




OPEN ACCESS

Feasibility of multiorgan risk prediction with routinely collected diagnostics: a prospective cohort study in the UK Biobank

Celeste McCracken ¹, Zahra Raisi-Estabragh,^{2,3}
Liliana Szabo,^{2,3,4} Michele Veldsman,⁵ Betty Raman,¹
Any Topiwala,⁶ Adriana Roca-Fernández,⁷ Masud Husain,^{5,8,9}
Steffen E Petersen,^{2,3,10,11} Stefan Neubauer,¹
Thomas E Nichols^{6,8}

10.1136/bmjebm-2023-112518

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjebm-2023-112518>).

For numbered affiliations see end of article.

Correspondence to:

Celeste McCracken, Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK; celeste.mccracken@cardiov.ox.ac.uk



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY. Published by BMJ.

To cite: McCracken C, Raisi-Estabragh Z, Szabo L, et al. *BMJ Evidence-Based Medicine* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjebm-2023-112518

Abstract

Objectives Despite rising rates of multimorbidity, existing risk assessment tools are mostly limited to a single outcome of interest. This study tests the feasibility of producing multiple disease risk estimates with at least 70% discrimination (area under the receiver operating curve, AUROC) within the time and information constraints of the existing primary care health check framework.

Design Observational prospective cohort study
Setting UK Biobank.

Participants 228 240 adults from the UK population.

Interventions None.

Main outcome measures Myocardial infarction, atrial fibrillation, heart failure, stroke, all-cause dementia, chronic kidney disease, fatty liver disease, alcoholic liver disease, liver cirrhosis and liver failure.

Results Using a set of predictors easily gathered at the standard primary care health check (such as the National Health Service Health Check), we demonstrate that it is feasible to simultaneously produce risk estimates for multiple disease outcomes with AUROC of 70% or greater. These predictors can be entered once into a single form and produce risk scores for stroke (AUROC 0.727, 95% CI 0.713 to 0.740), all-cause dementia (0.823, 95% CI 0.810 to 0.836), myocardial infarction (0.785, 95% CI 0.775 to 0.795), atrial fibrillation (0.777, 95% CI 0.768 to 0.785), heart failure (0.828, 95% CI 0.818 to 0.838), chronic kidney disease (0.774, 95% CI 0.765 to 0.783), fatty liver disease (0.766, 95% CI 0.753 to 0.779), alcoholic liver disease (0.864, 95% CI 0.835 to 0.894), liver cirrhosis (0.763, 95% CI 0.734 to 0.793) and liver failure (0.746, 95% CI 0.695 to 0.796).

Conclusions Easily collected diagnostics can be used to assess 10-year risk across multiple disease outcomes, without the need for specialist computing or invasive biomarkers. Such an approach could increase the utility of existing data and place multiorgan risk information at the fingertips of primary care providers, thus creating opportunities for longer-term multimorbidity prevention. Additional work is needed to validate whether these findings would hold in a larger, more representative cohort outside the UK Biobank.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Primary care health checks (like the National Health Service Health Check) present a crucial opportunity to assess underlying cardiovascular risk and to intervene to prevent or delay longer-term cardiovascular disease. Widely validated risk tools such as QRISK3 enable cardiovascular risk to be calculated easily at that appointment and inform targeted decision-making. There are validated risk scores to profile risk for other diseases, but there is not enough time during the health check to gather the various risk score inputs and handle separate calculator tools.

WHAT THIS STUDY ADDS

⇒ In this study, we show that information already being collected as part of the primary care health check could feasibly be combined into a single calculator providing 10-year risk estimates for multiple diseases across related organ systems of heart, brain, liver and kidney. Moreover, much of the essential information can be acquired remotely.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ When patients attend their health check, they could potentially receive risk scores for multiple disease outcomes in addition to cardiovascular risk. Having earlier access to multiorgan information has the potential to enable earlier intervention for risk factors, more targeted use of resources and more effective multimorbidity prevention.

Introduction

Multimorbidity presents an urgent and increasing health challenge for ageing populations,¹ with implications for health equity, disability and healthcare costs.²⁻³ Experts warn that effective handling of multimorbidity will require a multisystem approach⁴ prioritising proactive, rather than reactive, care,⁵⁻⁶ with primary care taking a leading role in chronic disease prevention.⁷⁻⁹

Assessment of atherosclerotic cardiovascular disease (CVD) risk is central to primary care and is now quick and easy due to widely available risk calculator tools such as QRISK3¹⁰ and Framingham Risk Score.¹¹ In the UK, primary care risk assessment has been codified in the form of the National Health Service (NHS) Health Check.¹²⁻¹³ At the health check, a number of clinical parameters are collected, CVD risk is assessed and the general practitioner directs personalised interventions that influence the long-term health trajectory of the patient. Despite urgent calls for more preventative attention to other diseases,¹⁴⁻¹⁵ there are currently no existing methods for multidisease risk prediction in primary care.

The primary targets of the NHS Health Check are heart disease, diabetes, stroke, dementia, kidney and liver disease, as laid out in the official guidance,¹² website¹⁶ and patient information.¹⁷ These conditions are known to share underlying mechanisms¹⁸⁻²⁰ and to co-occur in multimorbidity clusters.²¹⁻²³

The objective of this study is to examine the feasibility of expanding the primary care health check to include risk assessment across multiple diseases. We focus on the 10 most commonly occurring serious conditions across the heart, brain, kidney and liver, namely, myocardial infarction, atrial fibrillation, heart failure, stroke, all-cause dementia, chronic kidney disease, fatty liver disease, alcoholic liver disease, liver cirrhosis and liver failure. Having access to a wider panel of risk information could

lead to earlier disease detection, more targeted interventions and more effective prevention of longer-term multimorbidity.

However, there are several important challenges to consider. First, although risk scores have previously been developed for each additional condition (eg, dementia) there is simply not enough time within a 10–15 min consultation to gather all the required inputs and to calculate each risk score separately.²⁴ A preferred solution would involve a single pool of inputs, and a single data entry page, from which multiple risk estimates could be calculated simultaneously.

Second, individual risk scores differ by the people they exclude, depending on the cohort in which they were developed.²⁵ This leads to shifting sets of calculators (and required inputs) in the hands of the physician depending on the existing comorbidities of the patient. Instead, future solutions would include a person's medical history and existing diagnoses, and adjust risk estimates accordingly.²⁶⁻²⁷

Third, not all health measures are equally accessible. NHS England is actively exploring ways that remote healthcare solutions can be used effectively to ease health service usage and make primary care services more accessible to all.²⁸⁻³⁰ These objectives call us to reflect on the information that is easily obtained and consider whether simple metrics can be potentially powerfully combined.

In this study, we use the UK Biobank data resource to emulate the information available within the primary care setting. Our objective is to explore the feasibility of multidisease risk estimation with easily collected diagnostics (figure 1), setting the minimum acceptable performance of 0.70 area under the receiver operating curve (AUROC) across all outcomes (as per Fagerland and, Hosmer, p177³¹). We begin by evaluating a range of published

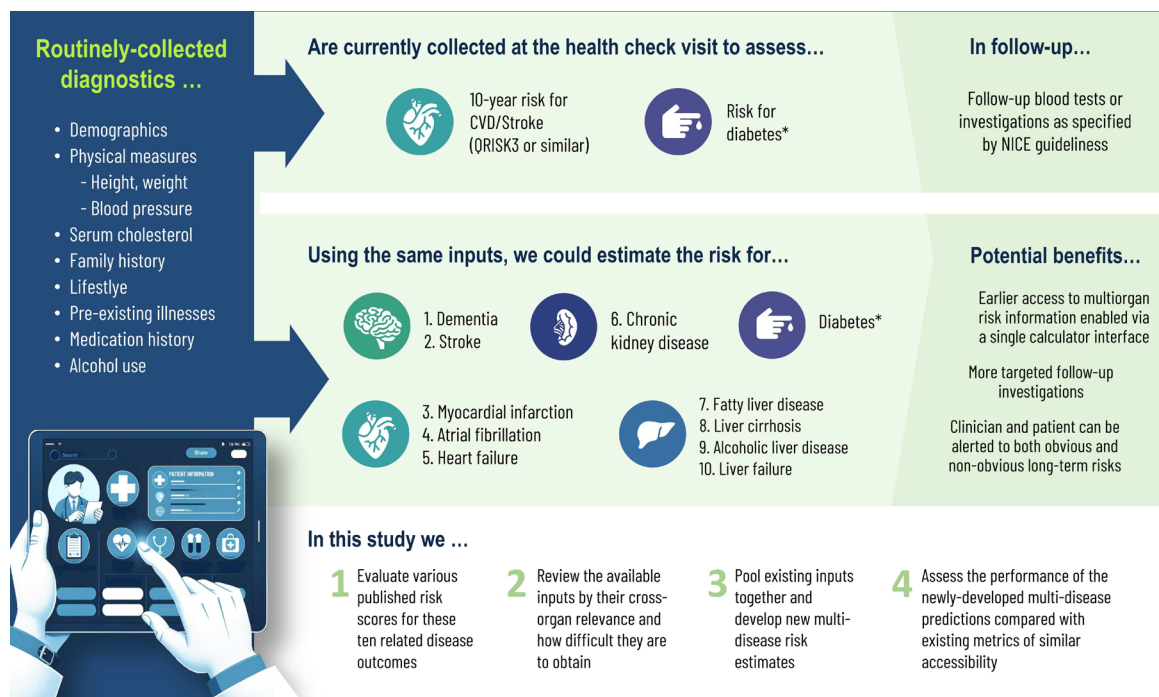


Figure 1 What can existing information tell us about multiorgan disease risk? In the context of increasing multimorbidity, we examine the feasibility of extending the existing primary care health check framework to include risk assessment for multiple diseases within the heart-brain-liver-kidney cluster. We evaluate a range of existing risk scores and consider whether and how they could be blended, and whether information already being collected could be effectively reused. If successful, this expansion could lead to earlier disease detection, more effective prevention and better resource allocation for multimorbidity prevention. *The diabetes screening component of the NHS Health Check protocol is not part of this analysis and would exist unchanged in both versions. CVD, cardiovascular disease; NICE, National Institute for Health and Care Excellence; NHS, National Health Service.

risk scores and assessing their performance in the UK Biobank cohort. We review the component inputs for each of those risk indices and identify scores that can be applied fully remotely (ie, without direct in-person contact) and those with a standard set of in-person inputs. Finally, we reuse the standard set of inputs to develop new risk scores compare their performance with existing risk scores.

Methods

Setting and study population

UK Biobank is a large prospective cohort study with participants drawn from the general population.³² UK residents aged 40–69 years old who are registered with a general practitioner, as identified from NHS registers, were invited to participate. Baseline data collection took place between 2006 and 2010, where registration date was used as the index date for the study. Follow-up events were ascertained via linked health records with latest censor date of 31 October 2022. Comprehensive details regarding linked primary care and hospital records are provided in official UK Biobank resources.^{33 34} To focus on 10-year risk estimation for all outcomes, follow-up was truncated at 10 years following baseline, giving a median follow-up time of 10 years (IQR=10–10). The clinical and demographic parameters collected in UK Biobank mimic those available in primary care and permit population-based modelling of multiorgan risk from baseline features.

From the overall UK Biobank cohort (n=502 386), 1298 participants were removed due to self-withdrawal or loss to follow-up, and 271 386 participants did not have primary care data available. There were 229 702 remaining participants who were confirmed to be present in both the main dataset and the primary care dataset. From these, a further 1462 participants were excluded due to missing values for height, weight, waist or hip circumference, leaving a final sample of 228 240 participants (see online supplemental figure 1A and online supplemental methods SM2).

NHS Health Check and easily collected diagnostics

The NHS Health Check is a preventative primary care initiative (<https://www.healthcheck.nhs.uk/>)¹² that forms the situational anchor for our study. Briefly, healthy people aged between 40 and 74 years are invited to visit their primary care team, where an inexpensive set of diagnostics are collected and 10-year risk of CVD is calculated using the widely validated QRISK3 calculator¹⁰ (<https://qrisk.org/three/>) or similar tool (figure 1).

Simple self-reported measures such as age, sex, family history, lifestyle factors, current medications and medical history are features that can be reported verbally and can be collected fully remotely (ie, without in-person contact). Physical measures such as height, weight, waist and hip circumference are easily measured without specialised technology. These are categorised as ‘remote features’ and are shown in the first two columns of online supplemental table 1. The term ‘remote’ is used to convey that remote collection of these parameters is possible, whether by phone or via an online form. Remote features can also be collected in person as part of the primary care visit.

Best practice guidelines¹² specify a minimum set of parameters to be collected as part of the standard NHS Health Check protocol, consistent with the use of QRISK3. These include a number of remote features, with the addition of blood pressure measurement and blood tests for total and high-density lipoprotein (HDL) cholesterol. In this study, we include all remote features plus the required NHS Health Check parameters (blood pressure and cholesterol) under the category of ‘standard features’.

Finally, there are various blood/biochemistry tests that are widely available but are not part of the existing first-line health check protocol. These measures were identified based on their inclusion in existing research or risk scores (see ‘Existing risk scores’ below) and have been included as an additional analysis to evaluate their potential incremental utility. These are shown in the fourth column of online supplemental table 1, and the full set of features including additional blood tests are referred to as ‘extended features’.

Ascertainment of outcomes

Diagnoses and dates for the 10 disease outcomes (itemised above) were collated across multiple UK Biobank sources including self-report, linked hospital and primary care records and deaths, using published code lists where available.^{35–37} Incident outcomes were defined by first occurrence of disease after baseline recruitment. Participants with a record of the same disease at baseline were excluded from modelling for that disease, and follow-up was censored at either death or the study end date. In addition to the defined outcomes, a wide selection of other potentially relevant diagnoses was collected using the same multisource approach (see online supplemental table 1). A full listing of UK Biobank codes for outcomes ascertainment is provided in online supplemental table 2.

Existing risk scores

We calculated QRISK3 for all study participants, along with 21 other published risk scores targeting disease risk across heart, brain, liver and kidney (see figure 2). These risk scores were selected based on a literature search for the most frequently used metrics for each outcome, the availability of published equations and the availability of online calculators for quality checking. Detailed information for all published indices is provided in online supplemental table 3. For incident stroke risk we considered QStroke³⁸ and CHA2DS2-VASc,³⁹ a score comprising congestive heart failure, hypertension, age, diabetes, prior stroke or transient ischaemic attack, vascular disease and sex. For all-cause dementia, we included three dementia risk scores; one developed using the CAIDE study (Cardiovascular Risk Factors, Aging and Dementia),⁴⁰ the Lifestyle for Brain Health score (LIBRA)⁴¹ and the recently developed UK Biobank Dementia Risk Score (UKB-DRS).³⁷ For myocardial infarction and heart failure, we considered Framingham Risk Score (with and without blood lipids),¹¹ the Pooled Cohort Equations to Prevent Heart Failure (PCP-HF risk score⁴²) and QRISK3.¹⁰ For atrial fibrillation, we applied the Cohorts for Heart and Aging Research in Genomic Epidemiology model for atrial fibrillation (CHARGE-AF⁴³). Chronic kidney disease risk (stage 3+) was predicted by two versions of QKidney⁴⁴ and a Kidney Risk Score developed by Nelson and colleagues.⁴⁵ For fatty liver disease, we considered the Fatty Liver Index⁴⁶ and the Dallas Steatosis Index.⁴⁷ Three diabetes risk scores (two versions of QDiabetes⁴⁸ and Cambridge Diabetes Score⁴⁹) were included as possibly useful predictors. AUDIT-C⁵⁰ (the Alcohol Use Disorders Identification Test) is a questionnaire designed to assess risk for alcoholic liver disease, however, only the first AUDIT question was available in the UK Biobank with a high degree of data completeness. Lastly, liver fibrosis was represented by three candidate scores; the Fibrosis-4 Index (FIB-4⁵¹), the nonalcoholic fatty liver disease (NAFLD) fibrosis score⁵² and APRI (the aspartate aminotransferase/platelet ratio⁵³). All published risk scores have been separately validated in their own studies. To assess their general utility, all risk scores were applied to the whole sample and to 10-year follow-up regardless of restrictions present in each respective derivation cohort.

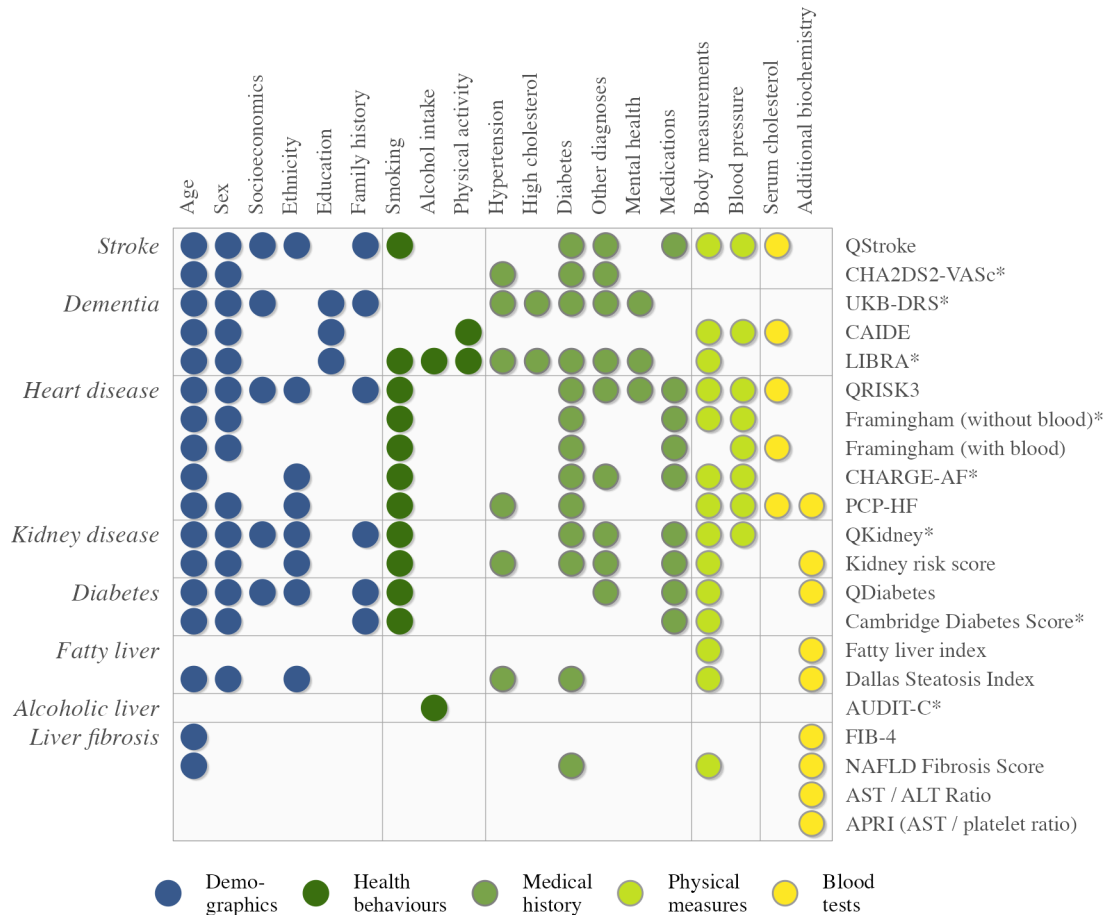


Figure 2 Risk scores across heart, brain, liver and kidney disease and their overlapping constituents. Coloured dots indicate measures that are included in the published risk scores shown on the right-hand side. Outcomes that are targeted by each risk score are shown on the left-hand side. Risk scores that can be implemented fully remotely are shown with an asterisk (*), in other words, risk scores that can be calculated without in-person contact. Physical measures include height, weight, waist circumference, hip circumference and resting heart rate. See online supplemental table 3 for a detailed listing of risk scores and their inputs. ALT, alanine aminotransferase; AST, aspartate aminotransferase; CAIDE, dementia risk score from the Cardiovascular Risk Factors, Aging and Dementia study; CHA2DS2-VASc, a score comprising congestive heart failure, hypertension, age, diabetes, prior stroke or transient ischaemic attack, vascular disease and sex; CHARGE-AF, Cohorts for Heart and Aging Research in Genomic Epidemiology model for atrial fibrillation; LIBRA, Lifestyle for Brain Health score; UKB-DRS, UK Biobank Dementia Risk Score.

Existing risk scores inherit their type from their inputs

At the high level, examination of score inputs (figure 2) shows significant overlap in feature topics, however, the finer level detail (online supplemental table 3) reveals significant variability in input requirements. Moreover, the inputs to each risk score differ by accessibility, in other words, risk score inputs are often a mixture of remote and standard features, with occasional additional blood tests. Extending the health information framework described above, we categorise any risk score that is comprised only of inputs that can be collected remotely as a remote risk score. UKB-DRS³⁷ (dementia) and Framingham risk score (using body mass index, BMI)¹¹ are examples of this. With the most restricted set of inputs, we would expect these models to be the least powerful and have the lowest predictive performance. Risk scores that include remote and standard features are considered standard risk scores (eg, QRISK,¹⁰ QStroke³⁸), while risk scores that require additional blood tests fall into the category of extended risk scores (eg, Nelson Kidney Risk Score,⁴⁵ Fatty Liver Index⁴⁶). We would expect risk models with access to the extended set of inputs (standard inputs plus additional biochemistry) to have the highest predictive performance. These categories are relevant to the process of building and comparing risk scores, such that performance comparisons are between models of the same type.

Ascertainment of other features

Age at baseline, self-reported sex, systolic blood pressure, pulse rate and anthropomorphic measurements were taken at baseline, along with a touchscreen questionnaire collecting information about ethnicity, education, family history, smoking, alcohol use and physical activity. Townsend Deprivation Index at baseline was assigned based on participant postcode. Ethnic groups and smoking categories were recoded to match QRISK3 specifications. Education was coded as binary—‘Do you have any postsecondary/college/university qualifications?’ Physical activity was dichotomised to greater than or equal to 600 summed metabolic equivalent task minutes per week,⁵⁴ approximately equivalent to 20 min exercise per day.⁵⁵ Family history was drawn from self-reported illnesses of mother, father and siblings, where age of illness was not specified. Descriptive statistics and source information for study features are provided in online supplemental table 4. Blood sampling was carried out as part of the baseline assessment, providing measures of total cholesterol, HDL cholesterol and additional biochemistry (online supplemental table 5). Missingness among remote features was very small (<1%) while missingness among blood test variables ranged between 3% and 14%. Covariate missingness was handled with multiple imputation, with details provided in

online supplemental methods SM3 and online supplemental table 6.

Statistical analysis

Statistical analysis was with R V.4.1.2 and RStudio V.2022.02.0. We randomly stratified the data to create a training set (70%, 159 768 participants) and an internal validation set (30%, 68 472 participants) following the Strengthening the Reporting of Observational studies in Epidemiology (STROBE) checklist (online supplemental table 7). Standard checks confirmed that the characteristics of the training and validation cohorts were not significantly different.

Pairwise modelling of cross-disease associations

Prior to full-scale modelling, we sought to describe the associations between diseases. We examined the association of prevalent conditions with the risk of incident conditions using Cox proportional hazards regression, adjusting by age, sex, postsecondary education, ethnicity, smoking, physical activity, alcohol intake frequency, BMI, Townsend Deprivation Index, family history (of heart disease, stroke and dementia), any cancer diagnosis, hypertension, high cholesterol and diabetes. Proportionality was checked with visualisation of residuals across all models. This analysis was performed using the whole cohort (n=228 240). Due to the large number of tests in this section, all coefficient tests were adjusted for multiple testing via the Benjamini-Hochberg method⁵⁶ with a false discovery rate of 5%.

Evaluation of existing risk scores

From the initial set of 22 published risk indices described above, we identified the best-performing score for each outcome across the 10 years of follow-up in the whole sample. Importantly, we evaluated all risk-score-outcome combinations, checking for potential predictive utility of each index beyond its original derivation cohort and intended outcome. We identified the risk score with highest discriminative performance, as measured by AUROC. AUROC was selected as the primary criterion for risk score performance because it does not depend on a specific prediction threshold and is more effective than simple accuracy in situations where rare events are being predicted.

New models for heart, brain, liver and kidney disease

Then, new models were developed for each of the 10 outcomes in the training set, and their performance for 10-year prediction was assessed in the internal validation set (online supplemental figure 1B). Model fitting was carried out (1) using remote features, (2) using standard features and (3) using the extended set of features. Crucially, at each level, we restricted all models (across the 10 related outcomes) to draw from the same pool of predictors. Feature selection was conducted using a stability selection approach,⁵⁷ combining lasso Cox regression with bootstrapping to systematically identify the predictors that show consistent importance, thereby simplifying the final model and mitigating the risk of overfitting (see online supplemental methods SM4).

For each outcome, the final set of predictors was placed into a single survival model, with coefficients and prediction thresholds calculated using the training set, and predictions scored for discriminative accuracy in the validation set. Across all models, prediction performance was assessed with multiple metrics

including AUROC, sensitivity, specificity, Somer's Dxy and Brier score. Differences in performance metrics were further bootstrapped with 1000 bootstrapped samples to derive uncertainty estimates. To calculate sensitivity and specificity, the prediction thresholds were set to maximise balanced accuracy, given by (sensitivity+specificity)/2.⁵⁸ Comparative performance was further evaluated with calibration plots to visually assess the alignment of predicted probabilities with actual outcomes, and reclassification statistics (integrated discrimination improvement (IDI) and continuous net reclassification improvement (cNRI)) to quantify any incremental improvements between existing and new models (online supplemental methods SM5 and SM6).

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Results

Participant characteristics

Overall, the study sample (n=228 240) was 45.3% male and 54.7% female, with an average age of 56.5 years at baseline (SD 8.1 years, table 1). Prevalence of hypertension, high cholesterol and diabetes at baseline was 32.6%, 20.8% and 5.5%, respectively. The most common incident event was atrial fibrillation (n=9997; 4.4%), and liver failure was the least common (n=340 events; 0.1%). Training and internal validation sets were similar across baseline variables and outcomes.

Associations between existing disease and future disease risk

Pairwise Cox analysis between heart-brain-liver-kidney outcomes and existing disease diagnoses revealed multiple cross-organ associations (figure 3, online supplemental table 8). All major heart diseases were significantly associated with increased risk of stroke, liver failure and the development of chronic kidney disease. All non-infective liver diseases at baseline were associated with increased risk of heart disease within 10 years, with an 88% increased risk of heart failure in participants with cirrhosis (HR 1.88, 95% CI 1.31 to 2.71, $p=7.05 \times 10^{-4}$), and a 39% increased risk of myocardial infarction in participants with fatty liver disease at baseline (1.39, 95% CI 1.08 to 1.80, $p=0.012$). Participants with alcoholic liver disease at baseline had a 4.5-fold risk for all-cause dementia (4.49, 95% CI 3.10 to 6.49, $p=1.46 \times 10^{-15}$) while rheumatoid arthritis at baseline conferred a 61% increased risk (1.61, 95% CI 1.30 to 1.99, $p=1.50 \times 10^{-5}$). Diagnosis of kidney or systemic inflammatory disease at baseline was associated with increased 10-year risk for disease across heart, brain and liver. Depression diagnosis at baseline had significant associations with disease across all four organs while serious mental illness (bipolar/schizophrenia/other psychosis) was associated with increased risk of alcoholic liver disease (HR 1.73, 95% CI 1.04 to 2.85, $p=0.033$), chronic kidney disease (HR 1.74, 95% CI 1.49 to 2.02, $p=6.78 \times 10^{-13}$) and all-cause dementia (HR 3.17, 95% CI 2.58 to 3.90, $p=9.66 \times 10^{-28}$).

Best existing risk scores

Details of the three best-performing existing risk scores of each type (by highest sample AUROC) for each level of accessibility are shown in online supplemental table 9. From here, the one risk score with the highest AUROC was applied as the comparator in the validation sample, shown by dark blue bars in figure 4 with additional details in online supplemental table 10. At least one existing risk score surpassed the minimum adequate AUROC (0.70)

Table 1 Sample characteristics

Characteristic	Whole sample (n=228 240)	Training set (n=159 768)	Internal validation set (n=68 472)
Age (years)	56.5 (±8.1)	56.5 (±8.1)	56.5 (±8.1)
Sex: female	124 891 (54.7%)	87 563 (54.8%)	37 328 (54.5%)
Sex: male	103 349 (45.3%)	72 205 (45.2%)	31 144 (45.5%)
Townsend Deprivation Index	-1.34 (±3.03)	-1.34 (±3.03)	-1.35 (±3.03)
Postsecondary education			
No	91 042 (39.9%)	63 739 (39.9%)	27 303 (39.9%)
Yes	135 207 (59.2%)	94 625 (59.2%)	40 582 (59.3%)
(Missing)	1991 (0.9%)	1404 (0.9%)	587 (0.9%)
White ethnicity	217 863 (95.5%)	152 469 (95.4%)	65 394 (95.5%)
All other ethnicity groups	10 377 (4.5%)	7299 (4.6%)	3078 (4.5%)
Current smoker	23 934 (10.5%)	16 716 (10.5%)	7218 (10.5%)
Physically active (METS≥600)	169 005 (74.0%)	118 544 (74.2%)	50 461 (73.7%)
Alcohol intake less than once per week	70 489 (30.9%)	49 341 (30.9%)	21 148 (30.9%)
Alcohol intake once a week or more	157 751 (69.1%)	110 427 (69.1%)	47 324 (69.1%)
Self-reported health			
Excellent	36 350 (15.9%)	25 469 (15.9%)	10 881 (15.9%)
Good	131 466 (57.6%)	91 985 (57.6%)	39 481 (57.7%)
Fair	48 542 (21.3%)	33 949 (21.2%)	14 593 (21.3%)
Poor	10 634 (4.7%)	7475 (4.7%)	3159 (4.6%)
(Missing)	1248 (0.5%)	890 (0.6%)	358 (0.5%)
Physical measurements			
Body mass index (BMI kg/m ²)	26.8(24.2, 30.0)	26.8(24.2, 30.0)	26.8(24.2, 30.0)
Obesity (BMI≥30 kg/m ²)	57 469 (25.2%)	40 375 (25.3%)	17 094 (25.0%)
Waist circumference (cm)	90.3 (±13.5)	90.4 (±13.5)	90.3 (±13.5)
Resting heart rate (bpm)	69.4 (±11.3)	69.4 (±11.3)	69.4 (±11.2)
Systolic blood pressure (mm Hg)	138.2 (±18.7)	138.2 (±18.6)	138.2 (±18.7)
Total/HDL cholesterol ratio	4.14 (±1.13)	4.14 (±1.13)	4.14 (±1.13)
Risk factors			
Hypertension	74 503 (32.6%)	52 240 (32.7%)	22 263 (32.5%)
High cholesterol	47 471 (20.8%)	33 315 (20.9%)	14 156 (20.7%)
Diabetes	12 543 (5.5%)	8917 (5.6%)	3626 (5.3%)
Chronic obstructive pulmonary disease	5288 (2.3%)	3706 (2.3%)	1582 (2.3%)
Any cancer	21 896 (9.6%)	15 307 (9.6%)	6589 (9.6%)
Existing conditions at baseline			
Stroke	4284 (1.9%)	2994 (1.9%)	1290 (1.9%)
Alzheimer's disease/dementia	346 (0.2%)	246 (0.2%)	100 (0.1%)
Myocardial infarction	6664 (2.9%)	4719 (3.0%)	1945 (2.8%)
Atrial fibrillation	4232 (1.9%)	2964 (1.9%)	1268 (1.9%)
Heart failure	1677 (0.7%)	1194 (0.7%)	483 (0.7%)
Chronic kidney disease	5610 (2.5%)	3886 (2.4%)	1724 (2.5%)
Fatty liver disease	1063 (0.5%)	718 (0.4%)	345 (0.5%)
Alcoholic liver disease	434 (0.2%)	298 (0.2%)	136 (0.2%)
Liver cirrhosis	498 (0.2%)	326 (0.2%)	172 (0.3%)
Liver failure	255 (0.1%)	170 (0.1%)	85 (0.1%)
Diagnoses after baseline			
Stroke	3953 (1.7%)	2728 (1.7%)	1225 (1.8%)
Dementia	2720 (1.2%)	1911 (1.2%)	809 (1.2%)
Myocardial infarction	5953 (2.6%)	4122 (2.6%)	1831 (2.7%)
Atrial fibrillation	9997 (4.4%)	6977 (4.4%)	3020 (4.4%)
Heart failure	4988 (2.2%)	3500 (2.2%)	1488 (2.2%)
Chronic kidney disease (stages 3, 4 and 5)	8698 (3.8%)	6116 (3.8%)	2582 (3.8%)
Fatty liver disease	3585 (1.6%)	2534 (1.6%)	1051 (1.5%)
Alcoholic liver disease	561 (0.2%)	402 (0.3%)	159 (0.2%)
Liver cirrhosis	951 (0.4%)	685 (0.4%)	266 (0.4%)
Liver failure	340 (0.1%)	236 (0.1%)	104 (0.2%)
Follow-up time (years, median, IQR)	10 (10–10)	10 (10–10)	10 (10–10)

Entries are either counts (percentages), mean (SD) or median (25th percentile, 75th percentile).

BMI, body mass index; HDL, high-density lipoprotein; METS, metabolic equivalent task, summed per week.

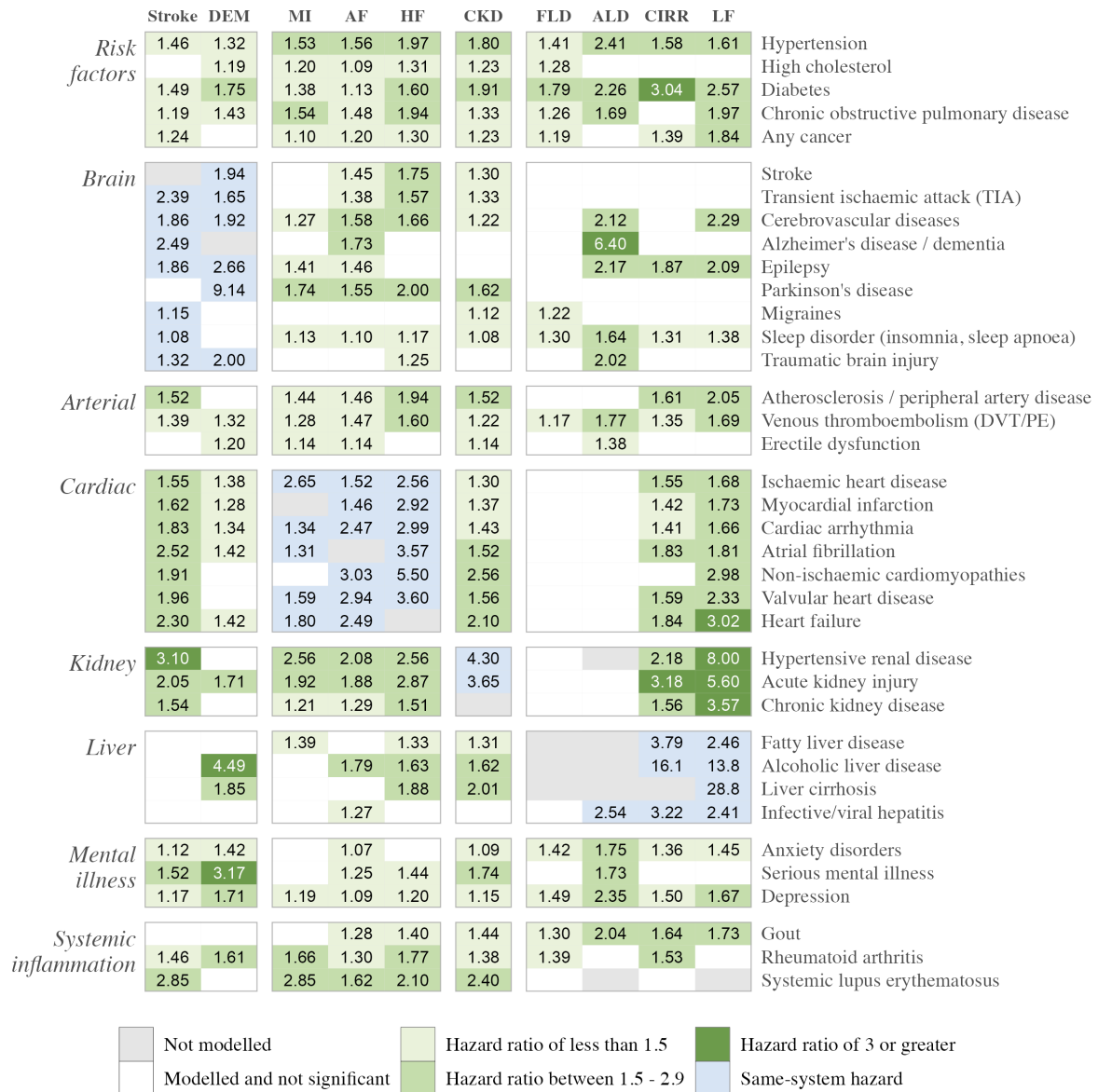


Figure 3 HRs for incident heart-brain-liver-kidney outcomes by existing disease diagnoses at baseline. Each entry shows the HR for incident outcomes (shown along the top) associated with the presence of existing risk factors, diagnoses and medication at baseline (shown down the right-hand side) in the whole cohort (n=228,240), using Cox-proportional hazards regression. For example, pre-existing hypertension increases the 10-year risk of stroke by 46%. Models are adjusted by age, sex, postsecondary education, ethnicity, smoking, physical activity, alcohol intake frequency, body mass index, Townsend Deprivation Index, family history (of heart disease, stroke and dementia), any cancer diagnosis, hypertension, high cholesterol and diabetes. HR significance was adjusted for multiple testing with a false discovery rate of 5%, where non-significant results are shown as empty white cells. Each result is from a different model. See online supplemental table 8 for detailed results. AF, atrial fibrillation; ALD, alcoholic liver disease; CKD, chronic kidney disease (stages 3, 4 or 5); CIRR, cirrhosis; DEM, all-cause dementia; DVT, deep vein thrombosis; FLD, fatty liver disease; HF, heart failure; LF, liver failure; MI, myocardial infarction; PE, pulmonary embolism.

for all outcomes except liver failure, with CHARGE-AF providing the best performance for atrial fibrillation across all levels (AUROC 0.759, 95% CI 0.751 to 0.767)), and UKB-DRS performing best for all-cause dementia (0.807, 95% CI 0.793 to 0.820)). Please note, that where a lower-level (remote or standard) model performs better than all more complex (extended) models for the same outcome, it will be retained as the best model at that level. Within the remote models, QKidney 5 had the highest AUROC for stroke (0.701 (95% CI 0.687 to 0.715)) and a surprisingly high remote model AUROC for heart failure (0.798 (95% CI 0.787 to 0.809)). Within standard models, we found several expected risk-score-outcome pairings, namely QStroke for stroke (0.727 (95% CI 0.714 to 0.741)), QRISK3 for myocardial infarction (0.757 (95%

CI 0.747 to 0.767)) and QKidney three for chronic kidney disease (0.760 (95% CI 0.751 to 0.769)). Unexpectedly, QStroke also had the highest AUROC for 10-year heart failure (0.806 (95% CI 0.795 to 0.817)).

Multiorgan risk prediction

The prediction performance of the newly developed risk scores is presented as green bars in figure 4 with additional details in online supplemental table 10. In general, the new risk score models performed as well or better than existing risk scores for all outcomes, with all new models achieving AUROC above 0.70. Using standard health check predictors, newly developed models performed significantly better than existing risk scores for some

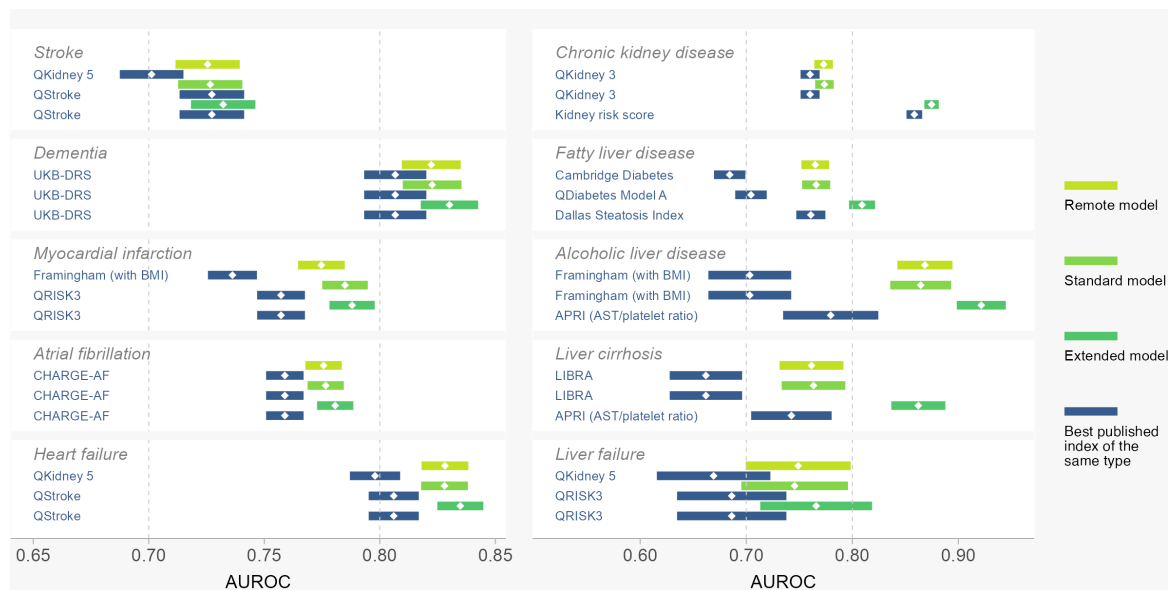


Figure 4 Performance of multiorgan risk scores in the validation set. Comparison of area under the receiver operating curve (AUROC, also known as concordance or C-statistic) for 10-year risk prediction across 10 outcomes. Horizontal bars show the 95% CI for validation set AUROC with uncertainty estimated from 1000 bootstrapped samples. Predictions from existing risk scores are shown in dark blue, while newly developed risk scores are shown in green. Remote models contain health metrics that can be answered verbally or self-measured easily by the patient. Standard models contain all remote metrics plus blood pressure and serum cholesterol tests. Extended models contain further blood tests. See online supplemental table 10 for detailed results. BMI, body mass index; UKB-DRS, UK Biobank Dementia Risk Score; APRI, the ratio of aspartate aminotransferase (AST) to platelet count.

outcomes, with higher AUROC for myocardial infarction of 0.785 with 95% CI (0.775 to 0.795), atrial fibrillation (0.777 (95% CI 0.768 to 0.785)), heart failure (0.828 (95% CI 0.818 to 0.838)), fatty liver disease (0.766 (95% CI 0.753 to 0.779)), alcoholic liver disease (0.864 (95% CI 0.835 to 0.894)) and liver cirrhosis (0.763 (95% CI 0.734 to 0.793)). Newly developed risk models had similar performance to existing scores for stroke (0.727 (95% CI 0.713 to 0.740)), dementia (0.823 (95% CI 0.810 to 0.836)) and chronic kidney disease (0.774 (95% CI 0.765 to 0.783)).

Importantly, for all outcomes studied, newly developed models using only remote features were able to achieve similar discriminative accuracy (AUROC) to their respective standard models.

When the set of additional biochemistry was added to the pool of potential predictors (in the extended models), there was very little incremental increase in predictive performance for heart and brain outcomes. In contrast, extended model features produced significantly better predictions for chronic kidney disease (AUROC 0.875, 95% CI 0.868 to 0.881), fatty liver (0.809, 95% CI 0.797 to 0.822), alcoholic liver (0.922, 95% CI 0.899 to 0.944) and cirrhosis (0.862, 95% CI 0.837 to 0.888). This improvement is substantial compared with existing risk scores and standard model estimates, suggesting that an approach with more blood biomarkers might be better at picking up these abnormalities. As expected, calibration statistics (online supplemental figure 3) and reclassification indices (online supplemental table 11) showed better calibration and significantly improved reclassification in the newly developed risk scores compared with existing risk scores, for example, standard myocardial infarction IDI=0.014, 95% CI (0.011 to 0.017) and cNRI=0.633, 95% CI (0.588 to 0.677). In nearly all comparisons, newly developed models had lower Brier scores (less average squared error) and higher Somers' Dxy (better rank correlation) than existing models. While there were some instances of improved sensitivity in the newer models (stroke, myocardial infarction, chronic kidney disease, alcoholic liver and cirrhosis), overall, the improvements in discrimination were mainly driven

by better specificity (fewer false positives, online supplemental table 10).

The heart-brain-liver-kidney risk model coefficients

The large number of coefficients for multioutcome models are provided as beta coefficients in online supplemental spreadsheet 1, and as HRs in online supplemental spreadsheet 2, with a visual overview in online supplemental tables 4–6.

Discussion

In this proof-of-concept analysis with 228 240 UK Biobank participants, we demonstrated that easily collected diagnostics can be used to assess risk across multiple disease outcomes. We have shown how this can be done without specialist computing or invasive biomarkers.

Pairwise modelling showed a complex pattern of cross-system associations, building on prior efforts to understand multimorbidity in the heart-brain-liver-kidney cluster.²³ We confirmed that disease risk across all four organs was significantly associated with well-known risk factors such as hypertension, diabetes and high cholesterol; as well as other factors that have not yet been incorporated into standard risk paradigms beyond QRISK3, such as mental illness, systemic inflammation, sleep quality, arterial health and medication use.^{59 60}

Recent studies have shown significant improvements in cardiovascular risk prediction using large data sets and machine learning methods.^{60–63} However, these studies still only target one organ (the heart), and when compared with conventional statistical models, deep learning or other 'black box' methods are not as readily explainable or easily translatable to clinical use.⁶⁴ Several studies have tackled multidisease prediction. Bayati *et al*⁶⁵ use multitask learning and group dimensionality reduction to identify a reduced pool of health check features to predict heart-brain-liver-kidney outcomes across 2 years follow-up. Most similar to the current work, Mahajan *et al*⁶⁶ used electronic health records

to derive multiple organ-specific risk scores, with a high degree of discrimination (AUROC > 0.80 across heart, brain, lung, kidney and digestive disease). However, this study predicted hospital readmission using previous admissions for the same disease, whereas our models predict risk of new-onset disease.

The current study has several important limitations. We recognise the importance of thoroughly evaluating existing risk instruments before moving forward with new risk score development. We have begun this process, but there is more work to be done. On the other hand, by restricting the pool of input variables, new risk score development may well be required to meet this constraint, particularly where remote risk scores of adequate quality do not yet exist.

We acknowledge that the internal validation performance of our scores (developed within UK Biobank) is not directly comparable with external validation performance of published risk scores developed outside UK Biobank. Furthermore, each published risk score has a range of validation values across published work. For example, the validation AUROC provided by the original QRISK3 paper¹⁰ was 0.88 for women and 0.86 for men. Since then other external validation performance has varied in a numerical range consistent with the current work, with values of 0.707 for women and 0.681 for men reported for the 45–64 age group in the Clinical Practice Research Datalink,⁶⁷ and values of 0.722 for women and 0.697 for men in a recent validation in UK Biobank.⁶⁸

It may seem unconventional to apply existing risk score instruments outside their intended cohort (eg, including people with comorbidities) and outside their intended outcome (eg, using QRISK3 to predict myocardial infarction rather than combined CVD). Our study is not the first to explore this approach^{25,69} and to provide essential evidence for whether existing scores hold unrealised potential in additional contexts.

Although large compared with some, we caution that this study is small compared with larger risk score development projects, and we have not provided an external validation cohort. Furthermore, UK Biobank participants are subject to self-selection bias, and as such, they are known to be healthier and less ethnically diverse than the UK population.⁷⁰ Therefore, any final models with this approach will require further recalibration and validation in large nationally representative cohorts. We acknowledge that there are some variables that are not well measured in the UK Biobank. Where these come up in multidisease risk equations, these are likely to be more accurately captured in a primary care-specific database.

In conclusion, this analysis demonstrates the feasibility of using standard health check predictors to produce multidisease risk estimates of reasonable quality. Such an approach has the potential to ease pressure on primary care, allowing physicians more time to focus on interpretation and follow-up⁷¹ thus providing new opportunities for multimorbidity prevention.

Author affiliations

¹Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, and Oxford University Hospitals NHS Foundation Trust, Oxford, UK

²William Harvey Research Institute, Queen Mary University of London, London, UK

³Barts Heart Centre, St Bartholomew's Hospital, Barts Health NHS Trust, London, UK

⁴Heart and Vascular Center, Semmelweis University, Budapest, Hungary

⁵Department of Experimental Psychology, University of Oxford, Oxford, UK

⁶Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, UK

⁷Perspectum Ltd, Oxford, UK

⁸Wellcome Centre for Integrative Neuroimaging (WIN FMRIB), University of Oxford, Oxford, UK

⁹Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford, UK

¹⁰Health Data Research UK, London, UK

¹¹Alan Turing Institute, London, UK

Contributors The study was conceived by TEN, SN, SEP and CM. CM and TEN formulated the statistical analysis plan. CM led and performed the analysis and led in writing of the manuscript. ZR-E and LS contributed to drafting of the manuscript. All coauthors including MV, BR, AT, AR-F and MH provided critical review of the work. All authors read and approved the final manuscript. TEN and SN provided overall supervision for the work. TEN and SN are the guarantors of the work. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding CM and SN are supported by the Oxford National Institute for Health and Care Research (NIHR) Biomedical Research Centre (IS-BRC-1215-20008) and the Oxford BHF Centre of Research Excellence. BR acknowledges support from the BHF Oxford CRE (RE/18/3/34214). LS has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825903 (euCanShare project). ZR-E recognises the NIHR Integrated Academic Training programme which supports her Academic Clinical Lectureship post and BHF Clinical Research Training Fellowship (FS/17/81/33318). This work acknowledges the support of the National Institute for Health and Care Research Barts Biomedical Research Centre (NIHR203330); a delivery partnership of Barts Health NHS Trust, Queen Mary University of London, St George's University Hospitals NHS Foundation Trust and St George's University of London. This work was supported by Health Data Research UK, an initiative funded by UK Research and Innovation, Department of Health and Social Care (England) and the devolved administrations Wellcome Trust, NIHR Oxford Biomedical Centre and leading medical research charities. AT is supported by a Wellcome Trust (<https://wellcome.org/>) fellowship (216462/Z/19/Z). MH is supported by the Wellcome Trust (206330/Z/17/Z) and NIHR Oxford Biomedical Research Centre (IS-BRC-1215-20008). TEN is supported by the Li Ka Shing Centre for Health Information and Discovery, a NIH grant (<https://www.nih.gov/>, TN: R01EB026859), the NIHR Oxford Biomedical Research Centre (BRC-1215-20014), and a Wellcome Trust award (TN: 100309/Z/12/Z). This work includes data provided by patients and collected by the NHS and NHS Digital as part of their care and support. The research was supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with funding from the NIHR Oxford BRC. This research used data assets made available by National Safe Haven as part of the Data and Connectivity National Core Study, led by Health Data Research UK in partnership with the Office for National Statistics and funded by UK Research and Innovation (MC_PC_20058).

Disclaimer The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests SEP provides consultancy to Cardiovascular Imaging, Calgary, Alberta, Canada. BR consulted for Axcella Therapeutics. AR-F is an employee and shareholder in Perspectum, Oxford, UK. SN is a founder, shareholder and

former board member of Perspectum. TEN provides consultancy to Perspectum, Oxford, UK. All other authors declare no conflicts of interest. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and this study complies with the Declaration of Helsinki; the work was covered by the ethical approval for UK Biobank studies from the National Health Service (NHS) National Research Ethics Service on 17 June 2011 (Ref 11/NW/0382) and extended on 18 June 2021 (Ref 21/NW/0157) with written informed consent obtained from all participants. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. This analysis was produced under UK Biobank Access Application 59867. The data in this study are owned by the UK Biobank (www.ukbiobank.ac.uk) and legal constraints do not permit public sharing of the data. The UK Biobank, however, is open to all bona fide researchers anywhere in the world. Thus, the data used in this communication can be easily and directly accessed by applying through the UK Biobank Access Management System (www.ukbiobank.ac.uk/register-apply). Results from this study will be returned to UK Biobank according to their published returns policy.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iD

Celeste McCracken <http://orcid.org/0000-0003-1285-2393>

References

- Vetrano DL, Calderón-Larrañaga A, Marengoni A, *et al*. An International perspective on chronic Multimorbidity: approaching the elephant in the room. *The Journals of Gerontology* 2018;73:1350–6.
- Canoy D, Tran J, Zottoli M, *et al*. Association between Cardiometabolic disease Multimorbidity and all-cause mortality in 2 million women and men registered in UK general practices. *BMC Med* 2021;19:258.
- Soley-Bori M, Ashworth M, Bisquera A, *et al*. Impact of Multimorbidity on Healthcare costs and utilisation: A systematic review of the UK literature. *Br J Gen Pract* 2021;71:e39–46.
- Barnett K, Mercer SW, Norbury M, *et al*. Epidemiology of Multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet* 2012;380:37–43.
- World Health Organization. Multimorbidity: Technical Series on Safer Primary Care, 2016. Available: <https://iris.who.int/bitstream/handle/10665/252275/9789241511650-eng.pdf> [Accessed 28 Nov 2023].
- Head A, Fleming K, Kypridemos C, *et al*. Multimorbidity: the case for prevention. *J Epidemiol Community Health* 2021;75:242–4.
- Lawrence W, Watson D, Barker H, *et al*. Meeting the UK government's prevention agenda: primary care practitioners can be trained in skills to prevent disease and support self-management. *Perspect Public Health* 2022;142:158–66.
- Pearson-Stuttard J, Ezzati M, Gregg EW. Multimorbidity—a defining challenge for health systems. *Lancet Public Health* 2019;4:e599–600.
- Bierman AS. Preventing and managing Multimorbidity by integrating behavioral health and primary care. *Health Psychol* 2019;38:851–4.
- Hippisley-Cox J, Coupland C, Brindle P. Development and validation of Qrisk3 risk prediction Algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017;357:j2099.
- D'Agostino RB Sr, Vasan RS, Pencina MJ, *et al*. General cardiovascular risk profile for use in primary care. *Circulation* 2008;117:743–53.
- Hylton K, Thompson K, Kearney M, *et al*. Public Health England; NHS Health Check Best practice guidance: For commissioners and providers, 2019. Available: <https://www.healthcheck.nhs.uk/secmsfile?id=1480> [Accessed 2 Jan 2024].
- NHS. The NHS Long Term Plan, 2020. Available: <https://www.longtermplan.nhs.uk/wp-content/uploads/2019/08/nhs-long-term-plan-version-1.2.pdf> [Accessed 2 Nov 2023].
- Karlsen TH, Sheron N, Zelber-Sagi S, *et al*. The EASL-lancet liver Commission: protecting the next generation of Europeans against liver disease complications and premature mortality. *The Lancet* 2022;399:61–116.
- Shlipak MG, Tummalaipalli SL, Boulware LE, *et al*. The case for early identification and intervention of chronic kidney disease: conclusions from a kidney disease: improving global outcomes (KDIGO) controversies conference. *Kidney Int* 2021;99:34–47.
- NHS England. NHS Health Check. NHS.uk, 2023. Available: <https://www.nhs.uk/conditions/nhs-health-check/> [Accessed 2 Feb 2024].
- Public Health England. NHS health check information leaflet. 2015. Available: <https://www.healthcheck.nhs.uk/secmsfile?id=401> [Accessed 2 Feb 2024].
- El Hadi H, Di Vincenzo A, Vettor R, *et al*. Relationship between heart disease and liver disease: A two-way street. *Cells* 2020;9:567.
- McCracken C, Raisi-Estabragh Z, Veldsman M, *et al*. Multi-organ imaging demonstrates the heart-brain-liver axis in UK Biobank participants. *Nat Commun* 2022;13:7839.
- Siriwardhana C, Lim E, Davis J, *et al*. Progression of diabetes, ischemic heart disease, and chronic kidney disease in a three chronic conditions Multistate model. *BMC Public Health* 2018;18:752.
- Prados-Torres A, Calderón-Larrañaga A, Hanco-Saavedra J, *et al*. Multimorbidity patterns: a systematic review. *J Clin Epidemiol* 2014;67:254–66.
- Soley-Bori M, Bisquera A, Ashworth M, *et al*. Identifying Multimorbidity clusters with the highest primary care use: 15 years of evidence from a multi-ethnic metropolitan population. *Br J Gen Pract* 2022;72:e190–8.
- Bisquera A, Gulliford M, Dodhia H, *et al*. Identifying longitudinal clusters of Multimorbidity in an urban setting: A population-based cross-sectional study. *Lancet Reg Health Eur* 2021;3:100047.
- Iacobucci G. Scrap 10 minute consultations, RCGP URGES. *BMJ* 2019;365:l2270.
- Badawy MAEMD, Naing L, Johar S, *et al*. Evaluation of cardiovascular diseases risk calculators for Cvd prevention and management: Scoping review. *BMC Public Health* 2022;22:1742.
- Poppe KK, Doughty RN, Wells S, *et al*. Developing and validating a cardiovascular risk score for patients in the community with prior cardiovascular disease. *Heart* 2017;103:891–2.

- 27 Alonso-Morán E, Nuño-Solinis R, Onder G, *et al*. Multimorbidity in risk stratification tools to predict negative outcomes in adult population. *Eur J Intern Med* 2015;26:182–9.
- 28 NHS England. Improving access for all: reducing inequalities in access to general practice services, 2018. Available: <https://www.england.nhs.uk/publication/improving-access-for-all-reducing-inequalities-in-access-to-general-practice-services/> [Accessed 27 Nov 2023].
- 29 Verma P, Kerrison R. Patients' and physicians' experiences with remote consultations in primary care during the COVID-19 pandemic: a multi-method rapid review of the literature. *BJGP Open* 2022;6:BJGP0.2021.0192.
- 30 Ladds E, Khan M, Moore L, *et al*. The impact of remote care approaches on continuity in primary care: a mixed-studies systematic review. *Br J Gen Pract* 2023;73:e374–83.
- 31 Fagerland MW, Hosmer DW. A goodness-of-fit test for the proportional odds regression model. *Stat Med* 2013;32:2235–49.
- 32 Sudlow C, Gallacher J, Allen N, *et al*. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Med* 2015;12:e1001779.
- 33 UK Biobank. Primary Care Linked Data (Version 1.0), 2019. Available: https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/primary_care_data.pdf [Accessed 30 Nov 2023].
- 34 UK Biobank. Hospital Inpatient Data (Version 4.0), 2023. Available: <https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/HospitalEpisodeStatistics.pdf> [Accessed 3 Feb 2024].
- 35 UK Biobank. Code lists for health outcomes, 2020. Available: <https://biobank.ndph.ox.ac.uk/showcase/refer.cgi?id=594> [Accessed 23 Mar 2022].
- 36 Hayward KL, Johnson AL, Horsfall LU, *et al*. Detecting non-alcoholic fatty liver disease and risk factors in health databases: accuracy and limitations of the ICD-10-AM. *BMJ Open Gastroenterol* 2021;8:e000572.
- 37 Anatóürk M, Patel R, Ebmeier KP, *et al*. Development and validation of a dementia risk score in the UK Biobank and Whitehall II cohorts. *BMJ Ment Health* 2023;26:e300719.
- 38 Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of Ostroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. *BMJ* 2013;346:f2573.
- 39 Ntaios G, Lip GYH, Makaritsis K, *et al*. Chads2, Cha2Ds2-Vasc, and long-term stroke outcome in patients without atrial fibrillation. *Neurology* 2013;80:1009–17.
- 40 Kivipelto M, Ngandu T, Laatikainen T, *et al*. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. *Lancet Neurol* 2006;5:735–41.
- 41 Schiepers OJG, Köhler S, Deckers K, *et al*. Lifestyle for brain health (LIBRA): a new model for dementia prevention. *Int J Geriatr Psychiatry* 2018;33:167–75.
- 42 Khan SS, Ning H, Shah SJ, *et al*. 10-year risk equations for incident heart failure in the general population. *J Am Coll Cardiol* 2019;73:2388–97.
- 43 Alonso A, Krijthe BP, Aspelund T, *et al*. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc* 2013;2:e000102.
- 44 Hippisley-Cox J, Coupland C. Predicting the risk of chronic kidney disease in men and women in England and Wales: prospective derivation and external validation of the Qkidney®scores. *BMC Fam Pract* 2010;11:49.
- 45 Nelson RG, Grams ME, Ballew SH, *et al*. Development of risk prediction equations for incident chronic kidney disease. *JAMA* 2019;322:2104–14.
- 46 Bedogni G, Bellentani S, Miglioli L, *et al*. The fatty liver index: A simple and accurate Predictor of hepatic steatosis in the general population. *BMC Gastroenterol* 2006;6:33.
- 47 McHenry S, Park Y, Browning JD, *et al*. Dallas steatosis index identifies patients with Nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2020;18:2073–80.
- 48 Hippisley-Cox J, Coupland C. Development and validation of Qdiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. *BMJ* 2017;359:j5019.
- 49 Griffin SJ, Little PS, Hales CN, *et al*. Diabetes risk score: towards earlier detection or type 2 diabetes in general practice. *Diabetes Metab Res Rev* 2000;16:164–71.
- 50 Bush K, Kivlahan DR, McDonnell MB. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking
- 51 Vallet-Pichard A, Mallet V, Nalpas B, *et al*. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and Fibrotest. *Hepatology* 2007;46:32–6.
- 52 Angulo P, Hui JM, Marchesini G, *et al*. The NAFLD fibrosis score: A noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* 2007;45:846–54.
- 53 De Matteis C, Cariello M, Graziano G, *et al*. AST to platelet ratio index (APRI) is an easy-to-use Predictor score for cardiovascular risk in metabolic subjects. *Sci Rep* 2021;11:14834.
- 54 The IPAQ group. Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) – Short and Long Forms, 2005. Available: https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/ipaq_analysis.pdf [Accessed 21 Oct 2021].
- 55 NHS England. Physical activity guidelines for adults aged 19 to 64. nhs.uk, 2022. Available: <https://www.nhs.uk/live-well/exercise/exercise-guidelines/physical-activity-guidelines-for-adults-aged-19-to-64/> [Accessed 1 Feb 2024].
- 56 Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 1995;57:289–300.
- 57 Meinshausen N, Bühlmann P. Stability selection. *Journal of the Royal Statistical Society Series B* 2010;72:417–73.
- 58 Thölke P, Mantilla-Ramos Y-J, Abdelhedi H, *et al*. Class imbalance should not throw you off balance: choosing the right classifiers and performance Metrics for brain decoding with imbalanced data. *Neuroimage* 2023;277.
- 59 Correll CU, Solmi M, Veronese N, *et al*. Prevalence, incidence and mortality from cardiovascular disease in patients with pooled and specific severe mental illness: a large-scale meta-analysis of 3,211,768 patients and 113,383,368 controls. *World Psychiatry* 2017;16:163–80.
- 60 Dolezalova N, Reed AB, Despotovic A, *et al*. Development of an accessible 10-year Digital cardiovascular (Dicava) risk assessment: a UK Biobank study. *Eur Heart J Digit Health* 2021;2:528–38.
- 61 Alaa AM, Bolton T, Di Angelantonio E, *et al*. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLOS ONE* 2019;14:e0213653.
- 62 Alsayegh F, Alkhamis MA, Ali F, *et al*. Anemia or other Comorbidities? using machine learning to reveal deeper insights into the drivers of acute coronary syndromes in hospital admitted patients. *PLOS ONE* 2022;17:e0262997.
- 63 Steinfeldt J, Buergel T, Loock L, *et al*. Neural network-based integration of Polygenic and clinical information: development and validation of a prediction model for 10-year risk of major adverse cardiac events in the UK Biobank cohort. *Lancet Digit Health* 2022;4:e84–94.
- 64 Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996;49:1225–31.
- 65 Bayati M, Bhaskar S, Montanari A. A low-cost method for multiple disease prediction. *AMIA Annu Symp Proc* 2015;2015:329–38.
- 66 Mahajan A, Deonarine A, Bernal A, *et al*. Developing the total health profile, a Generalizable unified set of Multimorbidity risk scores derived from machine learning for broad patient populations: retrospective cohort study. *J Med Internet Res* 2021;23:e32900.
- 67 Livingstone S, Morales DR, Donnan PT, *et al*. Effect of competing mortality risks on predictive performance of the Qrisk3 cardiovascular risk prediction tool in older people and those with Comorbidity: external validation population cohort study. *Lancet Healthy Longev* 2021;2:e352–61.
- 68 Parsons RE, Liu X, Collister JA, *et al*. Independent external validation of the Qrisk3 cardiovascular disease risk prediction model using UK Biobank. *Heart* 2023;109:1690–7.
- 69 Zhu W-G, Xiong Q-M, Hong K. Meta-analysis of Chads2 versus Cha2Ds2-Vasc for predicting stroke and thromboembolism in atrial fibrillation patients independent of anticoagulation. *Tex Heart Inst J* 2015;42:6–15.
- 70 Huang JY. Representativeness is not representative: addressing major inferential threats in the UK Biobank and other big data repositories. *Epidemiology* 2021;32:189–93.
- 71 Atkins L, Stefanidou C, Chadborn T, *et al*. Influences on NHS health check Behaviours: a systematic review. *BMC Public Health* 2020;20:1359.