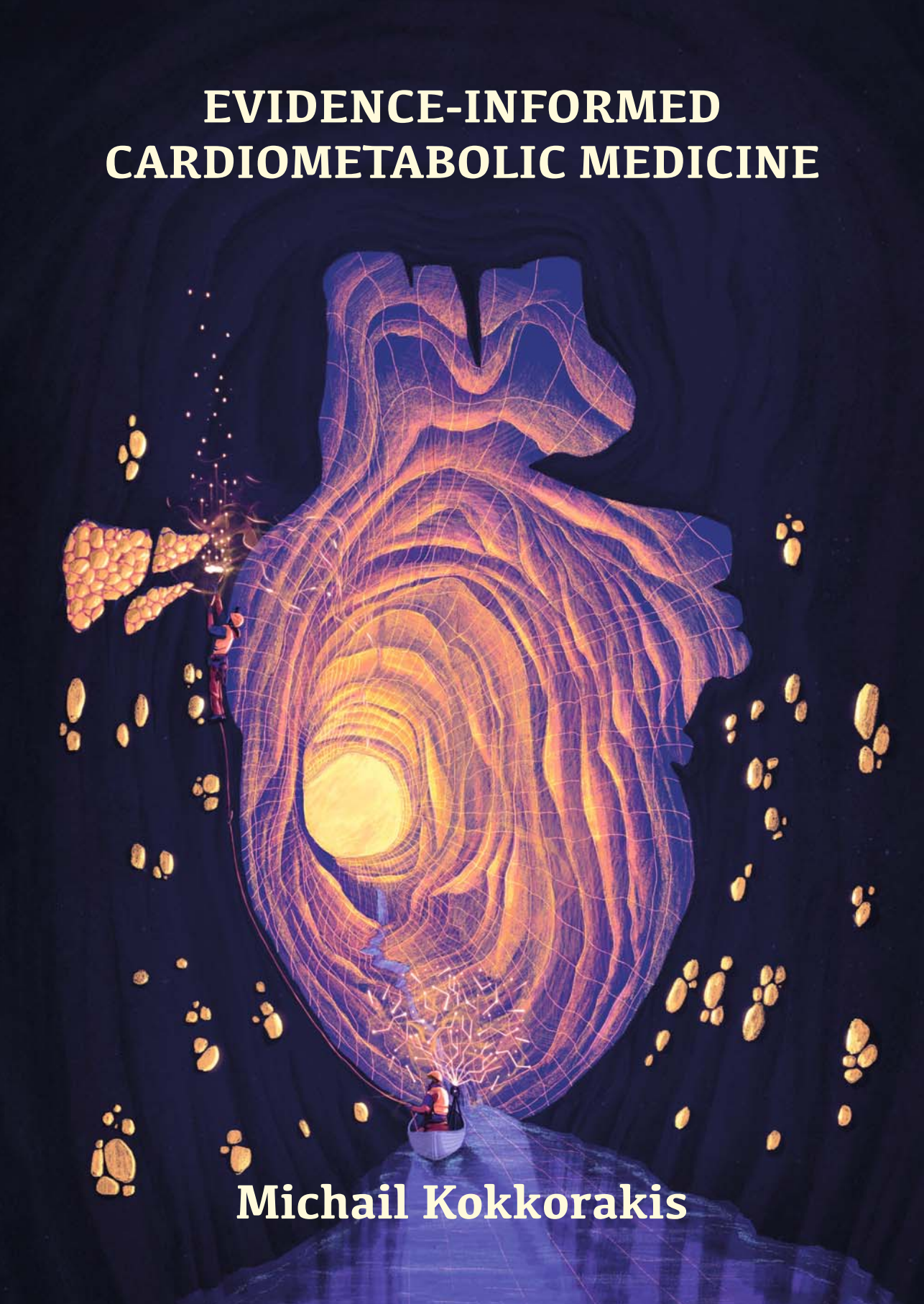
EVIDENCE-INFORMED CARDIOMETABOLIC MEDICINE



Michail Kokkorakis

EVIDENCE-INFORMED CARDIOMETABOLIC MEDICINE

Michail Kokkorakis

ISBN: 978-94-6522-590-6

Provided by thesis specialist Ridderprint, ridderprint.nl
Printing: Ridderprint
Layout and design: Dennis Hendriks, persoonlijkproefschrift.nl
Cover artwork: Simone Rotella
Chapter paintings artwork: generated by Michail Kokkorakis using OpenAI's DALL·E and subsequently refined with image editing software

The publication of this thesis was financially supported by the University of Groningen.

university of groningen

Evidence-Informed Cardiometabolic Medicine

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. J.M.A. Scherpen
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Monday 3 November 2025 at 11.00 hours

by

Michail Kokkorakis

born on 19 February 2001

Supervisors
Prof. R.H. Henning
Prof. C.S. Mantzoros

Co-supervisor
Dr. J.N. Alves Castela Cardoso Forte

Assessment Committee
Prof. R.S.N. Fehrmann
Prof. D.E. Atsma
Prof. H. Pijl

To my family.

«Ασφάλεια εστί το προνοείν και προλαμβάνειν.
Το δε προνοείν και προλαμβάνειν κρείττον εστί του θεραπεύειν.»

~ Ιπποκράτης

“Safety lies in foresight and prevention.
And foresight and prevention are better than cure.”

~ Hippocrates

Contents

Chapter 1. Introduction 9

Part I ~ Non-laboratory Predictive Modeling and Scalable Risk Stratification

Chapter 2. Effective questionnaire-based prediction models for type 2 diabetes across several ethnicities: a model development and validation study 29
EClinicalMedicine 2023;64:102235

Chapter 3. Simplifying coronary artery disease risk stratification: development and validation of a questionnaire-based alternative comparable to clinical risk tools 57
EBioMedicine 2025;111:105518

Part II ~ Omics Integration in Risk Stratification and Modeling Approaches

Chapter 4. GDF-15 improves the predictive capacity of steatotic liver disease non-invasive tests for incident morbidity and mortality risk for cardio-renal-metabolic diseases and malignancies 91
Metabolism 2024;163:156047

Chapter 5. Radiomic and proteomic signatures of body mass index on brain ageing and Alzheimer's-like patterns of brain atrophy 125
EBioMedicine 2025;116:105763 (pre-publication version)

Part III ~ Real-World Evidence and Comparative Effectiveness of Cardiometabolic Drugs in Neurocognitive Outcomes

Chapter 6. Comparative effectiveness of sodium-glucose cotransporter-2 inhibitors and glucagon-like peptide-1 receptor agonists for incident dementia risk: a retrospective multicohort study 163
Submitted manuscript

Chapter 7. Discussion 185

Appendices English summary 208
Nederlandse samenvatting 210
Acknowledgements 212
Curriculum vitae 213
List of publications 214


PROMETHEUS

Chapter 1

Introduction

Preface

Prometheus' Liberation: A mythological parallel to the struggle and hope in cardiometabolic medicine

In Greek mythology, the Titan Prometheus (Προμηθεύς; *etymology: forethought*) is portrayed as a figure of defiance, sacrifice, and eventual redemption. In the Mecone trick, which symbolized the "settling of accounts" between mortals and gods, Prometheus deceived Zeus (Ζεύς), the king of the gods, during a sacrificial meal. He presented two offerings to the Olympian: one was a portion of beef concealed inside an ox's stomach (nutritious food hidden in an unappealing exterior), and the other was the bull's bones wrapped in "glistening fat" (an inedible offering hidden in an attractive exterior). Zeus chose the latter and, angered by his choice, decided to punish humans by taking fire away from them. Prometheus felt compassion for humanity and defied Zeus by stealing fire from Olympus to give it back to mankind (*hubris*: the initial offense). This fire represented knowledge, technology, and the foundations of civilization, essential tools for human advancement and progress. As a *nemesis* for his daring act, Zeus sentenced Prometheus to eternal punishment. He was bound to a rock, where each day an eagle, symbolizing Zeus's anger, would eat his liver, which would regenerate overnight (being the only organ capable of regeneration), ensuring his suffering was endlessly repeated (*tisis*: vengeance). Yet, Prometheus's story does not conclude with perpetual punishment. Heracles (Ἡρακλῆς), the Greek divine hero, eventually intervened. On his journey to complete the Twelve Labors, Heracles encountered Prometheus, who was still bound to the rock. Moved by his suffering and recognizing the injustice of his eternal punishment, Heracles, with Zeus' permission, killed the eagle and freed Prometheus from his chains, granting him salvation. This myth serves as a powerful metaphor for the ongoing battle against cardiometabolic diseases, which are the focus of this doctoral thesis. Just as Prometheus was punished for his boldness in restoring fire, a tool for survival and progress, our quest to manage cardiometabolic diseases mirrors humanity's struggle to harness new knowledge for the improvement of health. In cardiometabolic medicine, the "fire" comprises an analogy for the ability to predict, prevent, and manage these interconnected diseases. Yet, much like Prometheus's eternal punishment, the complexity of cardiometabolic disorders continues to challenge us (and, in some cases, literally destroy the liver), requiring a continuous pursuit of innovation and a holistic approach to overcoming the cardiometabolic pandemic. Heracles' intervention in the story symbolizes the potential for breakthroughs in medicine. Just as he freed Prometheus, modern advancements and innovations, whether in preventive care, data science in medicine, or pharmacotherapy, offer the hope of liberation from the burden of cardiometabolic diseases. With continued research and innovation, we, too, may break free from the chains of fragmented care, offering holistic, *evidence-informed* solutions to prevent and manage these chronic conditions.

Introduction

From defective insulin action to Syndrome X to Metabolic syndrome – a historical perspective

In 1936, Harold Percival Himsworth first suggested that human disease could be secondary to a defect in insulin action (1). In 1977 and 1978, Gerald B. Phillips introduced the concept of a glucose-insulin-lipid defect in an attempt to describe various risk factors for myocardial infarction. These factors included glucose intolerance, hyperinsulinemia, hypercholesterolemia, hypertriglyceridemia, and hypertension, which tend to occur together as a "constellation of abnormalities." These abnormalities were also linked to aging, obesity, and other cardiovascular diseases (CVDs) (2). Phillips proposed that a common underlying factor was responsible for these associations and suggested that the identification of this factor could help prevent CVDs. Phillips speculated that an alteration in the sex hormone milieu is the major predisposing factor for myocardial infarction. Approximately 50 years after Himsworth's suggestion of insulin resistance as a driver of human disease, Gerald Reaven first formalized in 1988 the concept of metabolic syndrome (MetS) (3). Initially, "Syndrome X" concerned a spectrum of impairments that are of enormous importance in the genesis of coronary artery disease, in 1988, with insulin resistance being well-established as the main driver (3). Reaven later highlighted that insulin concentration, abdominal girth, and body mass index (BMI) together, but obesity in particular (estimated by BMI), increase the likelihood that a person will develop the clinical syndromes associated with the defect in insulin action (4). Nonetheless, Reaven questioned the pedagogic and clinical utility of MetS and suggested that the clinical emphasis should be on treating effectively any underlying CVD risk factor (4). Table 1 provides an overview of the MetS risk factors and cut-off scores.

Table 1. Metabolic syndrome risk factors and cut-offs according to the United States National Cholesterol Education Program's Adult Treatment Panel III (NCEP-ATPIII) (5).

Risk factors	Cut-off scores
(1) Impaired glucose tolerance	Fasting plasma glucose ≥100 mg/dl (5.6 mmol/l)
(2) Abdominal obesity	Waist circumference ≥102 cm in men ≥88 cm in women
(3) Hypertriglyceridemia	≥150 mg/dl (1.7 mmol/l) or drug treatment for high triglycerides
(4) Reduced high-density lipoprotein	<40 mg/dl (1 mmol/l) in men <50 mg/dl (1.3 mmol/l) in women or drug treatment for low high-density lipoprotein
(5) Hypertension	≥130/85 mmHg or drug treatment for hypertension

Paradigm shift – a need for a holistic approach to a cardiometabolic pandemic

Even though obesity was officially recognized as a chronic disease by the American Medical Association in 2013, the pathological adipose tissue function, also known as "adiposopathy," was only recently hypothesized to comprise the paradigm of cardiometabolic risk (6). More importantly, despite MetS having been linked to multiple adverse outcomes driven by an increased cardiometabolic risk for over 80 years, in 2023, a new comprehensive care model was proposed by the American Heart Association (AHA): cardiovascular–kidney–metabolic health. The aim of cardiovascular–kidney–metabolic health was to emphasize a holistic management approach to MetS, type 2 diabetes (T2D), chronic kidney disease (CKD), and CVD due to excess and/or dysfunctional adiposity (7). The defining feature of adiposopathy is the accumulation of visceral fat, which is an independent risk marker of cardiovascular and metabolic morbidity and mortality, as well as ectopic fat deposition (8, 9). Specifically, ectopic fat deposition sheds light on the full pathogenic potential of dysfunctional adipose tissue, offering a more comprehensive view of its critical interactions with other body systems (8). All the above risk factors are driving an ever-growing pandemic of cardiometabolic diseases, including but not limited to obesity, MetS, T2D, steatotic liver disease (metabolic dysfunction-associated steatotic liver disease [MASLD] or metabolic dysfunction-associated steatohepatitis [MASH]), CVD, and CKD. The prevalence of these conditions is alarming, with one in four adults in Western countries having obesity and the same number of individuals estimated to have MASLD. Similarly, global MetS and T2D prevalence is 34% and 10%, respectively, with many undiagnosed cases still under the radar (10). Such increases in prevalence contribute to major global health challenges, and cardiometabolic diseases are some of the world's leading causes of disability and mortality, accounting for 3.67 million (for MASLD) up to 428.32 million (for CVD) disability-adjusted life years (DALYs) (Figure 1) [https://vizhub.healthdata.org/gbd-results/, (11)]. Table 2 showcases key health quantification metrics and how DALYs are calculated.

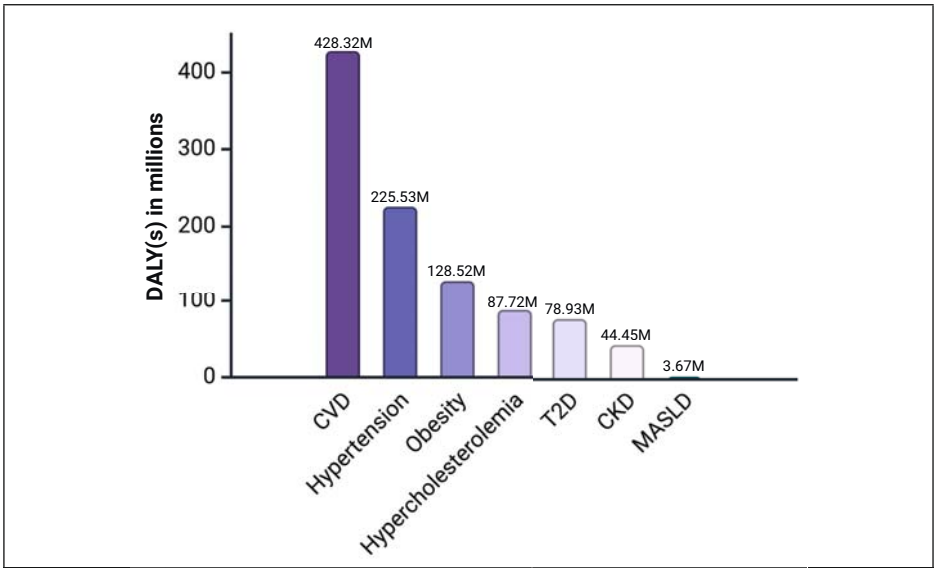


Figure 1. Disability-adjusted life years (DALYs) due to cardiometabolic diseases retrieved from the Global Burden of Disease, Injuries, and Risk Factors Study (GBD) Study 2021 [https://vizhub.healthdata.org/gbd-results/, (11)].
Abbreviations: M, million; CVD, cardiovascular disease; DALYs, disability-adjusted life years; T2D, type 2 diabetes; CKD, chronic kidney disease; MASLD, metabolic dysfunction-associated steatotic liver disease. Created with BioRender.com.

Table 2. Key health quantification metrics (15).

Term	Definition	Calculation
Years lived with disability (YLDs)	Years lived with disability, illness, or injury. One YLD represents the equivalent of one full year of healthy life lost due to disability or ill-health. YLDs capture health loss by considering both the severity and duration of a nonfatal condition.	YLDs are calculated by multiplying the prevalence of a condition by a disability weight, which reflects the severity of the condition on a scale from 0 (perfect health) to 1 (equivalent to death).
Years of life lost (YLLs)	Premature mortality that considers both the frequency of deaths and the age at which it occurs. One YLL represents the loss of one year of life. YLLs measure the difference between an individual's age at death and their life expectancy at the time of their death.	YLLs are calculated by multiplying the number of deaths at each age by the standard life expectancy for that age group.
Disability-adjusted life-years (DALYs)	The sum of the years of life lost to due to premature mortality (YLLs) and the years lived with a disability (YLDs) due to prevalent cases of the disease or health condition in a population.	DALYs = YLLs + YLDs

Of note, cardiometabolic diseases rise in parallel to the degree of westernization of populations, signifying the value of lifestyle and socioeconomic status in disease development (12). On top of that, cardiometabolic diseases are interconnected, sharing common underlying pathophysiology and risk factors. Therefore, when clustered together, cardiometabolic conditions exponentially increase the likelihood of developing additional (cardiometabolic) diseases. This necessitates a multidisciplinary and coordinated approach expected to improve the currently segregated care path of different experts working independently. For instance, in the case of MASLD, starting with simple steatosis, the accumulation of (ectopic) fat within the liver accelerates hepatocellular inflammation and hepatocyte damage (ballooning), proceeding to MASH, which is present in ~20% of individuals with MASLD (13, 14). The prevalence of MASLD in metabolic surgery candidates with severe obesity is >90%, in patients with CVD ~70%, and in patients with T2D >55% (14, 16). Notably, cardiometabolic diseases often remain "silent" for a long period of time, owing to their chronic nature, before (serious) events occur, due to late complications of one or the accumulation of multiple diseases. Almost all major organ systems are affected as a consequence of cardiometabolic diseases, with associated clinical challenges including kidney failure (17), premature cognitive decline (18, 19), MASLD (20, 21), and increased risk for cancer (22, 23). Therefore, it is of the utmost importance that clinicians, researchers, and especially patients realize the cardiometabolic timeline and embrace early preventive measures and screening initiatives (Figure 2).

In Phillips's words (1978), *"The solution to the problem of coronary heart disease is prevention."* Reflecting on this, we need to identify gaps in the current approach to managing cardiometabolic diseases in order to provide *evidence-informed* solutions.

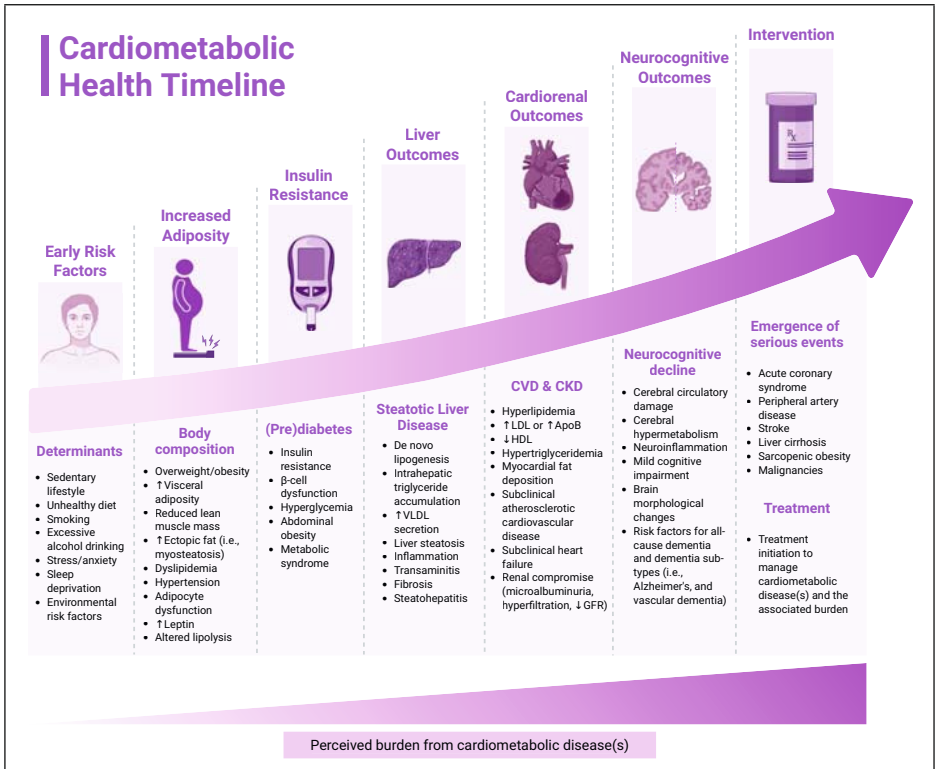


Figure 2. Cardiometabolic Health Timeline, depicting the progression of cardiometabolic disease across key life stages and illustrating the increasing burden of disease over time (7).
The timeline begins in young adulthood, where behaviors such as sedentary lifestyle, unhealthy diet, smoking, alcohol consumption, and environmental risk factors (i.e., air, noise, and light pollution) contribute to increased adiposity and body composition changes. This includes overweight/obesity, visceral fat accumulation, and ectopic fat deposition. These changes promote insulin resistance and the onset of (pre)diabetes, characterized by hyperglycemia, β-cell dysfunction, and metabolic syndrome. As the disease progresses, liver outcomes such as metabolic dysfunction-associated steatotic liver disease, metabolic dysfunction-associated steatohepatitis, and fibrosis emerge, driven by increased hepatic lipogenesis and triglyceride accumulation. Cardiorenal outcomes due to, among other factors, long-standing hyperlipidemia and myocardial fat deposition include subclinical cardiovascular disease and chronic kidney disease. Subsequently, or sometimes concurrently, neurocognitive decline may occur, increasing the risk for dementia and brain morphological changes. In the final stage, adverse outcomes such as acute coronary syndrome, stroke, liver cirrhosis, and obesity-related malignancies develop, requiring clinical intervention. The arrow indicates the increasing perceived burden of cardiometabolic diseases both by healthcare specialists and patients, emphasizing the escalating health impact over time and across organ systems. Abbreviations: ApoB, apolipoprotein B; CVD, cardiovascular disease; CKD, chronic kidney disease; GFR, glomerular filtration rate; HDL, high-density lipoprotein; LDL, low-density lipoprotein; VLDL, very low-density lipoprotein. Created with BioRender.com.

At a glance – key gaps in current cardiometabolic risk stratification and management addressed in this thesis

- **T2D risk assessment should be scalable and validated in diverse ethnic groups**

One issue with the current deployment of prevention strategies is that many individuals at high risk are unaware of this risk. Even more so, people who may have already developed chronic conditions are often also unaware of their disease(s) until they develop advanced complications or present with clear symptoms to their general practitioner. According to the International Diabetes Federation, one in two adults living with T2D is currently undiagnosed (https://diabetesatlas.org/). Given the high rate of undiagnosed individuals with T2D, as well as the increased risk of developing other cardiometabolic diseases due to insulin resistance, as discussed previously, the diagnosis and prediction of future T2D are crucial. Notably, three out of four individuals with T2D live in low- and middle-income countries, where population-wide screening strategies based on laboratory tests may be unfeasible [https://diabetesatlas.org/, (24)]. To address this issue, a lifestyle- or questionnaire-based prediction may inform interventions for lifestyle modification and early treatment initiation, thereby preventing complications and enhancing quality of life. Although the clinical value of non-laboratory incident T2D prediction tools is well established, they lack extensive validation in a wide variety of ethnicities. Prediction models are often created based on White populations without being validated on other ethnic groups, and thus, novel models are needed for these groups for a more accurate and scalable T2D screening strategy (16, 25-27).

- **CVD risk assessment needs a holistic approach, with a focus on cardiometabolic health and several ethnic populations**

Even though obesity/visceral adiposity are central components of the cardiometabolic continuum, the so far fragmented approach to CVD risk assessment seems to hold clinicians, researchers, and patients hostage. The most prominent examples are the European Society of Cardiology (ESC) guidelines for primary prevention of CVD, i.e., ESC Systematic COronary Risk Evaluation (SCORE) risk charts, which do not consider MetS (as a whole) in CVD risk estimation or BMI and/or waist circumference as risk modifiers (28). Likewise, the American Heart Association/American College of Cardiology (AHA/ACC) pooled cohort equations also do not include MetS as a whole or obesity and may not be generalizable to individuals of other ethnicity groups who were not included in the derivation cohort (29); however, the AHA Predicting Risk of CVD EVENTs (PREVENT) equations were very recently developed, specifically, to improve CVD risk prediction in cardiovascular–kidney–metabolic subgroups (i.e., obesity, T2D, and CKD) (30). The development of PREVENT equations signifies the necessity (for the so far unmet

clinical need) of incorporating predictors for kidney and metabolic health, as well as validating prediction models across multiple ethnic populations, reinforcing a comprehensive approach to screening, risk assessment, and prevention of CVD in patients with or at risk for cardiometabolic conditions. Additionally, despite the growing ethnic diversity in Europe, the existing prediction models have generally not been adequately validated for ethnic minority populations within European patient cohorts. These groups often exhibit a higher prevalence of traditional CVD risk factors, variable responses to treatment, increased CVD-related morbidity, higher rates of all-cause mortality, and a lower overall quality of life (31-35). Therefore, there is a pressing need for reliable, scalable prognostic tools that do not rely on physical exams or blood chemistry data, are scalable and cost-effective, and have been validated across large, multi-ethnic populations (35).

- **Coupling MASLD/MASH non-invasive identification with multiple adverse outcome screening**

Another largely unmet clinical need concerns MASLD/MASH, which has a complicated and largely unknown pathophysiology and, most importantly, ineffective screening measures. The current gold standard diagnostic for MASLD/MASH is liver biopsy, an invasive and complicated procedure unsuitable for screening or prognostic purposes (36). However, given the high prevalence of MASLD/MASH (~30%), performing liver biopsies in one-third of the Dutch population would be associated with an estimated incidence of approximately half a million major bleeding events (37, 38). Even though a gold standard diagnostic or prognostic non-invasive test (NIT) remains elusive for this disease, the realization that MASLD is the new member of the cardiometabolic block and the possibility to prescribe pharmacotherapies for MASH without the need for a liver biopsy will soon shift MASLD and MASH identification from liver excellence centers to endocrinology clinics and potentially to primary care (16, 39-42). Notably, MASLD/MASH significantly increases cardiometabolic disease risk, among other diseases, and its identification through NITs is becoming increasingly popular and internationally endorsed by practice guidelines. Considering the shared risk factors between MASLD and other cardiometabolic diseases, a logical next step is to fine-tune existing NITs and repurpose them for a multipurpose screening (of several adverse outcomes at once).

- **Interconnectivity between cardiometabolic risk factors and neurocognitive outcomes**

Numerous studies have demonstrated a complex connection between cardiometabolic health and neurocognitive decline (43). According to the 2024 report of the Lancet Commission on dementia prevention, intervention, and care, 45% of dementia cases could potentially be prevented, with 12% linked to factors such as elevated low-density lipoprotein cholesterol, T2D, hypertension, and

midlife obesity (44). Each of these conditions individually, as well as the presence of multiple cardiometabolic disorders (synergistically), has been associated with widespread structural brain changes and an increased risk of dementia as these conditions accumulate (45, 46). Nonetheless, the neuroimaging and proteomic signatures of brain age and Alzheimer's disease-like atrophy in individuals with overweight and obesity without diagnosed cognitive impairment have not been studied.

- **Real-world evidence and comparative effectiveness of cardiometabolic drugs in neurocognitive outcomes**

The final unmet clinical gap addressed in this thesis is the management of cardiometabolic diseases (41). Specifically, the key driver of the cardiometabolic pandemic, obesity, faces a lack of effective management interventions, with lifestyle interventions achieving a maximum of 7% weight loss reported in randomized controlled trials (RCTs) with at least five years of follow-up (47). Obesity pharmacotherapy has also been limited to low weight-loss percentages, and only recently, more effective options have emerged, i.e., tirzepatide (Food and Drug Administration-approved for obesity in November 2023), achieving higher percentages of ~21–25% (48). Nonetheless, (i) the long term side-effects of recently approved medications are still not fully known (i.e., effects of anti-obesity medications on muscle and bone mass – due to very limited availability of such data), and (ii) the protective effects of cardiometabolic medication other than the ones they are designed for (i.e., drug class differences regarding effects on neurocognitive outcomes) remain to be fully elucidated. This thesis specifically addresses the latter, which lacks robust real-world comparative effectiveness evidence for the most prescribed antihyperglycemic agents to inform decision-making. Notably, the Standards of Care in Diabetes, published in 2024 by the American Diabetes Association Professional Practice Committee, do not specify pharmacological adjustments for individuals at risk of developing dementia, despite some evidence suggesting favorable dementia outcomes with certain T2D drugs.

Solutions on the horizon – in the absence of *evidence-based* and in the need for *evidence-informed*

So far, evidence-based cardiometabolic medicine can be achieved through comprehensive RCTs validating data-driven ways to stratify risk, diagnose conditions, and test personalized interventions. In the absence of such RCTs, which have led to uncertainty about the effects of diagnostic and prognostic algorithms and therapies, there is an unmet clinical need for personalized and *evidence-informed* cardiometabolic medicine. This approach aims to prospectively validate advanced predictive models and biomarkers for cardiometabolic health and, by

extension, their impact on neurocognitive diseases in a real-world setting, which can be achieved by analyzing electronic health records (EHR) databases.

Thesis aims

Given the considerable progress in understanding the interconnectivity between cardiometabolic diseases and associated burdens, the field of cardiometabolic medicine warrants further investigation. Specifically, by applying data-driven algorithms in EHR databases, the following unmet clinical needs should be addressed in an evidence-informed way:

(i) there is an insufficient development and validation of Machine Learning risk-stratification models using non-invasive or non-laboratory features, allowing for scalability;
(ii) the emerging field of "-omics" is unexplored in improving and/or repurposing existing NITs for cardiometabolic outcomes and beyond. Also, there is a lack of insight into the discriminative/predictive role of -omics on brain aging and atrophy among individuals with cardiometabolic disease;
(iii) moving on to pharmacotherapies, even though large-scale EHR databases exist with extensive follow-up data, there is a lack of comparative effectiveness studies of existing cardiometabolic drug classes on neurocognitive outcomes. These outcomes are particularly relevant since they are not systematically examined in the current clinical trials of cardiometabolic drugs, even though the risk of neurocognitive impairment is increased in the presence of cardiometabolic disease.

With a main focus on T2D, CVD, MASLD/MASH, and obesity, this thesis addresses the above challenges through a multifaceted approach comprising modeling studies, biomarker research, and a real-world EHR-based comparative effectiveness study, ultimately aiming to set the basis for future design of optimal diagnostic tools and systematic evaluation of (emerging) pharmacotherapies. The three aims of this thesis, along with a conceptual framework that compares the contemporary approach to cardiometabolic patients and evidence-informed approaches, are presented in detail in Figure 3.

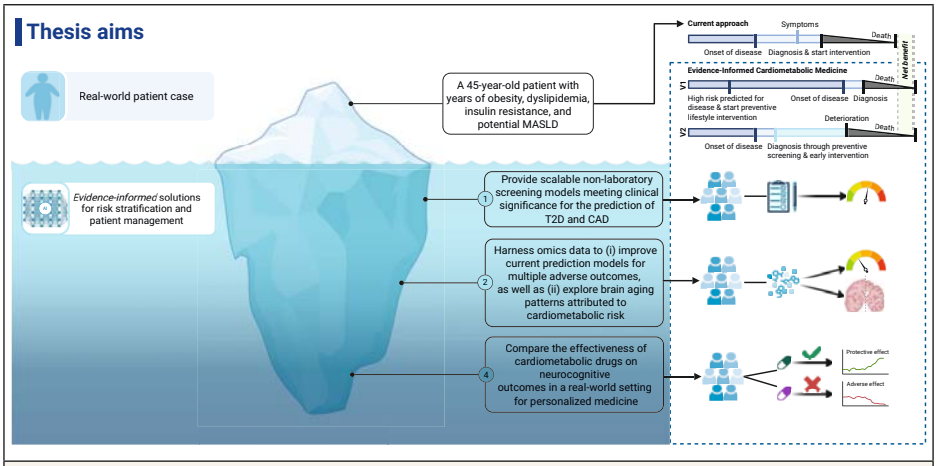


Figure 3. The thesis aims, along with a conceptual comparative model of health outcomes, to examine the differences between traditional and evidence-informed cardiometabolic medicine approaches.
In the current (reactive) approach, patients typically experience symptom onset years after disease onset, followed by diagnosis and later death, with little to no emphasis on preventive screening. In contrast, the evidence-informed approaches incorporate primary (V1) and secondary (V2) preventive screening, leading to earlier intervention initiation and timely diagnosis, respectively, thus extending the duration of health before disease onset, complications, and death. The net benefit of these strategies is indicated by the difference in the duration of healthy life between the two groups. Abbreviations: T2D, type 2 diabetes; CAD, coronary artery disease; MASLD, metabolic dysfunction-associated steatotic liver disease; V, version. Created with BioRender.com.

Thesis Overview

The present thesis comprises three parts, spanning innovative modeling approaches, biomarker/-omics research, and comparative effectiveness research.

Part I: Non-laboratory Predictive Modeling and Scalable Risk Stratification

Here, we focused on developing and validating non-laboratory-based models that are scalable and accurately forecast cardiometabolic risk.

- **Chapters 2 and 3** aimed to develop and externally validate questionnaire-based models in both the UK Biobank (total n = 502,359) and Lifelines (total n = 167,729). These models were designed to predict T2D and CAD using key lifestyle traits and medical history, while achieving performance levels comparable to clinically established models that are expensive to deploy for large-scale population screening due to logistical and financial challenges. This approach has the potential to inform targeted preventive strategies, particularly in resource-limited settings.

Part II: Omics Integration in Risk Stratification and Modeling Approaches
Here, we aimed to integrate omics data to enhance and repurpose traditional risk models for steatotic liver disease and to identify omics signatures associated with brain aging.

- **Chapter 4** sought to validate the use of steatotic liver disease NITs (non-invasive tests) for detecting multiple adverse outcomes in the UK Biobank and proposes significant improvements through the addition of proteomic features. The practical aim of this chapter was to translate the developed model into an online tool, thereby facilitating further research and clinical application.
- **Chapter 5** evaluates Machine Learning-based neuroimaging markers and proteomics of brain age and Alzheimer's disease-like atrophy in 46,288 individuals without diagnosed cognitive impairment and with overweight or obesity. This harmonized analysis of 15 cohorts sought to provide sex-specific evidence on the association between BMI states and brain aging and atrophy patterns, as well as plasma proteins associated with these conditions.

Part III: Real-World Evidence and Comparative Effectiveness of Cardiometabolic Drugs in Neurocognitive Outcomes
Here, we examined how real-world EHR data could inform the comparative effectiveness of specific cardiometabolic therapies on less well-studied outcomes, namely neurocognitive endpoints.

- **Chapter 6** aimed to compare head-to-head the real-world effectiveness of leading cardiometabolic drug classes, i.e., glucagon-like peptide-1 (GLP-1) receptor agonists and sodium-glucose cotransporter-2 (SGLT2) inhibitors on neurocognitive outcomes among individuals with T2D. For this study, we used one of the largest EHR networks available, with data from >3 million individuals with T2D.

Summary and Discussion

- **Chapter 7** synthesizes the main findings of this thesis and discusses their implications for clinical practice. It also outlines the remaining unmet clinical needs and proposes several directions for future research.

References

1. Himsworth HP. Diabetes mellitus: its differentiation into insulin-sensitive and insulin-insensitive types. The Lancet. 1936;227(5864):127-30.
2. Phillips GB. Relationship between serum sex hormones and glucose, insulin and lipid abnormalities in men with myocardial infarction. Proc Natl Acad Sci U S A. 1977;74(4):1729-33.
3. Reaven GM. Banting lecture 1988. Role of insulin resistance in human disease. Diabetes. 1988;37(12):1595-607.
4. Reaven GM. The metabolic syndrome: is this diagnosis necessary? Am J Clin Nutr. 2006;83(6):1237-47.
5. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the Metabolic Syndrome. Circulation. 2009;120(16):1640-5.
6. Bays HE. Adiposopathy is “sick fat” a cardiovascular disease? J Am Coll Cardiol. 2011;57(25):2461-73.
7. Ndumele CE, Neeland IJ, Tuttle KR, Chow SL, Mathew RO, Khan SS, et al. A Synopsis of the Evidence for the Science and Clinical Management of Cardiovascular-Kidney-Metabolic (CKM) Syndrome: A Scientific Statement From the American Heart Association. Circulation. 2023;148(20):1636-64.
8. Carbone F. CardioMetabolic medicine, one more last step forward. Eur Heart J. 2022;43(20):1895-6.
9. Neeland IJ, Ross R, Després JP, Matsuzawa Y, Yamashita S, Shai I, et al. Visceral and ectopic fat, atherosclerosis, and cardiometabolic disease: a position statement. Lancet Diabetes Endocrinol. 2019;7(9):715-25.
10. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2023;402(10397):203-34.
11. Zhang H, Zhou X-D, Shapiro MD, Lip GYH, Tilg H, Valenti L, et al. Global burden of metabolic diseases, 1990–2021. Metabolism - Clinical and Experimental. 2024;160.
12. Ferrari AJ, Santomauro DF, Aali A, Abate YH, Abbafati C, Abbastabar H, et al. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet. 2024;403(10440):2133-61.
13. Fox A, Feng W, Asal V. What is driving global obesity trends? Globalization or “modernization”? Global Health. 2019;15(1):32.
14. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology. 2018;67(1):328-57.
15. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology. 2016;64(1):73-84.
16. Muzurović E, Peng CC, Belanger MJ, Sanoudou D, Mikhailidis DP, Mantzoros CS. Nonalcoholic Fatty Liver Disease and Cardiovascular Disease: a Review of Shared Cardiometabolic Risk Factors. Hypertension. 2022;79(7):1319-26.
17. Thomas G, Sehgal AR, Kashyap SR, Srinivas TR, Kirwan JP, Navaneethan SD. Metabolic syndrome and kidney disease: a systematic review and meta-analysis. Clin J Am Soc Nephrol. 2011;6(10):2364-73.

18. Ng TP, Feng L, Nyunt MS, Feng L, Gao Q, Lim ML, et al. Metabolic Syndrome and the Risk of Mild Cognitive Impairment and Progression to Dementia: Follow-up of the Singapore Longitudinal Ageing Study Cohort. JAMA Neurol. 2016;73(4):456-63.
19. Ben Assayag E, Eldor R, Korczyn AD, Kliper E, Shenhar-Tsarfaty S, Tene O, et al. Type 2 Diabetes Mellitus and Impaired Renal Function Are Associated With Brain Alterations and Poststroke Cognitive Decline. Stroke. 2017;48(9):2368-74.
20. Yki-Järvinen H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. Lancet Diabetes Endocrinol. 2014;2(11):901-10.
21. Kokkorakis M, Boutari C, Katsiki N, Mantzoros CS. From non-alcoholic fatty liver disease (NAFLD) to steatotic liver disease (SLD): an ongoing journey towards refining the terminology for this prevalent metabolic condition and unmet clinical need. Metabolism. 2023;147:155664.
22. Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, et al. Diabetes and cancer: a consensus report. Diabetes Care. 2010;33(7):1674-85.
23. Esposito K, Chiodini P, Colao A, Lenzi A, Giugliano D. Metabolic syndrome and risk of cancer: a systematic review and meta-analysis. Diabetes Care. 2012;35(11):2402-11.
24. Goff LM, Ladwa M, Hakim O, Bello O. Ethnic distinctions in the pathophysiology of type 2 diabetes: a focus on black African-Caribbean populations. Proc Nutr Soc. 2020;79(2):184-93.
25. Chen L, Magliano DJ, Balkau B, Colagiuri S, Zimmet PZ, Tonkin AM, et al. AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures. Med J Aust. 2010;192(4):197-202.
26. Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care. 2003;26(3):725-31.
27. Davies MJ, Aroda VR, Collins BS, Gabbay RA, Green J, Maruthur NM, et al. Management of Hyperglycemia in Type 2 Diabetes, 2022. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care. 2022;45(11):2753-86.
28. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J. 2021;42(25):2439-54.
29. Diaz CL, Shah NS, Lloyd-Jones DM, Khan SS. State of the Nation's Cardiovascular Health and Targeting Health Equity in the United States: A Narrative Review. JAMA Cardiology. 2021;6(8):963-70.
30. Khan SS, Matsushita K, Sang Y, Ballew SH, Grams ME, Surapaneni A, et al. Development and Validation of the American Heart Association's PREVENT Equations. Circulation. 2024;149(6):430-49.
31. Leigh JA, Alvarez M, Rodriguez CJ. Ethnic Minorities and Coronary Heart Disease: an Update and Future Directions. Curr Atheroscler Rep. 2016;18(2):9.
32. Nasir K, Shaw LJ, Liu ST, Weinstein SR, Mosler TR, Flores PR, et al. Ethnic differences in the prognostic value of coronary artery calcification for all-cause mortality. J Am Coll Cardiol. 2007;50(10):953-60.
33. Mital R, Bayne J, Rodriguez F, Ovbiagele B, Bhatt DL, Albert MA. Race and Ethnicity Considerations in Patients With Coronary Artery Disease and Stroke: JACC Focus Seminar 3/9. J Am Coll Cardiol. 2021;78(24):2483-92.
34. Stronks K, Snijder MB, Peters RJ, Prins M, Schene AH, Zwinderman AH. Unravelling the impact of ethnicity on health in Europe: the HELIUS study. BMC Public Health. 2013;13:402.
35. Shroufi A, Chowdhury R, Anchala R, Stevens S, Blanco P, Han T, et al. Cost effective interventions for the prevention of cardiovascular disease in low and middle income countries: a systematic review. BMC Public Health. 2013;13:285.

36. Kouvari M, Valenzuela-Vallejo L, Guatibonza-Garcia V, Polyzos SA, Deng Y, Kokkorakis M, et al. Liver biopsy-based validation, confirmation and comparison of the diagnostic performance of established and novel non-invasive steatotic liver disease indexes: Results from a large multi-center study. Metabolism. 2023;147:155666.
37. Zhou F, Stueck A, McLeod M. Liver biopsy complication rates in patients with non-alcoholic fatty liver disease. Can Liver J. 2022;5(2):106-12.
38. Thampanitchawong P, Piratvisuth T. Liver biopsy:complications and risk factors. World J Gastroenterol. 1999;5(4):301-4.
39. Angelidi AM, Sanoudou D, Hill MA, Mantzoros CS. Management of patients with the cardio renal liver metabolic syndrome: The need for a multidisciplinary approach in research, education and practice. Metabolism. 2024;159:155997.
40. Kokkorakis M, Folkertsma P, Forte JC, Wolffenbuttel BHR, van Dam S, Mantzoros CS. GDF-15 improves the predictive capacity of Steatotic liver disease non-invasive tests for incident morbidity and mortality risk for cardio-renal-metabolic diseases and malignancies. Metabolism. 2024:156047.
41. Kokkorakis M, Boutari C, Hill MA, Kotsis V, Loomba R, Sanyal AJ, et al. Resmetirom, the first approved drug for the management of metabolic dysfunction-associated steatohepatitis: Trials, opportunities, and challenges. Metabolism. 2024;154:155835.
42. Kokkorakis M, Muzurović E, Volčanšek Š, Chakhtoura M, Hill MA, Mikhailidis DP, et al. Steatotic Liver Disease: Pathophysiology and Emerging Pharmacotherapies. Pharmacol Rev. 2024;76(3):454-99.
43. Jia L, Du Y, Chu L, Zhang Z, Li F, Lyu D, et al. Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. Lancet Public Health. 2020;5(12):e661-e71.
44. Livingston G, Huntley J, Liu KY, Costafreda SG, Selbæk G, Alladi S, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. Lancet. 2024;404(10452):572-628.
45. Dove A, Marseglia A, Shang Y, Grande G, Vetrano DL, Laukka EJ, et al. Cardiometabolic multimorbidity accelerates cognitive decline and dementia progression. Alzheimers Dement. 2023;19(3):821-30.
46. Tai XY, Veldsman M, Lyall DM, Littlejohns TJ, Langa KM, Husain M, et al. Cardiometabolic multimorbidity, genetic risk, and dementia: a prospective cohort study. Lancet Healthy Longev. 2022;3(6):e428-e36.
47. Lingvay I, Cohen RV, Roux CWL, Sumithran P. Obesity in adults. Lancet. 2024;404(10456):972-87.
48. Aronne LJ, Sattar N, Horn DB, Bays HE, Wharton S, Lin WY, et al. Continued Treatment With Tirzepatide for Maintenance of Weight Reduction in Adults With Obesity: The SURMOUNT-4 Randomized Clinical Trial. JAMA. 2024;331(1):38-48.


ASCLEPIUS

Part I

Non-laboratory Predictive Modeling and Scalable Risk Stratification

Chapter 2

Effective questionnaire-based prediction models for type 2 diabetes across several ethnicities: a model development and validation study

Michail Kokkorakis, Pytrik Folkertsma, Sipko van Dam, Nicole Sirotin, Shahrad Taheri, Odette Chagoury, Youssef Idaghdour, Robert H. Henning, José Castela Forte, Christos S. Mantzoros, Dylan H. de Vries, Bruce H. R. Wolffenbuttel

EClinicalMedicine 2023;64:102235

Summary

Background Type 2 diabetes disproportionately affects individuals of non-White ethnicity through a complex interaction of multiple factors. Therefore, early disease detection and prediction are essential and require tools that can be deployed on a large scale. We aimed to tackle this problem by developing questionnaire-based prediction models for type 2 diabetes prevalence and incidence for multiple ethnicities.

Methods In this proof of principle analysis, logistic regression models to predict type 2 diabetes prevalence and incidence, using questionnaire-only variables reflecting health state and lifestyle, were trained on the White population of the UK Biobank (n = 472,696 total, aged 37–73 years, data collected 2006–2010) and validated in five other ethnicities (n = 29,811 total) and externally in Lifelines (n = 168,205 total, aged 0–93 years, collected between 2006 and 2013). In total, 631,748 individuals were included for prevalence prediction and 67,083 individuals for the eight-year incidence prediction. Type 2 diabetes prevalence in the UK Biobank ranged between 6% in the White population to 23.3% in the South Asian population, while in Lifelines, the prevalence was 1.9%. Predictive accuracy was evaluated using the area under the receiver operating characteristic curve (AUC), and a detailed sensitivity analysis was conducted to assess potential clinical utility. We compared the questionnaire-only models to models containing physical measurements and biomarkers as well as to clinical non-laboratory type 2 diabetes risk tools and conducted a reclassification analysis.

Findings Our algorithms accurately predicted type 2 diabetes prevalence (AUC = 0.901) and eight-year incidence (AUC = 0.873) in the White UK Biobank population. Both models replicated well in the Lifelines external validation, with AUCs of 0.917 and 0.817 for prevalence and incidence, respectively. Both models performed consistently well across different ethnicities, with AUCs of 0.855–0.894 for prevalence and 0.819–0.883 for incidence. These models generally outperformed two clinically validated non-laboratory tools and correctly reclassified >3,000 additional cases. Model performance improved with the addition of blood biomarkers but not with the addition of physical measurements.

Interpretation Our findings suggest that easy-to-implement, questionnaire-based models could be used to predict prevalent and incident type 2 diabetes with high accuracy across several ethnicities, providing a highly scalable solution for population-wide risk stratification. Future work should determine the effectiveness of these models in identifying undiagnosed type 2 diabetes, validated in cohorts of different populations and ethnic representation.

Funding University Medical Center Groningen.

Research in context

Evidence before this study

Type 2 Diabetes (T2D) is an increasingly prevalent condition affecting more than 462 million individuals worldwide. Disease prevention and early detection are crucial to mitigate potentially life-threatening complications as well as healthcare costs. In this setting, using prediction tools is vital to foster population health, mainly through screening. A comprehensive literature search on PubMed (from January 1, 1996, to August 1, 2023) and Medline (from January 1971 to August 1, 2023) showed that there is a knowledge gap concerning T2D prediction models purely based on easy-to-collect questionnaire features. Besides, there is a lack of thorough validation of models trained on White populations among non-White ethnicities. Questionnaire data reflect lifestyle behaviours and health states that play a cardinal role in T2D. It is also evident that certain ethnicities are affected more than others by T2D, facing an earlier onset of the disease and potentially more complications.

Added value of this study

This proof of principle study demonstrates that models trained on the White population of the UK Biobank achieved clinically relevant performances for prevalence and incidence prediction across five non-White populations, as well as in the Lifelines external validation cohort. Furthermore, in most instances, these models significantly outperformed the concise Finnish Diabetes Risk Score (FINDRISC) and the Australian Type 2 Diabetes Risk Assessment Tool (AUSDRISK), two widely validated non-laboratory-based clinical risk prediction tools. This demonstrates the potential clinical implications of our models in a wide variety of settings, including non-White populations.

Implications of all the available evidence

Deploying these models at a large scale in the primary care setting can be a precise, scalable, and cost-effective means to diagnose positive cases and predict the risk of developing T2D, irrespective of ethnicity. Additionally, resource-limited settings will benefit from using our models by reducing the number of individuals needed to be screened while capturing a significant proportion of the ones developing T2D. To determine the effectiveness of these models in identifying undiagnosed T2D, a follow-up study is required using a cohort where undiagnosed cases can be correctly identified. This effectiveness should be validated in cohorts of different populations and ethnic makeups, as this may vary between these groups.

Introduction

The number of individuals living with type 2 diabetes mellitus (T2D) is rapidly increasing globally, driven by factors such as ageing, urbanisation, sedentarism, and the increasing prevalence of obesity (1, 2). In 2019, diabetes accounted for 66.3 million disability-adjusted life years (DALYs) and 4.2 million deaths among adults worldwide, (3) with disproportionately steep prevalence and complications among non-White ethnic minorities in low-income and middle-income countries (4).

Populations of non-White ethnic backgrounds are disproportionately affected by diabetes, with a three to five times higher prevalence of T2D than people of White-European background (5). South Asians, for instance, usually develop T2D five to ten years earlier and experience a two-to six-fold increased risk of developing T2D compared to White European individuals (6). Likewise, 23% of Black African-Caribbean individuals with T2D are diagnosed under the age of 40 years in comparison to only 9% of White Europeans (7). Among the predominantly Arab population of the Gulf Cooperation Council countries, T2D prevalence has been suggested to be as high as 25%–36% when undiagnosed case estimates are included and occur at a younger age (8). A previous study in the United Arab Emirates showed a prevalence rate of adult T2D and undiagnosed diabetes at 25% and 14.8%, respectively (9). Despite the greater incidence and prevalence of T2D and associated comorbidities in these populations, publicly available diabetic registries and validated prediction models for screening or early diagnosis remain scarce (10). Existing risk prediction tools in these populations have shown only moderate sensitivity and specificity and are not widely used in clinical practice (11). Because of the high rate of undiagnosed diabetics, the prediction of the presence of T2D (prevalence prediction) is essential in the aforementioned settings and highly relevant for lifestyle modification and early treatment initiation to avoid complications and reduced quality of life.

The clinical value of non-laboratory incident T2D prediction tools is well established; however, they lack extensive validation in a wide variety of ethnicities (12, 13). Data science, specifically Machine Learning (ML), has shown high potential to further improve risk stratification across a range of clinical applications, including early disease prediction in diabetes (14). More importantly, ML-based technologies can accommodate population-wide non-invasive screening, allowing for initial assessments and subsequent referrals (15–17). Large population cohorts, such as the UK Biobank and Lifelines, constitute a suitable platform for developing and validating data-driven population risk stratification algorithms. These biobanks comprise rich anthropometric, lifestyle, and medical information data, as well as long-term follow-up on disease outcomes of almost 700,000 individuals in total. Of the UK Biobank participants, circa 82% self-identified as “White” and

almost 18% self-identified as having a different ethnic background, henceforth referred to as "non-White", such as "East Asian or South Asian" ancestry, "Black, African, Caribbean, or other Black" ancestries, "Mixed" ancestries, and "Other" ancestries.

In this context, we aimed to develop ML models to predict the prevalence and an eight-year incidence of T2D that could be easily and widely implemented for population screening across multiple ethnicities. In this proof of principle study, we trained questionnaire-based algorithms on the White population of the UK Biobank and validated them internally within the non-White ethnic groups and externally in Lifelines. One challenge with models trained to predict health outcomes is that they can overfit the data they are trained on. This means that the generated models contain an inherent bias toward the training dataset, which can cause the models to perform poorly in practice. Therefore, we validated our models externally using Lifelines to test whether the produced models perform comparably outside the UK Biobank. Finally, we assessed the algorithms' potential clinical utility against two other ML-based models (containing additional features, i.e., physical measurements and biomarkers) and two gold-standard clinical risk models for the prediction of T2D incidence. Herewith, we showcase significantly enhanced prediction models that can transform primary diabetes care.

Methods

Study setting and participants

The UK Biobank is the largest longitudinal population-based cohort, consisting of 502,507 participants aged between 37 and 73 years old, recruited between 2006 and 2010 (18). For the UK Biobank, ethical procedures are controlled by a dedicated Ethics and Guidance Council (http://www.ukbiobank.ac.uk/ethics). All participants provided written informed consent prior to enrolment. The validation cohort, Lifelines, is a comprehensive and prospective White-European-based population cohort from the northern Netherlands. Lifelines contains data from 168,205 participants aged 0–93 years, with a mean age of 41 years, collected between 2006 and 2013 (19). Similarly, all participants provided written informed consent prior to enrolment. For a complete overview of the collected data, please see https://biobank.ndph.ox.ac.uk/showcase/catalogs.cgi and https://data-catalogue.lifelines.nl/.

Type 2 diabetes classification

In the UK Biobank, T2D diagnoses were assigned based on either self-reported T2D, diabetes diagnosed by a doctor, or T2D hospital record annotation based on the International Classification of Diseases (ICD-9 codes 250.X0, 250.X2, and ICD-10 codes E11.X). Supplementary Table S1A demonstrates the data fields

associated with the age of diagnosis that were employed to calculate the years until diagnosis from the initial assessment. In cases where more than one age of diagnosis was reported, the lowest reported age was used. We then classified all cases diagnosed before their visit to the assessment centre as prevalent cases, while cases diagnosed after their assessment were annotated as incident cases.

In Lifelines, participants were classified as having prevalent or incident T2D based on self-reported T2D (Supplementary Table S1B). Ages of diagnosis were not asked for during follow-up, and T2D follow-up was only asked for some assessments (2A, 3A, and 3B), while general diabetes follow-up was asked for all assessments (1B, 1C, 2A, 3A, and 3B). Therefore, we estimated the age of T2D diagnosis for every incident case by taking the mean of the age the participant had at the assessment reporting a T2D diagnosis and the age at the previous assessment. To calculate more specific ages of T2D diagnosis, if an incident case had reported a general diabetes follow-up diagnosis before their T2D diagnosis, the mean of the age during that assessment and the previous assessment was used instead to determine the age of T2D diagnosis. According to the National Institute for Health and Care Excellence (NICE) guidelines, the diagnosis of T2D is based on glycated haemoglobin (HbA1c) levels ≥48 mmol/mol, fasting plasma glucose levels ≥7 mmol/L, or random plasma glucose levels ≥11.1 mmol/L (20). Unless there are clinical symptoms, these values are not diagnostic of T2D and should be repeated for an individual to be considered as having T2D (20). Both in the UK Biobank and Lifelines, the thresholds for "potentially undiagnosed" T2D encompass a plasma glucose level surpassing 7 mmol/L or an HbA1c level exceeding 48 mmol/mol. We set this specific threshold for plasma glucose at 7 mmol/L due to the lack of specification in the UK Biobank records regarding whether glucose readings of individuals were taken while fasting or were random to prevent false negatives in the range of 7.0–11.1 mmol/L.

Input features

Input features concern the relevant variables used in the modelling procedure of our prediction analyses. Due to the large number of candidate features in the questionnaire, we performed feature selection: we started with an initial list containing all features and sub-selected those with an absolute correlation greater than 0.02 to the target outcome. We then reduced this list to ten features by iteratively extracting the top correlated feature and regressing this feature from the rest of the features. To allow for external validation, we mapped the input features from the UK Biobank to their associated or closest available Lifelines feature (Supplementary Table S2). During feature selection, missing values were imputed using the mean. To investigate whether adding basic measurement and biomarker features improved model performance, we added these features to the

questionnaire feature pool and performed feature selection and model training again.

Data preparation

For the prevalence analyses, everyone with "potentially undiagnosed" T2D was not included in our analysis to avoid bias. This is because, for a T2D diagnosis according to the NICE guidelines, a fasting plasma glucose test above 7 mmol/L, random plasma glucose levels exceeding 11.1 mmol/L, or HbA1c surpassing 48 mmol/mol are not diagnostic of T2D when the individual is asymptomatic and should be repeatedly positive (usually above 7 mmol/L, 11.1 mmol/L, or 48 mmol/mol, at least twice) (20). The participants of both the UK Biobank and Lifelines that surpass the aforementioned values have not repeated the tests for plasma glucose or HbA1c in a timely manner and, therefore, cannot be considered "undiagnosed cases of T2D". Besides, in the UK Biobank, individuals have greatly varying fasting times prior to enrolment, conferring uncertainty as to whether individuals with plasma glucose above 7 mmol/L have "potentially undiagnosed" T2D or did not fast long enough. Therefore, to ensure a clean dataset, these cases needed to be excluded from the analysis. For the incidence analyses, we first removed individuals with "potentially undiagnosed" T2D and anyone diagnosed with T2D by a doctor at baseline. Additionally, we removed all incident T2D cases with more than eight years until diagnosis and all persons not developing T2D but not returning to the assessment centre after eight years. Because the different inclusion criteria result in an under-representation of controls, we corrected the incidence in every ethnicity subset by oversampling the controls to obtain the incidence we observed when including remeasured participants only.

Model training and testing

We set out to predict prevalent and incident T2D across all ethnic groups of the UK Biobank and in Lifelines using questionnaire-based ML models. Self-reported ethnicity was extracted from the UK Biobank, and participants were divided into six different ethnicity groups (Supplementary Table S3). We used Sklearn's Logistic Regression with default settings for model training on the White ethnic population group of the UK Biobank using ten-fold cross-validation (21). The model's performance was internally validated in the five other ethnicity categories of the UK Biobank and externally validated in the independent Lifelines cohort. Even though Lifelines is comprised of 98% White individuals, it is imperative to validate our algorithms externally and show that the models can perform independently of the cohort (Supplementary Table S4). Additionally, since our models were trained on the White population of the UK Biobank, the ethnic makeup of Lifelines makes it an appropriate independent cohort for external model validation. All input features were normalised by fitting Sklearn's StandardScaler on the train set and then using this scaler to scale the features in both the train and test sets.

Moreover, we validated the non-laboratory clinical concise Finnish Diabetes Risk Score (FINDRISC) and the clinical Australian Type 2 Diabetes Risk Assessment Tool (AUSDRISK), which employ 9 and 13 features, respectively, spanning medical history, demographics, lifestyle, and anthropometrics, to predict incident T2D (12, 13).

Statistical analysis and risk stratification

The predictive performance of the models was assessed through the area under the receiver operating characteristic curve (AUC). AUC values and the associated confidence interval (CI) were calculated using DeLong's method from the R pROC package (22). Additionally, AUC values were compared to test for significant differences using the DeLong ROC test from the same package (22). To assess the potential clinical utility of the models across different populations, we took a three-step approach to risk stratification. First, we compared the ability of all models to identify individuals at high risk in the general population (including those with and without diabetes for prevalence and those who did and did not develop diabetes for incidence). Youden's method was used to find the risk threshold yielding the best sensitivity/ specificity balance. In addition to sensitivity and specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and the respective CI were calculated using the R epiR package (23). Then, we simulated another potential application of the incidence models across the different study populations. We stratified the population of every ethnic group into three risk strata based on the individuals' risk of incident T2D (high, medium, and low risk). Each risk stratum contains one-third of the incident T2D cases within each ethnic group. With this analysis, we aim to identify the greatest number of individuals that eventually developed T2D during the follow-up period while minimising the number of people who needed to be screened. Ultimately, to evaluate the improvement in risk prediction provided by our models compared to the abovementioned clinical tools, we conducted a reclassification analysis by calculating the reclassification of events and the categorical Net Reclassification Improvement (NRI) using the R Hmisc package (24). Reclassification analysis is a statistical technique that evaluates the effectiveness of a new diagnostic or predictive test compared to an already established one. This method involves classifying people into different risk categories based on the outcomes of both the new and existing tests. The purpose is to determine whether the new test enhances the accuracy of risk categorisation compared to the existing test. The NRI calculates the difference between the proportion of correctly reclassified individuals into higher-risk categories and those who are correctly reclassified into lower-risk categories; higher NRI values indicate that the new diagnostic model is more accurate at correctly predicting outcomes. Specifically, the NRI is the sum of the percentage of reclassified cases and the percentage of reclassified controls. To ensure fair comparisons between

models, we matched the sizes of the risk groups in the clinical models with our risk groups, which were determined based on the maximum Youden index.

Role of the funding source
The funder had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Results

Baseline characteristics
We set out to predict prevalent and incident T2D across all ethnic groups of the UK Biobank and in Lifelines using questionnaire-based ML models (Fig. 1). The included total group size for prevalent and incident T2D prediction models was 631,748 and 67,083 individuals, respectively. Baseline characteristics of the six ethnicity groups and Lifelines are presented in Table 1. Of note, the prevalence and incidence rates of T2D differed greatly between White and non-White populations, with non-White populations having between two-to almost four-fold higher prevalence (12.2–23.3%) and from half to as high as three-fold higher incidence (1.4–8.2%), than the White population of the UK Biobank (6% and 2.8%, respectively). In contrast, Lifelines had a lower prevalence (1.9%) and incidence (1.8%) of T2D compared to White UK Biobank, partly explained by the age differences between these two populations (Table 1).

Contribution of questionnaire features
The correlation between different questionnaire features pertaining to nutrition, smoking, physical activity, medication, and medical history and prevalent or incident T2D for each population are presented in detail in Supplementary Fig. S2A and B. The contribution of each feature to the prevalence and incidence model is shown in Fig. 2A and B. Both prevalence and incidence models put high importance on Body Mass Index (BMI) and the number of medications taken, positioning them in the top three features of both models. Furthermore, incidence includes a feature representing sedentarism (time spent watching television (TV)). We observe an evident performance saturation with five to six input variables, particularly for prevalence prediction.

Performance of type 2 diabetes prediction models
With ten questionnaire features, the performance of prevalence prediction models measured by their AUC ranged from 0.855 to 0.901 (Fig. 2C and Supplementary Fig. S3A) within the UK Biobank populations and an AUC of 0.917 in the independent validation cohort Lifelines. For models predicting incident T2D in the UK Biobank, AUCs ranged from 0.819 to 0.883 (Fig. 2D and Supplementary Fig. S3B),

while in Lifelines, the AUC was 0.817. The detailed performance metrics of the questionnaire-only models are shown in Supplementary Tables S5A and B.

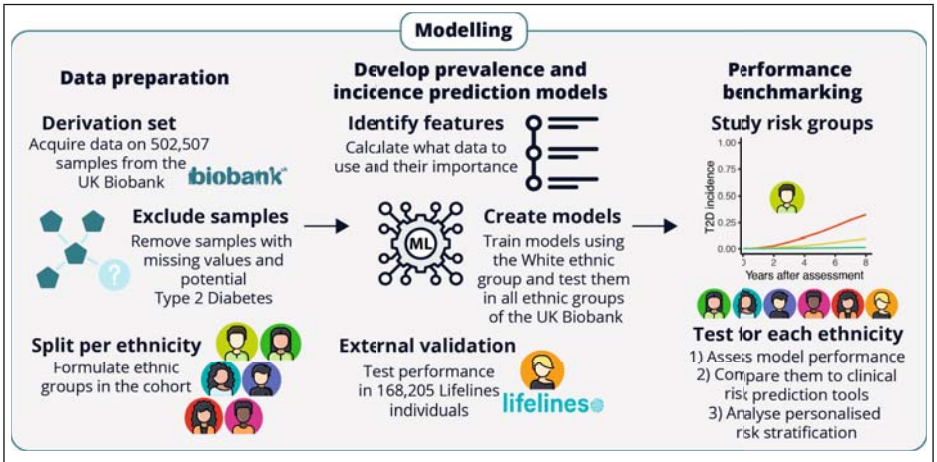


Figure 1. Workflow showing the steps taken to prepare the data and to create questionnaire-based prediction models for prevalent and incident type 2 diabetes.

Table 1. **Baseline characteristics of the internal and external study populations. Data are presented as the mean (SD) unless otherwise noted. The basic measurements are presented as median (IQR).**

	Internal training cohort	Internal testing cohort					External validation cohort
Ethnicities	**White** (n = 472,696)	**South Asian** (n = 8,024)	**Caribbean** (n = 5,137)	**East Asian** (n = 4,263)	**Black** (n = 3,969)	**Other** (n = 8,418)	**Lifelines** (n = 168,205)
Type 2 Diabetes							
Prevalence	6%	23.3%	15.6%	13%	15.9%	12.2%	1.9%
Incidence	2.8%	8.2%	5.6%	6%	1.4%	3.2%	1.8%
Questionnaire features							
Male	45.7%	53.8%	36.9%	45.8%	48.6%	45.7%	42.2%
Female	54.3%	46.2%	63.1%	54.2%	51.4%	54.3%	57.8%
Age	58 (13)	53 (14)	51 (12)	52 (14)	50 (13)	53 (14)	43 (18)
Aspirin	14%	18.5%	12.5%	11.6%	13.8%	13.8%	99.7%
Blood pressure medication	20.6%	27.1%	28.4%	19.3%	30.6%	20.6%	8.4%
Body mass index (BMI)	27.4 (4.8)	27.3 (4.5)	29.2 (5.6)	25.8 (4.3)	29.5 (5.2)	27.8 (5.1)	25.6 (4.6)
Bread intake (slices/week)	10 (10)	8 (8)	6 (6)	8 (10)	8 (10)	10 (9)	22.5 (19.5)
Coronary artery disease before the first assessment	3.3%	5.8%	1.6%	3.2%	1.6%	3.1%	1.5%
Cholesterol lowering medication	17.3%	26.5%	15.1%	18.5%	15.7%	17.6%	4.9%
Dentures	16.9%	9.7%	20.1%	15.5%	10%	14.4%	8.5%
Siblings history of diabetes	7.8%	25.9%	18.6%	18.4%	12.7%	14.6%	1.3%
Parents history of diabetes	16.5%	43.3%	45.8%	34.6%	23.5%	27.2%	6.4%
Number of medications taken	2 (4)	2 (4)	2 (4)	1 (3)	2 (4)	2 (4)	1 (2)
Pack years of smoking (0 years not included)	19.5 (22.5)	15 (15.5)	13.2 (14.2)	15 (16.6)	13.8 (15.1)	16.8 (19.3)	8.5 (13)
Slow walking pace	7.8%	17.7%	12.2%	15.4%	15.6%	15%	N/A

Table 1. **Baseline characteristics of the internal and external study populations. Data are presented as the mean (SD) unless otherwise noted. The basic measurements are presented as median (IQR). (continued)**

	Internal training cohort	Internal testing cohort					External validation cohort
Ethnicities	**White** (n = 472,696)	**South Asian** (n = 8,024)	**Caribbean** (n = 5,137)	**East Asian** (n = 4,263)	**Black** (n = 3,969)	**Other** (n = 8,418)	**Lifelines** (n = 168,205)
Time spent watching television (TV) (hours/day)	2.8 (1.7)	2.5 (1.7)	3.3 (2.2)	2.4 (1.8)	2.7 (2.1)	2.5 (1.9)	2.3 (1.6)
Unable to work because of sickness or disability	3.9%	7%	7.1%	4%	5.7%	6.9%	2.9%
Gained weight in the past year	27.8%	29.6%	39.7%	28%	36.1%	30.9%	20.9%
Lost weight in the past year	15%	15.3%	19.3%	14%	18%	17%	20.9%
Alcohol intake frequency	3 (2)	5 (3)	4 (2)	5 (3)	5 (3)	4 (2)	3 (1)
Daily or almost daily	21%	6.6%	9.2%	8.8%	4.9%	10.7%	10.9%
Three or four times a week	23.8%	8.1%	12.3%	9.8%	7.3%	12.6%	10.3%
Once or twice a week	26.3%	13.2%	23.3%	17%	16%	18.6%	39%
One to three times a month	11.2%	7.2%	15.8%	10.3%	10.5%	11%	19.5%
Special occasions only	10.9%	16.7%	26.1%	26.9%	28.9%	22%	N/A
Never	6.8%	47.6%	13.1%	27.1%	31.9%	23%	20.3%
Glucosamine intake	19.3%	11.6%	14.7%	17%	11.1%	14.1%	N/A
Basic measurements							
Seated height (cm)	138 (9)	133 (13)	135 (13)	134 (13)	134 (13)	136 (13)	NA (NA)
Waist circumference (cm)	90 (19)	92 (15)	91 (17)	86 (17)	93 (16)	90 (18)	88.5 (17)
Mean heart rate (bpm)	58 (10)	59 (9)	60 (8)	58 (9)	60 (9)	59 (9)	72 (11)
Mean diastolic blood pressure (mmHg)	81 (14)	82 (14)	84 (14)	81 (14)	84 (14)	81 (14)	72 (13)

Table 1. **Baseline characteristics of the internal and external study populations. Data are presented as the mean (SD) unless otherwise noted. The basic measurements are presented as median (IQR). (continued)**

Ethnicities	Internal training cohort	Internal testing cohort					External validation cohort
	White (n = 472,696)	South Asian (n = 8,024)	Caribbean (n = 5,137)	East Asian (n = 4,263)	Black (n = 3,969)	Other (n = 8,418)	Lifelines (n = 168,205)
Biomarker features							
Total cholesterol (mmol/L)	5.71 (1.14)	5.29 (1.12)	5.34 (1.09)	5.52 (1.12)	5.18 (1.1)	5.51 (1.15)	5.01 (1.02)
Gamma glutamyltransferase (U/L)	37.3 (42.2)	36.7 (39.2)	40.4 (40.3)	34.8 (35.3)	42.8 (42.4)	38.5 (43.3)	26.3 (25.7)
Glucose (mmol/L)	5.1 (1.2)	5.4 (1.9)	5.1 (1.5)	5.2 (1.5)	5.1 (1.5)	5.3 (1.7)	5 (0.8)
Glycated haemoglobin (HbA1c) (mmol/mol)	36 (6.5)	40.8 (10.6)	39.1 (9.5)	38.3 (8.4)	38.9 (10.1)	38 (9.3)	37.1 (4.9)
High light scatter reticulocyte (%)	0.4 (0.3)	0.4 (0.2)	0.5 (0.2)	0.4 (0.2)	0.5 (0.3)	0.4 (0.3)	N/A
Neutrophil count (10^9 cells/L)	4 (1.7)	4.2 (1.7)	3.1 (1.8)	3.9 (1.7)	2.8 (1.4)	3.9 (1.8)	3.1 (1.4)

Additionally, we performed an exploratory analysis of the potential added benefit of two other types of models: one including basic physical measurements and one including blood biomarkers (Supplementary Figs. S4A, S4B, S5A, S5B, S7A, S7B, S8A, and S8B). For prevalence prediction, including basic measurements significantly improved the performance of questionnaire-only models for all UK Biobank populations, except for Other, yet lowered the AUC of Lifelines (Supplementary Table S8A, Supplementary Fig. S10). In contrast, for incidence prediction, adding basic measurements significantly increased the performance of only two populations, UK Biobank White and Lifelines, though all populations showed higher AUCs. Including biomarkers led to a significant improvement in all instances except for incidence prediction among the Black population, where the Questionnaire-only models seem to yield a marginally higher performance (Supplementary Fig. S10 and Supplementary Tables S8A, S8B). The feature importance of these models is shown in Supplementary Figs. S4A, S4B, S7A, and S7B.

Comparison with non-laboratory clinical risk models

We also compared the questionnaire-only models to two clinically validated non-laboratory risk scores. First, we tested the performance of the concise FINDRISC, developed as a simple screening tool for individuals at high risk of developing T2D. We observed that the questionnaire-based models significantly outperformed FINDRISC for prevalence prediction in all populations, and they significantly outperformed FINDRISC in four out of seven populations for predicting incidence (Fig. 2C and D, and Supplementary Tables S9A, S9B). Similarly, the questionnaire-based models significantly outperformed the AUSDRISK models in all prevalence predictions as well as in three out of seven populations for incidence prediction (Fig. 2C and D, and Supplementary Tables S9A, S9B). In all other instances, there were no significant differences; however, our models yielded overall higher AUCs.

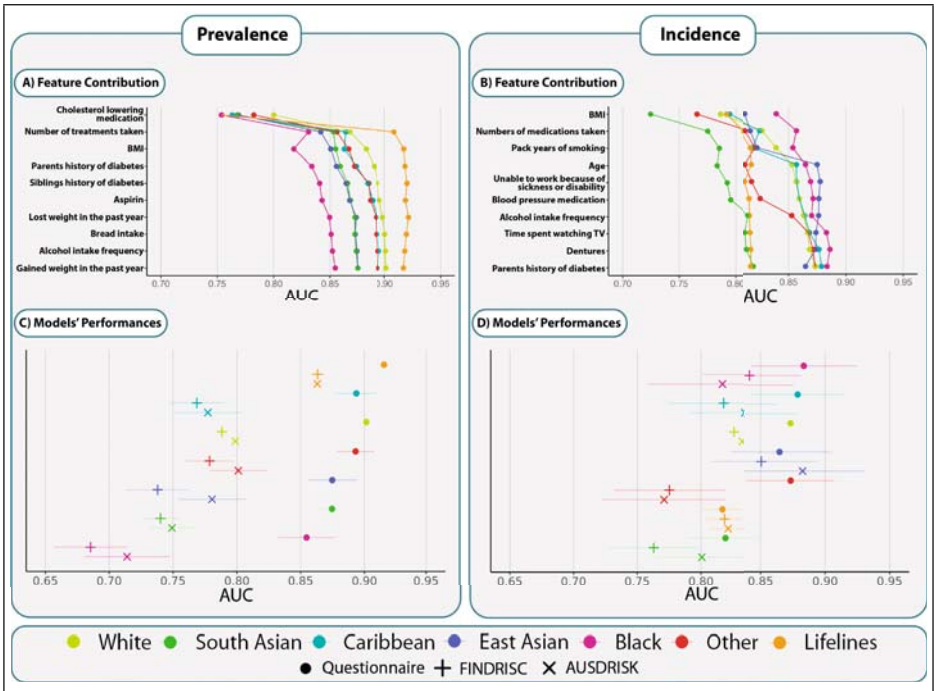


Figure 2. Feature contribution and performance of type 2 diabetes prediction models for prevalence and incidence.
A list of predicting features included in our models for prevalence (A) and incidence (B) prediction and their contribution to the models' performance is presented. Below, the performance of different models across populations for prevalence (C) and incidence (D) is shown. Each colour-symbol combination refers to a specific model and population, explained in detail in the bottom panel. The AUC and 95% CI are shown for all models. BMI, body mass index; AUC, area under the receiver operating characteristics; TV, television; FINDRISC, Finnish Diabetes Risk Score; AUSDRISK, Australian type 2 diabetes risk assessment tool; T2D, type 2 diabetes; CI, confidence interval.

Sensitivity analysis and clinical utility of risk stratification

Furthermore, we conducted an in-depth sensitivity analysis of the risk stratification for all models to assess their potential clinical utility (Supplementary Tables S5A, S5B, S6A, S6B, S7A, and S7B). Based on the thresholds provided by the Youden index, the questionnaire-only models obtained very high sensitivity-specificity balance, PPV, and NPV. Both sensitivity and specificity were consistently high (above 74% and 83% for prevalence and 75% and 68% for incidence, respectively) for all populations. The corresponding NPVs for all models were above 93% and 98% for prevalence and incidence, respectively. For the models including biomarkers, further improvement in the sensitivity-specificity balance was seen, with a lower proportion of individuals identified as high risk also translating to higher PPV across the populations for prevalence and incidence. All corresponding NPVs were above 97% and 99% for prevalence and incidence, respectively.

In the second step of the analysis, we separated each population into three risk strata (~33% of the T2D incident cases in each risk stratum) based on the individuals' risk of T2D eight-year incidence. We observed that the questionnaire-only models could identify small groups of very high-risk individuals who eventually developed T2D during the follow-up period (Fig. 3). By screening as little as 0.47% (Black population) to 7.6% (South Asian) of individuals from different ethnic populations (belonging to the high-risk strata), the questionnaire-only models identified 33% of individuals who developed T2D within each ethnic group. In the high-risk strata, the average incidence of T2D was at least ten-fold higher compared to the lowest-risk strata (Fig. 3). The models also identify 66% of all individuals who developed T2D (belonging to the high- and medium-risk strata) while screening only between 11.5% (Caribbean population) to 23.1% (South Asian population) of all individuals across different ethnic populations. These slightly larger high- and medium-risk strata also show at least a six-fold higher risk across all populations compared to lowest-risk population strata. For the two other types of models (with additional physical measurements and the ones with the addition of biomarkers), the highest-risk strata generally showed even higher average incidence despite the similar size (Supplementary Figs. S6 and S9). For all ethnicities, 66% of incident T2D cases (including high-risk and medium-risk individuals) could be identified by screening less than 10% of each ethnic population using the model, including biomarkers.

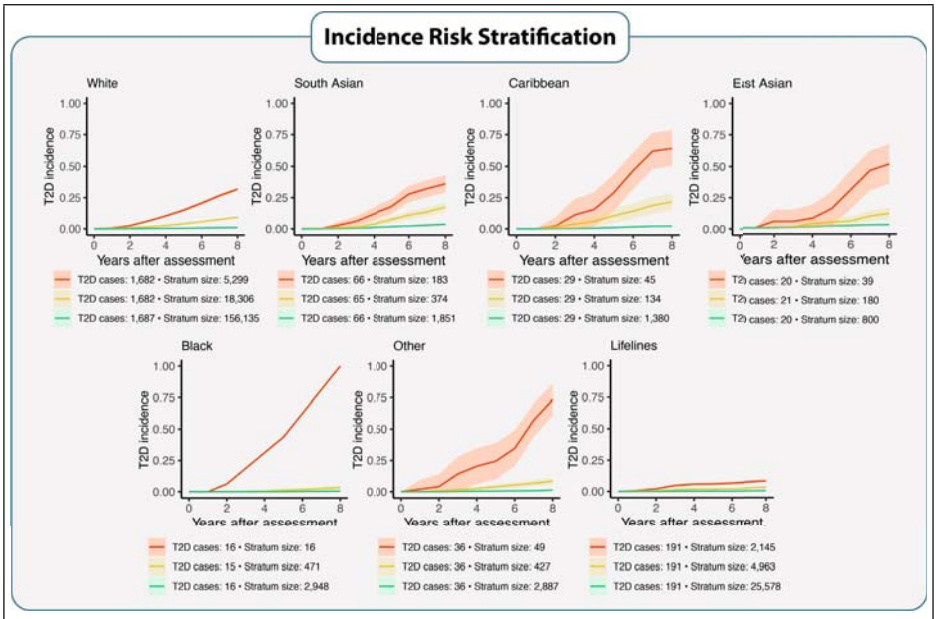


Figure 3. Risk identification for developing T2D per population.
Every population is separated into three risk strata, according to the individuals' risk of incident T2D (high risk = red, medium risk = yellow, low risk = green), with each risk stratum containing 33% of all T2D cases. The x-axis represents the interval of years between the biobank entry and the moment of receiving a diagnosis of 2D. The y-axis represents the incidence of T2D. The stronger-coloured lines represent the average T2D incidence within each risk stratum, and the lighter-coloured bands around the lines show the 95% CI. T2D cases correspond to the total number of T2D incident cases within each risk stratum. Stratum size corresponds to the number of individuals within each risk stratum. Stratum sizes show how many individuals must be screened to identify 33% of all T2D cases within each risk-stratum. T2D, type 2 diabetes; CI, confidence interval.

Reclassification analysis

Ultimately, the reclassification analysis demonstrates that in almost all cases, our models correctly reclassify more cases than the clinically established prediction tools FINDRISC and AUSDRISK (Table 2). Notably, for the White, Caribbean, Other, and South Asian populations, our models correctly reclassify more events reaching statistical significance compared to FINDRISC (Table 2). Compared to AUSDRISK, our models reach statistical significance among the White and Other populations in correctly reclassifying T2D cases, along with statistically significant NRI values (Table 2, Supplementary Table S10A). The addition of physical measurements overall reclassifies more events correctly and seems to perform better in Lifelines compared to the Questionnaire Models (Supplementary Table S10B). The models also including biomarkers, largely outperform the clinical tools by reaching statistical significance in almost all instances (Supplementary Table S10C). The high/low-risk group reclassifications, along with NRIs and reclassification of non-event percentages, are demonstrated in detail in Supplementary Tables 10A–C.

Table 2. Reclassification analysis comparing our questionnaire-based models to FINDRISC and AUSDRISK for incidence prediction.
Reclassification events % correspond to our models' net percentage of reclassified individuals with T2D compared to the clinically established tools. Reclassification of events per 10,000 events corresponds to the net number of T2D cases reclassified when screening 10,000 cases. Positive reclassification events indicate that our models correctly reclassify more cases than the other two clinical tools, whereas negative events indicate the opposite. The reclassification events percentages (%) and reclassification events N per 10,000 are presented along with the 95% CI. FINDRISC, Finnish Diabetes Risk Score; AUSDRISK, Australian type 2 diabetes risk assessment tool; T2D, type 2 diabetes; CI, confidence interval.

Risk model	Population	Reclassification events %	Reclassification events N per 10,000	P-value
FINDRISC	White	6.4 (5.2 – 7.6)	637 (519 – 756)	<0.001
FINDRISC	Black	2.2 (-5.2 – 9.5)	217 (-518 – 953)	0.6
FINDRISC	Caribbean	12.6 (3.7 – 21.5)	1,264 (374 – 2,154)	0.005
FINDRISC	East Asian	9.8 (-2.8 – 22.4)	984 (-278 – 2,245)	0.1
FINDRISC	Other	14.8 (6.4 – 23.3)	1,481 (637 – 2,326)	<0.001
FINDRISC	South Asian	12.7 (6.1 – 19.3)	1,269 (610 – 1,928)	<0.001
FINDRISC	Lifelines	-2.8 (-6.3 – 0.7)	-279 (-627 – 69)	0.1
AUSDRISK	White	5.9 (4.4 – 7.4)	591 (441 – 741)	<0.001
AUSDRISK	Black	3.4 (-8.2 – 15.1)	345 (-819 – 1,509)	0.6
AUSDRISK	Caribbean	5.7 (-3.9 – 15.3)	571 (-389 – 1,532)	0.2
AUSDRISK	East Asian	0 (-16.6 – 16.6)	0 (-1,656 – 1,656)	1
AUSDRISK	Other	25.6 (14.7 – 36.6)	2,564 (1,472 – 3,656)	<0.001
AUSDRISK	South Asian	7.8 (-0.9 – 16.4)	776 (-91 – 1,642)	0.08
AUSDRISK	Lifelines	0.4 (-3.7 – 4.4)	38 (-365 – 441)	0.9

Discussion

In this study of over 600,000 individuals for prevalence and over 67,000 for incidence prediction, we showed for the first time that questionnaire-based ML models can accurately predict T2D prevalence and eight-year incidence across all ethnicities present within the UK Biobank, as well as the Lifelines external validation cohort. For almost all ethnicities, these models outperformed two established clinically validated T2D risk assessment tools. Despite the performance improvement verified with the addition of blood biomarkers, the questionnaire-only models showed clinical utility for detecting prevalent and incident T2D.

Previous research on the performance of prediction models for incident T2D has shown substantial differences across ethnicities. A re-estimation of the Atherosclerosis Risk in Communities (ARIC) model for the prediction of five-year diabetes risk in the Coronary Artery Risk Development Study in Young Adults (CARDIA) cohort showed significant differences in performance between White

and African Americans (AUC 0.902 vs 0.816) (25). Another study of 12,043 Black and White individuals focusing on T2D prediction using anthropometric features and lipid levels reported an AUC of 0.79 (26). In this study, we observed less variation in the model performances between White and Black individuals for both prevalent and incident T2D prediction. The models developed herein overall outperform what has been previously demonstrated in Black populations, even without glucose as an input feature, and contradict the results of previous analyses that suggested that risk scores trained on the European-descent population are not applicable to other ethnic groups (26, 27). Additionally, our questionnaire-based models significantly outperformed FINDRISC and AUSDRISK across all seven populations for prevalent T2D detection. For incidence, our models outperformed the above-mentioned tools in four populations compared to FINDRISC and three populations compared to AUSDRISK. This is especially relevant since both FINDRISC and AUSDRISK have been shown to perform only moderately well in several non-White populations (28, 29) despite AUSDRISK including ethnicity as an input feature and being intended to be used in the ethnically diverse Australian population (30). As expected, adding blood biomarkers to the models resulted in further improvements in predictive performance with AUCs generally above 0.90, mainly due to high correlations conferred by these features (Supplementary Figs. S7A, S7B, S10). Despite being significant, these improvements in AUC were not substantial enough to unequivocally justify their deployment over the questionnaire-only models considering the practical challenges discussed further in detail below.

As such, the goal of population-level risk stratification is not merely to predict individual risk accurately but to clearly distinguish groups with different levels of risk (31). To assess the potential stratification utility of our models, we first optimised their sensitivity-specificity balance with the Youden index. We found that all models achieved high to very high sensitivity and specificity for both prevalence and incidence prediction across all ethnicities. Given the low prevalence and incidence of T2D in White populations, a high specificity and NPV were expected for the White UK Biobank population and Lifelines. However, specificity and NPV remained high even in other ethnicities with higher prevalence and incidence rates (Supplementary Tables S5A, S5B, S6A, S6B, S7A, and S7B). The main difference with the addition of biomarkers was the increase in PPV, stemming from the lower number of individuals identified as high risk (between 20% and 29% for questionnaire-only predictions and generally around 18% when biomarkers were included). However, we also aimed to assess the usefulness of the models in settings where resources are limited, or population health data is lacking, and where it is essential to accurately identify as many high-risk individuals as possible while minimising the number of screened individuals. In such instances, screening more than a quarter of the population might be prohibitive from a cost and logistics perspective, hampering the model's clinical utility. Herein, we demonstrated that

all models can also be applied to identify smaller groups of individuals at very high risk and that 33% and 66% of all incident diabetes cases can be identified by screening less than 10% and 23% of the population using the questionnaire-only models, respectively. Additionally, by demonstrating high predictive abilities for T2D prevalence, our models will be valuable for early diagnosis, especially in areas where T2D is underdiagnosed and often missed. This is essential for minimising complications and decreased quality of life associated with late T2D diagnosis.

The data from these two simulated scenarios suggests that while there is a benefit from including additional measurements in risk stratification models, questionnaire-only models predict prevalent and incident diabetes with high accuracy and clinical utility. By not being subject to the practical limitations associated with collecting physical measurements or biomarkers, a questionnaire-based tool comprises the first step towards identifying an initial high-risk population that could be referred for subsequent diagnostic or prognostic assessment in a primary care setting. At a sensitivity and specificity as high as 80%, we see that questionnaire-only models applied to the largest population we studied, with almost 180,000 White individuals in the UK Biobank training set for incidence prediction, would recommend follow-up for less than 40,000 individuals based on their eight-year T2D risk, and around 65,000 high-risk individuals with prevalent T2D (Supplementary Tables S5A and S5B). In the context of population health prevention programs, deploying more selective models brings about two advantages. On the one hand, it requires considerably fewer individuals to be screened to detect a substantial portion of high-risk individuals. On the other hand, in line with previous research, it has been shown that such programs are most effective when targeted at a specific outcome, such as T2D risk reduction, and when including high-risk individuals, as opposed to a non-stratified population (32). Based on our reclassification analyses, all models developed herein correctly reclassify predicted T2D cases and, in many instances, outperform the currently available models. Of note, our models have demonstrated significantly better net reclassification improvements and correctly reclassify more events when compared to available clinical tools. Specifically, when compared to FINDRISC, an additional 4,651 positive cases are correctly reclassified using our models per 40,000 events, reaching statistical significance. Likewise, for the comparisons with AUSDRISK, the respective number of positive cases that are correctly and significantly reclassified using our models is 3,155 per 20,000 events.

Eventually, translating the models presented in this proof of principle study into population health risk stratification tools for primary diabetes care is not without challenges. In fact, most digital health innovations fail to advance into clinical practice or fall short of their anticipated impact (33). This lack of adoption is often the result of a poor understanding of end-user needs and an inability to integrate

the solution into current care frameworks (33). We built questionnaire-only models to predict and diagnose T2D with the intent that individuals could complete them, potentially digitally, without requiring invasive biomarker collection or a visit to primary care facilities. While not replacing a trained clinician's evaluation, a patient-centred tool would facilitate timely screening and reach a larger audience by eliminating the need for primary care visits in the first phase. Policymakers have been encouraged to focus on prevention and innovation to enable large-scale diabetes awareness programmes (34). For such initiatives, another possible challenge in applying questionnaire-based models at scale is to ensure that all questions are answered. Therefore, we limited the number of questions included to ten.

Overall, our study has several strengths and certain inherent limitations. First, this study represents the largest hitherto reporting on the performance and potential clinical utility of a questionnaire-based risk stratification model for prevalent and incident T2D in two biobanks and across multiple ethnicities. From a modelling perspective, this minimises the chances of overfitting and provides evidence of the model's validity. Second, we applied strict inclusion and exclusion criteria, thereby minimising the risk of including individuals with undiagnosed T2D. Third, we validated two widely non-laboratory clinical tools, FINDRISC and AUSDRISK, in all ethnic groups of the UK Biobank and externally in Lifelines, which provides a comprehensive benchmark for the performance of our models. On the other hand, as with all self-reported biobank data, ethnicity data may only be partially accurate. Specifically, the self-reported ethnic background can be influenced by individual perceptions, cultural and social factors, and may not always accurately reflect an individual's ancestry and levels of admixture. Additionally, the categories used to describe ethnicity can differ between countries, making it difficult to compare results across studies. Another potential limitation lies in the categorisation of "potentially undiagnosed" T2D. To try to minimise the risk of including individuals who may have clinically high, although not repeatedly, plasma glucose or HbA1c concentrations without confirmed T2D diagnosis, we set the plasma glucose exclusion threshold at above 7 mmol/L and the HbA1c exclusion threshold at above 48 mmol/mol. These thresholds may not be realistic or indicative for "potentially undiagnosed" T2D since plasma glucose values are sometimes obtained in a non-fasted state or may not be reproduced if repeatedly tested. Thus, excluding "potentially undiagnosed" cases of T2D might have impacted the performance of the models presented herein. Besides, for prevalence prediction, in our study, individuals are already aware of their diagnosis, and if the questionnaire models were to be prospectively applied, the answers of individuals knowing they have T2D might be different from those unaware of it (undiagnosed cases). Moreover, these questionnaires were administered as part of a volunteer-led biobank cohort whose participants tend to be relatively healthier or younger individuals,

placing limitations around the age distribution to which they apply and potentially socioeconomic limitations. Lastly, due to the observational nature of this study, we cannot identify causal relationships between the variables included in the models and the predicted outcomes.

In conclusion, questionnaire-based ML models predict prevalent and incident T2D in multiple ethnicities with high accuracy and have the potential to enhance early diagnosis if deployed for population health screening in primary diabetes care. While biomarker-based models achieved enhanced performance, the questionnaire-only models produced significantly high and clinically useful predictions to be considered a valid alternative to these models and the challenges their large-scale deployment can pose. This is particularly important for populations of non-White ethnicity who are disproportionately impacted by T2D and regions with limited resources and access to primary diabetes care. While current prediction models show promise in diagnosing and predicting T2D, further research is needed to determine the effectiveness of these models in identifying undiagnosed type 2 diabetes. Specifically, a follow-up study is required using a cohort where undiagnosed cases can be correctly identified. This effectiveness should be validated in cohorts of different populations and ethnic makeups, as this may vary between these groups.

Contributors

MK, SvD, JCF, DdV, and BHRW conceived and designed the study. MK interpreted the data and analyses, conducted the literature search, made the figures, and wrote the manuscript. PF accessed the data, conducted data cleaning, statistical analyses, made the figures, and wrote the manuscript. SvD contributed to data interpretation and wrote the manuscript. MK and SvD accessed and verified the underlying data and analyses. NS, ST, OT, YI, RHH, and BHRW contributed to writing and reviewing the manuscript. JCF and CSM contributed to advising, writing and reviewing the manuscript and interpreting the analyses. DdV contributed to writing, reviewing, and interpreting the analyses and creating the figures. DdV and BHRW worked in supervisory capacities and contributed equally to the work presented herein. MK has full access to all the data in the study and had final responsibility for the decision to submit for publication.

Data sharing statement

Study data are available from UK Biobank and Lifelines but were used under license for the current study, which restricts their public availability. Data are, however, available from the authors upon reasonable request and when granted permission by the UK Biobank and Lifelines. All code is available and can be requested through the corresponding author.

Declaration of interests

MK, NS, ST, OC, YI, and RHH have no conflict of interest to declare. PF, SvD, JCF, and DdV are employed by Ancora Health B.V. and own shares of Ancora Health B.V. BHRW sits on the medical advisory board of Ancora Health B.V. CSM has been a shareholder of and reports grants through his institution and personal consulting fees from Coherus Inc., AltrixBio, grants through his institution from Merck, and grants through his institution personal consulting fees from Novo Nordisk, reports personal consulting fees and support with research reagents from Ansh Inc., reports personal consulting fees from Genfit, Lumos, Amgen, Corcept, Intercept, 89Bio, AstraZeneca and Regeneron, reports support (educational activity meals at and through his institution) from Amarin, Novo Nordisk and travel support and fees from TMIOA, Elsevier, the California Walnut Commission, College Internationale Research Servier, and the Cardio Metabolic Health Conference; none of which is related to the work presented herein.

Acknowledgements

All authors critically reviewed the manuscript for important intellectual content. The authors thank the UK Biobank data access granted through application 55495 and data access to the Lifelines data through application OV20_00020. The Lifelines Biobank initiative has been made possible by a subsidy from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs,

the University Medical Center Groningen, University Groningen, and the Northern Provinces of The Netherlands. This project was funded by the University Medical Center Groningen under project number: PPP-2019_023. The vectors used for Fig. 1 were retrieved from Flaticon.com.

Supplementary data

This article contains supporting information available online at: https://www.sciencedirect.com/science/article/pii/S2589537023004121?via%3Dihub#mmc1

References

1. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care. 2004;27(5):1047-53.
2. Kokkorakis M, Katsarou A, Katsiki N, Mantzoros CS. Milestones in the journey towards addressing obesity; Past trials and triumphs, recent breakthroughs, and an exciting future in the era of emerging effective medical therapies and integration of effective medical therapies with metabolic surgery. Metabolism. 2023;148:155689.
3. GBD. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1204-22.
4. Gujral UP, Narayan KMV. Diabetes in Normal-Weight Individuals: High Susceptibility in Nonwhite Populations. Diabetes Care. 2019;42(12):2164-6.
5. Goff LM, Ladwa M, Hakim O, Bello O. Ethnic distinctions in the pathophysiology of type 2 diabetes: a focus on black African-Caribbean populations. Proceedings of the Nutrition Society. 2020;79(2):184-93.
6. Banerjee AT, Shah BR. Differences in prevalence of diabetes among immigrants to Canada from South Asian countries. Diabet Med. 2018;35(7):937-43.
7. Paul SK, Owusu Adjah ES, Samanta M, Patel K, Bellary S, Hanif W, et al. Comparison of body mass index at diagnosis of diabetes in a multi-ethnic population: A case-control study with matched non-diabetic controls. Diabetes Obes Metab. 2017;19(7):1014-23.
8. Meo SA, Usmani AM, Qalbani E. Prevalence of type 2 diabetes in the Arab world: impact of GDP and energy consumption. Eur Rev Med Pharmacol Sci. 2017;21(6):1303-12.
9. Sulaiman N, Mahmoud I, Hussein A, Elbadawi S, Abusnana S, Zimmet P, et al. Diabetes risk score in the United Arab Emirates: a screening tool for the early detection of type 2 diabetes mellitus. BMJ Open Diabetes Res Care. 2018;6(1):e000489.
10. Davies MJ, Aroda VR, Collins BS, Gabbay RA, Green J, Maruthur NM, et al. Management of Hyperglycemia in Type 2 Diabetes, 2022. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care. 2022;45(11):2753-86.
11. Ng SW, Zaghloul S, Ali HI, Harrison G, Popkin BM. The prevalence and trends of overweight, obesity and nutrition-related non-communicable diseases in the Arabian Gulf States. Obes Rev. 2011;12(1):1-13.
12. Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care. 2003;26(3):725-31.
13. Chen L, Magliano DJ, Balkau B, Colagiuri S, Zimmet PZ, Tonkin AM, et al. AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures. Med J Aust. 2010;192(4):197-202.
14. Rigla M, Garcia-Saez G, Pons B, Hernando ME. Artificial Intelligence Methodologies and Their Application to Diabetes. J Diabetes Sci Technol. 2018;12(2):303-10.
15. Morgenstern JD, Buajitti E, O'Neill M, Piggott T, Goel V, Fridman D, et al. Predicting population health with machine learning: a scoping review. BMJ Open. 2020;10(10):e037860.
16. Kouvari M, Valenzuela-Vallejo L, Guatibonza-Garcia V, et al. Liver biopsy-based validation, confirmation and comparison of the diagnostic performance of established and novel non-invasive steatotic liver disease indexes: Results from a large multi-center study. Metabolism. 2023;147:155666.

17. Kouvari M, Mylonakis SC, Katsarou A, et al. The first external validation of the Dallas steatosis index in biopsy-proven non-alcoholic fatty liver disease: a multicenter study. Diabetes Res Clin Pract. 2023;203:110870.
18. Collins R. What makes UK Biobank special? Lancet. 2012;379(9822):1173-4.
19. Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. PLoS One. 2015;10(9):e0137203.
20. National Institute for Health and Care Excellence. Diabetes - type 2: Diagnosis in adults: NICE; 2023 [Available from: https://cks.nice.org.uk/topics/diabetes-type-2/diagnosis/diagnosis-in-adults/.
21. Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage. 2009;45(1 Suppl):S199-209.
22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.
23. Ramos-Louro P, Arellano Perez Vertti RD, Reyes AL, Martinez-Nava GA, Espinosa R, Pineda C, et al. mtDNA haplogroup A enhances the effect of obesity on the risk of knee OA in a Mexican population. Sci Rep. 2022;12(1):5173.
24. Harrell FE Jr. Hmisc: Harrell miscellaneous. 2023.25 Lacy ME, Wellenius GA, Carnethon MR, et al. Racial differences in the performance of existing risk prediction models for incident type 2 diabetes: the CARDIA study. Diabetes Care. 2016;39(2): 285-291.
25. Lacy ME, Wellenius GA, Carnethon MR, Loucks EB, Carson AP, Luo X, et al. Racial Differences in the Performance of Existing Risk Prediction Models for Incident Type 2 Diabetes: The CARDIA Study. Diabetes Care. 2016;39(2):285-91.
26. Wilkinson L, Yi N, Mehta T, Judd S, Garvey WT. Development and validation of a model for predicting incident type 2 diabetes using quantitative clinical data and a Bayesian logistic model: A nationwide cohort and modeling study. PLoS Med. 2020;17(8):e1003232.
27. Glumer C, Vistisen D, Borch-Johnsen K, Colagiuri S, Collaboration D-. Risk scores for type 2 diabetes can be applied in some populations but not all. Diabetes Care. 2006;29(2):410-4.
28. Dugee O, Janchiv O, Jousilahti P, Sakhiya A, Palam E, Nuorti JP, et al. Adapting existing diabetes risk scores for an Asian population: a risk score for detecting undiagnosed diabetes in the Mongolian population. BMC Public Health. 2015;15:938.
29. Rokhman MR, Arifin B, Zulkarnain Z, Satibi S, Perwitasari DA, Boersma C, et al. Translation and performance of the Finnish Diabetes Risk Score for detecting undiagnosed diabetes and dysglycaemia in the Indonesian population. PLoS One. 2022;17(7):e0269853.
30. Wu J, Hou X, Chen L, Chen P, Wei L, Jiang F, et al. Development and validation of a non-invasive assessment tool for screening prevalent undiagnosed diabetes in middle-aged and elderly Chinese. Prev Med. 2019;119:145-52.
31. Smith GD, Spiegelhalter D. Shielding from covid-19 should be stratified by risk. BMJ. 2020;369:m2063.
32. Zhou X, Siegel KR, Ng BP, Jawanda S, Proia KK, Zhang X, et al. Cost-effectiveness of Diabetes Prevention Interventions Targeting High-risk Individuals and Whole Populations: A Systematic Review. Diabetes Care. 2020;43(7):1593-616.
33. Mantena S, Celi LA, Keshavjee S, Beratarrechea A. Improving community health-care screenings with smartphone-based AI technologies. Lancet Digit Health. 2021;3(5):e280-e2.
34. Polyzos SA, Mantzoros CS. Diabetes mellitus: 100 years since the discovery of insulin. Metabolism. 2021;118:154737.

Chapter 3

Simplifying coronary artery disease risk stratification: development and validation of a questionnaire-based alternative comparable to clinical risk tools

Michail Kokkorakis, Pytrik Folkertsma, Filippos Anagnostakis, Nicole Sirotin, Manyoo Agarwal, Ronney Shantouf, Robert H. Henning, Hanno Pijl, Bruce H. R. Wolffenbuttel, Jeroen J. Bax, Douwe E. Atsma, José Castela Forte, Christos S. Mantzoros, Sipko van Dam

EBioMedicine 2025;111:105518

Summary

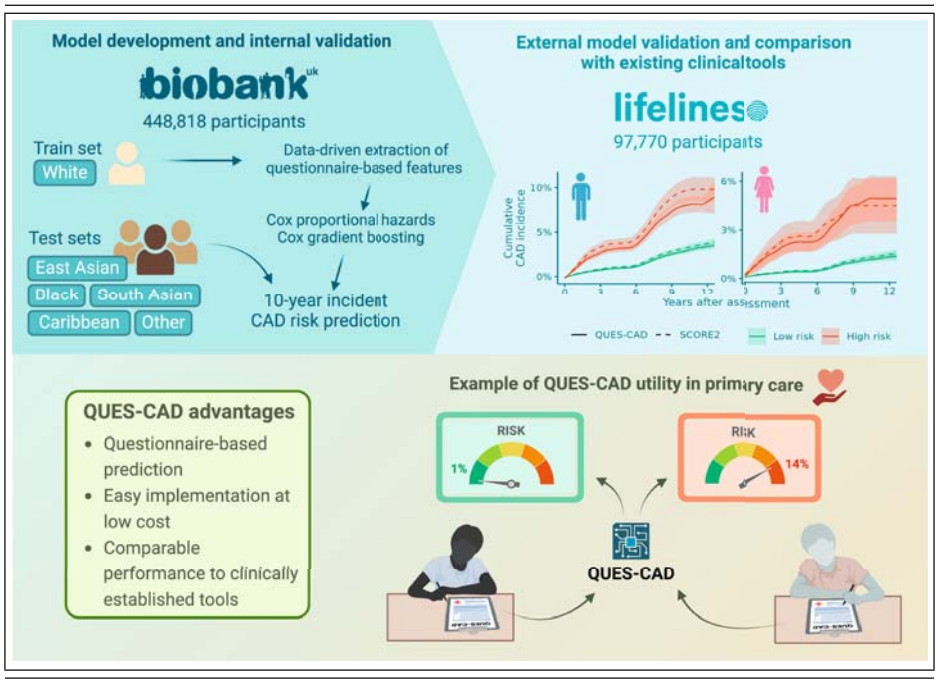
Background Coronary artery disease (CAD) comprises one of the leading causes of morbidity and mortality both in the European population and globally. All established clinical risk stratification scores and models require blood lipids and physical measurements. The latest reports of the European Commission suggest that attracting health professionals to collect these data can be challenging, both from a logistic and cost perspective, which limits the usefulness of established models and makes them unsuitable for population-wide screening in resource-limited settings, i.e., rural areas. Therefore, the aim of this study was to develop and externally validate a questionnaire-based risk stratification model on a population scale at minimal cost, i.e., the Questionnaire-Based Evaluation for Estimating Coronary Artery Disease (QUES-CAD) to stratify the 10-year incidence of coronary artery disease.

Methods Cox proportional hazards (CoxPH) and Cox gradient boosting (CoxGBT) models were trained with 10-fold cross-validation using combinations of ten questionnaire variables on the White population of the UK Biobank (n=448,818) and internally validated the models in all ethnic minorities (n=27,433). The Lifelines cohort was employed as an independent external validation population (n=97,770). Additionally, we compared QUES-CAD's performance, containing only questionnaire variables, to clinically established risk prediction tools, i.e., Framingham Coronary Heart Disease Risk Score, American College of Cardiology/ American Heart Association pooled cohort equation, World Health Organization cardiovascular disease risk charts, and Systematic Coronary Risk Estimation 2 (SCORE2). We conducted partial log-likelihood ratio (PLR) tests and C-index comparisons between QUES-CAD and established clinical prediction models.

Findings In the external validation set, QUES-CAD exhibited C-index values of CoxPH: 0.692 (95% Confidence Interval [CI]: 0.673–0.71) and CoxGBT: 0.699 (95% CI: 0.681–0.717) for the male population and CoxPH: 0.771 (95% CI: 0.748–0.794) and CoxGBT: 0.759 (95% CI: 0.736–0.783) for the female population. The addition of measurement-based variables and variables that require a prior medical examination (i.e., insulin use, number of treatments/medications taken, prevalent cardiovascular disease [other than CAD, and stroke diagnosed by a doctor]) and the further addition of biomarkers/other measurements (i.e., high-density lipoprotein [HDL] cholesterol, total cholesterol, and glycated haemoglobin) did not significantly improve QUES-CAD's performance in most instances. C-index comparisons and PLR tests showed that QUES-CAD performs and fits the data at least as well as the clinical prediction models.

Interpretation QUES-CAD performs comparably to established clinical prediction models and enables a population-wide identification of high-risk individuals for CAD. The model developed and validated herein relies solely on ten questionnaire variables, overcoming the limitations of existing models that depend on physical measurements or biomarkers.
Funding: University Medical Center Groningen.

Keywords Coronary artery disease, machine learning, data-driven prediction, risk stratification, discriminative abilities, population screening



Research in context

Evidence before this study

Coronary artery disease (CAD) is a highly prevalent condition and comprises one of the leading causes of morbidity and mortality both in the European population and globally. All clinical risk scores and models require data from laboratory analyses (blood lipids) and physical measurements (among others, blood pressure). This is a considerable burden since, according to the latest reports from the European Commission, attracting and obtaining health professionals to collect these data can be challenging, both from a logistic and cost perspective, limiting the usefulness of the currently available models, especially for population-wide risk stratification in resource-limited settings, i.e., rural areas. Additionally, even though Europe is becoming increasingly ethnically diverse, there is insufficient validation of the clinical risk prediction models among ethnic minorities of European cohorts, which present higher rates of conventional CAD risk factors, differing treatment response rates, excess CAD-related morbidity, mortality, and overall poor quality of life.

Added value of this study

Questionnaire-Based Evaluation for Estimating Coronary Artery Disease (QUES-CAD) is a novel Machine Learning questionnaire-based risk stratification tool that forecasts the 10-year incidence of CAD and performs comparable to established clinical risk scores, reflecting lifestyle choices, medical history, and social determinants of health.

Implications of all the available evidence

QUES-CAD illustrated risk stratification abilities and performs comparably to established clinical risk prediction tools, by detecting high-risk groups in the population that experience the highest incident rates of CAD, suggesting the potential utility for population-wide screening and identification of populations that might benefit from preventive interventions in a cost-effective and scalable manner.

Introduction

Coronary artery disease (CAD) is the most common type of heart disease and constitutes one of the leading causes of death globally, with a prevalence of over 7% (49). In addition to its significant mortality, CAD is also estimated to account for significant morbidity, accounting for 182 million disability-adjusted life years in 2019 and disproportionately affecting ethnic minority populations (50). Moreover, long-standing health inequalities and social disparities in Europe have widened the gap in cardiovascular care, which consequently impacts other disease outcomes (51, 52). In detail, CAD is especially implicated as a concomitant condition and complication in eminent cardiometabolic epidemics led by obesity, type 2 diabetes, and metabolic dysfunction-associated steatotic liver disease (21, 36, 53-57). Due to its subclinical disease progression, CAD continues to pose a significant challenge for health systems, which renders crucial the need to propose effective policies to narrow the aforementioned gap (58). Primary prevention and novel large-scale population screening programs could help reduce the premature mortality and burden associated with CAD (40, 58).

Identifying traditional risk factors of CAD has led to the development of scoring algorithms that stratify patients for the risk of incident CAD (59). Notably, several risk assessment tools have reached clinical significance and have been included to position themselves in the current cardiovascular disease (CVD) prevention guidelines (59). Specifically, the American Heart Association (AHA) recommends using the Framingham Coronary Heart Disease Risk Score (FRS) and the American College of Cardiology/AHA pooled cohort equation (ACC/AHA PCE) as first-line risk assessment tools for incident CVD (60, 61). Even though ACC/AHA PCE are currently the most widely recommended screening tool by the AHA, the Predicting Risk of Cardiovascular Disease Events (PREVENT) risk equations are now being introduced into practice as a potential replacement (62). Similarly, the European Society of Cardiology suggests using Systematic Coronary Risk Estimation 2 (SCORE2) for the same outcomes (63). Recently, revised World Health Organization (WHO) cardiovascular disease risk prediction models were developed, targeting global implementation, particularly in middle and low-income countries (64).

A recent report on inequalities in access to healthcare in 35 European countries identifies the inadequate availability of services, particularly in rural areas, as a major challenge (65). All the above-mentioned models require data from laboratory analyses (blood lipids) and physical measurements (systolic blood pressure). Therefore, obtaining these data can sometimes be challenging, both from a logistic and cost perspective, limiting the use of these models for population-wide risk stratification. Even though Europe is becoming increasingly ethnically diverse, there is overall insufficient validation of the above prediction models

among ethnic minorities in European patient cohorts, which present higher rates of conventional CAD risk factors, differing treatment response rates, excess CAD-related morbidity, all-cause mortality, and overall poor quality of life (31-34). Hence, there is a need for accurate prognostic tools that do not rely on physical or blood chemistry data, can be deployed in a scalable and cost-efficient way, and are validated in a large multi-ethnic population (35). This is particularly relevant in resource-limited settings, while sex-specific CAD risk assessment models have yet to receive the attention needed despite large risk and performance differences between female and male populations.

The aim of the current study is to develop and validate prognostic questionnaire-based models to stratify CAD risk (Questionnaire-Based Evaluation for Estimating Coronary Artery Disease [QUES-CAD]) in two independent European biobanks across various ethnic populations.

Methods

Study population

The UK Biobank is the largest longitudinal population-based cohort and includes 502,359 individuals aged 37 to 73 years recruited between 2006 and 2010, with follow-up data until October 2022, ranging between 12 and 17 years after the initial assessment with a median follow-up time of 14 years (Supplementary Figure S1A) (66). The Ethics and Guidance Council (http://www.ukbiobank.ac.uk/ethics) oversees ethical practices for the UK Biobank. Before enrollment, all individuals provided informed consent. Lifelines, here employed as a validation cohort, is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 167,729 persons living in the North of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioural, physical, and psychological factors that contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics. Lifelines comprises data from 168,205 people aged 0 to 93 years, gathered between 2006 and 2013 (67, 68). Before enrollment, each participant gave written informed consent. A detailed summary of the gathered data can be found in the respective feature catalogues: https://biobank.ndph.ox.ac.uk/showcase/catalogs.cgi and https://data-catalogue.lifelines.nl/. Follow-up times for Lifelines differ per individual, ranging between 0 and 15 years, with the median follow-up time being five years (Supplementary Figure S1B).

In both study populations, we only included individuals aged ≥40 years.

Assessment of outcomes

In the UK Biobank, CAD diagnosis was assigned based on ICD-9 codes (36, 410, 411, 412, 414, 429), ICD-10 codes (I21, I22, I23, I24.1, and I25.2), OPCS-4 codes (K401, K402, K403, K404, K411, K412, K413, K414, K451, K452, K453, K454, K45.5, K491, K492, K498, K499, K502, K751, K752, K753, 754 and K758, K759) and operation codes (UK Biobank field: 20004, codes 1070 and 1095); and "heart attack diagnosed by a doctor" (UK Biobank field: 6150, code 1), while controls were identified by the absence of all aforementioned diagnoses and procedure codes. Supplementary Tables S1A and S1B provide the data associated with the patient's age at the time of diagnosis, which were employed to calculate the years until diagnosis from the initial assessment. In cases where multiple ages of diagnosis were reported, the lowest reported age was used. All cases diagnosed before their assessment centre visit were then annotated as prevalent cases, while cases diagnosed after their assessment were annotated as incident cases.

In Lifelines, prevalent and incident CAD were annotated based on self-reported myocardial infarction (assessments 1A, 1B, 1C, 2A, 3A, and 3B), percutaneous coronary intervention (PCI) (assessments 1A, 1B, 1C, 3A, and 3B) and open self-reported diseases (1A, 1B, 1C, 2A, 3A). The age of diagnosis was not asked during follow-up; therefore, we estimated the age of diagnosis for every incident case by taking the mean of the age the participant had at the assessment reporting a CAD diagnosis and the age of the participant at the previous assessment (69).

All participants diagnosed with CAD at baseline were excluded from the datasets.

Predictors

We used a data-driven approach to select the best predictive features. First, all categorical features were transformed to one-hot encoding, maintaining their numerical format. Due to the large number of candidate features in the questionnaire pool, we performed a priori feature selection, starting with an initial list containing all features and sub-selecting those with an absolute correlation greater than 0.01 to the target outcome. Then, the final feature selection was performed by iteratively extracting the top correlated feature and removing its variance from the rest of the features until ten variables were selected. To facilitate external validation, we mapped the input features from the UK Biobank to their associated or closest available Lifelines feature (Supplementary Table S2). We integrated these variables into the existing variable pool to determine whether including physical measurement and biomarker variables enhanced the model's performance. Subsequently, we conducted feature selection and retrained the model to evaluate its performance with the augmented feature set.

Model development

Using self-reported questionnaire features, we built sex-specific risk stratification models for incident CAD across all ethnicities of the UK Biobank and externally in Lifelines. Self-reported ethnicity was extracted from the UK Biobank, and participants were divided into six different ethnicities (White, South Asian, Caribbean, East Asian, Black, and Other) based on self-perceived ethnicity (Supplementary Table S3). Models were trained on the White population and tested on the Ethnic minorities and Lifelines. Risk probabilities for White individuals themselves were obtained with 10-fold cross-validation (69). We used the Python package lifelines (version 0.29.0) to train Cox proportional hazards (CoxPH) regression models with default settings as a baseline model. The time variable was constructed as follows: for cases, the number of years until the event since the baseline assessment was used, while for controls, the years between the baseline assessment and the maximum follow-up date of October 2022 was used unless the participant had deceased prior to that point; in that case, the number of years from the baseline assessment to the date of death was used. Additionally, we trained Cox gradient boosting (CoxGBT) survival models using the Python library scikit-survival (version 0.23.0) (70). During the training process, we subsampled controls to two times the number of cases in order to speed up training time. The parameters used for training the gradient boosted survival models are described in Supplementary Table S4.

To ensure consistency, all input features were normalised prior to model training using a scaling reference generated on the training dataset by Sklearn's StandardScaler, which was then used to normalise features of both training and test datasets.

Handling of missing data

To ensure that our model inputs were complete, we excluded participants with missing values. This approach was expected to provide more accurate model performance since it avoids introducing potential biases from imputed data. However, as this approach decreases the sample size and the group sizes of some UK Biobank minorities are relatively small, we also trained the models while imputing missing values based on the train set using Sklearn's iterative imputer. This approach allowed us to perform the risk stratification analysis (Kaplan-Meier analysis; described below) with more subjects, though at the potential cost of reduced discriminative accuracy.

Implementation of existing models

We validated FRS with and without laboratory parameters, both laboratory and non-laboratory-based models developed by the revised WHO cardiovascular disease risk prediction approaches, the ACC/AHA PCE for atherosclerotic

cardiovascular disease risk estimation score, as well as SCORE2 (60, 61, 63, 64, 71).

Risk probabilities for FRS and ACC/AHA PCE were calculated using the R package Cvrisk (version 1.1.0). For SCORE2 and WHO, risk probabilities were calculated as described elsewhere (72, 73). We also included age as a standalone predictor (as a reference point).

Statistical analysis

The discriminatory performance of the models was evaluated by means of the concordance index (C-index) with a 95% confidence interval (CI) using the survival package (version 3.5.7). We performed partial log-likelihood ratio (PLR) tests between our models and the included existing models using the nonnestcox package (version 0.0.0) to compare the goodness of fit of the models, as well as C-index tests using the compareC package (version 1.3.2) to compare the discriminatory performance of the models. Significance thresholds were adjusted using Bonferroni correction.

We performed a more in-depth comparison between the QUES-CAD models and SCORE2. Typically, a SCORE2 threshold of ≥10% risk is used to classify individuals aged 50-69 years into very high- and lower-risk categories (https://www.escardio.org/Journals/E-Journal-of-Cardiology-Practice/Volume-22/rapid-online-self-assessment-of-individual-risk-for-cardiovascular-events-in-a). However, as some of the minority groups in the UK Biobank are relatively small, we opted for a lower SCORE2 cutoff of 0.075 (which is typically used to identify individuals at very high risk aged 40-49 years) to classify individuals into very high- and lower-risk groups to increase the confidence of the statistical analysis in the high-risk group. To ensure a meaningful, interpretable, and fair comparison, we extracted the probability thresholds for our models that returned the same group size for the White ethnic group as SCORE2's threshold of 0.075 did. This approach accounts for differences in model calibration, which reflect their development datasets and can skew direct comparisons at fixed probability thresholds. By aligning group sizes, we ensured that performance metrics, such as sensitivity and specificity, reflected the models' discriminatory capabilities rather than calibration-driven differences. We used the survivalROC package (version 0.3.1) to calculate the time-dependent sensitivity and specificity of our models for stratifying 10-year CAD risk. Overall, the 10-year incidence was estimated using the Kaplan-Meier method from the survival package. Time-dependent positive predictive value (PPV) and negative predictive value (NPV) were derived from the sensitivity, specificity, and incidence (Supplementary Table S5). 95% CI were calculated using 100 bootstraps. The R package ggsurvfit (version 1.0.0) was used to visualise cumulative incidence over

a 15-year follow-up period. All statistical analyses were conducted using R version 3.6.1 and for the model training analyses Python version 3.9 was used.

Role of the funding source

The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Ethics

UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval. This approval means that researchers do not require separate ethical clearance and can operate under the RTB approval. The Lifelines protocol was approved by the University Medical Center Groningen Medical Ethical Committee under number 2007/152. All participants signed an informed consent form. No participants were re-contacted during this project.

Results

Baseline characteristics

A brief overview of the methods is presented in Fig. 1, and detailed baseline characteristics are shown in Table 1A and Table 1B. This study population included 476,251 from the UK Biobank and 97,770 from Lifelines to train, validate, and externally test the machine learning (ML) models (Table 1A and Table 1B). The training set comprised 201,334 White individuals from the UK Biobank for men and 247,484 for women (Table 1A and Table 1B). The models were tested internally among 12,617 individuals for men and 14,816 for women of five different ethnic backgrounds and externally tested in Lifelines (n=40,580 for men, n=57,190 for women) (Table 1A, Table 1B, and Supplementary Table S6).

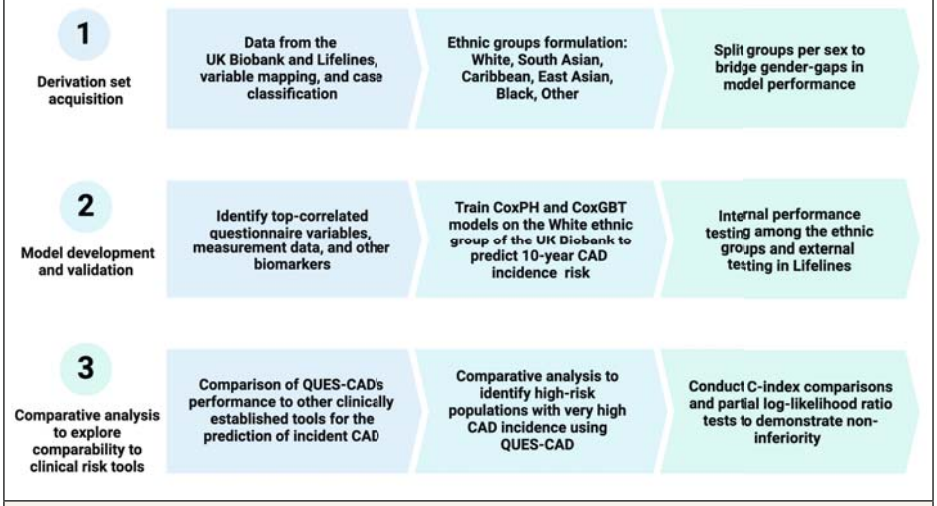


Figure 1. A brief overview of the methodology used in this study.
Abbreviations: CAD, coronary artery disease; QUES-CAD, Questionnaire-Based Evaluation for Estimating Coronary Artery Disease; CoxPH, Cox proportional hazards; CoxGBT, Cox gradient boosting. Created with BioRender.com.

Table 1A. Baseline characteristics of the internal and external study of male populations.
Data are presented as the mean (Standard Deviation [SD]) unless otherwise noted. The measurements are presented as median (Interquartile range [IQR]).

Population	White men		South Asian men		Caribbean men	
Total size (n)	cases 11093	controls 190241	cases 340	controls 3543	cases 71	controls 1698
Questionnaire variables						
Age (years)	62 (9.9)	58.4 (13.3)	58.7 (13)	52.4 (14.5)	57.5 (15.8)	51.4 (12.8)
Alcohol intake frequency (1: Daily or almost daily, 2: Three or four times a week, 3: Once or twice a week, 4: One to three times a month, 5: Special occasions only, 6: Never)	3 (2)	2 (2)	5 (3)	5 (3)	3 (2)	3 (3)
Average total household income before tax (1: <£18,000, 2: £18,000 –£30,999, 3: £31,000–£51,999, 4: ≥£52,000)	2 (2)	3 (2)	2 (2)	2 (3)	2 (2)	2 (2)
Aspirin use	28.60%	15.40%	34.40%	17.60%	35.20%	12%
Body mass index (kg/m^2)	28.2 (5.4)	27.2 (5)	27 (5.1)	26.5 (4.7)	28 (5.1)	28 (5.1)
Current smoking	17.30%	11.80%	13.80%	14.10%	23.90%	23.30%
Heart disease of father	34%	27.20%	34.10%	25.70%	17.20%	8.80%
Heart disease of mother	21.70%	15.40%	19.20%	14.10%	7.40%	8.50%
Number of cigarettes currently smoked daily (current cigarette smokers)	20 (8)	15 (10)	10 (8)	10 (8)	15 (7.5)	10 (9)
Pack years of smoking	27 (26.9)	21 (23.5)	17.5 (16)	14.7 (15.2)	20.1 (27.8)	14.2 (14.1)
Unable to work because of sickness or disability	7.90%	4%	12.60%	5.50%	16.90%	7.20%
Usual walking pace (1: Slow pace, 2: Steady average pace, 3: Brisk pace)	2 (1)	2 (1)	2 (1)	2 (0)	2 (0)	2 (1)
Dentures	25.80%	16.50%	17.10%	9.20%	26.80%	19.60%
Measurement-based variables and variables that require prior medical examination						
Insulin use	3.10%	1.10%	10.90%	2.40%	9.90%	2.90%
Number of treatments/ medications taken	3 (4)	1 (3)	3 (6)	1 (3)	3 (4.5)	1 (3)
Prevalent cardiovascular disease (other than coronary artery disease)	20%	6.40%	29.10%	6.90%	16.90%	4.10%
Stroke diagnosed by a doctor	3.60%	1.70%	3.80%	1.40%	7%	1.30%
Mean systolic blood pressure (mmHg)	143.5 (24.5)	139 (23)	140.8 (22.5)	133.5 (22)	142.5 (23.5)	137 (23)
Waist-to-hip ratio	1 (0.1)	0.9 (0.1)	1 (0.1)	0.9 (0.1)	0.9 (0.1)	0.9 (0.1)
Laboratory variables						
HDL cholesterol (mmol/L)	1.2 (0.4)	1.3 (0.4)	1.1 (0.3)	1.1 (0.3)	1.3 (0.4)	1.3 (0.4)
Total cholesterol (mmol/L)	5.5 (1.7)	5.5 (1.5)	5.2 (1.8)	5.3 (1.5)	5.2 (2.1)	5.2 (1.4)
Glycated haemoglobin (HbA1c) (mmol/mol)	36.4 (5.9)	35.1 (5.1)	40.6 (15.7)	38 (7.2)	39.6 (8.4)	38.1 (6.5)

	East Asian men		Black men		Other men		Lifelines men	
	cases 77	controls 1709	cases 43	controls 1752	cases 192	controls 3192	cases 869	controls 39711
	55.6 (14.8)	51 (14.2)	57.8 (12.9)	49.5 (11.9)	58.8 (11)	53.1 (15.2)	58.3 (16.5)	49.5 (13.2)
	4 (3)	4 (2)	5 (3)	5 (3)	4 (3)	3 (3)	3 (1)	3 (1)
	2 (2)	2 (3)	2 (2)	2 (2)	2 (2)	2 (2)	3 (2)	3 (2)
	23.40%	12.30%	41.90%	13.80%	24.60%	14.70%	99%	99.60%
	26.8 (5.1)	26 (4.5)	29.1 (5.9)	27.8 (5.1)	28.4 (5.2)	27.4 (5.1)	26.9 (4.2)	26.4 (4.3)
	26%	16.60%	18.60%	11.30%	24%	17.80%	21.90%	19.80%
	29.70%	19.50%	10.80%	5.80%	27.50%	21.20%	15.10%	7.30%
	16%	11.40%	7.70%	4.80%	19.30%	12.90%	31.20%	21.40%
	10 (10.2)	12 (9)	10 (4)	10 (8)	15 (8)	15 (10)	15 (10)	13 (12)
	16.8 (18.4)	16.5 (16.7)	17.7 (29.4)	13 (14.1)	20.4 (22.1)	18 (20)	15 (17.4)	12.5 (15.4)
	5.20%	4.60%	14%	4.70%	18.10%	6.40%	6.10%	3.20%
	2 (0)	2 (0)	2 (1)	2 (0)	2 (1)	2 (1)	2 (0)	2 (0)
	14.30%	13.70%	20.90%	9.30%	18.70%	13.90%	22.70%	13.80%
	5.20%	1.50%	2.30%	2.40%	8.80%	1.80%	5.60%	2.50%
	2 (5)	1 (3)	4 (4.5)	1 (3)	3 (4)	1 (3)	1 (3)	0 (2)
	19.50%	4%	25.60%	4.50%	21.40%	5.50%	21.70%	11.70%
	0%	0.70%	2.30%	1.50%	3.50%	0.90%	1.70%	1.10%
	135 (17.5)	133 (22.5)	143 (22.2)	137 (22.9)	143.5 (23)	135 (23)	137 (20)	130 (18)
	0.9 (0.1)	0.9 (0.1)	1 (0.1)	0.9 (0.1)	1 (0.1)	0.9 (0.1)	1 (0.1)	1 (0.1)
	1.2 (0.3)	1.2 (0.3)	1.3 (0.3)	1.2 (0.4)	1.1 (0.3)	1.2 (0.4)	1.2 (0.3)	1.3 (0.4)
	5.5 (2)	5.4 (1.5)	5 (2)	5.1 (1.5)	5.3 (1.6)	5.3 (1.5)	5.6 (1.3)	5.3 (1.2)
	38.5 (7.7)	37.1 (6.3)	38.5 (9.7)	37.8 (7.2)	38.8 (8.9)	36.3 (6.3)	39 (5)	38 (5)

Table 1B. Baseline characteristics of the internal and external study of female populations.
Data are presented as the mean (Standard Deviation [SD]) unless otherwise noted. The measurements are presented as median (Interquartile range [IQR]).

Population	White women		South Asian women		Caribbean women	
Total size (n)	cases 5321	controls 242163	cases 116	controls 3390	cases 60	controls 3001
Questionnaire variables						
Age (years)	62.9 (8.8)	58 (12.7)	59 (11)	52.8 (13.3)	58.6 (12.4)	50.9 (11.7)
Alcohol intake frequency (1: Daily or almost daily, 2: Three or four times a week, 3: Once or twice a week, 4: One to three times a month, 5: Special occasions only, 6: Never)	3 (3)	3 (2)	6 (1)	6 (1)	5 (2)	4 (2)
Average total household income before tax (1: <£18,000, 2: £18,000 –£30,999, 3: £31,000–£51,999, 4: ≥£52,000)	2 (2)	2 (1)	2 (2)	2 (2)	1 (1)	2 (2)
Aspirin use	21.70%	9%	31%	11.70%	33.30%	10.50%
Body mass index (kg/m^2)	27.7 (7.1)	26 (6.2)	28.3 (6.6)	26.7 (6.1)	30.7 (6.7)	28.6 (7.7)
Current smoking	17.70%	8.70%	3.40%	3%	21.70%	13%
Heart disease of father	38.70%	30.50%	39.10%	29.30%	20%	11.80%
Heart disease of mother	30.50%	20.60%	27.30%	18.20%	13.30%	12.70%
Number of cigarettes currently smoked daily (current cigarette smokers)	15 (10)	15 (10)	6 (7.5)	9 (6)	10 (8)	10 (9)
Pack years of smoking	24.4 (22.8)	16.8 (19.5)	17 (22.7)	10.1 (12.2)	17.8 (12.2)	11.9 (13.2)
Unable to work because of sickness or disability	7%	3%	11.20%	6.10%	16.70%	6.20%
Usual walking pace (1: Slow pace, 2: Steady average pace, 3: Brisk pace)	2 (1)	2 (1)	2 (1)	2 (0)	2 (1)	2 (1)
Dentures	30.40%	15.60%	15.50%	8.30%	35%	19.50%
Measurement-based variables and variables that require prior medical examination						
Insulin use	3.30%	0.70%	10.30%	1.80%	15%	1.90%
Number of treatments/ medications taken	3 (4)	2 (3)	4 (4)	2 (3)	4 (4.5)	2 (3)
Prevalent cardiovascular disease (other than coronary artery disease)	16%	3.60%	25%	5.40%	20%	4.50%
Stroke diagnosed by a doctor	3.20%	1%	2.60%	1%	1.70%	1.40%
Mean systolic blood pressure (mmHg)	141 (27.5)	132.5 (26)	140.8 (26.8)	130 (26.5)	144.2 (30.1)	132.5 (26.2)
Waist-to-hip ratio	0.8 (0.1)	0.8 (0.1)	0.9 (0.1)	0.8 (0.1)	0.9 (0.1)	0.8 (0.1)
Laboratory variables						
HDL cholesterol (mmol/L)	1.4 (0.5)	1.6 (0.5)	1.2 (0.3)	1.4 (0.4)	1.4 (0.5)	1.5 (0.5)
Total cholesterol (mmol/L)	6 (1.8)	5.8 (1.5)	5.4 (1.5)	5.4 (1.4)	5.3 (1.5)	5.3 (1.4)
Glycated haemoglobin (HbA1c) (mmol/mol)	36.7 (5.5)	35.1 (4.8)	42.1 (18.3)	37.8 (6.8)	42.4 (12.6)	37.3 (7.1)

East Asian women		Black women		Other women		Lifelines women	
cases 37	controls 2166	cases 27	controls 1880	cases 83	controls 4056	cases 433	controls 56757
61.5 (7.3)	52.7 (12.8)	57.2 (17.6)	50.5 (12.3)	60.8 (13.3)	53.1 (13.3)	59.3 (16.2)	49.4 (12.8)
5 (3)	5 (3)	5 (3)	5 (2)	5 (3)	5 (2)	4 (3)	3 (3)
2 (1)	2 (3)	1 (1)	2 (2)	2 (2)	2 (2)	2 (2)	3 (2)
27%	7.30%	33.30%	11.90%	24.30%	9.60%	100%	99.90%
25.7 (5.2)	24.5 (5.4)	31.1 (5.6)	30 (7.5)	28.8 (7)	26.6 (7.1)	26.6 (5.6)	25.5 (5.7)
5.40%	6.30%	7.40%	6%	14.90%	10.50%	25.40%	17.10%
41.70%	22.90%	12.50%	7.40%	25.40%	21.90%	20.10%	9.60%
35.10%	15.40%	8%	8.50%	31.40%	16.80%	36.70%	23.60%
10 (5)	10 (9)	20 (0)	10 (9)	10 (13)	10 (9)	12 (7.5)	10 (9)
32.6 (14.5)	12.5 (14.2)	12.3 (19.7)	12.6 (15.5)	21.9 (22.6)	13.8 (17.4)	12.8 (18.3)	9 (13.7)
5.40%	3.10%	14.80%	5.60%	10.80%	5.50%	9.90%	4.40%
2 (0)	2 (0)	2 (1)	2 (0)	2 (1)	2 (1)	2 (0)	2 (0)
18.90%	16.20%	22.20%	10.10%	18.90%	13.70%	26.60%	12.60%
5.40%	1%	3.70%	1.70%	4.10%	1.30%	5.80%	1.90%
4 (3)	1 (3)	4 (4)	2 (3)	2 (6)	2 (3)	2 (4)	1 (2)
16.20%	2.70%	25.90%	4.90%	18.10%	3.60%	25.20%	13.50%
0%	0.50%	7.40%	1%	2.70%	0.80%	2.30%	0.90%
141.5 (21)	128.5 (27)	146 (16.5)	134.5 (27.5)	143.8 (29.5)	129.5 (26.5)	132 (22.2)	123 (21)
0.9 (0.1)	0.8 (0.1)	0.9 (0.1)	0.8 (0.1)	0.9 (0.1)	0.8 (0.1)	0.9 (0.1)	0.9 (0.1)
1.5 (0.4)	1.5 (0.5)	1.3 (0.6)	1.5 (0.5)	1.4 (0.5)	1.5 (0.5)	1.5 (0.5)	1.6 (0.5)
5.4 (2.1)	5.7 (1.5)	5 (1.2)	5.2 (1.4)	5.9 (1.8)	5.6 (1.5)	5.6 (1.4)	5.2 (1.3)
38.8 (8.2)	36.5 (5.9)	42.8 (12.2)	37.1 (6.5)	39.2 (7.8)	36.1 (5.9)	39 (5)	38 (5)

Performance of questionnaire-based coronary artery disease risk stratification models (QUES-CAD) in comparison to established clinical models

The discriminative abilities of the QUES-CAD models and other clinical risk assessment scores in women and men across multiple ethnicities are illustrated in Fig. 2, Tables 2A and 2B, and Supplementary Table S7. In the models employing only questionnaire-based features, the three most significant contributing features for women were age, usual walking pace, and pack-years of smoking (Supplementary Fig. S2A and S2B), and for men, age aspirin use, and body mass index (Supplementary Fig. S2C and S2D). Concerning the discriminative ability of QUES-CAD, for the CoxGBT models in women, the C-index varied between 0.737 for the White population to 0.782 for the "Other" ethnic population. The CoxPH models demonstrated similar C-indices ranging from 0.7 in the Black population to 0.804 in the East Asian ethnic population. In men, CoxGBT models showed inferior performance accuracy compared to what we observed among females, with a C-index ranging from 0.653 in the East Asian population to 0.737 in the Black population (Supplementary Table S7). Similar performances were demonstrated with CoxPH C-index ranging from 0.697 in the Black population to 0.808 in the East Asian population for women and in men between 0.654 in the East Asian population and 0.761 in the Black population (Supplementary Table S7). In all instances, in the external validation set, Lifelines, we observed high performances, with a C-index ranging from 0.692 to 0.771 (Supplementary Table S7). For both men and women, the QUES-CAD models (CoxGBT and CoxPH) demonstrate comparable C-index values across ethnic groups and Lifelines and have a similar discriminative capacity (based on C-index) when compared to established clinical models, i.e., ACC/AHA PCE, WHO (with and without laboratory data), FRS (with and without laboratory data), and SCORE2 (Fig. 2).

Regarding PPV and NPV, QUES-CAD models yielded very high NPVs (≥95%) and very low PPVs (≤13%), minimizing false negatives while increasing the false positive results (Supplementary Table S7). Similarly, validating SCORE2 in our study, with a threshold of 7.5%, appears to return almost identical PPVs and NPVs compared to QUES-CAD across all ethnic groups and Lifelines (Supplementary Table S8).

The addition of measurement-based variables and variables that require prior medical examination (i.e., insulin use, number of treatments/medications taken, prevalent cardiovascular disease [other than CAD, and stroke diagnosed by a doctor]) and the further addition of biomarkers/other measurements (i.e., high-density lipoprotein [HDL] cholesterol, total cholesterol, and glycated haemoglobin) did not significantly improve QUES-CAD's performance in all instances, except for the White population (Supplementary Fig. S3, and Supplementary Tables S10, S11). The feature importance and hazard ratios for variables included in the models

with questionnaire & measurement-based variables (or variables that require prior medical examination)-based as well as the ones including biomarkers/other measurements, are presented in Supplementary Figures S4A-D and S5A-D.

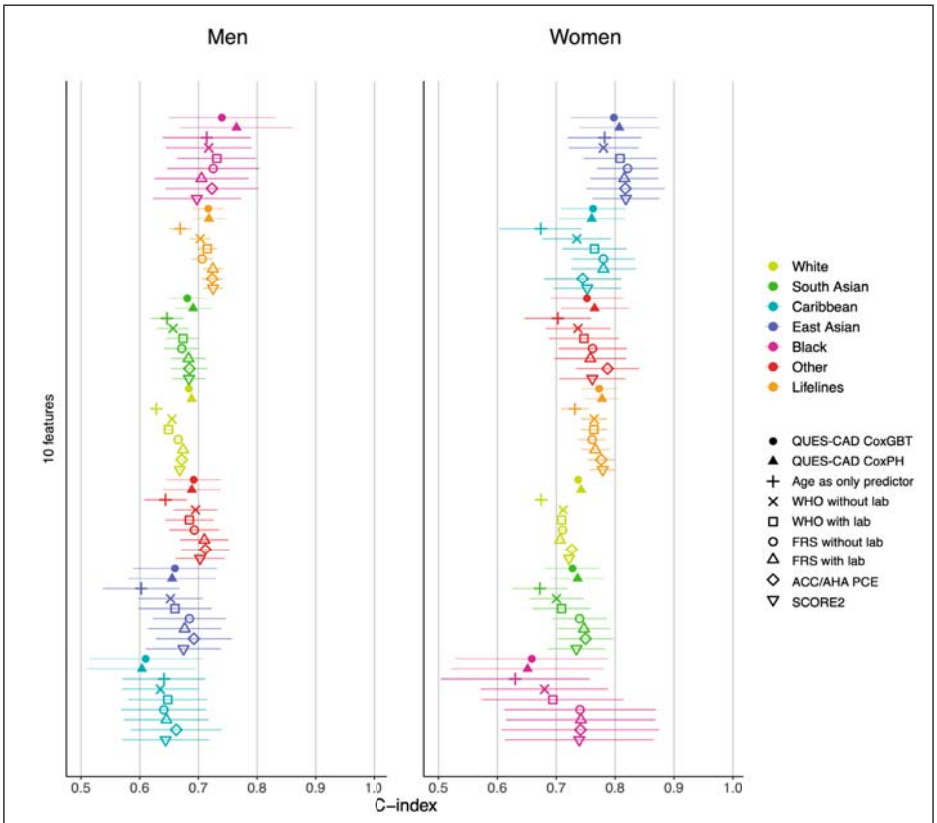


Figure 2. Discriminative abilities of several Coronary Artery Disease (CAD) forecasting models for men (left panel) and women (right panel).
The first filled circle and triangle refer to the QUES-CAD models we developed and validated herein. Each color-symbol combination refers to a specific model and population, explained in detail in the lateral panel. The C-index and 95% CI are presented for all models as horizontal lines. Abbreviations: CoxPH, Cox proportional hazards; CoxGBT, Cox gradient boosting; QUES-CAD, Questionnaire-Based Evaluation for Estimating Coronary Artery Disease; FRS, Framingham Coronary Heart Disease Risk Score; ACC/AHA PCE, American College of Cardiology/American Heart Association pooled cohort equation; WHO, World Health Organization; SCORE2, Systematic Coronary Risk Estimation 2.

Table 2A. Summary of significant results (based on the significant Bonferroni-adjusted p-values), as presented in Supplementary Tables S9-A-D, across all C-index comparisons between QUES-CAD vs established clinical risk tools.
This table illustrates the number of comparisons with statistically significant differences and shows how many of them resulted in QUES-CAD yielding significantly better outcomes. Specifically, each QUES-CAD vs existing model comparison was done in 6 ethnic subgroups of the UK Biobank and in the entire Lifelines dataset. The second number (after "/") in the table rows shows how many of these 7 results were significant (Bonferroni adjusted), and the first number (before "/") shows how many of these 7 results were significant in favour of QUES-CAD. All 196 comparisons are presented in greater detail in Supplementary Tables S9A-D. Abbreviations: PLR, partial log-likelihood ratio; SCORE2, Systematic Coronary Risk Estimation 2; WHO, World Health Organization; FRS, Framingham Coronary Heart Disease Risk Score; QUES-CAD, Questionnaire-Based Evaluation for Estimating Coronary Artery Disease; ACC/AHA, American College of Cardiology/American Heart Association; CAD, coronary artery disease; CI, confidence interval; CoxPH, cox proportional hazards; CoxGBT, cox gradient boosting.

	C-index women CoxPH	C-index men CoxPH	C-index women CoxGBT	C-index men CoxGBT
QUEST-CAD vs SCORE2	1/1	1/1	1/1	1/1
QUEST-CAD vs ACC/AHA PCE	1/1	1/1	0/0	1/1
QUEST-CAD vs FRS with lab	1/1	1/1	1/1	0/0
QUEST-CAD vs FRS without lab	1/1	1/1	1/1	1/1
QUEST-CAD vs WHO with lab	1/1	1/1	1/1	1/1
QUEST-CAD vs WHO without lab	1/1	2/2	1/1	1/1
QUEST-CAD vs Age as standalone marker	3/3	2/2	3/3	3/3

Table 2B. Summary of significant results (based on the significant Bonferroni-adjusted p-values), as presented in Supplementary Tables S9-A-D, across all PLR comparisons between QUES-CAD vs established clinical risk tools.
This table illustrates the number of comparisons with statistically significant differences and shows how many of them resulted in QUES-CAD yielding significantly better outcomes. Specifically, each QUES-CAD vs existing model comparison was done in 6 ethnic subgroups of the UK Biobank and in the entire Lifelines dataset. The second number (after "/") in the table rows shows how many of these 7 results were significant (Bonferroni adjusted), and the first number (before "/") shows how many of these 7 results were significant in favour of QUES-CAD. All 196 comparisons are presented in greater detail in Supplementary Tables S9A-D. Abbreviations: PLR, partial log-likelihood ratio; SCORE2, Systematic Coronary Risk Estimation 2; WHO, World Health Organization; FRS, Framingham Coronary Heart Disease Risk Score; QUES-CAD, Questionnaire-Based Evaluation for Estimating Coronary Artery Disease; ACC/AHA, American College of Cardiology/American Heart Association; CAD, coronary artery disease; CI, confidence interval; CoxPH, cox proportional hazards; CoxGBT, cox gradient boosting.

	PLR women CoxPH	PLR men CoxPH	PLR women CoxGBT	PLR men CoxGBT
QUEST-CAD vs SCORE2	0/1	0/0	2/2	1/1
QUEST-CAD vs ACC/AHA PCE	0/1	0/0	2/2	1/1
QUEST-CAD vs FRS with lab	0/0	0/0	1/1	1/1
QUEST-CAD vs FRS without lab	0/0	0/0	2/2	1/1
QUEST-CAD vs WHO with lab	0/0	1/1	2/2	1/1
QUEST-CAD vs WHO without lab	0/0	1/1	2/2	2/2
QUEST-CAD vs Age as standalone marker	0/0	1/1	2/2	1/1

Risk stratification of QUES-CAD is comparable to established clinical models

Subsequently, we performed a comparison between the performance of the easier-to-use QUES-CAD model and the comparator-established clinical models determining both the discriminatory power (C-index) and goodness-of-fit (PLR). Our results show that the QUES-CAD model achieved similar C-index and PLR values (not statistically different based on the external Lifelines cohort) to those of clinical models across multiple populations. Specifically, in the training set (using 10-fold cross-validation), White males and females, QUES-CAD CoxGBT showed overall significantly higher C-indices and PLRs (Bonferroni-adjusted significance thresholds) compared to all comparator clinical models (Tables 2A and 2B, Supplementary Tables S9A-D). For the rest of the UK Biobank ethnicities no significant differences were evident for either C-index or PLR (Tables 2A and 2B, Supplementary Tables S9A-D). Importantly, CoxGBT performed better than CoxPH, in terms of fitting the data (Tables 2A and 2B, Supplementary Tables S9A-D). The only instances where the comparator models appear to fit the data of the external validation set, Lifelines, better (significant Bonferroni-adjusted p-value for PLR comparisons) is when comparing SCORE2 and ACC/AHA PCE to CoxPH in women (Supplementary Table S9D); although, for the same comparisons with CoxGBT, QUES-CAD fit the data significantly better than both SCORE2 and ACC/AHA PCE (Table 2B and Supplementary Table S9B).

Regarding the 15-year CAD risk stratification, comparing QUES-CAD to the widely validated SCORE2 while maintaining the same high-risk group size for the White population, both QUES-CAD and SCORE2 perform comparably in all populations, and both high-risk groups (of QUES-CAD and SCORE2) have an almost identical incident risk, without significant differences (Fig. 3 and Supplementary Fig. S6). Similarly, low-risk groups also have an identical CAD incidence risk across all populations (Fig. 3 and Supplementary Fig. S6).

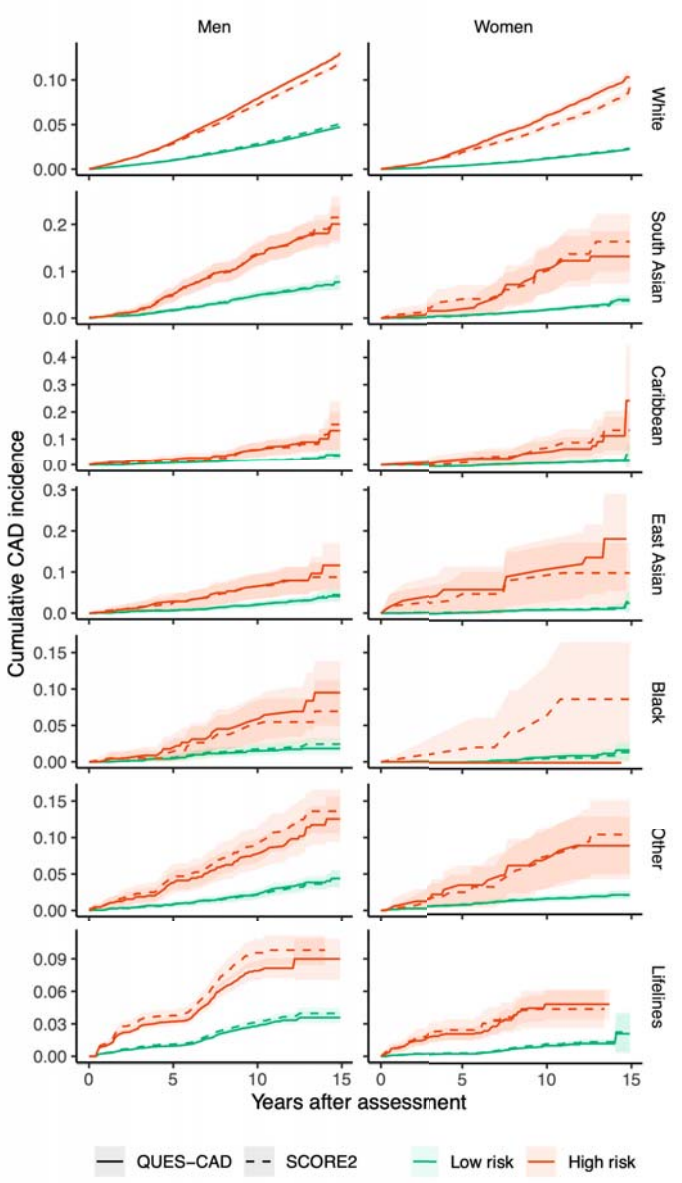


Figure 3. Cumulative incidence of CAD by ethnicity and sex over time.
The x-axis represents years after baseline (initial assessment), while the y-axis indicates the cumulative CAD incidence. Data are stratified by sex (men and women) and population, including White, South Asian, Caribbean, East Asian, Black, Other, and Lifelines (external validation cohort). Cumulative incidence curves are plotted for low and high-risk groups according to QUES-CAD CoxGBT and SCORE2 thresholds; for QUES-CAD we used the threshold that returned the same group size as SCORE2 in the White population. The lighter-colored lines represent the 95% CI. Abbreviations: CAD, coronary artery disease; QUES-CAD, Questionnaire-Based Evaluation for Estimating Coronary Artery Disease; CoxGBT, Cox gradient boosting; SCORE2, Systematic Coronary Risk Estimation 2; CI, confidence interval.

Discussion

In this study including 574,021 individuals, we sought to develop and evaluate the performance of our novel questionnaire-based non-laboratory risk stratification model for incident CAD. Furthermore, we internally validated QUES-CAD in ethnic minorities that are more prevalent than in other European countries; yet they approximate the predicted ethnic make-up of other European countries by 2030 (34). Additionally, we externally validated the developed models in Lifelines, one of the largest European biobanks, to examine the generalizability of our results. The primary finding was that the QUES-CAD model stratifies incident CAD as accurately as current clinical models such as FRS, SCORE2, ACC/AHA PCE, and WHO in all populations and also achieved relatively higher performance in the external validation cohort, suggesting its potential applicability for the multi-ethnic populations of Europe. The inclusion of physical measurements and/or biomarkers (variables presented in Table 1, Supplementary Figures S4A-D and S5A-D) did not increase the performance of QUES-CAD. At the same time, with the currently available tools and the present limitations of healthcare systems, creating a population risk stratification program that requires laboratory values for CAD may comprise an unrealistic goal. This is mainly due to the considerable costs and logistical challenges associated with the need for blood sample collection, but also for the organisation and management of "screening units" to perform these tests (74). Specifically, as described in the 2018 European Commission report regarding inequalities in healthcare access, attracting and retaining health professionals is problematic, and reaching particularly vulnerable communities with limited access to qualitative healthcare is burdensome (65). These results suggest that a questionnaire-based risk stratification algorithm performs at least as well as currently available tools (comprising a comparable number of features) that employ physical and blood biomarkers, enable new scalable avenues for risk screening with major implications for population health monitoring. A questionnaire-based model can be applied population-wide due to the limited cost and effort, as laboratory tests or time from medical personnel are no longer required. Notably, the specific variables selected to construct the QUES-CAD model exclude any variables that individuals find difficult to assess accurately, such as waist circumference or body fat percentage. Instead, only those questionnaire variables that were asked directly to the biobanks' participants were included, ensuring that these answers translate one-on-one to the answers that can be expected in a population-wide screening. This finding implies that these models are uniquely suited to be deployed without requiring assistance from medical personnel.

QUES-CAD may also enhance disease prevention in resource-limited settings where access to preventive cardiovascular care is limited (69). With the primary

goal to improve health equity and reduce health disparities while reducing the burden of CAD as a highly prevalent non-communicable disease (NCD), QUES-CAD is a readily scalable solution to accompany or replace the currently implemented risk models, especially in low socioeconomic status (SES) or resource-limited settings, i.e., in rural European areas. The Lancet NCD Action Group and the NCD Alliance suggested that cardiovascular risk reduction ranks among the five top priority interventions for NCDs (75). Notably, in the same study, tobacco use ranks first, while the category "obesity, unhealthy diet, and physical inactivity" ranked third (75). Therefore, a questionnaire-based risk stratification tool comprising such features, along with household income information, provides a solution to stratify individuals at a population level to facilitate the effective deployment of these interventions.

Despite growing efforts towards advancing ML initiatives in cardiovascular care, most of these applications focus on imaging, electrocardiography, and biomarker analyses (76). The development of QUES-CAD allows the novelty of reliable non-laboratory stratification methods to be implemented in the most recent and ever-increasing technological trends in clinical decision-making, such as the emerging use of large language models (LLM) for telemedicine applications (remote patient monitoring). LLMs can provide a suitable avenue to integrate QUES-CAD, making it user-friendly and widely accessible for virtual care (77, 78). For instance, users can provide input to a chatbot based on the questions included in QUES-CAD, and the chatbot can predict their ten-year risk of developing CAD. Furthermore, it is pertinent to mention that LLMs can receive both structured and unstructured data from the patient directly or from electronic health records, being able to adjust their prediction based on the inputs it receives across the patient's lifetime (79). Moreover, QUES-CAD showed greater uniformity in its discriminative performance across ethnicities compared to currently used clinical tools, and the absence of blood biomarkers enables individual risk calculation outside the "strict" healthcare setting, such as in a remote or hybrid environment. Even at primary care visits, QUES-CAD can be performed during consultation hours with the general practitioner via a website or mobile app and readily provides the individual risk for developing CAD over the next 15 years with ten simple-to-obtain questions, almost all of which are typically asked during a patient visit at a medical facility.

Since CoxGBT can capture non-linear relationships, our results suggest that the association between the covariates and the hazard is not entirely linear. Although CoxGBT performs better than the CoxPH model in Lifelines in terms of goodness-of-fit, similar results could potentially be achieved with a CoxPH model by incorporating interaction terms or non-linear transformations, such as quadratic terms. Given that GBT models are often regarded as "black box" models, it would be advantageous to focus on optimising the regression (CoxPH) models, as their

coefficients provide greater interpretability due to their linear structure. Examining the decision trees from the GBT model could also help uncover key interaction effects or inform necessary variable transformations, offering insights that might improve the regression models.

At the same time, we aim to bridge the sex gap by generating separate models for males and females that are optimised to yield the highest performance for each sex. Interestingly, the female version of QUES-CAD demonstrates higher C-indices, and besides this finding being also reported in other risk tools, different models (with discrete variables) for men and women are currently absent in clinical practice. The current study presents several strengths and limitations. First, this study achieves the highest validation standards in the ML field by showing the ML-based models' performance and potential clinical utility of a questionnaire-based risk stratification model for incident CAD in two large population cohorts across multiple ethnicities. From a modelling perspective, this minimises the chances of overfitting and provides evidence of the model's validity. Then, we further underpinned the reliability of our models by validating them in all ethnic populations of the UK Biobank and provided a comparison of the six major clinical risk stratification tools: FRS without lab, FRS with lab, ACC/AHA PCE, WHO without lab, WHO with lab, and SCORE2. One limitation is that ethnicity data may only be partially accurate, as with all self-reported biobank data. In particular, an individual's self-reported ethnicity may be shaped by their perceptions and cultural and societal influences and may not consistently be representative of their ancestral background. However, these biases are in some way desirable when the aim is to deploy the QUES-CAD on a population-wide scale, as in this case, these biases will be part of the assessment and accounted for by the models. Additionally, daily aspirin use (as a predictor) may be a marker of a clinical encounter and increased CVD risk. Lastly, since this study is observational, it is not possible to establish cause-and-effect relationships between the variables integrated into QUES-CAD and the anticipated outcomes.

Conclusion

In conclusion, QUES-CAD, a novel ML-based multi-ethnic CAD incident risk stratification tool, solely employs ten questionnaire-based variables and performs comparable to the established risk scoring systems (which require lab-based variables and other physical measurements) currently implemented in primary care cardiology guidelines. These questionnaire-based models reduce effort and cost to a minimum and can thereby revolutionise the monitoring of CAD risk, enabling population-wide screening to identify which individuals would benefit from preventive interventions, including both lifestyle and medical interventions, that target cardiometabolic risk in a cost-effective and scalable manner.

Contributors

MK contributed to the conceptualisation, data curation, validation, investigation, methodology, project administration, visualisation, manuscript writing, and manuscript review. PF contributed to the data analysis, data curation, investigation, methodology, validation, and visualisation. MK, PF, and SvD accessed and verified the underlying data. FA contributed to the visualisation, manuscript writing, and manuscript review. NS, MA, RS, RHH, HP, JJB, and DEA contributed to the manuscript review. BHRW contributed to the data curation, manuscript review, resources, and funding acquisition. JCF contributed to the conceptualisation, methodology, manuscript writing, and manuscript review. CSM and SVD contributed equally to the conceptualisation, data curation, validation, methodology, project administration, visualisation, manuscript writing, manuscript review, and supervision. All authors read and approved the final version of the manuscript.

Data sharing statement

Study data are available from UK Biobank and Lifelines but were used under license for the current study, which restricts their public availability. Data may be obtained from a third party and are not publicly available. Researchers can apply to use the UK Biobank and Lifelines data used in this study. More information about how to request UK Biobank data and the conditions of use can be found on their website (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access), and for Lifelines data, and the conditions of use can be found on their website (https://www.lifelines-biobank.com/researchers/working-with-us). The underlying code is available and can be requested from the corresponding author.

Declaration of Interest

PF, SvD, and JCF are employed by Ancora Health B.V. and own shares of Ancora Health B.V. BHRW sits on the medical advisory board of Ancora Health B.V., without being compensated for this position. All other authors have no conflict of interest to declare.

Acknowledgements

This project was funded by the University Medical Center Groningen under project number: PPP-2019_023. MK receives funding support from an MD-PhD grant from the University Medical Center Groningen. The funder had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript. The Lifelines initiative has been made possible by subsidy from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen (UMCG), Groningen University and the Provinces in the North of the Netherlands (Drenthe, Friesland, Groningen). The authors would like to thank Dr. Jonathan Dreyfuss, on behalf of Harvard Catalyst, for reviewing and

providing feedback regarding the analytical and statistical approaches followed in this manuscript. This work was conducted with support from UM1TR004408 award through Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health) and financial contributions from Harvard University and its affiliated academic healthcare centers. The content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard Catalyst, Harvard University and its affiliated academic healthcare centers, or the National Institutes of Health. The authors thank the UK Biobank data access granted through application 55495 and data access to the Lifelines data through application OV20_00020.

Supplementary data

This article contains supporting information available online at:
https://www.thelancet.com/cms/10.1016/j.ebiom.2024.105518/attachment/9e77195c-2615-4bb6-b708-8d1b24f56fd8/mmc1.docx

References

1. Himsworth HP. Diabetes mellitus: Its differentiation into insulin-sensitive and insulin-insensitive types. The Lancet. 1936;227(5864):127-30.
2. Phillips GB. Relationship between serum sex hormones and glucose, insulin and lipid abnormalities in men with myocardial infarction. Proc Natl Acad Sci U S A. 1977;74(4):1729-33.
3. Reaven GM. Banting lecture 1988. Role of insulin resistance in human disease. Diabetes. 1988;37(12):1595-607.
4. Reaven GM. The metabolic syndrome: is this diagnosis necessary? Am J Clin Nutr. 2006;83(6):1237-47.
5. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the Metabolic Syndrome. Circulation. 2009;120(16):1640-5.
6. Bays HE. Adiposopathy is “sick fat” a cardiovascular disease? J Am Coll Cardiol. 2011;57(25):2461-73.
7. Ndumele CE, Neeland IJ, Tuttle KR, Chow SL, Mathew RO, Khan SS, et al. A Synopsis of the Evidence for the Science and Clinical Management of Cardiovascular-Kidney-Metabolic (CKM) Syndrome: A Scientific Statement From the American Heart Association. Circulation. 2023;148(20):1636-64.
8. Carbone F. CardioMetabolic medicine, one more last step forward. Eur Heart J. 2022;43(20):1895-6.
9. Neeland IJ, Ross R, Després JP, Matsuzawa Y, Yamashita S, Shai I, et al. Visceral and ectopic fat, atherosclerosis, and cardiometabolic disease: a position statement. Lancet Diabetes Endocrinol. 2019;7(9):715-25.
10. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2023;402(10397):203-34.
11. Zhang H, Zhou X-D, Shapiro MD, Lip GYH, Tilg H, Valenti L, et al. Global burden of metabolic diseases, 1990–2021. Metabolism - Clinical and Experimental. 2024;160.
12. Ferrari AJ, Santomauro DF, Aali A, Abate YH, Abbafati C, Abbastabar H, et al. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet. 2024;403(10440):2133-61.
13. Fox A, Feng W, Asal V. What is driving global obesity trends? Globalization or “modernization”? Global Health. 2019;15(1):32.
14. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology. 2018;67(1):328-57.
15. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology. 2016;64(1):73-84.
16. Muzurović E, Peng CC, Belanger MJ, Sanoudou D, Mikhailidis DP, Mantzoros CS. Nonalcoholic Fatty Liver Disease and Cardiovascular Disease: a Review of Shared Cardiometabolic Risk Factors. Hypertension. 2022;79(7):1319-26.
17. Thomas G, Sehgal AR, Kashyap SR, Srinivas TR, Kirwan JP, Navaneethan SD. Metabolic syndrome and kidney disease: a systematic review and meta-analysis. Clin J Am Soc Nephrol. 2011;6(10):2364-73.

18. Ng TP, Feng L, Nyunt MS, Feng L, Gao Q, Lim ML, et al. Metabolic Syndrome and the Risk of Mild Cognitive Impairment and Progression to Dementia: Follow-up of the Singapore Longitudinal Ageing Study Cohort. JAMA Neurol. 2016;73(4):456-63.
19. Ben Assayag E, Eldor R, Korczyn AD, Kliper E, Shenhar-Tsarfaty S, Tene O, et al. Type 2 Diabetes Mellitus and Impaired Renal Function Are Associated With Brain Alterations and Poststroke Cognitive Decline. Stroke. 2017;48(9):2368-74.
20. Yki-Järvinen H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. Lancet Diabetes Endocrinol. 2014;2(11):901-10.
21. Kokkorakis M, Boutari C, Katsiki N, Mantzoros CS. From non-alcoholic fatty liver disease (NAFLD) to steatotic liver disease (SLD): an ongoing journey towards refining the terminology for this prevalent metabolic condition and unmet clinical need. Metabolism. 2023;147:155664.
22. Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, et al. Diabetes and cancer: a consensus report. Diabetes Care. 2010;33(7):1674-85.
23. Esposito K, Chiodini P, Colao A, Lenzi A, Giugliano D. Metabolic syndrome and risk of cancer: a systematic review and meta-analysis. Diabetes Care. 2012;35(11):2402-11.
24. Goff LM, Ladwa M, Hakim O, Bello O. Ethnic distinctions in the pathophysiology of type 2 diabetes: a focus on black African-Caribbean populations. Proc Nutr Soc. 2020;79(2):184-93.
25. Chen L, Magliano DJ, Balkau B, Colagiuri S, Zimmet PZ, Tonkin AM, et al. AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures. Med J Aust. 2010;192(4):197-202.
26. Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care. 2003;26(3):725-31.
27. Davies MJ, Aroda VR, Collins BS, Gabbay RA, Green J, Maruthur NM, et al. Management of Hyperglycemia in Type 2 Diabetes, 2022. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care. 2022;45(11):2753-86.
28. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J. 2021;42(25):2439-54.
29. Diaz CL, Shah NS, Lloyd-Jones DM, Khan SS. State of the Nation's Cardiovascular Health and Targeting Health Equity in the United States: A Narrative Review. JAMA Cardiology. 2021;6(8):963-70.
30. Khan SS, Matsushita K, Sang Y, Ballew SH, Grams ME, Surapaneni A, et al. Development and Validation of the American Heart Association's PREVENT Equations. Circulation. 2024;149(6):430-49.
31. Leigh JA, Alvarez M, Rodriguez CJ. Ethnic Minorities and Coronary Heart Disease: an Update and Future Directions. Curr Atheroscler Rep. 2016;18(2):9.
32. Nasir K, Shaw LJ, Liu ST, Weinstein SR, Mosler TR, Flores PR, et al. Ethnic differences in the prognostic value of coronary artery calcification for all-cause mortality. J Am Coll Cardiol. 2007;50(10):953-60.
33. Mital R, Bayne J, Rodriguez F, Ovbiagele B, Bhatt DL, Albert MA. Race and Ethnicity Considerations in Patients With Coronary Artery Disease and Stroke: JACC Focus Seminar 3/9. J Am Coll Cardiol. 2021;78(24):2483-92.
34. Stronks K, Snijder MB, Peters RJ, Prins M, Schene AH, Zwinderman AH. Unravelling the impact of ethnicity on health in Europe: the HELIUS study. BMC Public Health. 2013;13:402.
35. Shroufi A, Chowdhury R, Anchala R, Stevens S, Blanco P, Han T, et al. Cost effective interventions for the prevention of cardiovascular disease in low and middle income countries: a systematic review. BMC Public Health. 2013;13:285.

36. Kouvari M, Valenzuela-Vallejo L, Guatibonza-Garcia V, Polyzos SA, Deng Y, Kokkorakis M, et al. Liver biopsy-based validation, confirmation and comparison of the diagnostic performance of established and novel non-invasive steatotic liver disease indexes: Results from a large multi-center study. Metabolism. 2023;147:155666.
37. Zhou F, Stueck A, McLeod M. Liver biopsy complication rates in patients with non-alcoholic fatty liver disease. Can Liver J. 2022;5(2):106-12.
38. Thampanitchawong P, Piratvisuth T. Liver biopsy:complications and risk factors. World J Gastroenterol. 1999;5(4):301-4.
39. Angelidi AM, Sanoudou D, Hill MA, Mantzoros CS. Management of patients with the cardio renal liver metabolic syndrome: The need for a multidisciplinary approach in research, education and practice. Metabolism. 2024;159:155997.
40. Kokkorakis M, Folkertsma P, Forte JC, Wolffenbuttel BHR, van Dam S, Mantzoros CS. GDF-15 improves the predictive capacity of Steatotic liver disease non-invasive tests for incident morbidity and mortality risk for cardio-renal-metabolic diseases and malignancies. Metabolism. 2024:156047.
41. Kokkorakis M, Boutari C, Hill MA, Kotsis V, Loomba R, Sanyal AJ, et al. Resmetirom, the first approved drug for the management of metabolic dysfunction-associated steatohepatitis: Trials, opportunities, and challenges. Metabolism. 2024;154:155835.
42. Kokkorakis M, Muzurović E, Volčanšek Š, Chakhtoura M, Hill MA, Mikhailidis DP, et al. Steatotic Liver Disease: Pathophysiology and Emerging Pharmacotherapies. Pharmacol Rev. 2024;76(3):454-99.
43. Jia L, Du Y, Chu L, Zhang Z, Li F, Lyu D, et al. Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. Lancet Public Health. 2020;5(12):e661-e71.
44. Livingston G, Huntley J, Liu KY, Costafreda SG, Selbæk G, Alladi S, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. Lancet. 2024;404(10452):572-628.
45. Dove A, Marseglia A, Shang Y, Grande G, Vetrano DL, Laukka EJ, et al. Cardiometabolic multimorbidity accelerates cognitive decline and dementia progression. Alzheimers Dement. 2023;19(3):821-30.
46. Tai XY, Veldsman M, Lyall DM, Littlejohns TJ, Langa KM, Husain M, et al. Cardiometabolic multimorbidity, genetic risk, and dementia: a prospective cohort study. Lancet Healthy Longev. 2022;3(6):e428-e36.
47. Lingvay I, Cohen RV, Roux CWL, Sumithran P. Obesity in adults. Lancet. 2024;404(10456):972-87.
48. Aronne LJ, Sattar N, Horn DB, Bays HE, Wharton S, Lin WY, et al. Continued Treatment With Tirzepatide for Maintenance of Weight Reduction in Adults With Obesity: The SURMOUNT-4 Randomized Clinical Trial. JAMA. 2024;331(1):38-48.
49. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association. Circulation. 2022;145(8):e153-e639.
50. Javed Z, Haisum Maqsood M, Yahya T, Amin Z, Acquah I, Valero-Elizondo J, et al. Race, Racism, and Cardiovascular Health: Applying a Social Determinants of Health Framework to Racial/Ethnic Disparities in Cardiovascular Disease. Circ Cardiovasc Qual Outcomes. 2022;15(1):e007917.
51. Nadarajah R, Farooq M, Raveendra K, Nakao YM, Nakao K, Wilkinson C, et al. Inequalities in care delivery and outcomes for myocardial infarction, heart failure, atrial fibrillation, and aortic stenosis in the United Kingdom. Lancet Reg Health Eur. 2023;33:100719.

52. Cenko E, Manfrini O, Fabin N, Dorobantu M, Kedev S, Milicic D, et al. Clinical determinants of ischemic heart disease in Eastern Europe. Lancet Reg Health Eur. 2023;33:100698.
53. Kouvari M, Mylonakis SC, Katsarou A, Valenzuela-Vallejo L, Guatibonza-Garcia V, Kokkorakis M, et al. The first external validation of the Dallas steatosis index in biopsy-proven Non-alcoholic fatty liver Disease: A multicenter study. Diabetes Res Clin Pract. 2023;203:110870.
54. Kokkorakis M, Katsarou A, Katsiki N, Mantzoros CS. Milestones in the journey towards addressing obesity; Past trials and triumphs, recent breakthroughs, and an exciting future in the era of emerging effective medical therapies and integration of effective medical therapies with metabolic surgery. Metabolism. 2023;148:155689.
55. Boutari C, Kokkorakis M, Stefanakis K, Valenzuela-Vallejo L, Axarloglou E, Volčanšek Š, et al. Recent research advances in metabolism, clinical and experimental. Metabolism. 2023;149:155722.
56. Kokkorakis M, Chakhtoura M, Rhayem C, Al Rifai J, Ghezzawi M, Valenzuela-Vallejo L, et al. Emerging Pharmacotherapies for Obesity: A Systematic Review. Pharmacological Reviews. 2024:PHARMREV-AR-2023-001045.
57. Kokkorakis M, Muzurović E, Volčanšek Š, Chakhtoura M, Hill MA, Mikhailidis DP, et al. Steatotic Liver Disease: Pathophysiology and Emerging Pharmacotherapies. Pharmacological Reviews. 2024;76(3):454-99.
58. 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;76(25):2982-3021.
59. Pennells L, Kaptoge S, Wood A, Sweeting M, Zhao X, White I, et al. Equalization of four cardiovascular risk algorithms after systematic recalibration: individual-participant meta-analysis of 86 prospective studies. Eur Heart J. 2019;40(7):621-31.
60. 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: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation. 2019;140(11):e563-e95.
61. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. 1998;97(18):1837-47.
62. Khan SS, Matsushita K, Sang Y, Ballew SH, Grams ME, Surapaneni A, et al. Development and Validation of the American Heart Association's PREVENT Equations. Circulation. 2024;149(6):430-49.
63. group Sw, collaboration ESCCr. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J. 2021;42(25):2439-54.
64. Group WCRCW. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. Lancet Glob Health. 2019;7(10):e1332-e45.
65. Commission E, Directorate-General for Employment SA, Inclusion, Baeten R, Spasova S, Coster S, et al. Inequalities in access to healthcare – A study of national policies 2018: Publications Office; 2018.
66. Collins R. What makes UK Biobank special? Lancet. 2012;379(9822):1173-4.
67. Klijs B, Scholtens S, Mandemakers JJ, Snieder H, Stolk RP, Smidt N. Representativeness of the LifeLines Cohort Study. PLoS One. 2015;10(9):e0137203.
68. Sijtsma A, Rienks J, van der Harst P, Navis G, Rosmalen JGM, Dotinga A. Cohort Profile Update: Lifelines, a three-generation cohort study and biobank. Int J Epidemiol. 2022;51(5):e295-e302.

69. Kokkorakis M, Folkertsma P, van Dam S, Sirotin N, Taheri S, Chagoury O, et al. Effective questionnaire-based prediction models for type 2 diabetes across several ethnicities: a model development and validation study. EClinicalMedicine. 2023;64:102235.
70. Pölsterl S. scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn. Journal of Machine Learning Research. 2020;21(212):1-6.
71. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation. 2019;139(25):e1082-e143.
72. group Sw, collaboration ECr. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. European Heart Journal. 2021;42(25):2439-54.
73. Kaptoge S, Pennells L, De Bacquer D, Cooney MT, Kavousi M, Stevens G, et al. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. The Lancet Global Health. 2019;7(10):e1332-e45.
74. Spoyalo K, Lalande A, Rizan C, Park S, Simons J, Dawe P, et al. Patient, hospital and environmental costs of unnecessary bloodwork: capturing the triple bottom line of inappropriate care in general surgery patients. BMJ Open Qual. 2023;12(3).
75. Beaglehole R, Bonita R, Horton R, Adams C, Alleyne G, Asaria P, et al. Priority actions for the non-communicable disease crisis. Lancet. 2011;377(9775):1438-47.
76. Manlhiot C, van den Eynde J, Kutty S, Ross HJ. A Primer on the Present State and Future Prospects for Machine Learning and Artificial Intelligence Applications in Cardiology. Can J Cardiol. 2022;38(2):169-84.
77. Snoswell CL, Snoswell AJ, Kelly JT, Caffery LJ, Smith AC. Artificial intelligence: Augmenting telehealth with large language models. J Telemed Telecare. 2023:1357633X231169055.
78. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930-40.
79. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, et al. The future landscape of large language models in medicine. Commun Med (Lond). 2023;3(1):141.


ATHENA

Part II

Omics Integration in Risk Stratification and Modeling Approaches

Chapter 4

GDF-15 improves the predictive capacity of steatotic liver disease non-invasive tests for incident morbidity and mortality risk for cardio-renal-metabolic diseases and malignancies

Michail Kokkorakis, Pytrik Folkertsma, José Castela Forte, Bruce H. R. Wolffenbuttel, Sipko van Dam, Christos S. Mantzoros

Metabolism 2025;163:156047

Abstract

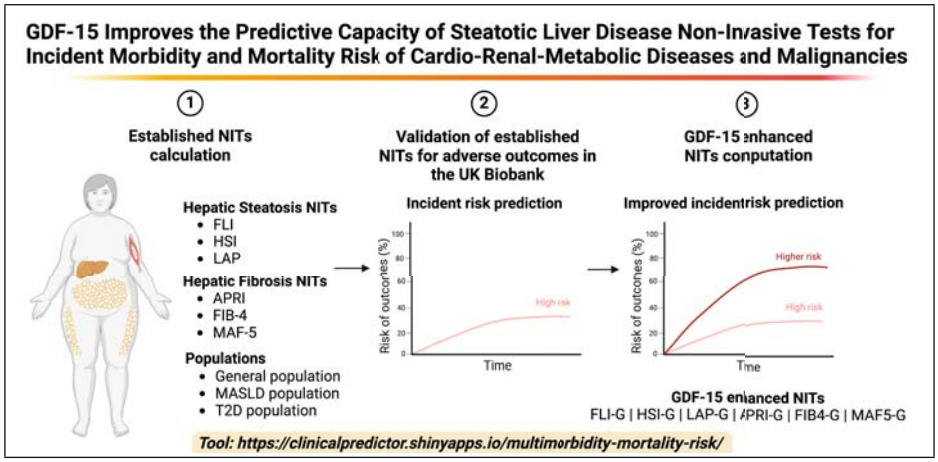
Background & Aims Noninvasive tools (NITs) are currently used to stratify the risk of having or developing hepatic steatosis or fibrosis. Their performance and a proteomic-enabled improvement in forecasting long-term cardio-renal-metabolic morbidity, malignancies, as well as cause-specific and all-cause mortality, are lacking. Therefore, the performance of established NITs needs to be investigated in identifying cardio-renal-metabolic morbidity, malignancies, cause-specific and overall mortality and improve their performance with novel, proteomic-enabled NITs, including growth differentiation factor 15 (GDF-15), allowing multipurpose utilization.

Methods 502,359 UK Biobank participants free of the study outcomes at baseline with a 14-year median follow-up were grouped into three categories: a) general population, b) potentially metabolic dysfunction-associated steatotic liver disease (MASLD) population, c) individuals with type 2 diabetes mellitus. The investigated NITs include Aspartate aminotransferase to Platelet Ratio Index (APRI), Fibrosis 4 Index (FIB-4), Fatty Liver Index (FLI), Hepatic Steatosis Index (HSI), Lipid Accumulation Product (LAP), and metabolic dysfunction–associated fibrosis (MAF-5) score.

Results Adding GDF-15 to the existing NITs led to significantly increased prognostic performance compared to the traditional NITs in almost all instances, reaching substantially high C-indices, ranging between 0.601 and 0.808, with an overall >0.2 improvement in C-index. Overall, with the GDF-15 enhanced NITs, up to more than seven times fewer individuals need to be screened to identify more incident cases of adverse outcomes compared to the traditional NITs. The cumulative incidence of all outcomes, based on the continuous value percentiles of NITs, is increasing exponentially in the upper quintile of the GDF-15 enhanced NITs.

Conclusions The herein-developed GDF-15 enhanced indices demonstrate higher screening effectiveness and significantly improved prognostic abilities, which are reduced to practice through an easy-to-use web-based calculator tool (https://clinicalpredictor.shinyapps.io/multimorbidity-mortality-risk/).

Keywords NITs, MASLD, AI, ML, omics



Highlights

- Incorporating GDF-15 significantly improved the prognostic accuracy of established NITs.
- Multipurpose GDF-15-enabled NITs led to a more efficient screening process, reducing the number of individuals needed to identify incident cases of adverse outcomes.
- The developed indices are accessible via a user-friendly web-based calculator, promoting easy integration into clinical practice.

Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD), previously known as non-alcoholic fatty liver disease (NAFLD), is the most prevalent chronic liver condition worldwide, impacting one-third of the global population and in 2021, was estimated to account for 3.67 million disability-adjusted life years across all ages (1-5). MASLD is defined as fat accumulation in the liver when excess alcohol consumption and other causes have been excluded (4). Starting with simple steatosis, the accumulation of fat within the liver accelerates hepatocellular inflammation and hepatocyte damage (ballooning), proceeding to metabolic dysfunction-associated steatohepatitis (MASH), which is approximately present in 20% of individuals with MASLD (6-8). The prevalence of MASLD in metabolic surgery candidates with severe obesity is >90%, in patients with type 2 diabetes mellitus (T2D) >55%, and in patients with cardiovascular disease (CVD) ~70% (4, 9). MASLD is also strongly associated with all-cause mortality, where CVD-related mortality is the most prevalent within this group (9, 10).

The current diagnostic gold standard for differentiating steatotic liver from MASH and staging liver fibrosis is liver biopsy, which is invasive, risky, and costly (4, 11, 12). Therefore, non-invasive tests (NITs) have been proposed to aid MASLD diagnosis (Supplementary Table 1) (13). Some NITs have been introduced into clinical diagnostic guidelines, i.e., by the American Gastroenterological Association (AGA); despite their suboptimal performance, they are widely accepted for first-line screening for MASLD, also in patients with T2D (14). Specifically, FIB-4 is the most widely validated NIT, being the main recommended index for suspected MASLD by both the AGA and National Institute for Health and Care Excellence, and has been shown to outperform, in accuracy, other indices in identifying individuals with a low probability of advanced fibrosis (15). A FIB-4 score <1.3 is strongly associated with a high negative predictive value for advanced hepatic fibrosis (F3-4). Since FIB-4 is influenced by age, another cut-off of FIB-4 <2.0 was suggested to more accurately exclude advanced hepatic fibrosis in individuals ≥65 years (16). Alternative NITs have been developed for identifying both hepatic fibrosis and hepatic steatosis without successful implementation in diagnostic guidelines for MASLD or MASH; however, there is a lack of comprehensive head-to-head comparison studies validating the most studied NITs (Supplementary Table 1) in stratifying risk for adverse outcomes.

Currently, there is a need for a wider healthcare perspective with the realization that MASLD is the new member of the cardiometabolic block, shifting MASLD and MASH identification from liver excellence centers to primary care (17, 18). In such settings, patients with (non-invasively identified) MASH may be eligible to receive the first-ever drug approved for noncirrhotic MASH with moderate to

advanced liver fibrosis – resmetirom (19). Additionally, in patients with obesity and other cardiometabolic diseases, emerging pharmacotherapies targeting obesity, such as glucagon-like peptide-1/glucagon receptor co-agonists, could also play a crucial role in reducing liver fat and improving metabolic outcomes, i.e., in patients with MASLD/MASH (20, 21). This will lead to an increased need for the implementation of simple NITs in primary care settings that would predict not only liver outcomes to guide therapeutics but also an expanded stratification for other cardio-renal-metabolic outcomes and malignancies, all of which are on the rise globally (22). Specifically, the recent practice guidance by the American Association for the Study of Liver Diseases encourages reassessing individuals through NITs every one to three years (based on the individual's cardiometabolic profile) regarding subjects' risk of adverse outcomes (14, 23, 24). Besides, several studies have investigated the prognostic role of NITs in cardiometabolic outcomes and mortality; however, notable limitations include small sample sizes, limited observation periods of ≤10 years, old terminology (linked to overlooking cardiometabolic risk factors and including moderate drinking), non-representative populations from hepatology clinics or tertiary care, cross-sectional analyses, and most importantly limited sensitivity, specificity, and accuracy of existing NITs (25-30). Importantly, no study has proposed a data-driven modified version of the clinically established NIT to significantly increase their prognostic performance for cardio-renal-metabolic morbidity, malignancies, and cause-specific and all-cause mortality. The improvement of established NITs in a data-driven way is essential, considering the suboptimal performance of available NITs in forecasting risk for adverse outcomes (31). For this purpose, large prospective population cohorts comprise a suitable platform for model development through the identification of highly predictive markers leading to novel models. In this study, we compare established NITs' prognostic performance for adverse outcomes and evaluate their performance against that of the improved NITs developed herein.

Methods

Data source

The UK Biobank comprises one of the world's largest longitudinal population-based biobanks, including 502,359 individuals (age range: 37-73 years) recruited between 2006 and 2010, with follow-up data until October 2022 (32). Ethical procedures are controlled by a dedicated Ethics and Guidance Council (http://www.ukbiobank.ac.uk/ethics), and all participants have given written informed consent before their enrolment. The complete catalog of all UK Biobank features and measurements is available on the following website: https://biobank.ndph.ox.ac.uk/showcase/catalogs.cgi. Patients and the public were not involved in any way.

Study population

In our study, simulating a real-life scenario, we included participants comprising the main target groups for cardiometabolic screening belonging to a) the general population, b) according to the latest clinical guidance, individuals eligible for a potential diagnosis of MASLD (suspicion of steatotic liver disease not verified by NITs or imaging/histology and excluding individuals with excessive drinking behavior [≥210 g for males or ≥140 g for females, per week and daily or almost daily alcohol intake frequency], secondary causes of liver steatosis [Supplementary Table 2], and individuals without cardiometabolic risk factors (4)), and c) individuals with T2D. Individuals having any of the included conditions of cardiometabolic morbidity and malignancies at baseline were also excluded from the analyses.

Endpoints

Five endpoints of incident morbidity were studied: (i) T2D, (ii) chronic kidney disease (CKD), (iii) composite hepatic events (CHE), (iv) CVD, and (v) malignancies. Three mortality outcomes were studied herein: (i) circulatory disease-related mortality, (ii) malignancies-related mortality, and (iii) all-cause mortality. Endpoints were identified by self-reported disease, disease diagnosed by a doctor, the International Classification of Diseases (ICD) 9 codes, and the ICD10 codes (Supplementary Table 3). The incident T2D endpoint was not included in the analyses of the individuals with T2D. T2D was annotated based on self-reported type 2 diabetes (data-field 20002), diabetes diagnosed by a doctor (data-field 2443), ICD9 (data-field 41271) codes 250.X0 and 250.X2 and ICD10 (data-field 41270) codes E11. CKD was annotated based on self-reported renal/kidney failure, ICD9 codes 581-586, and ICD10 codes I12-I13 and N18. CHE was annotated based on ICD10 codes K70-K76 and C22. CVD was annotated based on ICD10 codes I20-I25, I50, I60-64 and I70-74. Malignancies were annotated based on all ICD10 codes from the C category. For the mortality outcomes, the ICD10 death register (data-field 40001) was used for annotation. Circulatory disease-related mortality was annotated based on all ICD10 death register codes from the I category. Malignancies-related mortality was annotated based on all ICD10 death register codes from the C category. All-cause mortality was annotated based on any ICD10 code reported in the death register.

Proteomic analysis

The purpose of the proteomic analysis was to find the best predictive proteomic feature that could be added to the existing NIT's to better predict cardio-renal-metabolic diseases specific and overall morbidity and mortality, i.e., the adverse outcomes studied herein. For this analysis, we included everyone with available omics data. Participants and omics features with more than 20% of missing data were excluded from the analysis. The prediction target for these analyses was computed by taking the number of diseases mentioned in the "Endpoints"

section per participant multiplied by 10. The best feature to be added on top of the mentioned indices was found by first removing the variance explained by the index from the data and subsequently extracting the highest correlated omics feature to the prediction target. The final model scores were then computed by fitting a linear regression model on the existing NIT and the top omics feature using the Python statsmodels library (0.13.5) with 10-fold cross-validation. The cutoffs for classification into high-risk and low-risk were created so that the group sizes were the same as the group sizes of the existing NIT classifications based on the general population. Detailed quality control procedures can be found here (https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/PPP_Phase_1_QC_dataset_companion_doc.pdf), and the translation of plasma protein levels into Normalized Protein eXpression (NPX) values here (https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/Olink_1536_B0_to_B7_Normalization.pdf). Detailed information on proteomic measurements in the UK Biobank has been described elsewhere (33).

Statistical analyses

Statistical analyses were performed using R version 4.3.2 with the following packages: survival (3.5-7), survminer (0.4.9), ggsurvfit (1.0.0), ROCR (1.0-11), MLmetrics (1.1.1) and epiR (2.0.67). For every included condition, prevalent cases were removed before doing the analysis. For all analyses, individuals were followed from baseline until the time of emergence of one of our endpoints or the maximum follow-up time of 16.6 years, whichever came first. Hazard ratios (HRs), C-indices, and the 95% confidence interval (CI) were estimated using Cox proportional hazard models. Specifically, HRs were computed for the NIT classifications, and C-indices were computed for the NIT continuous values. To investigate the clinical utility of the existing NITs and novel scores developed herein, we calculated how many individuals need to be screened to capture a (true positive [TP]) case according to the following formula: (TP + false positive [FP] + true negative [TN] + false negative [FN]) / TP, as proposed by the authors “Screening Efficiency Ratio” (SER). This metric, however, is biased by the predicted risk group size (individuals surpassing the cut-off of an index); therefore, we also added a complementary metric, “False Negative Burden Ratio” (FNBR), to demonstrate the burden of missed cases as the number of individuals that are screened for every false negative (missed case): (TP + FP + TN + FN) / FN. Cumulative incidences were plotted by computing survival curves and then inverting the data. A cutoff of 15 years was used for the analyses due to the minimal number of events after 15 years of follow-up.

Results

The total number of individuals from the general population, from the potentially MASLD population, and from the T2D population that met our inclusion criteria across each outcome is presented in Supplementary Table 4, and the baseline population characteristics are presented in Table 1, Table 2, and Supplementary Table 5. FIB-4 captures older individuals overall, and Lipid Accumulation Product (LAP) captures individuals with the highest mean body mass index (BMI) and mean waist circumference, which were the highest among individuals with T2D (Table 1 and Table 2). While the ethnic makeup of the UK Biobank is not diverse and almost exclusively comprises a White population, the group of individuals with T2D was the most ethnically diverse (Table 2).

Table 1. Baseline characteristics of the general population.
Abbreviations: FIB-4, Fibrosis 4 Index; HSI, Hepatic Steatosis Index; FLI, Fatty Liver Index; LAP, Lipid Accumulation Product; MAF-5, metabolic dysfunction–associated fibrosis score; APRI, Aspartate aminotransferase to Platelet Ratio Index. Data are presented as the mean (standard deviation) unless otherwise noted or, if not normally distributed, as the median (interquartile range).

	FIB-4 > 1.3 or 2 (N=160454)	FIB-4 < 1.3 or 2 (N=295296)	FIB4-G > 8.9 (N=14162)	FIB4-G < 8.9 (N=26064)	HSI ≥ 36 (N=222889)	HSI < 35 (N=238617)	HSI-G > 7.2 (N=19745)	HSI-G < 7.2 (N=21138)	FLI ≥ 60 (N=178820)	FLI < 60 (N=287982)	FLI-G > 8.6 (N=15744)
Age (years)	61 (8)	56 (16)	64 (8)	55 (13)	59 (13)	58 (14)	62 (10)	54 (14)	59 (12)	58 (14)	63 (9)
Genetic sex (male)	52.70%	42.20%	55.80%	41.10%	47.60%	44.10%	53.70%	39%	63.60%	34.70%	58.20%
Waist circumference (cm)	89 (19)	90 (18)	94 (19)	87.2 (17)	98 (16)	82 (15)	95 (18)	85 (16)	101 (12)	83 (13)	97 (17)
Body mass index (kg/m²)	26.3 (5.5)	27 (5.8)	27.8 (6.3)	26.3 (5.3)	30 (4.7)	24.3 (3.2)	28.4 (6.2)	25.5 (4.8)	30.6 (5.3)	24.8 (3.8)	28.8 (6.1)
Platelet count (10^9 cells/L)	210.4 (53.6)	269.8 (67.5)	239.6 (77.3)	252 (71.8)	250.8 (75.5)	246 (72)	245.4 (75.7)	250 (71.7)	246.8 (75.5)	249 (72.7)	244 (77)
Alanine aminotransferase (U/L)	19.9 (12)	20.3 (12)	21.4 (13.2)	19.3 (11.1)	25 (15)	17.1 (7.9)	22.1 (13.5)	18.3 (9.8)	26.1 (15.4)	17.6 (8.4)	22.6 (13.9)
Aspartate aminotransferase (U/L)	26.9 (8.9)	23.2 (6.9)	25.9 (9.1)	23.8 (7.2)	25.2 (8.8)	23.8 (7)	25.6 (8.7)	23.5 (7)	26.3 (9.2)	23.4 (6.9)	26 (9)
Gamma-glutamyl transferase (U/L)	26 (23)	26.5 (22.2)	30.6 (27.6)	24.4 (19.4)	33 (28.5)	21.8 (14.9)	31.2 (27.3)	22.7 (17)	39.7 (33.2)	21.1 (13)	33.4 (29.7)
Triglycerides (mmol/L)	1.4 (1.04)	1.53 (1.13)	1.66 (1.2)	1.39 (1.01)	1.79 (1.25)	1.24 (0.84)	1.7 (1.2)	1.3 (0.93)	2.13 (1.35)	1.2 (0.72)	1.81 (1.27)
Type 2 diabetes prevalence before first assessment	5.20%	5.50%	13.20%	1.90%	9.60%	1.40%	10.70%	1.30%	9.90%	2.50%	12.60%
FIB-4	1.64 (0.59)	1.06 (0.35)	1.47 (0.71)	1.17 (0.52)	1.16 (0.55)	1.33 (0.62)	1.36 (0.65)	1.17 (0.54)	1.22 (0.6)	1.26 (0.59)	1.38 (0.66)
FIB4-G	8.16 (6.27)	6.46 (6.41)	12.1 (4.9)	5.06 (4.12)	7.7 (6.65)	6.58 (6.06)	10.7 (5)	4.26 (3.68)	8.53 (7.08)	6.34 (5.85)	11.7 (5.1)
HSI	33.3 (6.9)	35.6 (7.7)	35.8 (8.4)	34.2 (7)	39.1 (5.8)	31.5 (3.7)	36.8 (8.3)	33.1 (6)	39.9 (6.9)	32.4 (4.9)	37.3 (8.3)
HSI-G	7.45 (6.53)	6.83 (6.58)	12.2 (5.2)	4.98 (4.2)	8.75 (6.68)	5.62 (5.82)	10.8 (5.1)	4.15 (3.62)	9.57 (6.84)	5.69 (5.63)	11.8 (5)
FLI	41.7 (55.2)	48.4 (55.2)	61.7 (52.7)	38.6 (51.7)	74.7 (34.5)	22 (29.9)	65.8 (49.2)	30.2 (43.9)	81.8 (20)	25.3 (29.3)	71.7 (43.1)
FLI-G	7.57 (6.75)	6.79 (6.78)	12.4 (5.2)	4.94 (4.37)	8.82 (6.78)	5.55 (6.05)	10.9 (5.1)	4.08 (3.74)	10 (6.6)	5.41 (5.68)	12 (5)
LAP	25 (24.4)	23 (35.7)	26 (33.9)	22.6 (29.3)	29.7 (43.3)	21 (22.7)	27 (35.7)	21.7 (27)	33 (34.1)	19.4 (26.4)	28 (36.5)
LAP-G	7.81 (6.28)	6.6 (6.39)	12.1 (5)	5.07 (4.06)	7.78 (6.73)	6.43 (5.97)	10.7 (5.1)	4.24 (3.57)	8.6 (7.06)	6.22 (5.73)	11.7 (5.1)
APRI	0.32 (0.14)	0.22 (0.08)	0.27 (0.15)	0.24 (0.11)	0.25 (0.13)	0.24 (0.11)	0.26 (0.13)	0.24 (0.1)	0.27 (0.14)	0.24 (0.11)	0.27 (0.14)
APRI-G	7.89 (6.21)	6.6 (6.33)	12.1 (4.9)	5.09 (4.05)	7.78 (6.6)	6.46 (5.94)	10.6 (5)	4.28 (3.56)	8.61 (6.94)	6.24 (5.69)	11.7 (5)
MAF-5	-0.55 (2.15)	-1.31 (2.1)	-0.3 (2.48)	-1.35 (1.92)	-0.17 (2.21)	-1.71 (1.65)	-0.32 (2.34)	-1.56 (1.76)	0.28 (1.9)	-1.74 (1.51)	-0.07 (2.33)
MAF5-G	8.07 (6.67)	6.29 (6.55)	12.2 (5.6)	4.86 (4.22)	8.51 (7.07)	5.69 (5.84)	10.8 (5.4)	4.05 (3.69)	9.72 (7.06)	5.45 (5.56)	11.9 (5.3)
White	94.30%	94.30%	94.60%	92.80%	93.80%	94.90%	94.40%	92.60%	94.20%	94.40%	94.50%
East Asian	0.70%	0.90%	0.60%	0.90%	0.70%	1%	0.60%	1%	0.60%	1%	0.60%
South Asian	1.30%	1.70%	1.70%	1.30%	1.60%	1.40%	1.60%	1.30%	1.70%	1.50%	1.70%
Black	1%	0.60%	0.80%	1.70%	1%	0.50%	1%	1.70%	0.80%	0.70%	0.90%
Caribbean	1.10%	0.90%	0.80%	1.20%	1.20%	0.80%	0.90%	1.30%	1%	1%	0.80%
Other	1.50%	1.60%	1.40%	2.10%	1.70%	1.40%	1.50%	2.10%	1.60%	1.50%	1.40%

	FLI-G < 8.6 (N=25355)	LAP > 47.67 (N=26806)	LAP < 47.67 (N=437014)	LAP-G > 17.3 (N=2376)	LAP-G < 17.3 (N=38729)	APRI > 0.5 (N=20932)	APRI < 0.5 (N=434924)	APRI-G > 18.3 (N=1847)	APRI-G < 18.3 (N=38387)	MAF-5 ≥ 1 (N=60222)	MAF-5 < 1 (N=393749)	MAF5-G > 13.6 (N=5312)	MAF5-G < 13.6 (N=34729)
	55 (14)	58 (13)	58 (13)	64 (8)	58 (14)	60 (13)	58 (13)	64 (8)	58 (13)	60 (12)	58 (13)	64 (8)	58 (14)
	38.60%	69.10%	44.40%	65.70%	45%	72.30%	44.60%	64%	45.40%	75.10%	41.40%	68%	42.90%
	85 (17)	116 (11)	89 (18)	100 (20)	89 (18)	97 (18)	89 (18)	100 (20)	89 (18)	109 (14)	88 (17)	102 (19)	88 (18)
	25.7 (4.9)	34.6 (5.9)	26.5 (5.3)	29.4 (7.4)	26.6 (5.6)	28.4 (6.4)	26.7 (5.7)	29.5 (7.6)	26.7 (5.6)	32.4 (6.6)	26.1 (5)	30.1 (7.3)	26.4 (5.3)
	250 (71.8)	239 (73.6)	248.8 (74)	240.7 (89.6)	248.3 (73)	178.3 (66.5)	250.8 (72)	237.6 (89.8)	248.3 (73)	225.9 (70.6)	251.4 (73.2)	233.9 (79.4)	250 (72.6)
	18.6 (10.1)	25.4 (17.9)	19.9 (11.6)	21.8 (16.1)	19.9 (11.6)	42.8 (34.8)	19.7 (11.1)	22.2 (17.4)	19.9 (11.6)	33.1 (22.5)	19.1 (10.1)	25 (18.3)	19.5 (10.9)
	23.6 (7)	25.8 (10.1)	24.3 (7.7)	26.2 (12.8)	24.4 (7.7)	44.4 (21.6)	24.1 (7.3)	26.6 (14.4)	24.4 (7.7)	32 (14.7)	23.8 (7)	28 (12.7)	24.1 (7.2)
	23 (17.1)	35.9 (32.2)	25.8 (21.7)	35.8 (41.7)	25.9 (21.5)	51.8 (67.6)	25.7 (21.1)	36.7 (46)	26 (21.6)	44.2 (41.3)	24.5 (19)	37.9 (37.7)	25 (20)
	1.31 (0.91)	1.59 (1.67)	1.48 (1.07)	1.78 (1.27)	1.46 (1.08)	1.69 (1.42)	1.48 (1.09)	1.82 (1.33)	1.46 (1.08)	1.99 (1.42)	1.42 (1.02)	1.86 (1.36)	1.43 (1.04)
	1.60%	17.70%	4.60%	40.10%	3.80%	11.30%	5.10%	42.60%	4.10%	30.60%	1.50%	30.70%	2%
	1.19 (0.55)	1.25 (0.64)	1.25 (0.6)	1.49 (0.85)	1.25 (0.59)	2.31 (1.14)	1.22 (0.56)	1.54 (0.89)	1.25 (0.59)	1.46 (0.78)	1.22 (0.57)	1.53 (0.8)	1.23 (0.57)
	4.87 (4.06)	10 (8.5)	6.98 (6.24)	20.4 (5.5)	6.78 (5.86)	11.5 (8.4)	6.95 (6.24)	21.6 (5.4)	6.85 (5.97)	11.3 (8.5)	6.63 (5.9)	16.5 (5.7)	6.37 (5.31)
	33.4 (6.3)	43.7 (7.7)	34.4 (7)	38.1 (10.2)	34.6 (7.3)	37.3 (9.3)	34.7 (7.4)	38.2 (10.4)	34.6 (7.4)	42.4 (8.1)	34 (6.5)	39.1 (9.7)	34.2 (7)
	4.78 (4.02)	11.8 (8.7)	6.85 (6.35)	20.9 (5.7)	6.72 (6.03)	10.7 (8.8)	6.92 (6.41)	22 (5.5)	6.78 (6.12)	12.6 (8.3)	6.43 (5.91)	16.7 (5.6)	6.3 (5.4)
	31.7 (44.4)	94.3 (17.2)	43.8 (53)	77.3 (45.9)	44.6 (54.3)	75.8 (48.2)	44.8 (54.7)	78.7 (45.4)	44.9 (54.4)	91.4 (16.1)	39 (48.6)	81.7 (35.7)	41 (51.6)
	4.74 (4.19)	11.7 (8.4)	6.89 (6.55)	20.7 (5.3)	6.75 (6.24)	11 (8.9)	6.94 (6.62)	21.8 (5.2)	6.81 (6.32)	12.6 (8)	6.43 (6.1)	16.8 (5.4)	6.29 (5.59)
	21.3 (27.8)	53 (8.9)	22 (31.9)	31 (28.1)	23 (31.1)	31 (21)	23 (32)	30.3 (32.6)	23.3 (31.2)	41 (19)	21.7 (31.2)	34 (28)	22.5 (30.8)
	4.87 (3.96)	10.5 (8.5)	6.94 (6.2)	20.7 (5.4)	6.75 (5.84)	10.6 (8.3)	6.95 (6.22)	21.8 (5.5)	6.82 (5.92)	11.3 (8.6)	6.57 (5.83)	16.6 (5.8)	6.31 (5.24)
	0.24 (0.11)	0.27 (0.15)	0.25 (0.12)	0.28 (0.19)	0.25 (0.12)	0.61 (0.21)	0.24 (0.11)	0.28 (0.22)	0.25 (0.12)	0.37 (0.22)	0.24 (0.11)	0.31 (0.19)	0.24 (0.11)
	4.91 (3.95)	10.1 (8.5)	6.96 (6.15)	20.5 (5.5)	6.79 (5.78)	11.2 (8.4)	6.95 (6.16)	21.7 (5.5)	6.85 (5.89)	11.3 (8.5)	6.62 (5.8)	16.5 (5.7)	6.34 (5.2)
	-1.5 (1.78)	1.88 (2.76)	-1.13 (2.04)	1.03 (3.28)	-1.09 (2.08)	1.8 (2.39)	-1.13 (2.04)	1.18 (3.45)	-1.08 (2.09)	1.96 (1.54)	-1.3 (1.79)	1.21 (2.81)	-1.23 (1.92)
	4.68 (4.05)	12.8 (9.7)	6.76 (6.38)	21.2 (6.5)	6.63 (6.1)	12.9 (9.2)	6.77 (6.45)	22.5 (6.3)	6.7 (6.2)	14 (8.4)	6.22 (5.76)	17 (5.7)	6.17 (5.45)
	92.80%	91.40%	94.50%	92.10%	93.50%	92.80%	94.40%	92.40%	93.50%	93%	94.60%	93.30%	93.60%
	0.90%	0.40%	0.90%	0.60%	0.80%	0.80%	0.80%	0.70%	0.80%	0.70%	0.80%	0.70%	0.80%
	1.30%	1.30%	1.60%	3.20%	1.40%	1.50%	1.60%	3.20%	1.40%	2%	1.40%	2.40%	1.30%
	1.60%	2.40%	0.60%	1%	1.40%	1.60%	0.70%	0.80%	1.40%	1.10%	0.70%	1%	1.40%
	1.20%	2.50%	0.90%	1.30%	1.10%	1.40%	1%	1.10%	1%	1.30%	0.90%	0.90%	1.10%
	2.10%	2%	1.50%	1.80%	1.80%	1.80%	1.60%	1.70%	1.80%	1.80%	1.50%	1.70%	1.80%

Table 2. Baseline characteristics of the population with type 2 diabetes mellitus.
Abbreviations: FIB-4, Fibrosis 4 Index; HSI, Hepatic Steatosis Index; FLI, Fatty Liver Index; LAP, Lipid Accumulation Product; MAF-5, metabolic dysfunction–associated fibrosis score; APRI, Aspartate aminotransferase to Platelet Ratio Index. Data are presented as the mean (standard deviation) unless otherwise noted or, if not normally distributed, as the median (interquartile range).

	FIB-4 > 1.3 or 2 (N=8082)	FIB-4 < 1.3 or 2 (N=15242)	FIB4-G > 8.9 (N=1811)	FIB4-G < 8.9 (N=433)	HSI ≥ 36 (N=20445)	HSI < 35 (N=3083)	HSI-G > 7.2 (N=2060)	HSI-G < 7.2 (N=219)	FLI ≥ 60 (N=17214)	FLI < 60 (N=6547)	FLI-G > 8.6 (N=1933)
Age (years)	62 (6)	62 (12)	63 (9)	57 (14)	62 (10)	63 (10)	63 (9)	55 (14)	62 (10)	62 (11)	63 (9)
Genetic sex (male)	72.50%	58.70%	66.70%	53.90%	62.10%	71%	65.20%	49.30%	66.10%	55.80%	66.10%
Waist circumference (cm)	102 (19)	102 (17)	104 (19)	97 (20)	104 (16)	87 (13)	103 (19)	90 (18)	107 (15)	89 (11)	104 (18)
Body mass index (kg/m²)	30.3 (7.2)	30.6 (7.3)	31.1 (7)	28.8 (6.1)	31.4 (6.7)	24.1 (2.7)	31.1 (7)	26.6 (5.3)	32.5 (6.5)	25.8 (3.6)	31.3 (6.9)
Platelet count (10⁹ cells/L)	197.6 (57.4)	265.5 (72.9)	239 (86.8)	243.6 (80.2)	242.2 (81.3)	234 (80)	240.8 (85.4)	237.8 (78.4)	241 (81.3)	241.7 (81.1)	240 (86.3)
Alanine aminotransferase (U/L)	26.6 (19.6)	24.2 (14.1)	24.3 (15.9)	24.1 (13.3)	26.3 (16.4)	18.3 (8.5)	24.4 (15.7)	21.5 (11.1)	27.2 (17.3)	20.2 (10)	24.6 (15.8)
Aspartate aminotransferase (U/L)	29.5 (14.1)	23.3 (8.1)	25.3 (10.9)	24.4 (7.9)	25.3 (10.4)	24.1 (8.2)	25.2 (10.7)	23.9 (7.2)	25.9 (11.1)	23.5 (7.7)	25.3 (10.7)
Gamma-glutamyl transferase (U/L)	39.5 (42.9)	33.9 (27.2)	35.6 (32.5)	30.5 (22.6)	37.3 (32.8)	24.6 (18.2)	35.5 (31.8)	27.8 (18.9)	41.5 (36.4)	23.8 (14.7)	36.1 (32.7)
Triglycerides (mmol/L)	1.77 (1.38)	1.86 (1.31)	1.88 (1.34)	1.62 (1.3)	1.92 (1.35)	1.23 (0.92)	1.88 (1.33)	1.31 (1.06)	2.12 (1.38)	1.2 (0.76)	1.92 (1.35)
FIB-4	1.77 (0.77)	1.08 (0.39)	1.35 (0.75)	1.15 (0.58)	1.25 (0.66)	1.51 (0.74)	1.32 (0.73)	1.17 (0.62)	1.27 (0.68)	1.31 (0.67)	1.32 (0.73)
FIB4-G	15.8 (9.5)	14.7 (10.4)	17 (8.5)	6.21 (3.59)	15.5 (10)	12.8 (10.7)	16.1 (9.3)	4.26 (3.06)	16.1 (9.7)	12.2 (9.7)	16.7 (8.8)
HSI	40.4 (8.7)	42.3 (8.7)	41.9 (8.4)	39.6 (8.5)	42.8 (7.9)	33.1 (2.6)	41.9 (8.3)	37.5 (7.4)	43.9 (7.6)	35.8 (4.9)	42.1 (8.2)
HSI-G	16.4 (10.2)	16.4 (10.4)	18.2 (8.4)	7.27 (3.77)	17 (9.8)	12.1 (10.5)	17.3 (9.2)	5.26 (2.84)	17.8 (9.7)	12.2 (9.9)	17.8 (8.9)
FLI	82.5 (39.3)	81.4 (36.6)	84 (33)	65.6 (54.2)	85.7 (27.2)	28 (30.7)	83.7 (33.4)	44.8 (49.4)	89.5 (17.5)	37.7 (27.9)	84.7 (30.5)
FLI-G	16.1 (9.7)	15.8 (10.3)	17.7 (8.3)	6.9 (4.03)	16.6 (9.7)	11.9 (10.4)	16.9 (9.3)	4.68 (2.93)	17.4 (9.3)	11.5 (9.9)	17.5 (8.6)
LAP	35 (23)	32 (37.3)	35 (30)	30 (27.7)	35 (33.4)	24 (13)	34 (30)	25.7 (26.5)	38 (32.2)	25 (17.1)	35 (31)
LAP-G	15.6 (10)	15.1 (10.6)	17.3 (8.7)	6.43 (3.46)	15.8 (10.2)	12.9 (11)	16.3 (9.6)	4.4 (3.11)	16.4 (10.1)	12.3 (10.6)	16.9 (8.9)
APRI	0.38 (0.21)	0.22 (0.1)	0.27 (0.17)	0.25 (0.11)	0.27 (0.16)	0.26 (0.14)	0.27 (0.17)	0.24 (0.11)	0.27 (0.17)	0.25 (0.13)	0.27 (0.17)
APRI-G	15.6 (9.8)	14.9 (10.5)	17.1 (8.6)	6.27 (3.48)	15.6 (10.2)	12.8 (11)	16.2 (9.4)	4.5 (2.96)	16.2 (9.9)	12 (10)	16.8 (8.8)
MAF-5	3 (2.62)	1.81 (2.14)	2.41 (2.51)	1.5 (2.15)	2.43 (2.34)	0.8 (1.49)	2.39 (2.44)	1.13 (1.94)	2.79 (2.22)	0.8 (1.41)	2.43 (2.47)
MAF5-G	19.1 (10.2)	17.3 (9.8)	19.7 (8.2)	9.02 (4.03)	18.6 (9.7)	13.9 (9.6)	18.9 (9)	6.86 (3.07)	19.4 (9.4)	13.3 (8.9)	19.4 (8.4)
White	88.10%	86.90%	86.40%	83.40%	88.40%	81.10%	86.20%	83.10%	89.80%	81.10%	86.70%
East Asian	1.40%	1.60%	1.20%	2.10%	1.30%	3%	1.40%	0.90%	1%	2.80%	1.20%
South Asian	4.20%	5.60%	5.10%	5.30%	4.40%	9.10%	4.90%	5.90%	3.90%	7.90%	4.90%
Black	1.50%	1.30%	2.30%	3%	1.30%	1.70%	2.60%	2.70%	1.10%	2.20%	2.20%
Caribbean	2.20%	2%	2.30%	3%	2%	2.20%	2.50%	2.70%	1.80%	2.80%	2.30%
Other	2.70%	2.60%	2.70%	3.20%	2.60%	2.80%	2.50%	4.60%	2.40%	3.20%	2.60%

Given the similar performances observed between the general population and the potentially MASLD population, herein we report results for the general population and the T2D population; detailed performance of all analyses in the potentially MASLD population can be found in Supplementary Tables 5 and 6 and Supplementary Figures 1, 5, and 7. Initially, we compared established NITs for fibrosis and steatosis (Fig. 1A, Fig. 2A, Table 3, Table 4).

	FLI-G < 8.6 (N=357)	LAP > 47.67 (N=4608)	LAP < 47.67 (N=19105)	LAP-G > 17.3 (N=928)	LAP-G < 17.3 (N=1369)	APRI > 0.5 (N=2299)	APRI < 0.5 (N=21030)	APRI-G > 18.3 (N=769)	APRI-G < 18.3 (N=1475)	MAF-5 ≥ 1 (N=17792)	MAF-5 < 1 (N=5303)	MAF5-G > 13.6 (N=1583)	MAF5-G < 13.6 (N=636)
	56 (14)	61 (10)	62 (10)	64 (8)	61 (11)	62 (9)	62 (10)	64 (8)	61 (11)	62 (10)	62 (11)	63 (9)	59 (13)
	50.70%	86.50%	57.80%	67%	61.70%	77.50%	62%	66.10%	63.20%	70.40%	40.20%	67.90%	54.60%
	92 (17)	119 (11)	99 (15)	106 (18)	100 (19)	107 (18)	102 (18)	106 (20)	101 (19)	106 (17)	89 (12)	106 (19)	95 (18)
	27.2 (5.8)	36.1 (5.8)	29.4 (5.8)	31.7 (6.7)	29.9 (6.9)	31.9 (7.1)	30.4 (7.2)	32 (6.9)	30.1 (6.8)	31.8 (7)	26.5 (4.8)	31.8 (6.9)	28 (5.5)
	238.6 (76.7)	235.3 (76.3)	243 (82.3)	242 (94.4)	238.6 (80.1)	178.5 (68)	247 (77.5)	241.4 (96.9)	239 (81.8)	233 (76.4)	273 (84.9)	238 (86.4)	244 (83.2)
	22 (11.9)	27.7 (19.4)	24.3 (14.9)	23.5 (16)	24.6 (15.1)	48.9 (32.9)	23.7 (13.6)	23.6 (16.6)	24.6 (14.8)	27.5 (16.9)	18.4 (8.3)	25.1 (16.5)	23 (12.3)
	24 (7.4)	26.1 (12.1)	24.9 (9.6)	25 (12)	25.3 (9.3)	47.1 (22.4)	24.3 (8.5)	25.2 (12.7)	25.1 (9.2)	26.9 (10.8)	20.4 (5.7)	25.6 (11.5)	24 (7.7)
	27.5 (18.7)	41.1 (36.9)	34 (30.2)	35.5 (37.8)	33.8 (26.8)	72 (88.8)	33.6 (27.5)	35.7 (38.6)	33.8 (27.5)	39.2 (34.8)	25.5 (17.8)	37.1 (35.1)	29.6 (22.7)
	1.37 (1.06)	1.95 (1.56)	1.81 (1.29)	1.93 (1.33)	1.76 (1.34)	2.06 (1.61)	1.81 (1.3)	1.98 (1.37)	1.76 (1.31)	1.93 (1.38)	1.49 (1.11)	1.92 (1.39)	1.58 (1.2)
	1.21 (0.69)	1.3 (0.72)	1.28 (0.67)	1.35 (0.77)	1.28 (0.67)	2.37 (1.08)	1.23 (0.59)	1.39 (0.84)	1.27 (0.66)	1.37 (0.7)	1.05 (0.51)	1.34 (0.76)	1.21 (0.64)
	5.36 (3.89)	16.4 (10.1)	14.7 (10.4)	21.5 (5.9)	11 (6.1)	18 (9.9)	14.8 (10.1)	22.4 (5.9)	11.5 (6.8)	15.6 (9.9)	12.9 (11.2)	17.8 (7.8)	7.79 (4.41)
	37.9 (8)	47.1 (7.2)	40.3 (7.9)	42.1 (8.1)	41 (8.8)	43.1 (8.5)	41.5 (8.7)	42.2 (8.2)	41 (8.5)	43.1 (8.5)	37.1 (6.5)	42.4 (8.3)	38.7 (7.6)
	6.62 (3.39)	18.6 (9.4)	15.6 (10.5)	22.8 (6.1)	12.3 (6.5)	18.6 (9)	16.1 (10.3)	23.5 (6)	12.9 (7)	17.1 (9.7)	13.3 (11.4)	19.3 (7.7)	8.7 (4.12)
	47.9 (52)	96.8 (6)	75.3 (38.9)	85.7 (28.1)	76.8 (44.6)	92.8 (18.5)	80.1 (38.5)	86.8 (27.7)	78 (42.6)	87.8 (23.7)	43.6 (42.4)	86.9 (27.5)	59.7 (50.2)
	6.09 (3.25)	17.8 (9.4)	15.3 (10.3)	22.2 (5.9)	11.8 (6.5)	17.9 (9.1)	15.7 (10.1)	23 (5.8)	12.4 (6.9)	16.7 (9.6)	12.7 (11.5)	18.7 (7.5)	8.15 (4.4)
	26 (25.8)	55 (11)	28.9 (35.7)	37 (29.2)	32 (30)	39 (22)	33 (31)	36.5 (33)	33 (27.5)	37 (23)	20 (39.5)	37 (29)	28 (28.4)
	5.61 (3.48)	16.9 (10.1)	14.8 (10.6)	22 (6.1)	11.1 (6.4)	17 (9.4)	15 (10.4)	22.7 (6.1)	11.7 (6.8)	15.8 (10)	13.2 (11.5)	18.2 (8)	7.61 (4.28)
	0.25 (0.11)	0.28 (0.19)	0.26 (0.15)	0.26 (0.19)	0.27 (0.14)	0.64 (0.26)	0.25 (0.13)	0.27 (0.2)	0.27 (0.14)	0.29 (0.17)	0.19 (0.08)	0.28 (0.18)	0.25 (0.11)
	5.55 (3.41)	16.5 (9.8)	14.8 (10.4)	21.7 (6)	11 (6.2)	17.6 (9.6)	14.9 (10.3)	22.6 (6)	11.6 (6.7)	15.7 (10)	13.2 (11.6)	17.9 (7.8)	7.73 (4.28)
	1.21 (1.95)	4.25 (2.27)	1.82 (1.95)	2.5 (2.65)	2.05 (2.23)	4.85 (2.43)	1.98 (2.15)	2.62 (2.69)	2.07 (2.2)	2.71 (2.08)	0.35 (0.87)	2.64 (2.46)	1.38 (1.85)
	8.37 (3.44)	21.3 (9.9)	16.9 (9.6)	24.1 (6.4)	13.9 (6.3)	22.7 (8.8)	17.5 (9.7)	24.7 (5.9)	14.5 (6.8)	18.9 (9.5)	13.7 (10.2)	20.4 (7.5)	10.6 (4.2)
	80.70%	90.40%	86.60%	87%	84.90%	90.90%	86.90%	87.50%	84.90%	88.80%	83.10%	87.10%	83.50%
	2%	0.50%	1.80%	1.20%	1.50%	1.20%	1.50%	1.30%	1.40%	1.30%	2.30%	1.20%	1.70%
	5.60%	2.70%	5.70%	5.90%	4.50%	2.70%	5.40%	5.70%	4.90%	4.10%	8%	4.90%	5.20%
	4.80%	1.60%	1.30%	1.40%	3.50%	1.30%	1.40%	0.90%	3.30%	1.40%	1.30%	2%	3.60%
	3.40%	2.60%	1.90%	2.20%	2.70%	1.60%	2.10%	2.10%	2.60%	2%	2.10%	2.10%	3.30%
	3.60%	2.10%	2.70%	2.40%	2.90%	2.30%	2.60%	2.50%	3%	2.40%	3.20%	2.80%	2.70%

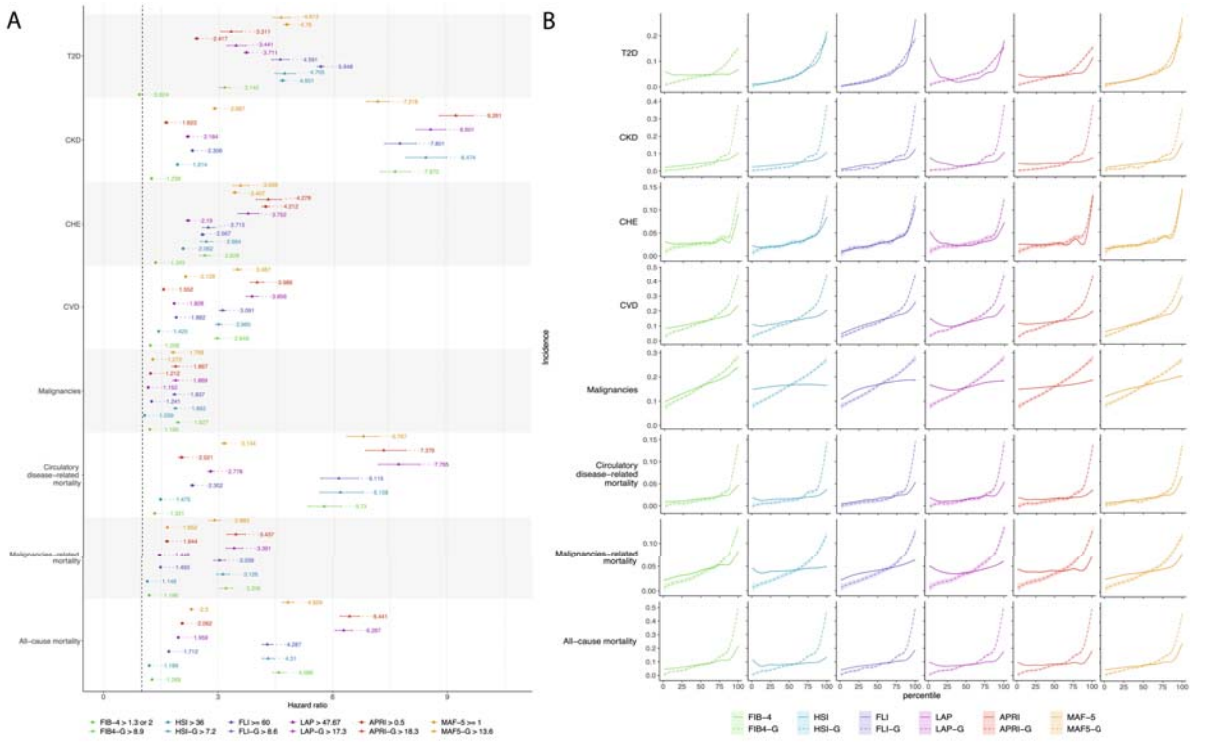


Figure 1. Panel A) Hazard ratios of incidence over a 15-year period of cardiometabolic morbidity, malignancies, specific-cause, and all-cause mortality according to the non-invasive indices' suggested thresholds for the general population. Panel B) Incidence of cardiometabolic morbidity, malignancies, specific-cause, and all-cause mortality over 15 years based on the continuous value percentiles of non-invasive tools for the general population.

The lighter-colored lines represent the 95% Confidence Interval. Abbreviations: T2D, type 2 diabetes mellitus; CKD, chronic kidney disease; CHE, composite hepatic events; CVD, cardiovascular disease; APRI, Aspartate aminotransferase to Platelet Ratio Index; FIB-4, Fibrosis-4; FLI, Fatty Liver Index; HSI, Hepatic Steatosis Index; LAP, Lipid Accumulation Product; MAF-5, metabolic dysfunction–associated fibrosis score.

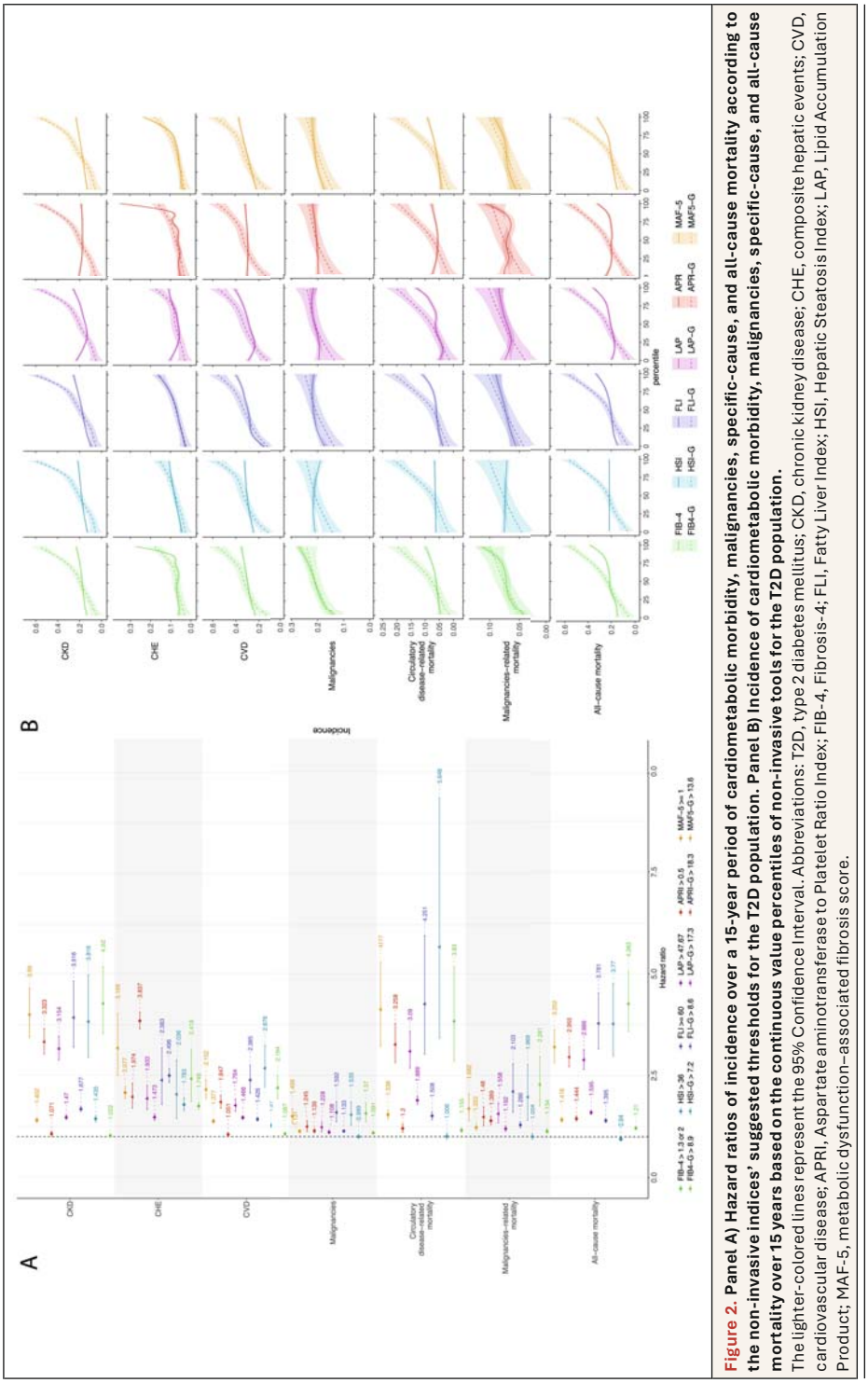


Figure 2. Panel A) Hazard ratios of incidence over a 15-year period of cardiometabolic morbidity, malignancies, specific-cause, and all-cause mortality according to the non-invasive indices' suggested thresholds for the T2D population. Panel B) Incidence of cardiometabolic morbidity, malignancies, specific-cause, and all-cause mortality over 15 years based on the continuous value percentiles of non-invasive tools for the T2D population.

The lighter-colored lines represent the 95% Confidence Interval. Abbreviations: T2D, type 2 diabetes mellitus; CKD, chronic kidney disease; CHE, composite hepatic events; CVD, cardiovascular disease; APRI, Aspartate aminotransferase to Platelet Ratio Index; FIB-4, Fibrosis-4; FLI, Fatty Liver Index; HSI, Hepatic Steatosis Index; LAP, Lipid Accumulation Product; MAF-5, metabolic dysfunction–associated fibrosis score.

Table 3. C-indices, hazard ratios, diagnostic metrics, absolute risk, absolute risk reduction, and number needed to screen for existing and novel non-invasive MASLD indices aiming to stratify risk for cardiometabolic morbidity, malignancies, cause-specific and all-cause mortality in the general population.
Risk group refers to the group of individuals having an index score above the threshold indicated. The control group refers to the individuals who do not have an index score above the threshold indicated. SER refers to the number of individuals needed to be screened to identify one case (lower is better), and FNBR refers to the number of individuals that are screened for every false negative identified (higher is better). Abbreviations: CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; SER, screening efficiency ratio; FNBR, false negative burden ratio; T2D, type 2 diabetes mellitus; CKD, chronic kidney disease; CHE, composite hepatic events; CVD, cardiovascular disease; FIB-4, Fibrosis 4 Index; HSI, Hepatic Steatosis Index; FLI, Fatty Liver Index; LAP, Lipid Accumulation Product; MAF-5, metabolic dysfunction–associated fibrosis score; APRI, Aspartate aminotransferase to Platelet Ratio Index; MASLD, metabolic dysfunction-associated steatotic liver disease.

Condition	Index and threshold	C-index (95% CI)	C-index P value	Hazard ratio (95% CI)	Hazard ratio P value	Sensitivity (95% CI)
T2D	FIB-4 > 1.3 or 2	0.51 (0.511 - 0.516)	3.00E-05	0.92 (0.911 - 0.938)	7.00E-08	33 (33-34)
	FIB4-G > 8.9	0.69 (0.686 - 0.698)	8.00E-235	3.14 (3.003 - 3.294)	1.00E-135	59 (57-61)
	HSI ≥ 36	0.74 (0.736 - 0.739)	0.00E+00	4.65 (4.573 - 4.73)	0.00E+00	79 (79-80)
	HSI-G > 7.2	0.74 (0.737 - 0.748)	0.00E+00	4.71 (4.451 - 4.975)	1.00E-170	79 (77-81)
	FLI ≥ 60	0.77 (0.77 - 0.773)	0.00E+00	5.65 (5.559 - 5.738)	0.00E+00	76 (75-76)
	FLI-G > 8.6	0.75 (0.741 - 0.752)	0.00E+00	4.59 (4.37 - 4.824)	4.00E-209	71 (69-73)
	LAP > 47.67	0.43 (0.427 - 0.431)	4.00E-22	3.71 (3.642 - 3.781)	0.00E+00	16 (15-16)
	LAP-G > 17.3	0.69 (0.688 - 0.7)	2.00E-253	3.44 (3.204 - 3.696)	6.00E-67	11 (10-13)
	APRI > 0.5	0.57 (0.567 - 0.571)	2.00E-20	2.42 (2.361 - 2.475)	6.00E-308	9 (9-10)
	APRI-G > 18.3	0.7 (0.693 - 0.704)	5.00E-256	3.31 (3.05 - 3.595)	6.00E-48	8 (7-10)
	MAF-5 ≥ 1	0.72 (0.723 - 0.726)	0.00E+00	4.76 (4.691 - 4.831)	0.00E+00	32 (32-33)
	MAF5-G > 13.6	0.74 (0.739 - 0.749)	0.00E+00	4.61 (4.393 - 4.845)	4.00E-214	32 (30-34)
CKD	FIB-4 > 1.3 or 2	0.62 (0.617 - 0.621)	2.00E-26	1.24 (1.222 - 1.256)	7.00E-54	40 (39-40)
	FIB4-G > 8.9	0.81 (0.804 - 0.813)	0.00E+00	7.67 (7.29 - 8.074)	0.00E+00	79 (78-81)
	HSI ≥ 36	0.61 (0.611 - 0.615)	0.00E+00	1.91 (1.887 - 1.941)	0.00E+00	64 (63-64)
	HSI-G > 7.2	0.81 (0.803 - 0.812)	0.00E+00	8.47 (7.955 - 9.028)	1.00E-249	88 (87-89)
	FLI ≥ 60	0.65 (0.647 - 0.651)	0.00E+00	2.31 (2.275 - 2.338)	0.00E+00	58 (58-59)
	FLI-G > 8.6	0.81 (0.803 - 0.812)	0.00E+00	7.8 (7.398 - 8.225)	0.00E+00	82 (80-83)
	LAP > 47.67	0.46 (0.462 - 0.466)	3.00E-06	2.18 (2.138 - 2.23)	7.00E-299	11 (11-12)
	LAP-G > 17.3	0.8 (0.8 - 0.809)	0.00E+00	8.6 (8.223 - 8.997)	0.00E+00	29 (27-31)
	APRI > 0.5	0.55 (0.543 - 0.547)	2.00E-08	1.62 (1.581 - 1.667)	1.00E-74	7 (7-7)
	APRI-G > 18.3	0.81 (0.802 - 0.81)	0.00E+00	9.26 (8.83 - 9.713)	0.00E+00	25 (23-27)
	MAF-5 ≥ 1	0.65 (0.653 - 0.656)	0.00E+00	2.89 (2.844 - 2.93)	0.00E+00	30 (29-30)
	MAF5-G > 13.6	0.81 (0.801 - 0.81)	0.00E+00	7.22 (6.925 - 7.524)	0.00E+00	49 (47-51)
CHE	FIB-4 > 1.3 or 2	0.57 (0.565 - 0.57)	4.00E-20	1.34 (1.322 - 1.367)	1.00E-69	41 (41-42)
	FIB4-G > 8.9	0.66 (0.654 - 0.67)	3.00E-163	2.63 (2.487 - 2.777)	8.00E-69	58 (55-61)
	HSI ≥ 36	0.62 (0.619 - 0.624)	0.00E+00	2.06 (2.026 - 2.098)	0.00E+00	66 (65-66)
	HSI-G > 7.2	0.68 (0.668 - 0.683)	1.00E-172	2.66 (2.51 - 2.827)	9.00E-61	71 (68-73)
	FLI ≥ 60	0.66 (0.66 - 0.665)	0.00E+00	2.57 (2.525 - 2.611)	0.00E+00	61 (60-62)
	FLI-G > 8.6	0.68 (0.672 - 0.687)	2.00E-180	2.72 (2.568 - 2.871)	1.00E-71	62 (59-65)
	LAP > 47.67	0.46 (0.46 - 0.466)	1.00E-04	2.19 (2.135 - 2.247)	4.00E-205	12 (11-12)
	LAP-G > 17.3	0.66 (0.652 - 0.668)	5.00E-153	3.75 (3.496 - 4.028)	8.00E-78	18 (16-20)
	APRI > 0.5	0.6 (0.596 - 0.601)	3.00E-40	4.21 (4.117 - 4.309)	0.00E+00	16 (15-16)
	APRI-G > 18.3	0.67 (0.657 - 0.673)	1.00E-166	4.28 (3.971 - 4.609)	1.00E-84	16 (14-18)
	MAF-5 ≥ 1	0.66 (0.662 - 0.667)	0.00E+00	3.41 (3.348 - 3.468)	0.00E+00	33 (32-34)
	MAF5-G > 13.6	0.68 (0.676 - 0.691)	9.00E-193	3.56 (3.357 - 3.768)	7.00E-107	34 (31-36)
CVD	FIB-4 > 1.3 or 2	0.59 (0.586 - 0.588)	2.00E-70	1.21 (1.198 - 1.218)	8.00E-109	38 (38-39)
	FIB4-G > 8.9	0.68 (0.681 - 0.688)	0.00E+00	2.95 (2.87 - 3.031)	0.00E+00	57 (55-58)
	HSI ≥ 36	0.56 (0.559 - 0.562)	0.00E+00	1.42 (1.413 - 1.437)	0.00E+00	56 (55-56)
	HSI-G > 7.2	0.69 (0.683 - 0.691)	0.00E+00	2.98 (2.898 - 3.073)	4.00E-304	70 (69-71)
	FLI ≥ 60	0.62 (0.614 - 0.616)	0.00E+00	1.88 (1.867 - 1.897)	0.00E+00	52 (51-52)
	FLI-G > 8.6	0.69 (0.689 - 0.696)	0.00E+00	3.09 (3.007 - 3.177)	0.00E+00	61 (60-62)
	LAP > 47.67	0.56 (0.559 - 0.561)	3.00E-114	1.83 (1.802 - 1.855)	0.00E+00	9 (9-9)
	LAP-G > 17.3	0.68 (0.681 - 0.688)	0.00E+00	3.86 (3.707 - 4.012)	5.00E-255	13 (12-14)
	APRI > 0.5	0.55 (0.546 - 0.549)	1.00E-32	1.55 (1.526 - 1.579)	6.00E-147	6 (6-7)

Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Risk group (n)	Cases in risk group (n)	Control group (n)	Cases in control group (n)	SER (lower is better)	FNBR (higher is better)
65 (65-65)	5 (5-5)	95 (95-95)	152372	7027	280054	14163	62	31
69 (68-69)	9 (9-10)	97 (97-97)	12351	1151	25631	790	33	48
55 (55-56)	8 (8-9)	98 (98-98)	202444	17009	235534	4396	26	100
56 (55-56)	9 (8-9)	98 (98-98)	17685	1561	20919	408	25	95
66 (65-66)	10 (10-10)	98 (98-98)	161606	16363	281435	5268	27	84
66 (66-67)	10 (10-11)	98 (97-98)	13811	1404	24998	579	28	67
95 (95-96)	15 (15-16)	96 (96-96)	22198	3361	417909	18198	131	24
97 (96-97)	15 (13-17)	95 (95-95)	1448	220	37360	1767	176	22
96 (96-96)	11 (10-11)	95 (95-95)	18633	1994	413894	19203	217	23
97 (97-98)	15 (13-17)	95 (95-95)	1078	161	36912	1782	236	21
91 (91-91)	16 (16-16)	96 (96-96)	42430	6827	388446	14189	63	30
91 (91-92)	16 (15-18)	96 (96-96)	3729	612	34093	1313	62	29
65 (65-65)	5 (5-5)	96 (95-96)	159890	8603	294703	13141	53	35
68 (67-68)	13 (13-14)	98 (98-98)	13966	1857	26055	484	22	83
53 (52-53)	6 (6-6)	97 (97-97)	222224	14022	238126	7952	33	58
54 (54-55)	11 (10-11)	99 (98-99)	19543	2104	21134	283	19	144
63 (63-63)	7 (7-7)	97 (97-97)	178130	12939	287505	9238	36	50
65 (64-65)	13 (12-13)	98 (98-98)	15548	1955	25345	435	21	94
95 (94-95)	9 (9-10)	96 (95-96)	26609	2527	436042	19594	183	24
96 (96-96)	31 (29-33)	96 (95-96)	2255	698	38641	1698	59	24
96 (95-96)	7 (7-8)	95 (95-95)	20822	1530	433877	20218	297	22
97 (97-97)	34 (32-36)	95 (95-96)	1737	590	38292	1752	68	23
88 (88-88)	11 (10-11)	96 (96-96)	59809	6382	393031	15200	71	30
89 (89-90)	22 (21-23)	97 (96-97)	5159	1135	34683	1193	35	33
65 (65-65)	4 (4-4)	97 (97-97)	159349	6066	294360	8554	75	53
66 (65-66)	6 (5-6)	98 (98-98)	13996	783	26023	568	51	70
52 (52-52)	4 (4-4)	98 (98-98)	221608	9673	237846	5075	47	91
53 (52-53)	5 (5-5)	98 (98-98)	19568	964	21110	398	42	102
63 (62-63)	5 (5-5)	98 (98-98)	177540	9068	287208	5794	51	80
63 (62-63)	5 (5-6)	98 (98-98)	15578	847	25317	518	48	79
94 (94-95)	6 (6-7)	97 (97-97)	26572	1719	435183	13118	269	35
95 (95-95)	10 (9-12)	97 (97-97)	2308	242	38590	1133	169	36
96 (96-96)	11 (11-12)	97 (97-97)	20400	2287	433415	12338	198	37
96 (96-96)	12 (11-14)	97 (97-97)	1782	214	38245	1138	187	35
88 (87-88)	8 (8-8)	98 (97-98)	59390	4818	392562	9705	94	47
88 (87-88)	9 (8-10)	97 (97-98)	5188	453	34650	885	88	45
66 (66-66)	15 (15-15)	87 (87-87)	150475	22407	282625	35877	19	12
71 (71-72)	25 (24-26)	91 (90-91)	12489	3095	25373	2365	12	16
54 (54-54)	16 (16-16)	89 (89-89)	208705	32934	230013	26082	13	17
58 (57-58)	22 (21-23)	92 (92-92)	17792	3895	20711	1658	10	23
65 (65-65)	19 (18-19)	90 (90-90)	165283	30735	278505	28885	14	15
68 (68-69)	24 (24-25)	91 (91-92)	13960	3411	24746	2165	11	18
95 (95-95)	22 (22-23)	87 (87-87)	23986	5343	416825	54040	83	8
97 (97-97)	41 (39-44)	87 (87-87)	1786	737	36914	4853	53	8
96 (96-96)	19 (19-20)	87 (87-87)	18912	3681	414289	54614	118	8

Table 3. C-indices, hazard ratios, diagnostic metrics, absolute risk, absolute risk reduction, and number needed to screen for existing and novel non-invasive MASLD indices aiming to stratify risk for cardiometabolic morbidity, malignancies, cause-specific and all-cause mortality in the general population. (continued)

Condition	Index and threshold	C-index (95% CI)	C-index P value	Hazard ratio (95% CI)	Hazard ratio P value	Sensitivity (95% CI)
	APRI-G > 18.3	0.68 (0.679 - 0.686)	0.00E+00	3.99 (3.815 - 4.164)	3.00E-218	11 (10-12)
	MAF-5 ≥ 1	0.62 (0.618 - 0.62)	0.00E+00	2.13 (2.107 - 2.15)	0.00E+00	22 (21-22)
	MAF5-G > 13.6	0.69 (0.688 - 0.695)	0.00E+00	3.49 (3.382 - 3.594)	0.00E+00	28 (26-29)
Malignancies	FIB-4 > 1.3 or 2	0.58 (0.575 - 0.578)	7.00E-34	1.2 (1.186 - 1.204)	5.00E-118	38 (38-39)
	FIB4-G > 8.9	0.61 (0.611 - 0.618)	9.00E-214	1.93 (1.88 - 1.976)	2.00E-151	48 (47-50)
	HSI ≥ 36	0.51 (0.508 - 0.51)	7.00E-11	1.06 (1.051 - 1.066)	2.00E-14	50 (49-50)
	HSI-G > 7.2	0.61 (0.603 - 0.61)	9.00E-184	1.86 (1.815 - 1.91)	3.00E-131	61 (60-63)
	FLI ≥ 60	0.54 (0.54 - 0.542)	5.00E-303	1.24 (1.232 - 1.251)	2.00E-183	43 (43-43)
	FLI-G > 8.6	0.61 (0.606 - 0.613)	8.00E-199	1.84 (1.792 - 1.883)	5.00E-133	51 (50-52)
	LAP > 47.67	0.52 (0.519 - 0.521)	4.00E-17	1.15 (1.135 - 1.169)	6.00E-21	7 (6-7)
	LAP-G > 17.3	0.61 (0.608 - 0.615)	1.00E-205	1.87 (1.79 - 1.951)	2.00E-47	9 (8-10)
	APRI > 0.5	0.52 (0.519 - 0.522)	9.00E-06	1.21 (1.192 - 1.233)	1.00E-30	5 (5-5)
	APRI-G > 18.3	0.61 (0.608 - 0.615)	2.00E-200	1.87 (1.778 - 1.96)	9.00E-38	7 (7-8)
	MAF-5 ≥ 1	0.54 (0.544 - 0.546)	0.00E+00	1.27 (1.259 - 1.286)	3.00E-121	16 (16-16)
	MAF5-G > 13.6	0.61 (0.603 - 0.61)	7.00E-175	1.8 (1.744 - 1.857)	1.00E-77	20 (19-21)
Circulatory disease-related mortality	FIB-4 > 1.3 or 2	0.63 (0.623 - 0.629)	1.00E-13	1.32 (1.292 - 1.351)	1.00E-35	41 (40-42)
	FIB4-G > 8.9	0.78 (0.775 - 0.79)	0.00E+00	5.73 (5.314 - 6.187)	1.00E-116	75 (72-78)
	HSI ≥ 36	0.57 (0.564 - 0.571)	2.00E-133	1.47 (1.442 - 1.508)	6.00E-69	58 (57-59)
	HSI-G > 7.2	0.78 (0.774 - 0.789)	0.00E+00	6.16 (5.62 - 6.748)	5.00E-88	85 (82-87)
	FLI ≥ 60	0.65 (0.648 - 0.654)	0.00E+00	2.3 (2.252 - 2.354)	0.00E+00	59 (58-60)
	FLI-G > 8.6	0.79 (0.778 - 0.793)	0.00E+00	6.11 (5.646 - 6.622)	3.00E-114	79 (76-81)
	LAP > 47.67	0.61 (0.611 - 0.618)	1.00E-100	2.78 (2.694 - 2.866)	2.00E-238	14 (14-15)
	LAP-G > 17.3	0.78 (0.777 - 0.792)	0.00E+00	7.77 (7.238 - 8.331)	1.00E-186	31 (28-34)
	APRI > 0.5	0.56 (0.557 - 0.564)	6.00E-06	2.02 (1.943 - 2.101)	1.00E-72	9 (8-9)
	APRI-G > 18.3	0.78 (0.772 - 0.788)	0.00E+00	7.38 (6.835 - 7.959)	4.00E-152	25 (22-28)
	MAF-5 ≥ 1	0.67 (0.669 - 0.675)	0.00E+00	3.13 (3.061 - 3.209)	0.00E+00	32 (31-33)
	MAF5-G > 13.6	0.78 (0.777 - 0.792)	0.00E+00	6.77 (6.333 - 7.232)	1.00E-182	50 (46-53)
Malignancies-related mortality	FIB-4 > 1.3 or 2	0.6 (0.593 - 0.597)	1.00E-18	1.2 (1.179 - 1.214)	1.00E-33	39 (38-40)
	FIB4-G > 8.9	0.7 (0.689 - 0.701)	1.00E-230	3.21 (3.051 - 3.369)	5.00E-122	63 (60-65)
	HSI ≥ 36	0.52 (0.518 - 0.523)	3.00E-21	1.15 (1.131 - 1.164)	9.00E-22	52 (51-52)
	HSI-G > 7.2	0.69 (0.683 - 0.695)	8.00E-207	3.13 (2.961 - 3.3)	2.00E-98	74 (72-76)
	FLI ≥ 60	0.57 (0.569 - 0.573)	1.00E-251	1.49 (1.471 - 1.514)	8.00E-173	48 (47-49)
	FLI-G > 8.6	0.69 (0.685 - 0.697)	8.00E-221	3.04 (2.892 - 3.193)	4.00E-111	65 (62-67)
	LAP > 47.67	0.53 (0.524 - 0.529)	0.2	1.45 (1.41 - 1.486)	4.00E-45	8 (8-8)
	LAP-G > 17.3	0.69 (0.686 - 0.699)	2.00E-233	3.39 (3.181 - 3.614)	2.00E-81	16 (15-18)
	APRI > 0.5	0.53 (0.525 - 0.529)	1.00E-06	1.64 (1.598 - 1.69)	2.00E-70	7 (7-8)
	APRI-G > 18.3	0.69 (0.686 - 0.699)	3.00E-226	3.44 (3.204 - 3.687)	4.00E-69	13 (12-15)
	MAF-5 ≥ 1	0.57 (0.57 - 0.574)	1.00E-306	1.65 (1.622 - 1.682)	1.00E-168	20 (19-20)
	MAF5-G > 13.6	0.68 (0.678 - 0.69)	5.00E-178	2.88 (2.735 - 3.039)	5.00E-90	30 (28-32)
All-cause mortality	FIB-4 > 1.3 or 2	0.62 (0.615 - 0.618)	2.00E-60	1.27 (1.257 - 1.282)	2.00E-120	40 (39-40)
	FIB4-G > 8.9	0.75 (0.744 - 0.751)	0.00E+00	4.59 (4.437 - 4.74)	0.00E+00	69 (68-71)
	HSI ≥ 36	0.53 (0.526 - 0.53)	4.00E-122	1.2 (1.187 - 1.211)	9.00E-74	53 (52-53)
	HSI-G > 7.2	0.74 (0.736 - 0.743)	0.00E+00	4.31 (4.153 - 4.472)	0.00E+00	79 (78-80)
	FLI ≥ 60	0.6 (0.594 - 0.597)	0.00E+00	1.71 (1.695 - 1.729)	0.00E+00	51 (50-51)
	FLI-G > 8.6	0.74 (0.739 - 0.746)	0.00E+00	4.29 (4.146 - 4.432)	0.00E+00	71 (70-72)
	LAP > 47.67	0.56 (0.555 - 0.558)	6.00E-78	1.96 (1.927 - 1.991)	0.00E+00	10 (10-11)
	LAP-G > 17.3	0.75 (0.743 - 0.751)	0.00E+00	6.29 (6.066 - 6.515)	0.00E+00	23 (22-25)
	APRI > 0.5	0.54 (0.542 - 0.545)	3.00E-28	2.06 (2.025 - 2.099)	0.00E+00	9 (8-9)
	APRI-G > 18.3	0.74 (0.74 - 0.748)	0.00E+00	6.44 (6.196 - 6.695)	0.00E+00	19 (18-20)
	MAF-5 ≥ 1	0.61 (0.611 - 0.614)	0.00E+00	2.3 (2.273 - 2.326)	0.00E+00	25 (25-25)
	MAF5-G > 13.6	0.74 (0.734 - 0.742)	0.00E+00	4.83 (4.679 - 4.983)	0.00E+00	39 (37-40)

Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Risk group (n)	Cases in risk group (n)	Control group (n)	Cases in control group (n)	SER (lower is better)	FNBR (higher is better)
98 (97-98)	43 (40-45)	87 (86-87)	1366	583	36504	4877	65	8
89 (89-89)	24 (23-24)	88 (88-88)	53127	12556	378437	45409	34	10
91 (91-92)	35 (33-36)	88 (88-89)	4303	1496	33408	3922	25	10
66 (66-66)	18 (18-18)	85 (84-85)	151669	27289	282793	43777	16	10
68 (68-69)	23 (23-24)	87 (86-87)	13183	3097	25131	3292	12	12
52 (52-52)	17 (17-17)	84 (84-84)	212527	35759	227353	36208	12	12
55 (55-56)	22 (21-22)	88 (87-88)	18512	3981	20420	2504	10	16
63 (63-63)	18 (18-19)	85 (85-85)	170025	31297	274908	41439	14	11
65 (65-66)	23 (22-23)	87 (86-87)	14700	3320	24444	3199	12	12
94 (94-94)	18 (18-19)	84 (84-84)	25610	4720	416498	67590	94	7
95 (95-95)	27 (26-29)	84 (84-84)	2136	587	37007	5930	67	7
96 (96-96)	19 (19-20)	84 (84-84)	19684	3768	414878	67311	115	6
96 (96-96)	28 (25-30)	84 (83-84)	1650	456	36670	5934	84	6
87 (87-87)	20 (19-20)	84 (84-84)	57068	11233	375704	59579	39	7
89 (88-89)	26 (24-27)	85 (84-85)	4881	1255	33259	5105	30	7
65 (65-65)	2 (2-2)	98 (98-98)	160441	3412	295284	4899	134	93
66 (65-66)	5 (5-5)	99 (99-99)	14161	697	26063	230	58	175
52 (52-52)	2 (2-2)	99 (98-99)	222875	4855	238605	3522	95	131
53 (52-53)	4 (4-4)	99 (99-99)	19743	795	21138	141	51	290
62 (62-62)	3 (3-3)	99 (99-99)	178806	4963	287971	3492	94	134
63 (62-63)	5 (4-5)	99 (99-99)	15743	741	25354	200	55	205
94 (94-94)	5 (4-5)	98 (98-98)	26803	1216	436992	7254	381	64
95 (95-95)	12 (11-14)	98 (98-98)	2376	292	38727	658	141	62
95 (95-96)	3 (3-4)	98 (98-98)	20928	719	434903	7592	634	60
96 (96-96)	12 (11-14)	98 (98-98)	1847	230	38385	697	175	58
87 (87-87)	4 (4-5)	99 (99-99)	60214	2621	393732	5589	173	81
88 (87-88)	9 (8-9)	99 (99-99)	5312	452	34727	457	89	88
65 (65-65)	5 (5-5)	96 (96-96)	160447	7449	295279	11720	61	39
66 (66-67)	8 (7-8)	98 (97-98)	14161	1091	26064	648	37	62
52 (52-52)	5 (4-5)	96 (96-96)	222877	10048	238603	9366	46	49
53 (52-53)	7 (6-7)	98 (98-98)	19744	1311	21138	461	31	89
62 (62-62)	5 (5-5)	96 (96-97)	178804	9404	287972	10212	50	46
63 (62-63)	7 (7-8)	98 (97-98)	15743	1151	25355	628	36	65
94 (94-94)	6 (6-6)	96 (96-96)	26804	1577	436990	17940	294	26
95 (94-95)	12 (11-14)	96 (96-96)	2375	293	38729	1492	140	28
96 (95-96)	7 (6-7)	96 (96-96)	20928	1372	434904	17803	332	26
96 (96-96)	13 (11-14)	96 (96-96)	1846	234	38387	1506	172	27
87 (87-87)	6 (6-7)	96 (96-96)	60214	3796	393733	15280	120	30
87 (87-88)	10 (9-11)	97 (96-97)	5311	514	34729	1212	78	33
65 (65-65)	10 (10-10)	92 (92-92)	160454	15955	295296	23967	29	19
69 (68-69)	21 (20-22)	95 (95-95)	14162	2985	26064	1323	13	30
52 (52-52)	10 (9-10)	92 (92-92)	222889	21261	238617	19034	22	24
55 (55-56)	17 (17-18)	96 (95-96)	19745	3437	21138	921	12	44
63 (63-63)	12 (11-12)	93 (93-93)	178820	20728	287982	19926	23	23
66 (65-66)	20 (19-20)	95 (95-95)	15744	3100	25355	1271	13	32
95 (95-95)	16 (15-16)	92 (92-92)	26806	4201	437014	36448	110	13
96 (96-97)	43 (41-45)	91 (91-92)	2376	1024	38729	3384	40	12
96 (96-96)	16 (16-17)	92 (92-92)	20932	3397	434924	36533	134	12
97 (97-97)	45 (43-47)	91 (91-91)	1847	829	38387	3481	49	12
88 (88-88)	16 (16-17)	92 (92-93)	60222	9870	393749	29657	46	15
90 (89-90)	31 (30-32)	93 (92-93)	5312	1650	34729	2596	24	15

Table 4. C-indices, hazard ratios, diagnostic metrics, absolute risk, absolute risk reduction, and number needed to screen for existing and novel non-invasive MASLD indices aiming to stratify risk for cardiometabolic morbidity, malignancies, cause-specific and all-cause mortality in the population with type 2 diabetes mellitus.

Risk group refers to the group of individuals having an index score above the threshold indicated. The control group refers to the individuals who do not have an index score above the threshold indicated. SER refers to the number of individuals needed to be screened to identify one case (lower is better), and FNBR refers to the number of individuals that are screened for every false negative identified (higher is better). Abbreviations: CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; SER, screening efficiency ratio; FNBR, false negative burden ratio; CKD, chronic kidney disease; CHE, composite hepatic events; CVD, cardiovascular disease; FIB-4, Fibrosis 4 Index; HSI, Hepatic Steatosis Index; FLI, Fatty Liver Index; LAP, Lipid Accumulation Product; MAF-5, metabolic dysfunction–associated fibrosis score; APRI, Aspartate aminotransferase to Platelet Ratio Index, MASLD, metabolic dysfunction-associated steatotic liver disease.

Condition	Index and threshold	C-index (95% CI)	C-index P value	Hazard ratio (95% CI)	Hazard ratio P value	Sensitivity (95% CI)
CKD	FIB-4 > 1.3 or 2	0.55 (0.545 - 0.554)	6.00E-05	1.02 (0.989 - 1.056)	0.5	35 (33-36)
	FIB4-G > 8.9	0.71 (0.698 - 0.722)	5.00E-66	4.26 (3.519 - 5.163)	4.00E-14	94 (91-96)
	HSI ≥ 36	0.54 (0.539 - 0.548)	2.00E-19	1.44 (1.362 - 1.513)	7.00E-12	90 (89-91)
	HSI-G > 7.2	0.71 (0.697 - 0.721)	2.00E-68	3.82 (2.937 - 4.963)	3.00E-07	97 (95-98)
	FLI ≥ 60	0.58 (0.576 - 0.585)	3.00E-60	1.68 (1.612 - 1.744)	2.00E-39	81 (80-82)
	FLI-G > 8.6	0.71 (0.702 - 0.725)	5.00E-71	3.92 (3.189 - 4.809)	3.00E-11	95 (93-97)
	LAP > 47.67	0.47 (0.466 - 0.476)	0.05	1.47 (1.418 - 1.523)	4.00E-27	25 (24-26)
	LAP-G > 17.3	0.71 (0.693 - 0.717)	6.00E-66	3.15 (2.871 - 3.465)	2.00E-34	63 (59-68)
	APRI > 0.5	0.51 (0.504 - 0.513)	1	1.07 (1.017 - 1.127)	0.2	10 (9-11)
	APRI-G > 18.3	0.71 (0.695 - 0.719)	5.00E-64	3.32 (3.028 - 3.647)	3.00E-38	58 (53-62)
	MAF-5 ≥ 1	0.56 (0.551 - 0.56)	1.00E-32	1.4 (1.345 - 1.46)	1.00E-16	82 (81-83)
	MAF5-G > 13.6	0.71 (0.695 - 0.719)	3.00E-62	3.99 (3.436 - 4.634)	2.00E-20	89 (86-92)
CHE	FIB-4 > 1.3 or 2	0.6 (0.592 - 0.607)	2.00E-16	1.75 (1.668 - 1.833)	3.00E-32	47 (44-49)
	FIB4-G > 8.9	0.64 (0.619 - 0.663)	1.00E-12	2.42 (1.86 - 3.144)	8.00E-04	91 (85-95)
	HSI ≥ 36	0.57 (0.568 - 0.581)	1.00E-23	1.78 (1.634 - 1.946)	4.00E-11	92 (91-93)
	HSI-G > 7.2	0.64 (0.62 - 0.662)	4.00E-12	2.04 (1.446 - 2.867)	0.04	95 (90-98)
	FLI ≥ 60	0.64 (0.63 - 0.643)	2.00E-63	2.5 (2.33 - 2.673)	1.00E-40	86 (85-88)
	FLI-G > 8.6	0.64 (0.623 - 0.665)	1.00E-12	2.38 (1.786 - 3.18)	0.003	93 (88-96)
	LAP > 47.67	0.47 (0.465 - 0.48)	0.09	1.47 (1.396 - 1.554)	5.00E-13	26 (24-28)
	LAP-G > 17.3	0.63 (0.607 - 0.65)	4.00E-11	1.93 (1.661 - 2.25)	1.00E-05	56 (48-63)
	APRI > 0.5	0.63 (0.627 - 0.641)	3.00E-37	3.84 (3.637 - 4.048)	3.00E-139	26 (24-28)
	APRI-G > 18.3	0.64 (0.616 - 0.659)	5.00E-12	1.97 (1.694 - 2.301)	9.00E-06	50 (42-57)
	MAF-5 ≥ 1	0.65 (0.646 - 0.659)	1.00E-139	2.08 (1.935 - 2.229)	5.00E-25	87 (85-89)
	MAF5-G > 13.6	0.66 (0.636 - 0.679)	8.00E-15	3.17 (2.497 - 4.021)	1.00E-06	88 (82-93)
CVD	FIB-4 > 1.3 or 2	0.54 (0.54 - 0.548)	0.001	1.07 (1.038 - 1.097)	0.02	35 (34-36)
	FIB4-G > 8.9	0.62 (0.607 - 0.63)	7.00E-24	2.19 (1.932 - 2.491)	6.00E-10	87 (84-90)
	HSI ≥ 36	0.52 (0.521 - 0.528)	4.00E-10	1.27 (1.218 - 1.324)	1.00E-08	89 (88-90)
	HSI-G > 7.2	0.62 (0.605 - 0.628)	1.00E-23	2.68 (2.212 - 3.238)	2.00E-07	95 (93-97)
	FLI ≥ 60	0.56 (0.557 - 0.565)	2.00E-54	1.43 (1.382 - 1.471)	5.00E-30	77 (76-78)
	FLI-G > 8.6	0.62 (0.611 - 0.634)	7.00E-26	2.39 (2.07 - 2.749)	9.00E-10	90 (88-93)
	LAP > 47.67	0.55 (0.549 - 0.556)	1.00E-10	1.47 (1.421 - 1.512)	1.00E-34	23 (22-24)
	LAP-G > 17.3	0.62 (0.608 - 0.631)	7.00E-25	1.76 (1.623 - 1.917)	1.00E-11	47 (43-52)
	APRI > 0.5	0.51 (0.504 - 0.512)	0.6	1.05 (1.005 - 1.099)	0.3	10 (9-11)
	APRI-G > 18.3	0.62 (0.604 - 0.628)	1.00E-22	1.85 (1.696 - 2.013)	7.00E-13	41 (37-46)
	MAF-5 ≥ 1	0.55 (0.551 - 0.559)	2.00E-42	1.38 (1.332 - 1.425)	2.00E-21	81 (79-82)
	MAF5-G > 13.6	0.62 (0.61 - 0.633)	1.00E-24	2.15 (1.936 - 2.392)	4.00E-13	80 (76-83)
Malignancies	FIB-4 > 1.3 or 2	0.55 (0.549 - 0.558)	3.00E-06	1.09 (1.057 - 1.125)	0.005	36 (35-38)
	FIB4-G > 8.9	0.56 (0.543 - 0.571)	3.00E-04	1.57 (1.363 - 1.809)	0.001	86 (83-89)
	HSI ≥ 36	0.51 (0.507 - 0.516)	0.005	1 (0.957 - 1.044)	1	87 (86-88)
	HSI-G > 7.2	0.55 (0.54 - 0.567)	5.00E-04	1.54 (1.271 - 1.855)	0.02	93 (90-95)
	FLI ≥ 60	0.52 (0.511 - 0.52)	3.00E-06	1.13 (1.096 - 1.172)	2.00E-04	75 (73-76)
	FLI-G > 8.6	0.56 (0.546 - 0.574)	8.00E-05	1.59 (1.364 - 1.858)	0.003	89 (86-92)
	LAP > 47.67	0.52 (0.512 - 0.52)	0.05	1.11 (1.068 - 1.148)	0.005	21 (20-22)
	LAP-G > 17.3	0.56 (0.541 - 0.569)	3.00E-04	1.23 (1.115 - 1.353)	0.03	44 (39-49)

Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Risk group (n)	Cases in risk group (n)	Control group (n)	Cases in control group (n)	SER (lower is better)	FNBR (higher is better)
66 (65-66)	18 (17-19)	82 (82-83)	7928	1419	15068	2684	16	9
24 (22-26)	25 (23-28)	93 (91-95)	1753	447	433	29	5	75
14 (13-14)	19 (18-19)	87 (86-88)	20180	3736	3022	398	6	58
12 (10-13)	24 (22-26)	93 (89-96)	2002	474	219	15	5	148
29 (29-30)	20 (19-21)	88 (87-88)	16958	3375	6477	800	7	29
19 (17-21)	25 (23-27)	93 (90-95)	1876	464	356	25	5	89
82 (81-83)	23 (22-25)	83 (83-84)	4506	1050	18875	3132	22	7
67 (65-70)	35 (32-39)	87 (85-89)	882	311	1355	179	7	12
90 (90-91)	19 (17-20)	82 (82-83)	2262	424	20739	3680	54	6
74 (72-76)	38 (34-42)	86 (84-88)	725	275	1461	201	8	11
24 (23-25)	19 (18-20)	86 (85-87)	17536	3319	5246	730	7	31
35 (32-37)	28 (26-30)	92 (90-94)	1530	425	634	50	5	43
67 (66-67)	11 (10-11)	94 (93-94)	7852	839	15090	963	27	24
20 (19-22)	9 (7-10)	96 (94-98)	1773	155	432	16	14	138
14 (13-14)	8 (8-9)	95 (95-96)	20113	1663	3035	142	14	163
10 (9-11)	8 (7-9)	96 (92-98)	2023	165	218	9	14	249
29 (28-30)	9 (9-10)	96 (96-97)	16903	1567	6478	246	15	95
17 (15-18)	8 (7-10)	96 (94-98)	1896	161	356	13	14	173
81 (81-82)	10 (10-11)	93 (92-93)	4520	471	18807	1351	50	17
61 (59-63)	11 (9-13)	94 (93-95)	904	98	1354	78	23	29
92 (92-92)	22 (20-24)	94 (93-94)	2165	474	20782	1328	48	17
67 (65-69)	11 (9-14)	94 (93-95)	749	85	1456	86	26	26
24 (23-25)	9 (8-9)	96 (95-96)	17465	1547	5256	230	15	99
31 (29-33)	10 (8-11)	97 (95-98)	1546	149	635	20	15	109
66 (66-67)	30 (29-31)	71 (70-72)	6506	1975	12667	3693	10	5
26 (24-29)	36 (33-38)	82 (77-85)	1385	493	387	71	4	25
14 (14-15)	30 (30-31)	75 (74-77)	16756	5094	2599	639	4	30
14 (12-16)	34 (32-36)	85 (80-90)	1598	544	199	29	3	62
31 (31-32)	32 (31-33)	76 (75-77)	13892	4448	5666	1341	4	15
22 (20-24)	35 (33-38)	83 (79-87)	1479	519	326	55	3	33
84 (83-85)	38 (36-40)	72 (71-73)	3544	1345	15935	4444	14	4
69 (66-71)	42 (38-45)	74 (71-76)	659	274	1150	304	7	6
91 (90-91)	30 (28-33)	71 (70-71)	1830	557	17347	5113	34	4
75 (72-77)	43 (39-48)	73 (71-76)	538	233	1234	331	8	5
26 (25-27)	31 (31-32)	76 (75-77)	14438	4517	4578	1094	4	17
38 (35-41)	38 (35-40)	80 (77-83)	1186	446	566	112	4	16
66 (65-67)	22 (21-23)	80 (79-80)	7585	1658	14346	2921	13	8
21 (19-23)	21 (19-23)	86 (82-89)	1697	363	410	58	6	36
13 (13-14)	21 (20-22)	79 (78-81)	19223	4034	2895	602	5	37
11 (9-12)	21 (19-23)	86 (80-90)	1926	401	211	30	5	71
28 (28-29)	22 (21-22)	81 (80-82)	16159	3489	6175	1186	6	19
17 (15-19)	21 (19-23)	86 (82-90)	1811	385	338	47	6	46
81 (80-82)	23 (21-24)	79 (79-80)	4336	976	17960	3682	23	6
61 (59-64)	22 (20-25)	81 (79-83)	856	191	1299	243	11	9

Table 4. C-indices, hazard ratios, diagnostic metrics, absolute risk, absolute risk reduction, and number needed to screen for existing and novel non-invasive MASLD indices aiming to stratify risk for cardiometabolic morbidity, malignancies, cause-specific and all-cause mortality in the population with type 2 diabetes mellitus. (continued)

Condition	Index and threshold	C-index (95% CI)	C-index P value	Hazard ratio (95% CI)	Hazard ratio P value	Sensitivity (95% CI)
	APRI > 0.5	0.51 (0.505 - 0.514)	0.1	1.14 (1.086 - 1.194)	0.006	11 (10-12)
	APRI-G > 18.3	0.55 (0.541 - 0.569)	6.00E-04	1.25 (1.126 - 1.377)	0.03	38 (33-43)
	MAF-5 ≥ 1	0.52 (0.513 - 0.522)	5.00E-05	1.13 (1.09 - 1.172)	8.00E-04	79 (78-80)
	MAF5-G > 13.6	0.56 (0.542 - 0.571)	3.00E-04	1.5 (1.332 - 1.687)	6.00E-04	78 (74-82)
Circulatory disease-related mortality	FIB-4 > 1.3 or 2	0.57 (0.562 - 0.578)	9.00E-05	1.16 (1.095 - 1.218)	0.007	37 (35-40)
	FIB4-G > 8.9	0.7 (0.677 - 0.714)	1.00E-20	3.83 (2.843 - 5.159)	7.00E-06	94 (90-97)
	HSI ≥ 36	0.51 (0.5 - 0.516)	0.2	1.01 (0.932 - 1.087)	0.9	87 (85-89)
	HSI-G > 7.2	0.69 (0.673 - 0.709)	2.00E-20	5.65 (3.409 - 9.359)	6.00E-04	98 (95-99)
	FLI ≥ 60	0.57 (0.565 - 0.58)	1.00E-17	1.51 (1.414 - 1.607)	1.00E-10	80 (78-82)
	FLI-G > 8.6	0.69 (0.676 - 0.712)	3.00E-21	4.25 (3.023 - 5.978)	2.00E-05	96 (92-98)
	LAP > 47.67	0.59 (0.587 - 0.602)	7.00E-15	1.89 (1.788 - 1.997)	1.00E-30	31 (29-33)
	LAP-G > 17.3	0.69 (0.675 - 0.71)	7.00E-21	3.09 (2.67 - 3.575)	1.00E-14	66 (59-73)
	APRI > 0.5	0.51 (0.501 - 0.517)	0.3	1.2 (1.107 - 1.301)	0.02	11 (10-13)
	APRI-G > 18.3	0.69 (0.674 - 0.71)	8.00E-20	3.26 (2.813 - 3.772)	8.00E-16	61 (54-68)
	MAF-5 ≥ 1	0.57 (0.562 - 0.577)	2.00E-21	1.54 (1.433 - 1.65)	1.00E-09	84 (82-85)
	MAF5-G > 13.6	0.7 (0.68 - 0.717)	1.00E-20	4.12 (3.214 - 5.274)	1.00E-08	91 (86-94)
Malignancies-related mortality	FIB-4 > 1.3 or 2	0.57 (0.559 - 0.574)	3.00E-05	1.13 (1.078 - 1.192)	0.01	37 (35-39)
	FIB4-G > 8.9	0.58 (0.559 - 0.602)	7.00E-04	2.28 (1.753 - 2.968)	0.002	90 (85-94)
	HSI ≥ 36	0.51 (0.508 - 0.522)	0.03	1 (0.934 - 1.078)	1	87 (85-89)
	HSI-G > 7.2	0.58 (0.554 - 0.597)	9.00E-04	1.97 (1.398 - 2.773)	0.05	95 (90-98)
	FLI ≥ 60	0.54 (0.53 - 0.544)	3.00E-08	1.29 (1.215 - 1.361)	1.00E-05	77 (75-79)
	FLI-G > 8.6	0.58 (0.563 - 0.606)	2.00E-04	2.1 (1.591 - 2.78)	0.008	92 (87-95)
	LAP > 47.67	0.52 (0.517 - 0.531)	0.2	1.19 (1.125 - 1.263)	0.002	22 (20-24)
	LAP-G > 17.3	0.58 (0.562 - 0.604)	3.00E-04	1.56 (1.337 - 1.815)	0.004	51 (43-59)
	APRI > 0.5	0.52 (0.509 - 0.524)	0.003	1.39 (1.292 - 1.493)	6.00E-06	13 (11-15)
	APRI-G > 18.3	0.58 (0.556 - 0.599)	0.001	1.48 (1.264 - 1.732)	0.01	43 (35-51)
	MAF-5 ≥ 1	0.54 (0.528 - 0.542)	9.00E-08	1.22 (1.15 - 1.3)	0.001	80 (78-82)
	MAF5-G > 13.6	0.58 (0.558 - 0.603)	4.00E-04	1.68 (1.381 - 2.049)	0.008	80 (73-86)
All-cause mortality	FIB-4 > 1.3 or 2	0.58 (0.578 - 0.586)	2.00E-26	1.21 (1.176 - 1.246)	4.00E-11	38 (37-39)
	FIB4-G > 8.9	0.68 (0.673 - 0.695)	3.00E-61	4.26 (3.589 - 5.063)	3.00E-17	94 (92-96)
	HSI ≥ 36	0.51 (0.501 - 0.509)	0.7	0.94 (0.903 - 0.98)	0.1	87 (86-87)
	HSI-G > 7.2	0.67 (0.663 - 0.685)	1.00E-54	3.77 (2.986 - 4.76)	1.00E-08	97 (95-98)
	FLI ≥ 60	0.56 (0.556 - 0.564)	2.00E-40	1.4 (1.349 - 1.443)	6.00E-23	78 (77-79)
	FLI-G > 8.6	0.68 (0.669 - 0.691)	2.00E-58	3.78 (3.153 - 4.535)	2.00E-13	95 (93-96)
	LAP > 47.67	0.56 (0.558 - 0.567)	2.00E-17	1.59 (1.545 - 1.646)	7.00E-50	27 (25-28)
	LAP-G > 17.3	0.68 (0.672 - 0.693)	1.00E-60	2.89 (2.655 - 3.136)	4.00E-37	62 (58-66)
	APRI > 0.5	0.52 (0.517 - 0.526)	1.00E-14	1.44 (1.384 - 1.506)	2.00E-18	13 (12-14)
	APRI-G > 18.3	0.68 (0.669 - 0.691)	5.00E-58	2.95 (2.72 - 3.21)	5.00E-39	56 (52-60)
	MAF-5 ≥ 1	0.56 (0.559 - 0.568)	3.00E-63	1.42 (1.364 - 1.469)	5.00E-21	82 (81-83)
	MAF5-G > 13.6	0.68 (0.67 - 0.693)	4.00E-58	3.2 (2.825 - 3.629)	2.00E-20	87 (84-90)

Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Risk group (n)	Cases in risk group (n)	Control group (n)	Cases in control group (n)	SER (lower is better)	FNBR (higher is better)
91 (90-91)	23 (21-25)	79 (79-80)	2133	493	19803	4088	44	5
67 (65-70)	22 (19-26)	81 (79-83)	707	159	1400	262	13	8
23 (23-24)	21 (21-22)	81 (80-82)	16735	3596	4976	951	6	23
31 (28-33)	22 (20-24)	85 (82-87)	1482	326	601	92	6	23
66 (65-66)	7 (6-8)	94 (93-94)	8080	565	15239	946	41	25
21 (19-22)	10 (9-12)	97 (95-99)	1811	184	433	12	12	187
13 (13-14)	6 (6-7)	94 (93-95)	20441	1312	3081	193	18	122
10 (9-12)	10 (9-11)	98 (95-100)	2060	201	219	4	11	570
28 (27-29)	7 (7-7)	95 (95-96)	17210	1215	6546	307	20	77
17 (15-18)	10 (9-12)	97 (95-99)	1933	197	357	9	12	254
81 (81-82)	10 (9-11)	94 (94-95)	4607	475	19101	1064	50	22
62 (60-64)	15 (13-17)	95 (94-96)	928	139	1369	71	17	32
90 (90-91)	7 (6-9)	94 (93-94)	2298	172	21026	1339	136	17
68 (66-70)	16 (13-18)	95 (94-96)	769	120	1475	76	19	30
23 (23-24)	7 (7-7)	95 (95-96)	17788	1230	5302	241	19	96
30 (28-33)	11 (9-13)	97 (96-98)	1583	174	636	18	13	123
66 (65-66)	8 (7-8)	93 (93-93)	8082	629	15239	1069	37	22
20 (18-22)	8 (7-10)	96 (94-98)	1811	149	433	16	15	140
13 (13-14)	7 (7-8)	93 (92-94)	20443	1506	3082	223	16	105
10 (9-11)	8 (7-9)	96 (92-98)	2060	162	219	9	14	253
28 (27-29)	8 (7-8)	94 (93-94)	17212	1348	6546	398	18	60
16 (15-18)	8 (7-9)	96 (94-98)	1933	156	357	14	15	164
81 (80-81)	8 (8-9)	93 (93-93)	4607	386	19103	1351	61	18
60 (58-63)	9 (8-11)	94 (92-95)	928	87	1369	84	26	27
90 (90-91)	10 (8-11)	93 (93-93)	2299	219	21027	1479	107	16
66 (64-68)	9 (7-12)	94 (92-95)	769	71	1475	94	32	24
23 (23-24)	8 (7-8)	94 (93-94)	17790	1354	5302	332	17	70
29 (27-31)	8 (7-10)	95 (93-97)	1583	131	636	32	17	69
66 (66-67)	24 (23-25)	79 (79-80)	8082	1932	15242	3160	12	7
24 (22-26)	31 (29-33)	92 (89-94)	1811	558	433	36	4	62
13 (13-14)	22 (21-22)	78 (76-79)	20445	4426	3083	687	5	34
12 (10-14)	28 (27-30)	91 (87-95)	2060	586	219	19	4	120
29 (29-30)	23 (23-24)	83 (82-84)	17214	4032	6547	1122	6	21
19 (17-21)	30 (28-32)	91 (88-94)	1933	573	357	32	4	72
83 (82-83)	30 (29-31)	80 (80-81)	4608	1381	19105	3804	17	6
68 (65-70)	41 (38-45)	83 (81-85)	928	383	1369	233	6	10
91 (91-91)	28 (26-30)	79 (78-79)	2299	650	21030	4442	36	5
74 (71-76)	43 (40-47)	82 (80-84)	769	333	1475	261	7	9
24 (24-25)	23 (22-24)	83 (82-84)	17792	4102	5303	894	6	26
34 (32-37)	32 (30-34)	89 (86-91)	1583	507	636	73	4	30

For the general population, overall, the metabolic dysfunction–associated fibrosis (MAF-5) score presented the highest C-indices in all outcomes except for T2D, where Fatty Liver Index (FLI) had a higher C-index, and except for malignancies, malignancies-related mortality, and all-cause mortality, for which FIB-4 demonstrated the highest C-indices (Table 3). For individuals with T2D, FLI has the highest C-indices for CKD and CVD, and MAF-5 has the highest C-index for CHE (Table 4). For the same population, FIB-4 is the highest-performing NIT for malignancies, malignancies-related mortality, and all-cause mortality, while LAP is the highest-performing NIT for predicting circulatory disease-related mortality (Table 4). Regarding Hazard Ratios (HRs), in the general population, FIB-4 appeared to perform very poorly in identifying incident T2D cases (HR=0.92), while MAF-5 for the general population demonstrated overall the highest HRs, while MAF-5's performance decreased in the T2D population (given the exclusion of prevalent T2D cases in this population) (Fig. 1A, Fig. 2A, Table 3, Table 4). In the T2D population, FLI presented the highest HRs for CKD and CHE, LAP for CVD, circulatory disease-related mortality, and all-cause mortality, and Aspartate aminotransferase to Platelet Ratio Index (APRI) for malignancies-related mortality, while Hepatic Steatosis Index (HSI) demonstrated the lowest HRs for malignancies and all mortalities (Fig. 2A).

Out of all available proteomic markers, growth differentiation factor 15 (GDF-15) emerged as the overall top correlated molecule among the proteomic data with all outcomes studied herein (R^2=0.222) (Supplementary Table 7). Adding GDF-15 to the existing NITs to create GDF-15 enhanced NITs (FIB4-G, HSI-G, FLI-G, LAP-G, APRI-G, and MAF5-G) led to significantly increased prognostic performance compared to the traditional NITs in almost all instances, reaching substantially high C-indices, ranging between 0.68 and 0.81, for the general population and between 0.67 and 0.71 for the T2D population, with an overall >0.2 improvement in C-index (Fig. 1A, Fig. 2A, Table 2, Table 4). Specifically, for CKD, the GDF-15 enhanced NITs present with more than three-fold improvements in HRs in the general population, and a similar increase in performance was observed in individuals with T2D (Fig. 1A, Fig. 2A, Table 3, Table 4). Notably, GDF-15, using the average of all six NITs' group sizes, by itself outperformed many NITs in predicting most outcomes studied herein, and a model containing only the top ten proteomic features (without any NITs included) yielded a similar performance to GDF-15 alone, while performing overall similarly to the GDF-15 enhanced NITs (Supplementary Figure 2).

Kaplan-Meier cumulative incidence curves for all outcomes studied herein showed that individuals with higher values of the GDF-15 enhanced NITs are at the highest risk of adverse outcomes with an overall significantly higher incidence of morbidity and mortality when compared with groups of individuals screened with the traditional NITs (Supplementary Figures 3-5).

The sensitivities and specificities of the GDF-15 enhanced NITs are overall higher than the traditional NITs in the general population (Table 3). In many instances, the sensitivity of the GDF-15 enhanced NITs is almost double that of the traditional NITs for both the general population and the T2D population (Table 3, Table 4). However, the differences in specificity remain overall constant between the traditional and GDF-15 enhanced NITs in the general population, and there seems to be an overall decrease in specificity among the GDF-15 enhanced NITs compared to the traditional NITs for the T2D population (Table 3, Table 4). To investigate clinical utility, the proposed metric herein, SER, demonstrates the number of individuals needed to be screened to identify one true positive case, and the FNBR, in turn, refers to the number of cases screened while missing one case (Table 3, Table 4, Supplementary Figures 6-8). The GDF-15 enhanced NITs present with improved clinical utility compared to the traditional NITs (Table 3, Table 4, Supplementary Figures 6-8). For instance, the benefits are clear in CKD, where 21 individuals need to be screened with FLI-G to identify one incident CKD case, while this number is 36 with FLI for the general population (Table 3, Supplementary Figure 6). Similarly, 22 individuals from the general population need to be screened with FIB4-G to identify one incident CKD case, with a corresponding number of 53 for FIB-4 (Table 3, Supplementary Figure 6). Notably, the GDF-15 enhanced NITs are overall less likely to miss a case of incident CKD compared to the traditional NITs, exemplified by the high efficacy of HSI-G observed in all populations (Table 3, Table 4, Supplementary Figures 6 and 8). For instance, HSI-G in the general population misses an incident CKD case for every 144 individuals screened, with the corresponding number for HSI being 58 (Table 3, Supplementary Figure 6). For the same comparison in individuals with T2D, HSI misses a case of CKD almost four times more frequently compared to HSI-G, and FIB-4 misses a case of CKD eight times more frequently than FIB4-G (Table 4, Supplementary Figure 8). Notably, for individuals with T2D, using the APRI-G reduces the number of individuals that need to be screened more than six times to identify a true positive case of CKD and more than seven times to identify a true positive case of circulatory disease-related mortality than using the traditional APRI (Table 4).

A sub-analysis investigating the continuous values (expressed in percentiles) of all established and GDF-15 enhanced NITs demonstrates an exponential increase in the incidence of adverse outcomes in the upper quantile of the GDF-15 enhanced NITs in almost all instances (except for malignancies in the general population, and CHE, malignancies as well as malignancies–related mortality for individuals with T2D) (Fig. 1B, Fig. 2B). These exponential increases translate to groups of high-risk individuals with two- to four-fold higher incidence of adverse outcomes (Fig. 1B, Fig. 2B).

Translating the aforementioned results into a practical application and demonstrating the simplicity of the scores developed herein, we constructed a web-based calculator tool (Supplementary Figure 9) that is easy to use by clinicians and researchers aiming to inform risk communication: https://clinicalpredictor.shinyapps.io/multimorbidity-mortality-risk/. In our tool, we have also demonstrated the predicted incident risk of all GDF-15 enhanced NITs for three follow-up time periods, i.e., 5, 10, and 15 years.

Discussion

This longitudinal analysis with a median follow-up of 14 years showed that among the established NITs, FLI and MAF-5 are overall superior as prognostic indices of cardio-renal-metabolic adverse events compared to HSI, APRI, LAP, and FIB-4, while FIB-4 is superior at forecasting malignancies, malignancies-related mortality, as well as all-cause mortality in the mainly Caucasian European population of the UK Biobank cohort. However, as demonstrated herein, currently available NITs are suboptimal for predicting incident cardio-renal-metabolic and malignancies-related morbidity, mortality, and cause-specific and all-cause mortality in the general population, the potentially MASLD population, and the T2D population.

Adding GDF-15, the major proteomics analysis-derived signal, to the existing NITs resulted in significant improvements in risk stratification. This translates to clinically meaningful reductions in the number of individuals needed to be screened to identify cases developing adverse outcomes. To the best of our knowledge, there are no other prognostic studies using large cohorts demonstrating the superior predictive performance of data-driven, omics-improved-indices compared to traditional NITs regarding secondary outcomes (other than MASLD, MASH, or respective disease progression and staging). Additionally, HSI, LAP, and MAF-5 have not been previously validated in the UK Biobank in forecasting adverse outcomes. The proteomic-enabled GDF-15 enhanced NITs appear to forecast all outcomes studied herein better compared to the established NITs, while they seem to be less accurate predictors of incident T2D in the general and potentially MASLD population and incident CHE among individuals with T2D. Since all GDF-15 enhanced NITs demonstrate very high negative predictive values, and the upper quantile seems to yield exponential increases in the incidence of adverse events, low scores can be discarded. Adding GDF-15 to routinely calculated indices reduces the number of individuals needing to be screened to identify true positive cases of adverse outcomes and substantially increases the number of individuals screened for every missed case of adverse outcomes. For instance, even for incident T2D in the general population, where FIB-4 seems to have the poorest performance, one would need to screen two times fewer individuals to identify

one case with FIB4-G compared to FIB-4. At a population level, these reductions in the number needed to be screened can prove to be significant and positively affect the utility of multipurpose non-invasive screening scores. Finally, we observed that GDF-15, as a standalone predictor, yielded high C-indices for all outcomes while using a proteomic model of only ten markers performed overall similarly to GDF-15 alone and comparably to the GDF-15 enhanced NITs in most outcomes. Nonetheless, a proteomics-only model increases complexity to be reduced to practice and, therefore, may not be, at least in the short term, clinically applicable.

Potential mechanisms for the observed results

Regarding the better performance of MAF-5 and FLI over the rest of the NITs among mainly cardio-renal-metabolic outcomes may be explained by the inclusion of both waist circumference and BMI in their formulas. A Consensus Statement from the International Atherosclerosis Society and International Chair on Cardiometabolic Risk Working Group on Visceral Obesity concluded that when incorporating both BMI and waist circumference as continuous variables in the same risk prediction model, waist circumference remains a positive predictor of mortality risk, but BMI is not related or presents a negative relation to this risk (34). The same Consensus Statement further supports that the enhanced ability of waist circumference in predicting health outcomes compared to BMI could be partly attributed to its effectiveness in identifying adults with elevated visceral adipose tissue mass (34). Similarly, our online tool supports the aforementioned statement by demonstrating the negligible contribution of BMI in MAF-5- and FLI-based risk prediction compared to waist circumference. Notably, the other hepatic fibrosis indices (i.e., FIB-4 and APRI), completely lacking anthropometric measurements in their formulas (Table 1), yield overall lower C-indices for the cardio-renal-metabolic outcomes studied herein. This is expected to possibly be population-specific, and thus, our data need to be comparatively assessed among Caucasians vs other ethnicities, given that BMI or waist circumference may play a different role in different populations (35).

Additionally, the identification of GDF-15 in this study as the most important predictor of adverse outcomes is in line with the current literature about this protein, which is available in the existing literature. The clinically relevant results achieved with the six novel proteomic-enabled NITs are well supported by GDF-15's pathophysiology as a metabolic stress-response cytokine belonging to the transforming growth factor β superfamily and being highly up-regulated in states characterized by metabolic stress and in inflammatory states (36). The relevance of visceral adiposity, as described above, is further supported by evidence that GDF-15 is elevated predominantly in visceral adipose tissue compared to subcutaneous adipose tissue in obesity (37). Increased GDF-15 expression is associated with macrophage infiltration within visceral fat, which is augmented

in obesity and T2D (37). Moreover, GDF-15 has been shown to correlate with myocardial injury, pressure cardiac overload, as well as malignancies, and other cardiometabolic diseases (36, 38, 39). It has also been reported that elevated levels of GDF-15 have been linked to an increased risk of incident CKD and a more rapid decline in kidney function in various renal disorders (40). GDF-15's prognostic roles have been previously demonstrated, even predicting survival with very high accuracy (41), but studies investigating the risk-prediction role of GDF-15 for multimorbidity outcomes in large prospective population cohorts were missing.

Even though circulating GDF-15 levels seem to be a promising marker of adverse cardiometabolic outcomes, GDF-15's physiological role in the cardiovascular systems is yet to be fully elucidated (42). Besides, preclinical evidence demonstrates a strong connection between GDF-15, liver, and lipids and prompts consideration as to whether GDF-15 is involved in the pathophysiology of liver diseases and, thus, whether it may potentially have a causal role in liver metabolic diseases (43). There is a lack of additional larger studies regarding the role of GDF-15 in glucolipid and cardiometabolic diseases (44).

Translation to clinical practice

The increased risk of adverse outcomes for all populations studied herein highlights the need for a multifaceted but simple screening strategy for multiple adverse outcomes at once using actionable tools to improve the quality of cardio-renal-liver-metabolic care. Especially with the recent approval of resmetirom for noncirrhotic MASH with moderate to advanced liver fibrosis (stages F2-F3), as well as the increasingly expanding list of potential uses of emerging weight-loss drugs, an unmet clinical need for risk stratification is emerging, leading to increased implementation of simple NITs in primary care settings to timely diagnose and subsequently treat individuals having MASH or at-risk MASH (4, 20, 45-48). Previous research suggests that individuals with MASLD present a significantly higher risk for cardio-renal-metabolic and malignancies-related morbidity as well as cause-specific and all-cause mortality, further signaling the need for multipurpose NITs concurrently screening for multiple outcomes (49-51). Considering that soon, primary care physicians will need easy and clear guidance on how to stratify risk for secondary outcomes, simpler solutions to current approaches based on smartphone-based artificial intelligence technologies are encouraged (52). Therefore, following the paradigm set by the simple bedside web-based calculation of FIB-4 or even other markers like the estimated glomerular filtration rate, the artificial intelligence-based web tool developed herein extends its use from a simple calculator app and automatically calculates the risk of adverse outcomes in a visually attractive and smartphone accessible way based on the percentile values of the GDF-15 enhanced NITs: https://clinicalpredictor.shinyapps.io/multimorbidity-mortality-risk/, facilitating risk communication to the

patients of the general population and translating biomarkers to comprehensible probabilities for adverse outcomes. Even though no causality can be inferred from this risk prediction tool, patients can easily understand the increased risk brought by the components of the GDF-15 enhanced NITs and realize the potential benefits, for instance, of lifestyle interventions targeting visceral adiposity and overall cardiometabolic health.

Limitations of this study

This study has several limitations that should be considered. The general population studied herein includes a potentially MASLD population that has not undergone screening with NITs, ultrasound, or liver biopsy and might differ from individuals with biopsy-proven MASLD and MASH. We have, however, also studied a potentially MASLD population and a T2D population. Besides, the UK Biobank population is almost exclusively made up of Caucasian individuals, and even though ethnic minority groups are available, there was insufficient data available for reliable results. Since the outcomes were identified using ICD codes and other data fields (for T2D and CKD), owning to biobank limitations, it was not possible to capture the severity of the disease or changes in the disease course during the maximum follow-up period of 15 years. Moreover, given the limited features available, we could not validate and improve any other widely implemented NITs that might be used in clinical practice for hepatic steatosis or fibrosis. Additionally, the currently limited availability of GDF-15 assays in clinical settings, as well as additional costs they might be associated with, limits the implementation of GDF-15 enhanced NITs in clinical practice. Lastly, due to the lack of proteomics data in other large population cohorts, the results presented herein have not been externally validated and need to be used in an independent cohort in similar and/or other populations.

Conclusions

After prospectively validating and comparing established NITs in a large population cohort, we identify FLI and MAF-5 as better prognostic NITs for adverse cardio-renal-metabolic outcomes, and FIB-4 for malignancies-related mortality, as well as all-cause mortality in the UK Biobank. Subsequently, we propose six GDF-15 enhanced NITs and show that these significantly outperform the existing NITs in predicting most adverse outcomes. These scores have potential clinical utility by minimizing the number of individuals needed to be screened. Guidelines regarding secondary risk stratification for both individuals with MASLD and/or T2D may use these results to adjust their recommendations. Future studies are needed to replicate and confirm the results presented herein.

Contributors
MK and PF contributed equally as shared first authors. All authors carried out acquisition, analysis, or interpretation of data. MK wrote the first draft. All authors read and critically revised the manuscript for important intellectual content. All authors approved the final version of the manuscript. MK is the guarantor. CSM supervised this work. The corresponding author attests that all listed authors meet authorship criteria and that no authors meeting the criteria have been omitted.

Funding
This project was funded by the University Medical Center Groningen under project number: PPP-2019_023. MK receives funding support from an MD-PhD grant from the University Medical Center Groningen. CSM receives funding support from a Massachusetts Life Sciences Center grant. The funder had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Data sharing
Study data are available from the UK Biobank but were used under license for the current study, which restricts their public availability. Data are, however, available from the authors upon reasonable request and when granted permission by the UK Biobank. All code is available and can be shared upon request from the corresponding author.

Competing interests
All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; MK declares no competing interests. PF, JCF, and SvD are employees of Ancora Health B.V. and own shares of Ancora Health B.V., and BHRW is a scientific advisory board member of Ancora Health B.V. without being compensated for this position. C.S.M. reports grants through his institution from Merck, Massachusetts Life Sciences Center, and Boehringer Ingelheim, has been a shareholder of and has received grants through his institution and personal consulting fees from Coherus Inc. and AltrixBio; he reports personal consulting fees and support with research reagents from Ansh Inc., collaborative research support from LabCorp Inc., reports personal consulting fees from Genfit, Lumos, Amgen, Corcept, Aligos, Intercept, 89 Bio, Madrigal, and Regeneron, reports travel support and fees from TMIOA, Elsevier, and the Cardio Metabolic Health Conference. None is related to the work presented herein.

Transparency
MK and CSM affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have

been omitted; and that any discrepancies from the study as originally planned have been explained.

Acknowledgments
All authors critically reviewed the manuscript for important intellectual content. The authors thank the UK Biobank data access granted through application 55495.

Supplementary data
This article contains supporting information available online at:
https://ars.els-cdn.com/content/image/1-s2.0-S0026049524002750-mmc1.pdf

References

1. Kokkorakis M, Boutari C, Katsiki N, Mantzoros CS. From non-alcoholic fatty liver disease (NAFLD) to steatotic liver disease (SLD): an ongoing journey towards refining the terminology for this prevalent metabolic condition and unmet clinical need. Metabolism. 2023;147:155664.
2. Boutari C, Polyzos SA, Mantzoros CS. Addressing the epidemic of fatty liver disease: A call to action, a call to collaboration, a call to moving the field forward. Metabolism. 2021;122:154781.
3. Polyzos SA, Kang ES, Tsochatzis EA, Kechagias S, Ekstedt M, Xanthakos S, et al. Commentary: Nonalcoholic or metabolic dysfunction-associated fatty liver disease? The epidemic of the 21st century in search of the most appropriate name. Metabolism. 2020;113:154413.
4. Kokkorakis M, Muzurović E, Volčanšek Š, Chakhtoura M, Hill MA, Mikhailidis DP, et al. Steatotic Liver Disease: Pathophysiology and Emerging Pharmacotherapies. Pharmacol Rev. 2024;76(3):454-499.
5. Zhang H, Zhou X-D, Shapiro MD, Lip GYH, Tilg H, Valenti L, et al. Global burden of metabolic diseases, 1990–2021. Metabolism - Clinical and Experimental. 2024;160.
6. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology. 2018;67(1):328-57.
7. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology. 2016;64(1):73-84.
8. Sanyal A, Charles ED, Neuschwander-Tetri BA, Loomba R, Harrison SA, Abdelmalek MF, et al. Pegbelfermin (BMS-986036), a PEGylated fibroblast growth factor 21 analogue, in patients with non-alcoholic steatohepatitis: a randomised, double-blind, placebo-controlled, phase 2a trial. Lancet. 2019;392(10165):2705-17.
9. Muzurović E, Peng CC, Belanger MJ, Sanoudou D, Mikhailidis DP, Mantzoros CS. Nonalcoholic Fatty Liver Disease and Cardiovascular Disease: a Review of Shared Cardiometabolic Risk Factors. Hypertension. 2022;79(7):1319-26.
10. Choe HJ, Moon JH, Kim W, Koo BK, Cho NH. Steatotic liver disease predicts cardiovascular disease and advanced liver fibrosis: A community-dwelling cohort study with 20-year follow-up. Metabolism - Clinical and Experimental. 2024;153.
11. Ando Y, Jou JH. Nonalcoholic Fatty Liver Disease and Recent Guideline Updates. Clin Liver Dis (Hoboken). 2021;17(1):23-8.
12. EASL-EASD-EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). J Hepatol. 2024;81(3):492-542.
13. Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease. Lancet. 2021;397(10290):2212-24.
14. ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, et al. 4. Comprehensive Medical Evaluation and Assessment of Comorbidities: Standards of Care in Diabetes-2023. Diabetes Care. 2023;46(Suppl 1):S49-s67.
15. Wattacheril JJ, Abdelmalek MF, Lim JK, Sanyal AJ. AGA Clinical Practice Update on the Role of Noninvasive Biomarkers in the Evaluation and Management of Nonalcoholic Fatty Liver Disease: Expert Review. Gastroenterology. 2023;165(4):1080-8.
16. McPherson S, Hardy T, Dufour JF, Petta S, Romero-Gomez M, Allison M, et al. Age as a Confounding Factor for the Accurate Non-Invasive Diagnosis of Advanced NAFLD Fibrosis. Am J Gastroenterol. 2017;112(5):740-51.

17. Handelsman Y, Anderson JE, Bakris GL, Ballantyne CM, Bhatt DL, Bloomgarden ZT, et al. DCRM 2.0: Multispecialty practice recommendations for the management of diabetes, cardiorenal, and metabolic diseases. Metabolism - Clinical and Experimental. 2024;159.
18. Angelidi AM, Sanoudou D, Hill MA, Mantzoros CS. Management of patients with the cardio renal liver metabolic syndrome: The need for a multidisciplinary approach in research, education and practice. Metabolism - Clinical and Experimental. 2024;159.
19. Kokkorakis M, Boutari C, Hill MA, Kotsis V, Loomba R, Sanyal AJ, et al. Resmetirom, the first approved drug for the management of metabolic dysfunction-associated steatohepatitis: Trials, opportunities, and challenges. Metabolism. 2024;154:155835.
20. Kokkorakis M, Chakhtoura M, Rhayem C, Al Rifai J, Ghezzawi M, Valenzuela-Vallejo L, et al. Emerging Pharmacotherapies for Obesity: A Systematic Review. Pharmacol Revi. 2025;77(1):100002.
21. Richter MM, Kemp IM, Heebøll S, Winther-Sørensen M, Kjeldsen SAS, Jensen NJ, et al. Glucagon augments the secretion of FGF21 and GDF15 in MASLD by indirect mechanisms. Metabolism - Clinical and Experimental. 2024;156.
22. Abeysekera KWM, Valenti L, Younossi Z, Dillon JF, Allen AM, Nourredin M, et al. Implementation of a liver health check in people with type 2 diabetes. Lancet Gastroenterol Hepatol. 2024;9(1):83-91.
23. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, Abdelmalek MF, Caldwell S, Barb D, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. Hepatology. 2023;77(5):1797-835.
24. Kanwal F, Shubrook JH, Adams LA, Pfotenhauer K, Wai-Sun Wong V, Wright E, et al. Clinical Care Pathway for the Risk Stratification and Management of Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology. 2021;161(5):1657-69.
25. Zou B, Yeo YH, Cheung R, Ingelsson E, Nguyen MH. Fatty Liver Index and Development of Cardiovascular Disease: Findings from the UK Biobank. Dig Dis Sci. 2021;66(6):2092-100.
26. Cheng YM, Wang CC, Kao JH. Metabolic associated fatty liver disease better identifying patients at risk of liver and cardiovascular complications. Hepatol Int. 2023;17(2):350-6.
27. Seo IH, Lee HS, Lee YJ. Fatty liver index as a predictor for incident type 2 diabetes in community-dwelling adults: longitudinal findings over 12 years. Cardiovasc Diabetol. 2022;21(1):209.
28. Kouvari M, Valenzuela-Vallejo L, Guatibonza-Garcia V, Polyzos SA, Deng Y, Kokkorakis M, et al. Liver biopsy-based validation, confirmation and comparison of the diagnostic performance of established and novel non-invasive steatotic liver disease indexes: Results from a large multi-center study. Metabolism. 2023;147:155666.
29. Kouvari M, Mylonakis SC, Katsarou A, Valenzuela-Vallejo L, Guatibonza-Garcia V, Kokkorakis M, et al. The first external validation of the Dallas steatosis index in biopsy-proven Non-alcoholic fatty liver Disease: A multicenter study. Diabetes Res Clin Pract. 2023;203:110870.
30. Polyzos SA, Mantzoros CS. Metabolic dysfunction-associated steatotic liver disease: Recent turning points for its diagnosis and management. Metabolism - Clinical and Experimental. 2024;157.
31. Kouvari M, Chrysohoou C, Damigou E, Barkas F, Kravvariti E, Liberopoulos E, et al. Non-invasive tools for liver steatosis and steatohepatitis predict incidence of diabetes, cardiovascular disease and mortality 20 years later: The ATTICA cohort study (2002-2022). Clin Nutr. 2024;43(3):900-8.
32. Collins R. What makes UK Biobank special? Lancet. 2012;379(9822):1173-4.

33. Gadd DA, Hillary RF, Kuncheva Z, Mangelis T, Cheng Y, Dissanayake M, et al. Blood protein assessment of leading incident diseases and mortality in the UK Biobank. Nature Aging. 2024;4(7):939-48.
34. Ross R, Neeland IJ, Yamashita S, Shai I, Seidell J, Magni P, et al. Waist circumference as a vital sign in clinical practice: a Consensus Statement from the IAS and ICCR Working Group on Visceral Obesity. Nat Rev Endocrinol. 2020;16(3):177-89.
35. Kokkorakis M, Folkertsma P, van Dam S, Sirotin N, Taheri S, Chagoury O, et al. Effective questionnaire-based prediction models for type 2 diabetes across several ethnicities: a model development and validation study. EClinicalMedicine. 2023;64:102235.
36. May BM, Pimentel M, Zimerman LI, Rohde LE. GDF-15 as a Biomarker in Cardiovascular Disease. Arq Bras Cardiol. 2021;116(3):494-500.
37. L'Homme L, Sermikli BP, Haas JT, Fleury S, Quemener S, Guinot V, et al. Adipose tissue macrophage infiltration and hepatocyte stress increase GDF-15 throughout development of obesity to MASH. Nat Commun. 2024;15(1):7173.
38. Fang L, Li F, Gu C. GDF-15: A Multifunctional Modulator and Potential Therapeutic Target in Cancer. Curr Pharm Des. 2019;25(6):654-62.
39. Siddiqui JA, Pothuraju R, Khan P, Sharma G, Muniyan S, Seshacharyulu P, et al. Pathophysiological role of growth differentiation factor 15 (GDF15) in obesity, cancer, and cachexia. Cytokine Growth Factor Rev. 2022;64:71-83.
40. Delrue C, Speeckaert R, Delanghe JR, Speeckaert MM. Growth differentiation factor 15 (GDF-15) in kidney diseases. Adv Clin Chem. 2023;114:1-46.
41. Kempf T, Björklund E, Olofsson S, Lindahl B, Allhoff T, Peter T, et al. Growth-differentiation factor-15 improves risk stratification in ST-segment elevation myocardial infarction. Eur Heart J. 2007;28(23):2858-65.
42. Tzikas S, Vassilikos V, Keller T. GDF-15 as a risk stratification biomarker for cardiovascular disease. Int J Cardiol. 2019;292:246-7.
43. Zhang M, Sun W, Qian J, Tang Y. Fasting exacerbates hepatic growth differentiation factor 15 to promote fatty acid β-oxidation and ketogenesis via activating XBP1 signaling in liver. Redox Biol. 2018;16:87-96.
44. Xiao QA, He Q, Zeng J, Xia X. GDF-15, a future therapeutic target of glucolipid metabolic disorders and cardiovascular disease. Biomed Pharmacother. 2022;146:112582.
45. Boutari C, Kokkorakis M, Stefanakis K, Valenzuela-Vallejo L, Axarloglou E, Volčanšek Š, et al. Recent research advances in metabolism, clinical and experimental. Metabolism. 2023;149:155722.
46. Kokkorakis M, Katsarou A, Katsiki N, Mantzoros CS. Milestones in the journey towards addressing obesity; Past trials and triumphs, recent breakthroughs, and an exciting future in the era of emerging effective medical therapies and integration of effective medical therapies with metabolic surgery. Metabolism. 2023;148:155689.
47. Huttasch M, Roden M, Kahl S. Obesity and MASLD: Is weight loss the (only) key to treat metabolic liver disease? Metabolism - Clinical and Experimental. 2024;157.
48. Lian L-Y, Targher G, Byrne CD, Liu W-Y, Zheng M-H. Resmetirom for MASH patients with diabetes: challenges and opportunities in the real world. Metabolism - Clinical and Experimental. 2024;156.
49. Roca-Fernandez A, Banerjee R, Thomaides-Brears H, Telford A, Sanyal A, Neubauer S, et al. Liver disease is a significant risk factor for cardiovascular outcomes - A UK Biobank study. J Hepatol. 2023;79(5):1085-95.
50. Kim KS, Hong S, Han K, Park CY. Association of non-alcoholic fatty liver disease with cardiovascular disease and all cause death in patients with type 2 diabetes mellitus: nationwide population based study. BMJ. 2024;384:e076388.
51. Bhala N. Liver steatosis linked to type 2 diabetes outcomes. BMJ. 2024;384:q309.

52. Mantena S, Celi LA, Keshavjee S, Beratarrechea A. Improving community health-care screenings with smartphone-based AI technologies. Lancet Digit Health. 2021;3(5):e280-e2.

Chapter 5

Radiomic and proteomic signatures of body mass index on brain ageing and Alzheimer's-like patterns of brain atrophy

Filippos Anagnostakis, **Michail Kokkorakis**, Keenan A. Walker, Ioanna Skampardoni, Junhao Wen, Guray Erus, Duygu Tosun, Vasiliki Tassopoulou, Yuhan Cui, Sindhuja T. Govindarajan, Dhivya Srinivasan, Randa Melhem, Elizabeth Mamourian, Haochang Shou, Ilya M. Nasrallah, Christos S. Mantzoros, Christos Davatzikos

EBioMedicine 2025;116:105763 (pre-publication version)

Summary

Background The impact of high body mass index (BMI) states and associated proteomic factors on brain ageing and Alzheimer's disease (AD) remains unclear.

Methods We sought to evaluate machine learning (ML)-based neuroimaging markers of brain age and AD-like brain atrophy in participants with obesity or overweight without diagnosed cognitive impairment (WODCI), in a harmonised study of 46,288 participants in 15 studies (the Imaging-Based Coordinate System for Aging and Neurodegenerative Diseases (iSTAGING) consortium). We also assessed the association between cognition, serum proteins, and brain ageing indices. Data were acquired between 1999 and 2020 and analysed from November 2024 onwards.

Findings The study comprised 46,288 participants, including 24,897 females and 21,391 males, with a mean age of 64.33 years (SD = 8.13) and a mean BMI of 26.81 kg/m^2 (SD = 4.49). The results demonstrate that the impact of obesity on brain ageing, and AD-like brain atrophy is weaker with increasing age and is significantly pronounced in males compared to females. Additionally, in males, obesity was significantly associated with approximately 2 additional years of brain ageing compared to normal weight and 1 additional year compared to the overweight group. Males with overweight also showed higher brain ageing values (8 additional months) than males with normal weight. Regarding AD-like brain atrophy, males with obesity displayed higher AD-like brain atrophy than males with normal weight, but females with normal weight showed higher AD-like brain atrophy than females with overweight. Sex differences within the same BMI categories were observed, with males exhibiting increased brain ageing compared to females, in obesity (1 additional year) and overweight groups (3 additional months). Higher AD-like brain atrophy was observed in males with overweight than in females with overweight. Females with normal weight displayed increased brain ageing (8 additional months) and AD-like brain atrophy relative to males with normal weight. In both retrospective and cross-sectional proteomics studies, five and eight proteins out of 1,463 proteins were significantly (positively or negatively) associated with brain ageing and 1-SD BMI change (SD=4.2 kg/m^2), respectively.

Interpretation The findings demonstrate that higher BMI states are associated with accelerated brain ageing and AD-like atrophy, particularly in males, while females with normal weight demonstrated higher brain ageing and AD-like atrophy than males with normal weight. Moreover, the impact of obesity on brain ageing and AD-like brain atrophy becomes weaker with increasing age. Further research is needed to investigate sex-specific mechanisms by which weight gain influences brain aging.

Funding National Institute on Aging's Intramural Research Program, National Institute on Aging, Intramural Research Program.

Research in context

Evidence before this study

We searched PubMed for articles in English published until November 10, 2024, examining the relation between body mass index (BMI) states and brain ageing. Search terms included "dementia", "brain age", "brain ageing", "BMI", "obesity", "brain age prediction", and "proteomics". Obesity, particularly midlife obesity, is known to be a modifiable risk factor for Alzheimer's disease (AD) and related dementias, with long-term obesity associated with cognitive decline and brain atrophy. However, the mechanisms linking obesity to brain ageing and dementia are not fully understood. Studies have demonstrated that high BMI is related to neurocognitive decline and compromised brain integrity, but the specific molecular markers, such as serum proteins, that mediate these effects are still unclear. Additionally, the relationship between obesity and brain ageing, particularly in individuals without cognitive impairments, remains underexplored.

Added value of this study

This study aims to investigate the effects of BMI and related proteomic markers on brain ageing and AD-like brain atrophy in participants without diagnosed cognitive impairment, utilising a comprehensive, diverse, and harmonised dataset along with machine learning methods. Additionally, the study seeks to explore the relationship between BMI, serum proteins, and brain ageing patterns.

Implications of all the available evidence

This study underscores that the impact of obesity and overweight on brain ageing and AD-like brain atrophy is more pronounced in males than females, highlighting potential sex-specific differences and the importance of early preventive measures. These findings contribute to our understanding of the neurocognitive consequences of obesity, paving the way for future studies elucidating mechanisms and pathways of obesity-related and sex-specific effects on the brain. The findings also highlight the potential role of brain age as an additional biomarker that can inform clinical trials and patient management that aim to control BMI.

Introduction

Obesity is considered a global epidemic of the 21st century (1-5). Notably, long-term increased body weight or abdominal obesity, as well as weight gain in midlife, is associated with a higher risk for incident dementia (6, 7). Specifically, in the United States, midlife obesity is the most prominent modifiable risk factor for Alzheimer's disease (AD) and related dementias, overtaking physical inactivity (8). The ageing of the population, along with the rapidly rising prevalence and incidence of overweight/obesity and obesity-related diseases, i.e., cardiovascular-kidney-metabolic diseases, are expected to result in a rise in morbidity, mortality and healthcare costs in the near future (9-11). The emerging recognition of neurocognitive decline as part of the cardiovascular-kidney-metabolic continuum, as well as addressing modifiable risk factors and unveiling the early pathophysiological changes before the onset of dementia, would enable personalised, preventive, and therapeutic interventions (12).

High body mass index (BMI), specifically ≥30 kg/m^2–indicating obesity, is postulated to impact brain integrity, relating to grey and white matter atrophy and dysregulated interregional circuitry, which may explain, in part, their higher risk for dementia (13, 14). Although only midlife obesity is associated with higher dementia incidence, the reason behind this remains unclear (15) and may be attributed to associations of dementia with weight loss in late life (12). A recent study revealed that individuals with increasing BMI in midlife and subsequently decreasing BMI in late life are at elevated risk for dementia, with BMI declining approximately seven to eight years preceding dementia diagnosis (14).

In the lack of conclusive evidence, consideration of molecular markers such as proteomics, may provide mechanistic insights into the link between BMI and brain atrophy across the human lifespan (16, 17). Recently, a large community-based prospective cohort study demonstrated that glial fibrillary acidic protein (GFAP), neurofilament light polypeptide (NfL), growth differentiation factor 15 (GDF15), and brevican core protein (BCAN) were associated with a 14-year dementia risk (18). However, the degree to which proteins modulated by metabolic risk factors, such as obesity, influence the pathophysiological process underlying dementia risk (brain atrophy and advanced ageing) still remains unclear. Therefore, differences in the cognitive state among individuals of the same age are encapsulated in the term brain ageing, which has emerged as a biomarker reflecting the individual's overall brain health. Brain ageing is typically employed to measure the deviation from the normative ageing trajectory, being strongly implicated in neurocognition (19).

In this multi-cohort study encompassing 46,288 harmonised brain magnetic resonance imaging (MRI) scans, we investigate brain ageing and AD-like brain

atrophy patterns (both quantified via machine learning indices), as well as neurocognitive changes in individuals with obesity or overweight. Second, we examine the links between BMI, proteomic alterations, deviations in brain ageing trajectories, and patterns of AD-like brain atrophy.

Methods

Study populations

MRI scans and clinical data used in this study were consolidated and harmonised by the Imaging-Based Coordinate System for Aging and Neurodegenerative Diseases (iSTAGING) study. The iSTAGING study comprises data acquired via various imaging protocols, scanners, data modalities, and pathologies, including 67,997 participants aged between 22 and 90 years from 17 studies on three continents. Specifically, the current study used MRIs from the following cohorts that met the inclusion criteria (Figure 1): Alzheimer's Disease Neuroimaging Initiative (ADNI), the Australian Imaging, Biomarkers and Lifestyle (AIBL), the Biomarkers of Cognitive Decline Among Normal Individuals (BIOCARD), the Human Connectome Project (HCP), the Multi-Ethnic Study of Atherosclerosis (MESA), the Open Access Series of imaging Studies (OASIS), the Penn study (PENN), the PResymptomatic EValuation of Experimental or Novel Treatments for AD (PREVENT-AD), the Women's Health Initiative Memory Study (WHIMS), the Wisconsin Registry for Alzheimer's Prevention (WRAP), the UK Biobank, the Baltimore Longitudinal Study of Aging (BLSA), the Coronary Artery Risk Development in Young Adults Study (CARDIA), the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS), and the Study of Health In Pomerania studies (SHIP).

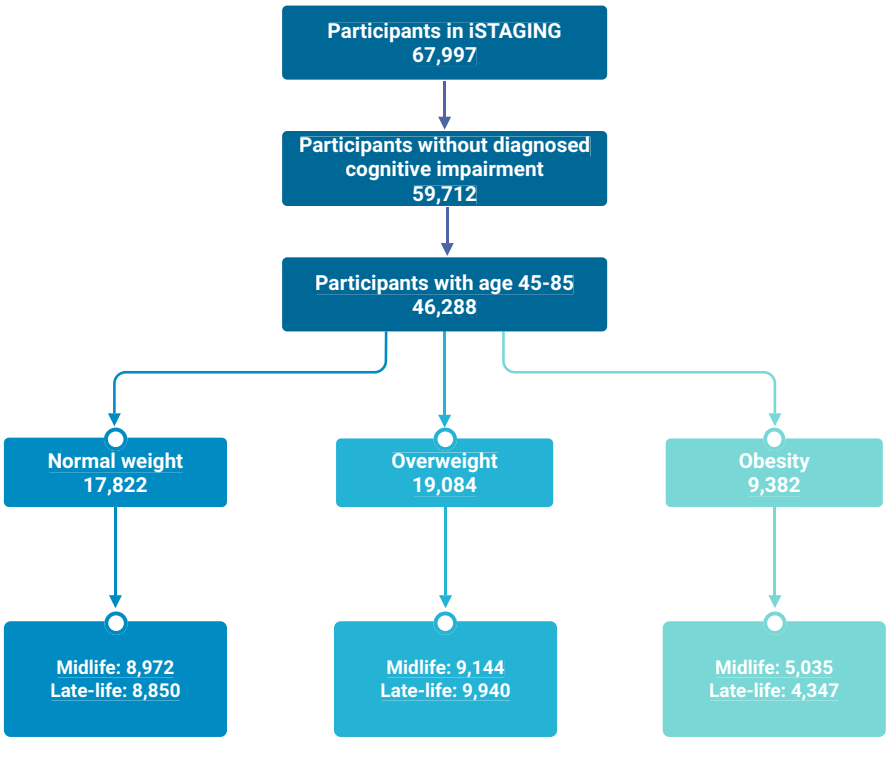


Figure 1. Pipeline.
Abbreviations: WODCI, Without Diagnosed Cognitive Impairment; BMI, Body Mass Index. Created with BioRender.com.

Subsequently, we grouped iSTAGING participants according to BMI categories (proposed by the World Health Organization) (20) as follows: (i) normal weight (BMI 18.5-24.9 kg/m^2), (ii) overweight (BMI 25.0-29.9 kg/m^2), and (iii) obesity (BMI ≥30 kg/m^2). The study included participants without diagnosed cognitive impairment (WODCI), defined as healthy controls or cognitively normal participants across the studies. Each study has its own criteria for defining cognitively normal or individuals without diagnosed cognitive impairment. The term 'without diagnosed cognitive impairment' encompasses participants labelled as either 'control (CN)', 'CN assumed by study criteria', 'cognitively unimpaired', 'cognitively normal', 'memory complainer (healthy control)', 'non-memory complainer (healthy control)', 'normal', or having 'normal cognition' based on the originating studies, which used study-specific criteria for such designations. The distribution of previous diagnoses across subgroups did not show significant differences. More information about each study separately can be found in our previous publications (21). Baseline characteristics for each population are presented in Tables 1–4.

Table 1. Demographic summary and volumetric measures of the WODCI sample (baseline scans).
N=46,288. Other races: Hispanic/Latino, Native American, Multiracial, unknown, other; information about races is presented as given in the originating studies.

Study	Sample size	Sex (%)	(%Asian)	(%Black)	(%Other)	(%White)	Age (years, SD)	APOE-e4 (% with one or two alleles)	Intracranial Volume (mm^3)
ADNI	908	Males: 44.05 Females: 55.95	0.77	4.3	32.49	62.33	72.01 +-5.98	28.74	1408323+-140698.81
AIBL	634	Males: 42.90 Females: 57.10	0	0	0	100	72.39+-5.81	21.81	1407718+-139867.99
BIOCARD	209	Males: 38.76 Females: 61.24	1.48	0.99	0.49	97.12	59.13+-7.75	34.45	1402292+-142628.68
BLSA	881	Males: 43.13 Females: 56.87	4.42	23.95	2.76	69.69	67.41 +- 10.39	24.86	1388496+-140222.09
CARDIA	860	Males: 46.63 Females: 53.37	0	40.81	0	59.19	51.63 +- 3.54	22.21	1213145+-126950.3
HANDLS	169	Males: 43.79 Females: 56.21	0	36.69	0	63.31	56.97+-6.65	0	1164199 +-115887.2
HCP	923	Males: 45.12 Females: 54.88	NA	NA	NA	NA	63.34+-11.42	0	1409371+-140042.73
OASIS	718	Males: 42.42 Females: 57.58	0.84	15.18	0.14	83.98	68.99+-8.34	32.55	1394015+-144311.51
PENN	12	Males: 50 Females: 50	0	25.00	0	75.00	69.13+-6.28	41.67	1395319+-152054.99
PreventAD	341	Males: 29.33 Females: 70.67	0	0	1.17	98.82	63.64+-5.06	38.42	1368309+-125066.79
SHIP	1417	Males: 47.92 Females: 52.59	0	0	0	100	59.07 +-9.21	NA	1225329 +-123612.5
UK BIOBANK	37894	Males: 47.41 Females: 52.59	1.34	0.61	0.90	97.12	64.08+-7.54	23.06	1449157 +-141275.97
WHIMS	592	Males: 0 Females: 100	1.35	2.53	2.36	93.75	73.09+-3.92	18.92	1332944+-97809.41
WRAP	217	Males: 30.41 Females: 65.59	0.46	2.30	2.30	94.93	63.6+-6.25	36.87	1423501+-127227.91

Table 2. **Demographic summary of the subgroups in each age group with normal weight.**

Age group	Sample size	Sex (%)	Study	Age (years, SD)	Intracranial brain volume
Midlife	8972	Males: 33.58 Females: 66.42	ADNI = 29 AIBL=24 BIOCARD = 65 BLSA = 123 CARDIA = 243 HANDLS = 30 HCP = 104 MESA = 24 OASIS = 46 PENN = 1 PreventAD = 88 SHIP = 232 UK BIOBANK = 7923 WHIMS = 1 WRAP = 39	57.66 +-4.80	1423498 +-136489.85
Late-life	7651	Males: 43.28 Females: 56.72	ADNI = 279 AIBL = 225 BIOCARD = 16 BLSA = 187 CARDIA = 0 HANDLS = 4 HCP = 87 MESA = 235 OASIS = 173 PENN = 3 PreventAD = 38 SHIP = 67 UK BIOBANK = 7368 WHIMS = 151 WRAP= 17	71.14 +- 4.22	1431580 +-139312.98

Table 3. Demographic summary of the subgroups in each age group with overweight.

Age group	Sample size	Sex (%)	Study	Age (years, SD)	DLICV
Midlife	9144	Males: 52.05 Females: 47.95	ADNI = 30 AIBL = 20 BIOCARD = 64 BLSA = 121 CARDIA = 296 HANDLS = 42 HCP = 106 MESA = 46 OASIS = 67 PENN = 1 PreventAD = 95 SHIP = 423 UK BIOBANK = 7787 WHIMS = 5 WRAP = 41	57.66 +-4.81	1453659 +-144234.41
Late-life	9940	Males: 56.23 Females: 43.77	ADNI = 343 AIBL = 249 BIOCARD = 22 BLSA = 234 CARDIA = 0 HANDLS = 13 HCP = 92 MESA = 322 OASIS = 214 PENN = 4 PreventAD = 49 SHIP = 208 UK BIOBANK = 7893 WHIMS = 252 WRAP = 45	71.17 +- 4.12	1452121 +-144148.25

Table 4. **Demographic summary of the subgroups in each age group with obesity.**

Age group	Sample size	Sex (%)	Study	Age (years, SD)	DLICV
Midlife	5035	Males: 44.09 Females: 55.91	ADNI = 26 AIBL = 11 BIOCARD = 35 BLSA = 92 CARDIA = 321 HANDLS = 70 HCP = 77 MESA = 45 OASIS = 70 PENN = 0 PreventAD = 48 SHIP = 329 UK BIOBANK = 3869 WHIMS=1 WRAP=41	57.23+-4.94	1431021 +-148293.75
Late-life	4347	Males: 48.03 Females: 51.97	ADNI = 201 AIBL = 105 BIOCARD = 7 BLSA = 124 CARDIA = 0 HANDLS = 10 HCP = 46 MESA = 251 OASIS = 149 PENN = 3 PreventAD = 23 SHIP = 158 UK BIOBANK = 3054 WHIMS = 182 WRAP = 34	70.84 +-4.08	1430786+-145615.76

Image Preprocessing

A fully automated processing pipeline was applied to extract region-of-interest (ROI) volumes from each participant's T1-weighted MRI scan. Preprocessing involved image intensity correction for inhomogeneities (22), and multi-atlas skull-stripping was applied to remove extra-cranial material (23). In this approach, a collection of pre-labelled atlases is warped in parallel to a participant's MRI. The warped atlases are combined via a weighted voting procedure, which ensures that the most suitable/similar atlases have the highest weights in this process (24). 145 anatomic ROIs, predominantly located in grey matter, were segmented using a multi-atlas, multi-warp label fusion-based method (24). These multi-atlas segmentation techniques provide robustness to errors in specific deformations by utilising several anatomical reference images (atlases), which are adjusted to align with the target image through deformable registration. Table 5 reports imaging parameters from the parent studies. Additional details about image preprocessing were reported previously (25, 26). Volumes of the 145 ROIs were used in the generation of ML-markers described below and were compared in the proteomic analysis (Table 6).

Table 5. Imaging parameters of the individual studies included in the iSTAGING consortium. Detailed information about each study protocol can be found elsewhere (26).

Study	Scanner	T1 Resolution
ADNI-1	1.5T General Electric, Phillips, and Siemens	0.94 x 0.94 x 1.2mm
ADNI-GO/ADNI-2	3T General Electric, Philips, and Siemens	1 x 1 x 1.2mm
ADNI-3	3T General Electric, Phillips, and Siemens	1 x 1 x 1/1.2mm
AIBL	1.5T and 3T Siemens, Avanto, TrioTim, Verio	1 x 1 x 1.2mm
BIOCARD-1.5T	1.5T General Electric	1 x 1 x 2mm
BIOCARD-3T	3T Philips Achieva	1 x 1 x 1.2mm
BLSA-1.5T	1.5T General Electric Signa	0.94 x 0.94 x 1.5mm
BLSA-3T	3T Philips	1 x 1 x 1.2mm
CARDIA	3T Siemens Tim Trio, Philips, Achieva	1 x 1 x 1mm
HANDLS	3T Siemens Trio	1 x 1 x 1mm
HCP	3T Siemens Prisma (VE11C)	0.7×0.7×0.7mm
MESA	3T Siemens Prisma (VE11C), Siemens Skyra (VD11B)	1 x 1 x 1mm
OASIS 1.5T	1.5T Siemens	1 x 1 x 1.25mm
OASIS 3T	3T Siemens	1 x 1 x 1mm
PENN	3T Siemens Trio	1 x 1 x 1mm
PreventAD	3T Siemens Trio	1 x 1 x 1mm
SHIP	1.5T Siemens Magnetom Avanto	1×1×1 mm
UK BIOBANK	3T Siemens Skyra (VD13)	1 x 1 x 1mm
WHIMS	1.5T General Electric, Philips, and Siemens	1 x 1 x 1.5mm
WRAP	3T General Electric, Discovery	1 x 1 x 1mm

Table 6. **ROIs used in general linear models. Abbreviations: GM, Grey Matter; WM, White Matter; DC, Diencephalon; (L), Left hemisphere; (R), Right hemisphere.**

Grey and white matter anatomic ROIs				
3rd Ventricle	Lateral orbital gyrus (R)	4th Ventricle	Lateral orbital gyrus (L)	Accumbens area (R)
Middle cingulate gyrus (R)	Accumbens area (L)	Middle cingulate gyrus (L)	Amygdala (R)	Medial frontal cortex (R)
Amygdala (L)	Medial frontal cortex (L)	Brain Stem	Middle frontal gyrus (R)	Caudate (R)
Middle frontal gyrus (L)	Caudate (L)	Middle occipital gyrus (R)	Cerebellum exterior (R)	Middle occipital gyrus (L)
Cerebellum exterior (L)	Medial orbital gyrus (R)	Cerebellum WM (R)	Medial orbital gyrus (L)	Cerebellum WM (L)
Postcentral gyrus medial segment (R)	Hippocampus (R)	Postcentral gyrus medial segment (L)	Hippocampus (L)	Precentral gyrus medial segment (R)
Inferior lateral ventricle (R)	Precentral gyrus medial segment (L)	Inferior lateral ventricle (L)	Superior frontal gyrus medial segment (R)	Lateral ventricle (R)
Superior frontal gyrus medial segment (L)	Lateral ventricle (L)	Middle temporal gyrus (R)	Pallidum (R)	Middle temporal gyrus (L)
Pallidum (L)	Occipital pole (R)	Putamen (R)	Occipital pole (L)	Putamen (L)
Occipital fusiform gyrus (R)	Thalamus (R)	Occipital fusiform gyrus (L)	Thalamus (L)	Opercular part of the inferior frontal gyrus (R)
Ventral DC (R)	Opercular part of the inferior frontal gyrus (L)	Ventral DC (L)	Orbital part of the inferior frontal gyrus (R)	Cerebellar Vermal Lobules I-V
Orbital part of the inferior frontal gyrus (L)	Cerebellar Vermal Lobules VI-VII	Posterior cingulate gyrus (R)	Cerebellar Vermal Lobules VIII-X	Posterior cingulate gyrus (L)
Basal forebrain (L)	Precuneus (R)	Basal forebrain (R)	Precuneus (L)	Frontal lobe WM (R)
Parahippocampal gyrus (R)	Frontal lobe WM (L)	Parahippocampal gyrus (L)	Occipital lobe WM (R)	Posterior insula (R)
Occipital lobe WM (L)	Posterior insula (L)	Parietal lobe WM (R)	Parietal operculum (R)	Parietal lobe WM (L)
Parietal operculum (L)	Temporal lobe WM (R)	Postcentral gyrus (R)	Temporal lobe WM (L)	Postcentral gyrus (L)
Fornix (R)	Posterior orbital gyrus (R)	Fornix (L)	Posterior orbital gyrus (L)	Anterior limb of internal capsule (R)

Table 6. ROIs used in general linear models. Abbreviations: GM, Grey Matter; WM, White Matter; DC, Diencephalon; (L), Left hemisphere; (R), Right hemisphere. (continued)

Grey and white matter anatomic ROIs				
Planum polare (R)	Anterior limb of internal capsule (L)	Planum polare (L)	Posterior limb of internal capsule inc. cerebral peduncle (R)	Precentral gyrus (R)
Posterior limb of internal capsule inc. cerebral peduncle (L)	Precentral gyrus (L)	Corpus callosum	Planum temporale (R)	Anterior cingulate gyrus (R)
Planum temporale (L)	Anterior cingulate gyrus (L)	Subcallosal area (R)	Anterior insula (R)	Subcallosal area (L)
Anterior insula (L)	Superior frontal gyrus (R)	Anterior orbital gyrus (R)	Superior frontal gyrus (L)	Anterior orbital gyrus (L)
Supplementary motor cortex (R)	Angular gyrus (R)	Supplementary motor cortex (L)	Angular gyrus (L)	Supramarginal gyrus (R)
Calcarine cortex (R)	Supramarginal gyrus (L)	Calcarine cortex (L)	Superior occipital gyrus (R)	Central operculum (R)
Superior occipital gyrus (L)	Central operculum (L)	Superior parietal lobule (R)	Cuneus (R)	Superior parietal lobule (L)
Cuneus (L)	Superior temporal gyrus (R)	Entorhinal area (R)	Superior temporal gyrus (L)	Entorhinal area (L)
Temporal pole (R)	Frontal operculum (R)	Temporal pole (L)	Frontal operculum (L)	Triangular part of the inferior frontal gyrus (R)
Frontal pole (R)	Triangular part of the inferior frontal gyrus (L)	Frontal pole (L)	Transverse temporal gyrus (R)	Fusiform gyrus (R)
Transverse temporal gyrus (L)	Fusiform gyrus (L)	Gyrus rectus (R)	Gyrus rectus (L)	Inferior occipital gyrus (R)
Inferior occipital gyrus (L)	Inferior temporal gyrus (R)	Inferior temporal gyrus (L)	Lingual gyrus (R)	Lingual gyrus (L)

Spatial Pattern of Abnormality for Recognition of Early Alzheimer's Disease (SPARE-AD) index

We derived machine learning metrics of brain age and AD-signature region brain atrophy, termed Spatial Pattern of Abnormality for Recognition of Early (SPARE) indices. All SPARE scores were obtained from models trained on different datasets, either via 5-fold cross-validation or via out-of-sample application of pre-trained models, ensuring no overlapping of the data. More specifically, for the data used in constructing the two SPARE models, 5-fold cross-validation was applied so

that all SPARE scores were derived when respective participants were in the test set, i.e., in the 20% not participating in model training. For the datasets that were not part of the training, SPARE scores were obtained by applying already-trained models to these external datasets. One SPARE model predicts brain age (SPARE-BA) using Support Vector Regression (27). The SPARE-BA model was previously trained on 145 single harmonised multi-atlas parcellation method (MUSE) ROI data from the iSTAGING consortium from four different studies with over 4000 participants and is detailed in Hwang et al. (28). Pooling imaging datasets from multiple studies is prone to study-related effects due to the lack of standardised image acquisition protocols and scanner variability. Systematic differences in sample demographics between studies may also confound analyses. To preserve study-related biological effects while correcting for systematic technical variance, we applied the Neuroharmonize toolbox (28) based on the multi-variate ComBat method combined with generalised additive cubic spline models to capture nonlinear age, sex, and deep learning-based intracranial volume measurement (DLICV) effects (29). This method reduces acquisition-related effects and preserves variability due to biological covariates. Each ROI volume was modelled as a nonlinear function of age, sex, and baseline DLICV. Based on the adjusted data, the remaining systematic differences in shift (location) and variance (scale) were attributed to site-specific acquisition settings and adjusted conservatively with an empirical Bayes regularisation. The harmonisation model was trained on each site's baseline scans of WODCI and then applied to the entire dataset. This method has already been validated in other works (28). A linear age bias correction was applied to both structural and functional SPARE-BA measures to remove the known systematic bias inherent in brain age calculations (30). Brain age gap, calculated by subtracting the chronological age from the predicted brain age (SPARE-BA) yielded a Pearson's correlation coefficient r=0.005 with chronological age. Higher SPARE-BA, relative to chronological age, is due to greater brain atrophy than typical for age, indicating advanced brain ageing. We measured brain atrophy in typical regions (regions from the temporal lobe, the cingulate and the insula) affected by AD using the SPARE Alzheimer's disease (SPARE-AD) model, which was derived using a support vector machine with a linear kernel trained to distinguish controls from participants with AD (27, 31). Detailed information about the SPARE-AD model can be found in previous publications (27, 31). Higher SPARE-AD values indicate a more pronounced presence of the AD-signature of regional brain atrophy, while negative values indicate brain patterns more similar to the control group; SPARE-AD values are unit-less, as they reflect the degree of expression of this imaging signature.

Blood proteomics

The UK Biobank blood sample collection was performed in 22 centres across the UK during the baseline visit in 2007-2010 and then during the imaging visit

to the UK Biobank in 2014. Blood samples from each participant were collected in ethylenediaminetetraacetic acid (EDTA) tubes and centrifuged at 2,500 g for 10 min at 4 °C to isolate plasma. The resulting supernatant was then portioned into aliquots and frozen at −80 °C until further analysis. Samples were subjected to quantification using the antibody-based Olink Explore Proximity Extension Assay. Proteomic profiling was done on plasma samples from 53,014 UK Biobank participants from the baseline visit in 2007-2010 and 1,173 UK Biobank participants from the imaging visit in 2014 (https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id=30900). After quality control, a total of 2922 unique proteins in the baseline visit in 2007-2010, and 1,463 unique proteins in the imaging visit in 2014 were measured across four panels focusing on cardiometabolic, inflammation, neurology, and oncology proteins (https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/PPP_Phase_1_QC_dataset_companion_doc.pdf). The inter- and intraplate coefficients of variation for all Olink panels were lower than 20% and 10%, respectively. Protein levels were represented as Normalised Protein eXpression (NPX) values, calculated by dividing sample and assay counts by extension control counts, followed by log transformation. Variation within and between plates was minimised by considering the median of extension control-normalised counts, batch-specific median NPX value, and the difference of the assay-specific median NPX value of each batch (https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/Olink_1536_B0_to_B7_Normalization.pdf).

Cognitive and CSF biomarkers

We included cognitive tests provided by eight studies (ADNI, BLSA, HANDLS, HCP, OASIS, PENN, UK Biobank, and WRAP), and CSF biomarkers from the ADNI study. More details about the participants of each study can be found in Table 7. Cognitive test measurements include Trail Making Test Parts A (TMT-A) and B (TMT-B), Digit Span Forward (DSF) and Backward (DSB), and Mini-Mental Status Exam (MMSE). CSF biomarkers contain phosphorylated tau (p-tau), total tau (t-tau), and beta-amyloid (Aβ) (Figure 2), and are measured in CSF aliquots using the multiplex xMAP Luminex platform (Luminex Corp, Austin, TX) or BioPlex 100 immunoassay platform (Bio Rad, Hercules, CA, USA) with the same Luminex platform, supplied by Innogenetics (kit: INNO-BIA AlzBio3, Ghent, Belgium; for research use only, assay lot #157353, calibrator lot #157379) immunoassay kit-based reagents. The kit reagents contain a mixture of three xMAP color-coded carboxylated microspheres, each containing a bead set coupled with well-characterized capture mAbs specific for Aβ1–42, t-tau, or p-tau181 (32). For ADNI, all measures were downloaded from the LONI website (adni.loni.ucla.edu).

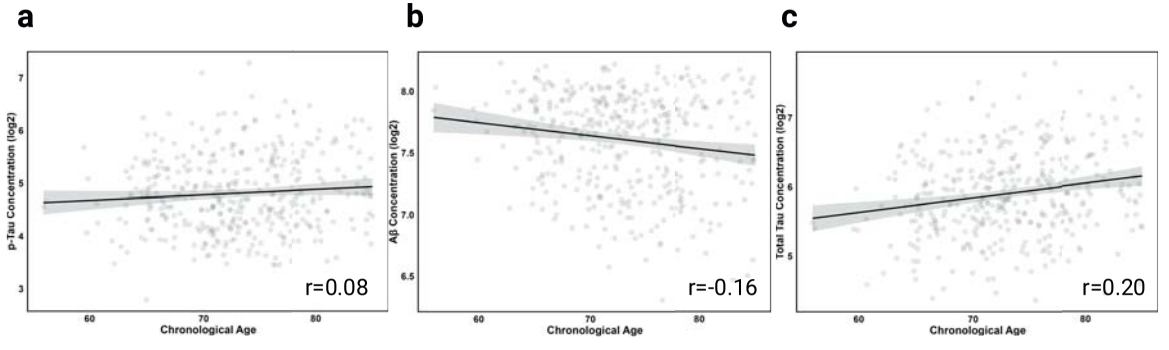


Figure 2. **a) Scatter plot of p-Tau concentration (log2) and chronological age. b) Scatter plot of Aβ concentration (log2) and chronological age. c) Scatter plot of total Tau concentration (log2) and chronological age. The r corresponds to Pearson's r coefficient.**

Table 7. **Cognitive and CSF biomarker status data availability for the N=46,288 WODCI participants within two years of the baseline brain MRI scan.**
If there was a mismatch between scan acquisition dates and those non-imaging feature measurement dates and BMI measurement dates, the search window for matching imaging and non-imaging data was set to 730 days (2 years). Abbreviations: CVLT, California Verbal Learning Test; MMSE, Mini-Mental State Examination; DSB, Digit Span Backward; DSF, Digit Span Forward; TMT-B, Trail Making Test B; TMT-A, Trail Making Test A.

	Total Tau	p-Tau	Beta Amyloid	CVLT	MMSE	DSB	DSF	TMT-B	TMT-A
Total participants	723	725	725	834	980	1,023	1,025	25,388	25,972
Participants per study	ADNI = 388	ADNI = 390	ADNI = 390	AIBL = 10 BLSA = 823 WHIMS = 1	ADNI =35 AIBL =10 BLSA = 755 HANDLS =150 OASIS =12 PENN =11 WRAP =7	ADNI = 3 AIBL = 6 BLSA = 845 HANDLS = 138 OASIS =13 PENN = 10 WHIMS =1 WRAP = 7	ADNI = 3 AIBL = 6 BLSA = 846 HANDLS = 139 OASIS =13 PENN =10 WHIMMS =1 WRAP =7	ADNI = 10 BLSA = 852 HANDLS = 157 HCP = 509 OASIS = 13 PENN = 10 UKBIOBANK = 23830 WRAP = 7	ADNI = 10 BLSA = 858 HANDLS = 158 HCP = 508 OASIS=13 PENN=10 UKBIOBANK = 24408 WRAP = 7

Statistics

Initially, we employed generalised linear models to analyse each machine learning index (SPARE-BA or SPARE-AD), with BMI groups, age, sex, intracranial volume, and interaction terms for BMI group × age, and BMI group × sex as predictors. We performed post-hoc comparisons between sex groups and BMI categories, adjusting *p* value by Tukey's test using the "emmeans" package in R. Cognitive test measurements were standardised, z-scores were calculated, and associations between BMI groups were derived from linear regression models. CSF biomarkers were log2 transformed and compared between BMI groups using linear regression. For the aforementioned regressions, BMI group, age, sex, and intracranial volume were used as covariates. These covariates are selected based on expertise and the existing literature.

The first proteomic analysis examined 2,922 proteins in 5,021 participants at the baseline visit and 1,460 proteins in 1,028 participants at the imaging visit. Then, we used generalised linear models with SPARE-BA as output and each protein, BMI (continuous), age, sex, and intracranial volume as predictors. Similarly, for each protein biomarker we used generalised linear models with SPARE-AD as output and each protein, BMI (continuous), age, sex, and intracranial volume as predictors. For each proteomic feature, adjusted standard deviation (SD) differences and 95% confidence interval (CI) associated with a one SD change in BMI (SD=4.2 kg/m^2) were estimated with BMI (continuous), age, and sex as predictors. Results with false discovery rate (FDR) values of <0.05 were considered significant.

We undertook mediation analysis to estimate the proportion of the effect of the BMI on brain ageing mediated by each significant protein from both visits using the mediation package in R. For this analysis, we used BMI measurements from the baseline visit and protein measurements from blood collected at the imaging visit from UKB participants and SPARE-BA from the imaging visit in 2014. In total, 4951 participants had both BMI baseline measurements and SPARE-BA values during the imaging visit. First, we estimated the effect of BMI separately on each protein using a linear regression model. Next, we fitted a second regression model to estimate the effect of BMI on brain ageing. The mediation function was then used to decompose the total effect of BMI on brain ageing into an average causal mediation effect (ACME) and an average direct effect (ADE), with nonparametric bootstrapping (1000 simulations) to calculate CIs. The proportion of the total effect mediated by each protein was calculated as the ratio of the ACME to the total effect. In all regressions age, sex, and intracranial volume were used as covariates. All analyses were done using R Studio (version 2022.12.0+353).

Ethics

The protocols of this study were approved by the University of Pennsylvania Institutional Review Board (reference number: 825722). Participants provided written informed consent to the corresponding studies.

Role of funders

Funders for the current study were not directly involved in the design, data collection, analysis, interpretation or writing of this manuscript.

Results

AD-like Brain Atrophy Indices and BMI Categories

This study included MRI scans from 17,822 participants with normal-weight, 19,084 with overweight, and 9,382 participants with obesity, with a mean (SD) age of 64.35 (8.11) years, 64.7 (8.09) and 63.54 (8.18), respectively. Figure 1 depicts a flowchart with inclusion and exclusion criteria. Participant characteristics at baseline scan can be found in Tables 1-4. Overall, participants with obesity presented a significant association with brain ageing when compared to participants with normal weight (β = 11.20; SE = 1.13; $p_{adj} < 0.001$, Tukey's test), and to participants with overweight (β = 8.17; SE = 1.11; $p_{adj} < 0.001$, Tukey's test). Participants with overweight displayed higher brain ageing values than participants with normal weight (β = 3.04; SE = 0.927; $p_{adj} = 0.003$, Tukey's test) (Figure 3). BMI group and age interaction: (Overweight: β = -0.072; SE = 0.113; p =0.525 generalised linear models, Obesity: β = -0.277; SE = 0.137; p = 0.043 generalised linear models). Among males, obesity was associated with increased SPARE-BA when compared to normal weight (β = 21.39; SE = 1.71; $p_{adj} < 0.001$, Tukey's test), and overweight (β = 12.63; SE = 1.59; $p_{adj} <0.001$, Tukey's test) (Figure 3c). Males with overweight exhibited higher SPARE-BA values than males with normal weight (β = 8.76; SE = 1.37; $p_{adj} < 0.001$, Tukey's test). Females with obesity showed higher SPARE-BA values in comparison to females with overweight (β = 3.70; SE = 1.55; $p_{adj} = 0.044$, Tukey's test). Furthermore, in the normal-weight group, females showed higher SPARE-BA values than males (β = 8.29; SE = 1.51; $p_{adj} < 0.001$, Tukey's test). Males compared to females showed increased SPARE-BA in the group with obesity (β = 12.08; SE = 1.94; $p_{adj} < 0.001$, Tukey's test), and in the overweight group (β = 3.15; SE = 1.44; $p_{adj} = 0.029$, Tukey's test) (Figure 3d).

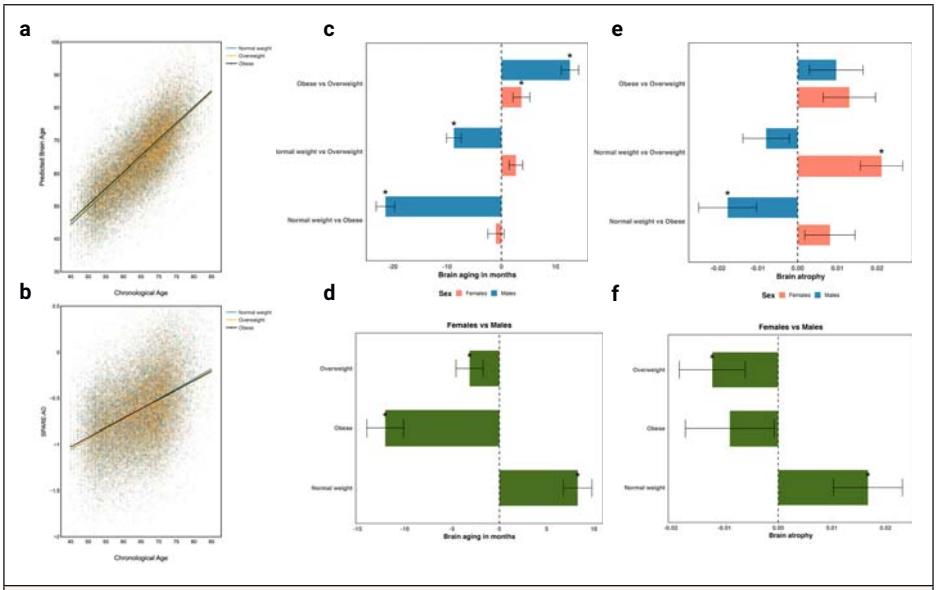


Figure 3. a) Scatter plot of chronological age and predicted age for each BMI category (normal weight, overweight, obesity). b) Scatter plot of chronological age and SPARE-AD for each BMI category (normal weight, overweight, obesity). c) Bar plot of SPARE-BA between BMI categories in males and in females. d) Bar plot of SPARE-BA of males versus females in each BMI category. e) Bar plot of SPARE-AD between BMI categories in males and in females. f) Bar plot of SPARE-AD of males versus females in each BMI category. Asterisk (*) is put above the bar plot where the FDR-adjusted p-value is <0.05.

In contrast, females in the normal-weight group exhibited higher SPARE-AD values than females in the overweight group (β = 0.02; SE = 0.005; p_{adj} = 0.002, Tukey's test) (Figure 3e). Males in the obesity group showed higher SPARE-AD than males with normal-weight (β = -0.02; SE = 0.007; p_{adj} = 0.041, Tukey's test) (Figure 3e). Additionally, females displayed higher values of SPARE-AD than males in the normal-weight group (β = 0.02; SE = 0.006; p_{adj} = 0.009, Tukey's test), but in the overweight group males showed higher SPARE-AD values (β = -0.01; SE = 0.006; p_{adj} = 0.045, Tukey's test) than females (Figure 3f). BMI group and age interaction: (Overweight: β = -0.006; SE = 0.006; p =0.303, generalised linear models, Obesity: β = -0.019; SE = 0.007; p = 0.008, generalised linear models). Age-stratified analysis of brain atrophy indices, and association with neurodegeneration biomarkers and cognitive tests can be found in the Supplementary Analysis s1-s2, Supplementary Tables s1-s2, and Supplementary Figures s1-s3.

Proteomic Markers and Brain Ageing

After adjusting for multiple comparisons, significant positive associations between brain ageing and 191 of the 2,922 protein biomarkers were identified at the baseline visit (at a 5% FDR; Figure 4a). Among these, 50 proteins showed a positive significant association with both brain ageing and BMI (per one standard

deviation increase in BMI as a continuous variable) (Figure 4b), and 43 proteins demonstrated a negative association with both brain ageing and BMI (Figure 4b). Growth/differentiation factor 15 (GDF15) was the only protein significantly associated with AD-like brain atrophy (Figure 4c).

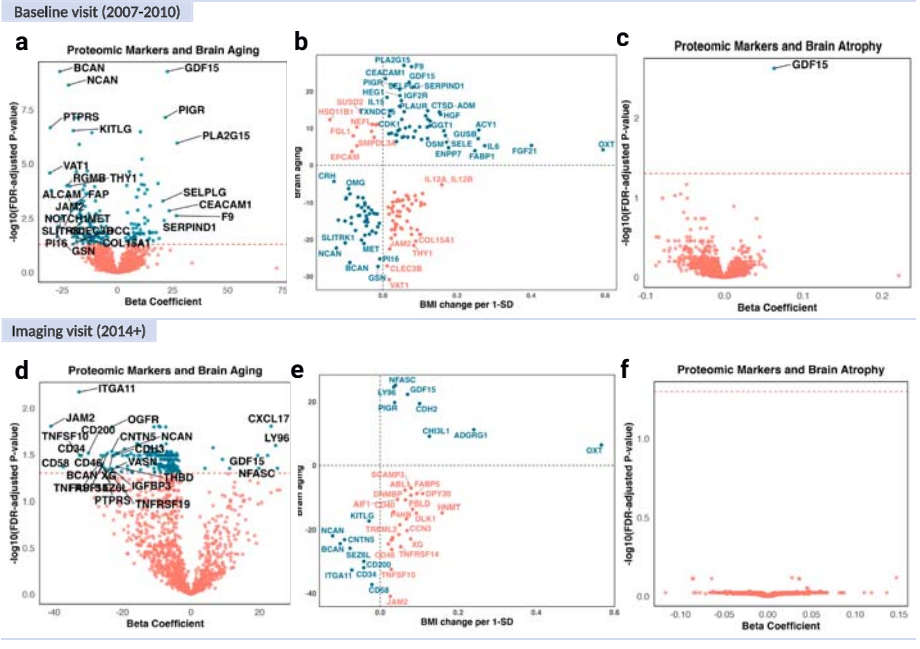


Figure 4. a) Volcano plot showing the proteins associated with brain ageing (expressed in months) at the baseline visit; the top 25 proteins are labeled. b) Scatter plot of the FDR<0.05 adjusted proteins associated with both brain ageing and BMI change per 1-SD at the baseline visit. c) Volcano plot showing the proteins associated with brain atrophy at the baseline visit. d) Volcano plot showing the proteins associated with brain ageing (expressed in months) at the imaging visit; the top 25 proteins are labeled. e) Scatter plot of the FDR<0.05 adjusted proteins associated with both brain ageing and BMI change per 1-SD at the imaging visit. f) Volcano plot showing the proteins associated with brain atrophy at the imaging visit.

At the imaging visit, 215 of the 1,460 protein biomarkers exhibited significant associations with brain ageing (at a 5% FDR; Figure 4d). Of these, 10 proteins were positively correlated, and 205 proteins were negatively correlated with brain ageing (Figure 4d). Among the 10 proteins positively associated with brain ageing, eight also showed a significant positive association with BMI. In contrast, of the 205 proteins negatively associated with brain ageing, nine proteins also demonstrated a negative association with BMI (Figure 4e). At the imaging visit, no protein displayed a significant association with AD-like brain atrophy after adjusting for multiple comparisons (Figure 4f).

At both visits, five proteins were positively associated with brain ageing, while 15 proteins were negatively associated with brain ageing, all at a 5% FDR correction level (Figure 5a). All of the five proteins that were associated positively with brain ageing, also demonstrated a significant positive association with BMI (per one SD increase in BMI as a continuous variable): Oxytocin-neurophysin 1 (OXT), chitinase-3-like protein 1 (CHI3L1), GDF15, Polymeric immunoglobulin receptor (PIGR), Lymphocyte antigen 96 (LY96) (at 5% FDR; Figure 5a, Table 8). On the other hand, eight proteins were associated negatively with both brain ageing and BMI decrease by 1-SD: CD58, Integrin a-11 (ITGA11), CD34, CD200, Seizure protein six homolog (SEZ6L), brevican core protein (BCAN), neurocan core protein (NCAN), Kit ligand (KITLG) (at 5% FDR; Figure 5a, Table 8). Proteins associated with brain ageing were further investigated regarding their association with BMI categories (Figure 5b).

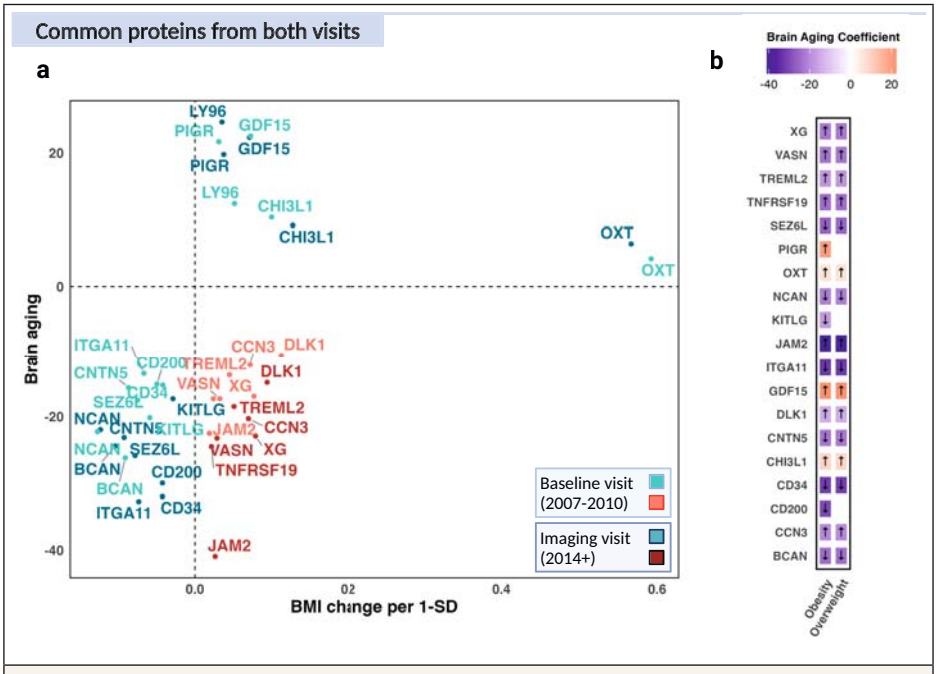


Figure 5. **a) Volcano plot of the proteins associated with brain ageing and BMI change per 1-SD at both baseline and imaging visit. b) Heatmap of the FDR<0.05 adjusted proteins associated with brain ageing, showing if they are expressed in obesity or in overweight in comparison with controls with normal weight.**

Table 8. Common proteins associated with both positive and negative brain ageing and BMI change per 1-SD in the baseline (2007-2010) and imaging visit (2014+).

Protein	Brain Ageing (months)	SE	FDR Adjusted p-value	BMI Adjusted SD (95% CI)	FDR Adjusted p-value	Brain Ageing (months)	SE	FDR Adjusted p-value	BMI Adjusted SD (95% CI)	FDR Adjusted p-value
Proteins associated positively with both brain ageing and BMI										
Baseline visit						**Imaging visit**				
LY96	12.4	3.3	5.41e-03	0.05 (0.04, 0.05)	8.73e-19	24.7	6.7	0.025	0.035 (0.029, 0.041)	6.96e-03
GDF15	22.5	3.1	5.30c-10	0.07 (0.07, 0.07)	3.58e-33	22.3	7.1	0.032	0.070 (0.065, 0.076)	3.22e-09
PIGR	21.7	3.4	7.07e-08	0.03 (0.03, 0.03)	4.38e-08	19.8	7.1	0.044	0.037 (0.032, 0.043)	2.20e-03
CHI3L1	10.4	1.7	3.28e-07	0.099 (0.094, 0.104)	8.91e-19	9.2	3.1	0.036	0.126 (0.113, 0.139)	5.90e-06
OXT	4.2	0.8	9.96e-07	0.59 (0.58, 0.60)	9.46e-153	6.4	1.8	0.027	0.567 (0.545, 0.589)	4.99e-30
Proteins associated negatively with both brain ageing and BMI										
ITGA11	-13.47	3.5	4.19e-03	-0.07 (-0.07, -0.06)	6.53e-34	-32.7	7.1	0.007	-0.072(-0.078, - 0.067)	2.4e-09
CD34	-15.2	4.9	3.22e-02	-0.04 (-0.04, -0.04)	6.87e-27	-32	10.1	0.032	-0.042 (-0.046, -0.038)	9.56e-07
CD200	-15.1	4.5	1.97e-02	-0.05 (-0.05, -0.05)	5.51e-32	-30	8.8	0.031	-0.042 (-0.046, -0.037)	2.04e-05
SEZ6L	-17.5	4.4	2.93e-03	-0.07 (-0.08, -0.07)	2.00e-66	-25.8	9.1	0.042	-0.077 (-0.081, -0.073)	6.67e-16
BCAN	-26.2	3.6	5.3e-10	-0.09 (-0.09, -0.09)	1.81e-63	-24.4	7.2	0.031	-0.102(-0.108, -0.097)	1.09e-16
CNTN5	-15.6	3.2	1.47e-06	-0.09 (-0.09, -0.08)	1.81e-48	-23	6.5	0.029	-0.091 (-0.097, -0.085)	8.56e-12
NCAN	-22.3	3.2	2.3e-09	-0.13 (-0.13, -0.12)	3.77e-98	-21.9	6.6	0.031	-0.122(-0.128, -0.116)	8.75e-20
KITLG	-20.1	3.3	2.9e-07	-0.06 (-0.06, -0.06)	2.99e-24	-17	6.4	0.048	-0.028 (-0.034, -0.022)	3.36e-02

We examined the mediating effects of each of the proteins that displayed significant association with brain ageing in both visits (LY96, GDF15, PIGR, CHI3L1, OXT, ITGA11, CD34, CD200, SEZ6L, BCAN, NCAN and KITLG) on the association between BMI and brain ageing, as measured by the SPARE-BA score (Supplementary Figure s4). The average causal mediation effect (ACME) was statistically significant for all proteins included, indicating that each protein mediates a portion of BMI's impact on brain ageing. In contrast, the average direct effect (ADE) was not significant, suggesting that BMI's influence on brain ageing may occur primarily through other pathways rather than direct pathways. The total effect of BMI on brain ageing was also nonsignificant, while the proportion of the total effect mediated was nonsignificant (Supplementary Tables s3-s14).

Discussion

This multi-cohort study, including 46,288 adults, aims to compare MRI-brain ageing indices in participants with obesity or overweight and elucidate the associations between proteomic features and these MRI-brain ageing indices. Our analysis of this large, diverse MRI dataset revealed that the impact of obesity on brain ageing, and AD-like brain atrophy is more pronounced in males than females, and its effects are weaker with increasing age. Moreover, higher BMI states are linked to accelerated brain ageing and AD-like brain atrophy, especially in males, whereas females with normal-weight exhibited greater brain ageing and AD-like brain atrophy compared to their male counterparts with normal-weight. Furthermore, proteomic investigations of two time-points revealed five proteins that are associated positively with both brain ageing and BMI, and eight proteins associated negatively with brain ageing and BMI.

The Lancet Commission estimated that 45% of global dementia cases could be attributed to 14 factors across three life stages: early life, midlife, and late life. Notable were cardiometabolic factors such as high LDL-C, diabetes, hypertension, and obesity primarily affecting midlife adults (12). Previous research has yielded conflicting results regarding the association between BMI and dementia risk, showing that midlife obesity displays a higher risk for dementia in a 27-year longitudinal population cohort (33), a result supported by a recent meta-analysis (34). On the contrary, a large UK cohort study of 2 million participants showed that dementia risk is lower in obesity and falls for every increasing BMI group in a median follow-up of 9.1 years (35). Shedding light on the existing contradictory evidence about the association of BMI with dementia, a US cohort of 1.3 million suggested that a higher BMI measured more than 20 years before dementia diagnosis is associated with higher risk, while a higher BMI measured within 10 years of dementia diagnosis is associated with a lower risk (36). This suggests that high BMI is directly linked with an increased risk for dementia and that the reduced

risk for dementia in obesity may be a result of reverse causation, influenced by the weight loss in the preclinical phase of dementia (35), which usually precedes 20-30 years before its clinical manifestation (37). Notably, BMI typically declines several years before a mild cognitive impairment diagnosis (17), and continues to decrease at a similar rate in both individuals who progress to dementia and those who do not (17). In alignment with the Lancet's Commission's findings and our own research using imaging and ML-derived brain ageing indices, obesity both in midlife and in late life, appears to contribute to the acceleration of brain ageing in individuals without diagnosed cognitive impairment (12, 26), displaying inferior performance in cognitive tests in comparison with individuals with normal-weight, and we showed that the effects of obesity on brain ageing, and AD-like brain atrophy are weaker with increasing age, with significant sex differences. Furthermore, weight loss in participants with obesity has been shown to decrease the risk for AD (38). Additionally, a combination of modest caloric restriction and the Mediterranean-Dietary Approaches to Stop Hypertension (DASH) diet intervention for neurodegenerative delay (MIND) diet in individuals with obesity showed improved cognitive outcomes, suggesting a potential reversal of the negative effects of obesity on the brain and cognition (39, 40). Recent studies have also highlighted the potential protective effects of weight loss medications like glucagon-like peptide 1 (GLP-1) analogues in reducing dementia incidence (41, 42).

Sex differences impact dementia risk, but their specific way of influencing it remains elusive. While men have higher rates of alcohol use, more years of education, and a stronger association between APOE4 carriage and dementia risk compared to women (43), the prevalence of Alzheimer's disease and related dementias (ADRDs) is higher in women. However, the population-attributable risk for dementia is higher in men (44). A US cohort study suggested that men have a higher midlife dementia risk, potentially due to women having fewer cardiovascular risk factors, such as coronary artery disease, myocardial infarction, and heart failure, which may mitigate their dementia risk (45). Our findings align with these observations, showing that patterns of brain ageing and AD-like brain atrophy are more prevalent in males than in females with obesity or overweight, but females with normal-weight displayed more brain ageing and AD-like brain atrophy compared to males with normal weight. This finding may suggest that weight gain influences brain ageing trajectories differently in males and females, with males exhibiting greater brain ageing and AD-like atrophy at higher BMI levels, while among participants with normal weight, females show greater brain ageing and AD-like atrophy.

So far, it remains unclear whether obesity and age synergistically or additively exacerbate structural brain ageing (46). While some studies suggest a borderline causal significance of BMI on the white matter-brain age gap (47), and a 10-year

brain age gap both in middle-aged individuals with obesity or overweight (48), other mechanisms likely contribute more significantly to cognitive decline in obesity. Specifically, metabolic syndrome, characterised by insulin resistance, is increasingly implicated in the development and progression of AD (49). Specifically, it is postulated that insulin resistance comprises the key link between metabolic syndrome and AD due to the several roles that brain insulin signalling plays, i.e., the regulation of food intake, body weight, reproduction, and learning and memory (50, 51). For instance, brain insulin resistance, referred to also as central insulin resistance, can manifest as impaired central and peripheral regulation of nutrient partitioning, cognitive and mood dysfunction, and brain-specific neuropathology and neurodegeneration (52), through an increase in neuronal vulnerability, the accumulation of amyloid-β plaques, tau phosphorylation, neurofibrillary lesions, α-synuclein lesions although evidence is yet inconsistent (52, 53). In our study, participants with obesity or overweight did not show any difference in total tau, p-tau, or amyloid-β measurements in their CSF compared to participants with normal-weight, except that amyloid-β appeared to be elevated in individuals with overweight compared to controls with normal-weight. Additionally, another potential pathophysiological explanation for the increased brain ageing in these states derives from the action of proinflammatory adipokines and cytokines that trigger, in addition to insulin resistance, an inflammatory response in microglia, inducing a self-sustaining feedback loop of additional cytokines and inflammation-associated with white matter changes (49, 54-60).

Our proteomic analysis identified GDF15 as significantly associated with AD-like brain atrophy in one of the two visits. GDF15, a cytokine in the transforming growth factor β superfamily, is linked to decreased brain volume in Alzheimer's-affected regions (61). Elevated in inflammatory states, GDF15 poses a high risk for AD-like brain atrophy patterns (62, 63) and influences food intake and energy balance, acting as an anorectic peptide (64). High serum levels correlate with poor neurocognitive performance and increased dementia risk, and plasma GDF15 levels have also shown a significant linear correlation with BMI (R^2=0.097; p=0.0049) (65, 66). GDF15 is upregulated in neurodegenerative disorders, suggesting its potential as a diagnostic marker (67) and predicting dementia up to 25 years before onset (18). In the brain, higher ectodysplasin A2 receptor (EDA2R) levels correlate with reduced total brain volume (68-70).

Additionally, we demonstrated that only five proteins were significantly associated with an increase both in brain ageing and in BMI per 1-SD in both proteomic investigations, i.e., LY96, GDF15, PIGR, CHI3L1, and OXT. LY96, a protein playing a crucial role in the innate immune response against bacterial lipopolysaccharides, age-related immune alterations and increased infiltration of peripheral immune cells like T cells and B cells into the aged brain were previously reported, however,

to the best of our knowledge, no study has specifically reported LY96 levels in brain ageing (71). GDF15 was shown herein to be associated with brain ageing, which aligns with previous research (72). Evidence of PIGR regarding neurodegeneration and ageing is lacking. Similarly, CHI3L1 is an inflammatory marker increased in obesity and diabetes and related to insulin resistance; it has been frequently investigated in body fluids as a surrogate marker of neuroinflammation in AD and other neurological disorders, with our results demonstrating an association with brain ageing of nine months (73, 74). Additionally, oxytocin, although it physiologically reduces food intake and increases energy expenditure, is found to be increased in plasma in obesity (75). With ageing, substantial changes emerge in the oxytocinergic system and neural circuitry mediating oxytocin's effects; with previous literature suggesting OXT's associations with age-related structural brain changes, herein it is demonstrated for the first time that OXT is associated with six months of brain ageing (76, 77).

In contrast, eight proteins, ITGA11, CD34, CD200, SEZL6, BCAN, CNTN5, NCAN, and KITLG, were identified as linked with a decrease in the brain age and BMI per 1-SD. Specifically, ITGA11 has been shown to be significantly associated with decreased odds of incident cognitive impairment (78). Additionally, studies have shown that in patients with AD, there is a marked decrease in CD200 and CD200 receptor expression, which may contribute to a pro-inflammatory state in human neurodegenerative diseases (79). Another study found significantly reduced counts of circulating CD34+ cells in early AD, which significantly correlated with age ($r = -0.66$; $p = 0.001$), cerebrospinal fluid β-amyloid (Aβ)1–42 ($r = -0.47$; $p = 0.025$) and most pronounced the Aβ42/40 ratio ($r = -0.69$; $p = 0.005$) (80). Besides, the rs1461684 G variant of the *CNTN5* gene has been associated with an increased risk of sporadic AD and faster disease progression in individuals without cognitive impairments (81). KITLG was found to be differently expressed between controls and participants with mild cognitive impairment (82). SEZL6, BCAN, and FGFBP1 have not been observed to decrease the risk of neurocognitive disorders. The mediation analysis of the effect of the significant proteins on the relationship between BMI and brain ageing suggested that these proteins have a significant effect separately, but the influence of BMI on brain ageing may occur primarily through other pathways.

This study has several strengths, including the large, diverse, multisite sample covering a wide age range and the use of advanced harmonisation methods. However, this study also presents limitations. First, the associations between the brain atrophy index (SPARE-AD) and BMI groups reflect brain atrophy patterns observed in AD dementia and not necessarily other dementia subtypes. Second, although the UK Biobank provides a comprehensive assessment of circulating proteins, not all of the human proteome is captured within this platform, and

biases may be present in the priority of measuring secreted proteins. Third, there was no external validation dataset from another country to replicate our findings, which could impact the results' generalisability. Fourth, the lack of long-term follow-up prevents the derivation of robust conclusions regarding the clinical progression. Furthermore, no causal relationship can be inferred from the proteomic associations, warranting additional genomic studies and the elucidation of potential sex-specific mechanisms. Lastly, the mediation analysis of the role of each protein in the relationship between BMI and brain ageing does not address whether other conditions are implicated in this relationship, such as atherosclerotic disease, vascular infarcts, and other cerebrovascular abnormalities related to brain ageing.

In this multi-cohort study, we showed that higher BMI states are associated with faster brain ageing and increased AD-like brain atrophy, particularly in males, while females with normal-weight showed more pronounced brain ageing and AD-like brain atrophy than males with normal-weight males. Second, proteomic analysis revealed 13 proteins postulated to be implicated in the pathophysiological mechanism of brain ageing and AD-like brain atrophy.

Contributors
FA and CD carried out the acquisition, analysis, or interpretation of data. FA and MK wrote the first draft. All authors read and critically revised the manuscript for important intellectual content. All authors approved the final version of the manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no authors meeting the criteria have been omitted.

Data sharing statement
Imaging and clinical data used in this study were provided by several individual studies via data-sharing agreements, which do not include permission for us to further share the data. Investigators must apply to the source data providers to access additional data and match their subject IDs to those used in this study under the current protocol (primarily for UKBB). Data from ADNI are available from the ADNI database (adni.loni.usc.edu) upon registration and compliance with the data usage agreement. Data from the UKBB are available upon request from the UKBB website (https://www.ukbiobank.ac.uk/). Data from the BLSA study are available upon request at https://www.blsa.nih.gov/how-apply. Data from the AIBL study are available upon request at https://aibl.org.au/. Data from the OASIS study are available upon request at https://www.oasis-brains.org/. Data requests for BIOCARD, HANDLS, PENN, WRAP, CARDIA, PreventAD, SHIP, and WHIMS datasets should be directed to the corresponding author (CD), who will direct them to the appropriate PIs. The R code can be accessed in the following link: https://github.com/FilipposAna/Radiomic_signatures.

Declaration of Interests
IMN is an expert reader in Clariti, has received payments from Subtle Medical, Inc for consulting, has served in the advisory board in Eisai and Premier Inc, and has done educational speaking for Peerview. IMN and DT declare that the University of Pennsylvania receives NIH grants RF1AG054409, U24NS130411. DT receives consulting fees from Roche, Veravas and Alzheon Imaging, and honoraria from the University of Pennsylvania. KW serves in the board of directors in the National Academy of Neuropsychology. All other authors declare no conflicts of interest.

Acknowledgements
The iSTAGING study is a multi-institutional effort funded by the National Institute on Aging (NIA) by RF1 AG054409 (C. Davatzikos). MK receives funding support from an MD-PhD grant from the University Medical Center Groningen. KAW is supported by the National Institute on Aging's Intramural Research Program by AG000348-01, and AG000349-01. HS is supported by RF1-AG054409, R01-MH123550, and U24-NS130411 grants. This study was supported, in part, by NIH grant RF1AG054409, U24NS130411 awarded to CD. This research has been conducted using the UK Biobank Resource under Application Number 35148.

Supplementary data

This article contains supporting information available online at:
https://drive.google.com/file/d/1ZT-Q5FKavR44bmhNJkIhn-jxnakO1vXU/view?usp=sharing

References

1. Kokkorakis M, Katsarou A, Katsiki N, Mantzoros CS. Milestones in the journey towards addressing obesity; Past trials and triumphs, recent breakthroughs, and an exciting future in the era of emerging effective medical therapies and integration of effective medical therapies with metabolic surgery. Metabolism. 2023;148:155689.
2. Kokkorakis M, Chakhtoura M, Rhayem C, Al Rifai J, Ghezzawi M, Valenzuela-Vallejo L, et al. Emerging pharmacotherapies for obesity: A systematic review. Pharmacological Reviews. 2025;77(1):100002.
3. Global, regional, and national prevalence of child and adolescent overweight and obesity, 1990-2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. Lancet. 2025;405(10481):785-812.
4. Global, regional, and national prevalence of adult overweight and obesity, 1990-2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. Lancet. 2025;405(10481):813-38.
5. Stefanakis K, Kokkorakis M, Mantzoros CS. The impact of weight loss on fat-free mass, muscle, bone and hematopoiesis health: Implications for emerging pharmacotherapies aiming at fat reduction and lean mass preservation. Metabolism. 2024;161:156057.
6. Wu Z, Owen A, Woods RL, Cribb L, Alharbi T, Zhou Z, et al. Associations of body habitus and its changes with incident dementia in older adults. J Am Geriatr Soc. 2024;72(4):1023-34.
7. Ma Y, Ajnakina O, Steptoe A, Cadar D. Higher risk of dementia in English older individuals who are overweight or obese. Int J Epidemiol. 2020;49(4):1353-65.
8. Slomski A. Obesity Is Now the Top Modifiable Dementia Risk Factor in the US. JAMA. 2022;328(1):10-.
9. Worldwide trends in underweight and obesity from 1990 to 2022: a pooled analysis of 3663 population-representative studies with 222 million children, adolescents, and adults. Lancet. 2024;403(10431):1027-50.
10. Hurd MD, Martorell P, Delavande A, Mullen KJ, Langa KM. Monetary costs of dementia in the United States. N Engl J Med. 2013;368(14):1326-34.
11. Kokkorakis M, Muzurović E, Volčanšek Š, Chakhtoura M, Hill MA, Mikhailidis DP, et al. Steatotic Liver Disease: Pathophysiology and Emerging Pharmacotherapies. Pharmacol Rev. 2024;76(3):454-99.
12. Livingston G, Huntley J, Liu KY, Costafreda SG, Selbæk G, Alladi S, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. Lancet. 2024.
13. Li G, Hu Y, Zhang W, Wang J, Ji W, Manza P, et al. Brain functional and structural magnetic resonance imaging of obesity and weight loss interventions. Mol Psychiatry. 2023;28(4):1466-79.
14. Singh-Manoux A, Dugravot A, Shipley M, Brunner EJ, Elbaz A, Sabia S, et al. Obesity trajectories and risk of dementia: 28 years of follow-up in the Whitehall II Study. Alzheimers Dement. 2018;14(2):178-86.
15. Fitzpatrick AL, Kuller LH, Lopez OL, Diehr P, O'Meara ES, Longstreth WT, Jr., et al. Midlife and late-life obesity and the risk of dementia: cardiovascular health study. Arch Neurol. 2009;66(3):336-42.
16. Li J, Liu C, Ang TFA, Au R. BMI decline patterns and relation to dementia risk across four decades of follow-up in the Framingham Study. Alzheimers Dement. 2023;19(6):2520-7.

17. Guo J, Wang J, Dove A, Chen H, Yuan C, Bennett DA, et al. Body Mass Index Trajectories Preceding Incident Mild Cognitive Impairment and Dementia. JAMA Psychiatry. 2022;79(12):1180-7.
18. You J, Guo Y, Zhang Y, Kang JJ, Wang LB, Feng JF, et al. Plasma proteomic profiles predict individual future health risk. Nat Commun. 2023;14(1):7817.
19. Kaufmann T, van der Meer D, Doan NT, Schwarz E, Lund MJ, Agartz I, et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. Nat Neurosci. 2019;22(10):1617-23.
20. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser. 2000;894:i-xii, 1-253.
21. Habes M, Pomponio R, Shou H, Doshi J, Mamourian E, Erus G, et al. The Brain Chart of Aging: Machine-learning analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition in the iSTAGING consortium of 10,216 harmonized MR scans. Alzheimers Dement. 2021;17(1):89-102.
22. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29(6):1310-20.
23. Doshi J, Erus G, Ou Y, Gaonkar B, Davatzikos C. Multi-atlas skull-stripping. Acad Radiol. 2013;20(12):1566-76.
24. Doshi J, Erus G, Ou Y, Resnick SM, Gur RC, Gur RE, et al. MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. Neuroimage. 2016;127:186-95.
25. Pomponio R, Erus G, Habes M, Doshi J, Srinivasan D, Mamourian E, et al. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. Neuroimage. 2020;208:116450.
26. Skampardoni I, Nasrallah IM, Abdulkadir A, Wen J, Melhem R, Mamourian E, et al. Genetic and Clinical Correlates of AI-Based Brain Aging Patterns in Cognitively Unimpaired Individuals. JAMA Psychiatry. 2024;81(5):456-67.
27. Habes M, Janowitz D, Erus G, Toledo JB, Resnick SM, Doshi J, et al. Advanced brain aging: relationship with epidemiologic and genetic risk factors, and overlap with Alzheimer disease atrophy patterns. Transl Psychiatry. 2016;6(4):e775.
28. Hwang G, Abdulkadir A, Erus G, Habes M, Pomponio R, Shou H, et al. Disentangling Alzheimer's disease neurodegeneration from typical brain ageing using machine learning. Brain Commun. 2022;4(3):fcac117.
29. Doshi J, Erus G, Habes M, Davatzikos C. DeepMRSeg: A convolutional deep neural network for anatomy and abnormality segmentation on MR images2019 July 01, 2019:[arXiv:1907.02110 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2019arXiv190702110D.
30. Smith SM, Vidaurre D, Alfaro-Almagro F, Nichols TE, Miller KL. Estimation of brain age delta from brain imaging. Neuroimage. 2019;200:528-39.
31. Davatzikos C, Xu F, An Y, Fan Y, Resnick SM. Longitudinal progression of Alzheimer's-like patterns of atrophy in normal older adults: the SPARE-AD index. Brain. 2009;132(Pt 8):2026-35.
32. Shaw LM, Vanderstichele H, Knapik-Czajka M, Figurski M, Coart E, Blennow K, et al. Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. Acta Neuropathol. 2011;121(5):597-609.
33. Whitmer RA, Gunderson EP, Barrett-Connor E, Quesenberry CP, Jr., Yaffe K. Obesity in middle age and future risk of dementia: a 27 year longitudinal population based study. BMJ. 2005;330(7504):1360.

34. Pedditzi E, Peters R, Beckett N. The risk of overweight/obesity in mid-life and late life for the development of dementia: a systematic review and meta-analysis of longitudinal studies. Age Ageing. 2016;45(1):14-21.
35. Qizilbash N, Gregson J, Johnson ME, Pearce N, Douglas I, Wing K, et al. BMI and risk of dementia in two million people over two decades: a retrospective cohort study. Lancet Diabetes Endocrinol. 2015;3(6):431-6.
36. Kivimäki M, Luukkonen R, Batty GD, Ferrie JE, Pentti J, Nyberg ST, et al. Body mass index and risk of dementia: Analysis of individual-level data from 1.3 million individuals. Alzheimers Dement. 2018;14(5):601-9.
37. Jansen WJ, Ossenkoppele R, Knol DL, Tijms BM, Scheltens P, Verhey FR, et al. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. JAMA. 2015;313(19):1924-38.
38. Kim KY, Ha J, Lee JY, Kim E. Weight loss and risk of dementia in individuals with versus without obesity. Alzheimers Dement. 2023;19(12):5471-81.
39. Arjmand G, Abbas-Zadeh M, Eftekhari MH. Effect of MIND diet intervention on cognitive performance and brain structure in healthy obese women: a randomized controlled trial. Sci Rep. 2022;12(1):2871.
40. Sawyer RP, Blair J, Shatz R, Manly JJ, Judd SE. Association of Adherence to a MIND-Style Diet With the Risk of Cognitive Impairment and Decline in the REGARDS Cohort. Neurology. 2024;103(8):e209817.
41. Tang B, Sjölander A, Wastesson JW, Maura G, Blotiere PO, Szilcz M, et al. Comparative effectiveness of glucagon-like peptide-1 agonists, dipeptidyl peptidase-4 inhibitors, and sulfonylureas on the risk of dementia in older individuals with type 2 diabetes in Sweden: an emulated trial study. EClinicalMedicine. 2024;73:102689.
42. Wium-Andersen IK, Osler M, Jørgensen MB, Rungby J, Wium-Andersen MK. Antidiabetic medication and risk of dementia in patients with type 2 diabetes: a nested case-control study. Eur J Endocrinol. 2019;181(5):499-507.
43. Gong J, Harris K, Lipnicki DM, Castro-Costa E, Lima-Costa MF, Diniz BS, et al. Sex differences in dementia risk and risk factors: Individual-participant data analysis using 21 cohorts across six continents from the COSMIC consortium. Alzheimers Dement. 2023;19(8):3365-78.
44. Nianogo RA, Rosenwohl-Mack A, Yaffe K, Carrasco A, Hoffmann CM, Barnes DE. Risk Factors Associated With Alzheimer Disease and Related Dementias by Sex and Race and Ethnicity in the US. JAMA Neurol. 2022;79(6):584-91.
45. Brady B, Zheng L, Kootar S, Anstey KJ. Sex and gender differences in risk scores for dementia and Alzheimer's disease among cisgender, transgender, and non-binary adults. Alzheimers Dement. 2024;20(1):5-15.
46. Bischof GN, Park DC. Obesity and Aging: Consequences for Cognition, Brain Structure, and Brain Function. Psychosom Med. 2015;77(6):697-709.
47. Wen J, Zhao B, Yang Z, Erus G, Skampardoni I, Mamourian E, et al. The genetic architecture of multimodal human brain age. Nat Commun. 2024;15(1):2604.
48. Ronan L, Alexander-Bloch AF, Wagstyl K, Farooqi S, Brayne C, Tyler LK, et al. Obesity associated with increased brain age from midlife. Neurobiol Aging. 2016;47:63-70.
49. Kim B, Feldman EL. Insulin resistance as a key link for the increased risk of cognitive impairment in the metabolic syndrome. Experimental & Molecular Medicine. 2015;47(3):e149-e.
50. van der Heide LP, Ramakers GM, Smidt MP. Insulin signaling in the central nervous system: learning to survive. Prog Neurobiol. 2006;79(4):205-21.

51. Kullmann S, Heni M, Hallschmid M, Fritsche A, Preissl H, Häring HU. Brain Insulin Resistance at the Crossroads of Metabolic and Cognitive Disorders in Humans. Physiol Rev. 2016;96(4):1169-209.
52. Arnold SE, Arvanitakis Z, Macauley-Rambach SL, Koenig AM, Wang HY, Ahima RS, et al. Brain insulin resistance in type 2 diabetes and Alzheimer disease: concepts and conundrums. Nat Rev Neurol. 2018;14(3):168-81.
53. Kellar D, Craft S. Brain insulin resistance in Alzheimer's disease and related disorders: mechanisms and therapeutic approaches. Lancet Neurol. 2020;19(9):758-66.
54. Arnoldussen IA, Kiliaan AJ, Gustafson DR. Obesity and dementia: adipokines interact with the brain. Eur Neuropsychopharmacol. 2014;24(12):1982-99.
55. Griffin WS. Inflammation and neurodegenerative diseases. Am J Clin Nutr. 2006;83(2):470s-4s.
56. Wilson CJ, Finch CE, Cohen HJ. Cytokines and cognition--the case for a head-to-toe inflammatory paradigm. J Am Geriatr Soc. 2002;50(12):2041-56.
57. Wisse BE. The inflammatory syndrome: the role of adipose tissue cytokines in metabolic disorders linked to obesity. J Am Soc Nephrol. 2004;15(11):2792-800.
58. Bolzenius JD, Laidlaw DH, Cabeen RP, Conturo TE, McMichael AR, Lane EM, et al. Impact of body mass index on neuronal fiber bundle lengths among healthy older adults. Brain Imaging Behav. 2013;7(3):300-6.
59. Kullmann S, Schweizer F, Veit R, Fritsche A, Preissl H. Compromised white matter integrity in obesity. Obes Rev. 2015;16(4):273-81.
60. Anagnostakis F, Kokkorakis M, Walker KA, Davatzikos C. Signatures and Discriminative Abilities of Multi-Omics between States of Cognitive Decline. Biomedicines. 2024;12(5).
61. Walker KA, Chen J, Zhang J, Fornage M, Yang Y, Zhou L, et al. Large-scale plasma proteomic analysis identifies proteins and pathways associated with dementia risk. Nat Aging. 2021;1(5):473-89.
62. May BM, Pimentel M, Zimerman LI, Rohde LE. GDF-15 as a Biomarker in Cardiovascular Disease. Arq Bras Cardiol. 2021;116(3):494-500.
63. Chrysafi P, Valenzuela-Vallejo L, Stefanakis K, Kelesidis T, Connelly MA, Mantzoros CS. Total and H-specific GDF-15 levels increase in caloric deprivation independently of leptin in humans. Nat Commun. 2024;15(1):5190.
64. Lockhart SM, Saudek V, O'Rahilly S. GDF15: A Hormone Conveying Somatic Distress to the Brain. Endocr Rev. 2020;41(4).
65. McGrath ER, Himali JJ, Levy D, Conner SC, DeCarli C, Pase MP, et al. Growth Differentiation Factor 15 and NT-proBNP as Blood-Based Markers of Vascular Brain Injury and Dementia. J Am Heart Assoc. 2020;9(19):e014659.
66. Sarkar S, Legere S, Haidl I, Marshall J, MacLeod JB, Aguiar C, et al. Serum GDF15, a Promising Biomarker in Obese Patients Undergoing Heart Surgery. Front Cardiovasc Med. 2020;7:103.
67. Xue XH, Tao LL, Su DQ, Guo CJ, Liu H. Diagnostic utility of GDF15 in neurodegenerative diseases: A systematic review and meta-analysis. Brain Behav. 2022;12(2):e2502.
68. Bilgic SN, Domaniku A, Toledo B, Agca S, Weber BZC, Arabaci DH, et al. EDA2R-NIK signalling promotes muscle atrophy linked to cancer cachexia. Nature. 2023;617(7962):827-34.
69. Harris SE, Cox SR, Bell S, Marioni RE, Prins BP, Pattie A, et al. Neurology-related protein biomarkers are associated with cognitive ability and brain volume in older age. Nat Commun. 2020;11(1):800.
70. Coffey CE, Ratcliff G, Saxton JA, Bryan RN, Fried LP, Lucke JF. Cognitive correlates of human brain aging: a quantitative magnetic resonance imaging investigation. J Neuropsychiatry Clin Neurosci. 2001;13(4):471-85.

71. Finger CE, Moreno-Gonzalez I, Gutierrez A, Moruno-Manchon JF, McCullough LD. Age-related immune alterations and cerebrovascular inflammation. Mol Psychiatry. 2022;27(2):803-18.
72. Casanova R, Walker KA, Justice JN, Anderson A, Duggan MR, Cordon J, et al. Associations of plasma proteomics and age-related outcomes with brain age in a diverse cohort. Geroscience. 2024;46(4):3861-73.
73. Zhao T, Su Z, Li Y, Zhang X, You Q. Chitinase-3 like-protein-1 function and its role in diseases. Signal Transduct Target Ther. 2020;5(1):201.
74. Sanfilippo C, Castrogiovanni P, Imbesi R, Kazakowa M, Musumeci G, Blennow K, et al. Sex difference in CHI3L1 expression levels in human brain aging and in Alzheimer's disease. Brain Res. 2019;1720:146305.
75. Lawson EA. The effects of oxytocin on eating behaviour and metabolism in humans. Nat Rev Endocrinol. 2017;13(12):700-9.
76. Huffmeijer R, van Ijzendoorn MH, Bakermans-Kranenburg MJ. Ageing and oxytocin: a call for extending human oxytocin research to ageing populations--a mini-review. Gerontology. 2013;59(1):32-9.
77. Freeman SM, Palumbo MC, Lawrence RH, Smith AL, Goodman MM, Bales KL. Effect of age and autism spectrum disorder on oxytocin receptor density in the human basal forebrain and midbrain. Transl Psychiatry. 2018;8(1):257.
78. Duggan MR, Butler L, Peng Z, Daya GN, Moghekar A, An Y, et al. Plasma proteins related to inflammatory diet predict future cognitive impairment. Mol Psychiatry. 2023;28(4):1599-609.
79. Walker DG, Dalsing-Hernandez JE, Campbell NA, Lue LF. Decreased expression of CD200 and CD200 receptor in Alzheimer's disease: a potential mechanism leading to chronic inflammation. Exp Neurol. 2009;215(1):5-19.
80. Maler JM, Spitzer P, Lewczuk P, Kornhuber J, Herrmann M, Wiltfang J. Decreased circulating CD34+ stem cells in early Alzheimer's disease: Evidence for a deficient hematopoietic brain support? Mol Psychiatry. 2006;11(12):1113-5.
81. Dauar MT, Labonté A, Picard C, Miron J, Rosa-Neto P, Zetterberg H, et al. Characterization of the contactin 5 protein and its risk-associated polymorphic variant throughout the Alzheimer's disease spectrum. Alzheimers Dement. 2023;19(7):2816-30.
82. Oh SL, Zhou M, Chin EWM, Amarnath G, Cheah CH, Ng KP, et al. Alzheimer's Disease Blood Biomarkers Associated With Neuroinflammation as Therapeutic Targets for Early Personalized Intervention. Front Digit Health. 2022;4:875895.


THEMIS

Part III

Real-World Evidence and Comparative Effectiveness of Cardiometabolic Drugs in Neurocognitive Outcomes

Chapter 6

Comparative effectiveness of sodium-glucose cotransporter-2 inhibitors and glucagon-like peptide-1 receptor agonists for incident dementia risk: a retrospective multicohort study

Michail Kokkorakis, Filippos Anagnostakis, Shrihari Nagarajan, Georgia Anastasiou, Christos S. Mantzoros

Submitted Manuscript

Summary

Objective This study provides comprehensive real-world evidence on the comparative effectiveness of glucagon-like peptide-1 receptor agonists (GLP-1 RAs) and sodium-glucose cotransporter-2 inhibitors (SGLT2i) on the risk of incident dementia in adults with type 2 diabetes mellitus (T2DM).

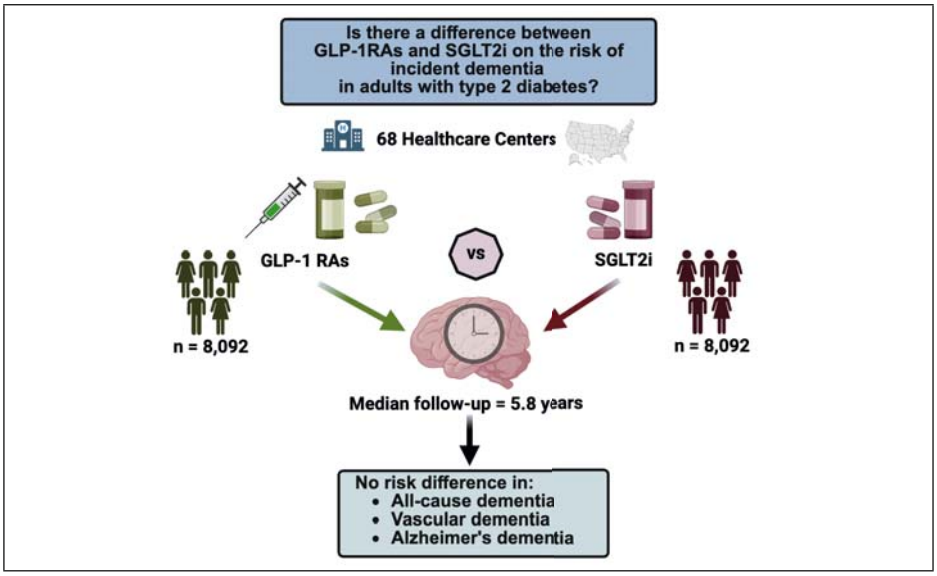
Research design and methods This cohort study utilised electronic health records from TriNetX (April 1, 2013, to December 31, 2019). Adults with T2DM on metformin who initiated GLP-1 RAs or SGLT2i were analysed using Cox proportional hazards models in 1:1 propensity score-matched cohorts. The primary outcome was all-cause dementia, with secondary outcomes including vascular and Alzheimer's dementia. Confounding variables were adjusted.

Results Among 25,814 participants (GLP-1 RAs: 16,951; SGLT2i: 8,863) over a median 5.8-year follow-up, 271 GLP-1 RA users and 238 SGLT2i users developed all-cause dementia. No significant differences were observed in all-cause dementia (HR: 1.12, 95% CI: 0.94–1.34), vascular (HR: 1.12, 95% CI: 0.75–1.67), or Alzheimer's dementia (HR: 1.15, 95% CI: 0.80–1.68). Subgroup analyses revealed increased dementia risk in female GLP-1 RA users when identifying dementia cases by dementia medication use.

Conclusions In this cohort study, the use of SGLT2i was not associated with an increased risk of dementia compared with the use of GLP-1RAs in patients with type 2 diabetes. Further research is warranted to explore the effects of newer incretin-based co-agonists on cognitive outcomes.

Funding This work did not receive funding.

Keywords Dementia, EHR, GLP-1 RA, SGLT-2i, incidence, real world.



Highlights

Why did we undertake this study?
GLP-1RAs and SGLT2i are the most promising antihyperglycemic agents to reduce dementia risk, and so far, there's no large comparative effectiveness study conducted between these two medications.

What is the specific question(s) we wanted to answer?
In this study, we aimed to determine whether there is any difference between GLP-1RAs and SGLT2i on the risk of incident dementia in adults with type 2 diabetes.

What did we find?
GLP-1 RAs did not exhibit any significant differences compared to SGLT2i in all-cause dementia, vascular dementia, and Alzheimer's dementia risk. Female users of GLP-1 RAs may have a higher risk for incident dementia compared to female SGLT2i users.

What are the implications of our findings?
With the imminent inclusion of neurocognitive outcomes in the type 2 diabetes guidelines, we do not suggest that there is any overall difference in the risk of dementia between GLP-1RAs and SGLT2i. However, sex-specific differences warrant further investigation.

Introduction

The rising prevalence of type 2 diabetes mellitus (T2DM) alongside the aging of the population globally poses a significant public health challenge. A systematic analysis for the Global Burden of Disease Study 2021 predicted an increase in T2DM cases by more than 60% in 2050 (1), intricately linked to cognitive decline and dementia. Notably, The Lancet Commission on Dementia suggested up to 40% of dementia cases may be preventable, with T2DM being a major modifiable risk factor, increasing dementia risk by 1.6 times (2). These findings emphasize the imperative need for comprehensive healthcare strategies integrating diabetes management and dementia prevention to mitigate the burgeoning societal burden posed by these interrelated conditions. Given the vast global population affected by T2DM, the considerably high risk of dementia among these individuals, and the potential of anti-diabetic medications to mitigate that risk, implementing a tailored pharmacological approach for older adults would have a profound impact on dementia prevention.

Although studies such as the CARMELINA phase 4 RCT focusing on linagliptin found that DPP4 inhibitors do not affect cognitive decline compared to placebo or sulfonylureas in patients with T2DM (3, 4), several reports based on randomized controlled trials (RCTs) and observational studies indicate a potential link between second-line antihyperglycemic medications and dementia (5). An exploratory analysis from the cardiovascular outcome phase 3 RCT, named “REWIND”, revealed that dulaglutide, a glucagon-like peptide-1 receptor agonist (GLP-1 RA), may reduce cognitive decline in individuals with T2DM (6). A recent large national cohort study indicated that GLP-1 RAs are associated with a lower risk of incident dementia compared with dipeptidyl peptidase-4 (DPP4) inhibitors and sulfonylureas (7) but did not include individuals using sodium-glucose transport cotransporter-2 inhibitors (SGLT2i), one of the most commonly used second-line antihyperglycemic drugs, which were recently shown to be the most effective when compared to DPP4 inhibitors and sulfonylureas in lowering, among other outcomes, mean HbA1c, body mass index, and systolic blood pressure (8). Thus, the critical question of whether GLP-1 RAs may have a differential effect on dementia when compared to SGLT2i remains to be addressed.

Despite the above evidence that some secondary antihyperglycemic drugs may have a protective role against dementia, the Standards of Care in Diabetes, released this year by the American Diabetes Association (ADA) Professional Practice Committee, do not specify pharmacological adjustments for individuals at risk of developing dementia. Nonetheless, the ADA recommends annual cognitive screening for adults 65 and older to facilitate early detection of mild cognitive impairment or dementia (9). Therefore, given the varying effects and risks

of cognitive dysfunction associated with different classes of antihyperglycemic drugs, there is a critical need for tailored and individualised pharmacological treatments. Specifically, T2DM or, by extension, cardiovascular-kidney-metabolic syndrome management could be adjusted according to not only disease-specific and overall mortality but also morbidity, including the cognitive status of older screened adults as cognitive status could become an axis of antihyperglycemic treatment selection, an area where clinical guidance on preventative strategies for dementia is presently limited.

To our knowledge, no comparative evaluation has been conducted between the newest and most promising second-line antihyperglycemic drug classes, SGLT2i and GLP-1 RAs, on the risk of incident dementia. To bridge this knowledge gap, we aimed to explore the effectiveness of the GLP-1 RAs and SGLT2i as second-line antihyperglycemic medications on top of metformin in a real-world setting in relation to the risk of all-cause dementia, vascular dementia, and Alzheimer's dementia.

Research design and methods

Data source

The study was conducted using data from the Research Network of the TriNetX Analytics platform. TriNetX is a platform that de-identifies and aggregates electronic health records (EHR) data, including more than 300 million patients from more than 120 healthcare organizations in countries in North and South America, Europe, the Middle East, Africa, and Asia-Pacific, including Japan.

The data analysis was performed between October 2024 and November 2024. As a federated network, research studies using the TriNetX research network do not require ethics approvals, as no patient identifiable information is received. Informed consent was waived because the study used de-identified secondary data. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

Study population

This new-user, active comparator cohort study (10-12) used data from commercially insured adults (aged ≥18 years) with T2DM who initiated treatment with SGLT2i or GLP-1 RAs between April 1st, 2013, and December 31st, 2019. Participants needed to take metformin within 6 months of their GLP-1 RA or SGLT2i prescription. Patients were included if they had undergone a healthcare visit within one year before their cohort entry. Patients were excluded if they had any dementia diagnosis at any time before cohort entry, any contraindication of GLP-1 RA prescription, or past GLP-1 RA or SGLT2i (Figure 1) (13, 14). We conducted a retrospective, multicohort

comparative effectiveness study of antihyperglycemic drugs using data and built-in analytic functions on the TriNetX Analytics platform (described in detail in Supplementary Tables 1–4).

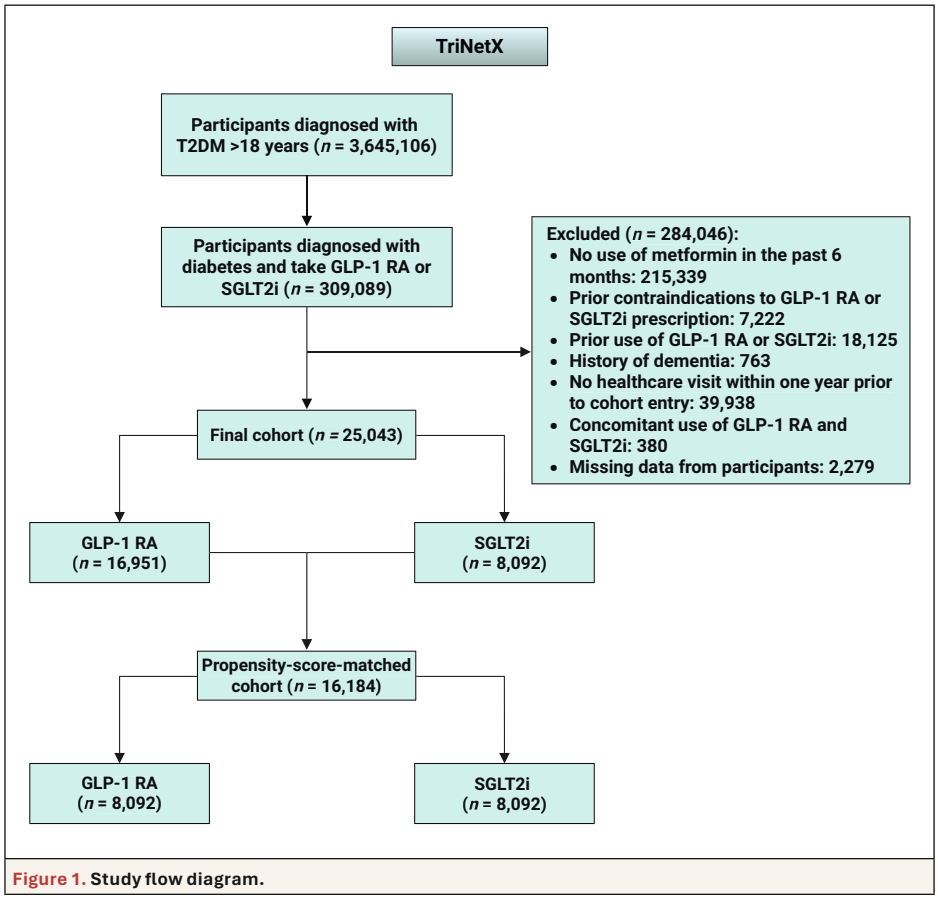


Figure 1. Study flow diagram.

Study outcomes

The primary outcome was dementia (International Classification of Diseases, 10th Revision, codes F01, F02, F03, G30, G31.0, G31.83). In additional analyses, we investigated dementia subtypes, including vascular dementia and Alzheimer's dementia. We also used an alternative outcome definition to attempt to increase specificity by defining dementia incidence based on the initiation of anticholinergic medications (donepezil, galantamine, rivastigmine, and memantine). Study outcomes are described in Supplementary Table 3.

Covariates

We measured potential confounders up to 1 year before the cohort entry date (Supplementary Table 4). Based on subject matter expertise and previous research on outcomes related to drug use in individuals with T2DMs, we identified variables that acted as confounders, proxies for confounders, or predictors of the outcome. The following variables were included in propensity score matching: age, gender, race, and from diagnoses: type 2 diabetes complications (neurological, kidney, and ophthalmic), overweight and obesity, hypothyroidism, essential hypertension, heart failure, atrial fibrillation and flutter, ischemic heart diseases, venous embolism and thrombosis, cerebrovascular diseases, unspecified soft tissue disorders, osteoarthritis, sleep disorders, migraine, pain, gastroesophageal reflux, metabolic dysfunction-associated steatotic liver disease, depression, anxiety disorders, mental and behavioural disorders due to psychoactive substance use, chronic kidney disease, acute kidney failure, asthma, chronic obstructive pulmonary disease, infectious and parasitic diseases, and neoplasms. The office or outpatient services, hospital inpatient observation care services, emergency visits, and preventive medicine services were included, together with the following medications: glipizide, sitagliptin, glimepiride, pioglitazone, glyburide, linagliptin, saxagliptin, adrenal corticosteroids, thyroid supplements, antilipemic agents, diuretics, beta-blockers, ace inhibitors, calcium channel blockers, angiotensin II inhibitors, antidepressants, analgesics, sedatives/hypnotics, anticonvulsants, central nervous system medications, anaesthetics, antipsychotics, central nervous system stimulants, gastroprotectors, antacids, antiemetics, laxatives, antirheumatics, skeletal muscle relaxants, anticoagulants, platelet aggregation inhibitors, and lastly lab values: total cholesterol, tricylglycerides, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, HbA1c, ALT, AST, ASP, and cobalamin. The cofounders included are based on the International Classification of Diseases (versions 9 and 10) diagnosis and procedure codes, and generic drug names, respectively.

Prespecified subgroup analyses

Subgroup analyses were performed to explore differences in risk concerning the desired outcomes between users of GLP-1 RAs and those of SGLT2i. These prespecified analyses considered factors such as age, sex, obesity, and hypertension. Notably, we further differentiated between long-acting GLP-1 RAs, including semaglutide, liraglutide, dulaglutide, and SGLT2i empagliflozin, canagliflozin, and dapagliflozin. This distinction is significant to understanding drug-specific influences on the observed clinical outcomes.

Statistical analysis

Baseline characteristics were compared with chi-square tests for categorical variables and independent-sample t-tests for continuous variables. The TriNetX

platform was used to run 1:1 propensity score matching using logistic regression. Once it yields scores that range between 0 and 1, the platform utilizes a greedy, nearest neighbour matching with a caliper of 0.1 pooled standard deviations once propensity scores have been calculated for each patient. Kaplan–Meier analysis was used to estimate the probability of outcome at daily time intervals with censoring applied. When the last fact (outcomes of interest or other medical encounters) in the patient's record was in the time window for analysis, the patient was censored on the day after the last fact in their record. This method adjusted for factors such as age, sex, obesity, and hypertension to evaluate their roles as risk factors in the outcomes of interest. We evaluated the relationships between GLP-1 RA users and the control group concerning primary and secondary outcomes using the Cox proportional hazards model, enabling the calculation of adjusted hazard ratios (HRs). HRs and 95% confidence intervals (CIs) were used to describe the relative hazard of the outcomes based on a comparison of time-to-event rates. Incidence rates (IRs) and rate differences (RDs), with 95% CIs, were calculated per 1,000 person-years. The proportional hazards assumption was tested using the generalised Schoenfeld approach on the TriNetX platform, with adjusted HRs recalculated for specific time intervals if the assumptions were violated. Survival probabilities were visualised using the Kaplan-Meier method. Sensitivity analyses included examining cases from different registration periods and conducting Cox proportional hazards analyses with varying covariates. Data were collected and analysed on November 15th, 2024, within the TriNetX Analytics platform. We used R software (version 3.2.2, Free Software Foundation, Inc, Boston, MA).

Role of the funding source
This work has not been funded.

Ethics
The de-identification of data adheres to the criteria outlined in section §164.514(b)(1) of the Health Insurance Portability and Accountability Act Privacy Rule, which takes precedence over TriNetX's waiver granted by the Western Institutional Review Board. Consequently, no additional ethical approval was necessary. Since the data are routinely gathered and completely anonymised, obtaining patient consent is not required. More detailed information is provided in the Supplementary Material section *"TrinetX platform description"*.

Results

Study population baseline characteristics
The cohort before propensity-score matching consisted of a total of 25,043 patients, of which 8,092 initiated treatments for the first time with SGLT2i and 16,951 with GLP-1 RAs (Table 1). Figure 1 shows the flowchart of participant

selection with their exclusion criteria for both included cohorts. The geographic location of patients initiating each drug class is distributed mainly in the Northeast and South United States regions. In the propensity-score matched analysis (n=8,092 per group), baseline characteristics were balanced between GLP-1RA and SGLT2i groups. Mean age was identical (57.7 years), with similar sex (47% male), ethnicity (8.1% Hispanic/Latinx), and racial distribution (~49.5% White, ~14.7% Black, ~8.3% Asian). Both groups had comparable rates of diabetes complications (11%), hypertension (66%), obesity (27%), heart failure (5%), and ischemic heart disease (14%). Healthcare utilization (e.g., outpatient visits ~63%), medication use (e.g., hypolipidemic agents ~65%), and laboratory measures (e.g., HbA1c ~8.8%) were also similar. Baseline characteristics of all cohorts before and after propensity-score matching are presented in detail in Table 1, and all covariates had a standardized mean difference <0.1 after weighting, indicating adequate balance. Baseline characteristics for all comparisons are presented in Supplementary Tables 6-15.

Table 1. Characteristics of before-and-after propensity-score matched GLP-1 RAs versus SGLT2i cohorts for the study population of patients with T2DM.

	Before propensity-score matching			After propensity-score matching		
	GLP-1 RAs	**SGLT2i**	**SMD**	**GLP-1 RAs**	**SGLT2i**	**SMD**
Total, n	16,951	8,863		8,092	8,092	
Age, Mean ± SD	55.4 +/- 12.1	58.1 +/- 11.3	0.233	57.7 +/- 11.5	57.7 +/- 11.4	0.002
Sex (%)						
Male	39.2	48.9	0.195	47.1	47.3	0.003
Female	53.3	39.2	0.285	41.0	41.2	0.006
Ethnicity (%)						
Hispanic/Latinx	9.3	7.9	0.050	8.1	8.1	<0.001
Not Hispanic/Latinx	68.8	68.2	0.012	67.1	68.0	0.020
Unknown	21.9	23.9	0.047	24.8	23.9	0.022
Race (%)						
White	54.0	47.2	0.136	49.4	49.5	0.002
Black	20.1	13.9	0.166	14.6	14.8	0.005
Asian	4.5	11.0	0.243	8.1	8.4	0.009
Unknown	14.6	21.6	0.184	21.3	20.9	0.011
Cardiovascular and other risk/conditions (%)						
Type 2 diabetes mellitus with neurological complications	13.0	11.2	0.056	11.3	11.4	0.003

Table 1. Characteristics of before-and-after propensity-score matched GLP-1 RAs versus SGLT2i cohorts for the study population of patients with T2DM. (continued)

	Before propensity-score matching			After propensity-score matching		
	GLP-1 RAs	SGLT2i	SMD	GLP-1 RAs	SGLT2i	SMD
Type 2 diabetes mellitus with kidney complications	6.5	8.4	0.071	7.6	7.7	0.005
Type 2 diabetes mellitus with opthalmic complications	5.4	5.1	0.015	5.2	5.0	0.010
Hypothyroidism	11.8	10.1	0.055	10.8	10.4	0.012
Overweight and obesity	36.5	25.3	0.244	27.1	26.8	0.007
Essential (primary) hypertension	61.4	67.0	0.118	65.9	65.6	0.006
Heart failure	4.5	5.9	0.065	5.0	5.3	0.011
Atrial fibrillation and flutter	3.7	5.2	0.073	4.6	4.9	0.013
Ischemic heart diseases	11.1	16.2	0.150	14.3	14.6	0.008
Cerebrovascular diseases	4.0	5.2	0.053	4.6	4.8	0.011
Venous embolism and thrombosis	1.3	1.2	0.009	1.2	1.2	0.001
Unspecified soft tissue disorders	16.9	13.7	0.088	14.6	14.1	0.012
Osteoarthritis	15.2	13.7	0.041	13.9	13.8	0.003
Asthma	10.2	7.3	0.102	7.3	7.6	0.010
Chronic obstructive pulmonary diseases	4.5	4.4	0.008	4.5	4.5	0.003
Gastro-esophageal reflux disease	17.8	16.4	0.037	17.0	16.7	0.007
Other liver diseases	7.1	7.0	0.005	6.8	7.1	0.009
Acute kidney failure	2.2	1.9	0.028	1.8	1.8	0.004
Chronic kidney disease	3.4	3.5	0.007	3.5	3.4	0.007
Neoplasms	13.2	14.4	0.032	14.1	14.1	0.002
Certain infectious and parasitic diseases	16.0	13.9	0.058	14.4	14.2	0.005
Depressive episode	14.9	10.7	0.127	11.6	11.3	0.007

Table 1. Characteristics of before-and-after propensity-score matched GLP-1 RAs versus SGLT2i cohorts for the study population of patients with T2DM. (continued)

	Before propensity-score matching			After propensity-score matching		
	GLP-1 RAs	SGLT2i	SMD	GLP-1 RAs	SGLT2i	SMD
Anxiety disorders	12.3	10.5	0.056	11.0	10.8	0.005
Mental and behavioral disorders due to psychoactive substance use	8.9	9.7	0.029	9.6	9.6	0.001
Sleep disorders	21.0	17.1	0.098	17.7	17.7	0.002
Pain, not elsewhere classified	11.4	11.1	0.008	11.3	11.1	0.006
Migraine	4.1	2.5	0.091	2.7	2.6	0.005
Pre-existing medical procedures (%)						
Office or other outpatient services	47.1	64.5	0.355	62.9	62.2	0.013
Hospital inpatient and observation care services	4.9	7.1	0.095	6.5	6.4	0.005
Emergency department services	17.8	15.9	0.049	15.7	15.6	0.004
Preventive medicine services	10.2	11.2	0.033	11.1	11.3	0.005
Medication prescriptions (%)						
Thyroid supplements	13.2	10.5	0.081	11.6	11.0	0.019
Antacids	27.3	28.3	0.023	27.5	27.7	0.004
Antiemetics	17.4	16.5	0.026	16.7	16.2	0.012
Laxatives	18.7	18.4	0.008	18.5	17.9	0.014
Other gastric medications	28.1	24.6	0.078	25.5	25.1	0.009
Antirheumatics	31.0	26.4	0.101	27.5	27.0	0.012
Skeletal muscle relaxants	13.9	12.0	0.058	12.1	12.1	0.002
Anticoagulants	14.1	14.0	0.003	13.8	13.8	0.001
Platelet aggregation inhibitors	23.6	25.4	0.043	24.4	24.7	0.007
Adrenal corticosteroids	31.6	29.5	0.044	29.8	29.9	0.002
Hypolipidemic agents	62.7	65.5	0.060	64.8	64.6	0.004
Diuretics	36.3	33.8	0.052	34.2	34.3	0.002

Table 1. Characteristics of before-and-after propensity-score matched GLP-1 RAs versus SGLT2i cohorts for the study population of patients with T2DM. (continued)

	Before propensity-score matching			After propensity-score matching		
	GLP-1 RAs	**SGLT2i**	**SMD**	**GLP-1 RAs**	**SGLT2i**	**SMD**
Beta blockers	29.2	31.4	0.048	30.5	30.6	0.001
ACE inhibitors	38.5	38.6	0.003	37.7	38.2	0.011
Angiotensin II inhibitor	23.4	25.3	0.043	24.5	24.9	0.008
Calcium channel blockers	22.0	23.4	0.032	22.4	23.0	0.014
Linagliptin	2.9	4.0	0.057	3.6	3.6	0.002
Sitagliptin	16.7	22.1	0.135	20.5	20.6	0.004
Saxagliptin	1.1	2.1	0.075	1.6	1.8	0.013
Glipizide	17.3	18.7	0.036	18.0	18.2	0.006
Glimepiride	10.3	11.2	0.028	11.2	11.2	0.003
Pioglitazone	5.7	5.1	0.025	5.0	5.1	0.006
Glyburide	3.8	5.6	0.085	5.1	5.1	0.001
Analgesics	52.8	49.8	0.060	50.0	49.8	0.004
Anesthetics	20.9	20.0	0.022	20.0	19.6	0.008
Sedatives/Hypnotics	24.0	22.5	0.034	22.3	22.4	0.001
Central nervous system medications	20.0	15.0	0.131	16.0	15.5	0.013
Central nervous system stimulants	3.0	1.9	0.071	2.2	2.0	0.009
Anticonvulsants	21.4	15.5	0.153	16.4	16.0	0.009
Antidepressants	31.2	21.7	0.215	23.5	23.0	0.010
Antipsychotics	4.7	3.3	0.067	3.6	3.5	0.005
Laboratory values (%)						
Hemoglobin A1C/ Hemoglobin total in blood [%]	8.8 +/- 2.0	8.6 +/- 1.8	0.058	8.8 +/- 2.0	8.7 +/- 1.8	0.059
Cholesterol in serum or plasma [mg/dL]	172.4 +/- 46.0	171.1 +/- 47.3	0.027	169.6 +/- 46.7	172.1 +/- 47.6	0.053
Cholesterol in LDL in serum or plasma [mg/dL]	91.5 +/- 36.5	89.7 +/- 36.7	0.049	88.5 +/- 36.8	90.4 +/- 36.7	0.050
Cholesterol in HDL in serum or plasma [mg/dL]	43.1 +/- 13.6	42.9 +/- 14.3	0.017	41.9 +/- 13.9	43.1 +/- 14.3	0.088

Table 1. Characteristics of before-and-after propensity-score matched GLP-1 RAs versus SGLT2i cohorts for the study population of patients with T2DM. (continued)

	Before propensity-score matching			After propensity-score matching		
	GLP-1 RAs	SGLT2i	SMD	GLP-1 RAs	SGLT2i	SMD
Triglycerides in serum, plasma, or blood [mg/dL]	201.4 +/- 205.0	203.9 +/- 189.1	0.013	206.1 +/- 210.1	205.3 +/- 192.1	0.004
Alanine aminotransferase in serum, plasma, or blood [U/L]	33.2 +/- 51.7	33.6 +/- 32.0	0.009	33.2 +/- 26.9	33.7 +/- 32.8	0.018
Aspartate aminotransferase in serum or plasma [U/L]	26.7 +/- 21.2	27.3 +/- 24.0	0.027	26.8 +/- 18.4	27.5 +/- 24.8	0.033
Alkaline phosphatase in serum, plasma, or blood [U/L]	84.3 +/- 31.2	82.3 +/- 39.0	0.058	83.1 +/- 31.7	83.1 +/- 39.8	0.001
Cobalamin (vitamin B12) in serum, plasma, or blood [pg/mL]	575.6 +/- 344.8	581.7 +/- 360.3	0.017	589.9 +/- 333.4	585.8 +/- 361.5	0.012
Systolic blood pressure [Hg]	130.8 +/- 16.9	131.0 +/- 17.6	0.009	131.2 +/- 16.9	131.1 +/- 17.6	0.009
Diastolic blood pressure [Hg]	77.0 +/- 10.7	77.0 +/- 11.0	0.006	77.0 +/- 10.5	77.1 +/- 11.0	0.011

Antihyperglycemic drug comparisons and incident dementia risk

A total of 509 cases dementia cases were developed during a median 5.8-year follow-up: 238 for SGLT2i initiators (IR: 5.20, 95% CI: 4.56–5.91; mean follow-up time 2,076 days), 271 for GLP-1 RA initiators (IR: 5.86, 95% CI: 5.19–6.59; per 1,000 person-years; mean follow-up time 2,104 days). Incidence (%) for all-cause dementia, vascular dementia, and Alzheimer's dementia are available in Supplementary Figure 1. Within 10 years, the risk for all-cause dementia did not differ between GLP-1 RA initiators and SGLT2i initiators (HR: 1.12, 95% CI: 0.94–1.34; RD 0.61, 95% CI -0.33 to 1.55) (Supplementary Figure 2). For secondary outcomes, similarly, no difference was observed regarding vascular dementia (HR: 1.12, 95% CI: 0.75–1.67, RD: 0.11, 95% CI: -0.29 to 0.52) and Alzheimer's dementia (HR: 1.15; 95% CI: 0.80–1.68, RD: 0.17, 95% CI: -0.27 to 0.61) (Supplementary Figures 3 and 4). Kaplan-Meier curves illustrating the cumulative incidence of the study outcomes in patients initiating GLP-1 RAs versus SGLT2i initiators showed consistent results (Figures 2 and 3).

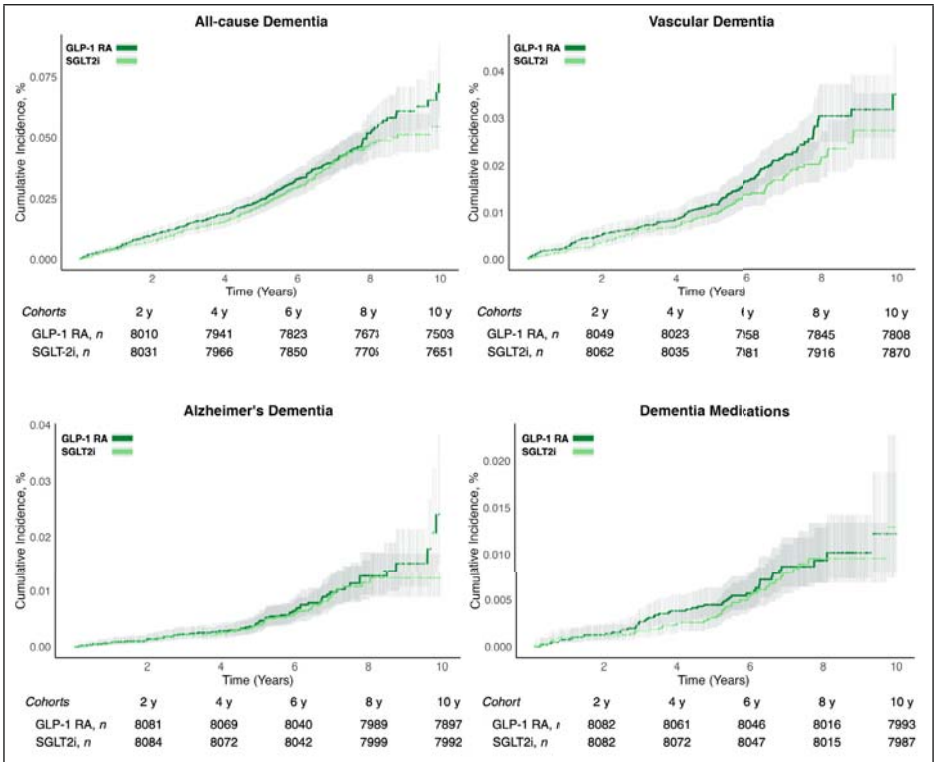
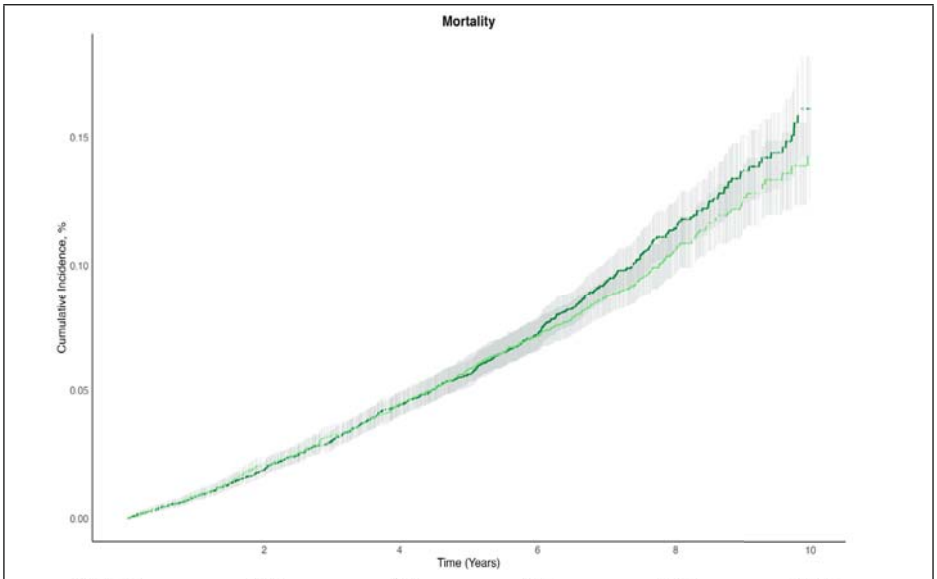


Figure 2. Kaplan Meier curves showing the cumulative incidence of all-cause dementia, vascular dementia, and Alzheimer's dementia in initiators of GLP-1 RAs and SGLT2i.



Cohorts	2 y	4 y	6 y	8 y	10 y
GLP-1 RA, *n*	7938	7735	7508	7165	6788
SGLT-2i, *n*	7925	7730	7514	7230	6937

Figure 3. Kaplan Meier curves showing the cumulative incidence of all-cause mortality in initiators of GLP-1 RAs and SGLT2i.

Subgroup analyses and sensitivity analyses

Estimates for all-cause dementia and Alzheimer's dementia were similar, comparing GLP-1 RAs with SGLT2i when the patients were stratified by age, obesity, and hypertension, without significant differences across groups (Supplementary Figures 2–4). Using medications for dementia as the outcome (dementia), among females, the risk with GLP-1 RAs is higher for dementia development compared to females using SGLT2i (HR: 1.70, 95% CI: 1.17–2.46; RD: 1.50, 95% CI: 0.42–2.58) (Supplementary Figure 5). Similarly, using dementia medications as the outcome for dementia, males using GLP-1 RAs showed an increased risk for all-cause mortality compared to males using SGLT2i (HR: 1.19, 95% CI: 1.02–1.38; RD: 2.62, 95% CI: 0.27–4.96) (Supplementary Figure 6). No other significant differences were shown (Supplementary Figures 7-10). For the drug vs. drug class-specific comparisons, semaglutide alone conferred a lower risk for all-cause mortality when compared to SGLT2i (HR: 0.71, 95% CI: 0.51–0.98; RD: -4.02, 95% CI: -7.04 to -1.00) (Supplementary Figure 11).

Subgroup analyses and sensitivity analyses with a one-year lag

After sensitivity analysis with a one-year lag, no significant differences were observed in any of the above groups or outcomes (Supplementary Figures 12-21)

except for the female group using medications for dementia as the outcome (dementia), with GLP-1 RAs, which persistently showed a higher for dementia development compared to females using SGLT2i (HR: 1.63, 95% CI: 1.14–2.35; RD: 1.38, 95% CI: 0.35–2.41) (Supplementary Figure 15).

Conclusions

This large cohort study of aggregate EHR data of more than 90 healthcare organisations was the first to compare head-to-head the effects of the newer antihyperglycemic drug classes – SGLT2i and GLP-1 RAs initiation in metformin users – on the risk of all-cause dementia, vascular, and Alzheimer's dementia in adult patients with T2DM. This study demonstrates that comparing two of the most effective antihyperglycemic drug classes, i.e., SGLT2i and GLP-1 RAs, as well as the main drugs of each class with the other class, does not reveal any difference in the risk for incident dementia. When identifying dementia cases through dementia medication use, female users of GLP-1 RAs persistently showed increased risk for incident dementia compared to female users of SGLT2i.

Previous population-based studies have shown results regarding the protective effect of GLP-1 RAs over DPP4 inhibitors and sulfonylureas (8) and the benefit of SGLT2i (mainly dapagliflozin and empagliflozin) compared to DPP4 inhibitors (15) and compared to non-SGLT2i users for the risk of dementia (16). Similarly, a recent Korean study also found that initiation of SGLT2i versus DPP4 inhibitor was associated with lower dementia risk, using a rigorous study design (17). In a pooled analysis of randomized trials, the use of GLP-1RAs, mainly liraglutide, displayed a reduced risk for dementia when compared to placebo (18). A recent South Korean study compared dulaglutide versus SGLT2i and did not observe any difference in all-cause dementia, Alzheimer's dementia, or vascular dementia (19). Consistent with previous research, our study focuses on a primarily non-Asian population that has not been adequately studied to date and suggests that the protective effects observed may be class-specific rather than solely drug-specific.

Our study has important clinical implications. Possible mechanisms that could explain the reduced risk of dementia with SGLT2i in humans remain elusive due to a lack of mechanistic or cohort studies or RCTs. Although underlying mechanisms could be either direct and/or indirect, preclinical studies on neuroimaging and neuropathology have demonstrated a reduced risk of developing dementia when treating mice with T2DM with SGLT2i, which are lipid-soluble and cross the blood-brain barrier, possibly due to their protective effects on cerebrovascular endothelial cells (20, 21). On the other hand, GLP-1 RAs have attracted more attention as a potential link between metabolic and brain impairment (22-24). GLP-1 RAs may exert their central effects on neuronal function by multiple mechanisms, including reducing inflammation, diminishing tau phosphorylation, and improving synaptic

function, in addition to their influence on insulin resistance (24, 25). Lastly, it is shown that GLP-1RAs may exert an indirect neuroprotective effect by drug-induced weight loss, which reduces Alzheimer's dementia risk in obesity (26). More in-depth mechanistic studies are needed before proceeding with large RCTs.

To the best of our knowledge, this is a novel study, offering the first direct comparison of the newer two antihyperglycemic drug classes (i.e., GLP-1 RAs vs SGLT2i) as second-line therapy among metformin users, with a median follow-up time of more than 5 years, concerning dementia risk, while having adjusted for a vast number of potential confounders.

Our study has several limitations. Given the retrospective observational nature of this study, no causal inferences can be drawn, but our data can form the basis for power calculations for specific randomised clinical trials in the future. Additionally, the TriNetX database consists of patients who have medical encounters with healthcare systems that contribute to the platform. Despite the platform encompassing over 120 million patients across more than 12 countries, it does not fully represent the global population. For instance, variables regarding socioeconomic status, a significant risk factor for the outcome of interest, are lacking, which may introduce a selection bias. Besides, alcohol consumption was not available as a variable to add to our sensitivity analyses, and this may have influenced the results. Due to privacy issues, the names of healthcare organisations were excluded from the analysis. Asian centres constituted less than 1% of the institutions, preventing assessments of comparisons within institutions or between continents. Moreover, as with most observational cohort studies, overdiagnosis, misdiagnosis, and underdiagnosis, as well as challenges in accounting for unmeasured or uncontrolled factors that could influence outcomes, are limiting factors. Such errors would be expected to lead to random misclassification and depression of the effect estimates towards the null, but could not have accounted for any statistically significant results demonstrated herein. Issues like self-selection bias and reverse causality also pose concerns, but it would seem unlikely that persons with dementia would pick one or the other medication. Specifically, until recently, there has been no information on GLP-1 RAs or SGLT2i being superior to each other in terms of dementia risk, and thus, this factor could not have biased the selection of one or the other medication by patients or prescribers. The assessment of patient adherence to medications solely through electronic health records was not available. Drug side effects or perceived lack of effectiveness may lead patients to discontinue medication, conditions that may affect study outcomes. However, any discontinuations, similar to any random misclassification, would be expected to suppress the effect estimates towards the null but could not have accounted for the statistically significant differences towards improving dementia outcomes reported herein.

Additionally, the absence of length of treatment use, i.e., for any medication, including metformin use, was not provided from the TriNetX platform, may have confounded the results and may be considered a limitation of this sensitivity analysis, but one would not expect differential length of treatment between the two classes studied herein. Nonetheless, there is no evidence that metformin duration is associated with the initiation of SGLT2i or GLP-1 RA therapy, and no evidence that the length of therapy by metformin is directly associated with dementia. There remains a possibility of residual and unmeasured confounding factors influencing the results, as is common in observational studies. It is possible that the observed relationships are influenced in part by improved management of diabetes or cardiorenal metabolism, although specific data on this were not presented. Besides, TriNetX does not contain claims data, potentially missing dementia cases, among, for example, patients using two hospitals (one of which may not be part of the TriNetX network). Finally, our results do not necessarily extend to the upcoming dual- and triple-agonists, which may have different effects on the brains of humans. Future long-term RCTs are essential to establish any causal relationships between antihyperglycemic medications and dementia or mortality.

In conclusion, our findings do not support a difference in all-cause, vascular, or Alzheimer's dementia with GLP-1 RAs compared with SGLT2i. Further studies with dual or triple incretin agonists are required to elucidate their potential to mitigate dementia risk.

Article Information

Author contributions

FA had access to and verified the underlying study data, was responsible for ensuring the integrity and accuracy of the data analysis, and is the guarantor. FA and MK drafted the manuscript. FA and MK are responsible for the study design, data analysis, and interpretation, and all authors contributed to the critical revision of the manuscript. CSM supervised the study. The corresponding author (MK) confirms that all listed authors meet the criteria for authorship and that no other authors meeting the criteria have been excluded.

Conflict of interest disclosures

FA, SN, and GA have no competing interests. MK receives funding support from an MD-PhD grant from the University Medical Center Groningen. Over the past three years, CSM reports grants through his institution from Merck, Massachusetts Life Sciences Center, and Boehringer Ingelheim; he reports personal consulting fees and support with research reagents from Ansh Inc., collaborative research support from LabCorp Inc., reports personal consulting fees from Nestle, Amgen, Corcept, Aligos, Intercept, 89 Bio, Madrigal, Novo Nordisk, and Regeneron, reports travel support and fees from TMIOA, Elsevier, and the Cardio Metabolic Health Conference. No consulting activity above is related to the work presented herein.

Data sharing statement

This study used population-level aggregate and de-identified data collected by the TriNetX Platform, which are available from TriNetX (https://trinetx.com/); however, third-party restrictions apply to the availability of these data. The data were used under license for this study, with restrictions that do not allow for the data to be redistributed or made publicly available. To gain access to the data, a request can be made to TriNetX (join@trinetx.com), but costs might be incurred, and a data-sharing agreement would be necessary. As the data are routinely collected and fully anonymized, patient consent is not required. Data specific to this study, including diagnosis codes and group characteristics in an aggregated format, are included in the paper as tables, figures, and supplementary online content.

Supplementary data

This article contains supporting information available online at:
https://drive.google.com/file/d/1WMtuqdq7ayjcCUN9_ZicyeNeHtnGVWyt/view?usp=drive_link

References

1. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2023;402(10397):203-34.
2. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet. 2020;396(10248):413-46.
3. Biessels GJ, Verhagen C, Janssen J, van den Berg E, Zinman B, Rosenstock J, et al. Effect of Linagliptin on Cognitive Performance in Patients With Type 2 Diabetes and Cardiorenal Comorbidities: The CARMELINA Randomized Trial. Diabetes Care. 2019;42(10):1930-8.
4. Biessels GJ, Verhagen C, Janssen J, van den Berg E, Wallenstein G, Zinman B, et al. Effects of linagliptin vs glimepiride on cognitive performance in type 2 diabetes: results of the randomised double-blind, active-controlled CAROLINA-COGNITION study. Diabetologia. 2021;64(6):1235-45.
5. Wium-Andersen IK, Osler M, Jørgensen MB, Rungby J, Wium-Andersen MK. Antidiabetic medication and risk of dementia in patients with type 2 diabetes: a nested case-control study. Eur J Endocrinol. 2019;181(5):499-507.
6. Cukierman-Yaffe T, Gerstein HC, Colhoun HM, Diaz R, García-Pérez LE, Lakshmanan M, et al. Effect of dulaglutide on cognitive impairment in type 2 diabetes: an exploratory analysis of the REWIND trial. Lancet Neurol. 2020;19(7):582-90.
7. Tang B, Sjölander A, Wastesson JW, Maura G, Blotiere P-O, Szilcz M, et al. Comparative effectiveness of glucagon-like peptide-1 agonists, dipeptidyl peptidase-4 inhibitors, and sulfonylureas on the risk of dementia in older individuals with type 2 diabetes in Sweden: an emulated trial study. eClinicalMedicine. 2024.
8. Bidulka P, Lugo-Palacios DG, Carroll O, O'Neill S, Adler AI, Basu A, et al. Comparative effectiveness of second line oral antidiabetic treatments among people with type 2 diabetes mellitus: emulation of a target trial using routinely collected health data. BMJ. 2024;385:e077097.
9. 13. Older Adults: Standards of Care in Diabetes-2024. Diabetes Care. 2024;47(Suppl 1):S244-s57.
10. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. Curr Epidemiol Rep. 2015;2(4):221-8.
11. Yoshida K, Solomon DH, Kim SC. Active-comparator design and new-user design in observational studies. Nature Reviews Rheumatology. 2015;11(7):437-41.
12. Schneeweiss S, Patorno E. Conducting Real-world Evidence Studies on the Clinical Outcomes of Diabetes Treatments. Endocrine Reviews. 2021;42(5):658-90.
13. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.
14. Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. JAMA. 2022;328(24):2446-7.
15. Wu CY, Iskander C, Wang C, Xiong LY, Shah BR, Edwards JD, et al. Association of Sodium-Glucose Cotransporter 2 Inhibitors With Time to Dementia: A Population-Based Cohort Study. Diabetes Care. 2023;46(2):297-304.
16. Kim HK, Biessels GJ, Yu MH, Hong N, Lee YH, Lee BW, et al. SGLT2 Inhibitor Use and Risk of Dementia and Parkinson Disease Among Patients With Type 2 Diabetes. Neurology. 2024;103(8):e209805.

17. Shin A, Koo BK, Lee JY, Kang EH. Risk of dementia after initiation of sodium-glucose cotransporter-2 inhibitors versus dipeptidyl peptidase-4 inhibitors in adults aged 40-69 years with type 2 diabetes: population based cohort study. BMJ. 2024;386:e079475.
18. Nørgaard CH, Friedrich S, Hansen CT, Gerds T, Ballard C, Møller DV, et al. Treatment with glucagon-like peptide-1 receptor agonists and incidence of dementia: Data from pooled double-blind randomized controlled trials and nationwide disease and prescription registers. Alzheimers Dement (N Y). 2022;8(1):e12268.
19. Hong B, Bea S, Ko HY, Kim WJ, Cho YM, Shin JY. Sodium-Glucose Cotransporter-2 Inhibitors, Dulaglutide, and Risk for Dementia : A Population-Based Cohort Study. Ann Intern Med. 2024;177(10):1319-29.
20. Bello-Chavolla OY, Antonio-Villa NE, Vargas-Vázquez A, Ávila-Funes JA, Aguilar-Salinas CA. Pathophysiological Mechanisms Linking Type 2 Diabetes and Dementia: Review of Evidence from Clinical, Translational and Epidemiological Research. Curr Diabetes Rev. 2019;15(6):456-70.
21. Lin B, Koibuchi N, Hasegawa Y, Sueta D, Toyama K, Uekawa K, et al. Glycemic control with empagliflozin, a novel selective SGLT2 inhibitor, ameliorates cardiovascular injury and cognitive dysfunction in obese and type 2 diabetic mice. Cardiovasc Diabetol. 2014;13:148.
22. Banks WA. The blood-brain barrier as an endocrine tissue. Nat Rev Endocrinol. 2019;15(8):444-55.
23. Drucker DJ. The benefits of GLP-1 drugs beyond obesity. Science. 2024;385(6706):258-60.
24. Grieco M, Giorgi A, Gentile MC, d'Erme M, Morano S, Maras B, et al. Glucagon-Like Peptide-1: A Focus on Neurodegenerative Diseases. Front Neurosci. 2019;13:1112.
25. Alford S, Patel D, Perakakis N, Mantzoros CS. Obesity as a risk factor for Alzheimer's disease: weighing the evidence. Obes Rev. 2018;19(2):269-80.
26. Kim KY, Ha J, Lee JY, Kim E. Weight loss and risk of dementia in individuals with versus without obesity. Alzheimers Dement. 2023;19(12):5471-81.


TOWER OF BABEL

Chapter 7

Discussion

At a glance

The emergence of large electronic health records (EHR) databases with prospectively collected data and long follow-up times, coupled with advanced computational tools, has markedly advanced our diagnostic and prognostic capabilities. In this thesis, we present a collection of studies that bridge cardiometabolic research and data science to enhance how we gauge and address cardiometabolic disorders. This concluding chapter takes a thorough look at our main findings, set against the broader context of evidence-informed cardiometabolic medicine. First, a brief historical overview of predictive models is provided, followed by a detailed examination of our own work, its contributions, and potential avenues for further investigation. Finally, the chapter highlights how digital innovations, biomarker-based assessments, and personalized medicine might steer future management of cardiometabolic conditions. Figure 4 provides a graphical summary of the findings of this thesis.

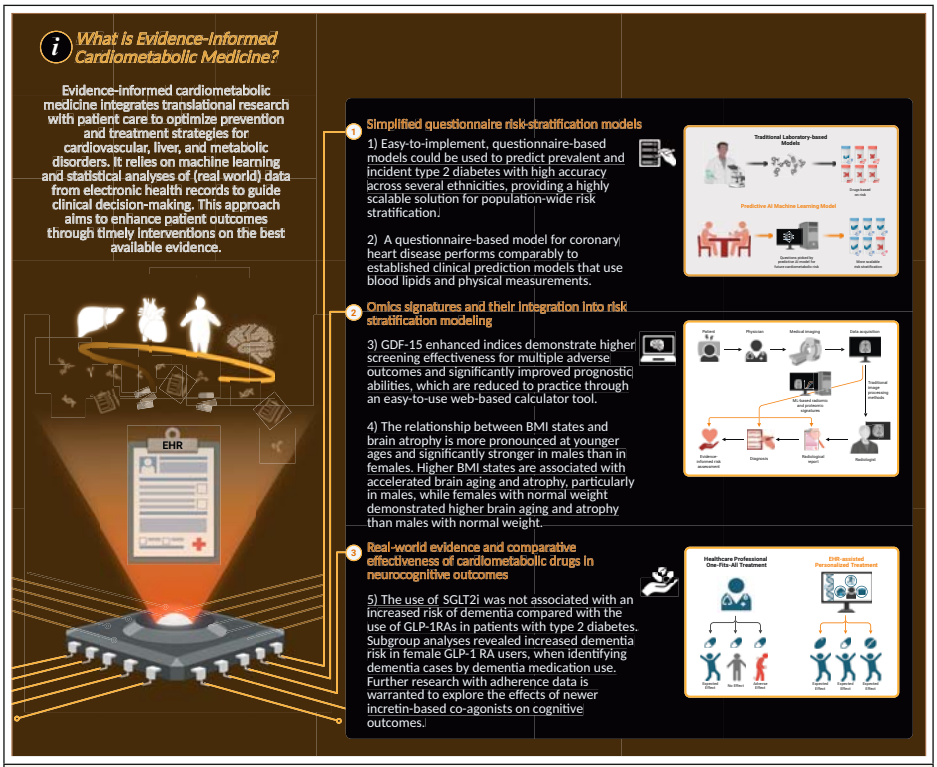


Figure 4. **Graphical summary of the thesis findings. Created with BioRender.com.**

From the set-out phase of Electronic Health Records and risk prediction in cardiometabolic disease to the state-of-the-art

In the earliest stages of predicting cardiometabolic risk, scientists and physicians searched for ways to predict risks tied to cardiovascular health and thus turned to constructing and studying population cohorts. As outlined in the introduction of this thesis, it was known that identifying the biggest cardiovascular threats was key, so research efforts focused on classifying individuals based on their likelihood of developing adverse cardiovascular events. More recently, digital technologies and artificial intelligence (AI) have enabled economical screening programs, a highly needed development in regions and settings with scarce resources, offering new care pathways in approaching cardiometabolic diseases.

This long-standing effort to refine metabolic risk measurement has been integral to the progress of preventive medicine as a whole. Groundbreaking work, such as the Framingham Heart Study launched in 1948, highlighted the major roles of hypertension, hypercholesterolemia, and tobacco use, insights that would inform the design of future predictive tools and shape our understanding of the development of non-communicable diseases (1). Soon after, in the late 1950s, the Seven Countries Study brought global attention to how dietary choices powerfully influence cardiovascular health, underscoring the importance of lifestyle changes in preventing atherosclerotic complications (2, 3). By the early 1980s, researchers had introduced the Framingham Risk Score, which is still used in contemporary medicine to assess the ten-year likelihood of a person developing coronary heart disease (4). Advances in technology continued to strengthen its predictive power in the following decades. Then, beginning in 2002, the PREDICT study utilized online decision support in primary care settings, automatically enrolling a large pool of patients for ongoing observation (5). More recently, data scientists have employed Machine Learning to uncover hard-to-detect risk patterns, while novel indicators, such as the Body Roundness Index, have emerged to address perceived shortcomings of traditional Body Mass Index (BMI) measurements (6).

In subsequent years, large population-focused projects recruiting volunteers took center stage. Since 2006, the UK Biobank has recruited around half a million participants, compiling an extensive database that reveals key phenotypic and genetic aspects of, among others, cardiometabolic disorders (7). Weaving multimodal data into existing risk models allows for a more precise quantification of an individual's vulnerability to type 2 diabetes (T2D) or coronary artery disease (CAD), thereby facilitating preventive measures and early interventions (8). Simultaneously, the expanded use of EHR systems in hospitals has yielded new sizable sources of acquiring patient information with minimal effort. One notable example is TriNetX, a worldwide research network that unites anonymized clinical

data from a multitude of healthcare providers and supports rapid access to millions of individual records (https://trinetx.com/).

So far, physicians have empirically trained their own "neural networks" to ask the right (hypothesis-testing) questions during history taking (anamnesis). Nonetheless, history taking, with the goal of establishing the proper diagnosis, may be improved by emerging data-driven approaches (i.e., by informing history taking from variables of questionnaire-driven risk-stratification models, such as those in **Chapters 2 and 3**). Specifically, for clinical implementation, large language models could facilitate triage, since they have already shown promising results in achieving physician-level performance or even being applied as standalone diagnostic tools in certain clinical scenarios (8, 9). Therefore, by merging the wealth of data from resources such as the UK Biobank and TriNetX with advanced computational methods, researchers have reached a turning point in forecasting cardiometabolic risks. Machine Learning techniques now appear to have the ability to map out the intricate interplay of a wide array of risk factors, often surpassing the predictive power of older cardiovascular models (10). In essence, the evolution of EHR-informed risk assessment has progressed from influential early epidemiological studies to cutting-edge, data-intensive approaches that integrate various streams of information, resulting in more accurate and tailored evaluations of cardiometabolic risk than ever before.

Although years of research have clarified cardiometabolic risk factors and established pioneering EHR-based approaches, questions remain regarding how to best harness novel data streams, integrate emerging biomarkers, and address settings with limited resources, as well as study diverse populations. Furthermore, real-world evidence (RWE), particularly on neurocognitive outcomes, is not yet comprehensively incorporated into comparative effectiveness research for cardiometabolic conditions.

In response, this thesis explores three main areas: it advances non-laboratory, scalable predictive models, assesses how omics data enhance risk stratification, and applies real-world data (RWD) to gauge the neurocognitive impact of commonly prescribed cardiometabolic therapies. By focusing on these underexplored aspects, the subsequent sections aim to fill gaps in conventional risk assessment and strengthen the link between research findings and clinical practice.

Thesis Contributions

I) Non-Laboratory Predictive Modeling and Scalable Risk Stratification
Risk stratification is crucial for identifying individuals at high risk of developing cardiometabolic conditions and enabling early interventions, as most

cardiometabolic diseases are largely modifiable (11). So far, disease prediction or risk stratification in a clinical setting has been primarily based on laboratory results or measurements, such as blood pressure and waist circumference. The first two chapters of this thesis address the insufficient development and validation of non-laboratory risk stratification models, particularly Machine Learning-based approaches. The availability of biobanks presents an unprecedented opportunity to select highly predictive variables that are easy to collect, such as answers to simple questions. This way, we aimed to simplify risk prediction while maintaining at least the same performance as the currently established clinical models, in a manner that allows for scalable implementation.

In **Chapter 2,** we developed a logistic regression-based questionnaire-based prediction model for T2D in the UK Biobank, achieving an area under the curve (AUC) of 0.901 for prevalent T2D and 0.873 for incident T2D. We also internally (among ethnic minorities) and externally validated the models in the Lifelines cohort, demonstrating an AUC of 0.917 for prevalence and 0.817 for incidence. Additive models incorporating basic physical measurements, such as BMI and waist circumference, as well as blood biomarkers like HbA1c and glucose, demonstrated minor AUC improvements, which would present significant feasibility challenges for large-scale deployment in population-wide screening. To further assess clinical applicability, risk-stratification analyses were conducted, categorizing individuals into high-, medium-, and low-risk groups. Our models identified risk phenotypes with a very high incident risk of T2D, while minimizing the number of individuals required to be screened to capture such high-risk phenotypes.

In **Chapter 3**, the development of the questionnaire-based model (QUES-CAD) for CAD employed Cox proportional hazards (CoxPH) and Cox gradient boosting (CoxGBT) models. The models were trained on 448,818 White UK Biobank participants and validated in 97,770 Lifelines participants. QUES-CAD achieved C-indices of 0.692 (95% CI: 0.673–0.71) in males and 0.771 (95% CI: 0.748–0.794) in females, performing comparably to established clinical risk scores such as SCORE2 and Framingham. Internal and external validations confirmed its robustness. A critical component of this study was ensuring that model performance metrics, such as sensitivity and specificity, reflected the models' discriminatory capabilities rather than calibration-driven differences. To achieve this, rigorous alignment of group sizes with SCORE2 was conducted prior to analysis. Partial log-likelihood ratio (PLR) tests were employed to compare the goodness of fit of the models, confirming that QUES-CAD achieved similar C-index and PLR values to those of clinical models across multiple ethnic populations and biobanks (i.e., UK Biobank vs. Lifelines). Furthermore, Kaplan-Meier incidence analyses demonstrated the ability of QUES-CAD to accurately stratify individuals into high- and low-risk categories over a 15-year period, further highlighting its

reliability. These results suggest that questionnaire-based models can serve as effective population-level screening tools, particularly in settings where laboratory testing is limited. However, further research is needed to prospectively test these models in real-world (clinical) scenarios among multi-ethnic populations.

II) Omics Integration in Risk Stratification and Modeling Approaches

The integration of omics data into risk prediction and stratification represents a significant advancement in precision cardiometabolic medicine (12, 13). The study on multipurpose growth/differentiation factor 15 (GDF-15)-enhanced non-invasive tests (NITs) highlights the added predictive value of proteomic markers in refining disease risk estimation, which has been overlooked in traditional models. Specifically, in **Chapter 4**, GDF-15 was identified as the strongest predictor of adverse cardiometabolic and malignancy outcomes, improving C-indices by more than 0.2 in models for mortality, cardiovascular disease, and chronic kidney disease. The incorporation of GDF-15 into existing NITs of liver fibrosis and steatosis reduced the number of required screenings by over seven times compared to traditional models, offering a more efficient approach for identifying high-risk individuals. Kaplan-Meier analyses further demonstrated that individuals in the highest quintile of GDF-15-enhanced NITs experienced an exponentially increasing incidence of adverse outcomes.

In **Chapter 5**, using one of the largest brain magnetic resonance imaging consortia, we analyzed 46,288 individuals to evaluate the impact of overweight and obesity on brain aging and Alzheimer's-like atrophy. Higher BMI states were associated with advanced brain aging (SPARE-BA), most strongly in males with obesity (~2 additional years vs. normal weight). Males with overweight also showed elevated SPARE-BA (+8 months). Within the same BMI groups, males generally exhibited greater brain aging than females, except among normal-weight individuals, where females had higher values. SPARE-AD (Alzheimer's-like atrophy) was increased in males with obesity and in females with normal weight. Proteomic profiling identified five and eight proteins out of 1,463 proteins that were significantly (positively or negatively) associated with brain ageing and a 1-standard-deviation change in BMI. These findings highlight the potential role of (biological) brain age as an additional biomarker of cardiometabolic health.

III) Real-World Evidence and Comparative Effectiveness of Cardiometabolic Drugs in Neurocognitive Outcomes

RWE allows for assessing the efficacy of pharmacotherapies or interventions while considering several variables and endpoints not accounted for in existing randomized controlled trials (RCTs). RCTs typically involve a specific population and examine how different subgroups of participants respond to novel drugs or interventions in a controlled setting. Even though RCTs are considered the "gold

standard" in clinical interventions and particularly in drug development, they present with considerable financial and logistical constraints, while the follow-up times to achieve the predetermined endpoints prolong the study duration. Rather than relying solely on RCTs, RWE provides insights into the efficacy of a treatment or intervention in a real-world setting for outcomes not studied and/or a more diverse population. There are several sources of RWD, including EHRs, patient registries, claims/billing data, and patient-generated data, along with data collected through mobile health applications and wearables. RWD from these sources can be collected and analyzed through various study designs, including prospective and retrospective cohort studies, case–control studies, and pragmatic clinical trials. Additionally, RWE has been utilized in several phases of the drug approval cycle and can even be used to optimize the design of RCTs (14). The latter has led to the realization that well-designed RWE studies closely replicate RCT results and may even predict outcomes of future RCTs.

In the case of cardiometabolic drugs, there is a lack of evidence from RCTs examining the most prescribed antihyperglycemic drug classes for neurocognitive outcomes, the risk of which is notably increased in the presence of cardiometabolic risk. Specifically, although it is well known that T2D is a risk factor for dementia, the comparative effectiveness of sodium-glucose cotransporter-2 inhibitors (SGLT2i) and glucagon-like peptide-1 receptor agonists (GLP-1 RAs), as drug classes, on dementia risk remains unclear (15). Therefore, **Chapter 6** provides RWE on the differences in dementia risk between these drug classes. Specifically, we studied 25,814 individuals with T2D receiving either of the above drug classes over a median follow-up of 5.8 years for dementia outcomes. The incidence of all-cause dementia was 3.3% in GLP-1 RA users and 2.9% in SGLT2i users, with no significant hazard ratio differences (HR: 1.12, 95% CI: 0.94–1.34). Vascular dementia and Alzheimer's dementia risks were comparable, though subgroup analyses revealed increased dementia risk in female GLP-1 RA users when identifying dementia cases by dementia medication use. However, future prospective studies are required to determine longer-term cognitive outcomes associated with newer incretin-based therapies, such as tirzepatide, a glucose-dependent insulinotropic polypeptide and GLP-1 RA.

Future Directions

I. What can be implemented now based on this thesis?

These proposed directions, outlined in detail below, directly build upon the foundations established in this thesis. The non-laboratory questionnaire models developed in **Chapters 2 and 3** can be directly implemented in community health settings with high cardiometabolic burden to stratify risk in the absence of laboratory data. **Chapters 4 and 5**, which highlighted the importance of

proteomic and neuroimaging biomarkers, foreshadow the move toward multi-omics integration for more nuanced models of (early) disease risk and progression. Likewise, the comparative effectiveness work in **Chapter 6** lays the groundwork for digital twin approaches, enabling cost-effective and rapid analyses that can facilitate personalized interventions (i.e., with improved clustering techniques to identify groups that benefit most from specific interventions). By layering these insights, a shift will be enabled toward a cardiometabolic strategy that starts with accessible risk stratification, incorporates advanced biomarkers, and culminates in AI-augmented, patient-centered therapies.

Implementation science: Bridging the gap between cardiometabolic research and clinical practice through AI-coupled Remote Patient Monitoring

The models from **Chapters 2–4** have been externally validated and can now be directly piloted via integration into telehealth platforms such as the HealthBox, a tool which uses at-home telemonitoring to improve care for chronic conditions (https://www.universiteitleiden.nl/en/research/research-projects/science/liacs-healthbox), allowing real-time, explainable risk feedback using wearable-linked data streams. A recent systematic review of 29 RCTs found that remote patient monitoring interventions consistently improved adherence, enhanced functional outcomes, and reduced hospital readmissions and outpatient visits (16). These effects were observed across multiple conditions, including cardiovascular diseases, and were facilitated by alert-driven and scheduled monitoring systems that enabled clinicians to detect health deterioration early and intervene proactively (16). Therefore, by coupling AI-driven models as developed in this thesis with remote patient monitoring, healthcare teams can shift from occasional patient visits to ongoing, real-time, and feedback-driven (digital) strategies for preventing and managing cardiometabolic diseases.

In this scenario, tracking of blood pressure, glucose levels, and physical activity allows AI tools to promptly identify changes that might signal a worsening condition or may increase the long-term risk of developing cardiometabolic diseases. This real-time tracking, in turn, enables personalized interventions. For instance, individuals with metabolic syndrome can benefit from wearable or sweat-based sensors that collect and share glucose data with an AI model, which can then detect early signs of insulin resistance (17). Instead of passively waiting for the next checkup at outpatient clinics or primary care, this ecosystem would send personalized and targeted dietary advice or arrange a virtual appointment with a provider (16). Additionally, through chatbots or mobile applications, the questionnaire-based algorithms we developed in **Chapters 2 and 3** could inform individuals of their future risk of developing T2D or CAD. Still, lancing AI-informed remote patient monitoring into clinical routines has hurdles, especially when it

comes to connecting these tools with EHR data (16). Many EHR platforms are isolated, limiting the seamless integration of data from wearables or smartphone apps. The Fast Healthcare Interoperability Resources framework includes mobile and cloud-based applications and may be a solution by facilitating standardized data exchange between AI systems and clinical infrastructure (https://digital.nhs.uk/services/fhir-apis). However, ensuring the adoption of such frameworks requires institutional approvals and thorough regulatory support.

While technological limitations present significant challenges, establishing clinician confidence in AI systems remains a fundamental obstacle to widespread implementation. Numerous predictive algorithms operate as "black box" solutions, producing clinical suggestions without providing intelligible reasoning behind their conclusions (18). This lack of interpretability poses challenges for healthcare providers seeking to evaluate and validate machine-generated assessments, particularly when making critical decisions in patient care. Emerging interpretable AI methodologies aim to bridge this gap by delivering contextualized explanations for algorithmic outputs (18). In clinical applications, this might involve an intelligent system not only identifying elevated cardiometabolic risk in a patient but also specifying contributing factors, such as persistent hypertension, inadequate restorative sleep patterns, and prolonged physical inactivity. Presenting these explanatory factors through interfaces that reference recognized medical protocols and diagnostic criteria may enhance practitioner acceptance and facilitate integration of AI-assisted technologies into care pathways. In line with this notion and to enhance explainability, we used linear models in **Chapters 2–4**, which can be translated into simple equations, and the coefficients of each predictor reflect their importance.

Enhancing patient participation stands as an equally vital consideration for the effective deployment of AI-enhanced remote patient monitoring (19). Many digital therapeutic platforms experience suboptimal engagement levels, particularly when designs neglect variations in technological proficiency, health education backgrounds, and intrinsic motivation. A particularly effective strategy combines algorithmic suggestions with human expertise through blended support models. Consider, for instance, a post-cardiac rehabilitation scenario: while Machine Learning analyzes heart rate recovery metrics to recommend activity thresholds, certified health coaches could offer personalized guidance, address psychosocial barriers, and reinforce positive behavioral changes, synergistically improving protocol adherence. Additionally, multisite implementation studies that track real-world effectiveness metrics, such as avoided hospitalizations and quality-adjusted life years, will be essential for understanding the value of AI algorithms in clinical practice and remote patient monitoring (16). Concurrently, regulatory bodies must establish adaptive frameworks that balance innovation with patient safety

and improved outcomes, particularly regarding continuous algorithm validation and informed consent processes for AI-mediated care recommendations (20, 21). Nonetheless, so far, medical device approval of AI technologies is typically granted based on proof of technical validity/accuracy and may not directly indicate whether AI demonstrates beneficial results in improving patient outcomes (21).

Achieving scalable integration of intelligent monitoring platforms necessitates unprecedented collaboration across computational, clinical, and operational domains. Overcoming challenges related to health data interoperability, provider acceptance, insurance coverage, and sustainable patient engagement will dictate the pace of progress in predictive healthcare analytics. Through thoughtful integration of validated AI models with EHR-connected monitoring ecosystems, health systems can cultivate prevention-oriented care models that simultaneously enhance individual outcomes and population health metrics (22). This transition demands ongoing investment in digital infrastructure, clinician training programs, and patient education initiatives to fully realize the paradigm-shifting potential of AI in modern cardiometabolic care.

II. What needs to be addressed in the future to advance further?

In the future, two main approaches will be needed: (i) diverse datasets and (ii) novel analysis methodologies. First, research in cardiometabolic medicine should aim to broaden data diversity for model development and validation by engaging larger, multi-ethnic cohorts for improved external validity. Next steps should also involve merging multiple data sources, i.e., genomics, proteomics, and real-time metabolomics, to enhance model adaptability. Synthetic data and digital twin technologies may be seen as a means to protect patient confidentiality while expanding the scope of AI-driven research. Ultimately, bridging these advances into daily clinical care will require the above-mentioned implementation science framework, which supports remote patient monitoring, AI-augmented workflows, earlier detection, and more personalized management of cardiometabolic conditions.

Longitudinal validation in varied ethnic groups

A persistent limitation within biomedical AI and predictive modeling is the insufficient representation of diverse ethnic populations. For example, in **Chapters 2 and 3,** we demonstrated the limited number of individuals self-identifying as not White, whereas in **Chapters 4 and 5,** we were unable to validate our results in ethnic minorities due to the scarcity of available data. Therefore, the ethnic homogeneity of the UK Biobank and Lifelines raises concerns about the external validity and equity of AI-informed risk assessments among ethnically diverse populations (23). Specifically, the key question is what the impact is of the non-

availability of data, i.e., how much harm (in terms of misclassification or reduced discriminative abilities) is caused.

To tackle these issues, upcoming investigations should train and/or validate models in large-scale, multi-ethnic cohorts, such as the China Kadoorie Biobank and the Million Veteran Program (23). These groups provide long-term data from diverse populations, thereby extending the relevance of predictive models to a broad spectrum of genetic lineages and cultural influences, which may impact the prevalence and risk factors associated with metabolic syndrome. Moreover, federated learning methods, where data remains within the original institution while models are collaboratively trained, may help strengthen generalizability without compromising patient confidentiality (24).

Crucially, longitudinal validation must consider socioeconomic and environmental determinants of health (25, 26). Genetic factors alone cannot fully explain variations in disease occurrence and outcomes (27). Therefore, AI algorithms should be built using datasets that account for factors such as diet, geographic location, healthcare accessibility, and exposure to environmental stressors. Integrating these variables can shift models away from purely genetic correlations and toward a more comprehensive picture of disease risk. Specifically, we showed that self-reported questionnaire data can predict adverse cardiometabolic outcomes and should be considered in multimodal AI tools or as standalone tools. Even though self-reported ratings may be biased, human bias may contain valuable information when training AI models and should not be overlooked. For instance, results from two long-term longitudinal cohorts (with follow-up periods of 10 years and 30 years) of women and men showed that higher optimism levels were associated with a longer lifespan, as well as a higher likelihood of achieving exceptional longevity, defined as survival to the age of 85 or older (28). Therefore, while an optimistic disposition may introduce bias into self-reported health assessments, these findings suggest that more optimistic individuals may benefit from psychosocial or other nonbiological factors that enhance survival.

Furthermore, multi-center projects that gather repeated measures rather than single-timepoint snapshots will be necessary. The frequency of follow-up depends on the outcome studied. Many cardiometabolic conditions unfold over years or even decades due to both genetic and lifestyle influences. Lengthy follow-up designs will thus be crucial for refining models of disease development and confirming their applicability to diverse populations.

Multi-omics integration for enhanced predictive power

Multimodal approaches, which bridge genomics, epigenomics, proteomics, metabolomics, radiomics, and microbiome studies, have demonstrated an

unprecedented potential for innovating disease prediction and personalized therapies by collecting the maximum amount of information from subjects (29, 30). Current knowledge indicates that merging multiple omics layers may reveal a superior account of disease processes, making risk assessment and treatment targeting more accurate (29). However, given the cost and technical barriers, i.e., proteomic-enabled screening should be reserved for tertiary cardiometabolic centers (as described later) with infrastructure for liquid chromatography, tandem mass spectrometry, and clinical bioinformatics.

The main challenge involves harmonizing these datasets, which are frequently generated from distinct technologies and can feature varying degrees of variability (29). Notable examples of advanced Machine Learning approaches to integrate distinct omics data, such as deep learning architectures and specialized dimensionality reduction methods (i.e., PHATE and Multiscale PHATE), have the ability to learn abstract representations of biological and clinical data at different levels of granularity and can predict clinical outcomes (31, 32). In **Chapter 5**, we applied advanced Machine Learning harmonization via generative adversarial network (GAN) methods, a promising tool for improving cross-site deep learning generalization, to address this challenge in the Imaging-Based Coordinate System for Aging and Neurodegenerative Diseases (iSTAGING) study (33).

Additionally, integrating real-time metabolomic measurements from wearable sensors could supply clinicians with timely information on the course of disease (34). By coupling ongoing monitoring data with standard biomarkers, AI-powered models might produce adaptive risk evaluations that adjust according to a patient's immediate physiological status. Such innovations are particularly valuable for malleable disorders, i.e., cardiometabolic disorders, where early preventive intervention can have a substantial impact on disease progression.

Nonetheless, it is pertinent to highlight that turning multi-omics results into actionable insights demands standardized protocols. Laboratories often rely on different analysis workflows, leading to inconsistencies in interpretation that can affect cross-biobank validation. Establishing global benchmarks for multi-omics data handling and standardizing assays, together with improving data compatibility, will be pivotal for incorporating these insights into routine care.

To ensure clinical and operational feasibility, a key priority is to develop decision frameworks for selecting which omics layers to integrate based on clinical use case, marginal predictive gain, and infrastructure. For example, genomic or methylation data may be sufficient for predicting early-life risk, whereas time-sensitive management decisions in advanced disease may require input from proteomic or metabolomic data. Omics inclusion should be justified by incremental

utility, not data availability. Future studies should prioritize benchmarking efforts that systematically quantify the added value of each omics layer per phenotype or disease stage, identifying when layers become redundant or negligible. This may also help guide the omission of low-yield layers, minimizing unnecessary cost and complexity.

Synthetic data and Digital Twin technology: Promising recipes for overcoming privacy and generalizability challenges

As seen above, i.e., regarding ethnic diversity, a recurring obstacle in risk stratification is the quest for comprehensive datasets while safeguarding individual privacy. Synthetic data creation, especially via GANs and latent diffusion approaches, has emerged as a novel potential response (35). These methods produce artificial datasets while preserving the statistical features of original and genuine data, and at the same time mitigating privacy risks. Unlike traditional predictive models that rely on static datasets, digital twin technology longitudinally integrates real-time patient data from EHR, wearables, and remote monitoring systems to construct individualized, dynamic simulations. Thus, these virtual patient models replicate physiological and metabolic processes, enabling real-time risk stratification, treatment simulations, and outcome predictions (35). Nonetheless, synthetic data must undergo thorough evaluation to ensure it does not introduce inaccuracies or weaken model performance. Beyond privacy safeguards, synthetic data could enhance model resilience. AI systems trained solely on actual datasets often overfit, performing well in training but faltering when encountering new populations (36). A pertinent example is the recently introduced “RETFound”, a foundational model for interpreting retinal images, capable of identifying ophthalmic pathology and predicting systemic disease. By incorporating synthetic data to supplement real-world datasets, a later model, DERETFound, achieved comparable performance to RETFound while utilizing only 16.7% of the real data required by the original model (37, 38).

Notably, EHR, such as TriNetX, aggregate structured and unstructured data, including medical history, diagnostic imaging, laboratory results, and treatment responses. Such variables may serve as the primary source of real-world clinical data used to train, validate, and refine predictive models for digital twin technologies. So far, within cardiovascular research, digital twins have been suggested to predict disease progression and guide therapy selection (39). Digital twin technologies refer to AI-generated, patient-specific simulations replicating an individual’s clinical, metabolic, and genetic profile. These simulations allow clinicians and researchers to test different treatment strategies in a virtual environment prior to real-world implementation. Figure 5 conceptualizes an overarching framework for this approach, in the future, wherein multi-modal data, including metabolomics, proteomics, imaging, and genomic information,

are integrated into an AI model that generates a digital twin. Within this metaverse, various treatment scenarios can be simulated and compared to identify the intervention most likely to result in a favorable clinical response. Ideally, training data should also include comparative effectiveness outcomes of cardiometabolic treatments that extend beyond neurocognitive outcomes, along with data on adherence for per-protocol evaluations (40).

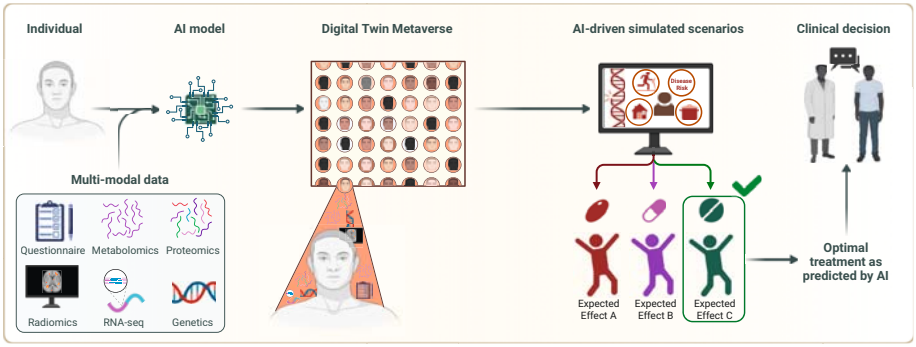


Figure 5. Digital Twin metaverse for AI-driven personalized cardiometabolic treatment prediction. The schematic illustrates an AI-driven approach to personalized cardiometabolic medicine using a digital twin framework. Multi-modal patient data, including questionnaire responses, metabolomics, proteomics, radiomics, RNA sequencing, and genetic information, are integrated into an AI model. This model generates a digital twin of the individual within a metaverse populated with diverse virtual representations. AI-driven simulations evaluate various treatment scenarios by incorporating disease risk factors and lifestyle parameters. The expected outcomes of different therapeutic interventions are compared, enabling the selection of an optimal treatment predicted to yield the most favorable response. The AI-derived recommendation informs the clinical decision-making process, facilitating personalized and evidence-informed patient management. Created with BioRender.com.

Specifically, digital twin methods utilize data from RCTs, extending the utility of RCT data by enabling the creation of synthetic cohorts, adjustment for prognostic covariates, and the simulation of treatment effects in new or real-world populations. This approach allows for addressing limitations observed in **Chapter 6** (i.e., predefined sensitivity analyses, missing adherence data, and suboptimal data for causal inference). In practice, digital twin platforms may benefit from EHR integration, as well as structured, longitudinal datasets with high-frequency inputs from wearable technologies that collect daily lifestyle trends, including sleep cycles, daily steps, and even metrics such as oxygen saturation and continuous glucose monitoring data. These are currently feasible mainly in academic hospitals with mature digital infrastructure for telemedicine and remote patient management. A phased pilot could prospectively enroll patients from outpatient clinics into a digital twin platform integrated with real-time EHR and wearable data, initially evaluating how accurately these models/technologies predict clinical outcomes and how readily healthcare providers incorporate digital twin insights into routine clinical practice. Subsequently, the pilot could expand to test

personalized treatment simulations and real-time patient monitoring in high-risk cardiometabolic populations.

The future of cardiometabolic medicine through the clinician's lens
Adapted from *Kokkorakis et al., Metabolism 148 (2023) 155689,* (41).

The future of healthcare must be optimized to deliver effective, multidisciplinary bariatric and cardiometabolic management (41). As obesity increasingly emerges as the central component of cardiometabolic medicine, the challenges of chronic weight management contribute to expanding the cardiometabolic patient population (41, 42). These individuals frequently present with multiple interconnected conditions, including T2D, CAD, chronic kidney disease, and metabolic dysfunction-associated steatotic liver disease (42, 43). Given the evolving understanding of obesity's systemic impact, the recently proposed subspecialty of cardiometabolic medicine underscores the necessity of a broader, more integrated specialty. This discipline encompasses lifestyle medicine, cardiovascular-kidney-liver-metabolic outcomes, and beyond (44), ensuring a holistic approach to disease prevention and management.

Given the rapid evolution of obesity pharmacotherapies and emerging evidence of multipurpose drugs, cardiometabolic care requires structural transformations akin to the oncological model. As we proposed previously, the ideal framework consists of a hierarchical network of specialized centers that provide standardized, evidence-based interventions (Figure 6). At the pinnacle of this structure, National Centers of Cardiometabolic Excellence will serve as hubs for research, innovation, policymaking, and training, fostering interdisciplinary collaboration. These centers will spearhead clinical trials, evaluate novel pharmacological and surgical interventions, and set standards for obesity management, ensuring optimal patient outcomes. Complementing these national centers, Regional Centers of Cardiometabolic Excellence will integrate expertise from endocrinology, cardiology, hepatology, nephrology, metabolic surgery, and clinical nutrition, thereby ensuring coordination of secondary and tertiary care.

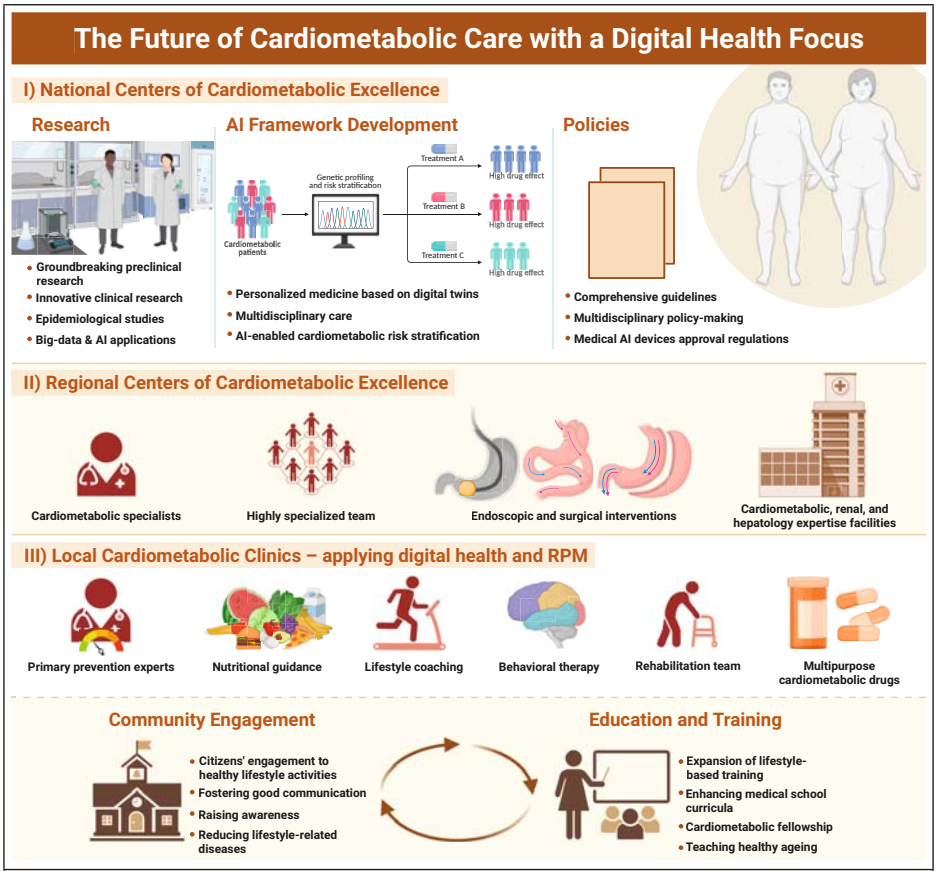


Figure 6. The proposed pyramid of care model for multidisciplinary management of cardiometabolic patients in the digital era.
Adapted from: *Kokkorakis et al., Metabolism 148 (2023) 155689*. Starting from primary prevention led by Local Cardiometabolic Clinics, progressing to secondary and tertiary care in Regional Centers of Cardiometabolic Excellence, and ultimately, to the National Centers of Cardiometabolic Excellence, which serve as powerhouses for research, comprehensive care, and policymaking. Abbreviation: AI, artificial intelligence; RPM, remote patient monitoring. Created with BioRender.com.

At the community level, Local Cardiometabolic Clinics will serve as the foundation for primary prevention, early detection, and chronic disease management. These centers will provide lifestyle-based weight management, transitioning to pharmacological interventions as needed. While limited in resources compared to national and regional centers, local clinics will play a crucial role in disseminating cardiometabolic education, promoting healthy aging, and incorporating emerging digital health solutions, as outlined above, such as digital twin technology and AI-powered decision support systems.

In the absence of evidence from RCTs, this thesis provides an evidence-*informed* approach in cardiometabolic medicine. Following this paradigm, the involvement of

this novel subspecialty of cardiometabolic medicine with a strong focus on digital medicine technologies is inevitable to achieve the above-mentioned pyramid of care. Although the pragmatic establishment of nationwide cardiometabolic and digital health multidisciplinary centers may seem an unfeasible goal in the short term, the continuous striving to promote comprehensive and multidisciplinary care may prove preferable and more effective than delayed perfection.

Conclusion

In this thesis, we studied large-scale EHR data, developed questionnaire-based screening tools, investigated omics profiling to refine the measurement and prediction of cardiometabolic risk and its associated outcomes, and conducted a RWE study. Our work underscores both the potential and the current challenges of applying AI-driven tools in diverse clinical contexts within the field of cardiometabolic medicine. By demonstrating the viability of non-laboratory questionnaires and illustrating how novel biomarkers and omics can enhance predictive models, we highlight pathways that could inform earlier and more targeted interventions. Moving forward, implementing inclusive AI frameworks will require multi-ethnic validation, multi-omics data, and synthetic data. Digital twin technologies and AI-driven remote patient management will enable the delivery of personalized cardiometabolic care in routine clinical practice. Ultimately, the management of cardiometabolic diseases should shift away from a compartmentalized, specialty-driven approach and move toward an integrated, multidisciplinary care model. Next-generation healthcare systems should incorporate the above technological frameworks to effectively manage the escalating prevalence of cardiometabolic diseases. In this context, *evidence-informed* medicine represents a viable strategy in an era marked by increasing workforce shortages and a concerning worldwide expansion in chronic disease prevalence.

References

1. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet. 2014;383(9921):999-1008.
2. Kromhout D. Serum cholesterol in cross-cultural perspective. The Seven Countries Study. Acta Cardiol. 1999;54(3):155-8.
3. Menotti A, Lanti M, Kromhout D, Blackburn H, Jacobs D, Nissinen A, et al. Homogeneity in the relationship of serum cholesterol to coronary deaths across different cultures: 40-year follow-up of the Seven Countries Study. Eur J Cardiovasc Prev Rehabil. 2008;15(6):719-25.
4. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. 1998;97(18):1837-47.
5. Pylypchuk R, Wells S, Kerr A, Poppe K, Riddell T, Harwood M, et al. Cardiovascular disease risk prediction equations in 400 000 primary care patients in New Zealand: a derivation and validation study. Lancet. 2018;391(10133):1897-907.
6. Thomas DM, Bredlau C, Bosy-Westphal A, Mueller M, Shen W, Gallagher D, et al. Relationships between body roundness with body fat and visceral adipose tissue emerging from a new geometrical model. Obesity (Silver Spring). 2013;21(11):2264-71.
7. Bešević J, Lacey B, Conroy M, Omiyale W, Feng Q, Collins R, et al. New Horizons: the value of UK Biobank to research on endocrine and metabolic disorders. J Clin Endocrinol Metab. 2022;107(9):2403-10.
8. Ding JE, Thao PNM, Peng WC, Wang JZ, Chug CC, Hsieh MC, et al. Large language multimodal models for new-onset type 2 diabetes prediction using five-year cohort electronic health records. Sci Rep. 2024;14(1):20774.
9. Katz U, Cohen E, Shachar E, Somer J, Fink A, Morse E, et al. GPT versus Resident Physicians — A Benchmark Based on Official Board Scores. NEJM AI. 2024;1(5):AIdbp2300192.
10. Goh E, Gallo RJ, Strong E, Weng Y, Kerman H, Freed JA, et al. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. Nature Medicine. 2025;31(4):1233-8.
11. Rakers M, van Hattem N, Plag S, Chavannes N, van Os HJA, Vos RC. Population health interventions for cardiometabolic diseases in primary care: a scoping review and RE-AIM evaluation of current practices. Front Med (Lausanne). 2023;10:1275267.
12. Sethi Y, Patel N, Kaka N, Kaiwan O, Kar J, Moinuddin A, et al. Precision Medicine and the future of Cardiovascular Diseases: A Clinically Oriented Comprehensive Review. J Clin Med. 2023;12(5).
13. Kidenya BR, Mboowa G. Unlocking the future of complex human diseases prediction: multi-omics risk score breakthrough. Front Bioinform. 2024;4:1510352.
14. Dang A. Real-World Evidence: A Primer. Pharmaceut Med. 2023;37(1):25-36.
15. Chatterjee S, Peters SA, Woodward M, Mejia Arango S, Batty GD, Beckett N, et al. Type 2 Diabetes as a Risk Factor for Dementia in Women Compared With Men: A Pooled Analysis of 2.3 Million People Comprising More Than 100,000 Cases of Dementia. Diabetes Care. 2016;39(2):300-7.
16. Tan SY, Sumner J, Wang Y, Wenjun Yip A. A systematic review of the impacts of remote patient monitoring (RPM) interventions on safety, adherence, quality-of-life and cost-related outcomes. npj Digital Medicine. 2024;7(1):192.

17. Lyzwinski L, Elgendi M, Shokurov AV, Cuthbert TJ, Ahmadizadeh C, Menon C. Opportunities and challenges for sweat-based monitoring of metabolic syndrome via wearable technologies. Communications Engineering. 2023;2(1):48.
18. Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nat Mach Intell. 2019;1(5):206-15.
19. Coffey JD, Christopherson LA, Williams RD, Gathje SR, Bell SJ, Pahl DF, et al. Development and implementation of a nurse-based remote patient monitoring program for ambulatory disease management. Front Digit Health. 2022;4:1052408.
20. Zhou K, Gattinger G. The Evolving Regulatory Paradigm of AI in MedTech: A Review of Perspectives and Where We Are Today. Ther Innov Regul Sci. 2024;58(3):456-64.
21. Park SH, Choi J, Byeon JS. Key Principles of Clinical Validation, Device Approval, and Insurance Coverage Decisions of Artificial Intelligence. Korean J Radiol. 2021;22(3):442-53.
22. Chen Y, Lehmann CU, Malin B. Digital Information Ecosystems in Modern Care Coordination and Patient Care Pathways and the Challenges and Opportunities for AI Solutions. J Med Internet Res. 2024;26:e60258.
23. Lewandowski AJ, Rutter MK, Collins R. Scientific and Clinical Impacts of UK Biobank in Cardiovascular Medicine. Circulation. 2024;150(24):1907-9.
24. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. npj Digital Medicine. 2020;3(1):119.
25. Aanegola R, Nakamura Sakai S, Kumar N. Longitudinal analysis of the determinants of life expectancy and healthy life expectancy: A causal approach. Healthcare Analytics. 2022;2:100028.
26. Kambeitz J, Meyer-Lindenberg A. Modelling the impact of environmental and social determinants on mental health using generative agents. npj Digital Medicine. 2025;8(1):36.
27. Hingorani AD, Gratton J, Finan C, Schmidt AF, Patel R, Sofat R, et al. Performance of polygenic risk scores in screening, prediction, and risk stratification: secondary analysis of data in the Polygenic Score Catalog. BMJ Med. 2023;2(1):e000554.
28. Lee LO, James P, Zevon ES, Kim ES, Trudel-Fitzgerald C, Spiro A, 3rd, et al. Optimism is associated with exceptional longevity in 2 epidemiologic cohorts of men and women. Proc Natl Acad Sci U S A. 2019;116(37):18357-62.
29. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nature Medicine. 2022;28(9):1773-84.
30. Muse ED, Topol EJ. Transforming the cardiometabolic disease landscape: Multimodal AI-powered approaches in prevention and management. Cell Metab. 2024;36(4):670-83.
31. Moon KR, van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, et al. Visualizing structure and transitions in high-dimensional biological data. Nature Biotechnology. 2019;37(12):1482-92.
32. Kuchroo M, Huang J, Wong P, Grenier J-C, Shung D, Tong A, et al. Multiscale PHATE identifies multimodal signatures of COVID-19. Nature Biotechnology. 2022;40(5):681-91.
33. Bashyam VM, Doshi J, Erus G, Srinivasan D, Abdulkadir A, Singh A, et al. Deep Generative Medical Image Harmonization for Improving Cross-Site Generalization in Deep Learning Predictors. J Magn Reson Imaging. 2022;55(3):908-16.
34. Wang M, Yang Y, Min J, Song Y, Tu J, Mukasa D, et al. A wearable electrochemical biosensor for the monitoring of metabolites and nutrients. Nature Biomedical Engineering. 2022;6(11):1225-35.

35. Arora A, Wagner SK, Carpenter R, Jena R, Keane PA. The urgent need to accelerate synthetic data privacy frameworks for medical research. Lancet Digit Health. 2025;7(2):e157-e60.
36. Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health. 2023;2(6):e0000278.
37. Sun Y, Tan W, Gu Z, He R, Chen S, Pang M, et al. A data-efficient strategy for building high-performing medical foundation models. Nat Biomed Eng. 2025;9(4):539-51.
38. Zhou Y, Chia MA, Wagner SK, Ayhan MS, Williamson DJ, Struyven RR, et al. A foundation model for generalizable disease detection from retinal images. Nature. 2023;622(7981):156-63.
39. Coorey G, Figtree GA, Fletcher DF, Snelson VJ, Vernon ST, Winlaw D, et al. The health digital twin to tackle cardiovascular disease-a review of an emerging interdisciplinary field. NPJ Digit Med. 2022;5(1):126.
40. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.
41. Kokkorakis M, Katsarou A, Katsiki N, Mantzoros CS. Milestones in the journey towards addressing obesity; Past trials and triumphs, recent breakthroughs, and an exciting future in the era of emerging effective medical therapies and integration of effective medical therapies with metabolic surgery. Metabolism. 2023;148:155689.
42. Kokkorakis M, Chakhtoura M, Rhayem C, Al Rifai J, Ghezzawi M, Valenzuela-Vallejo L, et al. Emerging pharmacotherapies for obesity: A systematic review. Pharmacol Rev. 2025;77(1):100002.
43. Kokkorakis M, Muzurović E, Volčanšek Š, Chakhtoura M, Hill MA, Mikhailidis DP, et al. Steatotic Liver Disease: Pathophysiology and Emerging Pharmacotherapies. Pharmacol Rev. 2024;76(3):454-99.
44. Wong ND. Preventive cardiology or cardiometabolic medicine: a new and emerging subspecialty? Cardiovasc Endocrinol Metab. 2020;9(3):66-9.

Appendices

English summary
Nederlandse samenvatting
Acknowledgements
Curriculum vitae
List of publications

English summary

Cardiometabolic diseases, comprising type 2 diabetes, obesity, cardiovascular disease, and metabolic dysfunction-associated steatotic liver disease, are the leading cause of global morbidity and mortality. Although early risk identification and targeted interventions have long been advocated, the current clinical framework remains reactive and fragmented. At the same time, the rapid rise of digital health technologies, i.e., remote patient monitoring and AI-driven models, opens new avenues to shift cardiometabolic care from episodic to continuous, personalized, and prevention-focused.

This thesis investigates how digital tools, novel biomarkers, and AI-informed methods can enable earlier, more equitable, and personalized cardiometabolic risk stratification. To this end, we utilized several data sources from large population-based cohorts, including Lifelines, UK Biobank, the iSTAGING consortium, and the TriNetX platform.

In **Part I**, we investigated **low-cost risk stratification models** that do not require laboratory measurements. In **Chapter 2**, we developed and validated a simple, non-laboratory risk score using self-reported variables to predict incident type 2 diabetes. The models demonstrated higher discriminatory accuracy than clinically available non-laboratory tools. In **Chapter 3**, we extended this framework to coronary artery disease, demonstrating that such models can stratify risk as accurately as clinical tools using blood-based markers and physical measurements, using variables obtainable from digital tools or remote consultations.

In **Part II**, we evaluated **emerging biomarkers**, including proteomics and radiomics. In **Chapter 4**, we investigated how proteomic biomarkers repurpose and improve current non-invasive indices of steatotic liver disease to predict future cardiometabolic and other adverse outcomes. In **Chapter 5**, we integrated neuroimaging data with multi-omics using advanced Machine Learning techniques to analyze the influence of cardiometabolic conditions on brain aging and brain atrophy.

In **Part III**, we explored the **comparative effectiveness of cardiometabolic drugs for dementia, a less well-studied outcome in clinical trials**. Specifically, in **Chapter 6**, we conducted a real-world retrospective multicohort study to assess the long-term comparative effectiveness of glucose-lowering treatments on neurocognitive outcomes. The findings do not report significant differences between the two major secondary antihyperglycemic drug classes.

In **Chapter 7**, I synthesized these insights into a roadmap for **real-world implementation** and discussed **future steps in evidence-informed cardiometabolic medicine**. Emphasis should be placed on equitable models trained on diverse datasets, multimodal AI systems, and digital twin technology.

Nederlandse samenvatting

Cardiometabole aandoeningen, waaronder diabetes type 2, obesitas, cardiovasculaire ziekten en metabool disfunctie-geassocieerde steatotische leverziekte, vormen wereldwijd de belangrijkste oorzaak van morbiditeit en mortaliteit. Hoewel vroegtijdige risicodetectie en gerichte interventies al lange tijd worden benadrukt, blijft het huidige klinische kader reactief en gefragmenteerd. Tegelijkertijd bieden de snelle opkomst van digitale gezondheidstechnologieën – zoals telemonitoring en door AI-modellen – nieuwe mogelijkheden om de cardiometabole zorg te verschuiven van episodisch naar continu, gepersonaliseerd en gericht op preventie.

Dit proefschrift onderzoekt hoe digitale instrumenten, nieuwe biomarkers en AI-geïnformeerde methoden kunnen bijdragen aan eerdere, meer rechtvaardige en meer gepersonaliseerde risicostratificatie van cardiometabole ziekten. Hiervoor maakten we gebruik van diverse databronnen uit grootschalige populatiecohorten, waaronder Lifelines, UK Biobank, het iSTAGING-consortium en het TriNetX-platform.

Deel I behandelt **risicostratificatiemodellen zonder laboratoriummetingen**. In **Hoofdstuk 2** ontwikkelden en valideerden we een eenvoudige risicoscore op basis van zelfgerapporteerde variabelen voor de voorspelling van incidentie van diabetes type 2. Deze modellen vertoonden een hogere discriminerende nauwkeurigheid dan de klinisch beschikbare niet-laboratoriummodellen. In **Hoofdstuk 3** breidden we dit raamwerk uit naar coronaire hartziekte en toonden we aan dat vergelijkbare modellen, met variabelen makkelijk verkrijgbaar vanuit digitale tools of digitale consulten, het risico net zo nauwkeurig konden inschatten als klinische tools gebaseerd op bloedmarkers en fysieke metingen.

Deel II richt zich op **opkomende biomarkers**, waaronder proteomics en radiomics. In **Hoofdstuk 4** onderzochten we hoe proteomics bestaande niet-invasieve indices van steatotische leverziekte kunnen herbestemmen en verbeteren voor de voorspelling van toekomstige cardiometabole en andere ongunstige uitkomsten. In **Hoofdstuk 5** integreerden we neuro-imaginggegevens met multi-omics door gebruik van geavanceerde machineleertechnieken om de invloed van cardiometabole aandoeningen op hersenveroudering en hersenatrofie te analyseren.

Deel III bespreekt de **vergelijkende effectiviteit van cardiometabole geneesmiddelen bij dementie, een in klinische studies minder bestudeerde uitkomst**. In **Hoofdstuk 6** voerden we een retrospectieve multicohortstudie uit in een real-world setting om de langetermijn-effectiviteit van glucoseverlagende

behandelingen op neurocognitieve uitkomsten te beoordelen. De resultaten laten geen significante verschillen zien tussen de twee belangrijkste secundaire klassen van antihyperglycemische middelen.

In **Hoofdstuk 7** heb ik deze bevindingen samengebracht in een **routekaart voor implementatie in de praktijk** en heb ik **toekomstige stappen in evidence-informed cardiometabole geneeskunde** besproken. De nadruk ligt op rechtvaardige modellen getraind op diverse datasets, multimodale AI-systemen en digital twin–technologie.

Acknowledgements

In my pursuit of knowledge, I have been fortunate to receive support, mentorship, and encouragement from many people. I am deeply grateful to all who have shaped my growth and contributed to my journey.

“If I have seen further, it is by standing on the shoulders of giants.”
~ Isaac Newton

Curriculum vitae

Michail Kokkorakis was born on February 19th, 2001, in Amarousio, Greece, and raised in Athens. He then began medical school at the University of Groningen in 2019. During his studies, he served as an elected member of the University Council (2021-2022), completed the Junior Scientific Masterclass Honors program, and was selected as the student commencement speaker for the BSc in Medicine graduation ceremony. He later secured a personal MD-PhD grant to study cardiometabolic medicine and digital health. He obtained his MD in September 2025 after completing his final clinical internships in Cardiology and Endocrinology at Leiden University Medical Center. While completing both his MD and PhD degrees in six years, he published 20 papers, accrued over 800 citations, and was nominated as one of the top three students of the Faculty of Medical Sciences for the 2024 Groningen University Fund (GUF) 100 Award. In recognition of his academic involvement, he was recently appointed Associate Editor of the journal Diabetology & Metabolic Syndrome. In July 2025, he was awarded the Rubicon Grant from the Dutch Research Council (NWO) and The Netherlands Organisation for Health Research and Development (ZonMw), one of the most prestigious early-career grants in the Netherlands, fully funding a two-year postdoctoral fellowship at Yale School of Medicine Section of Digestive Diseases, starting in late 2025.

List of publications

1. Anagnostakis F, **Kokkorakis M**, Asvestis C, Papadimopoulos I, Nagarajan S, Talbot K, Li L, Chen Y, Nasrallah IM, Wen J, Davatzikos C (2025) Impact of cardiometabolic conditions on the progression from mild cognitive impairment to dementia: a large cohort study. *Alzheimer's & Dementia* (accepted manuscript).
2. Global, regional, and national trends in routine childhood vaccination coverage from 1980 to 2023 with forecasts to 2030: a systematic analysis for the Global Burden of Disease Study 2023. *The Lancet* 406(10500):235-260.
3. Anagnostakis F, **Kokkorakis M**, Walker KA, Skampardoni I, Wen J, Erus G, Tosun D, Tassopoulou V, Cui Y, Govindarajan ST, Srinivasan D, Melhem R, Mamourian E, Shou H, Nasrallah IM, Mantzoros CS, and Davatzikos C (2025) Radiomic and proteomic signatures of body mass index on brain ageing and Alzheimer's-like patterns of brain atrophy. *EBioMedicine* 116:105763.
4. **Kokkorakis M***, Folkertsma P*, Anagnostakis F*, Sirotin N, Agarwal M, Shantouf R, Henning RH, Pijl H, Wolffenbuttel BHR, Bax JJ, Atsma DE, Forte JC, Mantzoros CS[§], and van Dam S[§] (2025) Simplifying coronary artery disease risk stratification: development and validation of a questionnaire-based alternative comparable to clinical risk tools. *EBioMedicine* 111:105518.
5. **Kokkorakis M***, Folkertsma P*, Castela Forte J, Wolffenbuttel BHR, van Dam S, and Mantzoros CS (2025) GDF-15 improves the predictive capacity of steatotic liver disease non-invasive tests for incident morbidity and mortality risk for cardio-renal-metabolic diseases and malignancies. *Metabolism* 163:156047.
6. GBD 2021 Adolescent BMI Collaborators (2025) Global, regional, and national prevalence of child and adolescent overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. *The Lancet* 405(10481):785-812.
7. GBD 2021 Adult BMI Collaborators (2025) Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. *The Lancet* 405(10481):813-838.
8. **Kokkorakis M***, Chakhtoura M*, Rhayem C[§], Al Rifai J[§], Ghezzawi M, Valenzuela-Vallejo L, and Mantzoros CS (2024) Emerging Pharmacotherapies for Obesity: A Systematic Review. *Pharmacological Reviews* 77:100002.
9. Stefanakis K*, **Kokkorakis M***, and Mantzoros CS (2024) The impact of weight loss on fat-free mass, muscle, bone and hematopoiesis health: Implications for emerging pharmacotherapies aiming at fat reduction and lean mass preservation. *Metabolism* 161:156057.
10. **Kokkorakis M***, Boutari C*, Hill MA, Kotsis V, Loomba R, Sanyal AJ, and Mantzoros CS (2024) Resmetirom, the first approved drug for the management of metabolic dysfunction-associated steatohepatitis: Trials, opportunities, and challenges. *Metabolism* 154:155835.
11. **Kokkorakis M**, Muzurović E, Volčanšek Š, Chakhtoura M, Hill MA, Mikhailidis DP, and Mantzoros CS (2024) Steatotic Liver Disease: Pathophysiology and Emerging Pharmacotherapies. *Pharmacological Reviews* 76:454–499.

12. Kim K*, Kim MM*, Skoufos G, Diffenderfer ES, Motlagh SAO, **Kokkorakis M**, Koliaki I, Morcos G, Shoniyozov K, Griffin J, Hatzigeorgiou AG, Metz JM, Lin A, Feigenberg SJ, Cengel KA, Ky B, Koumenis C, and Verginadis II (2024) FLASH proton radiotherapy mitigates inflammatory and fibrotic pathways and preserves cardiac function in a preclinical mouse model of radiation-induced heart disease. *International Journal of Radiation Oncology, Biology, Physics* 119(4):1234-1247.
13. Anagnostakis F, **Kokkorakis M**, Walker K, and Davatzikos C (2024) Signatures and Discriminative Abilities of Multi-Omics between States of Cognitive Decline. *Biomedicines* 12(5):941.
14. **Kokkorakis M**, Folkertsma P, van Dam S, Sirotin N, Taheri S, Chagoury O, Idaghdour Y, Henning RH, Castela Forte J, Mantzoros CS, de Vries DH*, Wolffenbuttel BHR* (2023) Effective questionnaire-based prediction models for type 2 diabetes across several ethnicities: a model development and validation study. *EClinicalMedicine* 64:102235.
15. **Kokkorakis M***, Katsarou A*, Katsiki N, and Mantzoros CS (2023) Milestones in the journey towards addressing obesity; Past trials and triumphs, recent breakthroughs, and an exciting future in the era of emerging effective medical therapies and integration of effective medical therapies with metabolic surgery. *Metabolism* 148:155689.
16. **Kokkorakis M***, Boutari C*, Katsiki N, and Mantzoros CS (2023) From non-alcoholic fatty liver disease (NAFLD) to steatotic liver disease (SLD): an ongoing journey towards refining the terminology for this prevalent metabolic condition and unmet clinical need. *Metabolism* 147:155664.
17. Boutari C, **Kokkorakis M**, Stefanakis K, Valenzuela-Vallejo L, Axarloglou E, Volčanšek Š, Chakhtoura M, and Mantzoros CS (2023) Recent research advances in metabolism, clinical and experimental. *Metabolism* 149:155722.
18. Kouvari M*, Mylonakis SC*, Katsarou A*, Valenzuela-Vallejo L, Guatibonza-Garcia V, **Kokkorakis M**, Verrastro O, Angelini G, Markakis G, Eslam M, George J, Papatheodoridis G, Mingrone G, and Mantzoros CS (2023) The first external validation of the Dallas steatosis index in biopsy-proven Non-alcoholic fatty liver Disease: A multicenter study. *Diabetes Research and Clinical Practice* 203:110870.
19. Kouvari M*, Valenzuela-Vallejo L*, Guatibonza-Garcia V*, Polyzos SA, Deng Y, **Kokkorakis M**, Agraz M, Mylonakis SC, Katsarou A, Verrastro O, Markakis G, Eslam M, Papatheodoridis G, George J, Mingrone G, and Mantzoros CS (2023) Liver biopsy-based validation, confirmation and comparison of the diagnostic performance of established and novel non-invasive steatotic liver disease indexes: Results from a large multi-center study. *Metabolism* 147:155666.
20. Anagnostakis F*, **Kokkorakis M***, Markouli M[§], and Piperi C[§] (2023) Impact of Solute Carrier Transporters in Glioma Pathology: A Comprehensive Review. *International Journal of Molecular Sciences* 24(11):9393.

*,[§] Indicate equal contribution.