Advancing Clinical Evidence

Evolving Roles of Randomized Trials and Observational Studies

David J. Cohen, M.D., M.Sc.

Director of Cardiovascular Reseach
Saint Luke's Mid America Heart Institute

University of Missouri- Kansas City
Kansas City, MO USA

Disclosures

Equity

- None

Grant Support/Drugs

- Eli Lilly - Schering-Plough
- Baxter

Grant Support/Devices

- Boston Scientific - Edwards Lifesciences
- MedRAD

Consulting/Advisory Boards

- Cordis - Medtronic

DJC: 4/09

Study Designs for Clinical Research

Weakest evid

- Single case report (anecdote)

study

trols

y

- Single randomized clinical trial
- Multiple large, randomized clinical trials

Strongest evidence

Challenge of Clinical Research:

To match each clinical question to the study design that will allow it to be answered in a practical, timely, and efficient manner

Randomized Clinical Trial

- Broadly defined as any controlled experiment involving human subjects, where treatment allocation is randomly assigned
- Originally designed for agricultural studies (Fisher)
- First medical RCT was a study of streptomycin treatment for pulmonary tuberculosis (BMJ 1948)

Why do we need RCTs?

- RCTs are the best available technique for eliminating bias in the assessment of a treatment effect
  - *Eliminates both measured and unmeasured confounding*
- With continued improvement in medical care, most treatment effects of interest in cardiovascular dz have only modest effects (RR reductions ~15-20%)
  - *Only RCTs can provide sufficient precision and confidence to reliably detect small benefits*
  - *Increasing emphasis on "large, simple trials" (>20K pts)*

Limitations of Clinical Trials

Only a finite # of clinical trials can be performed. Frequently, trial results may not apply to the particular patient or clinical situation in question

PCI: Anatomic/Patient Subsets

Anatomical factors

- 1,2,or 3-vessel disease
- Previous CABG
- Associated valve repair/replacement
- Aortic atherosclerosis/calcification

Patient factors

- Acute MI/Cardiogenic shock
- Comorbid conditions- renal failure, COPD, advanced age
- Diabetic

ORIGINAL CONTRIBUTION

Scientific Evidence Underlying the ACC/AHA Clinical Practice Guidelines

Pierluigi Tricoci, MD, MHS, PhD
Joseph M. Allen, MA
Judith M. Kramer, MD, MS
Robert M. Califf, MD
Sidney C. Smith Jr, MD

Context The joint cardiovascular practice guidelines of the American College of Cardiology (ACC) and the American Heart Association (AHA) have become important documents for guiding cardiology practice and establishing benchmarks for quality of care.

Objective To describe the evolution of recommendations in ACC/AHA cardiovascular guidelines and the distribution of recommendations across classes of recommendations and levels of evidence.

Data Sources and Study Selection Data from all ACC/AHA practice guidelines issued from 1984 to September 2008 were abstracted by personnel in the ACC Science and Quality Division. Fifty-three guidelines on 22 topics, including a total of 7196 recommendations, were abstracted.

Data Extraction The number of recommendations and the distribution of classes of recommendation (I, II, and III) and levels of evidence (A, B, and C) were determined. The subset of guidelines that were current as of September 2008 was evaluated to describe changes in recommendations between the first and current versions as well as patterns in levels of evidence used in the current versions.

Results Among guidelines with at least 1 revision or update by September 2008, the number of recommendations increased from 1330 to 1973 (+48%) from the first to the current version, with the largest increase observed in use of class II recommendations. Considering the 16 current guidelines reporting levels of evidence, only 314 recommendations of 2711 total are classified as level of evidence A (median, 11%), whereas 1246 (median, 48%) are level of evidence C. Level of evidence significantly varies across categories of guidelines (disease, intervention, or diagnostic) and across individual guidelines. Recommendations with level of evidence A are mostly concentrated in class I, but only 245 of 1305 class I recommendations have level of evidence A (median, 19%).

Conclusions Recommendations issued in current ACC/AHA clinical practice guidelines are largely developed from lower levels of evidence or expert opinion. The proportion of recommendations for which there is no conclusive evidence is also growing. These findings highlight the need to improve the process of writing guidelines and to expand the evidence base from which clinical practice guidelines are derived.

JAMA. 2009;301(8):831-841 www.jama.com

CLINICAL PRACTICE GUIDELINES are systematically developed statements to assist practitioners with decisions about appropriate health care for specific patients' circumstances.[1] Guidelines are often assumed to be the epitome of evidence-based medicine. Yet, guideline recommendations imply not only an evaluation of the evidence but also a value judgment based on personal or organizational preferences regarding the various risks and benefits of a medical intervention for a population.[2]

For more than 20 years, the American College of Cardiology (ACC) and the American Heart Association (AHA) have released clinical practice guidelines to provide recommendations on care of patients with cardiovascular disease. The ACC/AHA guidelines currently use a grading schema based on level of evidence and class of recommendation (available at http://www.acc.org and http://www.aha.org). The level of evidence classification combines an objective description of the existence and the types of studies supporting the recommendation and expert consensus, according to 1 of the following 3 categories:

- Level of evidence A: recommendation based on evidence from multiple randomized trials or meta-analyses
- Level of evidence B: recommendation based on evidence from a single randomized trial or nonrandomized studies
- Level of evidence C: recommendation based on expert opinion, case studies, or standards of care.

The class of recommendation designation indicates the strength of a recommendation and requires guideline writers not only to make a judgment

See also p 870 and Patient Page.

Author Affiliations: Division of Cardiology and Duke Clinical Research Institute (Dr Tricoci), Division of General Internal Medicine and Duke Center for Education and Research on Therapeutics (Dr Kramer), and Division of Cardiology and Duke Translational Medicine Institute (Dr Califf), Duke University, Durham, North Carolina; American College of Cardiology Science and Quality Division, Washington, DC (Mr Allen); and Center for Cardiovascular Science and Medicine, University of North Carolina, Chapel Hill (Dr Smith).
Corresponding Author: Pierluigi Tricoci, MD, MHS, PhD, Duke Clinical Research Institute, 2400 Pratt St, Room 0311, Terrace Level, Durham, NC 27705 (trico001@dcri.duke.edu).

- Reviewed all ACC/AHA practice guidelines from 1984-2008 (n=53 guidelines, 7196 recommendations)
- Levels of evidence in current guidelines
  - A (multiple RCTs)– 11%
  - B (single RCT or non- randomized studies only)– 41%
  - C (expert opinion or std of care)– 48%

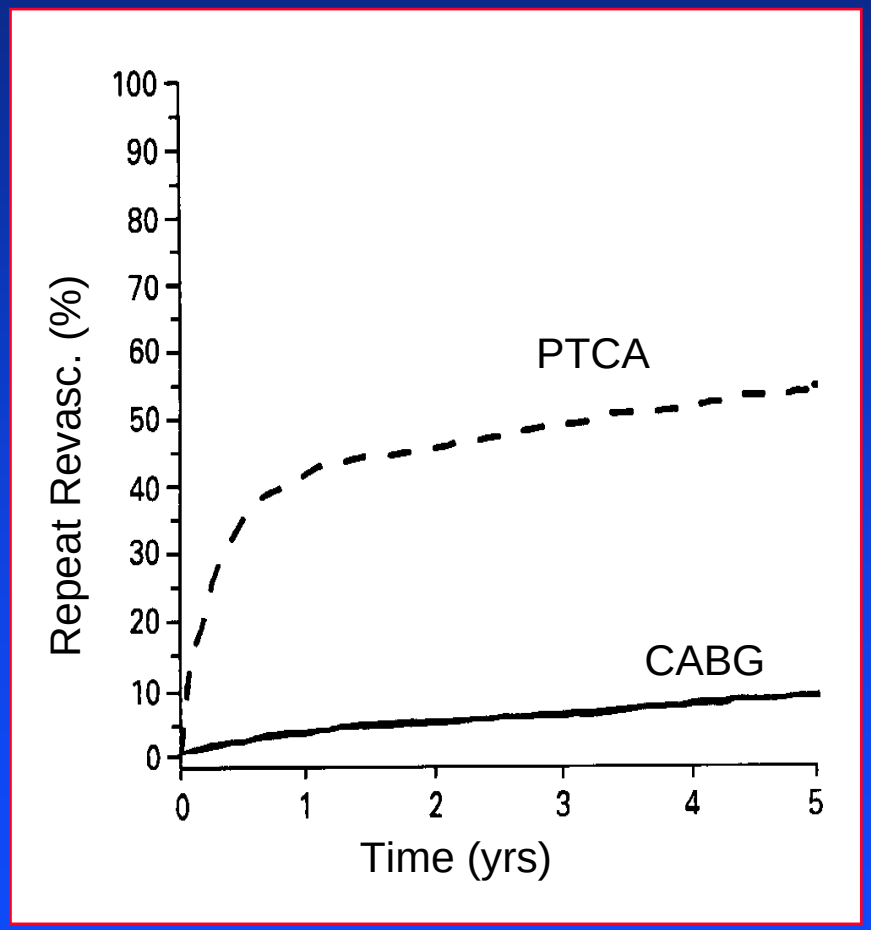
Trioci P, et al. JAMA 2009;301:831-41

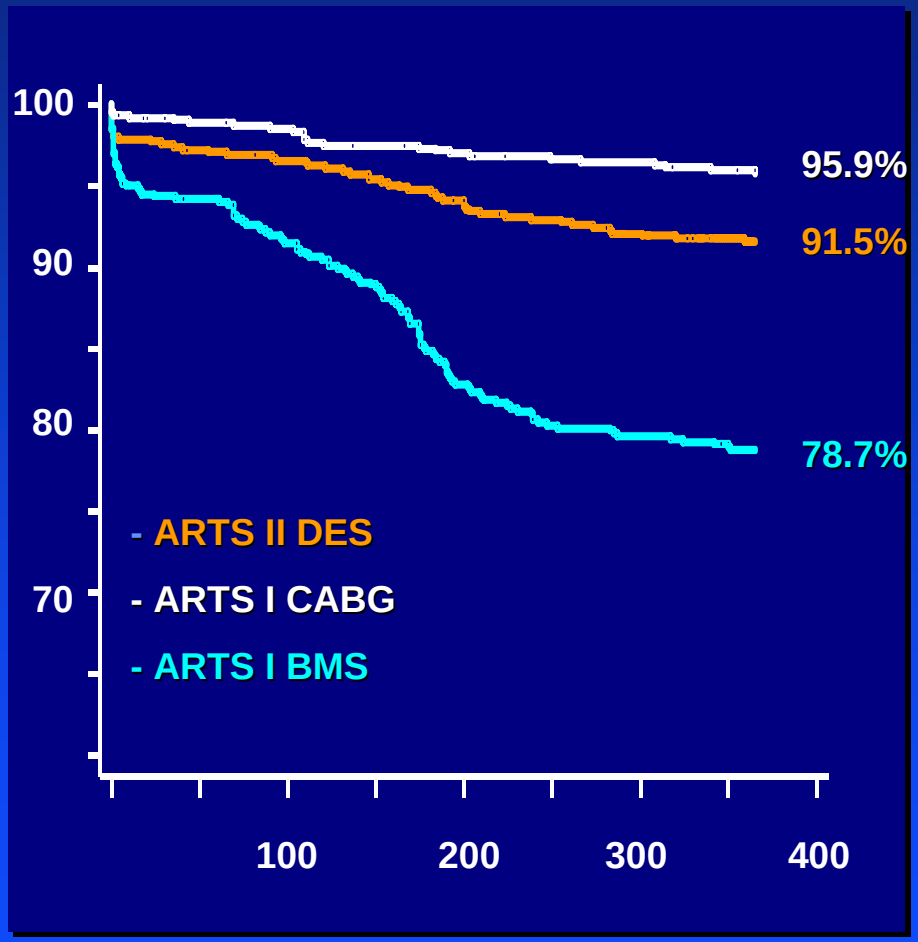
Limitations of Clinical Trials

Obsolescence

- RCT's are best suited to evaluation of "mature" treatments
- Clinical trials are a poor way to evaluate rapidly changing technologies and standards of care→ particularly problematic for medical devices
- Trials are particularly vulnerable when enrollment is slow or the follow-up duration is long

BARI: Repeat Revascularization



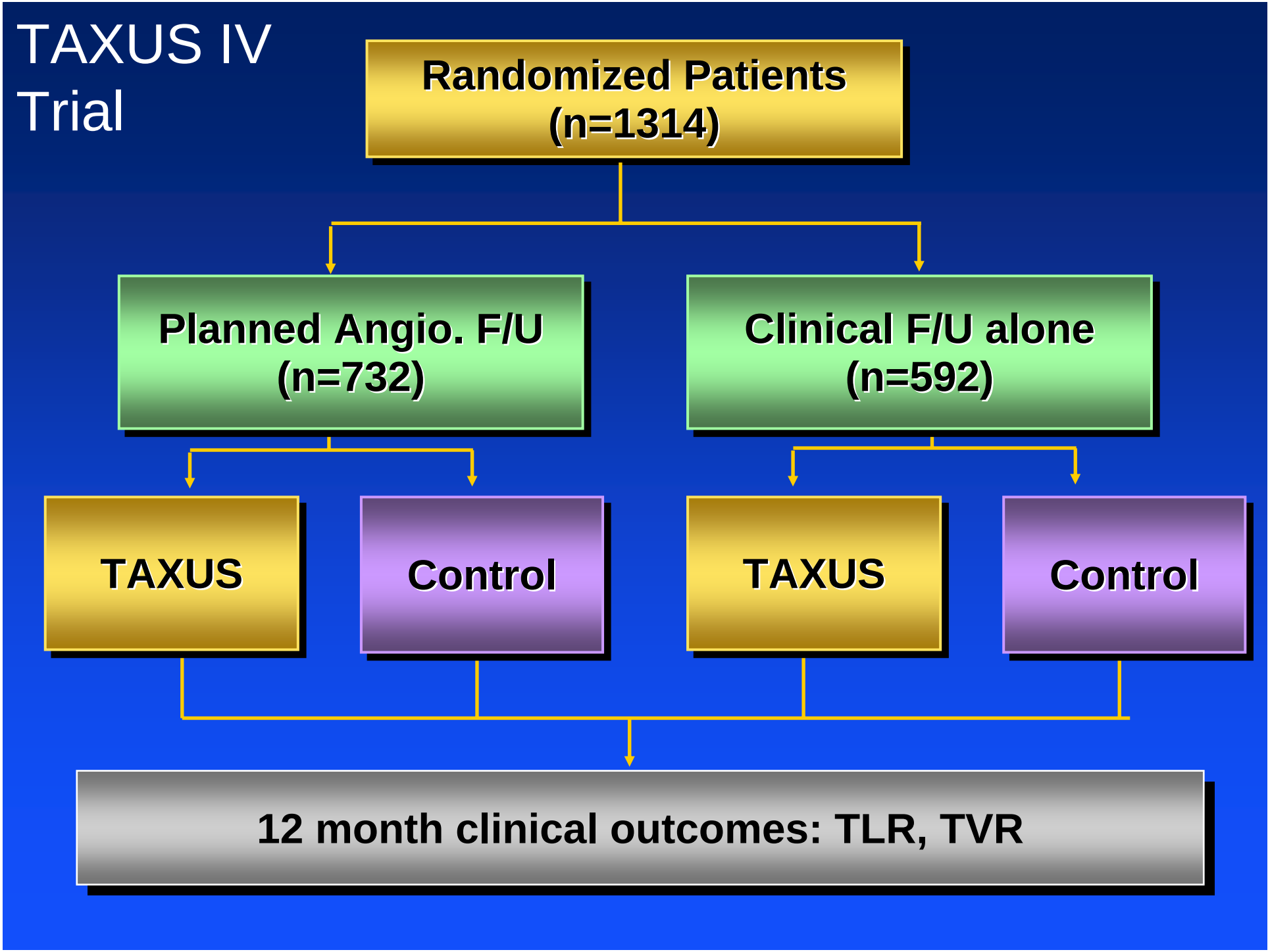


Limitations of Clinical Trials

Protocol-Driven Care

- For a variety of reasons (regulatory, safety, mechanistic research, etc.), clinical trials often impose additional diagnostic tests that do not occur in routine clinical practice
- Under certain circumstances, these tests may substantially bias the evaluation of clinical outcomes of interest

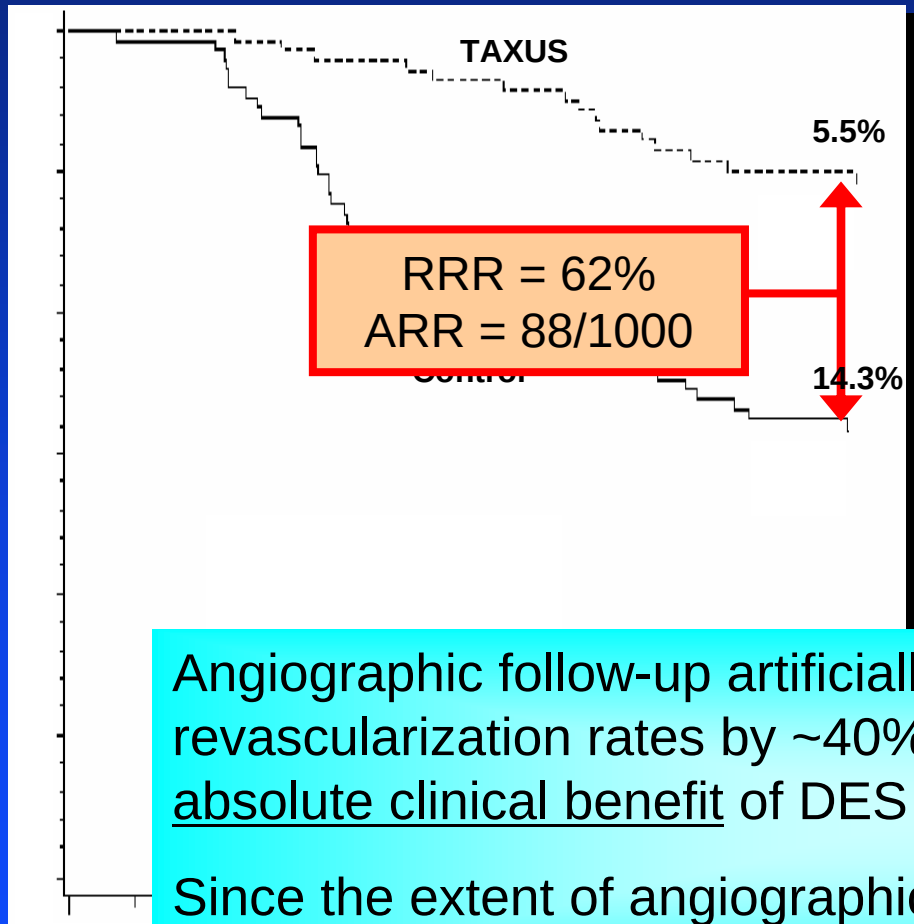
TAXUS IV Trial



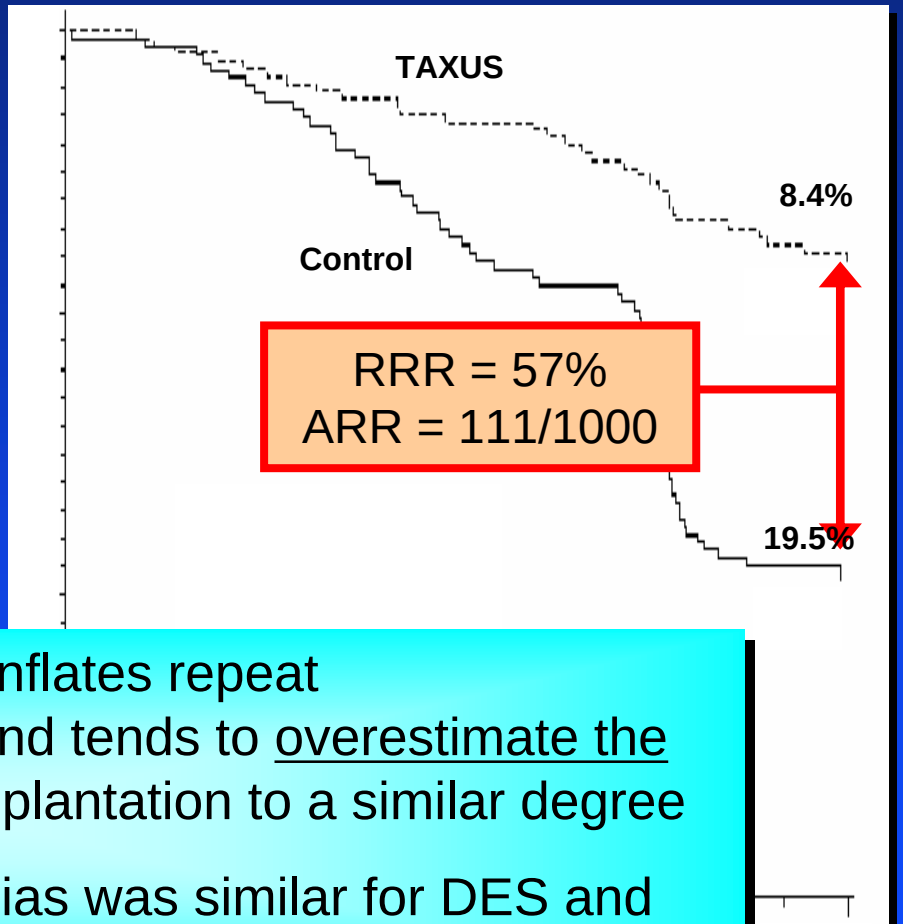
TAXUS IV

Impact of Angiographic F/U on Clinical Benefit of DES

Clinical F/U Alone



Angiographic F/U



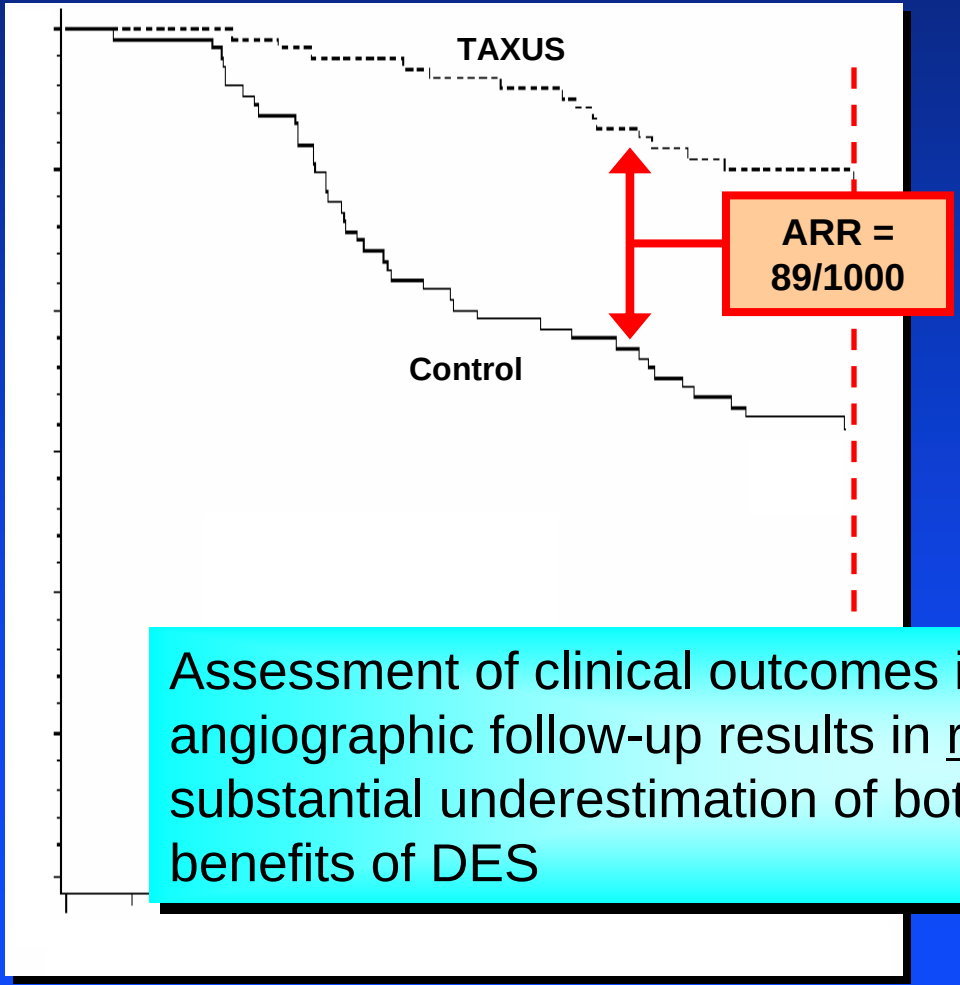
Angiographic follow-up artificially inflates repeat revascularization rates by ~40% and tends to overestimate the absolute clinical benefit of DES implantation to a similar degree

Since the extent of angiographic bias was similar for DES and BMS, however, the relative risk reduction is unaffected

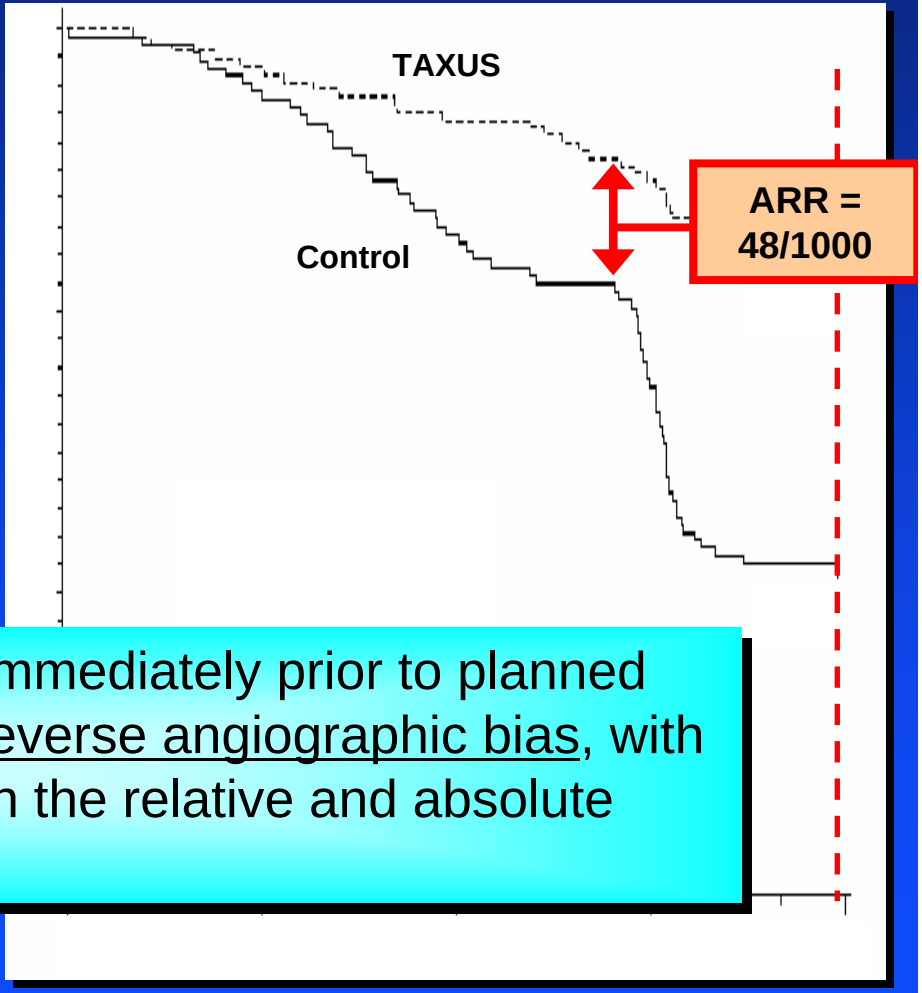
TAXUS IV

Impact of Angiographic F/U on Clinical Benefit of DES

Clinical F/U Alone



Angiographic F/U



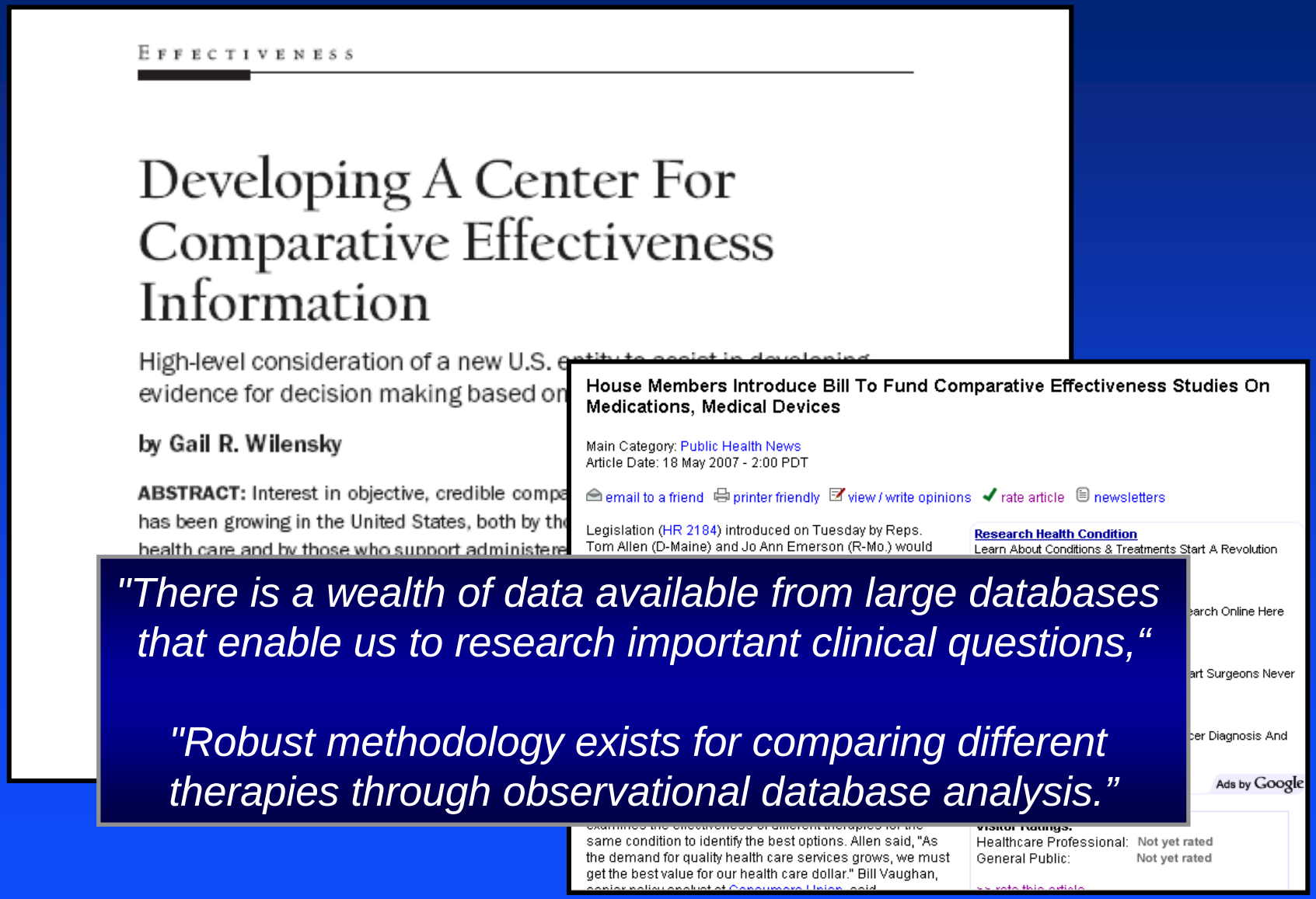
Assessment of clinical outcomes immediately prior to planned angiographic follow-up results in reverse angiographic bias, with substantial underestimation of both the relative and absolute benefits of DES

Additional Limitations of RCTs

- Often underpowered for modest treatment effects
  - *Still relevant from public health standpoint if affected population is large*
- Surrogate endpoints→ ? Clinical relevance
- Generalizability?
  - *Tend to study generally healthy patients*
  - *Treated with standardized protocols*
  - *By experienced providers*
- Certain questions not easily subject to RCT
  - *Unethical, impractical, no business case, or*
  - *Studies of harmful effects*

Can we use observational studies (registries) for clinical evidence development?

Comparative Effectiveness



EFFECTIVENESS

Developing A Center For Comparative Effectiveness Information

High-level consideration of a new U.S. e
evidence for decision making based on

by Gail R. Wilensky

ABSTRACT: Interest in objective, credible compa
has been growing in the United States, both by th
health care and by those who support administere

House Members Introduce Bill To Fund Comparative Effectiveness Studies On Medications, Medical Devices

Main Category: Public Health News
Article Date: 18 May 2007 - 2:00 PDT

email to a friend | printer friendly | view / write opinions | rate article | newsletters

Legislation (HR 2184) introduced on Tuesday by Reps. Tom Allen (D-Maine) and Jo Ann Emerson (R-Mo.) would

Research Health Condition
Learn About Conditions & Treatments Start A Revolution

earch Online Here

art Surgeons Never

cer Diagnosis And

Ads by Google

same condition to identify the best options. Allen said, "As the demand for quality health care services grows, we must get the best value for our health care dollar." Bill Vaughan,

Visitor Ratings:
Healthcare Professional: Not yet rated
General Public: Not yet rated

"There is a wealth of data available from large databases that enable us to research important clinical questions,"

"Robust methodology exists for comparing different therapies through observational database analysis."

Wilensky G Health Affairs Nov 2006:w572-w588

Prospective Multicenter Registry

- Study population- broad group of pts with same problem or undergoing same treatment
- Treatment according to local practice or physician preference
- All patients followed prospectively to assess for endpoints of interest

Registry Studies: Key Advantages

- Allows for *rapid enrollment* of large numbers of patients → accomodates changes in practice over time
- Broad inclusion criteria ensure that study's findings may be *applicable to most patients*
- Ideal for determining optimal procedural technique as well as for identifying appropriate patient subsets for treatment

Registry Studies: Key Disadvantages

Data quality and completeness

- Analysis results only as solid as the data ("Bad data in…")
- Particularly challenging with administrative datasets
- Incomplete data → rarely missing at random
- Not necessarily related to registry design, but more related to degree of rigor employed in data collection

Treatment selection bias

- Pt Level: risk factors, disease severity, comorbidity
- MD level: those selecting a specific treatment may differ in care process and quality
- Site-level: structural and quality of care differences

Techniques for Overcoming Selection Bias

- Regression modeling
  - *Adjust results directly for 'confounding factors' associated with treatment and outcome*
- Propensity adjustment
  - *Identify factors associated with treatment selection*
  - *Then adjust for the probability of treatment (propensity score) or match patients for this factor*
- Newer approaches
  - *Instrumental variables analysis*

Comparability of RCT and Observational Studies

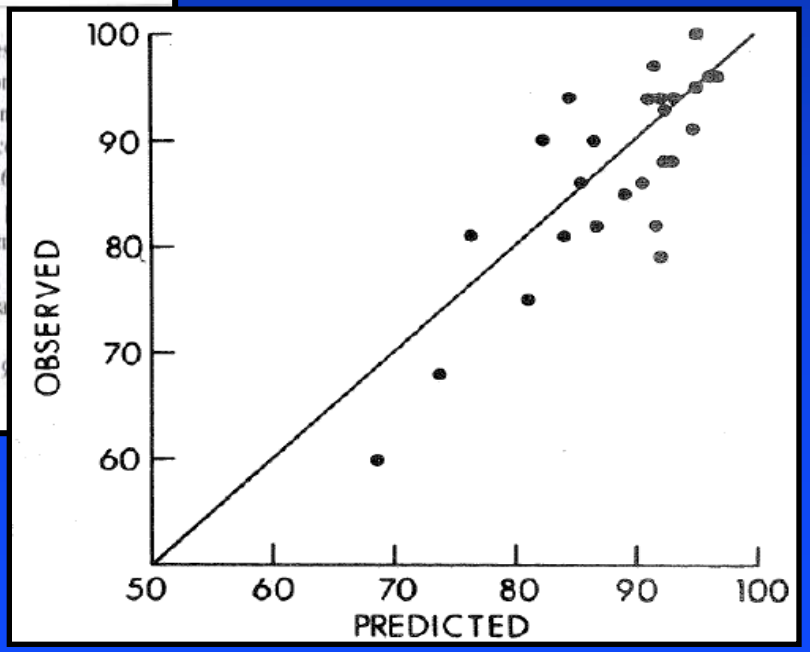
Comparison of Predictions Based on Observational Data With the Results of Randomized Controlled Clinical Trials of Coronary Artery Bypass Surgery

MARK A. HLATKY, MD, FACC,* ROBERT M. CALIFF, MD,* FRANK E. HARRELL, JR., PHD,†
KERRY L. LEE, PHD,† DANIEL B. MARK, MD, MPH,* DAVID B. PRYOR, MD, FACC*
Durham, North Carolina

Clinical decisions are most secure when based on findings from several large randomized clinical trials, but relevant randomized trial data are often unavailable. Analyses using clinical data bases might provide useful information if statistical methods can adequately correct for the lack of randomization. To test this approach, the findings of the three major randomized trials of coronary bypass surgery were compared with predictions of multivariable statistical models derived from observations in the Duke Cardiovascular Disease Databank. Clinical characteristics of patients at Duke University Medical Center who met eligibility requirements for each major randomized trial were used in the models to predict 5 year survival rate medical and surgical therapy in each rando

Model predictions agreed well with ra results and were within the 95% confidenc observed survival rates in 24 (92%) of 2 groups. The overall correlation between observed survival rates was good (Spearm 0.73, $p < 0.0001$). These results suggest performed analyses of observational data ca the results of randomized trials.

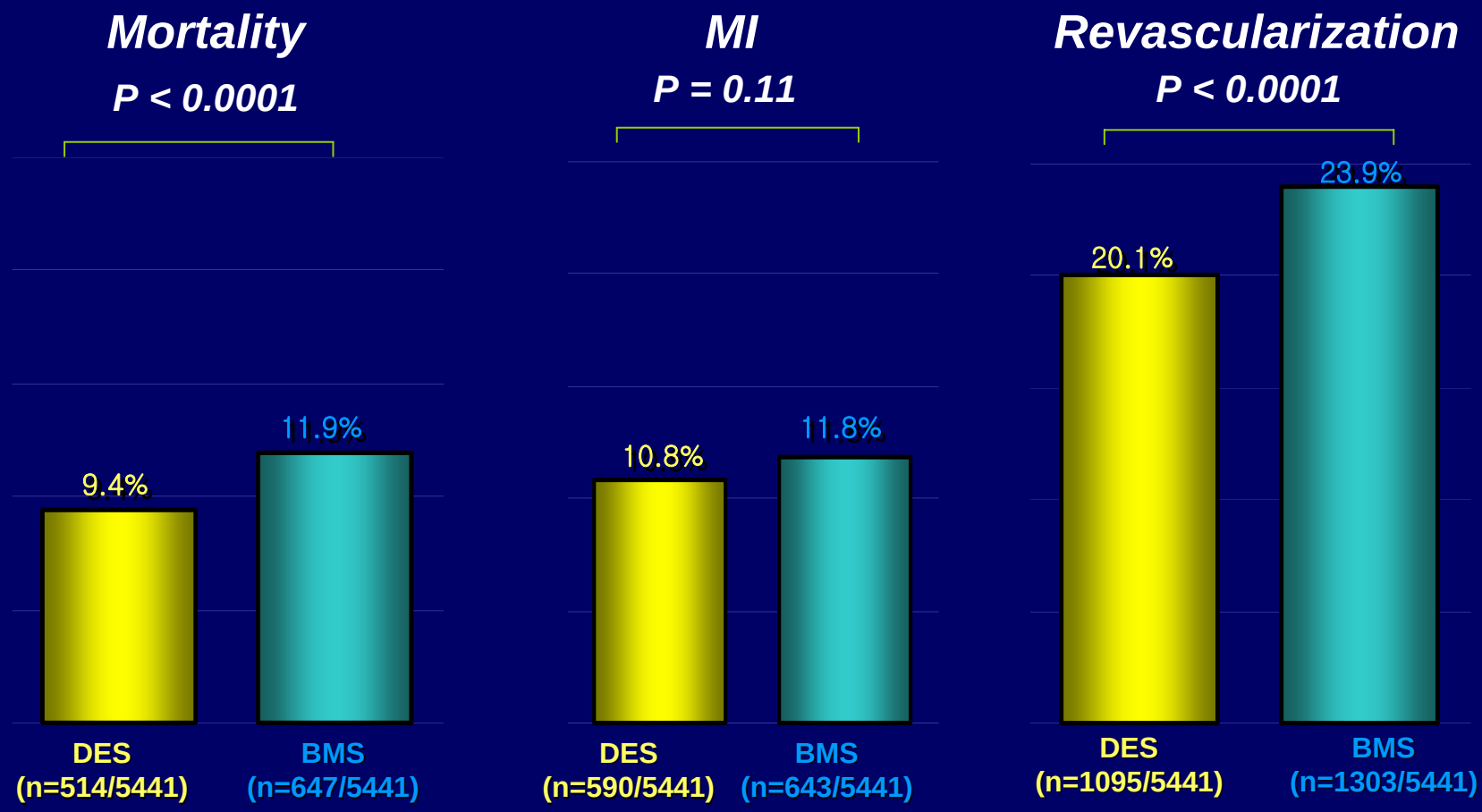
(J Am Coll Cardiol 1



JACC 1988:11:237-45

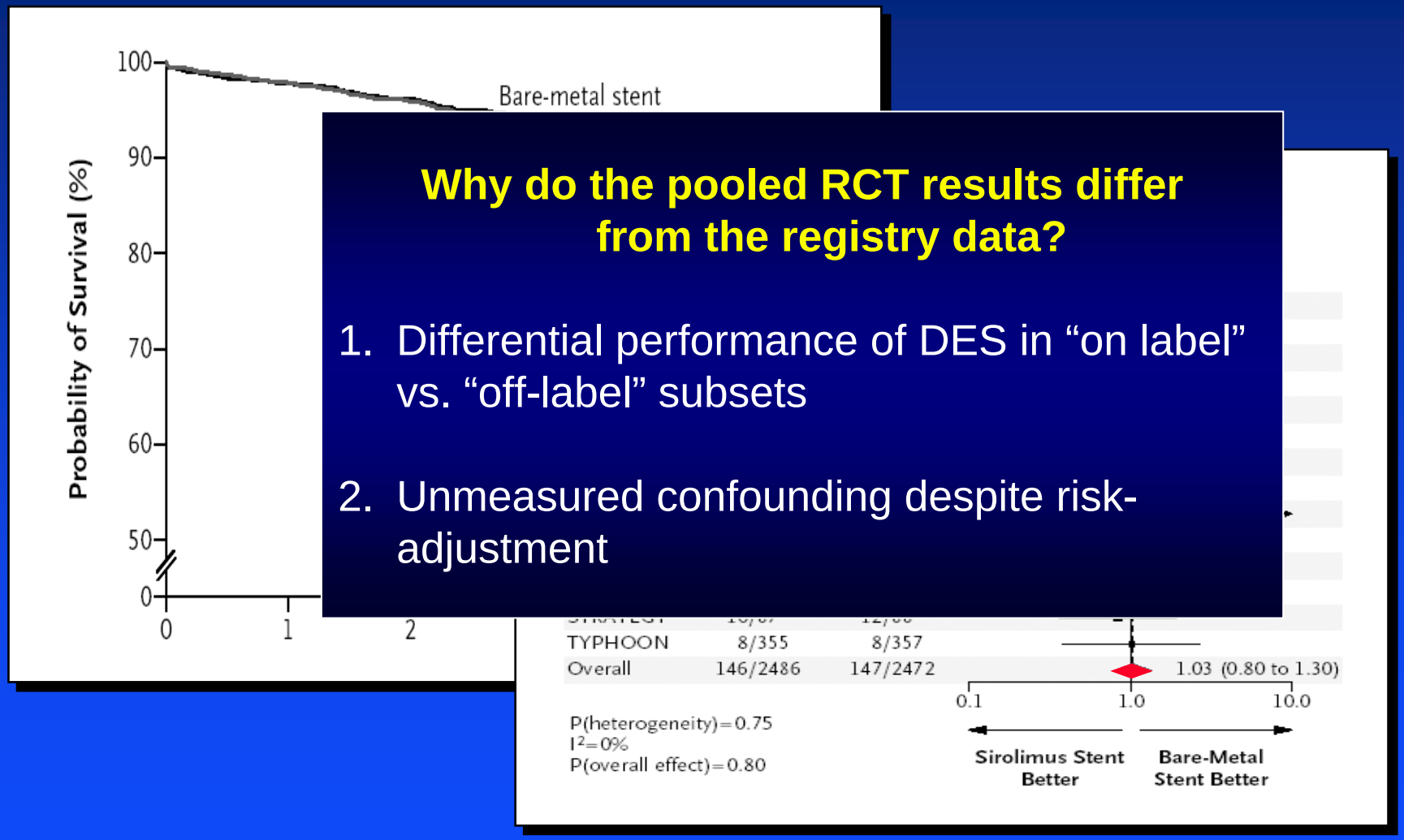
Drug-Eluting and Bare Metal Stenting in Massachusetts, Primary Results

Propensity Matched 2-Year Outcomes



Mauri L, et al. Circulation 2008;118:1817-27

Do Drug-Eluting Stents Save Lives?
Pooled RCT Results



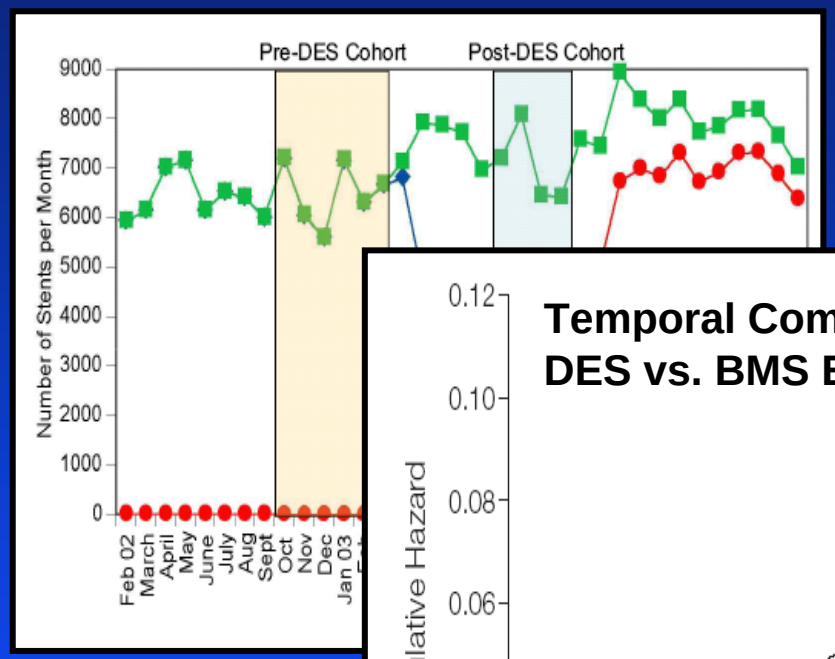
Why do the pooled RCT results differ from the registry data?

1. Differential performance of DES in “on label” vs. “off-label” subsets
2. Unmeasured confounding despite risk-adjustment

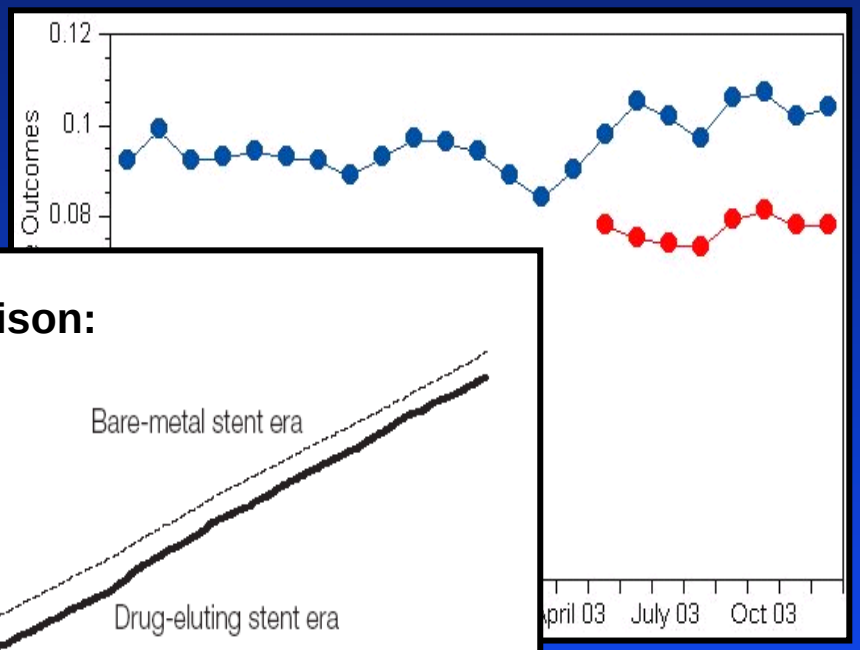
Kastrati et al. NEJM 2007; 356:1020-9

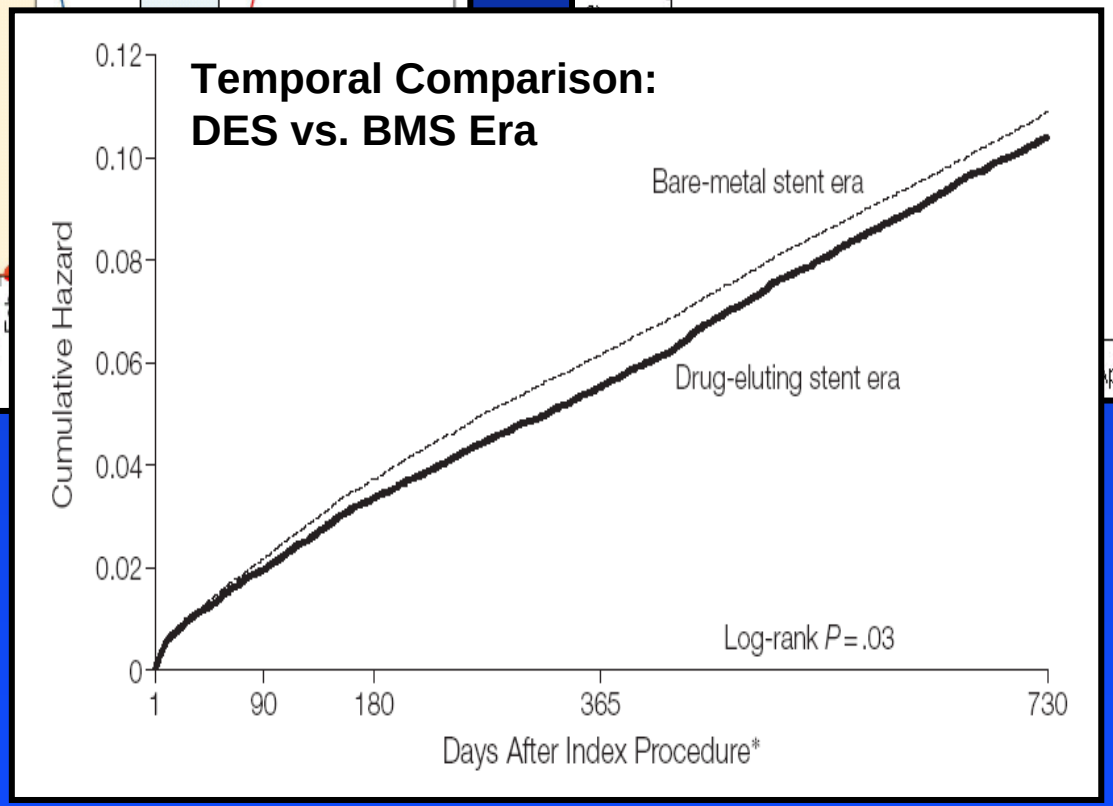
Alternate Approach: Time Series Comparison

Stent Type



2-Year Death or MI



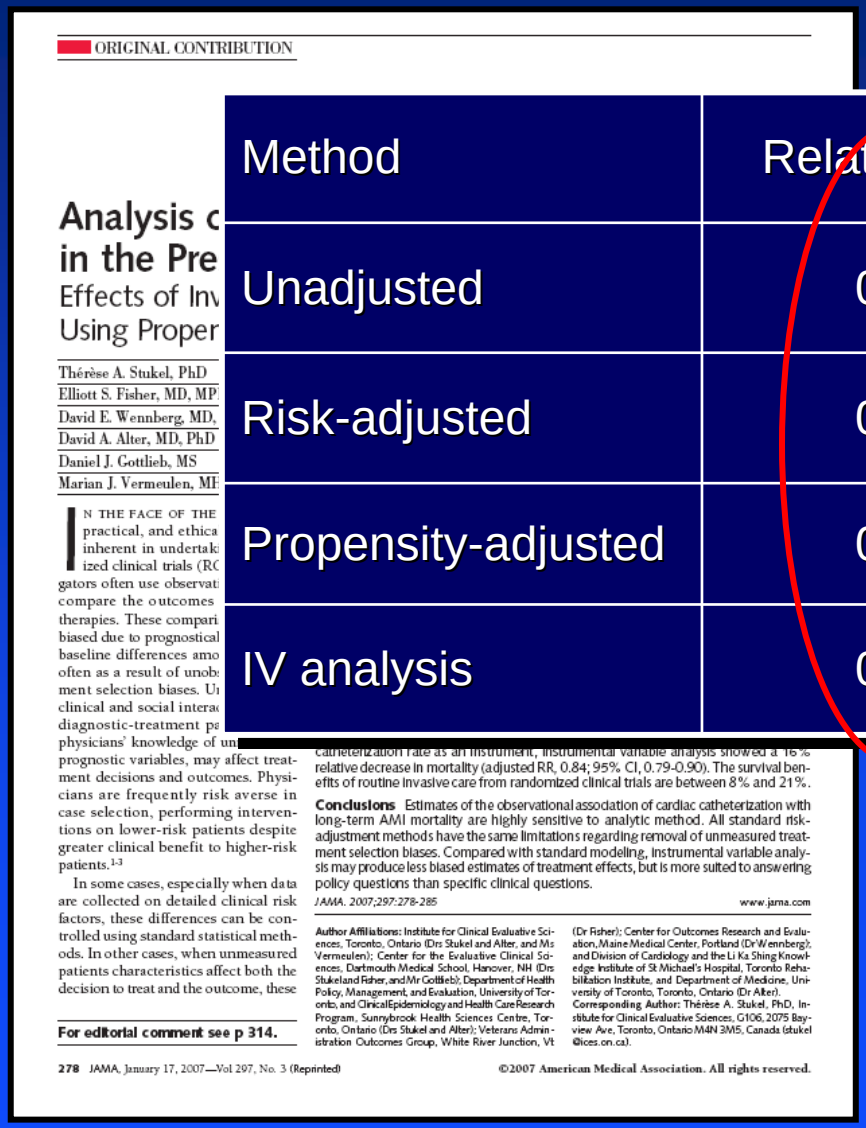


Malenka DJ, et al *JAMA.* 2008;299:2868-2876

Alternative Approach #2
Instrumental Variable Analysis

- Segregate patients by presence or absence of an "instrumental variable"
  - A factor which is correlated with the variable of interest, but is otherwise not associated with any other patient characteristic or treatment variable
  - "Natural experiment" or "Quasi-randomized" design
  - Main challenge: Can we identify an appropriate instrumental variable?

Impact of Residual Confounding:
Instrumental Variable Analysis vs. Risk-Adjustment



ORIGINAL CONTRIBUTION

Analysis c
in the Pre
Effects of Inv
Using Proper

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

IN THE FACE OF THE practical, and ethica inherent in undertak ized clinical trials (RC gators often use observat compare the outcomes therapies. These compari biased due to prognostical baseline differences amo often as a result of unob ment selection biases. U clinical and social intera diagnostic-treatment p physicians' knowledge of un prognostic variables, may affect treatment decisions and outcomes. Physicians are frequently risk averse in case selection, performing interventions on lower-risk patients despite greater clinical benefit to higher-risk patients.[1-3]

In some cases, especially when data are collected on detailed clinical risk factors, these differences can be controlled using standard statistical methods. In other cases, when unmeasured patients characteristics affect both the decision to treat and the outcome, these

For editorial comment see p 314.

catheterization rate as an instrument, instrumental variable analysis showed a 16% relative decrease in mortality (adjusted RR, 0.84; 95% CI, 0.79-0.90). The survival benefits of routine invasive care from randomized clinical trials are between 8% and 21%.

Conclusions Estimates of the observational association of cardiac catheterization with long-term AMI mortality are highly sensitive to analytic method. All standard risk-adjustment methods have the same limitations regarding removal of unmeasured treatment selection biases. Compared with standard modeling, instrumental variable analysis may produce less biased estimates of treatment effects, but is more suited to answering policy questions than specific clinical questions.

JAMA. 2007;297:278-285 www.jama.com

Author Affiliations: Institute for Clinical Evaluative Sciences, Toronto, Ontario (Drs Stukel and Alter, and Ms Vermeulen); Center for the Evaluative Clinical Sciences, Dartmouth Medical School, Hanover, NH (Drs Stukel and Fisher, and Mr Gottlieb); Department of Health Policy, Management, and Evaluation, University of Toronto, and Clinical Epidemiology and Health Care Research Program, Sunnybrook Health Sciences Centre, Toronto, Ontario (Drs Stukel and Alter); Veterans Administration Outcomes Group, White River Junction, Vt (Dr Fisher); Center for Outcomes Research and Evaluation, Maine Medical Center, Portland (Dr Wennberg); and Division of Cardiology and the Li Ka Shing Knowledge Institute of St Michael's Hospital, Toronto Rehabilitation Institute, and Department of Medicine, University of Toronto, Toronto, Ontario (Dr Alter).
Corresponding Author: Thérèse A. Stukel, PhD, Institute for Clinical Evaluative Sciences, G106, 2075 Bayview Ave, Toronto, Ontario M4N 3M5, Canada (stukel@ices.on.ca).

Method	Relative Risk	95% CI
Unadjusted	0.36	0.36-0.37
Risk-adjusted	0.51	0.50-0.52
Propensity-adjusted	0.52	0.51-0.53
IV analysis	0.84	0.79-0.90

Does
educe
after AMI?

AMI pts

et with >65

e analysis
ath rate,
3% to 65%
(quasi-natural experiment)

Implications: Even high-quality observational analyses often suffer from substantial residual confounding

Characteristics of an Ideal Registry

- Clinical (not just administrative) data
  - *Critical for risk adjustment*
- Prospective data collection
  - *Generally higher quality data; less subject to recall bias*
- All-inclusive population (rather than convenience sample) with clearly defined intake mechanism
  - *Critical to validity and generalizability*
- Problem-based or disease-based (as opposed to treatment/technology based)
  - *Cannot establish value of a technology by examining it in a vacuum*

Summary: RCTs vs. Registries

- If randomization an option, it is still by far the best and most definitive approach to developing unbiased, reliable evidence
- Nonetheless, gaps will continue to exist in our evidence base
  - *No trials*
  - *Non-representativeness (lack of generalizability)*
  - *Artificial nature of trial protocol*
- With careful planning and analysis, observational treatment comparisons can supplement our evidence development
  - *Hypothesis generating, confirmatory, extension of trials to understudied subsets*
  - *Must be careful consumers → some treatment comparisons may not be possible in observational data (at least with traditional methods to adjust for confounding)*