

Clinically relevant combined effect of polygenic background, rare pathogenic germline variants, and family history on colorectal cancer incidence

Emadeldin Hassanin^{1*}, Isabel Spier^{2,3,4*}, Dheeraj R. Bobbili⁵, Rana Aldisi¹, Hannah Klinkhammer^{1,6}, Friederike David², Nuria Dueñas^{7,8}, Robert Hüneburg^{3,9}, Claudia Perne^{2,3}, Joan Brunet^{4,7,8,10}, Gabriel Capella^{4,7,8}, Markus M. Nöthen², Andreas J. Forstner^{2,11,12}, Andreas Mayr⁶, Peter Krawitz¹, Patrick May⁵, Stefan Aretz^{2,3,4*}, Carlo Maj^{1*}

* shared

1 Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, Germany

2 Institute of Human Genetics, Medical Faculty, University of Bonn, Germany

3 National Center for Hereditary Tumor Syndromes, University Hospital Bonn, Germany

4 European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS) – Project ID No 739547

5 Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

6 Institute for Medical Biometry, Informatics and Epidemiology, Medical Faculty, University Bonn, Germany

7 Hereditary Cancer Program, Catalan Institute of Oncology-IDIBELL, ONCOBELL, Hospitalet de Llobregat, Barcelona, Spain

8 Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Instituto Salud Carlos III, Madrid, Spain

9 Department of Internal Medicine I, University Hospital Bonn, Germany

10 Hereditary Cancer Program, Catalan Institute of Oncology-IDBIGI, 17007 Girona, Spain

11 Centre for Human Genetics, University of Marburg, Germany

12 Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Germany

Corresponding author

Prof. Stefan Aretz, MD

Institute of Human Genetics

University of Bonn

Venusberg-Campus 1

53127 Bonn, Germany

Phone: + 49-(0)228-287-51009

Fax: + 49-(0)228-287-51011

e-mail: Stefan.Aretz@uni-bonn.de

37 **Conflict of interest statement:** No potential conflicts (financial, professional, or personal) rele-
38 vant to the manuscript.

39

40

41 **Author Contributions:**

42 Emadeldin Hassanin - Study concept and design, Statistical analysis, Analysis and interpreta-
43 tion of data, Drafting of the manuscript, Critical revision of the manuscript for important intellec-
44 tual content

45 Isabel Spier - Study concept and design, Analysis and interpretation of data, Drafting of the
46 manuscript, Critical revision of the manuscript for important intellectual content

47 Dheeraj R. Bobbili - Study concept and design, Critical revision of the manuscript for important
48 intellectual content

49 Rana Aldisi, Hannah Klinkhammer, Friederike David, Nuria Dueñas, Robert Hüneburg, Claudia
50 Perne, Joan Brunet, Gabriel Capella, Markus M. Nöthen, Andreas J. Forstner, Andreas Mayr -
51 Critical expert revision of the manuscript for important intellectual content

52 Peter Krawitz, Patrick May - Critical revision of the manuscript for important intellectual content,
53 Study supervision

54 Stefan Aretz, Carlo Maj - Study concept and design, Analysis and interpretation of data, Drafting
55 of the manuscript, Critical revision of the manuscript for important intellectual content, Study
56 supervision

57

58 **Abstract:** (260 words)

59
60 **Background and aims:** Summarised in polygenic risk scores (PRS), the effect of common, low
61 penetrant genetic variants associated with colorectal cancer (CRC), can be used for risk stratifi-
62 cation. **Methods:** To assess the combined impact of the PRS and other main factors on CRC
63 risk, 163,516 individuals from the UK Biobank were stratified as follows: 1. carriers status for
64 germline pathogenic variants (PV) in CRC susceptibility genes (*APC*, *MLH1*, *MSH2*, *MSH6*,
65 *PMS2*), 2. low (<20%), intermediate (20-80%), or high PRS (>80%), and 3. family history (FH) of
66 CRC. Multivariable logistic regression and Cox proportional hazards models were applied to
67 compare odds ratios (OR) and to compute the lifetime incidence, respectively. **Results:** De-
68 pending on the PRS, the CRC lifetime incidence for non-carriers ranges between 6% and 22%,
69 compared to 40% and 74% for carriers. A suspicious FH is associated with a further increase of
70 the cumulative incidence reaching 26% for non-carriers and 98% for carriers. In non-carriers
71 without FH, but high PRS, the CRC risk is doubled, whereas a low PRS even in the context of a
72 FH results in a decreased risk. The full model including PRS, carrier status, and FH improved
73 the area under the curve (AUC) in risk prediction (0.704). **Conclusion:** The findings demon-
74 strate that CRC risks are strongly influenced by the PRS for both a sporadic and monogenic
75 background. FH, PV, and common variants complementary contribute to CRC risk. The imple-
76 mentation of PRS in routine care will likely improve personalized risk stratification, which will in
77 turn guide tailored preventive surveillance strategies in high, intermediate, and low risk groups.

78

79

80 **Key words:**

81 Colorectal Cancer – Family History – Hereditary Cancer – Polygenic Risk – Risk Stratification

82 Introduction

83 Colorectal cancer (CRC) is the fourth leading cancer-related cause of death worldwide.
84 Major established exogenous risk factors are summarized as Western lifestyle. Howev-
85 er, an inherited disposition contributes significantly to the disease burden since up to
86 35% of interindividual variability in CRC risk has been attributed to genetic factors.^{1,2}

87 Around 5% of CRC occur on the basis of a monogenic, Mendelian condition (hereditary
88 CRC), in particular Lynch syndrome (LS) and various gastrointestinal polyposis syn-
89 dromes. Here, predisposing rare, high-penetrance pathogenic variants (PV, constitu-
90 tional / germline variants) result in a considerable cumulative lifetime risk of CRC and a
91 syndrome-specific spectrum of extracolonic tumors. The autosomal dominant inherited
92 LS is by far the most frequent type of hereditary CRC with an estimated carrier frequen-
93 cy in the general population of 1:300-1:500.³⁻⁵ It is caused by a heterozygous germline
94 PV in either of the mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6* or *PMS2* or, in
95 few cases, by a large germline deletion of the *EPCAM* gene upstream of *MSH2*. The
96 most frequent Mendelian polyposis syndrome is the autosomal dominant Familial Ade-
97 nomatous Polyposis (FAP) caused by heterozygous germline PV in the tumor suppres-
98 sor gene *APC*, followed by the autosomal recessive *MUTYH*-associated polyposis
99 (MAP) which is based on biallelic germline PV of the base excision repair gene
100 *MUTYH*.^{6,7} However, even in such monogenic conditions, the inter- and intrafamilial
101 penetrance and phenotypic variability is striking, pointing to modifying exogenous or
102 endogenous factors. Heterozygous (monoallelic) *MUTYH* germline PV may be associ-
103 ated with a slightly increased CRC risk (^{8,9}); the carrier frequency in northern European
104 populations is estimated to be 1:50-1:100.³

105 Approximately 20-30% of CRC cases are characterized by a suspicious, but unspecific
106 familial clustering of CRC (familial CRC). Around 25% of CRC cases occur before 50
107 years of age (early-onset CRC); in around one quarter of those a hereditary type (main-
108 ly LS) has been identified.¹⁰ Although further high-penetrance candidate genes have
109 been proposed,¹¹⁻¹³ the majority of familial and early-onset cases cannot be explained
110 by monogenic subtypes and instead are supposed to result from a multifactorial / poly-
111 genic etiology including several moderate-/intermediate penetrance risk variants and

112 shared environmental factors. A positive family history (FH) in first- and second-degree
113 relatives increases the risk of developing CRC by 2- to 9-fold (^{14;15}), which underpins
114 the hypothesis of shared genetic and non-genetic risk factors.

115 A variety of models to predict CRC risk has been developed and evaluated, which in-
116 clude clinical data, FH, lifestyle factors, and genetic information.¹⁶ For more than a dec-
117 ade, genome-wide association studies (GWAS) in large unselected CRC cohorts identi-
118 fied an increasing number of common, low-penetrance risk variants, mainly single nu-
119 cleotide polymorphisms (SNPs), which are significantly associated with CRC risk.^{17–20}
120 Each SNP risk allele individually contributes only little to CRC risk (OR 1.05 to 1.5),
121 however, summarised in quantitative polygenic risk scores (PRS), the combined effect
122 might explain a substantial fraction of CRC risk variability and can identify individuals at
123 several times lower and greater risk than the general population.^{21–23}

124 As such, it is expected that the genetic background defined by the common risk variants
125 may not only influence the occurrence of late-onset sporadic cases, but also modulate
126 the risk of familial, early-onset, and hereditary CRC.²⁴ Recent studies demonstrated that
127 high PRS values are associated with an increased risk of CRC and other common can-
128 cers in the general population up to an order of magnitude that is almost similar to he-
129 reditary tumor syndromes.^{25,26}

130 Based on these data, it can be hypothesized, that the identification of common genetic
131 CRC risk variants not only provides deep insights into the biological mechanisms and
132 pathways of tumorigenesis, but could improve personalized risk stratification for sporad-
133 ic, familial / early-onset, and hereditary CRC in the future by the implementation of SNP-
134 based PRS screening in routine patient care, which will in turn guide tailored preventive
135 strategies in high, moderate, and low risk groups.

136 However, even if previous studies provide promising results for a clinical benefit of a
137 PRS-based personalized risk stratification, the impact of common risk factors and their
138 interplay with high-penetrance variants and other unspecified factors, captured partly by
139 the FH, still has to be improved and validated in additional patient cohorts.

140 In the present work, we compare the prevalence and the lifetime risk of CRC among
141 163,516 individuals from a population-based European repository (UK Biobank, UKBB).
142 Individuals were stratified according to three major risk factors 1) their carrier status of
143 rare, high-penetrance pathogenic or likely pathogenic germline variants (hereafter de-
144 fined as PV) in the MMR and the *APC* genes, 2) a low, intermediate, or high PRS, and
145 3) a FH for CRC.

146 **Material and methods**

147 **Data Source**

148 UK Biobank (UKBB) genetic and phenotypic data were used in this study. UKBB is a
149 long-term prospective population-based study that has recruited volunteers mostly from
150 England, Scotland, and Wales, with over 500,000 participants aged 40 to 69 years at
151 the time of recruitment. For each participant, extensive phenotypic and health-related
152 data is available; genotyping data is accessible for 487,410 samples, and exome se-
153 quencing data is available for 200,643 people. All participants gave written consent, and
154 the dataset is available for research.²⁷

155 **Study participants**

156 CRC cases were defined based on self-reported code of 1022 or 1023 (in data field
157 20001), or ICD-10 code of C18.X or C20.X, D01.[0,1,2], D37.[4,5], or ICD-9 of 153.X or
158 154.[0,1] (in hospitalization records). Control samples were those that had no previous
159 diagnosis of any cancer. The study includes people of all ethnicities. Outliers for
160 heterozygosity or genotype missing rates, putative sex chromosome aneuploidy, and
161 discordant reported sex versus genotypic sex were excluded. Only individuals
162 (n=200,643) who had both genotyping and whole-exome sequencing (WES) data were
163 considered. If the genetic relationship between individuals was closer than the second
164 degree, defined as kinship coefficient > 0.0884 as computed by the UK Biobank, we
165 removed one from each pair of related individuals (cases were retained if exist).

166

167 **Variant selection**

168 We used ANNOVAR²⁸ to annotate the VCF files from the 200,643 WES samples. The
169 Genome Aggregation Database (gnomAD)²⁹ were used to retrieve variant frequencies
170 from the general population. We focused on rare PV for hereditary CRC (Lynch syn-
171 drome, polyposis) and considered the same variant filtering approach that was used in a
172 recent study aiming at selecting rare PV.³⁰ The following inclusion criteria were used: 1)
173 only *APC*, *MUTYH*, *MLH1*, *MSH2*, *MSH6*, *PMS2* variants in protein-coding regions were
174 included; 2) allele frequency (AF) <0.005 in at least one ethnic subpopulation of
175 gnomAD; 3) not annotated as “synonymous,” “non-frameshift deletion” and “non-
176 frameshift insertion”; 4) annotated as “pathogenic” or “likely pathogenic” based on
177 ClinVar.³¹ We did not include *MUTYH* in the pooled analysis since no biallelic (i.e. high
178 penetrance) case was identified in the cohort; however, we included the heterozygous
179 (monoallelic) carriers in the single gene analysis to compare the effect size with the oth-
180 er genes.

181 **PRS**

182 We applied a previously validated PRS for CRC with 95 variants to calculate the PRS.¹⁷
183 The PRS was computed using the PLINK 2.0³² scoring function through UKBB genotype
184 data. To reduce PRS distributions variance among genetic ancestries, we applied a
185 previous approach.³³ We used the first four ancestry principal components (PCs) to fit a
186 linear regression model to predict the PRS across the full dataset (pPRS~PC1 + PC2 +
187 PC3 + PC4). Adjusted PRS (aPRS) were calculated by subtracting pPRS from the raw
188 PRS and used for the subsequent analysis.

189 **Statistical analysis**

190 Individuals were divided into groups depending on 1) carrier status of PV, 2) PRS, and
191 3) FH. For FH, we considered participants’ reports of CRC in their parents and siblings
192 (data fields: 20110, 20107, 20111). For PRS, individuals were assigned into three
193 groups: low (<20% PRS), intermediate (20-80% PRS), and high (>80% PRS).

194 We conducted both an analysis specific to single genes and a combined analysis (i.e.,
195 carriers of PV in *APC*, *MLH1*, *MSH2*, *MSH6* and *PMS2*). First, we estimated the OR for
196 each carrier group based on a logistic regression adjusting for age at recruitment, sex
197 and the first four ancestry PCs. Afterwards, we additionally incorporated interactions

198 between PV carriers and FH with PRS by introducing an interaction term within the lo-
199 gistic regression model.

200 We calculated the lifetime risk by age 75 from carrier status of rare PV and the PRS
201 based on a Cox proportional hazards model. Individual's age served as the time scale,
202 representing the time to event, for observed cases (age at diagnosis), and censored
203 controls (age at last visit). Carrier status, PRS category, age, sex and the first four an-
204 cestry PCs were incorporated in the model, and adjusted survival curves were pro-
205 duced.

206 Model performance was assessed via the area under the receiver operating characteris-
207 tic curve (AUC), Nagelkerke's Pseudo-R², and the C-index for time-to-event data. R
208 3.6.3 with the corresponding add-on packages *survival* and *survminer* was used for all
209 statistical analyses.

210 Results

211 Stratification of UKBB individuals for CRC prevalence, FH, and PV carrier status

212 We identified 1,902 CRC cases (894 prevalent cases and 1,008 incident cases) among
213 the 163,516 UKBB individuals that retained after exclusion criteria, with a mean age at
214 diagnosis of 60.9 years. The remaining 161,614 individuals with no previous diagnosis
215 of any cancer were considered as controls, with a mean age of 56.9 years at last visit
216 (Table 1). The European population represents 92% of the analyzed cohort.

217 The fraction of individuals with a positive FH of CRC is significantly higher in cases
218 (19%) compared to controls (11%) (OR = 1.95 [1.73-2.19], P < 0.01) and ranges be-
219 tween 9% and 23% in the subgroups (Table 2). There is a significantly higher proportion
220 of individuals with a FH of CRC not only among carriers of PV in the selected cancer
221 susceptibility genes (OR = 1.96 [1.72-2.20], P < 0.01), but also among non-carriers with
222 high PRS (OR = 1.60 [1.31-1.94], P < 0.01).

223 In the analyzed CRC susceptibility genes *APC*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, we identi-
224 fied 399 heterozygous carriers of 111 PV. They were present in 30 (1.57%) cases and

225 369 (0.23%) controls, which is in line with published data. A list of the considered vari-
226 ants and annotations is shown in [Supplementary Table S1](#), a summary of the number of
227 PV carriers per gene is provided in [Supplementary Table S2](#). No individual with a ho-
228 mozygous PV was identified.

229 **PRS distribution within the UKBB cohort**

230 CRC PRS follow a normal distribution both regarding raw and PC-adjusted PRS ([Sup-](#)
231 [plemental Figure S1](#)) and is significantly higher in cases compared to controls ($P <$
232 $2.2e-16$) ([Supplemental Figure S2](#)).

233 The prevalence of CRC according to PRS percentiles demonstrates that values in the
234 extreme right tail of the PRS distribution are associated with a non-linear increase of
235 CRC risk, whereas in the left tail a less evident non-linear decrease can be observed
236 ([Supplemental Figure S3](#)). This supports the hypothesis of using PRS to stratify individ-
237 uals into risk classes (i.e., low, intermediate, and high risk) according to a liability
238 threshold model.

239 **Interplay between PV and PRS**

240 There was no overlap between the selected rare high penetrance PV and the common
241 SNPs used for PRS calculation, and thus, the PRS represents an additional genetic
242 signal. Notably, the PRS distributions showed that the mean and median of PRS is sig-
243 nificantly higher in affected carriers compared to unaffected carriers ($P = 0.004$) ([Sup-](#)
244 [plemental Figure S4](#)).

245 We assessed how CRC risk is influenced by PRS and carrier status for PV in high
246 penetrant CRC susceptibility genes (*APC*, *MLH1*, *MSH2*, *MSH6*, *PMS2*) by calculating
247 the ORs for CRC across groups compared to non-carriers with intermediate PRS as
248 reference group. Non-carriers with a low or high PRS are estimated to have a 0.5-fold or
249 2.1-fold change in the odds for CRC, respectively. We observed that the PRS also al-
250 ters the penetrance of PV in susceptibility genes considerably as PV carriers with high
251 PRS had four times higher OR than carriers with low PRS (OR = 17.5 and 3.9, respec-
252 tively; [Figure 1A](#)). We did not observe a significant interaction between PV carrier status
253 and PRS ($p = 0.87$).

254 The high PRS, which is by definition present in 20% of the non-carriers, is associated
255 with an almost doubled CRC risk (Figure 1A, Table 2). Since the vast majority (97.9%)
256 of non-carriers are controls (=healthy), almost the same percentage results if only
257 healthy non-carriers are considered.

258 Similarly, the lifetime cancer risk analysis shows a combined impact of PV and PRS:
259 Among carriers, the estimated cumulative incidence by age 75 increased from 40% in
260 case of a low PRS to 74% in case of a high PRS compared to 6% to 22% for non-
261 carriers (Figure 1B).

262 **Inclusion of family history on cancer risk stratification**

263 Taking individuals with no FH and intermediate PRS as a reference, both FH and PRS
264 are associated with a higher CRC risk (Figure 1C). The CRC risk for individuals having
265 low PRS and no FH (OR 0.6) is five times lower than for individuals having both positive
266 FH and high PRS (OR 3.1). We did not observe a significant interaction between FH
267 status and PRS ($p = 0.12$). Noteworthy, individuals without FH and high PRS and indi-
268 viduals with FH and intermediate PRS both have similar CRC risks with an OR of
269 around 2, whereas the CRC risk of individuals having low PRS even in the context of a
270 FH is decreased compared to the reference group.

271 Among individuals with FH, the cumulative CRC incidence by age 75 increases 3-fold
272 from 8% in case of a low PRS to 26% in case of a high PRS (Figure 1D). Noteworthy,
273 the cumulative CRC incidence of individuals with a positive FH and an intermediate
274 PRS is lower (16%) than for individuals with negative FH and a higher PRS category
275 (21%), respectively.

276 The full model integrating PRS, FH, and PV status shows that the CRC risk is strongly
277 influenced by PRS in all groups (Figure 2, Supplementary Table S3). Considering the
278 non-carriers with no FH and intermediate PRS group as reference, the CRC OR in low
279 PRS is 0.6 for non-carriers with no FH, while it is estimated more than 60 times higher
280 (OR 40) for carriers with FH and high PRS (Figure 2A). The corresponding cumulative
281 CRC incidences are 6% and 98%, respectively (Figure 2B). Although all PV carriers
282 showed a significantly increased CRC risk, both the PRS and FH modify these risks

283 considerably: depending on the FH and PRS, the OR in PV carriers vary between 4 and
284 40 and the cumulative incidence between 35% and 98%.

285 PRS improved model discrimination over carrier status and FH of CRC in first-degree
286 relatives. The AUC derived from PRS (0.688) was higher compared to those derived
287 using FH (0.654) and carrier status (0.646). The full model including PRS, carrier status,
288 and FH improved the AUC (0.704) in risk prediction by 1.6%, 5%, and 5.8%, respective-
289 ly, and was also better than any combination of two factors (Table 3).

290 **The impact of polygenic risk in single gene mutation carriers**

291 The gene-specific analysis revealed a strong variability in risk conferred by rare hetero-
292 zygous PV in the different genes. The largest effect sizes are attributable for *MLH1* and
293 *APC*, those for *MSH2* and *MSH6* are a bit less, while the effect size for *PMS2* is consid-
294 erably lower (Figure 3). When heterozygous *MUTYH* variants are included in this analy-
295 sis, the risks are very similar to the *PMS2*-related risks. Both the *PMS2* and heterozy-
296 gous *MUTYH* risks show a broad overlap with the non-carrier risks, while there is no
297 overlap between the risks of non-carriers and those with PVs in *MLH1*, *MSH2*, *MSH6*,
298 and *APC*.

299 We estimated how PRS and FH influence CRC prevalence among PV carriers in each
300 of the five susceptibility genes. Despite the different effect sizes, the PRS and FH modi-
301 fies the relative risk across all genes; however, the effect of PRS and FH is conversely
302 related to the penetrance of the gene with the smallest effects in *MLH1* PV carriers.

303 As for the overall analysis, in the gene-specific analysis a positive FH, a PV in a cancer
304 risk gene, and a high PRS are associated with an increased CRC risk. As such, an indi-
305 vidual with a low-penetrance *PMS2* PV, but high PRS and / or positive FH ends up with
306 an estimated CRC risk similar to a *MSH6* PV carrier without FH and / or low PRS.

307 **Discussion**

308 Recent studies demonstrated that the polygenic background, defined as PRS based on
309 disease-associated SNPs, modifies the risks for several cancers of the general popula-
310 tion including CRC considerably, both in terms of age at onset and cumulative lifetime

311 risks.^{11,22,26,34–36} In line with this, the risk alleles of those SNPs are found to also accu-
312 mulate in unexplained familial and early-onset CRC cases.^{24,37} Whereas a low polygenic
313 burden decreases the CRC risk down to one quarter on average, individuals with a high
314 PRS (>80%) doubles and those with a very high PRS (99%) almost quadruplicate their
315 risk and thus, reach a CRC risk in an order of magnitude almost comparable to carriers
316 of hereditary CRC with low PRS.³⁰ In a previous study, Jia et al. found that the risk of
317 CRC is significantly associated with its PRS: Compared with individuals in the lowest
318 PRS quintile those in the highest quintile had a greater than threefold risk (during a 5.8-
319 year follow-up period). Hazard Ratios estimated with the middle quintile as the refer-
320 ence resulted in a risk between 0.56-1.71, a threefold risk in those in the top 1% of
321 PRS, and a 70% reduced CRC risk for individuals in the bottom 1% of the PRS.³⁶

322 To extend these studies on how the CRC prevalence is influenced by genetic suscepti-
323 bility using, we used the sufficiently larger, more robust dataset of the most recent
324 UKBB cohort, incorporate the family history (FH) as an additional factor for risk stratifi-
325 cation, and include a single gene analysis. We considered both the genetic component
326 driven by rare high-penetrance PV associated with hereditary CRC and common low-
327 penetrance variants captured by the PRS.

328 Firstly, our results confirm that the polygenic background strongly modulates CRC risk
329 in the general population. Compared to the average polygenic burden, individuals with a
330 low (<20%) or high (>80%) PRS are estimated to have a 0.5-fold or 2.1-fold change in
331 the odds for CRC, respectively. The additional time-to-event analysis revealed a corre-
332 sponding cumulative lifetime risk of 6% and 22% by age 75. Hence, when the PRS is
333 included in risk calculation, around 20% of healthy individuals of the general population
334 with no FH of CRC have a doubled CRC risk, which is similar to those with a first de-
335 gree relative affected by CRC.³⁸ These so far unknown and otherwise unrecognisable
336 at-risk individuals might need surveillance 10-15 years earlier than usually recommend-
337 ed.³⁹ On the other hand, the around 20% of individuals with low PRS and no FH might
338 need less surveillance than the general population due to a considerably lowered risk,
339 while even those with low PRS and positive FH might not need a more intense surveil-
340 lance than the general population..

341 It is well known that among patients with hereditary CRC syndromes, the age of onset
342 and cumulative CRC incidence is very heterogeneous, even within PV carriers of the
343 same family. The estimated gene-specific, individual CRC lifetime risks of LS patients
344 with *MLH1* or *MSH2* PV can be lower than 10% but as high as 90%-100% in a consid-
345 erable fraction. In the past, the analysis of modifying effects based on common CRC-
346 associated variants in LS and other high-risk groups has been restricted to selected co-
347 horts and small subsets of SNPs.^{40,41} A recent study demonstrated that the polygenic
348 background also substantially influences the CRC risk in LS using UKBB data, even
349 though the ORs for CRC risks could only be predicted due to the small sample sizes.³⁰
350 In the present work, ORs could be calculated directly from the model since over three
351 times more UKBB individuals have been included with six times more CRC cases, and
352 five times more PV carriers.

353 So secondly, we were able to show that the PRS modifies the CRC risks not only in the
354 general population considerably, but also in carriers of a MMR gene PV identified in the
355 general population. For the first time we demonstrated, that this is also true for *APC* PV.
356 Depending on the PRS, the cumulative CRC lifetime incidence in PV carriers ranged
357 between 40% and 74%, and thus, the PRS is able to explain parts of the interindividual
358 variation in CRC risk among PV carriers.

359 However, the single-gene analysis revealed heterogeneous effects across genes and
360 therefore the modifying role of the polygenic background should be framed within the
361 absolute risk attributable to individual genes. As expected, the effect of the PRS seems
362 to be relevant in particular in less penetrant CRC risk genes such as *PMS2* where the
363 OR ranges between 0.94 and 5.43 respectively (Figure 3). This is in line with findings in
364 moderate breast cancer risk genes such as *CHEK2*, *PALB2* and *ATM*⁴²⁻⁴⁴ and suggests
365 that PRS inclusion in risk stratification may in particular be relevant to prevent excess of
366 surveillance measures in PV carriers of those genes.

367 In addition, our results provide evidence that the inclusion of FH can further and inde-
368 pendently improve the risk stratification in both carriers and non-carriers. Including PRS
369 and FH in risk assessment, the cumulative CRC lifetime incidence ranged between 8%
370 and 26%, and in PV carriers between 30% and 98%, and thus, outperformed the con-

371 sideration of a single risk factor. This suggests that familial clustering points to addition-
372 al risk factors besides those captured by common low-risk SNPs (PRS) and rare
373 PV.^{45,46} These might be common and rare structural genetic alterations including copy
374 number variants, rare non-coding variants, or other intermediate and low-impact risk
375 variants not included routinely in PRS models, and non-genetic contributors such as
376 environmental / lifestyle factors.

377 Only few PRS studies considered the FH. In line with our results, Jenkins et al. found no
378 correlation between SNP-based and FH-based risks and an improved risk stratification
379 when both PRS and FH are considered.⁴⁵ In the analyses by Jia et al., the AUC derived
380 from PRS (0.609) was substantially higher compared to the one derived using FH
381 (0.523). Adding PRS and FH of cancer in first-degree relatives improved the models
382 discriminatory performance (AUC 0.613).^{16,47} Our AUC calculations point in the same
383 direction with a higher AUC (0.704) when all three risk factors (PRS, FH, carrier status)
384 are considered.

385 Interestingly and in apparent contrast to our results and those of others, a study using
386 826 European-descent carriers of PV in the DNA MMR genes *MLH1*, *MSH2*, *MSH6*,
387 *PMS2*, and *EPCAM* (i.e. LS carriers) from the Colon Cancer Family Registry (CCFR)
388 did not find evidence of an association between the PRS and CRC risk, irrespective of
389 sex or mutated gene, although an almost identical set of SNPs was used for PRS calcu-
390 lations.⁴⁸ A reason which might partly explain different risk estimates between studies
391 using individuals from a population-based repository such as the UKBB and those using
392 curated clinical data registries, where patients / families with suspected hereditary dis-
393 ease are included (e.g. the CCFR), is a potentially different risk composition across co-
394 horts recruited in different ways (recruitment bias). That way, a familial clustering of
395 CRC might reflect the existence of several genetic and non-genetic risk factors as out-
396 lined above, which are not captured by the PRS and which may superimpose the poly-
397 genic impact.

398 In particular, the composition of cases and controls is different between the Jenkins et
399 al.⁴⁸ study on the one hand and the Fahed et al.³⁰ and present study on the other hand.
400 In the Jenkins et al. study, obviously both cases (i.e., PV carriers with CRC) and con-

401 trols (healthy PV carriers) derived from the same LS families, while the UKBB controls
402 are PV carriers not apparently related to the PV cases. This is also reflected by the dif-
403 ferent ratio between cases and controls (7.5% CRC cases among PV carriers in the
404 present study, but 61% in the Jenkins et al. study). Hence, the controls in the Jenkins et
405 al. study are relatives of the cases and thus, it is likely that they share parts of the poly-
406 genic background and other risk factors of their affected relatives (cases) to a certain
407 extent which may explain the observed missing effect of the PRS. The comparison be-
408 tween population-based and registry-based predictions indicates that the study design
409 and recruitment strategy may strongly influence the results and conclusions. Conse-
410 quently, the application of PRS in clinical practice should consider the familial back-
411 ground and ascertainment of the patient.

412 Our data analyses provide evidence that the PRS acts as a relevant risk modifier for
413 CRC among both the general population and population-based PV carriers in genes
414 causing hereditary CRC. The findings of us and others qualify the PRS as important
415 component of risk stratification and resulting risk-adapted surveillance strategies in
416 terms of age of onset and frequency. Given the risk distribution across PRS groups, the
417 PRS can define a considerable proportion of the general population at a CRC risk level
418 which is considered sufficient for a more or a less intensive surveillance. Importantly,
419 the non-carriers with high PRS are a much larger target group compared to PV carriers
420 and thus might generate an even higher preventive effect from a healthcare perspec-
421 tive. A small group of non-carriers with positive FH and high PRS even has CRC risks
422 almost in the same order of magnitude as LS carriers without additional risk factors and
423 thus may need similar intensive surveillance measures.

424 According to these findings, there should be a potential benefit for both the general
425 population and at-risk individuals carrying PV, from the inclusion of PRS in healthcare
426 prevention policies, as risk-stratified surveillance improves early disease detection and
427 prevention. A recent study demonstrated that individuals with a higher genetic risk
428 benefited more substantially from preventive measures than those with a lower risk:
429 CRC screening was associated with a significantly reduced CRC incidence and more
430 than 30% reduced mortality among individuals with a high PRS.^{49,50} Preliminary calcula-
431 tions indicate that polygenic-risk-stratified CRC screening could become cost-effective

432 under certain conditions including an AUC value above 0.65 which was reached in our
433 analyses.⁵¹

434 Based on the striking different penetrance between individual hereditary CRC genes,
435 very recent guidelines start to recommend a more gene-specific surveillance intensity in
436 LS and polyposis.^{52,53} Given the strong modifying effect, the inclusion of additional risk
437 factors will result in a more appropriate, clinically relevant risk stratification. Our results
438 demonstrate that a combined risk assessment including FH and PRS will likely improve
439 precise risk estimations and tailored preventive measures not only in the general popu-
440 lation, but also in patients with hereditary disease.

441 Our study has some limitations. Firstly, there is evidence of a “healthy volunteers” selec-
442 tion bias of the UKBB population (UKBB participants tend to be healthier than the gen-
443 eral population), and thus the results might not be completely generalizable in terms of
444 effect sizes.⁵⁴ Secondly, we cannot exclude that few carriers of APC PV who were clas-
445 sified as controls, are affected by a polyposis but have not been recognized as such or
446 did not develop CRC due to intensive surveillance and / or prophylactic surgery, so that
447 the calculated CRC risk of APC PV might be slightly underestimated. Thirdly, our risk
448 assessment was based solely on genetic variants and FH and did not include other risk
449 factors. Previous studies on UKBB showed that lifestyle modifiable risk factors play a
450 pivotal role in cancer prevalence, and a shared lifestyle within families could influence
451 FH with the disease.^{47,55} That might explain the partly independent association of the FH
452 and the genetic risk. Finally, although we performed the analysis on the whole UKBB
453 cohort, we could not test the risk stratification generalizability across different popula-
454 tions due to the limited sample size. PRS could be biased towards the European popu-
455 lation as PRS was constructed based on European reference GWAS. Thus, these PRS
456 might be a worse predictor in non-European or admixed individuals, as previously dis-
457 cussed in different studies.⁵⁶

458 In conclusion, we show the important role of PRS and FH on CRC risk in both the gen-
459 eral population and population-based carriers of a monogenic predisposition for CRC.
460 The combined effect of common variants can strongly alter the age-related penetrance
461 and life time risk of CRC. Thus, the PRS represents an additional, independent stratifi-

462 cation level to cancer risk besides the FH and likely increase the accuracy of risk esti-
463 mation. Consequently, PRS can define a relevant proportion within the general popula-
464 tion as a risk group, which should be considered as subjects for more intense surveil-
465 lance measures, and in addition point to a striking risk variability even among carriers of
466 hereditary CRC, which requires more personalized, risk-adapted surveillance strategies.
467 As expected, the modifying effect of the PRS seems to be relevant in particular for
468 moderate penetrant CRC risk genes. When important factors such as polygenic back-
469 ground, FH, and non-genetic modifiers are included in risk assessment, the dichoto-
470 mous risk division between sporadic and hereditary CRC will be partly replaced by a
471 more continuous risk distribution.

472

473 **Acknowledgements**

474 UK Biobank analyses were conducted via application 52446 using a protocol approved by the
475 Partners HealthCare Institutional Review Board. CM and EH are supported by the BONFOR-
476 program of the Medical Faculty, University of Bonn (O-147.0002). This study was supported (not
477 financially) by the European Reference Network on Genetic Tumour Risk Syndromes (ERN
478 GENTURIS) – Project ID No 739547. ERN GENTURIS is partly co-funded by the European
479 Union within the framework of the Third Health Programme “ERN-2016 – Framework Partner-
480 ship Agreement 2017–2021”. DRB and PM are supported by the FNR INTER IN-
481 TER/DFG/21/16394868. This research was also supported by the Instituto de Salud Carlos III
482 and co-funded by European Social Fund—ESF investing in your future—(grants CM19/00099
483 and PID2019-111254RB-I00) and from the European Union’s Horizon 2020 research and inno-
484 vation program under the EJP RD COFUND-EJP N° 825575.

485 **References**

- 486 1. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in
487 the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Fin-
488 land. *N Engl J Med* 2000;343:78–85.
- 489 2. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer
490 among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer*
491 2002;99:260–266.
- 492 3. Win AK, Jenkins MA, Dowty JG, et al. Prevalence and Penetrance of Major Genes
493 and Polygenes for Colorectal Cancer. *Cancer Epidemiol Biomarkers Prev* 2017;26:404–
494 412.
- 495 4. Biller LH, Syngal S, Yurgelun MB. Recent advances in Lynch syndrome. *Fam Cancer*
496 2019;18:211–219.
- 497 5. Grzymalski JJ, Elhanan G, Morales Rosado JA, et al. Population genetic screening
498 efficiently identifies carriers of autosomal dominant diseases. *Nat Med* 2020;26:1235–
499 1239.
- 500 6. Talseth-Palmer BA. The genetic basis of colonic adenomatous polyposis syndromes.
501 *Hered Cancer Clin Pract* 2017;15:5.

- 502 7. Kanth P, Grimmert J, Champine M, et al. Hereditary Colorectal Polyposis and Cancer
503 Syndromes: A Primer on Diagnosis and Management. *Am J Gastroenterol*
504 2017;112:1509–1525.
- 505 8. Vogt S, Jones N, Christian D, et al. Expanded extracolonic tumor spectrum in
506 MUTYH-associated polyposis. *Gastroenterology* 2009;137:1976–1985.
- 507 9. Win AK, Dowty JG, Cleary SP, et al. Risk of colorectal cancer for carriers of muta-
508 tions in MUTYH, with and without a family history of cancer. *Gastroenterology*
509 2014;146:1208–1211.
- 510 10. Stoffel EM, Koeppe E, Everett J, et al. Germline Genetic Features of Young Individ-
511 uals With Colorectal Cancer. *Gastroenterology* 2018;154:897–905.
- 512 11. Chubb D, Broderick P, Dobbins SE, et al. Rare disruptive mutations and their contri-
513 bution to the heritable risk of colorectal cancer. *Nat Commun* 2016;7:11883.
- 514 12. Schubert SA, Morreau H, Miranda NFCC de, et al. The missing heritability of familial
515 colorectal cancer. *Mutagenesis* 2020;35:221–231.
- 516 13. Yurgelun MB, Kulke MH, Fuchs CS, et al. Cancer Susceptibility Gene Mutations in
517 Individuals With Colorectal Cancer. *J Clin Oncol* 2017;35:1086–1095.
- 518 14. Brenner H, Hoffmeister M, Haug U. Family history and age at initiation of colorectal
519 cancer screening. *Am J Gastroenterol* 2008;103:2326–2331.
- 520 15. Butterworth AS, Higgins JP, Pharoah P. Relative and absolute risk of colorectal
521 cancer for individuals with a family history: a meta-analysis. *Eur J Cancer* 2006;42:216–
522 227.
- 523 16. McGeoch L, Saunders CL, Griffin SJ, et al. Risk Prediction Models for Colorectal
524 Cancer Incorporating Common Genetic Variants: A Systematic Review. *Cancer*
525 *Epidemiol Biomarkers Prev* 2019;28:1580–1593.
- 526 17. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk
527 variants for colorectal cancer. *Nat Genet* 2019;51:76–87.
- 528 18. Schmit SL, Edlund CK, Schumacher FR, et al. Novel Common Genetic Susceptibil-
529 ity Loci for Colorectal Cancer. *J Natl Cancer Inst* 2019;111:146–157.
- 530 19. Lu Y, Kweon SS, Tanikawa C, et al. Large-Scale Genome-Wide Association Study
531 of East Asians Identifies Loci Associated With Risk for Colorectal Cancer. *Gastroenter-*
532 *ology* 2019;156:1455–1466.

- 533 20. Law PJ, Timofeeva M, Fernandez-Rozadilla C, et al. Association analyses identify
534 31 new risk loci for colorectal cancer susceptibility. *Nat Commun* 2019;10:2154.
- 535 21. Thomas M, Sakoda LC, Hoffmeister M, et al. Genome-wide Modeling of Polygenic
536 Risk Score in Colorectal Cancer Risk. *Am J Hum Genet* 2020;107:432–444.
- 537 22. Hsu L, Jeon J, Brenner H, et al. A model to determine colorectal cancer risk using
538 common genetic susceptibility loci. *Gastroenterology* 2015;148:1330–1339.
- 539 23. Frampton MJ, Law P, Litchfield K, et al. Implications of polygenic risk for personal-
540 ised colorectal cancer screening. *Ann Oncol* 2016;27:429–434.
- 541 24. Mur P, Bonifaci N, Díez-Villanueva A, et al. Non-Lynch Familial and Early-Onset
542 Colorectal Cancer Explained by Accumulation of Low-Risk Genetic Variants. *Cancers*
543 (Basel) 2021;13.
- 544 25. Frampton M, Houlston RS. Modeling the prevention of colorectal cancer from the
545 combined impact of host and behavioral risk factors. *Genet Med* 2017;19:314–321.
- 546 26. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for com-
547 mon diseases identify individuals with risk equivalent to monogenic mutations. *Nat*
548 *Genet* 2018;50:1219–1224.
- 549 27. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep
550 phenotyping and genomic data. *Nature* 2018;562:203–209.
- 551 28. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and
552 wANNOVAR. *Nature Protocols* 2015;10:1556–1566.
- 553 29. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum
554 quantified from variation in 141,456 humans. *Nature* 2020;581:434–443.
- 555 30. Fahed AC, Wang M, Homburger JR, et al. Polygenic background modifies pene-
556 trance of monogenic variants for tier 1 genomic conditions. *Nat Commun* 2020;11:3635.
- 557 31. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among
558 sequence variation and human phenotype. *Nucleic Acids Res* 2014;42:D980–985.
- 559 32. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the chal-
560 lenge of larger and richer datasets. *Gigascience* 2015;4:7.
- 561 33. Khera AV, Chaffin M, Zekavat SM, et al. Whole-Genome Sequencing to Character-
562 ize Monogenic and Polygenic Contributions in Patients Hospitalized With Early-Onset
563 Myocardial Infarction. *Circulation* 2019;139:1593–1602.

- 564 34. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic
565 risk scores. *Nat Rev Genet* 2018;19:581–590.
- 566 35. Jenkins MA, Makalic E, Dowty JG, et al. Quantifying the utility of single nucleotide
567 polymorphisms to guide colorectal cancer screening. *Future Oncol* 2016;12:503–513.
- 568 36. Jia G, Lu Y, Wen W, et al. Evaluating the Utility of Polygenic Risk Scores in Identify-
569 ing High-Risk Individuals for Eight Common Cancers. *JNCI Cancer Spectr*
570 2020;4:pkaa021.
- 571 37. Archambault AN, Su YR, Jeon J, et al. Cumulative Burden of Colorectal Cancer-
572 Associated Genetic Variants Is More Strongly Associated With Early-Onset vs Late-
573 Onset Cancer. *Gastroenterology* 2020;158:1274–1286.
- 574 38. Fuchs CS, Giovannucci EL, Colditz GA, et al. A prospective study of family history
575 and the risk of colorectal cancer. *N Engl J Med* 1994;331:1669–1674.
- 576 39. Rex DK, Boland CR, Dominitz JA, et al. Colorectal Cancer Screening: Recommen-
577 dations for Physicians and Patients from the U.S. Multi-Society Task Force on Colorec-
578 tal Cancer. *Am J Gastroenterol* 2017;112:1016–1030.
- 579 40. Wijnen JT, Brohet RM, Eijk R van, et al. Chromosome 8q23.3 and 11q23.1 variants
580 modify colorectal cancer risk in Lynch syndrome. *Gastroenterology* 2009;136:131–137.
- 581 41. Talseth-Palmer BA, Wijnen JT, Brenne IS, et al. Combined analysis of three Lynch
582 syndrome cohorts confirms the modifying effects of 8q23.3 and 11q23.1 in MLH1 muta-
583 tion carriers. *Int J Cancer* 2013;132:1556–1564.
- 584 42. Hassanin E, May P, Aldisi R, et al. Breast and prostate cancer risk: The interplay of
585 polygenic risk, rare pathogenic germline variants, and family history. *Genetics in Medi-*
586 *cine* 2021. Available at:
587 www.sciencedirect.com/science/article/pii/S1098360021053879.
- 588 43. Mars N, Widén E, Kerminen S, et al. The role of polygenic risk and susceptibility
589 genes in breast cancer over the course of life. *Nat Commun* 2020;11:6383.
- 590 44. Gao C. Genome engineering for crop improvement and future agriculture. *Cell*
591 2021;184:1621–1635.
- 592 45. Jenkins MA, Win AK, Dowty JG, et al. Ability of known susceptibility SNPs to predict
593 colorectal cancer risk for persons with and without a family history. *Fam Cancer*
594 2019;18:389–397.

- 595 46. Biller LH, Horiguchi M, Uno H, et al. Familial Burden and Other Clinical Factors As-
596 sociated With Various Types of Cancer in Individuals With Lynch Syndrome. *Gastroen-*
597 *terology* 2021;161:143–150.
- 598 47. Kachuri L, Graff RE, Smith-Byrne K, et al. Pan-cancer analysis demonstrates that
599 integrating polygenic risk scores with modifiable risk factors improves risk prediction.
600 *Nat Commun* 2020;11:6084.
- 601 48. Jenkins MA, Buchanan DD, Lai J, et al. Assessment of a Polygenic Risk Score for
602 Colorectal Cancer to Predict Risk of Lynch Syndrome Colorectal Cancer. *JNCI Cancer*
603 *Spectr* 2021;5:pkab022.
- 604 49. Choi J, Jia G, Wen W, et al. Effects of Screenings in Reducing Colorectal Cancer
605 Incidence and Mortality Differ by Polygenic Risk Scores. *Clin Transl Gastroenterol*
606 2021;12:e00344.
- 607 50. Stanesby O, Jenkins M. Comparison of the efficiency of colorectal cancer screening
608 programs based on age and genetic risk for reduction of colorectal cancer mortality. *Eur*
609 *J Hum Genet* 2017;25:832–838.
- 610 51. Naber SK, Kundu S, Kuntz KM, et al. Cost-Effectiveness of Risk-Stratified Colorec-
611 tal Cancer Screening Based on Polygenic Risk: Current Status and Future Potential.
612 *JNCI Cancer Spectr* 2020;4:pkz086.
- 613 52. Monahan KJ, Bradshaw N, Dolwani S, et al. Guidelines for the management of he-
614 reditary colorectal cancer from the British Society of Gastroenterology
615 (BSG)/Association of Coloproctology of Great Britain and Ireland (ACPGBI)/United
616 Kingdom Cancer Genetics Group (UKCGG). *Gut* 2020;69:411–444.
- 617 53. Seppälä TT, Dominguez-Valentin M, Sampson JR, et al. Prospective observational
618 data informs understanding and future management of Lynch syndrome: insights from
619 the Prospective Lynch Syndrome Database (PLSD). *Fam Cancer* 2021;20:35–39.
- 620 54. Tyrrell J, Zheng J, Beaumont R, et al. Genetic predictors of participation in optional
621 components of UK Biobank. *Nat Commun* 2021;12:886.
- 622 55. Saunders CL, Kilian B, Thompson DJ, et al. External Validation of Risk Prediction
623 Models Incorporating Common Genetic Variants for Incident Colorectal Cancer Using
624 UK Biobank. *Cancer Prev Res (Phila)* 2020;13:509–520.

625 56. Martin AR, Kanai M, Kamatani Y, et al. Clinical use of current polygenic risk scores
626 may exacerbate health disparities. *Nat Genet* 2019;51:584–591.

627

628 **Figures Legends**

629 **Figure 1: Colorectal cancer (CRC) odds ratio (OR) and cumulative incidence**
630 **(CI) among individuals stratified for presence of pathogenic variant (PV) carri-**
631 **er status and family history.**

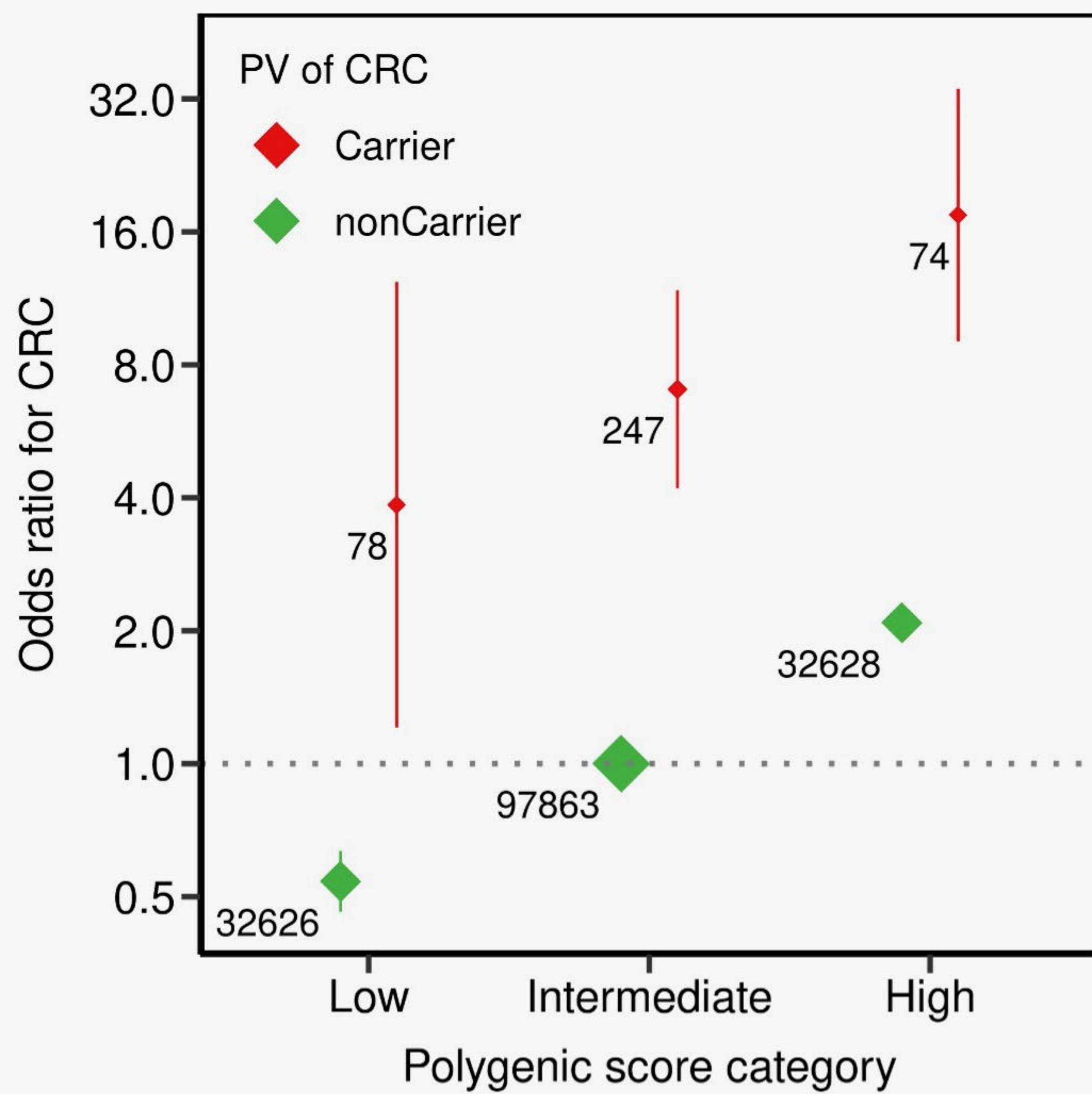
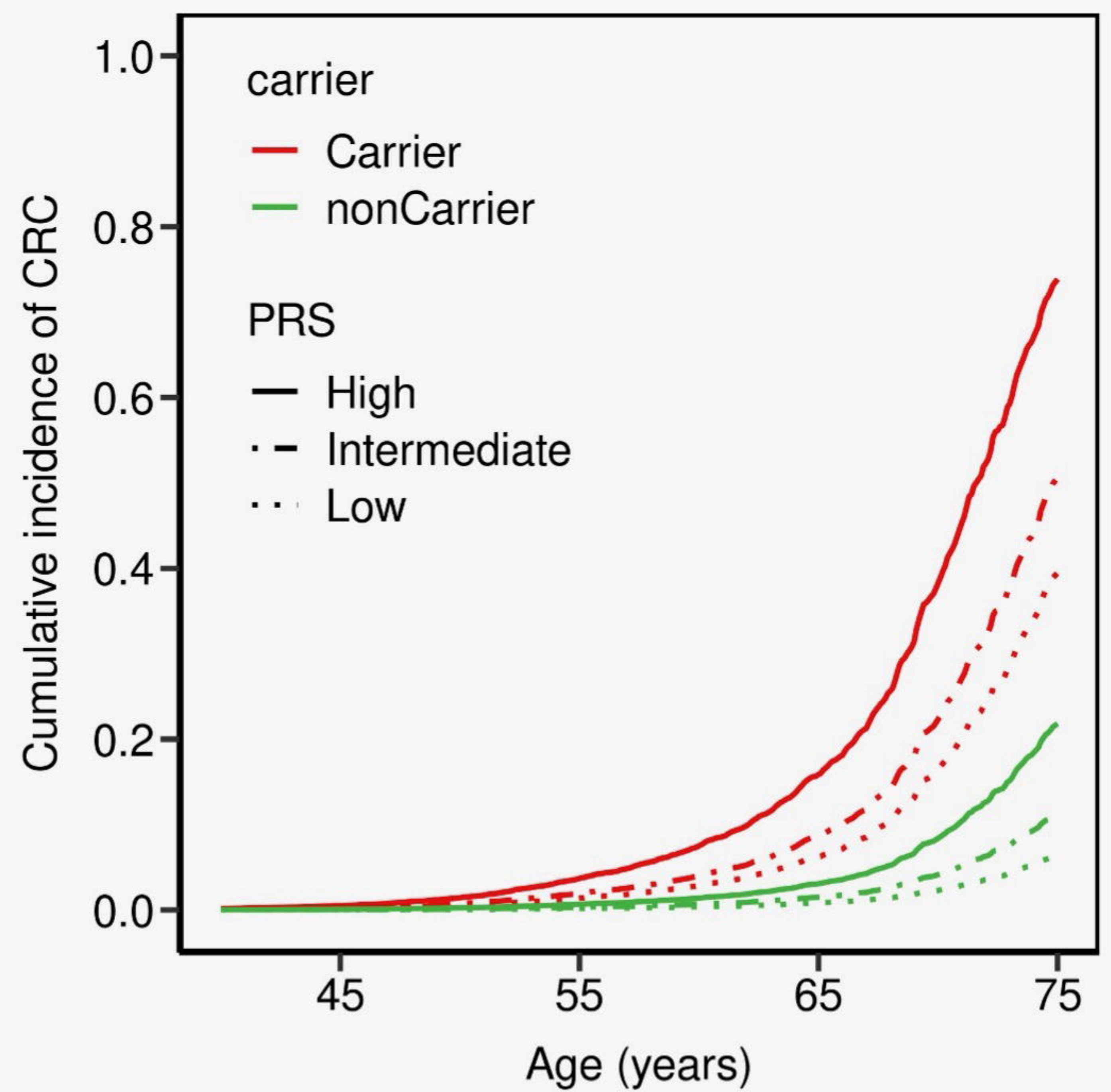
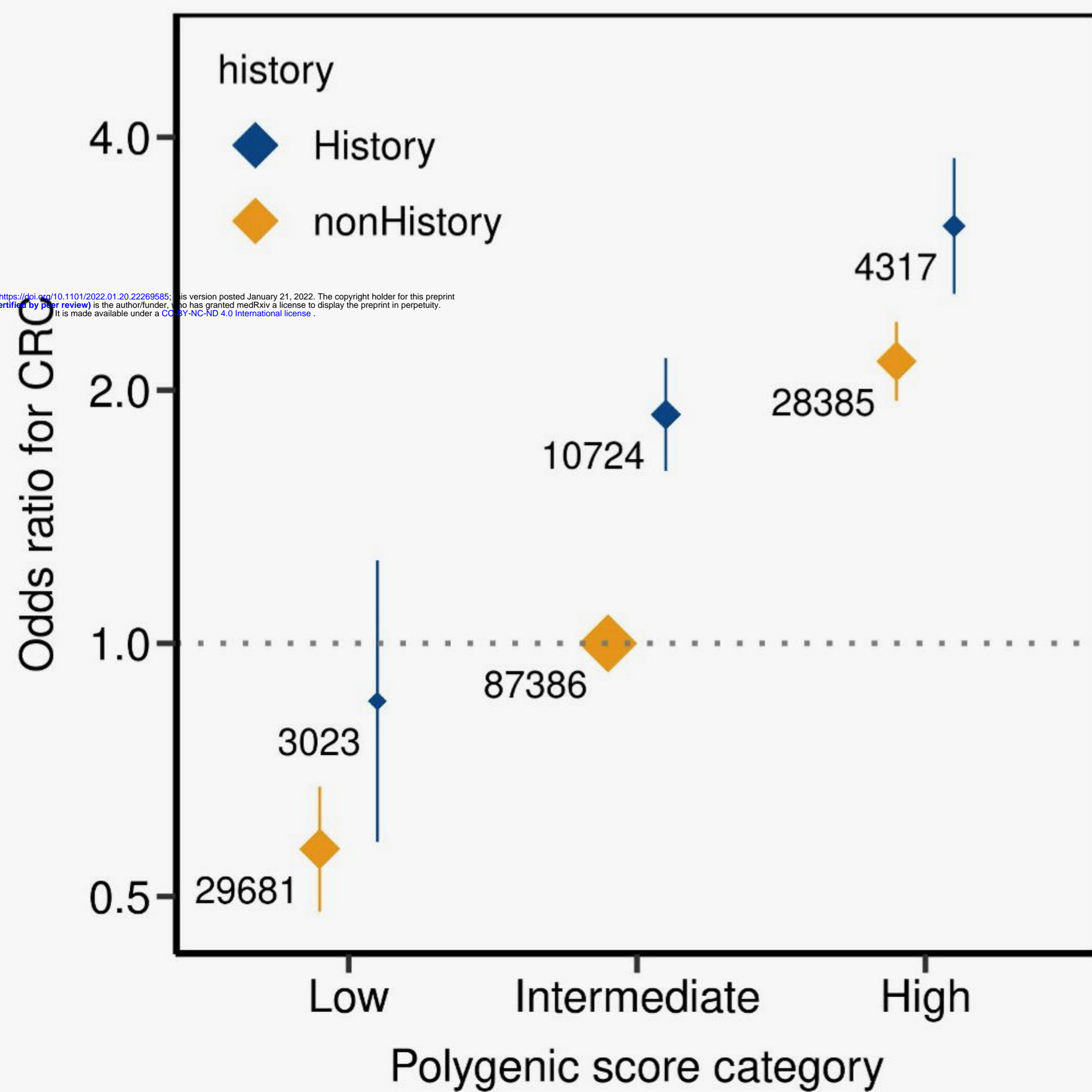
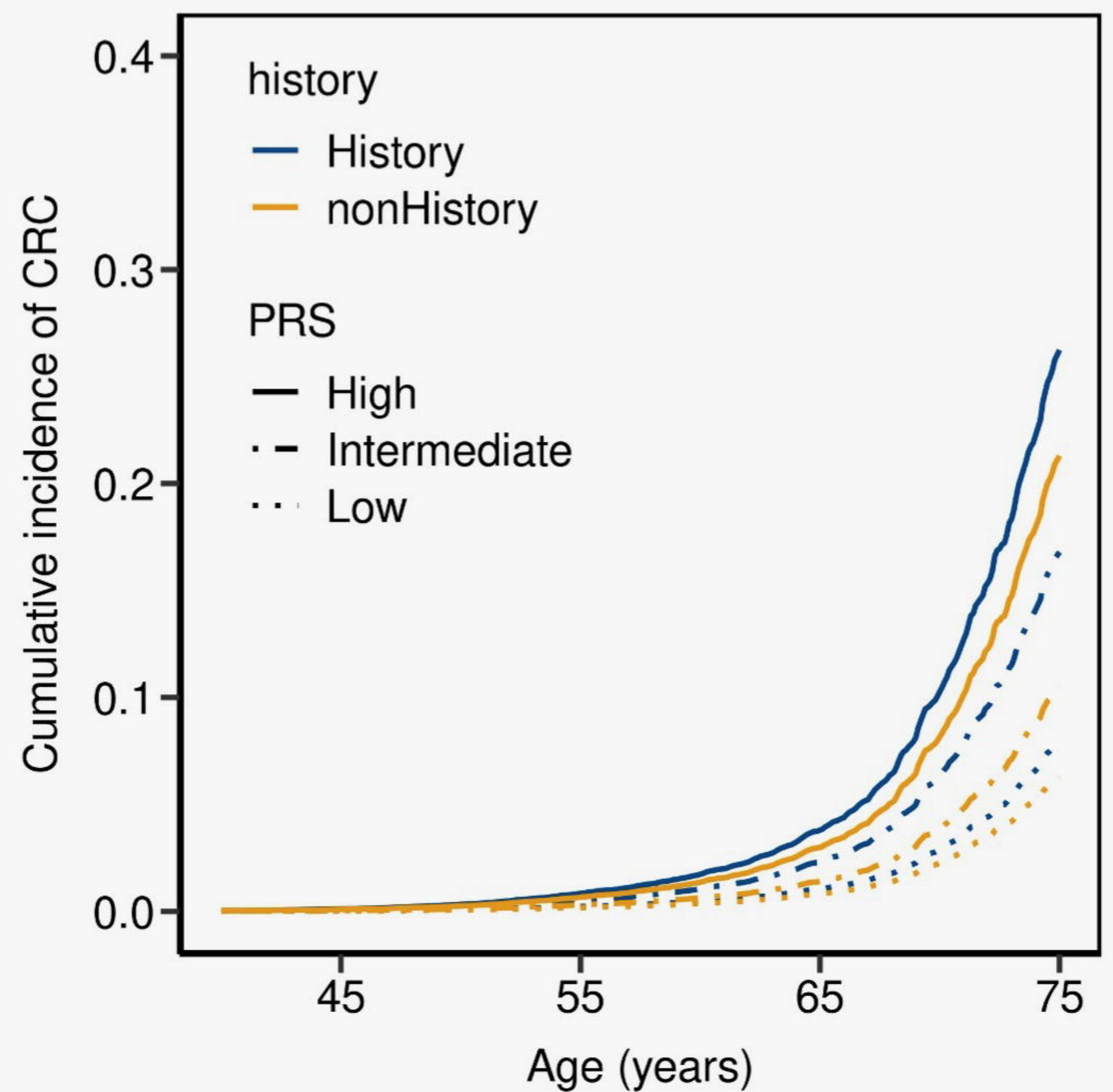
632 Individuals stratified for PV carrier status (A+B), and family history (first-degree rela-
633 tive with CRC) (C+D) into three strata based on their polygenic risk score (PRS):
634 Low (<20 % percentile), intermediate (20-80 % percentile), or high (>80 % percen-
635 tile) PRS. The OR was calculated from a logistic regression model with age, sex
636 and the first four principal components of ancestry as covariates. The reference
637 group was non-carriers with intermediate PRS (A), and no family history with inter-
638 mediate PRS (C). The adjusted OR is indicated by the colored boxes. The numbers
639 next to the ORs indicate the sample size of the corresponding group. The 95% con-
640 fidence intervals are indicated by the vertical lines around the boxes. Cumulative
641 incidence was estimated from a cox-proportional hazard model using age, sex and
642 the first four ancestry principal components as covariates.

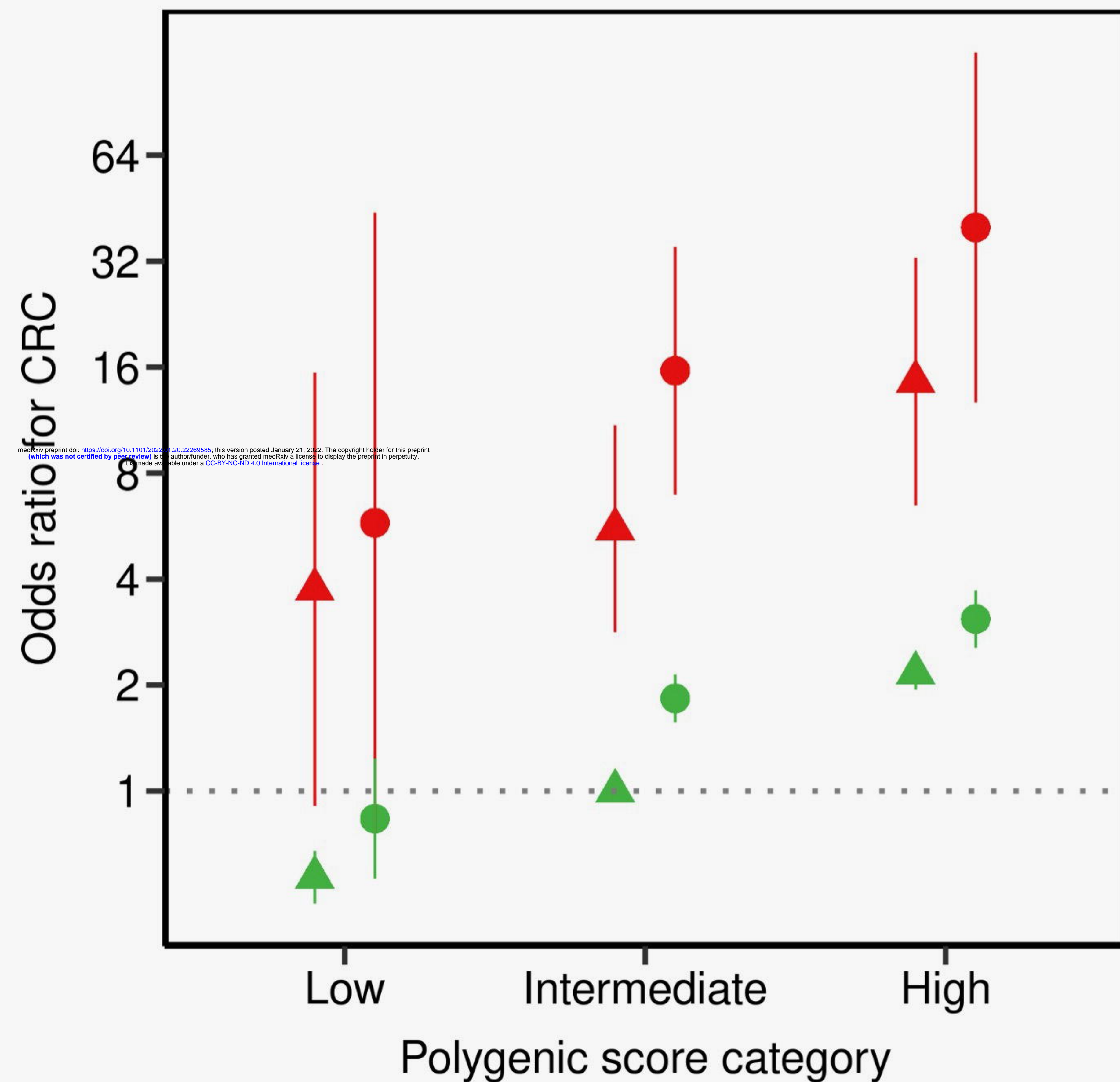
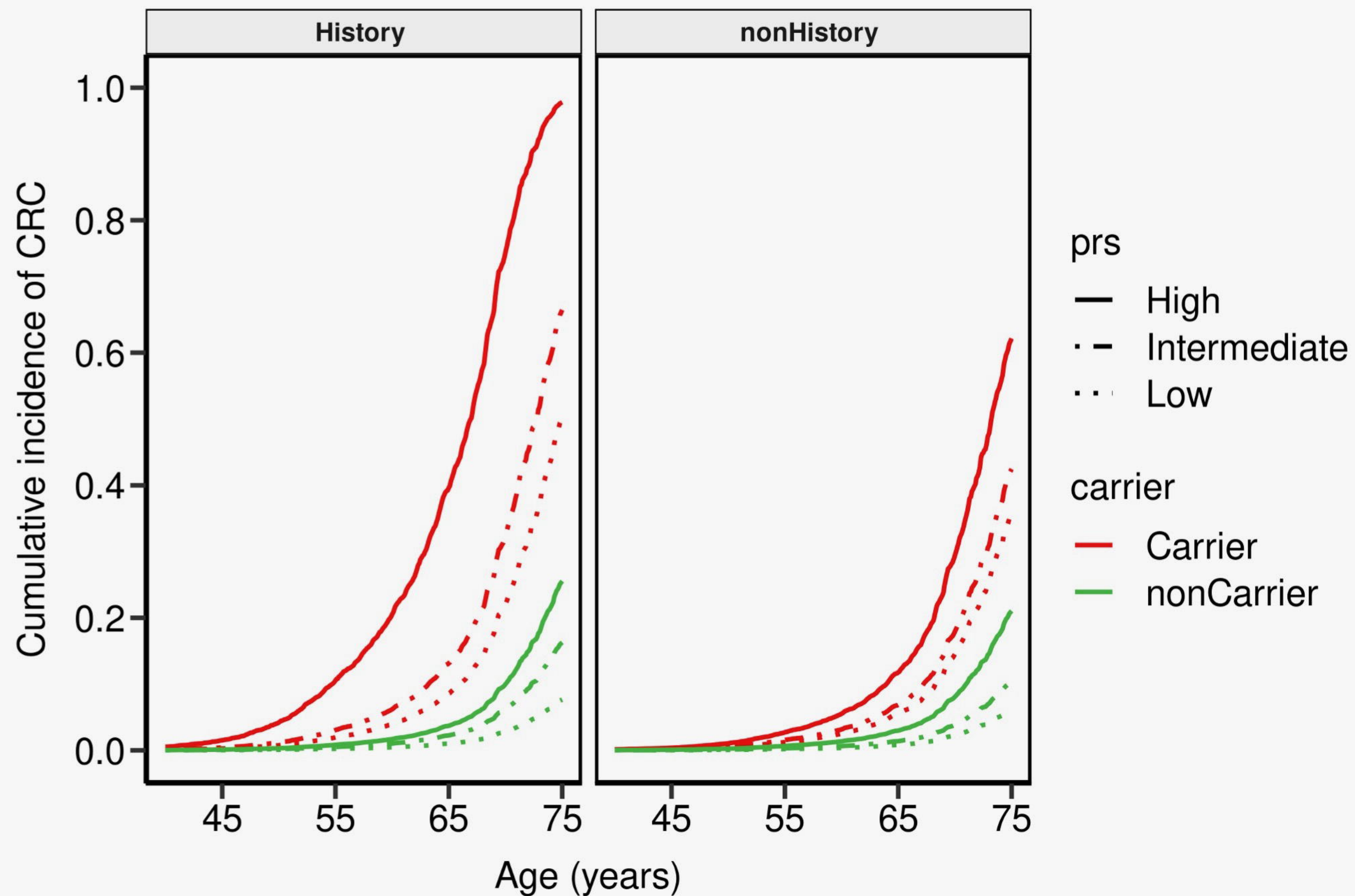
643 **Figure 2: Interplay of pathogenic variant (PV) carrier status, family history**
644 **(FH), and polygenic risk score (PRS).**

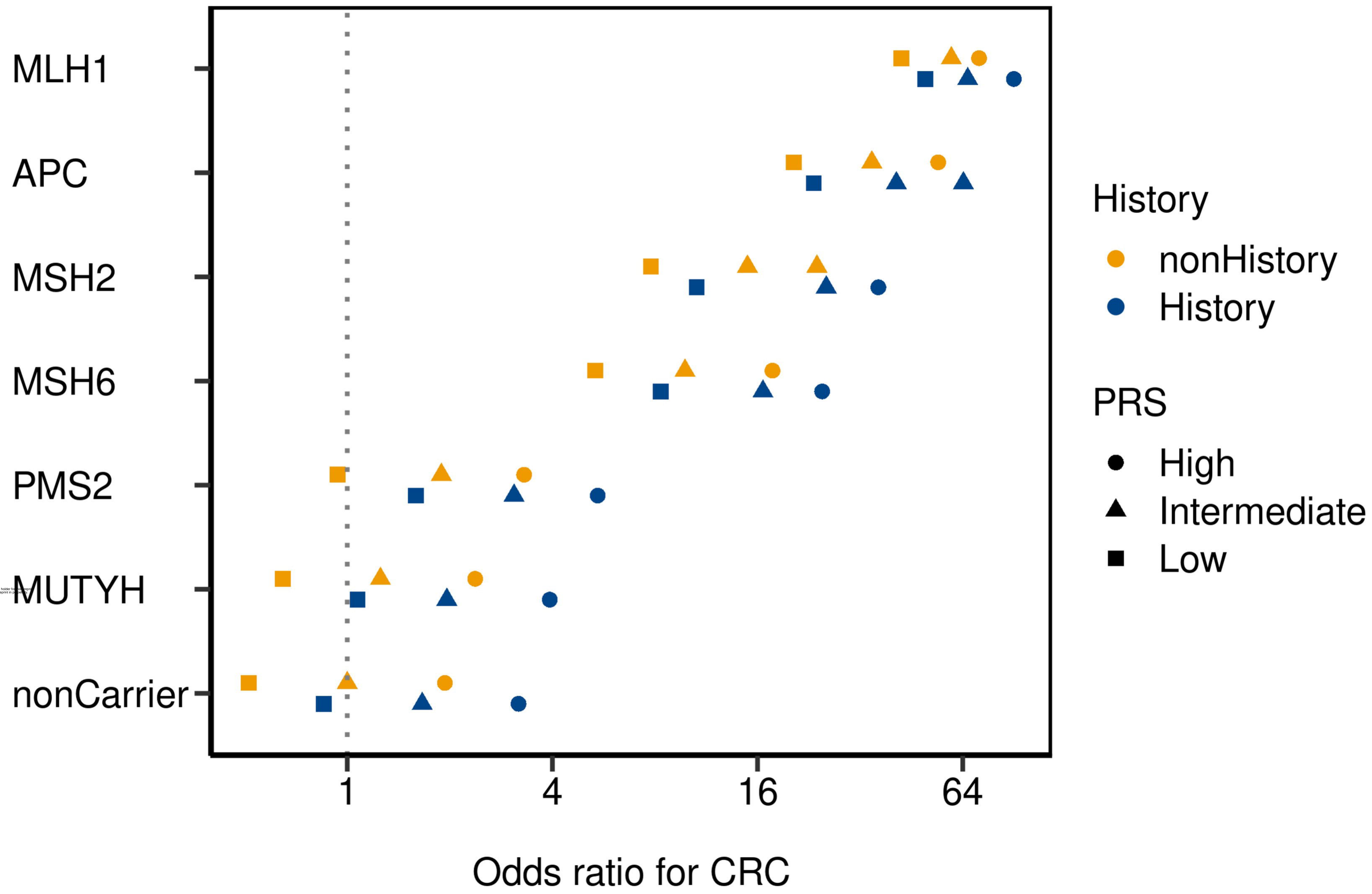
645 (A) Colorectal cancer (CRC) odds ratios (ORs) were estimated from logistic models
646 adjusted for age, sex and first four ancestry principal components. Non-carriers with
647 intermediate PRS and no family history served as the reference group. (B) Cumula-
648 tive incidence was estimated from a cox-proportional hazard model using age, sex
649 and the first four ancestry principal components as covariates.

650 **Figure 3: Interplay of heterozygous pathogenic variant (PV) carrier status,**
651 **family history (FH), and polygenic risk score (PRS) in single genes.**

652 Odds ratios (ORs) for colorectal cancer (CRC) were estimated from logistic models
653 adjusted for age, sex and first four ancestry principal components. Non-carriers with
654 intermediate PRS and no family history served as the reference group.

A**B****C****D**

A**B**



medRxiv preprint doi: <https://doi.org/10.1101/2022.01.20.22289695>; this version posted January 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Table 1: Characteristics of the 163,516 UK Biobank participants by colorectal cancer (CRC) status.

	Cases	Controls
Participants, n	1902	161614
Male, n (%)	1017 (53.47)	73979 (45.78)
Female, n (%)	885 (46.53)	87635 (54.22)
Age, mean (SD)	60.96 (8.56)	56.91 (8.51)
Carriers, n (%)	30 (1.58)	369 (0.23)
Family history of colorectal cancer, n (%)	368 (19.35)	17696 (10.95)

Table 2: Characteristics of the UK Biobank participants by carrier status and polygenic risk score (PRS) strata.

Carrier Status	Carrier			nonCarrier		
	High	Intermediate	Low	High	Intermediate	Low
PRS strata						
Participants,	74	247	78	32628	97863	32626
Cases, n (%)	11 (14.86)	16 (6.48)	3 (3.85)	686 (2.1)	1004 (1.03)	182 (0.56)
Controls, n (%)	63 (85.14)	231 (93.52)	75 (96.15)	31942 (97.9)	96859 (98.97)	32444 (99.44)
Male, n (%)	36 (48.65)	110 (44.53)	35 (44.87)	14824 (45.43)	45115 (46.1)	14876 (45.6)
Female, n (%)	38 (51.35)	137 (55.47)	43 (55.13)	17804 (54.57)	52748 (53.9)	17750 (54.4)
Age at assessment, mean (SD)	57.12 (8.89)	56.16 (9.15)	57.35 (8.42)	56.93 (8.52)	56.97 (8.51)	56.96 (8.53)
Family history of colorectal cancer, n (%)	17 (22.97)	53 (21.46)	18 (23.08)	4300 (13.18)	10671 (10.9)	3005 (9.21)

Table 3: Model discrimination assessed for combinations of polygenic risk score (PRS), family history of CRC (FH) and carrier

	AUC (C.I 95%)	C-index	Nagelkerke's Pseudo-R²
PRS + FH + carrier	0.704 (0.68-0.73)	0.657	0.055
PRS + FH	0.698 (0.67-0.72)	0.652	0.052
PRS + carrier	0.693 (0.66-0.71)	0.646	0.051
PRS	0.688 (0.66-0.71)	0.640	0.049
FH + carrier	0.660 (0.64-0.69)	0.580	0.032
FH	0.654 (0.63-0.68)	0.574	0.031
Carrier	0.646 (0.62-0.67)	0.556	0.030

Note: All models are adjusted for sex, age and first four ancestry principal components.