

MR-Base

An online platform for Mendelian randomization using summary data

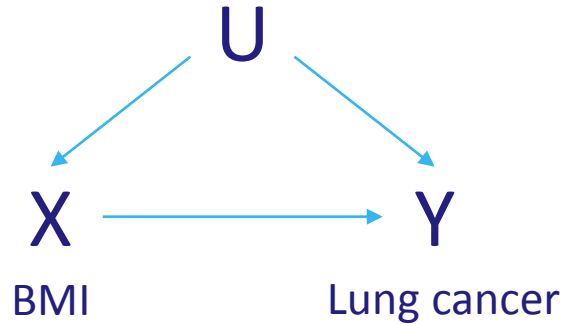
Kaitlin Wade

Integrative Cancer Epidemiology Programme (ICEP)

Integrative Epidemiology Unit (IEU)

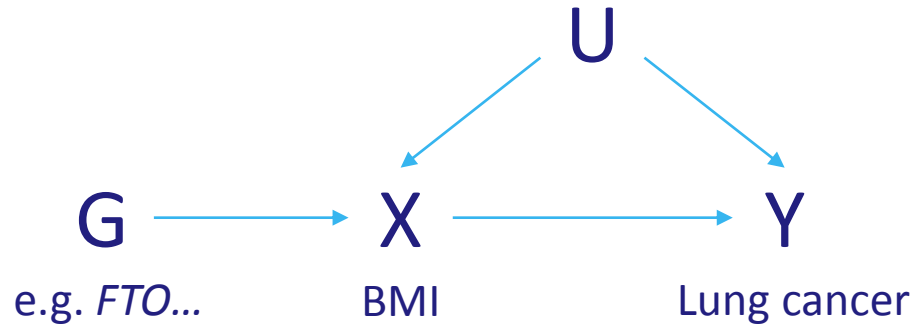
University of Bristol

Observational epidemiology

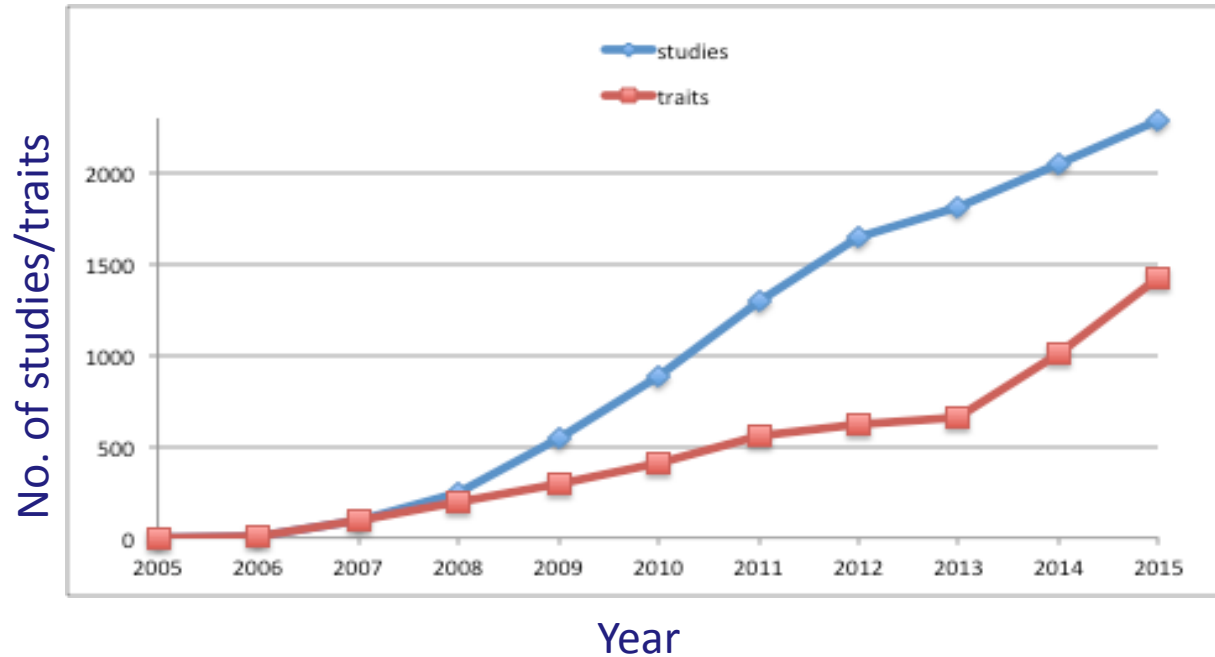


- Confounding
- Reverse causation
- Bias

Mendelian randomization



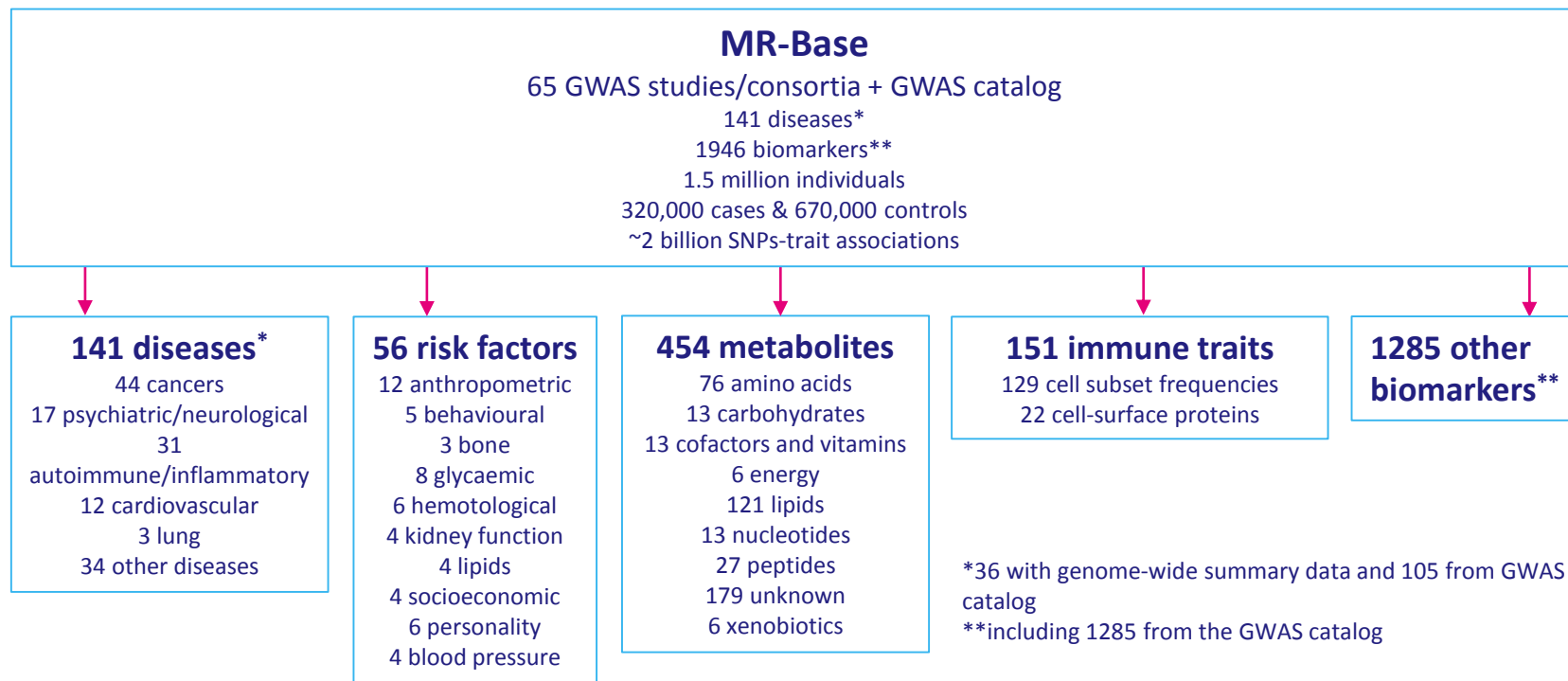
Impact of GWAS on Mendelian randomization



Maximise impact of existing data


Traditional design strategy	Novel design strategy
Individual level data	Summary genetic data
Single exposure/single outcome	Systematic approaches
Multiple databases	Central database
Multiple conventions/definitions	Harmonized data
Manual implementation of MR	Automation of the MR analysis pipeline

MR-Base: a database of harmonized summary genetic data



MR-Base access

- Web interface
 - www.mrbase.org
 - To be released soon (Summer 2016)
- R package
 - To be released soon (Summer 2016)



Welcome to MR Base

About

Acknowledgements

Data access agreement

Perform MR analysis

- Choose exposures
- Choose outcomes
- Run MR

Quick SNP lookup

Access agreement


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

[Review access agreement](#)

Current status









App version:
1.1.4 acff29

Last updated:
16 May 2016




General info of repository

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

	SNP-PHENOTYPE ASSOCIATIONS 3,417,657,704
	TRAITS WITH INSTRUMENTS 340,164
	GWAS CONSORTIA 36
	GWAS STUDIES WITH FULL SUMMARY DATA 990
	POSSIBLE CAUSAL RELATIONSHIPS 182,396,248
	NUMBER OF INDIVIDUALS REPRESENTED 1.5 million
	TWO SAMPLE MR METHODS 10
	SENSITIVITY ANALYSES 6

Performing MR analysis

Choose exposure



Welcome to MR Base

About

Acknowledgements

Data access agreement

Perform MR analysis

Choose exposures

Choose outcomes

Run MR

Quick SNP lookup

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

Drug proxies

Methylation level QTLs

Manual file upload

The file must be a plain text file.

To do simple SNP look ups it must have at least one column with the header **SNP**.

To do an MR analysis it must have the following column headers:

- SNP - rs IDs of the instruments for the exposure
- beta - effect sizes for each SNP
- se - standard errors
- effect_allele - Effect allele

It's useful to have these columns too:

- other_allele - Other allele
- eaf - Effect allele frequency

You can see an example file here: [telomere_length.txt](#)

Upload plain text file

Choose file

No file chosen

Preview of uploaded table

Separator

Comma

Space

Tab

Required columns

SNP column name

SNP

Required columns for MR

Effect column name

beta

Standard error column name

se

Effect allele column name

effect_allele

Useful columns for MR

P-value

pval

Phenotype name


Phenotype


Units

units

General info for exposure file

Exposure file types

 University of
BRISTOL

 CANCER
RESEARCH
UK

MRC

Integrative
Epidemiology
Unit

Choose exposure

e.g. MR-Base repository

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 exposure source

- ☐ Manual file upload
- ☐ NHGRI-EBI GWAS catalog
- ☒ MR Base GWAS catalog
- ☐ Gene expression QTLs
- ☐ Protein level QTLs
- ☐ Metabolite level QTLs
- ☐ Drug proxies
- ☐ 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 sentinel SNPs. These SNPs will be used as instruments in the MR analysis.

p-value threshold ☒ Perform clumping

LD Req

Clumping distance (kb)

Display columns


☒ ID ☒ Trait ☒ Consortium ☒ Number of cases ☒ Number of controls ☒ Sample size ☒ PubMedID ☒ Year ☒ cat ☒ ethnic ☒ filename ☒ gender ☒ note ☒ path ☒ priority ☒ public ☒ sd ☒ subcat ☒ trait_strict ☒ unit

Search: body mass index

ID	Trait	Consortium	Number of cases	Number of controls	Sample size	PubMedID	Year	cat	ethnic	filename	gender	note	path	priority	public	sd	subcat	trait_strict	unit
2	body mass index	GIANT			339224	26679413	2015	risk factor	mixed	All_ancestries_SNP_gwas_mc_merge_nogc.tbl.uniq.gz.tab	both gender	GWAS is in standard normal units (inverse normal transformation); SD is to help guide interpretation but is only approximate	/projects/MRC-IEU/publicdata/GWAS_summary_data/GIANT_2014_2015/BM2015	1	public	4.77	anthropometric		
84	body mass index	GIANT	123868	20935630	2010	risk factor	Caucasian	GIANT_BMI_Splicing2010_publicrelease_HapMapCeuFreq.txt.tab.all_pos.all_neg	both gender	GWAS is in standard normal units (inverse normal transformation); SD is to help guide interpretation but is only approximate	/projects/MRC-IEU/publicdata/GWAS_summary_data/GIANT_2014_2015/Waist2015	8	public	4.77	anthropometric				
94	body mass index	GIANT	60506	23754948	2013	risk factor	Caucasian	GIANT_Randall2013PlosGenet_stage1_publicrelease_HapMapCeuFreq_BMI_MEN_N.txt.tab	male	GWAS is in standard normal units (inverse normal transformation); SD is to help guide interpretation but is only approximate	/projects/MRC-IEU/publicdata/GWAS_summary_data/GIANT_2010_2012_2013	5	public	4.77	anthropometric				

Select study

Choose outcome



Welcome to MR Base

About

Acknowledgements

Data access agreement

Perform MR analysis

- Choose exposures
- Choose outcomes**
- Run MR

Quick SNP lookup

Select outcomes for analysis

The MR Base database houses a large collection of summary statistic data from hundreds of GWAS studies. In order to perform two sample MR, the SNPs that were selected for the exposures will be extracted from the outcomes that you select here. Please select the outcomes that you want to test for being causally influenced by the exposures.

Studies available in MR base

Display columns

☐ ID ☒ Trait ☒ Consortium ☒ Number of cases ☒ Number of controls ☒ Sample size ☒ PubmedID ☒ Year ☐ cat ☐ ethnic ☐ filename ☐ gender ☐ note ☐ path ☐ priority ☐ public ☐ sd ☐ subcat ☐ trait_strict ☐ unit

Show 10 entries

Search: ILCCO

Trait	Consortium	Number of cases	Number of controls	Sample size	PubmedID	Year
965 Lung adenocarcinoma	ILCCO	3442	14894	18336	24880342	2014
966 Lung cancer (all)	ILCCO	11348	15861	27209	24880342	2014
967 Lung cancer (squamous cell)	ILCCO	3275	15038	18313	24880342	2014
All	All	All	All	All	All	All

Showing 1 to 3 of 3 entries (filtered from 990 total entries)


Previous 1 Next

e.g. lung cancer

Select study

Run MR analysis

Prune SNPs



Welcome to MR Base

About

Acknowledgements

Data access agreement

Perform MR analysis

Choose exposures

Choose outcomes

Run MR

Quick SNP lookup

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?

☒ Use proxies?

Minimum LD Rsq value

0.6 0.4 0.2 0 0.2 0.4 0.6 0.8 1

☒ Allow palindromic SNPs?

MAF threshold for aligning palindromes

0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5

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:

- ☒ Wald ratio
- ☐ Fixed effects meta analysis (simple SE)
- ☐ Fixed effects meta analysis (delta method)
- ☐ Random effects meta analysis (delta method)
- ☐ Maximum likelihood
- ☒ MR Egger
- ☐ MR Egger (bootstrap)
- ☒ Weighted median
- ☐ Penalised weighted median
- ☒ Inverse variance weighted

Choose analysis method

Submit

Working: Extracting outcome data

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


Perform MR analysis

Submit!

MR Results

Results

Sensitivity analyses



Welcome to MR Base

About

Acknowledgements

Data access agreement

Perform MR analysis

Choose exposures

Choose outcomes

Run MR

MR Results

Quick SNP lookup

Select analysis

Exposure

body mass index || GIANT || 2015 || SD (kg/m²)

Outcome

Lung cancer (all) || ILCCO || 2014

Analysis summary

Exposure: body mass index || GIANT || 2015 || SD (kg/m²)

Outcome: Lung cancer (all) || ILCCO || 2014

Number of SNPs: 85

Save results

Generate HTML report

Downloads

Download harmonised summary statistics

Download MR results

Download leave-one-out sensitivity analysis

Download single SNP MR results

Analysis log

```
Getting selected variables...
Extracting instruments from MR Base...
Performing clumping...
Extracting outcome summary statistics from database...
Harmonising exposure and outcome summary statistics...

The following SNPs were removed due to allele harmonising issues:
rs7551597
rs9579883
rs17081654
rs146518
rs9914578
rs9928136

Performing MR analyses...
Performing sensitivity analyses...
Generating plots...
Analysis complete!
```

MR results

Heterogeneity statistics

Causal direction test

Horizontal pleiotropy

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	nnp	b	se	pval
MR Egger	85	0.04995	0.2407	0.8361
Weighted median	85	0.09709	0.1265	0.4429
Inverse variance weighted	85	0.1516	0.09828	0.1229

Single SNP analysis

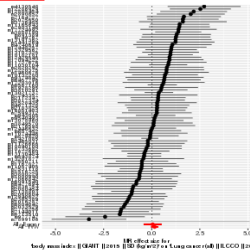
Method comparison plot

Leave-one-out analysis

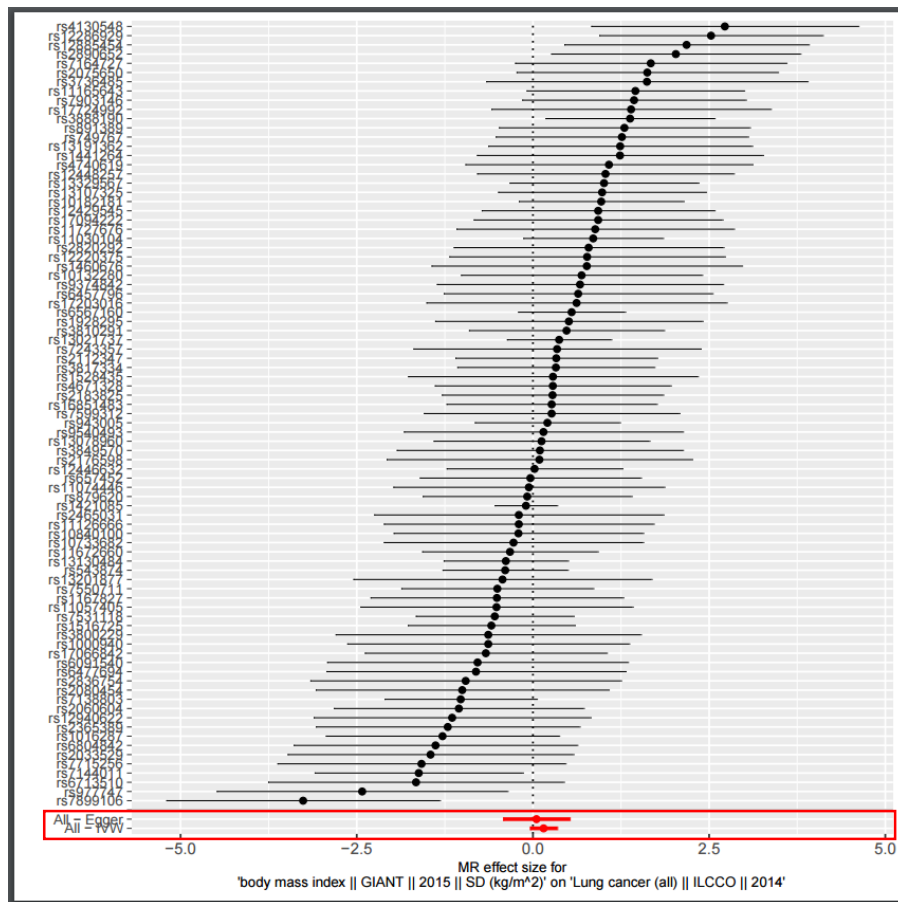
Funnel plot

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.

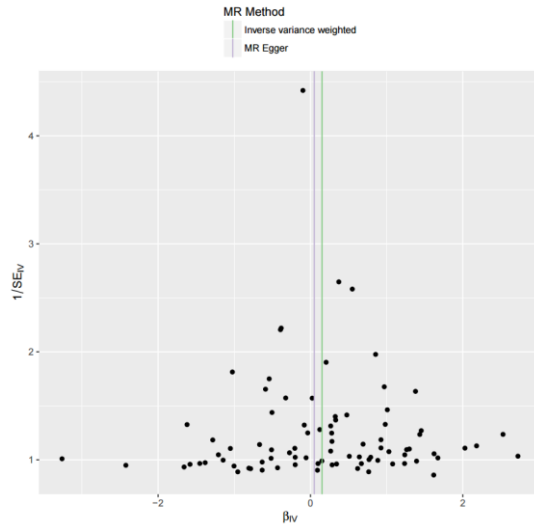
Download PDF of this graph



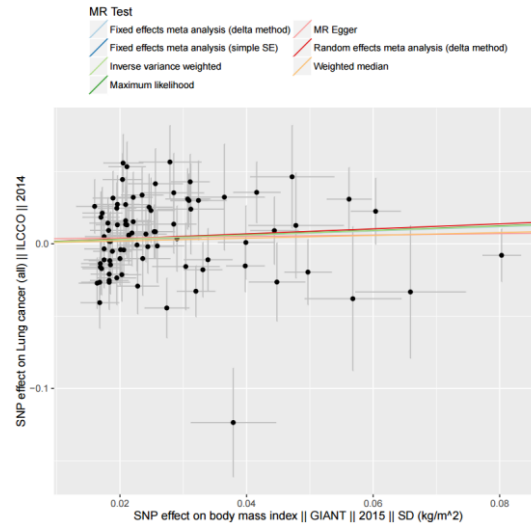
Forest plot



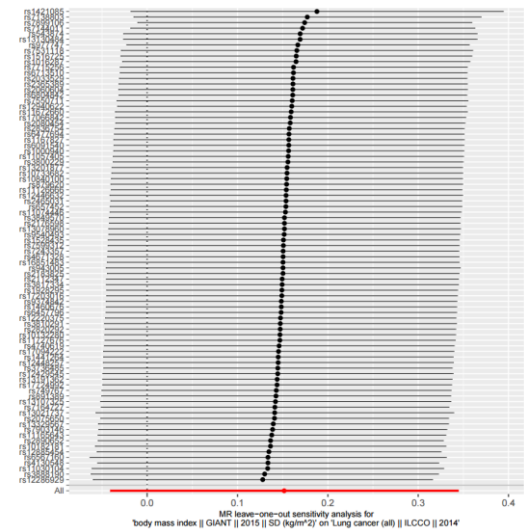
Funnel plot



Comparison of MR results



Sensitivity analysis



Summary and conclusions

- Large database of harmonized summary data
- Automates the MR analysis pipeline
- Increases efficiency of hypothesis-driven and hypothesis-free approaches
- Allows
 - Prioritization of intervention targets
 - Informing of clinical and public health guidelines and advice
 - Improvement of drug discovery pipeline
 - Analysis of secondary outcomes
- Fast and highly cost effective
- Enhanced translation of potential interventions into practice

Acknowledgments

MR-Base team

Dr Philip Haycock
Dr Chris Zheng
Dr Tom Gaunt
Dr Gibran Hemani
Dr Hashem Shihab
Dr Charles Laurin
Dr Benjamin Elsworth

Funders

Cancer Research UK
Medical Research Council

Collaborators

University of Bristol
Prof George Davey Smith
Prof Caroline Relton
Prof Richard Martin
Dr Jack Bowden

University of Cambridge
Dr Stephen Burgess

Databases

dbGAP:
<http://www.ncbi.nlm.nih.gov/gap>
NHGRI GWAS catalog:
<https://www.genome.gov/gwastudies/>

GWAS Consortia

ADIPOGen
ALSOGD
AMD Gene
C4D
CARDIoGRAM
CKDGen
DIAGRAM
EGG
GABRIEL
GCAN
GEFOS

GIANT
GLGC
GPC
GTEx
GUGC
HaemGen
ICBP
IGAP
IIBDGC
ILCCO
IMSGC
Immunobase
MAGIC
MDACC
PANSCAN
PGC
SSGAC
TAG

www.mrbase.org/alpha

