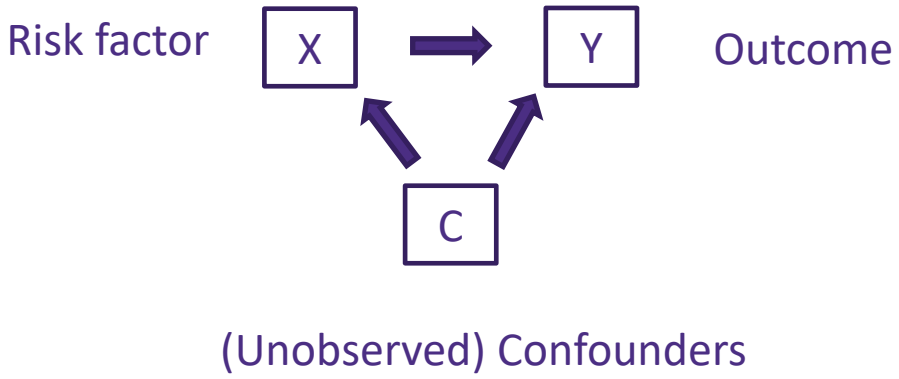
SCHOOL OF PUBLIC HEALTH
EPIDEMIOLOGY
UNIVERSITY of WASHINGTON

Session 12: Mendelian Randomization

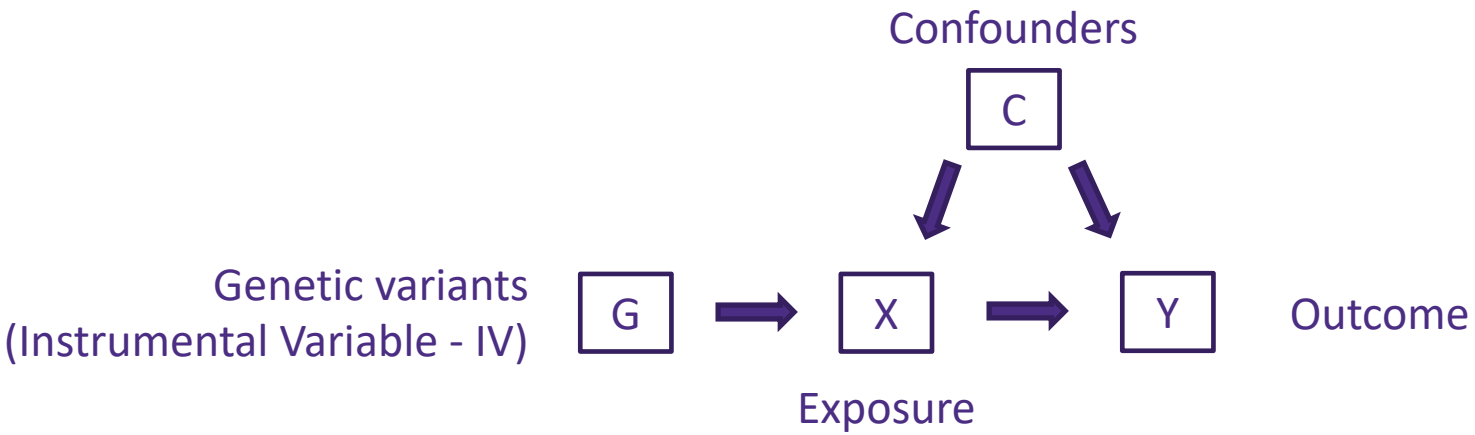
Drawback of observational studies

CONFOUNDING



REVERSE CAUSATION

We can leverage genetic variation to (partly) overcome these issues



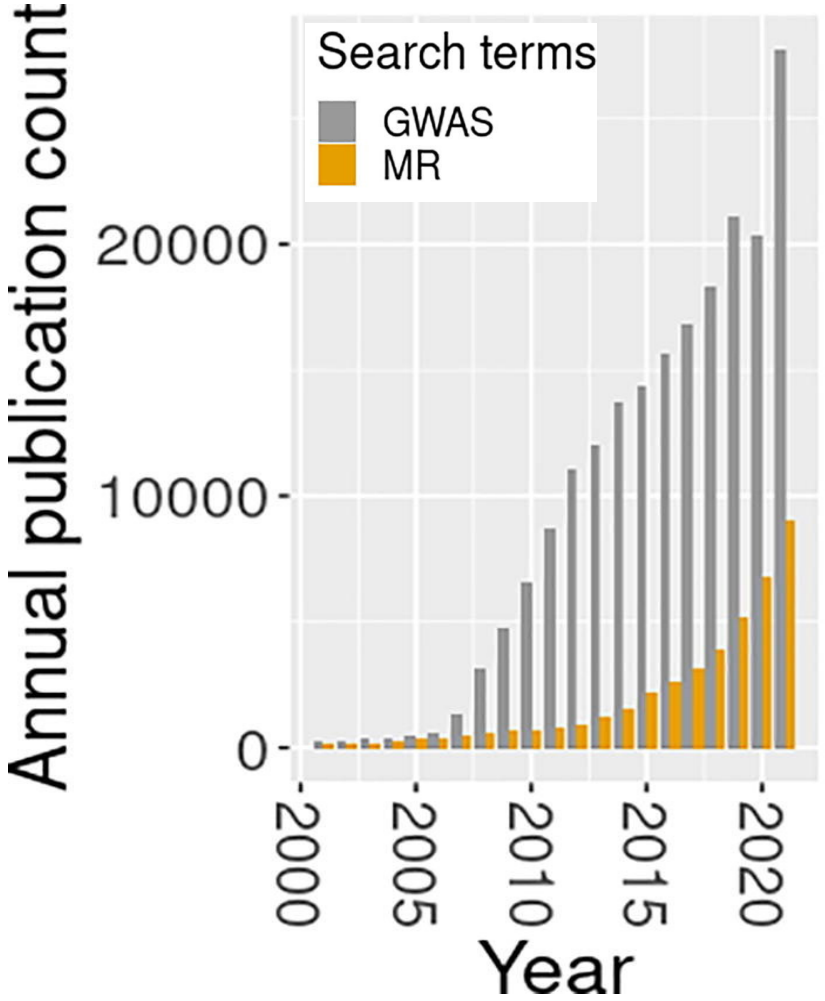
EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

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.

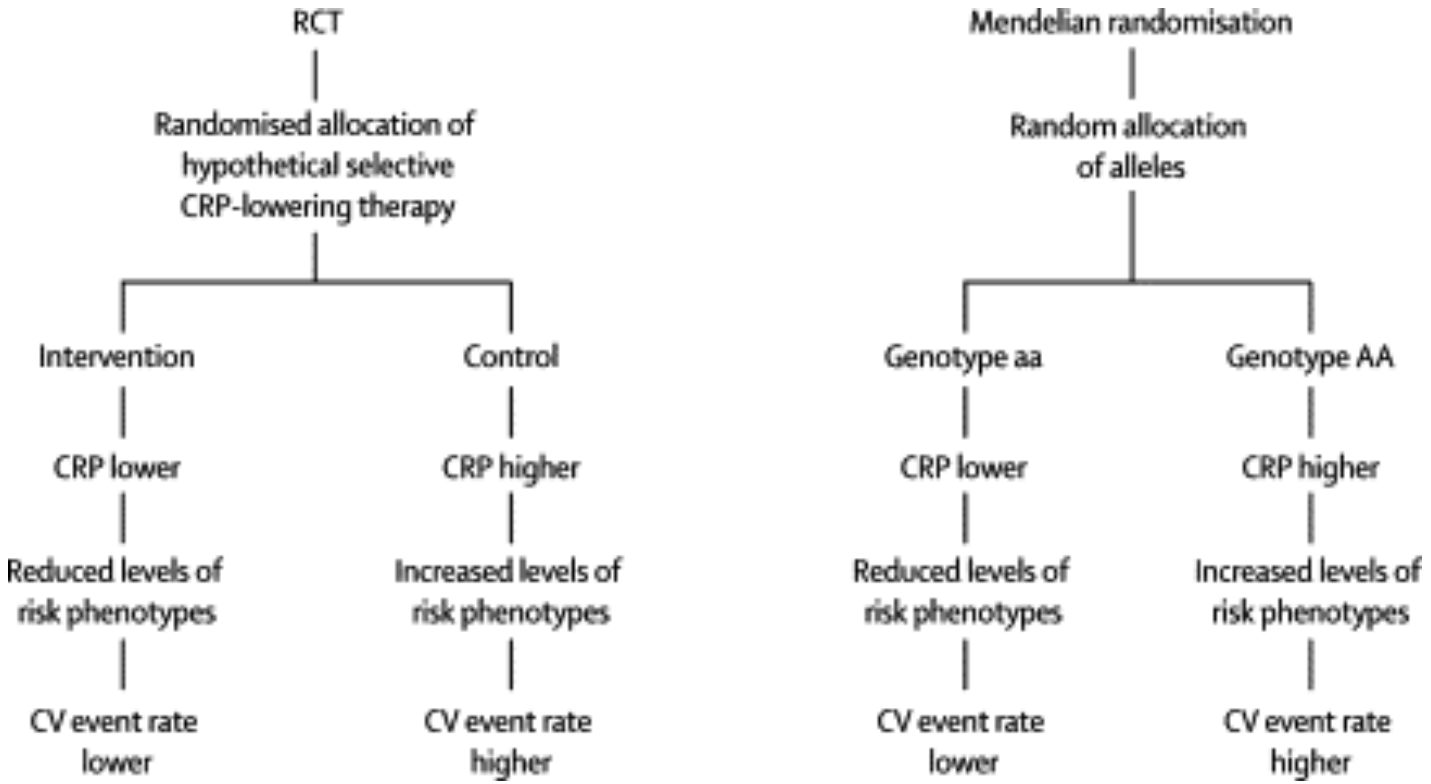
- Furthermore, associations between the genotype and the outcome will not be affected by reverse causation because disease does not affect genotype



Ebrahim & Davey Smith, Hum Genet 2008
Davey Smith & Ebrahim, Int J Epi 2004
Boehm & Zhou, Comp Str Biotech J 2022

Possible effects of C-reactive protein (CRP) concentrations 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.



Hingorani & Humphries, Lancet 2005

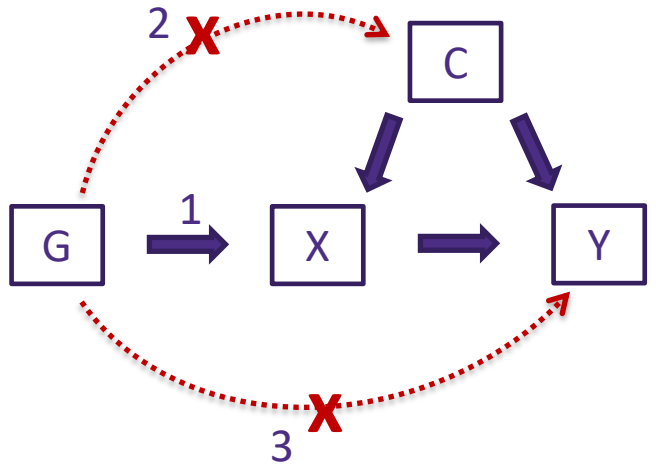
EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

The three key assumptions in MR analyses

1. **Relevance assumption**: G (SNP or a combination of SNPs) is robustly associated with X (exposure)

2. **Independence assumption**: 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. **Exclusivity assumption**: G is related to Y only through its association with X—i.e., G is not associated with Y either directly or indirectly through other traits (i.e., no pleiotropy)



Pleiotropy: The potential for genetic variants to associate with multiple phenotypes

MR Dictionary: https://mr-dictionary.mrcieu.ac.uk/

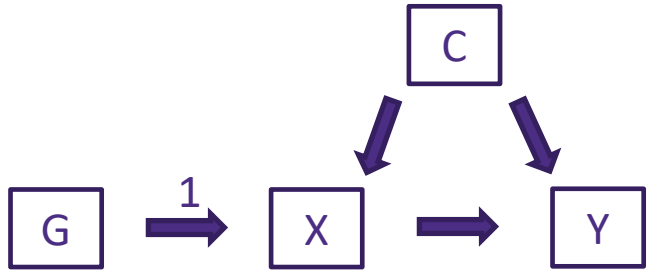
Assumption 1: Relevance assumption

> A “weak” instrument variable (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.



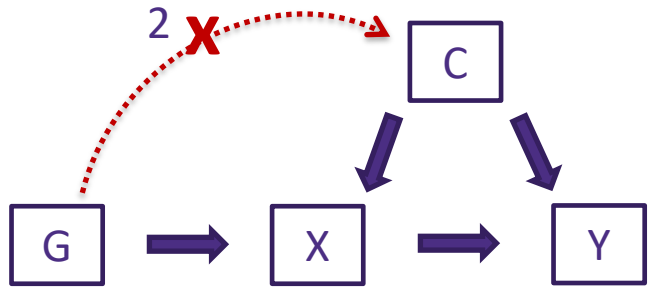
Pierce, IJE 2011

Assumption 2: Independence assumption

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

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

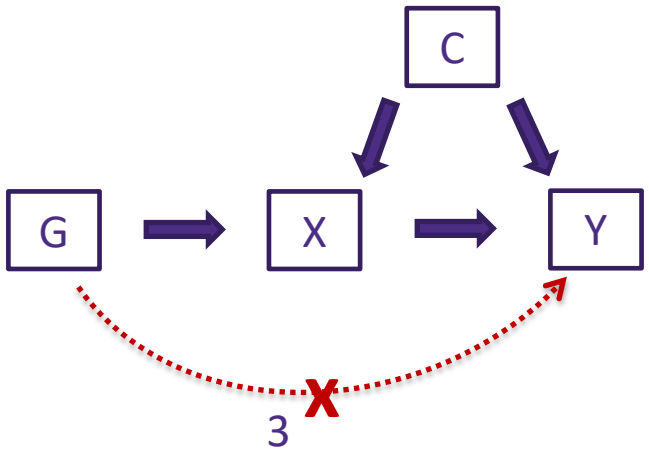
> If you have individual-level data, you can adjust for this (e.g., PCs)



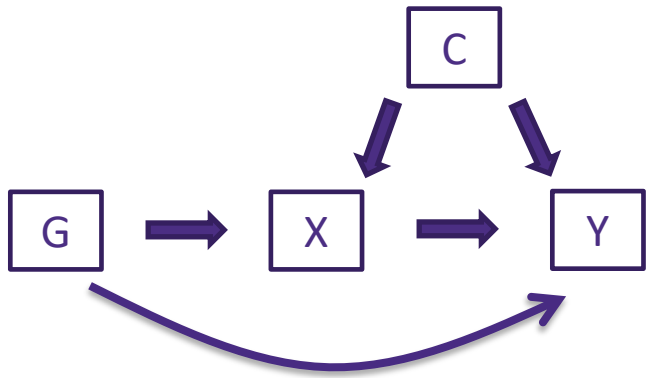
Assumption 3: Exclusivity assumption

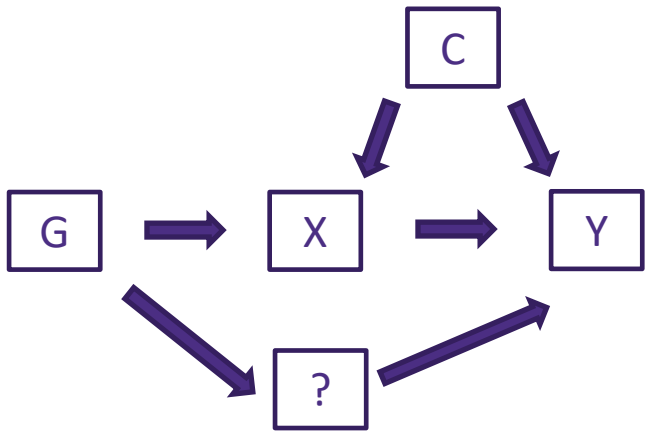
> This assumption is the trickiest

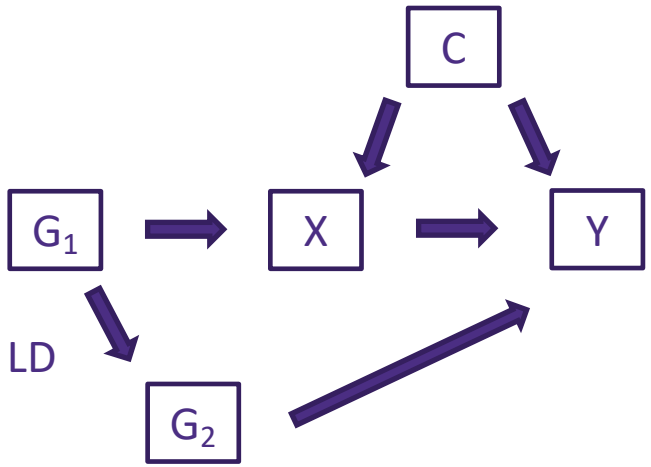
> Assumes that G is only associated with Y via X and thus the association between G and Y is fully mediated by X and not through any unmeasured factor(s).

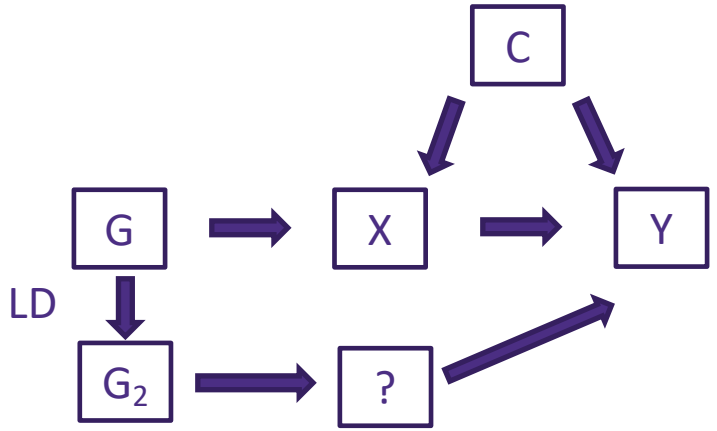


Scenarios invalidating assumption 3









BREAKOUT ACTIVITY

> In which examples (a-f) below do the MR assumptions not hold for assessing the association between exposure (X) and outcome (Y)? Why? Why not?

a. genetic variants → exposure → mediator → outcome

b. genetic variants → mediator → exposure → outcome

c. genetic variants → exposure → outcome
   ↓
   related variable

d. genetic variants → related variable → outcome
   ↓
   exposure

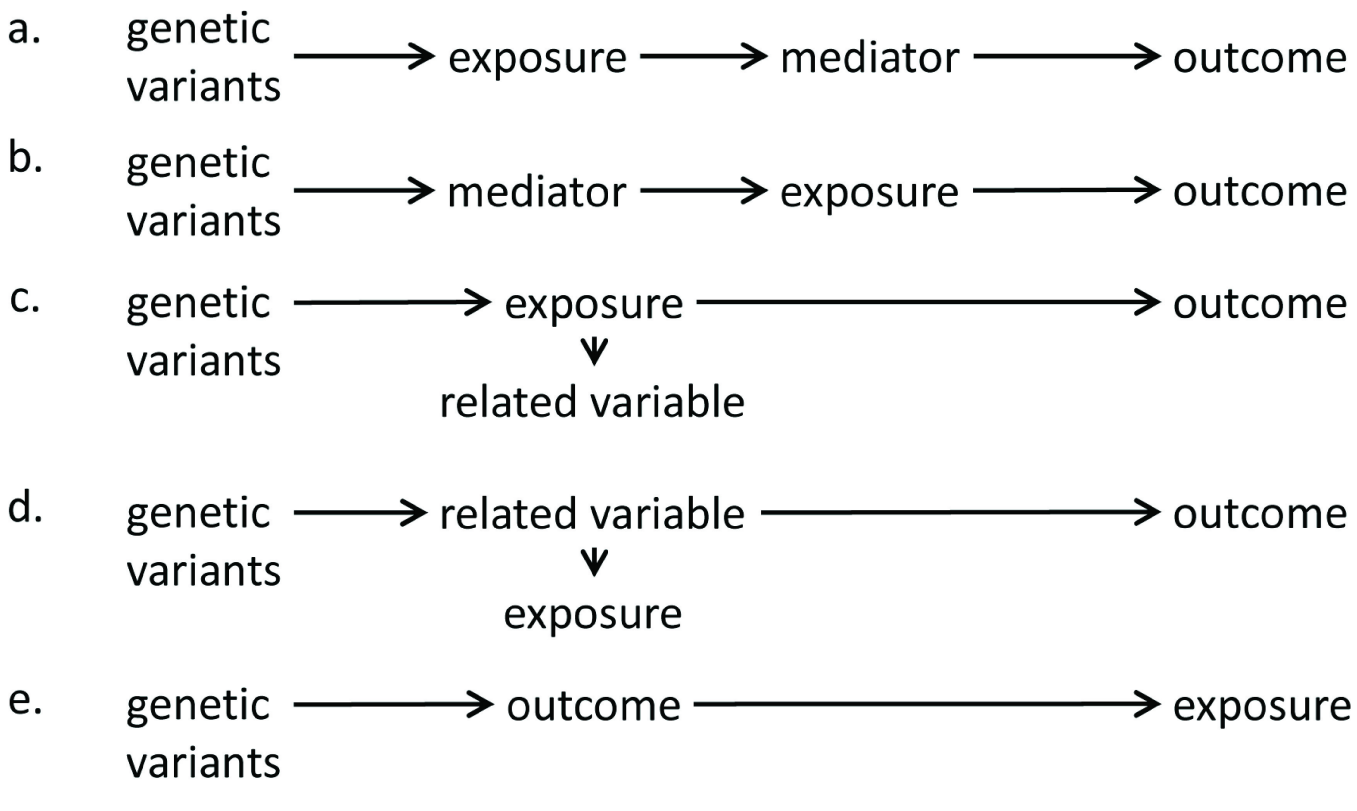
e. genetic variants → outcome → exposure

Burgess, Wellcome Open Res 2019

EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

BREAKOUT ACTIVITY

> In which examples (a-f) below do the MR assumptions not hold for assessing the association between exposure (X) and outcome (Y)? Why? Why not?



Ok! No alternate path from G to Y

Pathway from G to Y does not pass via the exposure; assumptions not met. However, assumptions are met for the related variable, which could instead be tested as the exposure.

Reverse causation; G incorrectly identified as primarily affecting the exposure

Burgess, Wellcome Open Res 2019

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

One-sample MR

> 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. $X = a + \gamma G$
2. $Y = c + \beta X^*$, 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 IVs cause bias towards the confounded X-Y association

Two-sample MR

> GWAS summary statistics for G-X and the G-Y associations are estimated in two non-overlapping samples.

> Assumes the two populations are similar (ancestry, age, etc.)

> Weak IVs cause bias towards the null

> Note: The G-X and G-Y associations need to be coded using the same effect allele

Two-sample MR

Example IV criteria (can vary) based on G-X

- Genome-wide significant variants ($P<5x10^{-8}$)
- Independent (e.g., $r^2<0.10$)
- Common (MAF>1%, if sufficiently large N)
- Exclude palindromic variants (e.g., A/T, G/C) if MAF~50%
- Exclude incompatible alleles between G-X and G-Y (e.g., if a variant has A/G alleles for exposure but A/C for outcome

*Extract these variants from G-X **and** G-Y GWAS summary statistics to conduct MR*

Hemani et al., eLife 2018

Two-sample MR: Inverse-variance weighted (IVW) method to estimate the causal effect of X on Y

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

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

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, $\sigma_{\beta 2k}^{-2}$ is the inverse variance for the G-Y association.

Hemani et al., eLife 2018

MR-base: An easy tool for Mendelian Randomization Analysis

- http://app.mrbase.org/

- A web-based platform (MR-Base) and an R-package "*TwoSampleMR*'.

- Has catalogued thousands of genotype-phenotype associations and also allows manual file upload.

MR BASE

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

TwoSampleMR R package

Logged in as
Hongjie Chen
hongjie.chen41@gmail.com

- Perform MR analysis
- Quick SNP lookup

University of BRISTOL | MRC Integrative Epidemiology Unit | CANCER RESEARCH UK

MR BASE

A platform for Mendelian randomisation using summary data from genome-wide association studies

Click on 'Perform MR analysis'

To begin analysis in the web application please review the data access agreement and accept by logging in with your google account.

Get started

All analyses, data extraction and more can be performed using the TwoSampleMR R package. Additionally, you can use the R package to analyse your own outcome datasets.

Get the R package

Current status

Beta phase release

App version:
1.2.2 3a435d (31 January 2019)

R version:
3.5.1

Host:
e4ec2116cb55

R/TwoSampleMR version:
0.4.18

Database version:
0.2.0 (17 December 2017)

To use MR-Base using the TwoSampleMR R package directly:
https://github.com/MRCIEU/TwoSampleMR

See our sister website LD Hub for automated LD score regression:
http://ldsc.broadinstitute.org/

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

TwoSampleMR R package

Logged in as
Hongjie Chen
hongjie.chen41@gmail.com

- Perform MR analysis
  - Choose exposures
  - Choose outcomes
  - Run MR
- Quick SNP lookup

A platform for Mendelian randomisation using summary data from genome-wide association studies

Select the exposure (Instrumental variable), outcome and analysis scheme here.

To begin analysis in the web application please review the data access agreement and accept by logging in with your google account.

Get started

All analyses, data extraction and more can be performed using the TwoSampleMR R package. Additionally, you can use the R package to analyse your own outcome datasets.

Get the R package

Current status

Beta phase release

App version:
1.2.2 3a435d (31 January 2019)

R version:
3.5.1

Host:
e4ec2116cb55

R/TwoSampleMR version:
0.4.18

Database version:
0.2.0 (17 December 2017)

To use MR-Base using the TwoSampleMR R package directly:
https://github.com/MRCIEU/TwoSampleMR

See our sister website LD Hub for automated LD score regression:
http://ldsc.broadinstitute.org/

Choosing instruments for the exposure

To use two sample MR to estimate the causal effect of an exposure on an outcome, the first step is to identify SNPs that are robustly associated with the exposure. These summary statistics for these SNPs can be taken from a sample from which there is no data on the outcome.

Please provide instruments by choosing from one of the data sources below, or by uploading your own data. You can choose multiple exposures to be analysed, and multiple instruments per exposure.

Choose instruments

Select exposure source

- ○ Manual file upload
- ○ NHGRI-EBI GWAS catalog
- ◉ MR Base GWAS catalog
- ○ Gene expression QTLs
- ○ Protein level QTLs
- ○ Metabolite level QTLs
- ○ Methylation level QTLs

Can either use the instruments provided by MR-base,
Or use the manually uploaded file.

MR Base GWAS catalog

The MR Base database holds a collection of the summary statistics from a large number of GWASs. It is possible to use this resource to manually identify instruments, and to therefore use these traits as exposures by finding the independent GWAS significant hits from these summary associations.

To use a trait as an exposure, highlight the relevant row in the table below (multiple traits can be selected). All SNPs with p values below the specified threshold will be extracted, and clumping will be used to remove SNPs in LD with sentinal SNPs. These SNPs will be used as instruments in the MR analysis.

p-value threshold: 5e-08

LD Rsq: 0.001

☑ Perform clumping

Clumping distance (kb): 10000

Selection criteria of SNPs

Display columns

☑ ID ☑ Trait ☑ Note ☑ First author ☑ Consortium ☑ Number of cases ☑ Number of controls ☑ Sample size ☑ Number of variants ☑ Year ☑ PubmedID ☑ Access ☑ Category ☑ Population ☑ Priority ☑ Sd ☑ Sex ☑ Subcategory ☑ Unit

Show 10 entries Search:

	ID	Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Year	PubmedID	Access	Category	Population	Priority	Sd	Sex	Subcategory	Unit
1	1	Adiponectin		Dastani Z	ADIPOGen			39883	2675209	2012	22479202	public	Risk factor	Mixed	1	0.57	Males and females	Protein	ln(mg/dL)
10	10	Crohn's disease		Jostins L	IIBDGC	14763	15977	30740	13898	2012	23128233	public	Disease	European	1		Males and females	Autoimmune / inflammatory	log odds
100	100	Hip circumference	Adjusted for BMI	Randall JC	GIANT			60586	2725796	2013	23754948	public	Risk factor	European	15	8.4548	Males	Anthropometric	SD (cm)
1000	1000	Depressive symptoms		Okbay	SSGAC			161460	6524475	2016	27089181	public	Risk factor	European	1		Males and females	Psychiatric / neurological	SD
1001	1001	Years of schooling		Okbay	SSGAC			293723	8146841	2016	27225129	public	Risk factor	European	1	3.71	Males and females	Education	SD (years)
1002	1002	Leptin	Adjusted for BMI; effect allele frequencies are missing	Kilpelainen				32161	2474010	2016	26833098	public	Risk factor	European	1		Males and females	Hormone	log ng/ml
1003	1003	Leptin	Effect allele frequencies are missing	Kilpelainen				32161	2473865	2016	26833098	public	Risk factor	European	2		Males and females	Hormone	log ng/ml
1004	1004	Age at menopause		Day	ReproGen			69360	2418696	2015	26414677	public	Risk factor	European	1	3.93	Females	Reproductive aging	years
1005	1005	Percent emphysema		Manichaikul A	MESA			7667	2968584	2014	24383474	public	Risk factor	Mixed	1	0.708	Males and females	Lung disease	log (% emphysema + 1)
1006	1006	Mean platelet volume	Effect allele frequencies are missing	Gieger C	HaemGen			66867	2690859	2011	22139419	public	Risk factor	European	1	0.109	Males and females	Haemotological	log fl
	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All	All

Specify the literature to be included here

Search for the exposure of interest here.

Showing 1 to 10 of 21,266 entries Previous 1 2 3 4 5 … 2127 Next

Choosing instruments for the exposure

To use two sample MR to estimate the causal effect of an exposure on an outcome, the first step is to identify SNPs that are robustly associated with the exposure. These summary statistics for these SNPs can be taken from a sample from which there is no data on the outcome.

Please provide instruments by choosing from one of the data sources below, or by uploading your own data. You can choose multiple exposures to be analysed, and multiple instruments per exposure.

Choose instruments

Select exposure source

- ○ Manual file upload
- ○ NHGRI-EBI GWAS catalog
- ◉ MR Base GWAS catalog
- ○ Gene expression QTLs
- ○ Protein level QTLs
- ○ Metabolite level QTLs
- ○ Methylation level QTLs

MR Base GWAS catalog

The MR Base database holds a collection of the summary statistics from a large number of GWASs. It is possible to use this resource to manually identify instruments, and to therefore use these traits as exposures by finding the independent GWAS significant hits from these summary associations.

To use a trait as an exposure, highlight the relevant row in the table below (multiple traits can be selected). All SNPs with p values below the specified threshold will be extracted, and clumping will be used to remove SNPs in LD with sentinal SNPs. These SNPs will be used as instruments in the MR analysis.

p-value threshold 5e-08

LD Rsq 0.001

☑ Perform clumping

Clumping distance (kb) 10000

Display columns

☑ ID ☑ Trait ☑ Note ☑ First author ☑ Consortium ☑ Number of cases ☑ Number of controls ☑ Sample size ☑ Number of variants ☑ Year ☑ PubmedID ☑ Access ☑ Category ☑ Population ☑ Priority ☑ Sd ☑ Sex ☑ Subcategory ☑ Unit

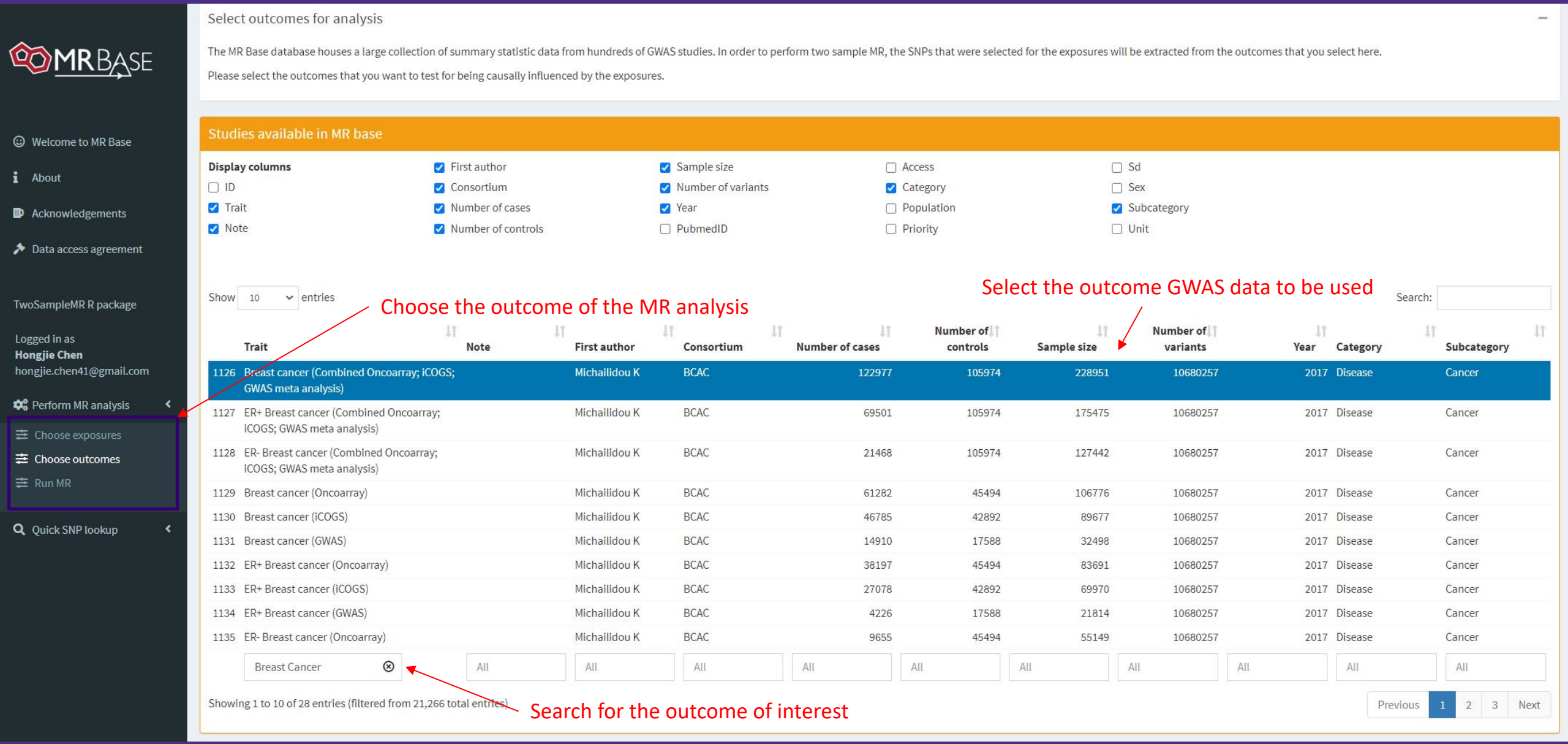
Show 10 entries Search:

	ID	Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Year	PubmedID	Access	Category	Population	Priority	Sd	Sex	Subcategory	Unit
1	1	Adiponectin		Dastani Z	ADIPOGen			39883	2675209	2012	22479202	public	Risk factor	Mixed	1	0.57	Males and females	Protein	ln(mg/dL)
10	10	Crohn's disease		Jostins L	IIBDGC	14763	15977	30740	13898	2012	23128233	public	Disease	European	1		Males and females	Autoimmune / inflammatory	log odds
100	100	Hip circumference	Adjusted for BMI	Randall JC	GIANT			60586	2725796	2013	23754948	public	Risk factor	European	15	8.4548	Males	Anthropometric	SD (cm)
1000	1000	Depressive symptoms		Okbay	SSGAC			51460	6524475	2016	27089181	public	Risk factor	European	1		Males and females	Psychiatric / neurological	SD
1001	1001	Years of schooling		Okbay	SSGAC			293723	8146841	2016	27225129	public	Risk factor	European	1	3.71	Males and females	Education	SD (years)
1002	1002	Leptin	Adjusted for BMI; effect allele frequencies are missing	Kilpelainen				32161	2474010	2016	26833098	public	Risk factor	European	1		Males and females	Hormone	log ng/ml
1003	1003	Leptin	Effect allele frequencies are missing	Kilpelainen				32161	2473865	2016	26833098	public	Risk factor	European	2		Males and females	Hormone	log ng/ml
1004	1004	Age at menopause		Day	ReproGen			69360	2418696	2015	26414677	public	Risk factor	European	1	3.93	Females	Reproductive aging	years
1005	1005	Percent emphysema		Manichaikul A	MESA			7667	2968584	2014	24383474	public	Risk factor	Mixed	1	0.708	Males and females	Lung disease	log (% emphysema + 1)
1006	1006	Mean platelet volume	Effect allele frequencies are missing	Gieger C	HaemGen			66867	2690859	2011	22139419	public	Risk factor	European	1	0.109	Males and females	Haemotological	log fl

Select the instruments
to be included
In your MR analysis

Showing 1 to 10 of 21,266 entries

Previous 1 2 3 4 5 … 2127 Next



	Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Year	Category	Subcategory
1126	Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)		Michailidou K	BCAC	122977	105974	228951	10680257	2017	Disease	Cancer
1127	ER+ Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)		Michailidou K	BCAC	69501	105974	175475	10680257	2017	Disease	Cancer
1128	ER- Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)		Michailidou K	BCAC	21468	105974	127442	10680257	2017	Disease	Cancer
1129	Breast cancer (Oncoarray)		Michailidou K	BCAC	61282	45494	106776	10680257	2017	Disease	Cancer
1130	Breast cancer (iCOGS)		Michailidou K	BCAC	46785	42892	89677	10680257	2017	Disease	Cancer
1131	Breast cancer (GWAS)		Michailidou K	BCAC	14910	17588	32498	10680257	2017	Disease	Cancer
1132	ER+ Breast cancer (Oncoarray)		Michailidou K	BCAC	38197	45494	83691	10680257	2017	Disease	Cancer
1133	ER+ Breast cancer (iCOGS)		Michailidou K	BCAC	27078	42892	69970	10680257	2017	Disease	Cancer
1134	ER+ Breast cancer (GWAS)		Michailidou K	BCAC	4226	17588	21814	10680257	2017	Disease	Cancer
1135	ER- Breast cancer (Oncoarray)		Michailidou K	BCAC	9655	45494	55149	10680257	2017	Disease	Cancer
	Breast Cancer	All	All	All	All	All	All	All	All	All	All

MR BASE

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

TwoSampleMR R package

Logged in as
Hongjie Chen
hongjie.chen41@gmail.com

- Perform MR analysis
  - Choose exposures
  - Choose outcomes
  - Run MR
- Quick SNP lookup

Move forward and set up the MR analysis

LD clumping

Most two sample MR methods require that the instruments do not have LD between them.

Linkage disequilibrium

- (•) Do not check for LD between SNPs
- ( ) Use clumping to prune SNPs for LD

LD proxies

If a particular exposure SNP is not present in an outcome dataset, should proxy SNPs be used instead through LD tagging?

- [x] Use proxies?

Minimum LD Rsq value

0.6 | 0.8 | 1

0.6 0.64 0.68 0.72 0.76 0.8 0.84 0.88 0.92 0.96 1

- [x] Allow palindromic SNPs?

MAF threshold for aligning palindromes

0.01 | 0.3 | 0.49

0.01 0.06 0.11 0.16 0.21 0.26 0.31 0.36 0.41 0.46 0.49

Allele harmonisation

An important step in two sample MR is making sure that the effects of the SNPs on the exposure correspond to the same allele as their effects on the outcome. This is potentially difficult with palindromic SNPs.

Handling reference alleles

- ( ) All effect alleles are definitely on the positive strand
- (•) Attempt to align strands for palindromic SNPs
- ( ) Exclude palindromic SNPs

Select methods for analysis

Many methods exist for performing two sample MR. Different methods have sensitivities to different potential issues, accommodate different scenarios, and vary in their statistical efficiency.

Choose which methods to use:

- [x] Wald ratio
- [ ] Maximum likelihood
- [x] MR Egger
- [ ] MR Egger (bootstrap)
- [ ] Simple median
- [x] Weighted median
- [ ] Penalised weighted median
- [x] Inverse variance weighted
- [ ] IVW radial
- [ ] Inverse variance weighted (multiplicative random effects)
- [ ] Inverse variance weighted (fixed effects)
- [ ] Simple mode
- [x] Weighted mode
- [ ] Weighted mode (NOME)
- [ ] Simple mode (NOME)
- [ ] Robust adjusted profile score (RAPS)
- [ ] Sign concordance test
- [ ] Unweighted regression

Submit

Once you have selected exposures, outcomes, and analysis options you are ready to perform the analysis.

Perform MR analysis

After setting up the analysis scheme, click here to submit the request to perform the MR analysis

MR BASE

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

TwoSampleMR R package

Logged in as
Hongjie Chen
hongjie.chen41@gmail.com

- Perform MR analysis
  - Choose exposures
  - Choose outcomes
  - Run MR
  - MR Results
- Quick SNP lookup

Results appear after the analysis is done

Select analysis

Exposure

Age at menopause || id:1004

Outcome

Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis) || id:1126

Generate HTML report

Exposure details

Name: Age at menopause
ID: 1004
Number of instruments used: 42
Units: years
Number of cases: NaN
Number of controls: NaN
Sample size: 69360
PubmedID: 26414677
First author: Day
Consortium: ReproGen
Year: 2015

Outcome details

Name: Breast cancer (Combined Oncoarray; iCOGS; GWAS meta analysis)
ID: 1126
SNPs in GWAS: 10680257
Number of instruments identified: 35
... of which are LD proxies: 0
Units: log odds
Number of cases: 122977
Number of controls: 105974
Sample size: 228951
PubmedID: 29059683
First author: Michailidou K
Consortium: BCAC
Year: 2017

# of instruments were found in the outcome GWAS, which were used in the MR analysis.

Downloads for all analyses

Download harmonised summary statistics

Download MR results

Download leave-one-out sensitivity analysis

Download single SNP MR results

Download the generated datasets or MR analysis results here.

MR results | Heterogeneity statistics | Causal direction test | Horizontal pleiotropy — **Tables**

This table shows the MR estimates from each method of the causal effect of the exposure on the outcome. The effects are reported in the units that were used to estimate the SNP effects.

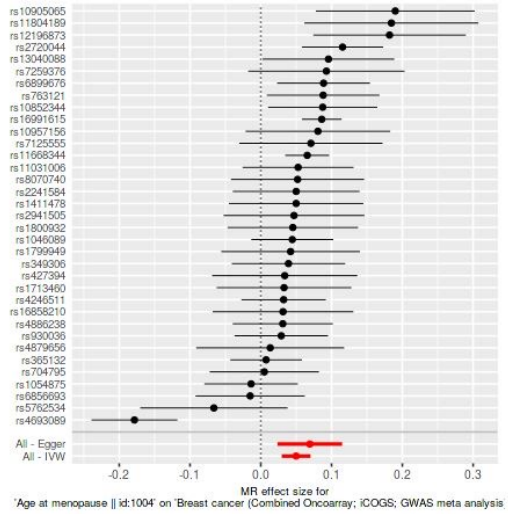
method	nsnp	b	se	pval
MR Egger	35	0.06926	0.02329	0.005452
Weighted median	35	0.05319	0.01036	2.815e-7
Inverse variance weighted	35	0.04993	0.01036	0.000001446
Weighted mode	35	0.06599	0.01204	0.000004092

Test statistics corresponding to MR analysis approaches selected.

Single SNP analysis | Method comparison plot | Leave-one-out analysis | Funnel plot — **Graphs**

The causal effect of exposure on outcome is estimated using each SNP singly using the Wald ratio, and represented in a forest plot. The MR estimate using all SNPs using the MR Egger and IVW methods are also shown. Formal estimates of heterogeneity are shown in the tables below.

Download PDF of this graph



Causal effect of exposure on outcome, by SNP

Evidence of heterogeneity of associations (confirmed in “Heterogeneity statistics” tab) suggests some SNPs exhibit horizontal pleiotropy.

…but MR-Egger is unbiased even if heterogeneity assumption is violated, and MR-Egger is significant here.

BREAKOUT ACTIVITY

> (Explore MR-Base (http://www.mrbase.org) to conduct your own MR study. Run an MR study of smoking pack years and lung cancer risk following the example in class.)

Details about MR-Base: Hemani et al., eLife 2018 (https://elifesciences.org/articles/34408)

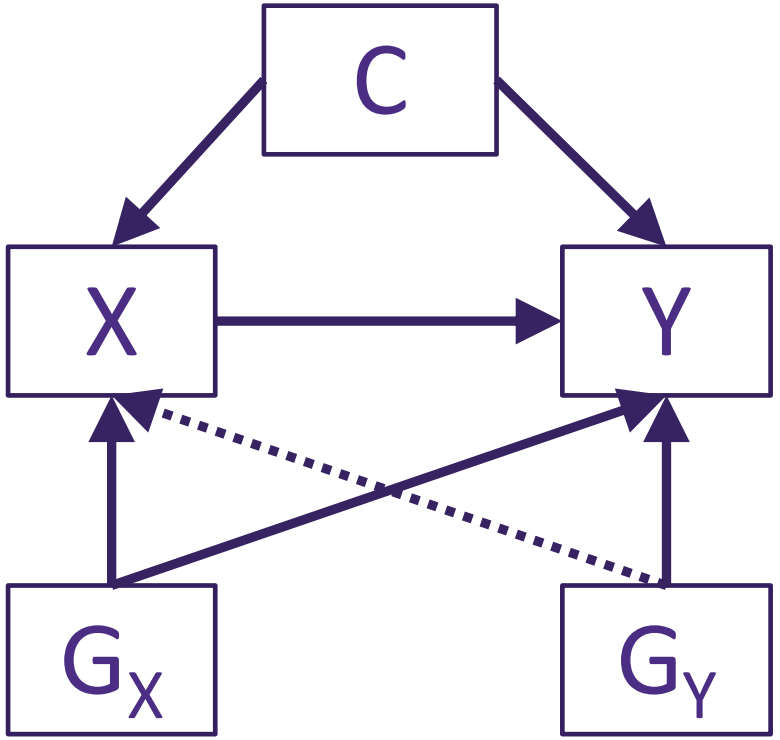
Bidirectional MR analysis

> Approach to overcome reverse causation (X←→Y)

> IVs for both X and Y are used to assess the causal association in both directions

1. Is G_X associated with Y?
2. Is G_Y associated with X?

(Also confirm that G_X is associated with X and that G_Y is associated with Y

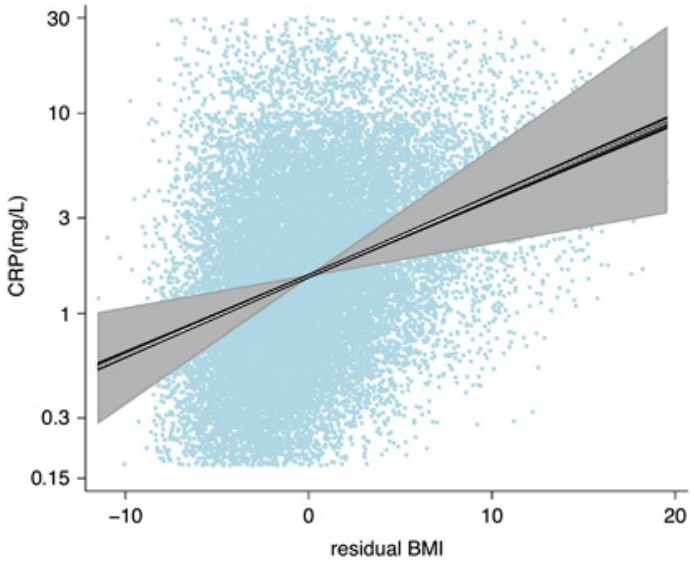


EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

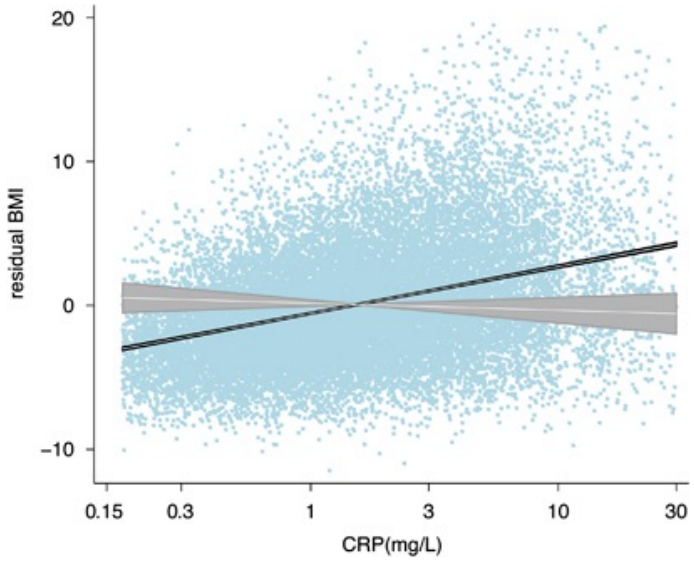
BMI and CRP – what causes what?

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

Gray: 95% CI for *FTO* loci as IV for residual BMI.



Gray: 95% CI for *CRP* locus rs3091244 as IV for CRP.



Blue dots 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

The observed association between circulating CRP and measured BMI is suggested to be driven by BMI, with adiposity causally influencing circulating CRP levels and not vice-versa

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

Timpson et al, Int J Obesity 2011

EPIDEMIOLOGY
SCHOOL OF PUBLIC HEALTH

Considerations for MR analyses

- 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.
- Consider the three core assumptions and how they apply in your study
  - In practice, this is very difficult, especially for the third assumption of no pleiotropy.
- Look for consistency across MR approaches
  - Tells you how robust your results are given the different assumptions
- In the end, it is a helpful tool to complement observational findings
  - You can never assume causality because you can never be 100% sure that all assumptions are met

Table 1. Summary of some methods proposed for Mendelian randomization: inverse-variance weighted method and robust methods.

Method	Consistency assumption	Strengths and weaknesses	Reference	Software
Inverse-variance weighted	All variants valid or balanced pleiotropy	Most efficient (greatest statistical power), biased if average pleiotropic effect differs from zero	18	*†
MR-Egger	InSIDE	Sensitive to outliers, sensitive to violations of InSIDE assumption, InSIDE assumption often not plausible, often less efficient	19	*†
Weighted median	Majority valid	Robust to outliers, sensitive to addition/removal of genetic variants	20	*†
Mode-based estimation	Plurality valid	Robust to outliers, sensitive to bandwidth parameter and addition/removal of genetic variants, generally conservative	21	*†
MR-PRESSO	Outlier-robust	Removes outliers, efficient with valid IVs, very high false positive rate with several invalid IVs	22	‡
MR-Robust	Outlier-robust	Downweights outliers, efficient with valid IVs, high false positive rate with several invalid IVs	23	*
MR-Lasso	Outlier-robust	Removes outliers, efficient with valid IVs, high false positive rate with several invalid IVs	23	
MR-RAPS	Balanced pleiotropy (except outliers)	Downweights outliers, sensitive to violations of balanced pleiotropy assumption	24	‡
Contamination Mixture	Plurality valid	Robust to outliers, sensitive to variance parameter and addition/removal of genetic variants	25	*
MR-Mix	Plurality valid	Robust to outliers, requires large numbers of genetic variants, very high false positive rate in several scenarios	26	‡

Each of the methods in the table can be implemented using summarized data. False positive rates refer to the simulation study by Slob and Burgess[27]. InSIDE is the Instrument Strength Independent of Direct Effect assumption.

Burgess, Wellcome Open Research 2020

Mendelian Randomization in R

- Encodes several methods for performing Mendelian randomization analyses with summarized data. Summarized data on genetic associations with the exposure and with the outcome can be obtained from large consortia. These data can be used for obtaining causal estimates using instrumental variable methods.
- **https://cran.r-project.org/web/packages/MendelianRandomization/index.html**

- **https://www.youtube.com/channel/UCHjMrVSqOu1rcrYQPAD_bNA**