




OPEN ACCESS

Original research

Independent external validation of the QRISK3 cardiovascular disease risk prediction model using UK Biobank

Ruth E Parsons ¹, Xiaonan Liu,¹ Jennifer A Collister,¹ David A Clifton,² Benjamin J Cairns,¹ Lei Clifton¹

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/heartjnl-2022-321231>).

¹Nuffield Department of Population Health, University of Oxford, Oxford, UK

²Institute of Biomedical Engineering, University of Oxford, Oxford, UK

Correspondence to

Lei Clifton;
lei.clifton@ndph.ox.ac.uk

Received 10 March 2023

Accepted 20 June 2023

Published Online First

9 July 2023

ABSTRACT

Objective To externally evaluate the performance of QRISK3 for predicting 10 year risk of cardiovascular disease (CVD) in the UK Biobank cohort.

Methods We used data from the UK Biobank, a large-scale prospective cohort study of 403 370 participants aged 40–69 years recruited between 2006 and 2010 in the UK. We included participants with no previous history of CVD or statin treatment and defined the outcome to be the first occurrence of coronary heart disease, ischaemic stroke or transient ischaemic attack, derived from linked hospital inpatient records and death registrations.

Results Our study population included 233 233 women and 170 137 men, with 9295 and 13 028 incident CVD events, respectively. Overall, QRISK3 had moderate discrimination for UK Biobank participants (Harrell's C-statistic 0.722 in women and 0.697 in men) and discrimination declined by age (<0.62 in all participants aged 65 years or older). QRISK3 systematically overpredicted CVD risk in UK Biobank, particularly in older participants, by as much as 20%.

Conclusions QRISK3 had moderate overall discrimination in UK Biobank, which was best in younger participants. The observed CVD risk for UK Biobank participants was lower than that predicted by QRISK3, particularly for older participants. It may be necessary to recalibrate QRISK3 or use an alternate model in studies that require accurate CVD risk prediction in UK Biobank.

INTRODUCTION

Cardiovascular diseases (CVDs) are the leading cause of global mortality,¹ and healthcare providers need to be able to identify patients with a high risk of CVD to allocate primary prevention measures accurately and reliably. Prognostic models can classify individuals into event risk groups, allowing decisions to be made about their healthcare. There are numerous prediction models designed to estimate the risk of developing CVD in use worldwide, including the Framingham,² SCORE³ and QRISK^{4–6} models. The QRISK models are Cox proportional hazard models that predict time to CVD events and are routinely used in the UK by healthcare providers to calculate the 10 year risk of CVD for patients during routine NHS health checks.⁷

The first QRISK model was published in 2007 with the aim of estimating the 10 year risk of CVD in females and males.⁴ QRISK was followed by an

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ The QRISK3 model is routinely used in the UK by healthcare providers to calculate the 10 year cardiovascular disease risk of patients.
- ⇒ Although QRISK3 was derived and validated using primary care data, it is used without validation in UK Biobank by researchers.

WHAT THIS STUDY ADDS

- ⇒ QRISK3 has moderate discrimination for UK Biobank participants but this declines with age.
- ⇒ QRISK3 systematically overpredicts cardiovascular disease risk in UK Biobank, particularly in older participants, by as much as 20%.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ Researchers that require the accurate prediction of cardiovascular events in UK Biobank may want to use strategies (eg, recalibration) to improve QRISK3's prediction.
- ⇒ Where the use of QRISK3 is not essential, researchers may benefit from developing tailored risk prediction models for their study aims using other risk factors available for UK Biobank participants.

updated model (QRISK2) in 2008, which included additional risk factors compared with its predecessor.⁶ Since 2008, QRISK2 has been updated continually, with the inclusion of type 1 diabetes as a risk factor, expansion of categories for the smoking variable, and updated Townsend deprivation score with new census data.^{5,6} The most up to date QRISK model, QRISK3, was derived in 2017 using health records of 1 283 174 patients aged between 35 and 74 years registered between 1995 and 2007 with practices across the UK,⁶ to incorporate risk factors that were outlined in the National Institute for Health and Care Excellence (NICE) 2014 clinical guideline.⁸ The risk factors included in the QRISK3 model can be seen in [box 1](#).

QRISK3 was externally validated using primary care data from the Clinical Practice Research Data-link (CPRD) in 2021 and was found to perform well at the overall population level.⁹ CPRD has a similar case-mix to the QRISK3 derivation cohort,



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY. Published by BMJ.

To cite: Parsons RE, Liu X, Collister JA, et al. *Heart* 2023;**109**:1690–1697.

Box 1 Risk factors included in the QRISK3 model⁶

1. Age at study entry (baseline)
2. Ethnic origin (nine categories)
3. Deprivation (as measured by the Townsend score, where higher values indicate higher levels of material deprivation)
4. Systolic blood pressure
5. Body mass index
6. Total cholesterol: high-density lipoprotein cholesterol ratio
7. Smoking status (non-smoker, former smoker, light smoker (1–9/day), moderate smoker (10–19/day) or heavy smoker (≥ 20 /day))
8. Family history of coronary heart disease in a first-degree relative aged less than 60 years
9. Diabetes (type 1, type 2 or no diabetes)
10. Treated hypertension (diagnosis of hypertension and treatment with at least one antihypertensive drug)
11. Rheumatoid arthritis (diagnosis of rheumatoid arthritis, Felty's syndrome, Caplan's syndrome, adult-onset Still's disease or inflammatory polyarthropathy not otherwise specified)
12. Atrial fibrillation (including atrial fibrillation, atrial flutter and paroxysmal atrial fibrillation)
13. Chronic kidney disease (general practitioner recorded diagnosis of chronic kidney disease stage 3, stage 4 or 5) and major chronic renal disease (including nephrotic syndrome, chronic glomerulonephritis, chronic pyelonephritis, renal dialysis and renal transplant)
14. Measure of systolic blood pressure variability (standard deviation (SD) of repeated measures)
15. Diagnosis of migraine (including classic migraine, atypical migraine, abdominal migraine, cluster headaches, basilar migraine, hemiplegic migraine, and migraine with or without aura)
16. Corticosteroid use (British National Formulary (BNF) chapter 6.3.2 including oral or parenteral prednisolone, betamethasone, cortisone, depo-medrone, dexamethasone, deflazacort, efcortisol, hydrocortisone, methylprednisolone or triamcinolone)
17. Systemic lupus erythematosus (SLE) (including diagnosis of SLE, disseminated lupus erythematosus or Libman-Sacks disease)
18. Second generation 'atypical' antipsychotic use (including amisulpride, aripiprazole, clozapine, lurasidone, olanzapine, paliperidone, netiquette, risperidone, sertindole or zotepine)
19. Diagnosis of severe mental illness (including psychosis, schizophrenia or bipolar affective disease)
20. Diagnosis of erectile dysfunction or treatment for erectile dysfunction (BNF chapter 7.4.5 including alprostadil, phosphodiesterase type 5 inhibitors, papaverine or phentolamine)

and therefore the external validation of QRISK3 using CPRD measures the reproducibility of the model's performance. The transportability of QRISK3 (ie, how well it performs in a population with different characteristics to the derivation cohort) must be explored in each independent population that differs considerably in setting to the derivation cohort¹⁰ before the model can be used reliably in that independent population.

We have identified multiple studies where the QRISK3 scores for UK Biobank participants are used in the analyses. The extent to which the authors of the identified studies address the

discrimination and the calibration of QRISK3 applied to UK Biobank data varies, and the way that this lack of validation may affect the conclusion of these publications depends on the study aims. Without considering the discrimination and calibration of the QRISK3 model applied to UK Biobank data, the accuracy of resulting CVD risk scores is unknown, and any subsequent conclusions that are drawn may be misleading. This study is the first to investigate the calibration of QRISK3 in UK Biobank, and provides an independent external validation of the QRISK3 model applied to this widely used cohort.

METHODS**Study population**

The study design and methods of the UK Biobank study have been described previously.^{11 12} Approximately 9.2 million people were invited to take part in the UK Biobank study between 2006 and 2010; these were individuals aged between 40 and 69 years, who were registered with the NHS and who lived in a 25 mile radius of 1 of 22 assessment centres in England, Wales and Scotland.¹¹ A total of 503 325 individuals attended an assessment centre, a response rate of 5.5%.^{11 12} At baseline, all participants completed a touch-screen questionnaire, a verbal interview and had physical measurements and samples of blood, urine and saliva collected.^{11 12}

To align with the exclusion criteria used in the derivation of QRISK3,⁶ we excluded UK Biobank participants who had a prior diagnosis of CVD, were using prescribed cholesterol-lowering medication at cohort entry or had missing Townsend deprivation scores. The participants in UK Biobank (40–69 years at baseline) were all within the age range required for the QRISK3 model.

We did not involve participants in the design, or conduct, or reporting, or dissemination plans of our research.

Definition of risk factors

All UK Biobank variables we have used in the QRISK model were measured at the baseline recruitment visit, which took place in 2006–2010. A similar approach was used in the original QRISK derivation, with 'the value closest to the entry date to the cohort for each patient' used for each risk factor.⁴

We matched the risk factors included in the QRISK3 model to the variables available in UK Biobank, including variables collected during the baseline assessment visit at study entry, and prevalent disease diagnoses determined through linked hospital records (further detail in online supplemental materials). For sociodemographic risk factors, data from the baseline assessment visit was used, and for diagnosis of diseases, linked healthcare records were used to identify prevalent diagnoses at baseline. When a risk factor required for the QRISK3 model could not be perfectly matched to a UK Biobank field, the closest matching field was used.

Outcomes

The outcome of interest is incident CVD defined in the QRISK3 derivation by a composite outcome of coronary heart disease, ischaemic stroke or transient ischaemic attack.⁶ We derived this outcome using International Classification of Diseases and the Office of Population Censuses and Surveys Classification of Interventions and Procedures version 4 codes from hospital episode statistics and death registration data (further detail in online supplemental materials). The follow-up time for each participant was calculated as being the number of years from date of baseline assessment until the earliest date of the following: CVD event date, death date by other causes, date of loss-to-follow-up

or UK Biobank administrative censoring date (England: 2020-11-30; Scotland: 2020-10-31; Wales: 2018-02-28).

Statistical analysis

As in the derivation of QRISK3, participants with missing Townsend deprivation scores were excluded and those with missing data on ethnicity were assumed to be white.⁶ We used multiple imputation (MI) with chained equations to impute missing data on total cholesterol/high-density lipoprotein cholesterol ratio, smoking status, weight, height, systolic blood pressure (SBP) and SBP variability, by gender. To assess the assumption of 'missing at random', we compared baseline characteristics before and after MI (online supplemental table 6). Further details on MI can be found in the online supplemental materials.

Model performance of QRISK3 was assessed by both discrimination and calibration. Discrimination measures the ability to distinguish between low-risk and high-risk patients¹³; patients with higher risk predictions should have higher event rates than those with lower risk predictions.^{14 15} We assessed discrimination overall and in each age group as used in the QRISK3 derivation (35–44, 45–54, 55–64, 65–74 years) at 10 years. We used Harrell's C-statistic, a measure which quantifies the correlation between ranked predicted and observed survival¹⁶ such that a Harrell's C-statistic of 0.5 would indicate that the risk prediction from the model is no better than chance in predicting patient outcomes and 1 would indicate that the model is approaching perfect separation of patient outcomes. Additionally, we used Royston and Sauerbrei's D-index¹⁷ which measures the amount of variation in risk between individuals with low and high predicted risks. Royston's D-index can be interpreted as the log hazard ratio between the low-risk and high-risk groups, with higher values showing greater discrimination; and an increase of 0.1 or more over other scores is an indicator of improved outcome discrimination.^{6 18}

Calibration assesses how well the predicted risk corresponds to the observed risk on a group level.¹⁴ We assessed calibration graphically by comparing the mean predicted risk with the mean observed risk at 10 years, by deciles of the QRISK3 predicted risk distribution. The observed risks were obtained using cumulative incidence Kaplan-Meier estimates at 10 years. Calibration was evaluated overall and in each age group.

All analyses were conducted in R (V.4.1.1).

RESULTS

Data from 502 488 participants in UK Biobank were reviewed for eligibility in this study (figure 1). In line with the QRISK3 derivation exclusion criteria, we excluded 623 participants with missing Townsend deprivation scores, a further 90 296 using statins at baseline and finally 8199 with previous diagnosis of CVD. The median follow-up time was 11.7 years and 92.4% participants had 10 years or more of follow-up. Our analysis population consisted of 233 233 female and 170 137 male participants from UK Biobank. The minimum age of the participants enrolled in UK Biobank was 39.7 years for women and 37.4 for men and the maximum 71.0 and 73.7, respectively. Table 1 shows the baseline characteristics for the included participants and table 2 shows the quantity of missing data by age and sex. Overall, we replaced all missing data with MI and pooled the statistical estimates from 10 imputed datasets using Rubin's rules (further detail in online supplemental materials).

The maximum follow-up time in our study population (after applying exclusion criteria) was 13.95 years, less than the maximum follow-up time of 15 years in the derivation cohort of QRISK3.

DISCRIMINATION

Discrimination of QRISK3 was moderate for both female (Harrell's C-statistic 0.72, D-index 1.28) and male UK Biobank participants (Harrell's C-statistics 0.70, D-index 1.11) (table 3).

The ability of QRISK3 to discriminate was attenuated as age increased (table 4), with the Harrell's C-statistic and D-index, respectively, decreasing from 0.72 and 1.39 in the youngest female participants to 0.62 and 0.67 in the oldest female participants. These measures decreased from 0.73 and 1.34 in the youngest male participants to 0.60 and 0.54 in the oldest male participants. Discrimination was better in female participants than male participants overall and by age.

Calibration

Figure 2 shows the agreement between the 10 year observed and QRISK3 predicted CVD risk, grouped by the decile of the participant's respective QRISK3 score, such that the 10% of participants with the lowest QRISK3 score were binned into the first decile and so forth. The predicted probability within each decile group is plotted as blue squares in figure 2 and was calculated

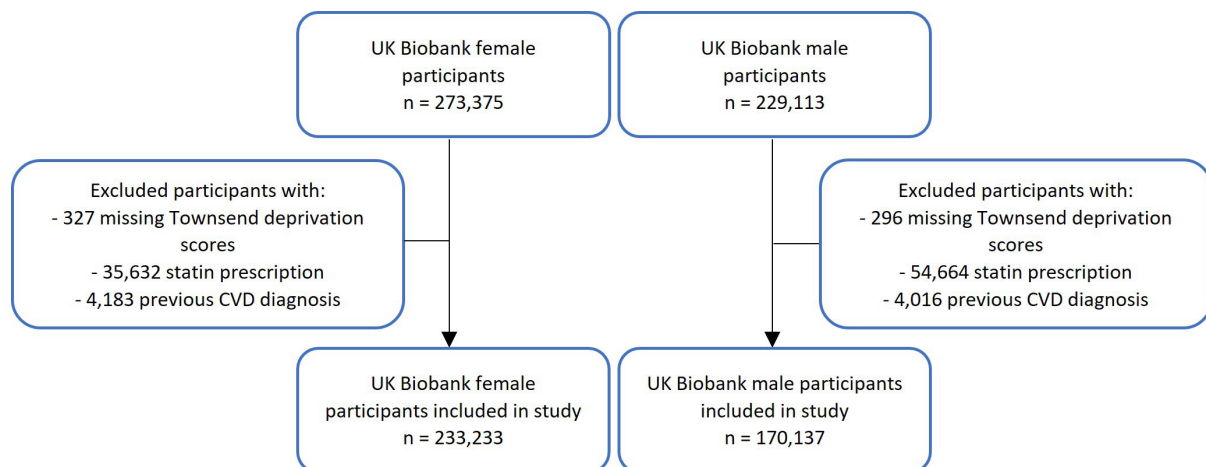


Figure 1 Flow chart of the female and male population used for this study after exclusions for missing Townsend deprivation scores, statin prescription and previous diagnoses of CVD. CVD, cardiovascular disease.

Table 1 Baseline characteristics of all participants in the UK Biobank cohort and the published QRISK3 derivation cohort⁶ by sex

	UK Biobank female (N=233 233)	QResearch derivation female (N=4 019 956)	UK Biobank male (N=170 137)	QResearch derivation male (N=3 869 847)
Age (years), mean (SD)	56.0 (8.0)	43.3 (15.3)	55.8 (8.2)	42.6 (14.0)
Systolic blood pressure (mm Hg), mean (SD)	134.4 (19.2)	123.2 (18.2)	140.4 (17.4)	129.2 (16.3)
Missing	897 (0.4)	692 511 (17.2)	510 (0.3)	1 225 165 (31.7)
Measure of systolic blood pressure variability (SD of repeated measures), mean (SD)	5.4 (4.5)	9.3 (6.2)	5.2 (4.3)	9.9 (6.8)
Missing	897 (0.4)	896 135 (22.3)	510 (0.3)	1 530 945 (39.6)
Total cholesterol: high-density lipoprotein cholesterol ratio (mmol/L), mean (SD)	3.9 (1.0)	3.7 (1.2)	4.6 (1.2)	4.4 (1.4)
Missing	35 880 (15.4)	2 421 398 (60.2)	22 694 (13.3)	2 402 100 (62)
Family history of coronary heart disease	99 896 (42.8)	481 628 (12.0)	61 453 (36.1)	357 987 (9.3)
Self-reported ethnicity				
White or not stated	221 300 (94.9)	3 564 651 (88.7)	161 300 (94.8)	3 435 408 (88.8)
Indian	2255 (1.0)	77 683 (1.9)	1929 (1.1)	81 805 (2.1)
Pakistani	538 (0.2)	39 541 (1.0)	703 (0.4)	46 948 (1.2)
Bangladeshi	46 (<0.1)	31 930 (0.8)	94 (0.1)	42 111 (1.1)
Other Asian	670 (0.3)	53 559 (1.3)	638 (0.4)	45 753 (1.2)
Black Caribbean	2337 (1.0)	37 781 (0.9)	1293 (0.8)	30 610 (0.8)
Black African	1380 (0.6)	77 813 (1.9)	1325 (0.8)	71 245 (1.8)
Chinese	879 (0.4)	33 767 (0.8)	486 (0.3)	23 730 (0.6)
Other ethnic group	3828 (1.6)	103 231 (2.6)	2369 (1.4)	92 237 (2.4)
Townsend deprivation score, mean (SD)	-1.4 (3.0)	0.4 (3.2)	-1.3 (3.1)	0.5 (3.3)
Body mass index (kg/m ²), mean (SD)	26.7 (5.0)	25.4 (5.1)	27.4 (4.02)	25.9 (4.2)
Missing	1148 (0.5)	1 093 554 (27.2)	1098 (0.6)	1 393 672 (36.0)
Smoking status				
Non-smoker	140 470 (60.2)	2 051 803 (51.0)	88 708 (52.1)	1 463 941 (37.8)
Ex-smoker	71 274 (30.6)	589 521 (14.7)	59 143 (34.8)	594 265 (15.4)
Light smoker (<10 a day)	2743 (1.2)	434 954 (10.8)	1656 (1.0)	507 523 (13.1)
Moderate smoker (10–19 a day)	7765 (3.3)	226 128 (5.6)	5705 (3.4)	251 170 (6.5)
Heavy smoker (20 or over a day)	4693 (2.0)	115 890 (2.9)	5927 (3.5)	188 857 (4.9)
Missing	6288 (2.7)	601 660 (15.0)	8998 (5.3)	864 091 (22.3)
Atrial fibrillation	669 (0.3)	15 177 (0.4)	1190 (0.7)	20 098 (0.5)
Erectile dysfunction	NA	NA	698 (0.4)	90 753 (2.3)
Migraine	10 421 (4.5)	257 825 (6.4)	2613 (1.5)	103 995 (2.7)
Rheumatoid arthritis	3239 (1.4)	45 700 (1.1)	1171 (0.7)	20 997 (0.5)
Chronic kidney disease	227 (0.1)	19 396 (0.5)	189 (0.1)	12 254 (0.3)
Severe mental illness	1013 (0.4)	274 069 (6.8)	881 (0.5)	167 115 (4.3)
Systemic lupus erythematosus	440 (0.2)	4010 (0.1)	45 (<0.1)	365 (<0.1)
Diabetes type 1	17 (<0.1)	10 060 (0.3)	16 (<0.1)	11 617 (0.3)
Diabetes type 2	1797 (0.8)	48 022 (1.2)	2161 (1.3)	58 395 (1.5)
Second-generation 'atypical' antipsychotic use	483 (0.2)	19 140 (0.5)	456 (0.3)	20 123 (0.5)
Corticosteroid use	1898 (0.8)	96 955 (2.4)	1253 (0.7)	56 533 (1.5)
Treated hypertension	29 089 (12.5)	223 494 (5.6)	19 167 (11.3)	164 255 (4.2)

Values are number (percentages) unless otherwise stated.

as the average QRISK3 score within that group. The observed 10 year CVD probability within each group is plotted as orange circles in figure 2 and was calculated using the Kaplan-Meier method (to account for right censoring). The plot suggests that QRISK3 systematically overpredicts CVD risk for UK Biobank participants, with the magnitude of overprediction increasing at higher risk deciles.

Figure 3 shows the agreement between the 10 year observed and QRISK3 predicted CVD risk, grouped by the decile of the participant's respective QRISK3 score and presented by age group and can be interpreted in the same way as figure 2. Figure 3 suggests that QRISK3's overprediction of CVD risk for UK Biobank participants may be driven by the increasing magnitude of overprediction for older participants.

DISCUSSION

Principal findings

In this external validation, QRISK3 had moderate ability to discriminate CVD risk for women and men in UK Biobank. The best discriminative accuracy of QRISK3 was seen in the youngest age group for both men and women and discriminative accuracy diminished with age. QRISK3 had poor calibration in UK Biobank, with overprediction of CVD events for both sexes and in all age groups at 10 years of follow-up.

Interpretation of findings

The poor calibration of QRISK3 in UK Biobank contrasts with the findings from the QResearch internal validation cohort⁶ and

Table 2 Quantity of missing data of all participants in the UK Biobank at baseline by sex and age

	Female				Male			
	<45 (N=27 022)	45–54 (N=74 928)	55–64 (N=96 788)	≥65 (N=34 495)	<45 (N=22 482)	45–54 (N=54 093)	55–64 (N=67 247)	≥65 (N=26 313)
Age group (years)								
Systolic blood pressure	136 (0.5)	287 (0.4)	354 (0.4)	120 (0.3)	80 (0.4)	143 (0.3)	187 (0.3)	100 (0.4)
Measure of systolic blood pressure variability	136 (0.5)	287 (0.4)	354 (0.4)	120 (0.3)	80 (0.4)	143 (0.3)	187 (0.3)	100 (0.4)
Total cholesterol: high-density lipoprotein cholesterol ratio	4283 (15.9)	11 794 (15.7)	14 667 (15.2)	5136 (14.9)	3080 (13.7)	7077 (13.1)	8970 (13.3)	3567 (13.6)
Body mass index	151 (0.6)	368 (0.5)	454 (0.5)	175 (0.5)	159 (0.7)	352 (0.7)	422 (0.6)	165 (0.6)
Smoking status	1098 (4.1)	2313 (3.1)	2221 (2.3)	656 (1.9)	1478 (6.6)	2904 (5.4)	3337 (5.0)	1279 (4.9)

Values are number (percentages) unless otherwise stated.

Table 3 Measures of discrimination performance for QRISK3 applied to UK Biobank participants, with 95% CIs

	Female	Male
C-statistic (95% CI)	0.72 (0.72 to 0.73)	0.70 (0.69 to 0.70)
D-index (95% CI)	1.28 (1.25 to 1.31)	1.11 (1.08 to 1.13)

CPRD external validation cohort⁹ that showed good overall calibration. This discrepancy may be due to differences in population characteristics, temporal sequence of risk factors and event occurrence (online supplemental figure 1), and the prevalence of risk factors between populations. The QRisk and CPRD cohorts were derived from primary care databases, with a different case mix to UK Biobank which is known to be less representative of the general UK population, with participants tending to be older, female, live in less socioeconomically deprived areas, be less obese, smoke less, drink less alcohol and have fewer health conditions on average in addition to the evidence for healthy volunteer bias in UK Biobank.¹⁹ Additionally, though the ICD-9 and 10 codes used were the same between QRISK3 and UK Biobank, there are differences in sources used to derive CVD outcomes between the derivation cohort (see online supplemental materials of the study by Hippisley-Cox *et al*⁶) and this study; specifically, we incorporated CVD-related operative procedures from hospital inpatient data and we only used hospital records and death registry to derive the CVD outcome not incorporating CVD diagnosis by GP.

The interactions between age and other predictor variables were already assessed and significant interactions were included in the QRISK3 model,^{4,5} but we still observed increased overprediction of CVD by age when applying the QRISK3 model to the UK Biobank cohort. The greatest overprediction of CVD risk by QRISK3 in UK Biobank was observed in the oldest age group in both sexes. Although it was not the case in the QRisk internal validation study,⁶ poor calibration for older participants was also found in the CPRD external validation cohort.⁹ This may reflect the UK Biobank cohort being considerably older than both the QRisk and CPRD cohorts and that the traditional risk factors in QRISK3 may not account for the unobserved heterogeneity in CVD risk resulting from individuals with the same values of risk factors having different instantaneous CVD risk distribution.²⁰ This heterogeneity may be due to the complex interaction between ageing and cardiovascular physiology, the higher prevalence of undiagnosed or untreated comorbidities in older people and/or the changes in the magnitude of risk factors over the life course.

The discrimination of QRISK3 in UK Biobank is consistent with studies that have applied the model to this population previously,^{21–24} where the overall discrimination is better in younger compared with older participants in UK Biobank. This is to be expected as older participants have generally higher CVD risk, hence a relatively reduced CVD risk range compared with their younger counterparts. Discrimination is a measure of how well the risk prediction equation ranks participants and since age is an important risk factor for CVD, most CVD risk equations are expected to have reduced ability to rank older participants compared with younger ones within the same cohort.

Where results are important for accurate risk prediction and decision-making, poor calibration is arguably more important than discrimination. The consequences of poorly calibrated risk prediction models have been explored more in clinical settings than in epidemiological studies, which may lead to

Table 4 Measures of discrimination performance for QRISK3 in the UK Biobank cohort in each age group, with 95% CIs

Age group (years)	Females				Males			
	<45	45–55	55–65	65–75	35–45	45–55	55–65	>65
Harrell's C-statistic (95% CI)	0.72 (0.69 to 0.75)	0.70 (0.69 to 0.71)	0.66 (0.65 to 0.67)	0.62 (0.61 to 0.63)	0.73 (0.70 to 0.75)	0.67 (0.66 to 0.68)	0.63 (0.63 to 0.64)	0.60 (0.59 to 0.61)
D-index (95% CI)	1.39 (1.20 to 1.58)	1.20 (1.12 to 1.28)	0.92 (0.87 to 0.97)	0.67 (0.61 to 0.73)	1.34 (1.20 to 1.48)	1.01 (0.95 to 1.07)	0.75 (0.71 to 0.79)	0.54 (0.49 to 0.60)

spurious findings in the latter. A model with poor calibration is unlikely to provide true risks when comparing risk estimates between groups in a cohort, and researchers need to be aware of this when reporting their results. The degree to which the performance abilities of a model (when applied to an independent external dataset) should be considered depends on the research objective of the specific study. For studies that require accurate prediction of cardiovascular events in cohorts that are independent from the derivation cohort, there are strategies for minimising the model's inaccuracy and improving model performance depending on the study aims. These strategies include model recalibration, updating the independent variables, considering using an alternative model and collecting the most appropriate risk factors in cohort studies, as discussed further by Parsons, *et al.*²⁵

Researchers working on CVD risk prediction have been moving towards including more personalised factors and using machine learning models, in conjunction with traditional risk factors and statistical models. Such advancements are typically observed in epidemiological cohorts where deep phenotypic and -omics data are available, for example, in UK Biobank. However, these complex models may not be practical when available data for CVD risk prediction are limited to routinely collected data, which is likely the case for healthcare providers in the UK. Future research can focus on improving the consistency and accuracy of CVD risk prediction in both epidemiological studies and routine healthcare.

Strengths and limitations

The major strengths of this study are the large sample size and low loss to follow-up (<0.01% participants). This study followed the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement,²⁶ covering all 22 checklist items that are essential for good reporting of studies that validate multivariable prediction models.

The coding of some variables in UK Biobank data fields does not exactly match the risk factors and outcomes required in QRISK3 which may limit its predictive abilities in UK Biobank. However, our study provides information in assessing QRISK3 in a population that is different from a primary care cohort, as QRISK3 is often applied outside of clinical settings.

CONCLUSION

QRISK3 overpredicts CVD risk for participants of UK Biobank, with the magnitude of this overprediction increasing by age. QRISK3 has moderate overall discrimination for UK Biobank participants, however, the discriminative accuracy of the model declines for older participants. Noting the differences in case-mix between UK Biobank and primary care data, researchers using UK Biobank data that require a CVD risk prediction model that is well calibrated or has good discriminatory prediction for older participants may want to consider recalibrating QRISK3 or using an alternative model.

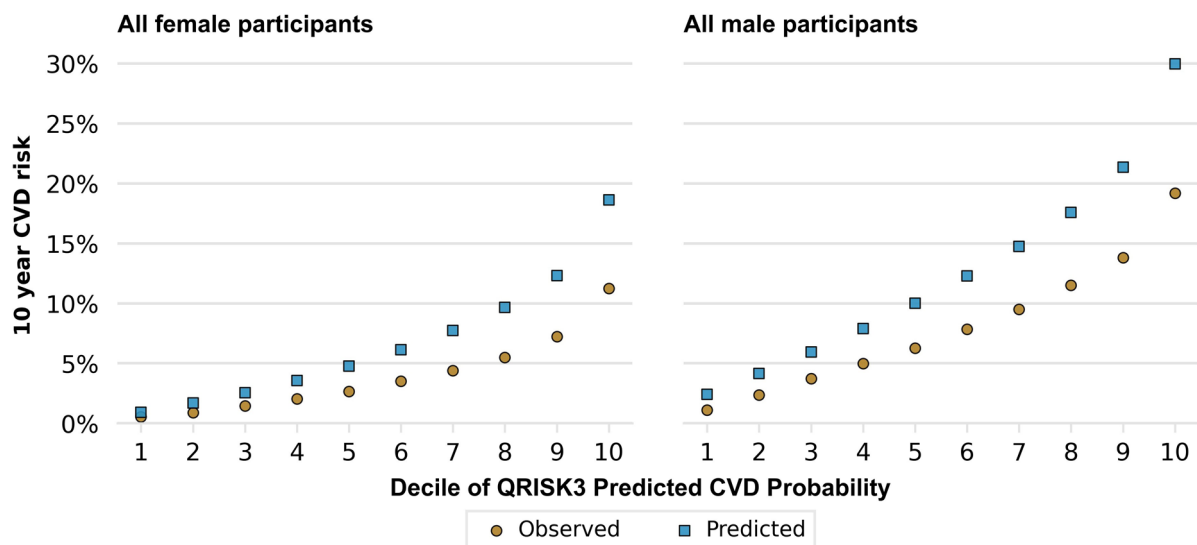


Figure 2 Calibration of QRISK3 at 10 years for female and male participants of UK Biobank overall. The cumulative Kaplan-Meier observed CVD probability in each 10th of risk is denoted by the orange circular markers and the mean predicted QRISK3 score in each 10th of risk is denoted by the blue square markers. CVD, cardiovascular disease.

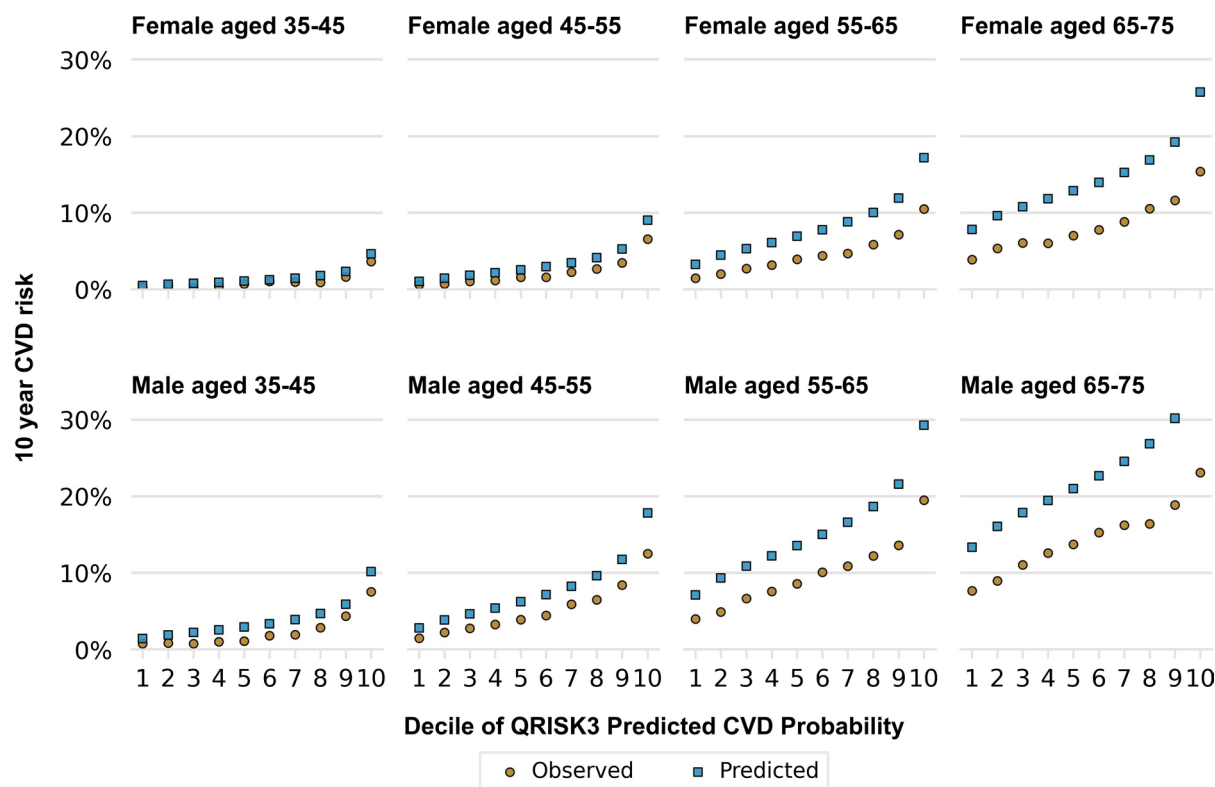


Figure 3 Calibration of QRISK3 at 10 years for female and male participants of UK Biobank in each age group. The cumulative Kaplan-Meier observed CVD probability in each 10th of risk is denoted by the orange circular markers and the mean predicted QRISK3 score in each 10th of risk is denoted by the blue square markers.

Acknowledgements We thank the participants of UK Biobank and the study team for enabling us to conduct this research.

Contributors Conceptualisation—RP, LC, BC; data curation—RP, XL, JC; formal analysis—RP, XL, JC; methodology—RP, LC, XL, JC; supervision—LC, BC, DC; original draft and writing—RP; review and editing—RP, XL, JC, LC, BC, DC; guarantor—RP.

Funding The UK Biobank study was supported by the Wellcome Trust, Medical Research Council, Department of Health, Scottish government, and Northwest Regional Development Agency. It has also received funding from the Welsh Assembly government and British Heart Foundation. DC declares academic grants from GlaxoSmithKline and personal fees from Oxford University Innovation, Biobeats and Sensyne Health, outside the context of this work. DC is funded by an RAEng Research Chair and an NIHR Research Professorship, in addition to support from the NIHR Oxford Biomedical Research Centre, the InnoHK Centre for Cerebro-cardiovascular Engineering, the Oxford Pandemic Sciences Institute and the Oxford-Suzhou Centre for Advanced Research (China).

Competing interests None declared.

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

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the National Health Service’s National Research Ethics Service North West (11/ NW/0382). 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 research has been conducted using the UK Biobank Resource under Application Number 33952. Requests to access the data should be made via application directly to the UK Biobank, <https://www.ukbiobank.ac.uk>.

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

Ruth E Parsons <http://orcid.org/0000-0001-7070-1505>

REFERENCES

- 1 World Health Organization (WHO). WHO reveals leading causes of death and disability worldwide: 2000-2019. World Health Organization (WHO), 2020.
- 2 Wilson PW, D’Agostino RB, Levy D, *et al*. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97:1837–47.
- 3 Conroy RM, Pyörälä K, Fitzgerald AP, *et al*. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003;24:987–1003.
- 4 Hippisley-Cox J, Coupland C, Vinogradova Y, *et al*. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ* 2007;335:136.
- 5 Hippisley-Cox J, Coupland C, Vinogradova Y, *et al*. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of Qrisk2. *BMJ* 2008;336:1475–82.
- 6 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.
- 7 Public Health. NHS health checks: Qrisk3 explained. London: Public Health England, 2021.
- 8 National Institute for Health and Care Excellence. Clinical guideline 181: lipid modification: cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. 2016. Available: <https://www.nice.org.uk/guidance/cg181> [Accessed 13 Aug 2021].
- 9 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.
- 10 Ramspek CL, Jager KJ, Dekker FW, *et al.* External validation of Prognostic models: what, Why, how, when and where. *Clin Kidney J* 2021;14:49–58.
 - 11 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.
 - 12 Allen N, Sudlow C, Downey P, *et al.* UK Biobank: Current status and what it means for epidemiology. *Health Policy and Technology* 2012;1:123–6.
 - 13 Steyerberg EW. Clinical prediction models. In: *Clinical prediction models*. Cham: Springer International Publishing, 2019.
 - 14 Royston P, Altman DG. External validation of a Cox Prognostic model: principles and methods. *BMC Med Res Methodol* 2013;13:33:1–5.
 - 15 Justice AC, Covinsky KE, Berlin JA. Assessing the Generalizability of Prognostic information. *Ann Intern Med* 1999;130:515–24.
 - 16 Harrell FE, Califf RM, Pryor DB, *et al.* Evaluating the yield of medical tests. *JAMA* 1982;247:2543.
 - 17 Royston P, Sauerbrei W. A new measure of Prognostic separation in survival data. *Stat Med* 2004;23:723–48.
 - 18 Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United kingdom: independent and external validation of an updated version of Qrisk2. *BMJ* 2012;344:e4181.
 - 19 Fry A, Littlejohns TJ, Sudlow C, *et al.* Comparison of Sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017;186:1026–34. 10.1093/aje/kwx246 Available: <https://doi.org/10.1093/aje/kwx246>
 - 20 Hougaard P. Frailty models for survival data. *Lifetime Data Anal* 1995;1:255–73.
 - 21 Welsh C, Welsh P, Celis-Morales CA, *et al.* Glycated hemoglobin, Prediabetes, and the links to cardiovascular disease: data from UK Biobank. *Diabetes Care* 2020;43:440–5.
 - 22 Elliott J, Bodinier B, Bond TA, *et al.* Predictive accuracy of a Polygenic risk score-enhanced prediction model vs a clinical risk score for coronary artery disease. *JAMA* 2020;323:636. 10.1001/jama.2019.22241 Available: <https://doi.org/10.1001/jama.2019.22241>
 - 23 Trinder M, Uddin MM, Finneran P, *et al.* Clinical utility of Lipoprotein(A) and LPA genetic risk score in risk prediction of incident Atherosclerotic cardiovascular disease. *JAMA Cardiol* 2021;6:287. 10.1001/jamacardio.2020.5398 Available: <https://doi.org/10.1001/jamacardio.2020.5398>
 - 24 Riveros-Mckay F, Weale ME, Moore R, *et al.* An integrated Polygenic and clinical risk tool enhances coronary artery disease prediction. *Circ Genom Precis Med* 2021;14:e003304.
 - 25 Parsons RE, Colopy GW, Clifton DA, *et al.* Clinical prediction models in Epidemiological studies: lessons from the application of Qrisk3 to UK Biobank data. *J Data Sci* 2022;1–13.
 - 26 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg* 2015;102:148–58.