Mendelian randomisation

Nic Timpson
Objectives:

Build on “epigenetic epidemiology” and “strengthening causal inference” sessions

Introduce MR specifically as a means to strengthening causal inference

Illustrate the process/underpinnings of MR give worked examples

Should should be able to:

Describe what MR is

Describe why MR allows causal inference

Describe an example of MR

Formulate a crude MR experiment
What is Mendelian randomisation?

*Genetic association study or observational epidemiology?*

- Conceptual approach to Mendelian randomisation should be taken from the point of view of conventional epidemiology.

- Observational epidemiology is limited, however problems can be avoided given the existence of “better measures”.

- In an observational framework what could we use as a “better measure”?

- Mendelian randomisation aims to use genetic variation related to risk factors of interest in efforts to re-assess observational estimates.
Assumptions of MR analysis

- Z associated with X
- Z is independent of U
- Z is independent of Y given U and X

X → Y
Why are genotypes good instruments?

An instrumental variable is a variable that is:

(i) Associated with an exposure of interest

(ii) NOT associated with the outcome of interest – except through its association with the exposure of interest.
Clustered Environments and Randomized Genes: A Fundamental Distinction between Conventional and Genetic Epidemiology

George Davey Smith¹,²*, Debbie A. Lawlor¹,², Roger Harbord¹, Nic Timpson¹,², Ian Day¹,², Shah Ebrahim³

¹ Department of Social Medicine, University of Bristol, Bristol, United Kingdom, ² Medical Research Council Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol, United Kingdom, ³ Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

<table>
<thead>
<tr>
<th></th>
<th>Number of comparisons</th>
<th>Expected number</th>
</tr>
</thead>
<tbody>
<tr>
<td>96 non-genetic variables</td>
<td>4560</td>
<td>228</td>
</tr>
<tr>
<td>23 candidate loci</td>
<td>253</td>
<td>12.7</td>
</tr>
</tbody>
</table>

Correlations here taken at $p \leq 0.05$
Total number of phenotypes: 121
Total number of pairs: 7260

<table>
<thead>
<tr>
<th>p value</th>
<th>0.05</th>
<th>0.01</th>
<th>0.0001</th>
<th>6.88 x 10^{-6} *</th>
<th>1.37 x 10^{-6} *</th>
<th>1.37 x 10^{-8} *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected</td>
<td>363</td>
<td>72</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Observed</td>
<td>1843</td>
<td>1281</td>
<td>684</td>
<td>568</td>
<td>504</td>
<td>421</td>
</tr>
</tbody>
</table>

*bonferroni corrected results of significance values 0.05, 0.01 and 0.0001, respectively

• Phenotypes are highly over correlated, as expected.
Conventional observational epidemiology

Modifiable exposure (Phenotype, affected by genotype and by environment, e.g. CRP)

Outcome (Usually a disease or health status, e.g. coronary heart disease)

Confounders (Factors associated with both exposure and outcome, including unmeasured confounders and confounders measured with error, e.g. smoking, socioeconomic position, diet, physical activity)

Conventional observational epidemiology

It is often impossible to exclude unmeasured or residual confounding as an explanation for observed exposure/outcome associations

Epigenetic Epidemiology 2012
**Mendelian randomisation approach**

**Confounders**
(Factors associated with both exposure and outcome, including unmeasured confounders and confounders measured with error, e.g. smoking, socioeconomic position, diet, physical activity)

**Instrumental variable**
(Genotype, randomly determined, e.g. CRP)

**Modifiable exposure**
(Phenotype, affected by genotype and by environment, e.g. CRP)

**Outcome**
(Usually a disease or health status, e.g. coronary heart disease)
Mendelian randomisation approach

**Instrumental variable**
(Genotype, randomly determined, e.g. CRP)

**Modifiable exposure**
(Phenotype, affected by genotype and by environment, e.g. CRP)

**Outcome**
(Usually a disease or health status, e.g. coronary heart disease)

**Confounders**
(Factors associated with both exposure and outcome, including unmeasured confounders and confounders measured with error, e.g. smoking, socioeconomic position, diet, physical activity)
Example – *using* instruments for alcohol consumption
Metabolism of alcohol

Ethanol $\rightarrow$ Acetaldehyde $\rightarrow$ Acetic acid

ADH

CYP2E1

ALDH

* Mainly occurs in the liver, but some activity is also present in the oral cavity and digestive tract
ALDH2 genotype by alcohol consumption, g/day: 5 studies, n=6815

Relationship between characteristics and ALDH2 genotype

Age

Smoker

BMI

Cholesterol

Epigenetic Epidemiology 2012
ALDH2 genotype and systolic blood pressure

(heterozygote vs heterozygote)

(homozygote vs homozygote)
Example – using instruments for adiposity

\[ BMI \rightarrow \text{Adiposity} \rightarrow Y \]
GIANT_3 (n~250,000) – Speliotes et al 2010
BUT, are these “genes for obesity”? 

Copenhagen data – n~38000 cross sectional 
FTO effect ~0.1SD across the distribution 

Effect mirrored in a simple Cumulative frequency plot 

Epigenetic Epidemiology 2012
Example – using instruments for adiposity

- **FTO** → **Adiposity**
- **U**
- **Traits of metabolism**
  - **CRP/BMI**
  - **IHD**
Common Variation in the FTO Gene Alters Diabetes-Related Metabolic Traits to the Extent Expected Given Its Effect on BMI

Rachel M. Freathy,1 Nicholas J. Timpson,2,3 Debbie A. Lawlor,3,4 Anneli Pouta,5 Yoav Ben-Shlomo,4 Aimo Ruokonen,2 Shah Ebrahim,6 Beverley Shields,1 Eleftheria Zeggini,5 Michael N. Weedon,1 Cecilia M. Lindgren,2,7 Hana Lango,1 David Meizner,1 Luigi Ferrucci,6 Giuseppe Paolisso,9 Matthew J. Neville,7 Fredrik Karpe,7 Colin N.A. Palmer,10 Andrew D. Morris,10 Paul Elliott,11 Marjo-Riitta Jarvelin,9,11 George Davey Smith,3,4 Mark I. McCarthy,2,7 Andrew T. Hattersley,1 and Timothy M. Frayling1

Diabetes, 2008
Does Greater Adiposity Increase Blood Pressure and Hypertension Risk?
Mendelian Randomization Using the FTO/MC4R Genotype

Nicholas Timpson, Roger Harbord, George Davey Smith, Jeppe Zacho, Anne Tybjærg-Hansen, Børge G. Nordestgaard

<table>
<thead>
<tr>
<th>Variable</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>OR</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BMI tertile</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% in lowest education</td>
<td>29.49</td>
<td>37.31</td>
<td>47.72</td>
<td></td>
<td>9.7-106</td>
</tr>
<tr>
<td></td>
<td>(28.58, 30.42)</td>
<td>(36.35, 38.28)</td>
<td>(46.74, 48.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FTO genotypes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>% in lowest education</td>
<td>38.5</td>
<td>38.37</td>
<td>38.15</td>
<td></td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>(37.56, 39.45)</td>
<td>(37.57, 39.17)</td>
<td>(36.77, 39.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MC4R genotypes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.7</td>
</tr>
<tr>
<td>% in lowest education</td>
<td>38.39</td>
<td>38.29</td>
<td>38.77</td>
<td></td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>(37.65, 39.14)</td>
<td>(37.38, 39.21)</td>
<td>(36.53, 41.05)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Timpson et al Hypertension 2009
**Components of analysis**

Observational effects/trend

Linear regression result

Adjusted linear regression result

Instrumental variable regression result

A test for difference ($p(DWH)$)
MR not just with continuous outcomes...

MR not just with simple instruments... 

Histograms of BMI by the two groups generated from the two groups from figure 1. Red lines indicate the mean BMI for each – these are ~20 and 23 kg/m$^2$ respectively.

** A change of threshold to a score using the top 50,000 markers from the GIANT GWAS for BMI takes advantage of extra variance explained (>3.5% - continuous)
Limitations to Mendelian randomisation

1- Heterogeneity
   
   - Genetic heterogeneity through *linkage disequilibrium* or by *pleiotrophy*.
   
   - Phenotypic heterogeneity & complexity (e.g. EC-SOD).
   
   - Population genetic heterogeneity in instrument choice.
LMS allowing adjustment for age-specific heteroscedasticity and skewness. Median (or \(M\)) curves were modelled with 9 degrees of freedom, with corresponding curves for the coefficient of variation (\(S\)) with 5 d.f. and skewness (\(L\)) with 3.
Limitations to Mendelian randomisation

1- Heterogeneity
   - Genetic heterogeneity through *linkage disequilibrium*
     or by *pleiotrophy*.
   - Phenotypic heterogeneity & complexity (e.g. EC-SOD).
   - Population genetic heterogeneity in instrument choice.

2- Population stratification
Limitations to Mendelian randomisation

1- Heterogeneity
   - Genetic heterogeneity through *linkage disequilibrium*
     or by *pleiotrophy*.
   - Phenotypic heterogeneity & complexity (e.g. EC-SOD).
   - Population genetic heterogeneity in instrument choice.

2- Population stratification

3- Canalisation

   ‘the cell can take specific permitted trajectories, leading to different cell fates’
Limitations to Mendelian randomisation

1- Heterogeneity
   - Genetic heterogeneity through linkage disequilibrium or by pleiotrophy.
   - Phenotypic heterogeneity & complexity (e.g. EC-SOD).
   - Population genetic heterogeneity in instrument choice.

2- Population stratification

3- Canalisation

4- Power (instrument “strength” and basic power)
Can Lactase Persistence Genotype Be Used to Reassess the Relationship between Renal Cell Carcinoma and Milk Drinking? Potentials and Problems in the Application of Modifiable Risk

<table>
<thead>
<tr>
<th>Nonpersistent</th>
<th>Persistent</th>
<th>$P_{\text{unadj}}$</th>
<th>$\text{OR}_{\text{adj}}$ (95% CI)</th>
<th>$P_{\text{adj}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{CC}$ $n$ (%)</td>
<td>$\text{CT+TT}$ $n$ (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cases</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romania</td>
<td>51 (59.3)</td>
<td>35 (40.7)</td>
<td>95 (55.2)</td>
<td>77 (44.8)</td>
</tr>
<tr>
<td>Poland</td>
<td>30 (37.0)</td>
<td>51 (63.0)</td>
<td>296 (37.4)</td>
<td>495 (62.6)</td>
</tr>
<tr>
<td>Russia</td>
<td>121 (42.5)</td>
<td>164 (57.5)</td>
<td>324 (41.2)</td>
<td>462 (58.8)</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>126 (27.2)</td>
<td>337 (72.8)</td>
<td>176 (29.5)</td>
<td>421 (70.5)</td>
</tr>
<tr>
<td><strong>All countries</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heterogeneity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Numbers and proportion of controls by genotype and milk-drinking category. $P_{\text{unadj}}$ represents an unadjusted $\chi^2$ test, whereas $\text{OR}_{\text{adj}}$ and $P_{\text{adj}}$ represent a logistic regression analysis for the odds of being in the persistence group by milk drinking status adjusted for age, sex, education, smoking, and drinking.
“Importantly, despite its size… study had low power…”

- relatively weak relationship between genotype and milk consumption

(an often ignored characteristic in the examination of lactase persistence genotypes & with weak instruments comes instability and bias - Burgess)
Objectives:

Build on “epigenetic epidemiology” and “strengthening causal inference” sessions

Introduce MR specifically as a means to strengthening causal inference

Illustrate the process/underpinnings of MR give worked examples

Should should be able to:

Describe what MR is

Describe why MR allows causal inference

Describe an example of MR

Formulate a crude MR experiment
References


- Freathy, R.M. *et al*. Common variation in the *FTO* gene has a metabolic impact consistent with its effect on BMI. *Diabetes* 57, 1419-1426 (2008).


