



OPTIMIZING THE VALUE OF OBSERVATIONAL DATA FOR COMPARATIVE EFFECTIVENESS RESEARCH

AN OVERVIEW

TAMI L. MARK, PHD

MAY 10, 2011



THOMSON REUTERS

TAMI L. MARK, PhD, MBA



- Director, Analytic Strategies, Thomson Reuters
 - Responsible for providing analysis and statistical consultation to outcomes research department
- Health Services Researcher/Health Economist
 - PhD from Johns Hopkins University
- Research Interests
 - Observational research methods
 - Mental health and substance abuse health services research



OUTLINE

1. Motivation and skepticism of observational data
2. Tools to address confounding
3. Transparency
4. Enhancing data elements
5. Validation of data elements
6. Methods of causal inference



MOTIVATION

- Expanding therapeutic choices and increasing costs has generated interest in comparative effectiveness research
- Interest in reflected in recent legislation
 - *Food and Drug Administration Amendment Act of 2007*
 - *American Recovery and Reimbursement Act of 2009*
 - *Patient Protection and Affordable Care Act of 2010*
- Simultaneously, electronic healthcare databases have been growing in size and depth.
- A variety of thought leaders argue this growing mass of healthcare observational data should be leveraged to address the growing comparative effectiveness research (CER) agenda
- Also much skepticism about observational data's ability to meet this challenge



INSTITUTE OF MEDICINE: LEARNING HEALTHCARE SYSTEMS CONCEPTS

“Contending with the changes in healthcare clearly requires doing things differently...”

The notion of dependence on expensive, time-consuming, and relatively narrow clinical experiments as a static and sole source of evidence on the constantly accelerating flow of diagnostic and treatment challenges, is unfeasible.”



STRENGTHS OF OBSERVATIONAL DATA

- Databases can be extremely large
 - Which is critical for studying rare events
- Populations are heterogeneous and include people often excluded from RCTs
 - e.g., those with comorbidities, elderly, children
- Interventions are applied under real-world conditions
- Patients can be followed longitudinally
- Studies are relatively inexpensive



WEAKNESSES RELATIVE TO RANDOMIZED CONTROLLED TRIALS

- Randomized controlled trials (RCTs) widely accepted as the strongest evidence for comparative effectiveness because of strong internal validity
- **Random assignment to treatment and control group renders groups probabilistically exchangeable**
- Confounding may still be introduced:
 - Noncompliance
 - Loss to follow-up
 - Selected samples
 - Random imbalances
 - Un-blinding



OBSERVATIONAL STUDIES CAN PROVIDE WRONG INFORMATION

- **Damning study on HRT leaves women in limbo**

The startling news that hormone replacement therapy may do more harm than good has caused worldwide confusion and panic.

Prospective Cohort Study [1985 Nurses Health Study]

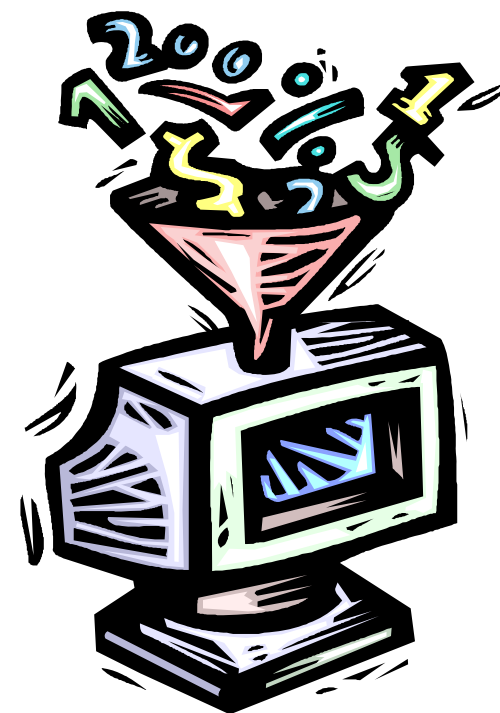
VERSUS

Randomized Control Trials [Heart and Estrogen-progestin Replacement Study and Women's Health Initiative]

Taubes, Gary. Do We Really Know What Makes Us Healthy? New York Times, September 16, 2007

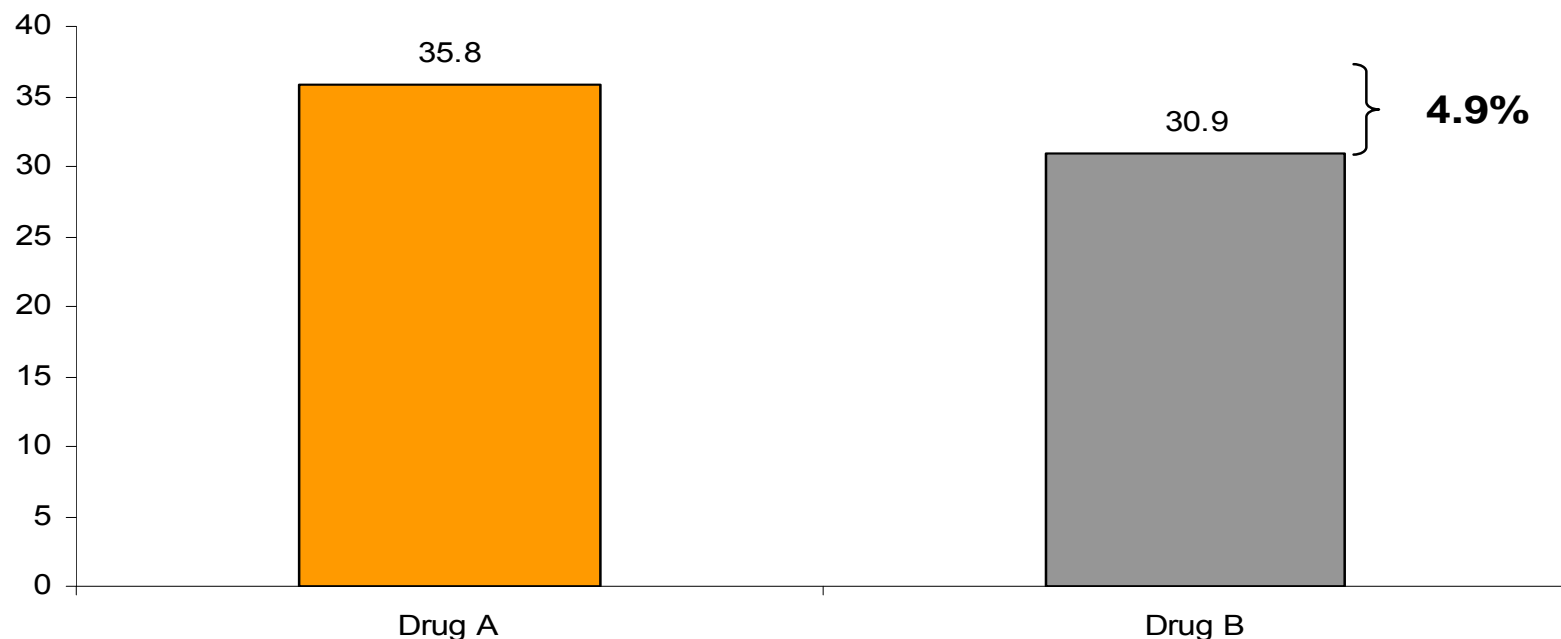


II. TOOLS TO ADDRESS CONFOUNDING



EXAMPLE OF SELECTION BIAS?: RELATIVE EFFECTIVENESS OF TWO PSYCHIATRIC DRUGS IN PREVENTING PSYCHIATRIC HOSPITALIZATIONS

Percent Hospitalized Over a Year After Starting Drug A or B



- Is Drug B **more effective** than Drug A?
- Are the patients who take Drug A **sicker** than the patients who take Drug B?



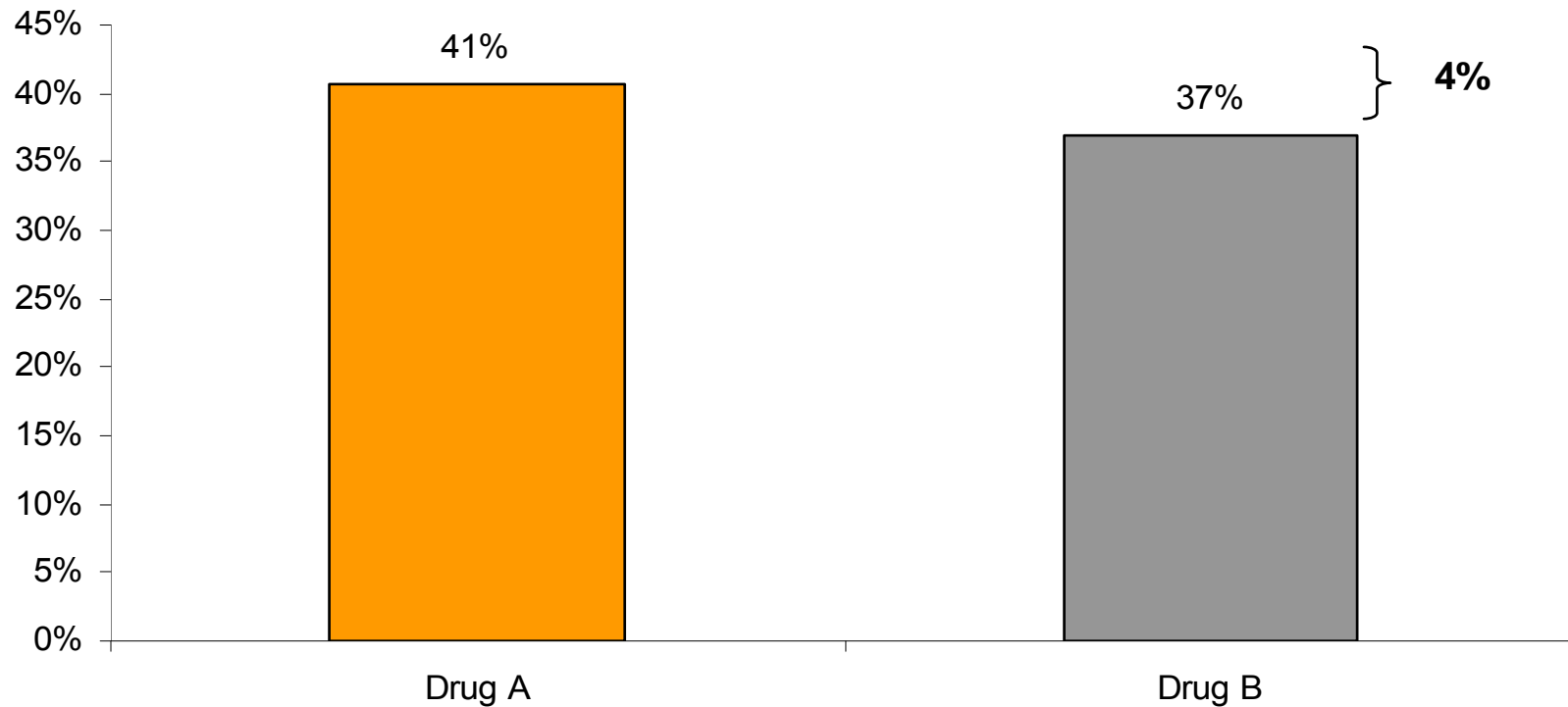
CONFOUNDING

- In statistics, a **confounding variable** is a variable that correlates (positively or negatively) with both the dependent variable and the independent variable
- Excluding these variables can lead to an erroneous “false positive” conclusion that the dependent variables are in a causal relationship with the independent variable
- Confounding is a major threat to the validity of inferences made about cause and effect (i.e., internal validity), as the observed effects should be attributed to the confounder rather than the independent variable



EXAMPLE CONT'D: PATIENTS WHO RECEIVE DRUG A HAD MORE HOSPITALIZATIONS PRIOR TO INITIATING A MEDICATION

Percent Hospitalized in Year Prior to Starting on Drug A or B



P < 0.001



HOW CAN WE ASSURE THAT THE PATIENTS RECEIVING DRUG A AND DRUG B ARE SIMILAR?

- Randomization → Flip a coin and assign people randomly to Drug A or Drug B
- Observational Study Methods → Balance differences between groups by manipulating data



METHODS OF ADJUSTING FOR CONFOUNDERS

1. Restriction
2. Stratification
3. Multivariate Regression
4. Matching
5. Propensity Score Analysis
6. Instrumental Variables
7. Difference-in-Difference Methods



WHAT FACTORS MIGHT BE CONFOUNDERS?

- **Key to avoiding confounding and selection bias due to pre-exposure variables is to identify in advance and measure as many confounders and selection factors as are practical using subject-matter expertise**
- **Possible confounders affecting hospitalization rates**
 - Severity of illness
 - Sociodemographic factors
 - Medical comorbidities
- **Often in observational studies must use proxies for underlying confounders**
 - Severity of illness might be proxied by past hospitalization rates



1. RESTRICTION

- Limit study to only people with certain characteristics to create a homogenous sample
- Sometimes used in concert with other methods
- Common restrictions in administrative data studies
 - Diagnoses
 - Exclude contraindications
 - Incident vs. prevalent/New user design



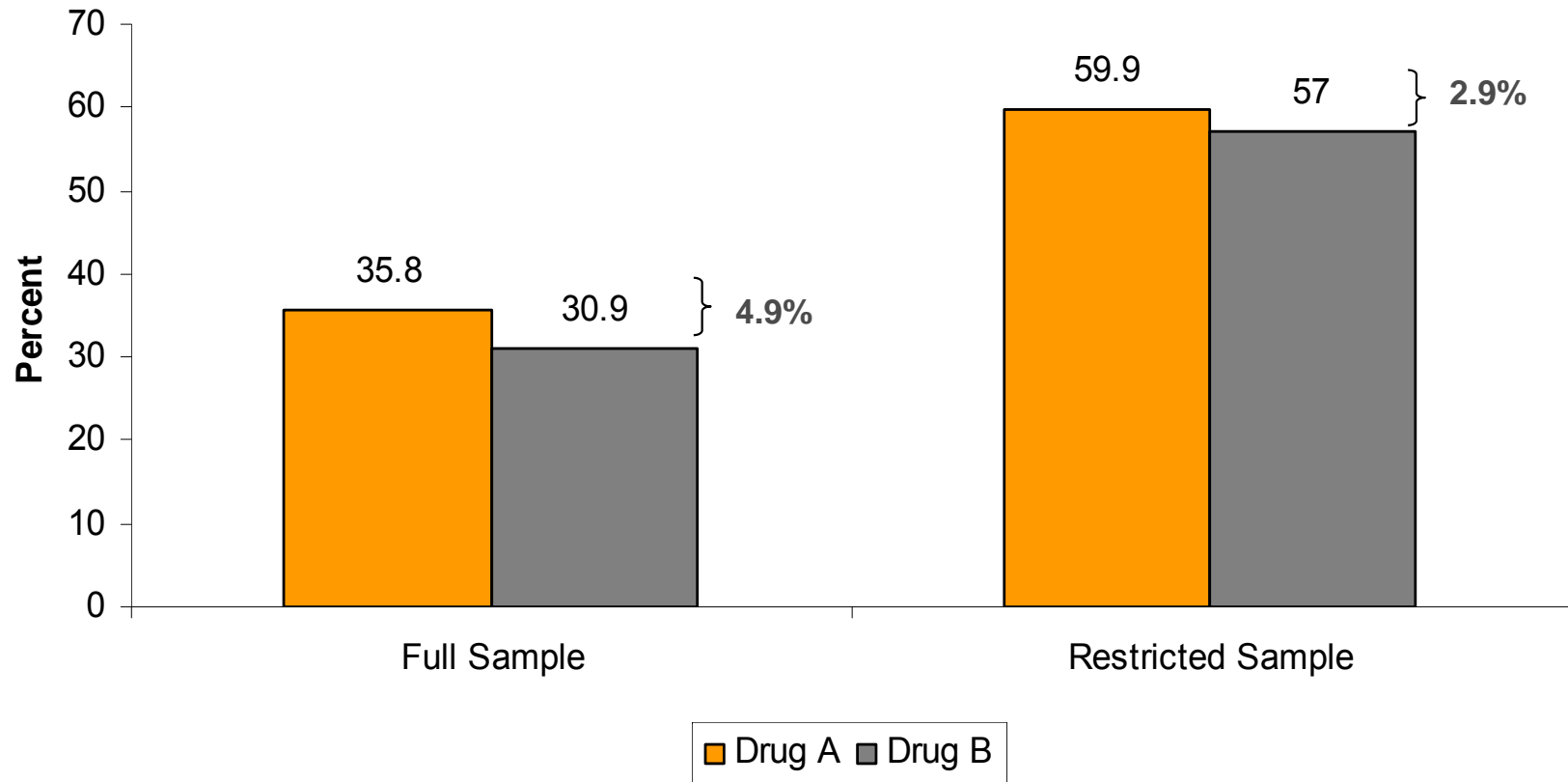
1. RESTRICTION, CONTINUED

- New user design
 - Limit study to people who are new to medications/interventions by using clean period
 - Allows one to adjust on pre-intervention characteristics (e.g., sicker or healthier people select drug)
 - Controls for the fact that continued drug utilization and better outcomes are often positively correlated



RESULTS FOR OVERALL SAMPLE COMPARED TO ONLY THOSE WITH PRE-PERIOD HOSPITALIZATION

Percent Hospitalized: Full and Sample Restricted to Patients with a Pre-Period Hospitalization



LIMITATIONS OF RESTRICTION

- Need to balance sample size and generalizability with homogeneity
 - e.g., what about the people who weren't hospitalized in the pre-period?
- May still be other confounding factors
 - e.g., age, gender, comorbidities
- May not control for unobservable differences

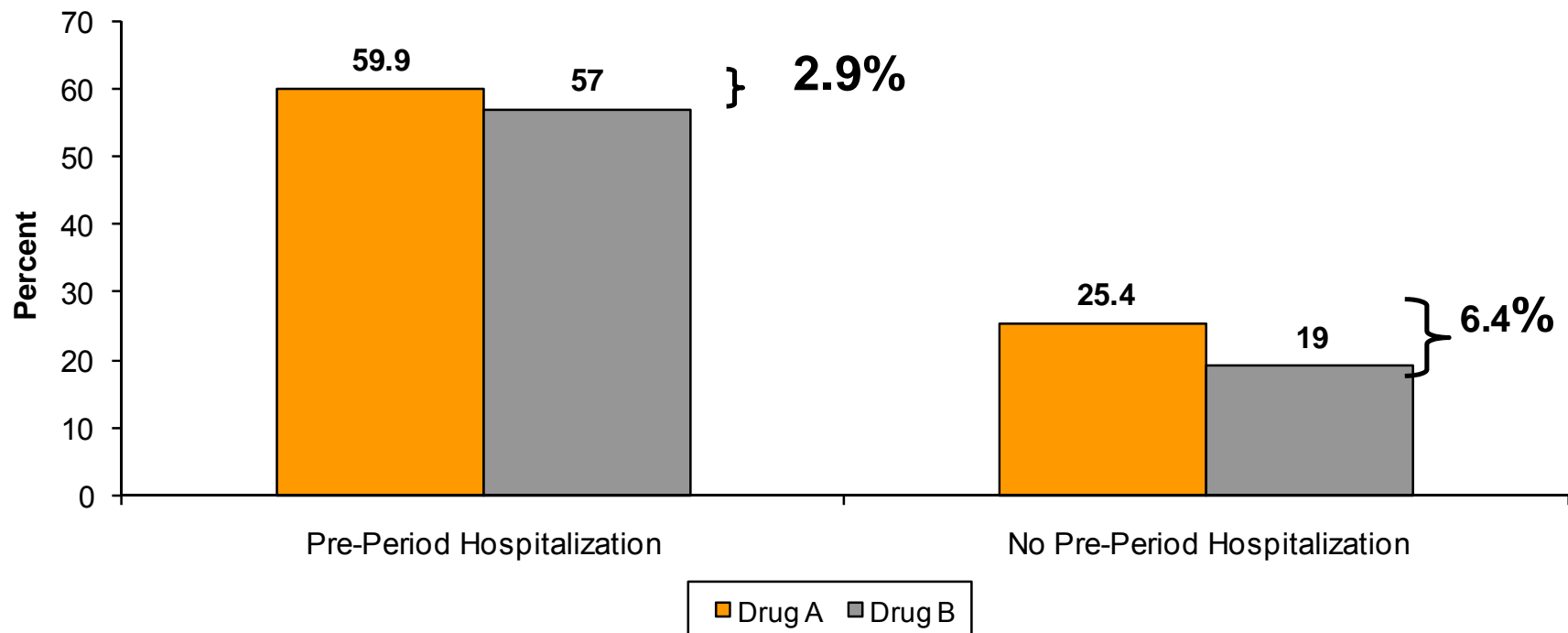


2. STRATIFICATION

- Involves placing data into subcategories, called strata, so that each subcategory can be observed separately

EXAMPLE: STRATIFIED BY PRE-PERIOD HOSPITALIZATIONS

Hospitalizations Outcomes for Sample Stratified by Pre-Period Hospitalization

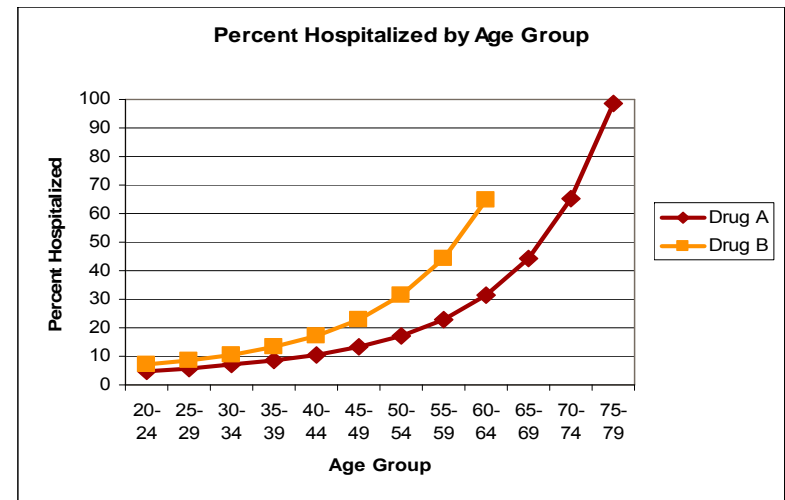


LIMITATIONS OF STRATIFICATION

- Ability to only control for few variables
- Ability to use only a limited number of strata
- May not control for unobservable differences

3. MULTIVARIATE REGRESSION

- $Y = \alpha + \beta X_1 + \beta X_2 + \beta X_3 + \beta X_4 + \beta X_5 + e$
- **Strengths**
 - Ability to control for multiple confounders simultaneously
- **Weakness**
 - Patients with little overlap on covariates are not identified and these patients are not excluded from the analysis, where as they are excluded in other approaches to be discussed
 - Assumes no unobservable confounders



- Example: Drug A patients are all < 65 while Drug B patients were < and > 65
- Predicting outside the range of the covariate leading to too much sensitivity to specification
- Examine normalized difference in distribution of variables to see if this is a problem



4. MATCHING

- Match persons with similar age, gender, race, etc. in treatment arms. Exclude persons without match.



- Limitation of matching: omitting significant portion of population that doesn't match may limit generalizability



- May not control for unobservable confounders



EXAMPLE: MANY POTENTIAL MATCH VARIABLES

- **Demographics:** gender, age, race, income, education, urban/rural
- **Comorbidities:** diabetes, hypertension, high cholesterol, obesity, bipolar disorder, anxiety, substance abuse, CCI, CDS, ICD-9-Count
- **Pre-period use of particular medications**
- **Pre-period use of side-effect medications**
- **Pre-period healthcare utilization**
 - Percent hospitalized
 - Percent with ER visit
 - Adherence rates
 - Visit to specialty mental health provider



5. PROPENSITY SCORE MATCHING

- Definition
 - Probability of Treatment A or Treatment B conditional on a subjects' baseline characteristics
- Advantages
 - Unlike directly matching on each covariate, can match on multiple criteria because the data are reduced into one propensity score
 - Groups are similar and balanced across covariates
 - Intuitive



STEPS IN PROPENSITY SCORE MATCHING

1. Determine propensity score

- Modeling the probability of receiving treatment (exposure) given a vector of selected covariates, usually using logistic regression
- The predicted value creates a propensity score for each person in the study
- Decision Point: What type of variables to include

2. Match persons with similar propensity score

- Decision Point: Select matching algorithm
- Decision Point: Determine matching ratio
- Decision Point: Determine if you have a treatment and control/donor population



STEPS IN PROPENSITY SCORE MATCHING

3. Check balance

- Decision Point: Type of balance statistics
- Decision Point: Decide if balance is adequate
- Decision Point: Decide how to adjust sample if balance is not adequate

4. Estimate effects

5. Sensitivity analysis



5. STEPS IN PROPENSITY SCORE MATCHING: SELECTING VARIABLES

- Selection of covariates is very important and should be given careful consideration
- Timing of covariate measurement is important
 - Typically include all *baseline* variables that might confound the relationship between treatment and outcome
 - Important that variables *caused* by the outcome (e.g., severity of illness after treatment) not be included in the model to avoid “over-controlling”
- Including variables that are correlated with the treatment but not with the outcome may reduce ability to match the group
- New methods for selecting covariates
 - *High dimensional propensity score analysis*: An algorithm that empirically identifies candidate covariates, prioritizes covariates, and integrates them into a propensity-score-based confounder adjustment model
 - *Open question*: Are more covariates better?



5. STEPS IN PROPENSITY SCORE MATCHING: CHOOSING MATCHING ALGORITHM AND MATCHING RATIOS

- Matching algorithms
 - Nearest neighbor matching
 - Radius matching
 - Kernel matching
 - Mahalanobis matching
- Matching with or without replacement
- Matching ratios
 - Generally get more efficiency with greater matching ratios but not much “bang for buck” after 1:3



5. STEPS IN PROPENSITY SCORE MATCHING: CHECK BALANCE

	Pre-Propensity Score		Post Propensity Score	
Pre-period Anxiety	8%	13%	8%	8%
Pre-period Substance Abuse	30%	31%	30%	30%
Pre-period Psych Related Hospitalization	42%	38%	42%	42%
Pre-period Psych Related ER	26%	25%	26%	26%



5. STEPS IN PROPENSITY SCORE MATCHING: BALANCE DIAGNOSTICS

- Significance testing
 - Influenced by sample size
 - May improve simply because matched sample is smaller than the unmatched sample

- Standardized difference
 - Not influenced by sample size

$$d = \frac{100 \times | \bar{x}_{treatment} - \bar{x}_{control} |}{\sqrt{\frac{s_{treatment}^2 + s_{control}^2}{2}}}$$

- Other methods
 - Comparisons of the variances for continuous measures
 - Plots of distributions



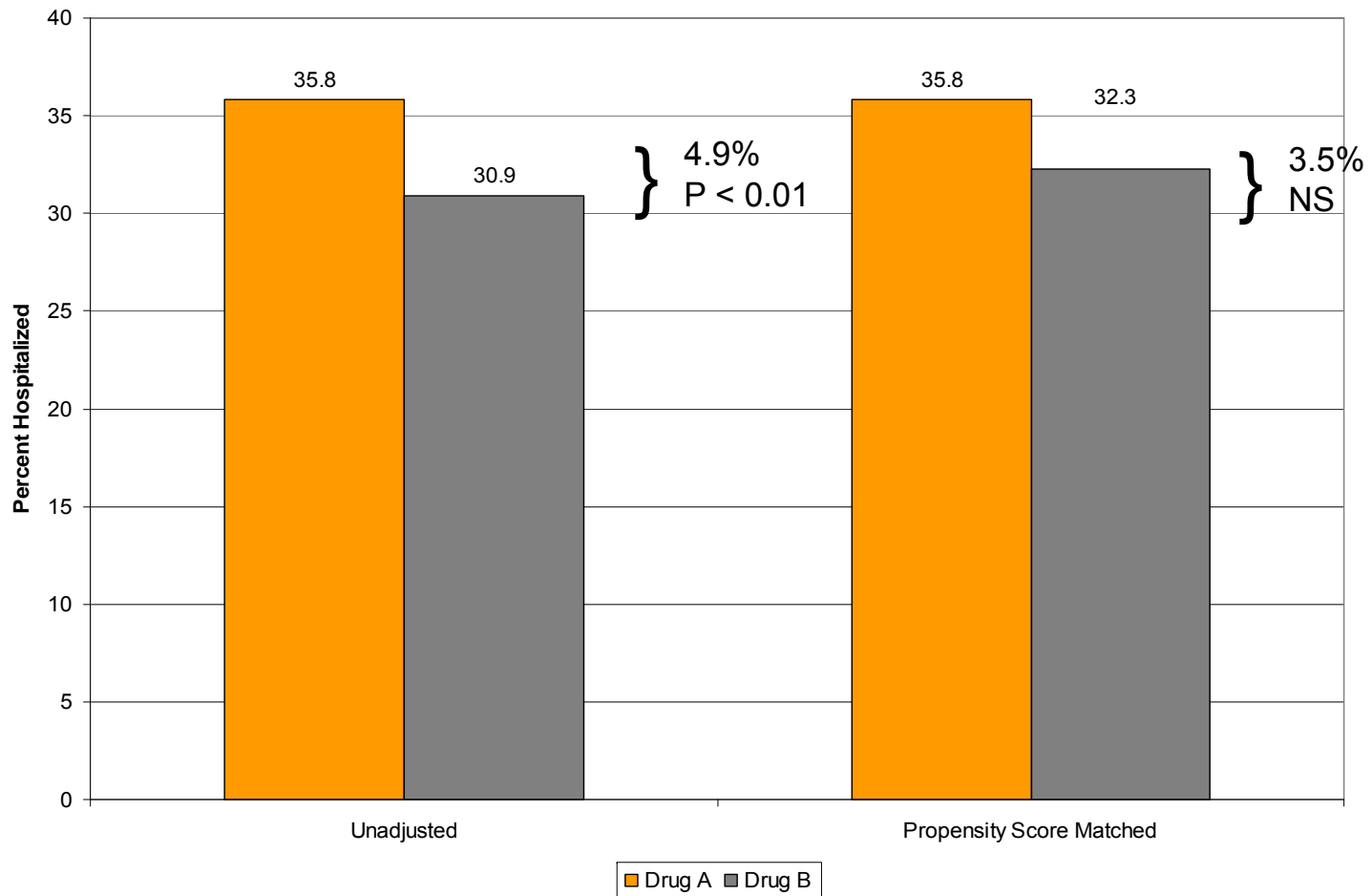
5. STEPS IN PROPENSITY SCORE MATCHING: BALANCE DIAGNOSTICS

- If Balance is not good may need to adjust matching algorithm such as by setting a narrower caliper
- May need to eliminate patients from the control or treatment group to obtain the best match
- Should be aware and document how sample has changed to obtain match



RESULTS AFTER PROPENSITY SCORE MATCHING

Percent Hospitalized: Unadjusted and Propensity Score Matched Samples



LIMITATIONS OF PROPENSITY SCORE MATCHING

- Limitations
 - Only balances on observable differences between cases and controls.
 - May balance on unobservables if they are correlated with observables (but unknown)
 - Bias may still be introduced by unobservable characteristics that are correlated with the outcomes and treatment selection
 - Matched samples may differ from unmatched in important ways influencing external validity
 - e.g., exclude most severe
 - Difficult to find matches when comparing on multiple interventions



ALTERNATIVE USES OF PROPENSITY SCORES

- Stratification: Use propensity score to create strata and compare groups across strata
 - Usually compare on 5 equally sized strata (quintiles)
 - Straightforward approach
- Regression: Use propensity score as covariate in regression model
 - Use with caution
 - Shown that if covariance matrices in the treated and untreated groups are unequal, bias may increase
- Inverse probability of treatment weighting



INVERSE PROBABILITY OF TREATMENT WEIGHTING: SIMPLE EXAMPLE

Suppose we have the following distribution of gender:

	Male	Female	Total
Treatment	40	60	100
Control	75	25	100
Total	115	85	200

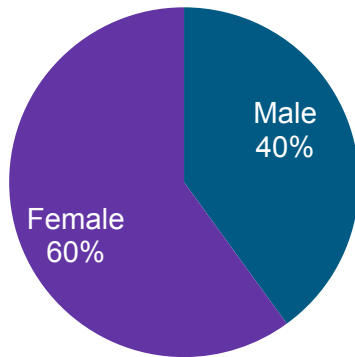
IPTW Weights

	Male	Female
Treatment	$(40+75)/40$	$(60+25)/60$
Control	$(40+75)/75$	$(60+25)/25$



SIMPLE EXAMPLE

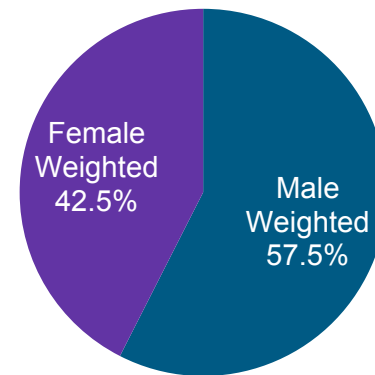
Gender Distribution in Treatment Group



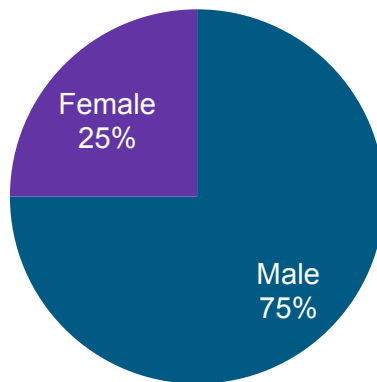
Weight for male
 $=1/(40/(40+75))=2.875$

Weight for female
 $=1/(60/(60+25))=1.417$

Weighted Gender Distribution in Treatment Group



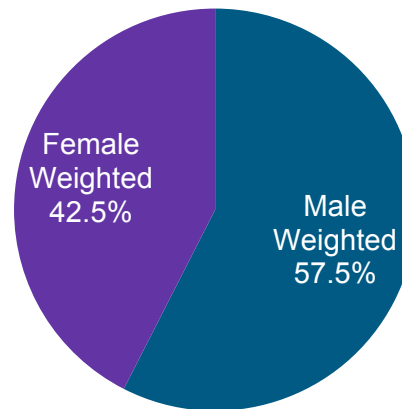
Gender Distribution in Control Group



Weight for male
 $=1/(75/(40+75))=1.533$

Weight for female
 $=1/(25/(60+25))=3.4$

Gender Distribution in Control Group



HOW TO OBTAIN IPTW

- A discrete choice model (e.g. probit, logit, multinomial logit) is implemented to obtain the estimated probability of being in the treatment groups for each individual, given the confounding factors
- The inverse of the predicted probability is applied as probability weights in the subsequent analysis



PROPENSITY SCORE MATCHING VS. WEIGHTING

- Imbens (2004):
 - IPTW – Average treatment effect
 - Matching – Average treatment effect among the treated
- Matching: Based on sub-sample that matched
- Weighting: Can be based on sample or sub-sample
- Differing results may be due to different treated and control populations included



AREAS FOR FUTURE RESEARCH

- How to select covariates for propensity score models?
- How to tell how much imbalance is too much?
 - And how to address imbalances
- How to select the correct matching method?
- Which propensity method is best (stratification, regression, IPTW, matching)?



6. INSTRUMENTAL VARIABLES

- Advantage is that have the potential to eliminate bias from observable and unobservable covariates
- First stage, regress treatment variable on all variables related to treatment or the outcome
- Next, use the regression estimates to construct a predicted value for the treatment variable (endogenous regressor)
- Substitute this predicted value for the actual value of the treatment in the outcome equation
- By using predicted rather than actual values of treatment, you are “breaking” its correlation with the error term, so you get an unbiased estimate of treatment effect



INSTRUMENTAL VARIABLE (IV) EXCLUSION RESTRICTION

- To “identify” the effect of treatment on outcome, we need at least one variable that affects receipt of treatment, but does not directly influence outcomes
- These “instruments” generate variation in the treatment variable that is not related to the outcome (similar to flipping a coin in random assignment)
- Correlation of the IV with the treatment (endogenous regressor) must be high *relative* to its correlation with the outcome
- Weak instruments may do more harm than good



EXAMPLES OF POTENTIAL IV INSTRUMENTS FOR CER

- Distance to a hospital
- Benefit design
 - Formulary restrictions
 - Co-payment levels
- Provider preferences
 - Percent of particular types of antidepressants by a physician
 - Percent of people in county with schizophrenia who received a particular medication
 - Regional catheterization rates



EXAMPLE: INSTRUMENTAL VARIABLE BASED ON PERCENT IN COUNTY PRESCRIBED DRUG

Correlation Between Instrumental Variable and Treatment and Outcome Variable

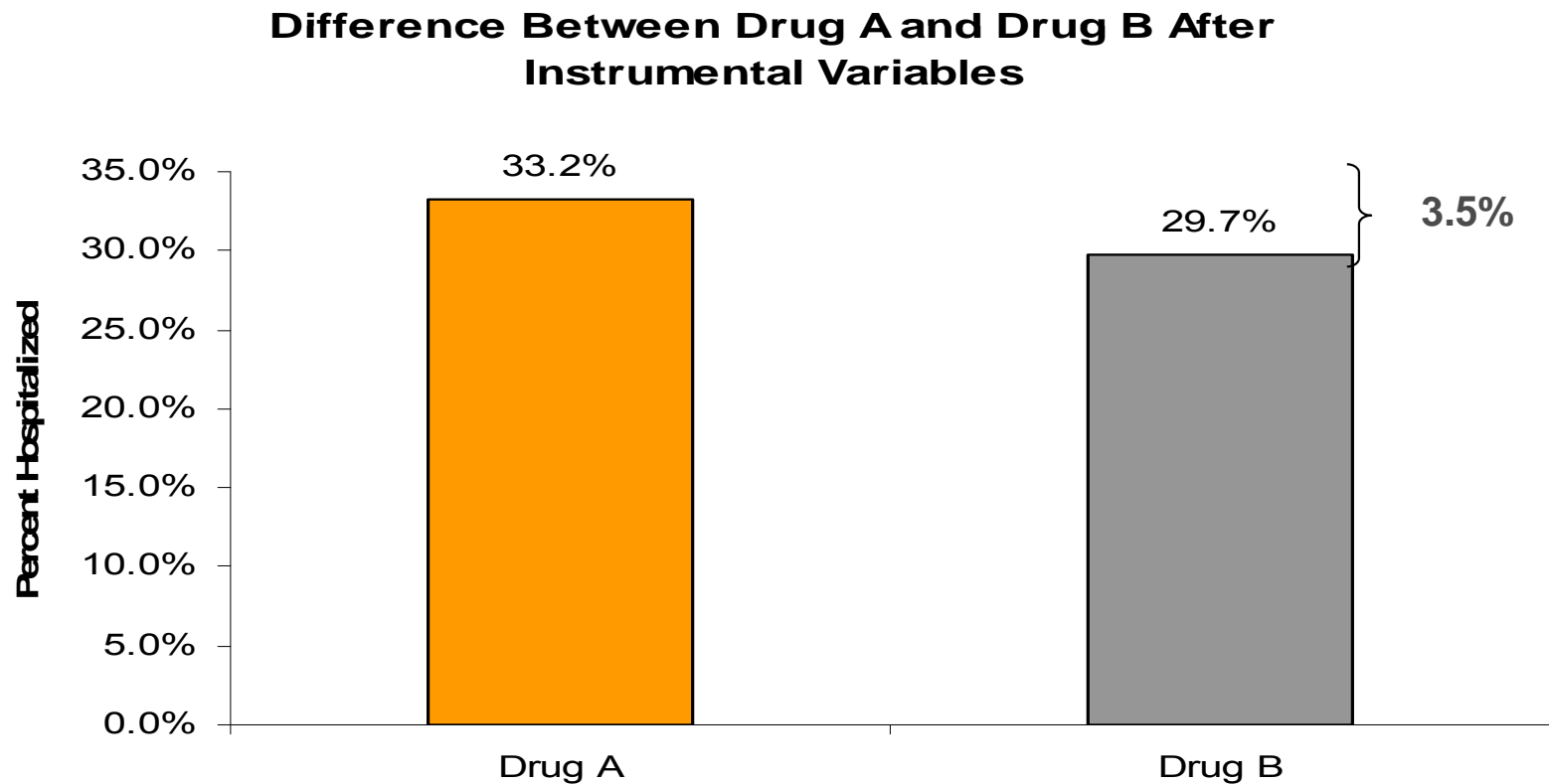
	Use Drug A	% Hospitalized
Percent Prescribed Drug A in County	0.245	-0.0093

- F-test to test whether IV are associated with treatment
- Over-restriction test to see if IV are not associated with outcomes



RESULTS AFTER INSTRUMENTAL VARIABLES

- As with propensity score, 3.5% difference between between Drug A and Drug B which is not statistically significant



LIMITATIONS OF INSTRUMENTAL VARIABLES

- Instrumental variables were established for linear variables
- Creating instruments for binary variables is a relatively new development
- Determining if instrumental variable is truly exogenous and strongly enough related to outcome is very difficult
- Reviewers may be skeptical of the “quality” of your instruments



7. DIFFERENCE-IN-DIFFERENCE (DID) METHODS

- Also known as “interrupted time series”
- General approach
 - Change from pre- to post-treatment in Group 1
 - Change from pre- to post-treatment in Group 2
 - Difference in change Group 1 from change in Group 2
- Notes
 - Focus of analysis is not on level of outcome but change
 - Used more typically to evaluate policy changes than to compare treatments



7. DIFFERENCE-IN-DIFFERENCE (DID) METHODS, CONT'D

- Example
 - Hospitals that implemented safety controls experienced a larger decline in adverse event rates as compared to hospitals that did not implement controls
- Advantages
 - Controls for unobservables such as trends in declining adverse events due to improved technology that affects both groups
- Limitation
 - Assumes two groups will exhibit the same trend over time in the absence of treatment. May be problematic if baseline is very different between two groups.
- Balancing act
 - Pre-period matching vs. generalizability
- Extensions
 - Panel data or multiple per observations per person
 - Propensity score matching at baseline



SUMMARY OF OBSERVATIONAL METHODS

- Many ways to control for confounding through design and statistics
- It's important to use the approaches thoughtfully
- All techniques still rely on having enough of the right information



III. TRANSPARENCY

- Lies, damn lies, and statistics
- Debate about how to increase trust in observational data methods
- Suggested approaches
 - Peer review
 - Sensitivity analyses
 - Conflict of interest statements
 - Pre-specified analysis plan
 - Making data available to non-authors



IV: ENRICH SECONDARY DATA BASES THROUGH LINKAGES

- Add outcome measures
- Add information that can control confounding

TYPES OF DATA ELEMENTS

- Claims: Diagnoses (ICD-9-CM), procedures (CPT/HCPC), utilization (hospitalizations, adherence), medications received
- BMI (weight and height)
- Vital signs (BP, temperature)
- Health behaviors (smoking, exercise, alcohol consumption)
- Standardized symptoms assessment (e.g., depression, pain)
- Death
- Laboratory tests (e.g., HbA1c levels, creatinine)
- EKG
- Pathology data
- Radiology
- Genomic data
- Nurse observations



ENHANCING MEASURES ON SECONDARY DATA THROUGH LINKAGES

1. Ongoing surveys + claims data
2. Absenteeism data + claims data
3. Laboratory data + claims data
4. Primary data collection + claims data
5. Registry data + claims data
6. Electronic medical records + claims data



CLAIMS DATABASES: E.G., MARKETSCAN DATABASES

- Proprietary U.S. databases providing individual-level, de-identified, healthcare claims fully integrated across care settings
- Care settings: Inpatient, emergency room, outpatient, and retail drug
- Sources: Commercial, insurance Medicare supplemental, insurance, and Medicaid populations
- Covered lives: About 30 million lives under age 65
- Cohort size: E.g., about 1 million people can be followed for 10 years



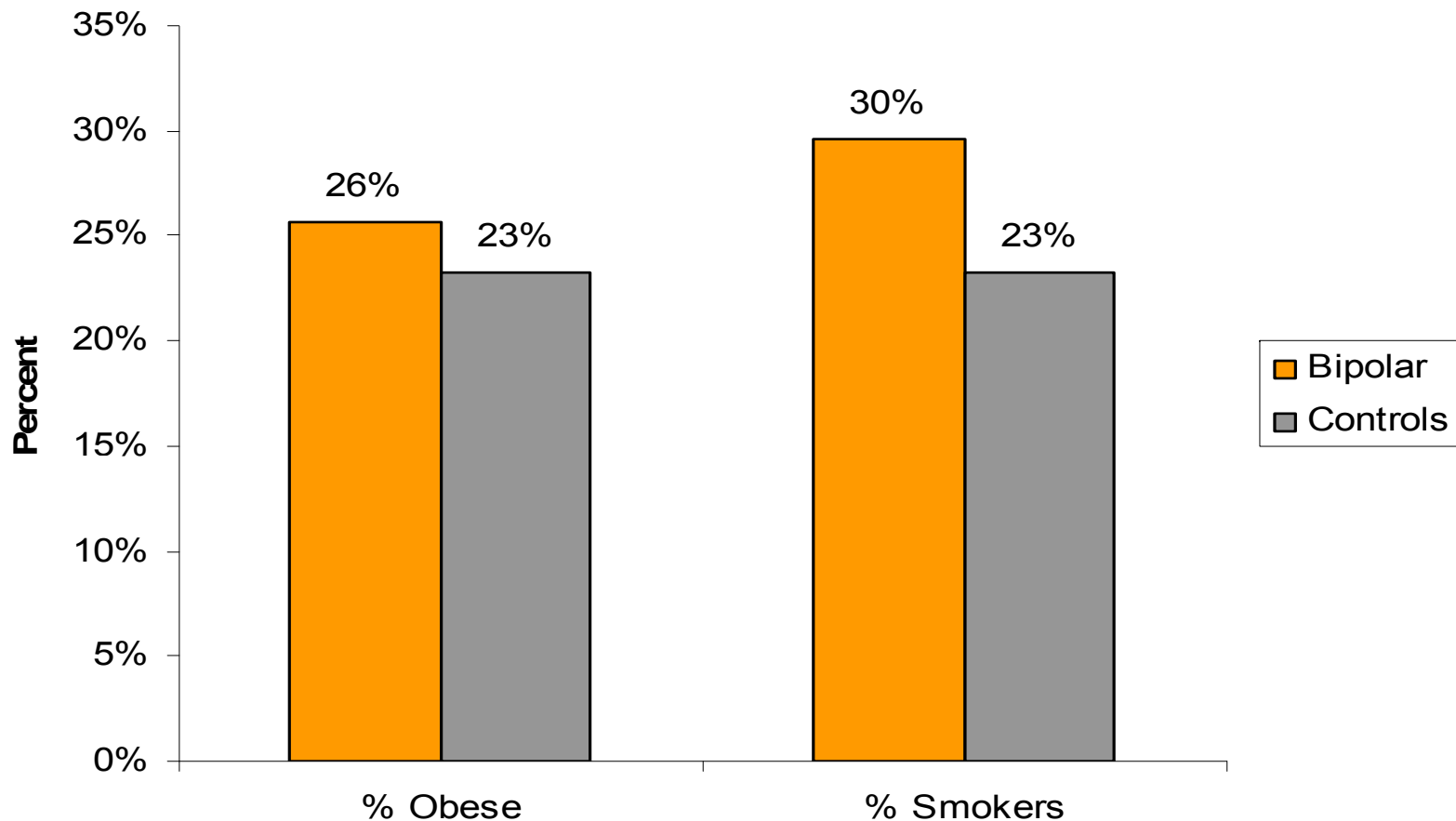
1. ONGOING SURVEYS + CLAIMS (HEALTH RISK ASSESSMENTS)

- Completed annually by employees of participating employers
- Types of variables
 - Height, weight
 - Health behaviors: smoking, drinking
 - Blood pressure, cholesterol
 - Sleep
 - Stress
- Uses of HRA data
 - Outcomes
 - e.g., effect of antipsychotics on weight
 - Controls
 - e.g., effect of antipsychotics on stroke controlling for smoking status



HYPOTHETICAL HRA EXAMPLE: IMPORTANT COVARIATES FOR SAFETY STUDIES OF ANTIPSYCHOTICS AND HEART DISEASE

Bipolar Vs. Controls: % Obese and % Smokers

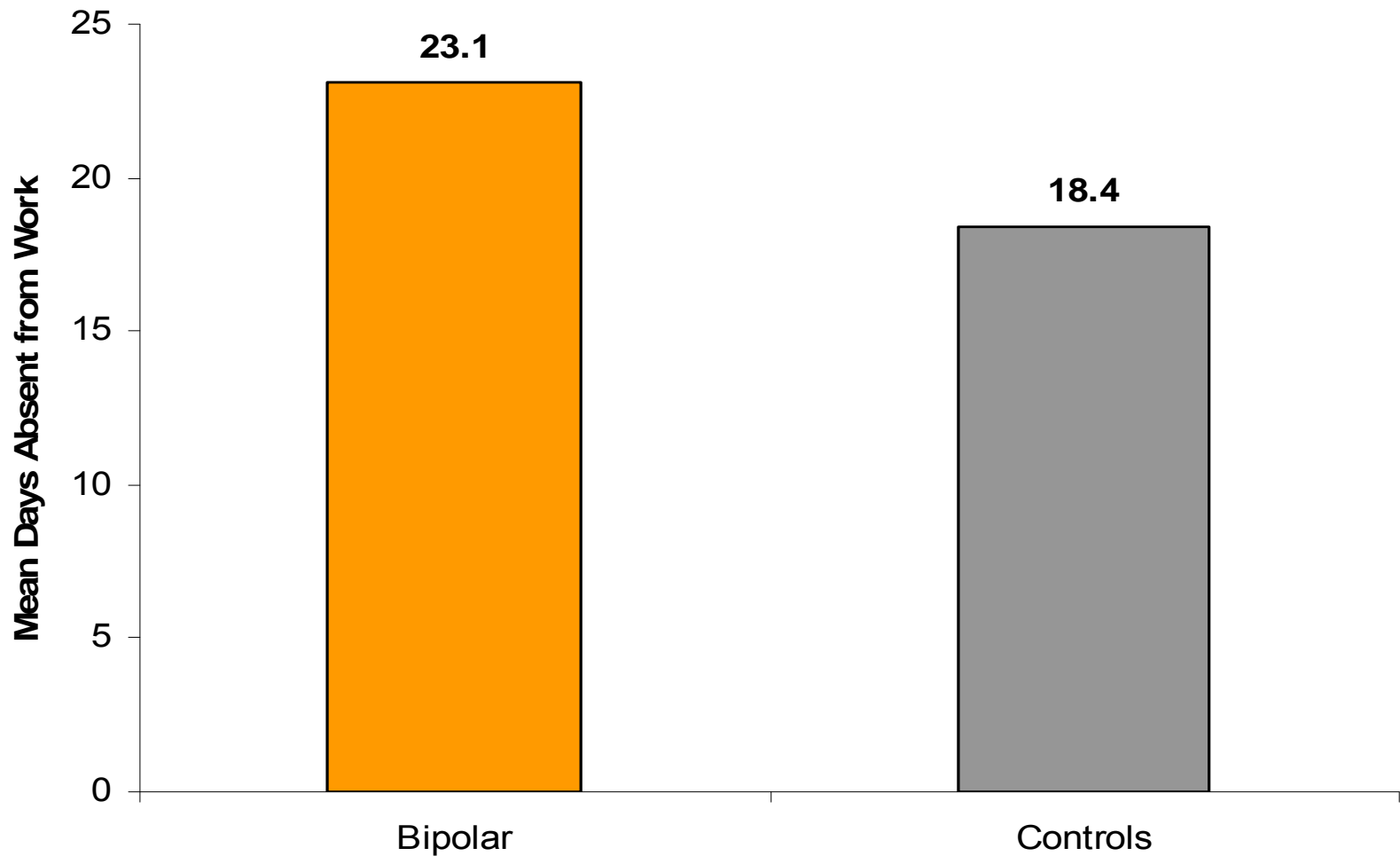


2. ABSENTEEISM DATA + CLAIMS (HEALTH AND PRODUCTIVITY DATA)

- Absenteeism, short-term disability, and workers' compensation data
- Allows one to go beyond medical outcome to functioning in workplace



EXAMPLE: LINKED HPM DATA ON WORKPLACE ABSENCE AMONG BIPOLAR PATIENTS VS. CONTROLS

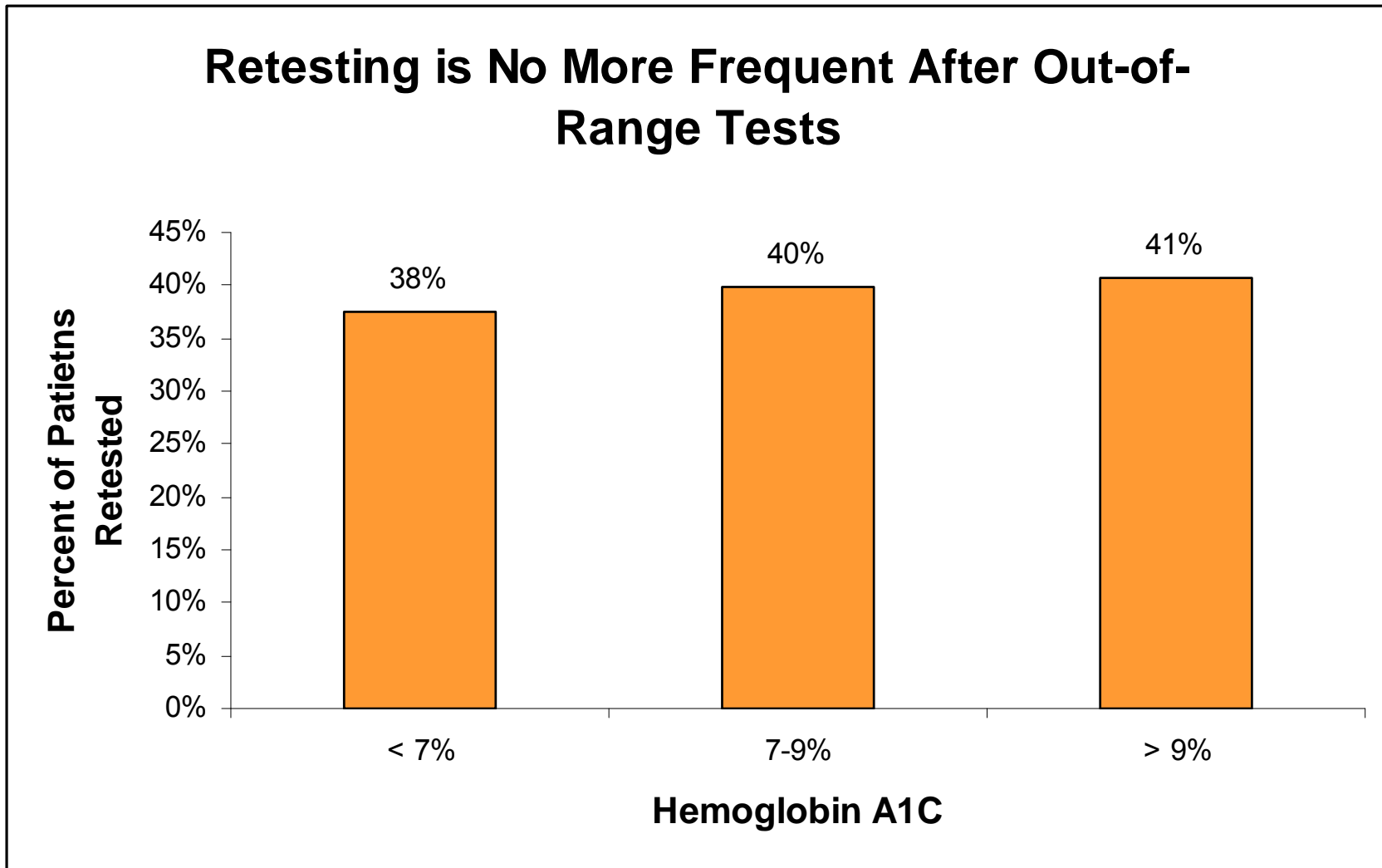


3. LABORATORY DATA + CLAIMS

- Hematocrit
- Hemoglobin
- Cholesterol
- Triglycerides
- HbA1C
- CD4 cell count
- Thyroid tests
- Liver function tests



EXAMPLE: LABORATORY DATA



4. PRIMARY DATA COLLECTION + CLAIMS DATA

- The *Child and Household Influenza-Illness and Employee Function Study (CHIEF)*
 - Survey of employees about examine how household member illness – specifically child flu – impacts employee productivity
 - Primary data collection on whether any of the household members experienced the flu each month; symptoms, work absence, presenteeism, vaccination
 - Claims data to provide information on healthcare resource utilization and costs



PRIMARY DATA COLLECTION + CLAIMS DATA

- Medicare Current Beneficiary Survey (MCBS)
 - A continuous, multipurpose survey of a nationally representative sample of aged, disabled, and institutionalized Medicare beneficiaries
 - Sponsored by the Centers for Medicare & Medicaid Services
 - Link Medicare claims to survey-reported events
- Nursing Home Minimum Data Set (MDS) linked to Medicare claims data
 - MDS is a standardized, clinically-based instrument used to assess residents in nursing homes
 - Collects data on demographic, functional, medical, psychological, and cognitive status



5. REGISTRY + CLAIMS DATA

- NCI Seer-Medicare Linked Database
 - Surveillance, Epidemiology and End Results (SEER) data
 - Cancer registries that collect clinical, demographic, and cause of death information for persons with cancer
 - Linked to Medicare claims
- US Renal Data System
 - Data reported by ESRD providers linked to Medicare claims
- National Cardiovascular Data Registry (NCDR®)
 - Initiative of the American College of Cardiology Foundation®, began in 1997 to document processes and outcomes of care in the cath lab setting
 - Registry linked to Medicare claims



6. EMR + CLAIMS DATA

- General practice EMR and claims data such as *MarketScan* claims linked to GE EMR data



OBSTACLES TO LINKAGES

- Privacy
- Incentives
- Standard data elements
- Population overlap



V. VALIDATION OF MEASURES ON OBSERVATIONAL DATA

- Increasing attention to validity of observational data, such as claims data
- In safety research, validation has focused on looking back to medical records
- Increasing efforts to develop standardized algorithms to identify health outcome (e.g. OMOP)



VI. OTHER “CRITERIA” FOR DRAWING CAUSAL INFERENCES

- Examples of causal “criteria”:
 - Sir Austin Bradford Hill (1965)
 - Naranjo criteria
- No well established “check-list”



EXAMPLES OF FACTORS FOR DRAWING CAUSAL INFERENCES

- **Consistency:** Repeated observation of an association in different studies, under different conditions
- **Specificity:** When effect is specific to one outcome but not another
 - e.g., raising copayments on statins lowers adherence to statins but not other medications
- **Temporality:** Cause precedes outcome, outcome occurs immediately after exposure
- **Biological Gradient:** Dose-response, exposure-response
- **Plausibility:** Scientific plausibility of an association
- **Experimental:** Eliminate causal agent to see if frequency of outcome declines



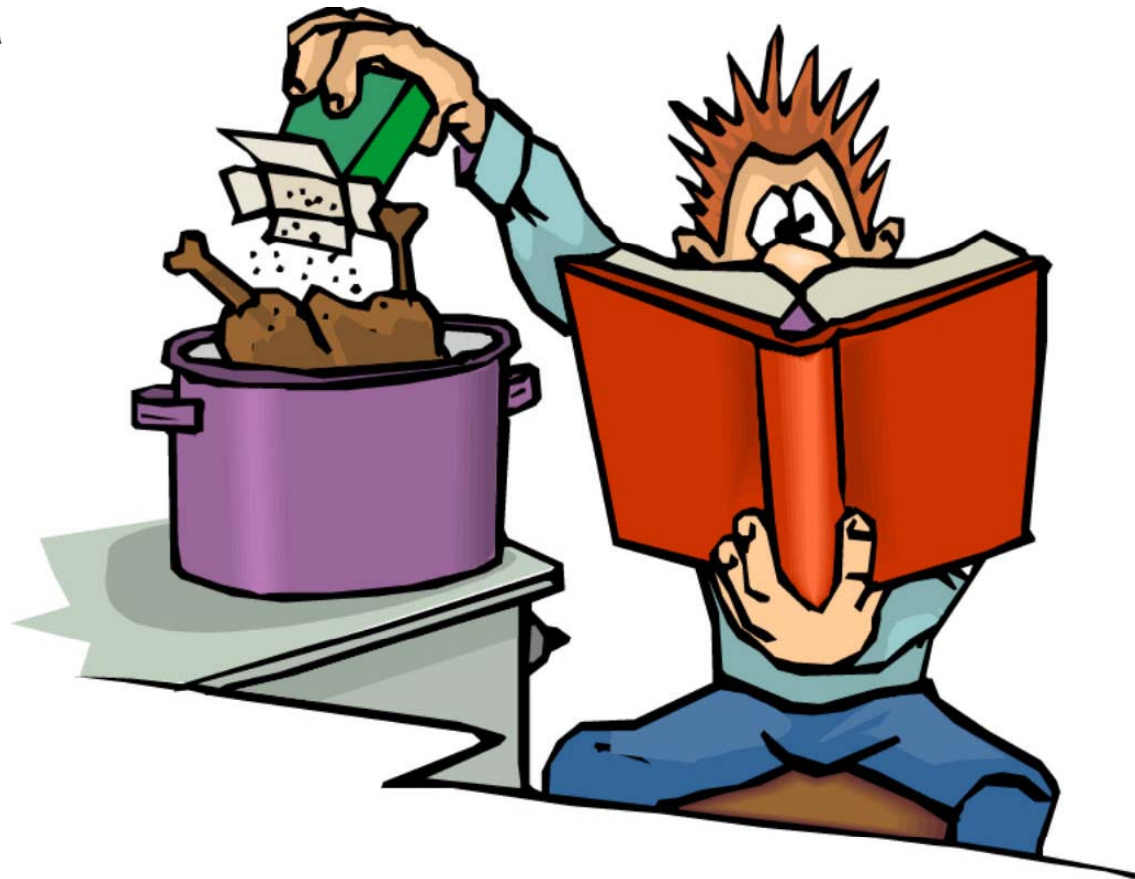
CONCLUSIONS

- Non-experimental observational data:
 - Challenges of using it for CER are not trivial
 - Opportunities for adding to comparative knowledge are great
- Confounding may be addressed through:
 - Research designs
 - Statistical techniques
 - Richer data sources
- Electronic data:
 - Are growing in size and clinical richness
 - Will enrich outcome measures
 - Aid in addressing confounding



“Overcoming the limitations of observational research is the most important frontier of research on study methods,” Harold C. Sox and Sheldon Greenfield

- We don't have a cookbook yet for conducting comparative effectiveness research with observational data



FOR MORE INFORMATION

Contact:

Tami.Mark@thomsonreuters.com



REFERENCES

- Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine*. 27: 2037-2049.
- Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med*. 2009 Nov 10;28(25):3083-107.
- Baser O. Too much ado about propensity score models? Comparing methods of propensity score matching. *Value Health*. 2006 Nov-Dec;9(6):377-85.
- Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*. 1995; 90 (430):443-450.
- D'Agostino RB. Propensity score methods for bias reduction in the comparison of the treatment to a non-randomized control group. *Statistics in Medicine*. 1998; 17: 2265-2281.
- Fu AZ, Dow WH, Liu GG. Propensity score and difference-in-difference methods: A study of second-generation antidepressant use in patients with bipolar disorder. *Health Serv Outcomes Res Method*. 2007



REFERENCES

- Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: Instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *Clin Epidemiol*. 2009 Apr 6.
- Ray WA. Evaluating medication effects outside of clinical trials: Newuser designs. *Am J Epidemiol*. 2003;158:915–920.
- Rosenbaum PR, Rubin DB. The central role of propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41-55
- Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009 Jul;20(4):512-22. PubMed PMID: 19487948

