

# 1 **Circulating white blood cell traits and colorectal cancer** 2 **risk: A Mendelian randomization study**

3 Andrei-Emil Constantinescu<sup>1,2,3\*</sup>, Caroline J Bull<sup>1,2,3</sup>, Nicholas Jones<sup>4</sup>, Ruth Mitchell<sup>1,2</sup>,  
4 Kimberley Burrows<sup>1,2</sup>, Niki Dimou<sup>5</sup>, Stéphane Bézieau<sup>6</sup>, Hermann Brenner<sup>7,8,9</sup>, Daniel D  
5 Buchanan<sup>10,110,12</sup>, Mauro D'Amato<sup>13,14,15</sup>, Mark A Jenkins<sup>16</sup>, Victor Moreno<sup>17,18,19,20</sup>, Rish K  
6 Pai<sup>21</sup>, Caroline Y Um<sup>22</sup>, Emily White<sup>4,24</sup>, Neil Murphy<sup>5</sup>, Marc Gunter<sup>5</sup>, Nicholas J Timpson<sup>1,2</sup>  
7 Jeroen R Huyghe<sup>22§</sup> and Emma E Vincent<sup>1,2,3§</sup>

8

9 <sup>1</sup>MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

10 <sup>2</sup>Bristol Medical School, Population Health Sciences, University of Bristol, Bristol, UK

11 <sup>3</sup>School of Translational Health Sciences, Bristol Medical School, University of Bristol,  
12 Bristol, United Kingdom

13 <sup>4</sup>Institute of Life Science, Swansea University Medical School, Swansea, United Kingdom

14 <sup>5</sup>Nutrition and Metabolism Branch, International Agency for Research on Cancer, World  
15 Health Organization, Lyon, France

16 <sup>6</sup>Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) Nantes, Nantes,  
17 France

18 <sup>7</sup>Division of Clinical Epidemiology and Aging Research, German Cancer Research Center  
19 (DKFZ), Heidelberg, Germany

20 <sup>8</sup>Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National  
21 Center for Tumor Diseases (NCT), Heidelberg, Germany

22 <sup>9</sup>German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ),  
23 Heidelberg, Germany

24 <sup>10</sup>Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of  
25 Melbourne, Parkville, Victoria 3010 Australia

26 <sup>11</sup>University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer  
27 Centre, Parkville, Victoria 3010 Australia

28 <sup>12</sup>Genetic Medicine and Family Cancer Clinic, The Royal Melbourne Hospital, Parkville,  
29 Victoria, Australia

30 <sup>13</sup>Department of Medicine and Surgery, LUM University, Casamassima, Italy

31 <sup>14</sup>Gastrointestinal Genetics Lab, CIC bioGUNE - BRTA, Derio, Spain

32 <sup>15</sup>Ikerbasque, Basque Foundation for Science, Bilbao, Spain

33 <sup>16</sup>Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global  
34 Health, The University of Melbourne, Melbourne, Victoria, Australia

35 <sup>17</sup>Oncology Data Analytics Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet  
36 de Llobregat, Barcelona, Spain

37 <sup>18</sup>CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

38 <sup>19</sup>Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona,  
39 Spain

40 <sup>20</sup>ONCOBEL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de  
41 Llobregat, Barcelona, Spain

42 <sup>21</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic Arizona, Scottsdale,  
43 Arizona, USA

44 <sup>22</sup>Department of Population Science, American Cancer Society, Atlanta, Georgia, USA

45 <sup>23</sup>Department of Epidemiology, University of Washington School of Public Health, Seattle,  
46 Washington, USA

47 <sup>24</sup>Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington,  
48 USA

49 <sup>§</sup>These authors contributed equally.

50 \*Corresponding author:

51 Andrei-Emil Constantinescu, Oakfield House, Oakfield Grove, Bristol, United Kingdom,

52 BS8 2BN, +44 (0) 117 928 7274, [andrei.constantinescu@bristol.ac.uk](mailto:andrei.constantinescu@bristol.ac.uk)

53

54 **Abstract**

55 Observational studies have suggested a protective role for eosinophils in colorectal cancer  
56 (CRC) development and implicated neutrophils, but the causal relationships remain unclear.  
57 Here, we aimed to estimate the causal effect of circulating white blood cell (WBC) counts (N  
58 = ~550,000) for basophils, eosinophils, monocytes, lymphocytes and neutrophils on CRC risk  
59 (N = 52,775 cases and 45,940 controls) using Mendelian randomization (MR). For  
60 comparison, we also examined this relationship using individual-level data from UK Biobank  
61 (4,043 incident CRC cases and 332,773 controls) in a longitudinal cohort analysis. The  
62 inverse-variance weighted (IVW) MR analysis suggested a protective effect of increased  
63 basophil count and eosinophil count on CRC risk [OR per 1-SD increase: 0.88, CI(95%):  
64 0.78-0.99,  $P=0.04$ ; OR: 0.93, CI(95%): 0.88-0.98,  $P=0.01$ ]. The protective effect of  
65 eosinophils remained [OR per 1-SD increase: 0.88, CI(95%): 0.80-0.97,  $P=0.01$ ] following  
66 adjustments for all other WBC subtypes, to account for genetic correlation between the traits,  
67 using multivariable MR. A protective effect of increased lymphocyte count on CRC risk was  
68 also found [OR: 0.84, CI(95%): 0.76-0.93,  $P=6.70e-4$ ] following adjustment. Consistent with  
69 MR results, a protective effect for eosinophils in the cohort analysis in the fully adjusted  
70 model [RR per 1-SD increase: 0.96, CI(95%): 0.93-0.99,  $P=0.02$ ] and following adjustment  
71 for the other WBC subtypes [RR: 0.96, CI(95%): 0.93-0.99,  $P=0.001$ ] was observed. Our  
72 study implicates peripheral blood immune cells, in particular eosinophils and lymphocytes, in  
73 CRC development, highlighting a need for mechanistic studies to interrogate these  
74 relationships.

75

76 **What is already known of this topic** – While previous observational studies have suggested  
77 a protective role for eosinophils in colorectal cancer development and implicated neutrophils,

78 whether changes in the levels of circulating white blood cells causes colorectal cancer has not  
79 been explored.

80 **What this study adds** – Our study is the first to use Mendelian randomization (MR) to  
81 investigate this relationship. In parallel, for comparison, we also conduct the largest cohort  
82 study to date on the topic. We found evidence to suggest that elevated eosinophil and  
83 lymphocyte count may have a protective effect on CRC risk, adding new insights into the  
84 pathogenesis of CRC.

85 **How this study might affect research, practice or policy** – Our findings will encourage  
86 further mechanistic exploration to understand the biological mechanisms underpinning our  
87 findings, which may lead to new therapeutic approaches or risk reduction strategies.

88

89

## 90 **Keywords**

91 Mendelian randomization, UK Biobank, Colorectal cancer, GWAS, White blood cell count,  
92 Eosinophils

93

## 94 **Introduction**

95 Colorectal cancer (CRC) accounts for over 10% of all worldwide cancer cases and is the  
96 second leading cause of cancer-related deaths globally [1–3]. Overall, the number of CRC  
97 cases is rising [4] and is alarmingly on the rise in younger people (aged <50 years) [5–7].  
98 Given current challenges, and the estimation that 50% of CRC cases may be preventable [8],

99 a focus on identifying novel risk factors, and subsequent prophylactic and treatment options  
100 is warranted to limit the future healthcare burden.

101

102 White blood cells (WBCs) are commonly measured in routine blood tests and are divided into  
103 five subtypes: basophils, eosinophils, lymphocytes, monocytes and neutrophils [9].  
104 Alterations to circulating WBC counts, have been found to play a role in disease risk,  
105 severity, and progression, CRC development and mortality [10–15]. For example, higher  
106 circulating counts of basophils and eosinophils, cells with a role in IgE-mediated immunity,  
107 have been associated with reduced CRC risk and increased survival [16–20]. Higher absolute  
108 lymphocyte count [such as T, B and natural killer (NK) cells] has also been associated with  
109 better overall survival of CRC [21,22]. By contrast, a high absolute monocyte count was  
110 found to be associated with worse CRC survival [21,22], consistent with their potential role  
111 in tumour progression and metastasis [23]. Finally, neutrophils, a critical part of the innate  
112 immune system [24], have also been associated with poor CRC overall survival [21,22].

113

114 Traditional (i.e. cross-sectional, case-control or cohort) studies account for the majority of  
115 epidemiologic analyses undertaken on CRC, which can suffer from certain limitations, such  
116 as confounding and reverse causation which can bias effect estimates [25,26]. Mendelian  
117 randomization (MR) is a method in genetic epidemiology which could overcome these issues  
118 by using single-nucleotide polymorphisms (SNPs) to proxy for an exposure of interest to  
119 estimate the effect of an exposure on an outcome [27,28]. It operates akin to a randomized  
120 control trial (RCT), as alleles are randomly assigned at birth [29]. We have used MR to  
121 estimate the total effect of each WBC trait on CRC [30], and, to account for correlation

122 between WBC traits, multivariable (MV) MR to estimate the direct effect of these exposures  
123 on CRC by adjusting for their shared genetic architecture [30].

124

125 In this study we aimed to investigate the relationship between circulating WBC subtypes and  
126 CRC incidence. Firstly, an MR analysis was undertaken using the most comprehensive  
127 genome-wide association study (GWAS) for both WBC subtype counts and incident CRC  
128 available. Secondly, for comparison, we ran the largest longitudinal cohort study between  
129 WBCs and CRC to date using UK Biobank. Together these analyses allow us to compare  
130 between genetically proxied and observational estimates to thoroughly explore the  
131 relationship between circulating WBC counts and CRC development.

132

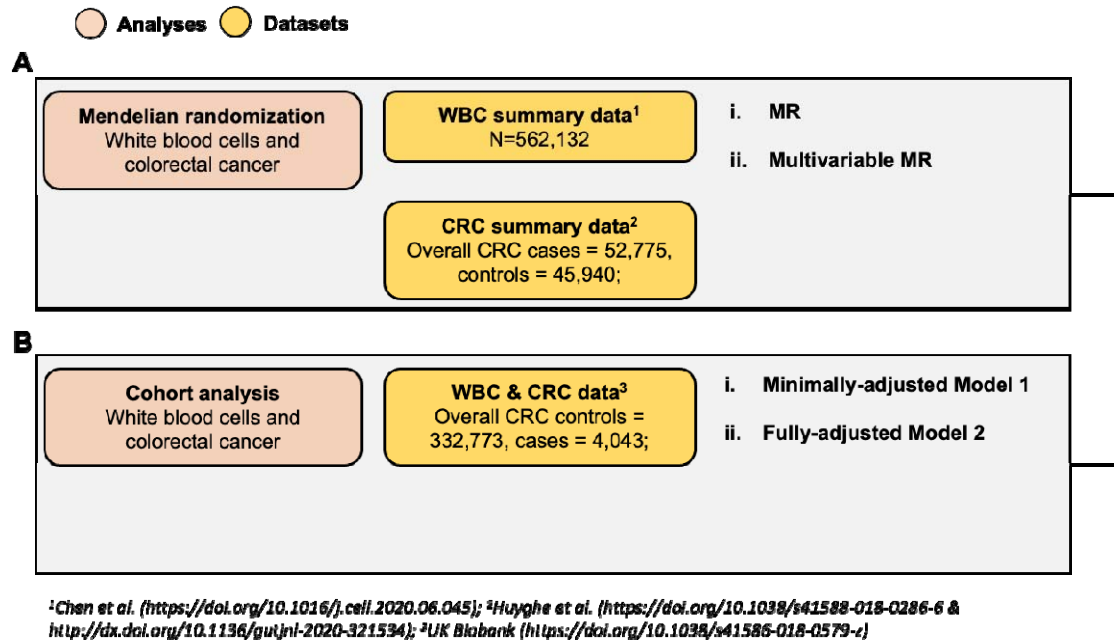
## 133 **Methods**

### 134 **Study design**

135 We aimed to investigate the relationship between circulating WBCs and CRC using a genetic  
136 epidemiologic and complementary observational approach. First, MR analyses were  
137 undertaken to estimate the effect of WBC subtype counts on CRC. We then performed a  
138 multivariable MR (MVMR) analysis where the direct effect of each WBC subtype count was  
139 estimated by adding all five WBC subtypes into the model (**Figure 1A**). STROBE-MR  
140 guidelines were followed (**STROBE-MR Supplement**) [31]. Here, units were interpreted as  
141 odds ratio (OR) for CRC per a normalized standard deviation (1-SD) increase in WBC count.  
142 Next, a prospective longitudinal cohort study was undertaken using UK Biobank individual-  
143 level data. Here, subtype specific WBC counts were studied individually, and then each was  
144 adjusted for each of the other traits (**Figure 1B**). STROBE guidelines were followed

145 (STROBE Supplement). Here, units were interpreted as risk ratio (RR) for CRC per a  
146 normalized standard deviation (1-SD) increase in WBC count.

147



148

149 **Figure 1. Study design.** We triangulated findings from two study designs: a Mendelian  
150 randomization analysis(A) and a longitudinal cohort analysis (B) to estimate the casual effect  
151 of WBC on CRC risk.

152

### 153 WBC count GWAS data

154 Summary statistics for WBC subtype counts were obtained from the “Blood Cell  
155 Consortium” (BCX) meta-analysis, the largest study being UKBB (N=~562,243) [32].  
156 Genetic sex, age, age<sup>2</sup>, study-specific covariates and PCs 1 to 10 were used as covariates. A  
157 brief description of each study included in the meta-analysis is available in **Supplementary**  
158 **Table 1**. Only variants which did not display heterogenous effect across studies were  
159 selected. Specific details on QC steps and association testing are available in the source

160 manuscripts. Summary statistics for WBC counts were downloaded from: [http://www.mhi-](http://www.mhi-humangenetics.org/en/resources/)  
161 [humangenetics.org/en/resources/](http://www.mhi-humangenetics.org/en/resources/).

162

### 163 **CRC GWAS data**

164 GWAS summary statistics for overall CRC risk and by anatomical subsite were taken from  
165 the most comprehensive meta-analyses to date [33]: colon cancer (split into proximal and  
166 distal colon cancer) and rectal cancer. Sixty-six studies, which include those of the Genetics  
167 and Epidemiology of Colorectal Cancer Consortium (GECCO), Colorectal Cancer  
168 Transdisciplinary Study (CORECT), and Colon Cancer Family Registry (CCFR) consortia  
169 were meta-analysed [34–36]. The final sample was predominantly of European ancestry with  
170 5.36% of East Asian ancestry. These were included due to their similar genetic architecture  
171 with regards to CRC risk [37]. Genetic sex, age, study-specific variables and PCs were used  
172 as covariates. An overview of all consortia included in the CRC meta-analysis is available in  
173 **Supplementary Table 2**, and a breakdown of the sample sizes for each set of CRC summary  
174 statistics is presented in **Supplementary Table 3**.

175

### 176 **Genetic data processing**

177 To select valid MR instruments, summary statistics for exposures were processed using the  
178 “TwoSampleMR” R package [38,39]. The presence of correlated instruments, i.e. those in  
179 linkage disequilibrium (LD), can bias MR estimates [40]. Therefore, the exposure SNPs were  
180 clumped ( $r^2=0.001$ , window=10Mb, P-value threshold= $5e-8$ ) using the 1000 Genomes  
181 European dataset [41] as a reference panel. Following this step, the exposure and outcome  
182 datasets were “harmonised” i.e. had their effect alleles placed on the same reference strand  
183 [42]. SNPs with incorrect but unambiguous strand references were corrected, while those



184 with ambiguous strand references were removed. Similarly, palindromic SNPs with inferable  
185 strands were kept.

186

### 187 **MR analysis of WBC counts on CRC risk**

188 Primary MR analyses were undertaken using the inverse-variance weighted (IVW) method,  
189 which is the fixed-effects meta-analysis of the estimated effect of all exposure SNPs on the  
190 CRC outcome [43]. Conditional F-statistics were calculated to detect weak instrument bias  
191 [44] for each exposure SNP as previously described [45]. Several sensitivity analyses were  
192 undertaken to compare with the main IVW estimates. The presence of vertical pleiotropy i.e.  
193 when a trait is downstream of the genetic variant but on the same biological pathway as the  
194 exposure [46], was measured using Cochran's Q heterogeneity test [47]. Horizontal  
195 pleiotropy, when some or all instruments for a trait acts through a different pathway to the  
196 exposure [46], can violate one of the main MR assumptions. A number of sensitivity MR  
197 analyses were undertaken to identify horizontal pleiotropy: MR-Egger (where the regression  
198 intercept is not constrained to zero) [48], Weighted median (the median of all SNP ratio  
199 estimates, where each ratio is weighted by the inverse of the variance) [49], Weighted mode  
200 (assumes that the most frequent estimate in a set of instruments is zero) [50] and MR-  
201 PRESSO (detects individual SNPs contributing to horizontal pleiotropy) [51]. The direction  
202 of the causal relationship between WBC traits and CRC risk was tested using the MR Steiger  
203 method, which uses Steiger's test to test the difference between the Pearson correlations of  
204 genetic variants with both the exposure and outcome [39].

205

### 206 **Multivariable MR analysis of WBC counts on CRC risk**

207 The IVW method was used for the MVMR analysis. First, a pair-wise analysis between all  
208 five WBC subtype counts was undertaken, where the proportion of variance explained (PVE)

209 for SNPs used to instrument a WBC trait was estimated in the other four WBC subtypes  
210 using previously described methodology [45]. The direct effect of each WBC subtype was  
211 estimated by adding in all five WBC subtypes into the MVMR model. Bias arising from  
212 weak instruments was also determined here. This was undertaken using methodology  
213 described by Sanderson et al., where a generalized version of Cochran's Q was employed to  
214 evaluate instrument strength [52]. Standard Cochran's Q statistic [47] was calculated to  
215 detect the presence of heterogeneity. For those traits with an F-statistic <10, a follow-up  
216 MVMR analysis was done accounting for the presence of weak instruments.

217

### 218 **UK Biobank phenotypic data**

219 Between the years of 2007 and 2010, UKBB participants visited assessment centres (N=22)  
220 throughout the UK [53,54]. The individuals had their health records linked, their genomes  
221 sequenced, and underwent multiple evaluations, such as self-report questionnaires and  
222 medical examinations [53,54]. The latter includes the analysis of blood cell samples using  
223 Beckman Coulter LH750 instruments designed for high throughput screening [55]. Total  
224 WBC count and WBC subtype percentage (%) were measured, with absolute WBC subtype  
225 count derived as "WBC subtype % / 100 x total WBC" and expressed as  $10^9$  cells/Litre [55].  
226 The blood sampling date variable was split into year, month, day, and minutes (passed since  
227 the start of the day of the appointment visit). Additional variables were gathered including  
228 recruitment centre, sampling device ID, age, genetic sex, principal components 1 to 10, BMI,  
229 Townsend deprivation index, smoking and alcohol drinker status (self-report questionnaire –  
230 UKBB codes 20116 and 20117). CRC cases were identified through hospital inpatient  
231 records coded to the 10<sup>th</sup> version of the International Classification of Disease (ICD-10).

232

### 233 **Filtering and selection criteria**

234 The UKBB dataset underwent a series of steps prior to further analyses. Withdrawn  
235 participants and those of non-European ancestry were excluded. Viable controls and incident  
236 CRC cases were defined using methodology previously described by Burrows et al. [56]  
237 (**Supplementary Table 4**). Here, we defined incident CRC cases as those diagnosed at least  
238 one year after blood sampling. Participants with no WBC measurement data or sampling date  
239 were removed, as were those who were known to be pregnant, have chronic conditions (e.g.  
240 HIV, blood cancers, thalassaemia), or undergoing erythropoietin treatment, as in Astle et al.  
241 [57] and Chen et al. [32], given the effects of these traits on WBC measurements. Those with  
242 acute conditions (e.g. upper respiratory infections) diagnosed less than 3 months prior to  
243 blood sampling were excluded. Finally, missing values in “Townsend Deprivation Index”,  
244 “Body mass index”, “Smoking status” and “Alcohol drinker status” variables were removed.

245

### 246 **Cohort study between WBC count and CRC**

247 We conducted a cohort analysis between circulating WBCs and incident CRC. WBC count  
248 values were log-transformed, after which they were adjusted for the following covariates:  
249 assessment centre, sex, age, age<sup>2</sup>, PCs 1 to 10, as in the Chen et al. GWAS [32]. The resulting  
250 residuals were rank-inverse normal transformed and then used in a logistic regression on  
251 CRC incidence. This main observational analysis was termed “Model 1”, a minimally  
252 adjusted model. An additional fully adjusted analysis was undertaken, termed “Model 2”,  
253 where BMI, Townsend DI, smoker status and alcohol drinker status were added as additional  
254 covariates. Following this, another pair of analyses was run, where all five WBC subtype  
255 counts were added together into the model to reduce potential bias due to their correlated  
256 values. Analyses where each WBC trait was studied individually were termed as  
257 “univariable”, while those where they were added together were termed as “multivariable”.

258

## 259 **Working environment**

260 All analyses were performed with R version 4.1.2 (Bird Hippie) in a Linux environment  
261 supported by the University of Bristol's Advanced Computing Research Centre (ACRC).  
262 Genetic data preparation, as well as the MR analyses, were undertaken with the  
263 "TwoSampleMR" R package [38,39]. The MVMR analyses were undertaken with the  
264 "MVMR" R package [58]. Scripts associated with this study are available on GitHub:  
265 <https://github.com/andrewcon/wbc-crc>.

266

## 267 **Results**

268

### 269 **Effect of WBC count on CRC**

270 Before performing the MR analysis, the average F-statistic for each WBC trait was estimated  
271 to detect the presence of weak instrument bias, which is generally indicated by an average F-  
272 statistic  $< 10$  [44]. For overall CRC, these were 64.48 (basophil count, 171 SNPs), 124.72  
273 (eosinophil count, 396 SNPs), 105.85 (lymphocyte count, 444 SNPs), 147.44 (monocyte  
274 count, 477 SNPs), 98.84 (neutrophil count, 387 SNPs), indicating strong MR instruments  
275 (**Supplementary Table 5**).

276

277 The IVW method showed evidence of a protective effect of basophil count (1-SD increase in  
278 WBC count OR: 0.88, 95% CI: 0.78-0.99, P-value: 0.037) for CRC risk (**Figure 2**,  
279 **Supplementary Figure 1**). For eosinophil count, the IVW method showed evidence of a  
280 protective effect for CRC risk (OR: 0.93, 95% CI: 0.88-0.98, P-value: 0.012) (**Figure 2**,  
281 **Supplementary Figure 1**). These results were confirmed by the MR-Egger, weighted  
282 median and MR-PRESSO sensitivity methods (**Supplementary Table 6**). The IVW method

283 also showed a protective effect of eosinophil count for CRC risk (OR: 0.91, 95% CI: 0.85-  
284 0.99, P-value: 0.021) in females. All results, including those for sex- and site-specific CRC,  
285 are available in **Supplementary Table 6**.

286

287 In site specific CRC analyses, the univariable IVW method demonstrated evidence for a  
288 protective effect of basophil count on colon (OR: 0.85, 95% CI: 0.74-0.98, P-value: 0.022)  
289 and distal colon (OR: 0.82, 95% CI: 0.70-0.97, P-value: 0.019) cancer (**Supplementary**  
290 **Table 6**). For eosinophil count, the main IVW analysis provided evidence for a protective  
291 effect for colon (OR: 0.90, 95% CI: 0.84-0.96, P-value: 0.001), proximal colon (OR: 0.89,  
292 95% CI: 0.82-0.96, P-value: 0.003) and distal colon (OR: 0.89, 95% CI: 0.82-0.97, P-value:  
293 0.007) cancer (**Supplementary Table 6**). These results were mostly supported by sensitivity  
294 analyses. There was a protective effect of total WBC count on colon (OR: 0.91, 95% CI:  
295 0.85-0.99, P-value: 0.02) and proximal colon (OR: 0.90, 95% CI: 0.82-0.98, P-value: 0.015)  
296 cancer (**Supplementary Table 6**).

297

298 Next, the presence of vertical and horizontal pleiotropy in the MR analyses was analysed.  
299 Cochran's heterogeneity test indicated the presence of heterogeneity in all but one WBC trait-  
300 CRC pair (basophil count-male CRC,  $P_{HET}=0.104$ ) (**Supplementary Table 7**). Following  
301 this, the MR-Egger test for horizontal pleiotropy was performed. Here, evidence for this type  
302 of pleiotropy was identified for eosinophil count and colon ( $P_{PLT}=0.018$ ), distal colon  
303 ( $P_{PLT}=0.015$ ), female ( $P_{PLT}=0.049$ ), male ( $P_{PLT}=0.041$ ) and overall ( $P_{PLT}=0.015$ ) CRC risk,  
304 suggesting a possible bias of MR estimates. This was also the case for lymphocyte count in  
305 female CRC ( $P_{PLT}=0.016$ ) (**Supplementary Table 7**). Although the MR-PRESSO method  
306 identified the presence of SNP horizontal-pleiotropic outliers, there was little evidence that

307 the removal of these outliers contributed to a notable shift in the point estimates  
308 (**Supplementary Table 8**).

309

### 310 **Multivariable MR of WBC count on CRC**

311 A pair-wise analysis of the proportion of variance explained (PVE) for SNPs instrumenting  
312 each WBC trait indicated a low PVE for each of the other WBC traits, with the exception of  
313 basophil count (2.44% vs. 2.39% when instrumenting neutrophil count) (**Supplementary**  
314 **Table 9**). The overall results indicated that statistical power should not suffer to a large  
315 degree by adding all five WBC subtype counts into the MVMR analysis. Here, the MVMR  
316 IVW method estimated a protective effect of eosinophil count (OR: 0.88, 95% CI: 0.80-0.97,  
317 P-value: 0.011) and lymphocyte count (OR: 0.84, 95% CI: 0.76-0.93, P-value: 0.0007) on  
318 CRC risk (**Figure 2**). Regarding sex-specific estimates, eosinophil count was protective for  
319 CRC risk (OR: 0.83, 95% CI: 0.73-0.94, P-value: 0.004) in females, this was also the case for  
320 lymphocyte count (OR: 0.76, 95% CI: 0.67-0.87, P-value: 6.46E-05).

321

322 Eosinophil count was estimated to have a direct protective effect on colon (OR: 0.84, 95%  
323 CI: 0.75-0.94, P-value: 0.002), proximal colon (OR: 0.92, 95% CI: 0.85-1.00, P-value: 0.042)  
324 and distal colon (OR: 0.88, 95% CI: 0.74-1.00, P-value: 0.049) cancer (**Supplementary**  
325 **Table 10**). Lymphocyte count was estimated to have a direct protective effect on colon (OR:  
326 0.85, 95% CI: 0.76-0.96, P-value: 0.007), distal colon (OR: 0.77, 95% CI: 0.67-0.88, P-value:  
327 0.0001) and rectal (OR: 0.86, 95% CI: 0.75-0.98, P-value: 0.022) cancer anatomical subsites  
328 (**Supplementary Table 10**).

329

330 Sensitivity analyses were undertaken to assess heterogeneity and the presence of weak  
331 instruments in the MVMR analysis. There was evidence of heterogeneity in all WBC trait-

332 CRC pairs (**Supplementary Table 11**). Conditional F-statistics showed evidence of weak  
 333 instruments ( $F < 10$ ) for basophil count (**Supplementary Table 11**). Based on these results, an  
 334 additional MVMR analysis was run adjusting for weak instruments for basophil count  
 335 (overall CRC  $OR_{Weak}$ : 1.3; male CRC  $OR_{Weak}$ : 1.2; female CRC  $OR_{Weak}$ : 1.7; colon cancer  
 336  $OR_{Weak}$ : 1.5; proximal colon cancer  $OR_{Weak}$ : 1.5; distal colon cancer  $OR_{Weak}$ : 1.5; rectal  
 337 cancer  $OR_{Weak}$ : 1.6) on CRC risk (**Supplementary Table 11**).

338

### 339 Cohort observational analysis between WBC count and CRC

340 336,816 UKBB participants remained after passing filtering and selection criteria  
 341 (Supplementary Figure 2). Of these, 332,773 were controls and 4,043 were incident CRC  
 342 cases. When split by genetic sex, there were 154,629 male and 178,144 female controls and  
 343 2,316 male and 1,727 female cases. Those with CRC were more likely to be male (57% vs.  
 344 46%), had a higher average age (60.7 vs. 55.8 years), slightly higher BMI (28.0 vs. 27.4  
 345  $kg/m^2$ ) and were more likely to have been cigarette smokers in the past (46% vs. 55% never  
 346 smokers and 44% vs. 34% previous smokers) (Table 1).

347

Characteristic	Control, N = 332,773 <sup>1</sup>	Case, N = 4,043 <sup>1</sup>
Sex		
Female	178,144 / 332,773 (54%)	1,727 / 4,043 (43%)
Male	154,629 / 332,773 (46%)	2,316 / 4,043 (57%)
Age (years)	55.8 (8.1)	60.7 (6.6)
Body mass index	27.4 (4.7)	28.0 (4.7)
Smoking status		
Never	183,987 / 332,773 (55%)	1,861 / 4,043 (46%)
Previous	114,349 / 332,773 (34%)	1,790 / 4,043 (44%)
Current	34,437 / 332,773 (10%)	392 / 4,043 (9.7%)
Alcohol drinker status		
Never	10,655 / 332,773 (3.2%)	122 / 4,043 (3.0%)
Previous	10,982 / 332,773 (3.3%)	140 / 4,043 (3.5%)
Current	311,136 / 332,773 (93%)	3,781 / 4,043 (94%)
Townsend deprivation index	-1.4 (3.0)	-1.6 (3.0)

Basophil count	0.0 (0.1)	0.0 (0.1)
Eosinophil count	0.2 (0.1)	0.2 (0.1)
Lymphocyte count	1.9 (0.6)	1.9 (0.6)
Monocyte count	0.5 (0.2)	0.5 (0.2)
Neutrophil count	4.2 (1.4)	4.3 (1.4)
Overall WBC count	6.9 (1.7)	7.0 (1.8)

---

1n / N (%); Mean (SD)

348 **Table 1. Baseline characteristics of UK Biobank study sample.**

349

350 As a percentage of the total WBC count based on the median values, basophils accounted for  
351 0.3%, eosinophils 2.11%, lymphocytes 28.16%, monocytes 6.78%, and neutrophils 60.39%  
352 (**Supplementary Table 12**). A pair-wise correlation matrix between each WBC subtype  
353 showed a high correlation between total WBC and neutrophil count, while for the other  
354 subtypes the correlation coefficients were equal or below 0.3 (**Supplementary Figure 3**).  
355 Batch variables (e.g. blood sample device and sampling date), Townsend DI, and alcohol  
356 drinker status explained some of the variance in WBC count (0% to 0.66%). Depending on  
357 the WBC subtype, genetic sex explained between 0.23% to 2.93% of the variance, BMI  
358 explained between 0.14% to 2.64%, and smoking status explained between 0.44% to 3.85%  
359 (**Supplementary Figure 4**).

360

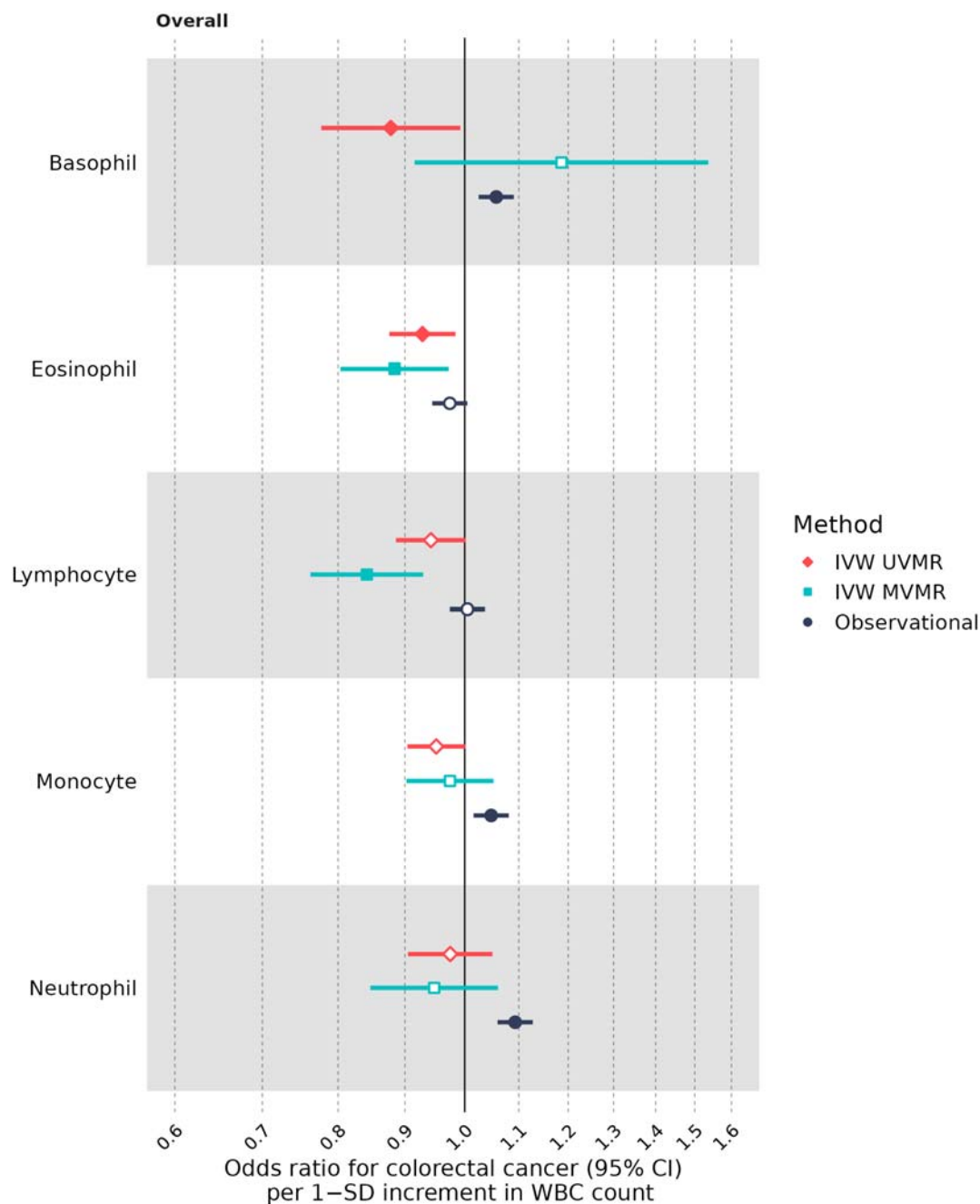
361 In the main analysis (“Model 1” - the minimally adjusted model), basophil count was  
362 positively associated with CRC risk (RR: 1.06, 95% CI: 1.02-1.09, P-value: 0.0005), as was  
363 monocyte (RR: 1.05, 95% CI: 1.02-1.08, P-value: 0.003) and neutrophil count (RR: 1.09,  
364 95% CI: 1.06-1.13, P-value: 1.94E-08) (**Figure 2, Supplementary Table 13**). In “Model 2”,  
365 eosinophil count was negatively associated with CRC risk (overall CRC RR: 0.96, 95% CI:  
366 0.93-0.99, P-value: 0.022) (**Supplementary Table 13**).

367



368 Observational associations were re-computed by adding all five WBC subtype counts  
369 together. In the minimally adjusted “Model 1”, eosinophil count (RR: 0.96, 95% CI: 0.93-  
370 0.99, P-value: 0.009) was associated with lower overall CRC odds, while basophil (RR: 1.04,  
371 95% CI: 1.01-1.08, P-value: 0.008) and neutrophil count (RR: 1.08, 95% CI: 1.05-1.12, P-  
372 value: 1.92E-06) were associated with an increase in overall CRC risk (**Supplementary**  
373 **Table 14**). These results largely coincided with those from the fully adjusted “Model 2”  
374 analyses (**Supplementary Table 14**). All results, including those for sex-specific CRC, are  
375 available in **Supplementary Table 14**.

376



377

378 **Figure 2. The relationship between WBC count and CRC risk based on UV, MV two-**  
379 **sample MR and cohort observational analyses.** Each WBC trait is presented on the X-axis.  
380 The estimated effect is presented on the Y-axis. Point estimates were filled where the P-value  
381 was less than 0.05. Results are interpreted as ORs (95% CI) for CRC risk per 1-SD  
382 normalized increment in WBC count.

383

## 384 **Discussion**

385 In this study we aimed to estimate the effects of five circulating WBC subtypes on CRC risk  
386 by using a combined genetic epidemiologic and longitudinal cohort framework. Through the  
387 aid of MVMR, we were able to assess the independent causal effect of WBC counts by  
388 adjusting for their shared genetic architecture. Taken together, the evidence across analyses  
389 suggests a potential protective effect of increased circulating eosinophil and lymphocyte  
390 count on CRC risk.

391

392 Consistent with our study, Prizment et al found that eosinophil count (tertiles Q3 and Q2 vs.  
393 Q1) was negatively associated with odds of developing colon, but not of rectal cancer [19].  
394 Similar results have been reported for other cancers; Wong et al. reported a negative trend  
395 between increasing eosinophil count quartiles and lung adenocarcinoma odds in a UKBB  
396 study [59], while a similar study looking at prostate cancer showed a negative association  
397 between eosinophil count quintiles Q3-5, as well as a per 1-SD increase in the trait (HR 0.96  
398 vs. OR 0.93 for CRC in our analysis) [60].

399

400 Eosinophils have a well-established role in allergic disease, including asthma and allergic  
401 rhinitis [61]. Indeed, MR analyses have also reported a causal effect of eosinophil count on  
402 allergic disease [57,62] and a recent systematic review investigating the relationship between  
403 allergies and cancer reported evidence for a reduced risk of CRC in those with allergic  
404 diseases [63].

405

406 Here, our results suggest that the immune response through by eosinophils provides  
407 protection against tumour development. Indeed, in several neoplasia, including CRC,

408 eosinophils have been found to play an anti-tumourigenic role and are a source of anti-  
409 tumourigenic molecules, such as eosinophil-derived neurotoxin (EDN) [64,65]. Experimental  
410 studies have also found a tumour-protective effect of IgE [66]. Increased eosinophil  
411 recruitment to the CRC tumour site has also been associated with better survival, even when  
412 adjusting for the effects of CD8<sup>+</sup> T-cells [13], and eosinophil-specific granule secretion of  
413 granzyme A has been linked with the killing of CRC cells [67].

414

415 In addition to eosinophil count, we also found a protective effect of lymphocyte count on  
416 CRC risk. While this was not apparent in the MR analysis, the MVMR estimates indicated  
417 lower ORs for CRC across all anatomical subsites (proximal colon cancer trended towards  
418 protective) with increased lymphocyte counts. The multivariable fully adjusted “Model 2” in  
419 the cohort analysis also indicated that there could be a negative association with CRC risk. It  
420 is not surprising that we found a protective effect of higher circulating levels of lymphocytes  
421 with CRC odds given their established role in combatting tumour development [68]. Tumour-  
422 infiltrating lymphocytes (TILs) like CD8<sup>+</sup> T-cells help antagonise tumour growth through  
423 direct action and recruitment of other immune cells [68]. High levels of TILs were previously  
424 associated with better CRC overall survival and disease-free survival [22,68]. In support of  
425 our findings, two observational studies found higher lymphocyte counts compared to cases  
426 vs. controls a year to six months prior to CRC diagnosis [69,70]. However, these results  
427 could indicate production and recruitment of lymphocytes to the site of pre-cancerous or  
428 undetected tumours rather than a causal effect.

429

### 430 **Limitations**

431 There are several limitations to this study. With regards to the cohort analysis, only baseline  
432 blood measurements were available. This assumes that WBC counts were constant and did

433 not allow us to establish a relationship between a trend in WBC count and its relationship  
434 with CRC odds. Nevertheless, baseline WBC count measurements have previously been  
435 shown to be associated with disease risk [59,60,71,72], making their study in relation to  
436 disease development a worthwhile endeavour. Also, in the cohort analysis, incident CRC  
437 cases were defined as those diagnosed at least one year after blood sampling, in order to not  
438 diminish the number of cases to a large degree. However, as CRC develops over a long  
439 period, our cohort analysis may have not excluded all participants with undiagnosed CRC.

440

441 With regards to the MR analysis, the genetic instruments used here proxied for lifetime  
442 variation of WBC count. Therefore, the MR analysis cannot be used to infer how large  
443 changes over a short timespan might affect CRC development. Regarding the MVMR  
444 method, caution should be applied when investigating traits with very weak instruments, as it  
445 cannot reliably adjust for those traits [58]. This was the case for basophil count, as the F-  
446 statistic was estimated to be between 4.7 and 4.8 (**Supplementary Table 12**). Therefore,  
447 despite pointing to an increased detrimental effect compared to the main MVMR analysis,  
448 ORs derived from the weak-MVMR analysis should be interpreted with this in mind.

449

## 450 **Conclusion**

451 In summary, the results generated here provide evidence for a protective causal effect of  
452 elevated levels of circulating eosinophil and lymphocyte counts on CRC risk. Going forward,  
453 additional research is needed to disentangle the biological mechanisms and pinpoint specific  
454 pathways through which eosinophils and lymphocytes might protect against CRC  
455 development.

456

## 457 **Availability of data and materials**

458 Summary statistics for WBC counts were downloaded from the following website:  
459 <http://www.mhi-humangenetics.org/en/resources/>. The summary-level GWAS data on  
460 outcomes used in this study were made available following an application to the Genetics and  
461 Epidemiology of Colorectal Cancer Consortium (GECCO):  
462 [https://www.fredhutch.org/en/research/divisions/public-health-sciences-  
463 division/research/cancer-prevention/genetics-epidemiology-colorectal-cancer-consortium-  
464 gecco.html](https://www.fredhutch.org/en/research/divisions/public-health-sciences-division/research/cancer-prevention/genetics-epidemiology-colorectal-cancer-consortium-gecco.html).

465

## 466 **Contributions**

467 AC, CJB, JH and EEV conceived the idea for the paper. AC conducted the analysis. All  
468 authors contributed to the interpretation of the findings. AC, CJB and EEV wrote the  
469 manuscript. All authors critically revised the paper for intellectual content and approved the  
470 final version of the manuscript.

471

## 472 **Funding information**

473 AC acknowledges funding from grant MR/N0137941/1 for the GW4 BIOMED MRC DTP,  
474 awarded to the Universities of Bath, Bristol, Cardiff and Exeter from the Medical Research  
475 Council (MRC)/UKRI. NJT is the PI of the Avon Longitudinal Study of Parents and Children  
476 (Medical Research Council & Wellcome Trust 217065/Z/19/Z) and is supported by the  
477 University of Bristol NIHR Biomedical Research Centre (BRC-1215-2001). NJT  
478 acknowledges funding from the Wellcome Trust (202802/Z/16/Z). EEV, CJB, ND and NJT  
479 acknowledge funding by the CRUK Integrative Cancer Epidemiology Programme  
480 (C18281/A29019). NJT, EEV and CJB work in a unit funded by the UK Medical Research  
481 Council (MC\_UU\_00011/1 & MC\_UU\_00011/4) and the University of Bristol. EEV and  
482 CJB are supported by Diabetes UK (17/0005587) and the World Cancer Research Fund  
483 (WCRF UK), as part of the World Cancer Research Fund International grant program  
484 (IIG\_2019\_2009). JRH acknowledges funding by the National Cancer Institute at the U.S.  
485 National Institutes of Health (R21CA230486). This work was also supported by the Elizabeth  
486 Blackwell Institute for Health Research, University of Bristol, and the Wellcome Trust  
487 Institutional Strategic Support Fund (ISSF 204813/Z/16/Z). The funders of the study had no  
488 role in the study design, data collection, data analysis, data interpretation or writing of the  
489 report.

490

## 491 **Ethics declaration**

492 All participants provided written informed consent, and each study was approved by the  
493 relevant research ethics committee or institutional review board.

494

495 **Consent for publication**

496 All authors consented to the publication of this work.

497

498 **Competing interests**

499 The authors declare no competing interests.

500

501 **Supplementary information**

502 Additional file 1

503 **Supplementary Figures.**

504 Additional file 2

505 **Supplementary Tables.**

506 Additional file 3

507 **STROBE-MR checklist.**

508 Additional file 4

509 **STROBE Cohort checklist.**

510 Additional file 5

511 **Full funding and acknowledgements details.**

512

513 **References**

- 514 [1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer  
515 statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36  
516 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.  
517 <https://doi.org/10.3322/CAAC.21492>.
- 518 [2] Sung H, Siegel RL, Rosenberg PS, Jemal A. Emerging cancer trends among young  
519 adults in the USA: analysis of a population-based cancer registry. *Lancet Public Heal*  
520 2019;4:e137–47. [https://doi.org/10.1016/S2468-2667\(18\)30267-](https://doi.org/10.1016/S2468-2667(18)30267-6/ATTACHMENT/EA91DB97-1C30-4C2F-B604-48BF68319153/MMC1.PDF)  
521 [6/ATTACHMENT/EA91DB97-1C30-4C2F-B604-48BF68319153/MMC1.PDF](https://doi.org/10.1016/S2468-2667(18)30267-6/ATTACHMENT/EA91DB97-1C30-4C2F-B604-48BF68319153/MMC1.PDF).
- 522 [3] Mauri G, Sartore-Bianchi A, Russo AG, Marsoni S, Bardelli A, Siena S. Early onset  
523 colorectal cancer in young individuals. *Mol Oncol* 2019;13:109.  
524 <https://doi.org/10.1002/1878-0261.12417>.
- 525 [4] Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl*  
526 *Oncol* 2021;14:101174. <https://doi.org/10.1016/J.TRANON.2021.101174>.
- 527 [5] Chambers AC, Dixon SW, White P, Williams AC, Thomas MG, Messenger DE.  
528 Demographic trends in the incidence of young onset colorectal cancer: a  
529 population-based study. *Br J Surg* 2020;107:595. <https://doi.org/10.1002/BJS.11486>.

- 530 [6] SEER. Surveillance, Epidemiology, and End Results (SEER) Program. 2019.
- 531 [7] Araghi M, Soerjomataram I, Bardot A, Ferlay J, Cabasag CJ, Morrison DS, et al.
- 532 Changes in colorectal cancer incidence in seven high-income countries: a population-
- 533 based study. *Lancet Gastroenterol Hepatol* 2019;4:511–8.
- 534 [https://doi.org/10.1016/S2468-1253\(19\)30147-5](https://doi.org/10.1016/S2468-1253(19)30147-5).
- 535 [8] Hull MA. Nutritional prevention of colorectal cancer. *Proc Nutr Soc* 2021;80:59–64.
- 536 <https://doi.org/10.1017/S0029665120000051>.
- 537 [9] Nicholson LB. The immune system. *Essays Biochem* 2016;60:275–301.
- 538 <https://doi.org/10.1042/EBC20160017>.
- 539 [10] Lee YJ, Lee HR, Nam CM, Hwang UK, Jee SH. White Blood Cell Count and the Risk
- 540 of Colon Cancer. *Yonsei Med J* 2006;47.
- 541 [11] Wu J, Ge X, Zhu W, Zhi Q, Xu M, Duan W, et al. Values of applying white blood cell
- 542 counts in the prognostic evaluation of resectable colorectal cancer. *Mol Med Rep*
- 543 2019;19:2330–40. <https://doi.org/10.3892/MMR.2019.9844>.
- 544 [12] Watt DG, Martin JC, Park JH, Horgan PG, McMillan DC. Neutrophil count is the most
- 545 important prognostic component of the differential white cell count in patients
- 546 undergoing elective surgery for colorectal cancer 2015;210:24–30.
- 547 [13] Prizment AE, Vierkant RA, Smyrk TC, Tillmans LS, Lee JJ, Sriramarao P, et al.
- 548 Tumor eosinophil infiltration and improved survival of colorectal cancer patients: Iowa
- 549 Women’s Health Study. *Mod Pathol* 2016;29:516–27.
- 550 <https://doi.org/10.1038/modpathol.2016.42>.
- 551 [14] Rosman Y, Hornik-Lurie T, Meir-Shafir K, Lachover-Roth I, Cohen-Engler A,
- 552 Munitz A, et al. Changes in peripheral blood eosinophils may predict colorectal cancer
- 553 – A retrospective study. *World Allergy Organ J* 2022;15:100696.
- 554 <https://doi.org/10.1016/J.WAOJOU.2022.100696>.
- 555 [15] Iijima K, Fujibayashi K, Okumura M, Sasabe N, Gunji T. Number of Eosinophils and
- 556 Incidence of Cancer in a Japanese Population□: A Single Institution Study. *Orig Artic*
- 557 *Ningen Dock Int* 2019;6:56–61.
- 558 [16] Liu Q, Luo D, Cai S, Li Q, Li X. Circulating basophil count as a prognostic marker of
- 559 tumor aggressiveness and survival outcomes in colorectal cancer. *Clin Transl Med*
- 560 2020 91 2020;9:1–12. <https://doi.org/10.1186/S40169-019-0255-4>.
- 561 [17] Wei Y, Zhang X, Wang G, Zhou Y, Luo M, Wang S, et al. The impacts of
- 562 pretreatment circulating eosinophils and basophils on prognosis of stage □–□
- 563 colorectal cancer. *Asia Pac J Clin Oncol* 2018;14:e243–51.
- 564 <https://doi.org/10.1111/AJCO.12871>.
- 565 [18] Harbaum L, Pollheimer MJ, Kornprat P, Lindtner RA, Bokemeyer C, Langner C.
- 566 Peritumoral eosinophils predict recurrence in colorectal cancer. *Mod Pathol* 2015 283
- 567 2014;28:403–13. <https://doi.org/10.1038/modpathol.2014.104>.
- 568 [19] Prizment AE, Anderson KE, Visvanathan K, Folsom AR. Inverse association of
- 569 eosinophil count with colorectal cancer incidence: Atherosclerosis Risk in
- 570 Communities study. *Cancer Epidemiol Biomarkers Prev* 2011;20:1861–4.
- 571 [https://doi.org/10.1158/1055-9965.EPI-11-0360/66619/AM/INVERSE-](https://doi.org/10.1158/1055-9965.EPI-11-0360/66619/AM/INVERSE-ASSOCIATION-OF-EOSINOPHIL-COUNT-WITH)
- 572 [ASSOCIATION-OF-EOSINOPHIL-COUNT-WITH](https://doi.org/10.1158/1055-9965.EPI-11-0360/66619/AM/INVERSE-ASSOCIATION-OF-EOSINOPHIL-COUNT-WITH).
- 573 [20] Ma W, Yang J, Li P, Lu X, Cai J. Association between allergic conditions and
- 574 colorectal cancer risk/mortality: a meta-analysis of prospective studies. *Sci Reports*
- 575 2017 71 2017;7:1–7. <https://doi.org/10.1038/s41598-017-04772-9>.
- 576 [21] Yamamoto T, Kawada K, Obama K. Inflammation-Related Biomarkers for the
- 577 Prediction of Prognosis in Colorectal Cancer Patients. *Int J Mol Sci* 2021, Vol 22,
- 578 Page 8002 2021;22:8002. <https://doi.org/10.3390/IJMS22158002>.
- 579 [22] Tanio A, Saito H, Uejima C, Takaya S, Yamamoto M, Tokuyasu N, et al. A prognostic



- 580 index for colorectal cancer based on preoperative absolute lymphocyte, monocyte, and  
581 neutrophil counts. *Surg Today* 2019;49:245–53. <https://doi.org/10.1007/S00595-018-1728-6/TABLES/2>.
- 582
- 583 [23] Olingy CE, Dinh HQ, Hedrick CC. Monocyte heterogeneity and functions in cancer. *J*  
584 *Leukoc Biol* 2019;106:309–22. <https://doi.org/10.1002/JLB.4RI0818-311R>.
- 585 [24] Quail DF, Amulic B, Aziz M, Barnes BJ, Eruslanov E, Fridlender ZG, et al.  
586 Neutrophil phenotypes and functions in cancer: A consensus statement. *J Exp Med*  
587 2022;219:39. <https://doi.org/10.1084/JEM.20220011/213202>.
- 588 [25] Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal  
589 inference in epidemiological studies. *Hum Mol Genet* 2014;23:R89–98.  
590 <https://doi.org/10.1093/hmg/ddu328>.
- 591 [26] Zheng J, Baird D, Borges M-C, Bowden J, Hemani G, Haycock P, et al. Recent  
592 Developments in Mendelian Randomization Studies. *Curr Epidemiol Reports*  
593 2017;4:330–45. <https://doi.org/10.1007/s40471-017-0128-6>.
- 594 [27] Smith GD, Ebrahim S. SEER. Surveillance, Epidemiology, and End Results (SEER)  
595 Program. *Int J Epidemiol* 2003;32:1–22. <https://doi.org/10.1093/ije/dyg070>.
- 596 [28] Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and  
597 limitations. *Int J Epidemiol* 2004;33:30–42. <https://doi.org/10.1093/ije/dyh132>.
- 598 [29] Davies NM, Holmes M V., Davey Smith G. Reading Mendelian randomisation  
599 studies: a guide, glossary, and checklist for clinicians. *BMJ* 2018;362:601.  
600 <https://doi.org/10.1136/bmj.k601>.
- 601 [30] Carter AR, Sanderson E, Hammerton G, Richmond RC, Davey Smith G, Heron J, et  
602 al. Mendelian randomisation for mediation analysis: current methods and challenges  
603 for implementation. *Eur J Epidemiol* 2021 365 2021;36:465–78.  
604 <https://doi.org/10.1007/S10654-021-00757-1>.
- 605 [31] Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA,  
606 Vanderweele TJ, et al. Strengthening the reporting of observational studies in  
607 epidemiology using mendelian randomisation (STROBE-MR): explanation and  
608 elaboration. *BMJ* 2021;375. <https://doi.org/10.1136/BMJ.N2233>.
- 609 [32] Chen MH, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, et al. Trans-  
610 ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5  
611 Global Populations. *Cell* 2020;182:1198-1213.e14.  
612 <https://doi.org/10.1016/j.cell.2020.06.045>.
- 613 [33] Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of  
614 common and rare genetic risk variants for colorectal cancer. *Nat Genet* 2019;51:76.  
615 <https://doi.org/10.1038/S41588-018-0286-6>.
- 616 [34] Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al.  
617 Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide  
618 Meta-analysis. *Gastroenterology* 2013;144.  
619 <https://doi.org/10.1053/j.gastro.2012.12.020>.
- 620 [35] Schumacher FR, Schmit SL, Jiao S, Edlund CK, Wang H, Zhang B, et al. Genome-  
621 wide association study of colorectal cancer identifies six new susceptibility loci. *Nat*  
622 *Commun* 2015 61 2015;6:1–7. <https://doi.org/10.1038/ncomms8138>.
- 623 [36] Schmit SL, Edlund CK, Schumacher FR, Gong J, Harrison TA, Huyghe JR, et al.  
624 Novel Common Genetic Susceptibility Loci for Colorectal Cancer. *JNCI J Natl Cancer*  
625 *Inst* 2019;111:146–57. <https://doi.org/10.1093/JNCI/DJY099>.
- 626 [37] Zhang B, Jia WH, Matsuda K, Kweon SS, Matsuo K, Xiang YB, et al. Large-scale  
627 genetic study in East Asians identifies six new loci associated with colorectal cancer  
628 risk. *Nat Genet* 2014;46:533. <https://doi.org/10.1038/NG.2985>.
- 629 [38] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base

- 630 platform supports systematic causal inference across the human phenome. *Elife*  
631 2018;7.
- 632 [39] Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between  
633 imprecisely measured traits using GWAS summary data. *PLoS Genet* 2017;13.  
634 <https://doi.org/10.1371/JOURNAL.PGEN.1007081>.
- 635 [40] Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but  
636 oft-forgotten) practices: the design, analysis, and interpretation of Mendelian  
637 randomization studies. *Am J Clin Nutr* 2016;103:965–78.  
638 <https://doi.org/10.3945/ajcn.115.118216>.
- 639 [41] Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al.  
640 A global reference for human genetic variation. *Nature* 2015;526:68–74.  
641 <https://doi.org/10.1038/nature15393>.
- 642 [42] Wootton RE, Sallis HM. Let’s call it the effect allele: a suggestion for GWAS naming  
643 conventions. *Int J Epidemiol* 2020;49:1734–5. <https://doi.org/10.1093/IJE/DYAA149>.
- 644 [43] Burgess S, Dudbridge F, Thompson SG. Combining information on multiple  
645 instrumental variables in Mendelian randomization: comparison of allele score and  
646 summarized data methods. *Stat Med* 2016;35:1880–906.  
647 <https://doi.org/https://doi.org/10.1002/sim.6835>.
- 648 [44] Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization  
649 studies with weak instruments. 2010.
- 650 [45] Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, et al. A  
651 Multivariate Genome-Wide Association Analysis of 10 LDL Subfractions, and Their  
652 Response to Statin Treatment, in 1868 Caucasians. *PLoS One* 2015;10:120758.  
653 <https://doi.org/10.1371/JOURNAL.PONE.0120758>.
- 654 [46] Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al.  
655 Guidelines for performing Mendelian randomization investigations. *Wellcome Open*  
656 *Res* 2020 4186 2020;4:186. <https://doi.org/10.12688/wellcomeopenres.15555.2>.
- 657 [47] Greco M F Del, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in  
658 Mendelian randomisation studies with summary data and a continuous outcome. *Stat*  
659 *Med* 2015;34:2926–40. <https://doi.org/10.1002/SIM.6522>.
- 660 [48] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid  
661 instruments: effect estimation and bias detection through Egger regression. *Int J*  
662 *Epidemiol* 2015;44:512–25. <https://doi.org/10.1093/ije/dyv080>.
- 663 [49] Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in  
664 Mendelian Randomization with Some Invalid Instruments Using a Weighted Median  
665 Estimator. *Genet Epidemiol* 2016;40:304–14.  
666 <https://doi.org/https://doi.org/10.1002/gepi.21965>.
- 667 [50] Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian  
668 randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*  
669 2017;46:1985–98. <https://doi.org/10.1093/ije/dyx102>.
- 670 [51] Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy  
671 in causal relationships inferred from Mendelian randomization between complex traits  
672 and diseases. *Nat Genet* 2018 505 2018;50:693–8. <https://doi.org/10.1038/s41588-018-0099-7>.
- 673
- 674 [52] Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of  
675 multivariable Mendelian randomization in the single-sample and two-sample summary  
676 data settings. *Int J Epidemiol* 2019;48:713–27. <https://doi.org/10.1093/ije/dyy262>.
- 677 [53] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an  
678 open access resource for identifying the causes of a wide range of complex diseases of  
679 middle and old age. *PLoS Med* 2015;12:e1001779–e1001779.

- 680 <https://doi.org/10.1371/journal.pmed.1001779>.
- 681 [54] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK  
682 Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–9.  
683 <https://doi.org/10.1038/s41586-018-0579-z>.
- 684 [55] Simon Sheard Rob Nicholls James Froggatt AM. UK Biobank Haematology Data  
685 Companion Document n.d.
- 686 [56] Burrows K, Bull CJ, Dudding T, Gormley M, Robinson T, Tan V, et al. Genome-wide  
687 Association Study of Cancer Risk in UK Biobank 2021.  
688 <https://doi.org/10.5523/bris.aed0u12w0ede20olb0m77p4b9>.
- 689 [57] Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The Allelic  
690 Landscape of Human Blood Cell Trait Variation and Links to Common Complex  
691 Disease. *Cell* 2016;167:1415-1429.e19. <https://doi.org/10.1016/j.cell.2016.10.042>.
- 692 [58] Sanderson E, Spiller W, Bowden J. Testing and correcting for weak and pleiotropic  
693 instruments in two-sample multivariable Mendelian randomization. *Stat Med*  
694 2021;40:5434–52. <https://doi.org/10.1002/SIM.9133>.
- 695 [59] Wong JYY, Bassig BA, Loftfield E, Hu W, Freedman ND, Ji BT, et al. White Blood  
696 Cell Count and Risk of Incident Lung Cancer in the UK Biobank. *JNCI Cancer Spectr*  
697 2020;4. <https://doi.org/10.1093/JNCICS/PKZ102>.
- 698 [60] Watts EL, Perez-Cornago A, Kothari J, Allen NE, Travis RC, Key TJ. Hematologic  
699 markers and prostate cancer risk: a prospective analysis in UK Biobank. *Cancer*  
700 *Epidemiol Biomarkers Prev* 2020;29:1615–26. [https://doi.org/10.1158/1055-9965.EPI-19-1525/70795/AM/HAEMATOLOGICAL-MARKERS-AND-PROSTATE-](https://doi.org/10.1158/1055-9965.EPI-19-1525/70795/AM/HAEMATOLOGICAL-MARKERS-AND-PROSTATE-CANCER-RISK-A)  
701 [CANCER-RISK-A](https://doi.org/10.1158/1055-9965.EPI-19-1525/70795/AM/HAEMATOLOGICAL-MARKERS-AND-PROSTATE-CANCER-RISK-A).
- 702
- 703 [61] Benson VS, Hartl S, Barnes N, Galwey N, Van Dyke MK, Kwon N. Blood eosinophil  
704 counts in the general population and airways disease: a comprehensive review and  
705 meta-analysis. *Eur Respir J* 2022;59. <https://doi.org/10.1183/13993003.04590-2020>.
- 706 [62] Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization  
707 accounting for correlated and uncorrelated pleiotropic effects using genome-wide  
708 summary statistics. *Nat Genet* 2020 527 2020;52:740–7.  
709 <https://doi.org/10.1038/s41588-020-0631-4>.
- 710 [63] Karim AF, Westenberg LEH, Eurelings LEM, Otten R, Gerth Van Wijk R. The  
711 association between allergic diseases and cancer: a systematic review of the literature  
712 n.d.
- 713 [64] Varricchi G, Galdiero MR, Loffredo S, Lucarini V, Marone GG, Mattei F, et al.  
714 Eosinophils: The unsung heroes in cancer? vol. 7. 2018.  
715 <https://doi.org/10.1080/2162402X.2017.1393134>.
- 716 [65] Arnold IC, Artola-Boran M, Gurtner A, Bertram K, Bauer M, Frangez Z, et al. The  
717 GM-CSF-IRF5 signaling axis in eosinophils promotes antitumor immunity through  
718 activation of type 1 T cell responses. *J Exp Med* 2020;217.  
719 <https://doi.org/10.1084/JEM.20190706/152117>.
- 720 [66] Platzer B, Elpek KG, Cremasco V, Baker K, Stout MM, Schultz C, et al. IgE/FcεRI-  
721 mediated antigen cross-presentation by dendritic cells enhances anti-tumor immune  
722 responses. *Cell Rep* 2015;10:1487. <https://doi.org/10.1016/J.CELREP.2015.02.015>.
- 723 [67] Legrand F, Driss V, Delbeke M, Loiseau S, Hermann E, Dombrowicz D, et al. Human  
724 Eosinophils Exert TNF-α and Granzyme A-Mediated Tumoricidal Activity toward  
725 Colon Carcinoma Cells. *J Immunol* 2010;185:7443–51.  
726 <https://doi.org/10.4049/JIMMUNOL.1000446>.
- 727 [68] Idos GE, Kwok J, Bonthala N, Kysh L, Gruber SB, Qu C. The Prognostic Implications  
728 of Tumor Infiltrating Lymphocytes in Colorectal Cancer: A Systematic Review and  
729 Meta-Analysis. *Sci Reports* 2020 101 2020;10:1–14. <https://doi.org/10.1038/s41598->

- 730 020-60255-4.
- 731 [69] Boursi B, Mamtani R, Hwang WT, Haynes K, Yang YX. A Risk Prediction Model for  
732 Sporadic CRC Based on Routine Lab Results. *Dig Dis Sci* 2016;61:2076–86.  
733 <https://doi.org/10.1007/S10620-016-4081-X/TABLES/4>.
- 734 [70] Goshen R, Mizrahi B, Akiva P, Kinar Y, Choman E, Shalev V, et al. Predicting the  
735 presence of colon cancer in members of a health maintenance organisation by  
736 evaluating analytes from standard laboratory records. *Br J Cancer* 2017 1167  
737 2017;116:944–50. <https://doi.org/10.1038/bjc.2017.53>.
- 738 [71] Siedlinski M, Jozefczuk E, Xu X, Teumer A, Evangelou E, Schnabel RB, et al.  
739 Siedlinski et al White Blood Cells and Blood Pressure Indices. *Circulation*  
740 2020;141:1307. <https://doi.org/10.1161/CIRCULATIONAHA.119.045102>.
- 741 [72] Welsh C, Welsh P, Mark PB, Celis-Morales CA, Lewsey J, Gray SR, et al. Association  
742 of total and differential leukocyte counts with cardiovascular disease and mortality in  
743 the UK Biobank. *Arterioscler Thromb Vasc Biol* 2018;38:1415–23.  
744 <https://doi.org/10.1161/ATVBAHA.118.310945>.  
745  
746

## FIGURES for “Investigating the effect of circulating immune cell counts on colorectal cancer risk”

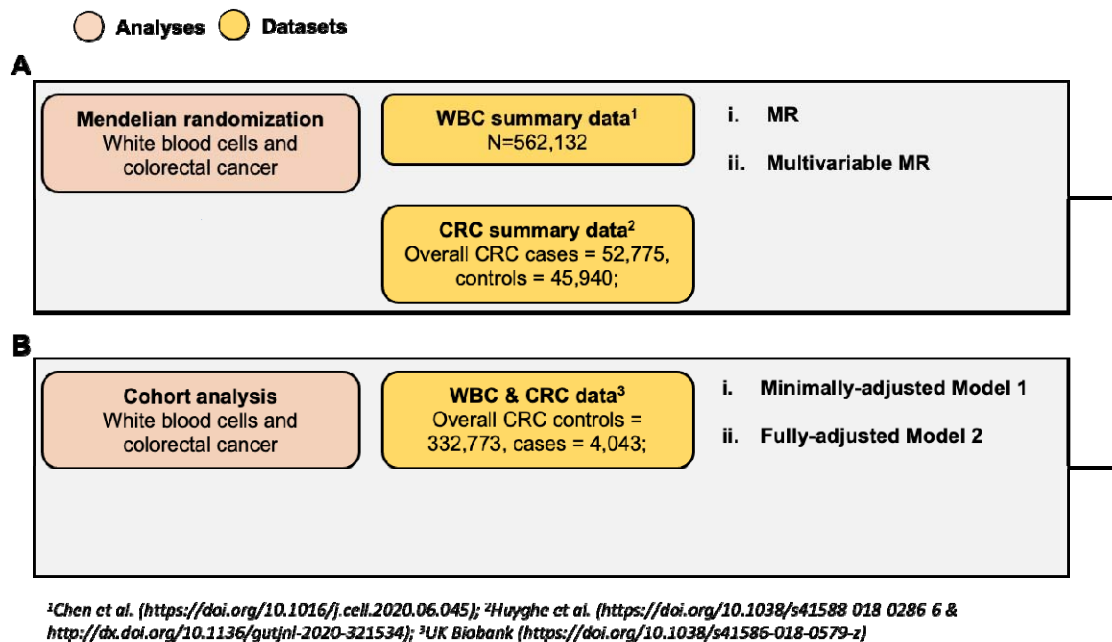


Figure 1: Study design

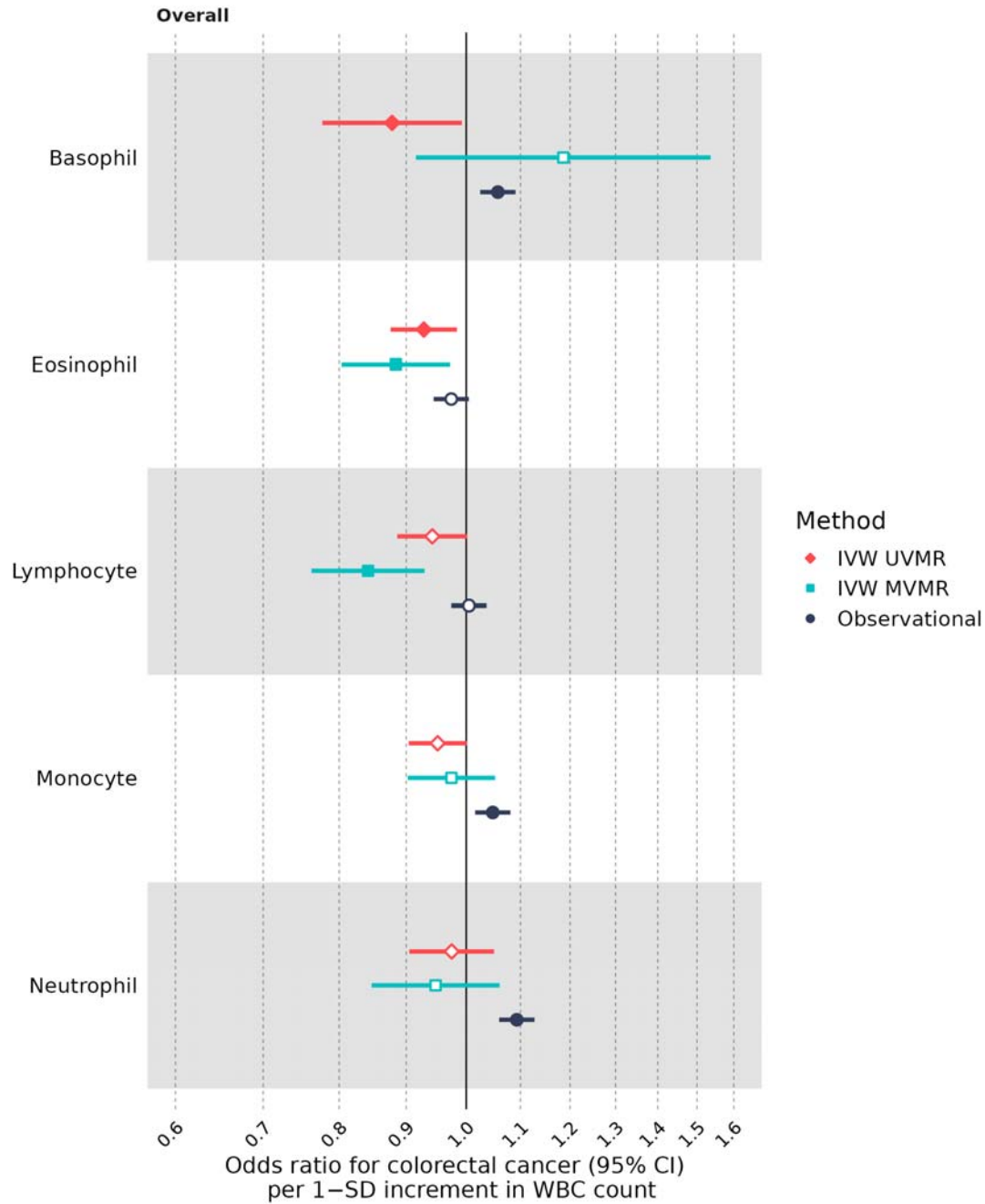


Figure 2: MR and Cohort analysis of WBC count on overall colorectal cancer