Observational comparative effectiveness research using large databases

Robert W. Yeh, MD MSc
Richard A. and Susan F. Smith Center for Outcomes Research in Cardiology
Beth Israel Deaconess Medical Center

November 21, 2017
Objectives

• Discuss strengths and limitations of large datasets.
• Describe methods of addressing confounding.
  – Propensity Score Methods
  – Alternatives, including Regression and Instrumental Variable approaches
• Discuss strengths and limitations of propensity score methods relative to alternatives.
• Use real examples from the literature throughout.
Examples of Large Datasets

• Administrative claims data
  – Medicare
  – State
  – Health system billing data
• EMR data
  – RPDR
  – Integrated health care system data
• Registries
  – National Cardiovascular Data Registry
  – Mass-DAC
Two major problems of large database CER

• Data accuracy and completeness
  – Are the outcomes measured accurately?
  – Are the patient characteristics captured accurately and comprehensively?
  – What is the impact when the answer is “no”.

• Observational comparisons are subject to confounding
Data Accuracy

• Many large databases use claims-based codes to identify outcomes.
  – Myocardial infarction (410.x1), eg.
  – Codes may not be accurate.
  • What is the positive predictive value of code? Negative predictive value?
  • What is the impact if “misclassification.”
Sensitivity analyses were performed to determine whether decreases in observed incidence rates for myocardial infarction were strongly influenced by misclassification because of possible miscoding of acute myocardial infarction or differential patterns of persons joining the health plan (in-migration) and existing members leaving the health plan (out-migration) during the study period. We addressed the former by expanding the definition of myocardial infarction to include ICD-9-CM codes 411.x, 413.x, and 414.x in association with elevated cardiac biomarker levels; we addressed the latter by restricting analyses to the subgroup with no in- or out-migration. Temporal trends in the incidence of myocardial infarction, ST-segment elevation myocardial infarction, and non–ST-segment elevation myocardial infarction in all sensitivity analyses were similar to results from the main analysis (see the Supplementary Appendix, available with the full text of this article at NEJM.org), so only the main results are presented here. We also observed that the proportion of 410.x hospital-discharge codes in the 410.x, 411.x, 413.x, and 414.x grouping increased during the study period, suggesting that miscoding of myocardial infarction was unlikely to have contributed to the decrease in rates of myocardial infarction.
greater than we observed. Our validation study showed that the positive predictive value of codes for ST-segment elevation myocardial infarction and non-ST-segment elevation myocardial infarction were not perfect, but the accuracy of these codes did not differ significantly over time, and our findings were consistent regardless of changes in positive predictive value within the 95% confidence limits (data not shown). Thus, although the absolute incidence would be affected by misclassification of the type of myocardial infarction, the observed trends in the overall incidence of myocardial infarction are unlikely to be biased by coding accuracy. Differential enrollment
Do you believe the data you are getting?

- Reviewer #1:
  Although I do understand the national inpatient sample, and that it is a 20% sampling that is extrapolated to the whole, I now have significant concerns despite previous substantiation of this methodology that it is merely in fact compounding initial sample errors by a factor of five.

- Reviewer #2:
  I still have a problem with a 20% sampling of the data and the major assumptions that are made from this.

- Reviewer #3:
  The limitations of the NIS have been recognized in this manuscript, but the dramatic changes in some of the data that emerged with this revision highlight the fact that the NIS is simply not a good resource for studying therapies in rapid transition. The weighting algorithms used for the NIS may be appropriate for steady state conditions or gradual evolution, but may break down when the topic is a therapy undergoing rapid transition or dissemination, such as TAVR shortly after approval. It is hard to know what to make of the dramatic differences in some of the findings that emerged from this revision. It calls into question not only the originally reported findings, but the current findings as well. I think this study might be more useful with a longer perspective, incorporating several years worth of data. The current sample is so brief in time that any lack of normalization will be magnified by the methodology. I worry that the methods (not the authors' methods per se, but the limitations of the NIS) may make this an inherently unstable data set.
CER: The Gold Standard

- Randomized Controlled Trials are the gold standard for comparing 2 treatments but...
  - May have limited generalizability.
  - May not include important subgroups, or lack power to examine those subgroups.
  - May lack feasibility to answer certain questions.

- Observational data are abundant, less expensive to acquire, and are often more representative of “real world” practice.
Observational Studies of Comparative Effectiveness are Subject to Confounding

Factors that lead to selection of treatment A vs. B

Treatment A Vs. Treatment B

Outcomes
Observational Studies of Comparative Effectiveness are Subject to Confounding

Factors that lead to selection of CABG vs. PCI

PCI vs. CABG

Myocardial Infarction

Death
Terminology

• Confounding by indication: confounding of the association between treatment and outcome that occurs because the indications that led to the selection of the treatment also are associated with the outcome.

• Also referred to as treatment selection bias

• Propensity scores are intuitively thought of as specifically addressing this form of confounding by modeling the actual treatment decision.
Randomized Clinical Trial

Randomization
Heads = surgery
Tails = stent

Factors that lead to selection of surgery vs. PCI

Bypass Surgery vs. PCI

Heart Attack Death

Richard A. and Susan F. Smith Center for Outcomes Research in Cardiology
Propensity Scores

- Propensity Score = the probability of receiving a certain treatment based on **observed** characteristics.

Factors that lead to selection of revascularization

Propensity Score

Factors that lead to:

- Bypass vs. Angioplasty
- Myocardial Infarction
- Death

Patients with the similar PS should be similar in terms of baseline characteristics.
Understanding Propensity Scores in Stages

• Stage 1: What are the predictors of treatment?
  – Treatment A vs. Treatment B is now the *outcome*
  – With 2 alternatives, this is now a dichotomous outcome → logistic regression
  – Create a model of variables that predict treatment assignment. (i.e. Table 1)

Model: Logistic
Outcome: Treatment A vs. B
Predictors: All factors that might be associated with treatment and/or outcome
Stage 1: Estimating the PS

Find patients that were “eligible” for both treatments.
### 1:1 Match

<table>
<thead>
<tr>
<th>CABG Patients</th>
<th>PCI Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>14%</td>
<td>63%</td>
</tr>
<tr>
<td>23%</td>
<td>99%</td>
</tr>
<tr>
<td>88%</td>
<td>56%</td>
</tr>
<tr>
<td>75%</td>
<td>21%</td>
</tr>
<tr>
<td>56%</td>
<td>14%</td>
</tr>
<tr>
<td>63%</td>
<td>35%</td>
</tr>
<tr>
<td>21%</td>
<td>38%</td>
</tr>
</tbody>
</table>
### Before and After PS “Adjustment” – Creating a Balanced Population (RCT-Like?)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Unadjusted Data</th>
<th>Data Adjusted with the Use of Inverse Probability Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CABG (N=86,244)</td>
<td>PCI (N=103,549)</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>73.1±5.6</td>
<td>74.7±6.5</td>
</tr>
<tr>
<td>Male sex (%)</td>
<td>68.6</td>
<td>57.8</td>
</tr>
<tr>
<td>History of heart failure (%)</td>
<td>11.5</td>
<td>10.2</td>
</tr>
<tr>
<td>History of myocardial infarction (%)</td>
<td>25.3</td>
<td>24.6</td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any</td>
<td>38.6</td>
<td>34.4</td>
</tr>
<tr>
<td>Requiring insulin</td>
<td>10.2</td>
<td>9.8</td>
</tr>
<tr>
<td>Hypertension (%)</td>
<td>84.8</td>
<td>83.4</td>
</tr>
<tr>
<td>Renal failure (%)</td>
<td>6.1</td>
<td>6.2</td>
</tr>
<tr>
<td>Chronic lung disease (%)</td>
<td>20.7</td>
<td>18.9</td>
</tr>
<tr>
<td>Cerebrovascular disease (%)</td>
<td>17.6</td>
<td>15.8</td>
</tr>
</tbody>
</table>
Concerns: Assessing Balance

• Important to assess Balance in all PS methods.
• Use Standardized Differences rather than p values
  – Mean difference divided by pooled standard error of the 2 groups
  – Accepted threshold of SD < 10% for all variables.

• Lack of Balance can reflect misspecification of the model
  – Tighter match?
  – Consider interaction terms?
Figure 3: Pre- and Post-matching standardized differences

- Maximum High
- In hospital PCI
- In hospital CAH
- Diabetes
- Age
- Heart failure
- GI bleed
- Dyslipidemia
- Other heart failure
- Bleeding event
- Stroke status
- GFR/RF
- History of CAD
- CKD
- Platelet inhibitors
- ST elevation MI
- Acute renal failure
- Non-ST elevation MI
- Bronchodilators
- Insulin
- Short acting insulins
- End stage renal disease
- Angina
- Dialysis procedure
- Female
- Long acting insulins
- Diabets
- Hepatitis
- Anti-inhibitors
- ACE/ARB
- History of PCI
- Early vs late
- Stains
- Calcium channel blockers
- Synthetis heart failure
- Hepatitis
- Acute respiratory failure
- In hospital mechanical ventilation
- Other shock
- Elective heart failure
- Sepsis shock
- PDA
- Any shock
- Oral glucose medications
- Diabetes Type 1
- Hypertension
- Miscellaneous bleed

---

Standardized Difference (%)
Concerns: Assessing Overlap

Can these groups be compared?

What is the c-stat?
Stage 2: Estimating the Treatment Effect

• Matching

• Regression using the PS as a patient-level covariate
  – The PS is a scalar value representing the clinical profile of the patient

• Inverse probability weighted estimators (IPWE) or Inverse probability treatment weighting (IPTW)
  – Weighted regression with weights = 1/probability of receiving the treatment received
Comparing Strategies

• Trim regions of non-overlap in most approaches

• In a matched sample, use paired statistics
  – McNemar’s Chi-square
  – Paired T-test
Figure 3. Rates of Survival in the CABG and PCI Populations, from an Analysis Adjusted with the Use of Inverse Probability Weighting.
Why Propensity Score Over Standard Regression

• Even when outcomes are more rare, can include a large number of covariates in propensity model.

characteristics (15). Propensity scores represent the estimated probabilities of patients receiving DES versus BMS in our population (15), in this case conditioned upon 102 observed covariates (Online Table). Inverse probability

• Traditional regression is limited to a certain number of predictors compared to number of events in sample (rule of thumb 10:1)

Summary of Propensity Score Advantages

• Separate stages divides the design from analysis

• Eliminates problem of low predictor: outcome ratio.

• Ability to examine overlap, trim tails, limit to a sample “eligible” for both treatments

• Intuitive presentation of data, especially for matching
Limitations of Propensity Scores

• How to choose variables for the model?
  – There may be hundreds of variables to choose from.
  – The “kitchen sink approach”? But one should only include prognostic variables in the model.

• Are all the important confounding variables captured in the data?

Major Limitation of All PS Approaches

• PS methods balance treatment groups with respect to measured covariates only.

• RCTs balance treatment groups with respect to measured and unmeasured covariates.
The primacy of clinical subject-matter knowledge

What the statistician can do:

• Examine distributions of PS and trim overlap or match appropriately
• Check balance and improve it
• Use appropriate paired testing
• Do sensitivity analyses to examine the sensitivity of findings to an unmeasured confounder
The primacy of clinical subject-matter knowledge

What the clinician can do:
Assess the plausibility of whether the PS model has appropriately accounted for unmeasured confounding.

**Surgical Candidacy and Selection Biases in Nonemergent Left Main Stenting**

*Implications for Observational Studies*

Edward J. McNulty, MD,*† William Ng, MBBS,* John A. Spertus, MD,§
Jonathan G. Zaroff, MD,* Robert W. Yeh, MD,∥ Xiushi M. Ren, MD,‡
Robert J. Lundstrom, MD*  
San Francisco and Redwood City, California; Kansas City, Missouri; and Boston, Massachusetts

**Conclusions** Surgical ineligibility dictating treatment selection is common in patients undergoing nonemergent ULM PCI, occurs on the basis of risk factors not captured by the ACC–NCDR, and is independently associated with worse long-term outcomes after adjusting for standard risk scores. (J Am Coll Cardiol Intv 2011;4:1020–7) © 2011 by the American College of Cardiology Foundation
A Typical Conversation in a “Heart Team” Meeting

Fellow: “Mr. Smith is 87, has 3VD, 60% LM, CKD. His anatomy looks surgical.”

Surgeon: “I don’t think he is going to rehabilitate well. He’s frail, uses a walker. Plus his touchdowns are terrible. Can you address it with PCI? The patient has a strong preference for PCI and I think it might be better.”

Interventionalist: “Yes, but it is a complex rotablator case, and his RCA will probably need to be staged.”

Consensus: Complex PCI, surgical non-candidate.
"I think PCI might be better" to the Epidemiologist

Diffuse severe CAD → Active Smoker → Uncontrolled Diabetes

Bad touchdowns

Kidney Disease

Neuropathy

Uses a Walker

Few studies are able to account for all of these variables!
1013 pts at MGH or BWH with multivessel or LM disease who underwent PCI

218 (22%) deemed “ineligible” for CABG
“Ineligibility” =
- 6-fold increase in in-hospital mortality after adjustment
- 3-fold increase in long term mortality

Figure 2. Mortality. Kaplan-Meier estimates comparing long-term mortality among those with surgical anatomy undergoing percutaneous revascularization stratified by documentation of surgical ineligibility. Surgical ineligibility (Inelig) was associated with a significant increase in mortality compared with those who were surgically eligible (Elig; log-rank \( P < 0.001 \)).
What can be done about confounding?
Falsification Hypotheses: A check for confounding

- Negative control hypotheses.
- Radial access vs. femoral access for coronary intervention example.

Baklanov et al. JACC 2013
Using the radial artery reduces GI bleeding?

Method 1: Logistic Regression
- Primary hypothesis: Access site bleeding
- Falsification hypothesis: Non-access site bleeding

Method 2: 1:1 Matching
- Primary hypothesis: Access site bleeding
- Falsification hypothesis: Non-access site bleeding

Method 3: IPTW
- Primary hypothesis: Access site bleeding
- Falsification hypothesis: Non-access site bleeding

Data from RIFLE STEACS randomized trial
- Primary hypothesis: Access site bleeding
- Falsification hypothesis: Non-access site bleeding

Odds ratio (95% CI)

- Method 1: Logistic Regression
  - Access site bleeding: 0.16 (0.05, 0.52)
  - Non-access site bleeding: 0.48 (0.26, 0.86)

- Method 2: 1:1 Matching
  - Access site bleeding: 0.20 (0.04, 0.71)
  - Non-access site bleeding: 0.44 (0.20, 0.93)

- Method 3: IPTW
  - Access site bleeding: 0.07 (0.02, 0.22)
  - Non-access site bleeding: 0.46 (0.25, 0.84)

- Data from RIFLE STEACS randomized trial
  - Access site bleeding: 0.34 (0.16, 0.68)
  - Non-access site bleeding: 0.96 (0.53, 1.74)

Wimmer et al. JACC 2014
Finding “Natural” Randomization

- IV approaches in observational studies rely on finding “randomization” that occurs outside of the randomized clinical trial.

- What is the effect of being a military veteran on mortality?
  - Problem of potential confounding.
  - The Vietnam War draft lottery was randomly assigned, but lower draft number meant one was much more likely to be a veteran than high draft number.
Examples of Instrumental Variables in the Literature: Weekend vs. Weekday Admission

• Question: What is the benefit of early vs. delayed catheterization after NSTEACS?
• Problem: Early catheterization may reflect ongoing chest pain or unstable symptoms (i.e. may lead to confounding).
• Instrumental Variable: Pts admitted on the weekend for NSTEMI are likely to wait longer for cath than patients admitted on weekdays.
• Assumptions: The day of presentation may be considered random. There are not other differences in the patient types or care on weekends vs. weekdays that influence outcomes apart from early vs. delayed cath.
• Findings: No benefit to early catheterization for NSTEACS

Physician Preferences—Can they be considered random?

• There is wide practice variation in physician selection of drugs and devices.
  – Not just based on the data.

Examples:
– Use of anticoagulants during stent procedures.
– Use of radial vs femoral access
– Use of arterial closure devices.
Anticoagulants during PCI

- Heparin vs. Bivalirudin
- Conflicting data from RCTs
- Wide practice variation in use
  - 100% users
  - 0% users
  - Everyone else in between
### Patient Characteristics by Operators

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>PCIs by High Bivalirudin Operators (N=230,447)</th>
<th>PCIs by Low Bivalirudin Operators (N=272,106)</th>
<th>Standardized Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yrs, median (IQR)</td>
<td>63 (54,73)</td>
<td>63 (54,73)</td>
<td>-0.68</td>
</tr>
<tr>
<td>Male (%)</td>
<td>158,433 (68.8)</td>
<td>186,690 (68.6)</td>
<td>-0.30</td>
</tr>
<tr>
<td>White (%)</td>
<td>198,079 (86.0)</td>
<td>236,289 (86.8)</td>
<td>2.58</td>
</tr>
<tr>
<td>BMI (kg/m²), median (IQR)</td>
<td>28.5 (25.2,32.7)</td>
<td>28.7 (25.3,32.9)</td>
<td>1.77</td>
</tr>
<tr>
<td>Current smoking (%)</td>
<td>82,208 (35.7)</td>
<td>100,706 (37.0)</td>
<td>2.78</td>
</tr>
<tr>
<td>Hypertension (%)</td>
<td>170,514 (74.0)</td>
<td>201,286 (74.0)</td>
<td>-0.04</td>
</tr>
<tr>
<td>Dyslipidemia (%)</td>
<td>154,880 (67.2)</td>
<td>182,155 (66.9)</td>
<td>-0.57</td>
</tr>
<tr>
<td>Family history of CAD (%)</td>
<td>48,487 (21.0)</td>
<td>61,304 (22.5)</td>
<td>3.61</td>
</tr>
<tr>
<td>Renal failure (currently on dialysis or GFR&lt;30) (%)</td>
<td>12,162 (5.3)</td>
<td>13,984 (5.1)</td>
<td>-0.62</td>
</tr>
<tr>
<td>Cerebrovascular disease (%)</td>
<td>24,335 (10.6)</td>
<td>29,388 (10.8)</td>
<td>0.78</td>
</tr>
<tr>
<td>Peripheral artery disease (%)</td>
<td>21,621 (9.4)</td>
<td>26,541 (9.8)</td>
<td>1.26</td>
</tr>
<tr>
<td>Chronic lung disease (%)</td>
<td>31,119 (13.5)</td>
<td>38,352 (14.1)</td>
<td>1.71</td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td>74,686 (32.4)</td>
<td>87,194 (32.0)</td>
<td>-0.78</td>
</tr>
</tbody>
</table>
## Endpoints by Bivalirudin Operators

<table>
<thead>
<tr>
<th></th>
<th>PCIs by High Bivalirudin Operators (N=230,447)</th>
<th>PCIs by Low Bivalirudin Operators (N=272,106)</th>
<th>Crude Risk Difference</th>
<th>Relative Risk Reduction</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-hospital bleeding (NCDR v4)</td>
<td>19,285 (8.4)</td>
<td>29,240 (10.7)</td>
<td>-2.3</td>
<td>21.5</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>In-hospital mortality</td>
<td>8,234 (3.6)</td>
<td>10,036 (3.8)</td>
<td>-0.20</td>
<td>5.26</td>
<td>0.020</td>
</tr>
<tr>
<td><strong>Secondary (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access-site bleeding</td>
<td>5,372 (2.4)</td>
<td>9,005 (3.3)</td>
<td>-0.90</td>
<td>27.3</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Non-access site bleeding</td>
<td>1,642 (0.72)</td>
<td>2,518 (0.93)</td>
<td>-0.21</td>
<td>22.6</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>RBC transfusion</td>
<td>8,574 (3.8)</td>
<td>12,578 (4.7)</td>
<td>-0.90</td>
<td>19.1</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>In-stent thrombosis</td>
<td>1,322 (0.57)</td>
<td>1,044 (0.38)</td>
<td>0.19</td>
<td>-50.0</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
What about choosing stents?

Instrumental Variable Analysis to Compare Effectiveness of Stents in the Extremely Elderly

Robert W. Yeh, MD, MSc; Samip Vasaiwala, MD, MSc; Daniel E. Forman, MD; Treacy S. Silbaugh, BSc; Katya Zelevinski, BA; Ann Lovett, RN, MA; Sharon-Lise T. Normand, PhD; Laura Mauri, MD, MSc

- DES use has shifted markedly over time.
- Time period itself is a strong predictor of any patient’s likelihood of receiving DES.
- Time period may not be strongly associated with outcomes via other paths.
Results of IV Comparison

Table 3. Primary Results—Unadjusted and Instrumental Variable–Based 1-Year Outcomes

<table>
<thead>
<tr>
<th>1-Year Outcomes</th>
<th>Unadjusted Outcomes</th>
<th>Unadjusted Instrumental Variable Outcomes</th>
<th>Adjusted Instrumental Variable Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DES (n=1507)</td>
<td>BMS (n=1183)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Risk Difference</td>
<td>Risk Difference</td>
<td>Risk Difference</td>
</tr>
<tr>
<td>Mortality, %</td>
<td>14.5</td>
<td>23.0</td>
<td>-2.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.35</td>
</tr>
<tr>
<td>Bleeding, %</td>
<td>10.3</td>
<td>12.4</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.68</td>
</tr>
<tr>
<td>Target vessel</td>
<td>4.3</td>
<td>9.3</td>
<td>-7.3</td>
</tr>
<tr>
<td>revascularization, %</td>
<td></td>
<td></td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.0001</td>
</tr>
</tbody>
</table>

Good Instrumental Variables Are Hard To Find!
Example: HDL and outcomes

Example: HDL and outcomes

- Could be confounded by:
  - Exercise
  - Diet
  - Non HDL lipid levels
  - Other?

How can we assess whether HDL is in the causal pathway with CV outcomes short of conducting a trial using a selective HDL-raising drug?
Mendelian Randomization

- Some genetic variants lead to alterations of the level of a biomarker.
  - If that biomarker (e.g. LDL) causes change in event risk, then the genetic variant should be associated with adverse events.
  - Any particular genetic variants is randomly inherited. (Mendel’s second law)
Nature’s Randomized Study

Conventional Trial

- Sample
  - Randomisation
    - Drug (e.g. statin)
      - LDL-C lower
      - CV event rate lower
    - Control
      - LDL-C unchanged
      - CV event rate unchanged

Mendelian Randomization

- Population
  - Random allocation of alleles
    - Genotype group A
      - LDL-C lower
      - CV event rate lower
    - Genotype group B
      - LDL-C unchanged
      - CV event rate unchanged

Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study

![Graph showing the relationship between HDL cholesterol and risk of myocardial infarction.](image)

**Table 3:** Instrumental variable analysis estimate of the association of genetically raised HDL cholesterol and risk of myocardial infarction using LIPG Asn396Ser as an instrument

<table>
<thead>
<tr>
<th>Number of individuals</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cases</strong></td>
<td><strong>Controls</strong></td>
</tr>
<tr>
<td>AngioGOCARD/KORA</td>
<td>0.76 (0.46-1.24)</td>
</tr>
<tr>
<td>IFS</td>
<td>1.25 (0.51-3.08)</td>
</tr>
<tr>
<td>deCODE</td>
<td>0.97 (0.60-1.56)</td>
</tr>
<tr>
<td>EPIC-NL</td>
<td>0.64 (0.27-1.55)</td>
</tr>
<tr>
<td>GemMIFS-II</td>
<td>1.36 (0.82-2.24)</td>
</tr>
<tr>
<td>GRACE</td>
<td>2.48 (1.10-5.56)</td>
</tr>
<tr>
<td>MAHA</td>
<td>1.08 (0.68-1.72)</td>
</tr>
<tr>
<td>PennATH</td>
<td>0.82 (0.37-1.83)</td>
</tr>
<tr>
<td>UCP</td>
<td>0.87 (0.44-1.77)</td>
</tr>
<tr>
<td>POPGEN</td>
<td>0.69 (0.42-1.14)</td>
</tr>
<tr>
<td>PROCARDIS</td>
<td>0.66 (0.45-0.98)</td>
</tr>
<tr>
<td>PROMIS</td>
<td>1.27 (0.74-2.16)</td>
</tr>
<tr>
<td>SHEEP</td>
<td>1.35 (0.85-2.14)</td>
</tr>
<tr>
<td>WTCCC</td>
<td>0.74 (0.49-1.12)</td>
</tr>
<tr>
<td><strong>All case-control studies</strong></td>
<td>0.84 (0.62-1.20)</td>
</tr>
</tbody>
</table>

**Meta-analysis of cohort studies**

| Per 0.03 mmol/L (1 mg/dL) increase in plasma HDL cholesterol | 0.98 (0.97-0.98) | 4×10⁻⁶⁴ | 1.02 (0.95-1.09) | 0.64 |
| Per 0.39 mmol/L (15 mg/dL) increase in plasma HDL cholesterol | 0.70 (0.66-0.74) | 4×10⁻⁶⁴ | 1.28 (0.46-3.61) | 0.64 |

Figure 2: Association of LIPG Asn396Ser with myocardial infarction in 116,320 participants from 20 studies. In each study, the HDL-cholesterol-raising serine allele was modeled.

Expanding graveyard of HDL-raising drugs
Summary

• Large databases have tremendous power but also many limitations for CER.

• Propensity score methods are commonly used to address confounding in non-randomized comparisons of treatments.
Summary

• Different methods of implementing propensity scores including regression, stratification, weighting and matching.

• Ultimately, clinical-subject expertise, in combination with knowledge of the limitations of the methods, are critical to understanding whether observational treatment comparisons are valid.
Thank you!

Questions?

ryeh@bidmc.harvard.edu