Integrating genome-wide polygenic risk scores and non-genetic risk factors to develop and validate risk prediction models for colorectal cancer

Sarah E. Briggs[1], Philip Law[2], James E. East[3,4], Sarah Wordsworth[4,5], Malcolm Dunlop[6], Richard Houlston[2], Julia Hippisley-Cox[7]*, Ian Tomlinson[8]*

*these authors contributed equally

Corresponding Authors:

sarah.briggs@ndm.ox.ac.uk; ian.tomlinson@igmm.ed.ac.uk

Affiliations:

[1]Nuffield Department of Medicine, University of Oxford, Oxford, UK; [2]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK; [3]Translational Gastroenterology Unit, Nuffield Department of Medicine, University of Oxford, Oxford, UK; [4]NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK; [5]Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK; [6]Colon Cancer Genetics Group, Medical Research Council Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK; [7]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK; [8]Cancer Research UK Edinburgh Centre, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK.

ABSTRACT

Objective

While population screening programs for cancer colorectal (CRC) have proven benefit, risk-stratified approaches may improve screening outcomes further. To date, genome-wide polygenic risk scores (PRS) for CRC have not been integrated with non-genetic risk factors. We aimed to evaluate several genome-wide approaches, and the benefit of adding PRS to the QCancer-10 (colorectal cancer) non-genetic risk model, to identify those at highest risk of CRC.

Design

Using UK Biobank we developed and compared six different PRS for CRC. The top-performing genome-wide and GWAS-significant PRS were then combined with QCancer-10 and performance compared to QCancer-10 alone.

Results

PRS derived using LDpred2 software performed best, with an odds-ratio per standard deviation of 1.58, and top age- and sex-adjusted C-statistic of 0.733 in logistic regression and 0.724 in Cox regression models in the Geographic Validation Cohort. Integrated QCancer-10+PRS models out-performed QCancer-10, with C-statistics of 0.730 and 0.693, and explained variation of 28.1% and 21.0% from QCancer-10+LDpred2 and QCancer-10 respectively in men; performance improvements in women were similar. Men in the top 20% of risk accounted for 47.6% of cases, and women 42.5% using QCancer-10+LDpred2 models, with a 3.49-fold increase in risk in men and 2.75-fold increase in women in the top 5% of risk, compared to average risk. Decision curve analysis showed that adding PRS to QCancer-10 improved net-benefit and interventions avoided across most probability thresholds.

Conclusion

Integrated QCancer-10+PRS models out-perform existing CRC risk prediction models. Evaluation of risk stratified screening using this approach in a bowel screening population could be warranted.

SUMMARY BOX

What is already known about this subject

- Risk stratification based on genetic or environmental risk factors may improve cancer screening outcomes
- Many polygenic risk scores (PRS) based on a limited number of genome-wide significant SNPs have been assessed in colorectal cancer (CRC), but just two studies have examined the use of genome-wide PRS methodologies
- No previously published study has examined integrated models combining genome-wide PRS and non-genetic risk factors beyond age
- QCancer-10 (colorectal cancer) is the top-performing non-genetic risk prediction model for CRC

What are the new findings?

- PRS derived using LDpred2 software outperform existing models, and other genome-wide and genome-wide significant models evaluated here
- Adding either LDpred2 PRS or genome-wide significant PRS improves the performance and clinical benefit of the QCancer-10 model, with greater gain from the LDpred2 model

How might it impact on clinical practice in the foreseeable future?

- The performance and clinical benefit of QCancer-10 is improved by adding PRS, to a level that suggests utility in stratifying CRC screening and prevention

INTRODUCTION

Colorectal cancer (CRC) is the fourth most common cancer in the UK, with increasing incidence in younger ages and countries with historically lower rates.[1,2] Population screening is effective in reducing CRC incidence and mortality, through detection and removal of pre-malignant adenomas, and earlier detection of cancers.[3] Screening modalities vary internationally. While colonoscopy is the gold-standard, it is expensive, invasive and time consuming. Many countries have adopted a staged process, with initial faecal blood testing, followed by colonoscopy for those who test positive.

Risk-stratified approaches to screening direct resources to those at highest risk have the potential to improve screening detection rates, reduce investigative burden of those at lower risk, and potentially improve cost-effectiveness.[4,5] Improved understanding of cancer risk could also improve informed consent and shared decision making around screening participation.

Both genetic and non-genetic factors contribute to an individual's risk of CRC, some of the latter being modifiable. Genetic variants known to predispose to CRC are mostly single nucleotide polymorphisms (SNPs) identified as significant in genome-wide association studies (GWAS). Genetic risk can be summarised in a polygenic risk score (PRS). Multiple genetic and non-genetic risk models have been developed to predict CRC risk in the general population, and many have been validated in the UK Biobank (UKB).[6,7]

Most existing PRS have combined GWAS-significant SNP genotypes weighted by their effect sizes. While including a greater number of SNPs generally produces better performance, discrimination has generally remained poor.[7] More recently, genome-wide PRS have incorporated many more SNPs than those reaching GWAS-significance. Several genome-wide PRS software tools are now available, with differences in performance across disease types,[8] but evaluation in CRC has been limited.[9,10] In a recent study, a genome-wide model derived using PRS software LDpred, incorporating 1.2 million SNPs, out-performed both machine learning approaches and a 140 GWAS-significant SNP PRS, with an age- and sex-adjusted area under the receiver operating characteristic curve (AUC) of 0.654.[9]

The top-performing non-genetic risk model in external validation is QCancer-10 (colorectal cancer), which has an AUC of 0.70 in men and 0.66 in women aged 40-70 in UKB.[6,11,12] QCancer-10 is a 15-year CRC prediction model, developed using the QResearch linked primary care database of almost 5 million individuals aged 25-84, registered at QResearch practices across England between 1998 and

2013.[11] It is based on age, ethnicity, family history, alcohol and smoking status, a small number of medical conditions, and for men, Townsend deprivation score and body mass index (BMI). As the predictors are derived from electronic health records, it could be embedded at point of care, and linked with screening records to facilitate risk stratification within the bowel screening programme. It thus forms a strong basis for development of an integrated risk prediction model.

Integrated models for CRC, which combine PRS with non-genetic risk factors, generally perform better than PRS alone.[7,13] The top-performing integrated model in external validation in UKB had an AUC of 0.71 in men and 0.67 in women, though this was largely attributable to the non-genetic component, the PRS based on 11 SNPs.[7]

We hypothesised that genome-wide PRS methods, and optimised SNP content based on recent European GWAS, would improve PRS performance, and that integrating this with QCancer-10 should provide for enhanced risk prediction to that afforded by existing models. We therefore developed PRS based on LDpred2, compared this to other genome-wide and GWAS-significant approaches, and validated findings in Geographic and Minority Ethnic Validation Cohorts. We validated QCancer-10, and then derived integrated QCancer-10+PRS risk models, which we internally validated and compared with QCancer-10 alone.

METHODS

Overview

We conducted a development and validation study of PRS utilising multiple methods and integrated PRS-epidemiological models, to predict risk of CRC in a set of UK individuals of bowel cancer screening age. We followed the PRS-RS and TRIPOD reporting guidelines for PRS and prediction modelling.[14,15]

We used UKB to derive and validate our risk models.[16] In brief, just over 500,000 participants (5.5% of invitees) were recruited to UKB from across the UK (between 2006 and 2010). Baseline demographics, medical, lifestyle and physical data, and blood, urine and saliva samples were collected at recruitment. Follow-up through linked hospital and registry data is ongoing. A detailed description of genetic resources including quality control measures can be found in Bycroft et al.[16] and Supplementary Methods.

We calculated age-specific and directly standardised CRC incidence rates in UKB and compared these with Office for National Statistics (ONS) data (Supplementary Methods).[17] See Figure 1 for sample exclusions for the modelling cohorts.

Polygenic risk scores

We meta-analysed 14 CRC GWAS cohorts (which did not include UKB) to provide SNP association effect sizes (see Supplementary Methods and Ref.[18]). The main outcome in all models was CRC diagnosis, identified through self-report at UKB enrolment visit and ICD-9 (153, 154.0, 154.1) and ICD-10 (C18-C20) codes in linked cancer and death registry and hospital data. For PRS development and evaluation in UKB, in logistic regression models we included incident and prevalent cases, with the remaining cohort used as controls. For Cox proportional hazards (Cox) models, prevalent cases with a diagnosis of CRC prior to cohort entry were excluded. Follow-up began at date of enrolment, and was censored at the earliest of date of incident CRC, loss to follow-up, death, or end of available registry follow-up (31st October 2015 for Scottish participants, 13th March 2016 for all other participants).

Three broad approaches to PRS development were evaluated (Supplementary Methods). Firstly, we used a 'standard' PRS (hereafter 'GWAS-sig'), which comprised a manually curated list of 50 sentinel SNPs shown in recent European GWAS-meta-analyses[18,19] to be independently and reproducibly associated with CRC risk at $P<5x10^{-8}$. This PRS was constructed as a log-additive sum of SNP dosages weighted by their betas. Betas were adjusted for winner's curse using FIQT correction.[20]

Secondly, genome-wide clumping and thresholding methodologies were evaluated using 'standard' (C+T) and 'stacked' (SCT) approaches.[21]

Thirdly, we used LDpred2,[8] which takes a Bayesian approach to SNP selection, accounting for linkage disequilibrium between the SNPs. We used three different LDpred2 options – an infinitesimal model (LDpred2-Inf), a non-sparse grid model (LDpred2-grid) and a sparse grid model (LDpred2-grid-sp).

Optimal PRS tuning parameters for genome-wide approaches were selected in the Training Cohort. For each optimal PRS, we then constructed both logistic regression and Cox risk models in the Test cohort with PRS, age, sex, genotyping array and the first four principal components (PCs) from UKB as predictors. We tested for interactions between age and PRS. Training and Test Cohorts included participants of white-British ancestry (identified through self-reported ethnicity and genetic

information)[16] from England and Wales (Figure 1). We compared performance to a 'Null' (reference) model based only on age, sex, genotyping array and four PCs. We also evaluated performance without age and sex in the model. Each model was internally validated, and shrinkage applied to adjust for optimism.

We reported the distribution of standardised PRS and adjusted odd ratios and hazard ratios per-standard deviation (Supplementary methods). We used the C-statistic and Somers' D_{xy} statistic to assess discrimination, in addition to Royston's D statistic and Kaplan-Meier curves across four risk groups for Cox models.[22] Nagelkerke's R^2 was used in logistic regression models and Royston and Sauerbrei's R_D^2 in Cox models to assess variance explained, and R^2attributable to the PRS was calculated by R^2 (full model) - R^2 (null model). Scaled Brier scores were used to assess overall model performance.[23] Confidence intervals and internal validation for all models used 500 bootstrap samples.

Polygenic risk score model validation

PRS models were externally validated in a Geographic Validation Cohort, comprising Scottish participants with European ancestry, and a Minority Ethnic Validation Cohort (from any region).

In addition to the performance metrics described above, calibration was assessed through the calibration slope and visual assessment of calibration plots, with calibration-in-the-large for logistic regression models. For Cox models, calibration plots were created over 5-8 years of follow-up.
In pre-specified subgroup analyses, we calculated performance statistics by sex and in those with a first degree family history of CRC. We evaluated calibration by age by plotting predicted and observed risk across 5-year age bands.

Development of QCancer-10+PRS combined models

The Integrated Modelling Cohort used for QCancer-10 validation and integrated model development comprised all individuals with imputed genetic data passing QC, excluding the 30000 individuals used for PRS hyper-parameter selection (Supplementary Methods), and with complete QCancer-10 predictor data (Figure 1).

Coding of QCancer-10 variables is described in Supplementary Methods. Since missingness was <5% for all predictors (Table S3), complete case analysis was used. The QCancer-10 score was calculated for males and females,[24] and performance evaluated using the published baseline survival functions.

We then recalibrated the models by re-estimating the baseline survival function (recalibration-in-the-large).

Sample size adequacy for integrated model development was calculated following Riley *et al.*[25] (Supplementary Methods). Integrated models included the risk score from QCancer-10 plus the top-performing PRS (based on the maximum C-statistic and R^2 in external validation) using Cox models, developed in men and women separately.

We used the same metrics and time periods to assess the original QCancer-10 model, and QCancer-10+PRS model performance as described for Cox PRS models. In pre-specified subgroup analyses we assessed expected to observed (E/O) ratios of risks and plotted calibration plots in those with a first degree family history of CRC and individuals from minority ethnic backgrounds, and plotted calibration by age.

Model sensitivities were evaluated by calculating the proportion of cases identified at centile thresholds for absolute risk and relative risk. Relative risks were calculated relative to an individual of the same age and sex, mean PRS (by sex), mean PCs, BMI of 25, white ethnicity, mean Townsend score, and no other CRC risk factors. We used decision curve analyses to compare the net benefit and interventions avoided using QCancer-10 and QCancer-10+PRS models.[26] For decision curve and subgroup analyses, QCancer-10+PRS models were first adjusted for optimism, and recalibrated QCancer-10 models were used.

Statistical analysis was performed using R/3.6.2.[27]

Ethics

The UK Biobank study has ethical approval from the North West Multi-centre Research Ethics Committee (16/NW/0274). This study was performed under UK Biobank application number 8508. All contributing GWAS studies were undertaken with ethical review board approval at respective study centres as detailed in Law *et al.*[18]

Patients and public were not involved in the design, conduct or reporting of this study.

RESULTS

Demographics for the UKB-derived Integrated Modelling Cohort are shown in Table 1. The characteristics of each PRS cohort are shown in Table S4. Age-standardised CRC incidence from linked cancer-registry data in the whole UKB cohort was 108.3 and 73.9 cases per 100000 person years at risk for men and women respectively, compared to 127.8 and 80.7 cases per 100000 person years at risk in ONS data.[17] Incidence in the Integrated Modelling Cohort, with cases identified through all linked data, was 118.0 CRCs per 100,000 years follow-up in men and 79.3 in women. Age-specific incidence rates in UKB (Figure S3) closely followed those from ONS until the age of 70, after which UKB rates were lower.

Table 1: Demographic data and medical conditions included in QCancer-10 models, in male and female Integrated Modelling Cohorts, and in cases. Values are numbers (%) unless otherwise indicated. CRC – colorectal cancer, IQR – interquartile range, NA – not applicable. *not included in model for females but provided for information.

	Male whole cohort (n = 196091)	Male cases (n = 1895)	Female whole cohort (n = 238946)	Female cases (n = 1458)
Follow-up (years), median (IQR)	7.1 (1.3)	3.7 (3.5)	7.1 (1.3)	3.8 (3.4)
Age (years), median (IQR)	58 (13)	63 (8)	57 (13)	61 (10)
Geographical region				
East Midlands	13254 (6.8)	127 (6.7)	16175 (6.8)	88 (6.0)
London	25843 (13.2)	203 (10.7)	33080 (13.9)	150 (10.3)
North East	22789 (11.6)	221 (11.7)	27688 (11.6)	174 (11.9)
North West	30259 (15.4)	325 (17.2)	35278 (14.8)	203 (13.9)
Scotland	14690 (7.5)	173 (9.1)	18729 (7.9)	150 (10.3)
South East	16812 (8.6)	165 (8.7)	21367 (9.0)	179 (12.3)
South West	16467 (8.4)	150 (7.9)	21086 (8.8)	136 (9.3)
Wales	8150 (4.2)	90 (4.7)	9942 (4.2)	63 (4.3)
West Midlands	18530 (9.4)	154 (8.1)	19783 (8.3)	128 (8.8)
Yorkshire and Humber	29297 (14.9)	287 (15.1)	35368 (14.8)	187 (12.8)
Ethnicity				
White/not recorded	185813 (94.8)	1848 (97.5)	224316 (94.6)	1399 (96.0)
Indian	2510 (1.3)	11 (0.6)	2601 (1.1)	12 (0.8)
Pakistani	903 (0.5)	1 (0.1)	616 (0.3)	4 (0.3)
Bangladeshi	132 (0.1)	0 (0.0)	61 (0.0)	0 (0.0)
Other Asian	841 (0.4)	2 (0.1)	748 (0.3)	2 (0.1)
Black African	1397 (0.7)	5 (0.3)	1412 (0.6)	6 (0.4)
Caribbean	1363 (0.7)	8 (0.4)	2498 (1.0)	10 (0.7)
Chinese	516 (0.3)	2 (0.1)	865 (0.4)	5 (0.3)
Other ethnic group	2616 (1.3)	18 (0.9)	3980 (1.7)	20 (1.4)
Townsend deprivation index, median (IQR)	-2.18 (4.19)	-2.33 (4.19)	-2.17 (4.09)*	-2.38 (3.96)*
BMI (kg/m2), median (IQR)	27.28 (5.04)	27.92 (5.11)	26.08 (6.23)*	26.42 (6.00)*

Smoking status				
Non-smoker	97088 (49.5)	739 (39.0)	142569 (59.8)	820 (56.2)
Ex-smoker	75100 (38.3)	935 (49.3)	74934 (31.4)	525 (36.0)
Light smoker	9361 (4.8)	84 (4.4)	8885 (3.7)	43 (2.9)
Moderate smoker	5816 (3.0)	43 (2.3)	7235 (3.0)	43 (2.9)
Heavy smoker	8726 (4.4)	94 (5.0)	4873 (2.0)	27 (1.9)
Alcohol intake				
Non-drinker	11985 (6.1)	89 (4.7)	22415 (9.4)	171 (11.7)
Trivial drinker	41810 (21.3)	335 (17.7)	96085 (40.3)	591 (40.5)
Light drinker	57817 (29.5)	521 (27.5)	76942 (32.3)	433 (29.7)
Moderate drinker	60694 (31.0)	624 (32.9)	37830 (15.9)	234 (16.0)
Heavy drinker	14960 (7.6)	205 (10.8)	3797 (1.6)	25 (1.7)
Very heavy drinker	8825 (4.5)	121 (6.4)	1427 (0.6)	4 (0.3)
Medical history				
Ulcerative colitis	1053 (0.5)	17 (0.9)	1211 (0.5)	12 (0.8)
Colorectal polyps	616 (0.3)	11 (0.6)	612 (0.3)	6 (0.4)
Diabetes	12893 (6.6)	184 (9.7)	7885 (3.3)	62 (4.3)
Breast cancer	NA	NA	9448 (4.0)	71 (4.9)
Uterine cancer	NA	NA	1030 (0.4)	16 (1.1)
Ovarian cancer	NA	NA	724 (0.3)	11 (0.8)
Cervical cancer	NA	NA	1711 (0.7)	10 (0.7)
Lung cancer	125 (0.1)	1 (0.1)	NA	NA
Blood cancers	1146 (0.6)	10 (0.5)	NA	NA
Oral cancer	483 (0.2)	12 (0.6)	NA	NA
Family history of CRC	19505 (9.9)	266 (14.0)	22252 (9.3)	169 (11.6)

Polygenic risk score models

Each of 6 PRS models assessed (Figure S4) improved performance over the Null model of age, sex, genotyping array and four PCs (Table 2). A weak interaction between age and PRS was noted (Table S5, Figure S5), but was not included in the models. LDpred2-grid and LDpred2-grid-sp performed best in logistic regression models across all metrics, with similar performance (Table 2). LDpred2-grid had the highest odds ratio per SD of PRS (1.584, 95% CI 1.536-1.633) and C-statistic (0.717, 0.711-0.725) (Table 2). Performance without adjustment for age and sex was considerably worse (Table S6). Internal validation showed low bias in all measures (Table 2).

In the Geographic Validation Cohort, discrimination and variation explained improved compared to the Test Cohort for all models. LDpred2-grid-sp performed best (C-statistic 0.733, 95% CI 0.710-0.753). All models under-predicted risk (CITL >0, Table 2) particularly in the highest PRS groups (Figure S6), and genome-wide models were slightly over-fitted (calibration slope >1, i.e. insufficient variation at the extremes of prediction, Table 2, Figure S6).

In subgroup analyses of logistic regression models (Table S7, Figure S7), discrimination and explained variation were better in males; models were better fitted in females but under-predicted risk to a greater extent, particularly in higher risk groups. Discrimination and variation explained were poorer in individuals with a first-degree family history of CRC, with models systematically underpredicting risk across PRS risk groups. All models tended to under-predict risk across age groups (Figure S8).

Performance was poor in the Minority Ethnic Validation Cohort, with little difference between models. Only LDpred2-grid and LDpred2-grid-sp showed improved performance over the Null model (Table 2). Models systematically under-predicted risk and were highly over-fitted (i.e. predictions were too extreme, Table 2), with modest improvement following recalibration (Figure S6).

In general, PRS performance in Cox models supported the logistic regression analysis (Tables 3 and S8, Figures S9-14). Best performance in external validation occurred with the LDpred2-grid-sp model (C-index 0.725, 95% CI 0.696-0.752).

Table 2: Apparent, internally and externally validated polygenic risk score (PRS) performance in logistic regression models (adjusting for age, sex, array and first 4 principal components). Values are performance indices plus 95% confidence intervals. Internal validation used 500 bootstrap samples. PRS OR per SD – odds ratio per standard deviation of polygenic risk score in the age- and sex-adjusted model; C – C statistic; Dxy – Somers' Dxy rank correlation; R2 – Nagelkerke's R^2(explained variation); Slope – Calibration Slope; CITL – calibration-in-the-large. *All values for R^2 in the Minority Ethnic Validation Cohort were <0 (indicating poorer performance than a model with no explanatory variables). Distributions of standardised PRS are shown in Figure S4.

	LDpred2-inf	LDpred2-grid	LDpred2-grid-sp	SCT	C+T	GWAS-sig	Null
Number of SNPs	1,104,409	1,104,409	616,956	194,756	306,912	50	NA
Apparent performance							
PRS OR per SD	1.435 (1.391 - 1.480)	1.584 (1.536 - 1.633)	1.571 (1.524 - 1.620)	1.417 (1.375 - 1.461)	1.425 (1.382 - 1.470)	1.390 (1.348 - 1.433)	NA
C	0.704 (0.697 - 0.712)	0.717 (0.711 - 0.725)	0.716 (0.710 - 0.723)	0.702 (0.695 - 0.711)	0.704 (0.697 - 0.711)	0.700 (0.693 - 0.707)	0.680 (0.672 - 0.687)
Dxy	0.407 (0.394 - 0.423)	0.435 (0.422 - 0.451)	0.432 (0.419 - 0.446)	0.404 (0.389 - 0.422)	0.407 (0.394 - 0.423)	0.400 (0.386 - 0.414)	0.359 (0.344 - 0.374)
R2 (%)	5.5 (5.1 - 5.9)	6.3 (5.9 - 6.8)	6.2 (5.8 - 6.7)	5.4 (5.0 - 5.9)	5.4 (5.1 - 5.9)	5.3 (4.9 - 5.7)	4.2 (3.8 - 4.6)
R2 PRS (%)	1.3 (1.1 - 1.5)	2.1 (1.9 - 2.4)	2.0 (1.8 - 2.3)	1.2 (1.0 - 1.4)	1.2 (1.0 - 1.5)	1.1 (0.9 - 1.3)	NA
Scaled Brier (%)	0.87	1.05	1.03	0.86	0.85	0.83	0.60
Internal validation							
C	0.703	0.717	0.716	0.701	0.703	0.700	0.679
Dxy	0.406	0.434	0.432	0.403	0.406	0.400	0.358
R2 (%)	5.4	6.3	6.2	5.4	5.4	5.3	4.2
Slope	0.996	0.997	0.998	0.996	0.995	0.999	0.996
Scaled Brier (%)	0.85	0.94	1.06	0.84	0.76	0.85	0.58
Geographic Validation							
C	0.726 (0.704 - 0.748)	0.732 (0.710 - 0.752)	0.733 (0.710 - 0.753)	0.718 (0.696 - 0.739)	0.719 (0.696 - 0.740)	0.703 (0.679 - 0.724)	0.677 (0.654 - 0.699)
Dxy	0.452 (0.408 - 0.496)	0.464 (0.420 - 0.504)	0.466 (0.421 - 0.507)	0.436 (0.392 - 0.477)	0.438 (0.392 - 0.480)	0.405 (0.358 - 0.447)	0.353 (0.308 - 0.397)
R2 (%)	7.0 (5.7 - 8.4)	7.6 (6.1 - 8.9)	7.6 (6.1 - 8.9)	6.4 (5.0 - 7.7)	6.6 (5.2 - 7.9)	5.4 (4.0 - 6.7)	3.8 (2.6 - 5.0)
Slope	1.137 (1.010 - 1.268)	1.091 (0.967 - 1.199)	1.104 (0.980 - 1.213)	1.076 (0.946 - 1.200)	1.098 (0.958 - 1.222)	0.994 (0.861 - 1.113)	0.936 (0.795 - 1.075)
CITL	0.206 (0.120 - 0.272)	0.198 (0.113 - 0.262)	0.199 (0.115 - 0.264)	0.194 (0.107 - 0.260)	0.195 (0.110 - 0.261)	0.191 (0.104 - 0.258)	0.191 (0.105 - 0.257)
Scaled Brier (%)	1.48	1.64	1.66	1.27	1.38	1.08	0.67
Minority Ethnic Validation*							
C	0.588 (0.545 - 0.627)	0.602 (0.558 - 0.640)	0.601 (0.559 - 0.640)	0.589 (0.546 - 0.626)	0.597 (0.554 - 0.636)	0.587 (0.543 - 0.624)	0.585 (0.542 - 0.623)
Dxy	0.176 (0.090 - 0.254)	0.203 (0.116 - 0.279)	0.203 (0.118 - 0.281)	0.179 (0.093 - 0.253)	0.195 (0.108 - 0.271)	0.174 (0.086 - 0.247)	0.171 (0.084 - 0.245)
Slope	0.175 (0.096 - 0.258)	0.204 (0.122 - 0.288)	0.208 (0.126 - 0.294)	0.161 (0.088 - 0.240)	0.195 (0.110 - 0.281)	0.143 (0.071 - 0.213)	0.144 (0.069 - 0.217)
CITL	1.299 (1.155 - 1.417)	1.336 (1.194 - 1.456)	1.325 (1.183 - 1.446)	1.360 (1.217 - 1.479)	1.310 (1.167 - 1.429)	1.392 (1.251 - 1.511)	1.343 (1.200 - 1.459)
Scaled Brier (%)	-0.02	0.04	0.03	-0.06	-0.06	-0.07	-0.15

Table 3: Apparent, internally and externally validated polygenic risk score (PRS) performance in Cox's proportional hazards models (adjusting for age, sex, array and first 4 principal components). Values are performance indices plus 95% confidence intervals are provided for each cohort. PRS HR per SD – adjusted hazard ratio of PRS in model per standard deviation of the PRS; C – Harrell's C index; Dxy – Somers' Dxy rank correlation; D – Royston's D statistic; R2D – Royston and Sauerbrei's R_D^2 (explained variation); Slope – Calibration Slope.

	LDpred2-inf	LDpred2-grid	LDpred2-grid-sp	SCT	C+T	GWAS-sig	Null
Apparent performance							
PRS HR per SD	1.368 (1.310 - 1.428)	1.563 (1.498 - 1.631)	1.545 (1.480 - 1.612)	1.378 (1.321 - 1.438)	1.397 (1.338 - 1.459)	1.377 (1.320 - 1.436)	NA
C	0.696 (0.685 - 0.707)	0.714 (0.704 - 0.726)	0.712 (0.702 - 0.723)	0.695 (0.685 - 0.706)	0.698 (0.689 - 0.709)	0.695 (0.685 - 0.706)	0.675 (0.665 - 0.687)
Dxy	0.391 (0.370 - 0.414)	0.427 (0.409 - 0.451)	0.424 (0.403 - 0.447)	0.391 (0.370 - 0.412)	0.396 (0.378 - 0.417)	0.390 (0.370 - 0.412)	0.350 (0.331 - 0.373)
D	1.085 (1.027 -1.150)	1.201 (1.143 - 1.268)	1.190 (1.132 - 1.255)	1.096 (1.034 - 1.163)	1.099 (1.043 - 1.162)	1.094 (1.031 - 1.162)	0.961 (0.902 - 1.021)
R2D (%)	22.0 (20.1 - 24.0)	25.6 (23.8 - 27.8)	25.3 (23.4 - 27.3)	22.3 (20.3 - 24.4)	22.4 (20.6 - 24.4)	22.2 (20.2 - 24.4)	18.1 (16.3 - 19.9)
Scaled Brier (%)	0.45	0.56	0.55	0.49	0.47	0.50	0.39
Internal validation							
C	0.694	0.713	0.711	0.694	0.697	0.694	0.674
Dxy	0.389	0.425	0.422	0.389	0.393	0.387	0.347
D	1.078	1.194	1.183	1.089	1.091	1.088	0.954
R2D	21.7	25.4	25.1	22.1	22.1	22.0	17.8
Slope	0.992	0.994	0.995	0.994	0.992	0.992	0.992
Scaled Brier (%)	0.44	0.55	0.54	0.47	0.46	0.49	0.38
Geographic Validation							
C	0.715 (0.686 - 0.743)	0.724 (0.696 - 0.751)	0.725 (0.696 - 0.752)	0.713 (0.686 - 0.740)	0.707 (0.681 - 0.734)	0.701 (0.675 - 0.729)	0.673 (0.644 - 0.702)
Dxy	0.430 (0.372 - 0.485)	0.448 (0.391 - 0.501)	0.450 (0.393 - 0.504)	0.426 (0.372 - 0.480)	0.415 (0.361 - 0.468)	0.402 (0.350 - 0.458)	0.345 (0.288 - 0.404)
D	1.243 (1.075 -1.406)	1.285 (1.124 - 1.448)	1.293 (1.130 - 1.460)	1.184 (1.029 - 1.346)	1.182 (1.023 - 1.348)	1.145 (0.992 - 1.319)	0.945 (0.790 - 1.113)
R2D	27.0 (21.7 - 32.1)	28.3 (23.2 - 33.3)	28.5 (23.4 - 33.7)	25.1 (20.2 - 30.2)	25.1 (20.0 - 30.3)	23.8 (19.0 - 29.4)	17.6 (13.0 - 22.9)
Slope	1.123 (0.950 - 1.291)	1.058 (0.911 - 1.204)	1.073 (0.925 - 1.220)	1.070 (0.919 - 1.234)	1.054 (0.897 - 1.223)	1.023 (0.869 - 1.204)	0.947 (0.774 - 1.142)
Scaled Brier (%)	0.75	0.76	0.78	0.63	0.61	0.59	0.37
Minority Ethnic Validation							
C	0.647 (0.593 - 0.700)	0.666 (0.610 - 0.720)	0.664 (0.609 - 0.718)	0.650 (0.596 - 0.705)	0.658 (0.606 - 0.710)	0.659 (0.605 - 0.715)	0.647 (0.595 -0.702)
Dxy	0.293 (0.185 - 0.399)	0.331 (0.221 - 0.440)	0.329 (0.219 - 0.437)	0.300 (0.192 - 0.410)	0.316 (0.212 - 0.420)	0.319 (0.210 - 0.430)	0.293 (0.189 - 0.403)
D	0.931 (0.650 - 1.273)	1.033 (0.736 - 1.374)	1.030 (0.734 - 1.363)	0.940 (0.640 - 1.281)	0.981 (0.682 - 1.320)	0.995 (0.693 - 1.335)	0.889 (0.610 - 1.229)
R2D	17.2 (9.2 - 27.9)	20.3 (11.5 - 31.1)	20.2 (11.4 - 30.7)	17.4 (8.9 - 28.1)	18.7 (10.0 - 29.4)	19.1 (10.3 - 29.9)	15.9 (8.1 - 26.5)
Slope	0.262 (0.161 - 0.397)	0.314 (0.205 - 0.452)	0.318 (0.207 - 0.455)	0.252 (0.154 - 0.384)	0.297 (0.188 - 0.442)	0.251 (0.151 - 0.389)	0.232 (0.136 - 0.366)
Scaled Brier (%)	0.16	0.26	0.26	0.16	0.21	0.19	0.14

QCancer-10 non-genetic model

QCancer-10 models risk in males and females separately. Comparative demographics of the original QCancer-10 derivation cohort and the Integrated Modelling Cohort are shown in Table S9. Performance of QCancer-10 in UKB was in line with previously published studies (Table 4).[6] As expected, the model for females performed less well than the model for males.[6] Both models tended to over-predict risk, which was corrected through recalibration, though in women the model continued to over-predict in the top risk decile (Figure S15). In subgroup analysis, models were well calibrated across age groups; they underpredicted risk in individuals from minority ethnic backgrounds; and the model for females tended to over-predict risk in those with a first-degree family history of CRC, particularly in higher risk groups (Table S10, Figures S16-S17).

Table 4: Apparent and internally validated performance of QCancer-10+LDP and QCancer-10+GWS models, compared with external validation of QCancer-10 in the same participants. Values are performance indices plus 95% confidence intervals. QCancer-10/PRS HR per SD – adjusted hazard ratio of QCancer-10 score or PRS in model per standard deviation of the PRS; C – Harrell's C statistic; Dxy – Somers' Dxy rank correlation; D – Royston's D statistic; R2D – Royston and Sauerbrei's R_D^2 (explained variation); Slope – Calibration Slope. *modelled using multiple fractional polynomial and therefore not presented

	QCancer-10+LDP		QCancer-10+GWS		QCancer-10
	Apparent	Internal Validation	Apparent	Internal Validation	
Males					
QCancer-10 HR per SD	2.309 (2.164 - 2.464)	NA	2.334 (2.187 - 2.490)	NA	NA
PRS HR per SD	1.592 (1.520 - 1.668)	NA	1.453 (1.388 - 1.521)	NA	NA
C	0.730 (0.720 - 0.741)	0.730	0.717 (0.707 - 0.727)	0.716	0.693 (0.682 - 0.704)
Dxy	0.460 (0.440 - 0.481)	0.460	0.433 (0.414 - 0.455)	0.433	0.847 (0.841 - 0.852)
D	1.282 (1.224 - 1.341)	1.280	1.209 (1.156 - 1.267)	1.207	1.058 (0.987 - 1.121)
R2D (%)	28.2 (26.3 - 30.0)	28.1	25.9 (24.2 - 27.7)	25.8	21.1 (18.9 - 23.1)
Scaled Brier (%)	0.81	0.80	0.81	0.80	0.59
Slope	NA	0.998	NA	0.998	0.995 (0.914 - 1.063)
Females					
QCancer-10 HR per SD	*	NA	1.764 (1.655 - 1.881)	NA	NA
PRS HR per SD	*	NA	1.359 (1.290 - 1.431)	NA	NA
C	0.686 (0.672 - 0.701)	0.685	0.668 (0.655 - 0.683)	0.668	0.645 (0.631 - 0.659)
Dxy	0.372 (0.345 - 0.402)	0.37	0.337 (0.310 - 0.366)	0.335	0.822 (0.816 - 0.830)
D	1.056 (0.979 - 1.136)	1.055	0.925 (0.850 - 1.000)	0.924	0.769 (0.695 - 0.847)
R2D (%)	21.0 (18.6 - 23.5)	21	17.0 (14.7 - 19.3)	16.9	12.4 (10.3 - 14.6)
Scaled Brier (%)	0.34	0.34	0.29	0.28	0.20
Slope	NA	0.996	NA	0.996	0.805 (0.724 - 0.899)

QCancer-10+PRS models

We selected LDpred2-grid-sp as the top-performing genome-wide PRS for integrated modelling with QCancer-10, favouring sparsity over the non-sparse model (see Supplementary Results for full model specifications and baseline hazards).[28] Cox models combining the QCancer-10 risk score with LDpred2-grid-sp (QCancer-10+LDP), and the GWAS-sig PRS (QCancer-10+GWS) both out-performed QCancer-10 (Table 4, Figure 2). Internal validation of the QCancer-10+PRS models showed very little optimism in performance estimates.

Models predicting risk in men had better discrimination, and explained more of the variation in risk than models for women (Table 4). Calibration by age was good (Figure S16), with slight under-prediction of risk in the top age group in women. As with QCancer-10, in those with a first degree family history of CRC, female QCancer-10+PRS models tended to over-predict risk, particularly in higher risk groups; male QCancer-10+PRS models were well calibrated (Table S10, Figure S17). In minority ethnicities, QCancer-10+PRS models underpredicted risk (Table S10) to a greater extent than QCancer-10, subject to the *caveat* of a low CRC case numbers (46 men, 58 in women) in this subgroup.

QCancer-10+LDP consistently provided the best risk prediction. Individuals predicted to be in the top 20% of absolute risk by QCancer-10+LDP accounted for 47.6% of male cases and 42.4% of female cases (Table 5). Men in the top 5% of risk had >3.49-fold increased absolute 5-year risk compared to the median, with a comparable 2.75-fold increase in women (see Table S12 for other models). QCancer-10, QCancer-10+GWS, and LDpred2-grid-sp had lower sensitivity in men than QCancer-10+LDP (Tables S13-S17). In women, QCancer-10+LDP and LDpred2-grid-sp models performed equally well, with higher sensitivity than QCancer-10+GWS and QCancer-10 (see Discussion, Tables S13-S17). Decision curve analyses confirmed that, across a wide range of probability thresholds, QCancer-10+LDP gave greater net benefit than QCancer-10+GWS and QCancer-10 for both men and women (Figure 3), and predicted a greater number of interventions avoided across clinically relevant thresholds.

Table 5: Sensitivity of QCancer-10+LDP models for CRC diagnosis over 5 years of follow-up across the top 25 centiles of absolute risk in males and females

Centile	Population per centile	Absolute 5-year risk centile cut-off (%)	Cases per centile	Cumulative % cases based on absolute risk (sensitivity)
Males				
1	1960	2.75	64	3.4
2	1961	2.34	61	6.6
3	1961	2.10	70	10.3
4	1961	1.94	56	13.3
5	1961	1.81	55	16.2
6	1961	1.71	57	19.2
7	1961	1.63	60	22.4
8	1961	1.55	41	24.6
9	1961	1.49	41	26.8
10	1961	1.43	48	29.3
11	1961	1.38	48	31.8
12	1960	1.33	35	33.6
13	1961	1.29	44	35.9
14	1961	1.25	31	37.5
15	1961	1.21	35	39.3
16	1961	1.17	33	41.0
17	1961	1.14	35	42.8
18	1961	1.11	32	44.5
19	1961	1.08	30	46.1
20	1961	1.05	29	47.6
21	1961	1.02	25	48.9
22	1961	1.00	32	50.6
23	1960	0.97	37	52.6
24	1961	0.95	27	54.0
25	1961	0.93	24	55.3
Females				
1	2384	1.53	58	4.0
2	2385	1.27	50	7.4
3	2385	1.13	48	10.7
4	2385	1.04	37	13.2
5	2385	0.97	37	15.7
6	2385	0.91	31	17.8
7	2385	0.87	39	20.5
8	2385	0.83	33	22.8
9	2385	0.80	29	24.8
10	2385	0.77	40	27.5
11	2385	0.74	26	29.3
12	2385	0.72	25	31.0
13	2385	0.70	19	32.3
14	2385	0.68	24	33.9
15	2385	0.66	22	35.4
16	2385	0.64	24	37.0

17	2385	0.63	14	38.0
18	2385	0.61	23	39.6
19	2385	0.60	22	41.1
20	2385	0.59	19	42.4
21	2385	0.57	17	43.6
22	2385	0.56	19	44.9
23	2385	0.55	20	46.3
24	2385	0.54	29	48.3
25	2385	0.53	15	49.3

By way of illustration, enhanced screening is frequently offered for those with a single first degree relative with CRC (FDRCRC), corresponding to a ~2.2-fold increased risk.[29] QCancer-10+LDP identified 18.2% of men (34.0% of cases) and 7.2% of women (16.5% of cases) as having a relative risk >2.2, of whom 76% and 70% respectively had no FDRCRC (Table S18).

DISCUSSION

We have undertaken the first study to develop and validate new prediction models for colorectal cancer that combine phenotypic risk with genome-wide PRS. Our findings demonstrate that LDpred2 significantly improves prediction of CRC above existing PRS models,[9] with top age- and sex-adjusted C-statistics of 0.733 in logistic regression models and 0.725 in Cox models in the Geographic Validation Cohort. We also show that combining the non-genetic QCancer-10 model with PRS improves model performance and clinical benefit, with greatest improvements seen in the QCancer-10+LDP model. To our knowledge the QCancer-10+LDP models have higher discrimination in UKB than any previously published CRC risk score.[7,12,13]

Our models could be used to improve or instigate risk-stratified CRC screening. QCancer-10[11] has recently been recommended to guide shared decision making around CRC screening.[30] However, our study predicts that QCancer-10+PRS models have a greater net-benefit and avoid more interventions than QCancer-10 across a wide range of clinically-relevant risk thresholds, with the greatest benefit from QCancer-10+LDP. The sensitivities achieved using QCancer-10+PRS exceed those of other integrated models recently validated in UKBioank.[7] The genome-wide SNP genotyping required for LDpred2 is reliably performed from saliva samples, and is rapid, inexpensive and straightforward to analyse. The sensitivities and decision curves provided by QCancer-10+LDP could therefore be used to inform clinical decision making.

Of the PRS methods evaluated, LDpred2-grid and LDpred-grid-sp models had highest discrimination, explained more of the variation in risk, and were well calibrated. The improvement in performance between the derivation and validation cohorts when using the PRS models probably results from lower genetic homogeneity in the latter. The Geographic Validation Cohort was well matched in age to the derivation cohort, but had a higher proportion of women; prevalence of CRC was higher, at 1.79% compared to 1.51% in the derivation cohort. We would expect performance in Northern European individuals in the general population to be similar to that of the Validation Cohort. Validation of the PRS models in a geographically external cohort demonstrates portability of the models.

Strengths of our study include our large GWAS meta-analysis (~68000 individuals) and non-overlap between this and modelling cohorts, thus reducing overfitting of the PRS and performance optimism.[31] We used expected genotype dosages rather than allele counts in each PRS,

incorporating uncertainty in genotype imputation, and applied correction for ascertainment bias to effect sizes in the genome-wide model.[20] Our GWAS-significant PRS used stringent inclusion criteria, including only SNPs which replicated in our base meta-analysis. The evaluation of multiple PRS methodologies, and examination of performance in both cross-sectional and prospective cohorts, with and without adjustment for age and sex,[28] facilitates comparison with previous models.

UKB provides a large sample size, extensive phenotyping, completeness of data recording, and linkage to external datasets. Linkage to cancer registry data in UKB ended in 2015/16 at the time of our study, so we have only been able to follow-up for a median 7 years; updating this will improve risk estimates and permit estimation of risks over a longer follow-up. The UKB age range of ~40-70 is similar to that of bowel cancer screening (50-74 in England and Scotland), but narrower than 25-84 used in the original Qcancer-10 study.[11] However, model performance in UKB is arguably unlikely to reflect relative performance in the general population, for several reasons. Model performance will vary between populations with different prevalence or risk of a disease – known as the 'spectrum effect'. As UKB has a lower incidence of disease than the general population of screening age, one might expect sensitivity to increase (which is of benefit in a screening test) when applied to a population with higher risk.[32] Furthermore, all of our models appeared to perform less well in females. For PRS models, wide confidence intervals in the Geographic Validation Cohort mean this finding should be interpreted with caution, but for models that include QCancer-10, this difference was not unexpected. The known healthy volunteer bias that exists in UKB is especially marked in women (for example, the reduction in all-cause mortality and overall cancer incidence in UKB relative to the general population is greater for women than men).[33] The QCancer-10 model has previously been shown to perform less well when validated in UKB than in the QResearch validation.[11] This is likely to be largely due to the differences in age distribution between the general population sample used to develop the original QCancer-10 score and the more restricted UKB sample in this study. External validation of a separate QCancer (colorectal) score for symptomatic patients (rather than the asymptomatic score evaluated here) in an independent population-based cohort showed comparable performance to the discovery study.[34] Overall, risk model performance should be validated in a population representative of the screening population, and we have shown that PRS calibration can be largely corrected in new (ethnically similar) populations by recalibration.

Further limitations of our study may include unknown differences in the demographics of the contributing base GWAS datasets and UKB. In addition, we did not include Mendelian CRC syndromes in the genetic model, and doing so would almost certainly provide improved

performance. Another limitation of our study, and PRS generally, is that most models are developed in individuals of European ethnicity. Although most CRC risk SNPs appear to be shared across ethnic groups, quantitative risk estimates cannot readily be transferred across populations,[35] and, as anticipated, the PRS performed poorly in the Minority Ethnic Validation Cohort.[36] As minority ethnic populations often have higher CRC associated mortality and lower screening uptake,[37-39] further work is urgently needed to improve PRS for CRC in these already disadvantaged populations.

Recent commentaries have been sceptical about the utility of PRS in cancer prevention and early diagnosis,[40] and implementation of PRS in clinical practice has been limited. However, our risk score predicts that ~10% of the population aged ~40-70 have relative risks of CRC high enough to warrant surveillance under current guidelines used in a familial risk context.[41] Furthermore, in a national population screening programme, a risk score with moderate predictive value has considerable potential for improving performance through risk stratification. For bowel cancer screening, a quantitative FIT score is used to decide who is investigated further by colonoscopy. A risk score could be applied alongside the FIT score to allocate colonoscopy more effectively, thus maintaining universal access to screening whilst improving performance. The risk models constructed here perform at a level that may well be clinically useful.[40] Alongside efforts to improve PRS performance in individuals of diverse ancestry, validation in a cohort representative of the screening population and evaluation in a screening trial are, we believe, warranted to assess the performance, acceptability and cost effectiveness of a mixed genetic and non-genetic risk model in a FIT-based bowel cancer screening programme.

Acknowledgements

This research has been conducted using data from UK Biobank, a major biomedical database, http://www.ukbiobank.ac.uk/. We thank all individuals who agreed to participate in the contributing GWAS studies and in UK Biobank, and the investigators, research associates and wider teams involved in these studies. We thank the authors of LDpred2 for their instructive PRS tutorial and code.

Declaration of interests

JHC is director of the QResearch database – a not-for-profit collaboration between University of Oxford and EMIS (commercial supplier of NHS computer systems). She is founder and shareholder of ClinRisk Ltd and was its medical director until June 2019. ClinRisk Ltd supplies free open-source software for research purposes. It also licenses other closed source software to implement risk prediction tools into NHS computer systems outside the submitted work. She is also an adviser to the CMO in England on cancer screening.

JEE has served on clinical advisory boards for Lumendi, Boston Scientific, and Paion; has served on the clinical advisory board and owns share options in Satisfai Health; and reports speaker fees from Falk. JEE serves on the ACPGBI / BSG guideline group for implementation FIT for the detection of CRC in patients with symptoms suspicious of CRC.

Funding

SEB is supported by an MRC Clinical Research Training Fellowship (MR/P001106/1). JEE and SW receive funding from the NIHR Oxford Biomedical Research Centre (BRC). This work of the Houlston Laboratory (PL, RH) is supported by a grant from Cancer Research UK (CR-UK) (C1298/A25514). JHC received funding from the John Fell Oxford University Press Research Fund, grants from CR-UK grant number C5255/A18085, through the Cancer Research UK Oxford Centre, grants from the Oxford Wellcome Institutional Strategic Support Fund (204826/Z/16/Z) and other research councils, during the conduct of the study. MD is funded by CR-UK Programme Grant C348/A12076. IT is funded by CR-UK Programme Grant C6199/A27327. The research was supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with funding from the NIHR Oxford BRC. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. Funders had no role in the study design; in the collection, analysis and interpretation of the data; in the writing of the report; or in the decision to submit the paper for publication.

Contributors

All authors contributed to study conception and design, with development of PRS and statistical analysis led by SEB, IT and JHC. IT, MD and RH provided data. SB and PL carried out primary data analysis, SEB completed the statistical analysis under supervision of IT and JHC. SB and IT wrote the first draft of the manuscript. All authors contributed to critical revision of the manuscript for important intellectual content, and have read and approved the final version.

Data Sharing Statement

UK Biobank data can be obtained through http://www.ukbiobank.ac.uk/. Genotype data are available in the European Genome-phenome Archive under accession numbers EGAS00001005412, EGAS00001005421, or from the Edinburgh University DataShare Repository (https://datashare.ed.ac.uk/). Finnish cohort samples can be requested from the THL Biobank https://thl.fi/en/web/thl-biobank. PRS SNP inclusion lists and model specifications will be deposited in the PGS catalogue repository (https://www.pgscatalog.org/). Risk scores for UKB participants will be returned to UK Biobank for use by approved researchers.

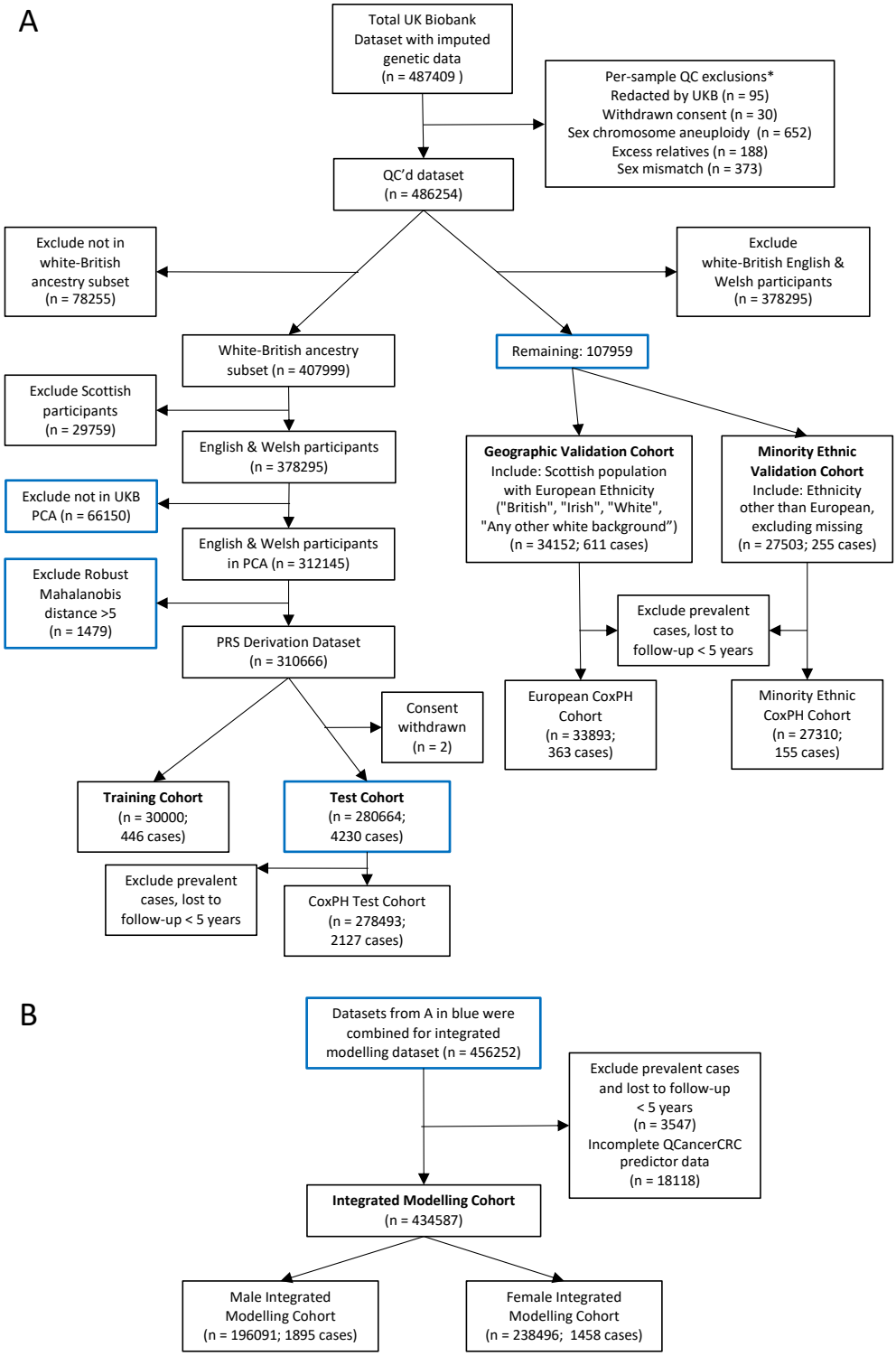

Figure 1. UK Biobank participant flow diagram. Panel A shows quality control and derivation of PRS modelling cohorts. Panel B shows participant selection for the Integrated Modelling Cohorts. *More than one exclusion may apply per person.

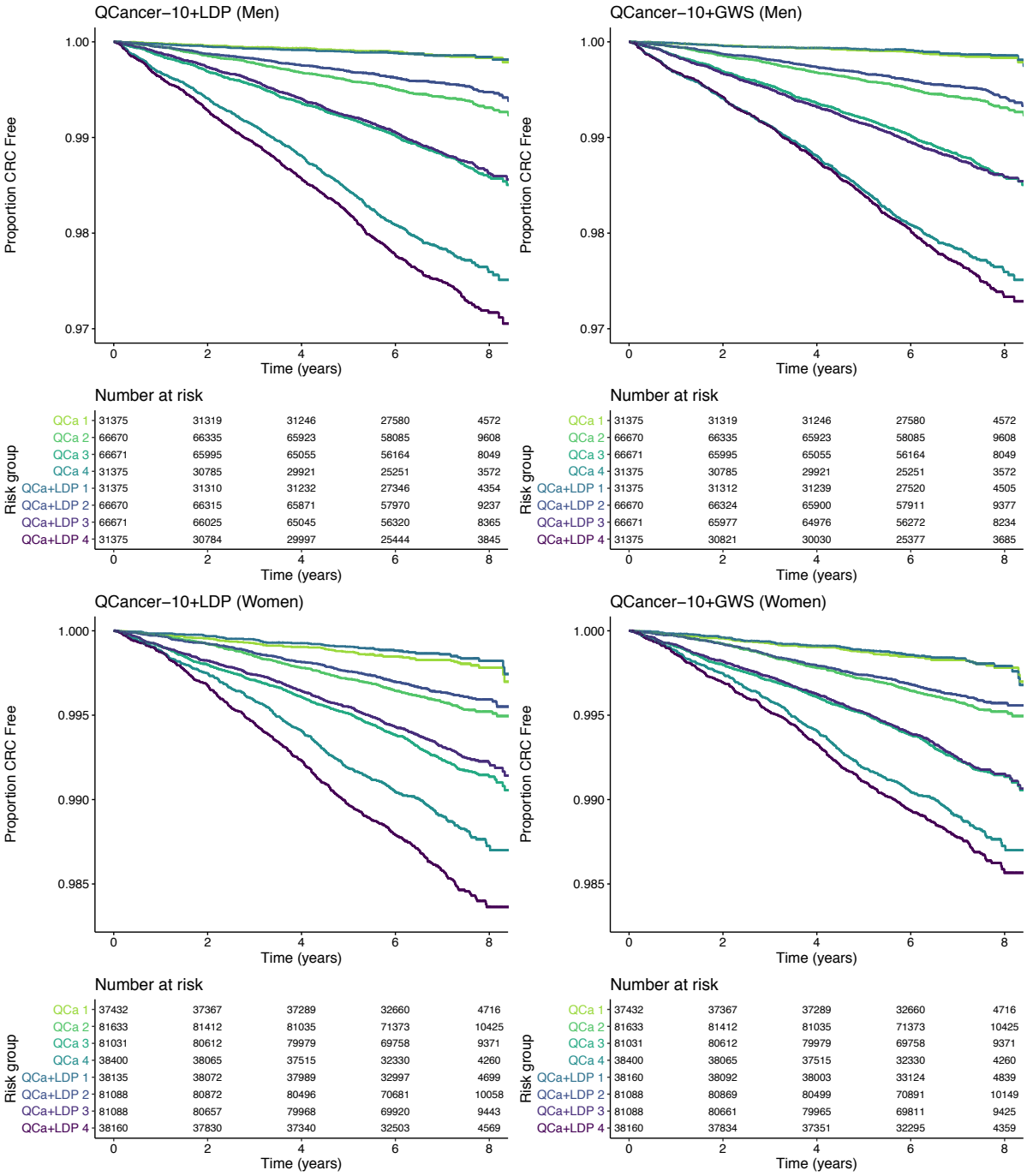

Figure 2. Kaplan-Meier curves across four risk groups (group 4 being highest risk) for QCancer-10+LDP and QCancer-10+GWS models compared to QCancer-10 in men and women.

QCa – QCancer-10 model; QCa+LDP – Qcancer-10+LDP model; QCa+GWS – QCancer-10+GWS model.

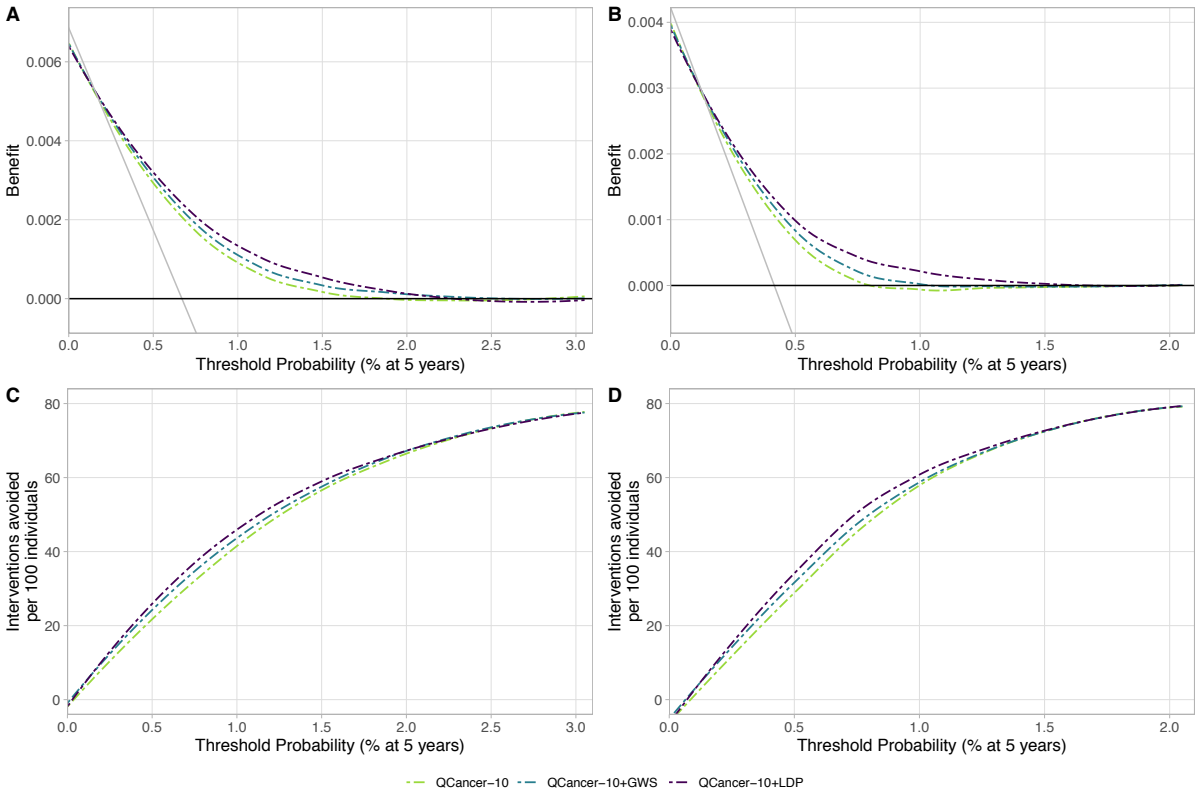

Figure 3. Decision Curve Analysis for QCancer-10+LDP, QCancer-10+GWS, and QCancer-10 models. Figures show net benefit in men (A) and women (B), and interventions avoided per 100 patients tested in men (C) and women (D). The thin grey line in net benefit curves indicates intervention for all, the thick black line no intervention.

REFERENCES:

1. GBD 2017 Colorectal Cancer Collaborators. The global, regional, and national burden of colorectal cancer and its attributable risk factors in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol*. 2019;**4**(12):913-33.DOI: 10.1016/S2468-1253(19)30345-0
2. Vuik FE, Nieuwenburg SA, Bardou M, et al. Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. *Gut*. 2019;**68**(10):1820-6.DOI: 10.1136/gutjnl-2018-317592
3. Cardoso R, Guo F, Heisser T, et al. Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. *Lancet Oncol*. 2021;**22**(7):1002-13.DOI: 10.1016/S1470-2045(21)00199-6
4. Pashayan N, Morris S, Gilbert FJ, Pharoah PDP. Cost-effectiveness and Benefit-to-Harm Ratio of Risk-Stratified Screening for Breast Cancer: A Life-Table Model. *JAMA Oncol*. 2018;**4**(11):1504–10.DOI: 10.1001/jamaoncol.2018.1901
5. Frampton MJ, Law P, Litchfield K, et al. Implications of polygenic risk for personalised colorectal cancer screening. *Ann Oncol*. 2015.DOI: 10.1093/annonc/mdv540
6. Usher-Smith JA, Harshfield A, Saunders CL, et al. External validation of risk prediction models for incident colorectal cancer using UK Biobank. *Br J Cancer*. 2018;**118**(5):750–9.DOI: 10.1038/bjc.2017.463
7. Saunders CL, Kilian B, Thompson DJ, et al. External Validation of Risk Prediction Models Incorporating Common Genetic Variants for Incident Colorectal Cancer Using UK Biobank. *Cancer Prev Res (Phila)*. 2020;**13**(6):509–20.DOI: 10.1158/1940-6207.CAPR-19-0521
8. Prive F, Arbel J, Vilhjalmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics*. 2020;**36**(22-23):5424–31.DOI: 10.1093/bioinformatics/btaa1029
9. Thomas M, Sakoda LC, Hoffmeister M, et al. Genome-wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. *Am J Hum Genet*. 2020;**107**(3):432–44.DOI: 10.1016/j.ajhg.2020.07.006
10. Fritsche LG, Patil S, Beesley LJ, et al. Cancer PRSweb: An Online Repository with Polygenic Risk Scores for Major Cancer Traits and Their Evaluation in Two Independent Biobanks. *Am J Hum Genet*. 2020;**107**(5):815–36.DOI: 10.1016/j.ajhg.2020.08.025
11. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithms to estimate future risk of common cancers in men and women: prospective cohort study. *BMJ Open*. 2015;**5**(3):e007825.DOI: 10.1136/bmjopen-2015-007825
12. Smith T, Muller DC, Moons KGM, et al. Comparison of prognostic models to predict the occurrence of colorectal cancer in asymptomatic individuals: a systematic literature review and external validation in the EPIC and UK Biobank prospective cohort studies. *Gut*. 2018.DOI: 10.1136/gutjnl-2017-315730
13. Kachuri L, Graff RE, Smith-Byrne K, et al. Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nat Commun*. 2020;**11**(1):6084.DOI: 10.1038/s41467-020-19600-4
14. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD). *Ann Intern Med*. 2015;**162**(10):735–6.DOI: 10.7326/L15-5093-2
15. Wand H, Lambert SA, Tamburro C, et al. Improving reporting standards for polygenic scores in risk prediction studies. *Nature*. 2021;**591**(7849):211–9.DOI: 10.1038/s41586-021-03243-6
16. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;**562**(7726):203–9.DOI: 10.1038/s41586-018-0579-z
17. Office for National Statistics. Cancer registration statistics, England [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland. Accessed September 2020].

18. Law PJ, Timofeeva M, Fernandez-Rozadilla C, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat Commun*. 2019;**10**(1):2154.DOI: 10.1038/s41467-019-09775-w
19. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet*. 2019;**51**(1):76–87.DOI: 10.1038/s41588-018-0286-6
20. Bigdeli TB, Lee D, Webb BT, et al. A simple yet accurate correction for winner's curse can predict signals discovered in much larger genome scans. *Bioinformatics*. 2016;**32**(17):2598–603.DOI: 10.1093/bioinformatics/btw303
21. Prive F, Vilhjalmsson BJ, Aschard H, Blum MGB. Making the Most of Clumping and Thresholding for Polygenic Scores. *Am J Hum Genet*. 2019;**105**(6):1213–21.DOI: 10.1016/j.ajhg.2019.11.001
22. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol*. 2013;**13**:33.DOI: 10.1186/1471-2288-13-33
23. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the Performance of Prediction Models A Framework for Traditional and Novel Measures. *Epidemiology*. 2010;**21**(1):128–38.DOI: 10.1097/EDE.0b013e3181c30fb2
24. ClinRisk Ltd. QCancer®(15yr,colorectal). 2015 https://qcancer.org/15yr/colorectal/.
25. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;**368**.DOI: ARTN m441 10.1136/bmj.m441
26. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;**26**(6):565-74.DOI: 10.1177/0272989X06295361
27. R Core Team. R: A language and environment for statistical computing. 2019 https://www.R-project.org/.
28. Janssens A, Joyner MJ. Polygenic Risk Scores That Predict Common Diseases Using Millions of Single Nucleotide Polymorphisms: Is More, Better? *Clin Chem*. 2019;**65**(5):609–11.DOI: 10.1373/clinchem.2018.296103
29. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol*. 2001;**96**(10):2992–3003.DOI: 10.1111/j.1572-0241.2001.04677.x
30. Helsingen LM, Vandvik PO, Jodal HC, et al. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. *BMJ*. 2019;**367**:l5515.DOI: 10.1136/bmj.l5515
31. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, Visscher PM. Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet*. 2013;**14**(7):507–15.DOI: 10.1038/nrg3457
32. Usher-Smith JA, Sharp SJ, Griffin SJ. The spectrum effect in tests for risk prediction, screening, and diagnosis. *BMJ*. 2016;**353**:i3139.DOI: 10.1136/bmj.i3139
33. 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**(9):1026–34.DOI: 10.1093/aje/kwx246
34. Collins GS, Altman DG. Identifying patients with undetected colorectal cancer: an independent validation of QCancer (Colorectal). *Br J Cancer*. 2012;**107**(2):260-5.DOI: 10.1038/bjc.2012.266
35. Wojcik GL, Graff M, Nishimura KK, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019;**570**(7762):514–8.DOI: 10.1038/s41586-019-1310-4
36. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;**51**(4):584–91.DOI: 10.1038/s41588-019-0379-x
37. Carethers JM, Doubeni CA. Causes of Socioeconomic Disparities in Colorectal Cancer and Intervention Framework and Strategies. *Gastroenterology*. 2020;**158**(2):354–67.DOI: 10.1053/j.gastro.2019.10.029

38. Campbell C, Douglas A, Williams L, et al. Are there ethnic and religious variations in uptake of bowel cancer screening? A retrospective cohort study among 1.7 million people in Scotland. *BMJ Open*. 2020;**10**(10):e037011.DOI: 10.1136/bmjopen-2020-037011
39. Hirst Y, Stoffel S, Baio G, McGregor L, von Wagner C. Uptake of the English Bowel (Colorectal) Cancer Screening Programme: an update 5 years after the full roll-out. *Eur J Cancer*. 2018;**103**:267–73.DOI: 10.1016/j.ejca.2018.07.135
40. Sud A, Turnbull C, Houlston R. Will polygenic risk scores for cancer ever be clinically useful? *NPJ Precis Oncol*. 2021;**5**(1):40.DOI: 10.1038/s41698-021-00176-1
41. Monahan KJ, Bradshaw N, Dolwani S, et al. Guidelines for the management of hereditary colorectal cancer from the British Society of Gastroenterology (BSG)/Association of Coloproctology of Great Britain and Ireland (ACPGBI)/United Kingdom Cancer Genetics Group (UKCGG). *Gut*. 2020;**69**(3):411–44.DOI: 10.1136/gutjnl-2019-319915