Mendelian randomization study of body mass index and colorectal cancer risk

Aaron P. Thrift[1,2], **Jian Gong**[1], **Ulrike Peters**[1,3], **Jenny Chang-Claude**[4], **Anja Rudolph**[4], **Martha L. Slattery**[5], **Andrew T. Chan**[6,7], **Adam E. Locke**[8], **Bratati Kahali**[9], **Anne E. Justice**[10], **Tune H. Pers**[11,12,13,14], **Steven Gallinger**[15], **Richard B Hayes**[16], **John A. Baron**[17], **Bette J. Caan**[18], **Shuji Ogino**[19,20,21], **Sonja I. Berndt**[22], **Stephen J. Chanock**[22], **Graham Casey**[23], **Robert W. Haile**[24], **Mengmeng Du**[1,3,7], **Tabitha A. Harrison**[1], **Mark Thornquist**[1], **David J. Duggan**[25], **Loïc Le Marchand**[26], **Noralane M. Lindor**[27], **Daniela Seminara**[28], **Mingyang Song**[21,29], **Kana Wu**[29], **Stephen N. Thibodeau**[30], **Michelle Cotterchio**[31], **Aung Ko Win**[32], **Mark A. Jenkins**[32], **John L. Hopper**[32], **Cornelia M. Ulrich**[33], **John D. Potter**[1,3,34], **Polly A. Newcomb**[1,3], **Michael Hoffmeister**[35], **Hermann Brenner**[35,36], **Emily White**[1,3], **Li Hsu**[1,37], and **Peter T. Campbell**[38]

[1]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA [2]Department of Medicine and Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA [3]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA [4]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany [5]Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT, USA [6]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA [7]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA [8]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA [9]Department of Internal Medicine, Division of Gastroenterology, and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA [10]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA [11]Divisions of Endocrinology and Genetics and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA [12]Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA [13]Department of Genetics, Harvard Medical School, Boston, MA, USA [14]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark [15]Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada [16]Division of Epidemiology, Department of Population Health, New York University School of Medicine, New York, NY, USA [17]Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, USA [18]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA [19]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA [20]Department of Medical Oncology, Dana-Farber

Corresponding author: Aaron P. Thrift, Ph.D., Dan L. Duncan Cancer Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas, 77030-3498. Tel: 713-798-9107; Fax: 713-798-3658; aaron.thrift@bcm.edu.

Cancer Institute, Boston, MA, USA [21]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA [22]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA [23]USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA [24]Department of Medicine, Division of Oncology, Stanford University, Stanford, CA, USA [25]Translational Genomics Research Institute, Phoenix, AZ, USA [26]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA [27]Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ, USA [28]Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA [29]Department of Nutrition, Harvard School of Public Health, Boston, MA, USA [30]Departments of Laboratory Medicine and Pathology and Laboratory Genetics, Mayo Clinic, Rochester, MN, USA [31]Prevention and Cancer Control, Cancer Care Ontario, Toronto, Canada [32]Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia [33]Population Sciences Huntsman Cancer Institute, Salt Lake City, UT, USA [34]Centre for Public Health Research, Massey University, Wellington, New Zealand [35]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany [36]German Cancer Consortium (DKTK), Heidelberg, Germany [37]Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA, USA [38]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA

Abstract

Background—High body mass index (BMI) is consistently linked to increased risk of colorectal cancer (CRC) for men, whereas the association is less clear for women. As risk estimates from observational studies may be biased and/or confounded, we conducted a Mendelian randomization study to estimate the causal association between BMI and CRC.

Methods—We used data from 10,226 CRC cases and 10,286 controls of European ancestry. The Mendelian randomization analysis used a weighted genetic risk score, derived from 77 genome-wide association study identified variants associated with higher BMI, as an instrumental variable (IV). We compared the IV odds ratio (IV-OR) with the OR obtained using a conventional covariate-adjusted analysis.

Results—Individuals carrying greater numbers of BMI-increasing alleles had higher CRC risk (per weighted allele OR, 1.31; 95% confidence interval [CI], 1.10–1.57). Our IV estimation results support the hypothesis that genetically influenced BMI is directly associated with risk for CRC (IV-OR per 5 kg/m^2, 1.50; 95% CI, 1.13–2.01). In the sex-specific IV analyses higher BMI was associated with higher risk of CRC among women (IV-OR per 5 kg/m^2, 1.82; 95% CI, 1.26–2.61). For men, genetically influenced BMI was not associated with CRC (IV-OR per 5 kg/m^2, 1.18; 95% CI, 0.73–1.92).

Conclusions—High BMI was associated with increased CRC risk for women. Whether abdominal obesity, rather than overall obesity, is a more important risk factor for men requires further investigation.

Impact—Overall, conventional epidemiologic and Mendelian randomization studies suggest a strong association between obesity and the risk of CRC.

Keywords

Colorectal cancer; Obesity; Epidemiology; Sex; Risk factors

Introduction

Observational studies consistently show that high body mass index (BMI) is associated with an increased risk of colorectal cancer (CRC) for men, whereas the magnitude of the association is less clear for women (1). Although these findings came from well designed and conducted epidemiologic studies, associations from observational studies could arise from bias and residual confounding and, therefore, may not necessarily reflect a causal relationship. First, the correlation between BMI and multiple other lifestyle and clinical factors suggests that adequate control for all potential confounding may be difficult. Conversely, over-adjustment in multivariable models can attenuate risk estimates towards the null and may explain the smaller associations reported for women (e.g., adjusting for menopausal hormone therapy) (2). Second, conventional epidemiologic analyses are often based on self-reports of participants' weights and heights at one point in time and thus may be subject to misclassification from inaccurate reporting and insufficient follow-up to capture changes in measurements over time (e.g., men over-report height more so than women, and heavier women tend to under-report weight more so than heavier men (3, 4)). Finally, early stages of cancer may result in weight loss, and so measures of body weight reported within a few years of CRC diagnosis may introduce bias, a particular problem in case-control studies. As the latent period between high BMI and CRC risk is unknown, a measure that reflects typical lifetime BMI may counteract some of these limitations. Determining whether or not the association is causal is important as BMI could be a key target for primary prevention of CRC.

The use of genetic variants under the Mendelian randomization framework is one approach to exploring the possible causal nature of the observed associations between BMI and risk of CRC (5). Mendelian randomization uses instrumental variables (e.g., genetic variants that proxy for directly measured environmental risk factors) to make causal inferences about the relationship between a risk factor and an outcome and this approach can overcome issues of confounding, recall bias and reverse causality inherent in observational studies (5–7). First, because alleles are randomly allocated from parents to offspring during gamete formation, the association between the instrumental variable and the outcome is not confounded by environmental exposures. Second, genetic variants are measured reliably and are not affected by disease status. Furthermore, as those with the risk alleles have been essentially allocated randomly to higher levels of exposure across the lifecourse, genetic variation may more accurately reflect lifetime exposure compared with a single measurement. A potential limitation of Mendelian randomization arises however for highly complex traits whereby genetics may explain only a small fraction of the phenotypic variability. Many early Mendelian randomization studies used a single variant as an instrumental variable and, in the case described above, were unable to test the full spectrum of genetically influenced

phenotype. By using multiple genetic variants as instrumental variables, we can increase statistical power and improve the precision of the instrumental variable estimate (8).

We therefore conducted a Mendelian randomization study using a weighted genetic risk score, derived from 77 genetic variants associated with higher BMI, as an instrumental variable to reassess the observational estimates of the associations between BMI and CRC risk.

Materials and Methods

Our analyses included individual participant data and genetic material from 10,226 CRC cases and 10,286 population-based controls of European ancestry in the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and Colon-Cancer Family Registry (C-CFR). The 11 studies (6 cohort and 5 case-control) used in our analyses (Supplementary Table S1) have been described in detail previously (9) and include the Health Professionals Follow-up Study (HPFS) (10); Nurses' Health Study (NHS) (11); Physician's Health Study (PHS) (12); Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) (13); VITamins and Lifestyle Study (VITAL) (14); Women's Health Initiative (WHI) (15); the Colon-Cancer Family Registry (C-CFR) (16); Ontario Familial Colon Cancer Registries (OFCCR) (17); Diet, Activity and Lifestyle Survey (DALS) (18, 19); Postmenopausal Hormone Study (PMH-CCFR) (20); and Darmkrebs: Chancen der Verhütung durch Screening (DACHS) (21). To avoid confounding by population stratification, we used principal components analysis (HapMap II populations as reference) to restrict our analyses to individuals of European ancestry (22). Cases were men and women with primary invasive colorectal adenocarcinoma (International Classification of Diseases, 9th revision (ICD-9) codes 153.0–153.4, 153.6–153.9, and 154.0–154.1) and histologic confirmation (by medical records, pathologic reports, or death certificates) of CRC was performed for all studies. All participants provided written, informed consent and each study was approved by the relevant Institutional Review Boards.

Data on demographic and lifestyle factors were collected using in-person interviews or self-completed questionnaires, and data were centrally harmonized as previously described (23). BMI was calculated as weight in kilograms divided by height in meters squared (kg/m^2). Although the studies obtained participants' weight (either self-report or measured) at different time points with respect to CRC diagnosis (Supplementary Table S1), we used weight that most accurately reflected usual adult weight prior to any disease-related weight loss. For the Mendelian randomization analysis, we used a weighted genetic risk score as an instrumental variable (IV) for BMI. A recent genome-wide association study (GWAS) conducted by the Genetic Investigation of Anthropometric Traits (GIANT) consortium involving over 300,000 individuals of European-descent identified 77 single nucleotide polymorphisms (SNPs) associated with BMI at genome-wide significance (e.g., $P<5\times10^{-8}$) (24). Genotype information on these 77 SNPs was used to construct the weighted genetic risk score. The GIANT GWAS for BMI found no evidence of non-additive effects among SNPs, thus we fit additive SNP effects. For each SNP, participants received a score of 0, 1 or 2 for carrying zero (wild-type homozygous), one (heterozygous) or two (homozygous for the risk allele) alleles associated with higher BMI. For imputed SNPs, participants received

scores that ranged from 0 to 2. SNPs were externally-weighted in the score by the per-allele change in BMI (the increase in BMI per 1 additional risk allele) reported for that SNP in the GIANT GWAS for BMI (24). Details on genotyping and quality assurance/quality control have been previously published (9). Call rates were 98% and all selected SNPs were in Hardy-Weinberg Equilibrium among controls (P 1×10^{-4}).

Statistical analysis

Associations between potential confounders and CRC risk were assessed using chi-square tests and Student t-tests. Associations among potential confounders, the IV, and BMI were assessed in controls using linear regression, Student t-tests and analysis of variance.

All participants were analyzed using a single-model (pooled) approach with adjustment for study. For the IV analysis, we used the two-stage control function IV approach, with the genetic risk score as an IV for BMI, and assumed an additive model (25). First, we fit a linear regression model within controls to predict BMI from the IV, adjusted for study and the top three principal components of genetic ancestry. Next, we fit a logistic regression model with BMI as a predictor and CRC case-control status as the outcome using robust standard errors and adjusted for the residuals from the first step (25). The coefficient for BMI from the second-stage model is the causal estimate for BMI. In Mendelian randomization studies, the IV for the exposure of interest is considered sufficiently strong if the first-stage F-statistic exceeds 10 (26). For the conventional covariate-adjusted analysis, BMI was the variable of interest and odds ratios (OR) and 95% confidence intervals (95%CI) were estimated using unconditional logistic regression models. We conducted minimally adjusted analyses (models were adjusted for age, sex, and study) and fully adjusted analyses (models were additionally adjusted for smoking status, family history of cancer, history of diabetes, use of aspirin or non-steroidal anti-inflammatory drugs [NSAIDs], consumption of fruit, vegetables, processed meat, and red meat, and [women only] use of menopause hormone therapy). Variables were retained in the fully adjusted analyses if they were statistically significantly associated with CRC risk after mutual adjustment for other risk factors. Participants with missing data on covariates were included in the analyses using an additional category for missing values. We assessed whether the estimates obtained using conventional epidemiologic and IV methods were different using the Durbin form of the Durbin-Wu-Hausman statistic (27). Rejection of the null hypothesis (i.e., risk estimates for BMI from both analyses are equal) indicates that the OR from the conventional covariate-adjusted analysis differs from the OR from the IV approach. We examined potential statistical interaction between BMI and sex by including a multiplicative term in the model, and conducted additional analyses for the BMI-CRC association stratified by sex and cancer sub-site. Furthermore, we performed a sensitivity analysis excluding studies with self-reported weight. In a supplementary analysis, we examined the association between a weighted genetic risk score for waist-to-hip ratio (WHR), derived using 47 variants associated with WHR adjusted for BMI (28), and CRC risk. Statistical significance was determined at $\alpha=0.05$ and all tests for statistical significance were two-sided. Analyses were performed using Stata 12.0 (StataCorp, College Station, TX).

Results

The mean age of study participants was 63 years and 56% were women (Supplementary Table S1). Compared with controls, cases were more likely to smoke cigarettes, have a family history of cancer, and have diabetes and less likely to use aspirin or NSAIDs and (women only) menopausal hormones (Table 1). Cases consumed more red and processed meats on average than controls, but less fruit and vegetables. On average, cases (women: mean BMI, 27.0; standard deviation [SD], 5.4; men: mean BMI, 27.3; SD, 4.0) had statistically significantly higher BMI than controls (women: mean BMI, 26.2; SD, 5.0; men: mean, 26.5; SD, 3.7) (Supplementary Table S1). BMI was statistically significantly associated with age, sex, history of diabetes, use of aspirin/NSAIDs, consumption of fruit, processed meat, and red meat and, for women only, use of menopause hormone therapy (Supplementary Table S2).

The allele frequencies for each of the 77 SNPs were consistent across the 11 studies (data not shown). The IV was normally distributed (mean, 1.97; median, 1.97; SD, 0.2; minimum, 1.37; maximum, 2.66) and the association between the IV and BMI was homogeneous across studies (test for heterogeneity, $I^2 = 0\%$; $P = 0.45$; Figure 1). From the pooled analysis, a one-unit increase in the IV was associated with a 3.23 kg/m^2 (95% CI, 2.66–3.79) increase in BMI. The IV explained only 1.2% of the variance in BMI, but was a sufficiently strong instrument for BMI (F-statistic, 126). There were no associations of the IV with age, sex, smoking status, family history of cancer, consumption of fruit, vegetables, processed meat, and red meat or use of menopause hormone therapy (Supplementary Table S2). However, the IV was associated with history of diabetes, and we found a modest positive association between the IV and use of aspirin/NSAIDs (Supplementary Table S2).

Individuals with greater numbers of (weighted) BMI-increasing alleles (i.e., those with a higher weighted genetic risk score) were at increased risk for CRC (per weighted allele OR, 1.31; 95%CI, 1.10–1.57; Supplementary Table S3). The IV analysis showed evidence that higher BMI was causally associated with increased risk of CRC (IV-OR per 5 kg/m^2, 1.50; 95% CI, 1.13–2.01). The IV point estimate was greater in magnitude than the point estimate from conventional covariate-adjusted analysis (minimally adjusted OR per 5 kg/m^2, 1.18; 95% CI, 1.15–1.22); however the 95% CIs overlapped and they were not statistically significantly different from one another (*P-difference* = 0.10). The findings were similar when we examined the BMI-CRC association by cancer sub-site in the colorectum (Table 2). Additional adjustment for confounding did not change the estimates in the conventional analysis (Table 2) and, likewise, adjustment of the IV analysis for age, sex, history of diabetes, and use of aspirin/NSAIDs did not change the IV risk estimate for BMI (IV-OR per 5 kg/m^2, 1.51; 95%CI, 1.13–2.02).

Using conventional methods, we found a statistically significant interaction between BMI and sex (CRC, $P_{interaction}$<0.001; Colon, $P_{interaction}$ <0.001, Rectal, $P_{interaction}$ =0.02). The IV was a sufficiently strong instrument for BMI in stratified analyses for men (F-statistic, 74.2; R^2, 1.6%) and women (F-statistic, 60.5; R^2, 1.0%). For men, there was no evidence that carriers of greater numbers of (weighted) BMI-increasing alleles were at increased risk for CRC and the risk estimate for CRC obtained from the IV analysis (IV-OR per 5 kg/m^2,

1.18; 95% CI, 0.73–1.92) contained 95% CIs that included the null; however, the IV results were in the same direction as the estimate from conventional covariate-adjusted analysis (minimally adjusted OR per 5 kg/m^2, 1.30; 95% CI, 1.23–1.38) and the two estimates were not statistically significantly different (*P-difference* = 0.70). In contrast, for women, carriers of greater numbers of (weighted) BMI-increasing alleles were at increased risk for CRC and the IV estimate for CRC risk (IV-OR per 5 kg/m^2, 1.82; 95% CI, 1.26–2.61) was statistically significantly different from the estimate obtained from conventional covariate-adjusted analysis (minimally adjusted OR per 5 kg/m^2, 1.14; 95% CI, 1.10–1.18; *P-difference* = 0.01).

For men, the OR for BMI associated with CRC was greater in magnitude when the conventional covariate-adjusted analysis was restricted to those studies with measured height and weight (minimally adjusted OR per 5 kg/m^2, 1.53; 95% CI, 1.33–1.76); there was no difference to the overall analysis for women (minimally adjusted OR per 5 kg/m^2, 1.13; 95% CI, 1.06–1.19).

In supplementary analyses, individuals with greater numbers of (weighted) WHR-increasing alleles were at increased risk for CRC (per weighted allele OR, 1.36; 95% CI, 1.08–1.73). In contrast to that seen for BMI, men with greater numbers of (weighted) WHR-increasing alleles were at increased risk for CRC, while there was no evidence that women carriers of greater numbers of (weighted) WHR-increasing alleles were at increased risk for CRC (Supplementary Table S3).

Discussion

Observational studies have consistently reported that high BMI is associated with increased risk of CRC; however, the magnitude of the BMI-CRC association may differ between men and women (1). The limitations of observational studies, such as residual confounding (e.g., by socio-economic status, smoking, alcohol, and other lifestyle and behavioral factors), reverse causation, and reporting bias, may contribute to the heterogeneity reported in previous studies. Using Mendelian randomization methods, we re-examined the association between BMI and CRC risk without the inherent limitations in observational studies and compared the IV and conventional estimates. Overall, we found that BMI was associated with CRC risk; however, the sex-specific associations in our study are in conflict with expectations based on the existing literature of the BMI-CRC association. For women, genetically influenced BMI was statistically significantly associated with risk of CRC and our findings suggest that BMI may confer greater risk of CRC than previously reported (1). On the other hand, for men, while the IV and conventional estimates of CRC risk associated with high BMI were not statistically significantly different, the IV for BMI was not associated with CRC risk for men.

Results from large meta-analyses indicate that the risk of CRC for men increases by 20–30% per 5 kg/m^2 increase in BMI. For women, the magnitude of the association was lower in all studies (~10% increased risk per 5 kg/m^2) and did not always reach statistical significance (1). Consistent with these results, in our conventional covariate-adjusted analysis, the

magnitude of the BMI-CRC association was substantially higher for men (32% increased risk per 5 kg/m^2) than for women (15% increased per 5 kg/m^2).

Using the Mendelian randomization approach, for women, we found evidence that genetically raised BMI was associated with increased risk of CRC. Importantly, the instrumental variable estimation for the BMI-CRC association for women was higher in magnitude than the risk estimate for BMI obtained from the conventional covariate-adjusted analysis. This suggests that the conventional risk estimate for women in our study, as well as those previously reported, may underestimate the true causal association between BMI and CRC risk for women. It is possible that previous estimates may have been biased toward the null if heavier women were more likely to under-report their weight than women with normal weight (3). This appears unlikely in our study as the results for women were similar between self-reported and measured BMI. There may also be sex-specific behavioral factors associated with BMI and CRC that may not be adequately measured and/or controlled for in a conventional epidemiologic analysis. This may explain why the risk estimates for BMI were different between self-reported and measured BMI for men, but not for women. It is also possible that the stronger effect in the IV analysis for women may be due to the IV for BMI reflecting genetic factors that express themselves early in life (e.g., BMI in childhood) and influence CRC risk more so than adult weight gain. An early effect of BMI on CRC risk for women may be plausible based on possible interactions between the different hormonal milieu associated with pre- and post-menopause. Mendelian randomization methods may better capture an early BMI effect on CRC risk for women than conventional epidemiologic methods. However, while our findings suggest that higher genetically influenced BMI is directly associated with increased CRC risk for women, the instrumental variable estimate is a valid test of the direction of the effect, and given the low variance explained, the exact magnitude of the effect could be in the range between the conventional estimates and our IV estimate.

For men, the conventional and IV estimates were not statistically significantly different from one another. However, the instrumental variable estimation gave evidence that genetically influenced BMI was not associated with CRC risk for men. It is unclear why the IV estimate for men was not statistically significant; however, it is unlikely that this was due to weak instrument bias as the IV was a strong instrument for BMI (F-statistic = 74.2). A possible explanation could be that the distribution of body fat, rather than total body fatness, may be an important predictor of CRC risk for men. In the supplementary analysis, we found that an IV for WHR was associated with CRC risk for men but not for women. Given that the IV was derived using SNPs associated with WHR adjusted for BMI, the two IVs in this study are largely distinct and these findings suggest that fat distribution may be important for CRC risk for men, while overall obesity may be more important for CRC risk for women.

The strengths of our study include the large sample size, centrally harmonized data, and the robust IV for the Mendelian randomization analysis. However, the findings from our Mendelian randomization study rely on three key assumptions, some of which are untestable: First, that the IV is associated with BMI. We constructed a weighted genetic risk score from 77 SNPs associated with higher BMI in an externally conducted genome-wide association study in order to maximize power (8) and provided evidence that the weighted

genetic risk score (the IV) is associated with BMI in our study (a sufficiently strong instrument with F-statistics >> 10). However, our IV analyses may have been underpowered as the IV explained less than 2% of total phenotypic variation for BMI, certainly a limitation for this study but one that is common when constructing IVs for highly complex traits – and seen in the original Locke paper (24) that established the relationship between the SNPs and BMI. The second assumption with Mendelian randomization is that the IV is associated with CRC only through its association with BMI. Although the biological effects of all the variants used in constructing the IV are not well known and this assumption cannot be tested directly, we feel that it is highly unlikely that a large subset of the 77 SNPs in the score would have pleiotropic associations with BMI and CRC. If the IV is associated with CRC risk through the same metabolic pathways (e.g., inflammation, oxidative stress, glucose/insulin dysregulation) that are suspected to connect BMI to carcinogenesis, then this second assumption would still hold, since both the IV and BMI would be considered markers of altered metabolic (or related) risk factors. The final assumption is that the IV is not associated with factors that confound conventional observational associations. We found that the weighted genetic risk score was not associated with age, sex, smoking status, family history of cancer, diet, and other factors that may confound the BMI-CRC association in conventional covariate-adjusted analyses. The IV was associated with diabetes and use of aspirin/NSAIDs. The correlation with diabetes was expected since high BMI is such a strong risk factor for diabetes, and indeed, this is probably further evidence that the IV is an indicator of obesity-associated altered metabolism. The association with aspirin/NSAID-use was not expected *a priori* but might relate to prophylactic or treatment related use of aspirin regimens among heavier weight persons, or it might be due to chance. Nonetheless, inclusion of diabetes and aspirin/NSAID use in the IV analyses did not change the IV risk estimates for BMI. The majority of studies included in our analyses used self-reported weight and height and we showed that this may bias conventional estimates. Imprecision on genotypes from imputed SNPs is a limitation and may have resulted in loss of power in the instrumental variable analysis. A final limitation in this study is the lack of data on tumor molecular phenotype (29–31) and further research is warranted to answer this question.

Unlike estimates obtained from conventional covariate-adjusted analysis of observational data, the estimates from our IV analysis are not limited by confounding and/or bias. Therefore, they may be interpretable as a measure of the causal effect of BMI on CRC risk. Several mechanisms have been proposed to explain the association between high BMI and CRC risk, including insulin and the insulin-like growth factor system, adipokines (e.g., leptin, adiponectin), inflammation (e.g., C-reactive protein), oxidative stress, altered immunity, and steroid hormones, as discussed in recent comprehensive reviews (1, 32, 33).

Overweight and obesity are good targets for primary prevention of CRC, and men and women should be encouraged to maintain a normal BMI (18.5–25 kg/m^2). Given the contrasting findings in this study for men and women, future studies are still needed to better understand the role of obesity (overall versus abdominal) in CRC carcinogenesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Bardou M, Barkun AN, Martel M. Obesity and colorectal cancer. Gut. 2013; 62:933–947. [PubMed: 23481261]

2. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. Epidemiology. 2009; 20:488–495. [PubMed: 19525685]

3. Shields M, Connor Gorber S, Tremblay MS. Estimates of obesity based on self-report versus direct measures. Health Rep. 2008; 19:61–76. [PubMed: 18642520]

4. Connor Gorber S, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. Obes Rev. 2007; 8:307–326. [PubMed: 17578381]

5. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003; 32:1–22. [PubMed: 12689998]

6. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008; 27:1133–1163. [PubMed: 17886233]

7. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000; 29:722–729. [PubMed: 10922351]

8. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res. 2012; 21:223–242. [PubMed: 21216802]

9. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. Gastroenterology. 2013; 144:799–807. [PubMed: 23266556]

10. Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Kearney J, et al. A prospective study of cigarette smoking and risk of colorectal adenoma and colorectal cancer in U.S. men. J Natl Cancer Inst. 1994; 86:183–191. [PubMed: 8283490]

11. Giovannucci E, Colditz GA, Stampfer MJ, Hunter D, Rosner BA, Willett WC, et al. A prospective study of cigarette smoking and risk of colorectal adenoma and colorectal cancer in U.S. women. J Natl Cancer Inst. 1994; 86:192–199. [PubMed: 8283491]

12. Christen WG, Gaziano JM, Hennekens CH. Design of Physicians' Health Study II--a randomized trial of beta-carotene, vitamins E and C, and multivitamins, in prevention of cancer, cardiovascular disease, and eye disease, and review of results of completed trials. Ann Epidemiol. 2000; 10:125–134. [PubMed: 10691066]

13. Prorok PC, Andriole GL, Bresalier RS, Buys SS, Chia D, Crawford ED, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Control Clin Trials. 2000; 21(6 Suppl):273S–309S. [PubMed: 11189684]

14. White E, Patterson RE, Kristal AR, Thornquist M, King I, Shattuck AL, et al. VITamins And Lifestyle cohort study: study design and characteristics of supplement users. Am J Epidemiol. 2004; 159:83–93. [PubMed: 14693663]

15. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control Clin Trials. 1998; 19:61–109. [PubMed: 9492970]

16. Newcomb PA, Baron J, Cotterchio M, Gallinger S, Grove J, Haile R, et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. Cancer Epidemiol Biomarkers Prev. 2007; 16:2331–2343. [PubMed: 17982118]

17. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet. 2007; 39:989–994. [PubMed: 17618283]

18. Slattery ML, Friedman GD, Potter JD, Edwards S, Caan BJ, Samowitz W. A description of age, sex, and site distributions of colon carcinoma in three geographic areas. Cancer. 1996; 78:1666–1670. [PubMed: 8859178]

19. Slattery ML, Potter J, Caan B, Edwards S, Coates A, Ma KN, et al. Energy balance and colon cancer--beyond physical activity. Cancer Res. 1997; 57:75–80. [PubMed: 8988044]

20. Newcomb PA, Zheng Y, Chia VM, Morimoto LM, Doria-Rose VP, Templeton A, et al. Estrogen plus progestin use, microsatellite instability, and the risk of colorectal cancer in women. Cancer Res. 2007; 67:7534–7539. [PubMed: 17671225]

21. Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M. Protection from colorectal cancer after colonoscopy: a population-based, case-control study. Ann Intern Med. 2011; 154:22–30. [PubMed: 21200035]

22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]

23. Hutter CM, Chang-Claude J, Slattery ML, Pflugeisen BM, Lin Y, Duggan D, et al. Characterization of gene-environment interactions for colorectal cancer susceptibility loci. Cancer Res. 2012; 72:2036–2044. [PubMed: 22367214]

24. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015; 518:197–206. [PubMed: 25673413]

25. Palmer TM, Sterne JA, Harbord RM, Lawlor DA, Sheehan NA, Meng S, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. Am J Epidemiol. 2011; 173:1392–1403. [PubMed: 21555716]

26. Staiger D, Stock JH. Instrumental variables regression with weak instruments. Econometrica. 1997; 65:557–586.

27. Baum CF, Schaffer ME, Stillman S. Instrumental variables and GMM: estimation and testing. Stata Journal. 2003; 3:1–32.

28. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015; 518:187–196. [PubMed: 25673412]

29. Campbell PT, Jacobs ET, Ulrich CM, Figueiredo JC, Poynter JN, McLaughlin JR, et al. Case-control study of overweight, obesity, and colorectal cancer risk, overall and by tumor microsatellite instability status. J Natl Cancer Inst. 2010; 102:391–400. [PubMed: 20208017]

30. Morikawa T, Kuchiba A, Lochhead P, Nishihara R, Yamauchi M, Imamura Y, et al. Prospective analysis of body mass index, physical activity, and colorectal cancer risk associated with beta-catenin (CTNNB1) status. Cancer Res. 2013; 73:1600–1610. [PubMed: 23442321]

31. Kuchiba A, Morikawa T, Yamauchi M, Imamura Y, Liao X, Chan AT, et al. Body mass index and risk of colorectal cancer according to fatty acid synthase expression in the nurses' health study. J Natl Cancer Inst. 2012; 104:415–420. [PubMed: 22312135]

32. Aleman JO, Eusebi LH, Ricciardiello L, Patidar K, Sanyal AJ, Holt PR. Mechanisms of obesity-induced gastrointestinal neoplasia. Gastroenterology. 2014; 146:357–373. [PubMed: 24315827]

33. Kant P, Hull MA. Excess body weight and obesity--the link with gastrointestinal and hepatobiliary cancer. Nat Rev Gastroenterol Hepatol. 2011; 8:224–238. [PubMed: 21386810]
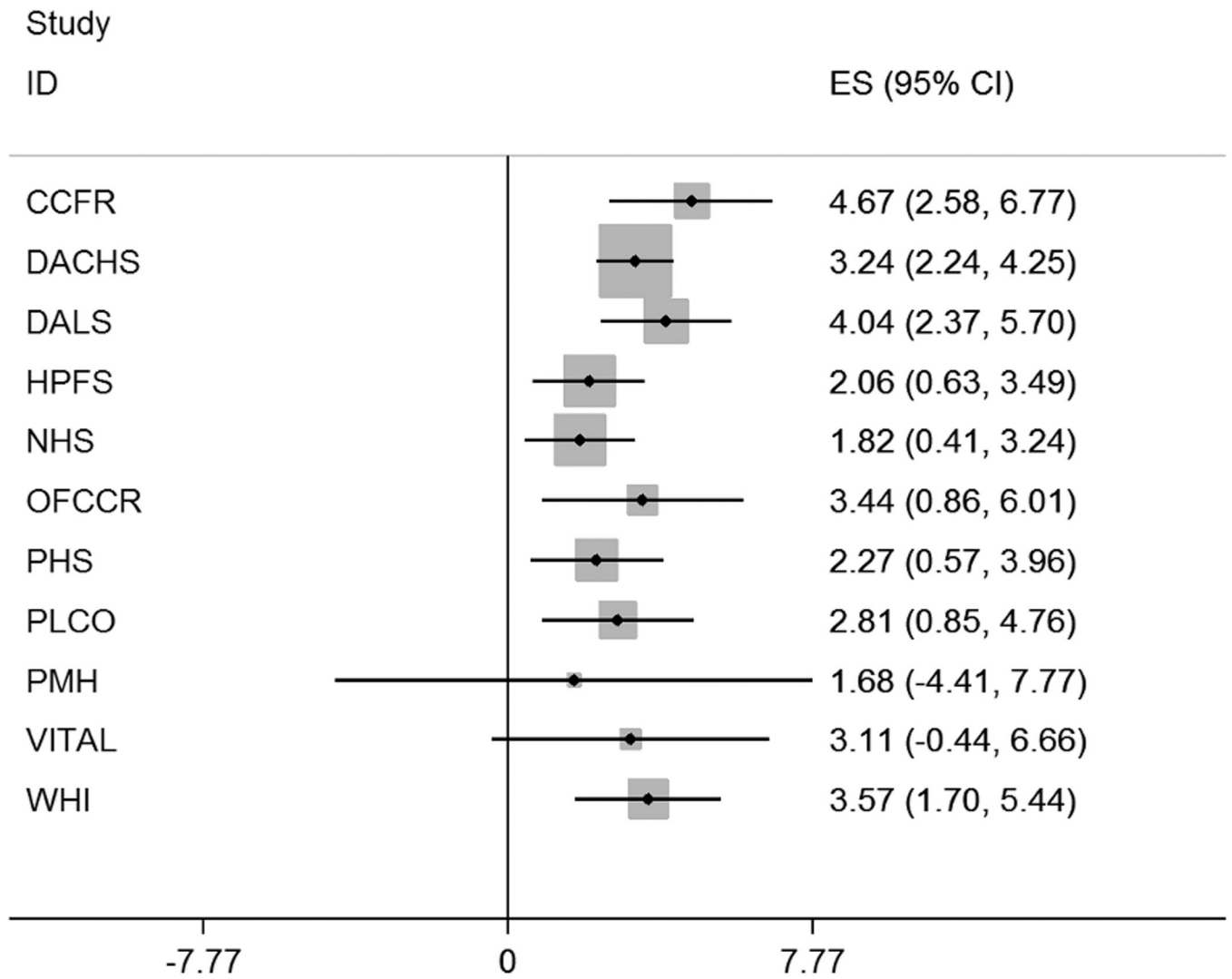
Figure 1.
The association between the weighted genetic risk score (the instrumental variable) and body mass index (BMI) across the 11 participating studies in GECCO and C-CFR.

Table 1

Characteristics of controls and cases with colorectal cancer (CRC), stratified by sex

Variable	Women			Men		
	Controls n=5747	CRC n=5644	p [a]	Controls n=4539	CRC n=4582	p [a]
Smoking status			<0.001			<0.001
Never	3029 (52.9)	2819 (50.2)		1706 (37.9)	1525 (33.6)	
Former	2145 (37.4)	2126 (37.9)		2365 (52.6)	2422 (53.3)	
Current	554 (9.7)	669 (11.9)		428 (9.5)	594 (13.1)	
Missing	19	30		40	41	
Body mass index (kg/m^2)			<0.001			<0.001
<25	2699 (47.0)	2347 (41.6)		1637 (36.1)	1324 (28.9)	
25–29.99	1935 (33.7)	1964 (34.8)		2223 (49.0)	2344 (51.2)	
30	1113 (19.4)	1333 (23.6)		679 (15.0)	914 (19.9)	
Family history			<0.001			<0.001
No	4806 (86.0)	4483 (82.0)		3685 (89.9)	3430 (82.1)	
Yes	782 (14.0)	986 (18.0)		414 (10.1)	749 (17.9)	
Missing	159	175		440	403	
Diabetes			<0.001			0.01
No	4742 (95.0)	4362 (91.1)		3262 (91.0)	3289 (89.2)	
Yes	249 (5.0)	428 (8.9)		323 (9.0)	400 (10.8)	
Missing	756	854		954	893	
Aspirin/NSAID use			<0.001			<0.001
No	3553 (62.0)	3895 (69.3)		2519 (55.7)	2928 (64.3)	
Yes	2176 (38.0)	1727 (30.7)		2004 (44.3)	1626 (35.7)	
Missing	18	22		16	28	
Fruit (servings/day)			<0.001			0.01
Q1	1052 (19.1)	1072 (20.3)		780 (18.0)	824 (19.6)	
Q2	2101 (38.2)	2226 (42.1)		1905 (43.9)	1917 (45.7)	
Q3	1397 (25.4)	1204 (22.7)		982 (22.6)	888 (21.2)	
Q4	952 (17.3)	790 (14.9)		673 (15.5)	566 (13.5)	

Variable	Women			Men		
	Controls n=5747	CRC n=5644	p [a]	Controls n=4539	CRC n=4582	p [a]
Missing	245	352		199	387	
Vegetables (servings/day)			<0.001			<0.001
Q1	1222 (22.2)	1328 (24.9)		805 (18.5)	863 (20.5)	
Q2	1825 (33.1)	1824 (34.3)		1791 (41.1)	1832 (43.5)	
Q3	1394 (25.3)	1209 (22.7)		1086 (24.9)	893 (21.2)	
Q4	1071 (19.4)	962 (18.1)		673 (15.5)	623 (14.8)	
Missing	235	321		184	371	
Processed meat (servings/day)			<0.001			<0.001
Q1	893 (19.1)	787 (19.0)		837 (23.4)	634 (18.6)	
Q2	1914 (41.0)	1514 (36.4)		1413 (39.5)	1281 (37.6)	
Q3	1117 (23.9)	1108 (26.7)		698 (19.5)	780 (22.9)	
Q4	742 (15.9)	745 (17.9)		630 (17.6)	714 (20.9)	
Missing	1081	1490		961	1173	
Red meat (servings/day)			0.001			<0.001
Q1	1489 (26.8)	1235 (23.7)		1228 (28.2)	949 (22.9)	
Q2	1439 (25.9)	1378 (26.4)		1325 (30.4)	1177 (28.4)	
Q3	1421 (25.6)	1371 (26.3)		1009 (23.1)	1077 (26.0)	
Q4	1207 (21.7)	1232 (23.6)		798 (18.3)	940 (22.7)	
Missing	191	428		179	439	
Menopause hormone therapy			<0.001			-
No	3062 (58.9)	3334 (67.6)		-	-	
Yes	2139 (41.1)	1599 (32.4)		-	-	
Missing	546	711		-	-	

NSAID, non-steroidal anti-inflammatory drug.

[a] P-value from Chi-square tests for categorical variables. Missing categories were excluded from comparison tests.

Table 2

Estimates of the effect of body mass index on risk of colorectal cancer (CRC), colon cancer and rectal cancer obtained using conventional covariate-adjusted and instrumental variable (IV) analysis (Odds ratios per 5 kg/m^2)

	Controls / Cases	Conventional covariate-adjusted analysis		Instrumental variable analysis	
		Minimally adjusted OR^a (95% CI)	Fully adjusted OR^b (95% CI)	IV-OR^c (95% CI)	P^d
All					
CRC	10286/10226	1.18 (1.15–1.22)	1.17 (1.14–1.21)	1.50 (1.13–2.01)	0.10
Colon cancer	10286/6864	1.19 (1.15–1.23)	1.18 (1.14–1.22)	1.50 (1.08–2.07)	0.17
Rectal cancer	10286/2365	1.15 (1.09–1.21)	1.14 (1.08–1.20)	1.26 (0.77–2.04)	0.73
Men					
CRC	4539/4582	1.30 (1.23–1.38)	1.29 (1.22–1.37)	1.18 (0.73–1.92)	0.70
Colon cancer	4539/2942	1.31 (1.23–1.39)	1.29 (1.21–1.37)	1.03 (0.60–1.78)	0.39
Rectal cancer	4539/1253	1.25 (1.14–1.36)	1.25 (1.13–1.37)	1.33 (0.62–2.85)	0.87
Women					
CRC	5747/5644	1.14 (1.10–1.18)	1.12 (1.08–1.17)	1.82 (1.26–2.61)	0.01
Colon cancer	5747/3922	1.15 (1.10–1.20)	1.13 (1.08–1.18)	1.94 (1.30–2.91)	0.01
Rectal cancer	5747/1112	1.11 (1.04–1.18)	1.08 (1.01–1.16)	1.24 (0.64–2.37)	0.74

CI, confidence interval; OR, odds ratio.

[a] Adjusted for age, sex, and study.

[b] Adjusted for age, sex, study, smoking status, family history of cancer, aspirin/NSAID use, diabetes, fruit, vegetable, processed meat, red meat, and (women only) menopause hormone therapy.

[c] Using weighted genetic risk score derived from 77 SNPs as an instrumental variable for BMI (Mendelian randomization), adjusted for study and the top three principal components of ancestry.

[d] For comparison of IV-OR with minimally adjusted OR (*P*<0.05 rejects the null hypothesis that they are equal).