BMJ Open

Omics Approach for Personalised Prevention of Type 2 Diabetes Mellitus for African and European Populations (OPTIMA): a protocol paper

Julia H Goedecke ,[1,2,3] Ina Danquah,[1,4,5] Carol Akinyi Abidha,[4] Charles Agyemang,[6,7] Hannah Maike Albers,[4] Stephen Amoah,[4] Carl Brunius,[8] Elin Chorell,[1] Fatima Hoosen,[2,9] Melony Fortuin-de Smidt,[1] Åsa Hörnsten,[10] Therese Karlsson,[8,11] Lars Lindholm,[12] Amy E Mendham ,[9,13] Lisa K Micklesfield ,[3] Kaspar Walter Meili,[12] Stefania Noerman,[8] Julia Otten ,[1] Stefan Söderberg ,[1] Eva L van der Linden ,[6] Clemens Wittenbecher,[8,14] Rikard Landberg,[8,15] Tommy Olsson[1]

To cite: Goedecke JH, Danquah I, Abidha CA, *et al*. Omics Approach for Personalised Prevention of Type 2 Diabetes Mellitus for African and European Populations (OPTIMA): a protocol paper. *BMJ Open* 2025;**15**:e099108. doi:10.1136/bmjopen-2025-099108

► Prepublication history for this paper is available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2025-099108).

JHG and ID are joint first authors.
RL and TO are joint senior authors.

Received 10 January 2025
Accepted 04 April 2025

Check for updates

© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to
Professor Julia H Goedecke; julia.goedecke@mrc.ac.za

ABSTRACT

Introduction The prevalence of type 2 diabetes (T2D) within sub-Saharan Africa (SSA) is increasing. Despite the pathophysiology of T2D differing by ethnicity and sex, risk stratification and guidelines for the prevention of T2D are generic, relying on evidence from studies including predominantly Europeans. Accordingly, this study aims to develop ethnic-specific and sex-specific risk prediction models for the early detection of dysglycaemia (impaired glucose tolerance and T2D) to inform clinically feasible, culturally acceptable and cost-effective risk management and prevention strategies using dietary modification in SSA and European populations.

Methods and analysis This multinational collaboration will include the prospective cohort data from two African cohorts, the Middle-Aged Soweto Cohort from South Africa and the Research on Obesity and Diabetes among African Migrants Prospective cohort from Ghana and migrants living in Europe, and a Swedish cohort, the Pre-Swedish CArdioPulmonary bioImage Study. Targeted proteomics, as well as targeted and untargeted metabolomics, will be performed at baseline to discover known and novel ethnic-specific and sex-specific biomarkers that predict incident dysglycaemia in the different longitudinal cohorts. Dietary patterns that explain maximum variation in the biomarker profiles and that associate with dysglycaemia will be identified in the SSA and European cohorts and used to build the prototypes for dietary interventions to prevent T2D. A comparative cost-effectiveness analysis of the dietary interventions will be estimated in the different populations. Finally, the perceptions of at-risk participants and healthcare providers regarding ethnic-specific and sex-specific dietary recommendations for the prevention of T2D will be assessed using focus group discussions and in-depth interviews in South Africa, Ghana, Germany (Ghanaian migrants) and Sweden.

Ethics and dissemination Ethical clearance has been obtained from all participating sites. The study results will be disseminated at scientific conferences and in journal publications, and through community engagement events and diabetes organisations in the respective countries.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Multinational collaboration between researchers in Europe (Sweden, Germany and the Netherlands) and Africa (South Africa and Ghana) enables the sharing and harmonisation of prospective cohort data from a European (Pre-Swedish CArdioPulmonary bioImage Study) cohort from Sweden and two African cohorts: Middle-Aged Soweto Cohort from South Africa and Research on Obesity and Diabetes among African Migrants Prospective including Ghanaians living in rural and urban Ghana and migrant Ghanaians living in Europe.

⇒ The study provides the unique opportunity to identify and compare metabolite and/or protein biomarkers reflecting sex and ethnic differences in the aetiology of dysglycaemia that may be used for the early detection of dysglycaemia, even before symptoms develop.

⇒ The dietary patterns associated with the biomarkers and dysglycaemia will build the basis for prototypes of sex-specific and ethnic-specific dietary interventions to prevent T2D, for which the cost-effectiveness will be assessed.

⇒ The study will adopt a user-centred and person-centred approach to explore the perceptions of individuals at risk of T2D, as well as healthcare providers regarding T2D risk and their willingness to engage in ethnic-specific and sex-specific dietary preventative strategies in different settings in sub-Saharan Africa and Europe.

⇒ Oral glucose tolerance tests were not performed in all cohorts, and the blood collection tubes used to determine the metabolomics and proteomic profiles were not standardised across all cohorts, which may be considered limitations of this study.

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

INTRODUCTION

The global prevalence of type 2 diabetes (T2D) is rising, and rates of dysglycaemia (including impaired glucose tolerance and T2D) are projected to double in sub-Saharan Africa (SSA) compared with Europe in the next 25 years.[1] Within SSA, South Africa (SA) has the highest number of people living with T2D,[1] and T2D was the second leading cause of death in 2016 (5.5% of deaths) and the highest among women (7.2% of deaths). Ghana is also experiencing an upsurge of T2D, mainly driven by rapid economic growth and urbanisation. In rural Ghana, the age-standardised prevalence of T2D is 3.6% in men and 5.5% in women, compared with 10.3% for men and 9.2% for women in urban Ghana.[2] In addition, first-generation Ghanaian migrants in Europe show a higher burden of T2D than their counterparts in their country of origin (12.8% among Ghanaian men in Amsterdam and 3.6% among Ghanaian men in rural Ghana).[2] Corresponding data from Northern Sweden show a prevalence for T2D of about 5%, with slightly more men than women being affected and a modest increase during recent years.[3] Notably, a significantly greater proportion of persons with diabetes are undiagnosed in SSA compared with Europe (54% vs 36%).[1] Accordingly, population-specific risk stratification is essential for the early detection of T2D to prevent or delay the progression of the disease.

Current evidence suggests that the pathogenesis of T2D in populations of African ancestry (here referred to as 'Africans') is different to that of populations of European ancestry (here referred to as 'Europeans').[4 5] For the same level of body fatness, Africans have less visceral adipose tissue[6] and liver fat[7] than their European counterparts, but paradoxically have lower whole-body insulin sensitivity and present with hyperinsulinaemia.[6 8] Notably, this phenotype is also observed in African adults and children living on the continent and in the diaspora, suggesting that this trait is highly conserved.[9] There is also evidence of sexual dimorphism in the pathogenesis of T2D in both SSA and Europe,[10 11] which may be explained by sex differences in body fat distribution, biological, sociocultural and lifestyle factors.[10 12 13] Notably, the interplay of biology and lifestyle factors is also key to understanding the ethnic differences in T2D risk.[4 14 15]

Despite these vast phenotypic differences between Africans and Europeans, risk stratification and the clinical guidelines for managing T2D globally are largely based on studies that predominantly include participants of European ancestry. In the first prospective study to examine waist circumference thresholds to predict incident dysglycaemia (impaired glucose tolerance (IGT) and T2D) and T2D in an SSA population, we showed that the optimal thresholds in an SA cohort differed from those in European populations.[16] However, these African-specific thresholds still showed low specificity even when combined with metabolic syndrome risk factors.[16] This highlights the critical need for studies to identify sex-specific and ethnic-specific biomarkers with high sensitivity and specificity. Such biomarkers could facilitate early detection of dysglycaemia, enhancing the effectiveness of preventive therapies before symptoms appear.

Studies are now using omics approaches to gain a mechanistic understanding of the pathophysiology of T2D as well as to improve diagnostic and risk prediction in clinical settings.[17–19] Indeed, in metabolomic studies, we have shown that two phosphatidylcholines containing odd-chain fatty acids and 2-hydroxyethanesulfonate were predictive biomarkers of T2D in a Swedish prospective cohort.[19] We have also shown that metabolites involved in phospholipid, bile acid and branched-chain amino acid pathways predicted the development of T2D in SA women 13 years later.[20] Although the metabolites only marginally improved overall risk prediction compared with traditional risk factors, such as age, body mass index (BMI), fasting plasma glucose, lipidemia, blood pressure, physical activity and smoking, these metabolites capture lifestyle exposures (diet) and contribute to the mechanistic understanding of T2D.[19] In contrast, using targeted proteomic analysis of 276 proteins in a study of ~12 000 Europeans, we showed that selected proteins clearly improved early prediction for T2D beyond traditional risk factors.[19] Furthermore, using a targeted proteomics (184 proteins) approach in middle-aged Black SA men and women, we recently identified 73 proteins associated with dysglycaemia in a cross-sectional study, of which 34 were validated in the EpiHealth cohort from Sweden and 39 were specific to this African cohort.[21] These proteins require verification as biomarkers for prediction of dysglycaemia in prospective studies and in other African cohorts.

The Omics Approach for Personalised Prevention of Type 2 Diabetes Mellitus for African and European Populations (OPTIMA) study provides the unique opportunity to identify and validate metabolites and proteins in prospective studies of incident dysglycaemia in African cohorts from SA and Ghana. SA and Ghana represent countries at various stages of the epidemiological transition. The results from these African cohorts will be compared with those from a Swedish cohort, providing a western European point of reference. The design will allow us to identify metabolite and/or protein biomarkers reflecting differences in T2D aetiology and associated dietary patterns for sex, ethnicity and geographic location. We hypothesise that these biomarkers will improve the prediction of T2D beyond the traditional anthropometric and glycaemic markers, which have been shown to have poor discriminatory ability, particularly in African populations.[16] We also hypothesise that some metabolite/protein predictors may mediate effects of diet on dysglycaemia. Thus, dietary patterns associated with biomarkers and incident T2D will be explored to build the prototypes for ethnic-specific and sex-specific dietary interventions to prevent T2D, for which the cost-effectiveness will be assessed.[22] Furthermore, due to the sexual dimorphism in T2D risk, we will develop accurate sex-specific models for T2D prediction in the different populations. It is also essential to understand the biopsychosocial aspects of

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

primary T2D prevention that vary significantly between countries and their respective healthcare systems. These factors are heavily influenced by socioeconomic, sociocultural and personal factors like illness perception.[23] Therefore, it is essential to use a person-centred approach to gain an understanding of the perceptions of healthcare providers and those at risk of T2D regarding preventative strategies, particularly across different settings.

The overarching aim of this study is to develop ethnic-specific and sex-specific risk prediction models for the early detection of dysglycaemia in both Africans and Europeans. These models will inform clinically feasible, culturally acceptable and cost-effective risk management and prevention strategies using dietary modification in both SSA and European populations.

To achieve this overall aim, the following specific objectives will be addressed by harnessing prospective cohort data from two African cohorts, the Middle-Aged Soweto Cohort (MASC) from SA and the Research on Obesity and Diabetes among African Migrants Prospective (RODAM-Pros) cohort from Ghana and migrants in Europe, and a cohort from Sweden, the Pre-Swedish CArdioPulmonary BioImage Study (Pre- SCAPIS):

1. To use an omics approach to identify protein and/or metabolite biomarkers that predict the development of dysglycaemia in the different cohorts.
2. To establish ethnic-specific and sex-specific algorithms including traditional clinical parameters and newly identified biomarkers for the early detection of dysglycaemia in European and SSA populations.
3. To determine dietary patterns that explain maximum variation in the biomarker profiles and identify the associations of these dietary patterns with dysglycaemia among European and SSA populations; and to use these dietary patterns to build the basis for prototypes of sex-specific and ethnic-specific dietary interventions for the prevention of T2D.
4. Compare the cost-effectiveness and distributional consequences of this dietary intervention against other strategies for T2D prevention based on improvements in population health in Ghana, SA, Germany and Sweden.
5. To use a user-centred and person-centred care approach to understand the perceptions of high-risk participants and healthcare providers regarding T2D risk and the acceptability of ethnic-specific and sex-specific preventative dietary recommendations.

A schematic diagram of the study objectives is presented in figure 1.

METHODS AND ANALYSIS

Description of the prospective cohort studies

Middle-Aged Soweto Cohort

The MASC participants were originally part of the African WITS-INDEPTH Partnerships for Genomic Research (AWI-Gen) Collaborative Centre, which is a Human Heredity and Health in Africa (H3A) Consortium study, as described in detail previously.[24] The AWI-Gen study included 1027 men and 1004 women aged 40–60 years living in Soweto, from which the MASC participants were randomly selected (n=1021).[16] Pregnant women, first-degree relatives of existing participants, recent immigrants (with <10 years of residence in the region) and individuals with physical impairments preventing measurement of blood pressure and other anthropometric indices were excluded from the study. MASC baseline assessments were conducted between 2017 and 2018

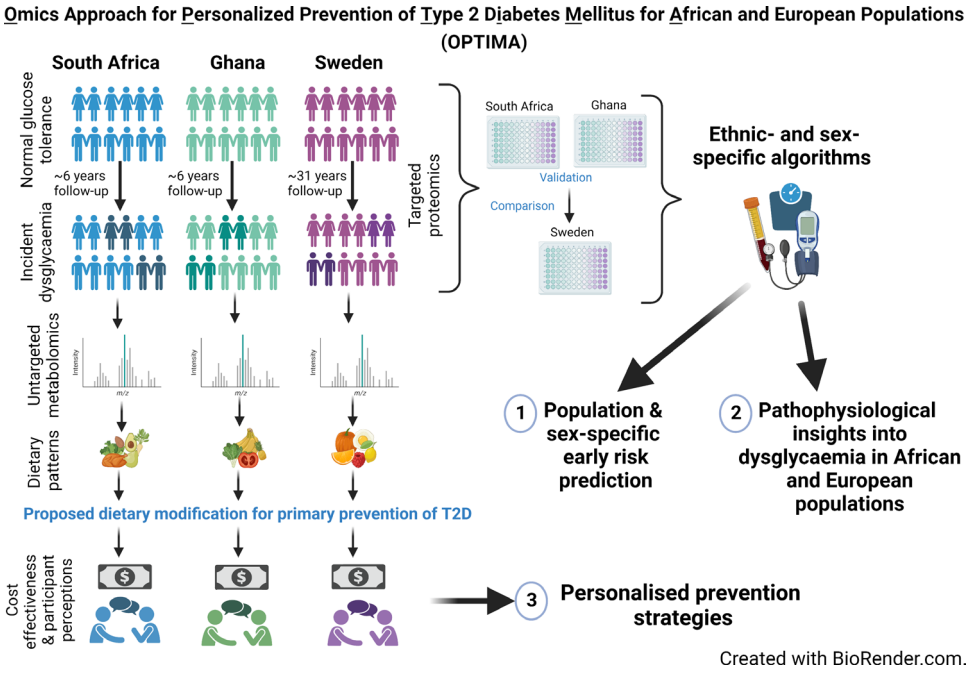

Figure 1 Schematic diagram of the OPTIMA study.

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

and the follow-up assessments were conducted between 2023 and 2024. The baseline and follow-up of the MASC were granted ethical clearance (clearance certificate no. M160604 and M230408, respectively) by the Human Research Ethics Committee (Medical) of the University of the Witwatersrand. Before participation in the research trial, all participants provided written informed consent.

Baseline and follow-up data collection included fasting blood sampling, measures of glucose tolerance (oral glucose tolerance test (OGTT)), body composition (anthropometry and dual X-ray absorptiometry), dietary intake (food frequency questionnaire (FFQ)), physical activity (accelerometry and Global Physical Activity Questionnaire (GPAQ)) and sociodemographic factors (questionnaires), as described previously.[13 16 25 26]

Research on Obesity and Diabetes among African Migrants Prospective cohort

Recruitment and baseline assessments of the multicentre, population-based RODAM cohort were performed between 2012 and 2015, with follow-up assessments being conducted between 2019 and 2021, as described previously.[27 28] The RODAM-Pros cohort includes participants from rural Ghana (n=638) and urban Ghana (n=608), and first-generation Ghanaian migrants residing in Amsterdam, the Netherlands (n=919). Also, individuals born in the Netherlands were recruited into RODAM-Pros but are not part of the OPTIMA project.[28] The participation rates at baseline were 76% in rural Ghana, 74% in urban Ghana and between 67% and 75% in Europe. Ethical approval was obtained from the respective ethics committees in Ghana (Committee on Human Research, Publication and Ethics, Kwame Nkrumah University of Science and Technology, Kumasi), and the Netherlands (Medical Ethics Review Committee, Academic Medical Centre, University of Amsterdam), both at baseline and follow-up. Written informed consent was obtained from all participants prior to participation.

Baseline (RODAM) and follow-up (RODAM-Pros) data collection included information on demographics, socioeconomic status, medical history, lifestyle factors, dietary intake (Ghana-specific Food Propensity Questionnaire (Ghana-FPQ), capturing the usual intake of 134 food items in 30 food groups), anthropometry and blood pressure, as described previously.[27 28] In addition, fasting blood samples were drawn for standard biochemical analyses, comprising but not limited to fasting glucose, haemoglobin A1c, insulin, blood lipids, liver enzymes, markers of kidney function and inflammation. The samples were centrally analysed at the laboratory of the Department of Endocrinology and Metabolism, Charité-Universitätsmedizin Berlin and at the laboratory at Amsterdam UMC, Amsterdam. No OGTT was performed in RODAM-Pros participants. All data were collected according to standard operating procedures, to allow for comparison between the different geographical locations and between baseline and follow-up.

The Pre-Swedish CArdioPulmonary bioImage Study

The Pre-Swedish CArdioPulmonary bioImage Study (Pre-SCAPIS) study includes individuals enrolled in SCAPIS-Umeå who were also enrolled in the Northern Sweden Health and Disease Study (NSHDS). SCAPIS is a general population-based prospective study (www.scapis.org). Between 2013 and 2018, men and women aged 50–64 years were randomly recruited from the census register at six sites in Sweden (Gothenburg, Linköping, Malmö/Lund, Stockholm, Umeå and Uppsala) and invited to a comprehensive examination as previously described.[29] In Umeå, 2508 participants were recruited between 2016 and 2017, of which 2072 (83%) had previously participated in the NSHDS. The NSHDS is the umbrella term for three cohorts, the Västerbotten Intervention Programme (VIP), the MOnitoring of Trends and Determinants in CArdiovascular Disease (MONICA) study and the Mammary Screening Programme (MSP). VIP and MONICA were initiated in the mid-1980s and are ongoing studies, whereas MSP stopped recruiting in the mid-1990s. For the purposes of this study, only participants from VIP and MONICA were included as those in the MSP did not complete an OGTT. The VIP invites inhabitants turning 40, 50 and 60 years for a health survey and MONICA is performed every fifth year and invites a random sample of the population aged 25–75 years. The VIP and MONICA studies used similar methodology, questionnaires including living conditions, lifestyle and health, dietary intake (64–66 or 84-item FFQ), physical activity, anthropometry, blood pressure and blood sampling for analysis of lipids and glucose as previously described.[30] Most participants in VIP and ~60% of the participants in MONICA had completed an OGTT after an overnight fast. In SCAPIS, a national protocol was followed with the focus on cardiopulmonary disease and its risk factors. In addition, an extensive questionnaire similar to the NSHDS questionnaire was used. For dietary intake, the FFQ Mini-Meal-Q was used. In SCAPIS, physical activity was evaluated by an accelerometer worn for 1 week. In addition to the national protocol, in Umeå, most participants had an OGTT.

The subsample of 2072 participants that underwent OGTT in SCAPIS had at least one previous examination in NSHDS, of which 1271 had at least two examinations and 476 had at least three examinations, altogether 3891 examinations, 94% in VIP, 3% in MONICA and 3% in MSP. This subsample with data and samples in NSHDS is hereafter referred to as pre-SCAPIS. The intervals between the first, second and third examinations in NSHDS and SCAPIS were 15.6, 9.0 and 5 years, respectively. At the first examination, 52% were women with a mean age of 42 years. The Swedish Ethics Authority approved the retrieval of data and samples from NSHDS (202105004) to perform biochemistry as described below (2023-02731-01). The SCAPIS study and the local extension (OGTT) were approved by the Regional Ethics committee in Umeå (2010-228-31M and 2016-151-3).

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Harmonisation of data between cohorts

As this study includes an opportunistic comparison of three independent cohorts, standardised methods for the collection of phenotypic data were not uniformly applied across cohorts. Nonetheless, most of the analysis will be performed within each cohort and comparisons between cohorts will be made by including only those variables that are comparable. These include the metabolomic and proteomic data that will be performed in a central laboratory (see below).

In terms of phenotypic data, for MASC and Pre-SCAPIS, fasting glucose was collected and a standard OGTT (75 g anhydrous glucose) was administered, whereas for RODAM-Pros, only fasting blood glucose was measured. Accordingly, dysglycaemia based on fasting glucose will be comparable across all three cohorts, while dysglycaemia based on the 2-hour glucose tolerance will be comparable only between MASC and Pre-SCAPIS. Anthropometry (weight, height and waist circumference) was measured at all sites using standard techniques and calibrated equipment. Similarly, blood pressure was measured at all sites in a seated position after 5 min rest using calibrated automated monitors (Pre-SCAPIS from 2000). Quality of life was measured using the 12-item Short Form Survey (SF-12) in all cohorts. Sociodemographic and lifestyle measures that are comparable between sites include age, marital status, ethnicity, education, employment status, smoking and alcohol intake. In addition, information on family history of T2D and medication use was recorded in all cohorts.

While FFQ has been used in all cohorts to capture dietary intake, these are different between cohorts, because they capture the usual intake of these different populations. To align with the goal of identifying ethnic-specific dietary strategies for T2D prevention and due to the fact that assessment instruments differed between the cohorts, the dietary analysis will be performed separately for each cohort. Physical activity has been assessed using both accelerometry[25] and GPAQ[31] in MASC, GPAQ in RODAM and the Cambridge Index[32] in Pre-SCAPIS. For the qualitative aspect of the project, standard interview guides will be developed in consultation with researchers from all cohorts, but these will be culturally adapted and specific to each setting. All methods are provided in detail in the sections below.

For metabolomics and proteomics data generation, the same analytical platforms (Waters Select MRT LC-QTOF) for metabolomics and proteomics (Olink, Uppsala, Sweden) will be used for all cohorts, ensuring the same quality control (QC) measures and sensitivity. However, some differences in metabolite and protein profiles are expected due to differences in sample matrices (EDTA plasma for MASC and Pre-SCAPIS, and lithium heparin plasma for RODAM-Pros). Such systematic differences between EDTA and heparin plasma are currently being evaluated on the analytical platform. Metabolomics and proteomics data will be analysed separately for each cohort, and metabolites associated with dysglycaemia will be replicated across cohorts, accounting for systematic differences between sample matrices.

Omics associated with the development of dysglycaemia and dietary patterns related to candidate biomarkers (objectives 1–3)

Study design

Within each cohort, a nested, case-control (1:1) design will be employed, analysing the baseline samples of incident cases and controls. Only participants with normal glucose tolerance (NGT) at baseline will be considered. Incident cases (those that develop dysglycaemia) and controls (those that remain NGT) will be matched based on age, sex and baseline BMI. Additionally, in the RODAM-Pros cohort, participants will also be matched based on site (rural, urban, Europe) and baseline fasting glucose concentrations.

Glycaemic status will be defined based on the American Diabetes Association criteria[33] as follows: NGT: fasting glucose <5.6 mmol/L and 2-hour OGTT <7.8 mmol/L; impaired fasting glucose (IFG): fasting glucose 5.6–6.9 mmol/L; IGT: 2-hour post glucose load: 7.8–11.0 mmol/L and T2D: fasting glucose ≥7 mmol/L and/or 2-hour postglucose load ≥11.1 mmol/L, or taking T2D medication. For MASC and Pre-SCAPIS, dysglycaemia will be defined as IFG and/or IGT or T2D, whereas for RODAM-Pros dysglycaemia will be defined as IFG or T2D based on the fasting glucose values.

Proteomic analysis

Proteomic analysis on baseline plasma samples will be performed with the multiplex immunoassays Olink Proseek Multiplex Metabolism, cardiovascular disease (CVD) II and CVD III panels (Olink, Uppsala, Sweden), measuring a total of 184 preselected protein biomarkers (www.olink.com/downloads). The method is based on proximity extension assay technology, and in each kit, 92 oligonucleotide-labelled antibody probe pairs can bind to their respective targets in the sample. Within each cohort, all samples will be randomised across plates, and four internal controls will be added to each sample to monitor the quality of assay performance. Intracoefficient and intercoefficient of variance are based on control samples (pooled plasma samples) included on each plate. Data are presented as normalised protein expression values, Olink Proteomics' arbitrary unit on Log2 scale (https://www.olink.com/key-links/).

Metabolomic analysis

Targeted and untargeted metabolomics analyses will be conducted on baseline plasma samples from the three cohorts. Untargeted metabolomics will follow the approach described by Zheng *et al*.[34] In brief, fasting plasma samples will undergo a simple protein precipitation step using methanol with a 1:9 ratio. Untargeted metabolomics analysis is performed using a high-performance liquid chromatography system (Waters Select MRT LC-QTOF) with reversed-phase (RP) and hydrophilic interaction

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

(HILIC) columns coupled with high-resolution mass spectrometry running in both modes in RP and only positive mode for HILIC. Quality assurance and analytical drift management are performed using study-specific QC and long-term QC samples injected regularly throughout each analytical batch. Untargeted metabolomics data will be preprocessed and analysed using in-house developed pipelines adapted to large-scale studies.[35–37]

Sample size estimation

We plan to include 600 participants (300 case/control pairs) per cohort in the studies related to the omics-related analyses (objectives 1–3). Sample size depends on different factors such as the minimum effect size to be detected, the estimated variability in the measurements, the desired statistical power, significance threshold and the test to be used for the analysis. In the case of omics such as metabolomics and proteomics, the traditional approaches for power estimations are not easily transferable, since a very large number of variables are of interest (metabolite features and proteins), and they are all measured with unknown variability and most importantly, the expected effect size is unknown. For metabolomics and proteomics data, the estimation of variation will be highly dependent on cohort and hence, using pilot data may not provide estimations that are easily transferable to another study, and may thus give limited guidance with regard to sample size estimations. There are data-driven approaches to estimate sample size for omics studies,[38–40] but they have seldom been implemented for the reasons mentioned above. From simulations, it was shown that a sample size of 20–200 individuals, depending on the setting, was sufficient to see meaningful (small) differences in common metabolites between cases and controls with a power of 0.8.[39] For proteomics, it has been shown that meaningful differences (2–3 SDs) in protein levels could be demonstrated in a case-control setting with a power of 0.8 among 50–250 case-control pairs.[41] For multi-omics studies—where, for example, metabolomics and proteomics data are combined—there are currently few established methods for power estimation, due to challenges such as differing noise levels, dynamic ranges, etc. Similarly, multi-omics datasets are increasingly collected to develop sample class predictors applying machine learning methods. In this case, the classification error rate, rather than the significance value, is used to assess performance, which requires different approaches to estimate formal power.[42] We have previously shown that such a sample size is sufficient to make meaningful targeted proteomic and metabolome-wide association analysis in the entire cohort to find metabolite signatures related to NGT, IGT and T2D.[19 21] We have conducted a similar analysis in a corresponding Swedish nested case-control analysis of comparable sample size, where we identified metabolites associated with incident T2D (502 case-control pairs) as well as their changes over a 10-year period in a subset (n=290). In addition, we identified markers related to insulin resistance and impaired glucose tolerance.[19] We identified a total of 39 metabolites associated with incident T2D, 38 of which had been previously reported in other studies, demonstrating the feasibility of our approach and the adequacy of the sample size used.[19] Moreover, we identified metabolites linking specific dietary exposures—such as fish intake, coffee intake and adherence to a Nordic diet—to T2D risk.[43 44] Given the challenges described—and considering that our datasets are larger than what has been suggested by separate power estimations for metabolomics and proteomics studies, as well as the meaningful and replicable findings observed in previous datasets of similar size (see above)—we are confident that the sample size is sufficient.

Data and statistical analysis

Protein and/or metabolite biomarkers that predict the development of dysglycaemia in the different cohorts (objective 1)

Within each cohort, participant characteristics will be reported after stratification by sex as mean (SD) or median (25th–75 percentiles) for normally distributed data and skewed data, respectively, and categorical variables will be reported as n (%). Differences between sex will be examined using a t-test, a Mann-Whitney U test and/or a χ^2 test, respectively.

Cohort-specific z-scores will be calculated for all proteins and metabolites based on the raw score minus the population mean, divided by the population SD. This allows for direct comparisons of the magnitude of the effect sizes of each individual protein/metabolite across cohorts. Missing data will be excluded pairwise for all analyses. Significance will be set at $p<0.05$.

We will evaluate different multivariate approaches including random forest, orthogonal partial least squares discriminant analysis (OPLS-DA) and logistic regression along with an in-house developed multilevel partial least squares regression (PLS) approach[19] to identify proteins, metabolites or metabolic pathways associated with incident dysglycaemia. The logistic models will explore any sex interactions and will be adjusted for confounding factors such as age, adiposity (BMI or waist circumference), sociodemographic (education, occupation and markers of wealth) and lifestyle characteristics (alcohol, smoking, physical activity and energy intake), known diseases and medication use. The level of significance will be set at a false discovery rate of 5%.

Ethnic-specific and sex-specific algorithms for the early detection of dysglycaemia in European and SSA populations (objective 2)

Using the omics data and traditional risk factors for each cohort, we will develop ethnic-specific and sex-specific algorithms that predict incident dysglycaemia in SSA and European populations. L1-regularised Cox regression implemented with the least absolute shrinkage and selection operator algorithm (LASSO-Cox)[45 46] (R package 'glmnet') will be applied to identify a parsimonious set of 'key' predictors from the unique proteins and metabolites selected in each cohort (objective 1), and traditional risk factors. Data will be auto-scaled prior to modelling.

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

To obtain a robust variable selection, the populations will be bootstrapped 200 times and subjected to LASSO-Cox modelling. The optimally selected variables will be ranked based on the frequency of their selection and considered 'key' predictors if selected in, for example, ≥70% of the bootstrap samples. Traditional risk factors known to be associated with IGT and T2D risk will be considered. They are age, anthropometric measurements (ie, waist circumference, weight and BMI), fasting biomedical measurements (ie, plasma glucose, low-density lipid-cholesterol and high-density lipid-cholesterol, triglycerides and total cholesterol), lifestyle factors (ie, smoking status, physical activity). Separate and combined models with proteins and metabolites along with traditional risk factors will be created. Missing values in the traditional risk factors will be imputed using the R package 'missForest'.

We will compare prediction performance between different panels of LASSO-Cox identified key predictors: (i) with proteins only; (ii) with metabolites only, (iii) protein and metabolites only; (iv) with traditional risk factors only; (v) with the addition of key proteins to key traditional risk factors; (vi) with the addition of key metabolites to traditional risk factors; (vii) with combined predictors selected from the combined dataset of 184 proteins, metabolite biomarkers and traditional risk factors. The performance of selected predictors will be assessed for men and women separately to account for potential differences due to sex. Harrell's concordance index and 95% CIs will be computed for the selected predictors (R package 'survival' and 'rms') and H-indices will be compared across models. We will also assess the time-dependent area under the receiver operating characteristic curves (AROCs) using the R package 'risksetROC', to evaluate how well the selected predictors model dysglycaemia incidence in the dataset using an incident/dynamic definition of the ROC curve at different follow-up periods.

Dietary patterns that explain maximum variation in the biomarker profiles and associate with dysglycaemia among SSA and European populations (objective 3)

In this project, we aim to derive sex-specific and ethnic-specific dietary recommendations for the prevention of T2D. Therefore, the analysis of dietary data in relation to the identified candidate biomarkers will be performed for each cohort separately. Nonetheless, the data analysis will follow the same protocol using the food frequency data from the three cohorts. For MASC, a 7-day quantitative food frequency questionnaire was used, consisting of 214 commonly eaten foods derived from analyses of 11 dietary surveys conducted in rural and urban SA, as described previously.[13] For the RODAM-Pros cohort, the Ghana-FPQ, capturing the usual intake of 134 food items in 30 food groups, was administered. For Pre-SCAPIS, a validated semi-quantitative FFQ was used, which initially consisted of 84 food items and since 1996 was reduced to 64–66 food items.[47] We will employ the respective national nutrient databases to translate semi-quantitative food group intakes into energy intake, as well as nutrient intakes (carbohydrate, fat, protein, fibre, micronutrients). These include the West African Nutrient Database[48] for RODAM-Pros, the South African food composition tables hosted by South African Medical Research Council[49] for MASC and the national food composition database at the Swedish Food Agency (https://soknaringsinnehall.livsmedelsverket.se/) for Pre-SCAPIS.

To identify the role of diet as a modifiable risk factor in the biomarker disease pathway, the food items will be grouped into common food groups in each cohort and then subjected to reduced rank regression (RRR). Through RRR, we will combine the exploratory nature of data-driven techniques of dietary pattern analysis (based on common food groups) with hypothesis-based approaches informed by the identified biomarker-disease associations (here: proteins and metabolites for dysglycaemia). RRR is a specific form of PLS analysis that maximises the variation in response variables (proteins and metabolites) through linear combinations of predictor variables (food group intakes). In each cohort, the participants will be assigned an RRR score that reflects the degree of adherence to this dietary pattern, that is, how strong an individual follows the identified dietary pattern. In the next step, we will calculate Cox regression models with age as the underlying timescale to determine the prospective associations of the baseline RRR dietary patterns with incident dysglycaemia. We will adjust for relevant lifestyle variables such as energy intake, physical activity, smoking and alcohol consumption, as well as for socio-economic variables, such as education, occupation and markers of wealth. Also, anthropometric measures will be considered in the final step of the models, as they might mediate the relationships between diet, biomarkers and dysglycaemia. The RRR-derived dietary patterns and the national food-based dietary guidelines will build the basis of ethnic-specific and sex-specific prototypes of dietary interventions in each country.

A comparative health economic analysis of strategies for the prevention of T2D (objective 4)

The health economic substudy focuses on a comparative cost-effectiveness analysis (CEA) between the participating countries, that is, Ghana (low), SA (medium) and Sweden (high), as well as migrant Ghanaians living in Germany (high), providing context at different stages of economic development. The CEA of the ethnic-specific and sex-specific prototypes of dietary intervention for T2D prevention for each country (objective 3) in combination with screening for T2D will be compared with two alternatives: (i) effects of sugar-sweetened beverages (SSB) tax and (ii) effects of both SSB tax and screening for T2D.

Normative principles applied in the health economic analysis are (a) cost-effectiveness and (b) availability, defined as the intervention capacity to reach those in need. Thus, we will adopt a population perspective and investigate the health benefits of the strategies and

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

determine whether the interventions influence the distribution of health, namely whether they close or widen inequality gaps along factors such as income, rural-urban setting, education, and gender.

We hypothesise that the effectiveness of the strategies is largely determined by factors outside the healthcare system such as:

- The disposable household income spent on food.
- Whether the society can effectively implement T2D screening and/or taxes on SSB.
- The individual's cost of participation in T2D screening. That includes loss of income during the screening, travel time and costs directly related to travel (eg, cost of tickets/taxi fares).
- If the population registration system is reliable and provides accurate information to healthcare providers.

The main data sources are the respective cohort studies (RODAM-Pros, MASC and Pre-SCAPIS) from the different countries. The analysis will primarily use data common to the three cohorts, such as SF-12, to calculate quality-adjusted life years (QALY) weights, and rely on estimates informed by the secondary literature. Costing data sources will be context-specific.

The analysis will estimate the extent to which the different strategies can effectively reach the target groups in the four countries. This will be based on an analysis of the external factors outlined above, and secondary data on the outcomes of SSB taxes and screening for T2D in other contexts. We will model cost and QALYs gained for the segments of the population reached by the interventions, based on a previously published Markov model by Neumann *et al*[22] to model T2D progression. To assess dietary lifestyle interventions and SSB taxes, the health effects of a changed diet will be modelled by relying on abstractions of dietary characteristics such as the amount of sugar consumed. Uncertainty will be assessed using probabilistic and scenario-based sensitivity analyses. The main outcome of the incremental cost-effectiveness analysis will be costs per QALY. To ensure comparability between the different countries, cost will be defined relatively, in units of disposable household income or will be standardised using purchasing power parity. Stratified analyses will be used to assess health equity consequences.

The goal is to provide context-specific recommendations that consider each strategy's capacity to improve outreach, reduce health gaps and achieve cost-effectiveness.

Perceptions regarding ethnic-specific and sex-specific dietary preventative strategies among at-risk European and SSA adults and healthcare providers (objective 5)

We will apply a user-centred and person-centred design and collaborate with patient organisations and social unions in Sweden, SA, Ghana and Germany (representing migrant Ghanaians as part of RODAM-Pros) (as described below in patient and public involvement) to get help to contextualise and culturally adapt semi-structured interview guides to be used among participants at risk for T2D. We will then perform focus group discussions (FGDs) with people at risk of T2D and in-depth interviews (IDIs) with healthcare providers (primary care nurses, dieticians and diabetes educators) in Sweden, SA, Ghana and Germany.

Two FGDs for both men and women (four in total), including 8–10 participants per focus group, will be hosted in each setting. Participants with impaired glucose metabolism (fasting glucose: 6.1–7 mmol/L or 2-hour glucose: 7.8–11.1 mmol/L during an OGTT) will be recruited from the Pre-SCAPIS and MASC cohorts. In both Ghana and Germany, participants >40 years of age who are living with obesity (BMI >30 kg/m^2) and have a family history of T2D will be recruited from the Diabetes Center of Komfo Anokye Teaching Hospital in Kumasi and from community centres and faith organisations in Germany, respectively. IDIs will be conducted with 8–15 healthcare providers from each country. In the interviews, acceptability and general willingness to participate in targeted dietary strategies to prevent T2D will be the focus among at-risk participants, while person-centred care and goal-centred approaches[23 50] will be the focus among the healthcare providers. The interviews will follow a culturally adapted semi-structured interview guide that is specific to each setting. The interview guide is based on earlier research such as the Theoretical Framework of Acceptability (TFA),[51] but the interviews strive to be inductive, where the participants' agendas and experience are the focus. The TFA systematically assesses how participants and healthcare providers perceive interventions, considering attitudes, beliefs and expectations about dietary changes for preventing dysglycaemia. It addresses practical aspects like perceived burden (time, effort, resources), ethical considerations and opportunity costs. The framework identifies barriers and facilitators pre-intervention, guiding adjustments for optimal effectiveness. Considering cultural and contextual factors ensures interventions are sensitive and appropriate. Overall, the framework informs decision-making by assessing stakeholders' perspectives, crucial for effective intervention implementation. Interviews will be digitally recorded, transcribed verbatim, translated into English when needed and analysed inductively using qualitative content analysis[52] by qualitative researchers in the team. To ensure trustworthiness, triangulation of parts of the data analysis from each country will be performed, where coding and categorisation of original texts will be discussed from different perspectives and in relation to the various contexts. This will be an iterative process where every person involved in the analysis will contribute.

Patient and public involvement

Patient and public involvement within SA, Ghana, Germany and Sweden will be through diabetes organisations, and through interactions with health professionals (dieticians, nurses, diabetes educators) and patients linked with clinical and community health facilities. The diabetes organisations in SA are represented by Sweet

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Life, Diabetes South Africa, Diabetes Education Society of South Africa and the Society for Endocrinology, Metabolism and Diabetes in South Africa. In Ghana and Germany, we collaborate with the Ghana Academy of Nutrition and Diabetics and the Ghana Women Association, respectively. Sweden is represented by local representatives of the Swedish Diabetes Association.

These regional groups will be invited in accordance with the person-centred design of the FGDs (objective 5). Person-centred usability methods can identify important issues and information that can enhance the participants' experience and thereby improve the quality of study tools, and also facilitate implementation on a larger scale. In the development of interview guides, reference groups of participants from patient organisations will be included. They will also be requested to respond to possible questions during the analytic process as well as comment on the result and its interpretation before it is published. In case of later implementation of self-management support on a larger scale, possibly in a digital form, these reference groups will be invited again to inform about their view on usability.[53]

ETHICS AND DISSEMINATION

Ethical clearance to complete the omics-related and dietary-related analyses (objectives 1–3) has been obtained for each site, as outlined in the cohort descriptions. Ethical clearance for the focus group discussions (objective 5) will be obtained from the Health Research Ethics Committees at each site prior to participant recruitment. Before participation in the research trial, all participants will provide written informed consent.

Data will be adequately pseudonymised to ensure that individuals cannot be identified within the dataset. All data will be stored on secure, encrypted and password-protected servers at each participating institution, with appropriate backup protocols in place. This study will adhere to the highest international information/data security standards including ISO/IEC 27018:2019, and ISO/IEC 27040:2015, which are in place at all the research sites, and continuously reviewed/revised based on current international practice. Prior to sending any samples or data, material transfer agreements and data transfer agreements will be signed between the four partner organisations. Data will then be transferred via a secure data link to the BIANCA cluster within the National Academic Infrastructure for Supercomputing in Sweden-Sensitive Data, hosted by Linköping University. This national computing resource is specifically designed for the management, storage and analysis of sensitive (personal) data, with access limited to authorised project members only. All analysis will be conducted within BIANCA, with only the output data exported.

The results of the study will be disseminated through scientific conferences and journal publications, as well as via community engagement events and diabetes organisations in the respective countries. Once published (including preprints), non-sensitive data will be made available in anonymised and aggregated form on reasonable request. Unpublished data will be available for sharing 3 years after the completion of the grant period (ie, from 2030). Access will be governed by the joint agreement of all project partners and in accordance with GDPR regulations.

Author affiliations
[1]Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden
[2]Biomedical Research and Innovation Platform, South African Medical Research Council, Cape Town, South Africa
[3]South African Medical Research Council/WITS Developmental Pathways for Health Research Unit (DPHRU), Department of Paediatrics, University of the Witwatersrand Johannesburg, Johannesburg, South Africa
[4]Transdisciplinary Research Area "Technology and Innovation for Sustainable Futures" and Center for Development Research (ZEF), University of Bonn, Bonn, Germany
[5]Heidelberg Institute of Global Health (HIGH), Medical Faculty and University Hospital, Heidelberg University, Heidelberg, Germany
[6]Department of Public and Occupational Health, Amsterdam UMC, Locatie AMC, Amsterdam, The Netherlands
[7]Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
[8]Department of Life Sciences, Division of Food and Nutrition Science, Chalmers University of Technology, Gothenburg, Sweden
[9]Health through Physical Activity, Lifestyle and Sport Research Centre (HPALS), Division of Physiological Sciences, Department of Human Biology, University of Cape Town, Cape Town, South Africa
[10]Department of Nursing, Umeå University, Umeå, Sweden
[11]Department of Internal Medicine and Clinical Nutrition, University of Gothenburg, Gothenburg, Sweden
[12]Department of Epidemiology and Global Health, Umeå University, Umeå, Sweden
[13]Riverland Academy of Clinical Excellence, Riverland Mallee Coorong Local Health Network, Berri, South Australia, Australia
[14]SciLifeLab, Stockholm, Sweden
[15]Wallenberg Laboratory, Department of Molecular and Clinical Medicine, Institute of Medicine, University of Gothenburg Sahlgrenska Academy, Gothenburg, Sweden

X Amy E Mendham @amy.mendham

Contributors Each author contributed to writing different aspects of the protocol and read and approved the final version of the protocol paper. JHG is the guarantor.

Funding This project is supported under the framework of the European Research Area Personalised Medicine Joint Transnational Call (ERA-PerMed JTC2022), by the South African Medical Research Council (SAMRC) with funds received from the Department of Science and Innovation, Bundesministerium für Bildung und Forschung (BMBF), Germany; Vinnova (Dnr: 2022-00547) and the Swedish Research Council (Vetenskapsrådet; Dnr: 2022-00924 for SN), Sweden. The handling of participants' sensitive data was enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) under project number SENS2023526, partially funded by the Vetenskapsrådet through grant agreement no. 2022-06725. This work was further supported by the SciLifeLab & Wallenberg Data Driven Life Science Programme (grant KAW 2020.0239; CW).

Disclaimer The SAMRC had no role in the design of the study; in the collection, analysis or interpretation of the data; in the writing of the report or in the decision to submit the paper for publication.

Competing interests JHG is employed by the SAMRC, which partly funded this research. The remaining authors declare no competing interests.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the 'Methods' section for further details.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; peer reviewed for ethical and funding approval prior to submission.

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Julia H Goedecke http://orcid.org/0000-0001-6795-4771
Amy E Mendham http://orcid.org/0000-0002-1959-6698
Lisa K Micklesfield http://orcid.org/0000-0002-4994-0779
Julia Otten http://orcid.org/0000-0001-9016-1139
Stefan Söderberg http://orcid.org/0000-0001-9225-1306
Eva L van der Linden http://orcid.org/0000-0002-4037-1717

REFERENCES

1 Saeedi P, Petersohn I, Salpea P, *et al*. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract* 2019;157:107843.
2 Agyemang C, Meeks K, Beune E, *et al*. Obesity and type 2 diabetes in sub-Saharan Africans - Is the burden in today's Africa similar to African migrants in Europe? The RODAM study. *BMC Med* 2016;14:166.
3 Jansson SPO, Fall K, Brus O, *et al*. Prevalence and incidence of diabetes mellitus: a nationwide population-based pharmaco-epidemiological study in Sweden. *Diabet Med* 2015;32:1319–28.
4 Goedecke JH, Olsson T. Pathogenesis of type 2 diabetes risk in black Africans: a South African perspective. *J Intern Med* 2020;288:284–94.
5 Goedecke JH, Mendham AE. Pathophysiology of type 2 diabetes in sub-Saharan Africans. *Diabetologia* 2022;65:1967–80.
6 Goedecke JH, Levitt NS, Lambert EV, *et al*. Differential effects of abdominal adipose tissue distribution on insulin sensitivity in black and white South African women. *Obesity (Silver Spring)* 2009;17:1506–12.
7 Goedecke JH, Keswell D, Weinreich C, *et al*. Ethnic differences in hepatic and systemic insulin sensitivity and their associated determinants in obese black and white South African women. *Diabetologia* 2015;58:2647–52.
8 Goedecke JH, Dave JA, Faulenbach MV, *et al*. Insulin response in relation to insulin sensitivity: an appropriate beta-cell response in black South African women. *Diabetes Care* 2009;32:860–5.
9 Osei K, Schuster DP, Owusu SK, *et al*. Race and ethnicity determine serum insulin and C-peptide concentrations and hepatic insulin extraction and insulin clearance: comparative studies of three populations of West African ancestry and white Americans. *Metab Clin Exp* 1997;46:53–8.
10 Goedecke JH, George C, Veras K, *et al*. Sex differences in insulin sensitivity and insulin response with increasing age in black South African men and women. *Diabetes Res Clin Pract* 2016;122:207–14.
11 Tramunt B, Smati S, Grandgeorge N, *et al*. Sex differences in metabolic regulation and diabetes susceptibility. *Diabetologia* 2020;63:453–61.
12 Micklesfield LK, Kagura J, Munthali R, *et al*. Demographic, socio-economic and behavioural correlates of BMI in middle-aged black men and women from urban Johannesburg, South Africa. *Glob Health Action* 2018;11:1448250.
13 Ratshikombo T, Goedecke JH, Soboyisi M, *et al*. Sex Differences in the Associations of Nutrient Patterns with Total and Regional Adiposity: A Study of Middle-Aged Black South African Men and Women. *Nutrients* 2021;13:4558.
14 Courville AB, Chung ST. Diet and Fat Metabolism-Uncovering Key Determinants of Diabetes Risk Prediction in Populations of African Descent. *J Clin Endocrinol Metab* 2021;106:e1900:e1900–1:.
15 Micklesfield LK, Lambert EV, Hume DJ, *et al*. Socio-cultural, environmental and behavioural determinants of obesity in black South African women. *Cardiovasc J Afr* 2013;24:369–75.
16 Goedecke JH, Nguyen KA, Kufe C, *et al*. Waist circumference thresholds predicting incident dysglycaemia and type 2 diabetes in Black African men and women. *Diabetes Obes Metab* 2022;24:918.
17 Thorand B, Zierer A, Büyüközkan M, *et al*. A Panel of 6 Biomarkers Significantly Improves the Prediction of Type 2 Diabetes in the MONICA/KORA Study Population. *J Clin Endocrinol Metab* 2021;106:e1647.
18 Abbasi A, Sahlqvist A-S, Lotta L, *et al*. A Systematic Review of Biomarkers and Risk of Incident Type 2 Diabetes: An Overview of Epidemiological, Prediction and Aetiological Research Literature. *PLoS One* 2016;11:e0163721.
19 Shi L, Brunius C, Lehtonen M, *et al*. Plasma metabolites associated with type 2 diabetes in a Swedish population: a case-control study nested in a prospective cohort. *Diabetologia* 2018;61:849.
20 Zeng Y, Mtintsilana A, Goedecke JH, *et al*. Alterations in the metabolism of phospholipids, bile acids and branched-chain amino acids predicts development of type 2 diabetes in black South African women: a prospective cohort study. *Metab Clin Exp* 2019;95:57–64.
21 Mendham AE, Micklesfield LK, Karpe F, *et al*. Targeted proteomics identifies potential biomarkers of dysglycaemia, beta cell function and insulin sensitivity in Black African men and women. *Diabetologia* 2023;66:174–89.
22 Neumann A, Lindholm L, Norberg M, *et al*. The cost-effectiveness of interventions targeting lifestyle change for the prevention of diabetes in a Swedish primary care and community based prevention program. *Eur J Health Econ* 2017;18:905–19.
23 Rutten G, Alzaid A. Person-centred type 2 diabetes care: time for a paradigm shift. *Lancet Diabetes Endocrinol* 2018;6:264–6.
24 Ali SA, Soo C, Agongo G, *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* 2018;11:1507133.
25 Micklesfield LK, Westgate K, Smith A, *et al*. Physical Activity Behaviors of a Middle-Age South African Cohort as Determined by Integrated Hip and Thigh Accelerometry. *Med Sci Sports Exerc* 2022;54:1493–505.
26 Kufe CN, Micklesfield LK, Masemola M, *et al*. Differences in Type 2 Diabetes Risk and the Association with Total and Regional Adiposity in Middle-Aged Black South African Men and Women. *JEMDSA* 2021;21.
27 Agyemang C, Beune E, Meeks K, *et al*. Rationale and cross-sectional study design of the Research on Obesity and type 2 Diabetes among African Migrants: the RODAM study. *BMJ Open* 2014;4:e004877.
28 Agyemang C, van der Linden EL, Antwi-Berko D, *et al*. Cohort profile: Research on Obesity and Diabetes among African Migrants in Europe and Africa Prospective (RODAM-Pros) cohort study. *BMJ Open* 2022;12:e067906.
29 Bergström G, Berglund G, Blomberg A, *et al*. The Swedish CArdioPulmonary BioImage Study: objectives and design. *J Intern Med* 2015;278:645–59.
30 Ljungberg J, Johansson B, Engström KG, *et al*. Traditional Cardiovascular Risk Factors and Their Relation to Future Surgery for Valvular Heart Disease or Ascending Aortic Disease: a case-referent study. *J Am Heart Assoc* 2017;6:e005133.
31 Bull FC, Maslin TS, Armstrong T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. *J Phys Act Health* 2009;6:790–804.
32 InterAct C, Peters T, Brage S. Validity of a short questionnaire to assess physical activity in 10 European countries. *Eur J Epidemiol* 2012;27:15–25.
33 American Diabetes Association. Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 2014;37:S81–90.
34 Zheng R, Brunius C, Shi L, *et al*. Prediction and evaluation of the effect of pre-centrifugation sample management on the measurable untargeted LC-MS plasma metabolome. *Anal Chim Acta* 2021;1182:338968.
35 Shi L, Westerhuis JA, Rosén J, *et al*. Variable selection and validation in multivariate modelling. *Bioinformatics* 2019;35:972–80.
36 Brunius C, Shi L, Landberg R. Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction. *Metabolomics (Los Angel)* 2016;12:173.
37 Schillemans T, Shi L, Liu X, *et al*. Visualization and Interpretation of Multivariate Associations with Disease Risk Markers and Disease Risk-The Triplot. *Metabolites* 2019;9:133.
38 Nyamundanda G, Gormley IC, Fan Y, *et al*. MetSizeR: selecting the optimal sample size for metabolomic studies using an analysis based approach. *BMC Bioinformatics* 2013;14:338.
39 Blaise BJ, Correia G, Tin A, *et al*. Power Analysis and Sample Size Determination in Metabolic Phenotyping. *Anal Chem* 2016;88:5179–88.
40 Saccenti E, Timmerman ME. Approaches to Sample Size Determination for Multivariate Data: Applications to PCA and PLS-DA of Omics Data. *J Proteome Res* 2016;15:2379–93.
41 Skates SJ, Gillette MA, LaBaer J, *et al*. Statistical design for biospecimen cohort size in proteomics-based biomarker discovery and verification studies. *J Proteome Res* 2013;12:5383–94.
42 Tarazona S, Balzano-Nogueira L, Gómez-Cabrero D, *et al*. Harmonization of quality metrics and power calculation in multi-omic studies. *Nat Commun* 2020;11:3092.

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

43 Shi L, Brunius C, Johansson I, *et al*. Plasma metabolite biomarkers of boiled and filtered coffee intake and their association with type 2 diabetes risk. *J Intern Med* 2020;287:405–21.
44 Shi L, Brunius C, Bergdahl IA, *et al*. Joint Analysis of Metabolite Markers of Fish Intake and Persistent Organic Pollutants in Relation to Type 2 Diabetes Risk in Swedish Adults. *J Nutr* 2019;149:1413–23.
45 Nowak C, Sundström J, Gustafsson S, *et al*. Protein Biomarkers for Insulin Resistance and Type 2 Diabetes Risk in Two Large Community Cohorts. *Diabetes* 2016;65:276–84.
46 Nowak C, Carlsson AC, Östgren CJ, *et al*. Multiplex proteomics for prediction of major cardiovascular events in type 2 diabetes. *Diabetologia* 2018;61:1748:1748–57:.
47 Johansson I, Hallmans G, Wikman A, *et al*. Validation and calibration of food-frequency questionnaire measurements in the Northern Sweden Health and Disease cohort. *Public Health Nutr* 2002;5:487–96.
48 Vincent A, Grande F, Compaore E, *et al*. *Food composition table for Western Africa (2019) user guide & condensed food composition table (2019)*. Rome: FAO/INFOODS, 2020.
49 Wolmarans P, Kunneke E, Laubscher R. Use of the South African Food Composition Database System (SAFOODS) and its products in assessing dietary intake data: Part II. South Afr J Clin Nutr 2009;22:59–67.
50 Boeykens D, Boeckxstaens P, De Sutter A, *et al*. Goal-oriented care for patients with chronic conditions or multimorbidity in primary care: A scoping review and concept analysis. *PLoS One* 2022;17:e0262843.
51 Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC Health Serv Res* 2017;17:88.
52 Graneheim UH, Lindgren BM, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Educ Today* 2017;56:29–34.
53 Walden A, Garvin L, Smerek M, *et al*. User-centered design principles in the development of clinical research tools. *Clin Trials* 2020;17:703–11.

BMJ Open: first published as 10.1136/bmjopen-2025-099108 on 22 April 2025. Downloaded from http://bmjopen.bmj.com/ on May 23, 2025 by guest . Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.