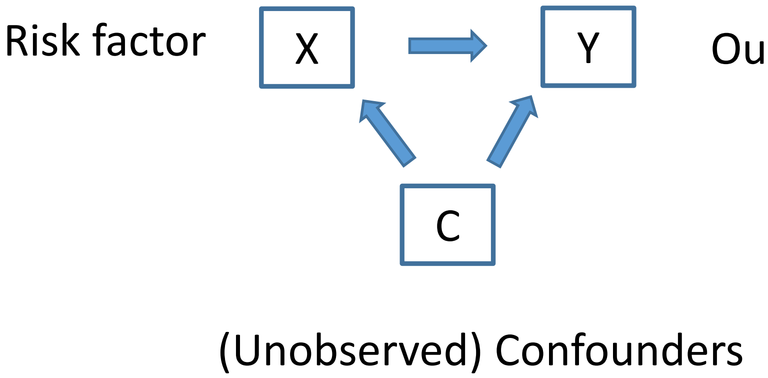
Mendelian Randomization

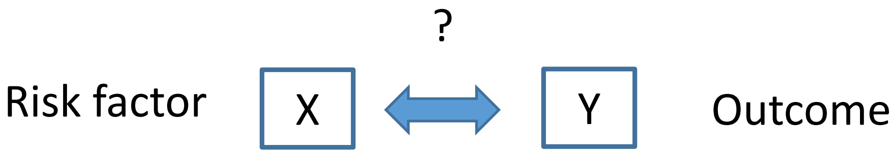
Drawback with observational studies



Risk factor X → Y Outcome

C

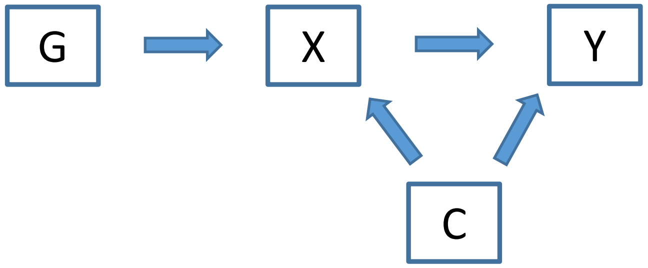
(Unobserved) Confounders



?

Risk factor X ⟷ Y Outcome

The power of genetics



Intermediate phenotype (risk factor)

Genetic variant (Instrumental Variable - IV)

G

X

Y

Outcome

C

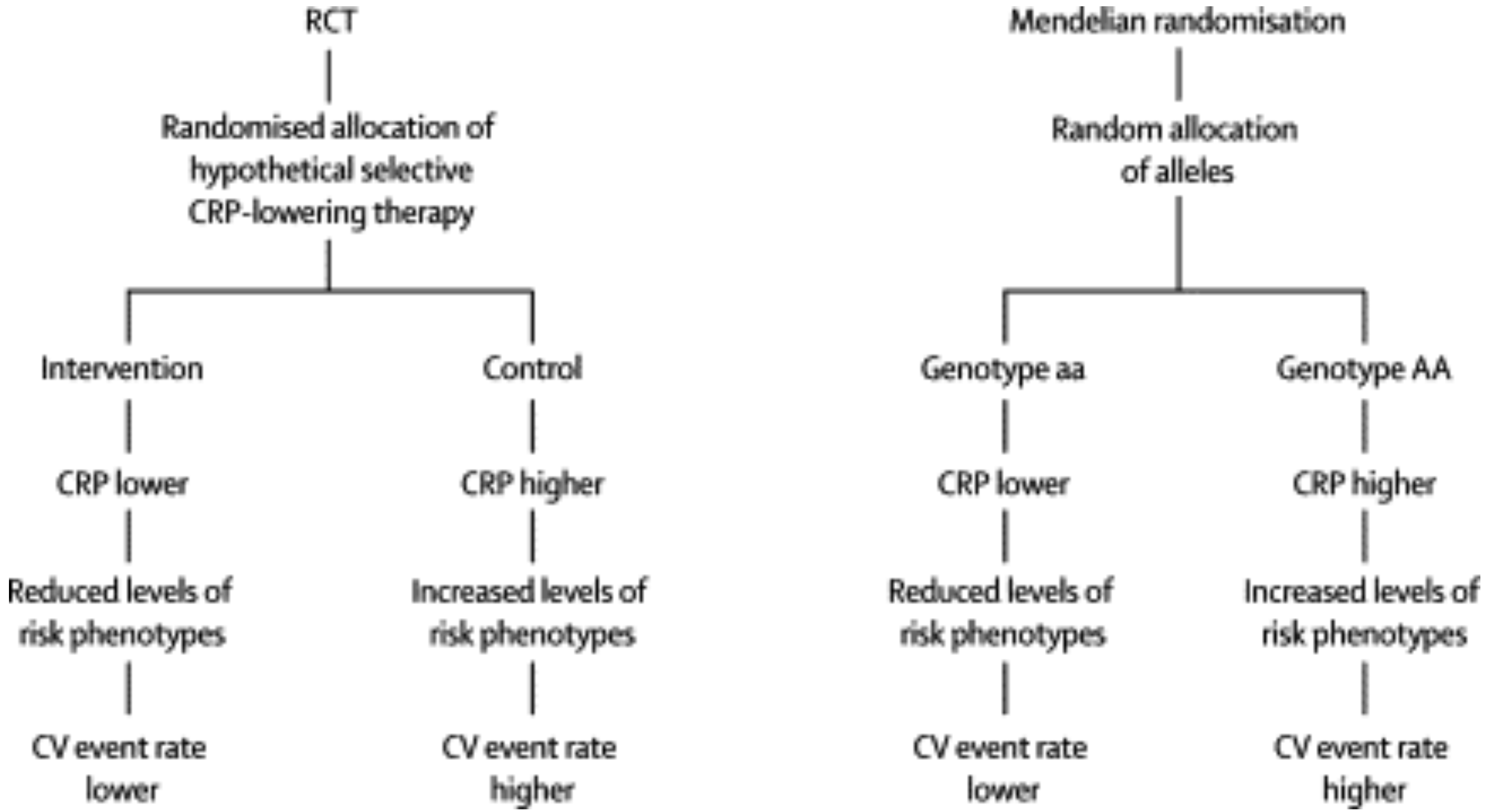
(Unobserved) Confounders

Mendelian Randomization

- *Basic principle: "genetic variants which mirror the biological effects of a modifiable environmental exposure and alters disease risk should be associated with disease risk to the extent predicted by their influence on exposure to the risk factor."*
- The random allocation of genetic variants from parents to offspring means these variants will generally be unrelated to other factors which affect the outcome.
- Furthermore, associations between the genotype and the outcome will not be affected by reverse causation because disease does not affect genotype

Ebrahim & Davey Smith, Hum Genet 2008
Davey Smith & Ebrahim, Int J Epi 2004

Possible effects of C-reactive protein (CRP) on cardiovascular (CV) events. Expected outcome from hypothetical randomized clinical trial of selective CRP-lowering intervention, and from Mendelian randomization analysis, if CRP were causal in developing CV.



RCT

Randomised allocation of hypothetical selective CRP-lowering therapy

Intervention

CRP lower

Reduced levels of risk phenotypes

CV event rate lower

Control

CRP higher

Increased levels of risk phenotypes

CV event rate higher

Mendelian randomisation

Random allocation of alleles

Genotype aa

CRP lower

Reduced levels of risk phenotypes

CV event rate lower

Genotype AA

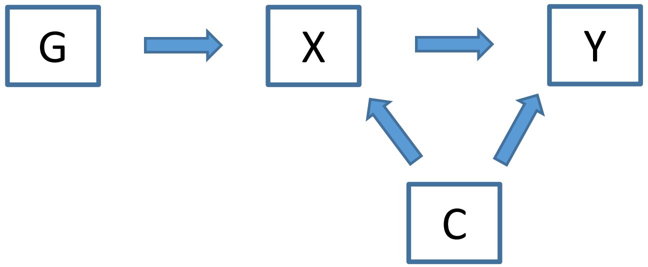
CRP higher

Increased levels of risk phenotypes

CV event rate higher

Hingorani & Humphries, Lancet 2005

Three key assumptions in MR analysis



1. G (SNP or a combination of multiple SNPs) is robustly associated with X (risk factor)

2. G is unrelated to any confounders C, that can bias the relationship between G and Y (outcome). In other words, there are no common causes of G and Y (e.g. population stratification)

3. G is related to Y only through its association with X (i.e. no pleiotropy)

Assumption 1: G is robustly associated with X

- Under certain conditions, the relative bias of the instrument variable (IV) estimate is ~1/F. A “weak” IV has been defined as having F<10, where

$$F = \frac{R^2(n-1-k)}{(1-R^2)k}$$

R^2 is variance in X explained by the IV(s), n is sample size and k is number of IVs

- Weak IVs can lead to biased effect estimates (in the direction of the observed X-Y association) in the presence of confounding of the X–Y relationship.

Pierce, IJE 2011

Assumption 2: No confounding

- G is independent of factors (measured and unmeasured) that confound the X-Y relation

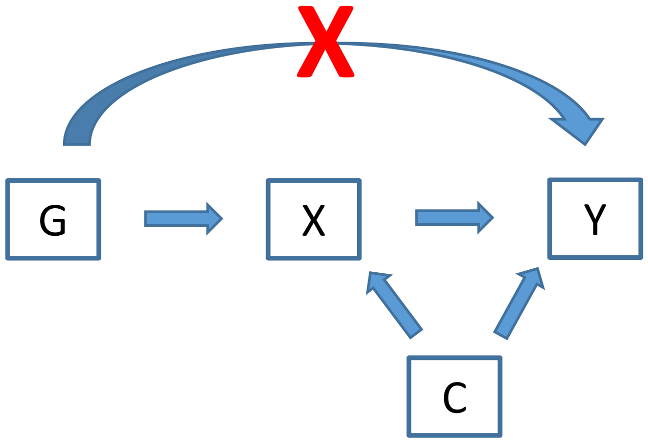
- Since G is randomized at birth and thus is independent of non-genetic confounders and is not modified by the course of disease, the one main concern here is population stratification – i.e. if ancestry is related both to G and Y.

- If you have individual-level data, you can test for this (e.g. PCA)

Assumption 3: No pleiotropy

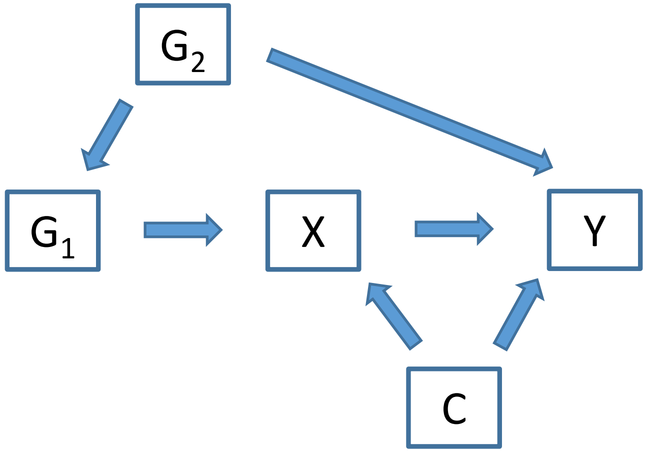
- This assumption is the trickiest

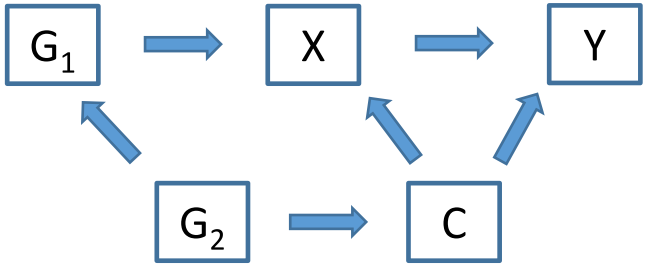
- Assumes that G is only associated with Y via X and thus the association between G and Y is fully mediated by X and not through any unmeasured factor(s). Needs to be true for SNPs in LD too



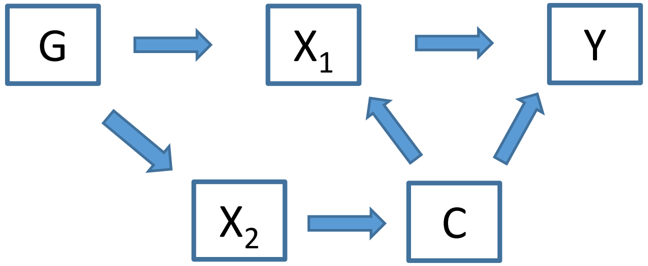
Scenarios invalidating assumption 3

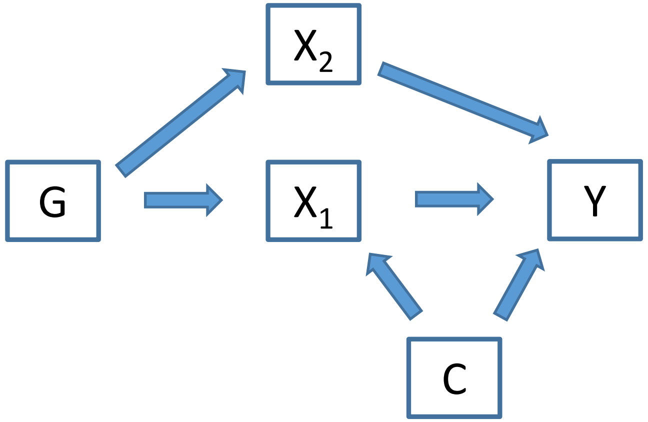
LD

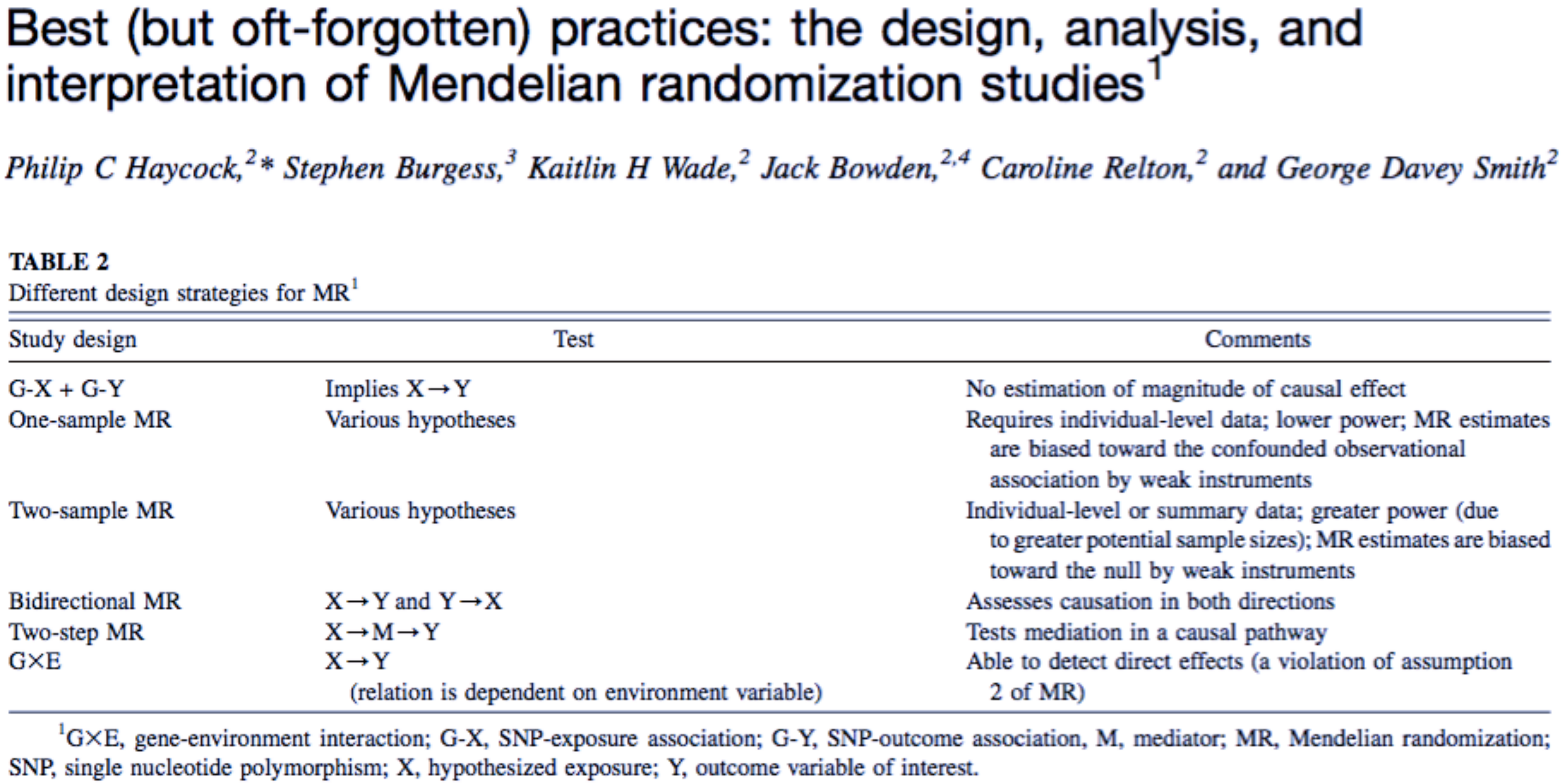




Pleiotropy







Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies[1]

Philip C Haycock,[2] Stephen Burgess,[3] Kaitlin H Wade,[2] Jack Bowden,[2,4] Caroline Relton,[2] and George Davey Smith[2]*

TABLE 2
Different design strategies for MR[1]

Study design	Test	Comments
G-X + G-Y	Implies X → Y	No estimation of magnitude of causal effect
One-sample MR	Various hypotheses	Requires individual-level data; lower power; MR estimates are biased toward the confounded observational association by weak instruments
Two-sample MR	Various hypotheses	Individual-level or summary data; greater power (due to greater potential sample sizes); MR estimates are biased toward the null by weak instruments
Bidirectional MR	X → Y and Y → X	Assesses causation in both directions
Two-step MR	X → M → Y	Tests mediation in a causal pathway
G×E	X → Y (relation is dependent on environment variable)	Able to detect direct effects (a violation of assumption 2 of MR)

[1] G×E, gene-environment interaction; G-X, SNP-exposure association; G-Y, SNP-outcome association, M, mediator; MR, Mendelian randomization; SNP, single nucleotide polymorphism; X, hypothesized exposure; Y, outcome variable of interest.

Haycock et al, Am J Clin Nutr 2016

Individual-level data in one sample

- Access to SNPs, risk factor, and outcome for all participants

- The causal effect of X on Y can be estimated using 2-stage least-squares (2SLS) regression:

1. X = $\mathrm{a} + \gamma G$
2. Y = $\mathrm{c} + \beta X^*$, where X^* are the genetically predicted exposure levels as measured in (1)

- The causal estimate is given by β
- Can be implemented in R using the “ivpack” package
- Weak instruments cause bias towards the observed confounded association

Summary data from two samples

- The G-X and the G-Y associations are estimated in two different samples.
- Assumes no overlap among samples and that the two populations are similar (ethnicity, age, sex, etc.)
- Here, bias due to weak IVs will be towards the null
- Note: The G-X and G-Y associations need to be coded using the same effect allele

Summary data from two samples

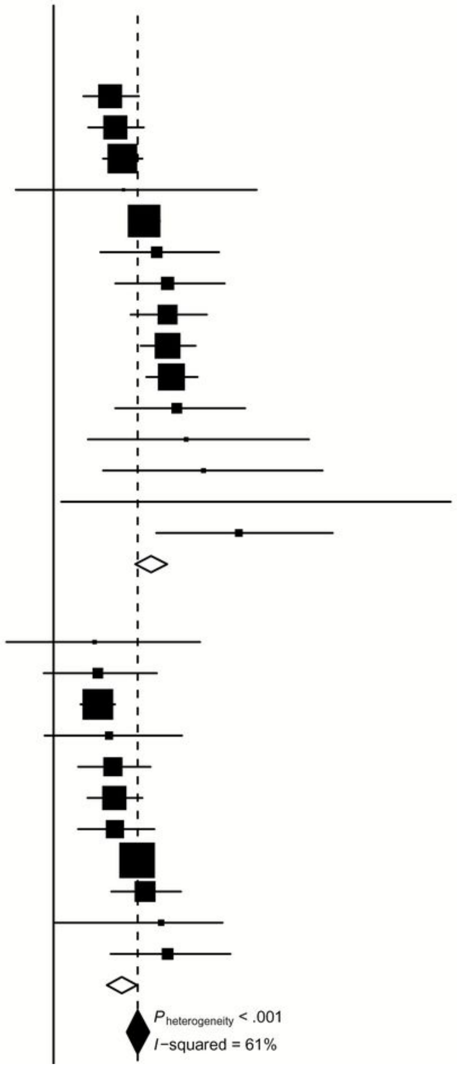
$$\hat{\beta} = \frac{\sum_k \beta_{1k}\beta_{2k}\sigma_{\beta_{2k}}^{-2}}{\sum_k \beta_{1k}\sigma_{\beta_{2k}}^{-2}}$$

$$se(\hat{\beta}) = \sqrt{\frac{1}{\sum_k \beta_{1k}^2\sigma_{\beta_{2k}}^{-2}}}$$

β_{1k} is the mean change in X per allele for SNP k, β_{2k} is the mean change in Y per allele for SNP k, σ_{2k}^{-2} is the inverse variance for the G-Y association.

Meta-analysis of associations between height and risk of breast cancer in prospective cohort studies.

Author (published year)	Total samples (N)	Events (n)	RR (95% CI) per 10 cm increase
Measured height			
Kabat (2013)	88 256	4 224	1.12 (1.06 to 1.17)
Wormser (2012)	522 257	3 926	1.13 (1.07 to 1.18)
Kabat (2013)	144 701	6 798	1.14 (1.10 to 1.18)
Kaaks (1998)	11 480	275	1.14 (0.92 to 1.41)
Ritte (2013)	306 600	9 307	1.18 (1.15 to 1.22)
Tornberg (1988)	46 570	1 182	1.21 (1.10 to 1.34)
Mellemkjaer (2012)	23 864	1 209	1.23 (1.13 to 1.35)
Ahlgren (2004)	117 415	3 340	1.23 (1.16 to 1.31)
Wiren (2014)	297 156	6 161	1.23 (1.18 to 1.29)
Tretli (1989)	567 333	8 427	1.24 (1.19 to 1.29)
Vatten (1992)	25 967	291	1.25 (1.13 to 1.39)
Macinnis (2004)	13 598	357	1.27 (1.07 to 1.52)
Tulinius (1997)	11 580	439	1.31 (1.10 to 1.55)
Freni (1996)	7 622	182	1.36 (1.02 to 1.81)
De Stavola (1993)	6 706	168	1.38 (1.21 to 1.57)
Subtotal	2 191 105	46 286	**1.20 (1.17 to 1.23)**
Self-reported height			
Sonnenschein (1999)	8 416	259	1.08 (0.90 to 1.30)
White (2012)	19 815	835	1.09 (0.98 to 1.21)
Kabat (2014)	192 514	9 169	1.09 (1.06 to 1.13)
Lacey (2009)	70 575	2 085	1.11 (0.98 to 1.26)
Canchola (2012)	52 642	2 321	1.12 (1.05 to 1.20)
van den Brandt (2000)	337 819	4 385	1.12 (1.07 to 1.18)
Michels−Blanck (1996)	428 653	2 226	1.12 (1.05 to 1.21)
Green (2011)	1 297 124	39 299	1.17 (1.15 to 1.19)
McCullough (2005)	409 093	4 265	1.19 (1.12 to 1.26)
Weiderpass (2004)	99 717	733	1.22 (1.00 to 1.34)
Baer (2006)	108 829	1 315	1.23 (1.12 to 1.36)
Subtotal	3 025 197	66 892	**1.14 (1.11 to 1.17)**
All studies combined	5 216 302	113 178	**1.17 (1.15 to 1.19)**



Ben Zhang et al. JNCI J Natl Cancer Inst 2015;107:djv219

JNCI

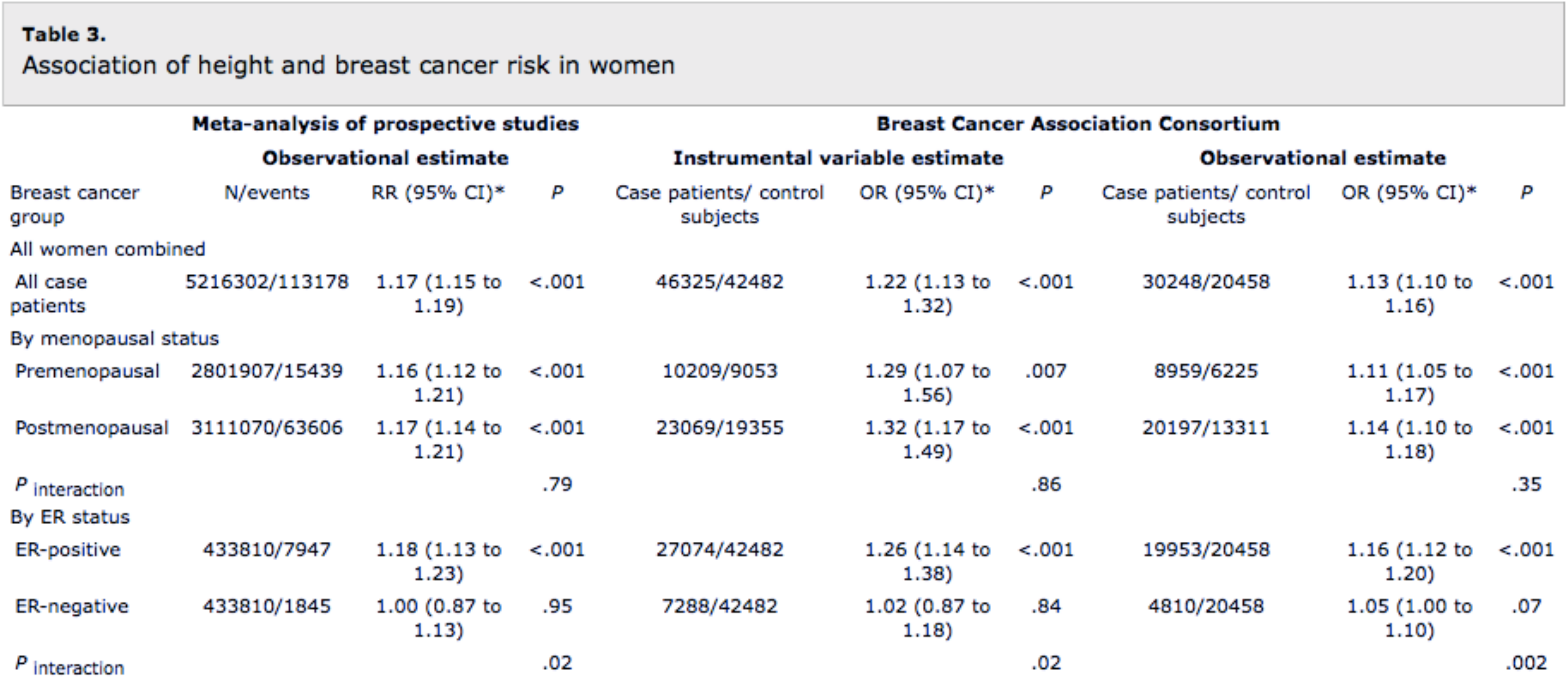


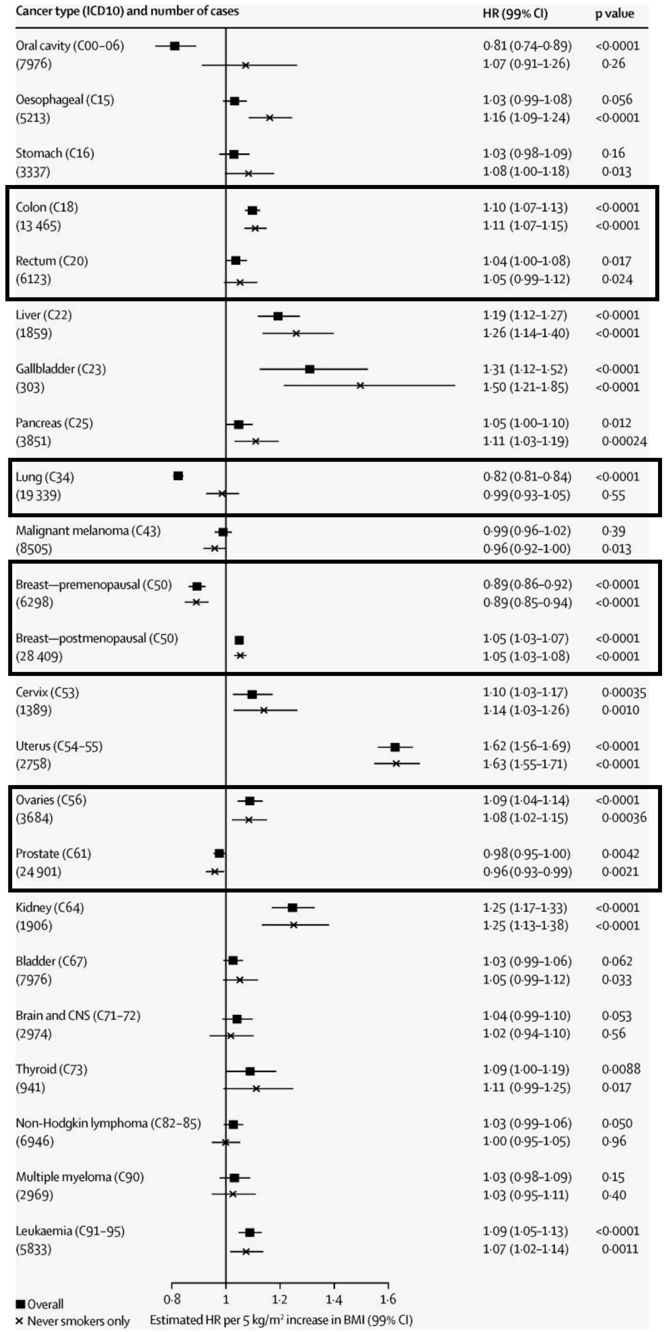
Table 3.

Association of height and breast cancer risk in women

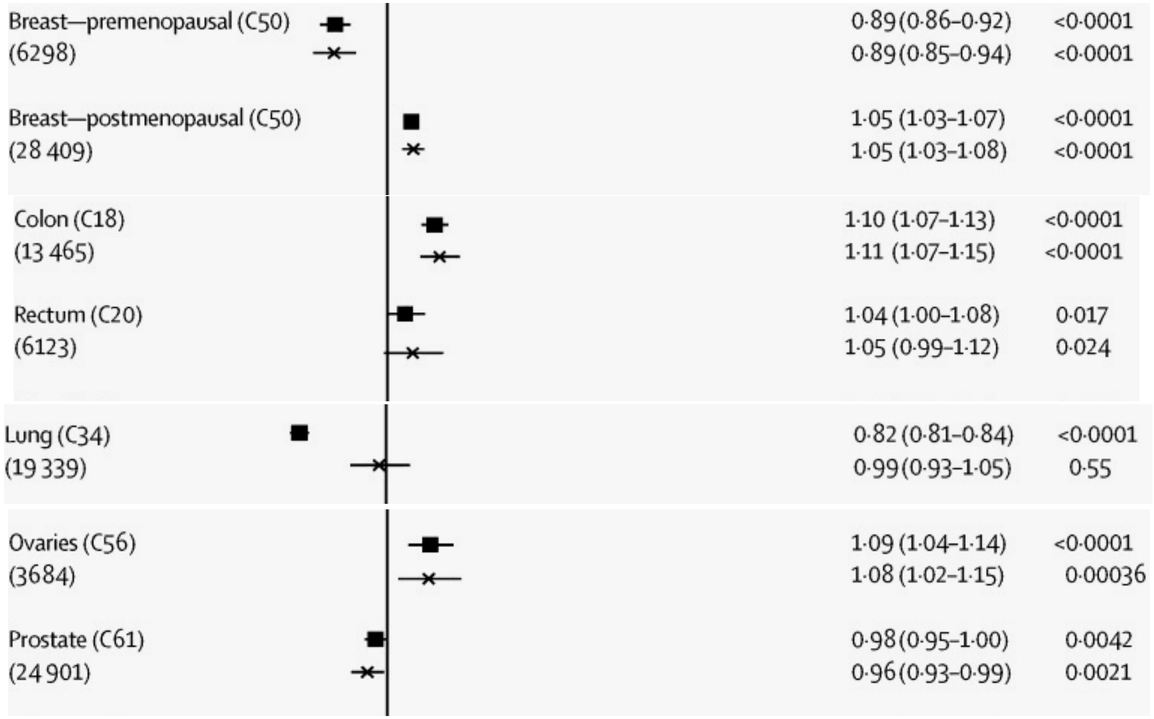
	Meta-analysis of prospective studies			**Breast Cancer Association Consortium**					
	Observational estimate			**Instrumental variable estimate**			**Observational estimate**		
Breast cancer group	N/events	RR (95% CI)*	*P*	Case patients/ control subjects	OR (95% CI)*	*P*	Case patients/ control subjects	OR (95% CI)*	*P*
All women combined									
All case patients	5216302/113178	1.17 (1.15 to 1.19)	<.001	46325/42482	1.22 (1.13 to 1.32)	<.001	30248/20458	1.13 (1.10 to 1.16)	<.001
By menopausal status									
Premenopausal	2801907/15439	1.16 (1.12 to 1.21)	<.001	10209/9053	1.29 (1.07 to 1.56)	.007	8959/6225	1.11 (1.05 to 1.17)	<.001
Postmenopausal	3111070/63606	1.17 (1.14 to 1.21)	<.001	23069/19355	1.32 (1.17 to 1.49)	<.001	20197/13311	1.14 (1.10 to 1.18)	<.001
P interaction			.79			.86			.35
By ER status									
ER-positive	433810/7947	1.18 (1.13 to 1.23)	<.001	27074/42482	1.26 (1.14 to 1.38)	<.001	19953/20458	1.16 (1.12 to 1.20)	<.001
ER-negative	433810/1845	1.00 (0.87 to 1.13)	.95	7288/42482	1.02 (0.87 to 1.18)	.84	4810/20458	1.05 (1.00 to 1.10)	.07
P interaction			.02			.02			.002

Ben Zhang et al. JNCI J Natl Cancer Inst 2015;107:djv219

- Association between BMI and cancer risk was assessed for 22 cancers
- 5.24 million individuals (166,996 cancer cases)



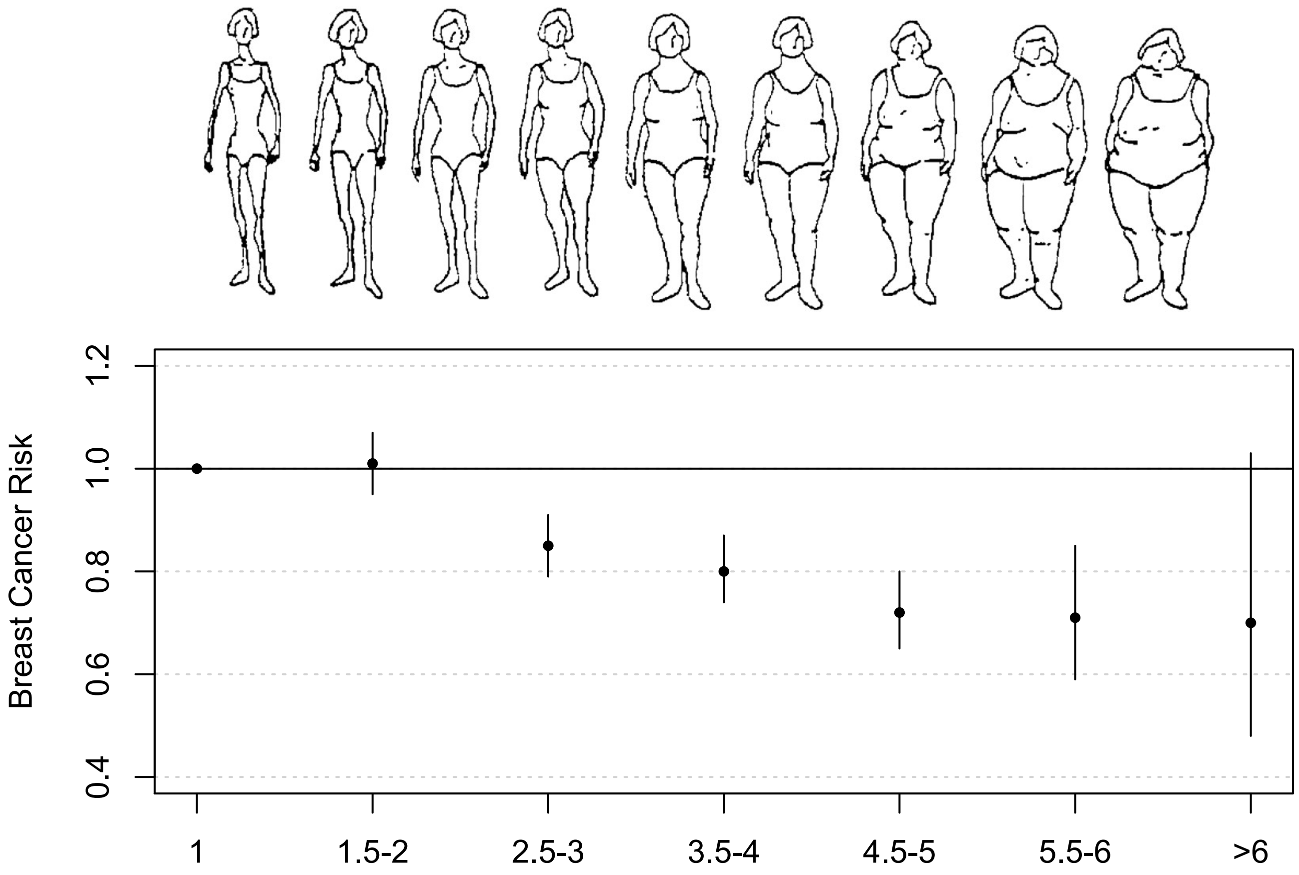
Cancer type (ICD10) and number of cases	HR (99% CI)	p value
Oral cavity (C00–06)	0·81 (0·74–0·89)	<0·0001
(7976)	1·07 (0·91–1·26)	0·26
Oesophageal (C15)	1·03 (0·99–1·08)	0·056
(5213)	1·16 (1·09–1·24)	<0·0001
Stomach (C16)	1·03 (0·98–1·09)	0·16
(3337)	1·08 (1·00–1·18)	0·013
Colon (C18)	1·10 (1·07–1·13)	<0·0001
(13 465)	1·11 (1·07–1·15)	<0·0001
Rectum (C20)	1·04 (1·00–1·08)	0·017
(6123)	1·05 (0·99–1·12)	0·024
Liver (C22)	1·19 (1·12–1·27)	<0·0001
(1859)	1·26 (1·14–1·40)	<0·0001
Gallbladder (C23)	1·31 (1·12–1·52)	<0·0001
(303)	1·50 (1·21–1·85)	<0·0001
Pancreas (C25)	1·05 (1·00–1·10)	0·012
(3851)	1·11 (1·03–1·19)	0·00024
Lung (C34)	0·82 (0·81–0·84)	<0·0001
(19 339)	0·99 (0·93–1·05)	0·55
Malignant melanoma (C43)	0·99 (0·96–1·02)	0·39
(8505)	0·96 (0·92–1·00)	0·013
Breast—premenopausal (C50)	0·89 (0·86–0·92)	<0·0001
(6298)	0·89 (0·85–0·94)	<0·0001
Breast—postmenopausal (C50)	1·05 (1·03–1·07)	<0·0001
(28 409)	1·05 (1·03–1·08)	<0·0001
Cervix (C53)	1·10 (1·03–1·17)	0·00035
(1389)	1·14 (1·03–1·26)	0·0010
Uterus (C54–55)	1·62 (1·56–1·69)	<0·0001
(2758)	1·63 (1·55–1·71)	<0·0001
Ovaries (C56)	1·09 (1·04–1·14)	<0·0001
(3684)	1·08 (1·02–1·15)	0·00036
Prostate (C61)	0·98 (0·95–1·00)	0·0042
(24 901)	0·96 (0·93–0·99)	0·0021
Kidney (C64)	1·25 (1·17–1·33)	<0·0001
(1906)	1·25 (1·13–1·38)	<0·0001
Bladder (C67)	1·03 (0·99–1·06)	0·062
(7976)	1·05 (0·99–1·12)	0·033
Brain and CNS (C71–72)	1·04 (0·99–1·10)	0·053
(2974)	1·02 (0·94–1·10)	0·56
Thyroid (C73)	1·09 (1·00–1·19)	0·0088
(941)	1·11 (0·99–1·25)	0·017
Non-Hodgkin lymphoma (C82–85)	1·03 (0·99–1·06)	0·050
(6946)	1·00 (0·95–1·05)	0·96
Multiple myeloma (C90)	1·03 (0·98–1·09)	0·15
(2969)	1·03 (0·95–1·11)	0·40
Leukaemia (C91–95)	1·09 (1·05–1·13)	<0·0001
(5833)	1·07 (1·02–1·14)	0·0011



Breast—premenopausal (C50)	0·89 (0·86–0·92)	<0·0001
(6298)	0·89 (0·85–0·94)	<0·0001
Breast—postmenopausal (C50)	1·05 (1·03–1·07)	<0·0001
(28 409)	1·05 (1·03–1·08)	<0·0001
Colon (C18)	1·10 (1·07–1·13)	<0·0001
(13 465)	1·11 (1·07–1·15)	<0·0001
Rectum (C20)	1·04 (1·00–1·08)	0·017
(6123)	1·05 (0·99–1·12)	0·024
Lung (C34)	0·82 (0·81–0·84)	<0·0001
(19 339)	0·99 (0·93–1·05)	0·55
Ovaries (C56)	1·09 (1·04–1·14)	<0·0001
(3684)	1·08 (1·02–1·15)	0·00036
Prostate (C61)	0·98 (0·95–1·00)	0·0042
(24 901)	0·96 (0·93–0·99)	0·0021

Bhaskaran et al, Lancet 2014

Childhood body fatness is inversely associated with breast cancer risk



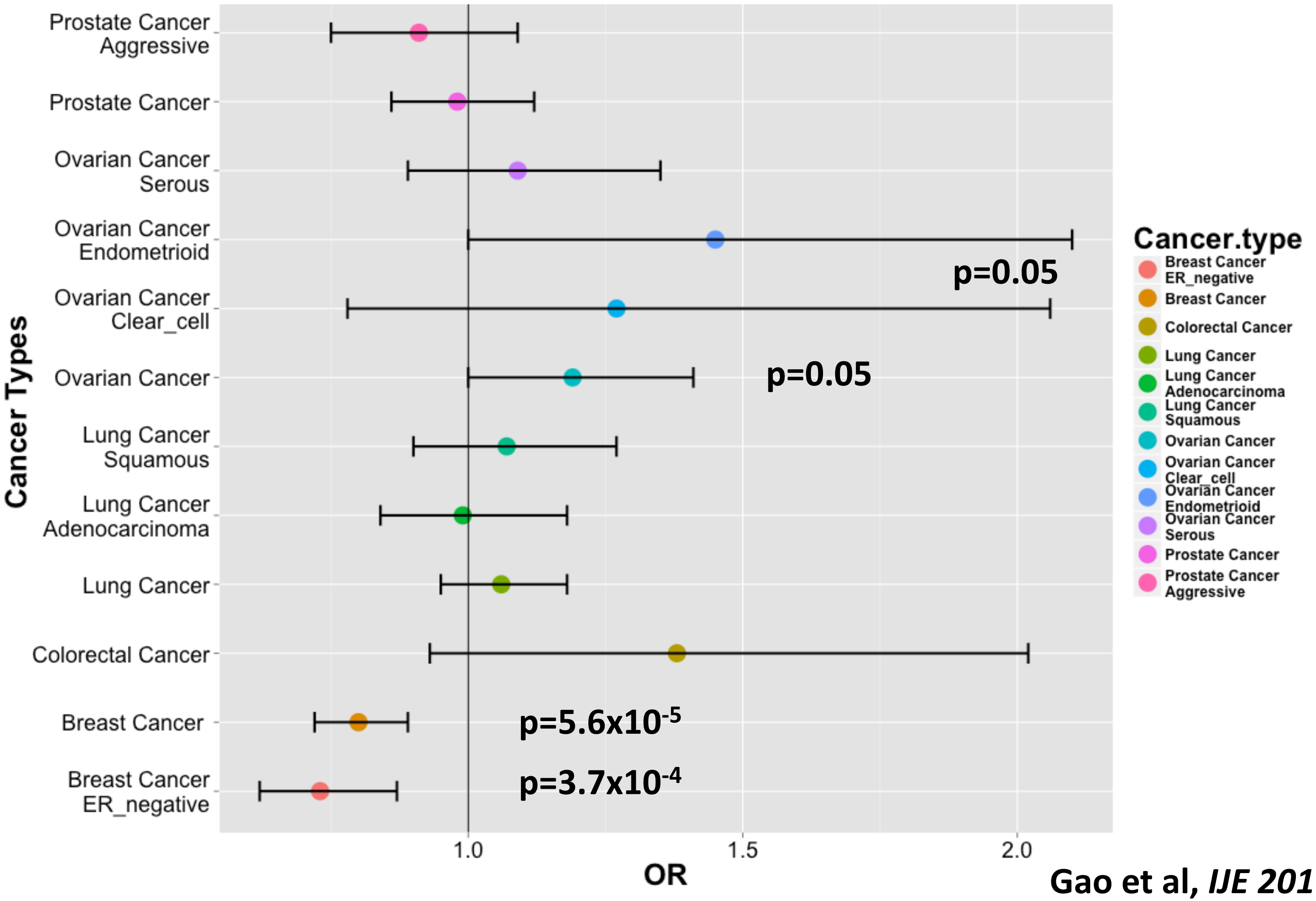
Baer et al, AJE 2010

Expansion to other cancer types within GAME-ON

Cancer Type	Cases	Controls	GWAS studies
Breast			
All	15,569	18,204	11
ER-negative	4,760	13,248	8
Colorectal	5,100	4,831	6
Lung[a]			
All	12,527	17,285	6
Adenocarcinoma	3,804	16,289	6
Squamous	3,546	16,434	6
Ovarian[a]			
All	4,369	9,123	3
Clear-cell	356	9,123	3
Endometrioid	715	9,123	3
Serous	2,556	9,123	3
Prostate			
All	14,160	12,712	6
Aggressive	4,446	12,724	6
Total	**51,725**	**62,155**	

Gao et al, *IJE 2016*

Childhood body fatness (9 SNPs)



$p=0.05$

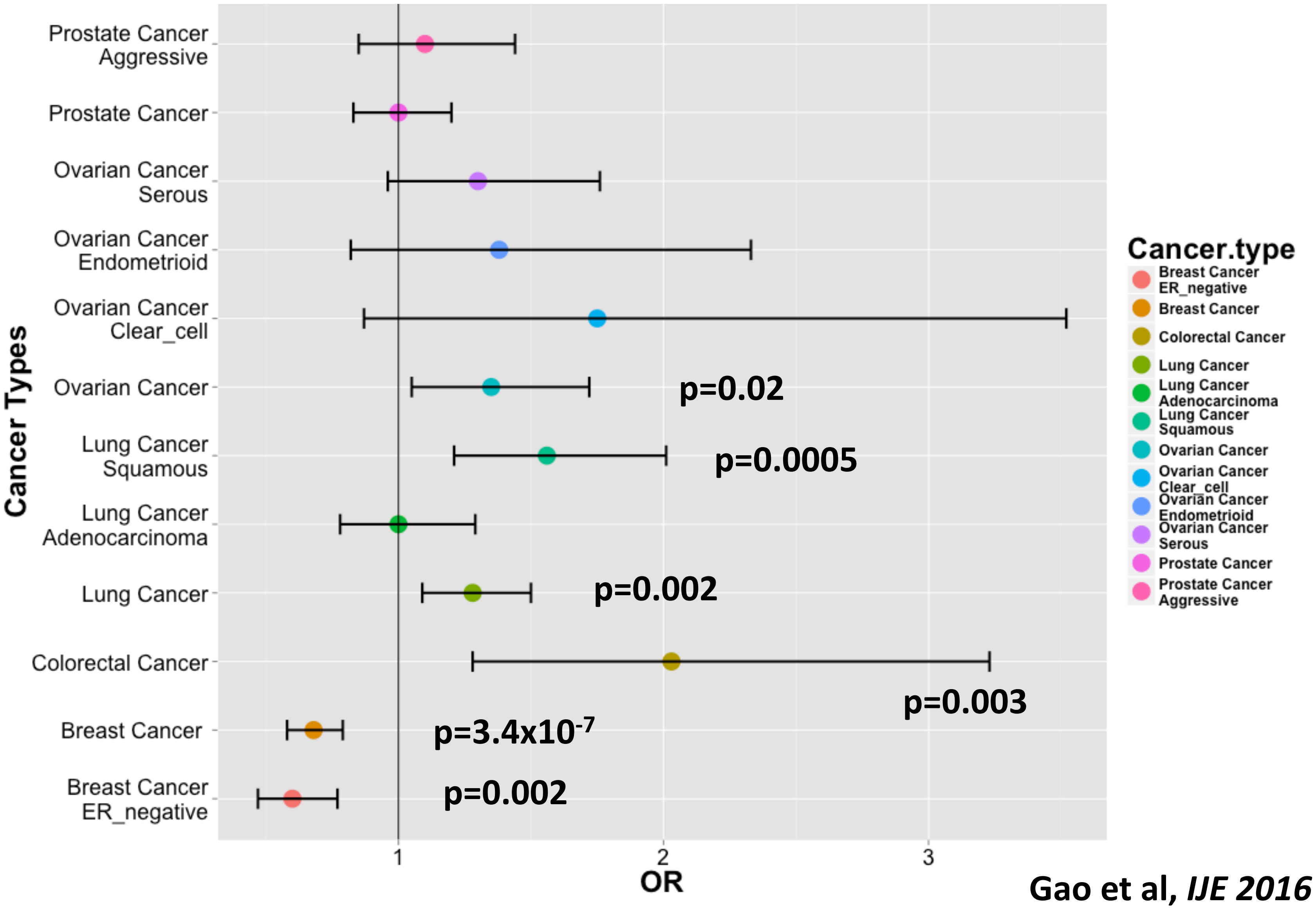
$p=0.05$

$p=5.6\text{x}10^{-5}$

$p=3.7\text{x}10^{-4}$

Gao et al, *IJE 2016*

Adult BMI (77 SNPs)



$p=0.02$

$p=0.0005$

$p=0.002$

$p=0.003$

$p=3.4 \times 10^{-7}$

$p=0.002$

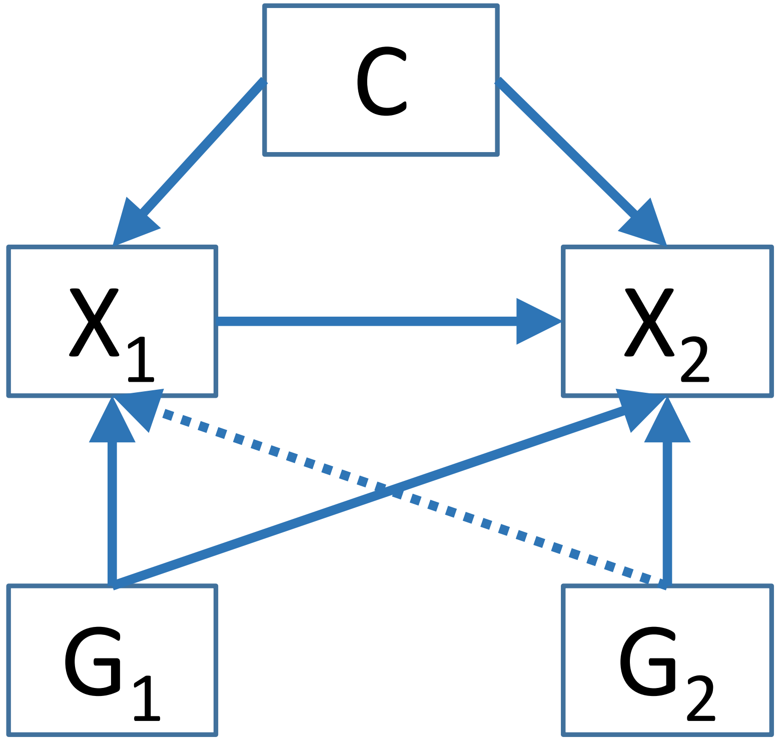
Gao et al, *IJE 2016*

Bidirectional MR analysis

- Approach to overcome reverse causation
- IVs for both X_1 and X_2 are used to assess the causal association in both directions

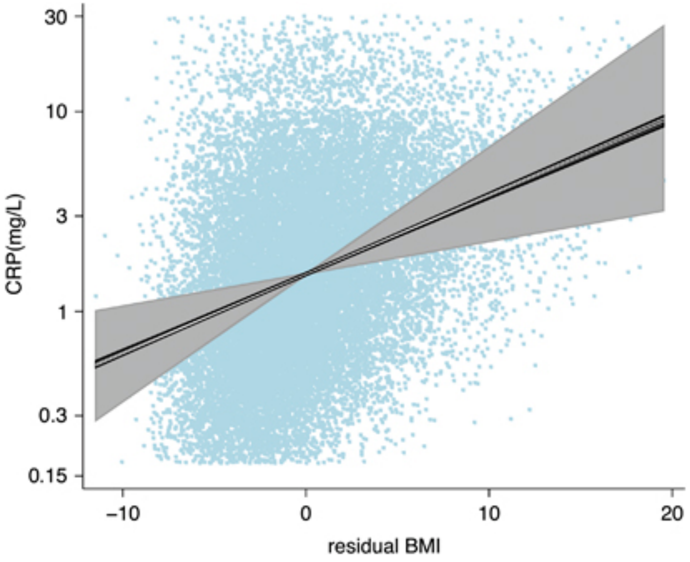
1. Is G_1 associated with X_2?
2. Is G_2 associated with X_1?

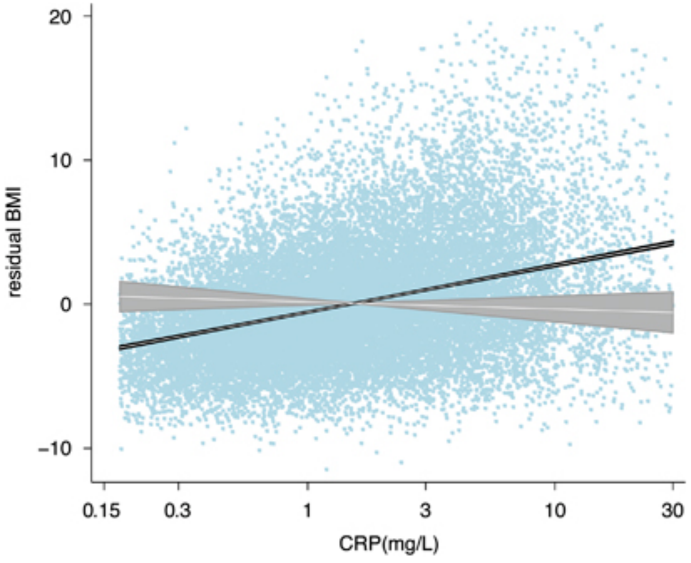
(Also confirm that G_1 is associated with X_1 and that G_2 is associated with X_2



BMI and CRP – what causes what?

- There is a consistent observed association between high BMI and high CRP levels



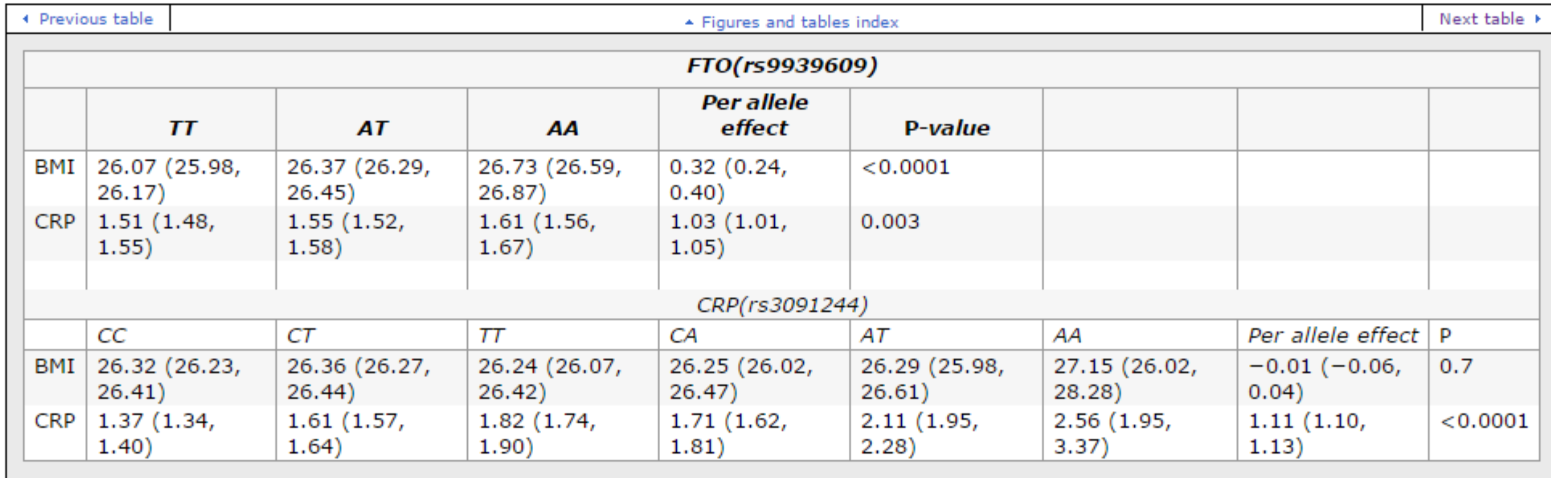


Light grey points represent a scatter plot of the correlation between circulating CRP and residual BMI. Gray areas represent 95% confidence regions around IV estimates. Black area represents 95% confidence regions around simple linear regression estimates.

Timpson et al, Int J Obesity 2011

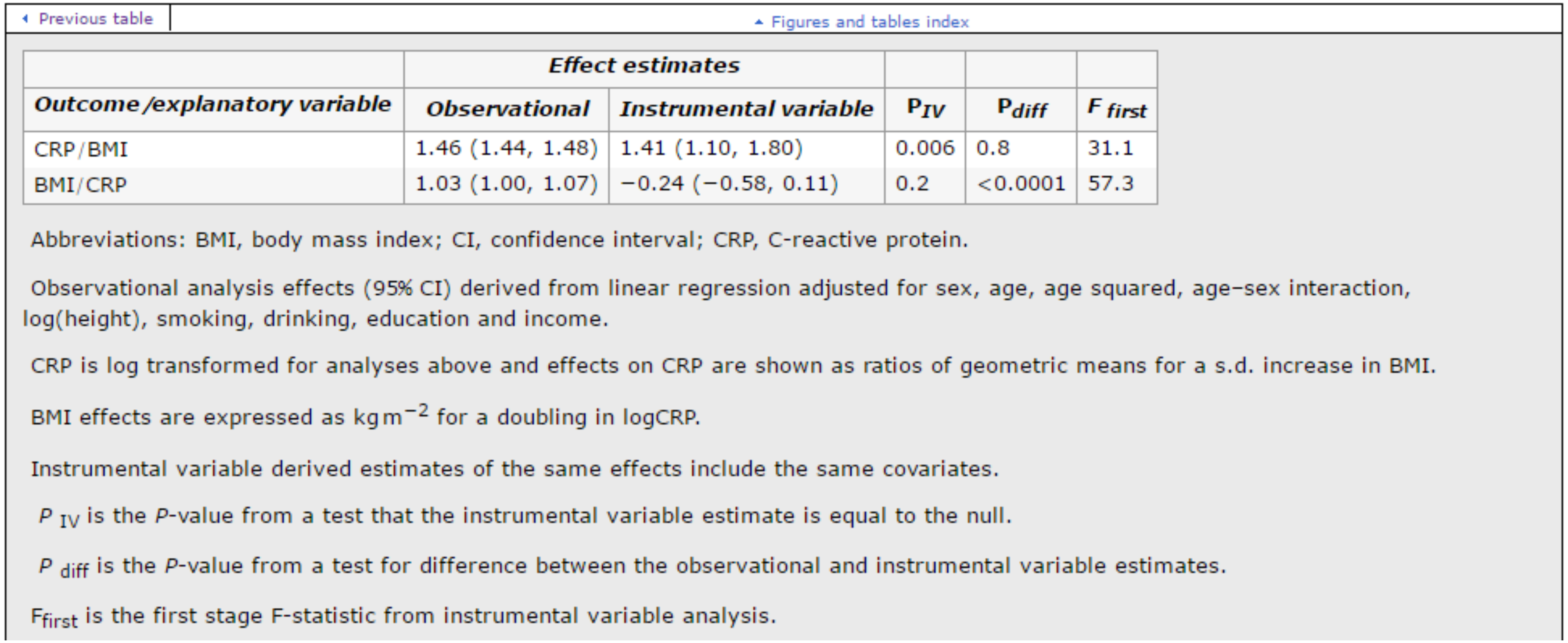
- These data suggest that the observed association between circulating CRP and measured BMI is likely to be driven by BMI, with CRP being a marker of elevated adiposity.

Table 4. Relationships between genotypic variation and BMI and circulating CRP


Previous table
Figures and tables index
Next table

	FTO(rs9939609)							
	TT	***AT***	***AA***	***Per allele effect***	***P-value***			
BMI	26.07 (25.98, 26.17)	26.37 (26.29, 26.45)	26.73 (26.59, 26.87)	0.32 (0.24, 0.40)	<0.0001			
CRP	1.51 (1.48, 1.55)	1.55 (1.52, 1.58)	1.61 (1.56, 1.67)	1.03 (1.01, 1.05)	0.003			
	CRP(rs3091244)							
	CC	*CT*	*TT*	*CA*	*AT*	*AA*	*Per allele effect*	P
BMI	26.32 (26.23, 26.41)	26.36 (26.27, 26.44)	26.24 (26.07, 26.42)	26.25 (26.02, 26.47)	26.29 (25.98, 26.61)	27.15 (26.02, 28.28)	−0.01 (−0.06, 0.04)	0.7
CRP	1.37 (1.34, 1.40)	1.61 (1.57, 1.64)	1.82 (1.74, 1.90)	1.71 (1.62, 1.81)	2.11 (1.95, 2.28)	2.56 (1.95, 3.37)	1.11 (1.10, 1.13)	<0.0001

Table 5. Observational and instrumental variable derived relationships between BMI and circulating CRP.


Previous table
Figures and tables index

	Effect estimates				
Outcome/explanatory variable	***Observational***	***Instrumental variable***	**P*IV***	**P*diff***	***F first***
CRP/BMI	1.46 (1.44, 1.48)	1.41 (1.10, 1.80)	0.006	0.8	31.1
BMI/CRP	1.03 (1.00, 1.07)	−0.24 (−0.58, 0.11)	0.2	<0.0001	57.3

Abbreviations: BMI, body mass index; CI, confidence interval; CRP, C-reactive protein.

Observational analysis effects (95% CI) derived from linear regression adjusted for sex, age, age squared, age–sex interaction, log(height), smoking, drinking, education and income.

CRP is log transformed for analyses above and effects on CRP are shown as ratios of geometric means for a s.d. increase in BMI.

BMI effects are expressed as kg m^{-2} for a doubling in logCRP.

Instrumental variable derived estimates of the same effects include the same covariates.

P_{IV} is the *P*-value from a test that the instrumental variable estimate is equal to the null.

P_{diff} is the *P*-value from a test for difference between the observational and instrumental variable estimates.

F_{first} is the first stage F-statistic from instrumental variable analysis.

Timpson et al, Int J Obesity 2011

Drawbacks with MR analysis

- Large sample sizes are needed!

- As genetic effects on risk factors are typically small, MR estimates of association have much wider confidence intervals than conventional epidemiological estimates.

- Make sure that the three key assumptions hold!