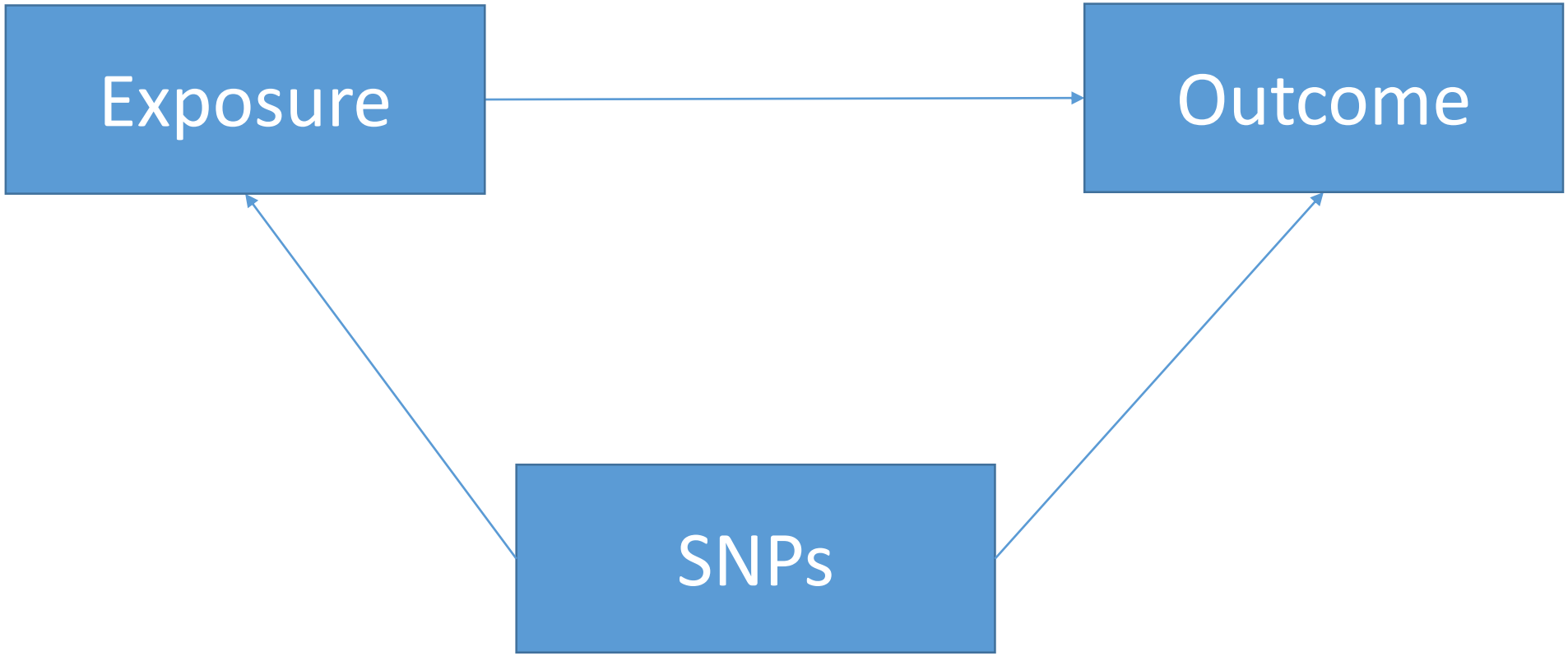
# MR

An overview

# What is MR



# One sample or Two sample

One Sample

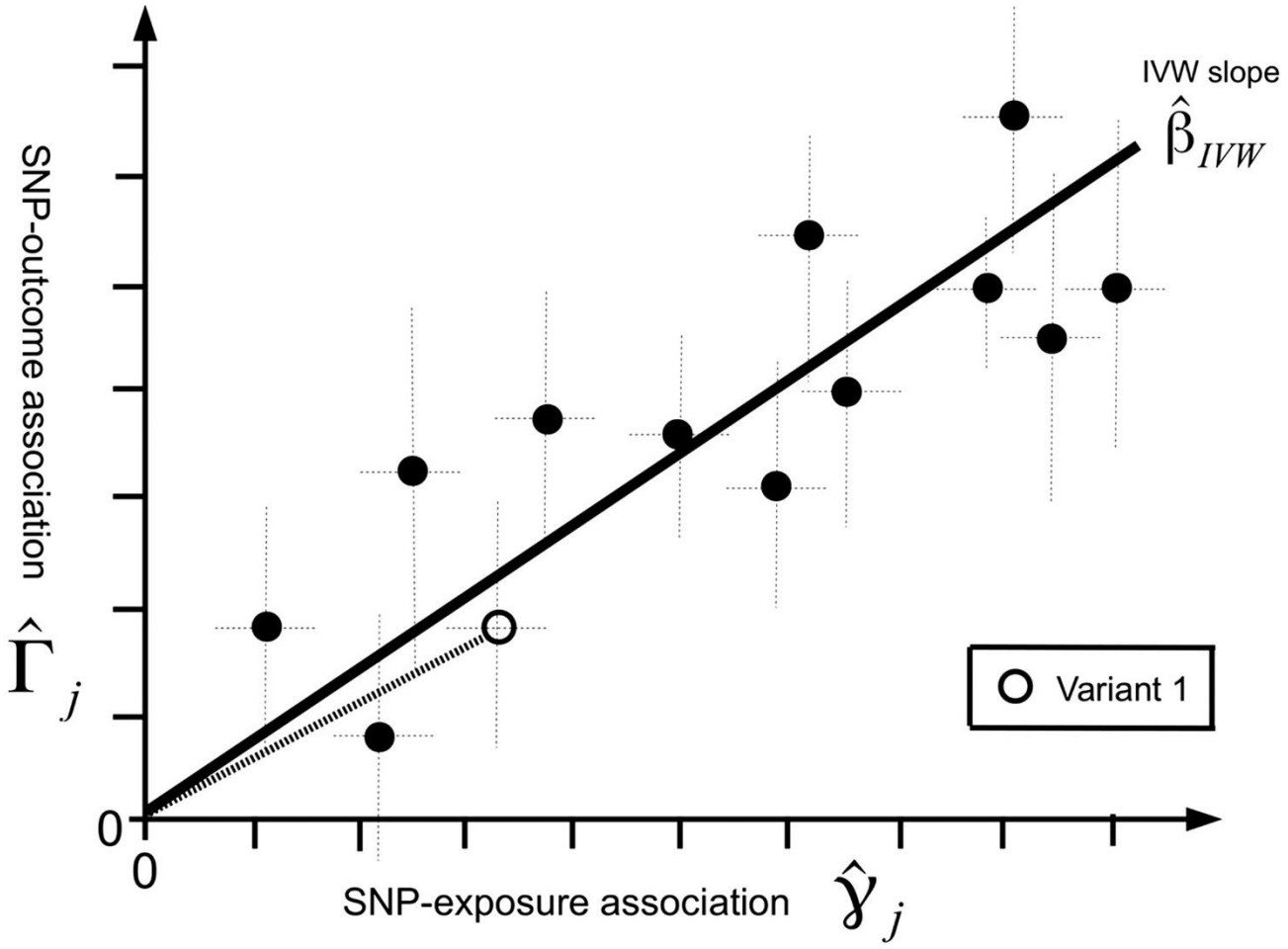
- Generate a score for each **person** in the analysis.
- Use instrumental variable methods to see if people with higher genetic scores have greater risk of disease.
- Sensitivity methods can be used (using two sample methods).
- Possible to control for other genetic scores (and phenotypic measurements).
- Potentially allows for non-linear associations.

Two sample

- Collate data for each **SNP** used in the analysis.
- Relies on a dose-response relationship between the exposure and outcome.
- Can leverage large GWAS studies to leverage SNP numbers and effect estimates.
- Has a range of methods to detect heterogeneity between SNP effects…
- And to detect outiers.
- Can perform multivariate methods, but only to control for genetic variables.

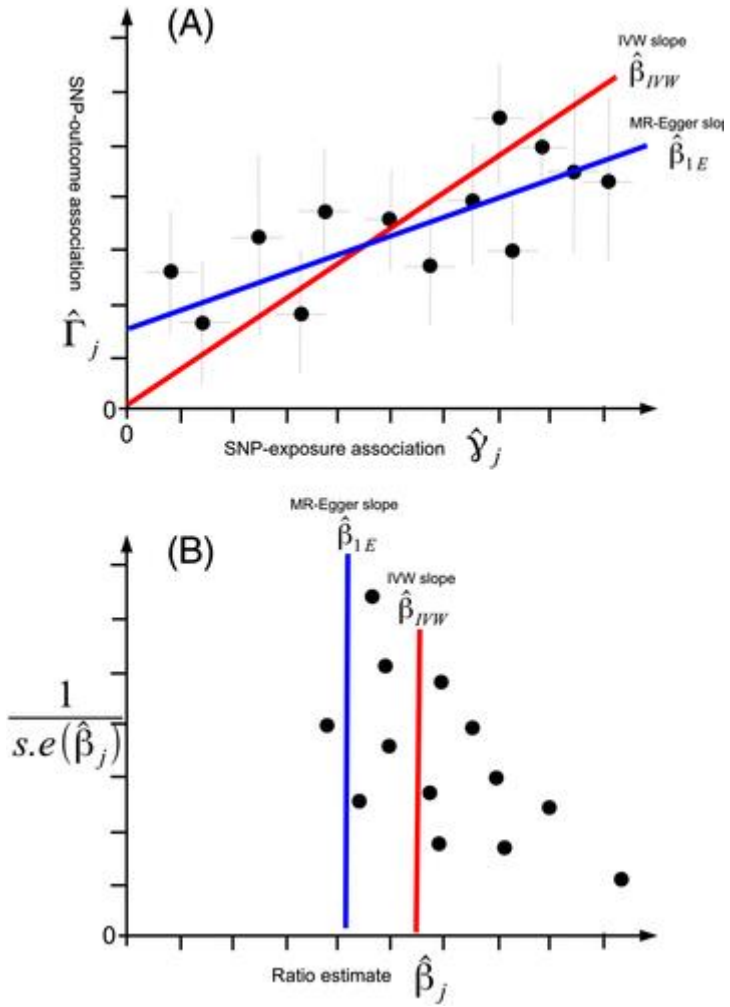
# Methods of MR - IVW

- IVW is the simplest of the MR methods – it draws a line of best fit through the variables.
- It makes the assumption that a value of zero for the exposure means a value of zero for the outcome.
- It also assumes that there is no pleiotropic effects across the SNP
- If these assumptions hold, this method has the most power to detect a signal.



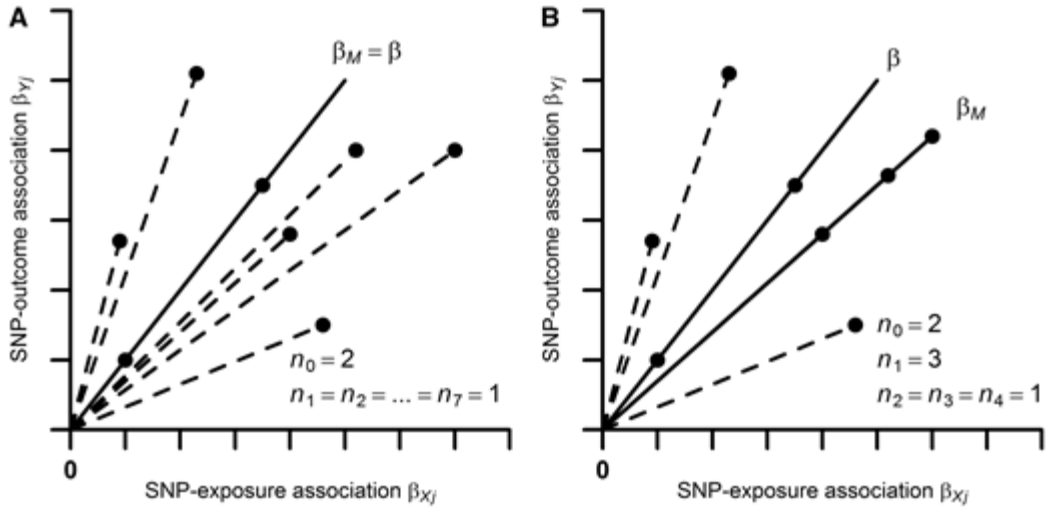
# Methods of MR - Eggers

- Second most common mendian randomisation method.
- Relaxes one of the assumptions of IVW to allow for some pleiotropic effects
- This is represented by the inclusion of the intercept.
- This is important in situation of “unbalanced heterogeneity”.
- The new assumption is that that effect the SNPs have on the confounding factor is not related to their effect on the exposure.



# Methods of MR - Weighted Median and Modal

- More general methods can be used when there is widespread pleiotropy, especially in both directions.
- These methods assume that as least some of the SNPs are valid – a majority in the median method, and a plurality in the mode based analysis.
- The Median method estimates the median effect, and generates standard errors by bootstrapping.
- The modal method generates bandwidths and selects the one that encompasses the greatest number of variants, standard errors also worked out by bootstrapping.



# MR outliers – inference from other phenotypes

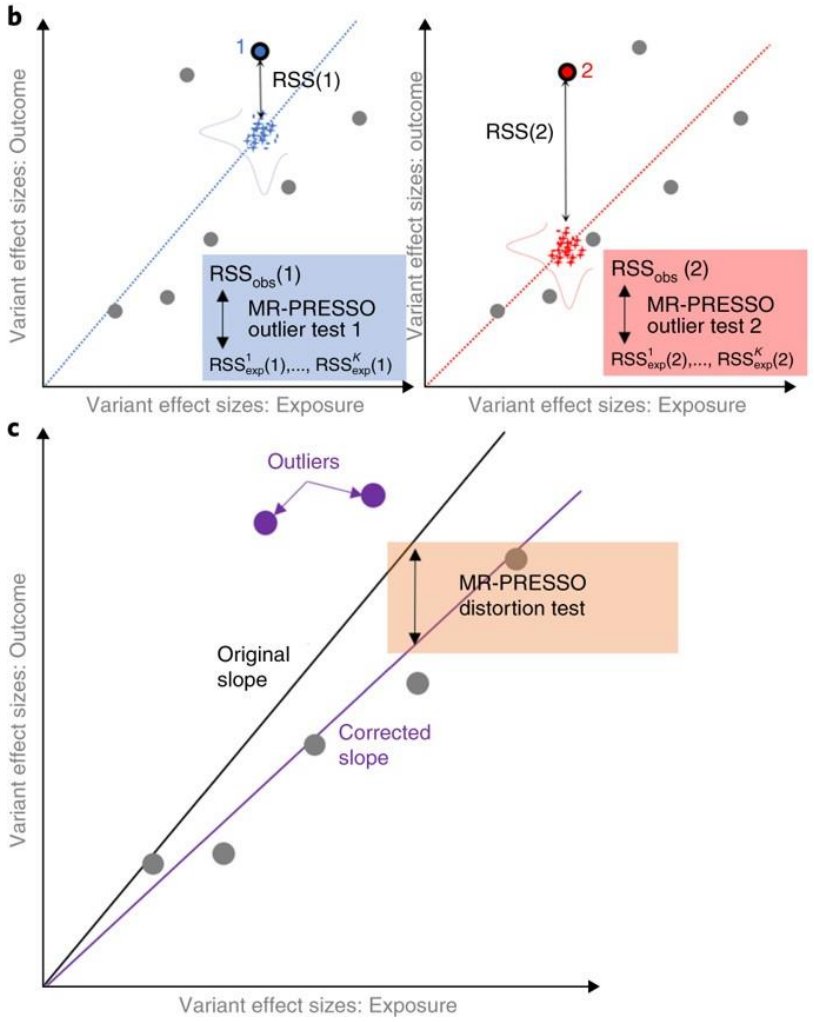
- Some times we might not want to include SNPs in an analysis.
- This can be based on biological knowledge.
  - We might exclude based on the nearest gene being in a pathway, or if the SNP is known to show very population specific effects.
- It can be based on association with other phenotypes
  - Ideally we would have the full data, and could run a proper multivariate MR analysis, but there might be occasions where the full data isn't available.
- Sometimes your outlier is the true effect.

# MR outliers – Stiger

- If a variable is associated with the outcome more strongly than with the exposure, there is likely reverse causality.
- You can't just judge this on p-value as the size of the discovery studies plays a role.
- Instead filter if the $r^2$ is higher for the exposure than the outcome.
- Models run without these SNPs are likely to be resistant to any effects of reverse confounding.

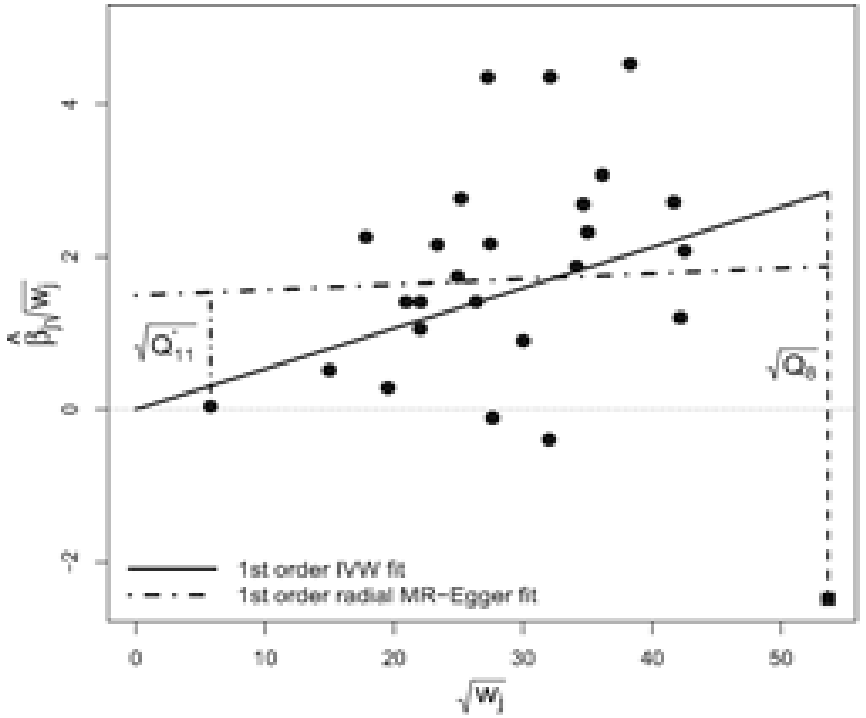
# MR outliers - PRESSO

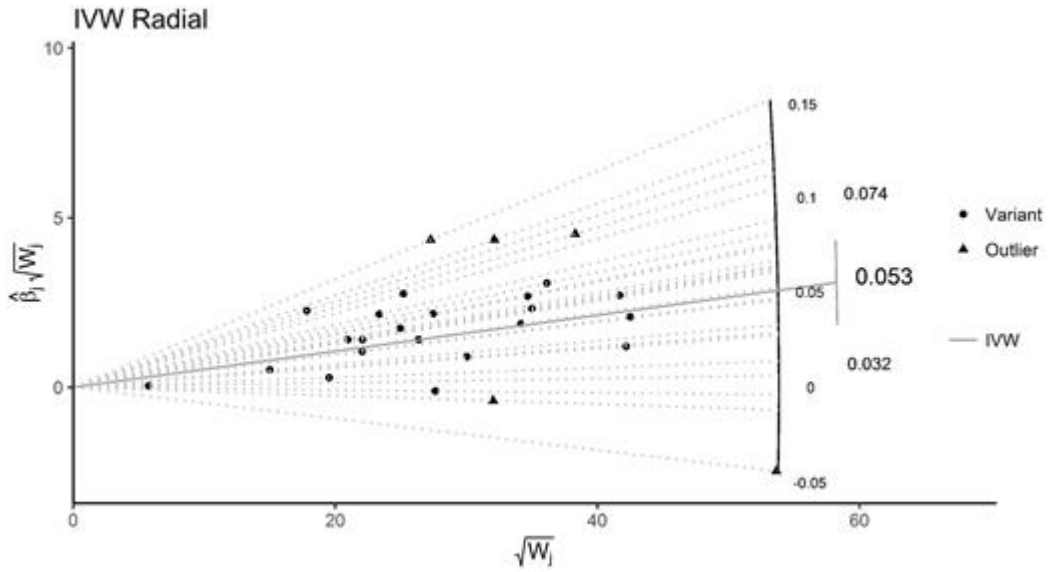
- PRESSO is a method that identifies and remove outlier SNPs.
- The first is that it calculates a Q-statistic as a global measure of heterogeneity.
- The second is to work out the residual sum of squares for each SNP, to a model generated with that SNP excluded.
- These are then compared to a simulated distribution of residual sum of squares
- This model can be performed in a multivariate frame-work
- There are a few issues, it sometimes finds a significant global Q, but does not identify any specific SNP. Also it sometimes fails.



# MR outliers - Radial

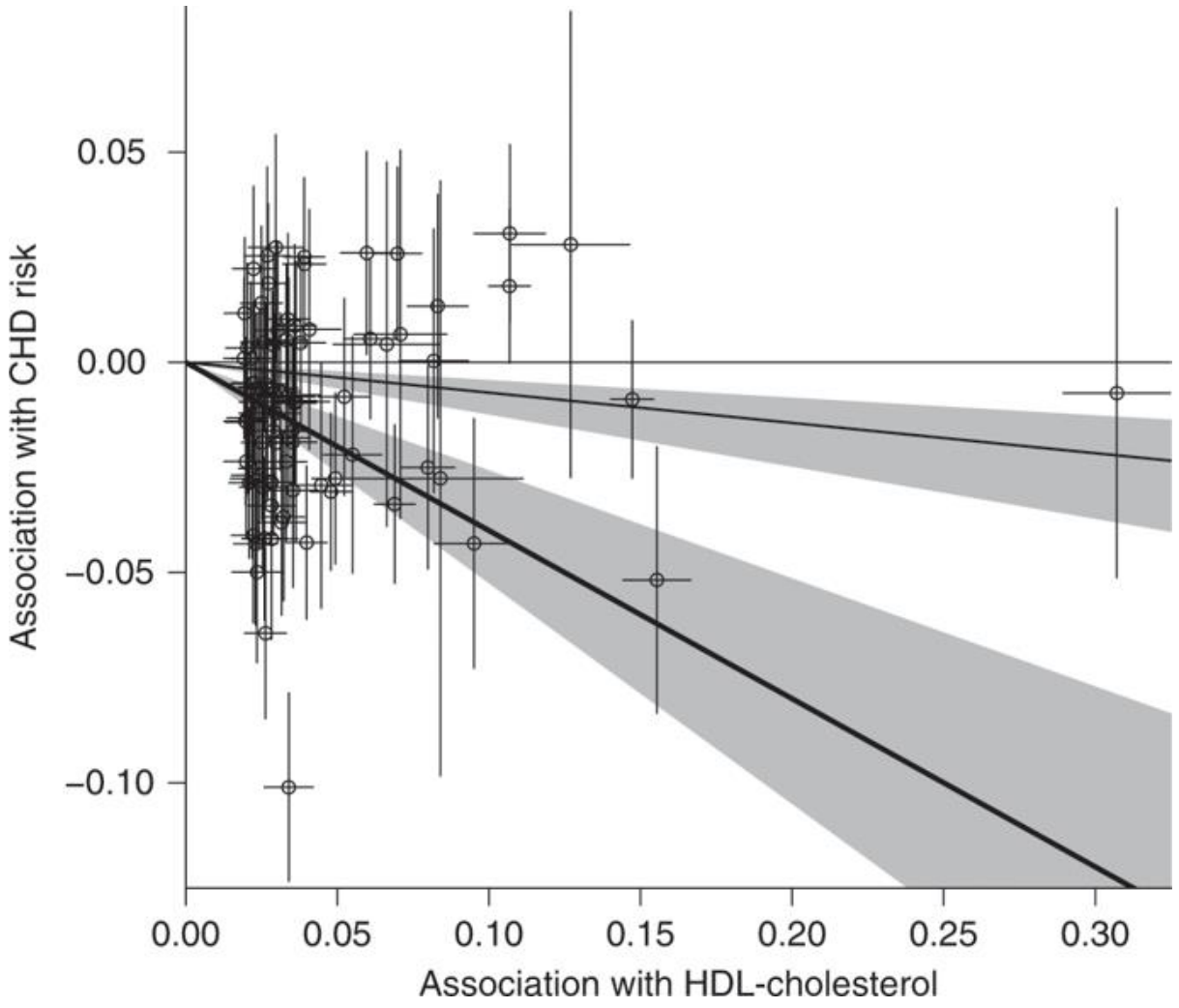
- Based on the radial plot method from meta-analysis.
- SNPs are plotted on this line as the weight (x-axis) vs the weight*beta (y-axis).
- The vertical distance of the SNP to the IVW line (or the EGGERS line using RÜcker's Q) is the contribution of that SNP to the Q value.
- This frame-work allows for a range of possible weights to be used.



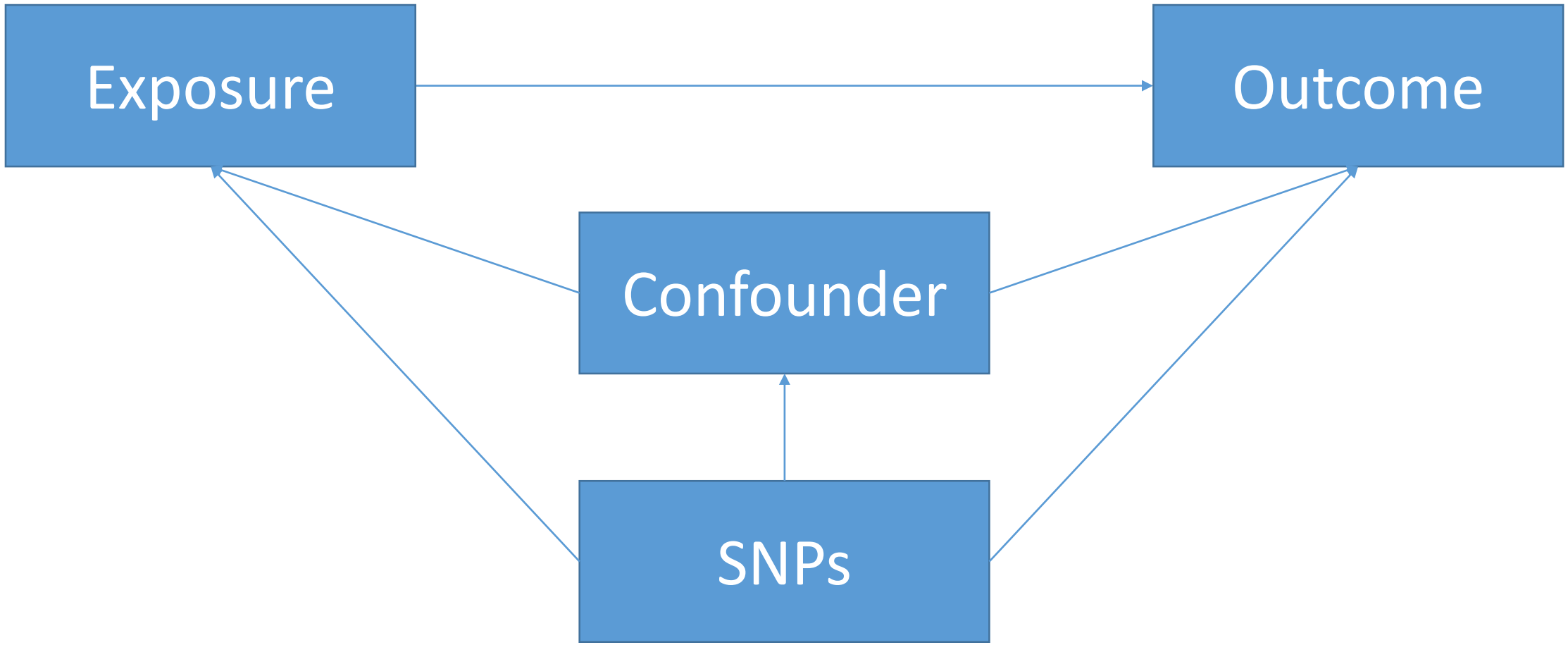



# MR Outliers - Contamination mixture method

- Allows for there being multiple true casual pathways represented in the data.
- Assumes that a plurality of instrument s are valid.
- Compares the likely hood that each variant is valid (is on a non-zero distribution with a small standard deviation) to it being invalid (on a distribution centred on zero with a wide standard deviation). It is this that is used as the weighting.
- It claims to be able to detect where there are multiple clusters as seen in the graph.



# Multivariable MR methods



# Pleiotropy

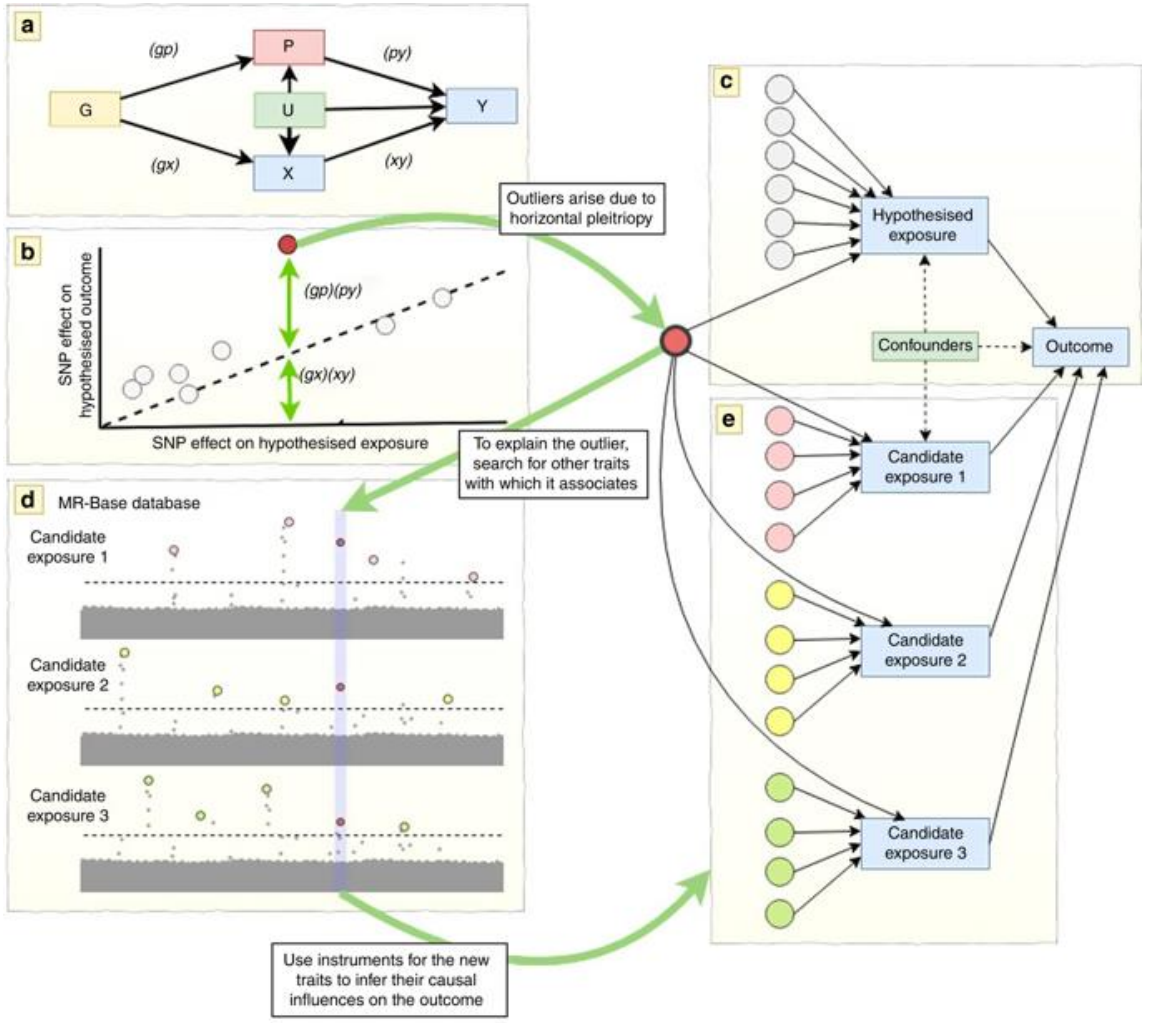
- “Horizontal” – the SNP has effects on two or more phenotypes that relate to different pathways that effect the outcome
- “Vertical” – The SNP ahs effects on two or more phenotypes that effect the outcome on the same pathway
- “Common Soil” – The SNPs have an effect on two or more phenotypes that effect an outcome, but the measured phenotypes are both downstream of the true casual factor

# Multivariable MR methods

- Because we are using a regression model frame-work, we can include other variables as covariates.
- We include the estimates of those SNPs on the covariant.
- There is a discussion about which variants to include
  - My personal take is that it depends on what you want to estimate.
- This can be done in the one sample frame works too, where a second score is used in the model
- Multivariate Mendelian randomisation can be run in both IVW, and MR-EGGER frameworks (as there may be confounding still unaccounted for).

# MR outliers - Tryx

- Works on the idea that outliers might be interesting
- Outliers are identified by the radial method.
- Then MR-TRYX performs a phe-was on the outlying SNPs, to identify phenotypes that might explain the pleiotropy.
- Scores for these phenotypes are then used in a multivariate frame-work to try to understand the causal web

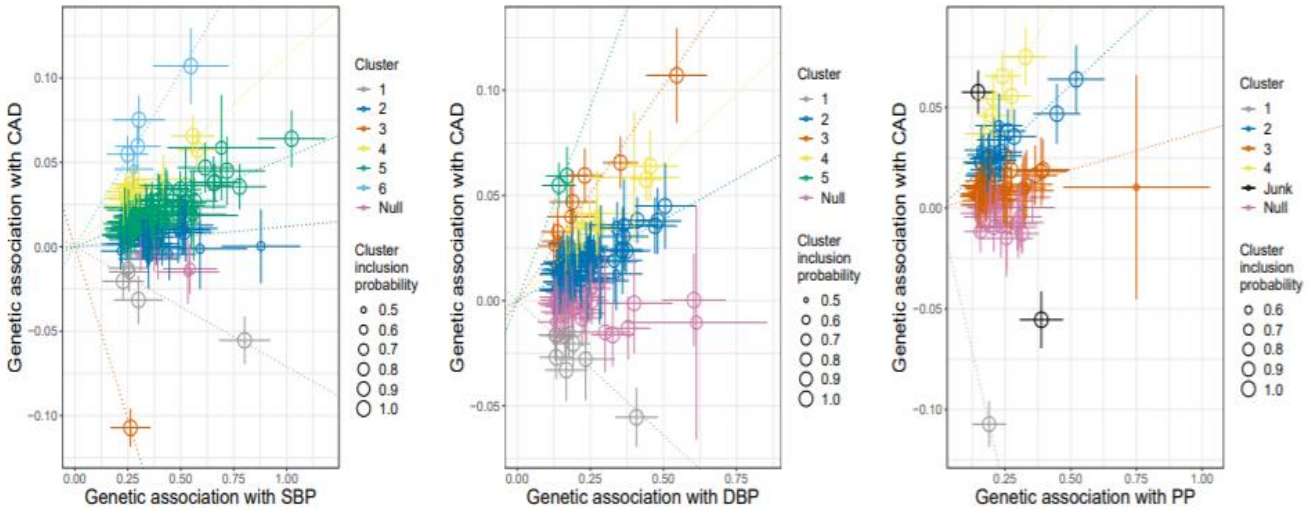


# Pleiotropy

- “Horizontal” – the SNP has effects on two or more phenotypes that relate to different pathways that effect the outcome
- “Vertical” – The SNP ahs effects on two or more phenotypes that effect the outcome on the same pathway
- “Common Soil” – The SNPs have an effect on two or more phenotypes that effect an outcome, but the measured phenotypes are both downstream of the true casual factor

# MR cluster, maybe a misnomer.

- This is based on identifying clusters of variants within the model.
- Do not run MR within the clusters (you'll be forcing it to have results)
- The idea behind this is more focused on biology - in the image you might find that the same SNPs are clustered together across related exposures, you could also do this with related outcomes.
- It might be possible to run wider clustering across a number of outcomes, using the casual estimate as the value driving the clusters.



# MR and bias

- MR studies are based on GWAS data, meaning there may be biases in the data.
- Collider bias might mean effect you exposure measurements
- GWAS studies, especially large scale studies are affected by recruitment bias.
- Remember that you exposure may not be what you think
  - A blood metabolite level might not mirror it’s bio-activity, and different SNPs might work in different ways