INTRODUCTION TO PHARMACOEPIDEMIOLOGY

9 September 2012
University of South Australia, Adelaide

M. Alan Brookhart, PhD
Associate Professor of Epidemiology
UNC Gillings School of Global Public Health

Adjunct Associate Professor of Medicine
UNC School of Medicine
University of North Carolina

With the growing availability of large healthcare databases, non-experimental studies of prescription medications are becoming increasingly common. However, appropriate design and analysis of such studies can be challenging. In this workshop we provide an intensive introduction to the field of pharmacoepidemiology. We review the data used in pharmacoepidemiology and the central threats to validity of such studies, including the healthy user bias, immortal person time bias, and various types of confounding bias. We then discuss approaches to mitigate these biases through design and analysis. The workshop will cover the comparative new user design, self-controlled designs, propensity score methods, and instrumental variable approaches. We will also discuss some additional topics in the field, including studies of medication adherence, disparities, and active safety surveillance of medical products.

Timetable

<table>
<thead>
<tr>
<th>Time</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:15-8:50</td>
<td>Registration</td>
</tr>
<tr>
<td>8:50-9:00</td>
<td>Welcome and introduction</td>
</tr>
<tr>
<td>9:00-9:30</td>
<td>A brief introduction to pharmacoepidemiology</td>
</tr>
<tr>
<td>9:30-10:40</td>
<td>Confounding and other biases in non-experimental studies</td>
</tr>
<tr>
<td>10:40-11:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:00-12:15</td>
<td>Propensity scores</td>
</tr>
<tr>
<td>12:15-1:15</td>
<td>Lunch</td>
</tr>
<tr>
<td>1:15-2:30</td>
<td>Instrumental variable methods and natural experiments</td>
</tr>
<tr>
<td>2:30-2:50</td>
<td>Coffee break</td>
</tr>
<tr>
<td>2:50-4:00</td>
<td>Studies of prescribing and adherence, and general discussion</td>
</tr>
</tbody>
</table>
Dr M. Alan Brookhart is an Associate Professor of Epidemiology and Medicine at the University of North Carolina at Chapel Hill. He completed a PhD in Biostatistics at the University of California, Berkeley, and held postdoctoral appointments at the Harvard Medical School and Brigham and Women's Hospital, Boston, before taking up his position at the University of North Carolina. His research is focused primarily on the development and application of new statistical methods and study designs for epidemiologic studies of medications using large clinical and healthcare utilization databases. In this area, he has made contributions to the development of quasi-experimental and instrumental variable approaches that can be used to estimate causal effects in the presence of unmeasured or poorly recorded confounding variables. He has also been involved with the development of propensity score and marginal structural model methodology and has also developed new epidemiologic approaches for studying medication adherence and use of healthcare services. Substantively, he is interested in the effects of medications in the elderly and patients with end-stage renal disease.

Date: Sunday 9 September 2012 9.00am – 4.00pm

Venue: Room C3-16, University of South Australia, City East Campus, Corner of North Terrace and Frome Road, Adelaide
Introduction to Pharmacoepidemiology

M. Alan Brookhart, Ph.D.
Department of Epidemiology,
UNC Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Learning Objectives

• To understand the scope of the field of pharmacoepidemiology
• To understand why we need observational/non-experimental studies of drugs
• To understand commonly used sources of data for pharmacoepidemiology

Pharmacoepidemiology

• Study of the use of and the effects of drugs in large numbers of people

Strom, Kimmel: Textbook of Pharmacoepidemiology 2006
Examples of Issues Addressed within Pharmacoepidemiology

• Drug utilization research/ quality of care
• Drug effects (effectiveness and safety)
• Analytic methods

Why do we need observational studies of drugs or medical products?

• Clinical trials provide gold standard evidence of drug effects
• Problems with clinical trials
  – Expensive
  – Small
  – Often drugs are compared against placebo
  – Exclude elderly, children, pregnant women, patients with important comorbidities
  – May be unethical
  – Not timely

MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20,536 high-risk individuals: a randomised placebo-controlled trial

• Enrolled patients 40-80 with some CV risk factors or diabetes
• Excluded patients with kidney disease, liver disease, life threatening condition (other than diabetes) such as COPD, cancer (other than non-malignant skin cancer)
• Excluded patient who might have a problem with compliance (psychiatric disorders, cognitive impairment, dementia, disabling stroke, etc)
• Less than 20% of patients were over 70
Enrolled patients 70-82 with some vascular risk factors
Excluded patients with cognitive impairment

Example of Need for Non-experimental CER: Antipsychotic Medications (APM) in the Elderly

- APMs approved to treat schizophrenia
- Widely used off-label to treat elderly patients with dementia
- Two broad classes: conventional (older drugs) versus atypical (newer drugs)
- Manufacturers of some of the atypicals conducted trials to assess effectiveness of the medications for controlling behavioral disturbances in elderly
- FDA meta-analysis: increased risk of mortality associated with atypical APMs (relative to placebo)
- FDA put a "black box" advisory on label of atypical APMs

Clinical Dilemma

- Should physicians switch patients to the first generation APMs?
- Older APMs have many known side effects, poor safety profile
- Head-to-head trial will never be not be done
  - Practically difficult
  - Ethically impossible
- Question must be answered by analyzing existing data
Increasing interest in “Comparative Effectiveness Research” in US

“Conduct and synthesis of research comparing the benefits and harms of different interventions and strategies to prevent, diagnose, treat and monitor health conditions in “real world” settings. The purpose of this research is to improve health outcomes by developing and disseminating evidence-based information to patients, clinicians, and other decision-makers, responding to their expressed needs, about which interventions are most effective for which patients under specific circumstances.”

--Report to President and Congress, Federal Coordinating Council For CER

Most trial are placebo-controlled, rather than comparative

• JUPITER trial randomized 17,800 people with elevated high-sensitivity C-reactive protein, but normal lipids
• Patients assigned to receive placebo or high-potency rosuvastatin therapy

JUPITER results

Would other less expensive statins provide a similar benefit in this population?

Ridker et al, Rosuvastatin to prevent vascular events in men and women with elevated C-reactive Protein. NEJM 2008
Randomized Comparative Studies

Major Outcomes in High-Risk Hypertensive Patients Randomized to Angiotensin-Converting Enzyme Inhibitor or Calcium Channel Blocker vs Diuretic

The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT)

- 33,357 patient randomized to one of three antihypertensives: ACEIs, Thiazides, CCBs
- Patients had hypertension and at least one CV risk factor
- Followed between 3-8 years
- Outcome: Blood pressure and major CVD events

- Thiazide diuretics as good as or superior to ACE Inhibitors and CCBs for all outcomes
- Established guideline for management of hypertension that are still used
- ALLHAT took 8 years to complete and cost $130 million

Why do we need observational studies of drugs or medical products?

- Problems with clinical trials
  - Expensive
  - Small
  - Often drugs are compared against placebo
  - Exclude elderly, children, pregnant women, patients with important comorbidities
  - May be unethical
  - Not timely
- > we need observational studies of medications
- 85% of CER is nonexperimental
What data can we use for non-experimental studies?

- Large cohort studies
  – Usually prospective or ongoing
- Healthcare and clinical database
- Disease registries
  – Cancer (SEER)
- Drug registries
  – E.g., antiretrovirals, biologics

Desired Qualities of a Database

- Representative
- Large
- Timely (i.e., up to date)
- Continuity
  – Individual observations
  – Calendar time
- Linkage on unique identifier
- Accessible
  – Without delay
  – Over prolonged periods (intimate knowledge of data)
  – To everyone

Desired Contents of Database

- All use of prescription drugs and over-the-counter (OTC) drugs
- Outpatient, inpatient, emergency care and reasons for visit
- Patient health-related behaviors
  – Smoking
  – Diet and exercise
- Indication for treatment
  – Clinical variables
  – Diagnoses
  – Laboratory
  – Radiographic
  – Function (RR, ejection fraction)
- Other determinants of treatment
  – Prescriber
  – SES
  – Frailty
- Cause-specific mortality
- Patient reported outcomes (QOL)
Healthcare and clinical databases
- Large N (often >> 1,000,000)
- Often population based
- No recall/interviewer bias
- Timely results
  - Regulatory
  - Commercial
  - Public Health
- Growing use to assess
  - Unintended and intended drug effects

Insurance Claims Databases
- Billing data from payors
- Closely audited
- Dispensed (filled) prescriptions
  - Best data on drug exposure in PE
- Diagnostic data potentially dependent on financial incentives (system/country specific!)
  - Inpatient DRGs
  - Outpatient procedures
- Age, sex
- Often race, income, mortality
- US e.g., MarketScan, IMS, i-3, Medicaid, Medicare

Examples of Claims Databases in US
- Medicaid
  - ~ 50 million lives
  - Low income pregnant women and families
  - Chronic disabilities (e.g., ESRF)
  - Low-income seniors
- Medicare
  - All 65+
  - Part D (drug insurance)
    - Since 1/1/2006
    - ~ 1/3 FPS (individual dispensed prescriptions)
    - Available to academic centers for research (UNC)
    - Pharmacy assistance programs
Limitations of Healthcare Databases

- Uncertain validity of diagnostic data
- Lack of data on confounders, but
  - Depending on specific hypothesis
  - Validation studies (external control)
  - Sensitivity analyses
- No OTC drugs
  - NSAIDs including aspirin
  - PPIs
  - Others (e.g., orlistat)
- US: High turnover of population < 65
- Formularies, deductibles
- Missing dispensing prescription drugs

Other Things to Worry About

- Outcome not reliably coming to medical attention
  - E.g., diabetes (vs. MI, stroke)
- Lethal outcomes (e.g., MI, suicide, injury)
- Immeasurable drug exposures
  - Inpatient
  - Nursing home
- Strong confounding
  - Association with exposure
  - Association with outcome
  - Prevalence
- Large OTC proportion
- Poorly defined outcomes

Electronic medical record databases

- Advantages
  - High validity of diagnostic data
  - Some information on lifestyle
  - Some test results (e.g., laboratory, RR)
- Disadvantages
  - Uncertain completeness of diagnostic data
    (out of system, hospital, specialist)
  - Prescribed drugs (not: filled – one step removed from taking)
  - Drug lists vs. e-prescribing
  - Various coding systems (including: none!)
EMR Databases US

- Group Health Cooperative (Washington)
  - ~ 500k lives
  - Health Maintenance Organization (HMO)
  - Pharmacy benefits management (PBM)
- Kaiser Permanente
  - ~ 8.2 million lives
- HMO Research Network
  - ~ 1 million lives (?)
- Regenstrief

Healthcare Databases from Outside US

- Canada
  - Saskatchewan
    - ~ 1 million lives (whole province)
    - Quebec
      - RAMQ (approx. 45% of adult population)
- Netherlands
  - PHARMO
    - ~ 500k lives covered
  - Rotterdam Study
    - Cohort with linked pharmacy records
- UK
  - GPRD
    - THIN
    - ~ 3 million lives covered
- Scotland
  - Tayside medicines monitoring unit (MEMO)
    - ~ 500k lives covered
- Scandinavia (Denmark, Sweden, Norway)
  - Whole population
  - Several millions

Disease, Device, and Drug Registries

- Systems that collect data on patients with diagnosed with a disease, who have received a certain procedure, medical device, or medication
- Sometimes these are simply include baseline data collected at the time of enrollment
- Sometimes these include detailed follow-up information, outcomes
<table>
<thead>
<tr>
<th><strong>Westphalian Stroke Registry</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Regional data bank in northwestern Germany</td>
<td></td>
</tr>
<tr>
<td>• All patients treated for stroke symptoms who were admitted to the participating 42 hospitals.</td>
<td></td>
</tr>
<tr>
<td>• Collected variables include sociodemographic characteristics, cerebrovascular risk factors, comorbidities, stroke type, and diagnostic data</td>
<td></td>
</tr>
<tr>
<td>• Treatment information</td>
<td></td>
</tr>
<tr>
<td>• Complications and discharge status</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>SEER Cancer Registry in US</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• SEER=Surveillance, Epidemiology, and End Results</td>
<td></td>
</tr>
<tr>
<td>• Collecting data since 1973 from regions covering about 28% of US</td>
<td></td>
</tr>
<tr>
<td>• Collects data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, and first course of treatment</td>
<td></td>
</tr>
<tr>
<td>• No follow up other than date of death obtained from vital statistics</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Many Other Examples</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Many countries have registries to track patients with artificial joints</td>
<td></td>
</tr>
<tr>
<td>• Many other device registries</td>
<td></td>
</tr>
<tr>
<td>• CABG and stent registries</td>
<td></td>
</tr>
<tr>
<td>• Transplant receipt registries</td>
<td></td>
</tr>
<tr>
<td>• Many drug registries in US are required as part of post-marketing surveillance</td>
<td></td>
</tr>
</tbody>
</table>
**Registry Strengths**

- Usually contain rich, clinically relevant baseline data
- Sometimes contain detailed clinical follow-up data

**Registry Limitations**

- Sometime these are simply include baseline data collected at the time of enrollment
- Follow-up data are often coarse, do not contain good information on treatment changes
- Drug device registries often lack a control group
- Available only on a segment of the population
- Often small

**Future Directions**

- Database linkage
  - Add claims data to cohort studies
  - Easy to get informed consent
  - E.g., ARIC, WHI, Rotterdam
- Internal validation studies
  - Add additional information for subgroup
  - E.g., Medicare Current Beneficiary Survey (MCBS)
- Add disease registries to EMR data
  - E.g., cancer registry
- Add PROs (collect during office visit)
-- notes --
Confounding and Other Source of Bias
The New User Design

M. Alan Brookhart, Ph.D.
Department of Epidemiology,
UNC Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Learning Objectives

• To understand how confounding bias arises in studies of therapeutics
• To understand the characteristics of the new user design and how they mitigate many forms of confounding bias
• To recognize immortal and unexposable person time bias and know how to avoid these problems

Confounding

Confounder

Medication Exposure

Outcome of Interest
Confounding During Treatment Initiation

- Confounder
- Medication Initiation → Outcome of Interest

Confounding by Indication / Disease Severity

- Disease Severity (clinical need)
- Initiation of Preventive Therapy → Outcome of Interest

Case Study: Statins and Primary Prevention of Myocardial Infarction

- Statins are safe and widely used cholesterol lowering agents
- Prescribed to patients at risk of CAD or with existing CAD
- Study among Medicare/PACE enrollees in PA, 1995-2002
  - All hospitalizations discharge data and physician office data (ICD-9 coded diagnoses and procedure codes)
  - Merged with pharmacy claims
- Identified 38,046 new users of statins (w/ no hx of MI)
- Matched these by calendar time 1-1 to non-users of statins (w/ no hx of MI)
- Outcome was time until hospitalization for acute MI (within one year)
Unadjusted Results

- 805 events in “control” arm
- 1123 events statin arm
- Unadjusted hazard ratio = 1.36
- Do statins increase the one-year risk of MI by 36%?

Multivariable Cox PH Model

- Next we adjusted for age, sex, and 30+ covariates abstracted from the claims data: history of co-morbid conditions, history of medication use, Charlson index, etc.
- Result: Hazard Ratio = 1.21 (95% CI 1.09-1.36)
- Clearly, residual confounding not controlled.

SSRI Antidepressants and Suicide

- Fluoxetine (Prozac) the first SSRI-type anti-depressant (AD)
- Released in the US in 1988 and marketed as being safer and more effective than older ADs
- There were reports of suicide and violent behavior among patients recently started on Prozac (from older ADs) (Teicher MH, Glod C, Cole JO. 1990 Am J Psychiatry)
- Newly initiated patients were likely those that had failed on an older treatment
- Confounding by disease severity
Protopathic Bias

• Closely related to CBI
• An early, undiagnosed form of disease leads to a treatment of early conditions
• Disease is subsequently recognized
• Exposure appears to cause disease

Confounding by The Healthy User Effect

Healthy Behaviors Often Associated with Benefits not Substantiated in RCTs

• Hormone Replacement Therapy
  – Observational Result: HRT associated with a 30% reduced risk of AMI
  – RCTs: HRT associated with a increased risk of MI, stroke, and breast cancer.

• Vitamin E in women
  – Observational research: 30%-40% decrease in risk of cardiovascular outcomes attributable to Vitamin E use
  – RCT: No benefit. (Lee et al, JAMA 2005)

• Many other examples
...women who use estrogen replacement therapy had a better cardiovascular risk profile than those who did not...

Confounding by Frailty / Serious Comorbidity

Frailty / Serious Illness

Physician Unlikely to Prescribe Preventive Meds

Initiation of Preventive Therapy

Outcome of Interest

Paradoxical Relations of Drug Treatment with Mortality in Older Persons

Robert J. Glynn, Eric L. Knight, Reesa Leek, and Jerry Avorn
Confounding by Functional / Cognitive Status

- Functional Cognitive Impairment
  - Patients may not be able to easily visit physician, pharmacy
  - Initiation of Preventive Therapy → Outcome of Interest

Functional status is a confounder of the association of influenza vaccine and risk of all cause mortality in seniors

- Influenza vaccine found to be associated with decreased mortality risk during the non-flu season
- Statistical adjustment for functional status attenuated this relation

Confounding During Treatment Changes

- Risk Factor
  - Medication Discontinuation/Change → Outcome of Interest
Informative Treatment Changes: The Sick Stopper Effect

Frailty / Nearness to Death

Discontinuation of Preventive Therapy

Outcome of Interest

Informative Treatment Changes: The Healthy Adherer Effect

“Compliance Bias”

Patient’s Concern About Health

Other Healthy Behaviors

Adherence to a Preventive Therapy

Outcome of Interest

Adherence to Placebo and Mortality Risk (Simpson, BMJ 2006)
• 145,000 new users of statins in British Columbia
• Examined association between statin adherence and both accidents and various clinical outcomes unlikely to affected by a statin

### Associations Between Adherence to Statin Treatment and Health-Related Events

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Medication Intolerance</td>
<td>1000</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>Medication Change</td>
<td>5000</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>Medication Change Outcome of Interest</td>
<td>10000</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
</tr>
</tbody>
</table>

### Informative Treatment Changes:

**Medication Intolerance / Treatment Failure**

- Adverse Effect
- Lack of Efficacy

![Diagram](attachment://diagram.png)

- Medication Change
- Outcome of Interest
Summary

- Relative to non-users, prevalent users are more likely...
  - to have an indication for treatment
  - to follow a healthy lifestyle
  - to be cognitively and functionally intact
  - to not have other, serious comorbidities
  - to tolerate the medication and derive benefit from it

New User Design

- New User Design proposed by Ray et. 2003
- Compare new users of a medication of interest to new users of a comparator drug/no treatment
- Requires no use of either therapeutic or comparator drug
- Pairs naturally with propensity score methods to control confounding by baseline factors

New User Design Mimics A RCT
New User Design

- Identify all people initiating treatment in a defined population (people and time)
- Define minimum period without drug exposure (wash-out) prior to $t_0$
  - Make sure you would see drug (in system)!
- Include everyone meeting these criteria
- Start follow-up as of this time $t_0$
- Define all covariates up to $t_0$
  - You may want to include $t_0$
  - Use same length interval for covariate definition for everyone (e.g., wash-out)

Permits Study of Early Events

- Period after initiation often associated with increased risk (Guess 89)
  - Benzodiazepines and falls
  - NSAID and peptic ulcer
  - ACE-inhibitors and angioedema
- Depletion of susceptibles
- Physiologic adaptation
- Selection (adherence) bias = healthy user
**New User vs. First Time User**

- First ever exposure would be ideal
- Possible with drugs new on the market
- Rarely ever possible with older drugs
- Wash-out period
  - Usually plausible
  - Not for serious acute events (anaphylaxis)
- Same problem as in RCT
- Make sure you mention that new users may not be first time users (drug naïve)

**New Users Design Separates Confounders from Intermediates**

- Confounders influence treatment choice
- Intermediates are affected by treatment and subsequently affect outcome. No way of separating these in prevalent users cohort
- Example:
  - Statins and LDL
  - Antihypertensives and blood pressure

**Disease Risk Factors**

- New user design
  - Everything up to \( t_0 \) is a potential confounder
- Control for measured confounders
- Even more obvious with propensity scores
  - What affects treatment choice?
  - What risk factors affect treatment choice?
- Everything after \( t_0 \) is a different animal
  - Ignore
  - Use other methods, e.g., MSM
Follow-up

- Obvious timescale ($t_0$)
- Reduce healthy adherer (sick stopper) bias by using comparator drug if possible
- Decide on censoring for stopping/switching
  - Last prescription + days supply + grace period
  - No censoring
    - First exposure carried forward
    - Intention-to-treat
- Stratify by time on drug to detect time-varying hazard ratios

New User Design with Active Comparator

- Can either compare new users of a drug of interest to users of a comparator drug (active comparator)
- Often specified by research question (comparative effectiveness)
- “Is drug A safer or more effective than drug B?”
- Or can be a mechanism to control confounding

Limiting confounding by design: Comparative New User Design

Baseline period:
No past use of medication

Treatment Prescribed

New Users of Drug A
New Users of Drug B

Washout Period

Treatment Randomized

Drug A
Drug B
Strengths of Active Comparator

- Reduce confounding by indication
  - Clinical alternative
  - Similar point in disease progression
  - Problem: step-up therapies (but reality often better than expected, e.g., TNF-α vs. MTX)
- Reduce confounding by frailty
  - Similar medicalization/access

Comparator Drug Examples

- Glargine vs. NPH insulin
- ARB vs. ACE
- TNF-α vs. MTX
- Rosiglitazone vs. Pioglitazone
- Sulfonylureas vs. metformin
- Etc.

Problems: Many drug may not have a logical comparators

- Statins
- Vaccines
- One approach: use a drug with a different indication (e.g., anti-glaucoma drugs comparator for statins)
  - Reduce confounding by frailty, healthy user effect, etc
  - Problem indications are different, may not reduce confounding by indication
- Another approach: use the date of a physician visit
Strengths of the New Users Design

- Both groups are new users and thus similar:
  - Health seeking behavior, cognitive and physical functioning, etc
- Proper choice of a control can minimize confounding by indication
- Can study events that occur immediately after follow-up
- Groups are not enriched patients tolerant of medication
- Temporal separation of covariates and exposure

Limitations of the New Users Design

- Ideal for healthcare databases
  - Exposure and covariate information on day to day basis
- Difficult in cohort studies where exposure not well ascertained
- Limits sample size considerably, but
  - Less bias, wider CI
  - Much better coverage probability!
- Limits ability to assess long term effects
- Gives more weight to short term users

Alternative Design: follow-up begins after an index event

- Typically index date is a sentinel event, e.g., a diagnosis or hospitalization
- Interested in assessing effects of medication in patients who have experienced the event
- Post-MI medication use
  - Index date: discharge from hospital
  - Assess use of statins, ACE Inhibitors, etc
  - Examine effect on outcome
Common Source of Bias in Study Design

• Hypothetical study design
  – Identify post-MI patients
  – Determine whether they start post-MI meds in the thirty days after hospital discharge, classify them as exposed or unexposed
  – Examine survival by treated versus untreated

Immortal Time Bias

• Study design creates time in which an outcome could not occur
• Usually occurs before a subject starts treatment
• Often unintentionally created by restricting on an event that happens during follow-up

Solution to Immortal Time Bias

• Do not select cohort based on events occurring during follow-up
• Or apply selection to everyone
  – Create an exposure ascertainment period that everyone must have
• Have a common index date and make exposure time-varying
<table>
<thead>
<tr>
<th>Immeasurable Time Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Time when exposure cannot occur or be observed</td>
</tr>
<tr>
<td>– Hospitalizations, acute care stays</td>
</tr>
<tr>
<td>• Often leads to exaggerated benefits of treatment</td>
</tr>
</tbody>
</table>
Counterfactuals and Propensity Score Methods

M. Alan Brookhart, Ph.D.
Department of Epidemiology,
UNC Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Learning Objectives

• To understand the concept of a counterfactual and a causal effect
• To understand how propensity scores can be used to estimate causal effects
• To understand a variety of practical issue involved with propensity score methods

Ex: Non-steroidal anti-Inflammatory drugs and peptic ulcer disease risk in routine practice

• Compare risk of GI outcomes in between
  – Non-selective NSAIDs
  – COX-2 selective NSAIDs (“Coxibs”) as they are used in a routine practice setting (the “real world”)
• In trials, coxibs were slightly less likely to cause GI problems
• What is the benefit of Coxibs in a real world patient population?
**Ideal Causal Experiment**

- NS NSAID Treatment → Outcome Under No Treatment, Y(0)
- Coxib Treatment → Outcome Under Treatment, Y(1)
- Go Back in Time

**Counterfactuals**

- Y(1) and Y(0) are “counterfactual” or potential outcomes
- If we knew Y(1) and Y(0) for all patients, we could identify optimal treatment for everyone
- Unfortunately, we only observe one potential outcome – fundamental problem of causal inference
- Causal inference is similar to analysis of censored data
- Denote observed outcome Y, and observed treatment with X

**Causal Parameters/Contrasts**

- Let Y be an indicator of whether a patient experienced the outcomes during follow-up (a zero or one variable)
- Causal risk difference E[Y(1)] - E[Y(0)]
  - Interpretation: risk of outcome if everyone had been treated minus risk of outcome if nobody had been treatment
- Causal risk ratio E[Y(1)] / E[Y(0)]
  - Interpretation: risk of outcome if everyone had been treated divided by the risk of outcome if nobody had been treatment
- These tell us about treatment effects in a population but not individuals
Estimating Distributions of Counterfactuals

- We can estimate distributions of counterfactuals in idealized RCTs (fully blinded, perfect compliance, etc)
- No systematic difference between experimental units across arms of the trial

Y(1), Y(0) are independent of (unrelated to) treatment arm assignment

- The distribution of Y(1) is the same as the distribution of Y among those randomized to receive treatment

Can estimate E[Y(1)] with the mean of Y among those assigned to treatment

Key Problem in Observational Studies

- In observational/non-randomized studies the key assumptions Y(1), Y(0) are independent of (unrelated to) treatment arm assignment do not hold.

- For example, Coxib treatment may be more likely to be assigned to patients at greater risk of GI complications

- We say that treatment is "confounded."

- E[Y(1)] not necessarily equal to E[Y|X=1]

Confounding by Indication

Notation: X=treatment (0,1), C=vector of confounders, and Y=outcome
Causal Inference

- Causal inference is concerned with estimating readily interpretable causal contrasts from observational data
- In other words, estimating parameters that we would (or could) estimate in a randomized controlled trial
- As we will see, sometimes these cannot be easily estimated and we must settle for alternative quantities

Key assumption for causal inference
No unmeasured confounders / exchangeability

$Y(1), Y(0)$ are independent of treatment ($X$) given the confounders ($C$)

$C$ is a set of variables (age, sex, history of GI bleed, etc)

Among people with the same values for the confounders, treatment is effectively randomized.

Estimating Causal Effects by Stratification

Within small subgroups/strata of confounders (patients with a specific set of characteristics, we denote with $C=c$, e.g. Age=72, Gender=female, History of GI bleed=0, etc)

Under no unmeasured confounding, we can estimate within-strata causal effects

$E[Y|X=1, C=c] = E[Y(1)|C=c]$  
$E[Y|X=0, C=c] = E[Y(0)|C=c]$  

We can then average these to get average causal effects, e.g., $E[Y(1) - Y(0)]$
Controlling Confounding with Statistical Models

Propensity Score and IPTW Methods

Confounders (GI risk factors) C

Notation: X=treatment (0,1), C=vector of confounders, and Y=outcome

Key Propensity Score Theory

Propensity score is the probability of receiving treatment given C

\[ PS(C) = P(X=1|C) \]

If all confounders are measured, Rosenbaum and Rubin show

\[ Y(1), Y(0) \text{ are independent of } X \text{ given } PS(C) \]

Among people with the same propensity score, treatment is effectively randomized.

Estimating the Propensity Score

Propensity scores are not know--must be estimated

\[ P(X=1|C)=\text{expit}(b_0+b_1\text{age } + b_2\text{sex } + b_3\text{CHD}+\ldots) \]

For each patient a predicted probability of receiving treatment is computed -- the estimated PS
Methods of Using the PS

- Covariate adjustment (not optimal)
- Stratification on PS
- Matching on the PS
- Weighting on the PS (e.g., IPTW)
- Hybrid approaches: combine matching with multivariable regression (Cochran and Rubin) & doubly robust estimators (Robins)

Stratification on the Propensity Score

- Treatment effects are estimated within strata of PS
- Treatment effects averaged across strata
- This yields an estimate of the average effect of treatment
- Subject to residual bias within strata
Matching on the PS

- Match exposed to unexposed with similar PS
- Subjects who cannot be matched discarded
- Creates good balance of measured covariates
- Greedy matching techniques
  (http://www2.sas.com/proceedings/sugi26/p214-26.pdf)

Matching on the PS, cont.

- Limitation of matching
  - May lose many participants
  - Individuals in the tails of the distribution can be difficult to match
  - Generalizability: The effect of treatment may be different in those participants that cannot be matched.
  - Interpretability—not always a causal parameters

Inverse Probability of Treatment Weighting (IPTW)

- Each subject weighted by the inverse of the probability that they received their observed treatment
- Inverse probability of treatment (IPTW) estimator
  - Fit a standard regression, but weight by
    $1/PS(X)$, in treated patients
    $1/(1 - PS(X))$, in untreated patients
IPTW creates “pseudopopulation” in which treatment is unrelated to covariates

Original population

No association between NSAID use (X) and GI risk in pseudopopulation

Experimental Treatment Assignment Assumption

- Everyone must have a non-zero probability of being treated or not
  \(0 < \text{Pr}(X=1|C) < 1\)
- Even small violations of this assumption can cause bias

IPTW estimates the average effect of treatment in the population

Absolute Scale (e.g., Risk Difference)
\[RD = E[Y(1)] - E[Y(0)]\]

Relative Scale (e.g., Risk Ratio)
\[RR = \frac{E[Y(1)]}{E[Y(0)]}\]

This contrasts with other treatment effects (treatment in the treated)
\[RD_{TT} = E[Y(1)|X=1] - E[Y(0)|X=1]\]
Poorly Defined Populations

- Populations in pharmacoepi are often ill-defined
- If patients with contraindications are treated, may get hugely up-weighted
- Cause IPTW to give peculiar results

Hypothetical Distribution of Propensity Scores

SMR Weight

- Weighting method uses a standardized mortality/morbidity ratio (SMR) weight:
  - Value of 1 in the treated
  - Propensity odds in the untreated, PS(X)/(1-PS(X))
- This weighting approach uses the treated group as the standard
- Yields the effect of “treatment among the treated.”
- E[Y(1)-Y(0)|X=1]
### Hypothetical Distribution of Propensity Scores

![Hypothetical Distribution of Propensity Scores](image)

### Doubly Robust Estimators

- Depends on both an outcome model and propensity score model
- More efficient than IPTW
- Estimate is consistent as long as at least one model is correctly specified!
- Does not depend on the experimental treatment assumption when outcome model is correct
- Emerging methodology: Targeted maximum likelihood
Motivating Example: Observational Study of Non-steroidal Anti-Inflammatory Drugs and GI bleeding risk in an elderly population

- Compare risk of GI outcomes in elderly between
  - Non-selective NSAIDs
  - COX-2 selective NSAIDs
- Coxibs are slightly less likely to cause GI problems
- Coxibs are likely to be selectively prescribed to patients at increased GI risk
- Classic problem of confounding by indication

Data

- Population: Medicare beneficiaries in Pennsylvania eligible for a state run pharmaceutical benefit program (PACE)
  - Low to moderate income elderly
- Cohort of new users of COX-2 inhibitors or non-selective NSAIDs between Jan. 1, 1999 and Jul. 31, 2002
  - Yielded N=49,919
- Drug exposure came from pharmacy claims data, ITT analog
- Outcomes and covariates were derived from Medicare hospital claims data
- Outcome was defined as a hospitalization for peptic ulcer disease or GI bleeding during follow-up (60-days)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coxib</th>
<th>NS NSAID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female Gender</td>
<td>86%</td>
<td>81%</td>
</tr>
<tr>
<td>Age &gt; 75</td>
<td>75%</td>
<td>65%</td>
</tr>
<tr>
<td>Charlson Score &gt; 1</td>
<td>76%</td>
<td>71%</td>
</tr>
<tr>
<td>History of Hospitalization</td>
<td>31%</td>
<td>26%</td>
</tr>
<tr>
<td>History of Warfarin Use</td>
<td>13%</td>
<td>7%</td>
</tr>
<tr>
<td>History of Peptic Ulcer Disease</td>
<td>6%</td>
<td>2%</td>
</tr>
<tr>
<td>History of GI Bleeding</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Concomitant GI drug use</td>
<td>5%</td>
<td>4%</td>
</tr>
<tr>
<td>History GI drug use</td>
<td>27%</td>
<td>20%</td>
</tr>
<tr>
<td>History of Rheumatoid Arthritis</td>
<td>9%</td>
<td>3%</td>
</tr>
<tr>
<td>History of Osteoarthritis</td>
<td>49%</td>
<td>33%</td>
</tr>
</tbody>
</table>
Counterfactuals

Y(0) outcome a patient would experience if given NS NSAIDs
Y(1) outcome a patient would experience if given Coxibs

Treatment Effects on Absolute Scale (e.g., Risk Difference)

\[ RD = E[Y(1)] - E[Y(0)] \]

Treatment Effects on Relative Scale (e.g., Risk Ratio)

\[ RR = \frac{E[Y(1)]}{E[Y(0)]} \]

Example: Analysis

- Estimated PS using logistics regression
- Using 17 a priori selected covariates: GI risk factors and measures of frailty. Also included calendar year.
- PS Model yielded a c-statistic of 0.67
- Matched on estimated PS using a greedy matching algorithm to create a PS matched cohort (N=33,526)

Distribution of PS within Exposure Groups
### Distribution of PS within Exposure Groups (Matched Cohort)

![Graph showing distribution of PS within exposure groups](image)

### Assessing Balance Using Matching

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unmatched (N=49,919)</th>
<th>PS Matched (N=33,526)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coxib Users</td>
<td>32,273</td>
<td>16,763</td>
</tr>
<tr>
<td>NSAID Users</td>
<td>17,646</td>
<td>16,763</td>
</tr>
<tr>
<td>Female Gender</td>
<td>86%</td>
<td>82%</td>
</tr>
<tr>
<td>Age &gt; 75</td>
<td>75%</td>
<td>68%</td>
</tr>
<tr>
<td>Charlson Score&gt;1</td>
<td>76%</td>
<td>71%</td>
</tr>
<tr>
<td>History of Hospitalization</td>
<td>31%</td>
<td>26%</td>
</tr>
<tr>
<td>History of Warfarin Use</td>
<td>13%</td>
<td>7%</td>
</tr>
<tr>
<td>History of Peptic Ulcer Disease</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td>History of GI Bleeding</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Concomitant GI Drug use</td>
<td>5%</td>
<td>4%</td>
</tr>
<tr>
<td>History of Rheumatoid Arthritis</td>
<td>5%</td>
<td>3%</td>
</tr>
<tr>
<td>History of Osteoarthritis</td>
<td>49%</td>
<td>35%</td>
</tr>
</tbody>
</table>

### NSAIDs & GI Bleeds: Results

<table>
<thead>
<tr>
<th>Statistical Method</th>
<th>RR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted (Crude)</td>
<td>1.09 (0.91-1.30)</td>
</tr>
<tr>
<td>Multivariable Logistic Regression</td>
<td>0.96 (0.79-1.15)</td>
</tr>
<tr>
<td>Including PS in Regression Model</td>
<td>0.95 (0.79-1.14)</td>
</tr>
<tr>
<td>PS Matching</td>
<td>0.95 (0.77-1.17)</td>
</tr>
<tr>
<td>Inverse Probability of Treatment Weighting</td>
<td>0.87 (0.71, 1.09)</td>
</tr>
<tr>
<td>SMR Weighted Estimator</td>
<td>0.83 (0.66, 1.03)</td>
</tr>
</tbody>
</table>
Example Paper with Treatment Effect Heterogeneity

Original Contribution

Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect

Tobias Kurth, Alexander M. Walter, Robert J. Glynn, K. Arnaud Chevret, J. Michael Gaziano, Klaus Berger, and James M. Foliad

![Image of a graph showing probability density function of the propensity score for the I-TRAQ database]

FIGURE 1. Probability density function of the propensity score for the I-TRAQ database. The x-axis represents the propensity score, and the y-axis represents the probability density. The two lines represent treated (1) and untreated (0) groups.

Thrombolysis and Mortality

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Treated (n = 25)</th>
<th>Not Treated (n = 50)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>99 to 100</td>
<td>0.3666 26</td>
<td>0.3333 25</td>
<td>0.674</td>
</tr>
<tr>
<td>96 to 99</td>
<td>0.3434 25</td>
<td>0.3111 25</td>
<td>0.123</td>
</tr>
<tr>
<td>91 to 95</td>
<td>0.3889 25</td>
<td>0.3444 25</td>
<td>0.098</td>
</tr>
<tr>
<td>86 to 90</td>
<td>0.4571 26</td>
<td>0.4074 25</td>
<td>0.080</td>
</tr>
<tr>
<td>75 to 85</td>
<td>0.4231 25</td>
<td>0.3778 25</td>
<td>0.201</td>
</tr>
<tr>
<td>66 to 74</td>
<td>0.2308 25</td>
<td>0.2173 25</td>
<td>0.102</td>
</tr>
<tr>
<td>25 to 50</td>
<td>0.1600 25</td>
<td>0.2333 25</td>
<td>0.014</td>
</tr>
<tr>
<td>10 to 24</td>
<td>0.0909 25</td>
<td>0.1571 25</td>
<td>0.0000007</td>
</tr>
</tbody>
</table>

All cause mortality in stroke patients; Kurth T et al. AJE 2006
Coxib Example: Unmeasured Confounding

- Many GI risk factors are unmeasured in health care claims data files
  - Tobacco use
  - BMI / Obesity
  - Alcohol consumption
  - Aspirin use
- PS, IPTW methods cannot address this problem

An abundance of codes

- Search through these data to find claims codes that serve as proxies for previously unmeasured confounders.
Sources of codes

- Inpatient services
- Outpatient services
- Inpatient diagnoses (3, 4, 5-digit ICD)
- Outpatient diagnoses (3, 4, 5-digit ICD)
- Pharmacy fills (generic drug, drug class)
- Lab tests
- Lab values
- ...

Proxies in Claims

- Claims may contain proxies for unobserved confounders
- Lipid-testing important confounder in studies of statins (Seeger, Med Care)
- Can we identify important proxies in healthcare claims?
High-dimensional PS (hd-PS) Algorithm

- The approach:
  - Collect as many codes as possible
  - Identify those codes that could possibly bias the exposure/outcome relationship
  - Combine variables identified a priori with the “best” of these codes in a propensity score.
  - Use this “high dimensional propensity score” to adjust for confounding.
- Currently implemented in a SAS macro.

Detailed results of coxib study using hd-PS

<table>
<thead>
<tr>
<th>#</th>
<th>Variable included in propensity score model</th>
<th>NO of variables adjusted</th>
<th>Variables per data source</th>
<th>Data source</th>
<th>CoxPH model 1</th>
<th>CoxPH model 2</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Univar model</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>age, sex, race, year</td>
<td>1</td>
<td></td>
<td></td>
<td>1.23</td>
<td>0.99~1.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>previous history (Y/N)</td>
<td>1</td>
<td></td>
<td></td>
<td>0.99</td>
<td>0.96~1.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>previous non-steroidal anti-inflammatory drugs</td>
<td>2</td>
<td></td>
<td></td>
<td>0.93</td>
<td>0.91~0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>admission source (inpatient, outpatient)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Possible Explanations?

- Coxibs are not GI protective in this elderly population
- High non-adherence
- NS NSAIDs are co-prescribed with GI protective drugs
- Unmeasured confounding
Practical Guidelines

1) Importance of variable selection
   – Avoid entering variables not associated with outcome
   – Report % of exposed that could be matched to unexposed

2) Look for non-uniform effects over range of PS
   – Consider matching, range restrictions, trimming
   – Discuss residual confounding vs. treatment heterogeneity

3) Implementation of PS (modeling, stratification, matching, weighting) minor issue given uniform effects

Discussion / Questions

Multivariable Outcome Models

Usually too many confounders to stratify over and we must use a model.
Multivariable outcome models are models of an expectation (mean/average value) of an outcome given covariates and treatment.

Linear Regression
\[ E[Y|X,C] = b_0 + b_1X + b_2C + b_3C^*X \]

Logistic Regression
\[ E[Y|X,C] = (1+exp(-b_0 - b_1X - b_2C - b_3C^*X))^{-1} \]
Causal Inference From Multivariable Outcome Models

If all confounders are measured (treatment if exchangeable) and model is correct, then model is estimating an expected value of a counterfactual given covariates:

\[ E[Y|X=1, C] = E[Y(1)|C], \quad E[Y|X=0, C] = E[Y(0)|C] \]

One can then average these to get average causal effects (not conditional on C) – see appendix.

Validity depends on getting the model right!

---

Getting a more meaningful/interpretable estimate

\[ E[Y(1)|C] \] is the expected value of \( Y(1) \) given a set of confounders

How do you get from a model for \( E[Y(1)|C] \) and \( E[Y(0)|C] \) to causal parameters/contrasts of interest?

For example, the causal risk difference

\[ E[Y(1)] - E[Y(0)] \]

---

Marginalization (“G-computation”)

If we have a single discrete covariate, \( C \)

\[
E[Y(0)] = \sum_c E[Y(1) | C = c] \Pr(C = c)
\]

\[
= \sum_c E[Y | A = 1, C = c] \Pr(C = c)
\]

Weighted average of “sub-group” effects, where the weights are the probability density.

Estimate this with our fitted model and the empirical (observed) distribution of \( C \)

\[
\sum_{i=1}^n \hat{E}[Y | A = 1, C = c_i] \frac{1}{n}
\]
In English, please?
We can estimate causal parameters of interest using a fit multivariable model.
0) Fitting the multivariable model to the observed data
1) Create a dataset but set A=1 for all patients,
2) Using fit model generate predicted outcomes for all patients
3) Take the average of these to estimate $E[Y(1)]$
4) Repeat 1)-3) but set A=0 for all patients to estimate $E[Y(0)]$
5) Estimate causal risk difference

\[ RD = \hat{E}[Y(1)] - \hat{E}[Y(0)] \]

Fitted Model Allows Us to Estimate Other Parameters of Interest
• Can estimate the effect of treatment in the treated (on a risk difference scale)
  \[ E[Y(1)|A=1] - E[Y(0)|A=1] \]
  (or relative scale)
  \[ \frac{E[Y(1)|A=1]}{E[Y(0)|A=1]} \]
• Fit model to all patients
• Set treatment to zero for the treated patients, use model to predict outcome in patients, average these to get an estimate of $E[Y(0)|A=1]$
• Estimate $E[Y(1)|A=1]$ using empirical (observed) rate of outcome in the treated

Can Estimate the Effects on a Population of “Treatment Rules”
• Define new counterfactuals
  $Y(\text{"treat on if on warfarin"})$=outcome for a patient if he is only treated if he is on warfarin
  \[ Y(1) = \text{outcome if treated} \]
  \[ Y(0) = \text{outcome if not treated} \]
• Estimate
  \[ E[Y(\text{"treat on if on warfarin"})] - E[Y(0)] \]
How do we estimate this?

- To estimate \( E[Y(\text{"treat on if on warfarin"})] \)
- Fit out multivariable model \( E[Y|X,C] \)
- Create a new dataset with treatment reassigned based on treatment rule
- Use fit model to generate predicted values of the outcome for all patients
- Average these to estimate \( E[Y(\text{"treat on if on warfarin"})] \)
- Compare this to \( E[Y(0)] \) as previously estimated

We can use model to estimate counterfactuals in different populations

\[
E[Y(1)] = \sum_c E[Y(1)|C=c] \Pr(C=c)
\]

\( E[Y(1)] \) depends on the distribution of the covariates, \( \Pr(C=c) \) …

What if the average age in the population were ten years older? We can plug-in an arbitrary distribution of \( C, \Pr^*(C=c) \), and estimate \( E[Y(1)] \)

\[
\hat{E}[Y(1)] = \sum_c \hat{E}[Y|A=1,C=c] \Pr^*(C=c)
\]

Major limitation of “G-computation” based on an outcome model

- Outcome model must be correctly specified
  - Include all confounders
  - Including interactions between covariates
- Easy to inadvertently extrapolate model in to region where there is little covariate data
- Propensity score / inverse-probability of weighting methods
Derivation of IPTW

\[
E \left[ \frac{XY}{Pr(X=1|C)} \right] = E \left[ \frac{XY(1)}{Pr(X=1|C)} \right] \quad \text{By consistency assumption}
\]

\[
= E \left[ \frac{XY(1)}{Pr(C=1|C)} \frac{1}{Pr(X=1|C)} \right] = E \left[ \frac{Y(1)}{Pr(X=1|C)} E[X|C,Y(1)] \right]
\]

\[
= E[Y(1)] \quad \text{By no unmeasured confounders}
\]

\[
RD = E[Y(1)] - E[Y(0)] = E \left[ \frac{XY}{Pr(X=1|C)} \right] - E \left[ \frac{(1 - X)Y}{1 - Pr(X=1|C)} \right]
\]

Derivation of IPTW, cont…

\[
RD = E \left[ \frac{XY}{Pr(X=1|C)} \right] - E \left[ \frac{(1 - X)Y}{1 - Pr(X=1|C)} \right]
\]

\[
RD_{\text{IPTW}} = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i Y_i}{PS(C_i)} - \frac{1}{n} \sum_{i=1}^{n} \frac{(1 - X_i)Y_i}{1 - PS(C_i)}
\]
Natural Experiments and Instrumental Variable Methods

M. Alan Brookhart, Ph.D.
Department of Epidemiology,
UNC Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Learning Objectives

• To understand the assumptions and mechanics underlying instrumental variable estimation
• To understand how to evaluate an interpret an instrumental variable analysis
• To learn about some instrumental variable estimators that have been used in practice

Donald Rumsfeld the Accidental Epidemiologist

“... there are known knowns; there are things we know we know. We also know that there are known unknowns; that is to say we know that there are some things we do not know. But there are also unknown unknowns – the ones we don’t know we don’t know. ... it is the latter category that tend to be the difficult ones.”
Instrumental Variable Methods

- Developed and widely used by economists
- Can be used to bound and estimate treatment effects even when confounders are unmeasured
- IV methods depend on the existence of an instrumental variable ("instrument")

Causal Diagram of Structural IV Assumptions

Example: Randomized Controlled Trial with Non-Compliance

Note: Z can be a valid IV under less restrictive conditions
Note 2: Double headed arrow represents association due to direct causal relation between Z and C or C and Z or an assoc. due to a common cause

IV Assumptions Informally

- Instrument should be correlated with treatment
- Instrument should be related to outcome only through association with treatment (often termed the exclusion restriction)
  - Empirically unverifiable, but can be explored in observed data.
**Intention-to-treat (ITT) Approach**

In RCTs with non-compliance, as-treated can be biased estimate of the effect of treatment.

ITT estimates the effect of Z on Y

\[ ITT = \hat{E}[Y | Z = 1] - \hat{E}[Y | Z = 0] \]

In placebo-controlled trials, ITT estimates tend to be biased towards the null when there is non-compliance.

---

**Classic IV estimator is a rescaled ITT estimator**

\[
\hat{a}_{IV} = \frac{\hat{E}[Y | Z = 1] - \hat{E}[Y | Z = 0]}{\hat{E}[X | Z = 1] - \hat{E}[X | Z = 0]}
\]

- X is received treatment

- Numerator is the intention to treat (ITT) estimate of the risk difference
- Denominator is estimate of the effect of the instrument on treatment on the risk difference scale

---

**Modeling Issues**

- IVs can also be motivated as a solution to systems of equations (allows one to include cov)
  - A linear model for treatment (first-stage) that includes IV and covariates
  - A linear model for the outcome that includes exposure and covariates
  - System is solved by two-stage least-squares
- Many other variations
  - IV probit (implemented in Stata), probit models for both first and second stages
Weak Instruments

- Consistent: Wald estimator / 2SLS converges to true parameter but still biased in finite samples
- When instruments are weakly related to treatment (as quantified using a first-stage F statistic),
  - Residual bias in IV due to violations of assumptions is amplified
  - Variance increased
  - 2SLS estimates biased toward OLS, even if IV is perfect
  - 2SLS confidence intervals are too narrow, particularly with many instruments and/or a first-stage F under 10.
  - Alternative estimation procedure (LIML: limited information maximum likelihood) is preferable.

See Staiger & Stock (1997)

Heterogeneous Treatment Effects

- Additional assumption required to justify IV estimator
- One example: “Monotonicity”
  - In RCT example: 4 latent causal classes: always takers, never-takers, defiers, compliers
  - Monotonicity -> no defiers
  - If you took treatment in the placebo arm, you would receive treatment in active arm
  - IV estimates the average effect of treatment in the compliers (‘marginal’ patients)

Goal: Use instrumental variable methods to estimate short-term risk of GI outcomes between
- COX-2 selective NSAIDs versus
- Non-selective NSAIDs

Confounding: Coxibs are likely to be selectively prescribed to patients at increased GI risk
Characteristics of Cohort

<table>
<thead>
<tr>
<th>Variable</th>
<th>COX</th>
<th>NS NSAID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female Gender</td>
<td>86%</td>
<td>81%</td>
</tr>
<tr>
<td>Age &gt; 75</td>
<td>75%</td>
<td>65%</td>
</tr>
<tr>
<td>Charlson Score&gt;1</td>
<td>76%</td>
<td>71%</td>
</tr>
<tr>
<td>History of Hospitalization</td>
<td>31%</td>
<td>26%</td>
</tr>
<tr>
<td>History of Warfarin Use</td>
<td>13%</td>
<td>7%</td>
</tr>
<tr>
<td>History of Peptic Ulcer Disease</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>History of GI Bleeding</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Concomitant GI drug use</td>
<td>5%</td>
<td>4%</td>
</tr>
<tr>
<td>History GI drug use</td>
<td>27%</td>
<td>20%</td>
</tr>
<tr>
<td>History of Rheumatoid Arthritis</td>
<td>5%</td>
<td>3%</td>
</tr>
<tr>
<td>History of Osteoarthritis</td>
<td>49%</td>
<td>31%</td>
</tr>
</tbody>
</table>

Unmeasured Variables

- Do not have data on
  - Lifestyle variables (e.g., diet, exercise, tobacco use)
  - Cognitive status
  - Physical functioning
  - Clinical variables (e.g., blood pressure, BMI)
  - Lab results (e.g., cholesterol levels)
  - Education level

Conventional Statistical Approach

- Parameter of interest is the risk difference
  - Risk of GI bleed if given COX-2 – Risk of GI bleed if given a NS NSAID
- Conventional linear regression
  - Crude RD
  - Multivariable adjusted RD
Conventional Analysis: Results

<table>
<thead>
<tr>
<th>Outcome Definition</th>
<th>Conventional Unadjusted</th>
<th>Conventional Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated RDx100 (95% CI)</td>
<td>Estimated RDx100 (95% CI)</td>
</tr>
<tr>
<td>GI Event within 60 days</td>
<td>0.03 (-0.12, 0.18)</td>
<td>-0.04 (-0.20, 0.10)</td>
</tr>
</tbody>
</table>

We report the risk difference x 100

Possible Explanations?

• Coxibs are not GI protective in this elderly population
• High non-adherence
• NS NSAIDs are co-prescribed with GI protective drugs
• Unmeasured confounding

Unmeasured Indications for COX-2 Treatment

• These are selectively prescribed to patients at risk of GI complications
• Many GI risk factors are unmeasured in health care claims data files
  – Tobacco use
  – BMI / Obesity
  – Alcohol consumption
  – Aspirin use
  – Complaints to MD about stomach problems
What can we do?

- Sensitivity analysis
  - Requires assumptions about distributions of unknown confounders
- External adjustment, two-stage designs, multiple imputation, propensity score calibration
- Find an instrument!

Physician as IV

- Coxib prescribing is driven strongly by MD preference (Solomon DH, et. al. 2003)
- Implication: Some patients would be treated with coxibs by some physicians and with non-selective NSAIDs by others
- Differences in coxib prescribing patterns is the natural experiment that we exploit

Patient’s GI Risk

- Low
- Moderate
- High

"Marginal Patient"

COX-2 Preferring Physician

NS NSAID  COXIB  COXIB

NS NSAID Preferring Physician

NS NSAID  NS NSAID  COXIB
Estimating Preference

– Volume of NSAID prescribing varies considerably among physicians
– Our approach: use the type of the last NSAID prescription written by each physician as a measure of current preference
– If for last patient, physician wrote a coxib prescription, for the current patient he is classified as a “coxib preferring physician”, other he is classified as an “non-selective NSAID preferring physician.”

Re-Analysis of NSAID Data
## Association between risk factors and treatment received

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coxib Users X=1</th>
<th>NS NSAID Users X=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female Gender</td>
<td>86%</td>
<td>81%</td>
</tr>
<tr>
<td>Age &gt; 75</td>
<td>75%</td>
<td>65%</td>
</tr>
<tr>
<td>Charlson Score &gt; 1</td>
<td>76%</td>
<td>71%</td>
</tr>
<tr>
<td>History of Hospitalization</td>
<td>31%</td>
<td>26%</td>
</tr>
<tr>
<td>History of Warfarin Use</td>
<td>13%</td>
<td>7%</td>
</tr>
<tr>
<td>History of Peptic Ulcer Disease</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>History of GI Bleeding</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Concurrent GI drug use</td>
<td>5%</td>
<td>4%</td>
</tr>
<tr>
<td>History of GI drug use (e.g., PPIs)</td>
<td>27%</td>
<td>20%</td>
</tr>
<tr>
<td>History of Rheumatoid Arthritis</td>
<td>5%</td>
<td>3%</td>
</tr>
<tr>
<td>History of Osteoarthritis</td>
<td>49%</td>
<td>33%</td>
</tr>
</tbody>
</table>

## Instrument should be unrelated to observed patient risk factors

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coxib Preference Z=1</th>
<th>NS NSAID Preference Z=0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female Gender</td>
<td>84%</td>
<td>84%</td>
</tr>
<tr>
<td>Age &gt; 75</td>
<td>73%</td>
<td>72%</td>
</tr>
<tr>
<td>Charlson Score &gt; 1</td>
<td>76%</td>
<td>76%</td>
</tr>
<tr>
<td>History of Hospitalization</td>
<td>29%</td>
<td>27%</td>
</tr>
<tr>
<td>History of Warfarin Use</td>
<td>12%</td>
<td>10%</td>
</tr>
<tr>
<td>History of Peptic Ulcer Disease</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>History of GI Bleeding</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Concurrent GI drug use</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>History of GI drug use (e.g., PPIs)</td>
<td>25%</td>
<td>24%</td>
</tr>
<tr>
<td>History of Rheumatoid Arthritis</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>History of Osteoarthritis</td>
<td>40%</td>
<td>41%</td>
</tr>
</tbody>
</table>

## Instrument should be related to treatment

<table>
<thead>
<tr>
<th>Last NSAID Prescription (IV)</th>
<th>Current Prescription (Actual Treatment)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coxib X=1</td>
</tr>
<tr>
<td>Coxib Z=1</td>
<td>(73%)</td>
</tr>
<tr>
<td>Non-Selective NSAID Z=0</td>
<td>(50%)</td>
</tr>
</tbody>
</table>
**IV estimate of the effect of coxib exposure on GI outcome**

\[
\frac{E[Y|Z=1]-E[Y|Z=0]}{E[X|Z=1]-E[X|Z=0]} = \frac{-0.21\%}{22.8\%} = -0.92\%
\]

- Numerator is the intention to treat (ITT) estimate of the risk difference
- Denominator is estimate of the effect of the instrument on treatment on the risk difference scale

---

**Results: Estimated Risk of GI Complication**

<table>
<thead>
<tr>
<th></th>
<th>Conventional Unadjusted</th>
<th>Conventional Adjusted</th>
<th>Instrumental Variable Unadjusted</th>
<th>Instrumental Variable Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimated RDx100 (95% CI)</td>
<td>Estimated RDx100 (95% CI)</td>
<td>Estimated RDx100 (95% CI)</td>
<td>Estimated RDx100 (95% CI)</td>
</tr>
<tr>
<td>GI Event within 60 days</td>
<td>0.03 (-0.12, 0.18)</td>
<td>-0.04 (-0.20, 0.10)</td>
<td>-0.92* (-1.74, -0.10)</td>
<td>-1.02* (-1.88, -0.16)</td>
</tr>
</tbody>
</table>

*We report the risk difference x 100  * Significant at α=0.05

---

**Comparison to RCT Results**

<table>
<thead>
<tr>
<th></th>
<th>Risk Difference per 100 patients (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60 days</td>
</tr>
<tr>
<td>IV Estimate (All Patients)</td>
<td>-0.22* (1.74, -0.10)</td>
</tr>
<tr>
<td>VIGOR trial (Patients with RA)</td>
<td>0.47 (0.63, 0.12)</td>
</tr>
<tr>
<td>CLASS trial (Patients with OA or RA)</td>
<td>Not Reported</td>
</tr>
</tbody>
</table>
**Limitation: Violations of Exclusion Restriction**

- IV should affect outcome only through its association with treatment
- IV weakly associated age, Charlson score, history of arthritis, hospitalizations
  -> Differences in patient case-mix
- IV weakly associated with past use of warfarin
  -> Differences in medical practice or case-mix
  Physicians who use coxibs see sicker patients, use medications that increase GI risk

**Limitation: Heterogeneous Treatment Effects**

- When treatment effects are heterogeneous, IV estimator may be biased for ATE
- Under ‘monotonicity’ IV estimates average treatment effect in ‘marginal’ patients

**Monotonicity Assumption**

- In a randomized trial, coin flip encourages patients to take drug A or drug B
- Monotonicity states that there are no patients who would always do the opposite of what they were encouraged to do
- Monotonicity will not strictly hold in our setting
If monotonicity doesn’t hold, what is IV estimating in the presence of treatment effect heterogeneity?

- Weighted average of treatment effects, where the weight in a sub-group depends on the strength of the IV in the subgroup
- Can use subject matter knowledge to interpret...

Treatment effect heterogeneity: overuse of medications

- Coxibs are thought to be over-used, given to many patients who may not benefit from added GI protection
- High risk patients treated by most physicians
- IV is affecting treatment more in low risk patients
  - IV estimate over-weights effect of treatment in low risk patients
- If low risk patients less likely to benefit, IV underestimates benefit of treatment at population-level (ATE)

Treatment effect heterogeneity: underuse of medications

- Statins are widely thought to be underused, not given to many patients who might benefit
- Low risk patients not being treated by most physicians
  - IV estimate over-weights effect of treatment in high risk patients
- If high risk patients more likely to benefit, IV overestimates benefit of treatment at population-level (ATE)
Treatment effect heterogeneity: misuse of medications (Contraindications)

- Physicians who infrequently use a medication may be more likely to misuse it
- Patients are at greater risk of adverse event if they see a physician who does not use medication
- Preference-based IV methods could make a drug appear to prevent a side effect that it causes

Exploring Possible Bias Due To Treatment Effect Heterogeneity

- Can look for evidence of possible treatment effect heterogeneity
- Does strength of the IV vary across sub-groups?
- Coxib study overall strength of IV was 24%
- In patients with a history of GI bleed, IV strength was 19%
- IV likely slightly underestimating average treatment effect (ATE)

Other examples of preference-based instrument

- Explicit clinician preference (Korn, Stat. Sci.)
- Clinic, hospital as IV (Johnston, J Clin Epi)
- Geographic region as instrument (Wen, J Clin Epi, Brooks et al, HSR, Stuckel T, et al JAMA)
-> All attempt to estimate treatment effects by using difference in practice patterns as a quasi-experiment
• Used claims data from Ontario
• Used geographic region as an IV
• Found significant variation in vaccination rates across regions
• Results compatible with recent studies
• Should have used pre-flu season as a negative control
• Differences between regions in vaccine assessment might have biased results to null
Distance to Specialized Care As An Instrumental Variable

"Does More Intensive Treatment of Acute Myocardial Infarction Reduce Mortality?"

- Medicare claims data identify admissions for AMI, 1987-91
- Treatment: Cardiac catheterization (marker for aggressive care)
- Outcome: Survival to 1 day, 30 days, 90 days, etc.
- Instrument: Indicator of whether the hospital nearest to a patient’s residence does catheterizations.

Are assumptions valid?

1. Is IV associated with treatment?
   26.2% get cath if nearest hospital does caths
   19.5% get cath if nearest hospital does not do caths

2. Is IV associated with outcome other than through its effect on treatment?
   Can’t be determined—but IV is unassociated with observed patient characteristics.

McClellan, et al. results

1. Conventional methods
   - 1-year mortality is 30% lower (17% vs. 47%) if catheterized
   - OLS estimate is -24%, adjusting for observable risk factors

2. IV estimator suggests catheterization associated with 10 percentage point reduction in mortality
   \[
   \frac{E[Y|Z=1]-E[Y|Z=0]}{E[X|Z=1]-E[X|Z=0]} = \frac{-0.7}{6.7} = -10.4\%
   \]
IVs can also be created

- ‘Randomized encouragement’ designs
- Randomized ‘academic detailing’ programs (Avorn and Soumerai)
- Designed delays (McClure M., Dormuth C; work in British Columbia)

Mendelian Randomization (Davey-Smith)

- Using genes as instruments for phenotypes or environmental exposures
- Mendel’s Law of Independent Assortment: during gamete formation, segregation of alleles from one allelic pair is independent of the segregation of the alleles of another allelic pair
The effect of alcohol on HDL and hypertension
Davey-Smith and Ebrahim, BMJ 2005

• Studies of the effect of alcohol consumption are difficult
• Alcohol related to many lifestyle characteristics exposures that are hard to measure
• Enzyme aldehyde dehydrogenase (AD) responsible for alcohol metabolism
• 50% of Japanese are homozygous or heterozygous for a non-functional variant of the AD gene

| Association between genotype and various characteristics |
|----------------|-------------|-------------|
|                | Homozygous | Heterozygous | Functional Variant |
| Mean Alcohol Consumption (ml/day) | 5.3 | 15.1 | 29.2 |
| Mean Age       | 61.3       | 61.5       | 60.6       |
| % Smokers      | 48.5       | 47.9       | 47.7       |
| Mean HDL (mmol/l) | 1.24 | 1.35 | 1.4 |
| % with Hypertension | 40.6 | 37.7 | 46.9 |

Mendelian Randomization
Discussion
• Does this genotype seem like a valid instrument for the effect of alcohol?
• Controversy about effectiveness of vaccine in the elderly and patients with ESRD
• Receipt of vaccine appears to be a marker of good health
• Reports finding 50% reduced risk of mortality in vaccinated patients
• Year-to-year variation in vaccine match represent a natural experiment that we can exploit

Sketch of Design and Analysis

• Identified all hemodialysis patients prevalent on Sept. 1st 1997 and 1998
• Standard Analysis:
  – Vaccination status is a time-varying covariate
• Alternative analysis
  – Compared vaccinated in 1997 to vaccinated in 1998
  – Follow-up started on date vaccine was administered

Antigenic Distance – Vaccine Match

Table 1. Description of the seasons

<table>
<thead>
<tr>
<th>Year</th>
<th>% Match</th>
<th>Predominant strain</th>
<th>Start of season</th>
<th>End of season</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>120%</td>
<td>A/H3N2</td>
<td>1/16/1999</td>
<td>4/10/1999</td>
</tr>
<tr>
<td>1999</td>
<td>120%</td>
<td>A/H3N2</td>
<td>1/18/1999</td>
<td>3/25/1999</td>
</tr>
</tbody>
</table>
Antigenic Distance – Vaccine Match

Table 2. Description of flu seasons

<table>
<thead>
<tr>
<th>Year</th>
<th>% Match</th>
<th>Predominant subtype</th>
<th>Start of flu season</th>
<th>End of flu season</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>8%</td>
<td>A(H3N2)</td>
<td>1/24/1998</td>
<td>2/21/1998</td>
</tr>
<tr>
<td>1998</td>
<td>90%</td>
<td>A(H3N2)</td>
<td>1/16/1999</td>
<td>4/10/1999</td>
</tr>
<tr>
<td>2000</td>
<td>100%</td>
<td>A(H3N2), B</td>
<td>1/12/2002</td>
<td>4/27/2002</td>
</tr>
<tr>
<td>2001</td>
<td>11%</td>
<td>A(H3N2)</td>
<td>1/26/2003</td>
<td>1/17/2004</td>
</tr>
</tbody>
</table>

Standard Analysis: Cox Proportional Hazards Model

Table 3. Estimates of Vaccine Effectiveness Comparing Vaccinated to Unvaccinated Populations by Year

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of Events</th>
<th>Estimated Vaccine Effectiveness</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>300</td>
<td>0.88 (0.83-0.93)</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>350</td>
<td>0.85 (0.80-0.90)</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>400</td>
<td>0.82 (0.78-0.86)</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>450</td>
<td>0.79 (0.75-0.83)</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>500</td>
<td>0.76 (0.72-0.80)</td>
<td></td>
</tr>
</tbody>
</table>

Antigenic Distance – Vaccine Match

Table 4. IME Estimates of Vaccine Effectiveness by Computing Methods in Misclassified Years among Vaccinated vs Unvaccinated Populations

<table>
<thead>
<tr>
<th>Year</th>
<th>Unvaccinated</th>
<th>Vaccinated</th>
<th>Adjusted Vaccine Effectiveness</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>150</td>
<td>250</td>
<td>0.90 (0.85-0.95)</td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>200</td>
<td>300</td>
<td>0.87 (0.82-0.92)</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>250</td>
<td>350</td>
<td>0.84 (0.79-0.89)</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>300</td>
<td>400</td>
<td>0.81 (0.76-0.86)</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>350</td>
<td>450</td>
<td>0.78 (0.73-0.83)</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

- Pharmacoepidemiology
  - Very large data sets
  - Limited ascertainment of confounders
- IV methods may be often indicated
- Key is finding good instruments!
- Care must be taken with
  - Study design
  - Evaluating assumptions
  - Interpreting/generalizing results

Discussion / Questions

Readings On Instrumental Variable Methods

Recommended Reading


Smith GD, Elrashidi S. What can mendelian randomisation tell us about modifiable behavioral and environmental exposures? BMJ 2005
Fitting IVs in Stata

Two-Stage Linear Model

\[ X = a_0 + a_1Z + a_2 \text{age} + a_3 \text{gender} + \ldots + e_x \]

\[ Y = b_0 + b_1 X + b_2 \text{age} + b_3 \text{gender} + \ldots + e_y \]

System is solved by two-stage least-squares

Stata Code

Unadjusted Model (no covariates)
ivreg y (x=z), first

Adjusted Model (with covariates)
xi: ivreg y bleeding ulcer i.year i.gender ost_arthrit (x=z), first
**Stata Code**

Robust standard errors for IV estimator to account for within-physician clustering

\[ \text{ivreg } y \text{ (x=z), first cluster(doctor)} \]

IV Probit Model

\[ \text{ivprobit } y \text{ (x=z), first} \]
Studying Medication Adherence and Outcomes

M. Alan Brookhart, Ph.D.
Department of Epidemiology,
UNC Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Learning Objectives

• To understand how to measure and model medication adherence using pharmacy claims data
• To understand some challenges and potential approach to estimating the effects of adherence on outcomes

Everyone benefits from good adherence

• Stakeholders
  – Pharmaceutical companies
  – Physicians
  – Pharmacies
  – Patients
Overview of Lecture

- Introduction
- Measuring adherence
- Example: Adherence with Osteoporosis Medications
- Dynamic patterns of adherence
- Example: Statins in British Columbia
- The healthy user/adherer effect
- Adherence and comparative safety/effectiveness research

The Consequences of Nonadherence

- 125,000 deaths per year in U.S.\(^1\)
- All medication-related hospital admissions in the United States, 33 to 69 percent are due to poor medication adherence.\(^4\)
- Total cost estimates range from $100 billion\(^2\) to $300 billion.\(^3\)

Why we need to study adherence

- To evaluate the magnitude of the problem
- To understand adherence
- To target interventions
- To help inform/interpret observational safety and effectiveness research of drugs

---

\(^2\) Cited by DiMatteo, Med Care, 2004.
What do we know / do not know

- Many papers on
  - How low adherence is
  - Patient groups at risk of becoming nonadherent (people of lower education, socioeconomic status, depressed patients)
  - Weak predictors of non-adherence (medication regimen complexity, cost)
  - Consequences of nonadherence (somewhat questionable validity)

- Very little is known about
  - Why patients stop specific treatments
  - How to predict nonadherence at the patient level
  - What interventions will cause meaningful improvements

<table>
<thead>
<tr>
<th>Method of Measuring Adherence</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct (pharmacy database)</td>
<td>Most accurate</td>
<td>Patients may be pill-poppers or those who return medicines unopened for another use</td>
</tr>
<tr>
<td>Indirect (biological markers)</td>
<td>Objective</td>
<td>Requires expensive equipment and skills to measure biological effects</td>
</tr>
<tr>
<td>Self-report measures</td>
<td>Single measures that reflect adherence in the clinical setting</td>
<td>Sensitivity to true adherence issues in a patient's care is unknown</td>
</tr>
</tbody>
</table>

Osterberg and Blaschke, NEJM 2005

Typical Pharmacy Claims Data

- Date filled
- Agent (NDC code) & dose
- Days Supply
- Physician identifier
- Pharmacy identifier
- "Refill" indicator
Typical Pharmacy Refill Data

Typical Pharmacy Refill Data

Percentage of Days Covered (PDC)

Percentage of Days Covered (PDC)

Medication Possession Ratio (MPR)

Medication Possession Ratio (MPR)

MPR = 150/164 = 0.91
Measuring Persistence

• Selected all new user of osteoporosis medications who were Medicare beneficiaries and eligible for PACE from January 1, 1996, through December 31, 2002.
• Osteoporosis medications were bisphosphonates, HRT, raloxifene, and calcitonin.
• Follow-up was broken into 60-day intervals, percentage of days covered by medication was computed for each interval (patients were dropped from the denominator at death/censoring).
• Discontinuation was defined 120 days with no medication available.

Compliance With Osteoporosis Medications

- Table 1. Characteristics of Patients in the 12 Months Before Initiating a Medication for Osteoporosis

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>43,902</td>
</tr>
<tr>
<td>Female sex</td>
<td>30,622</td>
</tr>
<tr>
<td>Age y</td>
<td>79.1 ± 6.6</td>
</tr>
<tr>
<td>White race</td>
<td>30,419</td>
</tr>
<tr>
<td>No. of major comorbid conditions</td>
<td>2.2 ± 2.9</td>
</tr>
<tr>
<td>No. of different medications</td>
<td>5.2 ± 4.5</td>
</tr>
<tr>
<td>No. of physician visits</td>
<td>9.1 ± 5.6</td>
</tr>
<tr>
<td>Acute care hospitalization</td>
<td>15.191</td>
</tr>
<tr>
<td>Nursing home residence</td>
<td>4806</td>
</tr>
<tr>
<td>Presence of the No. pain, fracture, or spine</td>
<td>7,690</td>
</tr>
<tr>
<td>Bone mineral density testing</td>
<td>6055</td>
</tr>
<tr>
<td>Starting medications (chemotherapy or combination)</td>
<td>15,750</td>
</tr>
<tr>
<td>Bisphosphonate</td>
<td>11,720</td>
</tr>
<tr>
<td>Calcitonin</td>
<td>1,270</td>
</tr>
<tr>
<td>Calcitriol</td>
<td>1,270</td>
</tr>
<tr>
<td>Raloxifene hydrochloride</td>
<td>2,818</td>
</tr>
<tr>
<td>Estradiol + third generation hormone</td>
<td>3,138</td>
</tr>
<tr>
<td>Other combinations</td>
<td>436</td>
</tr>
</tbody>
</table>
Persistence is very poor

One year after initiating treatment for osteoporosis, 45.2% of the 40,002 patients were not continuing treatment.

Statistical Modeling of Adherence

- Modeled adherence in each 60-day interval via a repeated measures model
  - PDC as a continuous variable
  - Adherence as a dichotomous variable (PDC>66%)
- One model with baseline variables, one with time-varying covariates
OP adherence study: Results

- Persistence is poor
- Clinical need predicts adherence
- Poor adherence was associated with non-white race, old age
60% of patients who stop treatment for 60 days have restarted within two years...

Use of OP medications appears to be dynamic.

Positive interpretation:
Adherence not quite as bad as we thought

Adherence as a Dynamic Process?

- The prevailing paradigm is that adherence is relatively static
- Many health-related behaviors are cyclical
  - Dieting
  - Exercise
- Is it useful to view adherence as a dynamic process?

A study of 239,911 new users of statins in British Columbia, of whom 129,167 (53.8%) had a period of nonadherence that lasted for at least 90 days.

- How many of these patients restart statin therapy?
- Can we identify predictors of re-initiation?
Re-initiation Rate

- Of patients who stopped therapy for at least 90 days, an estimated 38% restarted treatment within one year and 52% restarted within two years.
- Statin use is dynamic

Identifying Predictors of Re-initiation: a case crossover design

- Events
  - Cholesterol testing
  - Any physician visit
  - Visit with physician who started the patient on a statin
  - CAD-related hospitalization

Results

<table>
<thead>
<tr>
<th>Event</th>
<th>14-Day Delay Period</th>
<th>14-Day Baseline Period</th>
<th>28-Day Delay Period</th>
<th>28-Day Baseline Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholesterol testing</td>
<td>0.25</td>
<td>0.47</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>Any physician visit</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Physician who started the patient on a statin</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>CAD-related hospitalization</td>
<td>0.13</td>
<td>0.26</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event</th>
<th>14-Day Delay Period</th>
<th>14-Day Baseline Period</th>
<th>28-Day Delay Period</th>
<th>28-Day Baseline Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholesterol testing</td>
<td>1.65</td>
<td>0.64</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>Any physician visit</td>
<td>1.02</td>
<td>0.47</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>Physician who started the patient on a statin</td>
<td>1.00</td>
<td>0.47</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>CAD-related hospitalization</td>
<td>0.13</td>
<td>0.26</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Statin Adherence Dynamics Study: Results

- Statin use is dynamic, once stopped does not mean always stopped
- “Fire-and-forget” approach to treatment not optimal
- Physician follow-up and provider continuity appear to be important components of adherence

Limitations

- Don’t know why a patient stopped taking med
- Uncertainty about causal process
  - Do patients see a physician because they need a refill?
  - Physician urges patient to resume treatment

Adherence to Lipid-Lowering Therapy and the Use of Preventive Health Services: An Investigation of the Healthy User Effect


- Are patients who adhere to statins more likely to do other things that might affect outcomes?
- Sought to examine association between adherence and use of prevention-oriented health services
- Identified a cohort of new users of statins between 1996 and 2004 with no evidence of coronary heart disease (history of AMI, diabetes, angina, hypercholesterolemia)
Study Design

- Dropped patients who died (602), lost eligibility (1,937), entered a nursing home (1,269) during ascertainment period — 20,783 remained
- Outcomes: fecal occult blood test, influenza vaccination, pneumonia vaccination, mammogram, prostate specific antigen test, and bone mineral density test.

Healthy Adherer Results

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Univariable odds ratio</th>
<th>95% confidence interval</th>
<th>Multivariable-adjusted odds ratio</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bone mineral density test</td>
<td>1.04</td>
<td>0.94, 1.17</td>
<td>1.08</td>
<td>0.96, 1.10</td>
</tr>
<tr>
<td>Screening mammogram</td>
<td>1.22</td>
<td>1.20, 1.24</td>
<td>1.22</td>
<td>1.20, 1.26</td>
</tr>
<tr>
<td>More visits</td>
<td>1.60</td>
<td>1.59, 1.61</td>
<td>1.57</td>
<td>1.55, 1.60</td>
</tr>
<tr>
<td>Prostate-specific antigen test</td>
<td>1.08</td>
<td>0.97, 1.20</td>
<td>1.01</td>
<td>0.92, 1.10</td>
</tr>
<tr>
<td>Fecal occult blood test</td>
<td>1.20</td>
<td>1.19, 1.22</td>
<td>1.25</td>
<td>1.18, 1.32</td>
</tr>
<tr>
<td>Influenza vaccination</td>
<td>1.19</td>
<td>1.14, 1.26</td>
<td>1.20</td>
<td>1.15, 1.26</td>
</tr>
</tbody>
</table>

*p Values were assessed at the end of follow-up, using the Statistical Analysis System for Windows (SAS) (v9) software. Clinical significance is defined as *p < 0.05.

Conclusion

- Patients who adhere to statins more likely to receive a range of prevention-oriented clinical service
• Research Question: Are patients who are adherent to statins at lower risk of outcomes unlikely to be affected by statin exposure but likely to be related to healthy lifestyle?
• Population: All new users of statins in British Columbia with no evidence of existing heart disease

### Conclusions

• Patients who adhere to statins more likely to receive a range of prevention-oriented clinical service at decreased risk of accidents and adverse health outcomes
Medication nonadherence is associated with a broad range of adverse outcomes in patients with coronary artery disease

- 15,767 patient with CAD
- Adherence dichotomized: PDC>80% in first 180 days
- Adherence associated with various outcomes during a 1-year follow-up period

### Main Results

**Table 1: Characteristics of the study population according to adherence category for each of the 3 medications**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All (n = 11,053)</th>
<th>Nonadherent (n = 3,445)</th>
<th>Adherent (n = 7,608)</th>
<th>P (Adherent vs. Nonadherent)</th>
<th>P (All vs. Nonadherent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>60.2 (10.0)</td>
<td>61.8 (10.7)</td>
<td>58.5 (9.9)</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Heart failure</td>
<td>32.1</td>
<td>36.5</td>
<td>28.6</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>DM</td>
<td>46.7</td>
<td>52.2</td>
<td>41.2</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>CAD</td>
<td>40.9</td>
<td>47.6</td>
<td>34.2</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Hypertension</td>
<td>94.1</td>
<td>94.3</td>
<td>94.0</td>
<td>0.89</td>
<td>0.001</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>29.5</td>
<td>34.6</td>
<td>24.6</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Current smoker</td>
<td>10.0</td>
<td>12.2</td>
<td>8.8</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Adherence dichotomized</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adherent</td>
<td>22.7</td>
<td>26.3</td>
<td>22.3</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Cessation</td>
<td>18.9</td>
<td>19.0</td>
<td>18.7</td>
<td>0.2</td>
<td>0.001</td>
</tr>
<tr>
<td>Alcohol</td>
<td>4.5</td>
<td>5.7</td>
<td>4.1</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Smoking</td>
<td>29.7</td>
<td>34.3</td>
<td>28.5</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Hypertension</td>
<td>10.2</td>
<td>12.1</td>
<td>8.8</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>Total events</td>
<td>73.7</td>
<td>72.8</td>
<td>74.6</td>
<td>0.1</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Main Results**

- Very strong effects
- Effect weaker for more specific outcomes
**Sensitivity Analysis**

<table>
<thead>
<tr>
<th>Medication</th>
<th>Acid reflux</th>
<th>CV mortality</th>
<th>CV hospitalization</th>
<th>CHD mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placebo</td>
<td>1.03 (0.95, 1.11)</td>
<td>1.05 (0.97, 1.14)</td>
<td>1.02 (0.94, 1.10)</td>
<td>1.04 (0.96, 1.14)</td>
</tr>
<tr>
<td>PPI</td>
<td>1.02 (0.94, 1.10)</td>
<td>1.06 (0.98, 1.15)</td>
<td>1.03 (0.95, 1.11)</td>
<td>1.05 (0.97, 1.14)</td>
</tr>
<tr>
<td>H2 blockers</td>
<td>1.01 (0.93, 1.09)</td>
<td>1.05 (0.98, 1.14)</td>
<td>1.02 (0.94, 1.10)</td>
<td>1.04 (0.96, 1.14)</td>
</tr>
</tbody>
</table>

- Acid reflux disease is symptomatic
- PPI, H2 blockers often not used chronically
- Confounding: angina often confused for reflux disease
- Fewer people are adherent

---

**Effectiveness of Statin Therapy in Adults With Coronary Heart Disease**

- Meta-analysis of 19 placebo-controlled statin trials in secondary prevention
- All cause mortality reduced by 16% (vs 85%)
- CHD mortality and non-fatal MI by 25% (vs 35% CV hospitalization 62% CV Death)

---

- Estimation of the benefits of adherence appears to be overstated
- What else can we do to estimate the effect of adherence?
Approaches to Control the Healthy User Bias: Better Adjustment

- Variables
  - Healthy behaviors
  - Unhealthy behaviors
  - Education
  - Use of other medications
  - Cognitive and functional status
  - Access to care
- These variables are not available in most pharmacoepidemiologic databases in US
- High-dimensional “proxy” adjustment

Approaches to the Healthy User Bias: Active Comparator Group

- Compare adherent new initiators of statins to adherent new users of other preventive medications
  - What medications?
    - You want something that does not affect the outcome.
  - What about an instrumental variable?

Drug Copayment and Adherence in Chronic Heart Failure: Effect on Cost and Outcomes

- Hard to study effects of medication adherence
- Use copayment as an instrument for the effect of adherence of BB and ACEI in heart failure
- Does this seem like a reasonable IV?
- How would you interpret the results?
Half of Aetna beneficiaries who experienced an AMI were randomized to receive free cardiovascular drugs ($0 copay).

Other received usual benefit.

Hypothesis: Reducing copays will improve adherence, and save money and lives.

6,000 patients randomized.

Rates of adherence were slightly higher in the free drugs arm (6%).

Rates of primary outcome (first major vascular event) not reduced.

Rates of all vascular event were lower in free drugs arm HR=0.89.

Costs were not different between groups.

<table>
<thead>
<tr>
<th>Group</th>
<th>Predicted Medication Possession Ratio, 2002 (%)</th>
<th>Predicted Medical Cost, 2003 ($)</th>
<th>Predicted Frequency of Hospitalization, 2003 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE inhibitor 5</td>
<td>94.5</td>
<td>7583</td>
<td>13.0</td>
</tr>
<tr>
<td>10</td>
<td>93.2</td>
<td>7594</td>
<td>13.3</td>
</tr>
<tr>
<td>15</td>
<td>91.9</td>
<td>7524</td>
<td>13.7</td>
</tr>
<tr>
<td>20</td>
<td>90.7</td>
<td>7495</td>
<td>14.0</td>
</tr>
<tr>
<td>25</td>
<td>89.4</td>
<td>7406</td>
<td>14.4</td>
</tr>
<tr>
<td>30</td>
<td>88.1</td>
<td>7437</td>
<td>14.7</td>
</tr>
<tr>
<td>β-Blocker 5</td>
<td>94.3</td>
<td>8003</td>
<td>10.0</td>
</tr>
<tr>
<td>10</td>
<td>93.5</td>
<td>8079</td>
<td>10.4</td>
</tr>
<tr>
<td>15</td>
<td>92.6</td>
<td>8057</td>
<td>10.8</td>
</tr>
<tr>
<td>20</td>
<td>91.7</td>
<td>8536</td>
<td>11.2</td>
</tr>
<tr>
<td>25</td>
<td>90.8</td>
<td>8417</td>
<td>11.7</td>
</tr>
<tr>
<td>30</td>
<td>89.9</td>
<td>8300</td>
<td>12.1</td>
</tr>
</tbody>
</table>

ACE = angiotensin-converting enzyme.
Final Lesson:
Antipsychotic Medications (APM) in the Elderly

• APMs approved to treat schizophrenia
• Widely used off-label to treat elderly patients with dementia
• Two broad classes: conventional (older drugs) versus atypical (newer drugs)
• Manufacturers of some of the atypicals conducted trials to assess effectiveness of the medications for controlling behavioral disturbances in elderly
• FDA meta-analysis: increased risk of mortality associated with atypical APMs (relative to placebo)
• FDA put a “black box” advisory on label of atypical APMs
Ideally results from different approaches will agree

- APM Study comparing risk of death of new users atypicals to new users of conventionalals
- Used various analytical approaches
  - Used COX-PH regression adjusting covariates
  - Sub-classification on the propensity scores
  - Instrumental variables based on prescribing physician
  - Established a dose-response relation within each APM group
  - Restricted to different populations (with and without dementia diagnosis, current nursing home resident)
  - Sensitivity analysis suggested there would have to be tremendous unmeasured confounding to change our conclusion
- All results suggested that conventional APMs have a higher short-term risk of death than atypical APMs
In the current study, although we agree with the core findings of Wang et al., we find that additional analyses are needed to fully understand the associations. The primary objective of our study was to explore the relationship between different factors and the occurrence of complications. We found that in addition to the factors identified by Wang et al., other variables such as age and gender also play a significant role.

We conclude that further research is needed to fully understand the complex interplay between various factors and the occurrence of complications. Our findings suggest that a comprehensive approach to identification and prevention of complications is necessary. We recommend further studies to confirm our findings and to develop effective strategies for prevention.
Increased Mortality in Elderly Patients with Dementia-Related Psychosis — Elderly patients with dementia-related psychosis treated with antipsychotic drugs are at an increased risk of death. Analysis of seventeen placebo-controlled trials (modal duration of 10 weeks), largely in patients taking atypical antipsychotic drugs, revealed a risk of death in drug-treated patients of between 1.6 to 1.7 times the risk of death in placebo-treated patients. Over the course of a typical 10-week controlled trial, the rate of death in drug-treated patients was about 4.5%, compared to a rate of about 2.6% in the placebo group. Although the causes of death were varied, most of the deaths appeared to be either cardiovascular (e.g., heart failure, sudden death) or infectious (e.g., pneumonia) in nature. Observational studies suggest that, similar to atypical antipsychotic drugs, treatment with conventional antipsychotic drugs may increase mortality. The extent to which the findings of increased mortality in observational studies may be attributed to the antipsychotic drug as opposed to some characteristic(s) of the patients is not clear.

In the absence of regulatory action, the final jury is the prescriber

“…More analysis of these drugs (antipsychotics) clearly needs to be done before any firm conclusions emerge. In the meantime, we should temper our bias that older treatments are de facto safer because they have been on the market longer. As the old saying goes, you don’t know what you don’t know.”

Medical Progress, Dec. 9th 2005
If done well, non-experimental research can contribute useful information about comparative safety and effectiveness of therapeutics.

• “... While many clinicians have shied away from using atypical antipsychotics, this study offers strong (although not convincing) evidence that conventional antipsychotics are even more dangerous. ... it is wise to limit the use of antipsychotics in general, and if they are used, atypicals are likely to be safer.”

  • -Ashish K. Jha, MD MPH
  • Outcomes Research in Review