Instrumental variables & Mendelian randomization

Qingyuan Zhao

Department of Statistics, The Wharton School, University of Pennsylvania

May 9, 2018 @ JHU

\[ \hat{\gamma} = \text{lm}(X \sim Z) \]

\[ \hat{\beta}_0 = \text{lm}(Y \sim Z) \]
Outline

1. Instrumental variables (IV)
   - Why IVs?
     - Draw back of observational studies
     - Core assumptions and the promise of IVs
   - Examples of IVs
     - In Economics
     - In Public Health
     - In Human Genetics
   - Classical estimators
     - Two-stage least squares (TSLS)
     - Limited information maximum likelihood (LIML)

2. Mendelian randomization (MR)
   - Summary-data MR
     - Data structure
     - Modeling assumptions
   - Statistical methods
     - Meta-analysis methods
     - Likelihood methods
Outline

1. Instrumental variables (IV)
   - Why IVs?
     - Draw back of observational studies
     - Core assumptions and the promise of IVs
   - Examples of IVs
     - In Economics
     - In Public Health
     - In Human Genetics
   - Classical estimators
     - Two-stage least squares (TSLS)
     - Limited information maximum likelihood (LIML)

2. Mendelian randomization (MR)
   - Summary-data MR
     - Data structure
     - Modeling assumptions
   - Statistical methods
     - Meta-analysis methods
     - Likelihood methods
Hierarchy of evidence

When the goal is to infer causation...

- Expert opinions, case reports, animal studies
- Observational studies (case-control and cohort design)
- Natural experiments (Mendelian randomization)
- RCTs

Quality of evidence

Figure: (A rough) Hierarchy of evidence in medical studies.¹

Fundamental challenge of observational studies

“Correlation does not imply causation”.

Observational studies = Enumerating confounders

- Idea: Conditioning on possible sources of spurious correlation.
- Example: Possible confounders between smoking and lung cancer:
  - Age.
  - Sex.
  - Urban/Rural.
  - Working environment.
  - Socioeconomic class.
  - ...

- **Fundamental challenge:** We can never be sure this list is complete.

- The promise of instrumental variables: unbiased estimation of causal effect without enumerating confounders.
What is an instrument variable (IV)?

Causal diagram for IV

![Causal diagram showing instrument variable (IV) relationships](image)

Core IV assumptions

1. **Relevance**: $Z$ is associated with the exposure ($X$).
2. **Effective random assignment**: $Z$ is independent of the unmeasured confounder ($C$).
3. **Exclusion restriction**: $Z$ cannot have any direct effect on the outcome ($Y$).

Wald’s estimator based on Intention-to-treat (ITT) analysis

$$\text{Causal effect of } X \text{ on } Y \approx \frac{\text{ITT Effect of } Z \text{ on } Y}{\text{ITT Effect of } Z \text{ on } X}$$
Outline

1 Instrumental variables (IV)
   - Why IVs?
     - Draw back of observational studies
     - Core assumptions and the promise of IVs
   - Examples of IVs
     - In Economics
     - In Public Health
     - In Human Genetics
   - Classical estimators
     - Two-stage least squares (TSLS)
     - Limited information maximum likelihood (LIML)

2 Mendelian randomization (MR)
   - Summary-data MR
     - Data structure
     - Modeling assumptions
   - Statistical methods
     - Meta-analysis methods
     - Likelihood methods
IV in Economics: Effect of military service on earnings

- In 1970, the U.S. government conducted draft lottery to determine priority of conscription for the Vietnam war.
- Exercise: Justify the core IV assumptions.
- The draft lottery can be regarded as a “natural experiment” of military service.

---

### Results of the Vietnam-war lottery study

**Table 4.1.3: Wald estimates of the effects of military service on the earnings of white men born in 1950**

<table>
<thead>
<tr>
<th>Earnings year</th>
<th>Earnings Mean</th>
<th>Eligibility Effect Mean</th>
<th>Veteran Status Mean</th>
<th>Eligibility Effect Mean</th>
<th>Wald Estimate of Veteran Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1981</td>
<td>16,461</td>
<td>-435.8</td>
<td>0.267</td>
<td>0.159</td>
<td>-2,741</td>
</tr>
<tr>
<td></td>
<td>(210.5)</td>
<td></td>
<td>(0.040)</td>
<td></td>
<td>(1,324)</td>
</tr>
<tr>
<td>1971</td>
<td>3,338</td>
<td>-325.9</td>
<td>0.267</td>
<td>0.159</td>
<td>-2,741</td>
</tr>
<tr>
<td></td>
<td>(46.6)</td>
<td></td>
<td>(0.040)</td>
<td></td>
<td>(1,324)</td>
</tr>
<tr>
<td>1969</td>
<td>2,299</td>
<td>-2.0</td>
<td>0.267</td>
<td>0.159</td>
<td>-2,741</td>
</tr>
<tr>
<td></td>
<td>(34.5)</td>
<td></td>
<td>(0.040)</td>
<td></td>
<td>(1,324)</td>
</tr>
</tbody>
</table>

Notes: Adapted from Angrist (1990), Tables 2 and 3. Standard errors are shown in parentheses. Earnings data are from Social Security administrative records. Figures are in nominal dollars. Veteran status data are from the Survey of Program Participation. There are about 13,500 individuals in the sample.
IV in Public Health: Effectiveness of vaccine

This is also called randomized encouragement design.

The same idea can be applied to RCTs with non-compliance.

---

Compared to *trans*-SNPs, *cis*-SNPs are more likely to satisfy exclusion restriction (criterion 3).

This is a special case of “Mendelian randomization” where genetic variation is used as IV and typically $X$ is an epidemiological risk factor (more downstream).

---

Outline

1 Instrumental variables (IV)
   - Why IVs?
     Draw back of observational studies
     Core assumptions and the promise of IVs
   - Examples of IVs
     In Economics
     In Public Health
     In Human Genetics
   - Classical estimators
     Two-stage least squares (TSLS)
     Limited information maximum likelihood (LIML)

2 Mendelian randomization (MR)
   - Summary-data MR
     Data structure
     Modeling assumptions
   - Statistical methods
     Meta-analysis methods
     Likelihood methods
Linear IV model

- The Wald ratio estimator becomes inadequate when $Z$ and $X$ are multivariate.
- The most commonly used IV estimators are based on the following linear model:

$$Y_i = X_i^T \beta + Z_i^T \alpha + U_i,$$
$$X_i = Z_i^T \gamma + V_i.$$

IV assumptions in the linear model

1. Relevance: $\gamma \neq 0$;
2. Exogeneity: $Z_i \perp (U_i, V_i)$;
3. Exclusion restriction: $\alpha = 0$.

- The exposure variable $X_i$ is called *confounded* or *endogenous* if it is correlated with $U_i$ (or equivalently, if $V_i$ is correlated with $U_i$).
Identification of causal effect

Under the linear IV model, the causal effect $\beta$ satisfies

$$\mathbb{E}[Z_i(Y_i - X_i^T\beta)] = 0.$$ 

Notice how this is different from the usual normal equation

$$\mathbb{E}[X_i(Y_i - X_i^T\beta)] = 0.$$ 

To identify $\beta$, we need $\text{dim}(Z_i) \geq \text{dim}(X_i)$.

Just-identified case: When $\text{dim}(Z_i) = \text{dim}(X_i)$, we can estimate $\beta$ by solving

$$\sum_{i=1}^{n} Z_i(Y_i - X_i^T\beta) = 0.$$

The solution in matrix-form is

$$\hat{\beta} = (Z^T X)^{-1} Z^T Y.$$ 

Over-identified case: When $\text{dim}(Z_i) > \text{dim}(X_i)$, we have some freedom to choose which (linear combinations of) equations to solve.
Two-stage least squares (TSLs)

- In the over-identified case, here is a general class of IV estimator:
  Let \( f : \mathbb{R}^{\text{dim}(Z)} \mapsto \mathbb{R}^{\text{dim}(X)} \) be any function that maps from the space of \( Z \) to \( X \). Then \( \beta \) satisfies

  \[
  \mathbb{E}[f(Z_i) \cdot (Y_i - X_i^T \beta)] = 0.
  \]

- The most efficient choice of \( f \) is \( f(Z_i) = \mathbb{E}[X_i | Z_i] = Z_i^T \gamma \).

- The nuisance parameter \( \gamma \) is not known but can be estimated from the data. The most common estimator is least squares:

  \[
  \hat{\gamma} = (Z^T Z)^{-1} Z^T X.
  \]

- Thus the IV estimator of \( \beta \) is given by (let \( P_Z = Z(Z^T Z)^{-1} Z^T \))

  \[
  \hat{\beta} = [(Z\hat{\gamma})^T X]^{-1} (Z\hat{\gamma})^T Y = (X^T P_Z X)^{-1} (X^T P_Z Y).
  \]

- This is called two-stage least squares, because (let \( \hat{X} = P_Z X \))

  \[
  \hat{\beta} = \text{lm}(Y \sim \hat{X}) = \text{lm}(Y \sim \text{predict(lm}(X \sim Z)))
  \]

- However, standard error of \( \hat{\beta} \) cannot be obtained directly from \text{lm} because \( \hat{\gamma} \) is estimated from the data.
Limited information maximum likelihood (LIML)

- Recall the linear IV model:
  \[
  Y_i = X_i^T \beta + U_i, \\
  X_i = Z_i^T \gamma + V_i.
  \]

- The LIML estimator assumes the noise variables \((U_i, V_i)\) are jointly normal with mean \(0\) and covariance \(\Sigma\).

- LIML maximizes the log-likelihood of this problem:
  \[
  l(\beta, \gamma, \Sigma) = -\frac{1}{2} \sum_{i=1}^{n} \log |\Sigma^{-1}| + \left( Y_i - X_i^T \beta \right)^T \Sigma^{-1} \left( Y_i - X_i^T \beta \right) - \left( X_i - Z_i^T \gamma \right)^T \Sigma^{-1} \left( X_i - Z_i^T \gamma \right) .
  \]

- TSLS and LIML are asymptotically equivalent (when \(n \to \infty\) and \(\text{dim}(X_i)\) and \(\text{dim}(Z_i)\) are fixed).

- LIML is more robust to weak instruments (small \(\gamma\)).
Outline

1 Instrumental variables (IV)
   • Why IVs?
     Draw back of observational studies
     Core assumptions and the promise of IVs
   • Examples of IVs
     In Economics
     In Public Health
     In Human Genetics
   • Classical estimators
     Two-stage least squares (TSLS)
     Limited information maximum likelihood (LIML)

2 Mendelian randomization (MR)
   • Summary-data MR
     Data structure
     Modeling assumptions
   • Statistical methods
     Meta-analysis methods
     Likelihood methods
MR = Using genetic variation as IV

Examine the core IV assumptions

<table>
<thead>
<tr>
<th>Criterion</th>
<th>✔</th>
<th>Modern GWAS have identified many causal variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion</td>
<td>✔</td>
<td>Almost Comes for free due to Mendel’s Second Law</td>
</tr>
<tr>
<td>Minor concern: population stratification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td>?</td>
<td>Problematic because of wide-spread pleiotropy</td>
</tr>
<tr>
<td>(multiple functions of genes).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

▸ Exercise: what if there are two SNPs in LD?
Non-conventional challenges in MR

Weak instruments Many genetic variants are only weakly associated with the exposure.
  ▶ Solution: Use LIML-type estimator instead of TSLS.

Two-sample IV/MR Association data for $X$ and $Y$ often come from different population.
  ▶ Need to justify the causal structure is invariant.\(^5\)

Summary-data MR Most GWAS data come in summary-statistics format due to privacy.
  ▶ Solution: Develop statistical methods that can be applied to summary statistics.

Pleiotropy Exclusion restriction is likely violated for many genetic IVs.
  ▶ Solution: Use more robust methods that account for pleiotropy.

---

A general workflow of two-sample summary-data MR

1. Select independent IVs for the exposure (using **GWAS-E1**).
2. Extract GWAS summary statistics of the selected IVs for the exposure (using **GWAS-E2**).
3. Extract GWAS summary statistics of the selected IVs for the outcome (using **GWAS-O**).
4. **Harmonize** data in steps 2 and 3 so the reference allele is the same.
5. Perform statistical analysis.

Open-source software
A one-stop solution is being developed in the R package **TwoSampleMR**:\(^6\)
- A large number of public GWAS summary datasets being collected.
- Convenient wrapper of LD clumping and data harmonization.
- Functions for statistical analysis.

**Caveat**
TwoSampleMR does not differentiate between **GWAS-E1** and **GWAS-E2**, which may introduce selection bias (also called winner’s curse).

[^6]: https://github.com/MRCIEU/TwoSampleMR
Modeling assumptions for GWAS summary data

Dataset: estimated effects \((\hat{\gamma}, \hat{\Gamma})\) and standard errors \((\sigma_X, \sigma_Y)\).

**Assumption 1: Measurement error model**

\[
\begin{pmatrix}
\hat{\gamma} \\
\hat{\Gamma}
\end{pmatrix} \sim N \left( \begin{pmatrix}
\gamma \\
\Gamma
\end{pmatrix}, \begin{pmatrix}
\Sigma_X & 0 \\
0 & \Sigma_Y
\end{pmatrix} \right), \quad \Sigma_X = \text{diag}(\sigma_{X1}^2, \ldots, \sigma_{Xp}^2), \quad \Sigma_Y = \text{diag}(\sigma_{Y1}^2, \ldots, \sigma_{Yp}^2).
\]

**Pre-processing warrants Assumption 1**

- Large sample size \(\Rightarrow\) CLT.
- (Approximate) independence due to
  1. Non-overlapping samples (in GWAS-E1, GWAS-E2, GWAS-O).
  2. Independent SNPs.

**Assumption 2: Linking the genetic associations (ITT effects)**

The causal effect \(\beta\) satisfies \(\Gamma \approx \beta \gamma\). In particular, we have found a reasonable model for \(\alpha = \Gamma - \beta \gamma\) is universal pleiotropy with outliers

1. Most \(\alpha_j \perp \gamma_j\) and \(\alpha_j \overset{i.i.d.}{\sim} N(0, \tau^2)\) for some small \(\tau^2\).
2. A few \(|\alpha_j|\) might be very large.
Statistical problem

Genetic association

\((\hat{\gamma}_j, \hat{\Gamma}_j, \sigma_{X_j}, \sigma_{Y_j})_{j=1:p}\)

\(\Rightarrow\) Epidemiological causation

\[\beta_0\]

\[\hat{\gamma} = \text{lm}(X \sim Z)\]

\[\hat{\Gamma} = \text{lm}(Y \sim Z)\]

\(\hat{\gamma}_j\) → (Gene) → (HDL) → (Heart disease) → Epidemiological causation

\(\hat{\Gamma}_j\) → (Gene) → (HDL) → (Heart disease) → Epidemiological causation
Outline

1 Instrumental variables (IV)
   • Why IVs?
     Draw back of observational studies
     Core assumptions and the promise of IVs
   • Examples of IVs
     In Economics
     In Public Health
     In Human Genetics
   • Classical estimators
     Two-stage least squares (TSLS)
     Limited information maximum likelihood (LIML)

2 Mendelian randomization (MR)
   • Summary-data MR
     Data structure
     Modeling assumptions
   • Statistical methods
     Meta-analysis methods
     Likelihood methods
Meta-analysis methods

- Each SNP produces an independent Wald estimator: \( \hat{\beta}_j = \hat{\Gamma}_j/\hat{\gamma}_j \).
- Using Delta method and assuming \( \Gamma_j \equiv \beta_0 \gamma_j \), we can obtain
  \[
  \hat{\beta}_j = \frac{\hat{\Gamma}_j}{\hat{\gamma}_j} = \frac{\Gamma_j + \epsilon \gamma_j}{\gamma_j + \epsilon \chi_j} \approx N\left( \frac{\Gamma_j}{\gamma_j}, \frac{\sigma^2_{\chi_j} + \beta^2 \sigma^2_{\gamma_j}}{\gamma_j^2} : = \sigma_j^2 \right).
  \]
- Next combine the individual estimates using meta-analysis:
  1. Inverse-variance weighting:
     \[
     \hat{\beta}_{IVW} = \text{Mean}(\hat{\beta}_j, \text{weights} = 1/\sigma_j^2).
     \]
  2. Weighted median:
     \[
     \hat{\beta}_{WMed} = \text{Median}(\hat{\beta}_j, \text{weights} = 1/\sigma_j^2).
     \]
  3. MR-Egger regression:
     \[
     \hat{\beta}_{\text{Egger}} = \text{lm}(\hat{\Gamma}_j \sim \hat{\gamma}_j, \text{weights} = 1/\sigma_j^2).
     \]
Weak instrument bias

▶ The main issue of meta-analysis methods is that the Delta method approximation is not accurate if $|\gamma_j|/\sigma_{X_j}$ is small.

▶ In general, the distribution of $\hat{\beta}_j = \hat{\Gamma}_j/\hat{\gamma}_j$ is a mixture of Cauchy distribution and a bi-modal distribution.\(^7\)

```r
> p <- 100000; r <- (2 + rnorm(p))/(1 + rnorm(p));
> hist(r[abs(r) < 15], 100); abline(v = 2, lty = "dashed", col = "red", lwd = 3);
> median(r)
[1] 1.327066
```

▶ This problem is known as weak instrument bias in the IV literature.

Likelihood methods

- Using the ratio $\hat{\Gamma}_j / \hat{\gamma}_j$ is a little silly if $\hat{\gamma}_j$ is small.
- We can pool information from multiple weak IVs using the likelihood.
- Assuming $\Gamma_j \equiv \beta \gamma_j$, the log-likelihood of SNP $j$ is
  
  $$l_j(\beta, \gamma) = -\frac{(\hat{\gamma}_j - \gamma_j)^2}{(2\sigma_{Xj}^2)} - \frac{(\hat{\Gamma}_j - \gamma_j \beta)^2}{(2\sigma_{Yj}^2)}.$$

- Sufficient statistic for $\gamma_j$: $\hat{\gamma}_{j, MLE}(\beta) = \frac{\hat{\gamma}_j / \sigma_{Xj}^2 + \beta \hat{\Gamma}_j / \sigma_{Yj}^2}{1 / \sigma_{Xj}^2 + \beta^2 / \sigma_{Yj}^2}$.

- Conditional score is defined as
  
  $$C_j(\beta) = \frac{\partial}{\partial \beta} l_j(\beta, \gamma) - \mathbb{E}\left[ \frac{\partial}{\partial \beta} l_j(\beta, \gamma) \bigg| \hat{\gamma}_{j, MLE}(\beta) \right] = \frac{\gamma_j (\hat{\Gamma}_j - \beta \hat{\gamma}_j)}{\sigma_{Yj}^2 + \beta^2 \sigma_{Xj}^2}.$$

- Observation 1: $\gamma_j$ only appears as weight to “residual” $\hat{\Gamma}_j - \beta \hat{\gamma}_j$.
- Observation 2: $\hat{\gamma}_{j, MLE}(\beta)$ is independent of $\hat{\Gamma}_j - \beta \hat{\gamma}_j$.

---

Increased efficiency and robustness

- The observations above motivate a general class of unbiased estimating equations:

\[ \sum_{j=1}^{p} \frac{f_j(\hat{\gamma}_j, \text{MLE}(\beta)) \cdot \psi(\hat{\Gamma}_j - \beta \hat{\gamma}_j)}{\sigma_{Y_j}^2 + \beta^2 \sigma_{X_j}^2} = 0 \]

- Heuristic: choose \( f_j \) to increase efficiency; choose bounded \( \psi \) to be robust against outliers.
- Example: \( f_j(\hat{\gamma}_j, \text{MLE}) \) is the spike-and-slab shrinkage estimate of \( \gamma_j \).
- Example: \( \psi \) is Huber's score function.
An overdispersion phenomenon

- So far we have assumed $\Gamma_j = \beta \gamma_j$ (at least for most $j$).
- However we have found that $\alpha_j = \Gamma_j - \hat{\beta} \gamma_j$ seems to be approximately normal when $\hat{\beta}$ is obtained as above.

A real data example: Effect of BMI on SBP

- Left ($p = 25, p_{\text{sel}} < 5 \cdot 10^{-8}$): scatter-plot of GWAS summary data;
- Right ($p = 160, p_{\text{sel}} < 10^{-4}$): Q-Q plot of standardized residuals.
Adjusting the score of overdispersion parameter

- A reasonable model is most $\alpha_j \sim N(0, \tau^2)$.
- Statistical estimation of $\tau^2$ is non-trivial due to the Neyman-Scott phenomenon.

Neyman-Scott problem (a simplified scenario)

Suppose we observe independent pairs

$$X_{ij} \sim N(\gamma_i, \tau^2), \quad i = 1, \ldots, n, \quad j = 1, 2.$$ 

The goal is to estimate $\tau^2$, but the MLE is inconsistent:

$$\hat{\tau}^2 = \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{2} (X_{ij} - \bar{X}_i.)^2 = \frac{1}{4n} \sum_{i=1}^{n} (X_{i1} - X_{i2})^2 \xrightarrow{p} \tau^2 / 2.$$ 

An easy fix in this case is to use $2\hat{\tau}^2$. However the inconsistency of MLE is common in many other problems involving a large number nuisance parameters, and the fix is usually complicated.

- There is a relatively simple fix in the MR problem (details omitted).
Recap

1. **Instrumental variables (IV)**
   - **Why IVs?**
     - Draw back of observational studies
     - Core assumptions and the promise of IVs
   - **Examples of IVs**
     - In Economics
     - In Public Health
     - In Human Genetics
   - **Classical estimators**
     - Two-stage least squares (TSLS)
     - Limited information maximum likelihood (LIML)

2. **Mendelian randomization (MR)**
   - **Summary-data MR**
     - Data structure
     - Modeling assumptions
   - **Statistical methods**
     - Meta-analysis methods
     - Likelihood methods

Bonus slides: A MR study on blood lipids.
The Lipid Hypothesis

“Decreasing blood cholesterol significantly reduces the risk of cardiovascular diseases.”

1913 First evidence from a rabbit study.

1950s – 1980s Accumulation of evidence from observational studies. Transformation to the LDL hypothesis.


1990s Skepticism continue until landmark statin trials.

2010s Reaffirmation from Mendelian randomization.

However, the role of HDL cholesterol remains quite controversial.

---

9History based on: Academy of Medical Sciences Working Group (2007). “Identifying the environmental causes of disease: how should we decide what to believe and when to take action?” Academy of Medical Sciences.
The HDL Hypothesis

“HDL is protective against heart diseases.”

1960s Formulation of the hypothesis from observational studies. The inverse association has been firmly established over the years.

1980s Supporting evidence from animal studies.

But... 2000s Null findings from studies of Mendelian disorders.

2010s Failed RCTs, though each has its own caveats.

2010s Null findings from Mendelian randomization.

“I’d say the HDL hypothesis is on the ropes right now,” said Dr. James A. de Lemos . . . Dr. Kathiresan said. “I tell them, ’ It means you are at increased risk, but I don’t know if raising it will affect your risk.’” — New York Times, May 16, 2012.

Reasons of null findings: flawed design, lack of power, HDL function hypothesis . . .

We will reassess the evidence for HDL using a new design and new statistical methods of Mendelian randomization.

Application to HDL and coronary heart disease

Dataset

- Used a 2010 GWAS of blood lipids to select 1122 independent SNPs not associated with LDL or triglycerides ($p$-value $\geq 0.01$).
- 23 SNPs were genome-wide significant for HDL.
- HDL dataset: an non-overlapping 2013 GWAS of blood lipids.
- Coronary artery disease dataset: CARDIoGRAMplusC4D consortium.

Fitted prior for $\gamma_j/\sigma_{X_j}$

- Spike: $p_1 = 0.91$, $\sigma_1 = 0.73$;
- Slab: $p_2 = 0.09$, $\sigma_2 = 4.57$.

Increase of efficiency (rough estimates from simulation)

Conventional MR (23 SNPs) $\uparrow_{100\%}$ Genome-wide MR (1122 SNPs) $\uparrow_{20\%}$ Spike-and-slab shrinkage
Visualization of MR estimation

23 genome-wide significant SNPs in the selection GWAS.
- Rest 1099 SNPs.
Results

<table>
<thead>
<tr>
<th>Observational studies</th>
<th>LDL</th>
<th>HDL</th>
<th>Triglycerides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angelantonio et al. (2009)</td>
<td>1.50 (1.39–1.61)</td>
<td>0.78 (0.74–0.82)</td>
<td>0.99 (0.94–1.05)</td>
</tr>
<tr>
<td>Voight et al. (2012)</td>
<td>1.54 (1.45–1.63)</td>
<td>0.62 (0.58–0.66)</td>
<td>1.42 (1.31–1.52)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Previous MR studies</th>
<th>LDL</th>
<th>HDL</th>
<th>Triglycerides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voight et al. (2012)</td>
<td>2.13 (1.69–2.69)</td>
<td>0.93 (0.68–1.26)</td>
<td>Not reported</td>
</tr>
<tr>
<td>Holmes et al. (2014)*</td>
<td>1.92 (1.68–2.19)</td>
<td>0.96 (0.70–1.31)</td>
<td>1.26 (1.00–1.61)</td>
</tr>
<tr>
<td>White et al. (2016)</td>
<td>1.68 (1.51–1.87)</td>
<td>0.95 (0.85–1.06)</td>
<td>1.28 (1.13–1.45)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>New MR analysis</th>
<th>LDL</th>
<th>HDL</th>
<th>Triglycerides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using all SNPs</td>
<td>1.61 (1.45–1.80)</td>
<td>0.82 (0.73–0.91)</td>
<td>1.00 (0.84–1.21)</td>
</tr>
<tr>
<td>Using significant SNPs</td>
<td>1.76 (1.53–2.03)</td>
<td>0.88 (0.74–1.04)</td>
<td>Not available†</td>
</tr>
<tr>
<td>Using non-significant SNPs</td>
<td>1.25 (1.02–1.54)</td>
<td>0.75 (0.62–0.92)</td>
<td>1.53 (1.01–2.33)</td>
</tr>
</tbody>
</table>

Table: Reported: Estimated odds ratio and 95% confidence interval.
*The original results were reported in a different unit.
Disclaimer: These results are preliminary and not yet published.
IV methods are very useful tools to infer causality and are becoming widely used in public health and human genetics.

I hope this is not the end of your journey with IVs. Here are some suggested readings:

- Chapter 4 of Angrist and Pischke’s book *Mostly Harmless Econometrics: An Empiricist’s Companion*.