External Validation of Risk Prediction Models Incorporating Common Genetic Variants for Incident Colorectal Cancer Using UK Biobank

Catherine L. Saunders¹, Britt Kilian¹, Deborah J. Thompson², Luke J. McGeoch³, Simon J. Griffin¹, Antonis C. Antoniou², Jon D. Emery¹, Fiona M. Walter¹, Joe Dennis², Xin Yang², and Juliet A. Usher-Smith¹

ABSTRACT

The aim of this study was to compare and externally validate risk scores developed to predict incident colorectal cancer that include common genetic variants (SNPs), with or without established lifestyle/environmental (questionnaire-based/classical/phenotypic) risk factors. We externally validated 23 risk models from a previous systematic review in 443,888 participants ages 37 to 73 from the UK Biobank cohort who had 6-year prospective follow-up, no prior history of colorectal cancer, and data for incidence of colorectal cancer through linkage to national cancer registries. There were 2,679 (0.6%) cases of incident colorectal cancer. We assessed model discrimination using the area under the operating characteristic curve (AUC) and relative risk calibration. The AUC of models including only SNPs increased with the number of included SNPs and was similar in men and women: the model by Huyghe with 120 SNPs had the highest AUC of 0.62 [95% confidence interval (CI), 0.59–0.64] in women and 0.64 (95% CI, 0.61–0.66) in men. Adding phenotypic risk factors without age improved discrimination in men but not in women. Adding phenotypic risk factors and age increased discrimination in all cases (P < 0.05), with the best performing models including SNPs, phenotypic risk factors, and age having AUCs between 0.64 and 0.67 in women and 0.67 and 0.71 in men. Relative risk calibration varied substantially across the models. Among middle-aged people in the UK, existing polygenic risk scores discriminate moderately well between those who do and do not develop colorectal cancer over 6 years. Consideration should be given to exploring the feasibility of incorporating genetic and lifestyle/environmental information in any future stratified colorectal cancer screening program.

Introduction

Colorectal cancer is one of the most commonly diagnosed malignancies worldwide. Incidence is rising in many countries (1), particularly among younger individuals (2). There is good evidence that population screening reduces colorectal cancer incidence and mortality (3–6), and in most countries with a high frequency of colorectal cancer screening has now been introduced (7).

As for other cancers, including breast (8), lung (9), and ovarian (10), there is increasing interest in using risk models to predict who is at highest and lowest risk of developing colorectal cancer and guide screening decisions. By identifying those who are more likely to benefit from screening and inviting them earlier, more frequently or for different screening tests, this approach has the potential to increase net benefit of colorectal cancer screening at lower cost.

We, and others, have previously shown that several risk models based on phenotypic variables have relatively good discrimination in external validation and could, therefore, be used for this purpose (11, 12). Advances in genetic research and technology mean that it will soon be possible to provide a relatively cheap, quick, and accurate assessment of an individual’s genetic risk of colorectal cancer. Models incorporating genetic variables may, therefore, be simpler to implement. Previous work has also shown that using such genetic information to stratify screening has the potential to improve efficiency of screening (13) and reduce the number of individuals screened while still detecting as many cases (14). Of the 29 models identified in our review, however, many had only been assessed in the development population. The comparative performances of these models in the same population and the
added value of including phenotypic risk factors together with genetic risk factors in the UK population are not known.

In order to inform future risk stratified screening approaches in the UK, in this study we assessed the performance of risk scores that include common genetic variants and predict future colorectal cancer, with or without phenotypic risk factors, in a cohort of 443,888 individuals.

The identification of genetic risk factors for common diseases has been disproportionately focused in white/European populations (15). Consequently, there is concern that risk-stratified screening in the UK could increase inequalities in health. We therefore also additionally assessed the performance of genetic risk scores among ethnic minorities in an exploratory analysis.

Materials and Methods

We performed an external validation of genetic (GRS) and combined GRS plus phenotypic risk models for prediction of incident colorectal cancer, following the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) guideline (16).

Validation cohort

For our validation cohort, we used UK Biobank, the largest population-based cohort in the UK (17). Women and men aged 37 to 73 years, who were registered with the National Health Service and lived within approximately 25 miles of one of 22 study assessment centers across the UK, were invited to participate between 2006 and 2011. From 9.2 million invitations, 503,325 were recruited (5.5%) and attended a baseline assessment when data were collected using touchscreen questionnaires, interviews and physical measurements, and blood samples provided.

Colorectal cancer diagnosis in the validation cohort

Colorectal cancer diagnosis is recorded for UK Biobank participants through linkage to national cancer registries. The most recent cancer record in UK Biobank has a diagnosis date of October 27, 2016. We censored all follow-up at March 31, 2016, to ensure that late registrations were not missed. We included 482,089 participants and 842 cases of colorectal cancer in the 18-month follow-up period.

Selection of risk prediction models

In our systematic review (19), we identified 29 genetic only (GRS) or combined GRS and phenotypic models for colorectal cancer from 20 publications. After contacting authors if insufficient data were provided in the original publications, we excluded 11 models because either details of the risk alleles were not available (20, 21), the relative risk parameters associated with predictors included in the model were not published (22–24), they included variables not available in UK Biobank (25–27) or biochemical risk factors (28), or if the model was developed separately for proximal and distal colon and rectal cancers (29).

After these exclusions, we included 17 models from our systematic review (14, 18, 22, 24, 30–38). We additionally separately included the GRS component from five of these models (29–31, 35, 38) that were developed only as combined GRS plus phenotypic models. For the Hsu model (29), only the GRS was included as the GRS plus phenotypic versions were developed separately for proximal and distal colon, and rectal cancers. From the Smith publication, we considered only the model incorporating the Wells phenotypic risk score (39). We further included the GRS plus phenotypic risk model developed by Jenkins and colleagues (40), which was published separately from the genes-alone GRS developed by the same team.

Details of these 23 models, including the study design, development method, and the risk factors used in the prediction of colorectal cancer risk for each, are given in Table 1. Fourteen models were GRSs including only SNPs, four included SNPs plus phenotypic factors, but not age, and five a combination of SNPs, phenotypic factors, and age.

Genotyping

For all UK Biobank participants, blood samples were genotyped using Affymetrix UK BiLEVE Axiom array and Affymetrix UK Biobank Axiom array and imputed to the combined 1000 Genomes Project v.3 and UK10K reference panels using SHAPEIT3 and IMPUTE3. The lowest imputation info score for the SNPs used in these analyses was 0.86 (41).
<table>
<thead>
<tr>
<th>Author, year</th>
<th>Country</th>
<th>Number of SNPs</th>
<th>Method of development of GRS</th>
<th>Developed in original publication as a “genes-only” score</th>
<th>Method of development of combined model</th>
<th>Non-genetic risk factors included in score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abe 2017</td>
<td>Japan (d, v)</td>
<td>11</td>
<td>Unweighted allele counting model</td>
<td>Logistic regression</td>
<td>Age • •</td>
<td></td>
</tr>
<tr>
<td>Dunlop 2013</td>
<td>UK, Canada, Australia, USA and Germany (d), Sweden and Finland (v)</td>
<td>10</td>
<td>Weighted allele model weighted by published log odds</td>
<td>Stratified absolute risks calculated using logistic regression applied to Scottish data</td>
<td>Age • •</td>
<td></td>
</tr>
<tr>
<td>Frampton 2016</td>
<td>UK (v)</td>
<td>37</td>
<td>Weighted allele model weighted by published log odds</td>
<td>Logistic regression</td>
<td>• • •</td>
<td>Referral pattern, alcohol consumption, regular exercise, dietary folate intake</td>
</tr>
<tr>
<td>Hosono 2016</td>
<td>Japan (d, v)</td>
<td>6</td>
<td>Unweighted allele counting model</td>
<td>Logistic regression</td>
<td>Age • •</td>
<td></td>
</tr>
<tr>
<td>Hsu 2015</td>
<td>USA and Germany (d, v)</td>
<td>31</td>
<td>Weighted model weighted by published log odds</td>
<td>Logistic regression</td>
<td>Age • •</td>
<td>Alcohol use, physical exercise, red meat and vegetable intake, and NSAIDs/aspirin use</td>
</tr>
<tr>
<td>Huyghe 2019</td>
<td>European (91.7%) and East Asian (8.3%) (d)</td>
<td>120</td>
<td>Weighted allele model weighted by study derived weights</td>
<td>Logistic regression</td>
<td>Age • •</td>
<td></td>
</tr>
<tr>
<td>Ibanez-Sanz 2017</td>
<td>Spain (d, v)</td>
<td>21</td>
<td>Unweighted allele counting model (weighted allele models weighted by published log-odds and study derived log odds similar so not reported)</td>
<td>Logistic regression</td>
<td>Age • •</td>
<td></td>
</tr>
<tr>
<td>Iwasaki 2017</td>
<td>Japan (d, v)</td>
<td>6</td>
<td>Weighted allele model weighted by study derived log hazards transformed per allele HR</td>
<td>Developed in men only</td>
<td>Age • •</td>
<td>Alcohol</td>
</tr>
</tbody>
</table>

(Continued on the following page)
Table 1. Details of the development and factors included in each of the risk scores included in validation study. (Cont’d)

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Country</th>
<th>Number of SNPs</th>
<th>Method of development of GRS</th>
<th>Developed in original publication as a “genes-only” score</th>
<th>Method of development of combined model</th>
<th>Non-genetic risk factors included in score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenkins 2016 / 2019</td>
<td>Weighted allele model weighted by published log odds</td>
<td>49</td>
<td>Weighted allele model weighted by published log odds</td>
<td>-</td>
<td>Logistic regression</td>
<td>Height, education, history of type 2 diabetes mellitus, alcohol consumption, regular aspirin use, regular NSAID use, smoking, intake of fiber, calcium, folate, processed meat, red meat, fruit, vegetables, total energy, physical activity (both) HRT (women)</td>
</tr>
<tr>
<td>Jeon 2018</td>
<td>Australia, Canada, Germany, Israel, and USA.(d, v)</td>
<td>63</td>
<td>Weighted allele model weighted by study derived estimated regression coefficients</td>
<td>-</td>
<td>Developed separately in women and men</td>
<td>Diabetes, multi-vitamin usage, years of education, alcohol intake, physical activity, NSAID usage, red meat intake, smoking, and estrogen use (women only)</td>
</tr>
<tr>
<td>Smith 2018</td>
<td>UK (d, v)</td>
<td>42</td>
<td>Weighted allele model weighted by published log odds</td>
<td>-</td>
<td>Log GRS combined with predicted log hazard ratio from original model</td>
<td>-</td>
</tr>
<tr>
<td>Wang 2013</td>
<td>Taiwan (d, v)</td>
<td>16</td>
<td>Logistic regression/GRS based on genotypes not alleles</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Xin 2018</td>
<td>China (d, v)</td>
<td>14</td>
<td>Weighted allele models weighted by published log odds and by study derived weights</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Yarnall 2013</td>
<td>UK (v)</td>
<td>14</td>
<td>Weighted allele model/developed with a simulation based procedure using REGENT software</td>
<td>-</td>
<td>OR estimated from simulation based procedure using REGENT software</td>
<td>Alcohol, fiber intake, red meat intake, and physical activity</td>
</tr>
</tbody>
</table>

Abbreviations: BMI, body mass index; CRC, colorectal cancer; d, development; GRS, genetic risk score; NSAID, non-steroidal anti-inflammatory drug; OR, odds ratio; SNP, single-nucleotide polymorphism; v, validation.
Coding of risk models in UK Biobank

Full details of the definition of each risk factor and how we operationalized them in the UK Biobank dataset and handled missing data are given in Supplementary Tables S1 (SNPs) and S2 (phenotypic risk factors).

Ethnicity

Race/ethnicity, hereafter ethnicity, as this is the terminology used in the UK, was quantified using categories from the UK Office of National Statistics 2001 16 group classification (42), summarized in 5 groups: White/European (response options: English/Welsh/Scottish/Northern Irish/Irish, and Any other White background), Mixed (White and Black Caribbean, White and Black African, White and Asian, and Any other Mixed/multiple ethnic background), South Asian (Indian, Pakistani, Bangladeshi, and Any other Asian background), Black (African, Caribbean, and Any other Black/African/Caribbean background), Other (Chinese and Any other ethnic group).

Analysis

For all of the models, we computed the predicted risk score for developing colorectal cancer for each participant using data collected at baseline assessment. Only one score (Dunlop and colleagues) predicted an absolute risk of colorectal cancer over a specific time frame, and all other scores predicted relative, rather than absolute risk (RR).

We assessed model performance in terms of discrimination and relative risk calibration. For GRS scores, we calculated standardized versions of the scores and estimated the odds ratio (OR) for colorectal cancer associated with a 1 standard deviation (SD) increase in the standardized risk scores as a measure of discrimination; discrimination was further assessed using the area under the receiver operating characteristic curve (AUC) for all models.

Calibration was assessed graphically by comparing the expected with observed relative risk of developing colorectal cancer over the 6-year follow-up period. We stratified our analysis cohort by deciles of predicted relative risk (with decile 5 as the baseline). The observed relative risk in each decile was calculated as the proportion of cases in the decile divided by the proportion of cases in the entire sample, divided by the same estimate in decile 5. The expected RR in each decile was calculated as the geometric mean (because log RR rather than RR is normally distributed) of the predicted RR in the decile divided by the mean predicted RR in decile 5.

We also calculated sensitivity, specificity, positive, and negative likelihood ratios (LR+ and LR−) and the positive and negative predictive values (PPV and NPV) using a cutoff value for each risk score chosen such that 10% of the population had values above the cutoff; the procedure was then repeated using cutoffs for which 20%, 80%, and 90% had values above the cutoff.

In a further analysis, we compared the performance of the GRS scores with models additionally including age (categorized as above and below age 60, the current UK colorectal cancer screening threshold age) and family history in order to provide a preliminary evaluation to inform whether there would be a benefit to incorporate GRS into clinical practice.

Our first analysis considering the discrimination of GRSs alone was stratified by ethnicity, and considered all UK Biobank participants without stratification by sex, because of the relatively small numbers of people reporting non-white/European ethnicity in UK Biobank. Discrimination using AUC and RR calibration were assessed among people from all ethnic backgrounds and stratified by sex. All analyses were carried out in Stata 15 (StataCorp 2017).

Missing data

For our primary analyses, we used a “complete-case” approach including only those individuals for whom all of the risk factors in a particular prediction model were available. The sample size is therefore consistent across all genetic models, except for the model of Wang and colleagues that incorporated genotypes which, in a small number of samples, could not be ascertained from imputed allele counts, and the model of Huyghe and colleagues as described above. Sample sizes varied between models which included non–genetic risk factors, as the amount of missing data for phenotypic risk factors—particularly dietary ones—varied between models (Supplementary Table S3).

Sensitivity analyses

We primarily focused our sensitivity analyses on areas not previously addressed (11) and those particularly relevant to genetic risk models. Accordingly, we compared the performance of risk scores after excluding people with high degree of relatedness to other cohort members (people with >10 relatives, estimated using genetic data), and randomly excluding one of each pair of first-degree relatives. We also compared the performance in a cohort excluding all individuals with any cancer diagnosis prior to baseline. For both these analyses, we followed the methods from previous work (43). We carried out an additional sensitivity analysis comparing discrimination for the GRS models among people 60 and under, and over 60. In a further sensitivity analysis, we excluded all people with a history of colonoscopy at baseline.

For GRS plus phenotypic models, we carried out a further sensitivity analysis restricting our sample for the GRS models to those individuals in whom the GRS plus phenotypic version of the risk score could be estimated (i.e., only among those individuals with complete phenotype characterization for each score).

For the Huyghe model (18), we carried out a further sensitivity analysis including the 6 years of follow-up used for the other models. For the model from Abe (30), we additionally compared results including and excluding folate, because of the large amount of missing data for this variable.

Results

The characteristics of the study population are shown in Supplementary Table S4. Compared with those who did not
Saunders et al.

**Table 2.** Genetic risk score (GRS) discrimination/odds ratio (OR) for colorectal cancer per standard deviation (SD) increase in risk score, stratified by self-reported white/non-white ethnicity.

<table>
<thead>
<tr>
<th>Mean standardized GRS (SD), n = 430,511</th>
<th>OR (95% CI) per 1SD of GRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without colorectal cancer (n = 427,908)</td>
<td>Only self-reported white ethnicity (n = 419,579)</td>
</tr>
<tr>
<td>With colorectal cancer (n = 2,603)</td>
<td></td>
</tr>
</tbody>
</table>

- **Abe 2017** -0.01 (1.00) 0.16 (0.98) 1.20 (1.15–1.25) 1.23 (0.98–1.53) 1.18 (1.14–1.22)
- **Dunlop 2013** 0.00 (1.00) 0.22 (0.98) 1.25 (1.20–1.30) 1.22 (0.98–1.52) 1.25 (1.20–1.30)
- **Frampton 2016** 0.00 (1.00) 0.17 (1.00) 1.18 (1.13–1.23) 1.40 (1.12–1.76) 1.18 (1.14–1.23)
- **Franz 2016** 0.00 (1.00) 0.13 (0.97) 1.16 (1.12–1.21) 1.10 (0.88–1.36) 1.14 (1.10–1.19)
- **Hsu 2015** -0.01 (1.00) 0.26 (0.98) 1.30 (1.25–1.35) 1.27 (1.01–1.61) 1.30 (1.25–1.35)
- **Huyghe 2019** -0.01 (1.00) 0.47 (0.99) 1.60 (1.50–1.72) 1.10 (0.73–1.65) 1.60 (1.50–1.72)
- **Ibanez-Sanz 2017** 0.00 (1.00) 0.19 (1.00) 1.20 (1.16–1.25) 1.28 (1.02–1.60) 1.22 (1.17–1.26)
- **Iwasaki 2017** 0.00 (1.00) 0.13 (1.00) 1.15 (1.11–1.20) 0.96 (0.75–1.21) 1.14 (1.10–1.19)
- **Jenkins 2016** -0.01 (1.00) 0.23 (0.97) 1.28 (1.23–1.33) 1.10 (0.88–1.38) 1.26 (1.21–1.31)
- **Jean 2018** 0.00 (1.00) 0.29 (0.99) 1.33 (1.28–1.38) 1.42 (1.13–1.79) 1.33 (1.28–1.39)
- **Smith 2018** -0.01 (1.00) 0.23 (0.98) 1.27 (1.22–1.32) 1.39 (1.13–1.73) 1.26 (1.22–1.31)
- **Wang 2013** 0.00 (1.00) 0.02 (0.98) 1.01 (0.97–1.05) 1.08 (0.85–1.36) 1.02 (0.98–1.06)
- **Xin 2018** 0.00 (1.00) 0.13 (1.00) 1.15 (1.11–1.20) 1.24 (1.04–1.49) 1.14 (1.09–1.18)
- **Yarnali 2013** 0.00 (1.00) 0.19 (1.01) 1.21 (1.17–1.26) 1.22 (0.98–1.53) 1.22 (1.17–1.26)

develop colorectal cancer, those who did were more likely to be male, older, and have a family history of colorectal cancer. Although the number of SNPs varied across models, the median number of risk alleles among colorectal cancer cases was typically one higher than among those who did not develop cancer.

**Discrimination**

**GRS discrimination, stratified by ethnicity**

Among the 441,141 participants who self-reported their ethnicity, 419,579 (95%) were White. Mean standardized GRS among people with colorectal cancer varied between 0.02 and 0.47 in the whole validation cohort (Table 2), and from -1.44 to 0.85 when stratified by ethnicity (Supplementary Table S5). The Huyghe GRS had the highest OR per 1 SD of GRS, 1.60 (1.50–1.72) among all UK Biobank participants, and 1.61 (1.30–1.92) among people reporting a white/European ethnic background. Five of the GRSs showed some discrimination ability among people with non-white ethnicity (refs. 14, 22, 24, 29, 31; Table 2).

**GRS discrimination, stratified by sex**

**Figures 1 and Supplementary Table S6** show the discrimination, measured by AUC, for all models in the whole UK Biobank population, stratified by sex. Discrimination for the 14 GRS models is similar in men and women and ranges from 0.50 to 0.63. The AUC increased with the number of SNPs included within the risk models (Fig. 2), with the 120-SNP model by Huyghe having the highest AUC of 0.62 [95% confidence interval (CI), 0.59–0.64] in women and 0.64 (95% CI, 0.61–0.66) in men. In general, the GRSs based on newer genome-wide association (GWA) studies also performed better than those based on candidate genes or older GWA studies. These analyses stratified by White-European/non-White-European ethnicity are presented in Supplementary Table S7.

GRS-plus phenotype discrimination, stratified by sex

Among the four models that incorporated both genetic and phenotypic risk factors but not age (31, 34, 37, 40), adding phenotypic factors, including body mass index (BMI), smoking, family history, alcohol intake, red meat intake, and physical activity, either did not change the AUC (31, 34) or reduced it (37, 40) in women. In men, however, the AUC improved from 0.55 (0.54–0.57) to 0.59 (0.57–0.6) with the addition of smoking, BMI, alcohol, fiber intake, red meat intake, and physical activity in the Yarnall model (34) and from 0.55 (0.54–0.57) to 0.58 (0.56–0.59) with the addition of family history, BMI, alcohol intake, physical exercise, red meat and vegetable intake, and NSAIDs/aspirin use in the Ibanez-Sanz model (37).

Considering only people in whom it was possible to calculate a risk score based on both the GRS and GRS plus phenotypic versions of the same models confirmed these differences, with significant (P < 0.05) improvements in AUC observed only with the addition of phenotypic risk factors in the Yarnall and Ibanez-Sanz models in men (Supplementary Table S8).

By comparison, the discrimination of models including age was greater (P < 0.05) than for GRSs alone for all models in both women and men, with a relatively greater improvement in performance in men than women (Fig. 1). The models with highest AUC were those developed by Abe, Hosono, Dunlop, and Smith (24, 30, 32, 35), with AUCs between 0.67 and 0.71 in men and 0.64 and 0.67 in women.

The sensitivity, likelihood ratios, and PPV and NPV for thresholds at which varying percentages of the population are classified as high risk are presented in full in Supplementary Tables S9 and S10. Within the 10% of the population with the highest risk, the top performing GRS (18) identified 18.6% of women and 22.3% of men who went on to develop colorectal cancer. Among those with the highest 20% risk, this increased to 30.7% for women and 35.7% for men. For the four best performing GRS plus phenotypic models that also included age (24, 30, 32, 35), in women the top 10% of the population
included between 19.7% and 20.7% of those who went on to develop colorectal cancer and the top 20% 33.8% to 38.1%. In men, the corresponding values are 21.1% to 26.9% and 38.2% to 47.9%. The NPVs were high and comparable (>99.3) for all models.

Calibration

Results from the relative risk calibration are presented for women and men in Figs. 3 and 4 and Supplementary Tables S11 and S12. There is variation in calibration across models, with some models having very poor calibration. There is a
consistent pattern that models tend to overestimate relative risk at higher levels of risk. Calibration did not substantially improve in models based on a GRS plus phenotypic risk factors compared with those that include only a GRS, except for those which incorporated age.

The additional analysis comparing the AUC of models incorporating each GRS, age and family history, with age and family history alone found that for all models except Wang adding the GRS increased the AUC by between 0.01 and 0.03 (Supplementary Table S13), i.e., reducing between 3% and 9% of the remaining error.

Sensitivity analyses
The results from the all the other sensitivity analyses were consistent with the main analysis (Supplementary Tables S14–S19).

Discussion
Key findings
This is, to our knowledge, the first external validation to use a single cohort to directly compare multiple published risk prediction models for colorectal cancer that include common genetic markers. It shows that genetic information alone discriminates moderately well between those who do and do not develop colorectal cancer over a 6-year period. The best performing GRS by Huyghe and colleagues (18) had an AUC of 0.62 in women and 0.64 in men and OR per 1SD of 1.60, comparable with polygenic risk scores in breast cancer [AUC 0.63 (95% CI, 0.62–0.64), OR per 1SD 1.49–1.71 (43)] and coronary artery disease [0.62 (95% CI, 0.62–0.63; ref. 44)] and better than risk models including phenotypic risk factors without age (11, 24).

In contrast to risk models incorporating phenotypic risk markers for which performance in men is better than in women (11, 12), the performance of GRSs was equivalent. Consistent with this, adding phenotypic risk factors without age to a GRS improved the AUC in men but not in women, with the best performing models including SNPs, phenotypic risk factors, and age having AUCs between 0.67 and 0.71 in men and 0.64 and 0.67 in women.

The potential impacts of incorporating these risk scores into practice are best appreciated by the differences in risk classification between the models. Our previous work using UK Biobank showed that using the current age-based English bowel cancer screening threshold of 60 years to identify the population at high risk for colorectal cancer and then random sampling to account for ties, targeting people...
in the top 10% of risk would identify 17% of men and 16% of women who developed colorectal cancer (11). The equivalent proportions using the Huyghe model with the highest discrimination for genetic risk factors alone in this study would be 22% for men and 19% for women, representing an increase in detection of 2% to 5% of cases for the same number of people screened. This is consistent with an analysis by Jenkins and colleagues (13) who showed that, in a hypothetical population based on the Australian population in 2011, inviting individuals for screening based on their genetic risk rather than all those ages between 50 and 74 would result in a 3.1% overall improved efficiency.

Strengths and limitations
The main strengths of this analysis are the use of a large cohort with nearly 3 million person-years of follow-up, comprehensive phenotyping and genotyping and linkage to national cancer registries, and the inclusion of 23 risk models identified from a systematic review (19). We were not only able to perform the largest external validation to date of multiple colorectal cancer risk models in a single population but were also able to compare the performance of models including only SNPs to those including SNPs plus phenotypic risk factors with or without age and family history.

However, despite its many strengths, the UK Biobank has a number of limitations. In particular, the response rate to invitations to take part was only 5.5% (17). As a result, participants are more likely to be older, to be female, and to live in less socioeconomically deprived areas than nonparticipants; incidence rates of colorectal cancer are also lower than in the general population (45). We limited the effect of this “healthy volunteer” bias by restricting our analyses to relative risk and discrimination. Analysis of SNP data may also potentially be less likely to be confounded by this “healthy volunteer” bias than models including phenotypic risk factors as well. The large variation in age at baseline assessment also means that age has a disproportionately large impact on colorectal cancer risk compared with other phenotypic measures. A third limitation is that colorectal cancer cases registered at Scottish cancer registries may not all be complete after March 31, 2015. This may affect up to 10% of the cohort and result in an underestimate of colorectal cancer risk overall. Finally, colorectal cancer screening at age 60 in the UK was rolled out during the period of recruitment to UK Biobank. Some, but not all, members of the cohort will therefore have participated in the
national screening program during this period. Our sensitivity analysis excluding people with a history of colonoscopy at baseline, however, did not change the findings of this analysis.

Although similar to the 2011 UK Census (45), the predominantly White population within the UK Biobank cohort also means that AUC estimates when restricted to participants from ethnic minorities have wide CIs, and assessing the model performance is challenging. Any small differences between White and non-White populations should, therefore, not be overinterpreted. In our primary analysis, we have applied the GRS/GRS plus phenotypic risk models to everyone, regardless of the population in which they were developed. We acknowledge the issue of population stratification leading to spurious associations between genetic risk factors and disease when people from diverse ethnicities are analyzed in the same cohort. Nonetheless, in practice any policy for developing stratified screening will be applied across the whole country; these results present the performance of these models in this scenario.

There are also limitations with our analysis. Firstly, we excluded two models identified from our systematic review because they included variables not present in UK Biobank and we had to derive proxy variables if there were no exact matches for variables included in the risk models. We also only assessed RR calibration. Although this does not allow us to assess how closely the predicted absolute risk matches the observed absolute risk, it illustrates the variation in calibration across the models and highlights that models would need to be calibrated to the population in which they are going to be used.

Comparison with existing literature

When compared with findings of validation studies in external populations or nonrandom split samples (19), our results for discrimination are similar for the genetic risk models by Dunlop, Ibanez-Sanz, and Smith, but lower for the genetic risk models by Hosono and Xin. The lower discrimination seen in this cohort for the Hosono and Xin models likely reflects differences in the ethnicity of the populations: the model by Hosono was developed using logistic regression and subsequently validated in a Japanese cohort and the model by Xin used published genome-wide association studies (GWAS) from European and Asian populations and validated the model in a Chinese cohort. Our results for the Smith model that combined genetic risk factors with phenotypic risk factors and age from the model by Wells were also consistent with the evaluation of that model in an earlier release of UK Biobank data. For the remaining three models (Abe, Dunlop, and Hosono), our results showed better discrimination. In all three cases, the previous validation studies had been performed in case–control studies matched by age, effectively removing the effect of age from the models, and highlighting the strength of age as a risk factor. This effect of age may also explain why Smith and colleagues found in UK Biobank that addition of a GRS to either the models by Taylor (46) or Wells (39), both of which include age, did not improve discrimination and did not result in a substantive change in the predicted probability for the majority of participants (24).

Conclusions and implications for future research

This study shows that existing GRSs are able to discriminate moderately well between those who develop colorectal cancer and those that do not, and unlike phenotypic risk scores, they do not perform differentially in men and women. Genetic risk scores may also be easier to implement than phenotypic risk scores as they do not change over time and are not associated with the same degree of measurement error and bias as phenotypic risk factors. With on-going GWAS and increasingly comprehensive imputation panels that allow for improved low-frequency and rare genetic variant imputation, it is expected that new SNPs associated with colorectal cancer will be identified in the future (14, 47). Alongside these efforts to identify further SNPs, future studies are needed to develop and validate GRSs in non-European populations and model the potential impact of incorporating them into screening programs, as is being done for other cancers (8). There are a number of other issues that also need to be considered before stratified screening based on genetic risk can be implemented. These include the need for fundamental changes in the infrastructure and mechanisms for genetic data collection, storage and sharing (48), and ethical, legal, and social considerations such as equity of access to genetic testing, insurance issues, and whether screening programs should exclude people currently eligible as a result of their age on the basis of low genetic risk. Further research should focus on these areas as well as modeling the potential health benefits and cost effectiveness of implementing stratified genetic risk–based colorectal cancer screening. By demonstrating the current performance of risk scores including genetic risk information in a large UK population, this study supports the need for such research.

Disclosure of Potential Conflicts of Interest

B. Kilian and J.A. Usher-Smith report receiving grants from Cancer Research UK and Bowel Cancer UK. D.J. Thompson reports receiving grants from Cancer Research UK during the conduct of the study; and other support from Genomics plc outside the submitted work. S.J. Griffin reports receiving grants from Bowel Cancer UK. A.C. Antoniou reports receiving other support from Cambridge Enterprise outside the submitted work. J.D. Emery reports receiving grants from Genotype Pty Ltd. outside the submitted work. No potential conflicts of interest were disclosed by the other authors.

Ethical Approval

The UK Biobank study was approved by the North West Multi-Centre Research Ethics Committee (reference number 06/MRE09/65), and at recruitment all participants gave informed written consent to participate in UK Biobank and be followed up, using a signature capture device.

Data Sharing

All data on the risk models are available from the reports or authors of the primary research. All key data fields derived for this study and the underlying STATA code used to generate the main results of
the article will be made available through the UK Biobank access team (access@ukbiobank.ac.uk). No additional data are available.

Authors’ Contributions
Conception and design: S.J. Griffin, J.D. Emery, J.A. Usher-Smith
Development of methodology: C.L. Saunders, D.J. Thompson, A.C. Antoniou, J.D. Emery, J.A. Usher-Smith
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): B. Kilian, L.J. McGeoch, X. Yang
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): C.L. Saunders, B. Kilian, D.J. Thompson, S.J. Griffin, A.C. Antoniou, F.M. Walter, J. Dennis, J.A. Usher-Smith
Writing, review, and/or revision of the manuscript: C.L. Saunders, B. Kilian, D.J. Thompson, L.J. McGeoch, S.J. Griffin, A.C. Antoniou, J.D. Emery, F.M. Walter, X. Yang, J.A. Usher-Smith
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): B. Kilian
Study supervision: S.J. Griffin, J.A. Usher-Smith

Acknowledgments
This research has been conducted using the UK Biobank Resource under Application Number 28126. This work was funded by a grant from Bowel Cancer UK (18PG0008). J.A. Usher-Smith is funded by a Cancer Research UK Prevention Fellowship (C55650/A21464). A.C. Antoniou and X. Yang are supported by a Cancer Research UK grant (C12292/A20861). DJT was supported by a Cancer Research UK grant (C1287/A16563). The University of Cambridge has received salary support in respect of S.J. Griffin from the NHS in the East of England through the Clinical Academic Reserve. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service (NHS), the National Institute for Health Research (NIHR), or the Department of Health. All researchers were independent of the funding body, and the funder had no role in data collection, analysis, and interpretation of data: in the writing of the report; or decision to submit the article for publication.

We thank James Brimicombe for data management support and our patient and public representative Margaret Johnson for providing helpful comments on the findings.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 16, 2019; revised January 15, 2020; accepted February 11, 2020; published first February 18, 2020.

References
Saunders et al.


External Validation of Risk Prediction Models Incorporating Common Genetic Variants for Incident Colorectal Cancer Using UK Biobank

Catherine L. Saunders, Britt Kilian, Deborah J. Thompson, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1940-6207.CAPR-19-0521

Supplementary Material
Access the most recent supplemental material at:
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2020/02/18/1940-6207.CAPR-19-0521.DC1

Cited articles
This article cites 46 articles, 7 of which you can access for free at:
http://cancerpreventionresearch.aacrjournals.org/content/13/6/509.full#ref-list-1

Citing articles
This article has been cited by 4 HighWire-hosted articles. Access the articles at:
http://cancerpreventionresearch.aacrjournals.org/content/13/6/509.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerpreventionresearch.aacrjournals.org/content/13/6/509.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.