

RESEARCH ARTICLE

Ability of known colorectal cancer susceptibility SNPs to predict colorectal cancer risk: A cohort study within the UK Biobank

Aviv Gafni^{1*}, Gillian S. Dite¹, Erika Spaeth Tuff², Richard Allman^{1,3}, John L. Hopper³

1 Genetic Technologies Ltd., Fitzroy, Victoria, Australia, **2** Phenogen Sciences Inc, Charlotte, North Carolina, United States of America, **3** Centre for Epidemiology and Biostatistics, University of Melbourne, Parkville, Victoria, Australia

* aviv.gafni@gtglabs.com

OPEN ACCESS

Citation: Gafni A, Dite GS, Spaeth Tuff E, Allman R, Hopper JL (2021) Ability of known colorectal cancer susceptibility SNPs to predict colorectal cancer risk: A cohort study within the UK Biobank. PLoS ONE 16(9): e0251469. <https://doi.org/10.1371/journal.pone.0251469>

Editor: Philipp D. Koellinger, University of Wisconsin-Madison, UNITED STATES

Received: March 28, 2021

Accepted: September 2, 2021

Published: September 15, 2021

Copyright: © 2021 Gafni et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Access to the data used in this study can be obtained by applying directly to the UK Biobank at <https://www.ukbiobank.ac.uk/register-apply/>. The authors did not receive special access privileges to the data that others would not have. Interested researchers will be able to access the data in the same manner by applying directly to the UK Biobank.

Funding: The authors received no external funding for this work. AG, GSD, EST and RA are employed by a commercial company, Genetic Technologies

Abstract

Colorectal cancer risk stratification is crucial to improve screening and risk-reducing recommendations, and consequently do better than a one-size-fits-all screening regimen. Current screening guidelines in the UK, USA and Australia focus solely on family history and age for risk prediction, even though the vast majority of the population do not have any family history. We investigated adding a polygenic risk score based on 45 single-nucleotide polymorphisms to a family history model (combined model) to quantify how it improves the stratification and discriminatory performance of 10-year risk and full lifetime risk using a prospective population-based cohort within the UK Biobank. For both 10-year and full lifetime risk, the combined model had a wider risk distribution compared with family history alone, resulting in improved risk stratification of nearly 2-fold between the top and bottom risk quintiles of the full lifetime risk model. Importantly, the combined model can identify people ($n = 72,019$) who do not have family history of colorectal cancer but have a predicted risk that is equivalent to having at least one affected first-degree relative ($n = 44,950$). We also confirmed previous findings by showing that the combined full lifetime risk model significantly improves discriminatory accuracy compared with a simple family history model 0.673 (95% CI 0.664–0.682) versus 0.666 (95% CI 0.657–0.675), $p = 0.0065$. Therefore, a combined polygenic risk score and first-degree family history model could be used to improve risk stratified population screening programs.

Introduction

Colorectal cancer is the fourth deadliest cancer, causing nearly 900,000 deaths every year globally. Worldwide, colorectal cancer is the 2nd most common cancer in women and the 3rd in men, with men having around 25% higher incidence and mortality compared with women [1, 2]. Colorectal cancer has several non-modifiable risk factors, including age, family history, sex and genetic makeup. Roughly 5%–10% of colorectal cancer cases have an affected first-degree

Limited (EST is employed by Phenogen Sciences Inc, a subsidiary of Genetic Technologies Ltd), which provided support in the form of salaries but did not have any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of all authors are articulated in the Author Contributions section.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: AG, GSD, EST and RA are employees of Genetic Technologies Limited (EST is employed by Phenogen Sciences Inc, a subsidiary of Genetic Technologies Ltd). The model for colorectal cancer susceptibility analysed in this study is a commercial product of Genetic Technologies Limited. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

relative, and the strength of associated risk depends on the number and closeness of the relationship, and on the ages at diagnosis of the affected relative(s) and the age of the at risk consult [3–9]. Efforts to better understand heritability of colorectal cancer in family studies underscore the complex relationship with environmental components [10, 11].

Rare high penetrance mutations have been found to cause hereditary colorectal cancers, including those predisposing to Lynch syndrome and familial adenomatous polyposis, accounting for 5%–7% of all colorectal cancer cases. Known genetic mutations account for only half of the cases in persons with such family histories [12]. The unexplained causes of cases with a family history could be due to polygenic factors, such as common low penetrance single-nucleotide polymorphisms (SNPs) [13, 14] or lifestyle causes that are also shared by family members [15].

In recent years, an increasing number of susceptibility SNPs have been identified by genome-wide association studies, which examine vast numbers of variants across the genome for associations with disease risk [16, 17]. Although each susceptibility SNP has a weak association with colorectal cancer risk, the cumulative association of many SNPs combined as a polygenic risk score (PRS) can result in a substantial risk gradient (in both directions) and is potentially an effective risk stratification method [14, 18]. For example, Jenkins et al. [19] used a cohort enriched for family history to show the value of a PRS in stratifying individuals by risk, particularly those with a family history but not found to be carriers of mutations associated with Lynch syndrome or familial adenomatous polyposis. Importantly, their study confirmed that a 45 SNP panel in conjunction with having a family history of colorectal cancer could identify non-trivial proportions of the population who would likely benefit from earlier screening. The use of polygenic risk models to inform targeted screening has potential benefit in clinical genetics settings for families in which high-risk mutations cannot be identified [18].

Notwithstanding that observation, the reality is that about 90% of colorectal cancer cases have no family history in first-degree relatives and it is this group that could benefit from improved risk prediction [3]. Given the incidence of colorectal cancer diagnosed before age 50 years is increasing [20, 21], it is particularly important to focus on risk prediction to accurately identify at-risk adults who may not be identified by current standard screening guidelines. Therefore, there is an important justification for improved risk prediction tools to guide screening and risk reduction.

Our aim is to investigate whether better risk stratification can be achieved in the general population, using the UK Biobank, a prospective population-based cohort. To this end, we have investigated the ability of a model comprising 45 SNPs (PRS) and first-degree family history to stratify risk in the general population and the discriminatory performance and calibration of the model to inform the potential utility in broad application risk stratified screening.

Methods

Study sample

The UK Biobank is a major biomedical database, comprises of 500,000 volunteers who were aged 40–69 years when recruited in 2006–2010 from England, Scotland and Wales. The purpose of the UK Biobank is to assist researchers in studying disease prevention, diagnosis and treatment and investigate the determinants of a wide spectrum of diseases in middle and later life [22, 23]. The UK Biobank has Research Tissue Bank approval (REC #16/NW/0274) that covers analysis of data by approved researchers. All participants provided written informed consent to the UK Biobank before data collection began. This research has been conducted using the UK Biobank resource under Application Number 47401.

Table 1. Eligibility criteria.

N eligible	Criteria	N dropped
502,488	Active participants in UK Biobank	
487,869	Reported sex same as genetically determined sex	14,619
409,289	White British and genetically Caucasian	78,580
406,745	No previous diagnosis of colorectal cancer at baseline	2,544
404,715	Aged 40–69 years at assessment date	2,030
403,998	Genome-wide SNP data available	717

<https://doi.org/10.1371/journal.pone.0251469.t001>

Each participant has provided detailed personal and medical history information and has undergone physical and biological measurements. Samples provided include blood, urine and saliva. All participants who provided blood have been genotyped and genome-wide SNP data is available for each [24]. All participants have agreed to their health status being followed-up via linkage to health registries and general practice and hospital records. Therefore, the UK Biobank is a powerful resource to study genetic associations and disease risk due to being a prospective cohort, its large size, and the wealth of genetic and clinical information it has and will collect. The eligibility criteria for this study are described in Table 1.

Generation of PRS

A PRS was calculated for each UK Biobank participant using the 45 SNPs (S1 Table) that were found to be associated with colorectal cancer by previous studies [13, 25]. Using the method of Mealiffe et al [26], we computed for each SNP a (relative) risk score utilising previously published estimates of the odds ratio (OR) per risk allele and risk allele frequency (p) [19]. For each individual SNP, we calculated the unscaled population average risk using the formula:

$$\mu = (1 - p)^2 + 2p(1 - p)OR + p^2OR^2.$$

Weighted risk values were used to normalise the population average to 1, which were calculated as $1/\mu$, OR/μ and OR^2/μ for the three genotypes (defined by the number of risk alleles 0, 1, or 2). The polygenic risk score for each participant was generated by multiplying the weighted risk values for each of the 45 SNPs (assuming independent and additive risks on the log odds scale) [19].

Outcome

The outcome of interest was invasive colorectal cancer diagnosis after baseline assessment. Colorectal cancer was identified using linked cancer registry data using ICD-9 (1530–1539, 1540–1541), ICD-10 (C18–C20) codes or self-reported disease. Follow-up began at date of baseline assessment and observations were censored at the earliest of date of diagnosis, date of death or 31 March 2016 (the latest date for which linkage to cancer registries is complete), whichever occurred first. For analysis of standardised incidence ratios (SIR) for 10-years of follow-up, we ceased follow-up after 10 years.

Risk scores

We evaluated the following two models involving: (i) family history only (based on number of affected first-degree relatives) and (ii) a combination of both family history and the PRS (combined model). Relative risks for having 0, 1 or ≥ 2 first-degree relatives diagnosed with colorectal cancer were obtained from a previous study [27], and centred to have a population average of 1. In the study by Roos et al [27], the authors performed a systematic review and meta-

analysis of cohort and case-control studies to investigate the effect of family history on colorectal cancer relative risk and absolute risk values. We used the relative risk values for family history calculated from cohort studies, which were based mainly on white populations from Western and Northern Europe, United States and Australia. One study focused on Asian population. The PRS model used 45 independent SNPs described previously by Jenkins et al [13, 19]. For SNPs rs10904849, rs35509282, rs4925386 and rs10911251, we used surrogate SNPs rs10904850, rs11100443, rs11204472 and rs6669796 respectively, and for 19qhap (19q13.2) and 11qhap (11q12.2), we used the tag SNPs rs1800469 and rs174537 respectively (S1 Table).

The combined model was created by multiplying the family history and PRS relative risk terms. Calculation of absolute 10-year risk was performed using sex- and age-specific incidence rates for England in 2013, and took into account competing mortality, obtained from the UK Office for National Statistics [28]. For the calculation of the absolute full lifetime risk to age 85, mortality rates were excluded. Risk scores were centred to have a mean of 1. SIRs were calculated using the observed vs expected colorectal cancer incidence based on population gender- and age-specific incidence rates for England in 2006–2016 [29]. Confidence intervals for the SIR were calculated using the default method of a quadratic approximation to the Poisson log likelihood for the log-rate parameter [30].

Statistical analysis

Model performance. Model discrimination was determined using the area under the receiver operating characteristic curve (AUC). We assessed model calibration using logistic regression analysis [31], for which the observed colorectal cancer case status was the dependent variable and the log-odds of our model's predicted probability for the outcome of colorectal cancer during the follow-up time was the independent variable. The test for dispersion was performed by evaluating the null hypothesis that the estimated regression coefficient was equal to 1 in the model without a constant term [31]. Overdispersion occurs when the observed values have greater variability than the expected values produced by the model, while underdispersion occurs when the observed values show less variation than expected. This is measured using logistic regression where a slope >1 suggests predicted risks are too extreme and a slope <1 suggests predicted risks are too moderate. We used logistic regression with no intercept terms to assess dispersion for the 10-year risk and full lifetime risk for the combined model. The coefficients were $\beta = 1.088$ for 10-year risk and $\beta = 1.848$ for full lifetime risk.

Broad sense calibration was measured using 10-year follow-up data from the UK Biobank, for which the SIR (observed/expected incidence) was calculated for both models.

All statistical analyses were performed using Stata version 16.1 [30]. All statistical tests were two sided and $p < 0.05$ was considered nominally statistically significant.

Results

Characteristics of participants and the mean and median PRS (relative risk), and 10-year and full lifetime risks for the combined model are summarised in Table 2. The mean age at baseline of colorectal cancer cases and controls was 61.45 years (SD 6.33) and 57.28 years (SD 7.96), respectively.

Overall, the SIR of observed colorectal cancer compared with the number expected using age- and gender-specific population incidences was 0.92 (95% CI = 0.88–0.95) (Table 3), meaning that the colorectal cancer incidence in the UK Biobank data was 8% (95% CI = 5–12%) less than expected. Furthermore, the SIR broken down by gender showed that the expected incidence for females and males was 6% (95% CI = 1–11%) and 10% (95% CI = 6–14%), respectively, less than expected (Table 3). When the SIR was broken down by age, for ages 60–

Table 2. Summary statistics for the eligible UK biobank cohort.

	Unaffected Total 401,006	Affected (incident cases) Total 2,992
	N (%)	N (%)
Age at cohort entry (years)		
40–49	88,648 (22.11)	198 (6.62)
50–59	133,056 (33.18)	800 (26.74)
60–69	179,302 (44.71)	1,994 (66.64)
Age when diagnosed with colorectal cancer (years)		
40–49	–	79 (2.64)
50–59	–	496 (16.58)
60–69	–	1,647 (55.05)
70–79	–	770 (25.74)
Gender		
Female	217,501 (54.24)	1,275 (42.61)
Male	183,505 (45.76)	1,717 (57.39)
Number of first-degree relatives diagnosed with colorectal cancer		
0	356,437 (88.89)	2,544 (85.03)
1	42,129 (10.51)	412 (13.77)
2+	2440 (0.61)	36 (1.20)
	Mean (SD) median	Mean (SD) median
Full lifetime risk: combined model	0.068 (0.041) 0.057	0.080 (0.050) 0.067
10-year risk: combined model	0.011 (0.009) 0.008	0.016 (0.013) 0.013
Relative risk: PRS	0.999 (0.574) 0.848	1.112 (0.657) 0.934

<https://doi.org/10.1371/journal.pone.0251469.t002>

69 (the majority of cases), the colorectal cancer incidence was ~11% (95% CI = 7–15%) less than expected. This is consistent with the recognized healthy volunteer selection bias of the UK Biobank [32, 33]. For ages 50–59 the incidence was ~3% less than expected, and for ages 40–49, the incidence was 6% higher than expected, but the confidence intervals included 1.

Model performance

For full lifetime risk, the AUC for the combined model was 0.673 (95% CI 0.664–0.682 and the AUC for the family history model was 0.666 (95% CI 0.657–0.675). For 10-year risk, the AUC of the combined model was 0.674 (95% CI 0.665–0.683) and the AUC of the family history model was 0.668 (95% CI 0.659–0.677). The difference between the model fits was significant (10-year risk: $\chi^2 = 7.16$, $df = 1$, $p = 0.0075$; full lifetime risk: $\chi^2 = 7.42$, $df = 1$, $p = 0.0065$).

The 10-year risk combined model was slightly under-dispersed (dispersion coefficient 1.08, 95% CI 1.07–1.09), while the full life risk combined model was considerably under-dispersed (dispersion coefficient 1.84, 95% CI 1.83–1.86). Next, we performed broad sense calibration by analysing the observed and expected ratio (SIR) using 10-year follow-up data. Our data showed a small overestimation of risk for both the family history model (SIR = 0.94, 95% CI 0.91–0.98) and the combined model (SIR = 0.95, 95% CI 0.91–0.98), compared with the population incidence data.

Risk stratification

We investigated the risk distributions of the family history model (Fig 1A and 1C) and the combined model (Fig 1B and 1D) using the entire eligible UK Biobank cohort. Fig 1A shows the full lifetime risk distribution for the family history model, where there are six possible

Table 3. Standardised incidence ratios (SIR) using age- and gender-specific incidence rates—Overall and by subgroups.

	Total observations	O	E	SIR	95% CI
Overall risk	403,998	2992	3253.20	0.92	0.88–0.95
Risk by gender:					
Female	218,776	1275	1355.90	0.94	0.89–0.99
Male	185,222	1717	1897.29	0.90	0.86–0.94
Risk by age-group:					
40–49	88,846	198	186.61	1.06	0.92–1.22
50–59	133,856	800	824.73	0.97	0.90–1.04
60–69	181,296	1994	2241.86	0.89	0.85–0.93
Combined model—10-year risk					
Quintile 1 (lowest)	80,368	166	199.63	0.83	0.71–0.96
Quintile 2	80,575	374	446.53	0.83	0.75–0.92
Quintile 3	80,753	552	679.77	0.81	0.74–0.88
Quintile 4	80,958	757	868.69	0.87	0.81–0.93
Quintile 5 (highest)	81,344	1143	1058.57	1.08	1.01–1.14
Combined model—Full lifetime risk					
Quintile 1 (lowest)	80,568	366	543.82	0.67	0.60–0.74
Quintile 2	80,727	526	614.98	0.85	0.78–0.93
Quintile 3	80,762	561	657.90	0.85	0.78–0.92
Quintile 4	80,849	648	694.81	0.93	0.86–1.00
Quintile 5 (highest)	81,092	891	741.70	1.20	1.12–1.28
By number of affected first-degree relatives					
0	358,981	2544	2857.48	0.89	0.85–0.92
1	42,541	412	371.93	1.10	1.00–1.22
2	2,409	35	23.02	1.52	1.09–2.11
Combined model—10-year risk: (No first-degree family history)					
Quintile 1 (lowest)	71,412	124	171.16	0.72	0.60–0.86
Quintile 2	71,616	329	383.60	0.85	0.77–0.95
Quintile 3	71,753	465	592.75	0.78	0.71–0.86
Quintile 4	71,951	664	767.75	0.86	0.80–0.93
Quintile 5 (highest)	72,249	962	942.22	1.02	0.95–1.08
Top 10%	36,189	546	499.22	1.09	1.00–1.19
Top 5%	18,126	305	259.97	1.17	1.04–1.31
Bottom 10%	35,680	36	63.30	0.56	0.41–0.78
Bottom 5%	17,836	14	26.90	0.52	0.30–0.87
Combined model—Full lifetime risk: (no first-degree family history)					
Quintile 1 (lowest)	71,621	333	479.16	0.69	0.62–0.77
Quintile 2	71,725	438	541.21	0.81	0.73–0.88
Quintile 3	71,782	494	579.08	0.85	0.78–0.93
Quintile 4	71,834	547	609.60	0.89	0.82–0.97
Quintile 5 (highest)	72,019	732	648.44	1.12	1.05–1.21
Top 10%	36,051	408	330.10	1.23	1.12–1.36
Top 5%	18,035	214	167.53	1.27	1.11–1.46
Bottom 10%	35,795	151	229.02	0.66	0.56–0.77
Bottom 5%	17,901	79	111.42	0.71	0.56–0.88

SIR was calculated based on number of cases observed and expected using sex-specific UK population rates of colorectal cancer incidence rates, calculated for the entire eligible UK Biobank cohort or by gender, age group or by family history status, and stratified by full lifetime and 10-year risk categories for the combined model.

Abbreviations: O = observed, E = expected.

<https://doi.org/10.1371/journal.pone.0251469.t003>

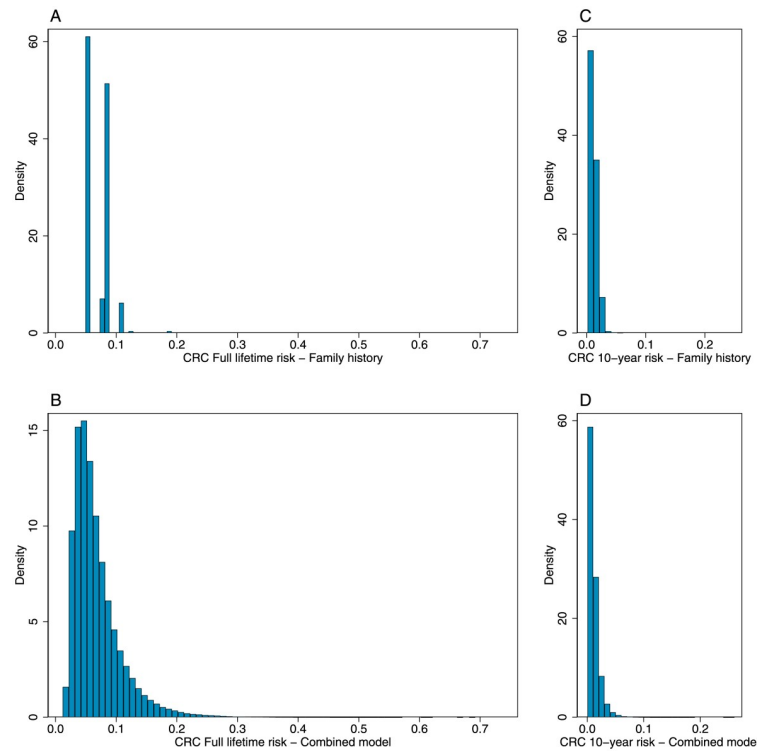


Fig 1. Risk distribution plots for the eligible UK Biobank participants. Plots show the Full lifetime risk distribution for a model with family history only (A) and the combined model (B), and 10-year risk for the family history model (C) and the combined model (D).

<https://doi.org/10.1371/journal.pone.0251469.g001>

categories (0, 1 and 2+ affected first-degree relatives by gender) (median = 0.073, inter-quartile range = 0.027, min = 0.053, max = 0.188). Fig 1B shows the full lifetime risk distribution for the combined model (median = 0.057, inter-quartile range = 0.042, min = 0.011, max = 0.688). Fig 1C and 1D show the 10-year risk distribution for the family history model (median = 0.010, inter-quartile range = 0.009, min = 0.001, max = 0.055) and the combined model (median = 0.008, inter-quartile range = 0.009, min = 0.0004, max = 0.251) respectively.

The SIRs by quintiles of full lifetime risk and 10-year risk for the combined model are shown in Table 3 and Fig 2. We observed an increase in risk gradient between full lifetime risk categories; persons in the top quintile of risk have ~35% higher colorectal cancer incidence than those in the middle quintile and ~53% higher colorectal cancer incidence than those in the bottom quintile. The 10-year risk quintile gradient was less than the full lifetime risk gradient, but showed the same trend. To compare risk stratification of persons with a family history to those without, we also broke down the SIR analysis by number of affected first-degree relatives and for people without any family history. We observed that the top quintile, decile and 95th percentile (for participants without family history) have similar risk values, compared to someone with 1 affected first-degree relative. Also, the risk for people with 2 affected first-degree relatives overlaps with the top risk categories (due to the large confidence interval) (Fig 2). Although the range of SIR is diminished in the 10-year risk graph, the trend is still visible (Fig 2D).

Discussion

Colorectal cancer is a major public health issue worldwide, with high incidence in many westernised countries [34], in addition to increasing incidence for young adults [35, 36]. Several

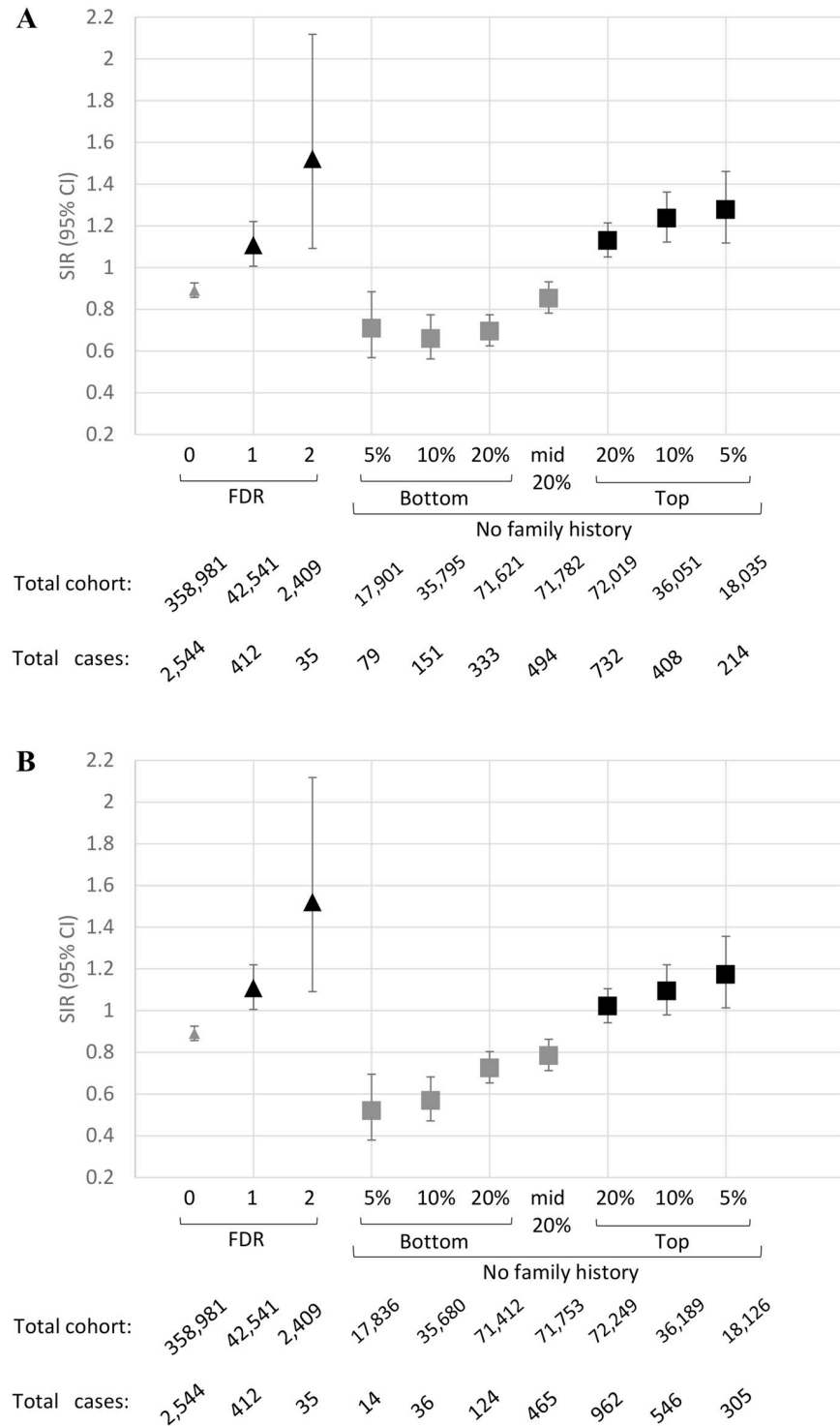


Fig 2. Comparison of the standardised incidence ratios (SIR) for different subgroups. SIR values were generated based on number of cases observed and expected using sex-specific UK population incidences for the number of affected first-degree relatives (FDR) vs the combined model for people without a family history. SIR values were plotted against number of affected first-degree relatives in comparison with full lifetime (A) and 10-year (B) risk categories for participants without family history.

<https://doi.org/10.1371/journal.pone.0251469.g002>

modalities for early detection exist including colonoscopy and faecal occult blood testing. Evident in long-term trends available from US Surveillance, Epidemiology, and End Results data, prevention of advanced colorectal cancer is feasible with screening programs based on colonoscopy as opposed to faecal occult blood testing. In comparison with UK and Australian data, US colorectal incidences are below their Western counterparts and this could be due to differences in screening programs. But every program comes at a cost (colonoscopy versus faecal occult blood testing, for example). The US approach (by the American Cancer Society) [37] to rising incidence in young adults is to lower colonoscopy screening age to 45—an approach that will surely be more effective at detecting early onset disease, but at a far greater cost and an increased risk of over screening thousands of adults. Furthermore, Ladabaum et al. [38] provided evidence that screening compliance would be a more efficacious approach to reduce colorectal incidence and death. The National Colorectal Roundtable announced in 2018 the goal of achieving 80% colorectal screening participation in every community in the US. Building on that compliance goal, and keeping early onset disease in mind, novel risk stratification approaches can only improve screening outcomes by enabling a focus on at-risk persons.

The majority of colorectal cancer cases do not have monogenic (Lynch syndrome and familial adenomatous polyposis) causes, but have [39] multifactorial causes due to genetic, environmental and lifestyle factors [1]. Risk stratification of the general population will assist in identifying those at higher risk and enable the implementation of targeted screening and risk reduction for this group. Currently, screening decisions in the general population are based on age and family history (in UK, USA, and Australia), and recommendations for early screening are based on the number (and age at diagnosis) of affected first-degree and second-degree relative(s) (as determined by each country's medical bodies) [40–42]. Basing screening decisions on family history alone has its caveats, including incorrect reporting of cases in relatives due to lack of knowledge of the cancer diagnosis, or of the site of the cancer [43, 44]. However, as the vast majority of colorectal cancer cases do not have a first-degree relative [3, 4, 45], current screening guidelines do not accurately identify persons above population average risk thresholds. There are many other factors involved in the risk of colorectal cancer, including genetic, environmental and lifestyle factors that, if measured and taken into account, can more accurately identify where people are with respect to risk-based screening thresholds.

Given that there is the potential for more than 40% of colorectal cancer cases to be prevented by behavioural modification, risk-stratification based on non-modifiable risk factors (like family history and polygenic risk) could allow for pre-emptive screening and, importantly, cost-effective risk-reduction options. Notably, the potential benefits of a so-called “healthy lifestyle” on colorectal cancer incidence appears to be evident across all polygenic risk categories [46, 47].

In this study, we evaluated how much a PRS based on 45 SNPs [19] improves colorectal cancer risk prediction when added to a simple family history model. By confirming the performance of a PRS originally constructed using a cohort enriched for family history [19], we have therefore demonstrated the clinical validity of this risk measure for the general population. We have shown that adding a PRS to a model that includes only family history results in modestly improved discriminatory performance. We have also shown that the variance of the risk distribution of the combined model is much greater than that of the family history alone model, Fig 1). As a consequence, for the UK Biobank participants, using the combined model there were more than 29,000 (~17% of) males with no affected first-degree relative but a full lifetime risk scores greater or equal to ~11% (the family history model risk score of a male with one affected first-degree relative in the UK Biobank). There were more than 34,000 (~17% of) females with no affected first-degree relative with full lifetime risk scores (combined model) greater or equal to ~7.3% (the family history model risk score of a female with one affected first-degree

relative in the UK Biobank). In agreement with previous data [48–50], using a PRS we are able to identify 72,019 participants with an increased risk equivalent to having an affected first-degree relative. Importantly, the combined model captures the crucial components of non-modifiable colorectal cancer risk.

In summary, we have found that stratifying colorectal cancer risk by including a PRS with first-degree family history results in an improved risk prediction compared with using family history alone for a sample that mirrors the general population, for which <15% had a family history. This is in agreement with previous studies that have also shown that a PRS adds substantial value in colorectal cancer risk stratification and explains a sizeable excess risk of colorectal cancer, independent of family history [17, 48]. Our new data strengthens the argument for clinical application of polygenic risk assessment in the general population, and especially for those without a family history, and supports the expansion of current recommendations that focus only on family history and age as the main criteria for screening.

Better colorectal cancer risk stratification in the general population will improve identification of at-risk individuals. A significant finding of our work is that 20% of participants based on PRS have a similar full lifetime risk of colorectal cancer as the ~11% identified solely by a first-degree family history, and therefore should thus be assessed with the same importance.

Reinforcing the importance of the polygenic risk score for assessing risk is the recent finding that four of the SNPs included in our PRS (rs12241008, rs2423279, rs3184504, and rs961253) have been shown to be associated with increasing adenoma count at colonoscopy [51]. Adenoma count is not only an indicator of risk itself but is a measure of colorectal cancer development. Identification of at-risk individuals based on a PRS-integrated model will allow for the improved screening and thus removal of such lesions prior to malignant transformation. Furthermore, there is evidence that the PRS association is stronger for proximal compared with distal disease [52] suggesting that risk assessment could help inform endoscopists' colonoscopy procedural plan, such as a slightly slower withdrawal time [53].

From a health economic perspective, the model used in the present study, which incorporates only non-modifiable risk factors, exceeds the benchmark discrimination threshold ($AUC \geq 0.67$) at which risk stratified colorectal cancer screening is thought to become cost effective [54]. Future iterations of the combined model to include additional risk factors should improve the calibration and discrimination and consequently improve the clinical utility of such a tool for colorectal cancer screening uptake, compliance and screening cessation, and post-polypectomy follow-up.

Conclusion

The practical clinical benefit of a risk assessment model that combines PRS and family history is to identify adults who are at an increased risk of colorectal cancer, sufficient to qualify for supplemental screening recommendations who would not otherwise be identified because they do not have a family history, or do not have a strong enough family history to meet screening thresholds.

Study limitations

Our study has several limitations. First, our model is under-dispersed, both for the 10-year risk and full lifetime risk models. This could be resolved in a future study by recalibrating the model or by updating the model with additional risk factors. We also found a small association between family history and PRS ($OR = 1.11$), which is a limitation of our study, as this association might explain the overestimation of risk we observed in the broad sense calibration for 10-year follow-up. This could be resolved by recalibration of the model in an independent

dataset. Furthermore, the model only takes into account first-degree relatives, and we do not break down the risk of first-degree relative by consultant's age or age at diagnosis of the first-degree relative. Given that familial risk, and the PRS associations, depend on these ages [55], there will be some underestimation of risks for young adults and some overestimation of risk for the majority of adults with mild family history, such as a first-degree relative diagnosed at 70 or older. In two recent meta-analysis, the overall colorectal cancer risk associated with family history was found to be lower than previously reported, suggesting we are likely overestimating familial risk in older adults [27, 56]. Because of the substantial environmental contribution to colorectal cancer, there remains unaccounted, modifiable risk not captured by this combined model. Calibration of the model could be improved by increasing the number of susceptibility SNPs and adding further clinical risk factors in the combined model, including smoking history, alcohol and processed meat consumption and BMI. Secondly, the UK Biobank recruited only between the ages of 40–69 years. There were few incident cases in participants in their 40s. This affected our ability to confirm published evidence [17, 48, 57], suggesting superior clinical utility in PRS to help detect early-onset colorectal cancer before age 50 years; we observed a not significant trend in the expected direction for the few young age at diagnosis cases (S2 Table). Furthermore, a recently published study [58] has identified, using exome data, 76 participants in the UK biobank who are potential Lynch syndrome carriers, 17 of whom are cases. Although these are small proportions of the cohort, they could still bias our results, causing an underestimation in some of our standard incidence ratio estimates. To investigate this, we excluded participants from the analysis based on the published pathogenic variants initially identified [58] and compared the SIR results to the original dataset. We found no difference in comparison with the original analyses, as the majority of these potential Lynch syndrome participants didn't pass our eligibility criteria for the analysis (Table 1), resulting in only two Lynch syndrome cases in the final dataset. Furthermore, as shown in Table 3, SIR estimates from the entire cohort were lower than expected (by ~8%), suggesting the UK biobank cohort is "healthier" with respect to colorectal cancer risk than the general population. "Healthy lifestyle" is associated with colorectal cancer incidence regardless of PRS, and because we do not yet incorporate modifiable risk factors, our model is not accounting for those who are at high risk based on the PRS but are at low risk based on modifiable risk factors, and vice versa. This could also affect the performance of our model in stratifying colorectal cancer risk categories. Additionally, we do not account for risk differences for those participants who underwent bowel screening. Therefore, we could be overestimating short-term risk for those who have had bowel screening.

Ten-year risk scores are meant to assess short-term risk of being diagnosed with colorectal cancer and would be more efficacious for the general population if modifiable risk factors were incorporated. This current model incorporates non-modifiable risk factors and is best suited for determining baseline colorectal cancer risk without the consideration of highly modifiable risk factors attributed to colorectal cancer [59]. Finally, we are aware of the population-specific limitations of this study which was restricted to white, Northern-European population. While there is evidence that many susceptibility SNPs are consistent in the strengths and direction of their associations across ethnicities [60–62], there are ethnic specific-loci and variants that have yet to be incorporated into this model.

Future directions

To improve the model calibration, we plan to perform future analysis using additional colorectal cancer susceptibility risk SNPs and create an expanded combined model with additional risk predictors to produce a more comprehensive colorectal cancer risk assessment tool,

applicable across multiple ethnicities. Improvement and validation of the predictive ability of such a colorectal cancer risk assessment tool will facilitate implementation and ultimately hopefully adoption into routine clinical care.

Supporting information

S1 Table. SNPs used to generate PRS.

(PDF)

S2 Table. Standardised incidence ratios (SIR) by age group for full lifetime and 10-year risk (combined model).

(PDF)

Acknowledgments

We wish to thank Mr Lawrence Whiting for his invaluable expertise in the management of large data files from the UK Biobank.

Author Contributions

Conceptualization: Gillian S. Dite, Erika Spaeth Tuff, Richard Allman.

Data curation: Gillian S. Dite.

Formal analysis: Aviv Gafni.

Project administration: Richard Allman.

Software: Aviv Gafni.

Supervision: Gillian S. Dite.

Writing – original draft: Aviv Gafni.

Writing – review & editing: Aviv Gafni, Gillian S. Dite, Erika Spaeth Tuff, John L. Hopper.

References

1. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet*. 2019; 394(10207):1467–80. [https://doi.org/10.1016/S0140-6736\(19\)32319-0](https://doi.org/10.1016/S0140-6736(19)32319-0) PMID: 31631858
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018; 68(6):394–424. <https://doi.org/10.3322/caac.21492> PMID: 30207593
3. Henrikson NB, Webber EM, Goddard KA, Scrol A, Piper M, Williams MS, et al. Family history and the natural history of colorectal cancer: systematic review. *Genet Med*. 2015; 17(9):702–12. <https://doi.org/10.1038/gim.2014.188> PMID: 25590981
4. Schoen RE, Razzak A, Yu KJ, Berndt SI, Firl K, Riley TL, et al. Incidence and mortality of colorectal cancer in individuals with a family history of colorectal cancer. *Gastroenterology*. 2015; 149(6):1438–45.e1. <https://doi.org/10.1053/j.gastro.2015.07.055> PMID: 26255045
5. Fuchs CS, Giovannucci EL, Colditz GA, Hunter DJ, Speizer FE, Willett WC. A prospective study of family history and the risk of colorectal cancer. *N Engl J Med*. 1994; 331(25):1669–74. <https://doi.org/10.1056/NEJM199412223312501> PMID: 7969357
6. Ramsey SD, Yoon P, Moonesinghe R, Khoury MJ. Population-based study of the prevalence of family history of cancer: implications for cancer screening and prevention. *Genet Med*. 2006; 8(9):571–5. <https://doi.org/10.1097/01.gim.0000237867.34011.12> PMID: 16980813
7. Kune GA, Kune S, Watson LF. The Melbourne Colorectal Cancer Study. Characterization of patients with a family history of colorectal cancer. *Dis Colon Rectum*. 1987; 30(8):600–6. <https://doi.org/10.1007/BF02554806> PMID: 3622163

8. St John DJ, McDermott FT, Hopper JL, Debney EA, Johnson WR, Hughes ES. Cancer risk in relatives of patients with common colorectal cancer. *Ann Intern Med.* 1993; 118(10):785–90. <https://doi.org/10.7326/0003-4819-118-10-199305150-00005> PMID: 8470852
9. Stephenson BM, Finan PJ, Gascoyne J, Garbett F, Murday VA, Bishop DT. Frequency of familial colorectal cancer. *Br J Surg.* 1991; 78(10):1162–6. <https://doi.org/10.1002/bjs.1800781005> PMID: 1958974
10. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer.* 2002; 99(2):260–6. <https://doi.org/10.1002/ijc.10332> PMID: 11979442
11. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med.* 2000; 343(2):78–85. <https://doi.org/10.1056/NEJM200007133430201> PMID: 10891514
12. Aaltonen L, Johns L, Jarvinen H, Mecklin JP, Houlston R. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. *Clin Cancer Res.* 2007; 13(1):356–61. <https://doi.org/10.1158/1078-0432.CCR-06-1256> PMID: 17200375
13. Jenkins MA, Makalic E, Dowty JG, Schmidt DF, Dite GS, MacInnis RJ, et al. Quantifying the utility of single nucleotide polymorphisms to guide colorectal cancer screening. *Future Oncol.* 2016; 12(4):503–13. <https://doi.org/10.2217/fon.15.303> PMID: 26846999
14. 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):1330–9.e14. <https://doi.org/10.1053/j.gastro.2015.02.010> PMID: 25683114
15. Jenkins MA, Ait Ouakrim D, Boussioutas A, Hopper JL, Ee HC, Emery JD, et al. Revised Australian national guidelines for colorectal cancer screening: family history. *Med J Aust.* 2018; 209(10):455–60. <https://doi.org/10.5694/mja18.00142> PMID: 30359558
16. Montazeri Z, Li X, Nyiraneza C, Ma X, Timofeeva M, Svinti V, et al. Systematic meta-analyses, field synopsis and global assessment of the evidence of genetic association studies in colorectal cancer. *Gut.* 2020; 69(8):1460–71. <https://doi.org/10.1136/gutjnl-2019-319313> PMID: 31818908
17. Li X, Timofeeva M, Spiliopoulou A, McKeigue P, He Y, Zhang X, et al. Prediction of colorectal cancer risk based on profiling with common genetic variants. *Int J Cancer.* 2020. <https://doi.org/10.1002/ijc.33191> PMID: 32638365
18. 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; 62(6):871–81. <https://doi.org/10.1136/gutjnl-2011-300537> PMID: 22490517
19. Jenkins MA, Win AK, Dowty JG, MacInnis RJ, Makalic E, Schmidt DF, et al. Ability of known susceptibility SNPs to predict colorectal cancer risk for persons with and without a family history. *Fam Cancer.* 2019; 18(4):389–97. <https://doi.org/10.1007/s10689-019-00136-6> PMID: 31209717
20. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin.* 2017; 67(3):177–93. <https://doi.org/10.3322/caac.21395> PMID: 28248415
21. Weinberg BA, Marshall JL. Colon Cancer in Young Adults: Trends and Their Implications. *Curr Oncol Rep.* 2019; 21(1):3. <https://doi.org/10.1007/s11912-019-0756-8> PMID: 30659375
22. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015; 12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779> PMID: 25826379
23. Centre UBC. UK Biobank: Protocol for a large-scale prospective epidemiological resource. 2007.
24. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018; 562(7726):203–9. <https://doi.org/10.1038/s41586-018-0579-z> PMID: 30305743
25. 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. <https://doi.org/10.1038/ejhg.2017.60> PMID: 28488675
26. Mealiffe ME, Stokowski RP, Rhees BK, Prentice RL, Pettinger M, Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst.* 2010; 102(21):1618–27. <https://doi.org/10.1093/jnci/djq388> PMID: 20956782
27. Roos VH, Mangas-Sanjuan C, Rodríguez-Girondo M, Medina-Prado L, Steyerberg EW, Bossuyt PMM, et al. Effects of Family History on Relative and Absolute Risks for Colorectal Cancer: A Systematic Review and Meta-Analysis. *Clin Gastroenterol Hepatol.* 2019; 17(13):2657–67.e9. <https://doi.org/10.1016/j.cgh.2019.09.007> PMID: 31525516
28. ONS. Cancer Registration Statistics, England. 2013.
29. ONS. Cancer Registration Statistics, England. 2006–2016.

30. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC. 2019.
31. MacInnis RJ, Bickerstaffe A, Apicella C, Dite GS, Dowty JG, Aujard K, et al. Prospective validation of the breast cancer risk prediction model BOADICEA and a batch-mode version BOADICEACentre. *Br J Cancer*. 2013; 109(5):1296–301. <https://doi.org/10.1038/bjc.2013.382> PMID: 23942072
32. Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004; 58(8):635–41. <https://doi.org/10.1136/jech.2003.008466> PMID: 15252064
33. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. 2017; 186(9):1026–34. <https://doi.org/10.1093/aje/kwx246> PMID: 28641372
34. Rabeneck L, Chiu HM, Senore C. International Perspective on the Burden of Colorectal Cancer and Public Health Effects. *Gastroenterology*. 2020; 158(2):447–52. <https://doi.org/10.1053/j.gastro.2019.10.007> PMID: 31622620
35. Siegel RL, Jemal A, Ward EM. Increase in incidence of colorectal cancer among young men and women in the United States. *Cancer Epidemiol Biomarkers Prev*. 2009; 18(6):1695–8. <https://doi.org/10.1158/1055-9965.EPI-09-0186> PMID: 19505901
36. Patel SG, Boland CR. Colorectal Cancer in Persons Under Age 50: Seeking Causes and Solutions. *Gastrointest Endosc Clin N Am*. 2020; 30(3):441–55. <https://doi.org/10.1016/j.giec.2020.03.001> PMID: 32439081
37. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. *CA Cancer J Clin*. 2018; 68(4):250–81. <https://doi.org/10.3322/caac.21457> PMID: 29846947
38. Ladabaum U, Mannalithara A, Meester RGS, Gupta S, Schoen RE. Cost-Effectiveness and National Effects of Initiating Colorectal Cancer Screening for Average-Risk Persons at Age 45 Years Instead of 50 Years. *Gastroenterology*. 2019; 157(1):137–48. <https://doi.org/10.1053/j.gastro.2019.03.023> PMID: 30930021
39. Syngal S, Brand RE, Church JM, Giardiello FM, Hampel HL, Burt RW, et al. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am J Gastroenterol*. 2015; 110(2):223–62; quiz 63. <https://doi.org/10.1038/ajg.2014.435> PMID: 25645574
40. Parkin CJ, Bell SW, Mirbagheri N. Colorectal cancer screening in Australia: An update. *Aust J Gen Pract*. 2018; 47(12):859–63. <https://doi.org/10.31128/AJGP-01-18-4472> PMID: 31212405
41. Lin JS, Piper MA, Perdue LA, Rutter C, Webber EM, O'Connor E, et al. Screening for Colorectal Cancer: A Systematic Review for the US Preventive Services Task Force. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Rockville (MD)2016.
42. Koo S, Neilson LJ, Von Wagner C, Rees CJ. The NHS Bowel Cancer Screening Program: current perspectives on strategies for improvement. *Risk Manag Healthc Policy*. 2017; 10:177–87. <https://doi.org/10.2147/RMHP.S109116> PMID: 29270036
43. Mitchell RJ, Brewster D, Campbell H, Porteous ME, Wyllie AH, Bird CC, et al. Accuracy of reporting of family history of colorectal cancer. *Gut*. 2004; 53(2):291–5. <https://doi.org/10.1136/gut.2003.027896> PMID: 14724166
44. Mai PL, Garceau AO, Graubard BI, Dunn M, McNeel TS, Gonsalves L, et al. Confirmation of family cancer history reported in a population-based survey. *J Natl Cancer Inst*. 2011; 103(10):788–97. <https://doi.org/10.1093/jnci/djr114> PMID: 21562245
45. Lowery JT, Ahnen DJ, Schroy PC 3rd, Hampel H, Baxter N, Boland CR, et al. Understanding the contribution of family history to colorectal cancer risk and its clinical implications: A state-of-the-science review. *Cancer*. 2016; 122(17):2633–45. <https://doi.org/10.1002/cncr.30080> PMID: 27258162
46. Carr PR, Weigl K, Jansen L, Walter V, Erben V, Chang-Claude J, et al. Healthy Lifestyle Factors Associated With Lower Risk of Colorectal Cancer Irrespective of Genetic Risk. *Gastroenterology*. 2018; 155(6):1805–15.e5. <https://doi.org/10.1053/j.gastro.2018.08.044> PMID: 30201362
47. Carr PR, Weigl K, Edelmann D, Jansen L, Chang-Claude J, Brenner H, et al. Estimation of Absolute Risk of Colorectal Cancer Based on Healthy Lifestyle, Genetic Risk, and Colonoscopy Status in a Population-Based Study. *Gastroenterology*. 2020; 159(1):129–38.e9. <https://doi.org/10.1053/j.gastro.2020.03.016> PMID: 32179093
48. Weigl K, Hsu L, Knebel P, Hoffmeister M, Timofeeva M, Farrington S, et al. Head-to-Head Comparison of Family History of Colorectal Cancer and a Genetic Risk Score for Colorectal Cancer Risk Stratification. *Clin Transl Gastroenterol*. 2019; 10(12):e00106. <https://doi.org/10.14309/ctg.000000000000106> PMID: 31800541
49. Weigl K, Chang-Claude J, Knebel P, Hsu L, Hoffmeister M, Brenner H. Strongly enhanced colorectal cancer risk stratification by combining family history and genetic risk score. *Clin Epidemiol*. 2018; 10:143–52. <https://doi.org/10.2147/CLEP.S145636> PMID: 29403313

50. Thomas M, Sakoda LC, Hoffmeister M, Rosenthal EA, Lee JK, van Duijnhoven FJB, et al. Genome-wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. *Am J Hum Genet.* 2020; 107(3):432–44. <https://doi.org/10.1016/j.ajhg.2020.07.006> PMID: 32758450
51. Sullivan BA, Qin X, Redding TSt, Gellad ZF, Stone A, Weiss D, et al. Genetic Colorectal Cancer and Adenoma Risk Variants Are Associated with Increasing Cumulative Adenoma Counts. *Cancer Epidemiol Biomarkers Prev.* 2020; 29(11):2269–76. <https://doi.org/10.1158/1055-9965.EPI-20-0465> PMID: 32928932
52. 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. <https://doi.org/10.1053/j.gastro.2018.03.030> PMID: 29574091
53. Patel VD, Thompson WK, Lapin BR, Goldstein JL, Yen EF. Screening Colonoscopy Withdrawal Time Threshold for Adequate Proximal Serrated Polyp Detection Rate. *Dig Dis Sci.* 2018; 63(11):3084–90. <https://doi.org/10.1007/s10620-018-5187-0> PMID: 29974376
54. Naber SK, Kundu S, Kuntz KM, Dotson WD, Williams MS, Zauber AG, et al. Cost-Effectiveness of Risk-Stratified Colorectal Cancer Screening Based on Polygenic Risk: Current Status and Future Potential. *JNCI Cancer Spectr.* 2020; 4(1):pkz086. <https://doi.org/10.1093/jncics/pkz086> PMID: 32025627
55. Li S, Hopper JL. Age dependency of the polygenic risk score for colorectal cancer. *Am J Hum Genet.* 2021; 108(3):525–6. <https://doi.org/10.1016/j.ajhg.2021.02.002> PMID: 33667395
56. Wong MCS, Huang J, Huang JLW, Pang TWY, Choi P, Wang J, et al. Global Prevalence of Colorectal Neoplasia: A Systematic Review and Meta-Analysis. *Clin Gastroenterol Hepatol.* 2020; 18(3):553–61.e10. <https://doi.org/10.1016/j.cgh.2019.07.016> PMID: 31323383
57. Archambault AN, Su YR, Jeon J, Thomas M, Lin Y, Conti DV, et al. Cumulative Burden of Colorectal Cancer-Associated Genetic Variants Is More Strongly Associated With Early-Onset vs Late-Onset Cancer. *Gastroenterology.* 2020; 158(5):1274–86.e12. <https://doi.org/10.1053/j.gastro.2019.12.012> PMID: 31866242
58. Patel AP, Wang M, Fahed AC, Mason-Suares H, Brockman D, Pelletier R, et al. Association of Rare Pathogenic DNA Variants for Familial Hypercholesterolemia, Hereditary Breast and Ovarian Cancer Syndrome, and Lynch Syndrome With Disease Risk in Adults According to Family History. *JAMA Netw Open.* 2020; 3(4):e203959. <https://doi.org/10.1001/jamanetworkopen.2020.3959> PMID: 32347951
59. Kachuri L, Graff RE, Smith-Byrne K, Meyers TJ, Rashkin SR, Ziv E, et al. Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nat Commun.* 2020; 11(1):6084. <https://doi.org/10.1038/s41467-020-19600-4> PMID: 33247094
60. Schmit SL, Schumacher FR, Edlund CK, Conti DV, Ihenacho U, Wan P, et al. Genome-wide association study of colorectal cancer in Hispanics. *Carcinogenesis.* 2016; 37(6):547–56. <https://doi.org/10.1093/carcin/bgw046> PMID: 27207650
61. Wang H, Schmit SL, Haiman CA, Keku TO, Kato I, Palmer JR, et al. Novel colon cancer susceptibility variants identified from a genome-wide association study in African Americans. *Int J Cancer.* 2017; 140(12):2728–33. <https://doi.org/10.1002/ijc.30687> PMID: 28295283
62. Wang H, Haiman CA, Burnett T, Fortini BK, Kolonel LN, Henderson BE, et al. Fine-mapping of genome-wide association study-identified risk loci for colorectal cancer in African Americans. *Hum Mol Genet.* 2013; 22(24):5048–55. <https://doi.org/10.1093/hmg/ddt337> PMID: 23851122