International Journal of Epidemiology, 2015, 662–672
doi: 10.1093/ije/dyv082
Advance Access Publication Date: 20 May 2015
Original article

Mendelian Randomization Causal Analysis

Mendelian randomization study of height and risk of colorectal cancer

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] Tonu Esko,[8,9,10,11] Andrew R Wood,[12] Jian Yang,[13,14,15] Sailaja Vedantam,[9,11] Stefan Gustafsson,[16] Tune H Pers,[9,10,11] GIANT Consortium, John A Baron,[17]Stéphane Bezieau,[18] Sébastien Küry,[18] Shuji Ogino,[19,20,21] Sonja I Berndt,[22] Graham Casey,[23] Robert W Haile,[24] Mengmeng Du,[1] Tabitha A Harrison,[1] Mark Thornquist,[1] David J Duggan,[25] Loic Le Marchand,[26] Mathieu Lemire,[27] Noralane M Lindor,[28] Daniela Seminara,[29] Mingyang Song,[30] Stephen N Thibodeau,[31] Michelle Cotterchio,[32] Aung Ko Win,[33] Mark A Jenkins,[33] John L Hopper,[33] Cornelia M Ulrich,[34] John D Potter,[1,3,35] Polly A Newcomb,[1,3] Robert E Schoen,[36] Michael Hoffmeister,[37] Hermann Brenner,[37] Emily White,[1,3] Li Hsu[1,38] and Peter T Campbell,[39]

[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]Estonian Genome Center, University of Tartu, Tartu, Estonia, [9]Divisions of Endocrinology and Genetics and Center for Basic Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA, [10]Department of Genetics, Harvard Medical School, Boston, MA, USA, [11]Broad Institute, Cambridge, MA, USA, [12]Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK, [13]Queensland Brain Institute, University of Queensland, Brisbane, QLD, Australia, [14]University of Queensland Diamantina Institute, Translation Research Institute, Brisbane, QLD, Australia, [15]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby, Denmark, [16]Department of Medical Sciences, Uppsala University, Uppsala, Sweden, [17]Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC, USA, [18]CHU Nantes, Service de Génétique Médicale, Nantes, France, [19]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [20]Department of Medical Oncology, Dana-Farber 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]Ontario Institute for Cancer Research, Toronto, ON, Canada, [28]Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ, USA, [29]Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, [30]Department of Nutrition, Harvard School of Public Health, Boston, MA, USA, [31]Departments of Laboratory Medicine and Pathology and Laboratory Genetics, Mayo Clinic, Rochester, MN, USA, [32]Prevention and Cancer Control, Cancer Care Ontario, Toronto, ON, Canada, [33]Melbourne School of Population Health, University of Melbourne, Melbourne, VIC, Australia, [34]Population Sciences Huntsman Cancer Institute, Salt Lake City, UT, USA, [35]Centre for Public Health Research, Massey University, Wellington, New Zealand, [36]Department of Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA, [37]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany, [38]Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA, USA and [39]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA

*Corresponding author. Dan L. Duncan Cancer Center, Baylor College of Medicine, One Baylor Plaza, MS: BCM305, Houston, TX 77030-3498, USA. E-mail: aaron.thrift@bcm.edu

Accepted 23 April 2015

Abstract

Background: For men and women, taller height is associated with increased risk of all cancers combined. For colorectal cancer (CRC), it is unclear whether the differential association of height by sex is real or is due to confounding or bias inherent in observational studies. We performed a Mendelian randomization study to examine the association between height and CRC risk.

Methods: To minimize confounding and bias, we derived a weighted genetic risk score predicting height (using 696 genetic variants associated with height) in 10 226 CRC cases and 10 286 controls. Logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (95% CI) for associations between height, genetically predicted height and CRC.

Results: Using conventional methods, increased height (per 10-cm increment) was associated with increased CRC risk (OR = 1.08, 95% CI = 1.02–1.15). In sex-specific analyses, height was associated with CRC risk for women (OR = 1.15, 95% CI = 1.05–1.26), but not men (OR = 0.98, 95% CI = 0.92–1.05). Consistent with these results, carrying greater numbers of (weighted) height-increasing alleles (per 1-unit increase) was associated with higher CRC risk for women and men combined (OR = 1.07, 95% CI = 1.01–1.14) and for women (OR = 1.09, 95% CI = 1.01–1.19). There was weaker evidence of an association for men (OR = 1.05, 95% CI = 0.96–1.15).

Conclusion: We provide evidence for a causal association between height and CRC for women. The CRC-height association for men remains unclear and warrants further investigation in other large studies.

Key words: Body height, colorectal cancer, epidemiology

Key Messages

- Observational studies have consistently reported an association between height and risk of colorectal cancer for women; however, the association of height with colorectal cancer for men is unclear.
- We used genetically determined height to re-examine whether the association between height and risk of colorectal cancer is causal.
- The results strongly suggest that height is causally associated with colorectal cancer risk for women.
- There is weaker evidence for the association between height and increased risk of colorectal cancer for men.

Introduction

Greater attained adult height has been consistently associated with higher risk of developing colorectal cancer (CRC);[1] however, the results from these epidemiological studies have suggested that this association may be stronger for women than for men.[2–12] It is possible that the positive association observed between height and CRC risk for women is due to residual confounding by factors that are difficult to measure and control for using conventional epidemiological methods. Conversely, misclassification of height may partly attenuate a true positive and causal association between height and CRC risk for men. Assuming a causal association exists, since shorter men may tend to over-report their height more than do taller men,[13] and those shorter men would have inherently lower CRC risk, then their over-reporting of height would cause them to migrate upward to taller height categories. This influx of men inherently at lower risk of CRC, due to their shorter height, would lower the rate of CRC in this otherwise higher risk group. On the other hand, women tend to more accurately report their height, potentially explaining why such bias may not occur (or may be less) for women. Height may be a marker of biological factors (e.g. steroid hormones and other growth factors) or a marker of factors such as nutrition and energy intake that could be causally associated with the risk of developing CRC; however, these potential biases hamper the interpretation of findings from observational studies.

Mendelian randomization offers an alternative method to investigate the association between height and the risk of CRC. An advantage of using Mendelian randomization and instrumental variable methods to investigate this link is that height is highly genetically determined.[14] Mendelian randomization uses instrumental variables (e.g. genetic variants that proxy for directly measured environmental, behavioural or social factors) to make causal inferences about the relationship between a risk factor and an outcome. Using height-related genetic variants to predict height, the instrumental variable approach can help overcome issues of confounding, recall bias and reverse causality inherent in observational studies(for example, where observational associations for height may be confounded by early life exposures, diet or reporting bias) and the resulting risk estimate should be a better estimate of the true causal effect of height on CRC risk.[15–17]

Here, we used genetic variation related to height and instrumental variable methods to reinvestigate the height-CRC association. We examined the association for all persons combined and for men and women separately to clarify sex-specific associations.

Methods

All participants provided written, informed consent and studies were approved by their respective institutional review boards.

Study population and sources of data

We used epidemiological and genetic data from 10 226 CRC cases and 10 286 population-based controls of European ancestry from 11 studies (6 cohort studies and 5 case-control studies) participating in the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and Colon Cancer Family Registry (C-CFR) (Supplementary Table 1, available as Supplementary data at *IJE* online). Full details on the consortium (GECCO + C-CFR) have been published elsewhere.[18] The 11 studies included in our analysis were the Health Professionals Follow-up Study (HPFS);[19] Nurses' Health Study (NHS);[20] Physician's Health Study (PHS);[21] Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO);[22] VITamins And Lifestyle Study (VITAL);[23] Women's Health Initiative (WHI);[24] Colon-Cancer Family Registry (C-CFR);[25] Ontario Familial Colon Cancer Registries (OFCCR);[26] Diet, Activity and Lifestyle Survey (DALS);[27,28] Postmenopausal Hormone Study (PMH-CCFR);[29] and Darmkrebs: Chancen der Verhütungdurch Screening (DACHS).[30] There was no overlap of participants between the 11 studies.

Outcomes

CRC cases were men and women with histologically confirmed (in each study by medical records, pathological reports or death certificates) invasive adenocarcinoma of the colon or rectum (International Classification of Disease Code, 9th revision: 153-154). We calculated risk estimates associated with height for CRC overall and separately for colon and rectal cancers.

Genotyping and the instrumental variable for height

Full details on genotyping, quality assurance/quality control and imputation have been previously reported.[18] In brief, to avoid confounding by population stratification, we used principal components analysis to restrict our analyses to individuals of European ancestry.[31] Genotyped single nucleotide polymorphisms (SNPs) were excluded based on call rate (<98%), lack of Hardy–Weinberg equilibrium in controls ($P < 1 \times 10^{-4}$) and low minor allele frequency (MAF $\leq 1\%$). Because imputation of genotypes is established as standard practice in the analysis of genotype array

data, we imputed the autosomal SNPs of all studies to the Utah residents with Northern and Western European ancestry from the Centre d'Etude du Polymorphisme Humain (CEPH) collection (CEU) population in HapMap II. Imputed SNPs were restricted based on MAF and overall imputation accuracy ($R^2 > 0.3$).

A recent genome-wide association study (GWAS) of height conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium identified 697 genetic variants associated with height at genome-wide significance ($P < 5 \times 10^{-8}$).[32] We excluded one variant associated with height in Ref #32 due to low imputation quality (rs17499117, mean $R^2 = 0.27$, range 0.14–0.36).

We created a weighted genetic risk score predicting height by summing the number of height-increasing alleles in each person across the 696 variants. For each variant, we assigned participants a value of 0, 1 or 2 for carrying zero (wild-type homozygous), one (heterozygous) or two (homozygous for the risk allele) alleles associated with greater height. When a variant was imputed, each participant was assigned a value between 0 and 2. Each variant in the genetic risk score was weighted by the per-allele change in height (the increase in cm per one additional risk allele) reported in the GIANT GWAS (i.e. we used external weights in the risk score).[32]

Covariates

The individual studies collected information on demographic and lifestyle factors through either in-person interviews or structured self-completed health and lifestyle questionnaires. The data harmonization process for the consortium has been described in full previously.[33]

Statistical analysis

We examined associations between height and risks of CRC overall, and its two components separately (colon cancer and rectal cancer) for women and men combined, and then separately for women and men. All statistical analyses were performed using Stata 13.0 (StataCorp LP, College Station, TX, USA).

First, we assessed associations with height (continuous and in sex-specific quartiles) using a two-step meta-analytic approach. In the first stage, study-specific odds ratios (OR) and corresponding 95% confidence intervals (95% CI) were estimated using logistic regression models, adjusted for age (continuous), sex, family history of CRC (no, yes), history of diabetes (no, yes), smoking status (never, former, current), aspirin or non-aspirin non-steroidal anti-inflammatory drugs (NSAID) use (no, yes), consumption of vegetables (sex-specific quartiles), consumption of red meat (sex-specific quartiles) and menopause hormone therapy (Women only; no, yes). To estimate the independent effect for height, instead of adjusting for weight or body mass index (BMI) which are associated with the outcome and are appreciably correlated with height, we adjusted for a weight-for-height variable (W/H^x) for such values of x that W/H^x was highly correlated with weight (Pearson correlation coefficient, r, close to 1) but not correlated with height (r close to 0). In our population, the optimal factor was $x = 1.8$ for all, $x = 1.5$ for women and $x = 1.7$ for men. We considered weight, smoking status, family history of cancer, diabetes, use of aspirin or NSAIDs, fruit, vegetable, processed meat and red meat consumption, sedentary lifestyle and menopause hormone therapy as potential confounding variables. Variables were retained in the final fully adjusted models if they were associated with CRC risk after mutual adjustment for other risk factors ($P < 0.05$). In the second step, the study-specific adjusted ORs were pooled to create a summary OR, using random effects models.[34] We used the I^2 statistic to assess the heterogeneity between studies.[35] I^2 values of 25%, 50% and 75% were used as evidence of low, moderate or high levels of heterogeneity, respectively.

Using controls only (representing the population from which the cases arose), we then examined the association between the weighted genetic risk score and height in each study and meta-analysed the results. In the first step, study-specific regression coefficients and corresponding 95% CIs were estimated using linear regression models, adjusted for age and the first three principal components that reflected the population structure to control for population stratification. In the second step, the study-specific regression coefficients were pooled to create a summary coefficient using random effects models. The F-statistic and R^2 for the regression of height on genetically predicted height (i.e., the weighted genetic risk score) were obtained; an F-statistic > 10 suggests the genetic instrument is associated with the exposure and is unlikely to suffer from weak instrument bias.[36]

Third, we examined whether genetically determined height was directly associated with risk of CRC, colon cancer and rectal cancer. We used the same two-step meta-analytic approach: first, using logistic regression models adjusted by age and the first three principal components to estimate study-specific ORs for the associations between genetically determined height and CRC, colon cancer and rectal cancer; and second, meta-analysing the results (using random effects models) to generate summary ORs.

Finally, we used linear regression to estimate the associations of genetically determined height and height with a set of potential confounders.

Results

Characteristics of cases and controls included in this analysis are presented in Supplementary Tables 1 and 2 (available as Supplementary data at *IJE* online). All potential confounders were associated with height and/or with risk of CRC (Supplementary Tables 2 and 3, available as Supplementary data at *IJE* online). The weighted genetic risk score was largely not associated with factors that may confound the observational association between height and risk of CRC. Dietary factors were somewhat associated with the weighted genetic risk score, but associations were not consistent across categories and tended to vary in sex-specific analyses (Supplementary Table 3).

Height, CRC, colon cancer and rectal cancer

In a conventional covariate-adjusted analysis (Table 1 and Figure 1A), a 10-cm increase in height was associated with an 8% increase in the risk of CRC (95% CI = 1.02–1.15), with moderate between-study heterogeneity ($I^2 = 45\%$). The point estimate was larger when we examined the association in studies involving measured height (OR per 10-cm increase in height = 1.12, 95% CI = 1.03–1.23) vs studies involving self-reported height (OR per 10-cm increase in height = 1.07, 95% CI = 1.00–1.16), although the CIs overlapped. The observational association with height did not vary according to cancer sub-site in the colorectum (Table 1). However, the observational associations differed by sex. Whereas height was associated with risks of CRC, colon cancer and rectal cancer for women in conventional covariate-adjusted analyses, there were no associations with height for men (Table 1).

Instrumental variable and height

When we regressed height on the weighted genetic risk score, individuals carrying greater numbers of (weighted) height-increasing alleles had higher attained adult height. For example, persons in the highest quartile of the weighted genetic risk score were 5.89 cm taller on average than persons in the lowest quartile. The association was homogeneous across studies (Figure 1B). The weighted genetic risk scores were strong instrumental variables for height (F-statistics: all, adjusted for sex, partial F-statistic = 1544; women, 916.4; men, 635.5) and the weighted genetic risk score explained 13.3% (partial R^2), 13.8% and 12.3% of the variation in height in all persons, women and men, respectively.

Instrumental variable, CRC, colon cancer and rectal cancer

We found evidence that persons carrying greater numbers of (weighted) height-increasing alleles were at greater risk for CRC (OR per 1-unit increase in the weighted genetic risk score = 1.07, 95% CI 1.01-1.14; Table 2 and Figure 1C). Corresponding ORs were 1.05 (95% CI = 0.98–1.12) for colon cancer, and 1.09 (95% CI = 0.98–1.20) for rectal cancer. Consistent with the conventional covariate-adjusted analysis (Table 1), women carrying greater numbers of (weighted) height-increasing alleles were at greater risk for CRC (OR per 1-unit increase in the weighted genetic risk score = 1.09, 95% CI 1.01–1.19). In contrast to the conventional covariate-adjusted analysis, we found some evidence that men carrying greater numbers of (weighted) height-increasing alleles have greater risk for CRC (OR per 1-unit increase in the weighted genetic risk score = 1.05, 95% CI 0.96–1.15). The point estimate for men was of similar order and in the same direction as that seen for women. We observed similar findings when we examined separately colon and rectal cancers (Table 2). Additionally adjusting for red and processed meats (and other CRC risk factors) did not change the ORs for CRC, colon cancer or rectal cancer associated with the weighted genetic risk score.

Discussion

In this large population-based study, we used a Mendelian randomization approach and genetically predicted height to examine for the presence of a causal association between height and CRC. Due to the study design, we did not however estimate the magnitude of this causal effect. The primary finding of this study was that greater height is causally associated with increased risk of CRC. The evidence was however somewhat weaker for men than for women. In the analyses involving men and women combined and those involving women only, the associations with height and genetically predicted height were consistent and demonstrated a strong causal association between height and CRC. For men, the conventional covariate-adjusted and instrumental variables analyses found little or weak evidence of an effect. However, because the point estimates and confidence intervals for men associated with genetically predicted height were similar to those found for women, it is possible that we did not find strong evidence of a causal effect for men simply due to chance or lack of statistical power.

Observational studies have consistently reported higher risk of CRC for taller women. Conversely, an association between height and the risk of CRC for men has been observed in some, but not all, studies and the magnitude of this association is generally lower than that for women.[2–12] Consistent with these observational associations, in our conventional covariate-adjusted analysis of height and the risk of CRC, we found that greater height

Table 1. Summary odds ratios (OR) and 95% confidence intervals (CI) for the association between height and risk of colorectal cancer (CRC), colon cancer and rectal cancer[a]

Site	All					Women					Men				
	Cases/ Controls	OR[b] (95% CI)	I^2	OR[c] (95% CI)	I^2	Cases/ Controls	OR[b] (95% CI)	I^2	OR[c] (95% CI)	I^2	Cases/ Controls	OR[b] (95% CI)	I^2	OR[c] (95% CI)	I^2
Colorectal cancer															
Q1	2652/2852	1.00 (Ref)		1.00 (Ref)		1338/1528	1.00 (Ref)		1.00 (Ref)		1314/1324	1.00 (Ref)		1.00 (Ref)	
Q2	2399/2351	1.13 (1.00–1.28)	52%	1.15 (1.00–1.32)	55%	1161/1126	1.17 (1.03–1.33)	13%	1.19 (1.04–1.36)	15%	1238/1225	1.04 (0.88–1.24)	50%	1.05 (0.89–1.25)	43%
Q3	2475/2582	1.07 (0.94–1.23)	58%	1.08 (0.96–1.22)	46%	1570/1657	1.10 (0.97–1.26)	29%	1.14 (0.99–1.32)	36%	905/925	0.98 (0.80–1.21)	55%	0.94 (0.80–1.10)	17%
Q4	2700/2501	1.19 (1.05–1.36)	57%	1.19 (1.03–1.36)	58%	1575/1436	1.27 (1.07–1.51)	55%	1.29 (1.06–1.56)	59%	1125/1065	1.07 (0.89–1.28)	49%	1.03 (0.86–1.24)	45%
Per 10 cm	10226/10286	1.08 (1.01–1.15)	50%	1.08 (1.02–1.15)	45%	5644/5747	1.13 (1.04–1.24)	47%	1.15 (1.05–1.26)	48%	4582/4539	0.99 (0.93–1.06)	0%	0.98 (0.92–1.05)	0%
Colon cancer															
Q1	1800/2852	1.00 (Ref)		1.00 (Ref)		948/1528	1.00 (Ref)		1.00 (Ref)		852/1324	1.00 (Ref)		1.00 (Ref)	
Q2	1654/2351	1.14 (1.00–1.31)	48%	1.15 (0.99–1.33)	51%	861/1126	1.17 (1.03–1.33)	0%	1.20 (1.05–1.36)	0%	793/1225	1.07 (0.86–1.34)	60%	1.08 (0.85–1.37)	61%
Q3	1675/2582	1.09 (0.95–1.24)	46%	1.08 (0.95–1.23)	36%	1095/1657	1.12 (0.97–1.29)	23%	1.13 (0.97–1.33)	33%	580/925	1.01 (0.82–1.25)	44%	0.94 (0.80–1.10)	3%
Q4	1735/2501	1.20 (1.06–1.35)	38%	1.19 (1.03–1.37)	48%	1018/1436	1.25 (1.05–1.50)	47%	1.27 (1.03–1.57)	56%	717/1065	1.11 (0.88–1.38)	56%	1.04 (0.81–1.33)	58%
Per 10 cm	6864/10286	1.08 (1.02–1.15)	29%	1.08 (1.02–1.15)	28%	3922/5747	1.13 (1.03–1.24)	46%	1.14 (1.03–1.27)	53%	2942/4539	1.01 (0.94–1.09)	5%	0.99 (0.92–1.07)	0%
Rectal cancer															
Q1	622/2852	1.00 (Ref)		1.00 (Ref)		257/1528	1.00 (Ref)		1.00 (Ref)		365/1324	1.00 (Ref)		1.00 (Ref)	
Q2	547/2351	1.08 (0.92–1.27)	15%	1.14 (0.93–1.39)	30%	209/1126	1.11 (0.87–1.43)	18%	1.10 (0.85–1.44)	15%	338/1225	1.06 (0.88–1.27)	0%	1.07 (0.88–1.30)	0%
Q3	585/2582	1.06 (0.87–1.30)	39%	1.13 (0.92–1.39)	36%	314/1657	1.09 (0.88–1.36)	16%	1.14 (0.93–1.40)	0%	271/925	0.92 (0.69–1.22)	32%	0.96 (0.71–1.30)	26%
Q4	611/2501	1.23 (1.03–1.47)	29%	1.26 (1.06–1.49)	16%	332/1436	1.32 (1.04–1.68)	26%	1.33 (1.05–1.68)	15%	279/1065	1.08 (0.81–1.43)	38%	1.15 (0.85–1.54)	35%
Per 10 cm	2365/10286	1.06 (0.99–1.14)	2%	1.08 (1.00–1.17)	0%	1112/5747	1.12 (1.00–1.24)	3%	1.14 (1.01–1.27)	0%	1253/4539	1.03 (0.90–1.18)	33%	1.07 (0.92–1.25)	41%

Summary odds ratios were estimated using a random-effects meta-analytic model.

Quartile definitions: Women (≤158 cm, 158.1–162.5 cm, 162.6–167.6 cm, > 167.6 cm); men (< 173 cm, 173 -< 178 cm, 178 -≤ 182 cm, > 182 cm).

[a]Not all cases from CRC analyses were included in the site-specific analyses (where we did not have site data for analysis).

[b]Adjusted for age (continuous) and sex (all only).

[c]Adjusted for age (continuous), sex (all only), family history, diabetes, smoking status, aspirin/NSAID use, consumption of vegetables, consumption of red meat, menopause hormone therapy (women only) and weight/height^x; where $x = 1.8$ for all, $x = 1.5$ for women and $x = 1.7$ for men.

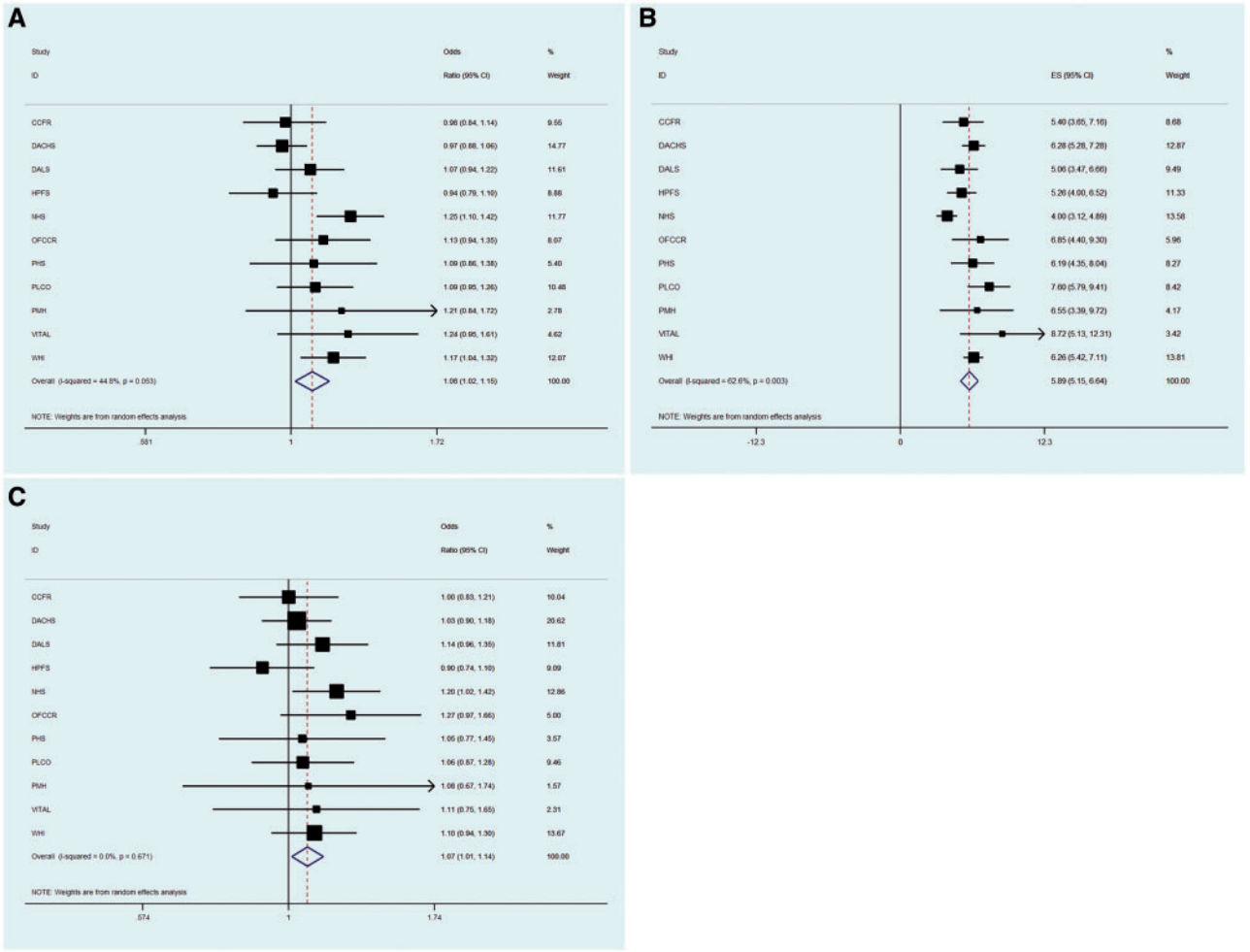

Figure 1. (A) Summary odds ratio for the association between a 10-cm increase in height and the risk of CRC. (B) Summary estimate comparing Q4 with Q1 for the association between the weighted genetic risk score (quartiles) and height (continuous). (C) Summary odds ratio for the association between a one-unit increase in the weighted genetic risk score and the risk of CRC.

was associated with increased risk of CRC for women. Height was not associated with CRC risk for men in this study. It is possible that selective over-reporting of height in men[13] may explain the lack of association with CRC risk in the conventional analyses. We found some evidence to support this hypothesis with the observational association greater in magnitude in studies involving measured height compared with those involving self-reported height. It is worth noting that whereas there was modest heterogeneity between the studies (both measured and self-reported height studies) for the observational association among women ($I^2 = 48\%$), there was no evidence of heterogeneity in the analyses among men ($I^2 = 0\%$); however, only one of eight studies in the men-only analysis included measured height. The same patterns of association were seen for risk of colon and rectal cancers.

Using instrumental variable methods, genetically predicted height was associated with risk of CRC for women. This novel finding provides strong support for a causal association between height and CRC risk for women; however, the magnitude of this effect could not be estimated in this study. The association between height and CRC risk for men remains unclear. Our study provides little or weak evidence for a causal association between height and CRC risk for men. However, because the point estimate for men associated with genetically predicted height was of similar order and in the same direction as that seen for women, we cannot completely rule out a causal association between height and CRC for men.

Several mechanisms by which greater height may confer higher risk of cancer have been proposed, and include higher exposure to steroid hormones and other growth factors, leading to higher cell turnover and greater risk of malignant transformation.[37] However, assuming height is causally associated with CRC risk for women and not for men, the mechanisms through which height would confer an effect on CRC risk for women only are unknown. Interestingly, although the relationships between adiposity,

Table 2. Summary odds ratios (OR) and 95% confidence intervals (CI) for the association between a 1-unit increase in the weighted genetic risk score (the instrumental variable for height) and risk of colorectal cancer (CRC), colon cancer and rectal cancer[a]

Site	All			Women			Men		
	Cases/ Controls	OR[b] (95% CI)	I^2	Cases/ Controls	OR[b] (95% CI)	I^2	Cases/ Controls	OR[b] (95% CI)	I^2
Colorectal cancer	10226/10286	1.07 (1.01–1.14)	0%	5644/5747	1.09 (1.01–1.19)	0%	4582/4539	1.05 (0.96–1.15)	0%
Colon cancer	6864/10286	1.05 (0.98–1.12)	0%	3922/5747	1.04 (0.95–1.14)	0%	2942/4539	1.06 (0.96–1.17)	0%
Rectal cancer	2365/10286	1.09 (0.98–1.20)	0%	1112/5747	1.05 (0.90–1.22)	0%	1253/4539	1.13 (0.98–1.31)	0%

Summary odds ratios were estimated using a random effects meta-analytical model.

[a]Not all cases from CRC analyses were included in the site-specific analyses (i.e. where we did not have site data for analysis).

[b]Adjusted for age (continuous) and the top three principal components of ancestry.

sex and risk of CRC are complex and unresolved, it has been suggested in some studies that BMI and weight gain are more strongly associated with CRC risk for men than for women.[38] Assuming this to be true, we can speculate that the contrast between men and women for height and adiposity may suggest different susceptibility periods to cancer development due to positive energy balance. For example, it is possible that positive energy balance early in life, reflected by attained adult height, confers higher risk of CRC for women, whereas positive energy balance later in life, as reflected by higher BMI or greater weight gain, confers higher risk of CRC for men. Sex is clearly a strong modifier for the influence of body size on CRC risk and requires further investigation in future studies.

The large sample size and homogeneous study population are strengths of the present study. We were able to adjust our analyses for a standard set of harmonized variables that may confound observational associations between height and CRC risk, and we found little evidence of between-study heterogeneity. In the instrumental variable analysis, we used multiple genetic variants to examine the complex relationship between height and CRC risk. Because genotype is randomly allocated at conception, associations between genetically determined height and CRC risk are unlikely to be biased, confounded or affected by reverse causality. Additionally, the association between genetically determined height and CRC risk is likely to reflect the effect of lifelong exposure to greater height.

There are limitations to our study. Because our study was restricted to participants of European descent, our results may not necessarily apply to other races. However, this fact also minimizes the risk of population stratification affecting the results of our instrumental variable analyses. Mendelian randomization relies on three key assumptions. First, that the instrumental variable is associated with height. We derived a weighted genetic risk score predicting height based on 696 genetic variants associated with height. The weighted genetic risk score model compared large numbers of people with large differences in genetically influenced height (e.g. almost 6 cm difference in height between persons in the highest vs lowest quartile of the score) and was a strong instrument for height (F-statistics $\gg 10$), reducing the chance of weak instrument bias. However, as the phenotypic variation in height explained by the weighted genetic risk score was less than 14%, we did not test the full spectrum of genetically influenced height. The second assumption is that the instrumental variable is not associated with potential confounders of the conventional height-CRC association. The association between the genetic risk score and red and processed meats was unexpected a priori, and may be due to chance. Nonetheless, inclusion of these factors in the analyses of genetically determined height and CRC did not change the risk estimates. Finally, the third assumption, that no other pathway exists between the instrumental variable and CRC risk, is difficult to prove, and may be violated by pleiotropy [e.g. through a joint mechanism, such as the insulin-like growth factor (IGF) axis]. It is possible that a number of the individual genetic variants used to derive the weighted genetic risk score are associated with height as well as other factors if height is causally associated with these secondary traits (i.e. vertical pleiotropy).[39] If the weighted genetic risk score is associated with CRC through biological factors related to height (e.g. the indirect effect, mediated through a secondary trait) rather than directly through height, this would still be the same pathway by which the marker height affects CRC risk. Furthermore, assuming the only pathway by which the variants are associated with secondary traits is through height, our findings hold in spite of the apparent pleiotropic association. A variant may also be associated with multiple pathways (horizontal pleiotropy), including those not involving height.[39] To the best of our knowledge, none of the 696 height variants used here have overly potent pleiotropic

effects and none was associated with CRC in recent GWASs of CRC.[18] Therefore, it seems unlikely that a substantial fraction of pleiotropic SNPs in the cumulative risk score would explain the associations with height.

In conclusion, height was causally associated with CRC risk; however, the sex-specific effects of height on CRC risk require further investigation. This is one of the first studies to use Mendelian randomization to reinvestigate the observed association between height and cancer risk. Our results show the value of instrumental variable methods to assess the existence of a causal association between height and cancer risk.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

This work is supported by grants to GECCO and CCFR, as well as by grants to the individual studies (as noted below). The funders of the study had no role in the design, analysis or interpretation of the data, or in writing or publication decisions related to this article.

GECCO: National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088; R01 CA059045).

CCFR: National Institutes of Health (RFA # CA-95-011) and through cooperative agreements with members of the Colon Cancer Family Registry and principal investigators. This genome-wide scan was supported by the National Cancer Institute, National Institutes of Health by U01 CA122839. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centres in the CFRs, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government or the CFR. The following Colon CFR centres contributed data to this manuscript and were supported by National Institutes of Health: Australasian Colorectal Cancer Family Registry (U01 CA097735), Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783) and the Seattle Colorectal Cancer Family Registry (U01 CA074794).

DACHS: German Research Council (Deutsche Forschungsgemeinschaft, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4 and CH 117/1-1), and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814).

DALS: National Institutes of Health (R01 CA48998 to M.L.S).

H.P.F.S. is supported by the National Institutes of Health (P01 CA 055075, UM1 CA167552, R01 137178, R01 CA 151993 and P50 CA 127003), N.H.S. by the National Institutes of Health (R01 CA137178, P01 CA 087969, R01 CA151993 and P50 CA 127003,) and P.H.S. by the National Institutes of Health (R01 CA042182).

OFCCR: National Institutes of Health, through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783); see CCFR section above. Additional funding toward genetic analyses of OFCCR includes the Ontario Research Fund, the Canadian Institutes of Health Research and the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation.

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Additionally, a subset of control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS (Yeager M *et al. Nat Genet* 2007;**39:**645–49), Colon CGEMS pancreatic cancer scan (PanScan) (Amundadottir *L et al. Nat Genet* 2009;**41:**986–90 and Petersen GM *et al. Nat Genet* 2010;**42:**224–28), and the Lung Cancer and Smoking study. The prostate and PanScan study datasets were accessed with appropriate approval through the dbGaP online resource (http://cgems.cancer.gov/data/) accession numbers phs000207.v1.p1 and phs000206.v3.p2, respectively, and the lung datasets were accessed from the dbGaP website (http://www.ncbi.nlm.nih.gov/gap) through accession number phs000093.v2.p2. Funding for the Lung Cancer and Smoking study was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438. For the lung study, the GENEVA Coordinating Center provided assistance with genotype cleaning and general study coordination, and the Johns Hopkins University Center for Inherited Disease Research conducted genotyping.

PMH: National Institutes of Health (R01 CA076366 to P.A.N.).

VITAL: National Institutes of Health (K05 CA154337).

WHI: The WHI programme is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN 268201100046C, HHSN268201100001C, HHSN2682011000 02C, HHSN268201100003C, HHSN268201100004C and HHS N271201100004C.

Acknowledgements

DACHS: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Renate Hettler-Jensen, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent technical assistance.

GECCO: The authors would like to thank all those at the GECCO Coordinating Center for helping bring together the data and people that made this project possible.

HPFS, NHS and PHS: We would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core, who assisted in the genotyping for NHS, HPFS and PHS under the supervision of Immaculata Devivo and David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for the PHS. We would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-Up Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. In addition, this study was approved by the Connecticut Department of Public Health (DPH) Human Investigations Committee. Certain data used in this publication were obtained from the DPH. The authors assume full responsibility for analyses and interpretation of these data.

PLCO: The authors thank Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff or the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Tom Riley and staff, Information Management Services Inc., Barbara O'Brien

and staff, Westat Inc and Bill Kopp, Wen Shao and staff, SAIC-Frederick. Most importantly, we acknowledge the study participants for their contributions to making this study possible. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

PMH: The authors would like to thank the study participants and staff of the Hormones and Colon Cancer study.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the programme possible. A full listing of WHI investigators can be found at: [https://cleo.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf].

Related reference list

Landi M., Chatterjee N, Yu K *et al*. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am *J Hum Gene*t 2009;**85:**679–91.

Yeager M, Orr N, Hayes RB *et al*. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 2007;**39:**645–49.

Conflict of interest: None declared.

References

1. World Cancer Research Fund/American Institute for Cancer Research. *Food, Nutrition, Physical Activity, and the Prevention of Cnacer: A Global Perspective*. Washington, DC: Cancer Research Fund/American Institute for Cancer Research, 2007.
2. Hughes LA, Simons CC, van den Brandt PA, Goldbohm RA, van Engeland M, Weijenberg MP. Body size and colorectal cancer risk after 16.3 years of follow-up: an analysis from the Netherlands Cohort Study. *Am J Epidemiol* 2011;**174**: 1127–39.
3. Campbell PT, Cotterchio M, Dicks E, Parfrey P, Gallinger S, McLaughlin JR. Excess body weight and colorectal cancer risk in Canada: associations in subgroups of clinically defined familial risk of cancer. *Cancer Epidemiol Biomarkers Prev* 2007;**16**:1735–44.
4. Oxentenko AS, Bardia A, Vierkant RA *et al*. Body size and incident colorectal cancer: a prospective study of older women. *Cancer Prev Res (Phila)* 2010;**3**:1608–20.
5. Bowers K, Albanes D, Limburg P *et al*. A prospective study of anthropometric and clinical measurements associated with insulin resistance syndrome and colorectal cancer in male smokers. *Am J Epidemiol* 2006;**164**:652–64.
6. Pischon T, Lahmann PH, Boeing H *et al*. Body size and risk of colon and rectal cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *J Natl Cancer Inst* 2006;**98**:920–31.
7. Engeland A, Tretli S, Austad G, Bjorge T. Height and body mass index in relation to colorectal and gallbladder cancer in two million Norwegian men and women. *Cancer Causes Control* 2005; **16**:987–96.
8. Bostick RM, Potter JD, Kushi LH *et al*. Sugar, meat, and fat intake, and non-dietary risk factors for colon cancer incidence in Iowa women (United States). *Cancer Causes Control* 1994;**5**: 38–52.
9. Chute CG, Willett WC, Colditz GA *et al*. A prospective study of body mass, height, and smoking on the risk of colorectal cancer in women. *Cancer Causes Control* 1991;**2**:117–24.
10. Wei EK, Giovannucci E, Wu K *et al*. Comparison of risk factors for colon and rectal cancer. *Int J Cancer* 2004;**108**:433–42.
11. Shimizu N, Nagata C, Shimizu H *et al*. Height, weight, and alcohol consumption in relation to the risk of colorectal cancer in Japan: a prospective study. *Br J Cancer* 2003;**88**: 1038–43.
12. MacInnis RJ, English DR, Hopper JL, Haydon AM, Gertig DM, Giles GG. Body size and composition and colon cancer risk in men. *Cancer Epidemiol Biomarkers Prev* 2004;**13**:553–59.
13. Shields M, Connor Gorber S, Tremblay MS. Estimates of obesity based on self-report versus direct measures. *Health Rep* 2008; **19**:61–76.
14. Yang J, Benyamin B, McEvoy BP *et al*. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010;**42**:565–69.
15. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003;**32**:1–22.
16. 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**: 1133–63.
17. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;**29**:722–29.
18. Peters U, Jiao S, Schumacher FR *et al*. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology* 2013;**144**:799–807.
19. Giovannucci E, Rimm EB, Stampfer MJ *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**: 183–91.
20. Giovannucci E, Colditz GA, Stampfer MJ *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**: 192–99.
21. 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**:125–34.
22. Prorok PC, Andriole GL, Bresalier RS *et al*. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Control Clin Trials* 2000;**21**:273S–309S.
23. White E, Patterson RE, Kristal AR *et al*. VITamins And Lifestyle cohort study: study design and characteristics of supplement users. *Am J Epidemiol* 2004;**159**:83–93.
24. 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.
25. Newcomb PA, Baron J, Cotterchio M *et al*. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol Biomarkers Prev* 2007;**16**:2331–43.
26. Zanke BW, Greenwood CM, Rangrej J *et al*. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007;**39**:989–94.

27. 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**: 1666–70.
28. Slattery ML, Potter J, Caan B *et al*. Energy balance and colon cancer – beyond physical activity. *Cancer Res* 1997;**57**:75–80.
29. Newcomb PA, Zheng Y, Chia VM *et al*. Estrogen plus progestin use, microsatellite instability, and the risk of colorectal cancer in women. *Cancer Res* 2007;**67**:7534–39.
30. 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.
31. 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**:904–09.
32. Wood AR, Esko T, Yang J *et al*. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014;**46**:1173–86.
33. Hutter CM, Chang-Claude J, Slattery ML *et al*. Characterization of gene-environment interactions for colorectal cancer susceptibility loci. *Cancer Res* 2012;**72**:2036–44.
34. Smith-Warner SA, Spiegelman D, Ritz J *et al*. Methods for pooling results of epidemiologic studies: the Pooling Project of Prospective Studies of Diet and Cancer. *Am J Epidemiol* 2006; **163**:1053–64.
35. Higgins JP, Thompson SG, Deeks JJ *et al*. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557–60.
36. Staiger D, Stock JH. Instrumental variables regression with weak instruments. *Econometrica* 1997;**65**:557–86.
37. Albanes D, Winick M. Are cell number and cell proliferation risk factors for cancer? *J Natl Cancer Inst* 1988;**80**:772–74.
38. Bardou M, Barkun AN, Martel M. Obesity and colorectal cancer. *Gut* 2013;**62**:933–47.
39. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* 2015;**181**:251–60.