Establishing a valid approach for estimating familial risk of cancer explained by common genetic variants

Korbinian Weigl, Jenny Chang-Claude, Li Hsu, Michael Hoffmeister and Hermann Brenner

Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center of Tumor Diseases (NCT), Heidelberg, Germany
Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
Division of Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany
Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 581, 69120 Heidelberg, Germany

We critically examined existing approaches for the estimation of the excess familial risk of cancer that can be attributed to identified common genetic risk variants and propose an alternative, more straightforward approach for calculating this proportion using well-established epidemiological methodology. We applied the underlying equations of the traditional approaches and the new epidemiological approach for colorectal cancer (CRC) in a large population-based case–control study in Germany with 4,447 cases and 3,480 controls, who were recruited from 2003 to 2016 and for whom interview, medical and genomic data were available. Having a family history of CRC (FH) was associated with a 1.77-fold risk increase in our study population (95% CI 1.52–2.07). Traditional approaches yielded estimates of the FH-associated risk explained by 97 common genetics variants from 9.6% to 23.1%, depending on various assumptions. Our alternative approach resulted in smaller and more consistent estimates of this proportion, ranging from 5.4% to 14.3%. Commonly employed methods may lead to strongly divergent and possibly exaggerated estimates of excess familial risk of cancer explained by associated known common genetic variants. Our results suggest that familial risk and risk associated with known common genetic variants might reflect two complementary major sources of risk.

Introduction
In the era of genome wide association studies (GWAS), many single nucleotide polymorphisms (SNPs) have been found to be associated with a higher risk for various types of cancers. With increasing sample size of the GWAS consortia, the power to detect common variants with small effects has rapidly increased, and discovery of several SNPs at once is now the rule rather than the exception. Many of the SNPs discovery studies included estimates of how much of the familial risk of cancer can be attributed to previously known and the newly discovered genetic variant(s). They mostly employed an equation first proposed by Cox et al. in

Author contributions: Statistical analysis and interpretation of data, drafting of the manuscript: Weigl K; DACHS study design and supervision, critical revision of the manuscript for important intellectual content, obtained funding: Chang-Claude J; Imputation of genetic data within the GECCO consortium, critical revision of the manuscript for important intellectual content: Hsu L; DACHS study design and supervision, critical revision of the manuscript for important intellectual content, obtained funding: Hoffmeister M; Concept and design of analyses, DACHS study design and supervision, drafting of the manuscript, critical revision of the manuscript for important intellectual content, obtained funding: Brenner H.

Additional Supporting Information may be found in the online version of this article.

Key words: genetic epidemiology, familial risk, common genetic variants, risk proportion, excess risk

Abbreviations: CI: confidence interval; CRC: Colorectal cancer; FH: family history; GRS: genetic risk score; GWAS: genome-wide association study; OR: odds ratio; SNP: single nucleotide polymorphism

Conflict of interest: The authors have nothing to disclose.

Grant sponsor: German Federal Ministry of Education and Research; Grant numbers: 01ER0814, 01ER0815, 01ER1505A, 01ER1505B, 01GL1712, 01KH0404; Grant sponsor: German Research Council; Grant numbers: BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, BR 1704/6-6, CH 117/1-1; Grant sponsor: National Cancer Institute; Grant number: U01-CA185094

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/ijc.32664

History: Received 21 May 2019; Accepted 14 Aug 2019; Online 4 Sep 2019

Correspondence to: Korbinian Weigl, Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 581, 69120 Heidelberg, Germany, Tel.: 0049-6221-42-1340, Fax: 0049-6221-42-1302, E-mail: k.weigl@dkfz.de

Int. J. Cancer: 146, 68–75 (2020) © 2019 The Authors. International Journal of Cancer published by John Wiley & Sons Ltd on behalf of UICC
What’s new?
Today’s methods to evaluate excess familial risk of cancer explained by associated known common genetic variants may lead to strongly divergent and possibly exaggerated estimates. This paper presents an alternative, more straightforward approach using well-established epidemiological methodology. Application in a large population-based case–control study in colorectal cancer supports suggestions that this proportion may be substantially smaller than previously assumed and highly dependent on SNP pruning methods, the assumed risk for having a family history of CRC, and the number of identified SNPs. Rather than reflecting a major subcomponent of familial risk, common genetic variants appear to reflect substantial complementary risk.

2007, or slight modifications thereof" that includes two major components:

1. The relative risk attributable to a given SNP, commonly denoted as $\lambda^*$, is given by $\lambda^* = \frac{\hat{p}(1-\hat{p})}{\hat{q}(1-\hat{q})}$, where $p$ is the population frequency of the minor allele, $q = 1-p$, and $r_1$ and $r_2$ are the relative risks for heterozygotes and rare homozygotes, respectively. If applied to multiple independent SNPs in a multiplicative model, the numerator consists of the sum of the log(λ*) across all SNPs.

Although this approach seems straightforward, there are a number of issues that deserve critical discussion.

- First, the relative risk estimates typically come from the large GWAS that detected the SNPs, and will—if winner’s curse is not appropriately addressed—typically be lower in independent samples, which may lead to substantial overestimation of the proportion of family history (FH) risk explained by the SNPs.

- Second, if the contribution of multiple SNPs is not completely independent but these SNPs are in some linkage disequilibrium (LD), the numerator of the equation may be substantially overestimated. On the other hand, if too restrictive “inclusion criteria” are employed that would consider only totally independent SNPs, the complementary information that correlated SNPs may still convey would be lost, leading to potential underestimation of the numerator.

- Third, the estimator for relative risk associated with FH in the denominator is commonly taken from pooled estimates from epidemiological studies and may differ from the relative risk in the populations used to derive the genetic risks.

- Fourth, the implicit partitioning of the excess risk of family history into some proportion that is explained by known genetic variants and some proportion that is explained by yet to be identified genetic factors neglects the fact that a substantial proportion of familial aggregation may be due to other reasons, such as familial aggregation of environmental or lifestyle risk factors.

- Fifth, carriers of common, low penetrance risk alleles are not restricted to persons with a FH. In fact, given the limited proportion of persons with a FH, the common, low penetrance risk alleles occur more often in persons without FH. These risk alleles are hence not restricted to familial risk, but in fact convey risk independent of FH.

In this article, we propose an alternative, straightforward “epidemiological approach,” using well-established epidemiological methodology that is unaffected by these concerns. We will use colorectal cancer (CRC), the third most common cancer globally, whose heritability was estimated to be 35%,[10] as an example to demonstrate our approach. Next to others, FH has been identified as a major CRC risk factor.[12] In the past decade, more and more SNPs associated with CRC risk have been discovered by GWAS (e.g.[13–16]). Risk increases associated with single SNPs are mostly very small, with odds ratios (OR) for the risk alleles ranging between 1 and 1.1, but polygenic scores based on multiple SNPs were shown to be highly predictive of CRC risk.[17–20]

Methods
Alternative epidemiological approach
In the proposed approach both genomic data and familial risk are derived from the same data set. The underlying concept is graphically depicted in Supporting Information Figure S1. The concept reflects the causal role of both genetic and environmental factors for CRC risk. The association of FH with CRC risk reflects clustering of both types of factors within families. Based on this model, we propose to estimate the proportion of the CRC risk that is associated with having a FH (which is most commonly defined as FH in a first-degree relative) that can be explained by common genetic variants by

\[
\text{Prop(SNPs)} = \left(1 - \frac{RR_a - 1}{RR_b - 1}\right) \times 100,
\]

where $RR_a$ is the relative risk (RR) for FH that is adjusted for common genetic variants, $RR_b$ is the RR for FH that is not adjusted for common genetic variants and both RRs are adjusted for environmental factors affecting CRC risk. Should the genetic variants not explain any of the FH-associated risk, $RR_a$ would equal $RR_b$, in which case Prop(SNPs) would be 0. Should the genetic variants completely (100%) explain the excess CRC risk for a FH beyond the excess risk explained by environmental factors, $RR_a$ would equal 1. In practice, one would expect $RR_a$ to be between 1.0 and $RR_b$ and Prop(SNPs)
Familial risk explained by common genetic variants

Study population

Data for the current analyses were taken from the DACHS study (Darmkrebs: Chancen der Verhütung durch Screening), which has been described in detail elsewhere. Briefly, DACHS is an ongoing population-based case–control study in southwest Germany. Patients with a first diagnosis of CRC aged at least 30 years are eligible for participation. Recruitment is conducted by all 22 hospitals in the study area that offer first line treatment to patients with CRC. Controls are randomly selected from population registries using frequency matching with respect to sex, age and county of residence. This analysis was based on 4,447 cases and 3,480 controls who were recruited from 2003 to 2016 and for whom genomic data were available. The ethics committees of the Medical Faculty at the University of Heidelberg and the Medical Chambers of Baden-Württemberg and Rhineland-Palatinate approved the study. Written informed consent was obtained from all participants.

Data collection

Standardized in-person interviews were conducted with both cases (typically during their hospital stay) and controls (at their homes) by trained interviewers. Detailed information about the participants’ family history and a variety of other risk and preventive factors was collected, and blood or buccal samples were taken. All CRC cases were histologically confirmed.

Genotyping

DNA was extracted from blood samples (in 99.1% of participants) or from buccal cells (in 0.9% of participants) using conventional methods. Details about genotyping and imputation are provided in Supporting Information Table S1.

Table 1. Selected characteristics of the study population, imputed data

| Characteristic                  | Group          | Cases n = 4,447 | Controls n = 3,480 | p-value¹ |
|-------------------------------|----------------|----------------|-------------------|---------|
| **Sex**                       |                |                |                   |         |
| Female                        |                | 1,749 (39.3)   | 1,344 (38.6)      | 0.5206  |
| Male                          |                | 2,698 (60.7)   | 2,136 (61.4)      |         |
| **Age**                       |                |                |                   |         |
| <50                           |                | 220 (5.0)      | 146 (4.2)         | 0.6458² |
| 50-59                         |                | 681 (15.3)     | 521 (15.0)        |         |
| 60-69                         |                | 1,333 (30.0)   | 1,080 (31.0)      |         |
| 70-79                         |                | 1,519 (34.2)   | 1,165 (33.5)      |         |
| ≥80                           |                | 694 (15.6)     | 568 (16.3)        |         |
| **Personal history of colonoscopy** |            | 1,127 (25.4)   | 2027 (58.3)       | <0.0001 |
| **Family history**            |                |                |                   |         |
| No FH                         |                | 3,487 (78.4)   | 2,896 (83.2)      | <0.0001 |
| SDR only³                     |                | 352 (7.9)      | 236 (6.8)         |         |
| FDR⁴                         |                | 608 (13.7)     | 348 (10.0)        |         |
| **Genetic risk score⁵**       |                |                |                   |         |
| Very low                      |                | 274 (6.2)      | 348 (10.0)        | <0.0001 |
| Low                           |                | 335 (7.5)      | 348 (10.0)        |         |
| Low-medium                    |                | 739 (16.6)     | 696 (20.0)        |         |
| Medium                        |                | 880 (19.8)     | 696 (20.0)        |         |
| Medium-high                   |                | 961 (21.6)     | 696 (20.0)        |         |
| High                          |                | 611 (13.7)     | 348 (10.0)        |         |
| Very high                     |                | 647 (14.6)     | 348 (10.0)        |         |
| **Cancer stage**              |                |                |                   |         |
| I                             |                | 1,035 (23.3)   | –                 | –       |
| II                            |                | 1,343 (30.2)   | –                 | –       |
| III                           |                | 1,423 (32.0)   | –                 | –       |
| IV                            |                | 647 (14.5)     | –                 | –       |

Note: Values are expressed as n (%).
¹Test for differences between cases and controls.
²p-value for t-test with continuous variable.
³FH in any SDR only.
⁴FH in at least one FDR.
⁵Classification of GRS: very low, ≤10th percentile; low, 10th–20th percentile; low-medium, 20th–40th percentile; Medium, 40th–60th percentile; medium-high, 60th–80th percentile; high, 80th–90th percentile; very high, >90th percentile; GRS generated with weighted risk alleles (weights equaling the beta-coefficient as found in the respective discovery study) of SNPs not in LD (cutoff 0.95).
Abbreviations: FDR, first-degree relative; FH, family history; SDR, second-degree relative.
Identification and selection of SNPs for the genetic risk score

A literature review was conducted to find SNPs that were found to be associated with a higher risk for persons of European descent as reported in detail elsewhere. Of 105 identified SNPs (Supporting Information Table S2), 97 could be reliably measured or imputed. For six SNPs, the risk allele in our sample was not the same as reported in the respective discovery study (rs1957636, rs72647484, rs7259371, rs2696839, rs11884596, rs2516420). The correlation between the SNPs measured with the correlation coefficient from the “genetics” package in R is depicted in Supporting Information Figure S2.

Statistical analyses

A genetic risk score (GRS) was derived with various approaches. It was both calculated unweighted, as mere sum of risk alleles, or as weighted sum of risk alleles, with weights equal to the log of the per-risk-allele-OR as reported in the discovery study. We further applied various LD thresholds (no LD, D' ≥ 0.95, D' ≥ 0.5, D' ≥ 0.1, or max. one SNP per locus) for inclusion of SNPs in the GRS (only including the most significant SNP of all SNPs in LD in logistic regression models within our sample), which resulted in different numbers of included SNPs (97, 90, 80, 71 and 59, respectively). Additionally, both continuous GRS and categorized GRS were analyzed, with GRS categories defined by percentiles of the GRS distribution in controls: 0–10, 10–20, 20–40, 40–60, 60–80, 80–90, 90–100.

Proportions of FH risk explained by common genetic variants were estimated according to 4 traditional approaches, using the formulas given by the respective paper, for 4 different ORs for FH, for 5 different inclusion criteria for common genetic variants under observation and both with relative risk estimates and risk allele frequencies for the genetic variants from the discovery studies and with relative risk estimates and risk allele frequencies from the DACHS study.

Figure 1. Scatter plot and box plot showing the distribution of GRS. The bottom and top of the box indicate the 25th and 75th percentile, with the middle line in the box representing the median. The p-value was estimated using Mann–Whitney U test. FDR, 1st-degree relative; FH, family history; GRS, genetic risk score.
| Assumed risk increase by family history (odds ratio) | Estimated proportion of explained family history risk (%) based on risk estimates for SNPs from Discovery study | DACHS study | Max. 1 SNP per locus |
|--------------------------------------------------|-------------------------------------------------|-------------|---------------------|
|                                                  | All SNPs<sup>2</sup> (N<sub>SNPs</sub> = 97) | Excluding SNPs in LD | Max. 1 SNP per locus |
|                                                  | D' ≥ 0.95 (N<sub>SNPs</sub> = 90) | D' ≥ 0.5 (N<sub>SNPs</sub> = 80) | D' ≥ 0.1 (N<sub>SNPs</sub> = 71) | All SNPs<sup>2</sup> (N<sub>SNPs</sub> = 97) | D' ≥ 0.95 (N<sub>SNPs</sub> = 90) | D' ≥ 0.5 (N<sub>SNPs</sub> = 80) | D' ≥ 0.1 (N<sub>SNPs</sub> = 71) | |
| Cox<sup>6</sup>                                  | 25.1                                      | 21.2                                      | 18.4                                      | 15.7                                      | 14.2                                      | 15.0                                      | 13.9                                      | 11.9                                      | 11.2                                      | 10.5                                      |
| Pharoah<sup>7</sup>                              | 29.9                                      | 25.5                                      | 22.0                                      | 18.8                                      | 17.0                                      | 15.8                                      | 14.5                                      | 12.4                                      | 11.7                                      | 10.9                                      |
| Zheng<sup>8</sup>                                | 27.8                                      | 24.0                                      | 20.7                                      | 17.6                                      | 16.0                                      | 15.0                                      | 13.9                                      | 11.9                                      | 11.2                                      | 10.5                                      |
| Jenkins<sup>9</sup>                              | 26.4                                      | 22.5                                      | 19.5                                      | 16.7                                      | 15.1                                      | 14.0                                      | 12.8                                      | 11.0                                      | 10.3                                      | 9.6                                       |
| 2.2                                              | Cox<sup>6</sup>                            | 27.8                                      | 23.5                                      | 20.3                                      | 17.5                                      | 15.7                                      | 16.7                                      | 15.4                                      | 13.2                                      | 12.5                                      | 11.7                                      |
| Pharoah<sup>7</sup>                              | 31.2                                      | 26.6                                      | 23.0                                      | 19.7                                      | 17.8                                      | 16.5                                      | 15.1                                      | 13.0                                      | 12.2                                      | 11.4                                      | 11.7                                      |
| Zheng<sup>8</sup>                                | 30.9                                      | 26.6                                      | 23.0                                      | 19.6                                      | 17.8                                      | 16.7                                      | 15.4                                      | 13.2                                      | 12.4                                      | 11.7                                      | 11.7                                      |
| Jenkins<sup>9</sup>                              | 29.3                                      | 25.0                                      | 21.6                                      | 18.5                                      | 16.7                                      | 15.5                                      | 14.2                                      | 12.2                                      | 11.5                                      | 10.7                                      | 10.7                                      |
| 2.0                                              | Cox<sup>6</sup>                            | 31.6                                      | 26.7                                      | 23.2                                      | 19.9                                      | 17.9                                      | 19.0                                      | 17.5                                      | 15.1                                      | 14.2                                      | 13.3                                      | 13.3                                      |
| Pharoah<sup>7</sup>                              | 32.7                                      | 27.9                                      | 24.1                                      | 20.6                                      | 18.7                                      | 17.3                                      | 15.9                                      | 13.6                                      | 12.8                                      | 12.0                                      | 12.0                                      | 12.0                                      |
| Zheng<sup>8</sup>                                | 35.1                                      | 30.3                                      | 26.2                                      | 22.3                                      | 20.3                                      | 19.0                                      | 17.5                                      | 15.1                                      | 14.2                                      | 13.3                                      | 13.3                                      | 13.3                                      |
| Jenkins<sup>9</sup>                              | 33.3                                      | 28.5                                      | 24.6                                      | 21.0                                      | 19.0                                      | 17.7                                      | 16.2                                      | 13.9                                      | 13.0                                      | 12.2                                      | 12.2                                      | 12.2                                      |
| 1.77<sup>3</sup>                                 | Cox<sup>6</sup>                            | 38.4                                      | 32.5                                      | 28.1                                      | 24.1                                      | 21.7                                      | 23.1                                      | 21.3                                      | 18.3                                      | 17.2                                      | 16.2                                      | 16.2                                      | 16.2                                      |
| Pharoah<sup>7</sup>                              | 34.8                                      | 29.7                                      | 25.7                                      | 21.9                                      | 19.8                                      | 18.4                                      | 16.9                                      | 14.5                                      | 13.6                                      | 12.7                                      | 12.7                                      | 12.7                                      | 12.7                                      |
| Zheng<sup>8</sup>                                | 42.7                                      | 36.8                                      | 31.7                                      | 27.0                                      | 24.6                                      | 23.1                                      | 21.2                                      | 18.3                                      | 17.2                                      | 16.2                                      | 16.2                                      | 16.2                                      | 16.2                                      |
| Jenkins<sup>9</sup>                              | 40.5                                      | 34.6                                      | 29.9                                      | 25.5                                      | 23.1                                      | 21.5                                      | 19.7                                      | 16.9                                      | 15.8                                      | 14.8                                      | 14.8                                      | 14.8                                      | 14.8                                      |

<sup>1</sup>Estimates based on odds ratios and allele frequencies as observed in the DACHS study.<br><sup>2</sup>In the DACHS study, 97 out of 105 SNPs could be analyzed (see Supporting Information Table S2).<br><sup>3</sup>FH estimate obtained in the DACHS study.
Next, we employed the newly suggested epidemiological approach and estimated the proportion of FH risk explained by common genetic variants by \( \text{Prop(SNPs)} = \left(1 - \frac{\text{OR}_{1}^{\text{FH}} / \text{OR}_{0}^{\text{FH}}}{\text{OR}_{1}^{\text{Ref}} / \text{OR}_{0}^{\text{Ref}}} \right) \times 100 \), where both ORs were estimated from multiple logistic regression models as outlined earlier. Different methods of handling the genetic information (GRS vs. separate variables) and, again, different LD thresholds were used for calculating the proportion of explained familial risk. Confidence intervals were computed using bootstrapping methods (\( n = 1,000 \)).

### Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### Results

Table 1 shows some main characteristics of the study population used for this analysis. Sex and age were equally distributed among cases and controls, reflecting matching. FH in a first-degree relative was more common in cases than controls (13.7% vs. 10.0%, \( p \)-value <0.0001), and a higher proportion of cases had a higher GRS compared to controls (\( p \)-value <0.0001). Figure 1 depicts the distribution of the GRS among cases and controls with relation to FH. A major difference in the distribution of the GRS was seen between cases and controls, but not between participants with and without FH, neither in cases nor in controls.

Traditional estimates of the proportion of FH-associated risk that can be statistically explained by common genetic variants are shown in Table 2. The estimated proportions ranged widely, depending on the combination of the employed criteria (LD, assumed OR for FH, estimation approach). Estimates of the proportion explained by the genetic variants were generally much lower when they were based on the relative risk estimates and risk allele frequencies from the DACHS study rather than the discovery studies, but substantial variation was observed even within both groups of estimates, ranging from 9.6% to 23.1% and from 14.2% to 42.5%, respectively.

In the DACHS study, the OR for FH derived from a model adjusted for the matching variables sex and age and a number of environmental factors was 1.77 (95% confidence interval (CI) 1.52 to 2.07), and this OR estimate was only slightly altered by additional adjustment for the GRS, regardless of how restrictive inclusion of SNPs was with respect to LD (Table 3). The proportion of FH associated risk explained by common genetic variants could be quantified as between 6.4% (95% CI: 1.5 to 12.3%) and 8.9% (95% CI: 4.0 to 14.8%) by the epidemiological approach. Very similar results were obtained in sensitivity analyses using the log(ORs) instead of the ORs (Supporting Information Table S3).

In the application of our alternative approach, we considered a number of design options, such as (i) adjustment for all common genetic variants as single variables or (ii) adjustment for common genetic variants through one variable, i.e. a GRS. We calculated the FH-associated risk proportion explained by common genetic variants with both approaches, and furthermore with a large variety of GRS model options and they all consistently yielded estimates in a low range between 5% and 14% (Table 4, Supporting Information Table S3).

### Discussion

Our analyses support suggestions from theoretical considerations that traditional approaches commonly employed in GWAS might overestimate the proportions of familial risk that can be explained by common genetic variants, due to the underlying assumptions. Furthermore, estimates obtained by these approaches strongly varied even with relatively minor variations in the inclusion criteria of SNPs. By contrast, our proposed epidemiological approach yielded much lower, but consistent and robust estimates of the familial proportion explained by known common genetic variants.

---

**Table 3. Association between family history and the risk for colorectal cancer with and without adjustment for various genetic risk scores**

| FH of CRC\(^1\) | No FH | FH in a FDR | % of explained familial risk\(^2\) |
|-----------------|------|-------------|---------------------------------|
| Cases, n (%)    | 3,487 (85.2) | 608 (14.8) | – |
| Controls, n (%) | 2,896 (89.3) | 348 (10.7) | – |
| Model 1, OR (95% CI)\(^3\) | Ref. | 1.77 (1.52–2.07) | – |
| Model 2, OR (95% CI)\(^4\) | Ref. | 1.70 (1.46–1.99) | 8.94 (4.02–14.81) |
| Model 3, OR (95% CI)\(^5\) | Ref. | 1.70 (1.46–1.99) | 8.94 (3.16–15.45) |
| Model 4, OR (95% CI)\(^6\) | Ref. | 1.71 (1.46–2.00) | 7.81 (2.18–13.91) |
| Model 5, OR (95% CI)\(^7\) | Ref. | 1.71 (1.47–2.00) | 7.38 (2.52–13.33) |
| Model 6, OR (95% CI)\(^8\) | Ref. | 1.72 (1.47–2.01) | 6.42 (1.50–12.27) |

\(^1\)Participants with a FH of CRC in a second-degree relative were excluded for this analysis (\( n = 588 \)).
\(^2\)95% confidence intervals for explained familial risk was calculated using bootstrapping methods (\( n = 1,000 \)).
\(^3\)Regression model 1 adjusted for sex, age, education, smoking, hormone replacement therapy among women, BMI and previous colonoscopy.
\(^4\)Continuous, weighted, no LD threshold.
\(^5\)Continuous, weighted, LD threshold \( D' \geq 0.95 \).
\(^6\)Continuous, weighted, LD threshold \( D' \geq 0.5 \).
\(^7\)Continuous, weighted, max. 1 SNP per locus.
\(^8\)Continuous, weighted, max. 1 SNP per locus.

---

© 2019 The Authors. *International Journal of Cancer* published by John Wiley & Sons Ltd on behalf of UICC.
Traditional approaches of calculating the explained proportion of familial risk by common genetic variants have been widely employed in the GWAS literature.\textsuperscript{3,5,15,23–26} Manuscripts reporting on the newly identified susceptibility loci commonly estimate the incremental and total proportion of familial risk explained by the respective SNPs. For example, so far published estimates for the familial risk for CRC allegedly explained by common genetic variants ranged from \( \sim 6\% \) for 10 SNPs\textsuperscript{3} to \( \sim 8\% \) for 20 SNPs under observation\textsuperscript{15} to \( \sim 12\% \) for 76 SNPs\textsuperscript{5} to \( \sim 22\% \)\textsuperscript{6} for 45 SNPs in a simulation study. Most studies took some LD measure into account\textsuperscript{5,6}, while others simply added up contributions of SNPs without LD-pruning.\textsuperscript{15} However, as shown by our analyses, such estimates may be quite sensitive to the underlying assumptions. In our example, the estimates obtained by traditional approaches had a wide range. In particular, the following patterns were observed:

First, all estimates were substantially higher when the relative risk estimates and the risk allele frequencies of the discovery study rather than those of the DACHS study were employed, which might to some degree reflect the well-known phenomenon of “winner’s curse.”\textsuperscript{27} The obtained relative risk estimates and the risk allele frequencies are subject to variation within populations, which might contribute to differing explained risk proportions.

Second, all of the traditional estimates were strongly dependent on the inclusion criteria of SNPs, being approximately 50% higher for the least restrictive approach (including all 97 SNPs) compared to the most restrictive approach (59 SNPs). Using a less restrictive cutoff might lead to overestimation of the explained familial risk by ignoring the redundancy of the risk information conveyed by SNPs in high linkage disequilibrium.

Third, the traditional estimates strongly depend on the assumed risk increase by family history, a direct consequence of the underlying equations, and may be biased if the risk associated with family history in the analyzed study differs from the assumed familial risk.

Although our suggested alternative approach is less susceptible to violations of these assumptions, possible disadvantages of the “epidemiological approach” also have to be kept in mind. Foremost, it requires availability of epidemiological data, especially information on FH of CRC and possible confounding risk factors. For our approach the risk estimate of having a FH of CRC needs to be obtained from the same data set from which the genetic data are analyzed. Many large genomic data sets might lack this kind of information, but could nevertheless conduct our proposed approach in subsets for which FH information is available. Furthermore, some of the considerations that are mentioned earlier might also apply for our approach, such as the issue of winner’s curse if the cohort under investigation was included in deriving the GRS without appropriate correction for overoptimism or the appropriate threshold of the linkage disequilibrium. However, estimates of the familial risk explained by known genetic variants were found to be much

| Linkage disequilibrium | Linkage disequilibrium | Linkage disequilibrium |
|------------------------|------------------------|------------------------|
| \( D' \geq 0.5 \)       | \( D' \geq 0.5 \)       | \( D' \geq 0.5 \)       |
| \( \text{NSNPs} = 90 \) | \( \text{NSNPs} = 80 \) | \( \text{NSNPs} = 71 \) |

Table 4. Proportions of explained familial risk (and 95% confidence intervals) with epidemiological approach, using different adjustment possibilities

Adjustment for common genetic variants

| Adjustment | Categorized \textsuperscript{1} | Continuous \textsuperscript{1} |
|------------|-----------------------------------|-----------------------------|
| Included as separate variables | 13.8 (6.1–23.2) | 11.6 (5.9–17.9) |
| Included as a GRS \textsuperscript{1} | 10.6 (4.8–16.9) | 9.9 (4.5–16.0) |
| Weighted \textsuperscript{2} | 8.6 (3.9–16.9) | 7.7 (3.6–14.7) |
| Continuous | 8.9 (4.0–18.8) | 8.9 (3.2–16.2) |

Note: 95% confidence intervals of explained familial risk was calculated using bootstrapping methods (\( n = 1,000 \)). \textsuperscript{1}Categories defined by percentiles of distribution in controls: 0–10, 10–20, 20–40, 40–60, 60–80, 80–90, 90–100. \textsuperscript{2}Weights equal to log(OR) from discovery study.
less affected by choice of the threshold for the linkage disequilibrium with the proposed epidemiological approach than with the traditional approaches (range for epidemiological approach 5.4–14.3% vs. range for traditional approaches 9.6–23.1%).

It has to be kept in mind that our analysis, like most previous analyses that estimated explained familial risk, was restricted to common variants. In principle, however, analogous approaches could also be applied in analyses additionally encompassing rare, highly penetrant germline mutations.

A specific strength of our study was the availability and use of comprehensive genetic and environmental data from the DACHS study, one of the largest case–control studies on CRC, with an unselected population-based multicenter recruitment.

In summary, we suggest an alternative, straightforward approach based on standard epidemiological methodology to estimate the proportion of FH-associated risk of cancer that is attributable to common genetic risk variants. Application of this approach in a large population-based case–control study on CRC supports suggestions that this proportion may be substantially smaller than previously assumed and highly dependent on SNP pruning methods, the assumed risk for having a FH of CRC and the number of identified SNPs. On the other hand, relevant prevalences of those variants in both people with and without FH imply that, for the example of CRC, polygenetic risk scores based on meanwhile identified common genetic variants explain quite a substantial proportion of overall risk far beyond FH in the total population. Rather than reflecting a major subcomponent of familial risk, common genetic variants appear to reflect substantial complementary risk. Vice versa, rather than exclusively or even primarily reflecting genetic factors, familial aggregation of risks may also reflect to a large extent other components, such as shared environmental or lifestyle factors. While the alternative epidemiological approach was illustrated using CRC as an example, it should be equally applicable to other cancer outcomes, and, in fact, any other disease outcome.

References

1. Eeles RA, Olama AA, Benlloch S, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet 2013;45:385–91. 91e1–2.
2. Goh CI, Schumacher FR, Easton D, et al. Genetic variants associated with predisposition to prostate cancer and potential clinical implications. J Intern Med 2012;271:353–65.
3. Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet 2008;40:1426–35.
4. Michaelidou K, Lindström S, Dennis J, et al. Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551:92–4.
5. Schmit SL, Edlund CK, Schumacher FR, et al. Novel common genetic susceptibility loci for colorectal cancer. J Natl Cancer Inst 2019;111(2):146–157.
6. Jenkins MA, Makalic E, Dowty JG, et al. Genotyping the utility of single nucleotide polymorphisms to guide colorectal cancer screening. Future Oncol 2016;12:503–13.
7. Pharoah PD, Antoniou AC, Easton DF, et al. Polygenes, risk prediction, and targeted prevention of breast cancer. N Engl J Med 2008;358:2796–803.
8. Zheng W, Zhang B, Cai Q, et al. Common genetic determinants of breast-cancer risk in east Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. Hum Mol Genet 2013;22:2539–50.
9. Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in CDH1 is associated with breast cancer risk. Nat Genet 2007;39:352–8.
10. Bray F, Ferlay J, Soerjomataram I, et al. 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.
11. Lichtenstein P, Holm NV, Verkasalo PK, et al. Inheritance of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85.
12. Lowery JT, Ahnen DJ, Schroy PC 3rd, et al. Understanding the contribution of family history to colorectal cancer risk and its clinical implications: a state-of-the-science review. Cancer 2016;122:2633–45.
13. Dunlop MG, Dobbins SE, Farrington SM, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet 2012;44:770–6.
14. Tomlinson IPM, Webb E, Carvajal-Carmona I, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Hum Mol Genet 2007;16:2836–43.
15. Al-Tassan NA, Whiffin N, Hocking FL, et al. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. Sci Rep 2015;5:10442.
16. Peters U, Jiao S, Schumacher FR, et al. Identification of genetic susceptibility loci for colorectal cancer in a genome-wide meta-analysis. Gastroenterology 2013;144:799–807.e24.
17. Hsu L, Jean J, Brenner H, et al. A model to determine colorectal cancer risk using common genetic susceptibility loci. Gastroenterology 2015;148:1330–9.e14.
18. Frampton MJ, Law P, Litchfield K, et al. Implications of polygenic risk for personalised colorectal cancer screening. Ann Oncol 2015;27:429–34.
19. Weigl K, Thomsen H, Balavarca Y, et al. Genetic risk score is associated with prevalence of advanced neoplasms in a colorectal cancer screening population. Gastroenterology 2018;155:88–98.e10.
20. Weigl K, Chang-Claude J, Knebel P, et al. Strongly enhanced colorectal cancer risk stratification by combining family history and genetic risk score. Clin Epidemiol 2018;10:9.
21. Brenner H, Chang-Claude J, Jansen L, et al. Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. Gastroenterology 2014;146:709–17.
22. Brenner H, Chang-Claude J, Seiler CM, et al. Protection from colorectal cancer after colonoscopy: a population-based, case-control study. Ann Intern Med 2011;154:22–30.
23. Broderick P, Carvajal-Carmona I, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet 2007;39:1135–7.
24. Jaeger F, Webb E, Howarth K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet 2008;40:26–8.
25. Tomlinson IP, Webb E, Carvajal-Carmona I, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 2007;39:984–8.
26. Tomlinson IP, Carvajal-Carmona LG, Dobkins SE, et al. Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. PLoS Genet 2011;7:e1002105.
27. Zhong H, Prentice RL. Correcting “winner’s curse” in odds ratios from genome-wide association findings for major complex human diseases. Genet Epidemiol 2010;34:78–91.
28. Tian Y, Kharazmi E, Sundquist K, et al. Familial colorectal cancer risk in half siblings and siblings: nationwide cohort study. BMJ 2019;364:l803.