Module 3H
Working with Observational Data

Jack Needleman
Department of Health Policy and Management
Fielding School of Public Health
March 6, 2014
Outline of Session

- Why use observational data
- Data sources
  - Clinical records
  - Administrative claims data
  - Registries
  - Survey data
- Extracting, merging and coding observational data
- Design of studies for using observational data
  - Addressing selection in design and sampling
  - Addressing selection using statistical adjustment
    - Propensity scoring
    - Instrumental variables
    - Sample selection models
Randomized control trial is considered gold standard for clinical research

Yet not always appropriate or necessary
Critique of the view that no treatment or intervention should be adopted without first testing it using an RCT

- Purported systematic review of randomized controlled trials to determine whether parachutes are effective in preventing major trauma related to gravitational challenge (!)
- “Advocates of evidence-based medicine have criticized the adoption of interventions evaluated by using only observational data. We think that everyone might benefit if the most radical protagonists of evidence-based medicine organized and participated in a double blind, randomized, placebo-controlled, crossover trial of the parachute.”

- Learn from non-RCT data all the time
So what makes RCTs so attractive

- Relatively few assumptions
  - Once pick endpoints, mainly statistical theory
  - Few conceptual or theoretical modeling challenges
  - Causality can be assumed

- Well defined intervention/treatment

- Restricts selection into treatment
  - Biggest concern with observational data
  - Individuals or organizations select into treatment based on expectations about value or outcomes
    - Those most likely to benefit most likely to choose treatment

- Balance in other factors that might influence effectiveness
  - Observables and unobservables

*Non-RCT studies must address selection and balance issues*
RCTs: Not a Panacea!

- **Attrition=Selection**
  - Can introduce nonrandomness

- **Control group behavior also matters**
  - Demoralization may overstate effectiveness
  - Competitive rivalry may understate effectiveness

- “**Intent to treat**” (ITT) effect: Difference between originally assigned groups
  - Non-compliance among intervention group or contamination among control group may lead to treatment looking less effective than it is
  - Often interested in treatment of the treated (TOT) or “as-treated” estimates; TOT > ITT
RCTs - Disadvantages

- Time consuming, expensive, complex
- May require a large number of clusters (units of randomization)
- Restrictive inclusion and exclusion criteria limit generalizability
- Often not appropriate for health plan or other system level interventions
  - Fidelity of intervention hard to assure
- Unlikely to tell you whether an intervention will improve routine practice
Why Observational Data?

- **Necessity**
  - Legal, ethical, or practical considerations make it impossible to conduct a true experimental design
  - Random assignment is not possible
  - Control/comparison groups cannot be incorporated into the design

- **Absolute advantages**
  - Large samples
  - Observe treatment in more typical settings and populations
    - Effectiveness studies
  - Where real world implementation matters
Models for designs using observational data

- **Non-experimental or pre-experimental**
  - Use all available data
  - Often just pre-post or treatment-control
  - Statistical methods to control for selection or balance covariates
  - Example of large nonexperimental study
    - Framingham heart study

- **Quasi-experimental**
  - Non-random assignment to treatment and control
  - Exploit conditions that sort individuals intro treatment and control that minimize threats of selection and attrition
    - Thus achieving equivalence of randomization
      - Reduced selection
      - Balanced covariates

- Will return to designing observational studies later
Outline of Session

- Why use observational data
- **Data sources**
  - Clinical records
  - Administrative claims data
  - Registries
  - Survey data
- Extracting, merging and coding observational data
- **Design of studies for using observational data**
  - Addressing selection in design and sampling
  - Addressing selection using statistical adjustment
    - Propensity scoring
    - Instrumental variables
    - Sample selection models
Data sources for observational data

- Clinical records
- Claims databases
- Registry data
- Survey data
Clinical records

- Paper and EHR
  - Physician, hospital, other providers, pharmacy
- Details of treatments, tests, progress notes, medical history, clinical indications, measures of outcomes and impacts of treatment
- Challenges
  - Decentralized
  - Paper requires manual extracting
  - Extracting data from EHRs
    - Non-comparable coding
    - Extracting data from open text fields
  - Incomplete
Claims data bases

- **Extracts from records**
- **Standardized, electronic but limited**
  - Diagnosis, treatment, some demographics
- **From providers:** Hospital discharge, ED
- **From payers:** Claims
  - All payer claims data bases being developed
- **Challenges**
  - Incomplete descriptions of treatment & diagnosis
  - Virtually no clinical data, histories, or outcomes of treatment
  - Linkages across records to compile patient history
    - Confidential identifiers
    - Carve-outs & multiple providers or payers
  - Incomplete data on costs or charges
  - Data errors
- **Widely used**
Disease registries

- E.g., Cancer SEER registries
- Include data on disease stage, demographic, history and other items
- May be linked to clinical or claims records, or national death index
  - E.g., Medicare
  - E.g., linked maternal and infant records linked to death records and treatment records

Challenges
- Incompleteness
Surveys

- E.g., National Ambulatory Medical Care Survey or National Hospital Ambulatory Medical Care Survey
  - Sample of providers, with some information available on provider
  - Sample of patients over fixed time period
    - More information than in claim record
  - Allows analysis of selected elements of patient treatment and patient and provider characteristics
    - E.g., choice of drugs, decision to hospitalize
- Challenges
  - Limited sample and time frame
  - No linkages among patient records
  - Data limitations
Some Sources for Identifying Secondary Datasets

- National Center for Health Statistics (NCHS)
  - http://www.cdc.gov/nchs/
- Inter-University Consortium for Political and Social Research (ICPSR)
  - http://www.icpsr.umich.edu/icpsrweb/ICPSR/
  - More info about ICPSR during Friday’s lab
- Research Data Assistance Center (ResDAC)
  - http://www.resdac.umn.edu/
  - Preferred source for researching CMS datasets
- AcademyHealth website
  - http://www.hsrmethods.org/DataSources.aspx
<table>
<thead>
<tr>
<th>Survey Name</th>
<th>Purpose</th>
<th>Sponsor and History</th>
<th>Method</th>
<th>Data Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Health Interview Survey (NHIS)</td>
<td>To assess health status and behaviors, including use of health care</td>
<td>National Center for Health Statistics (NCHS). Began in 1957</td>
<td>Household interview survey of about 35,000 households annually</td>
<td>Public use microdata are available to download from the NCHS</td>
</tr>
<tr>
<td>National Health and Nutrition Examination Survey (NHANES)</td>
<td>To measure health and nutritional status, based on physical examinations and laboratory tests</td>
<td>National Center for Health Statistics (NCHS). Began in 1959</td>
<td>Household interview, followed by a mobile examination center. About 9,000 individuals annually</td>
<td>Public use microdata are available to download from the NCHS. Limited access to some parts of the data because of confidentiality</td>
</tr>
<tr>
<td>National Health Care Surveys (NHCS)</td>
<td>A family of surveys to track characteristics and practices of various health care provider organizations</td>
<td>National Center for Health Statistics (NCHS). Began in 1973</td>
<td>Organization surveys using various probability sampling methods, depending on the type of organization being surveyed</td>
<td>Public use microdata are available to download from the NCHS</td>
</tr>
<tr>
<td>Behavioral Risk Factor Surveillance System (BRFSS)</td>
<td>To understand health behaviors and risks, such as diet, smoking, exercise, and many more</td>
<td>Conducted by state health agencies, under guidance from the Centers for Disease Control and Prevention (CDC). Began in 1984</td>
<td>Telephone interview survey of about 350,000 individuals annually</td>
<td>CDC’s Web Enabled Analysis Tool (WEAT) provides online analysis. Public use microdata are available to download</td>
</tr>
<tr>
<td>Youth Risk Behavior Survey (YRBS)</td>
<td>To understand health behaviors and risks of the nation’s youth</td>
<td>Conducted by the CDC. State and local health agencies conduct their own YRBSs, with guidance from the CDC. Began in 1991</td>
<td>Group-administered survey of a multistage random sample of schools. About 15,000 ninth through twelfth graders surveyed nationally. There are also about 40 state and 20 local YRBSs</td>
<td>CDC’s Youth Online provides online data access tools. Public use microdata are available to download</td>
</tr>
</tbody>
</table>
Outline of Session

- Why use observational data
- Data sources
  - Clinical records
  - Administrative claims data
  - Registries
  - Survey data
- Extracting, merging and coding observational data
- Design of studies for using observational data
  - Addressing selection in design and sampling
  - Addressing selection using statistical adjustment
    - Propensity scoring
    - Instrumental variables
    - Sample selection models
Extracting, merging and coding observational data

- Data entry and merging
  - Linking
  - Recoding
    - Interpretation
  - Data cleaning
  - Missing data
  - Free text
Linking and merging data

- Linking data through a geographic information system (GIS)
  - Example: A survey of health behaviors (including exercising) can be linked with GIS data on the availability of parks and recreation facilities

- Quantitative and qualitative data can be linked
  - Example: Researchers can conduct focus groups with respondents to a quantitative survey
Examples of linking and merging data

- Build multi-year panel data from single year files
  - AHA
  - Medicare

- Link records across data sets
  - SEER and death records/hospital records
    - Multiple fields to create unique identifiers
    - Probabilistic matching

- Link records within data sets when coding is imperfect
  - OSHPD discharge records
    - Probabilistic matching ignoring errors
Challenge of creating a multi-year AHA data set

- **1983-2003**
- **Challenges**
  - Changing variable names
  - New variables
  - Changing variable coding from numeric to alpha or added codes
  - Changing names/addresses
  - Changing AHA numbers
    - Closures/openings
  - Missing data
    - Interpolation
    - Verifying changes in variables (e.g., ownership)
Choices in coding up primary data

- **Example: TCAB**
  - 67 hospitals, 2 units/hospital (control, treatment)
  - CNO and unit manager surveys (2 units/hosp) (3 waves)
  - Unit staff surveys (3 waves, some different questions, categorical coding and some open ended, ~1800 responses/wave)
  - Monthly quality measures with numerators & denominators

- **Linkage variables**
- **Variable names**
- **Alphanumeric or numeric**
- **Coding missing data**
- **Creating working data sets**
  - Multirecord data to single record – name conventions
EHRs as sources of data for CER

- More widely available
  - VA
- Include information from history, physical and tests and lab results claims data lacks
  - Data collected in the course of usual care
- Examples:
  - Cleveland Clinic study of risks of coronary artery disease and CHF associated with different oral diabetes agents
  - Study found Intensity Modulated Radiation Therapy (IMRT) had fewer GI complications than proton therapy
  - Effects of statin use on pneumonia mortality
  - Comparison of rates of preventive screening services among patients with/without rheumatoid arthritis.
EHR in CER: Issues and challenges

- Data warehousing as organizational strategy
- Structured and unstructured fields
  - interoperability and informatics standards for data standardization
  - Retrieving and interpreting free text entries
- Other issues
  - Completeness
  - Reliability
  - Data quality
Natural Language Processing

- Being applied to clinical text (records) and biomedical literature
- Recent review found:
  - Uses are “numerous and far-reaching”
  - Clinical text harder to analyze
    - Abbreviations (std & non-std), errors in spelling and abbreviation, less context
  - Performance improving, exceeding 90% specificity and sensitivity in several cases
  - Current methods require annotated texts to train systems
    - Focus on narrow range of textual analysis
  - Current systems rarely applied outside labs developed in, due to scalability and generalizability issues
  - Use of competitive challenges to spur development
50 discharge summaries to be analyzed

15 common obesity comorbidities and four states (present, absent, questionable, unmentioned)

- Unmentioned further processed and reclassified – an intuitive judgment, in contrast with textual judgment
  - Done via references to examination and test results (e.g., blood sugar measurements), physical characteristics (e.g., BMI), medications and other diseases discussed in summary

- Systems performed best on most factual and objective, more difficulties in cases where only medical experts could infer
  - Particular difficulty extracting correct Absent assessment

- Training materials critical and likely to be a bottleneck for future progress
Outline of Session

- Why use observational data
- Data sources
  - Clinical records
  - Administrative claims data
  - Registries
  - Survey data
- Extracting, merging and coding observational data
- Design of studies for using observational data
  - Addressing selection in design and sampling
  - Addressing selection using statistical adjustment
    - Propensity scoring
    - Instrumental variables
    - Sample selection models
Purpose of Study Design

- Most CER studies involve comparison or contrast
  - Between groups or over range of variation in some variable
  - Over time or space
- To help ensure that a data set can provide a convincing test of a conceptual model and/or hypotheses
  - Helps bolster the case that any effects found from an intervention or policy were in fact the result of it, rather than due to extraneous factors
  - Makes it easier to understand any reasons to believe that this isn’t the case (i.e., “validity”)
- Key questions:
  - What else could explain the observed association (or absence of association)?
  - How are these competing hypotheses ruled out by design, sample choice, other aspects of the research?
Nonexperimental and Quasi-experimental study designs are critical elements of the research firmament.

- While they may offer weaker or more challenged evidence of causality, when well done routinely relied upon and trusted

**Question:** How can one make non-experimental or quasi experimental designs stronger?
Tools/dimensions in study design
What can you manipulate in design?

- Sample
  - Presence/absence of control group or other group to compare if analyzing treatments
  - Choose where possible to reduce selection biases
- Frequency/time of data collection
- Scope of measures/covariates/data included in analysis
- Statistical methods
Selection of Study Design

- Should build on theoretical, qualitative, and modeling work
- Should be closely linked to the purpose of the evaluation
- Goals are to:
  - Minimize bias
  - Increase confidence in causal interpretation
  - Rule out competing interpretations
  - Maximize generalizability
- See Eccles, et al., Quality and Safety in Health Care, 2003: 12:47-52 for further discussion
Quasi-Experimental Study Designs

- Studies that take an experimental approach without having full experimental control
  - Between true experimental designs and pre-experimental designs
  - Example: Pre-post study with (non-randomized) comparison group to examine the impact of mandating mental health and substance abuse parity coverage for Federal employees
- Try to incorporate sample strategies that reduce likelihood of selection and make assignment look “quasi-random”
  - E.g., use formulary differences across health plans to compare relative performance of different drugs
  - E.g., use waiting lists to compare WIC recipients to nonrecipients
- May also use statistical methods, like matching or propensity scoring to assure balance
Quasi-Experimental Designs

- **Separate sample pre-post test design**
  - Same or different groups believed to be equivalent
- **Nonequivalent control group design**
  - Treatment/control
- **Pre-post with control**
  - Same group or equivalent groups in pre & post studies

- **Simple pre-post can be expanded into interrupted time series designs if have multiple measurements**
  - More convincing because changes in trend more convincingly associated with intervention
Internal Validity (Aneshensel, 2002)

- **Definition**: The extent to which conclusions about cause and effect can be drawn from the research, i.e., did the intervention or policy make a difference in that particular instance?
  - In experimental research (generally conducted in lab settings), internal validity achieved through design (control or hold constant factors that could influence the outcome of interest).
  - In non-experimental research (conducted in naturally occurring settings or on observational study), internal validity achieved through design and analysis.
Threats to Internal Validity

- **Selection**
- History
- Maturation
- Testing
- Instrumentation
- Statistical Regression
- Attrition
- Diffusion or compensation
Selection

• Different types of individuals (or units of analysis) are partly responsible for the differential effects observed in $O_1$ vs. $O_2$
  
  • Example: A comparison of end-of-life costs may be confounded by unobserved preferences for aggressive vs. palliative care
  
  • E.g., younger patients chosen for surgery vs. medical treatment
  
  • E.g., physician visits: Medicaid vs. uninsured
  
• Common problem in CER
  
  • Other terms
    
    • Omitted variable bias (variables that predict entry into treatment)
    
    • Treatment selection
Threats to Internal Validity

- Selection
- History
- Maturation
- Testing
- Instrumentation
- Statistical Regression
- Attrition
- Diffusion or compensation
Something occurring between observation 1 ($O_1$) and observation 2 ($O_2$) in addition to the predictor ($X$) is partly responsible for the observed effects

- Example: The estimated impact of changes in Medicare reimbursement under the 1997 BBA on home healthcare use may have been confounded by the effect of Operation Restore Trust (OIG fraud investigation of HH agencies) during the same time period.

- Pre-post designs particularly susceptible to this
  - Pre-post with control can adjust (Diff in diff) if controls experienced same history

- Unique events occurring over limited group can provide basis for analysis of natural experiments

*How do you identify threats due to other events?*
Threats to Internal Validity

- Selection
- History
- Maturation
- Testing
- Instrumentation
- Statistical Regression
- Attrition
- Diffusion or compensation
Maturation

- The passage of time between $O_1$ and $O_2$ is partly responsible for the observed effects
  - Example: In a pre/post study, the impact of psychotherapy on the outcomes of depressed patients may be confounded if patients would have spontaneously remitted over time anyway, due to the natural course of the disease
- **Pre-post particularly susceptible**
  - Treatment/control less if subjects at same point of maturation
  - Pre-post with control, diff in diff help control
Threats to Internal Validity

- Selection
- History
- Maturation
- **Testing**
- Instrumentation
- Statistical Regression
- Attrition
- Diffusion or compensation
Testing effects

- Familiarity with the process of being measured at $O_1$ is partly responsible for the value of $O_2$.
  - “Hawthorne Effect”: The process of being “watched” affects the response.
  - Familiarity with test or awareness of testing contributes to change
  - Example: A study of whether providing educational materials and diet plans can promote weight loss among obese patients may be confounded by the fact that study participants knew they would be weighed regularly to measure their progress

  - **One option is to add cohorts to treatment and controls that aren’t initially tested**
    - Diff between tested and untested at end is measure of testing effect

  - **Testing effects can be part of program design**
Threats to Internal Validity

- History
- Maturation
- Testing
- Instrumentation
- Statistical Regression
- Selection
- Attrition
- Diffusion or compensation
• Differences in the measuring instrument between $O_1$ and $O_2$ are partly responsible for the observed effects

• Example: A new neighborhood clinic is built and patient satisfaction with their health care is evaluated. Four interviewers are used but only two were used in both periods.
• Comparing patient satisfaction with care when different providers use different instruments or sample selection methods
Threats to Internal Validity

- Selection
- History
- Maturation
- Testing
- Instrumentation
- Statistical Regression
- Attrition
- Diffusion or compensation
“Regression to the mean” between $O_1$ and $O_2$ is partly responsible for the observed effects.

Example: You run the claims processing group at an HMO and study number of claims processed per employee over a one-week period. You then take the bottom decile and provide them with additional training; their productivity goes up by 20% the next month.

Regression to mean due to selecting cases with $E(e)>0$ or $<0$

Issue of sampling strategy

One option:

Adjust by doing second test that forms baseline for measurement
Threats to Internal Validity

- Selection
- History
- Maturation
- Testing
- Instrumentation
- Statistical Regression
- Attrition
- Diffusion or compensation
• Differential loss of subjects over the course of the study is partly responsible for the differential effects observed in $O_1$ vs. $O_2$
  • E.g. A group of older at-risk drinkers agreed to participate in a randomized trial in which some are assigned to an educational intervention and others receive usual care; those in the intervention had lower rates of at-risk drinking at 1 year, but also had more dropout between baseline and followup

• Other side of selection

• Major threat to RCTs and other randomized studies

• Approaches
  • Aggressive efforts to retain subjects
  • Intent to treat analysis
  • Sensitivity analysis of cases lost to followup
Threats to Internal Validity

- Selection
- History
- Maturation
- Testing
- Instrumentation
- Statistical Regression
- Attrition
- Diffusion or compensation
Diffusion or compensation

- Control group exposed to some/all of treatment or alternative/compensatory treatment with similar impacts to treatment being studied
  - E.g., TCAB AONE controls
- Need to check/test
- Strategies to reduce risk
  - Seek samples in which treatments/controls are isolated or separated by time or space
    - Need to consider differences in history, etc., if do this
  - Discourage diffusion/compensation
    - Alternative intervention
    - Stepped wedge design (delayed entry into treatment)
Disciplines can differ in their approaches to causality and design
<table>
<thead>
<tr>
<th>No.</th>
<th>Common Term</th>
<th>Synonymous Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Panel data study</td>
<td>Longitudinal or cohort study</td>
</tr>
<tr>
<td>2</td>
<td>Time series study</td>
<td>Longitudinal study</td>
</tr>
<tr>
<td>3</td>
<td>Cross-section, time series</td>
<td>Longitudinal study</td>
</tr>
<tr>
<td>4</td>
<td>Choice-based sampling</td>
<td>Case-Control study</td>
</tr>
<tr>
<td>5</td>
<td>Dependent variable</td>
<td>Outcome, response, endogenous variable</td>
</tr>
<tr>
<td>6</td>
<td>Explanatory variable of interest</td>
<td>Dose, treatment, exposure, intervention, Exogenous variable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of interest, predictor variable</td>
</tr>
<tr>
<td>7</td>
<td>Explanatory variable</td>
<td>Confounder, independent variable, regressor, exogenous variable, covariate</td>
</tr>
<tr>
<td>8</td>
<td>Interaction</td>
<td>Effect modification</td>
</tr>
<tr>
<td>9</td>
<td>Parameter estimate</td>
<td>Beta, regression coefficient, treatment effect</td>
</tr>
</tbody>
</table>

**STUDY DESIGNS**

**ELEMENTS OF AN EQUATION**

Table 1: Study Design and Statistical Terms in Health Services Research (cont’d)

<table>
<thead>
<tr>
<th></th>
<th>REGRESSION MODELS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Partitioned model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Multiple regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Qualitative analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Logit (or probit) model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Ordered logit regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Multinominal logit regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Conditional logit regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Survival analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Stratified model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Multivariate regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Categorical data analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Binomial logistic regression, logistic regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Ordinal logistic regression, ordinal log-linear regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Polytomous logistic regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Conditional logistic regression, McFaddens logit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Cox regression, hazard model, duration model failure-time analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Event history analysis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TYPES OF BIAS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>Omitted variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Unmeasured covariate, unmeasured confounder, unobservable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Sample selection bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Censoring, selection bias, incidental truncation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Selection bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Unmeasured confounding, omitted variable bias, confounding by indication or contraindication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>TYPES OF ESTIMATORS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Multiple multivariate regression</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Quasi-Experimental Designs

- **Separate sample pre-post test design**
  - Same or different groups believed to be equivalent
- **Nonequivalent control group design**
  - Treatment/control
- **Pre-post with control**
  - Same group or equivalent groups in pre & post studies

- **Simple pre-post can be expanded into interrupted time series designs if have multiple measurements**
  - More convincing because changes in trend more convincingly associated with intervention
Separate-Sample Pre-Test/Post-Test Design

- Also called serial cross-sectional design
  - “R” is random assignment to the time of measurement:

    \[
    \begin{array}{c}
    \text{O1} \\
    \text{R} \\
    \text{X} \\
    \text{O2}
    \end{array}
    \]

- Compare pre-test (1st group) scores with post-test (2nd group) scores
Nonequivalent Control Group (2)

- **Other names for this design**
  - “Controlled before and after”
  - “Difference-in-differences” (after the statistical technique used for such study designs)

- **Often used when experimental treatment is administered to whole groups, therefore random assignment of subjects is not possible**
  - A control population with similar characteristics is identified, eg:
    - Separation in place or time
    - Waiting lists or other sorting mechanisms
  - Baseline and post-intervention data are collected on both the control and intervention populations
Nonequivalent Control Group Design

- Design: \[ 0_1 \quad X \quad 0_2 \]
  \[ 0_3 \quad 0_4 \]

- Tends to be in between pre-experimental and experimental designs in terms of dealing with threats to internal validity.

- **Key difference from experimental design:**
  - No randomization

- **Key difference from pre-experimental (one-group pretest posttest) design:**
  - Control group added
Comparing Study Designs

- Study designs can be compared on the basis of how well they address threats to validity
- Definition: Threats to validity are reasons that may invalidate conclusions one draws from a data analysis (i.e., the conclusions are unsupported by the data, or wrong)
- Two overall categories:
  - Threats to “internal validity”
  - Threats to “external validity”
- Some threats standard, others specific to study or topic
  - But specific may be special case of generic category
    - E.g., unique events may fall under general category of history
Differences-in-Differences, Graphically

Pre  

Post  

Treatment

Control
Differences-in-Differences, Graphically

Effect of program difference-in-difference (taking into account pre-existing differences between T & C and general time trend).
Interrupted Time Series Design

Design:

\[ 0_1 0_2 0_3 0_4 \times 0_5 0_6 0_7 0_8 \]

Does the intervention improve care more than the observed secular trend?

Requires the collection of data multiple times both before and after so that you understand the magnitude of the secular trend.

Analysis must account for the auto-correlation of data collected at multiple time points

Advantage is that you do not need a control group; disadvantage is that you need data over multiple time points.
When is Interrupted Time Series Used?

- When there is a “natural experiment” or when the treatment or intervention can be systematically introduced by the researcher
- When the outcome is measured at a periodic interval (blood pressure, Hb1Ac)
- Greatest threat to internal validity is history, since some other simultaneous unmeasured change may account for changes in observed outcomes
- Multiple time-series design may get around this
Quasi-Experimental Designs

- Separate sample pre-post test design
- Nonequivalent control group design
- Interrupted time series design
- Multiple time-series design
Multiple Time Series Design

- **Design:** Multiple Time-Series Design
  
  \[ 0_1 \ 0_2 \ 0_3 \ 0_4 \ X \ 0_5 \ 0_6 \ 0_7 \ 0_8 \]
  
  \[ 0_1 \ 0_2 \ 0_3 \ 0_4 \ 0_5 \ 0_6 \ 0_7 \ 0_8 \]

- **Best way to control for history in a time series design**
  
  - Extension of the nonequivalent control group design (longer time series for each group)
  
  - One group is exposed to the treatment of interest and the other is not
  
  - There are multiple measures of the outcome in both groups in a pre-post design
Multiple Time Series Design (cont’d)

- **Strength of the design again depends on similarity of the experimental and control groups (time trends need to be the same)**
  - Sample selection critical
  - Search for natural experiments, external policies or circumstances that minimize selection
    - Change in formularies
    - Variations in diffusion
    - State mandates in insurance coverage
    - Variations in implementation
      - E.g. disease management strategies
Role of covariates and conceptual models in quasi-experimental studies

- **Attraction of RCTs is simplification of research design**
  - Causal pathways controlled
  - Balance among factors influencing outcomes achieved by randomization (but should be checked)

- **Given these are missing without randomization**
  - Should model causal pathways to identify risks of selection, other factors that might influence results or provide alternative explanation of findings
  - Need to identify additional variables to (ideally) include in analysis
    - Covariates to improve precision
    - Meditators (intervening variables in causal pathways)
    - Moderators (moderate influence of independent variable of interest)
Conceptual Models, Mediation

Work environment

Nursing processes, eg discharge education

Readmission

If don’t have data on work environment, nursing process coefficient inflated.

If don’t have data on nursing processes, Reduced form for all paths by which Work environment affects readmission
**Conceptual models, mediators and moderators**

- **Work environment**
- **Nurse Staffing**
- **Readmission**
- **Nursing processes, eg discharge education**
- **Patient Dx & Demog**

Focus on nurse staffing as both a direct and indirect variable affecting readmission.

Relation of work environment and nurse staffing complex, but focus on potential for work environment to moderate association of staffing and readmission.
The difference in the odds on dying in hospitals with 8:1 and 4:1 patient/nurse ratios is -
0 percent in hospitals with poor environments;
16 percent in hospitals with mixed environments;
46 percent in hospitals with good environments.
More on covariates later in discussing risk adjustment
Controlling for sample selection via statistical controls

- Propensity scoring
- Instrumental variables
- Sample selection models
Examples of Treatment Effects

- Comparison of health outcomes among patients who receive more intensive (e.g., ECT among depressed patients) vs. less intensive treatment (e.g., psychotherapy)
- Comparison of health outcomes among patients who see a specialist for their condition vs. those who see generalists
- Comparison of health care utilization among patients with generous vs. minimal insurance coverage
Summary of statistical methods to address selection

- **Match on observables**
  - Propensity scoring methods
    - Matching
    - Weighting
- **Use predicted value**
  - `ivregress`, `ivprobit`, etc.
- **Match on unobservables**
  - Treatment selection/treat prob
  - Bivariate probit
Concepts behind propensity scoring

- **RCTs control selection via randomization and assure balance among covariates**
- **Propensity scoring – Predict entry into treatment based on observable characteristics**
  - Propensity scores, developed by Rosenbaum and Rubin, are used in situations where omitted-variable bias or selection is a problem.
  - With the propensity score approach, try to “balance out” the groups being compared in terms of their covariates, based on predicted likelihood of obtaining treatment.
  - What allows this method to operate is that the process of matching patients to treatments in practice is not fully consistent.
    - Some patients who look like those who would get Treatment A do not, and some patients who look unlikely to get Treatment A do.
Propensity scoring

- Approach appears to have the greatest acceptance by clinicians because the adjustment process is or should be based on the factors that clinicians use in practice to determine choice of treatment.
  - The problem of applying propensity methods has been having the data to model treatment choices.
  - Where data to closely model this choice are available, the outcomes of propensity modeling can match the results of randomized trials.
  - Where these data are not available and the propensities are modeled using available data such as gender, age, or insurance status that may or may not be closely correlated with the treatment decision, the methods can produce inaccurate estimates of treatment effects.

- The challenge in using propensity scores is not methodological, but having good data to model selection
Research question: Do specialists achieve better CD4 counts among individuals with HIV infection
But factors influencing who sees specialist lead to large differences in $p$(specialist) and may also influence CD4.

Graphs by hivspec

- Graph 0: Shows a higher distribution of $Pr$(hivspec) compared to Graph 1.
- Graph 1: Displays a lower distribution with a narrower range.
Propensity methods attempt to balance factors influencing both treatment and outcomes

- Propensity methods
  - Matching
  - Stratification
  - Weighted regression
- All based on predicting probability of treatment via logit or probit
- No gold standard
  - Test for balance in covariates
Propensity score methods – key assumptions

- **Given propensity score, other covariates balance**
  - Can be tested for observables

- **Assignment to treatment is random for persons with same propensity score**
  - Cannot be tested
  - Need to assess treatment model against clinical understanding of how treatment decisions made
Advantages of propensity score methods

- The main advantage of the propensity score approach over the usual regression adjustment is that it avoids out-of-sample prediction due to linearity assumptions.
  - For example, if everybody in the first group is young and everybody in the second group is old, then simply controlling for age won’t eliminate the bias.

- Continuing work on when and how propensity score methods work well
Propensity score methods - Matching

- Divide patients into groups with a similar probability of receiving the treatment, regardless of whether or not they actually did.
  - That way the comparison of outcomes between treated and untreated patients is “quasi-randomized.” Formalizing selecting a “matched sample.”
  - First, you randomly order the treated subjects.
  - Next, you use the propensity score to match the first treatment patient to all control patients within a given caliper around the propensity score; if more than one possible match is made, you would pick the closest match.
  - Both the treatment and matched control observations put into the new dataset process is repeated for all treated subjects.
  - Propensity scoring weights matching criteria based on association with likelihood of treatment.
Note that patients in either group who do not match to patients in the other group are eliminated from the sample altogether.

Although eliminating observations in this way leads to concerns about generalizability, statisticians will argue that if you retain observations, the comparison between the treated and untreated patients will be spurious anyway.

So it’s better to restrict the sample, since you get an unbiased comparison while understanding clearly which patients the comparison applies to.
But factors influencing who sees specialist lead to large differences in $p(\text{specialist})$ and may also influence CD4.
Propensity score methods - Stratification

- Divide sample into quintiles based on similar propensity scores
  - Check for balance in covariates, increase number of divisions to assure balance in covariates
- Compute average difference in outcome within each quintile
- Take average difference across quintiles
  - Statistics for estimating average difference and variance/standard error for sample as a whole well established.
- Some subsamples may have only treated or untreated and without match these are not used
But factors influencing who sees specialist lead to large differences in $p(\text{specialist})$ and may also influence CD4.
Third approach to propensity score methods is to weight observations in data to make average weighted propensity for treatment equivalent in both treatment and control groups comparable:

- Downweight observations who were treated who have high propensity for treatment
- Upweight observations who were not treated with low propensity for treatment

Actual formula for weights $1/p(\text{treatment})$ for treated, $1/(1-p(\text{treatment}))$ for control.
Instrumental Variables approaches to controlling for selection

- **The problem**
  - $X \rightarrow Y$, but $Y \rightarrow X$ as well
  - Reverse causality
    - Endogeneity
    - Selection equivalent to reverse causality
  - Want estimate of $X \rightarrow Y$ unaffected by reverse causality

- **Solution**
  - If there is are variables $Z$ which predicts $X$ ($Z \rightarrow X$) but are unrelated to $Y$, except through the $X$, then use the predicted value of $X$ ($X_{\text{hat}}$) to predict $Y$
    - The prediction of the association of $X$ and $Y$ from this regression will be unbiased and without endogeneity
Condition!!s needed for IV to work

- $Y = f(X)$
- $X = g(Y)$
- $X = h(Z)$, monotonically
  - Relationship needs to be strong enough
- $Y$ not predicted by $Z$ when $X$ is controlled for
  - Cannot be tested if have only 1 instrument and tests imperfect when have multiple instruments
- Therefore, $Y = j(X_{\text{hat}})$ provides estimate of effect of $X$ on $Y$ with endogeneity controlled

- Need to adjust standard errors on coefficient estimated using $X_{\text{hat}}$
  - Stat programs will do this automatically
- Good instruments are hard to find!!
Suggested sources of instruments

- Geography
  - Distance, rivers, small area variation
- Legal/political institutions
  - Laws, election dynamics
- Administrative/program rules
  - Wage/staffing rules, reimbursement rules, eligibility rules, mandates
- “Natural” randomization
  - Draft, birth date, lottery, roommate assignment, weather
Examples of IV used in CER

  - Use distance to hospital with cath as instrument for whether patient received catheterization
  - Use regional cardiac catheterization rate as instrument
  - Also do propensity scoring
Analysis of Observational Studies in the Presence of Treatment Selection Bias
Effects of Invasive Cardiac Management on AMI Survival Using Propensity Score and Instrumental Variable Methods

Thérèse A. Stukel, PhD
Elliott S. Fisher, MD, MPH
David E. Wennberg, MD, MPH
David A. Alter, MD, PhD
Daniel J. Gottlieb, MS
Marian J. Vermeulen, MHSc

Context  Comparisons of outcomes between patients treated and untreated in observational studies may be biased due to differences in patient prognosis between groups, often because of unobserved treatment selection biases.

Objective  To compare 4 analytic methods for removing the effects of selection bias in observational studies: multivariable model risk adjustment, propensity score risk adjustment, propensity-based matching, and instrumental variable analysis.

Design, Setting, and Patients  A national cohort of 122,124 patients who were elderly (aged 65-84 years), receiving Medicare, and hospitalized with acute myocardial infarction.
Results Patients who received cardiac catheterization (n = 73 238) were younger and had lower AMI severity than those who did not. After adjustment for prognostic factors by using standard statistical risk-adjustment methods, cardiac catheterization was associated with a 50% relative decrease in mortality (for multivariable model risk adjustment: adjusted relative risk [RR], 0.51; 95% confidence interval [CI], 0.50-0.52; for propensity score risk adjustment: adjusted RR, 0.54; 95% CI, 0.53-0.55; and for propensity-based matching: adjusted RR, 0.54; 95% CI, 0.52-0.56). Using regional catheterization rate as an instrument, instrumental variable analysis showed a 16% relative decrease in mortality (adjusted RR, 0.84; 95% CI, 0.79-0.90). The survival benefits of routine invasive care from randomized clinical trials are between 8% and 21%. 
From: Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive Cardiac Management on AMI Survival Using Propensity Score and Instrumental Variable Methods


Table 1. Select Baseline Characteristics According to Receipt of Cardiac Catheterization*

<table>
<thead>
<tr>
<th>Overall Cohort</th>
<th>Propensity-Based Matched Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Received Cardiac Catheterization</td>
</tr>
<tr>
<td></td>
<td>Within 30 Days</td>
</tr>
<tr>
<td></td>
<td>n = 48866</td>
</tr>
<tr>
<td>Predicted 1-year mortality</td>
<td>32.3 (18.3)</td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
</tr>
<tr>
<td>Age range, y</td>
<td></td>
</tr>
<tr>
<td>66-74</td>
<td>40.2</td>
</tr>
<tr>
<td>75-84</td>
<td>50.8</td>
</tr>
<tr>
<td>Men</td>
<td>49.7</td>
</tr>
<tr>
<td>Black</td>
<td>7.5</td>
</tr>
<tr>
<td>Social Security income &gt;=$2000</td>
<td>30.0</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
</tr>
<tr>
<td>History of angina</td>
<td>44.1</td>
</tr>
<tr>
<td>Previous myocardial infarction</td>
<td>32.9</td>
</tr>
<tr>
<td>Previous revascularization</td>
<td>17.8</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>27.2</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>35.6</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>12.8</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>24.9</td>
</tr>
<tr>
<td>Smoker†</td>
<td>15.1</td>
</tr>
<tr>
<td>AMI clinical presentation characteristics</td>
<td></td>
</tr>
<tr>
<td>Non-ST-segment elevation AMI</td>
<td>41.8</td>
</tr>
<tr>
<td>Shock</td>
<td>1.9</td>
</tr>
<tr>
<td>Hypotension</td>
<td>3.5</td>
</tr>
<tr>
<td>Received CPR</td>
<td>1.8</td>
</tr>
<tr>
<td>Peak creatinine kinase &gt;1000 U/L</td>
<td>29.1</td>
</tr>
<tr>
<td>Hospital characteristics</td>
<td></td>
</tr>
<tr>
<td>Annual AMI volume &gt;200 patients</td>
<td>20.1</td>
</tr>
<tr>
<td>Mortality</td>
<td></td>
</tr>
<tr>
<td>Died within 1 y</td>
<td>38.6</td>
</tr>
<tr>
<td>Died within 4 y</td>
<td>62.0</td>
</tr>
</tbody>
</table>

*All data are presented as percentages. Standardized difference is the mean difference divided by the pooled SD, expressed as a percentage.
†Predicted 1-year mortality was computed using the Cox proportional hazards regression model, including all baseline patient characteristics of age, sex, socioeconomic status, comorbidities, and clinical presentation.
‡Defined as current smoker.
§Derived by Kaplan-Meier method.
Table 2. Distribution of Select Covariates by Propensity Score Deciles, According to Receipt of Cardiac Catheterization

<table>
<thead>
<tr>
<th>Decile (Range) of Propensity Score*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.00-0.26</td>
<td>0.26-0.40</td>
<td>0.40-0.50</td>
<td>0.50-0.58</td>
<td>0.58-0.65</td>
<td>0.65-0.70</td>
<td>0.70-0.75</td>
<td>0.75-0.80</td>
<td>0.80-0.85</td>
<td>0.85-0.98</td>
</tr>
<tr>
<td>No. of patients</td>
<td>10,021</td>
<td>8,219</td>
<td>6,873</td>
<td>5,763</td>
<td>4,834</td>
<td>3,997</td>
<td>3,283</td>
<td>2,628</td>
<td>2,060</td>
<td>1,208</td>
</tr>
<tr>
<td>No cardiac catheterization</td>
<td>2,191</td>
<td>3,993</td>
<td>5,340</td>
<td>6,449</td>
<td>7,378</td>
<td>8,215</td>
<td>8,930</td>
<td>9,585</td>
<td>10,151</td>
<td>11,006</td>
</tr>
<tr>
<td>Cardiac catheterization</td>
<td>54.5</td>
<td>39.2</td>
<td>31.8</td>
<td>27.5</td>
<td>23.4</td>
<td>20.0</td>
<td>17.3</td>
<td>15.3</td>
<td>14.0</td>
<td>13.6</td>
</tr>
<tr>
<td>Predicted 1-year mortality, %†</td>
<td>51.2</td>
<td>38.9</td>
<td>31.8</td>
<td>27.4</td>
<td>23.5</td>
<td>20.0</td>
<td>17.3</td>
<td>15.3</td>
<td>13.5</td>
<td>12.8</td>
</tr>
<tr>
<td>Mean age, y‡</td>
<td>79.4</td>
<td>78.0</td>
<td>77.0</td>
<td>75.5</td>
<td>74.3</td>
<td>72.9</td>
<td>71.9</td>
<td>70.8</td>
<td>70.1</td>
<td>70.0</td>
</tr>
<tr>
<td>History of congestive heart failure, %</td>
<td>79.3</td>
<td>77.9</td>
<td>76.8</td>
<td>75.7</td>
<td>74.3</td>
<td>73.0</td>
<td>71.8</td>
<td>70.9</td>
<td>70.0</td>
<td>69.9</td>
</tr>
<tr>
<td>No cardiac catheterization</td>
<td>59.8</td>
<td>40.0</td>
<td>27.0</td>
<td>18.8</td>
<td>10.8</td>
<td>7.3</td>
<td>4.2</td>
<td>2.7</td>
<td>2.0</td>
<td>2.1</td>
</tr>
<tr>
<td>Cardiac catheterization</td>
<td>61.4</td>
<td>40.0</td>
<td>26.5</td>
<td>16.7</td>
<td>10.5</td>
<td>5.7</td>
<td>3.6</td>
<td>2.5</td>
<td>2.0</td>
<td>1.7</td>
</tr>
</tbody>
</table>

*Propensity scores were rounded to 2 decimal points. There was no overlap across deciles.
†Predicted 1-year mortality was computed using the Cox proportional hazards regression model, including all baseline patient characteristics of age, sex, race, socioeconomic status, comorbidities, and clinical presentation.
‡SD for age was 4.3 years.

Figure Legend:
### Table 3. Adjusted Relative Mortality Rate Associated With Receipt of Cardiac Catheterization Among Patients With AMI Using Standard Risk-Adjustment Methods

<table>
<thead>
<tr>
<th>Risk-Adjustment Method</th>
<th>Relative Mortality Rate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted survival model</td>
<td>0.364 (0.358-0.370)</td>
</tr>
<tr>
<td>Multivariable survival model (65 covariates)</td>
<td>0.510 (0.502-0.519)</td>
</tr>
<tr>
<td>Survival models using simple propensity score*</td>
<td></td>
</tr>
<tr>
<td>Propensity deciles alone</td>
<td>0.538 (0.529-0.547)</td>
</tr>
<tr>
<td>Propensity deciles plus all covariates</td>
<td>0.520 (0.511-0.529)</td>
</tr>
<tr>
<td>Survival models using complex propensity score†</td>
<td></td>
</tr>
<tr>
<td>Propensity deciles alone</td>
<td>0.540 (0.531-0.549)</td>
</tr>
<tr>
<td>Propensity deciles plus all covariates</td>
<td>0.522 (0.513-0.531)</td>
</tr>
<tr>
<td>Survival models using propensity-based matching cohort</td>
<td></td>
</tr>
<tr>
<td>Match within ±0.05 of propensity score and 5 y of age</td>
<td>0.538 (0.518-0.558)</td>
</tr>
<tr>
<td>Match within ±0.10 of propensity score and 5 y of age</td>
<td>0.528 (0.514-0.542)</td>
</tr>
<tr>
<td>Match within ±0.15 of propensity score and 5 y of age</td>
<td>0.511 (0.499-0.523)</td>
</tr>
</tbody>
</table>

**Figure Legend:**

Abbreviations: AMI, acute myocardial infarction; CI, confidence interval.
*Simple propensity score included all 65 patient, hospital, and ZIP code characteristics.
†Complex propensity score included all patient, hospital, and ZIP code characteristics and all interactions of age, sex, and race with the other characteristics (750 variables).
From: *Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive Cardiac Management on AMI Survival Using Propensity Score and Instrumental Variable Methods*


---

**Table 5.** Adjusted Mortality Differences Associated With Cardiac Catheterization Among Patients With AMI Using Linear Regression and Instrumental Variable Methods

<table>
<thead>
<tr>
<th>Risk-Adjustment Method</th>
<th>Absolute Mortality Difference (Δ) (SE)</th>
<th>Adjusted Relative Mortality Rate (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1-Year mortality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unadjusted</td>
<td>-0.244 (0.002)</td>
<td>0.37 (0.35-0.38)</td>
</tr>
<tr>
<td>Multiple linear regression†</td>
<td>-0.162 (0.002)</td>
<td>0.58 (0.57-0.59)</td>
</tr>
<tr>
<td>Instrumental variable, adjusted‡</td>
<td>-0.054 (0.015)</td>
<td>0.86 (0.78-0.94)</td>
</tr>
<tr>
<td><strong>4-Year mortality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unadjusted</td>
<td>-0.339 (0.003)</td>
<td>0.45 (0.44-0.46)</td>
</tr>
<tr>
<td>Multiple linear regression†</td>
<td>-0.207 (0.003)</td>
<td>0.67 (0.66-0.68)</td>
</tr>
<tr>
<td>Instrumental variable, adjusted‡</td>
<td>-0.097 (0.016)</td>
<td>0.84 (0.79-0.90)</td>
</tr>
</tbody>
</table>

Abbreviations: AMI, acute myocardial infarction; CI, confidence interval.

*Adjusted relative mortality rate is approximately $1 + \frac{\Delta}{m_{\text{noCATH}}}$, where $\Delta$ is the adjusted absolute mortality difference between patients with and without cardiac catheterization, and $m_{\text{noCATH}}$ is the Kaplan-Meier mortality rate among those patients without cardiac catheterization.

†Linear regression of mortality (binary variable) against all 65 observed patient, hospital, and ZIP code characteristics.

‡Instrumental variable analysis using mortality (binary variable) as the dependent variable and instrumental variable as regional cardiac catheterization rate for the 566 coronary angiography service areas, adjusted for all 65 observed patient, hospital, and ZIP code characteristics.

---

**Figure L**
Sample Selection Models – A third way to control for selection

The problem:
- \( Y = bT + e \)
- \( T = cY + w \), so endogeneity and therefore bias in estimated \( b \)

Solution
- Model entry into treatment with whatever variables you have other than \( Y \) and including at least one variable not included in predicting \( Y \) (this looks like an instrument in IV models)
  - \( T = dZ + u \)
  - \( u \) and \( e \) will be correlated, and the bias in \( b \) turns out to be a function of the correlation of \( u \) and \( e \)
  - By adding this correction factor into the prediction equation for \( Y \), one gets an unbiased estimate \( b' \) of the association of \( Y \) and \( T \)
Sample selection models

- Precise method of controlling for selection based on unobservables depends on whether Y is continuous or dichotomous
  - If continuous, use Heckman sample selection model
  - If dichotomous, use bivariate probit
Example of sample selection

THE APPLICATION OF SAMPLE SELECTION MODELS TO OUTCOMES RESEARCH: THE CASE OF EVALUATING THE EFFECTS OF ANTIDEPRESSANT THERAPY ON RESOURCE UTILIZATION

WILLIAM H. CROWN¹*, ROBERT L. OBENCHAIN², LUELLA ENGLEHART³, TAMRA LAIR⁴, DON P. BUESCHING² AND THOMAS CROGHAN².⁵
Abstract
Non-randomized studies of treatment effects have come under criticism because of their failure to control for potential biases introduced by unobserved variables correlated with treatment selection and outcomes. This paper describes the basic concepts of sample selection models—a technique used widely in the economics evaluation literature for nearly two decades—and discusses the potential role of these models in outcomes research. In addition, it presents a case study of the application of the sample selection modelling approach to evaluation of the effects of antidepressant therapies on medical expenditures for physician services. This case study presents empirical comparisons of alternative model specifications and discusses practical issues in evaluation of sample selection models. We demonstrate that, in this particular case, sample selection models yield very different conclusions regarding treatment effects than traditional ordinary least squares regression.
<table>
<thead>
<tr>
<th>First-stage variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td>Age</td>
</tr>
<tr>
<td>Prior psychiatric hospitalization</td>
<td>Included to model psychiatric medical history and severity of illness, both likely factors in medication prescribing decisions</td>
</tr>
<tr>
<td>Prior non-psychiatric hospitalization</td>
<td>Included to model general medical history and severity of comorbid illnesses, both likely factors in medication prescribing decisions</td>
</tr>
<tr>
<td>Type of depression diagnosis</td>
<td>Includes six possible depression diagnoses corresponding to ICD-9-CM codes: (1) Major depression, single episode; (2) Major depression, recurrent; (3) Neurotic depression; (4) Brief depressive reaction; (5) Prolonged depressive reaction and (6) Depression, not elsewhere classified</td>
</tr>
<tr>
<td>Diagnosis for a contraindicated condition</td>
<td>Includes ICD-9-CM codes associated with all listed contraindications associated with each study drug as stated in labelling</td>
</tr>
<tr>
<td>Pre-study major diagnostic indicators</td>
<td>Major ICD-9-CM code in pre-study period</td>
</tr>
<tr>
<td>Geographic region</td>
<td>Included to account for area variation in practice patterns and pricing</td>
</tr>
<tr>
<td>Provider type</td>
<td>Included to account for differences in prescribing patterns across provider types (for example, acute care hospitals, mental health facilities, family practitioners)</td>
</tr>
<tr>
<td>Six month time interval including month and year</td>
<td>Included to account for publicity and marketing strategies that vary over time and could affect prescribing decisions</td>
</tr>
<tr>
<td>Dependent variable*</td>
<td>Predictor variables</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Total health care costs during the study period</td>
<td>Mills ratio from stage-one models</td>
</tr>
<tr>
<td>Total health care costs during the study period for:</td>
<td>Age</td>
</tr>
<tr>
<td>(a) Physician visits</td>
<td>Gender</td>
</tr>
<tr>
<td>(b) Prescriptions</td>
<td>Any observed psychiatric hospitalization in the study period</td>
</tr>
<tr>
<td>Number of health care services during the study period for:</td>
<td>Any observed non-psychiatric hospitalization in the study period</td>
</tr>
<tr>
<td>(a) Physician visits</td>
<td>Type of depression</td>
</tr>
<tr>
<td>(b) Prescriptions</td>
<td>Geographic indicator</td>
</tr>
<tr>
<td>Total mental health care costs during the study period</td>
<td>Drug sub-cohort membership</td>
</tr>
<tr>
<td>Total mental healthcare costs during the study period for:</td>
<td>Study period major diagnostic category (MDC)</td>
</tr>
<tr>
<td>(a) Physician visits</td>
<td></td>
</tr>
<tr>
<td>(b) Prescriptions</td>
<td></td>
</tr>
<tr>
<td>Number of mental health care services during the study period for:</td>
<td></td>
</tr>
<tr>
<td>(a) Physician visits</td>
<td></td>
</tr>
<tr>
<td>(b) Prescriptions</td>
<td></td>
</tr>
</tbody>
</table>

* Other dependent variables modelled in analyses but not reported on in this paper included psychiatric and non-psychiatric hospitalization costs and ancillary outpatient services.
<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Sample selection</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter estimate</td>
<td>Parameter estimate</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.33</td>
<td>4.53</td>
</tr>
<tr>
<td>Patient age</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Gender</td>
<td>0.00</td>
<td>−0.33</td>
</tr>
<tr>
<td>Depression diagnostic indicators:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major depressive disorder, 1 episode</td>
<td>0.35</td>
<td>0.36</td>
</tr>
<tr>
<td>Major depressive disorder, &gt; 1 episode</td>
<td>0.70</td>
<td>0.67</td>
</tr>
<tr>
<td>Neurotic depression</td>
<td>0.25</td>
<td>0.31</td>
</tr>
<tr>
<td>Brief depressive reaction</td>
<td>0.49</td>
<td>0.39</td>
</tr>
<tr>
<td>Prolonged depressive reaction</td>
<td>0.66</td>
<td>0.66</td>
</tr>
</tbody>
</table>
Risk adjustment

- In CER, interest in not only average effects across typical patient but variations in effects by patient characteristics
  - If can randomize, then this is additional analysis
  - If cannot randomize, part of controlling for differences in groups receiving different treatments
    - Part of modeling selection into treatment but also variations in response
    - If selection cannot be easily modeled, risk adjustment is minimum
- Risk adjustment model should vary for outcome being studied
- Risk adjustment model also limited by available variables that can be included
  - Less rich for claims data than clinical data
- Wide range of standard risk adjustment models available
Saw example of risk adjustment model in cardiac cath study

From: **Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive Cardiac Management on AMI Survival Using Propensity Score and Instrumental Variable Methods**


---

**Table 1.** Select Baseline Characteristics According to Receipt of Cardiac Catheterization

<table>
<thead>
<tr>
<th>Overall Cohort</th>
<th>Propensity-Based Matched Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Received Catheterization Within 30 Days</td>
<td>Unmatched Patients Receiving Cardiac Catheterization (n = 42,045)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>No (n = 48,866)</th>
<th>Yes (n = 73,238)</th>
<th>Standardized Difference</th>
<th>No (n = 31,193)</th>
<th>Yes (n = 31,193)</th>
<th>Standardized Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted 1-year mortality (AMI severity, mean [SD])†</td>
<td>32.3 (18.3)</td>
<td>20.9 (13.3)</td>
<td>73.7</td>
<td>26.8 (15.5)</td>
<td>27.8 (12.5)</td>
<td>6.3</td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>40.2</td>
<td>64.4</td>
<td>49.9</td>
<td>45.2</td>
<td>45.3</td>
<td>0.1</td>
</tr>
<tr>
<td>75-84</td>
<td>50.8</td>
<td>35.6</td>
<td>49.9</td>
<td>54.8</td>
<td>54.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Men</td>
<td>49.7</td>
<td>58.4</td>
<td>17.8</td>
<td>53.2</td>
<td>49.6</td>
<td>7.2</td>
</tr>
<tr>
<td>Black</td>
<td>7.5</td>
<td>4.8</td>
<td>11.3</td>
<td>5.7</td>
<td>6.6</td>
<td>3.7</td>
</tr>
<tr>
<td>Social Security income =$2600</td>
<td>30.0</td>
<td>29.7</td>
<td>0.9</td>
<td>30.2</td>
<td>30.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History of angina</td>
<td>44.1</td>
<td>49.9</td>
<td>11.8</td>
<td>46.0</td>
<td>45.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Previous myocardial infarction</td>
<td>32.9</td>
<td>26.4</td>
<td>14.3</td>
<td>28.7</td>
<td>31.9</td>
<td>6.8</td>
</tr>
<tr>
<td>Previous revascularization</td>
<td>17.8</td>
<td>20.9</td>
<td>7.7</td>
<td>18.0</td>
<td>20.2</td>
<td>5.7</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>27.2</td>
<td>10.4</td>
<td>45.7</td>
<td>16.6</td>
<td>15.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>38.6</td>
<td>38.6</td>
<td>17.1</td>
<td>31.8</td>
<td>34.1</td>
<td>4.9</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>12.8</td>
<td>9.1</td>
<td>12.0</td>
<td>10.6</td>
<td>11.5</td>
<td>2.8</td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>24.9</td>
<td>17.6</td>
<td>18.3</td>
<td>20.9</td>
<td>23.3</td>
<td>5.9</td>
</tr>
<tr>
<td>Smoker†</td>
<td>15.1</td>
<td>18.0</td>
<td>5.0</td>
<td>16.5</td>
<td>17.0</td>
<td>1.2</td>
</tr>
<tr>
<td>AMI clinical presentation characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-ST-segment elevation AMI</td>
<td>41.8</td>
<td>39.9</td>
<td>5.9</td>
<td>39.8</td>
<td>40.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Shock</td>
<td>1.9</td>
<td>1.5</td>
<td>3.0</td>
<td>1.8</td>
<td>2.3</td>
<td>3.4</td>
</tr>
<tr>
<td>Hypotension</td>
<td>3.5</td>
<td>2.3</td>
<td>7.4</td>
<td>3.1</td>
<td>3.6</td>
<td>2.8</td>
</tr>
<tr>
<td>Received CPR</td>
<td>1.8</td>
<td>1.6</td>
<td>1.6</td>
<td>2.3</td>
<td>3.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Peak creatinine kinase &gt;1000 U/L</td>
<td>29.1</td>
<td>32.4</td>
<td>7.2</td>
<td>31.7</td>
<td>31.8</td>
<td>0.2</td>
</tr>
<tr>
<td>Hospital characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual AMI volume &gt;200 patients</td>
<td>20.1</td>
<td>30.4</td>
<td>23.6</td>
<td>22.9</td>
<td>20.5</td>
<td>5.6</td>
</tr>
<tr>
<td>Mortality‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Died within 1 y</td>
<td>38.6</td>
<td>14.2</td>
<td>24.6</td>
<td>34.6</td>
<td>19.0</td>
<td>10.6</td>
</tr>
<tr>
<td>Died within 4 y</td>
<td>62.0</td>
<td>27.8</td>
<td>35.4</td>
<td>55.4</td>
<td>36.3</td>
<td>21.4</td>
</tr>
</tbody>
</table>

**Figure Legend:**

Saw example of risk adjustment model in cardiac cath study.
Saw example of risk adjustment model in cardiac cath study

- **Note inclusion of SES variables**
  - Race/ethnicity
  - Gender
  - Social security income

- **For some purposes, not considered appropriate**
  - E.g., National Quality Forum allows stratification by SES variables in analysis of quality measures but not risk adjusting

- **Also not availability of clinical data beyond claims information**
Risk adjustment models should vary by outcome being studied

- **Cardiac bypass mortality rate variations might be influenced by**
  - Chronic conditions (e.g., lung or kidney disease)
  - Clinical indicators
- **Length of stay needs a different model**
  - Diagnosis
  - Comorbidities (but different list than cardiac surgery mortality)
  - SES or neighborhood factors?
- **Need to consider factors expect to influence outcome and then test empirically**
  - Typically asking what proportion of variance is being explained by including specific variables
  - Model interactions and non-linear transformations
  - R-squares of risk adjusters can be low to moderate
Variables to include in risk adjustment limited by data

- **Claims data**
  - No clinical indicators
  - If only have record for current visit/admission, limited to data on comorbidities and complications for that visit/admission
  - If have linked data set, can look backward to enrich data

- **Sometimes use proxies when data incomplete**
  - Zip code average income, education and ethnicity when SES data not available
  - Statistical name matching to estimate ethnicity if name but not ethnicity in data
Many commercial risk adjusters available, sometimes for a fee, but they have extensive validation

- **Examples**
  - APR-DRG, by 3M
  - Diagnostic Cost Groups, developed at Boston University, which has evolved into Hierarchical Coexisting Conditions model (HCC)
  - Ambulatory Care Groups and Ambulatory Diagnosis Groups, developed at Johns Hopkins University
  - AHRQ Patient Safety Indicator risk adjustment model
  - Ingenix Symmetry Episode Risk Groups

- **Vary by focus (inpatient, outpatient, both), type of data required, algorithm and weighting of individual factors**
Outline of Session

- Why use observational data
- Data sources
  - Clinical records
  - Administrative claims data
  - Registries
  - Survey data
- Extracting, merging and coding observational data
- Design of studies for using observational data
  - Addressing selection in design and sampling
  - Addressing selection using statistical adjustment
    - Propensity scoring
    - Instrumental variables
    - Sample selection models