Risk prediction models for colorectal cancer incorporating common genetic variants: a systematic review

Luke McGeoch[1], Catherine L Saunders[2], Simon J Griffin[2], Jon D Emery[2,3], Fiona M Walter[2,3], Deborah J Thompson[4], Antonis C. Antoniou[4], Juliet A Usher-Smith[2]

[1]School of Clinical Medicine, University of Cambridge, Cambridge, UK

[2]The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge School of Clinical Medicine, Box 113 Cambridge Biomedical Campus, Cambridge CB2 0SR, UK

[3]Department of General Practice and Centre for Cancer Research, Faculty of Medicine, Dentistry & Health Sciences The University of Melbourne, Victorian Comprehensive Cancer Centre, 305 Grattan St, Melbourne Victoria, 3010, Australia

[4]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Strangeways Research Laboratory, Wort's Causeway, Cambridge, CB1 8RN, UK

Running title: Genetic risk prediction models for colorectal cancer

Key words: Risk, prediction, colorectal cancer, genetics, review

Correspondence to: J Usher-Smith, The Primary Care Unit, Department of Public Health and Primary Care, University of Cambridge School of Clinical Medicine, Box 113 Cambridge Biomedical Campus, Cambridge CB2 0SR, UK Email: jau20@medschl.cam.ac.uk Tel: 01223 748693 Fax: 01223 768412

Competing Interests: The authors declare no potential conflicts of interest

Word count: 5195
Number of tables/figures: 5

ABSTRACT

Colorectal cancer (CRC) screening reduces CRC incidence and mortality. Risk models based on phenotypic variables have relatively good discrimination in external validation and may improve efficiency of screening. Models incorporating genetic variables may perform better. In this review we updated our previous review by searching Medline and EMBASE from the end date of that review (January 2014) to February 2019 to identify models incorporating at least one single nucleotide polymorphism (SNP) and applicable to asymptomatic individuals in the general population. We identified 23 new models, giving a total of 29. Of those in which the SNP selection was based on published GWASs, in external or split-sample validation the AUROC was 0.56-0.57 for models including SNPs alone, 0.61-0.63 for SNPs in combination with other risk factors and 0.56 to 0.70 when age was included. Calibration was only reported for four. The addition of SNPs to other risk factors increases discrimination by 0.01-0.05. Public health modelling studies suggest that, if determined by risk models, the range of starting ages for screening would be several years greater than using family history alone. Further validation and calibration studies are needed alongside modelling studies to assess the population-level impact of introducing genetic risk-based screening programme.

INTRODUCTION

Colorectal cancer (CRC) is the second leading cause of cancer-related death in Europe and the United States (1). There is good evidence that screening adults in the general population who are at average risk with faecal occult blood testing, flexible sigmoidoscopy or colonoscopy reduces CRC incidence and mortality (2–7). However, as with all screening programmes, CRC screening has the potential to cause harm, both directly to those screened and indirectly through diversion of resources away from other services. Targeted or stratified screening could potentially provide a way of reducing complication rates and demand on services while still ensuring those at greatest risk are effectively screened. For example, the U.S. Multi-Society Task Force on Colorectal Cancer endorse a risk-stratified approach with faecal immunochemical testing (FIT) screening in populations with an estimated low prevalence of advanced neoplasia and colonoscopy screening in high prevalence populations(8).

We have previously published a systematic review of risk prediction models for CRC and identified 40 models that have been developed and could potentially be used for risk stratification(9). These range from models including only data routinely available from electronic health records, such as age, sex and body mass index, to more complex models containing detailed information about lifestyle factors and genetic information. Using the UK Biobank cohort for external validation we have shown that several of those including only phenotypic risk factors and/or family history exhibit reasonable discrimination in a UK population (10). At the time of the literature search for that review (January 2014) only six risk models incorporating genetic risk factors and predicting future risk of developing CRC had been published, and their performance was similar to models including only phenotypic information. Since then, findings from genome-wide association studies have resulted in a

rapid rise in the number of published risk models incorporating genetic information. Simulation studies have also shown that using genetic information to stratify screening has the potential to improve efficiency (11) by reducing the number of individuals screened while still detecting as many cases (12). It is not clear, however, which genetic risk models perform best, how much combining common genetic variants with phenotypic risk factors improves model performance, or the potential public health impact of incorporating these models into screening programmes.

In order to inform future stratification of CRC screening using genetic data, we have updated our previous systematic review to identify and synthesize the performance of all published CRC prediction risk models that include common genetic variants and estimates of the potential public health impact of stratifying populations for screening based on genetic risk.

MATERIALS AND METHODS

We updated a previous systematic review following a published study protocol (PROSPERO 2018 CRD42018089654 Available from:

http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018089654).

Search strategy

We searched Medline, EMBASE and the Cochrane Library from January 2014 (the end date of the search in our previous review) to February 2019 applying the same search strategy used in our previous review, with no language limits (see Supplementary Materials and Methods S1, for complete search strategy for Medline and EMBASE). We subsequently manually screened the reference lists of all included papers.

Study selection

We included studies if they met all of the following criteria: (i) were published as a primary research paper in a peer-reviewed journal; (ii) provided a measure of relative or absolute risk using a combination of two or more risk factors, including at least one single-nucleotide polymorphism (SNP), that allows identification of individuals at higher risk of colon, rectal or colorectal cancer, or advanced colorectal neoplasia; (iii) reported a measure of discrimination (e.g. C-statistic, area under the receiver operating characteristic curve (AUROC)), or calibration (e.g. Hosmer-Lemeshow statistic, Observed/Expected ratio) , or a quantitative estimate of the implications of using the risk model for stratified screening; and (iv) included data applicable to the general population (i.e. the risk model was not specifically designed for individuals known to carry specific high-risk mutations or from families with a known cancer syndrome, such as familial adenomatous polyposis or hereditary nonpolyposis colorectal cancer). As in our previous review, studies including only highly selected groups, for example immunosuppressed patients, organ transplant recipients, or those with a previous history of colon and/or rectal cancer were excluded. We also included studies published prior to January 2014 that had been identified in our previous review if they met the above criteria.

One reviewer (LM) performed the search and screened 67% of the titles and abstracts to exclude papers that were clearly not relevant. The remaining 33% of titles and abstracts were divided between four reviewers (JUS, SG, JE, FW) for screening. The four reviewers also each independently assessed a random selection of 3% of the papers screened by LM. The full-text of all papers for which a definite decision to reject could not be made from the title and abstract alone were independently assessed by two reviewers (LM and JUS/SG/JE/FW). Those assessed as not meeting the inclusion criteria by both researchers were excluded. Those for which it was not clear were discussed with the wider research team. One paper was translated into English for assessment and subsequent data extraction.

Data extraction and synthesis

Data were extracted independently by two researchers (LM and JUS/SG/JE) directly into data tables to minimize bias. These tables included details on: (i) the development of the model, including potential sources of bias such as the selection processes for participants and SNPs; (ii) the risk model itself, including the variables included; (iii) the methods of model development (genetic and phenotypic components); (iv) the performance measures (discrimination (e.g. C-statistic, AUROC), or calibration (e.g. Hosmer-Lemeshow statistic, Observed/Expected ratio) of the risk model in the development population; (v) any external validation studies of the risk model, including the study design and performance of the risk model; and (vi) any public health modelling of the potential impact of using the risk models in practice. In papers that reported performance data for multiple step-wise models developed in the same population we included only the best performing model in our main analysis. If performance data were presented separately for a model including only SNPs and a model including both SNPs and phenotypic variables in the same paper, these were considered as two models. If performance data were presented separately for models that incorporated the same SNPs but were developed using unweighted allele counting or with allele weights derived either from the literature or the study population, we extracted both sets of data. To assess the incremental effect on performance of incorporating SNPs into the risk models, we additionally extracted data on the performance of the models including only phenotypic risk factors and/or family history, where they were reported.

At the same time as data extraction, an overall assessment of risk of bias was performed using four domains from the CHARMS checklist (study population, predictors, outcome and

sample size and missing data)(13). We also classified studies into the following groups according to the TRIPOD guidelines(14):

- development only (1a);

- development and validation using resampling (1b);

- random (2a) or non-random (2b) split-sample development and validation;

- development and validation using separate data (3); or

- validation only (4).

For the models including only SNPs, a model developed using SNPs selected from the literature, either with unweighted allele counting, or with allele weights derived from the literature, was considered as group 3 (development and validation using separate data). However, if the model used weights derived from the study population, or if the model included only the SNPs found to be significantly associated with CRC in the study population, we assigned it to either group 1b, 2a, 2b or 3, depending on the relationship between the study population and the testing population. Simulated populations were considered external populations.

RESULTS

From 12,394 papers we excluded 12,277 at title and abstract level and a further 103 after full-text assessment. After title and abstract screening by the first reviewer, no additional papers met the inclusion criteria in the random 12% screened by a second reviewer. There was also complete agreement amongst researchers at the full-text level with the most common reasons for exclusion being that the papers did not include a risk score (*n*=43), were conference abstracts (*n*=19) or did not include any performance measures (*n*=23) (Supplementary Figure

S1). Four were also excluded as they described models that were developed to detect prevalent undiagnosed disease rather than estimate future incident disease risk.

A further four papers were identified through citation searching. The addition of four papers (six risk models) which had been included in our previous systematic review gave a total of 22 papers describing 29 risk models for inclusion in the analysis. Table 1 summarizes these 29 risk models. Except for the model by Weigl *et al*., (15) that included CRC or advanced adenoma as the outcome, all had CRC as the outcome. The paper by Jung *et al*., (16) developed separate models for colorectal, colon and rectal cancer. As these were the only models for colon and rectal cancer, we included only the model for colorectal cancer in the analysis. Nine models included only SNPs, six included SNPs plus phenotypic factors but not age, and 14 a combination of SNPs, phenotypic factors and age. The number of SNPs included in the models ranged from 3 to 95.

Development of the risk models and risk of bias

Details of the methods used to select the predictors and develop each of the risk models are given in Table 1, with additional details of the setting, design, participants, outcome and sample size for each study in Supplementary Table S1. The majority of the risk models (*n* = 18) were developed or validated in white or European individuals. The others were developed or validated in Japanese (n = 4), Korean (n = 3), Chinese (n = 3) and Taiwanese (n = 1) populations.

A summary of the assessment of the risk of bias based on the four domains from the CHARMS checklist (study population, predictors, outcome and sample size and missing

data) is shown in Table 2. Overall we found 12 risk models to be at low risk of bias, 10 at unclear risk and five at high risk.

Risk of bias within the study participant domain was variable between studies. Those judged to be at unclear or high risk of bias reflected limited or missing details on the inclusion and exclusion criteria used to define study participants and/or use of cases or controls not representative of the general population, for example recruiting spouses or individuals attending outpatient hospital clinics as controls, or recruiting cases from adjuvant chemotherapy clinical trials.

When considering selection of predictors, the majority of the models ($n = 18$) included SNPs identified for inclusion from new or previously published genome-wide association studies (GWAS) in European or Asian-ancestry populations. In six, the authors had used GWAS studies from European or Asian populations to identify SNPs associated with CRC risk and then selected a subset of these SNPs for inclusion in the risk model on the basis of the associations with disease risk in an independent Japanese or Taiwanese population. Although this method was used to identify SNPs that may be associated with risk in non-European populations, given the small sample sizes of many of the studies and low statistical power this approach potentially excludes SNPs that are associated with risk in these populations. Two models[17] were developed on the basis of a GWAS study in a Korean population by selecting SNPs with evidence of association at the $p<10^{-6}$ significance level (which is less conservative than the conventionally accepted genome-wide level of significance, $p<5\times10^{-8}$ level for a GWAS study). A further three studies [18–20] selected SNPs based on plausible biological mechanisms leading to CRC and epidemiological studies (folate metabolism, DNA repair and breakdown of carcinogenic compounds, insulin-like growth factor and insulin).

One of these, the model by Jung et al.,(20) included both SNPs related to insulin metabolism and dietary fatty acids, potentially overestimating the risk for individuals with the risk allele.

Of the 20 models which include phenotypic risk factors, with or without age, in addition to SNPs, four used regression analyses to select which factors to include(15,18,20,21), one a bootstrap forest prediction model(19), and three(22,23) used risk factors identified from previous risk models. However, for the majority (*n* = 12) of models the publications included few details about how phenotypic factors were selected, and whether all those that had been considered were included in the final model. As a consequence, many do not include established risk factors for CRC.

The outcome (CRC) was defined histologically or from cancer registries in all studies, reducing the risk of bias due to case misclassification All studies reported numbers of cases and controls used in their development and/or validation analyses. Three included fewer than 150 cases (and hence had low statistical power). Only five studies adequately described how they dealt with missing data, so we cannot be certain that this was done appropriately in the remaining studies.

Discrimination and calibration of the risk models

Discrimination, as measured by the AUROC or C-statistic, was reported for 27 of the 29 risk models and calibration reported for four. The discrimination values are summarized graphically in Figure 1 and given in Supplementary Table S2, in which models are divided into those that include SNPs only and those that combine SNPs with phenotypic variables with or without age and whether the discrimination was assessed in the development population, bootstrap or a random-split sample, or in an external population or non-random

split sample. Where multiple AUROCs or C-statistics for the same model were reported for more than one method, measurement in the development populations always gave the highest discrimination, followed by that in bootstrapping or random split-sample validation studies and then in external populations. Where model performance was included for both men and women, discrimination was higher in men (0.59 in men compared with 0.56 in women(24), 0.63 in men compared with 0.62 in women(25), and 0.70 in men compared with 0.60 in women(17)).

Among the eight models that include only SNPs, the discrimination of seven was reported in external populations. This ranged between 0.56 and 0.60 in real-life populations and 0.63 in simulated populations. Of those assessed in real-life populations, the three considered at low risk of bias (Dunlop et al.,(26) Ibanez-Sanz et al.(21), and Smith et al.(23)) all have reported AUROCs of 0.56-0.57. Of the 19 risk models incorporating both SNPs and phenotypic variables, the models created by Procopciuc *et al*.(18), Jung *et al*., (20) and Shiao *et al*., (19), have the highest reported discrimination with AUROCs of 0.90 (95% CI, 0.86 - 0.93) in the development population, 0.93 in the development population and 0.85 in cross-validation respectively. In all three cases the SNPs were selected on the basis of candidate-gene association studies as opposed to GWAS studies. The models by Procopciuc *et al*. and Shiao *et al* were also developed in a small case-control studies with only 150 and 53 cases and 162 and 53 controls respectively, thus the resulting models are likely subject to a high degree of overfitting.

In the remaining models, in which the SNP selection was based on published GWASs, the AUROC in split sample validation or external validation in independent datasets ranged between 0.61 to 0.63 in models excluding age and 0.56 to 0.70 in those including age. The

best performing model in an independent validation population was the model by Smith *et al.* (23). Calibration was reported for only four of the 29 risk models. In three, the numbers of predicted colorectal cancers were in line with the observed numbers with non-significant p values of 0.086(18) and 0.336(27) under a Hosmer–Lemeshow statistic and 0.09 under a Grønnesby and Borgan test(22) respectively. Smith *et al.*,(23) assessed calibration graphically and found that the genetic risk score alone (Smith 2018a) was poorly calibrated, with over-estimation of risk for those in the top decile of risk. After re-calibration, however, both the genetic risk score alone and the genetic plus phenotypic models were well calibrated.

Incremental improvement of genetic over family history and/or phenotypic risk factors
Of the models that combined SNPs with family history and/or phenotypic risk factors, 15 compared the discrimination of models including SNPs, family history and phenotypic risk factors either alone or in combination(Table 3). Together these showed that adding SNPs to family history and/or phenotypic variables, and vice versa, leads to an increase in the AUROC of between 0.01 to 0.06. For example, in a cross-validation sample of a Spanish population, Ibanez-Sanz *et al.*, report an AUROC of 0.61 (95% CI, 0.59-0.64) for their environmental risk score comprising alcohol use, family history of CRC, BMI, physical exercise, red meat and vegetable intake, and NSAIDs/aspirin use and an AUROC of 0.56 (95% CI, 0.54-0.58) for their genetic risk score comprising 21 SNPs. For the combined risk score, they report an AUROC of 0.63 (95% CI, 0.60-0.66)(21). Iwasaki *et al.*, (22), Xin *et al.*, (27) and Weigl et al.,(15) additionally reported that adding genetic risk factors to a model including phenotypic risk factors increased the mean integrated discrimination improvement (IDI) by 0.015 (95% CI 0.0044 to 0.027), 0.031 (95% CI 0.023 to 0.039) and 0.04 (95% CI 0.03-0.05) respectively and the mean continuous net reclassification index (NRI) by 0.39 (95% CI 0.17 to 0.58), 0.317 (95% CI 0.225 to 0.408) and 0.29 (95% CI 0.14 to 0.43)

respectively. The study by Smith *et al*., in which a genetic risk score incorporating 41 SNPs identified from previous GWAS studies was added to two previously published phenotypic risk scores including age and family history of CRC (28,29) found that the genetic risk score did not meaningfully improve model discrimination. They did not report the IDI or NRI but overall the addition of genetic information resulted in 4-5% of individuals having a change in absolute risk of ≥ 0.3%. For those with an initial estimated absolute risk of <1%, this percentage was 3% and for those with an estimated absolute risk ≥1% 25-33% had a change in absolute risk of ≥ 0.3%.

Impact of stratifying populations for screening based on genetic risk

Eight studies assessed the potential impact of using the risk models to determine the starting age for screening. Seven of these calculated either the difference in recommended starting age for those at low or high risk or the years earlier those at high risk would be invited. These are summarised in Table 4. Considering SNPs alongside family history would result in individuals in the highest quintile of risk, for example, being invited between 13 and 21 years earlier, with the difference between the invitation ages of the highest quintile being and lowest quintile between 13 and 27 years. In all cases where estimates were provided for SNPs alone, family history alone, or SNPs and family history combined, the range was greater for SNPs than family history and greater for both combined than for either individually. Jenkins *et al*., (30) additionally estimated that if those in the highest quintile of risk were invited for screening at age 46 and those in the lowest quintile at age 59, 3.32 million people would be screened earlier, of which 8000 of those would be diagnosed with CRC, and 8.76 million would be screened later, of which 18,000 would be diagnosed with CRC.

The eighth study compared the size of the English population eligible for screening and the number of CRC cases potentially detectable using age-based screening and personalised screening in which eligibility is determined by absolute risk calculated using age and the Frampton *et al.* risk score(12). In a simulated population aged 55-69, 61% of men and 62% of women would be eligible for age-based screening (≥ 60 years) and 79% and 77% respectively of CRC cases would be diagnosed in this subset. With screening based on the genetic risk score (≥ average risk for an individual aged 60 (men 1.96%, women 1.19%)), 45% of men and 45% of women would be eligible for screening with 69% and 69% of CRC cases being identified. This translates into 16% fewer men and 17% fewer women being eligible for screening at the cost of detecting 10% and 8% fewer cases respectively.

DISCUSSION

Key findings

We have identified 29 risk models that incorporate common genetic variants to estimate future incidence of CRC in average-risk populations and that have either published measures of performance or estimates of the implications of using them for stratified screening. In external independent validation datasets, the three models considered at low risk of bias that include SNPs identified from GWAS studies all had similar discrimination (AUROC 0.56-0.57) (Dunlop et al.,(26) Ibanez-Sanz et al.(21), and Smith et al.(23)). Among the models that included SNPs in combination with other risk factors, the AUROC in split sample or external validation ranged between 0.61 to 0.63 in models excluding age and 0.56 to 0.70 in those including age. The model with the highest reported discrimination in an independent validation population was the model by Smith *et al*. that included 41 SNPs alongside age, diabetes, multi-vitamin usage, family history, years of education, BMI, alcohol intake, physical activity, NSAID usage, red meat intake, smoking and oestrogen use in women(23).

Only four reported data on model calibration. The addition of SNPs to risk scores already including family history and/or phenotypic variables increased discrimination by 0.01 to 0.06. Although this represents a modest increase in discrimination measured in terms of the AUROC, such differences can lead to substantial changes in risk stratification in the population, as illustrated by continuous NRI values of 0.3 to 0.4 seen in this review and demonstrated in the context of other diseases(31). Public health modelling within the studies suggest that if the models were used to determine the starting age for screening, this would result in individuals in the top 20% for risk being invited up to 23 years earlier than if determined by age-based criteria only, with the difference in age at invitation between the highest and lowest risk quintiles being several years greater for models including SNPs alone than for models including family history alone, and the difference for models including both SNPs and family history greater than that for models including either SNP or family history.

Strengths and Limitations

The main strengths of this review are the comprehensive literature search that included both subject headings and free text, and the systematic approach we used to screen papers for inclusion. The inclusion of more than one risk model from many of the published papers also enabled us to make comparisons between models that included different groups of risk factors or had been developed using different statistical methods. Although this approach enabled us to identify 23 risk models that have been published since our earlier review, we cannot exclude the possibility that there are others that we did not identify. Genetic research is also a rapidly advancing field with new papers reporting new genetic variants that could be incorporated into future risk scores being published regularly.

Other limitations of this review relate to the studies themselves. Most of the risk models were developed and/or tested in case-control studies. Estimates of absolute risk of developing CRC are therefore not possible and the collection of phenotypic risk factors will be subject to both recall and responder bias, potentially increasing the apparent discrimination. Conversely, in many, the matching variables were not included as covariates within the risk models and this may have resulted in underestimation of discrimination(32). The risk models also varied substantially in relation to size, selection of cases and controls and variables considered for inclusion. This heterogeneity meant it was not possible to assess whether, for example, the number of SNPs affected the performance of the models. Furthermore, most risk models were developed and/or tested in either European, Chinese or Japanese populations. The risk models in this review may therefore not be applicable to other population groups.

There was also heterogeneity in how the SNPs and phenotypic factors were selected and combined into risk scores, which ultimately impacts their performance in independent samples. For several models SNP selection was based on small sizes and/or there was limited detail on how lifestyle/hormonal risk factors were selected. Similarly, several models did not include well-established risk factors for CRC. Almost all, however, assumed that the associations of the SNPs are independent from each other and that risk follows an additive model on the log-Risk scale. These assumptions are generally considered to be robust(33) and many of the authors describe how they had sought to remove SNPs in linkage disequilibrium or associated with factors on the genetic pathway. In the absence of evidence of interactions, the models also assume that the strengths of associations for each SNP with CRC are constant with age. This may not be true and further studies are needed to assess for possible interactions.

Finally, in relation to the performance measures for the models, discrimination for many had only been assessed in the development population, no data on discrimination has been published for the genetic model with the largest number of SNPs(34), only four models reported data on calibration, and only two included estimates of net reclassification. As illustrated by the lower AUROCs seen in development populations when compared with the performance of the same models from bootstrapping or cross-validation, the performance of all prediction models is overestimated due to overfitting when both model development and performance assessment use the same data set, particularly in studies with small sample sizes(35). Additionally, while the AUROC or other measures of discrimination are important when considering how well individuals can be ranked in terms of predicted risk, without measures of calibration or reclassification it is not possible to assess how closely the estimated risks match the observed risks, how much including different factors in the risk scores influences the classification of individuals or whether the models stratify correctly into high/low categories of absolute risk that are of clinical importance.

Implications for future research

This review shows that a large number of risk scores incorporating common genetic markers have been developed to estimate future risk of CRC and suggests that many of these are better at discriminating between those at higher and lower risk of CRC than age alone, family history alone, or risk scores incorporating only phenotypic risk factors. As has been described previously(9,36), risk models such as these could be used to stratify the general population into risk categories, based either on estimates of absolute risk for those models including age or relative risk for those excluding age, to allow screening and preventive strategies to be targeted at those most likely to benefit. While the findings of this review therefore suggest that future risk prediction in colorectal cancer will improve with the inclusion of polygenic

risk factors, it remains uncertain how these models would perform in real-life settings and whether the increase in discriminatory performance and wider range of ages at which individuals would become eligible for screening that could be achieved through the inclusion of genetic variables translates into improved health of the population or the cost effectiveness of a screening programme.

Firstly, many of these models have not been externally validated and very few have had calibration assessed. As described above, these steps are essential before risk models can be incorporated into practice. To enable direct comparisons between the models, ideally the models identified in this review with the greatest number of SNPs and those with the highest reported discrimination would be assessed in a single independent cohort. However, the predictive ability of risk models is known to vary between populations and the risk of developing CRC varies substantially worldwide(37). The choice of models for independent validation will therefore depend on the population of interest and these analyses should be performed in populations similar to those in which use of the model is being considered. This is particularly important in the context of genetic risk models.   Comparisons between the population genetics of different ethnic groups have shown that the estimated associated risks and population frequencies of SNPs can vary substantially with ethnicity(38,39) and the overall magnitude of association of polygenic risk scores derived from GWAS in European-ancestry populations, as is the case for most models for CRC, may differ when applied to other populations(40). As highlighted by De La Vega and Bustamante, to avoid further inequities in health outcomes, the inclusion of diverse populations in CRC research, unbiased genotyping, and methods of bias reduction in genetic risk scores are critical(41).

Secondly, further methodological studies are required to improve genome-wide risk prediction in order to understand the potential benefits of including increasing numbers of SNPs, together with other rare moderate/high risk genetic variants and established or new lifestyle/environmental risk factors, as has been done for other cancers(42). These also include exploring more sophisticated statistical methods for developing polygenic risk scores(43), and novel methods such as machine learning approaches for combining the effects of diverse risk factors(40). Thirdly, there was substantial variation in the reporting of the studies in this review. Encouraging the use of reporting guidelines, such as the Genetic Risk Prediction Studies (GRIPS) statement(44,45) that includes a checklist of 25 items, would improve the transparency, quality, and completeness of the reporting of new models and facilitate future syntheses in this field.

Finally, the assessment of model performance is only one component when considering whether risk models are ready for clinical use; the context in which the model will be used, including the costs of measuring additional risk factors and the risk-benefit of any interventions offered, and the wider ethical, legal and social issues around implementation must also be considered. To our knowledge, only one study has modelled the potential impact of CRC screening based on age and SNPs on preventing deaths from CRC (11). Using age-specific crude rate of deaths due to CRC in a hypothetical population based on the Australian population in 2011 and assuming a 100% attendance rate at screening, that study showed that the net effect of inviting individuals for biennial FOBT based on their genetic risk would be 0.4% more colorectal cancer deaths and 0.2% more years of life lost per person invited to screen than inviting those aged between 50 and 74, against a background of 4.9% fewer screens, resulting in a 3.1% overall improved efficiency. The risk model used in that study was the model by Jenkins *et al.*, 2006 that includes 45 SNPs and had an AUROC of 0.63 in a

simulated population. It is likely, therefore, that similar improvements in efficiency would be seen with other models, many of which have reported AUROCs of greater than 0.63. However, that study did not consider the costs of implementing stratified screening, competing risks of death or the psychological harms associated with screening, uniform attendance across risk groups was assumed, and no data was included on the calibration of the model. Further modelling studies are therefore needed to assess the cost-effectiveness and differences in quality adjusted life years (QALYs) and implementation studies to asses risk-appropriate screening participation and the psychosocial consequences of this approach.

By identifying the published risk models for CRC that include common genetic variants and demonstrating the potential public health benefits of using such models to determine the starting age for screening, this study provides valuable evidence to support investment in this further research.

REFERENCES

1. Stewart B., Kleihues P. (Eds): World Cancer Report. IARC Press Lyon. 2003;

2. Lin JS, Piper MA, Perdue LA, Rutter CM, Webber EM, O'Connor E, et al. Screening for Colorectal Cancer. An Updated Systematic Review for the U.S. Preventive Services Task Force. JAMA. 2016 Jun 21;315(23):2576.

3. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet. 1996 Nov 30;348(9040):1472–7.

4. Holme Ø, Bretthauer M, Fretheim A, Odgaard-Jensen J, Hoff G. Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in

asymptomatic individuals. Cochrane database Syst Rev. 2013 Jan;9:CD009259.

5.   Kronborg O, Fenger C, Olsen J, Jørgensen OD, Søndergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet. 1996 Nov 30;348(9040):1467–71.

6.   Lindholm E, Brevinge H, Haglind E. Survival benefit in a randomized clinical trial of faecal occult blood screening for colorectal cancer. Br J Surg. 2008 Aug;95(8):1029–36.

7.   Brenner H, Stock C, Hoffmeister M. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. BMJ. 2014 Jan;348:g2467.

8.   Rex DK, Boland CR, Dominitz JA, Giardiello FM, Johnson DA, Kaltenbach T, et al. Colorectal cancer screening: Recommendations for physicians and patients from the U.S. Multi-Society Task Force on Colorectal Cancer. Gastrointest Endosc. 2017;153(1):307–23.

9.   Usher-Smith JA, Walter FM, Emery J, Win AK, Griffin SJ. Risk prediction models for colorectal cancer: a systematic review. Cancer Prev Res (Phila). 2015 Oct 13;

10.   Usher-Smith J, Harshfield A, Saunders C, Sharp S, Emery J, Walter F, et al. External validation of risk prediction models for incident colorectal cancer using UK Biobank. Br J Cancer. 2018;1–10.

11.   Stanesby O, Jenkins M. Comparison of the efficiency of colorectal cancer screening programs based on age and genetic risk for reduction of colorectal cancer mortality. Eur J Hum Genet. 2017;25(7):832–8.

12.   Frampton MJE, Law P, Litchfield K, Morris EJ, Kerr D, Turnbull C, et al. Implications of polygenic risk for personalised colorectal cancer screening. Ann Oncol. 2016;27(3):429–34.

13.   Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist. PLoS Med. 2014;11(10).

14.   Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Ann Intern Med. 2015;162(1):55–63.

15.   Weigl K, Thomsen H, Balavarca Y, Hellwege JN, Shrubsole MJ, Brenner H. Genetic Risk Score Is Associated With Prevalence of Advanced Neoplasms in a Colorectal Cancer Screening Population. Gastroenterology. 2018;155(1):88-98.e10.

16.   Jung KJ, Won D, Jeon C, Kim S, Kim T Il, Jee SH, et al. A colorectal cancer prediction model using traditional and genetic risk scores in Koreans. BMC Genet. 2015;16:49.

17.   Jo J, Nam CM, Sull JW, Yun JE, Kim SY, Lee SJ, et al. Prediction of Colorectal Cancer Risk Using a Genetic Risk Score: The Korean Cancer Prevention Study-II (KCPS-II). Genomics Inform. 2012;10(3):175–83.

18.   Procopciuc LM, Osian G, Iancu M. Colorectal Cancer Carcinogenesis: a Multivariate

Genetic Model in a Cohort of Romanian Population. Clin Lab. 2017;63(4):647–58.

19. Shiao SPK, Grayson J, Lie A, Yu CH. Personalized nutrition—genes, diet, and related interactive parameters as predictors of cancer in multiethnic colorectal cancer families. Nutrients. 2018;10(6):1–19.

20. Jung S, Yon PhD MPH, Zhang Z-F, MD P. The effects of genetic variants related to insulin metabolism pathways and the interactions with lifestyles on colorectal cancer risk. Menopause. 2019;

21. Ibanez-Sanz G, Diez-Villanueva A, Alonso MH, Rodriguez-Moranta F, Perez-Gomez B, Bustamante M, et al. Risk Model for Colorectal Cancer in Spanish Population Using Environmental and Genetic Factors: Results from the MCC-Spain study. Sci Rep. 2017;7:43263.

22. Iwasaki M, Tanaka-Mizuno S, Kuchiba A, Yamaji T, Sawada N, Goto A, et al. Inclusion of a genetic risk score into a validated risk prediction model for colorectal cancer in Japanese men improves performance. Cancer Prev Res. 2017;

23. Smith T, Gunter MJ, Tzoulaki I, Muller DC. The added value of genetic information in colorectal cancer risk prediction models: development and evaluation in the UK Biobank prospective cohort study. Br J Cancer. 2018;(May).

24. Hsu L, Jeon J, Brenner H, Gruber SB, Schoen RE, Berndt SI, et al. A model to determine colorectal cancer risk using common genetic susceptibility loci. Gastroenterology. 2015;148(7):1330.

25. Jeon J, Du M, Schoen RE, Hoffmeister M, Newcomb PA, Berndt SI, et al. Determining Risk of Colorectal Cancer and Starting Age of Screening Based on Lifestyle, Environmental, and Genetic Factors. Gastroenterology. 2018;154(8):2152-2164.e19.

26. Dunlop MG, Tenesa A, Farrington SM, Ballereau S, Brewster DH, Koessler T, et al. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. Gut. 2013 Jun 1;62(6):871–81.

27. Xin J, Chu H, Ben S, Ge Y, Shao W, Zhao Y, et al. Evaluating the effect of multiple genetic risk score models on colorectal cancer risk prediction. Gene. 2018;673:174–80.

28. Taylor DP, Stoddard GJ, Burt RW, Williams MS, Mitchell J a, Haug PJ, et al. How well does family history predict who will get colorectal cancer? Implications for cancer screening and counseling. Genet Med. 2011;13(5):385–91.

29. Wells BJ, Kattan MW, Cooper GS, Jackson L, Koroukian S. ColoRectal Cancer Predicted Risk Online (CRC-PRO) Calculator Using Data from the Multi-Ethnic Cohort Study. J Am Board Fam Med. 27(1):42–55.

30. Jenkins M, Win AK, Dowty J, MacInnis R, Makalic E, Schmidt D, et al. Ability of known susceptibility SNPs to predict colorectal cancer risk for persons with and without a family history. bioRxiv. 2018;267666.

31. Garcia-Closas M, Gunsoy NB, Chatterjee N. Combined associations of genetic and environmental risk factors: implications for prevention of breast cancer. J Natl Cancer Inst. 2014;106(11):1–6.

32. Pepe M, Fan J, Seymour C. Estimating the ROC Curve in Studies that Match Controls

to Cases on Covariates. Acad Radiol. 2013;20(7):863–73.

33. Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet. 2006 Oct;7(10):781–91.

34. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet. 2019;51(1):76–87.

35. Steyerberg EW, Moons KGM, Windt DA Van Der, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3 : Prognostic Model Research. PLoS Med. 2013;10(2):e1001381.

36. Roberts M. Implementation Challenges for Risk-Stratified Screening in the Era of PrecisionMedicine. JAMA Oncol. 2018;E1-2.

37. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2016 Jan 27;1–9.

38. He J, Wilkens LR, Stram DO, Kolonel LN, Henderson BE, Wu AH, et al. Generalizability and Epidemiologic Characterization of Eleven Colorectal Cancer GWAS Hits in Multiple Populations. Cancer Epidemiol Biomakers Prev. 2011;20(1):70–81.

39. Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG. Measures of human population structure show heterogeneity among genomic regions. Genome Res. 2005 Nov 1;15(11):1468–76.

40. Kim MS, Patel KP, Teng AK, Berens AJ, Lachance J. Genetic disease risks can be misestimated across global populations. Genome Biol. 2018;19(1):179.

41. Vega FMD La, Bustamante CD. Polygenic risk scores: a biased prediction? Genome Med. 2018;10(1):100.

42. Lee A, Mavaddat N, Wilcox AN, Cunningham AP, Carver T, Hartley S, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genet Med. 2019;0(0):1.

43. Chatterjee N, Shi J. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet. 2016;17(7):392–406.

44. Janssens ACJW, Ioannidis JPA, van Duijn CM, Little J, Khoury MJ, GRIPS Group. Strengthening the reporting of genetic risk prediction studies: the GRIPS statement. BMJ. 2011 Mar 16;342:d631.

45. Janssens ACJW, Ioannidis JPA, Bedrosian S, Boffetta P, Dolan SM, Dowling N, et al. Strengthening the reporting of genetic risk prediction studies (GRIPS): explanation and elaboration. Eur J Hum Genet. 2011;19.

TABLES

Table 1. Summary of risk models

Author, year	Country	Outcome	Factors included in score	Selection of SNPs	Method of development of GRS	Selection of phenotypic factors	Method of development of combined model	TRIPOD level*
Genetic risk factors alone								
Dunlop 2013a	UK, Canada, Australia, USA and Germany (d) Sweden and Finland (v)	CRC	10 SNPs	Published GWAS studies from European populations	Unweighted allele counting model	---	---	3
Frampton 2016	UK (v)	CRC	37 SNPs	Published GWAS studies from European populations	Weighted allele model weighted by published log odds	---	---	3
Hosono 2016a	Japan (d, v)	CRC	6 SNPs	Published GWAS studies from European populations followed by logistic regression	Unweighted allele counting model	---	---	2b
Huyghe 2019	European (91.7%) and East Asian (8.3%) (d)	CRC	95 SNPs	GWAS study	Weighted allele model weighted by study derived weights	---	---	1a
Ibanez-Sanz 2017a	Spain (d, v)	CRC	21 SNPs	Published GWAS studies included within European Bioinformatics Institute	Unweighted allele counting model (weighted allele models weighted by published log-odds and study derived log-odds similar so not reported)	---	---	3
Jenkins 2016	Australia, Canada, USA (v)	CRC	45 SNPs	Published GWAS studies from European populations	Weighted allele model weighted by published log odds	---	---	3**, 4**
Smith 2018a	UK (d, v)	CRC	41 SNPs	Published GWAS studies from predominantly European and white populations	Weighted allele model weighted by published log odds	---	---	3
Wang 2013	Taiwan (d, v)	CRC	16 SNPs	Published GWAS studies from Asian populations followed by replication analysis and jack-knife selection	Logistic regression	---	---	1b
Xin 2018a	China (d, v)	CRC	14 SNPs	Published GWAS studies from European or Asian populations	Unweighted allele counting model;	---	---	3

25

					Weighted allele model weighted by published log odds; Weighted allele model weighted by study derived weights			
Genetic plus phenotypic risk factors excluding age								
Ibanez-Sanz 2017b	Spain (d, v)	CRC	21 SNPs, family history of CRC, alcohol use, BMI, physical exercise, red meat and vegetable intake, NSAIDs/aspirin use	Published GWAS studies included within European Bioinformatics Institute	Unweighted allele counting model (weighted allele models weighted by published log-odds and study derived log-odds similar so not reported)	Logistic regression	Logistic regression	1b
Jeon 2018a	Australia, Canada, Germany, Israel and USA (d, v)	CRC (female)	63 SNPs, height, BMI, education, history of type 2 diabetes mellitus, smoking status, alcohol consumption, regular aspirin use, regular NSAID use, regular use of postmenopausal hormones, smoking, intake of fibre, calcium, folate, processed meat, red meat, fruit, vegetables, total-energy, physical activity	Published GWAS studies from predominantly European and Asian populations	Weighted allele model weighted by study derived estimated regression coefficients	No details given – all considered included	Logistic regression	2a
Jeon 2018b	Australia, Canada, Germany, Israel and USA. (d, v)	CRC (male)	63 SNPs, height, BMI, education, history of type 2 diabetes mellitus, smoking status, alcohol consumption, regular aspirin use, regular NSAID use, smoking, intake of fibre, calcium, folate, processed meat, red meat, fruit, vegetables, total-energy, physical activity	Published GWAS studies from predominantly European and Asian populations	Weighted allele model weighted by study derived estimated regression coefficients	No details given – all considered included	Logistic regression	2a
Procopciuc 2017	Romania (d)	CRC	7 SNPs, gender, alcohol, fried red meat	Candidate genes on metabolic pathway	Logistic regression	Logistic regression	Logistic regression	1a
Xin 2018b	China (d, v)	CRC	14 SNPs, smoking status	Published GWAS studies from European or Asian populations	Unweighted allele counting model	No details given – all considered included	Logistic regression	3
Yarnall 2013	UK (v)	CRC	14 SNPs, BMI, smoking, alcohol, fibre intake, red meat intake, physical activity	Published GWAS studies from predominantly European populations	Simulation based procedure using REGENT software	Literature review – all considered included	Simulation based procedure using REGENT software	3**
Genetic plus phenotypic risk factors including age								
Abe 2017	Japan (d, v)	CRC	11 SNPs, age, sex, referral pattern, current BMI, smoking, alcohol consumption, regular exercise, family history of colorectal cancer in a first degree relative, and dietary folate intake	Published GWAS studies from European and East Asian populations followed by logistic regression	Unweighted allele counting model	No details given – all considered included	Logistic regression	2b
Dunlop 2013b	UK, Canada,	CRC	10 SNPs, age, gender, first degree relative with CRC	Published GWAS studies from European populations	Unweighted allele counting model	No details given – all considered	Logistic regression	3

	Australia, USA and Germany (d) Sweden and Finland (v)						included		
Hosono 2016b	Japan (d, v)	CRC	6 SNPs, age, referral pattern, current BMI, smoking, alcohol consumption, regular exercise, family history of CRC, dietary folate intake	Published GWAS studies from European populations followed by logistic regression	Unweighted allele counting model	No details given – all considered included	Logistic regression	2b
Hsu 2015	USA and Germany (d, v)	CRC	27 SNPs, age, sex, family history of CRC, history of endoscopic examinations	Previous GWAS studies from European and East Asian populations	Unweighted allele counting model (weighted model weighted by published log-odds similar so not reported)	No details given – all considered included	Logistic regression	3
Iwasaki 2017	Japan (d, v)	CRC (male)	6 SNPs, age, BMI, alcohol, smoking status	Previous published model and GWAS from European and East Asian populations followed by cox proportional hazards modelling	Weighted allele model weighted by study derived log-transformed per allele HR	From previous model (Ma) except for physical activity	Weighted cox proportional hazards regression	1b
Jo 2012a	Korea (d, v)	CRC (female)	5 SNPs, age, family history of CRC	GWAS study in Korean population with significance level of $p<10^{-6}$	Unweighted allele counting model; weighted allele model weighted by study derived beta-coefficients	No details given – all considered included	Logistic regression	1b
Jo 2012b	Korea (d, v)	CRC (male)	3 SNPs, age, family history of CRC	GWAS study in Korean population with significance level of $p<10^{-6}$	Unweighted allele counting model; weighted allele model weighted by study derived beta-coefficients	No details given – all considered included	Logistic regression	1b
Jung 2015	South Korea (d)	CRC, colon and rectal cancer	7 SNPs, age, sex, smoking status, exercise status, fasting serum glucose, family history of CRC	Published GWAS studies from predominantly European and Asian populations followed by logistic regression	Unweighted allele counting model; weighted allele model weighted by study derived beta-coefficients	No details given – all considered included	Cox proportional hazards regression	1a
Jung 2019	USA (d)	CRC	4 SNPs, age, percentage calories from saturated fatty acids	Candidate genes related to insulin-growth like factor and insulin	Weighted allele model weighted by predictive value assessed via minimal depth method in nested random survival forest models	Multi-collinearity testing and univariate and stepwise regression analyses for final set to be included.	Random survival forest analysis	1a
Li 2015	China (d)	CRC	7 SNPs, age, sex, smoking, drinking	NHGRI GWAS database	Unweighted allele counting model;	No details given – all considered	Logistic regression	1a

27

					weighted allele model weighted by study derived beta-coefficients	included		
Shiao 2018 (d, v)	USA	CRC	5 SNPs, age, gender, BMI, thiamine, MTHFRR 677 expression level, HEI score (calories, total fruit, whole fruit, vegetables, dark green, total grains, whole grains, dairy, protein, oil and nuts, saturated fat, sodium, empty calories)	Candidate genes related to folate metabolism	Unweighted allele counting model	Bootstrap forest prediction modelling	Generalised regression elastic net model (penalised regression)	1b
Smith 2018b	UK (d, v)	CRC	41 SNPs, age, family history	Published GWAS studies from predominantly European and white populations	Weighted allele model weighted by published log odds	Factors included in Taylor et al. model	Standard model: log GRS combined with predicted log hazard ratio original model.	3
Smith 2018c	UK (d, v)	CRC	41 SNPs, age, diabetes, multi-vitamin usage, family history, years of education, BMI, alcohol intake, physical activity, NSAID usage, red meat intake, smoking, oestrogen use (women only)	Published GWAS studies from predominantly European and white populations	Weighted allele model weighted by published log odds	Factors included in Wells et al. model	Standard model: log GRS combined with predicted log hazard ratio original model.	3
Weigl 2018	Germany (d)	CRC or advanced adenoma	48 SNPs, age, sex, previous colonoscopy, physical activity, BMI	Published GWAS studies from European populations	Unweighted allele counting model	Factors statistically associated with genetic risk categories in controls	Logistic regression	1a

*Tripod level - 1a – Development only; 1b – Development and validation using resampling; 2a – Random split-sample development and validation; 2b – Non-random split-sample development and validation; 3 – Development and validation using separate data; 4 – external validation. CRC – colorectal cancer, SNP - single-nucleotide polymorphism, BMI – body mass index, NSAID – non-steroidal anti-inflammatory drug, wGRS – weighted genetic risk score. d = development; v - validation

** Simulated population

Table 2. Assessment of risk of bias of included articles

Author, year	Study Participants	Predictors	Outcome	Sample size and missing data	Overall
Genetic risk factors alone					
Dunlop 2013a	+	+	+	+	+
Frampton 2016	?	+	+	?	?
Hosono 2016a	?	?	+	?	?
Huyghe 2019	+	+	+	?	+
Ibanez-Sanz 2017a	+	+	+	?	+
Jenkins 2016, 2019	+	+	+	?	+
Smith 2018a	+	+	+	+	+
Wang 2013	?	-	+	?	-
Xin 2018a	?	+	+	?	?
Genetic plus phenotypic risk factors excluding age					
Ibanez-Sanz 2017b	+	+	+	?	+
Jeon 2018a and b	+	+	+	?	+
Procopciuc 2017	-	?	+	-	-
Xin 2018b	?	?	+	?	?
Yarnell 2013	?	+	+	?	?
Genetic plus phenotypic risk factors plus age					
Abe 2017	?	+	+	?	?
Dunlop 2013b	+	?	+	+	+
Hosono 2016b	?	?	+	?	?
Hsu 2015b	+	?	+	?	?
Iwasaki 2017b	+	+	+	?	+
Jo 2012 a and b	?	-	+	-	-
Jung 2015	+	?	+	?	?
Jung 2019	+	-	+	?	-
Li 2015	?	?	+	?	?
Shiao 2018	-	?	+	-	-
Smith 2018b	+	+	+	+	+
Smith 2018c	+	+	+	+	+
Weigl 2018	+	+	+	?	+

+ = low risk; ? = unclear risk; - = high risk

Table 3. Discriminatory performance of models including genomic risk factors only with those including family history and/or phenotypic risk factors only or genetic and family history and/or phenotypic risk factors combined

Author, year	Genetic risk factors only (AUROC (95% CI))	Family history alone (AUROC (95% CI))	Phenotypic risk factors only (AUROC (95% CI))	Genetic risk factors and family history (AUROC (95% CI))	Phenotypic risk factors and family history (AUROC (95% CI))	Genetic and phenotypic risk factors combined (AUROC (95% CI))	Genetic risk factors, family history and phenotypic risk factors combined (AUROC (95% CI))
Dunlop 2013	0.57						0.59
Hosono 2016	0.60				0.70		0.72
Hsu 2015	Women 0.55 Men 0.60	Women 0.52 Men 0.51					Women 0.56 Men 0.59
Ibanez-Sanz 2017	0.56 (0.54-0.58)		0.60 (0.57-0.61)		0.61 (0.59-0.64)		0.63 (0.60-0.66)
Iwasaki 2017	0.63[a]		0.60[a]			0.66[a]	
Jeon 2018a (female)		0.54 (0.52-0.55)		0.59 (0.58-0.60)	0.60 (0.59-0.61)		0.62 (0.61-0.63)
Jeon 2018b (male)		0.53 (0.52-0.54)		0.59 (0.58-0.60)	0.60 (0.59-0.61)		0.63 (0.62-0.64)
Jo 2012					Women: 0.60 (0.57-0.64) Men: 0.69 (0.65-0.73)		Women:0.65 (0.62-0.68) Men: 0.73 (0.68-0.77)
Jung 2015					0.73 (0.69-0.78)		0.74 (0.70-0.78)
Smith 2018a and b	0.56 (0.55-0.58)				0.67 (0.65-0.68) Excluding age: 0.52 (0.51-0.53)		0.68 (0.66-0.69)
Smith 2018a and c	0.57 (0.55-0.58)				0.68 (0.67-0.69) Excluding age: 0.58 (0.57-0.60)		0.69 (0.67-0.70)
Li 2015			0.57 (0.55-0.59)			0.59 (0.57-0.61)	
Weigl 2018			0.62			0.67	
Xin 2018b			0.52 (0.50-0.54)			0.61 (0.58-0.63)	

[a] All models include age in addition to genomic and/or phenotypic risk factors

Table 4. Results of population modelling studies showing the difference in recommended starting age or estimated number of years earlier that individuals would be invited to screening if the age of invitation was determined by a risk threshold based on a genetic or phenotypic model

Author, year	Model specific risk threshold used to determine starting age for screening	Type of risk model / included risk factors	Difference in years in recommended starting age for screening between those in the highest and lowest percentiles of risk			
			Papers selecting the top and bottom 1% of risk for comparison	Papers selecting the top and bottom 10% of risk for comparison	Papers selecting the top and bottom 20% of risk for comparison	Papers selecting the top and bottom 33% of risk for comparison
Hsu 2015	Average 10 year risk of a 50 year old (0.91%)	FH	---	Men: 5 (44 to 49)* Women: 4 (50 to 54)*	---	---
		FH + SNPs	---	Men: 10 (42 to 52) Women: 11 (47 to 58)	---	---
Jenkins 2019	0.3% 5 year estimated risk	SNPs	---	---	Men: 10 (45-55) Women: 14 (47 to 61)	---
		FH + SNPs	---	---	Men: 22 (35 to 57) Women: 27 (35 to 62)	---
Jenkins 2016 (USA)	1% 5 year estimated risk	FH + SNPs	---	Men: 27 (46 to 73) Women: 32 (48 to 80)	Men: 18 (48 to 66) Women: 21 (52 to 73)	---
Jenkins 2016 (Australia)	1% 5 year estimated risk	FH + SNPs	---	Men: 17 (46 to 63) Women: 23 (53 to 76)	Men: 13 (48 to 61) Women: 17 (55 to 72)	---
Jeon 2018	Average 10 year risk of a 50 year old (0.97%)	FH + SNPs + phenotypic	Men:17 (38 to 55) Women:21 (43 to 64)	Men: 11 (40 to 51) Women: 13 (46 to 59)	---	---
Huyghe 2018	Average 10 year risk of a 50 year old (1.13% for men and 0.68% for women)	SNPs	Men: 18 (41 to 59) Women: 24 (45 to 69)	Men: 10 (range 44 to 54) Women: 12 (range 49 to 61)	---	---
Weigl 2018	Average relative risk for a 60 year old with medium genetic risk	SNPs	---	---	---	17.5 (56 to 73)
Author, year	Risk threshold	Risk factors	Years earlier for recommended starting age for those in the highest percentiles			
			1%	10%	20%	33%
Dunlop 2013	5% 10 year estimated risk	FH	Men: >15 (from >75) Women: > 12 (from >80)	---	---	---

		FH + SNPs	Men: > 23 (from >75) Women: >22 (from >80)	---	---	---
Jenkins 2016 (USA)	1% 5 year estimated risk	FH	---	Men: 12 (from 67)* Women: 12 (from 73)*	---	---
		SNPs	---	Men: 14 (from 67) Women: 14 (from 73)	Men: 10 (from 67) Women: 11 (from 73)	---
		FH + SNPs	---	Men: 21 (from 67) Women: 25 (from 73)	Men: 19 (from 67) Women 21 (from 73)	---
Jenkins 2016 (Australia)	1% 5 year estimated risk	FH	---	Men: 9 (from 61)* Women: 12 (from 71)*	---	---
		SNPs	---	Men: 9 (from 61) Women: 12 (from 71)	Men: 6 (from 61) Women: 9 (from 71)	---
		FH + SNPs	---	Men: 15 (from 61) Women: 18 (from 71)	Men: 13 (from 61) Women 16 (from 71)	---

* based on presence or absence of family history (FH), not top and/or bottom 10%.

FIGURE LEGENDS

Figure 1. Relative discriminative performance of the risk scores grouped by those including only SNPs, those including SNPs plus family history and/or phenotypic risk factors without age and those including SNPs plus family history and/or phenotypic risk factors and age. Within each of these groups, the models are ordered according to sample size, with larger studies being those towards the bottom of each risk model category.
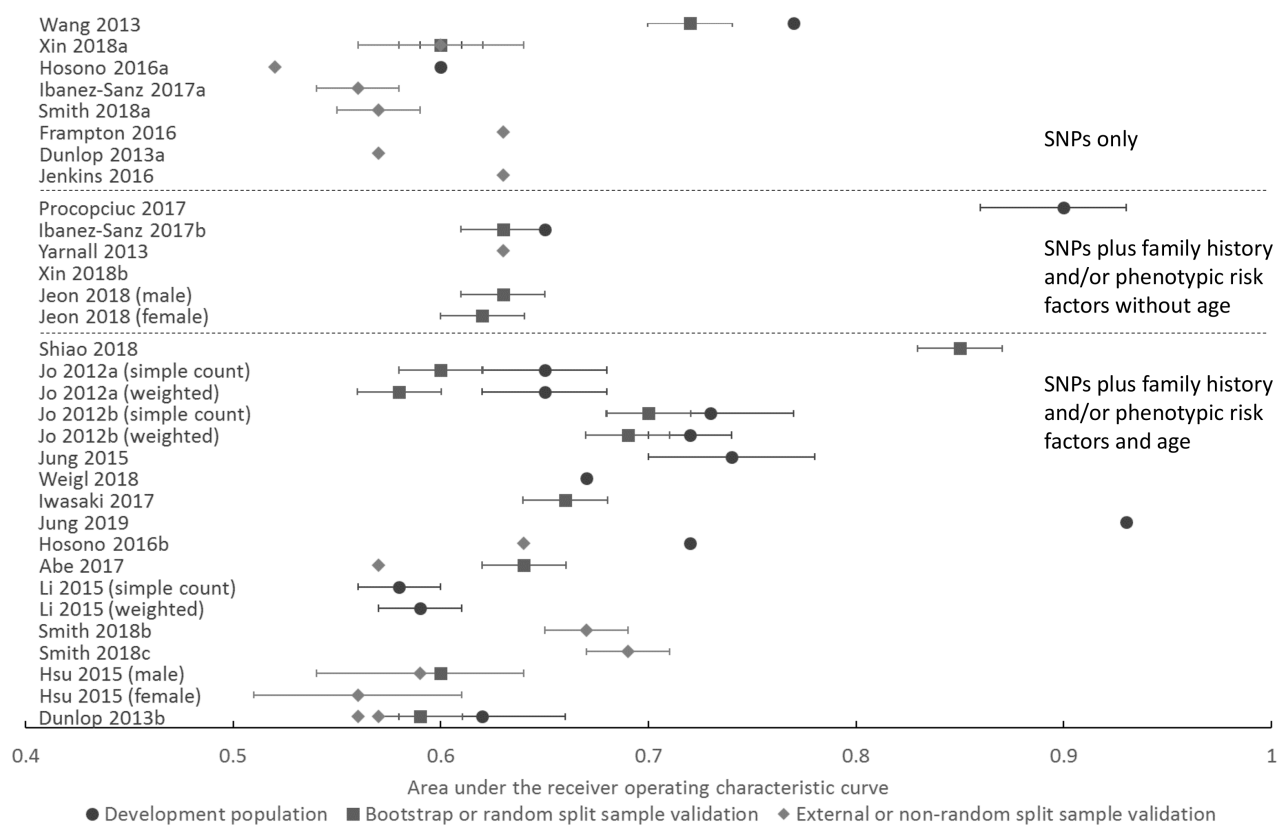
Figure 1