# Instrumental Variable Analysis and Interrupted Times Series Analysis in Health Policy Research

## "You Can't Fix by Adjustment What You Bungled by Design"

**ISPE's 10th Asian Conference on Pharmacoepidemiology**
**Brisbane, Australia**
**October 29, 2017**

**Stephen B. Soumerai**
**Professor of Population Medicine**
**Harvard Medical School /Harvard Pilgrim Health Care Institute**

# Presentation Agenda

1. Case study: a “bad” instrumental variable (IV): advanced life support vs. basic life support ambulances “leads” to increased mortality

2. Systematic review: validity of the four most common IVs in studies of the effects of health care interventions on mortality

3. Comparing the validity of cross-sectional adjustment with controlled interrupted time series designs in studies of benzodiazepine cessation and hip fracture

# Common Threats to Internal Validity

**Selection:** Pre-intervention differences between people in one experimental group vs. another

- **Confounding by Indication:** Physicians choose to preferentially treat or avoid pts who are sicker, older, or have had an illness longer

---

History

Maturation

Regression to the mean, etc.

# Hierarchy of Strong and Weak Designs: Capacity to Control for Biases

**Strong Design: Often Trustworthy Effects**

**Intermediate Design: Sometimes Trustworthy Effects**

**Weak Designs: Rarely Trustworthy Effects (No Controls for Common Biases.)**

# Hierarchy of Strong and Weak Designs: Capacity to Control for Biases

| Strong Design: Often Trustworthy Effects | |
|---|---|
| Multiple RCTs | The "gold standard" of evidence, incorporating systematic review of all studies. |
| Single RCT | A single, strong randomized experiment, but sometimes not generalizable. |
| Interrupted time series with control series (CITS) | Baseline trends often allow visible effects and control for biases. Two controls. |

# Hierarchy of Strong and Weak Designs: Capacity to Control for Biases

| Intermediate design: Sometimes Trustworthy Effects | |
|---|---|
| Single ITS | Controls for trends, but no comparison. |
| Before and after with comparison group | Pre-post change using two single observations. Comparability of baseline unclear. |

| Weak Designs: Rarely Trustworthy Effects (No Controls) | |
|---|---|
| Uncontrolled pre-post | Single observations before and after intervention, no baseline or control group. |
| Cross-sectional designs | Simple correlation, no baseline, no measure of change. |

# Background on IV Analysis

- IV analyses: weak cross-sectional designs
  - Assumes that IVs (e.g., distance to the hospital) randomizes tx ("ignorable tx assignment")

- Many IVs do not protect against bias
  - Heroic statistical adjustments do not control for differences between the study groups

"You can't fix by analysis what you bungled by design."

Source: Soumerai SB and Koppel R. Health Serv Res. 2017 Feb; 52(1):9-15.

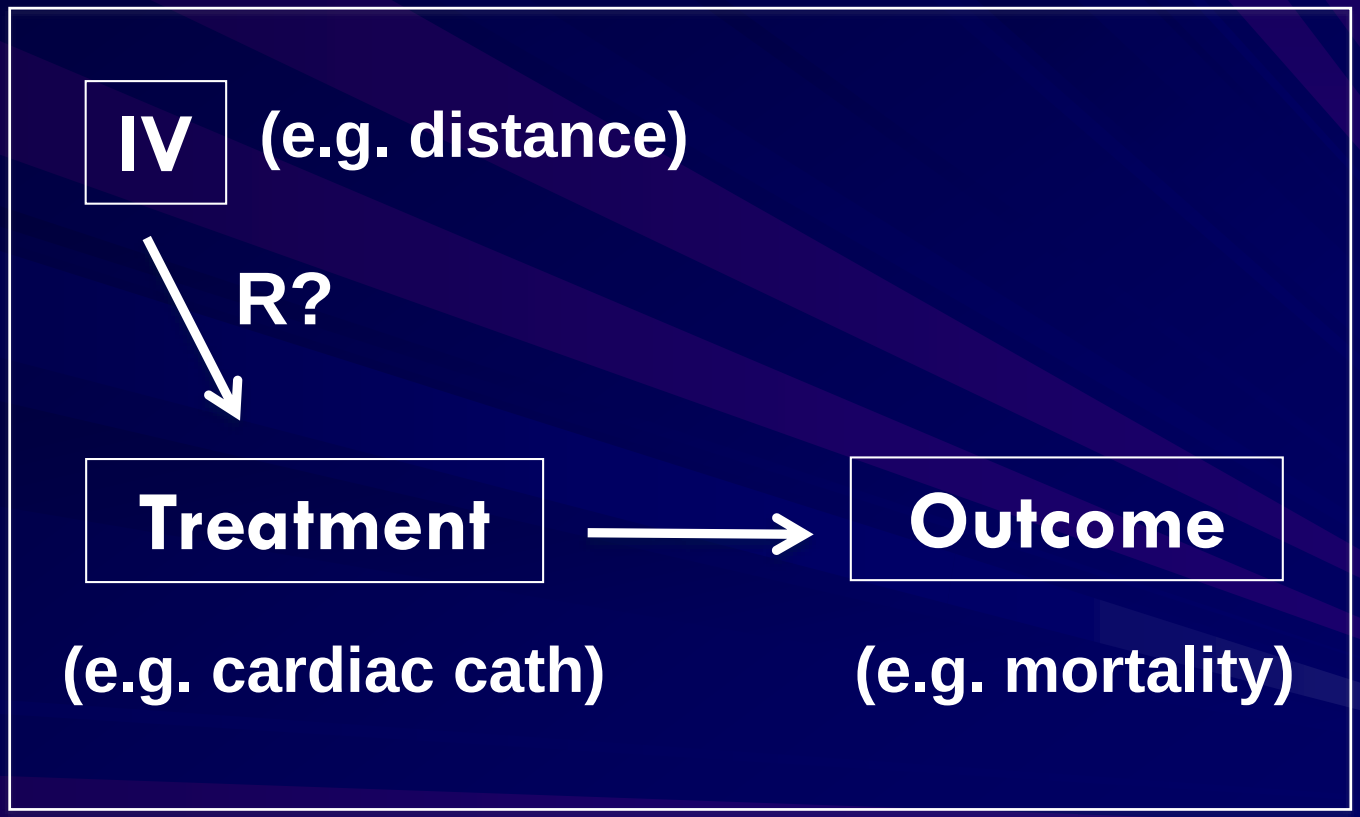
# Illustration of IV Analysis

In theory, IV controls for unobserved and observed patient characteristics that impact the outcome

- Predicts tx assignment
- Unrelated to factors influencing outcome (exclusion assumption)

Illustrative ex: distance to hospital "randomizes" cardiac cath to MI patients

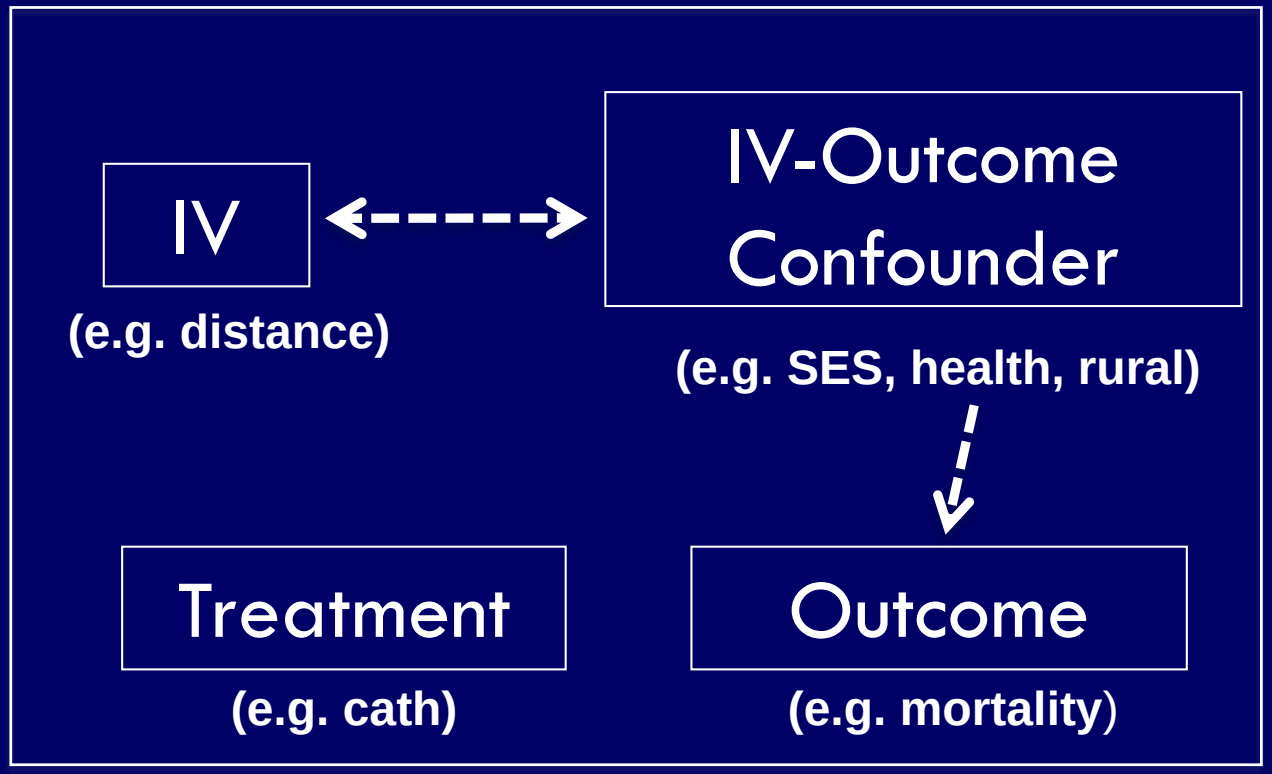
# Illustration of IV (cont.)



# Violation of IV Assumptions

**IV biased if IV outcome related through unadjusted 3rd variable: IV-outcome confounder**

**Exclusion restriction**



# Landmark 1994 IV CER article (JAMA)

Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality?

Analysis Using Instrumental Variables

Mark McClellan, MD, PhD; Barbara J. McNeil, MD, PhD; Joseph P. Newhouse, PhD

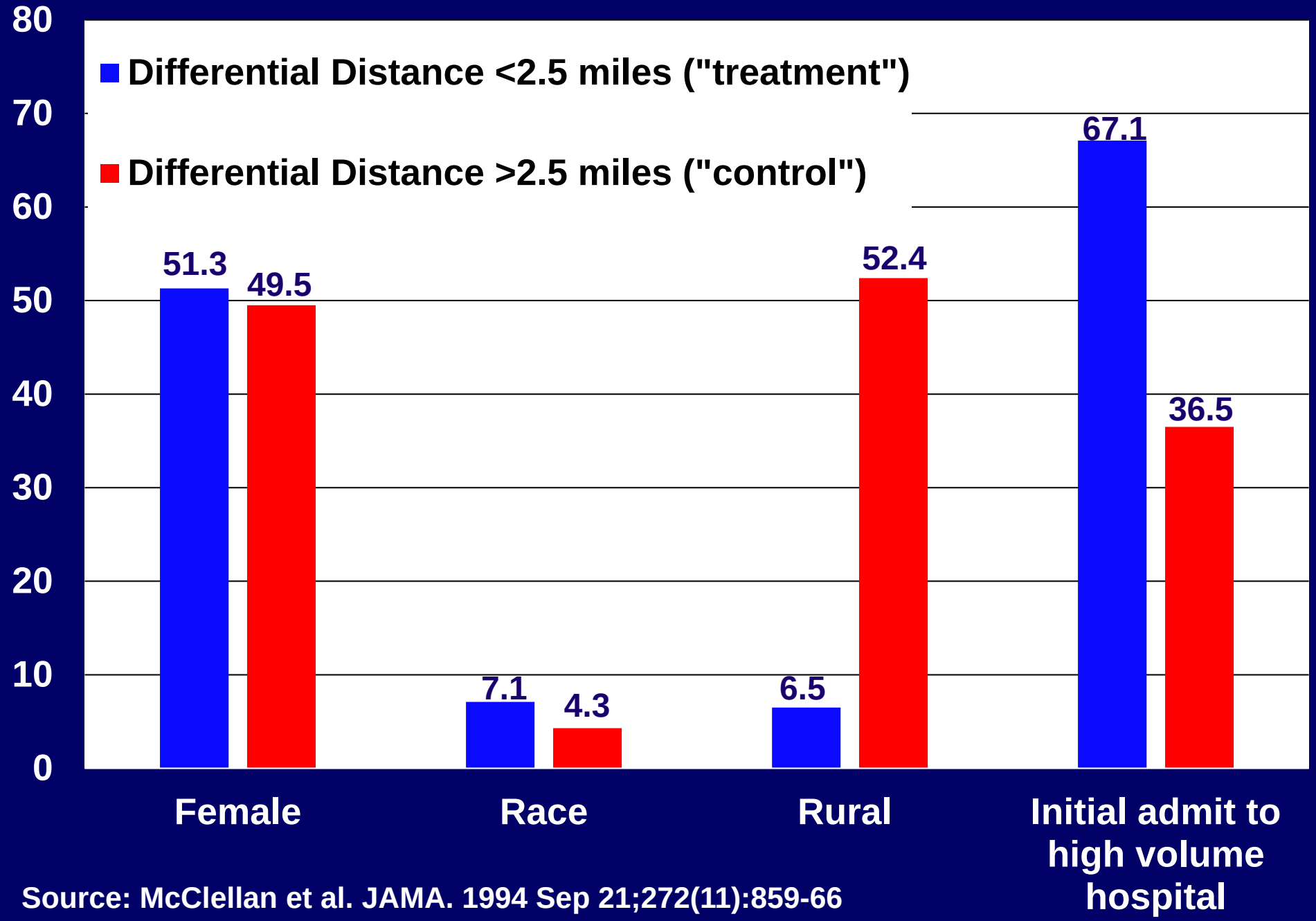
**Treatment:** cardiac catheterization

**Outcome:** mortality (survival)

**IV =** differential distance to catheterization hospital

**Cited 835 times**

# Patient Characteristics by Differential Distance




Source: McClellan et al. JAMA. 1994 Sep 21;272(11):859-66

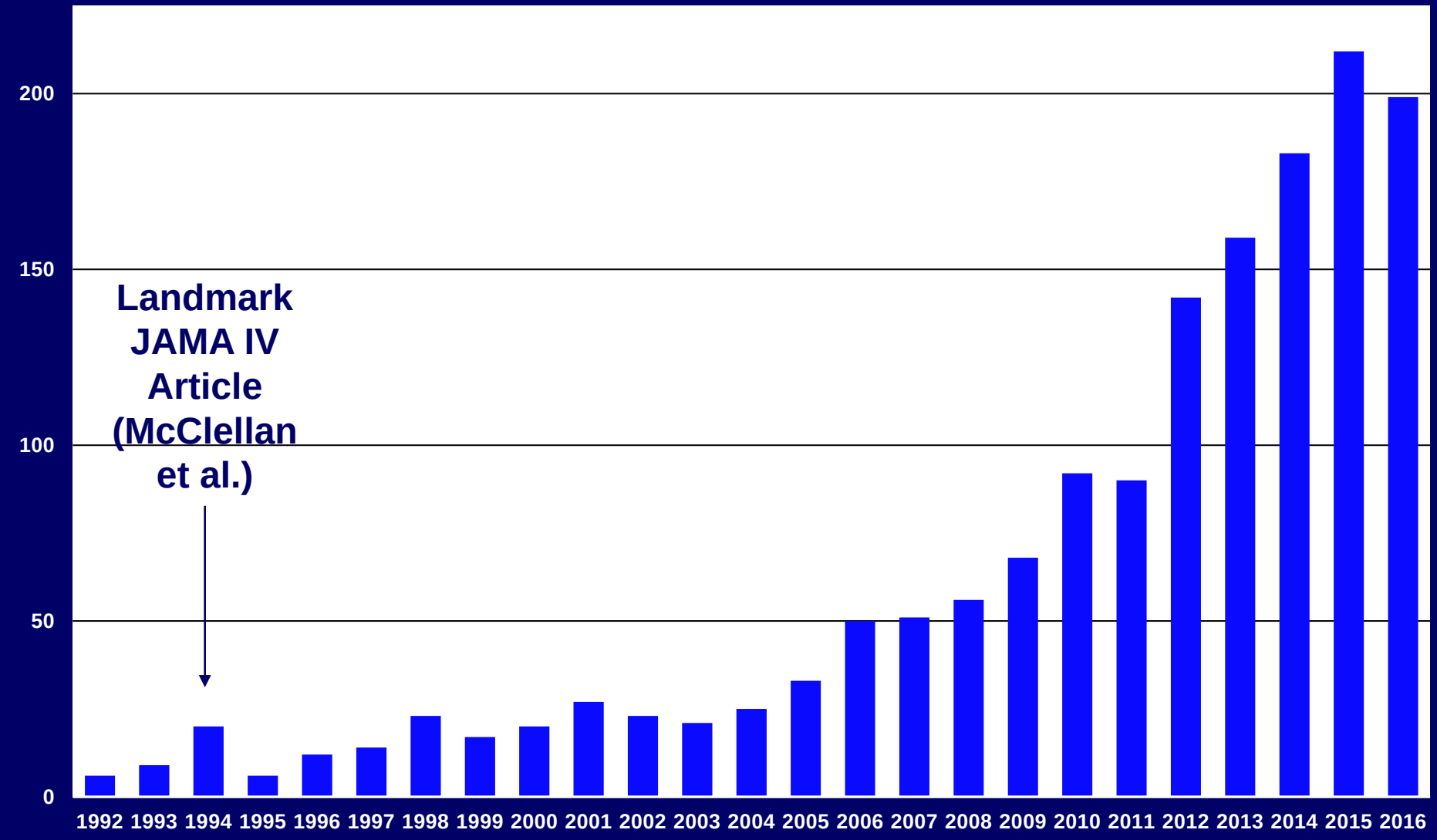
# Evidence of Unmeasured Confounding

***"…the beneficial effect of catheterization appears at day 1, before the catheterization…"***

***"Thus, aspects of acute care other than…invasive procedures" are responsible for better outcomes at cath hospitals***

**Source: McClellan et al. JAMA. 1994 Sep 21;272(11):859-66**

# Citation Search of Instrumental Variables: No. of Published Articles Per Year



# 1. Case Study: A bad instrumental variable (IV): advanced life support vs. basic life support ambulances “leads” to increased mortality

The Washington Post
*Democracy Dies in Darkness*

**PostEverything** • Perspective

# How bad science can lead to bad science journalism – and bad policy

This is what happens when news organizations don't catch lousy studies.

**By Stephen Soumerai and Ross Koppel** June 7

Follow @StephenSoumerai

About the authors

Annals of Internal Medicine

ORIGINAL RESEARCH

# Outcomes of Basic Versus Advanced Life Support for Out-of-Hospital Medical Emergencies

Prachi Sanghavi, PhD; Anupam B. Jena, MD, PhD; Joseph P. Newhouse, PhD; and Alan M. Zaslavsky, PhD

Source: Sanghavi P et al. Ann Intern Med. 2016 Jul 5;165(1):69-70.

# Causal Interpretation of IV Correlations

**Abstract Conclusion:** "Advanced life support (ALS) ambulances associated with substantially higher mortality…

**Final Sentence:** "In conclusion, our findings suggest that survival is longer with BLS and BLS may offer benefits for nonfatal outcomes."

# The Study

Cross-sectional analysis of mortality in Medicare claims data

Compared those picked up by basic vs advanced ambulances

- Adjustment with propensity scores and IVs
- No collaboration w/ emerg. med specialists

Survival at 90 days 4-7% higher with basic (BLS)

# Confusing Cause and Effect

**IV assumption:**

- Severely ill patients "randomized" to ALS
  - 1.Direct contrast, or 2. Counties with more/less BLS

**Not the case.**

- ALS sent to sicker patients, further away

**It's not random selection** (like RCTs); it's triage

# Typical EMT reactions

“We don’t send basic life support ambulances to a head-on car crash on a freeway.”

“A basic ambulance…won’t be activated for an elderly person who’s difficult to arouse, complaining of chest pain.”

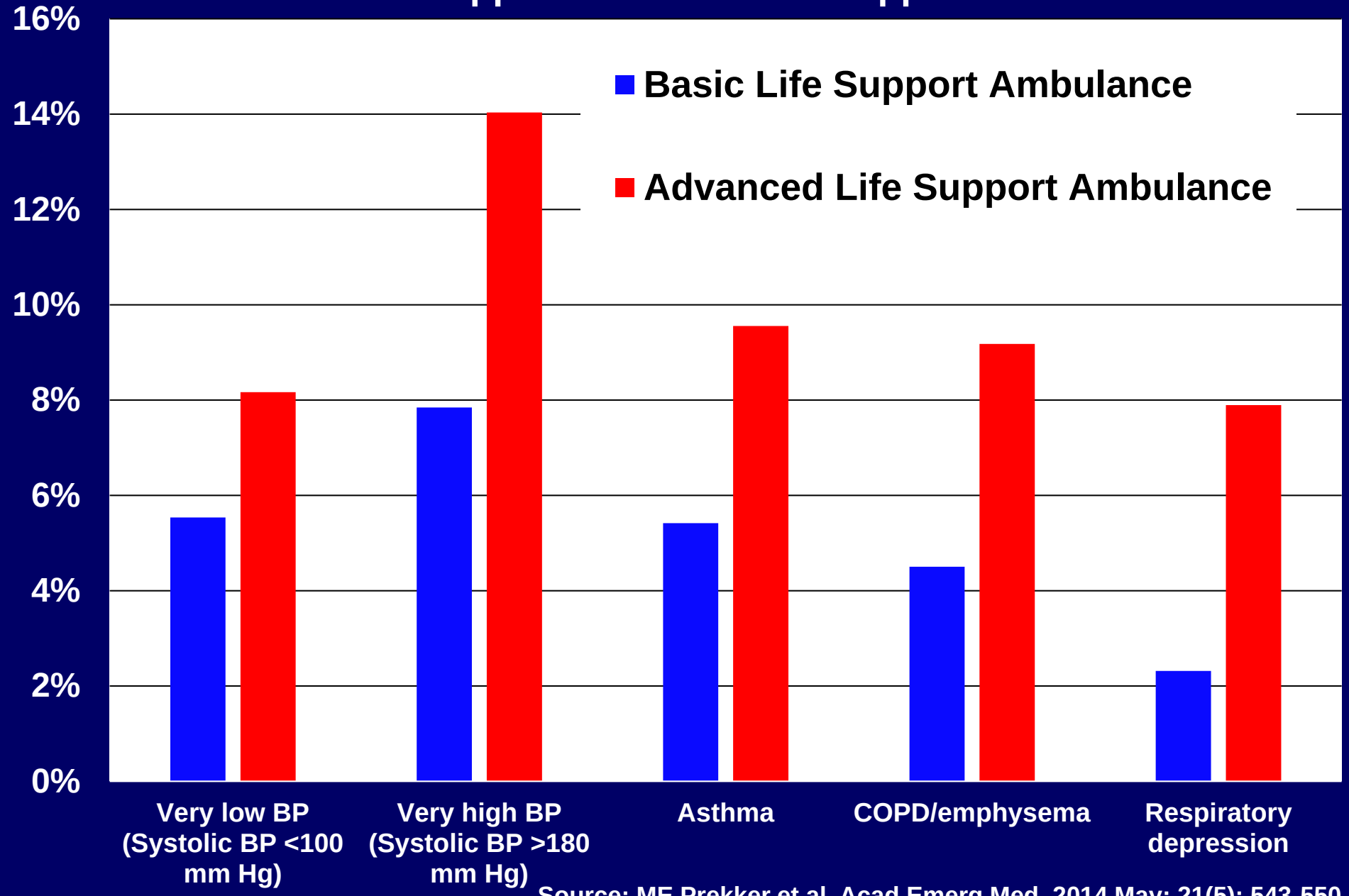
# Difference in Risk Factors for Mortality before Pickup

ALS is twice as likely to pick up people with respiratory distress

- Result: more deaths.

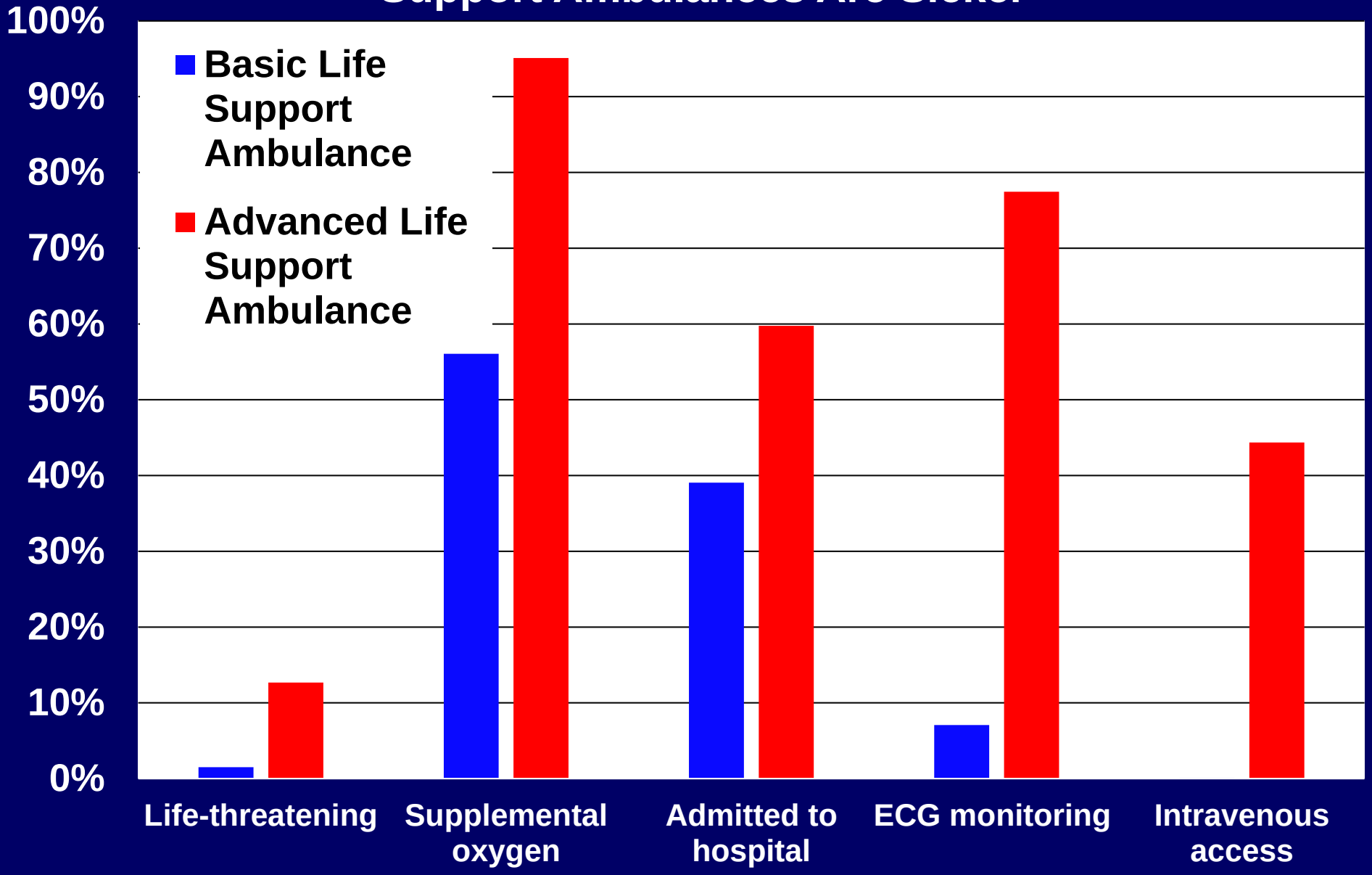
Source: Prekker ME et al. Acad Emerg Med. 2014 May; 21(5): 545-550.

# Several Serious Conditions of Patients Transported in Advanced Life Support vs. Basic Life Support Ambulances




Source: ME Prekker et al. Acad Emerg Med. 2014 May; 21(5): 543-550.

# Patients Transported in Advanced vs. Basic Life Support Ambulances Are Sicker




Source: ME Prekker et al. Acad Emerg Med. 2014 May; 21(5): 543-550.

# National Impact

**The article's authors exaggerated their single weak study, even calculating national savings of $320 million by abandoning ALS ambulances.**

The Washington Post
*Democracy Dies in Darkness*

To Your Health

## Need an ambulance? Why you may not want the more sophisticated version.

By Lena H. Sun October 12, 2015

# 2. Systematic review of bias in most common IVs in comparative effectiveness research

# Our Study

Annals of Internal Medicine | Research and Reporting Methods

## Potential Bias of Instrumental Variable Analyses for Observational Comparative Effectiveness Research

Laura Faden Garabedian, PhD; Paula Chu, MS; Sengwee Toh, ScD; Alan M. Zaslavsky, PhD; and Stephen B. Soumerai, ScD

Source: Garabedian LF et al. Ann Intern Med. 2014 Jul 15;161(2):131-8.

# Systematic Review Study Objectives

1. Evaluate the trend in the use of IVs for CER

2. Determine the most commonly used IVs

3. Identify potential IV-outcome confounders

4. Determine the proportion of IV CER studies that are potentially biased by IV-outcome confounders

# Majority of IV Studies Used 1 of 4 Most Common IVs (n=65; 61%)

**Regional Variation:** 49 studies (26.2%)

**Distance to Facility:** 38 (20.3%)

**Facility Variation:** 22 (11.8%)

**Provider Variation:** 14 (7.5%)

***Mortality** was the most common outcome for each IV type**

# Evidence in Literature of IV-Outcome Confounding (of 4 IVs and Mortality)

**Patient characteristics: race, SES, risk factors for mortality, health status, and urban/rural**

**Health system characteristics: facility and procedure volume, facility characteristics (e.g., teaching hospital)**

**Treatment characteristics: time to treatment, receipt of other lifesaving treatments**

# Did authors discuss or control for the potential IV-outcome confounders?

**83%** (54/65) stated the assumption of no IV-outcome confounding

**63%** (41/65) provided additional analyses or discussion to determine if the assumption was met

**6%** (4/65) considered potential IV-outcome confounders outside of study data

**NONE of the studies** in our review controlled for all of the IV-outcome confounders we identified

# Percent of Studies that Controlled for Confounders by IV Category

| Confounders | Distance (n=27 studies) | Regional Variation (n=23) | Facility Variation (n=14) | Physician Variation (n=9) |
|---|---|---|---|---|
| Patient Income | 44% | 70% | 14% | 0% |
| Patient Education | 15% | 22% | 14% | 0% |
| Urban/Rural | 44% | 52% | 7% | 22% |
| Volume (procedure) | 4% | 0% | 27% | 11% |
| Volume (facility) | 41% | 41% | 39% | 11% |

# Quantitative Assessments of Bias

**An IV-outcome confounder can lead to overestimation, underestimation or complete reversal of the true treatment effect**

***See Brookhart MA, Schneeweiss S. Int J Biostat. 2007;3(1):14**

# Study Conclusions

IV analysis is an increasingly popular method for CER

In practice, most IV CER studies are cross-sectional; overconfident in asserting that key IV assumptions are met

Most common IVs should be used cautiously because their results are potentially biased

# When less is more

"Don't use a lot where a little will do." -Proverb

# A Strong IV?

## 1 December 1969 | The Military Draft Lottery for the Vietnam War


On the evening of **December 1, 1969**, officials of the Selective Service and the U.S. Congress post birth dates in ascending order, as they are drawn in the first Vietnam-era draft lottery at Selective Service headquarters in Washington. The broadcast of the event was carried "live" over the radio and on CBS TV.

Alexander Pirnie draws the first of the 366 blue capsules

1970 RANDOM SELECTION SEQUENCE, BY MONTH AND DAY

| | JAN | FEB | MAR | APR | MAY | JUN | JUL | AUG | SEP | OCT | NOV | DEC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 305 | 086 | 108 | 032 | 330 | 249 | 093 | 111 | 225 | 359 | 019 | 129 |
| 2 | 159 | 144 | 029 | 271 | 298 | 228 | 350 | 045 | 161 | 125 | 034 | 328 |
| 3 | 251 | 297 | 267 | 083 | 040 | 301 | 115 | 261 | 049 | 244 | 348 | 157 |
| 4 | 215 | 210 | 275 | 081 | 276 | 020 | 279 | 145 | 232 | 202 | 266 | 165 |
| 5 | 101 | 214 | 293 | 269 | 364 | 028 | 188 | 054 | 082 | 024 | 310 | 056 |

C3i e-Magazine

# Vietnam Draft Lottery: Caveats

- Draft dodgers were generally young, well educated healthy men.
- So use intention to treat (include the draft dodgers in the comparative analysis)

**Source: Berinsky AJ and Chatfield S. Political Analysis. 2015;23:449-454.**

## 3. Comparing the validity of cross-sectional adjustment with stronger controlled interrupted time series designs studies of benzodiazepine cessation and hip fracture

# The Bias: Confounding by Indication

Plagues the field of observational comparative effectiveness of health care treatments.

Physicians choose to preferentially treat or avoid patients who are sicker, older, or have had an illness longer.

The trait (e.g., dementia) causes the adverse event (e.g., hip fracture), not the treatment itself (e.g., sedatives).

**“Landmark studies that failed to control for this bias nevertheless influenced worldwide drug safety programs for decades, despite better controlled longitudinal time-series studies that debunked the early dramatic findings…”**

**Source: Soumerai SB et al. Prev Chronic Dis. 2015 Jun 25;12:E101.**

# Background

One of the oldest and most accepted "truths" in medication safety research:

- Benzodiazepines (Valium and Xanax) that are prescribed for sleep and anxiety) may cause hip fractures among the elderly
- Because the drugs' sedating effects might cause falls and fractures

# Common designs: benzodiazepine/fx research

Weakest non-experimental, cross-sectional designs

CBI problematic in studies of benzodiazepines because physicians Rx them to elderly patients who are sick and frail

Because sickness and frailty are often unmeasured, their biasing effects are hidden

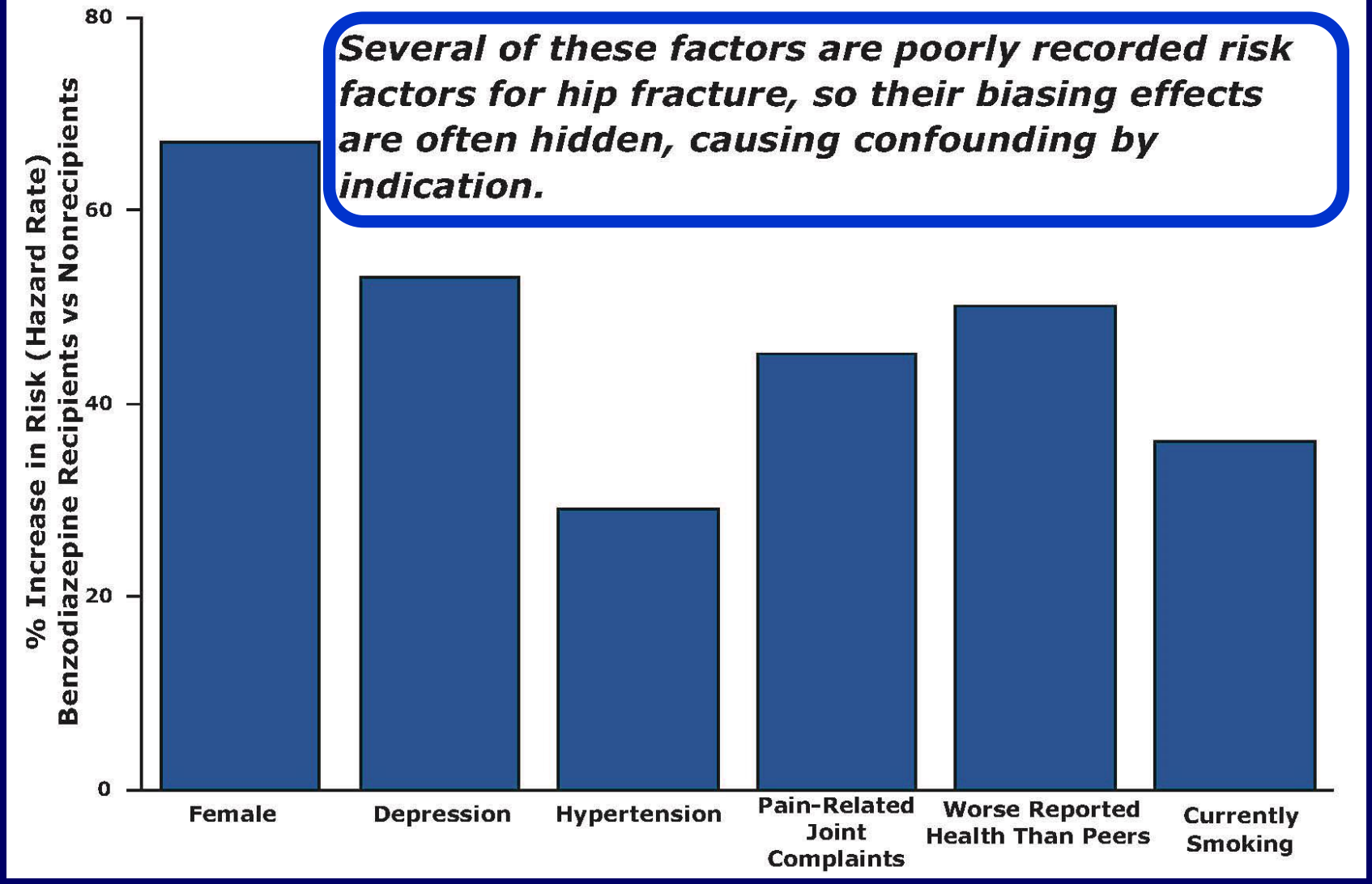



**Figure. Elderly people who begin benzodiazepine therapy (recipients) are already sicker and more prone to fractures than non recipients.**

**Source: Lujendijk et al. Br J Clin Pharmacol 2008:65(4)593-9**

# A Weak Design that does not control for Confounding by Indication

Thirty years ago, a landmark study used Medicaid claims data to show a relationship between benzodiazepine use and hip fracture in the elderly

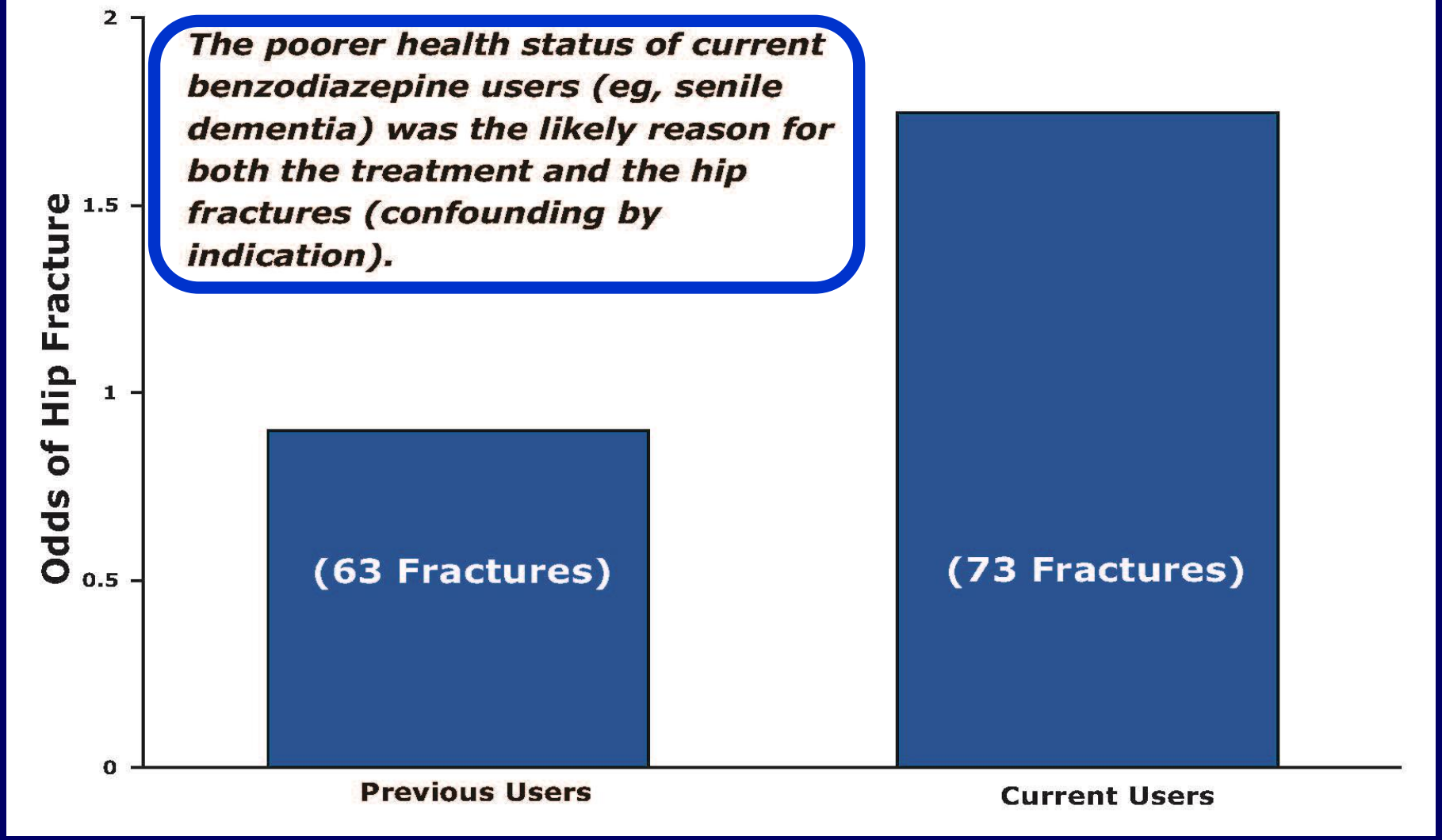
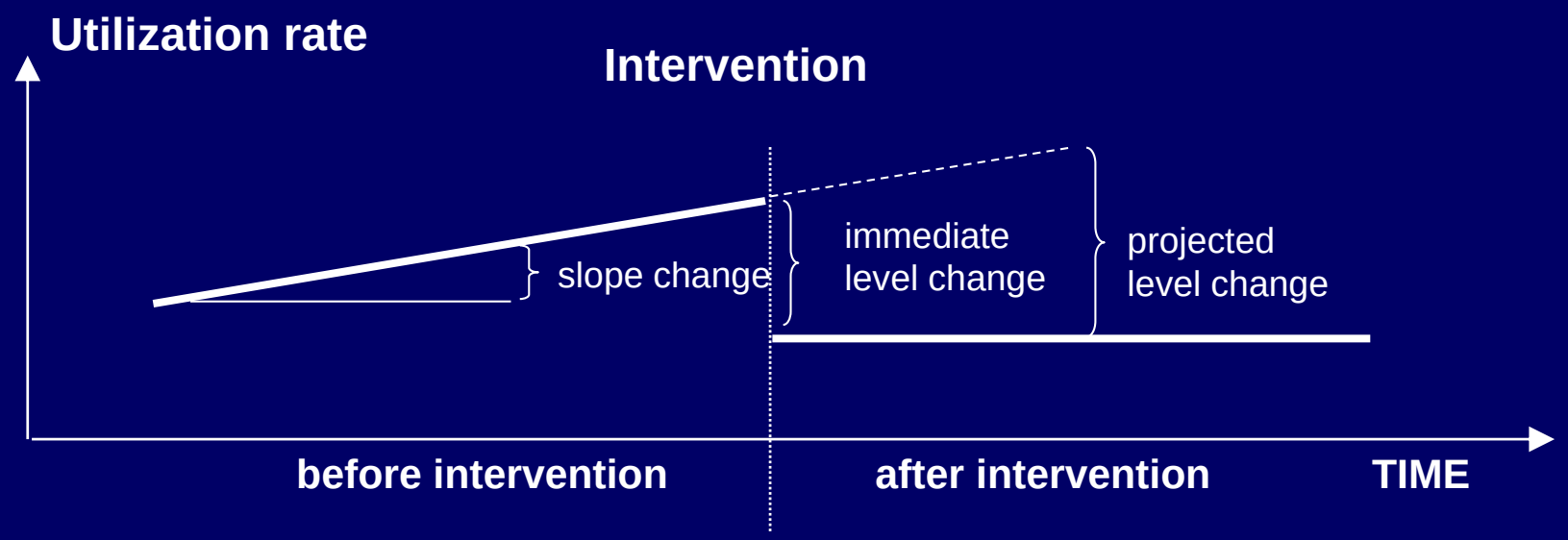



**Figure. Weak post-study epidemiological study suggesting that current users of Benzodiazepines are more likely than previous users to have hip fractures.**

Source: Ray et al. N Engl J Med 1987;316(7):363-9.

# Hypothetical Changes in Level and Slope of in a Stronger Time-Series Design

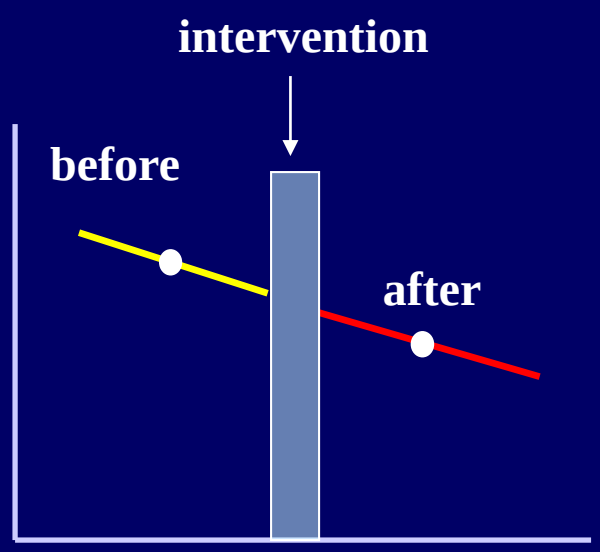
**Analysis of a health policy intervention by interrupted (segmented) linear regression.**

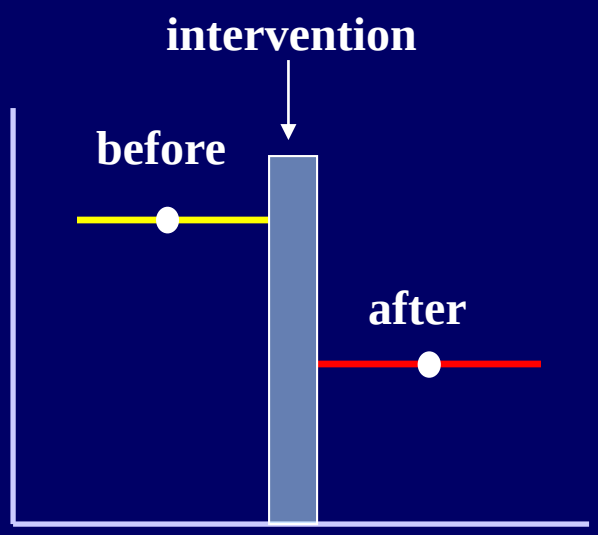


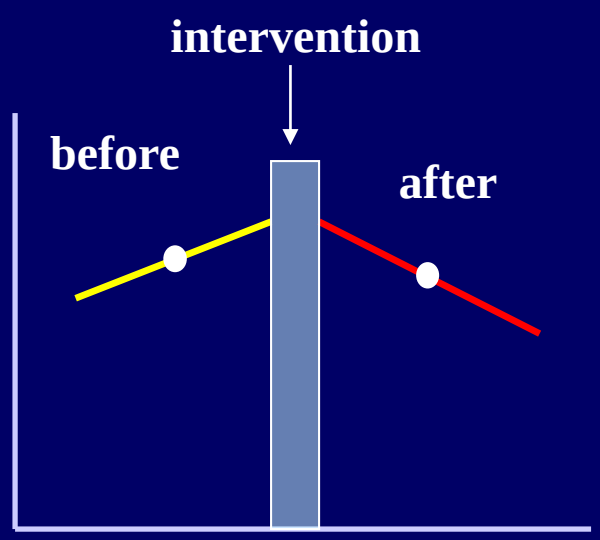

**Assumption: The (counterfactual) experience of patients had the policy not been implemented is correctly reflected by the extrapolation of the pre-policy trend**

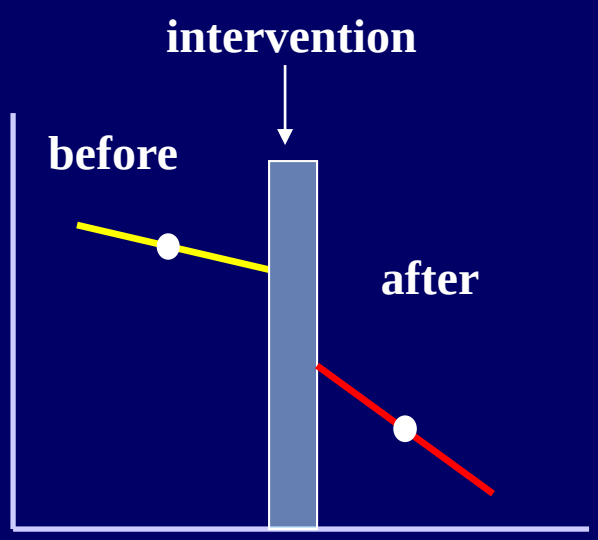
**Source: Schneeweiss S. Harvard Medical School**

# Different Effects That Can Be Observed in Time Series

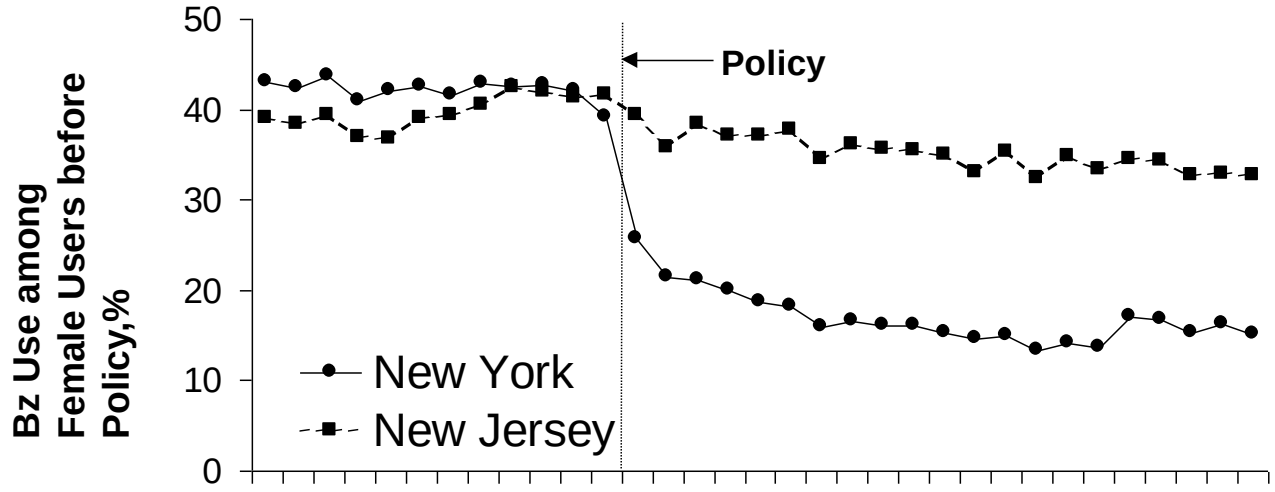


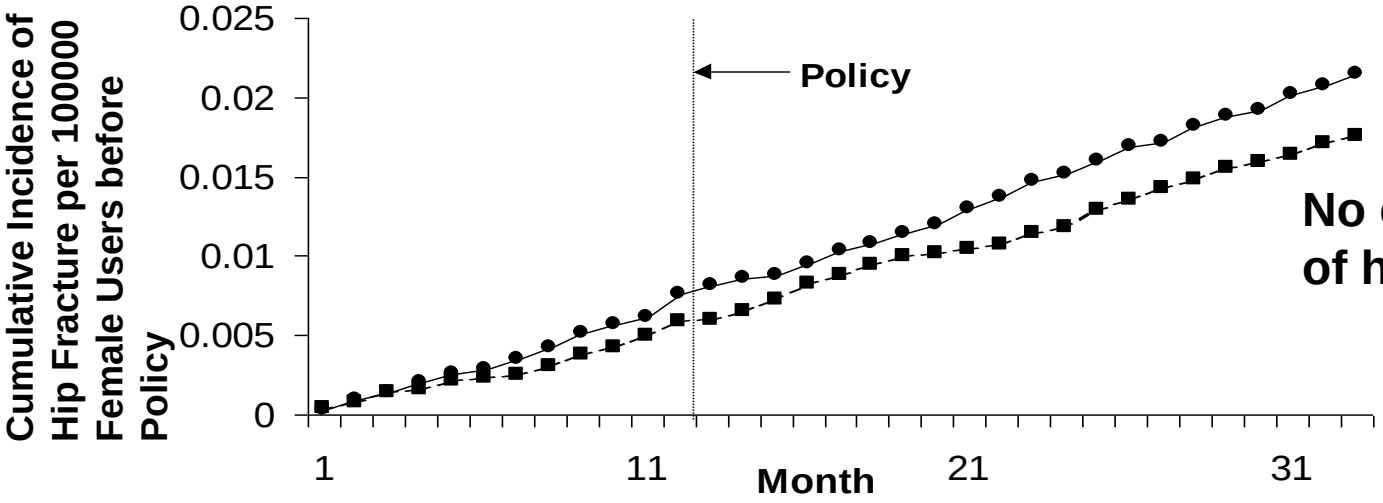



# Benzodiazepine (BZ) Use and Risk of Hip Fracture among Women in Medicaid Before and After NY Regulatory Surveillance Restricting BZ use




**60% decrease in BZ use in NY**




**No change in risk of hip fracture**

Source: Wagner AK et al. Ann Intern Med. 2007;146(2):96–103.

**Contrary to decades of previous studies, the Annals editors of this study concluded that:**

**"controlling benzodiazepine prescribing may not reduce hip fractures, possibly because the 2 are not causally related."**

- **ITS study by Briesacher et al confirmed above findings in long-term care (Arch Intern Med, 2010)**

# News Coverage

The findings of the early, landmark studies:

- hyped by the media, affecting MDs, policy makers.

Most reporters simply accepted authors' conclusions.

The New York Times stated that elderly people were

- "70% more likely to fall and fracture their hips"
- "thousands of hip fractures could be prevented each year if use of the drugs were discontinued."

# Coverage of New York ITS Study

**The Washington Post, January 15, 2007**

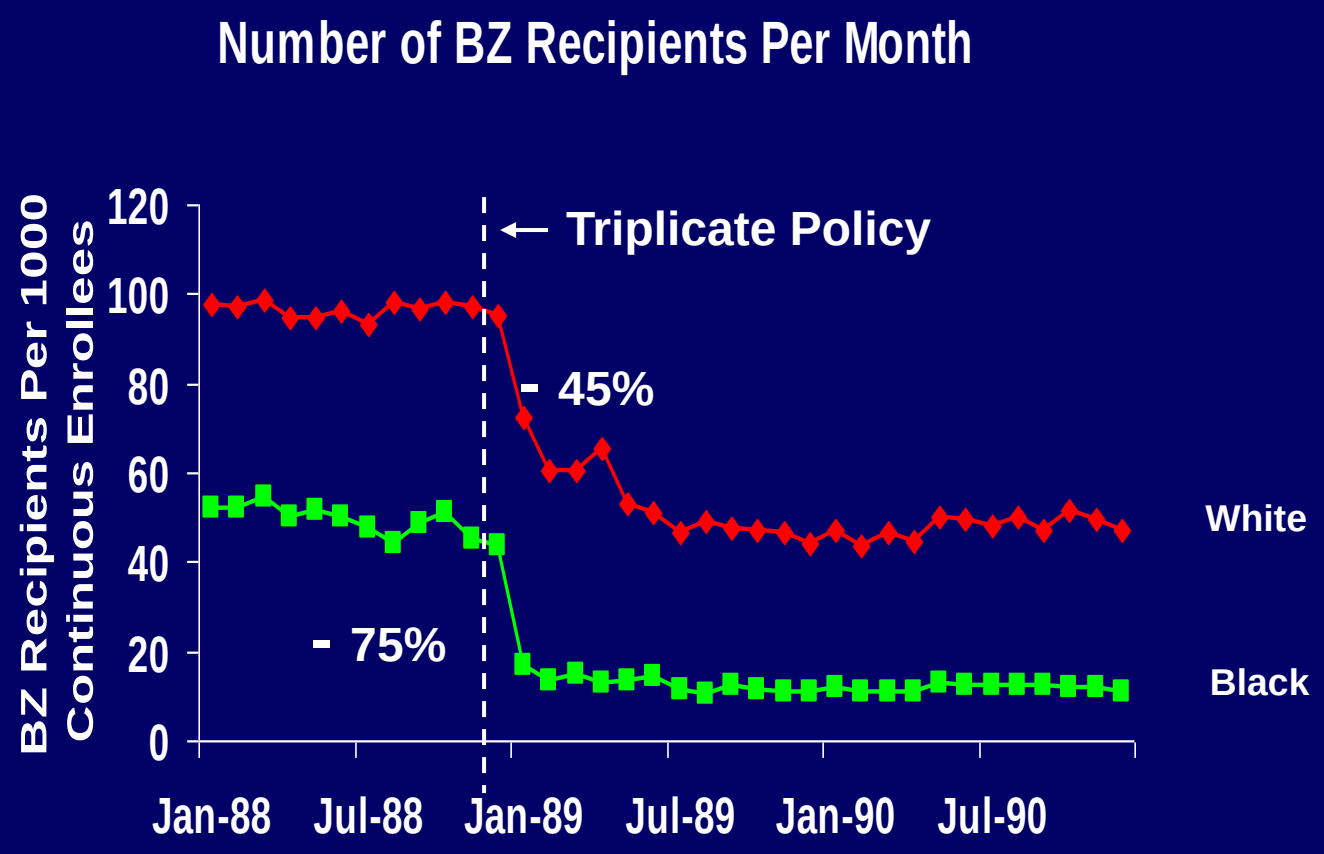
**Study Debunks Sedatives Link to Hip Facture In Elderly**

**"Sedative drugs called benzodiazepines (such as Valium) don't increase the risk of hip fractures in the elderly, a Harvard Medical School study said."**

**"US.. policies that restrict access to these drugs among the elderly need to be re-examined..."**

# Use of Longitudinal ITS to Measure Subgroup Effects

## Race Disparity: Impact of NY TPP on BZ Use




Source: Pearson SA et al. Arch Intern Med. 2006 Mar 13;166(5):572-9

# Conclusions

- Scientists, journalists, and policy makers don't appreciate the effect of bias on research.
- Common, weak designs either fall prey to biases or fail to control for their effects.
- We encourage the use of more visual data.
- Without some corrections, our field could lead to poor policy advice and adverse health outcomes.

Source: Soumerai SB et al. Prev Chronic Dis. 2015:12:E101.

PREVENTING CHRONIC DISEASE
PUBLIC HEALTH RESEARCH, PRACTICE, AND POLICY
Volume 12, E101 JUNE 2015

EDITOR'S CHOICE

# How Do You Know Which Health Care Effectiveness Research You Can Trust? A Guide to Study Design for the Perplexed

Stephen B. Soumerai, ScD; Douglas Starr, MS; Sumit R. Majumdar, MD, MPH

Soumerai et al. Prev Chronic Dis. 2016 Jun 23;13:E82.