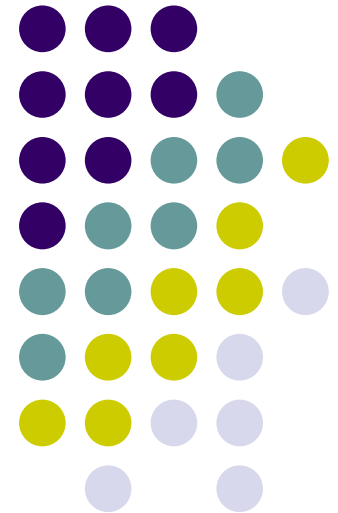
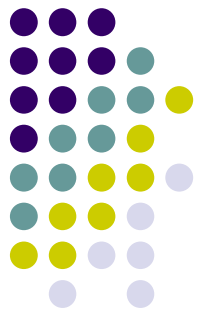


1. Comparative effectiveness

2. Propensity Score Method

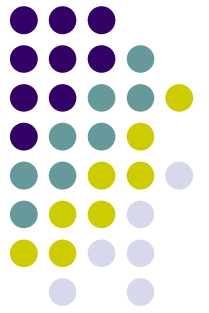
K. Arnold Chan 陳建煒, MD, ScD, FISPE
Harvard School of Public Health
i3 Drug Safety





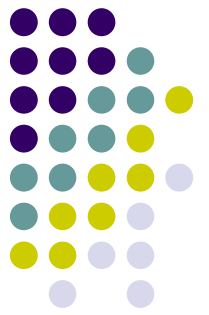
Potential Conflict of Interest

- Part time employee of i3 Drug Safety
- Adjunct Associate Professor, Harvard School of Public Health
- Co-editor of a book, no royalty received
- Public health worker



Outline

- Comparative effectiveness in the evaluation of medical products
 - History and some examples
 - Recent development
- Using propensity score to control for confounding in observational studies
- Q & A throughout

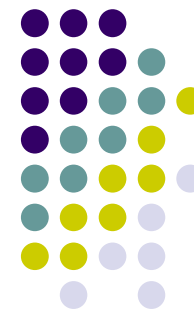



Some terms, old and new

- Patient Outcomes Research Team (PORT)
- Comparative effectiveness
- Patient-centered research

- A lot of U.S. politic involved

<http://www.hhs.gov/recovery/programs/os/cerbios.html>



 Friday Oct 22, 2010

Print

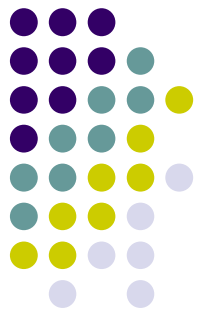


[Home](#) | [Overview](#) | [Programs](#) | [Plans & Reports](#) | [Grants & Contracts](#) | [Announcements](#) |

[HHS Home](#) > [Recovery](#) > Programs

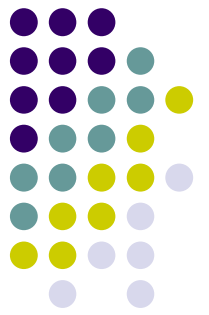
Federal Coordinating Council for Comparative Effectiveness Research Membership

Recovery Act Allocates \$1.1 Billion for Comparative Effectiveness Research



Comparative effectiveness

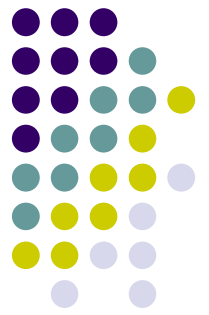
- The scientific and public health goal: move beyond the FDA placebo-controlled-trial paradigm
 - Inform clinical practice to improve quality of care
 - Optimal allocation of limited health care resources
 - Consider all treatment modalities
 - More than Drug A vs. Drug B head-to-head comparison
 - Relevant data for important subgroups



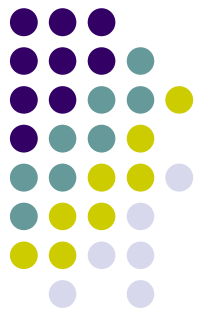
FDA approval requirement

- The general principle is placebo-controlled trials for non-life-threatening diseases
- Legal basis and years of guidelines development
- Not necessarily relevant for clinical practice
- Major disagreement with the Declaration of Helsinki revision

Effectiveness in real-life practice



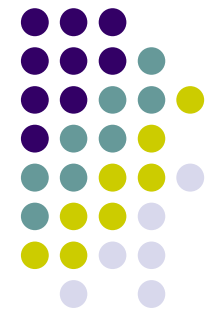
- A better than placebo, B better than placebo, C better than placebo, D better than placebo, ...Are they the same?
- Who wants to know the answer?
 - Clinicians / specialty groups
 - Organizations that develop treatment guidelines
 - Health insurance companies
 - Some manufacturers



Who should fund these studies?

- Government
 - U.S. National Institute of Health
 - U.S. Agency for Healthcare Research and Quality
- Manufacturers of medical products
- Health insurance agencies / companies (may be)

Treatment for Age-Related Macular Degeneration



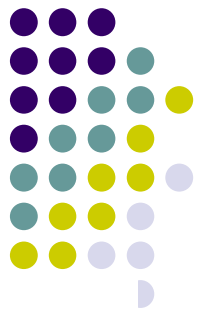
- Bevacizumab vs. Ranibizumab

Comparison of Age-related Macular Degeneration Treatments Trials: Lucentis-Avastin Trial

This study is ongoing, but not recruiting participants.

First Received: January 3, 2008 Last Updated: March 23, 2010 [History of Changes](#)

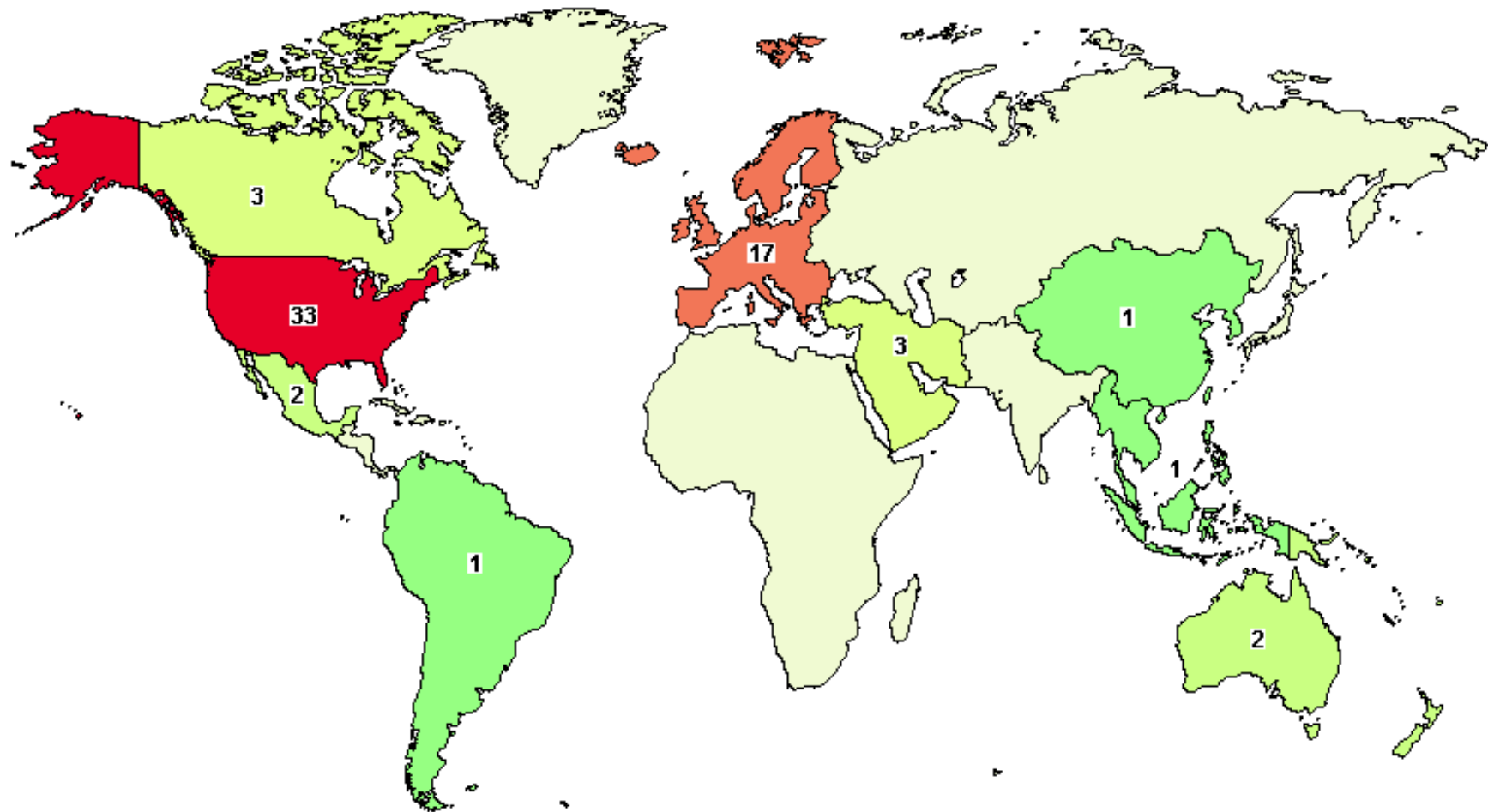
Sponsor:	National Eye Institute (NEI)
Information provided by:	National Eye Institute (NEI)
ClinicalTrials.gov Identifier:	NCT00593450

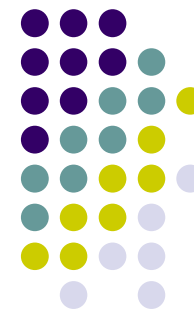


From clinicaltrials.gov

Map of 58 studies found by search of: bevacizumab and ranibizumab

Click on the map below to show a more detailed map (when available) or search for studies (when map not available).

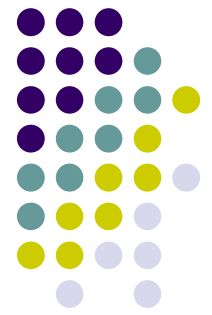




The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial -- ALLHAT (allhat.org)

- “The ALLHAT study was the largest trial for treating high blood pressure, and the second largest for lipid-lowering treatment. It included large numbers of patients over 65, women, African-Americans, and patients with diabetes. The trial was conducted in over 600 office-based practices and clinics throughout the U.S.A., Puerto Rico, the Virgin Islands, and in Canada.”
- “The purpose of ALLHAT was to compare four commonly used blood pressure drugs to reduce the risk of heart disease, stroke, and early death. Also, some of the ALLHAT patients were in a lipid-lowering trial to treat people with high blood pressure and mildly high cholesterol to see if this would further reduce these risks.”

Head-to-head comparison trial sponsored by a drug company PROVE-IT



- Cannon et al. N Engl J Med 2004; 350: 1495-1504 (also known as the TIMI-22 study)
- High dose atorvastatin (80mg/day) vs. standard dose pravastatin (40mg/day) for patients after acute coronary syndrome
- Results: the lower the LDL the better



Methods and objectives

Randomized trial

Observational study

Benefit

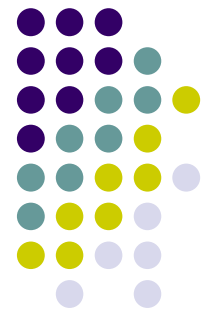
???

Risk

Safety trial

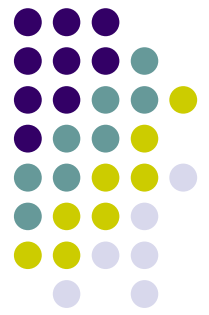
Simplified Controlled
Trial

Comparative Effectiveness in the U.S.



- Comparison of different treatment modality
- Evaluate treatment effects in important subgroups
 - Non-Caucasians
 - Older patients
 - Very young patients
 - Others
- The term has been shifted to Patient-Centered Research

<http://thehill.com/blogs/healthwatch/health-reform-implementation/118913-white-house-releases-funds-for-patient-centered-research-among-minorities>

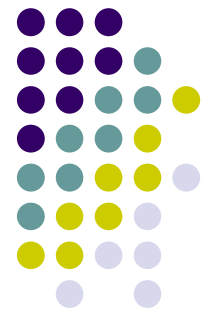


White House releases funds for patient-centered research among minorities

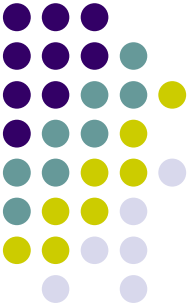
By Mike Lillis - 09/15/10 11:27 AM ET

The Department of Health and Human Services (HHS) on Wednesday released \$14 million to bolster programs designed to identify which healthcare treatments work best on ethnic minorities.

Methods for comparative effectiveness

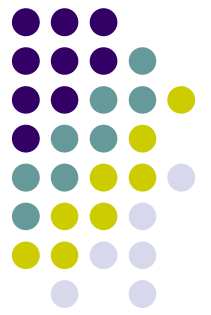


- Traditional randomized controlled trial
- Simplified controlled trial
 - Small number of collected data elements
 - Potential utilization of health insurance claims and electronic health records
- Cluster randomized trial
- Observation method
 - Very challenging because of confounding by indication

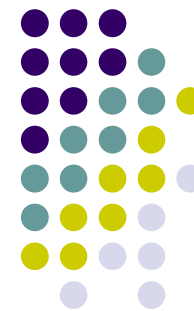


Questions?

References for propensity score methods



- Rubin. Estimating causal effects from large datasets using propensity scores. *Annals of Internal Medicine* 1997; 127: 757-63.
- Kurth & Seeger. Propensity score analyses in pharmacoepidemiology (Chapter 12). In: Hartzema, Tilson, & Chan (eds). *Pharmacoepidemiology and Therapeutic Risk Management*. Harvey Whitney Books, Cincinnati, OH, USA, 2008, page 301-324.



Control for confounding

- Is observational study always inferior to clinical trials?
- *Statistics in Medicine* 1983; 2: 267-271

Article

The need for randomization in the study of intended effects

Olli S. Miettinen

Departments of Epidemiology and Biostatistics, School of Public Health, Harvard University, Boston, Massachusetts 02115, U.S.A.

Funded by:

- U.S. National Cancer Institute; Grant Number: 5 P 01 CA 06373

KEYWORDS

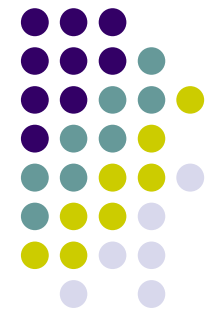
Randomization ▪ Confounding ▪ Efficacy of treatment ▪ Toxicity of treatment

ABSTRACT

The need for randomization as a means of controlling confounders is accentuated in the study of intended effects (efficacy) as compared with unintended ones (toxicity). The basic reason is that the indication for intervention is inherently a confounder in the study of efficacy but not of toxicity, whereas contraindications represent only a minor confounder even in toxicity research. Moreover, control of the indication in non-experimental terms is commonly infeasible owing to the complexity and subtlety of the indication.

Received: September 1982

Intended effects vs. Un-intended effects



Intended
(potential for confounding
by indication)

Un-intended

Benefit

Clinical trials
Observational
Studies (??)

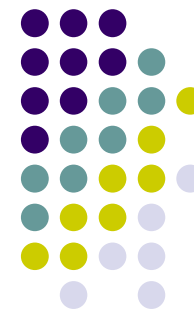
Observational studies
Clinical trials

Risk

Large Simple Trials
Observational studies

Observational studies

Jan P Vandenbroucke



- Lancet 2004; 163: 1728-31

VIEWPOINT

When are observational studies as credible as randomised trials?

- PLoS Medicine 2008; 5: e67

OPEN  ACCESS Freely available online

PLoS MEDICINE

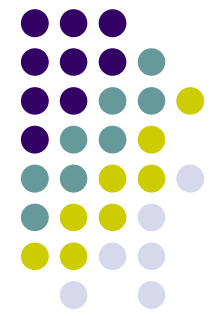
Essay

Observational Research, Randomised Trials, and Two Views of Medical Science

Jan P. Vandenbroucke

Vandenbroucke. PLoS 2008

Rethinking the hierarchy of evidence

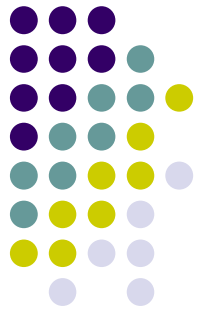


Box 1. Hierarchy of Study Designs for Intended Effects of Therapy

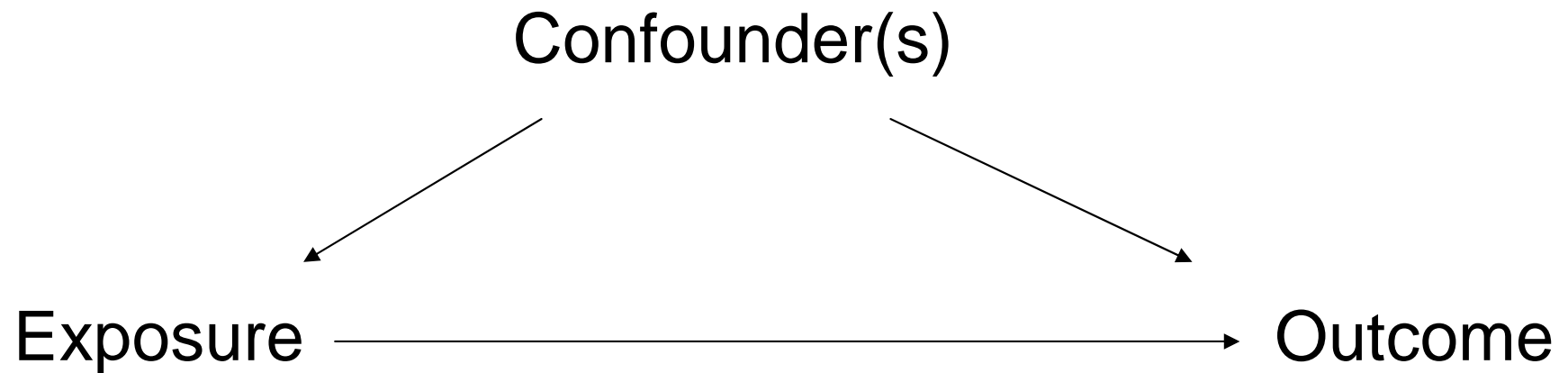
1. Randomised controlled trials
2. Prospective follow-up studies
3. Retrospective follow-up studies
4. Case-control studies
5. Anecdotal: case report and series

Box 2. Hierarchy of Study Designs for Discovery and Explanation

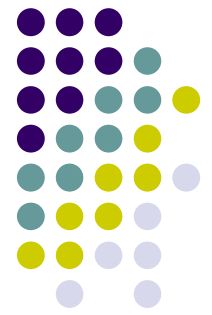
1. Anecdotal: case reports and series, findings in data, literature
2. Case-control studies
3. Retrospective follow-up studies
4. Prospective follow-up studies
5. Randomised controlled trials



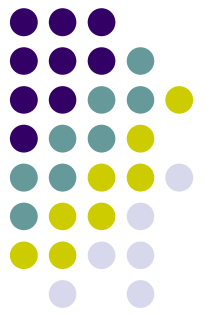
Review of confounding



Case-by-case evaluation of confounding

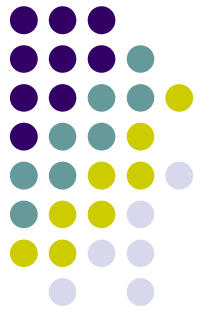


- Unintended safety outcomes
 - Antibiotic and acute liver failure
- Anticipated safety outcomes
 - Oral hypoglycemic agents and coronary heart disease
- Anticipated beneficial outcomes
 - Statin use and acute myocardial infarction



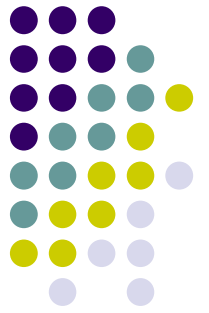
Examples of confounding in pharmacoepidemiology

- Confounding by indication
 - H₂ blockers, upper abdominal discomfort, and gastric cancer
 - Oral hypoglycemic agent, cardiovascular risk factors, and ischemic heart disease
- An informative example
 - High dose beta agonist, asthma severity, and asthma death



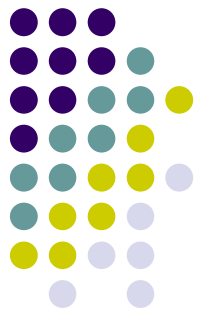
Confounding

- Prevention through study design
 - Randomization
 - Matching
 - Restriction
- Control through statistical analysis
 - Stratified analysis
 - Multiple regression
 - Propensity score method
 - Instrumental variable
 - ...



Confounders

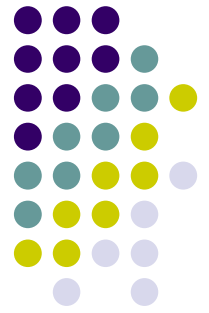
- A study of thiazolidenediones and coronary heart disease based on large linked health insurance database as an example
- Measured confounder
 - Cardiovascular disease history
- Unmeasured confounder
 - Smoking and Body Mass Index information are not available in the database
 - Solution: indirect control through measured confounders, validation through a subset of subjects, sensitivity analysis
- Unknown and unmeasured confounders
 - Randomization in clinical trial



Propensity score methods

- The goal is to estimate the probability of drug use
 - Could be based on logistic regression or any other methods to predict a binary outcome
- Not necessarily more valid than multiple regression methods
- The advantage is transparency and easy interpretation

The problem with multiple regression



- Statistical assumptions in regression models are not always met
- Insufficient information provided in reports in most medical and public health journals
- Which one of the adjusted RR and its 95% confidence interval to use?
- Propensity score method is not necessarily a 'better' method with regards to validity, but the process is transparent and easily interpretable.

Arch Intern Med 2002; 162: 936-42.

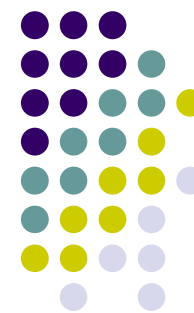


ORIGINAL INVESTIGATION

Gastric and Duodenal Safety of Daily Alendronate

James G. Donahue, DVM, PhD; K. Arnold Chan, MD, ScD; Susan E. Andrade, ScD; Arne Beck, PhD; Myde Boles, PhD; Diana S. M. Buist, PhD; Vincent J. Carey, PhD; Julie M. Chandler, PhD; Gary A. Chase, PhD; Bruce Ettinger, MD; Paul Fishman, PhD; Michael Goodman, PhD; Harry A. Guess, MD, PhD; Jerry H. Gurwitz, MD; Andrea Z. LaCroix, PhD; T. R. Levin, MD; Richard Platt, MD, MS

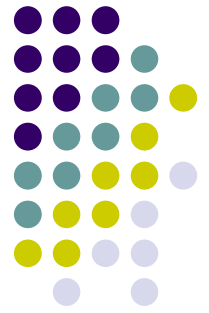
- Comparing alendronate users and unexposed
 - Crude RR = 3.0, 95% confidence interval 1.6 – 5.5
 - Adjusted RR = 1.8, 95% confidence interval 0.8 – 3.9
 - How did that happen?



The data structure and terminology

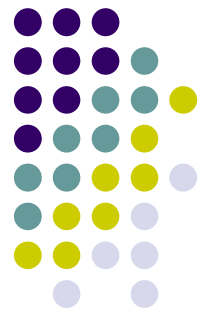
- Exposure: X
 - Binary exposure: drug A and drug B
- Outcome: Y
 - Binary outcome: first acute myocardial infarction (MI) vs. no MI
- Covariates: $C_1, C_2, C_3, C_4, \dots$
 - Hypertension, Diabetes, Use of nitrates, Use of calcium channel blockers, ...
- Multiple regression approach
 - $g(Y) = f(X, C_1, C_2, C_3, C_4, \dots)$
 - g and f corresponds to the statistical model (logistic, Poisson, or Cox model)
 - Estimated coefficient for X can be transformed to the RR estimate

Construction of a propensity score model

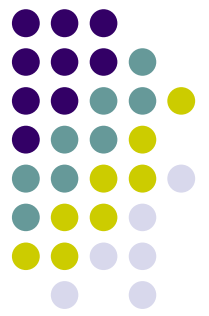


- Evaluate the relationship between C_i and X with a logistic regression model (X is binary, a subject received either A or B)
 - What determine the choice of therapy (A vs. B) among the study subjects?
 - $\text{Logit}(X) = \exp(a_1C_1 + a_2C_2 + a_3C_3 + a_4C_4 + \dots)$
 - Y (outcome) is not involved in the equation
- Estimate $a_1, a_2, a_3, a_4, \dots$
- For each study subject, estimate the probability of receiving A through the regression equation and covariate values.
 - The estimated propensity score, a value between 0 and 1

Rubin's suggestion for propensity score analysis



- Rank the subjects according to estimated propensity score
- Stratify the subjects into groups according to the estimated propensity score (e.g. into quintiles or deciles)
 - Those who are in the same stratum have similar probability of receiving the same drug, confounding by $C_1, C_2, C_3, C_4, \dots$ is minimized
- Direct evaluation of X and Y (crude RR) in each stratum
- Evaluate effect modification and combine the RRs through the Mantel-Haenszel method in the absence of effect modification



Propensity score method and matching

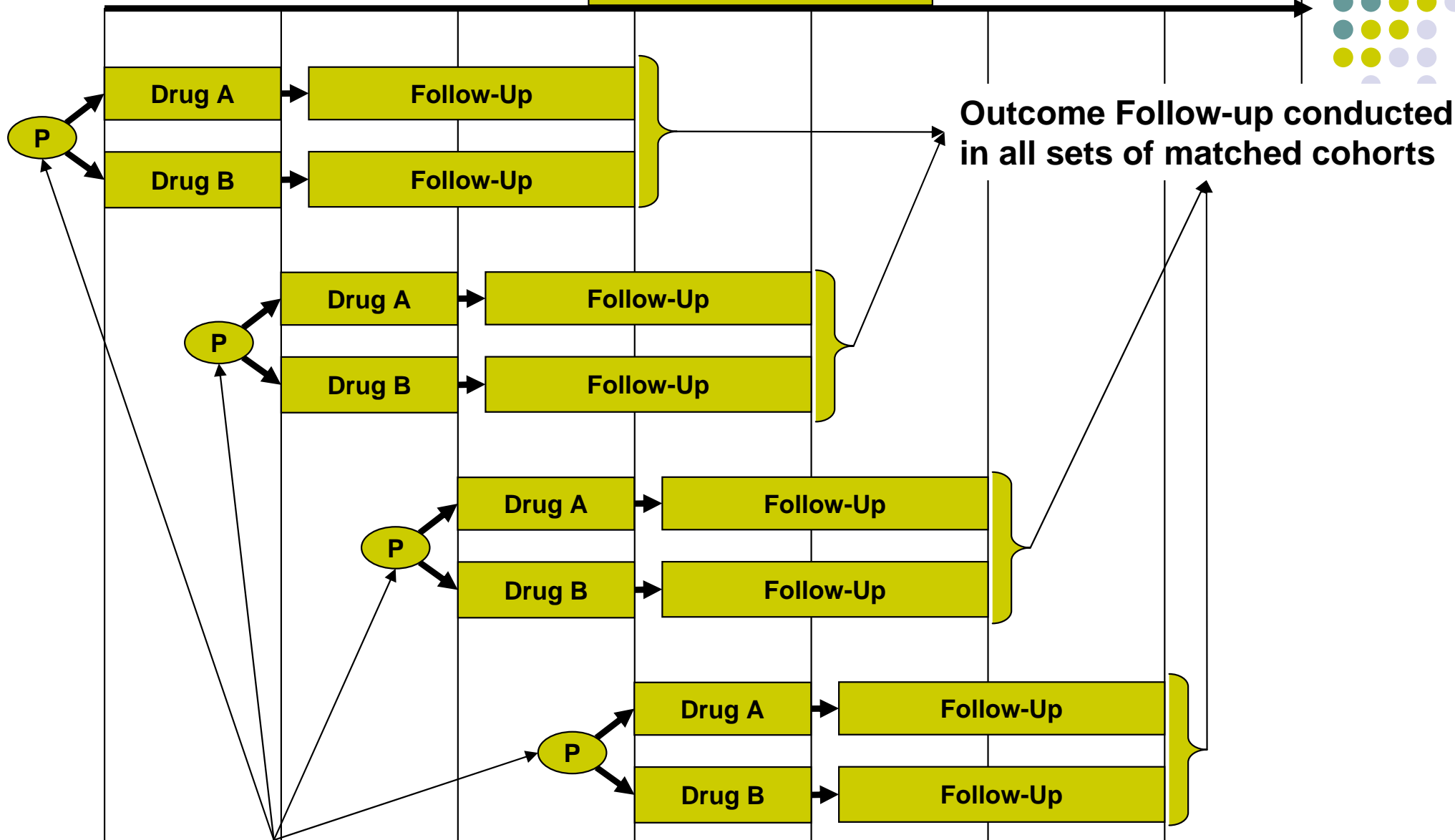
- An alternative is the ‘greedy match’ approach
- Start with a subject on drug A
- Find another subject on drug B who had the closest estimated propensity score
- Repeat for the next subject on drug A
- Result
 - Some subjects on drug A 1-1 matched to subjects on drug B, they are comparable with respect to $C_1, C_2, C_3, C_4, \dots$
 - Some subjects on drug A may not be matched

Matching within calendar time period



Drug Launch

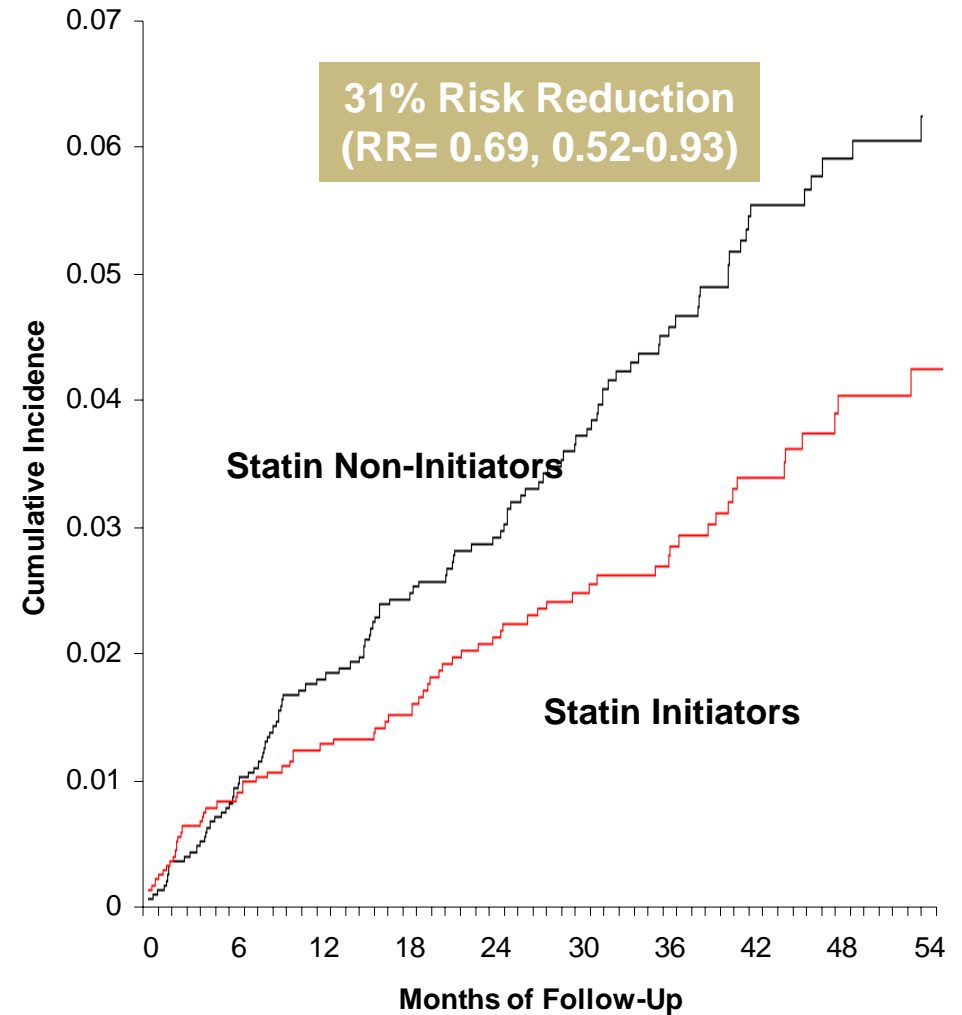
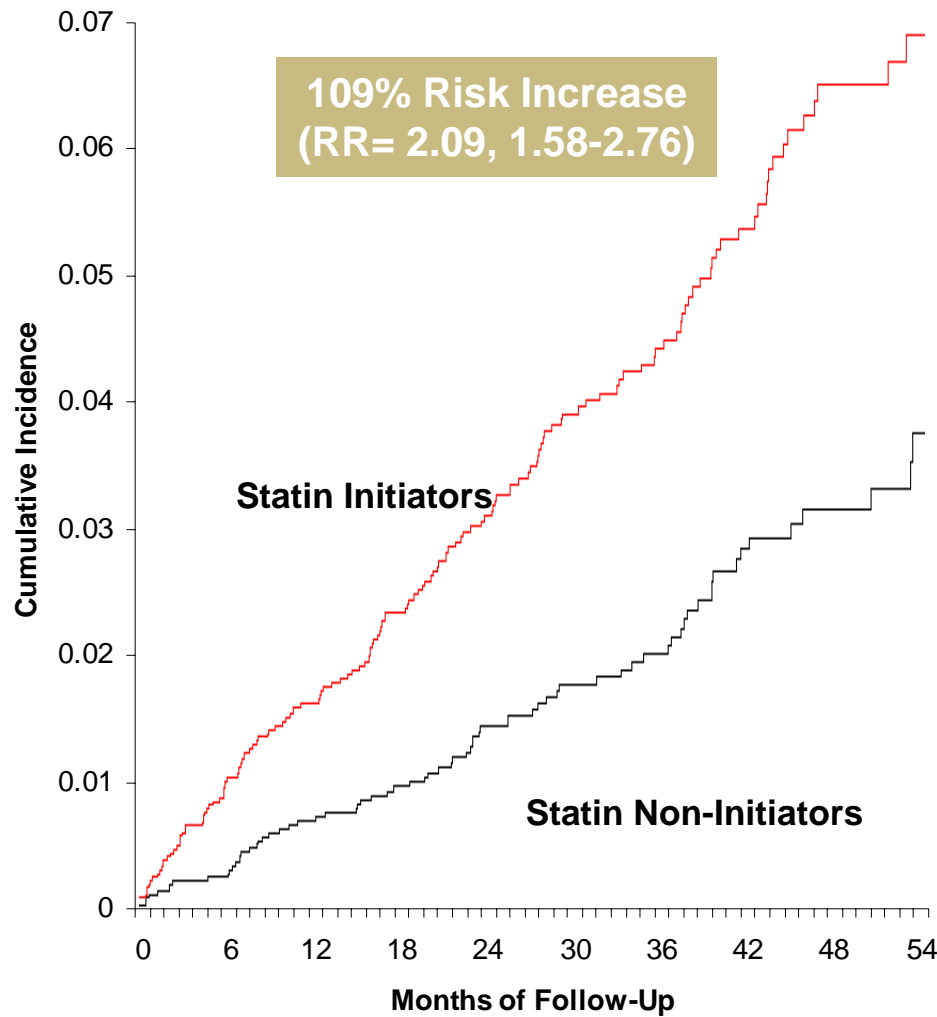
Time



Propensity score matching occurs repeatedly across time

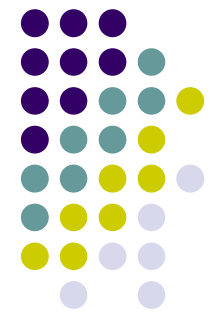
TASPOR

An example of matching by propensity score ... The Effect of Balancing



Seeger JD, et al. *Am J Cardiol* 2003;92:1447-1451

Ziyadeh et al. article as an example



Clinical Therapeutics/Volume 31, Number 11, 2009

The Thiazolidinediones Rosiglitazone and Pioglitazone and the Risk of Coronary Heart Disease: A Retrospective Cohort Study Using a US Health Insurance Database

Najat Ziyadeh, MA, MPH¹; Andrew T. McAfee, MD, MSc^{1,2}; Carol Koro, PhD^{3,4}; Joan Landon, MPH¹; and K. Arnold Chan, MD, ScD^{1,5}

Ziyadeh et. Al. Clinical Therapeutics 2009; 31: 2665-77.

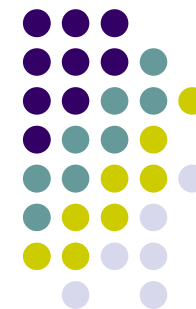


Table I. Demographic characteristics of study patients. Values are given

Variable	Exposure/Study Drugs	
	Rosiglitazone (n = 47,501)	Pioglitazone (n = 47,501)
Sex		
Male	27,226 (57.3)	27,433 (57.8)
Female	20,275 (42.7)	20,068 (42.2)
Age at drug regimen initiation, y		
<35	3204 (6.7)	3176 (6.7)
35-44	8519 (17.9)	8410 (17.7)
45-54	15,659 (33.0)	15,626 (32.9)
55-64	15,028 (31.6)	15,154 (31.9)
65-74	3763 (7.9)	3761 (7.9)
≥75	1328 (2.8)	1374 (2.9)
Year of drug regimen initiation		
2000-2002	12,553 (26.4)	11,982 (25.2)
2003-2004	12,037 (25.3)	12,535 (26.4)
2005-2007	22,911 (48.2)	22,984 (48.4)
Geographic region		
Northeast	3558 (7.5)	3535 (7.4)
South	24,034 (50.6)	24,069 (50.7)
Midwest	14,666 (30.9)	14,653 (30.8)
West	5243 (11.0)	5244 (11.0)

Ziyadeh et. Al. Clinical Therapeutics 2009; 31: 2665-77.

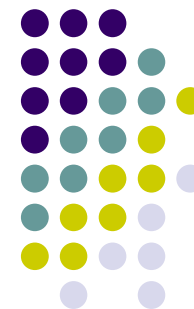
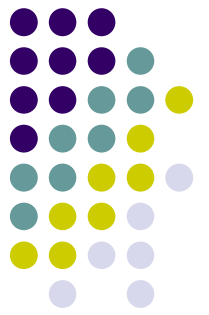
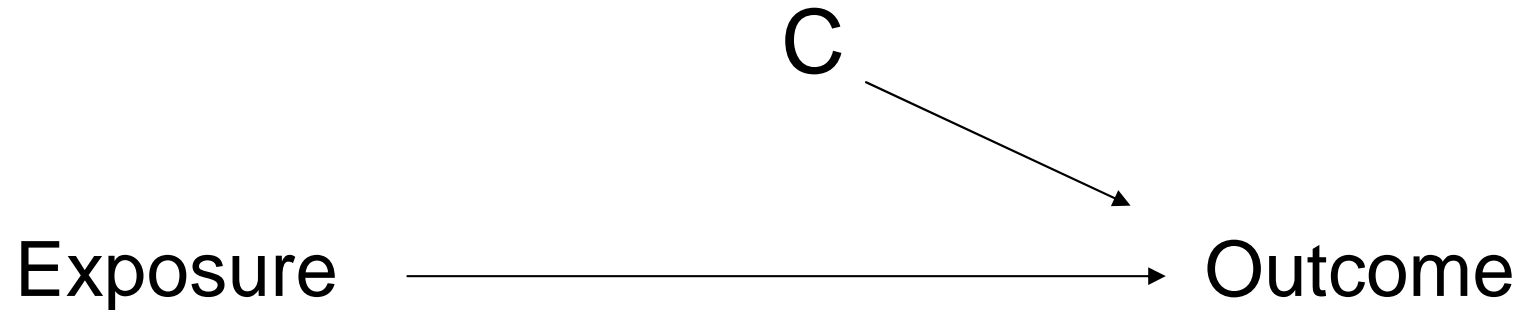


Table II. Diagnoses and patient drug use in the 6 months before study drug re

Variable	Exposure/Study Drugs	
	Rosiglitazone (n = 47,501)	Pioglitazone (n = 47,501)
Claims-based diagnoses		
Hyperlipidemia	30,251 (63.7)	30,503 (64.2)
Acute coronary syndrome	666 (1.4)	731 (1.5)
Myocardial infarction	769 (1.6)	782 (1.6)
Coronary revascularization	582 (1.2)	620 (1.3)
Angina	906 (1.9)	930 (2.0)
Coronary heart disease	4933 (10.4)	4998 (10.5)
Congestive heart failure	1226 (2.6)	1194 (2.5)
Hypertension	25,079 (52.8)	25,183 (53.0)
Smoking	1611 (3.4)	1681 (3.5)
Obesity	3904 (8.2)	3974 (8.4)
Antidiabetic drug use		
Metformin	26,762 (56.3)	26,332 (55.4)
Sulfonylureas	15,916 (33.5)	15,895 (33.5)
Other antidiabetic drugs	1364 (2.9)	1371 (2.9)
Insulin	346 (0.7)	319 (0.7)
Prescription drug use		
Nitrates	1517 (3.2)	1499 (3.2)
Digitalis preparations	898 (1.9)	848 (1.8)
β-Blockers	8265 (17.4)	8353 (17.6)
Calcium channel blockers	5990 (12.6)	6020 (12.7)
Diuretics	8053 (17.0)	8028 (16.9)
Antiplatelet agents	1793 (3.8)	1913 (4.0)
Statins	17,800 (37.5)	18,021 (37.9)
Fibrates	3786 (8.0)	3929 (8.3)
ACE inhibitors	15,291 (32.2)	15,422 (32.5)
Angiotensin receptor blockers	7258 (15.3)	7439 (15.7)

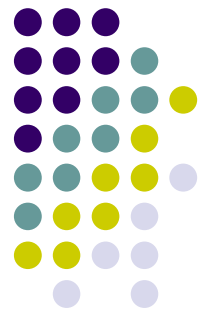


Propensity score methods



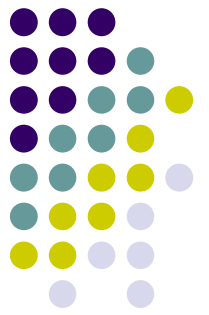
- Sound statistical basis
- Generate homogeneous subgroups through stratified analysis or
- Generate comparable cohorts through matching
- May indirectly control for unmeasured confounders
- Straightforward analysis and interpretation

JAMA 2007; 297: 278-285



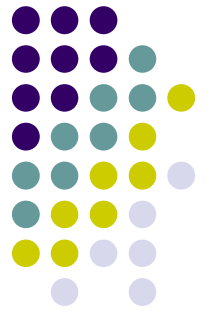
 ORIGINAL CONTRIBUTION

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



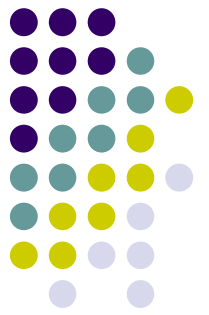
Stukel *et al.* JAMA 2007; 297: 278-285

- Data source: U.S.Medicare + additional data collection (Cooperative Cardiovascular Project)
- Study population: patients hospitalized with acute myocardial infarction
- Exposure: cardiac catheterization
- Outcome: mortality
- Challenge: confounding by indication



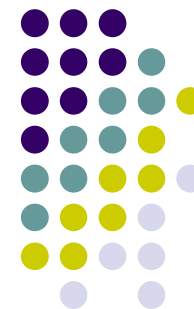
Stukel *et al.* JAMA 2007; 297: 278-285

- Four analytic methods
 - Multivariable model risk adjustment
 - Propensity score risk adjustment
 - Propensity-based matching
 - Instrumental variable analysis



Instrumental variable (IV)

- Highly correlated with treatment
- Does not independently affect the outcome
- Outcome of interest is a continuous variable
- In the Stukel et al. study
 - Regional cardiac catheterization rate was chosen as the IV



Stukel *et al.* JAMA 2007; 297: 278-285

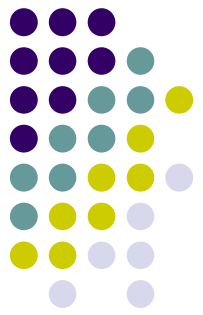
Table 3. Adjusted Relative Mortality Rate Associated With Receipt of Cardiac Catheterization Among Patients With AMI Using Standard Risk-Adjustment Methods

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

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

*Simple propensity score included all 65 patient, hospital, and ZIP code characteristics.

†Complex propensity score included all patient, hospital, and ZIP code characteristics and all interactions of age, sex, and race with the other characteristics (750 variables).



Stukel *et al.* JAMA 2007; 297: 278-285

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

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

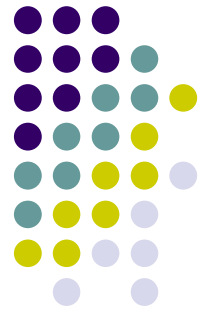
Abbreviations: AMI, acute myocardial infarction; CI, confidence interval.

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

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

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

Another good example

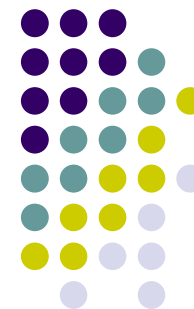


American Journal of Epidemiology
Copyright © 2005 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved; printed in U.S.A.

Vol. 163, No. 3
DOI: 10.1093/aje/kwj047
Advance Access publication December 21, 2005

Original Contribution

Results of Multivariable Logistic Regression, Propensity Matching, Propensity Adjustment, and Propensity-based Weighting under Conditions of Nonuniform Effect



Kurth et al. Am J Epidemiology 2006; 163: 262-270

TABLE 5. Comparison of the estimated treatment effect of tissue plasminogen activator on death using multivariable logistic regression, propensity score–matched analysis, regression adjustment with the propensity score, inverse-probability-of-treatment-weighted, and standardized mortality ratio-weighted analyses for ischemic stroke patients registered in a German stroke registry between 2000 and 2001, after restriction to participants whose propensity score is ≥ 0.05

	No.	OR*	95% CI*
Crude model	978	1.36	0.84, 2.19
Multivariable model†	978	1.30	0.74, 2.31
Matched on propensity score	338	0.89	0.49, 1.63
Regression adjusted with propensity score			
Propensity score, continuous	978	0.99	0.58, 1.68
Multivariable†	978	1.29	0.73, 2.29
Propensity score, deciles	978	1.24	0.75, 2.03
Multivariable†	978	1.31	0.74, 2.33
Weighted models			
IPTW*	978	1.09	0.62, 1.93
SMR* weighted	978	0.82	0.47, 1.44