamental to the successful implementation of genomic medicine in Africa as a functional and well-resourced healthcare setting will provide the platform necessary to elevate Africa's genetic services (5). Factors identified as necessary for health system strengthening include capacity building, increased funding for health, and more political support and commitment (38,57). These factors, along with patient education and awareness, need to be directed specifically to improving clinical services and genomic and clinical research on the continent (5,52,54). Additional accelerators unique to the research and services domain will also be required to advance PPH in Africa. These above-mentioned cross-cutting factors will be described further in the following sections and are illustrated in Figure 2.

Capacity and infrastructure

African countries vary significantly in the availability of genetics and genomics clinical training programs and infrastructure (44,48). Genomics research and clinical genetics capacity building, both on a human resources and infrastructure level, is

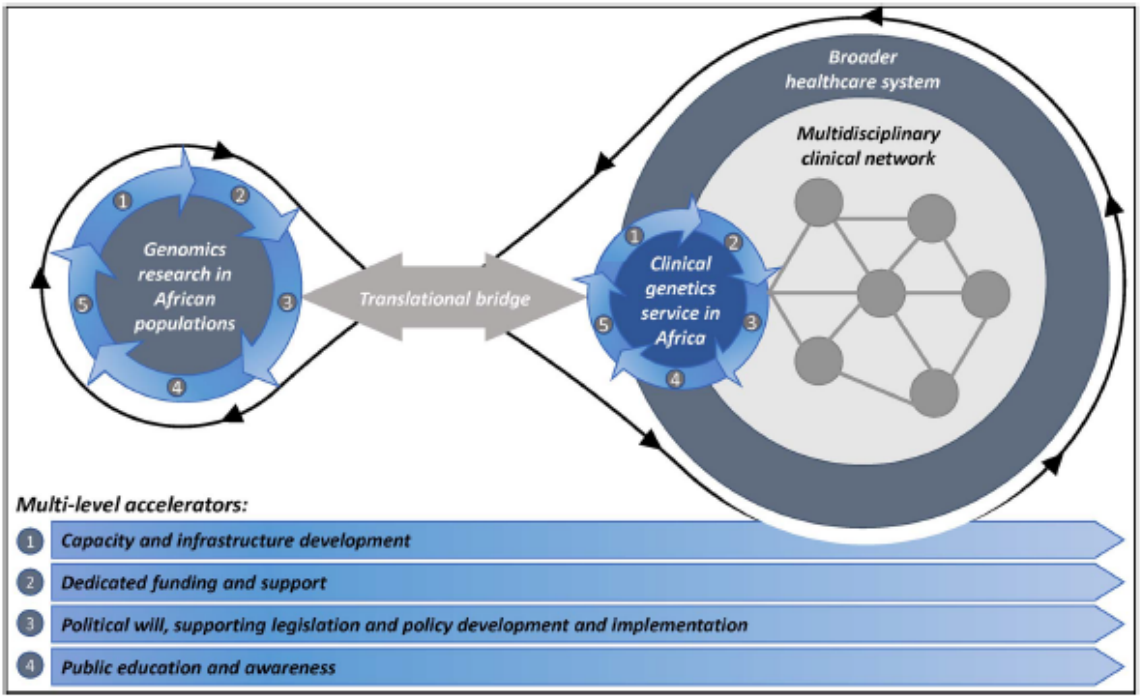


Figure 2. Accelerating the PPH agenda in Africa requires a multi-dimensional approach: In addition to the multi-dimensional accelerators (capacity and infrastructure, dedicated funding and support, political will and supporting legislation), accelerators are required at the genomic research level (Population genetics and variant profiles, Monogenic disease mutation profiles, Complex disease genomics, Infectious diseases and host response, Pharmacogenomics and Databases & data sharing) and at the clinical genetics level (Patient registries, Prevalence data on genetic disorders, Broad open access, Mutation databases and Healthcare integration). Research discoveries should translate into clinical practice that is beneficial to a broader multidisciplinary clinical network. A virtuous cycle between the research environment and healthcare system should exist where genomic learning is active and continuous, and each domain self-improves after successive advances.

urgently required to develop the critical mass of genomics and healthcare expertise needed for translational genomics and precision medicine (2,3,39,48).

Few African countries provide formalized training in clinical genetics and even fewer for genetic counselling (32,33,45). Most clinical geneticists in central and west Africa have been trained at international academic institutions, mainly in Europe. However, the tide is changing, and more African talent is being developed within Africa. Most local and continental genomics training initiatives, such as the H3Africa Consortium (www.h3africa.org) and the African Academy of Sciences (www.aasciences.africa), have promoted and accelerated training opportunities across the continent and contribute to the growing genetics knowledge and exposure amongst the medical fraternity. The African Genomic Medicine Training Initiative (AGMT) (48) and the West African Genetic Medicine Centre (WAGMC) (https://wagmc.org) promise additional training opportunities, providing genomic medicine training for nurses and targeted education on sickle cell anaemia and cancers. These initiatives have not, to date, been involved in training genetic health care professionals, who can deliver dedicated broad high-level genetic counselling and medical genetics services.

Despite this upward trajectory, significant challenges remain. Few efforts train specifically for the healthcare sector or for patient care, but rather for a research setting. In addition, language has been identified as a significant barrier to the uptake of cross-continental training initiatives (58), and the ability of Africa to retain scientific leaders who are capable of developing and maintaining sustainable research programs remains difficult (4).

Clinical and genomics infrastructure shortages further impede the continent's adoption of genomic medicine (3,44,54,59). Robust clinical infrastructure is required for adequate and equitable access to service and to collect, process and store clinical samples destined for diagnostic genomic analysis (60). Additionally, the infrastructure required for analysis and subsequent genomic medicine implementation is expensive and requires considerable expertise and a sophisticated operating environment (54,61). As resources for infrastructure projects remain scarce in Africa (59), developing innovative approaches that leverage existing infrastructure, promote the sharing of facilities at the national and regional level or prioritize subcontracting services to allow efforts to be concentrated on developing services are paramount (5,54). Furthermore, access is but the first step; proper integration of high-throughput technologies on affordable technical platforms and based on an African knowledge-base is needed if Africans are to reap the health benefits (62).

On a more positive note, Africa's genomics capacity is not restricted to human genetics, and this provides an opportunity for leveraging shared infrastructure, and genetics and bioinformatics expertise. There is significant capacity in infectious disease genomics with institutions such as the African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), the Africa Centre for Disease Control and Prevention (Africa CDC), Nigerian CDC, the Kwazulu-Natal Research Innovation

and Sequencing Platform (KRISP) and the National Institute for Communicable Diseases in South Africa. These initiatives spearheaded the response to COVID-19 on the continent (63–65). Furthermore, there are pockets of excellence for agricultural genomics, such as the Agricultural Research Centre (ARC) in South Africa (www.arc.agric.za) and the African Orphan Crops Consortium (http://africanorphancrops.org).

The use of digital tools could provide cost-effective solutions that broaden the reach and efficiency of genomic medicine. Distance-based training initiatives (66) and telegenetics platforms (67) enable increased access to training, testing and counselling resources by reducing traditional cost and time barriers (68). Such approaches would need to be tailored for Africa to assess efficacy and uptake.

While all these initiatives would accelerate the growth of genomics in Africa, dedicated healthcare professionals with broad genetics and genomics training will be required to translate this knowledge into appropriate clinical services. Furthermore, while shared equipment facilities are helpful, human genetics diagnostic laboratories have specific requirements to ensure results are returned to individual patients with high accuracy and appropriate counselling. Interpretation of individual genomic data for feedback to patients needs to be ring-fenced as a distinct requirement.

Dedicated funding and support

Most countries in Africa do not have formal medical specialties in clinical genetics and therefore lack dedicated funding for this medical discipline. South Africa and several North African countries, including Egypt, Morocco and Tunisia, are the exceptions, but still have challenges in providing broad and equitable services throughout their respective countries (33–36). Genetic services are often considered non-essential in resource-constrained health services and are then only offered where dedicated and informed medical practitioners seek out or develop diagnostic tests and offer patient and family counselling. This hampers access to diagnosis, treatment and prevention strategies, as well as opportunities for family planning across the genetic disease spectrum (52).

There is a need to create new funding opportunities from within the continent, including philanthropic, government and industry funding, that allow for African genomics questions, priorities and intended outcomes to be framed locally, nationally and regionally (59). A variety of co-operative arrangements, between the government and private sector, are becoming increasingly popular in health care due to their ability to mobilise resources, finance and reduce service delivery inefficiencies (69). A venture capital start-up company in Nigeria, 54gene, is developing such a partnership (https://54gene.com) and some genetic researchers in Africa have argued that public–private partnerships are necessary for expansion of genomic medicine on the continent (44).

Political will and supporting legislation

The movement toward integrating genomics into routine medical practice is gaining momentum with country-level, government-funded initiatives underway in at least 14 countries internationally (70). Such large-scale national efforts drive the adoption of genomic medicine through financial support (50), increasing genetic literacy, fostering national and regional collaborations (5), and limiting unnecessary duplication of effort and infrastructure (70). However, no such initiatives exist in any African country as yet (70). The potential for national and regional collaboration across Africa is large.

Some African governments are beginning to recognize the value of local genomic data and have supported research sequencing initiatives (71–73). Recently a call was made to sequence three million African genomes (3MAG) to capture the extent of genetic variation across the continent, which would not only improve global health and research but significantly reduce envisioned health disparities (74). Although such large scale public health genomic projects may not be feasible for now, African governments must still plan for genomic medicine and prioritise legislation, frameworks and policies that facilitate adequate and equitable genetic service provision and genomics research (4,50,59), with specific goals to translate knowledge gained into improved clinical services.

Public education and awareness

Genomic literacy and awareness are critical in managing genetic diseases, including diagnosis, prevention of complications and therapy (36,42). In Africa, public knowledge of genetics and genomics is still very poor, and few initiatives aim to improve genomic literacy amongst the public despite the potential health benefits (36,48,49,75) and indirect motivation for government buy-in and genomics medicine investment (44). Africa is characterized by great cultural and language diversity, thereby requiring a multidisciplinary approach to improving public and community genomics literacy and engagement (76). The low literacy rates in some Sub-Saharan African countries add a further complication (35,49).

The catalytic role of patient advocacy groups as a driver for change should not be underestimated (24,37,41). For example, the Albinism Society of South Africa (ASSA), which was started in 1994, is very active in providing information on the condition, as well as support services for people with albinism and advocating for their rights (particularly for adequate health care) (77). Similar groups are organized in Kenya, Tanzania, Cameroon and elsewhere (46).

Genomic and public health economics research relevant to Africa

There is unanimous agreement that understanding population genetic variation as it links to health is the gateway to genomic medicine in Africa (5,9). For PPH strategies to be applicable to Africans, African scientists must direct the research agenda and conduct high-quality genomics research that aims to understand health problems relevant to African populations (4,59).

There are almost no large-scale population-based studies in Africa. African participants in GWAS studies contribute <3% of the world genetic data (78), despite harbouring the greatest genetic diversity. This unequal weighting has not only biased our understanding of genetic disease but has also reduced the potential transferability of translational genomic research. There are also missed opportunities to recognize the value of African data in fine-mapping causal variants given the high diversity and low linkage disequilibrium in African populations (as reviewed in (44)).

Furthermore, there are extremely limited data on mutation profiles for monogenic diseases in Africa, with minimal understanding of the genetic epidemiology. Genomic technologies offer the potential to increase this knowledge rapidly and systematically. Key areas of focus where such interventions offer

possibilities for rapid acceleration include baseline characterization of population variant profiles, monogenic disease mutation profiles, infectious diseases and host responses, pharmacogenomics and precision therapy for complex diseases. Broad open-access databases and data sharing are critical components to ensure accessibility of the data across the continent, for the benefit of individuals on the continent but also for the many individuals distributed across the world who have their origins in Africa.

Concurrent clinical service accelerators are required. Apart from ensuring adequate capacity and training, the establishment of patient registries, biorepositories, electronic health records and mutation databases with broad access are essential for translational genomic medicine. This translation must encompass the key areas of analytical validity, clinical validity, clinically utility and ethical soundness.

Although the economic factors in the establishment of genomics services and their impact on healthcare in low and middle-income countries have not been assessed, it is reasonable to assume that they may have similar value to those demonstrated elsewhere (79–83). On an individual level, families value causal diagnoses to minimize the financial and emotional effects of the diagnostic odyssey. Further value arises from directed management strategies that improve patient health and well-being while expanding evidence-based research to reduce evidentiary uncertainty. The value at a continent-wide level needs to be explored further (84).

Conclusion

The need for globalization in research and the recognition of the critical role of African genetic diversity in unravelling the aetiology of disease states (5,9,85) is a useful starting point to justify the development of clinical genetic services in Africa. Informal discussions with colleagues in different African countries revealed several distinct approaches to genetics service delivery but practicing clinical geneticists could not be identified in most African countries. In some countries, there were no ring-fenced, structured genetic services. Many relied heavily on academic and research environments and funding to support sporadic clinical genetic services. Only a handful had formal training for clinical geneticists and genetic counsellors with dedicated government-supported services.

Service could be expedited by fostering intra- and inter-African institution collaborations to develop and expand the necessary infrastructure and capacity required to promote genomics-based health interventions in Africa. However, there is a need to create new funding opportunities from within the continent, including philanthropic, government and industry funding, for African researchers so that more research questions, priorities and intended outcomes can be framed locally, nationally and regionally.

Genetic services in Africa need to be culturally competent by considering and respecting the culture, values and traditions of the individuals and communities accessing the genetic services. Much research is needed to explore potential barriers and develop innovative approaches to genetic education and language and the mechanisms best suited to prevent and treat genetic disorders in Africa. Intensive community engagement is required to obtain informed opinions and use these to develop appropriate services. Genetic services need to be based on scientific validity obtained from studies in Africans and should have clinical utility in low resourced settings to ensure that they lead to health benefits. Overall, this will only be possible with strengthening of underlying general health systems.

Considering the limited and fledgling genetic services across most of Africa, the continent has perhaps not yet come of 'genomics' age but has a strong nucleus of determined and optimistic clinicians and scientists with a clear vision. There is hope for innovative solutions and technological leapfrogging, but in the absence of political will, and investment in translation through strong health care systems, the people of the continent will need to be patient.

Acknowledgements

We gratefully acknowledge the time and attention of medical geneticists in African countries in the north, west, east and south, who generously shared their knowledge and expertise during discussion sessions and who have informed the conclusions we reached. We thank Jennifer Kromberg for reading and commenting on an early draft and for her additional insights and suggestions.

Conflict of Interest statement. The authors declare that they have no conflicting interests.

Funding

MR is a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Innovation, and administered by the National Research Foundation.

References

1. Ndiaye Diallo, R., Gadji, M., Hennig, B.J., Guèye, M.V., Gaye, A., Diop, J.P.D., Sylla Niang, M., Lopez Sall, P., Guèye, P.M., Dem, A. *et al.* (2017) Strengthening human genetics research in Africa: report of the 9th meeting of the African Society of Human Genetics in Dakar in may 2016. *Glob. Heal. Epidemiol. Genom.*, 2, 1–4.
2. El-Kamah, G.Y., Mohamed, A.M., Gad, Y.Z., Abdelhak, S., Hennig, B.J., Ramesar, R.S., Landouré, G., Gaye, A., Newport, M.J., Williams, S.M. and Ramsay, M. (2020) Developing a road map to spread genomic knowledge in Africa: 10th conference of the African Society of Human Genetics, Cairo. *Egypt. Am. J. Trop. Med. Hyg.*, 102, 719–723.
3. Musanabaganwa, C., Mihigo, B., Tumusime, R., Uwanyirigira, M., Da Rocha, J., Hayat, M., Govender, M., Buto, P., Nyunga, T., Ramesar, R.S. *et al.* (2020) Building skills and resources for genomics, epigenetics, and bioinformatics research for Africa: report of the joint 11th conference of the African Society of Human Genetics and 12th H3 Africa consortium, 2018. *Am. J. Trop. Med. Hyg.*, 102, 1417–1424.
4. Rotimi, C., Abayomi, A., Abimiku, A., Adabayeri, V.M., Adebamowo, C., Adebiyi, E., Ademola, A.D., Adeyemo, A., Adu, D., Affolabi, D. *et al.* (2014) Research capacity. Enabling the genomic revolution in Africa. *Science*, 344, 1346–1348.
5. Tekola-Ayele, F. and Rotimi, C.N. (2015) Translational genomics in low- and middle-income countries: opportunities and challenges. *Public Health Genomics*, 18, 242–247.
6. Mulder, N., Abimiku, A., Adebamowo, S.N., de Vries, J., Matimba, A., Olowoyo, P., Ramsay, M., Skelton, M. and Stein, D.J. (2018) H3Africa: current perspectives. *Pharmgenomics. Pers. Med.*, 11, 59–66.

7. Wonkam, A. and de Vries, J. (2020) Returning incidental findings in African genomics research. *Nat. Genet.*, **52**, 17–20.
8. Mitropoulos, K., Al Jaibeji, H., Forero, D.A., Laissue, P., Wonkam, A., Lopez-Correa, C., Mohamed, Z., Chantratita, W., Lee, M.T.M., Llerena, A. *et al.* (2015) Success stories in genomic medicine from resource-limited countries. *Hum. Genomics*, **9**, 1–7.
9. Pereira, L., Mutesa, L., Tindana, P. and Ramsay, M. (2021) African genetic diversity and adaptation inform a precision medicine agenda. *Nat. Rev. Genet.*, **22**, 284–306.
10. Williams, T.N., Mwangi, T.W., Wambua, S., Peto, T.E.A., Weatherall, D.J., Gupta, S., Recker, M., Penman, B.S., Uyoga, S., MacHaria, A. *et al.* (2005) Negative epistasis between the malaria-protective effects of $\alpha+-$thalassemia and the sickle cell trait. *Nat. Genet.*, **37**, 1253–1257.
11. Goldman, A., Jenkins, T. and Krause, A. (1998) Molecular evidence that fragile X syndrome occurs in the south African black population. *J. Med. Genet.*, **35**, 878–879.
12. Essop, F.B. and Krause, A. (2013) Diagnostic, carrier and prenatal genetic testing for fragile X syndrome and other FMR-1-related disorders in Johannesburg, South Africa: a 20-year review. *S. Afr. Med. J.*, **103**, 994–998.
13. Lubala, T.K., Lumaka, A., Mbuyi-Musanzayi, S., Kayembe, T., Shongo, M.Y.P., Mukuku, O., Lubala, N., Malamba-Lez, D., Luboya, O.N. and Lukusa-Tshilobo, P. (2018) Fragile X syndrome with mosaic size mutation in a bantu patient from Central Africa. *Clin. Dysmorphol.*, **27**, 66–69.
14. Kengne Kamga, K., Nguefack, S., Minka, K., Wonkam Tingang, E., Esterhuizen, A., Nchangwi Munung, S., De Vries, J. and Wonkam, A. (2020) Cascade testing for fragile X syndrome in a rural setting in Cameroon (sub-Saharan Africa). *Genes (Basel)*, **11**, 136.
15. Kabahuma, R.I., Ouyang, X., Du, L.L., Yan, D., Hutchin, T., Ramsay, M., Penn, C. and Liu, X.Z. (2011) Absence of GJB2 gene mutations, the GJB6 deletion (GJB6-D13S1830) and four common mitochondrial mutations in nonsyndromic genetic hearing loss in a south African population. *Int. J. Pediatr. Otorhinolaryngol.*, **75**, 611–617.
16. Wonkam, A., Bosch, J., Noubiap, J.J., Lebeko, K., Makubalo, N. and Dandara, C. (2015) No evidence for clinical utility in investigating the Connexin genes GJB2, GJB6 and GJA1 in non-syndromic hearing loss in black Africans. *S. Afr. Med. J.*, **105**, 23–26.
17. Lebeko, K., Sloan-Heggen, C.M., Noubiap, J.J.N., Dandara, C., Kolbe, D.L., Ephraim, S.S., Booth, K.T., Azaiez, H., Santos-Cortez, R.L.P., Leal, S.M., Smith, R.J.H. and Wonkam, A. (2016) Targeted genomic enrichment and massively parallel sequencing identifies novel nonsyndromic hearing impairment pathogenic variants in Cameroonian families. *Clin. Genet.*, **90**, 288–290.
18. Adadey, S.M., Manyisa, N., Mnika, K., de Kock, C., Nembaware, V., Quaye, O., Amedofu, G.K., Awandare, G.A. and Wonkam, A. (2019) GJB2 and GJB6 mutations in non-syndromic childhood hearing impairment in Ghana. *Front. Genet.*, **10**, 1–8.
19. Arab, S. Ben, Masmoudi, S., Beltaief, N., Hachicha, S. and Ayadi, H. (2004) Consanguinity and endogamy in northern Tunisia and its impact on non-syndromic deafness.
20. Baine, F.K., Krause, A. and Greenberg, L.J. (2016) The frequency of Huntington disease and Huntington disease-like 2 in the south African population. *Neuroepidemiology*, **46**, 198–202.
21. Krause, A., Mitchell, C., Essop, F., Tager, S., Temlett, J., Stevanin, G., Ross, C., Rudnicki, D. and Margolis, R. (2015) Junctophilin 3 (JPH3) expansion mutations causing Huntington disease like 2 (HDL2) are common in south African patients with African ancestry and a Huntington disease phenotype. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **168**, 573–585.
22. Rebbeck, T.R., Friebel, T.M., Friedman, E., Hamann, U., Huo, D., Kwong, A., Olah, E., Olopade, O.I., Solano, A.R., Teo, S.H. *et al.* (2018) Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations. *Hum. Mutat.*, **39**, 593–620.
23. Zheng, Y., Walsh, T., Gulsuner, S., Casadei, S., Lee, M.K., Ogundiran, T.O., Ademola, A., Falusi, A.G., Adebamowo, C.A., Oluwasola, A.O. *et al.* (2018) Inherited breast cancer in Nigerian women. *J. Clin. Oncol.*, **36**, 2820–2825.
24. Krause, A., Seymour, H. and Ramsay, M. (2018) Common and founder mutations for monogenic traits in sub-saharan African populations. *Annu. Rev. Genomics Hum. Genet.*, **19**, 149–175.
25. Romdhane, L., Mezzi, N., Hamdi, Y., El-Kamah, G., Barakat, A. and Abdelhak, S. (2019) Consanguinity and inbreeding in health and disease in north African populations. *Annu. Rev. Genomics Hum. Genet.*, **20**, 155–179.
26. Sankoh, O. and Ijsselmuiden, C. (2011) Sharing research data to improve public health: a perspective from the global south. *Lancet*, **378**, 401–402.
27. Popejoy, A.B. (2019) Diversity in precision medicine and pharmacogenetics: methodological and conceptual considerations for broadening participation. *Pharmgenomics. Pers. Med.*, **12**, 257–271.
28. Kraft, S.A., Cho, M.K., Gillespie, K., Halley, M., Varsava, N., Ormond, K.E., Luft, H.S., Wilfond, B.S. and Soo-Jin Lee, S. (2018) Beyond consent: building trusting relationships with diverse populations in precision medicine research. *Am. J. Bioeth.*, **18**, 3–20.
29. Sabatello, M., Callier, S., Garrison, N.A. and Cohn, E.G. (2018) Trust, precision medicine research, and equitable participation of underserved populations. *Am. J. Bioeth.*, **18**, 34–36.
30. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M. and Daly, M.J. (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.*, **51**, 584–591.
31. Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R. and Domingue, B. (2019) Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.*, **10**, 1–9.
32. Kromberg, J.G., Sizer, E.B. and Christianson, A.L. (2013) Genetic services and testing in South Africa. *J. Community Genet.*, **4**, 413–423.
33. Kromberg, J.G.R. and Krause, A. (2013) Human genetics in Johannesburg, South Africa: past, present and future. *S. Afr. Med. J.*, **103**, 957–961.
34. Temtamy, S.A. and Hussen, D.F. (2017) Genetics and genomic medicine in Egypt: steady pace. *Mol. Genet. Genomic Med.*, **5**, 8–14.
35. Belhassan, K., Ouldim, K. and Sefiani, A.A. (2016) Genetics and genomic medicine in Morocco: the present hope can make the future bright. *Mol. Genet. Genomic Med.*, **4**, 588–598.
36. Elloumi-Zghal, H. and Chaabouni Bouhamed, H. (2018) Genetics and genomic medicine in Tunisia. *Mol. Genet. Genomic Med.*, **6**, 134–159.
37. Chaabouni-Bouhamed, H. (2008) Country report Tunisia: communities and community genetics. *Community Genet.*, **11**, 313–323.
38. Adeyemo, A.A., Amodu, O.K., Ekure, E.E. and Omotade, O.O. (2018) Medical genetics and genomic medicine in Nigeria. *Mol. Genet. Genomic Med.*, **6**, 314–321.

1

39. Uwineza, A. and Mutesa, L. (2015) Medical genetics and genomic medicine in Rwanda. *Mol. Genet. Genomic Med.*, **3**, 486–489.
40. Malherbe, H.L., Christianson, A.L., Woods, D. and Aldous, C. (2017) The case for the genetic nurse in South Africa. *J. Community Genet.*, **8**, 65–73.
41. Kromberg, J. and Jenkins, T. (1997) Cultural Influences on the Perception of Genetic Disorders in the Black Population of Southern Africa. In Clarke, A.P.E. (ed), *Culture, Kinship and Genes*. Macmillan Press, London, pp. 147–157.
42. Wonkam, A., Njamnshi, A.K. and Angwafo, F.F. (2006) Knowledge and attitudes concerning medical genetics amongst physicians and medical students in Cameroon (sub-Saharan Africa). *Genet. Med.*, **8**, 331–338.
43. Zhong, A., Darren, B., Loiseau, B., He, L.Q.B., Chang, T., Hill, J. and Dimaras, H. (2018) Ethical, social, and cultural issues related to clinical genetic testing and counseling in low- and middle-income countries: a systematic review. *Genet. Med.*, **0**, 1–11.
44. Munung, N.S., Mayosi, B.M. and de Vries, J. (2018) Genomics research in Africa and its impact on global health: insights from African researchers. *Glob. Heal. Epidemiol. Genom.*, **3**, 1–8.
45. Abacan, M.A., Alsubaie, L., Barlow-Stewart, K., Caanen, B., Cordier, C., Courtney, E., Davoine, E., Edwards, J., Elackatt, N.J., Gardiner, K. *et al.* (2019) The global state of the genetic Counseling profession. *Eur. J. Hum. Genet.*, **27**, 183–197.
46. Kromberg, J. (2018) Genetic counseling and albinism. In Kromberg, J. and Manga, P. (eds), *Albinism in Africa. Historical, Geographical, Medical, Genetic and Psychosocial Aspects*. Elsevier Academic Press, London, pp. 203–233.
47. Gillham, A., Greyling, B., Wessels, T.M., Mbele, B., Schwyzer, R., Krause, A. and Mahlangu, J. (2015) Uptake of genetic Counseling, knowledge of bleeding risks and psychosocial impact in a south African cohort of female relatives of people with Hemophilia. *J. Genet. Couns.*, **24**, 978–986.
48. Nembaware, V. and Mulder, N. (2019) The African genomic medicine training initiative (AGMT): showcasing a community and framework driven genomic medicine training for nurses in Africa. *Front. Genet.*, **10**, 1–14.
49. Mboowa, G., Sserwadda, I. and Aruhomukama, D. (2021) Genomics and bioinformatics capacity in Africa: no continent is left behind. *Genome*, **64**, 1–11.
50. Manolio, T.A., Abramowicz, M., Al-Mulla, F., Anderson, W., Balling, R., Berger, A.C., Bleyl, S., Chakravarti, A., Chantratita, W., Chisholm, R.L. *et al.* (2015) Global implementation of genomic medicine: we are not alone. *Sci. Transl. Med.*, **7**, 1–7.
51. Kerr, R. and Kromberg, J. (2018) Genetic testing, prenatal and postnatal diagnosis for albinism. In Kromberg, J.G.R. and Manga, P. (eds), *Albinism in Africa. Historical, Geographic, Medical, Genetic and Psychosocial Aspects, London*, London, pp. 235–256.
52. Alwan, A. and Modell, B. (2003) Recommendations for introducing genetics services in developing countries. *Nat. Rev. Genet.*, **4**, 61–68.
53. Bornstein, M.H. and Hendricks, C. (2013) Screening for developmental disabilities in developing countries. *Soc. Sci. Med.*, **97**, 307–315.
54. AESA (2020) A Framework for the Implementation of Genomic Medicine for Public Health in Africa. *A Framework for the Implementation of Genomic Medicine for Public Health in Africa*, Nairobi, Vol. 2020.
55. Green, E.D., Gunter, C., Biesecker, L.G., Di Francesco, V., Easter, C.L., Feingold, E.A., Felsenfeld, A.L., Kaufman, D.J., Ostrander, E.A., Pavan, W.J. *et al.* (2020) Strategic vision for improving human health at the forefront of genomics. *Nature*, **586**, 683–692.
56. Zeggini, E., Gloyn, A.L., Barton, A.C. and Wain, L.V. (2019) Translational genomics and precision medicine: moving from the lab to the clinic. *Science*, **365**, 1409–1413.
57. Oleribe, O.E., Momoh, J., Uzochukwu, B.S., Mbofana, F., Adebiyi, A., Barbera, T., Williams, R. and Taylor Robinson, S.D. (2019) Identifying key challenges facing healthcare systems in Africa and potential solutions. *Int. J. Gen. Med.*, **12**, 395–403.
58. Adebamowo, S.N., Francis, V., Tambo, E., Diallo, S.H., Landouré, G., Nembaware, V., Dareng, E., Muhamed, B., Odutola, M., Akeredolu, T. *et al.* (2018) Implementation of genomics research in Africa: challenges and recommendations. *Glob. Health Action*, **11**, 1–6.
59. Wonkam, A. and Mayosi, B.M. (2014) Genomic medicine in Africa: promise, problems and prospects. *Genome Med.*, **6**, 1–3.
60. Christianson, A., Zimmern, R., Kristoffersson, U., Schmidtke, J., Kent, A., Raouf, R., Barreiro, C. and Nippert, I. (2013) Health needs assessment for medical genetic services for congenital disorders in middle- and low-income nations. *J. Community Genet.*, **4**, 297–308.
61. Mulder, N.J., Adebiyi, E., Adebiyi, M., Adeyemi, S., Ahmed, A., Ahmed, R., Akanle, B., Alibi, M., Armstrong, D.L., Aron, S. *et al.* (2017) Development of bioinformatics infrastructure for genomics research. *Glob. Heart*, **12**, 91–98.
62. Knoppers, B.M., Sénécal, K., Borry, P. and Avard, D. (2014) Whole-genome sequencing in newborn screening programs. *Sci. Transl. Med.*, **6**, 229cm2.
63. Tessema, S.K., Inzaule, S.C., Christoffels, A., Kebede, Y., de Oliveira, T., Ouma, A.E.O., Happi, C.T. and Nkengasong, J.N. (2020) Accelerating genomics-based surveillance for COVID-19 response in Africa. *Lancet Microbe*, **1**, e227–e228.
64. Wibmer, C.K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Oosthuysen, B., Lambson, B.E., de Oliveira, T., Vermeulen, M., van der Berg, K. *et al.* (2021) SARS-CoV-2 501Y.V2 escapes neutralization by south African COVID-19 donor plasma. *Nat. Med.*, **27**, 622–625.
65. Tegally, H., Wilkinson, E., Lessells, R.J., Giandhari, J., Pillay, S., Msomi, N., Mlisana, K., Bhiman, J.N., von Gottberg, A., Walaza, S. *et al.* (2021) Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat. Med.*, **27**, 440–446.
66. Gurwitz, K.T., Aron, S., Panji, S., Maslamoney, S., Fernandes, P.L., Judge, D.P., Ghouila, A., Domelevo Entfellner, J.-B., Guerfali, F.Z., Saunders, C. *et al.* (2017) Designing a course model for distance-based online bioinformatics training in Africa: the H3ABioNet experience. *PLoS Comput. Biol.*, **13**, 1–11.
67. Bombard, Y. and Hayeems, R.Z. (2020) How digital tools can advance quality and equity in genomic medicine. *Nat. Rev. Genet.*, **21**, 505–506.
68. Cohen, S.A., Huziak, R.C., Gustafson, S. and Grubs, R.E. (2016) Analysis of advantages, limitations, and barriers of genetic Counseling service delivery models. *J. Genet. Couns.*, **25**, 1010–1018.
69. Meier, F., Schöffski, O. and Schmidtke, J. (2013) Public-private partnership as a solution for integrating genetic services into health care of countries with low and middle incomes. *J. Community Genet.*, **4**, 309–320.
70. Stark, Z., Dolman, L., Manolio, T.A., Ozenberger, B., Hill, S.L., Caulfied, M.J., Levy, Y., Glazer, D., Wilson, J., Lawler, M. *et al.* (2019) Integrating genomics into healthcare: a global responsibility. *Am. J. Hum. Genet.*, **104**, 13–20.
71. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A. *et al.* (2015) The African genome

1

variation project shapes medical genetics in Africa. *Nature*, 517, 327–332.
72. Choudhury, A., Ramsay, M., Hazelhurst, S., Aron, S., Bardien, S., Botha, G., Chimusa, E.R., Christoffels, A., Gamieldien, J., Sefid-Dashti, M.J. *et al.* (2017) Whole-genome sequencing for an enhanced understanding of genetic variation among south Africans. *Nat. Commun.*, 8, 1–12.
73. Wohlers, I., Künstner, A., Munz, M., Olbrich, M., Fähnrich, A., Calonga-Solís, V., Ma, C., Hirose, M., El-Mosallamy, S., Salama, M., Busch, H. and Ibrahim, S. (2020) An integrated personal and population-based Egyptian genome reference. *Nat. Commun.*, 11, 1–10.
74. Wonkam, A. (2021) Sequence three million genomes across Africa. *Nature*, 590, 209–211.
75. Muzoriana, N., Gavi, S., Nembaware, V., Dhoro, M. and Matimba, A. (2017) Knowledge, attitude, and perceptions of pharmacists and pharmacy students towards pharmacogenomics in Zimbabwe. *Pharmacy*, 5, 36–46.
76. Solomon, G., Greenberg, J., Futter, M., Vivian, L. and Penn, C. (2012) Understanding of genetic inheritance among Xhosa-speaking caretakers of children with hemophilia. *J. Genet. Couns.*, 21, 726–740.
77. Mazibuko, N.G. and Kromberg, J. (2018) A personal perspective: living with albinism. In Kromberg, J. and Manga, P. (eds), *Albinism in Africa. Historical, Geographic, Medical Genetic and Psychosocial Aspects*. Elsevier Academic Press, London, pp. 295–307.
78. Popejoy, A.B. and Fullerton, S.M. (2016) Genomics is failing on diversity. *Nature*, 538, 161–164.
79. Dragojlovic, N., Elliott, A.M., Adam, S., van Karnebeek, C., Lehman, A., Mwenifumbo, J.C., Nelson, T.N., du Souich, C., Friedman, J.M. and Lynd, L.D. (2018) The cost and diagnostic yield of exome sequencing for children with suspected genetic disorders: a benchmarking study. *Genet. Med.*, 20, 1013–1021.
80. Soden, S.E., Saunders, C.J., Willig, L.K., Farrow, E.G., Smith, L.D., Petrikin, J.E., LePichon, J.B., Miller, N.A., Thiffault, I., Dinwiddie, D.L. *et al.* (2014) Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci. Transl. Med.*, 6, 1–14.
81. Tsiplova, K., Zur, R.M., Marshall, C.R., Stavropoulos, D.J., Pereira, S.L., Merico, D., Young, E.J., Sung, W.W.L., Scherer, S.W. and Ungar, W.J. (2017) A microcosting and cost-consequence analysis of clinical genomic testing strategies in autism spectrum disorder. *Genet. Med.*, 19, 1268–1275.
82. Stark, Z., Schofield, D., Alam, K., Wilson, W., Mupfeki, N., Macciocca, I., Shrestha, R., White, S.M. and Gaff, C. (2017) Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement. *Genet. Med.*, 19, 867–874.
83. Chrystoja, C.C. and Diamandis, E.P. (2014) Whole genome sequencing as a diagnostic test: challenges and opportunities. *Clin. Chem.*, 60, 724–733.
84. Pollard, S., Weymann, D., Dunne, J., Mayanloo, F., Buckell, J., Buchanan, J., Wordsworth, S., Friedman, J.M., Stockler-Ipsiroglu, S., Dragojlovic, N. *et al.* (2021) Toward the diagnosis of rare childhood genetic diseases: what do parents value most? *Eur. J. Hum. Genet.*
85. Choudhury, A., Aron, S., Botigué, L.R., Sengupta, D., Botha, G., Bensellak, T., Wells, G., Kumuthini, J., Shriner, D., Fakim, Y.J. *et al.* (2020) High-depth African genomes inform human migration and health. *Nature*, 586, 741–748.

1

1

APPENDIX H: ORIGINAL PUBLICATION 3

Journal of *Personalized Medicine*

MDPI

Article

Clinicians' Perceptions towards Precision Medicine Tools for Cardiovascular Disease Risk Stratification in South Africa

Michelle Kamp [1,2], Oliver Pain [3], Andrew May [2], Cathryn M. Lewis [4,5] and Michèle Ramsay [1,2,*]

[1] Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, The University of the Witwatersrand, Johannesburg 2001, South Africa
[2] Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa
[3] Maurice Wohl Clinical Neuroscience Institute, Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 9RT, UK
[4] Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK
[5] Department of Medical and Molecular Genetics, Faculty of Life Sciences and Medicine, King's College London, London SE1 9RT, UK
* Correspondence: michele.ramsay@wits.ac.za

Abstract: Cardiovascular diseases (CVDs) are a leading cause of mortality and morbidity in South Africa. Risk stratification is the preferred approach to disease prevention, but identifying patients at high risk for CVD remains challenging. Assessing genetic risk could improve stratification and inform a clinically relevant precision medicine (PM) approach. Clinicians are critical to PM adoption, thus, this study explores practicing clinicians' perceptions of PM-based CVD risk stratification in South Africa's public health setting. Practicing clinicians (n = 109) at four teaching hospitals in Johannesburg, South Africa, completed an electronic self-administered survey. The effect of demographic and professional characteristics on PM-based CVD risk stratification perceptions was assessed. Fewer than 25% of respondents used clinical genetic testing, and 14% had formal genetics training. 78% had a low mean knowledge score, with higher scores associated with genetic training ($p < 0.0005$) and research involvement ($p < 0.05$). Despite limited knowledge and resources, 84% perceived PM approaches positively. 57% felt confident in applying the PM-based approach, with those already undertaking CVD risk stratification more confident ($p < 0.001$). High cost and limited access to genetics services are key barriers. Integrating genetic information into established clinical tools will likely increase confidence in using PM approaches. Addressing the genetics training gap and investment into the country's genomics capacity is needed to advance PM in South Africa.

Keywords: precision medicine; polygenic risk scores; clinician attitudes; clinical utility; clinical implementation; cardiovascular disease

check for updates

Citation: Kamp, M.; Pain, O.; May, A.; Lewis, C.M.; Ramsay, M. Clinicians' Perceptions towards Precision Medicine Tools for Cardiovascular Disease Risk Stratification in South Africa. *J. Pers. Med.* **2022**, *12*, 1360. https://doi.org/10.3390/jpm12091360

Academic Editor: William Duddy

Received: 7 June 2022
Accepted: 17 August 2022
Published: 24 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Cardiovascular diseases (CVDs) are the leading cause of mortality globally [1]. South Africans are at risk as they have a high and increasing prevalence of obesity, hypertension and undiagnosed diabetes [2]. In 2019, CVDs were the leading cause of mortality (15.85% of deaths) and the third leading cause of morbidity (7.08% of DALYs) in South Africa [3]. This growing epidemic is exacerbated by the lack of preventative screening, early diagnosis, and inadequate access to healthcare services. The burden associated with CVDs emphasizes the need for intensified efforts to identify high-risk patients and improve prevention, diagnosis and management.

Globally, and within South Africa, a risk-based prevention strategy is the most widely accepted approach to disease prevention, with high-risk individuals offered preventative medication and encouraged to adopt healthier lifestyles [4,5]. Clinical practice guidelines

advocate risk calculators to estimate the 10-year risk of disease in order to identify and target high-risk individuals [5–8]. Risk stratification is calculated from multiple demographic and clinical factors, such as age, sex, smoking status and blood lipid levels. Despite widespread use, it is estimated that these calculators fail to identify up to 40% of high-risk patients [9,10], an underestimation likely greater in Africa, where scores are yet to be validated in local populations [11]. Moreover, although genetics is a known risk factor for CVD [12,13], current clinical tools do not assess genetic risk directly.

The summation of genetic variants into a polygenic risk score (PRS), a single value estimate of an individual's total genetic risk burden for disease, is able to stratify individuals according to their genetic risk [13,14]. Integrating PRSs with traditional risk factors can improve the prediction of CVD and associated traits [10,15–17]. Such integrated tools, or precision medicine (PM) approaches, where individual genetic variation, lifestyle, and health history are leveraged to tailor prevention and treatment strategies, have the potential to improve patient care and reduce the burden of CVDs on healthcare systems [18–20].

Despite the anticipated benefits of PM, significant barriers to integrating genomics into clinical practice remain [18–22]. In addition to the scientific, educational, ethical, legal, and social barriers, the type of testing and physician-based characteristics may also play a role [23]. Given that clinicians are the interface between healthcare systems and patients, understanding their perceptions and willingness to utilize PM-based approaches is critical to ensuring patients reap the benefits of PM in a timely and cost-effective manner, especially where healthcare systems are fragile and under-resourced [23,24].

Little is known about healthcare providers' perceptions of PM tools in Africa [25,26]. Most research has explored medical professionals' views in higher and middle-income countries with robust healthcare settings [23,27–35]. This study aimed to address this gap by exploring medical doctors' perceptions of a PM-based tool for CVD risk prediction in South Africa. We report our assessment of the findings and discuss clinicians' needs and the broader barriers that need to be addressed to introduce PM-based approaches in the South African public health setting.

2. Materials and Methods

2.1. Study Design and Participants

Study participants were drawn from clinicians primarily affiliated with four public academic hospitals in Johannesburg, South Africa (Charlotte Maxeke Johannesburg Academic Hospital, Chris Hani Baragwanath Academic Hospital, Helen Joseph Academic Hospital, and Rahima Moosa Mother and Child Hospital). Four hospitals were selected to ensure an adequately sized and representative sample. Clinicians at all levels (intern, medical officer, registrar, and specialist) across all medical specializations were eligible for inclusion. Clinicians were excluded if they had not completed their medical training and/or were not affiliated with the selected hospitals.

The study was approved by the University of the Witwatersrand Human Research Ethics Committee (Medical), Johannesburg, South Africa (Approval #M210355).

2.2. Study Questionnaire

To assess clinician experiences and attitudes towards PM-based tools for CVD risk stratification, a 107-item electronic survey was developed. The survey was based on similar validated surveys [33–35], with adaptations to the local context, and was developed in consultation with a multidisciplinary group of academics and clinicians. Four specialists, two clinicians, and two geneticists assessed the survey face validity and ensured question clarity. The survey instrument was tested amongst five clinicians for content, design, and readability, and revised accordingly based on feedback.

The survey was organized into eight sections (Table 1 and Table S1) and administered in identical order (i.e., no section/item randomization). Items consisted of close-ended questions formatted using a five-point Likert-type rating scale. Items were scored in ascending order for all participants, with higher scores on rating scales implying increased

self-perceived knowledge, perception, and confidence of PM-based stratification tools. A higher rating indicated a greater degree of benefit or concern associated with the specific item in the case of benefits and concerns, and an increased level of impact indicated that the barrier would be a greater hindrance to tool implementation. *'I prefer not to answer'* was available for all questions, and free text options were available for respondents wishing to provide additional clarification.

Table 1. Survey organization and associated scales used.

Section	Theme	No. of Items	Scale
1	Demographic and professional information	10	Not applicable
2	Exposure: CVD risk stratification, and genetics training and testing	21	Not applicable
3	Knowledge: Genetics and PM, and educational video	15	Understanding: *None, Little, Some, Moderate, Expert* Agreement: *True, False, I do not know*
4	Perception toward PM-based CVD risk stratification	5	Agreement: *Strongly disagree, Disagree, Neutral, Agree, Strongly agree*
5	Benefits and concerns of PM-based CVD risk stratification	19	Benefit/concern: *Not, A little, Somewhat, Moderately, Very*
6	Confidence in applying PM-based CVD risk stratification	8	Confidence: *None, A little, Somewhat, Moderately, Very*
7	Expectations in applying PM-based CVD risk stratification	9	Role: *None, Supporting, Primary, Not sure* Adoption likelihood: *Very unlikely, Unlikely, No difference, Likely, Very likely*
8	Barriers to implementing PM-based CVD risk stratification	15	Impact: *None, Little, Some, Moderate, Strong*

Clinicians working in South Africa have diverse educational backgrounds, with education and training from multiple institutions within South Africa and abroad. To ensure that all respondents had a similar baseline understanding of PM and related concepts, a compulsory 15-min educational video was included in the questionnaire (available on request). To determine whether the video had successfully relayed the pertinent topics, Section 3 included nine true or false questions. A score of 78% (7 out of 9) and above was interpreted as successfully understanding the topics.

2.3. Data Collection and Analyses

Invitation to participate was extended via email to clinicians over a six-month period, from 10 September 2021 to 31 March 2022. Consenting respondents completed the survey, which was hosted online (REDCap). Due to low initial uptake, researchers requested permission to attend departmental clinical academic meetings, where the study aim and survey could be better outlined prior to inviting participation. Of the 15 departments approached, nine accepted (Internal Medicine, Cardiology, Clinical Genetics, Clinical Pathology, Emergency Medicine, Pediatrics and Child Health, and Surgery). Study data were collected and managed using REDCap (Research Electronic Data Capture) (version 12.5.4, Vanderbilt University, Nashville, TN, USA) tools hosted at the University of the Witwatersrand [36,37].

Self-perceived knowledge, perception and confidence (Sections 3, 4 and 6; Table 1) were calculated as continuous scores by converting responses to a numerical scale and summing the participants' responses, e.g., Knowledge: level of understanding scale converted to numeric scale (No understanding = 0, Little understanding = 1, Some understanding = 2, moderate understanding = 3, Expert understanding = 4) and the total score calculated by

summing scores for each question (maximum score = 24). Mean knowledge score (MKS) was derived by dividing the participants' total score by 6 (number of questions). The mean perception (MPS) and confidence (MCS) scores were calculated similarly, with a maximum obtainable score of 4. However, MPS was calculated over five items. Regarding the nine items for assessing comprehension following the educational video, one point was given if the correct answer was chosen, and a score of zero was given if the wrong answer or *'I do not know'* was chosen.

For descriptive comparisons, participants were dichotomized into high/low categories for the MKS and MCS scores using a breakpoint mean score of 2 (with low being ≤ 2, high > 2). Similarly, respondents were dichotomized into negative/positive categories for MPS scores at a breakpoint mean score of 2. As categorizing a continuous variable causes a loss of appreciable information, inferential analysis was restricted to mean scores. Similarly, age and number of years of clinical experience were dichotomized by the median.

Differences in MKS, MPS, and MCS distributions between groups were assessed with parametric (*t*-test and ANOVA) or non-parametric tests (Mann–Whitney or Kruskal–Wallis), where appropriate, followed by multivariate regression analysis (see Table 2 for study variables). Practice level and medical specialty were removed from the model due to collinearity with other predictors (Figure S1). All analyses were done using RStudio [38] (version 1.1.456), the psych (version 1.8.4; [39]) and ggplot2 (version 3.0.0; [40]) packages. Study responses were presented as percentages (95% confidence intervals; CI) and means (standard deviation; SD). Statistical significance was accepted at a p-value < 0.05.

Table 2. Categorical and continuous predictor variables used for analysis in this study.

Variable	Description	Type (Range)	Categories
Outcome variables			
Mean knowledge score (MKS)	Average self-reported knowledge relating to genetics and PM	Continuous (0–4)	Not applicable
Mean perception score (MPS)	Average self-reported perception relating to value of PM-based CVD risk stratification tool	Continuous (0–4)	Not applicable
Mean confidence score (MCS)	Average self-reported confidence relating to implementation of PM-based CVD risk stratification tool	Continuous (0–4)	Not applicable
Predictor variable			
Sex	Reported sex	Categorical	Female Male
Medical group	Practitioner type based on self-reported practice level and years of experience	Categorical	Consultant Trainee
Clinical experience	Number of years conducting clinical duties	Categorical	<8 years ≥8 years
Postgraduate qualifications	Have a postgraduate qualification in addition to a medical degree	Categorical	Yes No
Involvement in research	Are involved in medical research activities	Categorical	Yes No
Genetics or PM training	Have training specifically relating to genetics and/or PM	Categorical	Yes No
CVD stratification in clinical practice	Conducts CVD screening in their clinical practice	Categorical	Yes Sometimes No

CVD—Cardiovascular disease, PM—Precision Medicine.

3. Results

3.1. Participant Characteristics and Exposure to Genetics and Cardiovascular Disease Screening

A total of 114 participants completed the survey. Five respondents were excluded as the respondents were non-patient-facing, resulting in a final sample of 109 participants.

An accurate response rate could not be calculated, as the exact number of clinicians who received the questionnaire is unknown. Power calculations using Raosoft® showed a sample size of at least one hundred was required for an 8% margin of error and 95% confidence interval, assuming a response distribution of 50%. Table 3 presents respondent characteristics and exposure to study-related clinical variables. One-quarter of respondents were interns, designated 'trainees' in subsequent analyses, with a maximum of two years clinical experience and no medical specialization. Over three-quarters were consultant clinicians. 25% of these practiced within internal medicine disciplines, such as Cardiology and Endocrinology, and 70% within non-internal medicine disciplines, including Pediatrics and Clinical Genetics. The mean (SD) age of the study population was 37.1 years (12.5), with a mean of 11.8 years (12.1) of clinical experience. Age and clinical experience were highly correlated (r = 0.97, SE = 0.03). Consequently, inferential analysis was restricted to clinical experience as to avoid multicollinearity.

Table 3. Respondents' demographics, professional characteristics and study-related clinical exposures (n = 109).

Variable	Detail	N (%)
Hospital affiliation	Chris Hani Baragwanath Academic Hospital	23 (21.1)
	Charlotte Maxeke Johannesburg Academic Hospital	36 (33.0)
	Helen Joseph Hospital	13 (11.9)
	Rahima Moosa Mother and Child Hospital	32 (29.4)
	I prefer not to answer	5 (4.6)
Medical specialty	Internal medicine	21 (19.3)
	Non-internal medicine	57 (52.3)
	Not yet specialized (trainee)	27 (24.8)
	I prefer not to answer	4 (3.7)
Sex	Female	67 (61.5)
	Male	42 (38.5)
Age	<33 years	57 (51.4)
	≥33 years	51 (46.8)
	I prefer not to answer	1 (0.9)
Medical group	Consultant	82 (75.2)
	Trainee	27 (24.8)
Clinical experience	<8 years	55 (50.5)
	≥8 years	50 (45.9)
Postgraduate qualifications	Yes	66 (60.6)
	No	41 (37.6)
	I prefer not to answer	2 (1.8)
Involvement in medical research	Yes	59 (54.1)
	No	50 (45.9)
CVD stratification in clinical practice	Yes	41 (37.6)
	Sometimes	31 (28.4)
	No	36 (33.0)
	I prefer not to answer	1 (0.9)
CVD stratification using a risk score (n = 72)	Yes	25 (34.7)
	Sometimes	19 (26.4)
	No	28 (38.9)
Genetics training	Yes	17 (15.6)
	No	91 (83.5)
	I prefer not to answer	1 (0.9)
Genetics testing in clinical practice	Yes	31 (28.4)
	No	78 (71.6)

Table 3. *Cont.*

Variable	Detail	N (%)
Conditions genetics testing discussed * ($n = 31$)	Monogenic disorders	30 (96.8)
	Cancers	16 (51.6)
	CVD	6 (19.4)
	Nutrigenomics	1 (3.2)
	Pharmacogenomics	2 (6.5)
Genetic testing frequency ($n = 31$)	1 patient every 2 to 3 months	13 (41.9)
	1 patient per month	2 (6.5)
	2 to 5 patients per month	6 (19.4)
	6 to 10 patients per month	2 (6.5)
	11 to 20 patients per month	3 (9.7)
	20+ patients per month	3 (9.7)
	I prefer not to answer	2 (6.5)

* Each respondent could select multiple options.

Two-thirds of respondents undertake CVD risk stratification in their clinical practice. Sixty percent of those use a stratification tool to determine risk, with the Framingham Coronary Heart Disease Risk Score being the most used (80%). 58% undertake CVD screening in response to risk factors, including family history and abnormal laboratory results. Clinicians are confident in the tool, with the majority sharing results with patients (86%) and using them to guide treatment (96%). Most of those who do not undertake screening (68%) indicate that screening is not standard practice in their patient population, e.g., pediatrics and clinical genetics.

Overall exposure to genetics and PM was low. Fifteen percent of respondents indicated they had formal genetics training, and approximately one quarter had performed genetic testing in their clinical practice, although testing is infrequent (Table 3)

3.2. Genetics and Precision Medicine Knowledge

Participants rated their level of understanding of six genetics and PM-related topics. The mean knowledge scores per individual across all six items were low (mean (SD) = 1.57 (0.64); max score = 4) (Figure 1) resulting in 80% ($n = 108$; 95% CI 70.5–86.5) of respondents categorized as having 'low knowledge'. A single individual indicated they had 'extensive knowledge' across all six themes, and no respondents answered all six items as 'expert'.

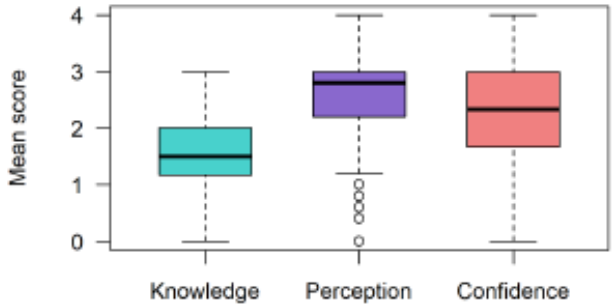


Figure 1. Respondents' overall mean knowledge, perception and confidence scores. Mean scores calculated per individual across items relating to knowledge, perception and confidence.

Variability in understanding shifted towards little or no knowledge for newer genetic concepts, with approximately two-thirds of individuals indicating little or no understanding of Genome-Wide Association Study (GWAS) and PRS (Figure 2A). Regarding the assessment of knowledge following the introductory video, 97% of respondents obtained a score of 78% or higher, suggesting the educational lecture successfully relayed the study's pertinent points and assisted in providing context to survey questions.

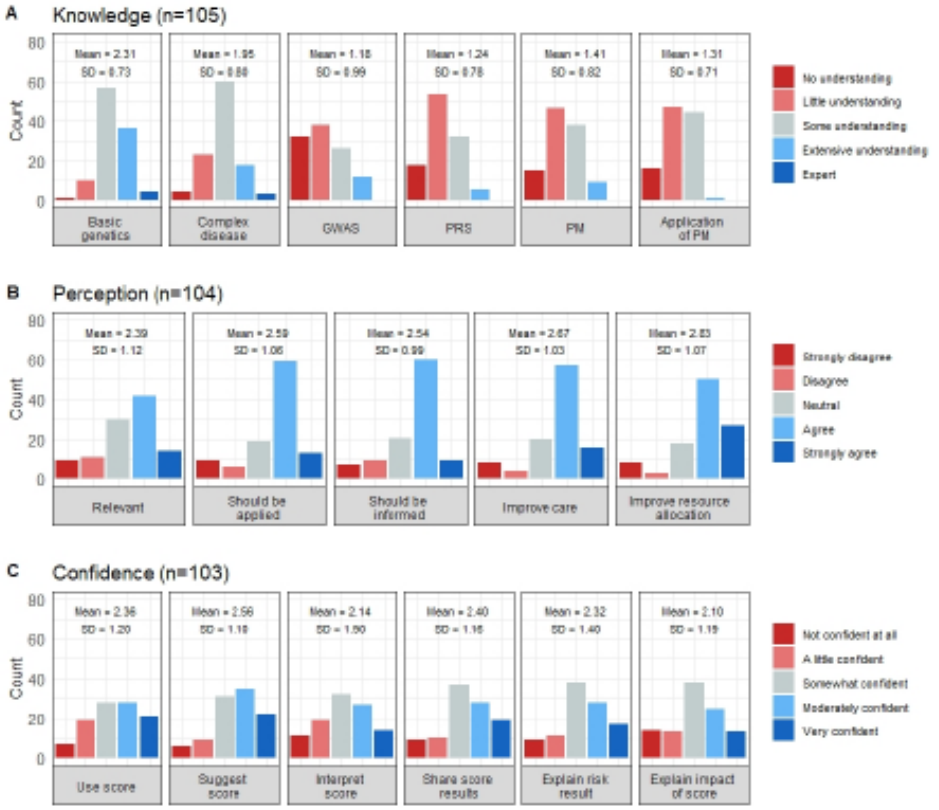


Figure 2. Respondents' knowledge, perception and confidence of items relating to PM-based CVD risk stratification. The distribution of the clinicians' responses regarding (**A**) Genetics and PM knowledge, (**B**) Perceptions towards PM-based CVD risk stratification and (**C**) Confidence in applying PM-based CVD risk stratification in clinical practice. GWAS—Genome-wide association studies, PRS—Polygenic risk scores, PM—Precision Medicine.

3.3. *Perceptions toward Precision Medicine-Based CVD Risk Stratification*

Overall, clinicians had a positive perception of a PM-based tool for CVD risk stratification (mean (SD) = 2.60 (0.90); max score = 4). Three-quarters of respondents had a positive mean perception score when categorized as positive or negative (n = 104; 76%, 95% CI 67.0–83.7) (Figure 1). Despite just over half of respondents believing such a tool would be relevant in their practice, at least two-thirds believed it should be applied in clinical practice (69%) and that it would improve care and prevention of CVD (70%) and national resource allocation (73%) (Figure 2B).

3.4. *Confidence in Applying a Precision Medicine-Based CVD Risk Stratification Tool in Their Practice Settings*

Confidence (self-perceived) scores showed considerable variation (mean (SD) = 2.31 (1.04); max score = 4). Approximately half of respondents (55%) considered themselves confident across all confidence items (Figure 1). Otherwise, confidence depended on the action required within the clinical pathway. Clinicians were most confident in suggesting and using the risk score, whereas 29% had little to no confidence in interpreting the risk score and explaining any potential adverse insurance policy impacts that may arise from risk information (27% little to no confidence) (Figure 2C). Forty percent of respondents indicated they would be moderately or very comfortable with adapting prevention and treatment strategies based on score results. However, a greater proportion of respondents were comfortable with adapting their treatment strategies (22%) than their prevention approaches (13%).

3.5. Factors Influencing Knowledge, Perceptions, and Confidence

We assessed how respondents' characteristics influenced their knowledge, perceptions, and self-confidence toward PM-based CVD risk stratification (Table 4). Medical specialty, sex, and clinical experience did not significantly influence knowledge levels. Medical research involvement had a small effect on knowledge, with those involved in research having increased self-perceptions of knowledge [1.71 (0.61) vs. 1.38 (0.61); $p < 0.01$; effect size r = 0.26]. Previous genetics training had a large effect on knowledge, with those having had training having significantly higher mean knowledge scores [2.26 (0.62) vs. 1.42 (0.54); $p < 5 \times 10^{-5}$; effect size R = 0.76]. Similarly, medical research involvement influenced overall self-confidence in applying PM-based risk stratification [2.57 (0.96) vs. 2.01 (1.06); $p < 0.05$; effect size R = 0.26]. Self-confidence scores were also influenced, with a moderate effect, by whether respondents currently undertook CVD risk stratification in their clinical practice or not [2.65 (1.10) vs. 1.93 (0.96); $p < 0.05$; effect size $\eta^2 = 0.07$]. Conducting CVD screening on a regular or infrequent basis resulted in a significantly higher mean perception score, and with a large effect [Yes: 2.89 (0.63), Sometimes: 2.65 (0.88) vs. No: 2.11 (1.03); $p < 0.0005$; effect size $\eta^2 = 0.13$].

Table 4. Predictors' influence on respondents' mean (SD) knowledge, confidence, and perception scores. *p* value is assessing differences in mean scores across variable subgroups.

Variable	Mean (SD) Knowledge Score (n = 107)	*p* Value	Mean (SD) Confidence Score (n = 104)	*p* Value	Mean (SD) Perception Score (n = 106)	*p* Value
Medical specialty						
Internal medicine	1.67 (0.67)		2.70 (0.94)		2.95 (0.64)	
Non-internal medicine	1.47 (0.62)	*0.310*	2.12 (1.09)	*0.098*	2.40 (0.98)	*0.052*
Not yet specialized	1.66 (0.62)		2.30 (0.92)		2.58 (0.85)	
Sex						
Female	1.59 (0.67)	*0.682*	2.41 (0.98)	*0.404*	2.45 (1.11)	*0.632*
Male	1.54 (0.61)		2.26 (1.08)		2.64 (0.74)	
Medical group						
Clinician	1.52 (0.64)	*0.318*	2.26 (1.08)	*0.933*	2.54 (0.93)	0.969
Trainee	1.66 (0.62)		2.30 (0.92)		2.58 (0.85)	
Practice level						
Specialist	1.61 (0.66)		2.25 (1.27)		2.48 (0.90)	
Registrar	1.41 (0.65)	*0.470*	2.32 (0.81)	*0.897*	2.71 (1.05)	*0.526*
Medical officer	1.48 (0.53)		2.56 (0.86)		2.60 (0.82)	
Intern	1.66 (0.62)		2.30 (0.92)		2.58 (0.85)	
Clinical experience						
<8 years	1.59 (0.60)	*0.580*	2.32 (0.84)	*0.651*	2.59 (0.86)	*0.945*
≥8 years	1.53 (0.67)		2.31 (1.23)		2.55 (0.95)	
Postgrad. qualifications						
Yes	1.62 (0.68)	*0.195*	2.32 (1.17)	*0.490*	2.62 (0.890)	*0.317*
No	1.45 (0.53)		2.27 (0.78)		2.48 (0.93)	
Medical research involvement						
Yes	1.71 (0.61)	*0.007* **	2.57 (0.96)	*0.007* **	2.69 (0.87)	*0.249*
No	1.38 (0.61)		2.01 (1.06)		2.43 (0.92)	
Genetics training						
Yes	2.26 (0.62)	3.518×10^{-5} ***	2.46 (0.86)	*0.720*	2.96 (0.61)	*0.146*
No	1.42 (0.54)		2.29 (1.17)		2.50 (0.93)	

Table 4. *Cont.*

Variable	Mean (SD) Knowledge Score (n = 107)	p Value	Mean (SD) Confidence Score (n = 104)	p Value	Mean (SD) Perception Score (n = 106)	p Value
CVD screening						
Yes	1.75 (0.59)		2.65 (1.10)		2.89 (0.63)	
No	1.46 (0.63)	*0.074*	1.93 (0.96)	*0.010* *	2.11 (1.03)	4.162×10^{-4} ***
Sometimes	1.45 (0.65)		2.33 (0.94)		2.65 (0.88)	

* Postgrad.—Postgraduate, CVD—cardiovascular diseases; p-value thresholds: *** = $p \leq 0.001$, ** = $p \leq 0.01$, * = $p \leq 0.05$.

3.6. Multivariable Analysis of the Factors Affecting Knowledge, Perceptions, and Confidence

When adjusting for study predictors, medical research involvement and genetics training remained positively associated with mean knowledge score, with the model explaining 30% of the overall variance. Similarly, for confidence and perception, medical research involvement (confidence only) and CVD screening status remained associated. However, the influence of CVD screening increased, and the model explained 14% and 13%, respectively. The model revealed the medical group to be associated with the mean knowledge and mean confidence score when adjusting for other predictors (Table S2).

3.7. Benefits and Concerns of a PM-Based CVD Risk Stratification

Figure 3 presents the distribution of the clinicians' responses regarding perceived benefits (A) and concerns (B) of a PM-based CVD risk stratification in the South African public setting. Respondents believed the main benefits associated with PM-based CVD risk stratification was the potential availability of a CVD risk score tailored to African populations (very beneficial and moderately beneficial n = 91; 86%; 95% CI: 77.4–91.6), followed by the adaption of prevention and treatment strategies (very beneficial and moderately beneficial (n = 89; 84%: 95% CI: 75.3–90.1). The most concerning issue for clinicians was the potential adverse impact that score results may have on a patient's life and insurance policies (very concerned and moderately concerned n = 77; 73%; 95% CI: 63.0–80.6), and the lack of transferability of current PM risk that scores have across populations (very concerned and moderately concerned n = 56; 53%; 95% CI: 42.9–62.5).

3.8. Clinician's Expectations: Time Horizon, Funding, and Expected Role

Assessing clinicians' expectations of PM-based tools for CVD risk stratification in South Africa's public revealed that clinicians anticipate a clinically relevant tool for African populations to be available in the longer term. Over one third (36%) expect the tool within the next six to ten years, and another third in more than ten years (39%). Half of the respondents feel that funding would be the responsibility of multiple players, but at least a third (35%) believe that national and provincial governments should be the primary funder. Despite the expected wait and funding complexities, 85% of respondents indicate that they are likely or very likely to alter their CVD screening approach if a tool becomes available (Table S3).

When asking clinicians what they believe their expected role will be in the implementation of PM-based CVD risk stratification, approximately two-thirds identify themselves as having the primary role in traditional patient-facing activities (referring the patient for testing (73%), sharing results with the patient (62%), and using results to guide prevention and treatment approaches (70%). In contrast, at least half of the respondents recognize that interpreting (55%) and explaining results to patients (54%) will require additional expertise and thus a multidisciplinary team (Table S3).

1

4. Discussion

As little is known about healthcare providers' perceptions of PM tools in Africa, this study aimed to explore medical doctors' perceptions of a PM-based tool for CVD risk prediction in South Africa. Results revealed clinical exposure to, and knowledge of genetics is limited in the South African public health setting, but most perceive PM approaches positively. Confidence in applying PM-based CVD risk stratification is variable, with those involved in research currently undertaking CVD risk stratification being more confident.

4.1. Positive Perceptions of PM-Based Tools despite Knowledge Gaps and Resource Constraints

Our study identified low rates of genomic testing, coupled with low self-perceived knowledge of genetics and PM-related concepts among clinicians and trainees practicing in South Africa's public health setting. Despite limited understanding and resource challenges, over two-thirds of respondents expressed positive perceptions of PM-based stratification tools. Such findings are not dissimilar to existing literature, which reveals low self-perceived knowledge but positive perceptions amongst healthcare providers across medical specializations, geographies and economic regions [23,25,26,30,32–35,41–45]. Assessment across five Implementing GeNomics In pracTicE (IGNITE) sites in the USA involving differing precision medicine applications revealed most clinicians believed genetic testing to be clinically useful. However, only a third thought they were adequately trained to provide care for genetically "high-risk" patients [23].

Although the value of integrated CVD risk scores was recognized, the perceived relevance of PM-based risk stratification to individual practice was variable and linked to varying practice standards across medical specialties, highlighting the need for clinician inclusion in the development of appropriate PM tools. The availability of risk estimates tailored to African populations was considered to be of particular value to providers, but the current limited transferability of PM-based approaches across ancestries [46] was also a key concern.

4.2. Addressing the Genetics Education Gap Is Paramount to Successful Adoption

Given that exposure to genetic testing and PM approaches is likely to increase with decreasing costs and increasing efforts to understand African genetic variation, these findings support prior calls for increased training and educational resources [23,30,41,47,48]. Furthermore, the limited genetic literacy amongst South African patients further necessitates additional support from healthcare providers to benefit from genetics and PM.

Lower self-perceived knowledge amongst consultants compared to trainees suggests that exposure to PM and related concepts in undergraduate training may have shifted in recent years. Alternatively, inexperienced clinicians may not be aware of the complexities of testing in a clinical setting as yet. Although different, mean knowledge scores for both groups were low, and highlight training requirements at undergraduate and practice levels. Genetics curricula can go unnoticed by clinicians as it is often integrated into the content of a course [49]. Simply increasing the volume and duration of genetic curricula is unlikely to be helpful, as advances in genomics do not occur in isolation and compete with other compelling learning demands [47]. Specialized courses and qualifications, although increasingly available, require substantial commitments and have relatively low uptake amongst non-genetic specialists [49]. Genomics education incorporated into healthcare providers' usual work activities, including departmental presentations and clinical meetings, has been identified as a preferred training strategy [50,51]. The most appropriate training approach is debated. However, there is consensus that education should address common misconceptions related to genetics, emphasize the clinical applications of genomics while not negating the complexity of the associated basic science, and ensure that clinicians can correctly utilize genetic-based risk estimations in their clinical practice [47,49,52].

Practicing clinicians have indicated a preference for online continued medical education, conferences, peer-reviewed literature, and in-clinic training [28,51,53]. Historically, most continued medical education has been passive and didactic, with limited impact

on changing practice. The most impactful approaches are those that are interactive, use a variety of instructional techniques, and focus on outcomes considered important by healthcare professionals. The appropriateness of such activities in South Africa will need to be further investigated, prioritizing culturally sensitive approaches that consider language differences and delivery format challenges.

4.3. Integrated Models Compatible with Current Clinical Practices Are Needed

In this study, confidence in utilizing the approach was variable and was directly influenced by previous CVD risk stratification exposure. Literature suggests that clinicians lack confidence in using and implementing new genomic tests and precision medicine approaches, and that this low confidence is linked to limited knowledge of genetics and related topics [28,30,33,42,50,54]. However, increased confidence has been associated with increased exposure to testing, or when the application is clearly defined and given within the context of a specific disease [23,55]. For example, oncologists already using genetic testing in clinical practice reported the highest confidence in using multimarker tumor panel results to guide patient care [55]. Increased confidence amongst those already using CVD stratification algorithms suggests that successfully integrating genetic risk assessment into existing clinical frameworks requires strategies compatible with current practice.

The areas clinicians lacked the most confidence in were related to the interpretation of score results and explaining potential impacts on insurance policies, which may reflect the lack of understanding of the genetic components of the score and the legislative framework. Respondents indicated a strong preference for a multidisciplinary model approach to testing, particularly interpreting score results and explaining such to patients. Multidisciplinary teams require specialized genetics skillsets, such as clinical geneticists and genetic counselors, which may not be feasible in the context of a skills shortage in South Africa. Increasing clinicians' comfort with using genomics in routine care may reduce the reliance on genetic specialists and improve the feasibility of genomic medicine over the long term. Additionally, the development of guidelines for PM-based scoring may help improve understanding in these areas and would support testing practices.

4.4. Funding Shortfalls and Skills Shortages Hinder PM Adoption in Resource-Scarce Settings

Similar to previous research and other settings, surveyed clinicians identified the affordability and accessibility of genetic services and genetics skills shortage as major impediments to PM implementation in South Africa's public health settings [18,25,30,41,42,56]. These challenges are not unique to PM and successful innovative models in funding and delivering care, especially within limited resource settings, should be explored.

Additional barriers raised from studies in well-resourced settings include those related to organizational structures (management, organizational culture and/or decentralized care), lack of clinical guidelines and the poor coordination of tests relative to treatment decisions [33]. Nephrologists in Australia highlighted barriers relating to their specific practice setting, e.g., leadership endorsement and goal setting [42], and medical oncologists in Canada identified patient genomic literacy and the current clinical utility of genomics as key challenges [35]. Although not identified as primary barriers in this South African study, these challenges are likely to be critical in the South African public setting as system-level barriers are addressed. Stakeholders will need to consider these challenges when developing a PM implementation roadmap, and leverage lessons learned as resource challenges are overcome and healthcare systems stabilize.

This study was novel in that it included a compulsory educational video before responding to the questionnaire. The purpose was to provide context and some exposure to PM across the responder group. However, several limitations to the study design merit discussion. Over and above the limited sample size and potential responder biases associated with online convenience sampling strategies [57], the study makes use of a newly developed instrument which has not yet been extensively tested for reliability and validity. There may have been issues with biased response sets (e.g., "faking good") that would

be difficult to detect in the absence of reverse-scored items [58]. Reassuringly, however, our results are intuitive and align with literature trends, diminishing concerns around the psychometric rigor of the instrument.

Lastly, for convenience, the study survey was administered to clinical practitioners based in large, relatively well-funded urban medical centers. Consequently, respondents in this study are unlikely to be representative of health professionals across South Africa. Understanding primary care providers' perceptions of and willingness to utilize PM-based tools in the primary care setting should be explored. Primary care provision is the foundation of care in South Africa, serving most communities, especially in rural areas, and where CVD risk stratification could have the greatest benefit [59–61]. Successful PM utilization requires multi-stakeholder co-operation, and thus the views of patients, insurers, policy making bodies and governments, should also be explored in follow up studies.

This study sheds light on an under-researched area of South African healthcare and contributes to a greater understanding of what will be required to successfully implement PM in South Africa, and potentially other resource-constrained environments.

5. Conclusions

Practicing clinicians have a positive perception of PM-based CVD risk stratification despite limited genomic knowledge, and they have a desire to adopt such tools in their practices. Limiting adoption barriers requires developing tools that leverage existing clinical approaches. The advancement of PM in South Africa requires an active effort to address the gap in genetic training whilst simultaneously investing in the country's genomics capacity, including both skill force and infrastructure.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jpm12091360/s1, Figure S1: Correlation matrix of predictors: Mean knowledge score, mean perception score and mean confidence score; Table S1: Survey questions, Table S2: Multivariate analysis, Table S3: Clinicians' expectations of PM-based risk stratification.

Author Contributions: Conceptualization, M.K. and M.R.; methodology, M.K. and M.R.; formal analysis, M.K., O.P. and A.M.; investigation, M.K.; resources, M.R.; data curation, M.K.; writing—original draft preparation, M.K.; writing—review and editing, M.K.; M.R., C.M.L., O.P. and A.M.; visualization, M.K. and O.P.; supervision, M.R., C.M.L. and O.P.; project administration, M.K.; funding acquisition, M.R. All authors have read and agreed to the published version of the manuscript.

Funding: M.R. holds a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Innovation, and administered by the National Research Foundation. M.K. is supported, in part, by a supplement to a National Institutes of Health grant (H3Africa Consortium grant U24HG009780). The APC was funded by NIH Grant U54HG006938 (AWI-Gen).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of UNIVERSITY OF THE WITWATERSRAND (protocol code M210355 and approval date 10/06/2021).

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: Available on reasonable request to the authors.

Acknowledgments: We gratefully acknowledge the time and attention of clinicians who completed the survey and shared their perceptions of PM-based risk stratification in South Africa, and Prof F. Raal at the Division of Endocrinology and Metabolism, Faculty of Health Sciences, University of the Witwatersrand for his invaluable subject expert support relating to cardiovascular diseases and associated clinical practice in South Africa. This paper represents independent research part-funded by the NIHR Maudsley Biomedical Research Centre at South London, the Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care in the UK, the US National Institutes of Health or the South African National Research Foundation.

Conflicts of Interest: Cathryn M. Lewis is a member of the Research and Development SAB at Myriad Neuroscience. The remaining authors declare that there are no competing interests. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Roth, G.A.; Mensah, G.A.; Johnson, C.O.; Addolorato, G.; Ammirati, E.; Baddour, L.M.; Barengo, N.C.; Beaton, A.Z.; Benjamin, E.J.; Benziger, C.P.; et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990–2019: Update from the GBD 2019 Study. *J. Am. Coll. Cardiol.* **2020**, *76*, 2982–3021. [CrossRef] [PubMed]
2. Gómez-Olivé, F.X.; Ali, S.A.; Made, F.; Kyobutungi, C.; Nonterah, E.; Micklesfield, L.; Alberts, M.; Boua, R.; Hazelhurst, S.; Debpuur, C.; et al. Regional and Sex Differences in the Prevalence and Awareness of Hypertension: An H3Africa AWI-Gen Study Across 6 Sites in Sub-Saharan Africa. *Glob. Heart* **2017**, *12*, 81–90. [CrossRef] [PubMed]
3. Abbafati, C.; Abbas, K.M.; Abbasi-Kangevari, M.; Abd-Allah, F.; Abdelalim, A.; Abdollahi, M.; Abdollahpour, I.; Abegaz, K.H.; Abolhassani, H.; Aboyans, V.; et al. Global Burden of 87 Risk Factors in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *Lancet* **2020**, *396*, 1223–1249. [CrossRef]
4. Califf, R.M.; Armstrong, P.W.; Carver, J.R.; D'Agostino, R.B.; Strauss, W.E. Task Force 5. Stratification of Patients into High, Medium Risk Subgroups for Purposes of Risk Factor Management General Perspectives on Risk and Effect on Intervention. *J. Am. Coll. Cardiol.* **1996**, *27*, 964–1047. [CrossRef]
5. Klug, E.Q.; Raal, F.J.; Marais, A.D.; Taskinen, M.R.; Dalby, A.J.; Schamroth, C.; Rapeport, N.; Jankelow, D.; Blom, D.J.; Catsicas, R.; et al. South African Dyslipidaemia Guideline Consensus Statement. *S. Afr. Med J.* **2012**, *102*, 178–188. [CrossRef]
6. Arnett, D.K.; Blumenthal, R.S.; Albert, M.A.; Buroker, A.B.; Goldberger, Z.D.; Hahn, E.J.; Himmelfarb, C.D.; Khera, A.; Lloyd-Jones, D.; McEvoy, J.W.; et al. 2019 ACC/AHA Guideline on the primary prevention of cardiovascular disease: A report of the american college of cardiology/American heart association task force on clinical practice guidelines. *Circulation* **2019**, *140*, e596–e646. [CrossRef]
7. Piepoli, M.F.; Hoes, A.W.; Agewall, S.; Albus, C.; Brotons, C.; Catapano, A.L.; Cooney, M.-T.; Corrà, U.; Cosyns, B.; Deaton, C.; et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts). Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur. Heart J.* **2016**, *37*, 2315–2381. [CrossRef]
8. Hippisley-Cox, J.; Coupland, C.; Brindle, P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: Prospective cohort study. *BMJ* **2017**, *357*, j2099. [CrossRef]
9. Lindbohm, J.V.; Sipilä, P.N.; Mars, N.J.; Pentti, J.; Ahmadi-Abhari, S.; Brunner, E.; Shipley, M.J.; Singh-Manoux, A.; Tabak, A.G.; Kivimaki, M. 5-year versus risk-category-specific screening intervals for cardiovascular disease prevention: A cohort study. *Lancet Public Health* **2019**, *4*, e189–e199. [CrossRef]
10. Tamlander, M.; Mars, N.; Pirinen, M.; Palotie, A.; Daly, M.; Riley-Gills, B.; Jacob, H.; Paul, D.; Runz, H.; John, S.; et al. Integration of questionnaire-based risk factors improves polygenic risk scores for human coronary heart disease and type 2 diabetes. *Commun. Biol.* **2022**, *5*, 158. [CrossRef]
11. Wagner, R.G.; Crowther, N.J.; Micklesfield, L.K.; Boua, P.R.; Nonterah, E.A.; Mashinya, F.; Mohamed, S.F.; Asiki, G.; Tollman, S.; Ramsay, M.; et al. Estimating the burden of cardiovascular risk in community dwellers over 40 years old in South Africa, Kenya, Burkina Faso and Ghana. *BMJ Glob. Health* **2021**, *6*, e003499. [CrossRef] [PubMed]
12. Nikpay, M.; Goel, A.; Won, H.H.; Hall, L.M.; Willenborg, C.; Kanoni, S.; Danish, S.; Kyriakou, T.; Nelson, C.P.; Hopewell, J.C.; et al. A Comprehensive 1000 Genomes-Based Genome-Wide Association Meta-Analysis of Coronary Artery Disease. *Nat. Genet.* **2015**, *47*, 1121–1130. [PubMed]
13. Khera, A.V.; Chaffin, M.; Aragam, K.G.; Haas, M.E.; Roselli, C.; Choi, S.H.; Natarajan, P.; Lander, E.S.; Lubitz, S.A.; Ellinor, P.T.; et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **2018**, *50*, 1219–1224. [CrossRef] [PubMed]
14. Lewis, C.M.; Vassos, E. Prospects for using risk scores in polygenic medicine. *Genome Med.* **2017**, *9*, 96. [CrossRef] [PubMed]
15. Inouye, M.; Abraham, G.; Nelson, C.P.; Wood, A.M.; Sweeting, M.J.; Dudbridge, F.; Lai, F.Y.; Kaptoge, S.; Brozynska, M.; Wang, T.; et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. *J. Am. Coll. Cardiol.* **2018**, *72*, 1883–1893. [CrossRef] [PubMed]
16. Riveros-Mckay, F.; Weale, M.E.; Moore, R.; Selzam, S.; Krapohl, E.; Sivley, R.M.; Tarran, W.A.; Sørensen, P.; Lachapelle, A.S.; Griffiths, J.A.; et al. Integrated Polygenic Tool Substantially Enhances Coronary Artery Disease Prediction. *Circ. Genom. Precis. Med.* **2021**, *14*, 192–200. [CrossRef]
17. Kullo, I.J.; Jouni, H.; Austin, E.E.; Brown, S.-A.; Kruisselbrink, T.M.; Isseh, I.N.; Haddad, R.A.; Marroush, T.S.; Shameer, K.; Olson, J.E.; et al. Incorporating a Genetic Risk Score into Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). *Circulation* **2016**, *133*, 1181–1188. [CrossRef]
18. Pereira, L.; Mutesa, L.; Tindana, P.; Ramsay, M. African genetic diversity and adaptation inform a precision medicine agenda. *Nat. Rev. Genet.* **2021**, *22*, 284–306. [CrossRef]

1

19. Chikowore, T.; Kamiza, A.B.; Oduaran, O.H.; Machipisa, T.; Fatumo, S. Non-communicable diseases pandemic and precision medicine: Is Africa ready? *eBioMedicine* **2021**, *65*, 103260. [CrossRef]
20. Kamp, M.; Krause, A.; Ramsay, M. Has translational genomics come of age in Africa? *Hum. Mol. Genet.* **2021**, *30*, R164–R173. [CrossRef]
21. Tekola-Ayele, F.; Rotimi, C.N. Translational Genomics in Low- and Middle-Income Countries: Opportunities and Challenges. *Public Health Genom.* **2015**, *18*, 242–247. [CrossRef] [PubMed]
22. Rotimi, C.; Abayomi, A.; Abimiku, A.; Adabayeri, V.M.; Adebamowo, C.; Adebiyi, E.; Ademola, A.D.; Adeyemo, A.; Adu, D.; Affolabi, D.; et al. Research capacity. Enabling the genomic revolution in Africa. *Science* **2014**, *344*, 1346–1348. [CrossRef] [PubMed]
23. Obeng, A.O.; Fei, K.; Levy, K.D.; Elsey, A.R.; Pollin, T.I.; Ramirez, A.H.; Weitzel, K.W.; Horowitz, C.R. Physician-Reported Benefits and Barriers to Clinical Implementation of Genomic Medicine: A Multi-Site IGNITE-Network Survey. *J. Pers. Med.* **2018**, *8*, 24. [CrossRef] [PubMed]
24. Hartzler, A.; McCarty, C.A.; Rasmussen, L.V.; Williams, M.S.; Brilliant, M.; Bowton, E.A.; Clayton, E.W.; Faucett, W.A.; Ferryman, K.; Field, J.R.; et al. Stakeholder engagement: A key component of integrating genomic information into electronic health records. *Genet. Med.* **2013**, *15*, 792–801. [CrossRef] [PubMed]
25. Abdela, O.A.; Bhagavathula, A.S.; Gebreyohannes, E.A.; Tegegn, H.G. Ethiopian health care professionals' knowledge, attitude, and interests toward pharmacogenomics. *Pharm. Pers. Med.* **2017**, *10*, 279–285. [CrossRef] [PubMed]
26. Muzoriana, N.; Gavi, S.; Nembaware, V.; Dhoro, M.; Matimba, A. Knowledge, Attitude, and Perceptions of Pharmacists and Pharmacy Students towards Pharmacogenomics in Zimbabwe. *Pharmacy* **2017**, *5*, 36. [CrossRef]
27. Haga, S.; Burke, W.; Ginsburg, G.; Mills, R.; Agans, R. Primary care physicians' knowledge of and experience with pharmacogenetic testing. *Clin. Genet.* **2012**, *82*, 388–394. [CrossRef]
28. Carroll, J.C.; Morrison, S.; Miller, F.A.; Wilson, B.J.; Permaul, J.A.; Allanson, J. Anticipating the primary care role in genomic medicine: Expectations of genetics health professionals. *J. Community Genet.* **2021**, *12*, 559–568. [CrossRef]
29. Chambers, C.; Axell-House, D.; Mills, G.; Bittner-Fagan, H.; Rosenthal, M.; Johnson, M.; Stello, B. Primary Care Physicians' Experience and Confidence with Genetic Testing and Perceived Barriers to Genomic Medicine. *J. Fam. Med.* **2015**, *2*, 1024.
30. Yu, M.W.C.; Fung, J.L.F.; Ng, A.P.P.; Li, Z.; Lan, W.; Chung, C.C.Y.; Li, Y.; Liu, Y.; Chung, B.H.Y.; Wong, W.C.W. Preparing genomic revolution: Attitudes, clinical practice, and training needs in delivering genetic counseling in primary care in Hong Kong and Shenzhen, China. *Mol. Genet. Genom. Med.* **2021**, *9*, e1702. [CrossRef]
31. Rubanovich, C.K.; Cheung, C.; Mandel, J.; Bloss, C.S. Physician preparedness for big genomic data: A review of genomic medicine education initiatives in the United States. *Hum. Mol. Genet.* **2018**, *27*, R250–R258. [CrossRef] [PubMed]
32. Addis, A.; Trotta, F.; Tafuri, G.; de Fiore, L. Information Needs on Precision Medicine: A Survey of Italian Health Care Professionals. *Ann. Ist. Super. Sanita* **2018**, *54*, 316–323. [CrossRef] [PubMed]
33. Albassam, A.; Alshammari, S.; Ouda, G.; Koshy, S.; Awad, A. Knowledge, perceptions and confidence of physicians and pharmacists towards pharmacogenetics practice in Kuwait. *PLoS ONE* **2018**, *13*, e0203033. [CrossRef] [PubMed]
34. Selkirk, C.G.; Weissman, S.M.; Anderson, A.; Hulick, P.J. Physicians' Preparedness for Integration of Genomic and Pharmacogenetic Testing into Practice within a Major Healthcare System. *Genet. Test. Mol. Biomark.* **2013**, *17*, 219–225. [CrossRef] [PubMed]
35. Chow-White, P.; Ha, D.; Laskin, J. Knowledge, attitudes, and values among physicians working with clinical genomics: A survey of medical oncologists. *Hum. Resour. Health* **2017**, *15*, 42. [CrossRef] [PubMed]
36. Harris, P.A.; Taylor, R.; Thielke, R.; Payne, J.; Gonzalez, N.; Conde, J.G. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **2009**, *42*, 377–381. [CrossRef]
37. Harris, P.A.; Taylor, R.; Minor, B.L.; Elliott, V.; Fernandez, M.; O'Neal, L.; McLeod, L.; Delacqua, G.; Delacqua, F.; Kirby, J.; et al. The REDCap consortium: Building an international community of software platform partners. *J. Biomed. Inform.* **2019**, *95*, 103208. [CrossRef]
38. RStudio Team. *RStudio: Integrated Development for R.*; RStudio: Boston, MA, USA, 2020.
39. Revelle, W.R. *Psych: Procedures for Personality and Psychological Research. Software*; Northwestern University: Evanston, IL, USA, 2017.
40. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016; ISBN 978-3-319-24277-4.
41. Spiech, K.M.; Tripathy, P.R.; Woodcock, A.M.; Sheth, N.A.; Collins, K.S.; Kannegolla, K.; Sinha, A.D.; Sharfuddin, A.A.; Pratt, V.M.; Khalid, M.; et al. Implementation of a Renal Precision Medicine Program: Clinician Attitudes and Acceptance. *Life* **2020**, *10*, 32. [CrossRef]
42. Jayasinghe, K.; Quinlan, C.; Mallett, A.J.; Kerr, P.G.; McClaren, B.; Nisselle, A.; Mallawaarachchi, A.; Polkinghorne, K.R.; Patel, C.; Best, S.; et al. Attitudes and Practices of Australian Nephrologists Toward Implementation of Clinical Genomics. *Kidney Int. Rep.* **2021**, *6*, 272–283. [CrossRef]
43. Carroll, J.C.; Allanson, J.; Morrison, S.; Miller, F.A.; Wilson, B.J.; Permaul, J.A.; Telner, D. Informing Integration of Genomic Medicine Into Primary Care: An Assessment of Current Practice, Attitudes, and Desired Resources. *Front. Genet.* **2019**, *10*, 1189. [CrossRef]

1

44. Alharbi, A.A.; Shaqran, T.M.; Eltobgy, A.A.E.; Albalawi, A.R.; Alnawmasi, W.S. Physicians' Perspective on Diabetes Mellitus Management within the Context of Personalized Medicine Era in Tabuk Governorate, Saudi Arabia. *Open Access Maced. J. Med Sci.* **2019**, *7*, 1706–1711. [CrossRef] [PubMed]
45. Albitar, L.; Alchamat, G.A. Pharmacogenetics: Knowledge assessment amongst Syrian pharmacists and physicians. *BMC Health Serv. Res.* **2021**, *21*, 1031. [CrossRef]
46. Martin, A.R.; Kanai, M.; Kamatani, Y.; Okada, Y.; Neale, B.M.; Daly, M.J. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **2019**, *51*, 584–591. [CrossRef] [PubMed]
47. Doll, B.; de Castro, M.J.; Fries, M.H.; Pock, A.R.; Seibert, D.; Yang, W. Precision Medicine—A Demand Signal for Genomics Education. *Mil. Med.* **2021**, *187*, 40–46. [CrossRef] [PubMed]
48. Crellin, E.; McClaren, B.; Nisselle, A.; Best, S.; Gaff, C.; Metcalfe, S. Preparing Medical Specialists to Practice Genomic Medicine: Education an Essential Part of a Broader Strategy. *Front. Genet.* **2019**, *10*, 789. [CrossRef]
49. Cornel, M.C. Evidence-Based Genetic Education of Non-Genetic-Expert Physicians: Experiences Over Three Decades in Amsterdam. *Front. Genet.* **2019**, *10*, 712. [CrossRef]
50. Nisselle, A.; King, E.A.; McClaren, B.; Janinski, M.; Metcalfe, S.; Gaff, C. Measuring physician practice, preparedness and preferences for genomic medicine: A national survey. *BMJ Open* **2021**, *11*, e044408. [CrossRef]
51. Laplace, B.; Calvet, B.; Lacroix, A.; Mouchabac, S.; Picard, N.; Girard, M.; Charles, E. Acceptability of Pharmacogenetic Testing among French Psychiatrists, a National Survey. *J. Pers. Med.* **2021**, *11*, 446. [CrossRef]
52. Wilcox, R.L.; Adem, P.V.; Afshinnekoo, E.; Atkinson, J.B.; Burke, L.W.; Cheung, H.; Dasgupta, S.; DeLaGarza, J.; Joseph, L.; LeGallo, R.; et al. The Undergraduate Training in Genomics (UTRIG) Initiative: Early & active training for physicians in the genomic medicine era. *Pers. Med.* **2018**, *15*, 199–208. [CrossRef]
53. Reed, E.K.; Johansen Taber, K.A.; Ingram Nissen, T.; Schott, S.; Dowling, L.O.; O'Leary, J.C.; Scott, J.A. What works in genomics education: Outcomes of an evidenced-based instructional model for community-based physicians. *Genet. Med.* **2016**, *18*, 737–745. [CrossRef]
54. Johnson, L.-M.; Valdez, J.M.; Quinn, E.A.; Sykes, A.D.; McGee, R.B.; Nuccio, R.; Hines-Dowell, S.J.; Baker, J.N.; Kesserwan, C.; Nichols, K.E.; et al. Integrating next-generation sequencing into pediatric oncology practice: An assessment of physician confidence and understanding of clinical genomics. *Cancer* **2017**, *123*, 2352–2359. [CrossRef] [PubMed]
55. De Moor, J.S.; Gray, S.W.; Mitchell, S.A.; Klabunde, C.N.; Freedman, A.N. Oncologist Confidence in Genomic Testing and Implications for Using Multimarker Tumor Panel Tests in Practice. *JCO Precis. Oncol.* **2020**, *4*, 620–631. [CrossRef] [PubMed]
56. Manolio, T.A.; Abramowicz, M.; Al-Mulla, F.; Anderson, W.; Balling, R.; Berger, A.C.; Bleyl, S.; Chakravarti, A.; Chantratita, W.; Chisholm, R.L.; et al. Global implementation of genomic medicine: We are not alone. *Sci. Transl. Med.* **2015**, *7*, 1–7. [CrossRef] [PubMed]
57. Fricker, R.D. Sampling Methods for Web and E-Mail Surveys. In *The SAGE Handbook of Online Research Methods*; SAGE Publications Ltd.: London, UK, 2008.
58. Tourangeau, R.; Rips, L.; Rasinski, K. *The Psychology of Survey Response*; Cambridge University Press: Cambridge, UK, 2009.
59. Ojji, D.B.; Lamont, K.; Ojji, O.I.; Egenti, B.N.; Sliwa, K. Primary care in the prevention, treatment and control of cardiovascular disease in sub-Saharan Africa. *Cardiovasc. J. Afr.* **2017**, *28*, 251–256. [CrossRef]
60. Beaglehole, R.; Epping-Jordan, J.A.; Patel, V.; Chopra, M.; Ebrahim, S.; Kidd, M.; Haines, A. Improving the prevention and management of chronic disease in low-income and middle-income countries: A priority for primary health care. *Lancet* **2008**, *372*, 940–949. [CrossRef]
61. Mash, R.; Almeida, M.; Wong, W.C.W.; Kumar, R.; Von Pressentin, K.B. The roles and training of primary care doctors: China, India, Brazil and South Africa. *Hum. Resour. Health* **2015**, *13*, 93. [CrossRef]

APPENDIX I: ORIGINAL PUBLICATION 4

Original research

BMJ Global Health

Multimorbidity in African ancestry populations: a scoping review

Michelle Kamp ,[1,2] Okechinyere Achilonu ,[2] Isaac Kisiangani ,[3] Daniel Maina Nderitu ,[3] Phelelani Thokozani Mpangase ,[2] Girmaw Abebe Tadesse ,[4] Kayode Adetunji,[2] Samuel Iddi ,[3] Skyler Speakman,[4] Scott Hazelhurst ,[2,5] Gershim Asiki ,[3,6] Michèle Ramsay ,[1,2] as members of the MADIVA Research Hub and the DS-I Africa Consortium

To cite: Kamp M, Achilonu O, Kisiangani I, *et al*. Multimorbidity in African ancestry populations: a scoping review. *BMJ Glob Health* 2023;**8**:e013509. doi:10.1136/bmjgh-2023-013509

Handling editor Seye Abimbola

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjgh-2023-013509).

Received 24 July 2023
Accepted 1 November 2023

For numbered affiliations see end of article.

Correspondence to
Professor Michèle Ramsay; Michele.Ramsay@wits.ac.za

ABSTRACT

Objectives Multimorbidity (MM) is a growing concern linked to poor outcomes and higher healthcare costs. While most MM research targets European ancestry populations, the prevalence and patterns in African ancestry groups remain underexplored. This study aimed to identify and summarise the available literature on MM in populations with African ancestry, on the continent, and in the diaspora.

Design A scoping review was conducted in five databases (PubMed, Web of Science, Scopus, Science Direct and JSTOR) in July 2022. Studies were selected based on predefined criteria, with data extraction focusing on methodology and findings. Descriptive statistics summarised the data, and a narrative synthesis highlighted key themes.

Results Of the 232 publications on MM in African-ancestry groups from 2010 to June 2022—113 examined continental African populations, 100 the diaspora and 19 both. Findings revealed diverse MM patterns within and beyond continental Africa. Cardiovascular and metabolic diseases are predominant in both groups (80% continental and 70% diaspora). Infectious diseases featured more in continental studies (58% continental and 16% diaspora). Although many papers did not specifically address these features, as in previous studies, older age, being women and having a lower socioeconomic status were associated with a higher prevalence of MM, with important exceptions. Research gaps identified included limited data on African-ancestry individuals, inadequate representation, under-represented disease groups, non-standardised methodologies, the need for innovative data strategies, and insufficient translational research.

Conclusion The growing global MM prevalence is mirrored in African-ancestry populations. Recognising the unique contexts of African-ancestry populations is essential when addressing the burden of MM. This review emphasises the need for additional research to guide and enhance healthcare approaches for African-ancestry populations, regardless of their geographic location.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Multimorbidity (MM) is a growing global health challenge. While much research focuses on high-income countries and mostly European-ancestry populations, people with African-ancestry are largely under-represented. Even when included, African Americans are often the primary focus and are known to be poorly representative of continental African populations. The definition of MM and study methodologies vary across studies making global comparisons challenging.

WHAT THIS STUDY ADDS

⇒ This study is the first comprehensive literature review on MM in African-ancestry populations. It shows that African-ancestry populations are not homogeneous and treating them as such may lead to erroneous conclusions and inappropriate public health interventions. The analysis of the literature identified crucial differences between MM studies and outcomes when comparing continental and diaspora African-ancestry populations and highlighted the pitfalls of using studies on African American populations as sole representatives for all African-ancestry populations.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ By emphasising the under-representation of studies focused on African and African-ancestry populations, this study encourages more inclusive research. It cautions against over-relying on diaspora African-ancestry populations as proxies for continental African groups and makes a plea for prioritising more nuanced and larger studies from Africa. Furthermore, it underscores significant gaps in MM research and calls on the global scientific and health communities to include more diverse African-ancestry communities from the continent and the diaspora, and to be more context-aware in their methodologies, interpretations and conclusions.

INTRODUCTION

Multimorbidity (MM), the co-occurrence of two or more chronic conditions in an

individual, is a growing global challenge.[1 2] MM is associated with increased premature mortality,[2] lowered quality of life,[3] diminished mental health,[4 5] increased polypharmacy risk[6] and intensified health services utilisation and associated costs, particularly in resource-poor settings.[7]

There are, however, different interpretations of the MM definition and a lack of global consensus. Some use the above-described definition but highlight that there is no index condition or dominant disease in the affected individual.[8 9] This is consistent with the WHO's definition emphasising the chronic and long-term nature of conditions and the absence of hierarchy or clustering of diseases.[10] Comorbidity, on the other hand, refers to additional conditions in patients with a primary condition of interest. The distinction between MM and comorbidity is important but not always clear, despite its critical importance for accurate phenotyping, research analyses and in healthcare settings. Additionally, there is variability in the nature and number of chronic diseases included in MM studies and the measurement methods used to assess MM across studies, as highlighted by a recent Delphi study.[11]

Improved life expectancy resulting from global health efforts has led to a rise in MM due to ageing populations and shifting lifestyle risk factors, including obesity and physical inactivity.[1 12–14] MM encompasses various combinations of chronic conditions, with concordant MM sharing origins and treatment requirements, and discordant MM involving unrelated or differently managed conditions.[15] As reviewed by Roomaney *et al*,[16] in low and middle-income countries (LMICs), MM is further complicated by overlapping infectious disease burdens, poverty-related environmental stressors, limited social infrastructure and under-resourced healthcare settings. It is estimated that at least one-third of adults residing in LMICs have multiple chronic conditions,[17] yet the epidemiology of MM in these regions remains relatively unknown.

Despite the substantial growth in MM studies, most have been conducted in high-income countries or among those of European-ancestry, limiting the generalisability of findings.[16 18] Although studies involving diverse ancestries, particularly European and African American (AA) populations, indicate increased MM risk and adverse health outcomes among individuals of African-ancestry,[19 20] the analyses are primarily focused on AAs and often based on small sample sizes. Moreover, most studies examine single disease outcomes or comorbidities with an index disease, leaving the true burden of MM, its epidemiological characteristics and healthcare strategies in African-ancestry populations largely unknown.[21]

This scoping review is timely as the global health community is advocating for diversity and inclusion in research studies, a viewpoint strongly echoed by researchers, healthcare professionals, policymakers and pharma. It is well recognised that African and African-ancestry populations are under-represented and understudied[19] and this review and critical analysis of the published studies are, therefore, important to raise awareness. When global studies report continent-wide prevalence, the projections for African populations are often based on a few small studies on selected communities. They do not reflect the heterogeneity across Africa and in diaspora populations, providing false impressions that could lead to disastrous public health interventions in communities where the data and findings do not apply.

In addition to the limited representation of diverse populations, MM studies face several methodological challenges, including the lack of international consensus on the definition of MM and the frequent use of non-standardised phenotype and laboratory-based measurement methods. These inconsistencies hinder research, comparative assessments and the translation of findings into guidelines and interventions.[22]

This scoping review seeks to identify, evaluate and summarise existing evidence on MM prevalence, patterns, associated factors, and research gaps specific to African-ancestry populations. It provides insight into similar and divergent trends among continental and diaspora African-ancestry individuals, in order to inform the most appropriate and relevant future research priorities.

The following research questions were addressed to guide the scoping review: (1) How many studies have been done on this topic, and what are the geographical locations of the studied populations? (2) Which definitions of MM and what study designs are commonly used in MM studies that include African-ancestry populations? (3) What are the common MM disease clusters among African-ancestry populations living on the continent and in the diaspora? (4) How does MM prevalence differ across sociodemographic characteristics and are these patterns consistent across continental and diasporic African populations?

METHODS

To assess MM within African-ancestry populations, a scoping review methodology was employed rather than a systematic review or meta-analytic approach as a scoping review is more appropriate when aiming to assess an emerging field of literature and identify research gaps.[23] This review began with the establishment of a research team consisting of members with expertise in epidemiology, demography, clinical medicine, genetics and data science research in Africa. Members were from several African countries (Kenya, Nigeria, South Africa, Ghana and Uganda) and advised on the broad study aim and protocol, including search terms, databases and thematic synthesis. A review protocol for this study is not available.

Search strategy and database search

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews.[24] The initial search was implemented in July 2022 and included a comprehensive literature search by one researcher (MK) in consultation with a librarian and field experts. Five electronic databases (PubMed,

1

Table 1 Search term strategy

	Search themes	MeSH terms
1	Multimorbidity	'comorbidity' OR 'co-morbidity' OR 'multimorbidity' OR 'multi-morbidity' OR 'multiple chronic conditions' OR 'chronic conditions' OR 'chronic disease'
2	African populations	'Africa' OR 'African' OR 'African population' OR 'African ancestry' OR 'African-ancestry'

Scopus, Science Direct, Web of Science and JSTOR) were searched to identify relevant publications on MM in individuals of African-ancestry. The search strategy incorporated a combination of MeSH (Medical Subject Headings) terms and keywords, including 'multimorbidity', 'comorbidity' and 'African population' and their synonyms. The final search terms are listed in table 1 and detailed in online supplemental materials. Search terms relating to MM were derived from previous systematic reviews, and search themes 1 and 2 were combined using the Boolean operator 'AND'. The search was limited to the title and abstract fields only, as a trial search found searching full text reduced the ability to detect relevant papers. Similarly, for this reason, a list of African countries was not included in the search strategy. Although the search terms were consistent across all databases, the search query was tailored to the specific requirements of each database.

The publication time frame was limited to 1 January 2010–31 June 2022, as research on MM has increased significantly since 2010, with approximately 80% of publications published after this date.[18]

Citation management and study selection

After conducting database searches, the search output citations were downloaded in BibTeX format and uploaded to an electronic screening and data management website, Rayyan.[25] The deduplication function was employed to remove duplicates. For the first level of screening, only the title and abstracts of unique entries were reviewed to avoid wasting resources on articles that did not meet the minimum inclusion criteria. Entries were screened independently and blindly by MK and three other reviewers (IK, DMN and GAT), and studies deemed irrelevant were discarded. A third group (GA and MR) assisted with conflict resolution.

Eligibility criteria

The inclusion criteria were studies reporting epidemiological data on MM or comorbidity in individuals of African-ancestry aged 18 years and above from continental Africa and studies including populations of African-ancestry in other global regions (eg, Caribbean, Europe, North America). Studies written in French were included to mitigate missing data from Francophone Africa.

The definition of MM used in this review was based on the Academy of Medical Sciences definition—the

Table 2 Study inclusion/exclusion criteria for the scoping review

	Inclusion	Exclusion
Methodology	► MM defined as a combination of recognised diseases/conditions (eg, self-report or International Classification of Disease 9th revision (ICD-9) or 10th revision (ICD-10) codes)	► MM defined as a combination of symptoms or pre-disease conditions, not defined as ICD-9 or ICD-10 diseases (eg, predisease, frailty, disability and quality of life) ► Transitions or trajectories within a single disease or from one condition into another (eg, cancer progression)
Study design	► Cross-sectional studies ► Longitudinal quantitative studies, including retrospective and prospective cohort studies ► Qualitative studies	► Systematic reviews ► Meta-analyses ► Case studies ► Expert opinion/committee reports
Population	► Adult humans (18+ years) ► African continent ► African-ancestry diaspora	► Infants, children or adolescents (<18 years) ► Animal research
Publication date	► 01 January 2010–31 June 2022	► Prior to 01 January 2010 ► Post 31 June 2022
Publication type	► Peer-reviewed journal articles ► Clinical trials	► Case reports ► Reviews ► Editorials ► Letters
Language	► English ► French	► Other languages

MM, multimorbidity.

1

coexistence of two or more chronic conditions, which included conditions in the following categories: physical non-communicable disease (NCD) of long duration, such as cardiovascular disease or cancer, a mental health condition of long duration, such as a mood disorder or dementia, or an infectious disease of long duration, such as HIV or hepatitis.[21] The term 'comorbidity' was included, as it has been used interchangeably with MM in the past, although it is now recognised as a distinct concept.[22] As per the Academy of Medical Sciences definition, acute conditions (<6 months) were not included in our definition. We use the term diaspora to specifically refer to populations of African-ancestry who no longer reside on the African continent, contrasting them to continental African populations residing in Africa. Table 2 provides a detailed summary of our inclusion and exclusion criteria.

Studies were excluded when both initial reviewers indicated that they did not meet the eligibility criteria. When their responses were discordant, a third reviewer assessed the abstract, and in some cases, reviewed the full publication before deciding whether the publication met inclusion criteria.

Data extraction and analyses

All citations deemed relevant after title and abstract screening underwent full-text review. Four reviewers (IK, MK, DMN and GAT) independently assessed the full texts (more than one person evaluated some citations), and study characteristics were captured electronically using a template to extract relevant data (online supplemental table S1). Data on MM definition, the approach to measuring MM, the conditions included, and the characteristics of the population under investigation were extracted. Studies were categorised by geographical location, participant ancestry and study design. The chronic conditions assessed to define MM were collected and organised into categories based on the body system affected.[11] Five reviewers independently extracted data and updated the extraction spreadsheet (OA, IK, MK, DMN and GAT). For the extracted variables (online supplemental table S1), the mean and SD, the absolute number or the percentage were recorded as appropriate.

To effectively analyse and interpret the extensive range of diseases evaluated, a categorisation approach was employed. Diseases were grouped into nine broader clusters (table 3). These clusters were derived using a body system approach with nuances to reflect the major global mortality and morbidity disease burdens over the last decade[26] and through consultation with clinicians.

Data summary and synthesis

The data extraction spreadsheet was imported into Microsoft Excel[27] for validation and coding. The data were then exported into R Statistical Computing Tool (V.4.3)[28] for analysis. Descriptive statistics were calculated

Table 3 Disease clusters and the associated definition and diseases

Cluster	Cluster detail	Example conditions
Cancer	Diseases in which abnormal cells divide uncontrollably and can invade nearby tissues.	Breast, lung, colon, cervical, and prostate cancers, Kaposi Sarcoma, etc.
Cardiovascular	Chronic diseases that affect the heart and blood vessels.	Arrhythmias, atrial fibrillation, cardiomyopathy, coronary artery disease, heart failure, hypertension, peripheral artery disease, stroke, etc.
Infectious	Conditions caused by persistent infections; including bacterial, viral or parasitic infections.	Chronic viral hepatitis (hepatitis D, E and G), hepatitis B, hepatitis C, HIV, human papillomavirus (HPV), long-COVID, tuberculosis, etc.
Metabolic	Chronic diseases that affect the body's metabolic system.	Dyslipidaemia, fatty liver disease, metabolic syndrome, obesity, type 2 diabetes mellitus, etc.
Musculoskeletal	Chronic conditions that affect the bones, joints and muscles.	Chronic back pain, osteoarthritis, osteoporosis, rheumatoid arthritis, back syndrome with radiating pain, back syndrome with non-radiating pain, etc.
Neurological	Conditions that affect the brain and nervous system.	Alzheimer's disease, dementia, epilepsy, multiple sclerosis, Parkinson's disease, autistic spectrum disorder, etc.
Psychiatric	Disorders that affect mood, thinking and behaviour.	Attention deficit hyperactivity disorder (ADHD), bipolar, depression, generalised anxiety disorder (GAD), substance use disorders, post-traumatic stress disorder (PTSD), schizophrenia, etc.
Renal	Chronic diseases that affect the kidneys.	Chronic kidney disease, renal failure, end-stage renal disease, etc.
Respiratory	Chronic diseases that affect the lungs.	Asthma, bronchiectasis, chronic obstructive pulmonary disease (COPD), cystic fibrosis, sleep apnoea, pneumonia, etc.

1

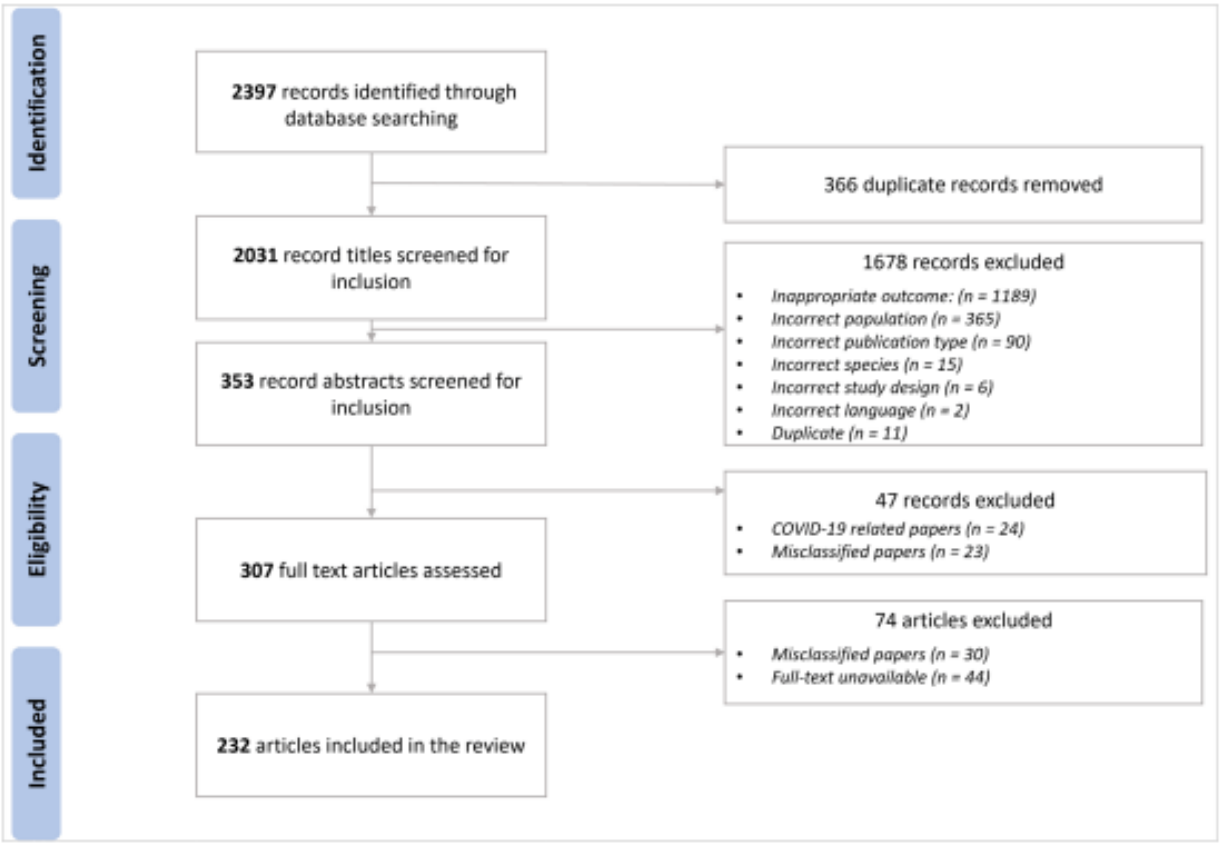


Figure 1 PRISMA-ScR flow diagram of the study selection process. Misclassified papers refer to those that on further analysis did not meet the inclusion criteria, for example, inappropriate outcome (did not meet the criteria for MM), Incorrect publication type (not original research). MM, multimorbidity; PRISMA-ScR, Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews.

to summarise the data, with frequencies and percentages utilised to describe nominal data. An UpSet plot[29] for the two groups of African-ancestry populations was constructed to visualise the disease clusters. A thematic synthesis was used to identify key themes emanating from the included articles.

Patient and public involvement

Patients and the public were not involved in this study. Clinicians known by the authors and working in MM across Africa were invited to comment.

RESULTS

Below we describe the results of our article search, the analyses of the included papers, and the themes identified through content analyses.

Search results

Figure 1 provides an overview of the search results. The initial search yielded 2397 articles, and after removing duplicates (366), 2031 unique records remained. Out of these, 1678 articles were excluded after screening. During the title and abstract screening phase, reviewers had disagreements on the exclusion/inclusion of 15% of the articles, which were resolved through discussion. A total of 353 abstracts were evaluated for eligibility, with 47 articles excluded due to misclassification (n=23) and the exclusion of papers focusing on MM in COVID-19 severity (n=24). 307 full-text articles were assessed, and an additional 74 exclusions were made due to misclassification (n=30) or the full text being unavailable (n=44). The BibTex citation file of analysed articles is available on GitHub (https://github.com/MADIVA-DSI/MM-in-African-ancestry-pops.git).

While multiple criteria led to paper exclusion, only one reason was assigned to each study. The most common exclusion reason was 'inappropriate outcome', indicating the research question did not address MM (as defined by the Academy of Medical Sciences) in African-ancestry populations, the focus of this review. These studies often examined comorbidity or broader chronic disease prevalence, prevention, and treatment strategies.

Study characteristics

The number of publications increased over the past decade from 5 in 2010 to 11 in 2022 (figure 2). While

1

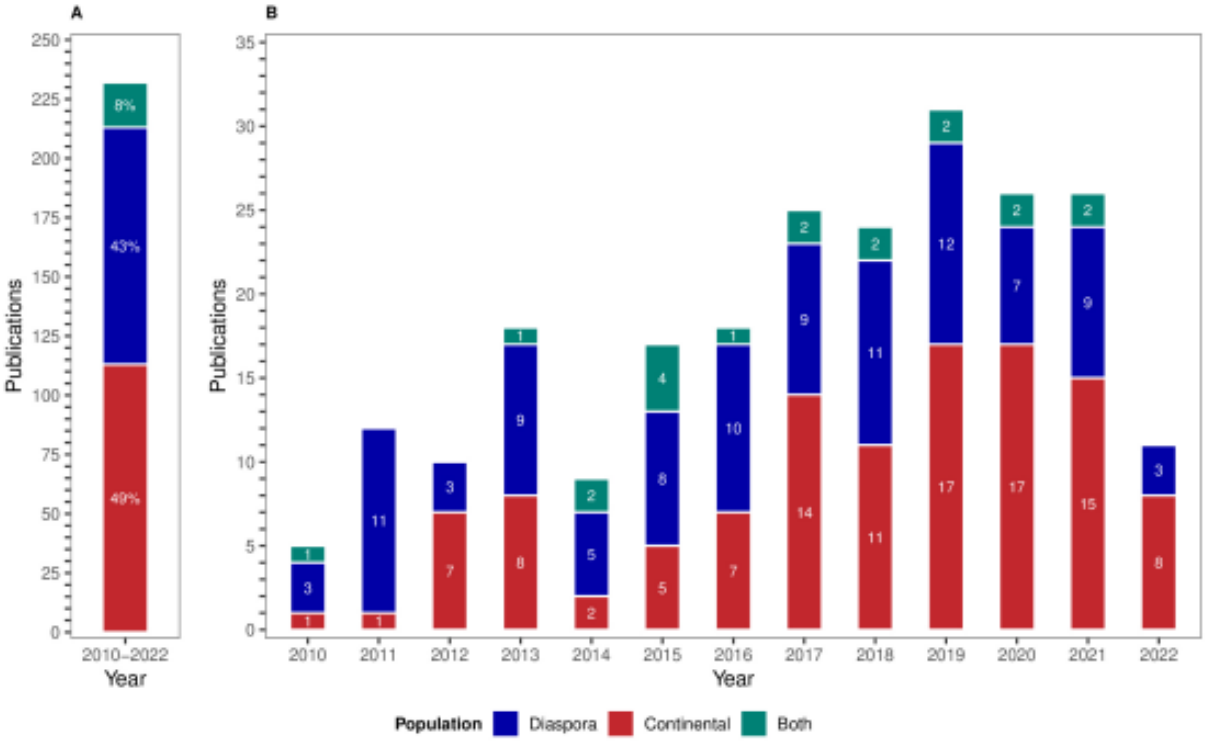


Figure 2 Number of publications by African-ancestry population geographic origin. (A) Proportion of publications by population group over the full assessed timeframe (January 2010–June 2022) (n=232). (B) Proportion of papers by African-ancestry population origin per year.

there was a decline in papers addressing MM in African-ancestry populations from 2020 to 2022, the upward trend persisted when including papers related to MM and COVID-19. During this period, we identified 23 publications associating MM as a risk factor for severe COVID-19 in African populations (2020: 8 studies, 2021: 10 studies, 2022: 5 studies).

Geographical location of African-ancestry populations studied

Figure 2 shows the number of publications per year by African ancestry population, distinguishing between those from the diaspora and the continent. The data show there were slightly more publications among continental populations (49%, n=113) than diaspora (43%, n=100), with 8% conducted in multiple countries involving both populations (n=19).

Figure 3 illustrates the global distribution of studies investigating MM in African-ancestry populations. Figure 3A presents the overall coverage of papers, while figure 3B displays the proportion of papers originating from different regions. Among the 113 papers focused on continental populations, the majority (55%) were conducted in Southern Africa (n=61), followed by East Africa (22%, n=24) and West Africa (15%, n=17). Less than 5% of studies included multiple African regions (n=3).

Diaspora studies were concentrated in North America (95%, n=94), specifically among AA, who were often part of diverse ancestry studies examining the influence of race on MM-related health outcomes. In Europe, diaspora studies typically involved recently migrated individuals (born in Africa or first-generation migrants).

In continental and diaspora groups, epidemiological studies focused on determining disease clusters and prevalence in specific populations. The most common study designs were cross-sectional (67%, n=155) and longitudinal (31%, n=73). Incidence measurement of MM was primarily observed in diaspora studies. Around two-thirds of the studies were conducted in hospital settings, including primary care and clinical settings, while 36% involved the general population. In the diaspora, secondary analysis of large cohorts or databases was prominent (64% of diaspora studies), while continental studies often utilised prospective hospital-based cohorts (32%). Continental studies had smaller sample sizes (median (range)=990 (14 to 417 786)) compared with diaspora studies (median (range)=2022 (20 to 3 739 528)).

Studies examined various risk factors for MM, including sociodemographic variables like age, sex and socioeconomic status (SES). Research conducted in North America was particularly interested in comparing

1

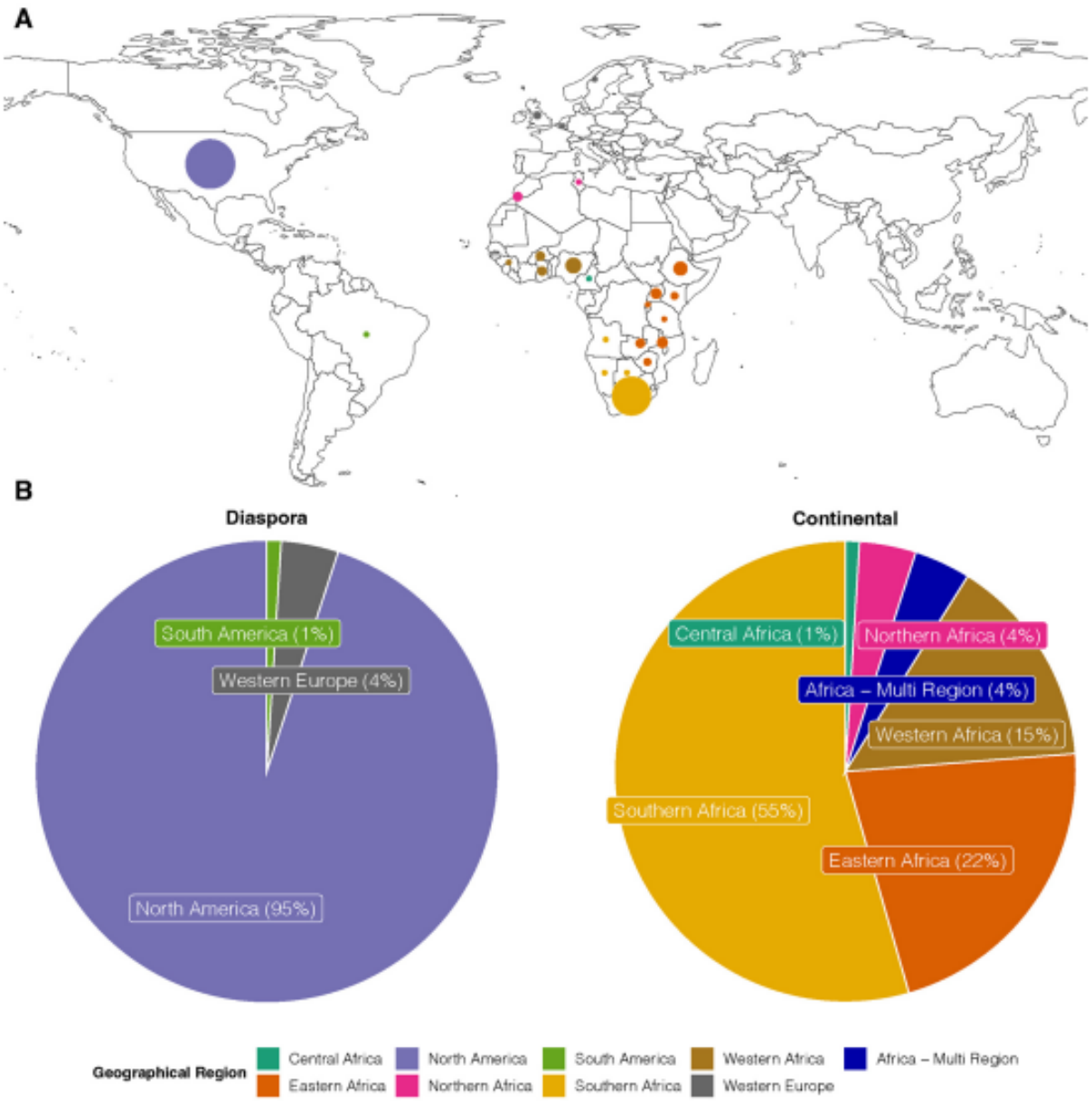


Figure 3 (A) Global location of participants from 213 multimorbidity studies in populations of African-ancestry. Studies including representation from both continental and diaspora populations (n=19) are not shown. (B) Distribution of papers in diaspora (n=100) and continental (n=113) African-ancestry populations by geographical region (n=213). Regional distribution is based on the five regions of Africa described by the United Nations Statistics Division (online supplemental table S2). 'Africa' includes countries from multiple regions.

MM between what they identified as race/ethnic groups. Some studies explored the impact of MM on mortality, morbidity, hospital resource utilisation, polypharmacy, adverse drug reactions and care integration.

Disease and MM ascertainment

We considered a study to have a clear definition of MM when it was stated in the objectives, methodology or discussion of the study. Our literature search found that most studies lacked clear MM definitions and showed considerable variability in disease inclusion. The terms MM and comorbidity were often used interchangeably, particularly in earlier studies.

Studies commonly employed disease counts to identify MM, but diseases included varied widely from 2 to over 200. Many studies used an author-developed disease list or, in some cases, a tool used to determine MM, such as

1

the Charlson Comorbidity Index.[30 31] These lists were available in 105 continental studies and 93 diaspora studies. In studies employing electronic health records or chart review, patient disease was identified mostly by ICD9 or ICD10 codes, based on clinical diagnosis. In smaller cohort studies, primarily in Africa, disease was determined through self-reporting or proxy variables such as medication use, or when clinical biomarkers were outside normal ranges. Although more cost-effective and feasible in resource-constrained research environments, the latter, particularly self-reporting, can introduce biases like recall errors and misclassification, potentially skewing the accuracy and prevalence of diagnoses.

Figure 4 illustrates the frequency of MM combinations in continental and diaspora studies. The most studied cluster in continental studies was cardiovascular-infectious-metabolic diseases (n=16). In the diaspora, psychiatric conditions (represented by more than one condition but grouped for simplicity) were most frequent (n=11), followed by cardiovascular-metabolic (n=10) and

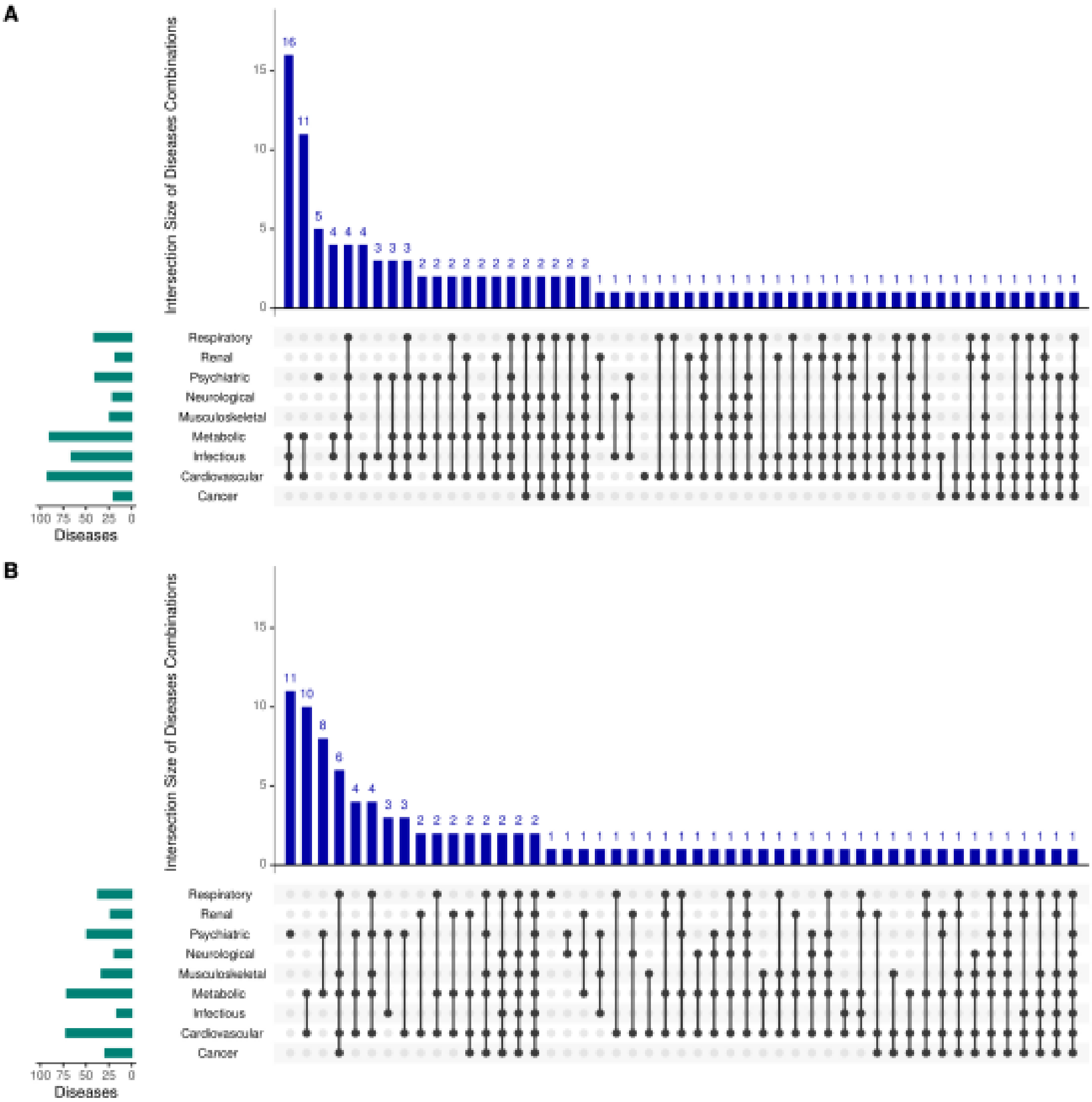


Figure 4 Distribution of disease cluster combinations (cancer, cardiovascular, infectious, metabolic, musculoskeletal, neurological, psychiatric, renal and respiratory) across 213 studies, stratified by (A) continental (113 studies) and (B) diaspora African-ancestry populations (100 studies). upSet plots generated using RStudio.

metabolic-psychiatric (n=8) clusters. Cardiovascular (eg, hypertension) and metabolic diseases (eg, diabetes) were the most widely included in both groups. In continental populations, 80% of studies assessed cardiovascular (n=92) or metabolic (n=90) diseases, while in the diaspora, at least 70% focused on these conditions (cardiovascular (73%, n=72); or metabolic (72%, n=71)) (see figure 4).

In contrast, the diaspora and continental studies differ greatly in their assessment of infectious diseases (predominantly HIV, Tuberculosis (TB) and malaria). Infectious diseases were more prevalent in continental studies (58%) compared with the diaspora (16%), whereas renal diseases, cancer, neurological and musculoskeletal disorders were under-represented in both groups.

Sociodemographic characteristics such as age, sex, household income and SES were examined in relation to MM prevalence and cumulative disease counts in studies involving continental and diaspora populations. Older individuals, particularly those aged 60 years and above, were found to have a higher risk of MM in both continental and diaspora populations. However, among continental populations, MM was also observed in relatively younger adults, approximately 45 years and below.[32–35]

Differences in MM by sex were identified, with women often showing a higher prevalence of MM, particularly in cardiovascular-metabolic clusters.[35–41] Low SES and income were generally associated with MM, but there were some population-specific differences. In the diaspora, lower SES was linked to higher MM prevalence and increased disease count,[42–44] while in continental studies, the impact of SES was variable. Lower SES was associated with higher MM[45–50]; but differences were observed. For example, studies in South Africa,[51 52] Burkina Faso[41] and Ghana[53] found that higher income was associated with concordant cardiometabolic MM, while lower incomes were associated with infectious disease-related MM.[46 48] In both populations, some studies revealed no effect of SES.[32 54–56] Ancestry effects in diaspora studies varied depending on recent migrations to Europe or AA in North America. In the USA, AA ethnicity consistently posed a risk factor for MM across most MM clusters.[19 57–61] One study assessing MM risk among men in a racially balanced and economically homogenous cohort in the USA found no difference in MM risk between AA and whites when in the same social environment.[62] Conversely, studies on recent migrations indicated lower MM prevalence among African-ancestry migrant populations.[63]

Studies also identified significant geographic variations in MM prevalence and patterns.[40 64] Using a population-based survey in South Africa, Akindele *et al*[40] observed distinct provincial differences, with lower rates of multiple chronic lifestyle diseases among residents of Limpopo, and Wong *et al*[64] demonstrated the intricate geospatial complexities of MM disease clusters within a rural community in Kwa-Zulu Natal.

The impact of behavioural risk factors, such as smoking, alcohol consumption and physical activity, on MM and treatment outcomes was assessed in some studies. When assessed, physical activity was noted as being largely negatively associated with MM, particularly within cardiometabolic MM, due to its impact on cardiometabolic risk factors, such as obesity and waist circumference.[65] Alcohol and substance use were largely controlled for or considered as disorders themselves in most studies and, therefore, their effect on MM was not well described.[32 46] However, assessment of studies examining alcohol use and smoking behaviour showed conflicting effects. For example, smoking and alcohol consumption was positively associated with multimorbid patients suffering from post-traumatic stress disorder in South Africa[66] but negatively associated with multimorbid women newly diagnosed with breast cancer; however, these bivariate associations did not persist in multinomial models.[67]

Thematic analysis

An in-depth review of included papers was used to extract recurring themes from the data. These themes were further reviewed and refined in consultation with the authorship team. While the quality of individual studies was not assessed in this review, caution was exercised in comparing the studies due to their diverse approaches in exploring MM in different settings. The themes below should not be viewed in isolation from one another, as they overlap somewhat and need to be integrated when considering solutions. First, we describe the limitations in terms of non-standardised terminology and data collection, a feature common across many disciplines, but we then highlight the heterogeneity among the combinations of chronic conditions that have been included in MM, but that in both African and diaspora studies the cardiovascular and metabolic clusters are most commonly studied.

Heterogeneity and inconsistency in terminology and methodological approaches

Similar to previous systematic reviews,[16 33 68] our findings highlight substantial heterogeneity in the terminology and methodological approaches used in the identified studies. This variability is not surprising given the relatively recent emergence of MM as a concept in healthcare research and the ongoing debates surrounding its definition and measurement.[11] Most studies used cross-sectional designs to estimate MM prevalence, while fewer employed longitudinal designs, possibly due to convenient and cost-effective participant recruitment. Continental African studies, with smaller sample sizes and reliance on self-reported data, may lead to biased prevalence and outcome estimates.

MM can be measured by counting the number of morbidities or using an MM index that considers both the number and severity of diseases, but the latter may be limited by data requirements.[68] Variability in MM measurement is influenced by national health priorities and data availability. Furthermore, debates around disease clusters and groupings persist, with some advocating conditions

on the same spectrum or targeting the same organ system collectively rather than separately,[69] particularly for well-managed conditions like hypertension.[70]

Dominance of cardiovascular-metabolic disease clusters

Distinct regional variation in disease clusters was observed. Continental populations exhibited a co-occurrence of cardiometabolic clusters and chronic infectious diseases, while the diaspora showed a dominance of psychiatric conditions, often combined with cardiometabolic diseases. This aligns with previous research highlighting depression and cardiometabolic disorders as commonly observed MM clusters.[16 71–73] Nonetheless, it is important to acknowledge that the available data suggest that the prevalence of condition clusters is strongly influenced by the context and population in which the research is conducted. Thus, when electronic health records were available, or when utilising large international cohorts, the number of diseases assessed to determine MM were much greater than smaller hospital or community-based cross-sectional cohorts.

In LMICs, such as those in Africa, chronic infectious illnesses like HIV/AIDS and TB are more prevalent. This is partially attributable to the longer life expectancies made possible by antiretroviral therapy, and the subsequent increase in susceptibility to NCDs at older ages.[21] Additionally, the cardiometabolic-infectious cluster may be exacerbated by the adverse effects of antiretroviral therapies on cardiovascular health.[74–76]

The dominance of specific clusters limits the understanding of the prevalence of concordant and discordant conditions and their associations with other diseases and health outcomes. However, one continental study found similar prevalence rates for concordant and discordant MM (approximately 12% each),[41] suggesting their comparable significance. Another study examining the association between physical MM and depression in six LMICs, including Ghana and South Africa, demonstrated that physical MM significantly increases the odds of depression, with a dose-dependent relationship as the number of diseases increases.[77]

Diverse effects of age, sex and SES across populations

Factors associated with MM risk in African-ancestry populations, such as age, sex, education, SES and comorbidities, were examined. Similar to studies conducted in high-income countries, older age, female sex and lower SES were identified as risk factors for MM. However, nuances were observes, highlighting the importance of context-specific research.

Older age consistently showed a higher prevalence of MM in both populations; however, in continental populations, infectious disease burden contributed to younger adults being diagnosed with MM.[32 35 64] Women of all ages were at increased risk of MM, potentially influenced by biological differences or social factors such as living and working environments, care-seeking behaviour and income inequalities.[21] The relationship between sex and MM varies depending on the specific diseases considered; as suggested in previous research. We noted men to be more at risk when evaluating psychiatric disorders such as substance abuse disorders. Low SES, typically measured by education, income, occupation or a composite of these, was identified as a risk factor for MM, with some disease-group intricacies. Also, SES may influence other factors such as sedentary lifestyles, health awareness and access to healthcare.[48 54] Furthermore, disentangling social context and ethnicity in MM risk among diaspora populations remains challenging.

Diaspora studies often include race as a covariate to compare MM prevalence among different racial groups. However, variations in study methodology and terminology make it challenging to synthesise and interpret interethnic differences in MM risk. It remains unclear whether genuine disparities exist beyond differences in diagnosis or survival rates. USA-based diaspora studies suggested significant differences in MM prevalence and health outcomes across ethnicities, with AAs often having a higher prevalence. For example, Mochari-Greenberger and Mosca,[78] reported higher readmission rates among hypertensive Hispanic and AA patients with comorbid diabetes than Whites. Clements *et al*[79] noted that AA had increased odds of most multiple chronic condition combinations compared with Whites as well as and an increased risk of mortality even when adjusting for disease combinations. However, psychiatric disorders, such as depression and anxiety, were noted as being more prevalent in European-ancestry populations,[19 80] likely due to under-reporting or underdiagnosing, particularly among men who may be more reticent to seek help from mental health services, compared with non-AA. A recent migration study in the Netherlands revealed that ethnic inequalities in MM persisted when adjusting for the lower SES of ethnic minority groups.[81]

Integrating care to address complex health outcomes associated with MM

The literature consistently showed that MM is associated with higher levels of health resource utilisation (eg, medications, primary care, emergency services) and adverse health outcomes (eg, mortality, hospitalisations and reduced quality of life), irrespective of disease clusters. Several studies, including those by Aparasu *et al*,[82] Bhagavathula *et al*[50] and Simakoloyi *et al*,[83] support these findings. Although patients with MM require frequent healthcare access, social determinants such as SES, geographic location and healthcare system factors can hinder access. Perry *et al*[84] described these factors as significant barriers to quality patient-centred care for older multimorbid AA men, despite their higher MM burden.

Multiple studies have highlighted the limitations of traditional healthcare systems that focus on single diseases and have recommended a shift towards an integrated care model (ICM) where practitioners from various specialties collaborate in managing patients with multiple conditions.[64 85–87] Wong *et al*[64] emphasised the

need for integrated management after observing high rates of uncontrolled hypertension (57.5%), diabetes (70.4%) and untreated TB in a rural South African population despite successful HIV management—78% of their cohort had undetectable HIV-1 RNA viral loads. However, implementing ICM poses challenges due to the increased complexity and time required for treating unrelated conditions. Limited evidence exists on the effective management of multiple conditions, particularly in resource-constrained settings like Africa. Nonetheless, Khabala *et al*[87] noted that ICM could improve the flexibility of care delivery and reduce clinician workload among patients with mixed chronic conditions in Kenya, and, in South Africa, ICM reduced HIV stigma in health facilities due to the non-segregation of patients,[88] and with consolidated guidelines, patient education materials and information systems, this could improve treatment and care for MM patients.[89 90] Implementing ICM in South Africa is estimated to have minimal additional costs, approximately USD1.06 (SD:0.33) per patient visit, over the current mean cost of USD4.94 (SD:0.70).[91]

Insufficient data exist to quantify the burden of NCDs in health facilities in sub-Saharan Africa (SSA) and their readiness to manage the growing NCD epidemic. Some studies, mainly in South Africa, have investigated the possibility of expanding HIV clinics to address cardiovascular diseases like hypertension and diabetes.[92] Despite the potential of integrated care to improve patient care and decrease individual and healthcare system costs, a significant gap in clinic and hospital staff training remains a barrier to effective implementation.[93–96]

Challenges in statistical analysis approaches

Studies primarily employed univariate descriptive analysis, while logistic regression was the preferred method for assessing the relationship between MM and exposure variables or MM's impact on an outcome. While simple modelling techniques offer interpretability advantages,[97] there is room for additional analyses through advanced machine learning (ML) and robust computational approaches. These approaches could reduce dependence on hand crafting by domain experts and scalability for the growing size of data and the number of variables.[97 98] Few studies reported model goodness-of-fit assessment or variable selection approaches to refine multivariate models. Using the best subset of variables that contribute to the outcome of interest can lead to more useful and informative models.

Despite the advantages of imputing missing data and available software for implementation, few studies addressed missingness. There was limited assessment of the spatial and temporal distribution of coexisting diseases, which could provide valuable information on MM hotspots and associated factors. Such studies can also offer insights into the future MM population profile and the likely change in MM prevalence.

DISCUSSION

This scoping review identified 232 relevant articles and used thematic analysis to address the aim of the review; to identify, evaluate and summarise the available published evidence on MM in African-ancestry populations and identify research gaps and future recommendations. The themes are often interdependent and should not be viewed in isolation.

Overview of included articles

The review revealed significant heterogeneity in MM terminology, methodology and disease measurement. While cardiovascular-metabolic disease clusters were dominant, regional variations were observed. Risk factors for MM in African-ancestry populations were similar to those found in European populations and high-income countries, including older age, female sex and lower SES,[99 100] with important exceptions. Therefore, the nuanced differences, highlighting the need for context-specific research and the identification of major gaps, need to be considered in the interpretation of MM in these populations.

Research gaps and recommendations for future research

Based on the included literature, we identified six research gaps and below we describe how they could be mitigated. These interconnected areas describe crucial interventions that would advance evidence-based research on MM and thereby potentially enhance patient care and health outcomes.

Paucity of data globally (sample size and complexity of data)

Limited data availability and a lack of large-scale epidemiological studies on MM, particularly in African-ancestry populations, hinder our understanding of MM prevalence and patterns in diverse populations. Future studies should increase sample size and undertake longitudinal population-based studies to better understand MM's longitudinal trajectory. Additionally, including genetic data in these studies can offer valuable insights into MM causation and the interplay between genetic and environmental factors. Gathering comprehensive data will improve our understanding of MM and facilitate more targeted interventions.

Lack of representation (many African countries and ethnic groups have no published data)

African-ancestry populations are under-represented in the MM literature, both within and outside of Africa. Furthermore, other socioeconomic and health factors impact MM prevalence in groups and marginalised communities. More research is needed in African-ancestry individuals in Central and South America as well as recent and growing migrant populations in Europe and elsewhere, who may have unique MM experiences. The Research on Obesity and Diabetes among African Migrants (RODAM) study highlights the influence of migration on MM risk factors in people originating from Ghana and emphasises the importance of understanding

1

the sociodemographic, cultural, and healthcare contexts of African-ancestry migrants for a comprehensive understanding of MM.[101]

Underrepresented disease areas (health priority mismatch)

On assessing the Global Burden of Disease (GBD) morbidity estimates both globally and within SSA, a mismatch between health priorities and the diseases investigated is evident (GBD morbidity estimates are found in online supplemental figure S1). Cardiometabolic and infectious diseases are the leading disease burdens in SAA,[25] which aligns with the focus of continental studies. This alignment is also true for cardiovascular diseases globally. However, mental health conditions among adults younger than 50 years and cancer among individuals older than 50 also account for a significant health burden in SSA, yet these clusters, particularly cancer, appear to be rarely assessed in continental MM studies. Additionally, as life expectancy rises in Africa, musculoskeletal and neurological diseases associated with the elderly will become more prevalent and should be investigated.

Moreover, there is a paucity of evidence on how clusters of conditions develop and change over time, making it challenging to predict how disease burdens might change within an individual's life, and to identify the best interventions and care approaches. Key factors to consider when assessing long-term health in SSA include not only infectious and NCD but also the combined effects of childhood malnutrition, and high rates of maternal HIV infection.[64]

Lack of standardised study designs and data collection protocols

The absence of standardised definitions and measurements and uncertainties about the applicability and validity of MM assessment tools in African-ancestry populations must be addressed. Moving forward, it is crucial to establish standardised criteria and methods to ensure consistent and accurate identification of MM across populations and regions while recognising the importance of context-specific assessments. This should be applied both to study communities that remain unexplored and to all the different disease clusters.

Need to leverage additional data analysis approaches

In addition to traditional statistical approaches to epidemiological research, recent studies beyond African populations have utilised emerging ML techniques to understand the complex nature of MM.[102 103] These ML techniques can be categorised into pairwise, probabilistic, factorisation and temporal methods.[102] Pairwise methods are relatively simple but limited in capturing complex interactions and multiple co-occurring diseases common in ageing populations. Probabilistic methods, such as Hidden Markov Models, provide an overall view of disease relationships. Factorisation methods, like non-negative matrix factorisation, offer interpretability but may struggle with encoding suppressiveness among diseases. Advanced ML techniques, including deep learning, show promise in unravelling the complexity of MM.[102] These techniques can infer co-occurring diseases, assess data trustworthiness, identify patterns, analyse longitudinal aspects, integrate multiple data sources and provide interpretable insights.[104–106] Future studies should also consider employing ML and prediction modelling approaches, building on successful collaborations between domain experts and ML researchers.

Invest in translational research

To comprehensively understand the interplay between MM, the healthcare system and the cultural/religious context, diverse perspectives must be considered. The gaps described above, therefore, need to be addressed in the specific milieu of the study communities and adapted to suit the context, while striving to generate interoperable data fit for use in comparative studies. Despite it being crucial to explore patients' experiences, caregivers' challenges and clinicians' perspectives regarding MM in African-ancestry populations, there is currently limited research in these domains. Additionally, cultural and religious beliefs will likely influence perceptions and behaviours related to MM. Assessing the financial and social costs of MM is also necessary. This knowledge can inform policy shifts, treatment approaches and clinical guidelines, leading to improved diagnosis, risk prediction, and treatment strategies.

Strengths and limitations of this scoping review

This scoping review is the first to summarise the literature on MM prevalence and patterns in African-ancestry populations worldwide. Dividing studies into continental and diaspora African-ancestry populations revealed important differences. By involving a multidisciplinary team of epidemiologists, demographers, clinicians, geneticists and data scientists, with representation from multiple African countries, we optimised the inclusion of relevant studies and ensured cultural sensitivity during the writing process. However, limitations such as limiting our search strategy to title and abstract only, and assessing open access texts or those available through institutional logins may have resulted in some literature being excluded. The literature did not lend itself to a systematic review or meta-analysis and, therefore, we did not perform quality assessment on the publications using standardised tools. Nevertheless, our results emphasise the necessity for additional research on MM in African-ancestry populations and the development of context-specific guidelines for prevention and management across different geographic locations.

CONCLUSION

In reviewing the literature on MM in populations with African-ancestry, we observed distinct disease clusters and varying effects of social determinants of health on MM. We provide strong evidence that Black Americans (also referred to as AAs), the most widely studied

1

African-ancestry group, are not a good proxy for all Africans. Furthermore, continental African regions and populations are highly diverse in their health profiles, exposures and behaviours. To address MM effectively in African-ancestry populations, we identified six research gaps: limited data availability, inadequate representation, under-represented disease clusters, absence of standardised study designs and data collection protocols, the need for innovative data analysis approaches and the need for more translational research. Additionally, considering the complexities observed among populations, it is crucial to account for location, resources and environmental context when developing interventions to mitigate the burden of MM, especially in African-ancestry populations.

Author affiliations
[1]Division of Human Genetics, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[3]African Population and Health Research Center (APHRC), APHRC Campus, Nairobi, Kenya
[4]IBM Research Africa, Nairobi, Kenya
[5]School of Electrical and Information Engineering, Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, South Africa
[6]Department of Women's and Children's Health, Karolinska Institute, Stockholm, Sweden

Twitter Michelle Kamp @michelleukamp and Girmaw Abebe Tadesse @girmaw_abebe

Acknowledgements We would like to thank the broader MADIVA research hub for their input into study design and interpretation.

Contributors MR, GA, SH and MK conceived the study and participated in the design of the study. MK, OA, IK, DMN, GAT, KA and SI undertook the literature review process. MK, OA and PTM undertook data analysis and visualisation. MK drafted the manuscript with input from OA. MR is acting as guarantor. All authors read and approved the final manuscript.

Funding Research reported in this publication was supported by the Fogarty International Center of the National Institutes of Health under Award Number U54TW012077. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This is a review of already published material; therefore, no ethics approval needed.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Data sharing is not applicable as no datasets were generated and/or analysed for this study. The BibTex file containing the articles included in this review is available on GitHub (https://github.com/MADIVA-DSI/MM-in-African-ancestry-pops.git).

ORCID iDs
Michelle Kamp http://orcid.org/0000-0002-4334-6883
Okechinyere Achilonu http://orcid.org/0000-0003-2911-0011
Isaac Kisiangani http://orcid.org/0000-0002-1107-4334
Daniel Maina Nderitu http://orcid.org/0000-0001-9800-2431
Phelelani Thokozani Mpangase http://orcid.org/0000-0001-8280-8940
Girmaw Abebe Tadesse http://orcid.org/0000-0002-2648-9102
Samuel Iddi http://orcid.org/0000-0002-2366-2774
Scott Hazelhurst http://orcid.org/0000-0002-0581-149X
Gershim Asiki http://orcid.org/0000-0002-9966-1153
Michèle Ramsay http://orcid.org/0000-0002-4156-4801

REFERENCES

1 Navickas R, Petric V-K, Feigl AB, *et al*. Multimorbidity: what do we know? What should we do. *J Comorb* 2016;6:4–11.
2 Nunes BP, Flores TR, Mielke GI, *et al*. Multimorbidity and mortality in older adults: a systematic review and meta-analysis. *Arch Gerontol Geriatr* 2016;67:130–8.
3 Fortin M, Lapointe L, Hudon C, *et al*. Multimorbidity and quality of life in primary care: a systematic review. *Health Qual Life Outcomes* 2004;2:51.
4 Cheng GJ, Wagner AL, O'Shea BQ, *et al*. Multimorbidity and mental health trajectories among middle-aged and older U.S. adults during the COVID-19 pandemic: longitudinal findings from the COVID-19 coping study. *Innov Aging* 2022;6:igac047.
5 Fortin M, Bravo G, Hudon C, *et al*. Psychological distress and multimorbidity in primary care. *Ann Fam Med* 2006;4:417–22.
6 Calderón-Larrañaga A, Poblador-Plou B, González-Rubio F, *et al*. Multimorbidity, polypharmacy, referrals, and adverse drug events: are we doing things well. *Br J Gen Pract* 2012;62:e821–6.
7 Wolff JL, Starfield B, Anderson G. Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *Arch Intern Med* 2002;162:2269–76.
8 Salisbury C, Johnson L, Purdy S, *et al*. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *Br J Gen Pract* 2011;61:e12–21.
9 Boyd CM, Fortin M. Future of multimorbidity research: how should understanding of multimorbidity inform health system design *Public Health Rev* 2010;32:451–74.
10 World Health Organization. *Multimorbidity: Technical Series on Safer Primary Care*. Geneva, Switzerland, 2016.
11 Ho ISS, Azcoaga-Lorenzo A, Akbari A, *et al*. Measuring multimorbidity in research: Delphi consensus study. *BMJ Med* 2022;1:e000247.
12 Skou ST, Mair FS, Fortin M, *et al*. Multimorbidity. *Nat Rev Dis Primers* 2022;8:48.
13 Hajat C, Stein E. The global burden of multiple chronic conditions: a narrative review. *Preventive Medicine Reports* 2018;12:284–93.
14 Marengoni A, Angleman S, Melis R, *et al*. Aging with multimorbidity: a systematic review of the literature. *Ageing Res Rev* 2011;10:430–9.
15 Barnett K, Mercer SW, Norbury M, *et al*. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012;380:37–43.
16 Roomaney RA, Wyk BV, Wyk V-V. Decolonising multimorbidity? Research gaps in low and middle-income countries. *Pan Afr Med J* 2022;41:140.
17 Nguyen H, Manolova G, Daskalopoulou C, *et al*. Prevalence of multimorbidity in community settings: a systematic review and meta-analysis of observational studies. *J Comorb* 2019;9:2235042X1987093.

18 Xu X, Mishra GD, Jones M. Mapping the global research landscape and knowledge gaps on multimorbidity: a bibliometric study. *J Glob Health* 2017;7:10414.
19 Alshakhs M, Jackson B, Ikponmwosa D, *et al*. Multimorbidity patterns across race/ethnicity as stratified by age and obesity. *Sci Rep* 2022;12:1–16.
20 Kalgotra P, Sharda R, Croff JM. Examining multimorbidity differences across racial groups: a network analysis of electronic medical records. *Sci Rep* 2020;10:13538.
21 Academy of Medical Sciences. *Multimorbidity: a priority for global health research*. London, UK, 2018.
22 Johnston MC, Crilly M, Black C, *et al*. Defining and measuring multimorbidity: a systematic review of systematic reviews. *Eur J Public Health* 2019;29:182–9.
23 Peters MDJ, Godfrey CM, Khalil H, *et al*. Guidance for conducting systematic scoping reviews. *Int J Evid Based Healthc* 2015;13:141–6.
24 Tricco AC, Lillie E, Zarin W, *et al*. PRISMA extension for scoping reviews (PRISMA-SCR): checklist and explanation. *Ann Intern Med* 2018;169:467–73.
25 Ouzzani M, Hammady H, Fedorowicz Z, *et al*. Rayyan-a web and mobile app for systematic reviews. *Syst Rev* 2016;5:210.
26 Institute for Health Metrics and Evaluation (IHME). GBD compare data visualization. 2020. Available: https://www.healthdata.org/data-visualization/gbd-compare [Accessed 04 May 2023].
27 Microsoft Corporation. *Microsoft Excel*. 2018.
28 R Core Team. R: A language and environment for statistical computing. 2021. Available: https://www.r-project.org/
29 Lex A, Gehlenborg N, Strobelt H, *et al*. Upset: visualization of intersecting SETS. *IEEE Trans Vis Comput Graph* 2014;20:1983–92.
30 Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
31 Sundararajan V, Henderson T, Perry C, *et al*. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004;57:1288–94.
32 Chang AY, Gómez-Olivé FX, Payne C, *et al*. Chronic multimorbidity among older adults in rural South Africa. *BMJ Glob Health* 2019;4:e001386.
33 Sharman M, Bachmann M. Prevalence and health effects of communicable and non-communicable disease comorbidity in rural Kwazulu-natal, South Africa. *Trop Med Int Health* 2019;24:1198–207.
34 Hopkins KL, Hlongwane KE, Otwombe K, *et al*. The substantial burden of non-communicable diseases and HIV-comorbidity amongst adults: screening results from an integrated HIV testing services clinic for adults in Soweto, South Africa. *EClinicalMedicine* 2021;38:101015.
35 Wade AN, Payne CF, Berkman L, *et al*. Multimorbidity and mortality in an older, rural black South African population cohort with high prevalence of HIV findings from the HAALSI study. *BMJ Open* 2021;11:e047777.
36 Boake M, Mash R. Diabetes in the Western Cape, South Africa: a secondary analysis of the diabetes cascade database 2015 - 2020. *Prim Care Diabetes* 2022;16:555–61.
37 Ntiyani N, Letamo G, Keetile M. Prevalence of and factors associated with hypertension, diabetes, stroke and heart attack multimorbidity in Botswana: evidence from STEPS 2014 survey. *PLoS One* 2022;17:e0265722.
38 Gaziano TA, Pandya A, Steyn K, *et al*. Comparative assessment of absolute cardiovascular disease risk characterization from non-laboratory-based risk assessment in South African populations. *BMC Med* 2013;11.
39 Mohamed SF, Haregu TN, Uthman OA, *et al*. Multimorbidity from chronic conditions among adults in urban slums: the AWI-Gen Nairobi site study findings. *Glob Heart* 2021;16.
40 Akindele MO, Useh U. Multimorbidity of chronic diseases of lifestyle among South African adults. *Pan Afr Med J* 2021;38:332.
41 Odland ML, Payne C, Witham MD, *et al*. Epidemiology of multimorbidity in conditions of extreme poverty: a population-based study of older adults in rural Burkina Faso. *BMJ Glob Health* 2020;5:e002096.
42 Cook JA, Burke-Miller JK, Steigman PJ, *et al*. Prevalence, comorbidity, and correlates of psychiatric and substance use disorders and associations with HIV risk behaviors in a multisite cohort of women living with HIV. *AIDS Behav* 2018;22:3141–54.
43 Vennu V, Abdulrahman TA, Alenazi AM, *et al*. Associations between social determinants and the presence of chronic diseases: data from the osteoarthritis initiative. *BMC Public Health* 2020;20:1323.
44 Diaz E, Poblador-Pou B, Gimeno-Feliu L-A, *et al*. Multimorbidity and its patterns according to immigrant origin. A nationwide register-based study in Norway. *PLoS One* 2015;10:e0145233.
45 Mthoko NFN, Pazvakawambwa L, Leonhardt M, *et al*. Physical comorbidity among patients attending mental health services at Windhoek central hospital, Namibia. *Pan Afr Med J* 2022;41:270.
46 Peltzer K. Tuberculosis non-communicable disease comorbidity and multimorbidity in public primary care patients in South Africa. *Afr J Prim Health Care Fam Med* 2018;10.
47 Weimann A, Dai D, Oni T. A cross-sectional and spatial analysis of the prevalence of multimorbidity and its association with socioeconomic disadvantage in South Africa: a comparison between 2008 and 2012. *Soc Sci Med* 2016;163:144–56.
48 Ataguba JEO. Inequalities in multimorbidity in South Africa. *Int J Equity Health* 2013;12:64.
49 Chikezie UE, Otakpor AN, Kuteyi OB, *et al*. Depression among people living with human immunodeficiency virus infection/acquired immunodeficiency syndrome in Benin city, Nigeria: a comparative study. *Niger J Clin Pract* 2013;16:238–42.
50 Bhagavathula AS, Seid MA, Adane A, *et al*. Prevalence and determinants of multimorbidity, polypharmacy, and potentially inappropriate medication use in the older outpatients: findings from Euroageism H2020 ESR7 project in Ethiopia. *Pharmaceuticals (Basel)* 2021;14:844.
51 Mutyambizi C, Chola L, Groot W, *et al*. The extent and determinants of diabetes and cardiovascular disease comorbidity in South Africa - results from the South African national health and nutrition examination survey (SANHANES-1). *BMC Public Health* 2017;17:745.
52 Alaba O, Chola L. The social determinants of multimorbidity in South Africa. *Int J Equity Health* 2013;12:63.
53 Kunna R, San Sebastian M, Stewart Williams J. Measurement and decomposition of socioeconomic inequality in single and Multimorbidity in older adults in China and Ghana: results from the WHO study on global ageing and adult health (SAGE). *Int J Equity Health* 2017;16:79.
54 Nimako BA, Baiden F, Sackey SO, *et al*. Multimorbidity of chronic diseases among adult patients presenting to an inner-city clinic in Ghana. *Global Health* 2013;9:61.
55 Zelnick LR, Katz R, Young BA, *et al*. Echocardiographic measures and estimated GFR decline among African Americans: the Jackson heart study. *Am J Kidney Dis* 2017;70:199–206.
56 Gabriel A, Zare H, Jones W, *et al*. Evaluating depressive symptoms among low-socioeconomic-status African American women aged 40 to 75 years with uncontrolled hypertension: a secondary analysis of a randomized clinical trial. *JAMA Psychiatry* 2021;78:426–32.
57 Dibato JE, Montvida O, Zaccardi F, *et al*. Association of cardiometabolic multimorbidity and depression with cardiovascular events in early-onset adult type 2 diabetes: a multiethnic study in the U.S. *Diabetes Care* 2021;44:231–9.
58 Erving CL, Frazier C. The association between multiple chronic conditions and depressive symptoms: Intersectional distinctions by race. *J Health Soc Behav* 2021;62:599–617.
59 Roth GA, Abate D, Abate KH, *et al*. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;392:1736–88.
60 Yang J, Ju X, Liu F, *et al*. Prediction for the risk of multiple chronic conditions among working population in the United States with machine learning models. *IEEE Open J Eng Med Biol* 2021;2:291–8.
61 Heckman BD, Merrill JC, Anderson T. Race, psychiatric comorbidity, and headache characteristics in patients in headache subspecialty treatment clinics. *Ethn Health* 2013;18:34–52.
62 Thorpe RJ Jr, Bell CN, Kennedy-Hendricks A, *et al*. Disentangling race and social context in understanding disparities in chronic conditions among men. *J Urban Health* 2015;92:83–92.
63 Marzà-Florensa A, Boateng D, Agyemang C, *et al*. Multimorbidity among migrant and non-migrant Ghanaians: the RODAM study. *Int J Public Health* 2021;66:1604056.
64 Wong EB, Olivier S, Gunda R, *et al*. Convergence of infectious and non-communicable disease epidemics in rural South Africa: a cross-sectional, population-based multimorbidity study. *Lancet Glob Health* 2021;9:e967–76.
65 Romano E, Ma R, Vancampfort D, *et al*. Multimorbidity and obesity in older adults from six Low- and middle-income countries. *Preventive Medicine* 2021;153:106816.
66 Pengpid S, Peltzer K. Mental morbidity and its associations with socio-behavioural factors and chronic conditions in rural

1

middle- and older-aged adults in South Africa. *J Psychol Afr* 2020;30:257–63.
67 Ayeni OA, Norris SA, Joffe M, *et al*. The multimorbidity profile of South African women newly diagnosed with breast cancer. *Int J Cancer* 2020;147:361–74.
68 Stirland LE, González-Saavedra L, Mullin DS, *et al*. Measuring multimorbidity beyond counting diseases: systematic review of community and population studies and guide to index choice. *BMJ* 2020;368:m160.
69 Fortin M, Stewart M, Poitras M-E, *et al*. A systematic review of prevalence studies on multimorbidity: toward a more uniform methodology. *Ann Fam Med* 2012;10:142–51.
70 Lee ES, Koh HL, Ho EQ-Y, *et al*. Systematic review on the instruments used for measuring the association of the level of multimorbidity and clinically important outcomes. *BMJ Open* 2021;11:e041219.
71 Sinnige J, Braspenning J, Schellevis F, *et al*. The prevalence of disease clusters in older adults with multiple chronic diseases--a systematic literature review. *PLoS One* 2013;8:e79641.
72 Violan C, Foguet-Boreu Q, Flores-Mateo G, *et al*. Prevalence, determinants and patterns of multimorbidity in primary care: a systematic review of observational studies. *PLoS One* 2014;9:e102149.
73 Prados-Torres A, Calderón-Larrañaga A, Hancco-Saavedra J, *et al*. Multimorbidity patterns: a systematic review. *J Clin Epidemiol* 2014;67:254–66.
74 Gottlieb S. Antiretroviral therapy increases risk of heart attack. *BMJ* 2003;327:1186–0.
75 Jarwani B. Cardiovascular disease and antiretroviral therapy. *J Glob Infect Dis* 2019;11:91–2.
76 Dubé MP, Lipshultz SE, Fichtenbaum CJ, *et al*. Effects of HIV infection and antiretroviral therapy on the heart and vasculature. *Circulation* 2008;118:e36–40.
77 Smith L, Shin JI, Butler L, *et al*. Physical multimorbidity and depression: a mediation analysis of influential factors among 34,129 adults aged ≥50 years from low- and middle-income countries. *Depress Anxiety* 2022;39:376–86.
78 Mochari-Greenberger H, Mosca L. Differential outcomes by race and ethnicity in patients with coronary heart disease: a contemporary review. *Curr Cardiovasc Risk Rep* 2015;9:20.
79 Clements JM, West BT, Yaker Z, *et al*. Disparities in diabetes-related multiple chronic conditions and mortality: the influence of race. *Diabetes Res Clin Pract* 2020;159:107984.
80 Breslau J, Aguilar-Gaxiola S, Kendler KS, *et al*. Specifying race-ethnic differences in risk for psychiatric disorder in a USA national sample. *Psychol Med* 2006;36:57–68.
81 Verest WJGM, Galenkamp H, Spek B, *et al*. Do ethnic inequalities in multimorbidity reflect ethnic differences in socioeconomic status? the HELIUS study. *Eur J Public Health* 2019;29:687–93.
82 Aparasu RR, Mort JR, Brandt H. Polypharmacy trends in office visits by the elderly in the United States, 1990 and 2000. *Res Social Adm Pharm* 2005;1:446–59.
83 Simakoloyi N, Erasmus E, van Hoving DJ. The characteristics of geriatric patients managed within the resuscitation unit of a district-level emergency centre in Cape town. *Afr J Emerg Med* 2022;12:39–43.
84 Perry RG, Mitchell JA, Hawkins J. The role of age and multimorbidity in shaping older African American men's experiences with patient–provider communication. *Geriatrics (Basel)* 2018;3:74.
85 Myers B, Joska JA, Lund C, *et al*. Patient preferences for the integration of mental health counseling and chronic disease care in South Africa. *Patient Prefer Adherence* 2018;12:1797–803.
86 Segafredo G, Kapur A, Robbiati C, *et al*. Integrating TB and non-communicable diseases services: pilot experience of screening for diabetes and hypertension in patients with tuberculosis in Luanda, Angola. *PLoS One* 2019;14:e0218052.
87 Khabala KB, Edwards JK, Baruani B, *et al*. Medication adherence clubs: a potential solution to managing large numbers of stable patients with multiple chronic diseases in informal settlements. *Trop Med Int Health* 2015;20:1265–70.
88 Ameh S, D'Ambruoso L, Gómez-Olivé FX, *et al*. Paradox of HIV stigma in an integrated chronic disease care in rural South Africa: viewpoints of service users and providers. *PLoS One* 2020;15:e0236270.
89 Matima R, Murphy K, Levitt NS, *et al*. A qualitative study on the experiences and perspectives of public sector patients in Cape town in managing the workload of demands of HIV and type 2 diabetes multimorbidity. *PLoS One* 2018;13:e0194191.
90 Godongwana M, De Wet-Billings N, Milovanovic M. The comorbidity of HIV, hypertension and diabetes: a qualitative study exploring the challenges faced by healthcare providers and patients in selected urban and rural health facilities where the ICDM model is implemented in South Africa. *BMC Health Serv Res* 2021;21:647.
91 Lebina L, Kawonga M, Oni T, *et al*. The cost and cost implications of implementing the integrated chronic disease management model in South Africa. *PLoS One* 2020;15:e0235429.
92 Mayosi BM, Lawn JE, van Niekerk A, *et al*. Health in South Africa: changes and challenges since 2009. *The Lancet* 2012;380:2029–43.
93 Adjei KK, Kikuchi K, Owusu-Agyei S, *et al*. Women's overall satisfaction with health facility delivery services in Ghana: a mixed-methods study. *Trop Med Health* 2019;47:41.
94 Lebina L, Oni T, Alaba OA, *et al*. A mixed methods approach to exploring the moderating factors of implementation fidelity of the integrated chronic disease management model in South Africa. *BMC Health Serv Res* 2020;20:617.
95 Peck CC, Goulet J-P, Lobbezoo F, *et al*. Expanding the taxonomy of the diagnostic criteria for temporomandibular disorders. *J Oral Rehabil* 2014;41:2–23.
96 Peck R, Mghamba J, Vanobberghen F, *et al*. Preparedness of Tanzanian health facilities for outpatient primary care of hypertension and diabetes: a cross-sectional survey. *Lancet Global Health* 2014;2:e285–92.
97 Ley C, Martin RK, Pareek A, *et al*. Machine learning and conventional statistics: making sense of the differences. *Knee Surg Sports Traumatol Arthrosc* 2022;30:753–7.
98 Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* 2006;2:117693510600200.
99 Bezerra de Souza DL, Oliveras-Fabregas A, Espelt A, *et al*. Multimorbidity and its associated factors among adults aged 50 and over: a cross-sectional study in 17 European countries. *PLoS One* 2021;16:e0246623.
100 Ofori-Asenso R, Chin KL, Curtis AJ, *et al*. Recent patterns of multimorbidity among older adults in high-income countries. *Popul Health Manag* 2019;22:127–37.
101 Asogwa OA, Boateng D, Marzà-Florensa A, *et al*. Multimorbidity of non-communicable diseases in low-income and middle-income countries: a systematic review and meta-analysis. *BMJ Open* 2022;12:e049133.
102 Hassaine A, Salimi-Khorshidi G, Canoy D, *et al*. Untangling the complexity of multimorbidity with machine learning. *Mech Ageing Dev* 2020;190:111325.
103 Uddin S, Wang S, Lu H, *et al*. Comorbidity and multimorbidity prediction of major chronic diseases using machine learning and network analytics. *Expert Systems with Applications* 2022;205:117761.
104 Koedel U, Schuetze C, Fischer P, *et al*. Challenges in the evaluation of observational data trustworthiness from a data producers viewpoint (FAIR+). *Front Environ Sci* 2022;9:737.
105 Wang H, Tang J, Wu M, *et al*. Application of machine learning missing data imputation techniques in clinical decision making: taking the discharge assessment of patients with spontaneous supratentorial intracerebral hemorrhage as an example. *BMC Med Inform Decis Mak* 2022;22:1–14.
106 Polessa Paula D, Barbosa Aguiar O, Pruner Marques L, *et al*. Comparing machine learning algorithms for multimorbidity prediction: an example from the Elsa-Brasil study. *PLoS One* 2022;17:e0275619.

1