Mendelian Randomization: Genes as Instrumental Variables

David Evans
University of Queensland
This Session

• What is Mendelian Randomization (MR)?

• Examples of MR in research

• Some ideas

• Using R to perform MR
Some Criticisms of GWA Studies...

• So you have a new GWAS hit for a disease... so what!?
• You can’t change people’s genotypes (at least not yet)
• You can however modify people’s environments...
• Mendelian Randomization is a method of using genetics to inform us about associations in traditional observational epidemiology
RCTs are the Gold Standard in Inferring Causality

**RANDOMISED CONTROLLED TRIAL**

**RANDOMISATION METHOD**

**EXPOSED:** INTERVENTION

**CONTROL:** NO INTERVENTION

**CONFOUNDERS EQUAL BETWEEN GROUPS**

**OUTCOMES COMPARED BETWEEN GROUPS**
Observational Studies

• RCTs are expensive and not always ethical or practically feasible

• Association between environmental exposures and disease can be assessed by observational epidemiological studies like case-control studies or cohort studies

• The interpretation of these studies in terms of causality is problematic
CHD risk according to duration of current Vitamin E supplement use compared to no use.

Rimm et al. NEJM 1993; 328: 1450-6
May 20, 1993

Vitamin E Greatly Reduces Risk Of Heart Disease, Studies Suggest

By JANE E. BRODY

Two new studies of more than 120,000 men and women strongly suggest that supplements of vitamin E can significantly reduce the risk of disease, researchers and other experts cautioned against rushing out to buy the vitamin supplements before further clinical trials confirm that they are beneficial.

The studies, by researchers at the Harvard School of Public Health and Brigham and Women's Hospital in Boston, showed that initially healthy participants were less likely to develop coronary disease at a rate about 40 percent lower than comparable men and women whose intake of this vitamin was lowest. The preventive effect held when the participants' blood levels of cholesterol.

The greatest protection was found at levels of about 100 international units of vitamin E a day for more than two years. The Federal recommendation is to consume fewer than 25 units from foods like vegetable oils, wheat germ, seeds, whole grains and nuts.

The researchers said vitamin E, as an antioxidant, might reduce heart disease by having an effect on low-density lipoprotein cholesterol, or LDL, which type of cholesterol damages arteries primarily after it has been oxidized.

The new findings, which appear today in The New England Journal of Medicine, are some of the first to find health benefits from taking large-dose, or "megadoses" of vitamins as a popular remedy whose value is unproven. Expert Urge Caution

While a person might conclude from the findings that it would be wise to take large doses of vitamin E supplements daily, their long-term safety has not been established.
The average American lifespan has increased nearly 3 years over the last 2 decades.*

We’ve been selling vitamins at a discount since 1977.

Coincidence? We don’t think so.

At VitaminShoppe.com we see vitamins as an essential part of a healthy life – not a luxury. And our pricing reflects that philosophy. Right now we are taking 40% off every item we stock. After 23 years in the vitamin business, we’ve learned how to assemble the finest vitamins, minerals, and supplements at the lowest prices...all 18,000 of them.
Use of vitamin supplements by US adults, 1987-2000

Vitamin E levels and risk factors: Women’s Heart Health Study

- Childhood SES ↓
- Manual social class ↓
- No car access ↓
- State pension only ↓
- Smoker ↓
- Daily alcohol ↑
- Exercise ↑
- Low fat diet ↑
- Obese ↓
- Height ↑
- Leg length ↑

Lawlor et al, Lancet 2004
Vitamin E supplement use and risk of Coronary Heart Disease

Stampfer et al NEJM 1993; 328: 144-9; Rimm et al NEJM 1993; 328: 1450-6; Eidelman et al Arch Intern Med 2004; 164:1552-6
"Well, so much for antioxidants."
Classic limitations to “observational” science

• Confounding

• Reverse Causation

• Bias
An Alternative to RCTs: Mendelian randomization

In genetic association studies the laws of Mendelian genetics imply that comparison of groups of individuals defined by genotype should only differ with respect to the locus under study (and closely related loci in linkage disequilibrium with the locus under study)

Genotypes can proxy for some modifiable risk factors, and there should be no confounding of genotype by behavioural, socioeconomic or physiological factors (excepting those influenced by alleles at closely proximate loci or due to population stratification)
Mendelian randomisation and RCTs

**Mendelian Randomisation**

- **Random Segregation of Alleles**
  - Exposed: Functional Alleles
  - Control: Null Alleles
  - Confounders Equal Between Groups
  - Outcomes Compared Between Groups

**Randomised Controlled Trial**

- **Randomisation Method**
  - Exposed: Intervention
  - Control: No Intervention
  - Confounders Equal Between Groups
  - Outcomes Compared Between Groups
Assumptions of Mendelian randomisation analysis

\[ X \rightarrow Y \]
Examples – *using* instruments for adiposity
Examples – *using* instruments for adiposity
In a Nutshell

• If adiposity DOES NOT causally affect metabolic traits, then the FTO variant should NOT be related to these metabolic traits

• If adiposity causally affects metabolic traits, then the FTO variant should also be related to these metabolic traits

• In this situation, the causal effect of adiposity can be estimated using an “instrumental variables analysis” as fitted by two stage least squares
Do intermediate metabolic traits differ as one would expect given a *FTO*-BMI effect?

Given the per allele *FTO* effect of ~0.1SD and known observational estimates one can derive an expected, per allele, effect on metabolic traits

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Expected Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting insulin</td>
<td>0.038 (0.033, 0.043)</td>
</tr>
<tr>
<td>Fasting Glucose</td>
<td>0.018 (0.014, 0.021)</td>
</tr>
<tr>
<td>Fasting HDL</td>
<td>-0.026 (-0.029, -0.023)</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

N~12,000 samples of European ancestry
Examples – using instruments for adiposity

FTO → Adiposity → Traits of metabolism

U

CRP
Bidirectional MR
CRP and BMI

- C-Reactive Protein (CRP) is a biomarker of inflammation
- It is associated with BMI, metabolic syndrome, CHD and a number of other diseases
- It is unclear whether these observational relationships are causal or due to confounding or reverse causality
- This question is important from the perspective of drug development
“Bi-directional Mendelian Randomization”
<table>
<thead>
<tr>
<th>Outcome / explanatory variable</th>
<th>Observational</th>
<th>Instrumental variable</th>
<th>$P_{IV}$</th>
<th>$P_{diff}$</th>
<th>$F_{first}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP/BMI</td>
<td>1.075</td>
<td>1.06</td>
<td>0.002</td>
<td>0.6</td>
<td>50.2</td>
</tr>
<tr>
<td></td>
<td>(1.073, 1.077)</td>
<td>(1.02, 1.11)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Informative Interactions
Light drinking in pregnancy may be good for baby boys, says study
Researchers find fewer behavioural problems and higher test scores at age 3

Sarah Boseley, health editor
The Guardian, Friday 31 October 2008
Article history
Mom's light drinking doesn't harm baby

REUTERS, Oct 6, 2010, 10.15am IST

Pregnant drinkers: drinkaware.co.uk - Want to know how alcohol may be affecting your baby? Find out today

Women who have one or two alcoholic drinks a week during pregnancy do not harm their children’s behavioural or intellectual development, according to a study by British scientists.

The researchers found that pregnant women who drank up to a glass (175 millilitres) of wine, up to 50 ml of spirits or just under a pint of beer a week did not affect their children. But children whose mothers were heavy drinkers were more likely to be hyperactive and have behavioural and emotional problems than those whose mothers did not drink during pregnancy, the scientists said.
Total difficulties in top 10% of scores by mother’s drinking category

## Maternal Alcohol Dehydrogenase and Offspring IQ

### Table 2. Results for adjusted model including 4 child variants.

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Maternal drinking during pregnancy</th>
<th></th>
<th>Non-drinkers N = 1375</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>&lt;1-6 units per week N = 2792</td>
<td>&lt;1-6 units per week N = 2792</td>
<td>&lt;1-6 units per week N = 2792</td>
<td>&lt;1-6 units per week N = 2792</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Per allele effect on WISC score &amp; 95% confidence intervals</td>
<td>P-value</td>
<td>Per allele effect on WISC score &amp; 95% confidence intervals</td>
<td>P-value</td>
</tr>
<tr>
<td>Child</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADH1A</td>
<td>rs2866151 &amp; 95% confidence intervals</td>
<td>0.004</td>
<td>−0.38 &amp; 95% confidence intervals</td>
<td>0.004</td>
<td>−0.38 &amp; 95% confidence intervals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>−1.95 &amp; 95% confidence intervals</td>
<td></td>
<td>(−2.47 to 1.71)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>−3.29 to −0.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADH1A</td>
<td>rs975833 &amp; 95% confidence intervals</td>
<td>0.03</td>
<td>−0.66 &amp; 95% confidence intervals</td>
<td>0.03</td>
<td>−0.66 &amp; 95% confidence intervals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>−1.72 &amp; 95% confidence intervals</td>
<td></td>
<td>(−2.90 to 1.59)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>−3.23 to −0.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADH1B</td>
<td>rs4147536 &amp; 95% confidence intervals</td>
<td>0.05</td>
<td>−0.71 &amp; 95% confidence intervals</td>
<td>0.05</td>
<td>−0.71 &amp; 95% confidence intervals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>−1.47 &amp; 95% confidence intervals</td>
<td></td>
<td>(−2.92 to 1.50)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(−2.97 to 0.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADH7</td>
<td>rs284779 &amp; 95% confidence intervals</td>
<td>0.003</td>
<td>−0.11 &amp; 95% confidence intervals</td>
<td>0.003</td>
<td>−0.11 &amp; 95% confidence intervals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>−1.27 &amp; 95% confidence intervals</td>
<td></td>
<td>(−1.12 to 1.35)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(−2.10 to −0.44)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Alcohol dehydrogenase (ADH) risk allele score in offspring and offspring IQ, stratified by maternal alcohol intake during pregnancy

P value for interaction of risk allele score and drinking during pregnancy equals 0.009

Any wine and kid’s a plonker

MUMS WARNED

By EMMA LITTLE, Health and Science Editor

MUMS-to-be who drink just ONE GLASS of wine give birth to kids with a lower IQ, researchers have claimed.
A study found any amount of alcohol during pregnancy can hit a baby’s developing brain. Doctors at Oxford and Bristol universities tracked
Association of LDL-C, HDL-C, and risk for coronary heart disease (CHD)

302K participants in 68 prospective studies

Emerging Risk Factors Collaboration, JAMA 2009
LDL and CHD Risk

Ference et al, JACC 2012
HDL: endothelial lipase Asn396Ser

Loss-of-function variants in endothelial lipase are a cause of elevated HDL cholesterol in humans

Andrew C. Edmondson,1 Robert J. Brown,1 Sekar Kathiresan,2,3 L. Adrienne Cupples,4 Serkalem Demissie,4 Alisa Knodle Manning,4 Majken K. Jensen,5 Eric B. Rimm,5,6 Jian Wang,7 Amrith Rodrigues,1 Vaneeta Bamba,1 Sumeet A. Khetarpal,1 Megan L. Wolfe,1 Stephanie DerOhannessian,1 Mingyao Li,8 Muredach P. Reilly,1,9 Jens Aberle,10 David Evans,10 Robert A. Hegele,7 and Daniel J. Rader1,9

- 2.6% of population carry Serine allele
- higher HDL-C
- No effect on other lipid fractions
- No effect on other MI risk factors

Edmondson, J Clin Invest 2009
**LIPG N396S and plasma HDL-C**

396S carriers have **5.5 mg/dl higher** HDL-C  
**P<10^{-8}**
After testing in 116,320 people, summary OR for \textit{LIPG} Asn396Ser is 0.99
Individuals who carry the HDL-boosting variant have the same risk for heart attack as those who do not carry the variant.
Effects of Torcetrapib in Patients at High Risk for Coronary Events

Philip J. Barter, M.D., Ph.D., Mark Caulfield, M.D., M.B., B.S., Mats Eriksson, M.D., Ph.D., Scott M. Grundy, M.D., Ph.D., John J.P. Kastelein, M.D., Ph.D., Michel Komajda, M.D., Jose Lopez-Sendon, M.D., Ph.D., Lori Mosca, M.D., M.P.H., Ph.D., Jean-Claude Tardif, M.D., David D. Waters, M.D., Charles L. Shear, Dr.P.H., James H. Revkin, M.D., Kevin A. Buhr, Ph.D., Marian R. Fisher, Ph.D., Alan R. Tall, M.B., B.S., and Bryan Brewer, M.D., Ph.D., for the ILLUMINATE Investigators*

RESULTS

At 12 months in patients who received torcetrapib, there was an increase of 72.1% in high-density lipoprotein cholesterol and a decrease of 24.9% in low-density lipoprotein cholesterol, as compared with baseline (P<0.001 for both comparisons), in addition to an increase of 5.4 mm Hg in systolic blood pressure, a decrease in serum potassium, and increases in serum sodium, bicarbonate, and aldosterone (P<0.001 for all comparisons). There was also an increased risk of cardiovascular events (hazard ratio, 1.25; 95% confidence interval [CI], 1.09 to 1.44; P=0.001) and death from any cause (hazard ratio, 1.58; 95% CI, 1.14 to 2.19; P=0.006). Post hoc analyses showed an increased risk of death in patients treated with torcetrapib whose reduction in potassium or increase in bicarbonate was greater than the median change.
Effects of Dalcetrapib in Patients with a Recent Acute Coronary Syndrome

Gregory G. Schwartz, M.D., Ph.D., Anders G. Olsson, M.D., Ph.D., Markus Abt, Ph.D., Christie M. Ballantyne, M.D., Philip J. Barter, M.D., Ph.D., Jochen Brunn, Ph.D., Bernard R. Chaitman, M.D., Ingar M. Holme, Ph.D., David Kallend, M.B., B.S., Lawrence A. Leiter, M.D., Eran Leitersdorf, M.D., John J.V. McMurray, M.D., Hardi Mundl, M.D., Stephen J. Nicholls, M.B., B.S., Ph.D., Prediman K. Shah, M.D., Jean-Claude Tardif, M.D., and R. Scott Wright, M.D., for the dal-OUTCOMES Investigators

RESULTS

At the time of randomization, the mean HDL cholesterol level was 42 mg per deciliter (1.1 mmol per liter), and the mean low-density lipoprotein (LDL) cholesterol level was 76 mg per deciliter (2.0 mmol per liter). Over the course of the trial, HDL cholesterol levels increased from baseline by 4 to 11% in the placebo group and by 31 to 40% in the dalcetrapib group. Dalcetrapib had a minimal effect on LDL cholesterol levels. Patients were followed for a median of 31 months. At a prespecified interim analysis that included 1135 primary end-point events (71% of the projected total number), the independent data and safety monitoring board recommended termination of the trial for futility. As compared with placebo, dalcetrapib did not alter the risk of the primary end point (cumulative event rate, 8.0% and 8.3%, respectively; hazard ratio with dalcetrapib, 1.04; 95% confidence interval, 0.93 to 1.16; P=0.52) and did not have a significant effect on any component of the primary end point or total mortality. The median C-reactive protein level was 0.2 mg per liter higher and the mean systolic blood pressure was 0.6 mm Hg higher with dalcetrapib as compared with placebo (P<0.001 for both comparisons).
OPINION

HDL—is it too big to fail?

Dominic S. Ng, Norman C. W. Wong and Robert A. Hegele

Abstract | The HDL hypothesis has suffered damage in the past few years. Clinical trials have shown that raising HDL cholesterol levels does not improve cardiovascular disease (CVD) outcomes. In addition, Mendelian randomization studies have shown that DNA variants that alter HDL cholesterol levels in populations are unrelated to incident CVD events. Balancing this deluge of negative data are substantial basic science data supporting the concept that raising HDL cholesterol levels reduces CVD risk. Also, functionally relevant HDL subfractions might be more important determinants of risk than overall HDL cholesterol levels. But, while wobbly, the HDL hypothesis is still standing, seemingly too big to fail owing to past intellectual, economic and psychological investments in the idea.

Ng, D. S. et al. Nat. Rev. Endocrinol. 9, 308–312 (2013); published online 15 January 2013; doi:10.1038/nrendo.2012.238
Using Multiple Genetic Variants as Instruments

- Allelic scores

Figure 1. DAG for a Mendelian randomisation analysis using four genetic variants as instrumental variables for the effect of fat mass on bone mineral density.


- Testing multiple variants individually
Mining the Human Phenome Using Allelic Scores That Index Biological Intermediates

David M. Evans¹,²,³*, Marie Jo A. Brion¹,²,³,⁴,⁵*, Lavinia Paternoster¹,², John P. Kemp¹,², George McMahon¹,², Marcus Munafò⁶, John B. Whitfield⁷, Sarah E. Medland⁷, Grant W. Montgomery⁷, The GIANT consortium⁸, The CRP consortium⁹, The TAG Consortium⁹, Nicholas J. Timpson¹,², Beate St. Pourcain¹,², Debbie A. Lawlor¹,², Nicholas G. Martin⁷, Abbas Dehghan⁸, Joel Hirschhorn⁴,⁹,¹⁰, George Davey Smith¹,²

A)

GWAS disease → Single SNP → Biological Intermediate or Pathway

B)

Biological Intermediate → Several SNPs → GWAS disease

Figure 1. (A) The usual paradigm in genome-wide association studies is to perform a GWAS of a trait/disease and then follow up any SNPs that reach genome-wide significance one marker at a time for putative biological function. The new paradigm (B) takes allelic scores of several SNPs that are known to proxy for a biological intermediate, and then tests to see whether these allelic scores are correlated with disease in GWAS datasets.
Figure 1. Association between polygene score and BMI measured at age nine in the ALSPAC cohort. Association between polygene score and BMI measured at age nine using different p-value thresholds for the construction of the score in ALSPAC children (N=5819). The lines joining the circles display the results for allelic scores calculated by using genotyped variants from across the genome in either a weighted (unbroken line) or an unweighted (dashed line) fashion. The lines joining the triangles display scores calculated similarly but excluding all variants +/-1 MB around 32 known BMI variants, and using either a weighted (unbroken line) or unweighted (dashed line) strategy. The histogram in the background displays the number of SNPs involved in construction of the allelic score at each corresponding SNP inclusion threshold for the “All variants” condition.
Mining the Phenome Using Allelic Scores

Table 1. Association between case-control status in the WTCCC and either a weighted genome-wide score consisting of all SNPs across the genome (“GW Score”), a weighted allelic score consisting of highly significant SNPs (p<5×10^{-8}) from known regions only (“Known”), or a weighted genome-wide score consisting of all SNPs across the genome with SNPs from known regions removed from its construction (“Complement”).

<table>
<thead>
<tr>
<th></th>
<th>BMI</th>
<th></th>
<th></th>
<th>CRP</th>
<th></th>
<th></th>
<th>LDLc</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GW Score</td>
<td>Known</td>
<td>Complement</td>
<td>GW Score</td>
<td>Known</td>
<td>Complement</td>
<td>GW Score</td>
<td>Known</td>
</tr>
<tr>
<td>Dir</td>
<td>P value</td>
<td>Dir</td>
<td>P</td>
<td>Dir</td>
<td>P value</td>
<td>Dir</td>
<td>P value</td>
<td>Dir</td>
</tr>
<tr>
<td>BD</td>
<td>0.051</td>
<td>-0.62</td>
<td>0.026</td>
<td>0.37</td>
<td>0.11</td>
<td>0.96</td>
<td>-0.049</td>
<td>-0.88</td>
</tr>
<tr>
<td>CHD</td>
<td>0.37</td>
<td>0.17</td>
<td>0.57</td>
<td>0.028</td>
<td>0.80</td>
<td>0.079</td>
<td>1.7×10^{-3}</td>
<td>9.2×10^{-3}</td>
</tr>
<tr>
<td>HT</td>
<td>0.76</td>
<td>-0.58</td>
<td>0.76</td>
<td>0.20</td>
<td>0.23</td>
<td>0.53</td>
<td>-0.011</td>
<td>-0.75</td>
</tr>
<tr>
<td>CD</td>
<td>0.97</td>
<td>0.90</td>
<td>0.99</td>
<td>2.9×10^{-4}</td>
<td>0.051</td>
<td>0.011</td>
<td>-0.73</td>
<td>-0.76</td>
</tr>
<tr>
<td>RA</td>
<td>0.18</td>
<td>0.15</td>
<td>-0.085</td>
<td>0.17</td>
<td>0.028</td>
<td>0.69</td>
<td>-0.26</td>
<td>-0.25</td>
</tr>
<tr>
<td>T1D</td>
<td>0.97</td>
<td>0.77</td>
<td>0.85</td>
<td>0.020</td>
<td>0.15</td>
<td>0.033</td>
<td>-0.018</td>
<td>0.58</td>
</tr>
<tr>
<td>T2D</td>
<td>&lt;2×10^{-16}</td>
<td>4.3×10^{-7}</td>
<td>1.8×10^{-12}</td>
<td>7.6×10^{-8}</td>
<td>0.50</td>
<td>2.1×10^{-7}</td>
<td>0.66</td>
<td>-0.12</td>
</tr>
</tbody>
</table>

See Tables S1 through S3 for a complete list of results.
BD = Bipolar Disorder; CHD = Coronary Heart Disease; HT = Hypertension; CD = Crohn’s Disease; RA = Rheumatoid Arthritis; T1D = Type 1 Diabetes; T2D = Type 2 Diabetes.
Dir = Direction of effect; P = P value.

- Could be applied to hundreds of thousands of molecular phenotypes simultaneously (gene expression, methylation, metabolomics etc)
Limitations to Mendelian Randomisation

1. Pleiotropy

2. Population stratification

3. Canalisation

4. Power (also “weak instrument bias”)

5. The existence of instruments
Mendelian Randomization in R

• There is a positive observational association between body mass index (BMI) and bone mineral density (BMD)

• It is unclear whether this represents a causal relationship

• We will use two stage least squares as implemented in R to address this question and estimate the causal effect of BMI on BMD
### Fitting in R - Datafile

<table>
<thead>
<tr>
<th>BMI</th>
<th>BMD</th>
<th>BMI_SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.371031158022524</td>
<td>0.860471934</td>
<td>0.0554687</td>
</tr>
<tr>
<td>-0.77975167453452</td>
<td>0.862923791</td>
<td>0.06125</td>
</tr>
<tr>
<td>-0.697738302042461</td>
<td>0.86130172</td>
<td>0.0684375</td>
</tr>
</tbody>
</table>

...
library(sem)

#Load BMI and BMD Data
x <- read.table(file="BMI_BMD_known.txt", header=TRUE, na.strings=-9)

#Observational regression of BMD on BMI
print("Observational regression of BMD on BMI")
summary(lm(x$BMD ~ x$BMI))

#Regression of BMI on BMI score – Check Instrument Strength
print("Regression of BMI on BMI score")
summary(lm(x$BMI ~ x$BMI_SCORE))

#Perform two stage least squares analysis
print("Two stage least squares analysis of BMD on BMI score")
summary(tsls(x$BMD ~ x$BMI, ~ x$BMI_SCORE))
# Observational regression of BMD on BMI
print("Observational regression of BMD on BMI")
summary(lm(x$BMD ~ x$BMI))

[1] "Observational regression of BMD on BMI"

Call:
lm(formula = x$BMD ~ x$BMI)

Coefficients:
         Estimate Std. Error  t value  Pr(>|t|)  
(Intercept)  0.90205  0.000625  1442.38 <2e-16 ***  
x$BMI      0.019523  0.000646  30.23  <2e-16 ***
# Regression of BMI on BMI score – Check Instrument Strength

```r
print("Regression of BMI on BMI score")
summary(lm(x$BMI ~ x$BMI_SCORE))
```

## OUTPUT:

[1] "Regression of BMI on BMI score"

Call:
`lm(formula = x$BMI ~ x$BMI_SCORE)`

Coefficients:

|                | Estimate | Std. Error | t value | Pr(>|t|)  |
|----------------|----------|------------|---------|-----------|
| (Intercept)    | -1.30067 | 0.09708    | -13.4   | <2e-16 *** |
| x$BMI_SCORE    | 20.56254 | 1.53438    | 13.4    | <2e-16 *** |

Residual standard error: 0.9532 on 5552 degrees of freedom
Multiple R-squared: 0.03133,   Adjusted R-squared: 0.03116
F-statistic: 179.6 on 1 and 5552 DF,  p-value: < 2.2e-16
COMMAND:

#Perform two stage least squares analysis
print("Two stage least squares analysis of BMD on BMI score")
summary(tsls(x$BMD ~ x$BMI, ~ x$BMI_SCORE))

OUTPUT:

[1] "Two stage least squares analysis of BMD on BMI score"

2SLS Estimates

Model Formula: x$BMD ~ x$BMI

Instruments: ~x$BMI_SCORE

|             | Estimate  | Std. Error | t value | Pr(>|t|) |
|-------------|-----------|------------|---------|----------|
| (Intercept) | 0.90199   | 0.0006302  | 1431.368| 0.000e+00|
| x$BMI       | 0.01442   | 0.0036687  | 3.931   | 8.551e-05|
Acknowledgments

• George Davey Smith
• Nic Timpson
• Sek Kathiresan