Presentation Notes

Background

* An Aging Population: This is an infographic I found on the US Census website and I am sure this comes as no surprise to those of you in this room that the projected number of individuals over 65 are expected to increase, here, we can see that in the United States alone the number of individuals over 65 is expected to almost double by 2060.
* As the population ages, the need to manage and provide healthcare for this distinct population grows
* To this end, can we use already published observational data to provide insight into the what genetic factors contribute causally to healthy aging?

Methods

* It has traditionally been difficult to evaluate observational studies for causal associations in risk factors in human longevity due to confounding , etc.
  + Solution: Mendelian Randomization approach. A statistical approach to find causal associations among observational data
  + Uses genetic variants /SNP/instrumental variable as a proxy for an exposure
  + Mendel’s Law of Inheritance: genes are inherited from parent to child are random.
    - We can utilize this randomness (e.g. an assumption of no confounding associated with SNPs in population) to approximate a RCT using observational data.
      * Genes are the ‘’random” assignments for an exposure or exposure levels
    - Temporality: a requirement for determining causality
      * Genetic factors are determined at birth, therefore our exposure precedes our outcome
      * Can be used in case studies where this can sometimes be an issue
  + Risk factors: limit by genetic associations of risk factors
* Specifically, for this project, we used Two Sample MR, which obtains significant SNP-Exposure from other studies and compares SNP-outcome relationships in another study. In this way we can make causal association among the Exposure-Outcome relationship.
  + Analysis/ Theory

Advantages

* Two Sample MR allows us to use data from publicly resources to investigate the relationship between and exposure and outcome
  + The exposure and outcome do not need to come from the same dataset
* Analysis is fairly simple
* Cost-efficient

Assumption #1 : SNPs truly associated with Exposure, as seen from independent studies

As we saw in the diagram, it is necessary that the SNP is associated with the exposure in order to infer the relationship between the exposure and outcome, when we evaluate the relationship between the SNP and outcome. This is to ensure that the SNP-outcome association is **mediated** through the exposure of interest.

However, how do we know the SNP-Exposure are truly associated? Usually we do this by limiting the SNPs to those that are significantly associated with an exposure from previous studies because we have significant evidence that a true relationship exists between SNP and Exposure

If we suspect that the p-value cut off is not strict enough we might run into several issues. The first is the presence of weak SNPs or genetic instrument variables (gIV) . This will dampen the ability of our analysis to provide an accurate estimate.

??? clarify the logic

Secondly, including SNPs not strongly associated with an exposure will reduce the power of MR analysis. ?? again clarify the logic

* What happens when assumption is violated , towards or away from null
* #1 :

## Assumption 2 SNPs not associated with E-O confounders

This is an important assumption of MR, as the lack of confounders is partly what allows us assume that the randomness of allele inheritance is equivalent to random assignment in an RCT and the subsequent equal distribution of confounders or lack thereof allows us to make causal estimates of the association between E and O.

The problem arises from trying to prove this positively, since most biological pathways are complex and it is unlikely that we know all the confounders in the E-O relationship.

However, this can be tested to some degree with known confounders to detect the presence of confounding. If there is none, we would feel more confident in the estimates determine in our analysis.

citation: Zheng, J., Baird, D., Borges, M. et al. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep 4, 330–345 (2017). https://doi.org/10.1007/s40471-017-0128-6

## Assumption 3 : SNPs are associated with outcome through Exposure only

By making this assumption we are saying the estimate from MR analysis is due completely to this SNP. If this is not true that the SNP-outcome is mediated through some other exposure and muddle what we can say about the true association between E and O. However, there are statistical techniques that are more robust to horizontal pleiotropy which we can use if we suspect that it is present. In this way we can begin our analysis with the assumption of no horizontal pleiotropy and use techniques to try to detect its presence and adjust for it.

As I mentioned when this assumption is violated we assume the presence of horizontal pleiotropy, that is the estimate we see between SNP and outcome, does not occur completely through our exposure, so we need to adjust our analysis to capture how much variation in our outcome is due to the exposure. The most common way to do this is via MR Egger.

- MR Egger

- InSIDE(Instrument Strength Independent of Direct Effect)

assumes no correlation between SNP-exposure association and direct efect of SNP-outcome

- suffers from low power

- weak instruments can influence the regression and therefore the estimate

- Weighted Median approach

- allows up to 50% of SNPs to be weak or invalid instruments

- Mode based approach

- futher relaxes pleiotropy assumption requirements

- weak instruments

\*\*Violations\*\*

- trait heterogeneity

- difficult to determine without further understanding of biological pathway

citation: Zheng, J., Baird, D., Borges, M. et al. Recent Developments in Mendelian Randomization Studies. Curr Epidemiol Rep 4, 330–345 (2017). https://doi.org/10.1007/s40471-017-0128-6

## Other Limitations

- developmenal compensation /canalization

- solution: careful interpretation of results, realizing that both environmental and genetic factors influence an outcome.

- bias towards null

- winner's curse

- no real solution, must view results in conjunction with other studies and what we already know

- due to underpowered GWAS studies 🡪 first rsid that was significant

- not typically a problem with current GWAS studies

- move to assumption 1, SNPs are not truly associated with E even though that is what is reported

- collider bias

- be careful in how we interpret the results of our analysis. Based on current knowledge, could the relationship we see in our study be due to other biases like loss to follow up.

- heterogeneity of effect

- estimates are not consistent across independent instruments

- use Cochran's Q statistic(IVW) or Rucker's Q statistic(Egger) to test for this

Limitations

Developmental compensation – unlike RCTs genetic variants reflect a life-long exposure, (not one of short duration), and associations might be due to developmental compensations that arise due this exposure, not the exposure itself

Table 1 . Common rsids

Table 2-5 : Rsids are unique to each dataset

Table 6-7: exposures that are common by UK biobank, and Exposures that are common with Deleen

Table of counts : number of exposures that are unique or common to each dataset

Venn Diagram of four datasets