



Must We Be Misled By Observational Studies?

Thomas E. Love, Ph.D.

thomaslove@case.edu

May 19, 2005 - Grand Rounds

The view is widely held that randomized controlled trials are the "gold standard" for evaluation and that observational methods (cohort and case control studies) have little or no value. This ignores the limitations of randomized trials, which may prove unnecessary, inappropriate, impossible, or inadequate.... The false conflict (between the two approaches) needs to be replaced with mutual recognition of their complementary roles. ... Researchers should be united in their quest for scientific rigor in evaluation, regardless of the method used. Black (1996)

HRT and Dementia

- August 2001:
Am J Obstet Gynecol
“Estrogen and HRT users have ... a 20% to 60% reduction in the risk of Alzheimer’s disease.”
- March 2005:
Lancet Neurol
“Estrogen with or without progestin, given to women 65 years and older ... substantially increases the risk of dementia of any cause and cognitive decline.”

HRT and Cardiovascular Disease

- 1985 [NHS]
“... estrogen reduces the risk of severe CHD.”
- 1997 JAMA
“HRT should increase life expectancy for nearly all postmenopausal women...”
- July 2002 JAMA [WHI Trial]
“(HRT) should not be initiated or continued for primary prevention of coronary heart disease.”

Pulmonary Artery Catheterization in the Critically Ill

- RHC saves lives.
Many MDs believe RHC is an necessary guide for therapy, and that RHC leads to better outcomes.
Can't even execute an RCT (equipoise)...
- SUPPORT study:
 - Careful OS with special attention to potential biases.
 - RHC “associated with increased mortality and utilization.”

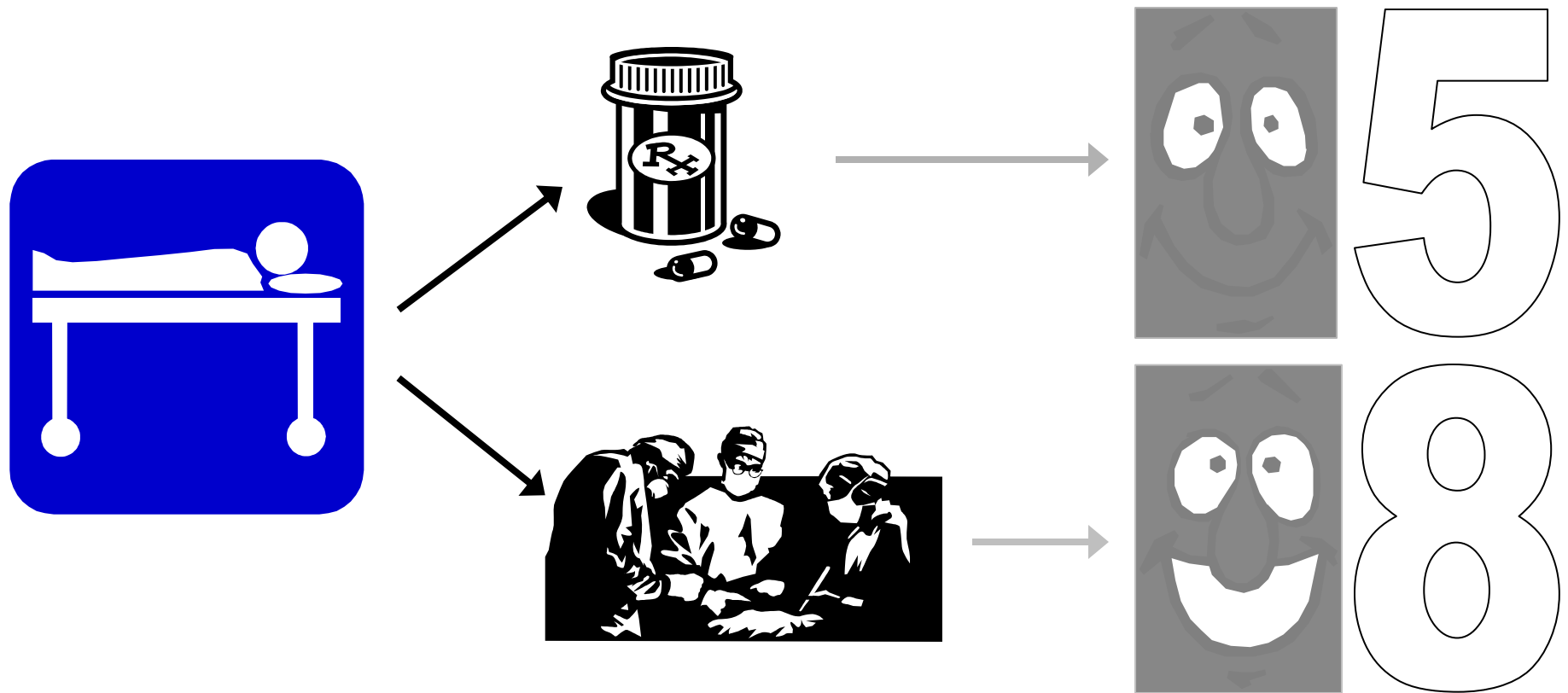
How Can We Avoid Being Misled?

1. What differentiates an observational study from a randomized controlled trial?
2. What is selection bias, and why should I care about it?
3. What can be done to deal with selection bias in observational studies?

Teasing Out Cause and Effect: Comparison of Potential Outcomes

- The causal effect of a treatment is based on a comparison of two potential outcomes.
 - Outcome patient would have if treated.
 - Outcome patient would have if untreated.
 - Causal effect = Treated – Untreated difference
- Problem: We only get to observe **one** of these two potential outcomes.

Causal Treatment (Exposure) Effects in terms of Potential Outcomes



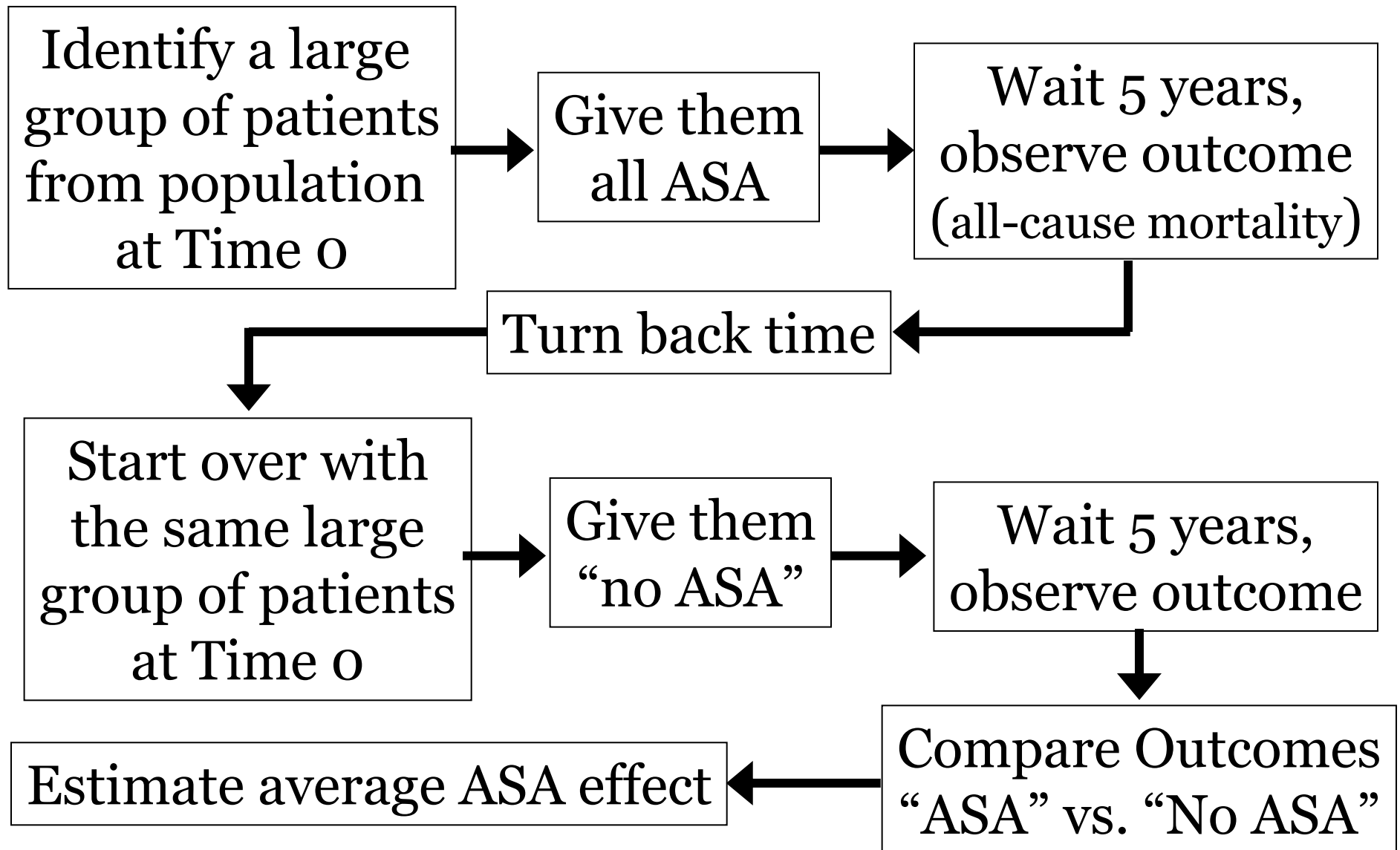
Surgery-Drug Treatment Effect =

3

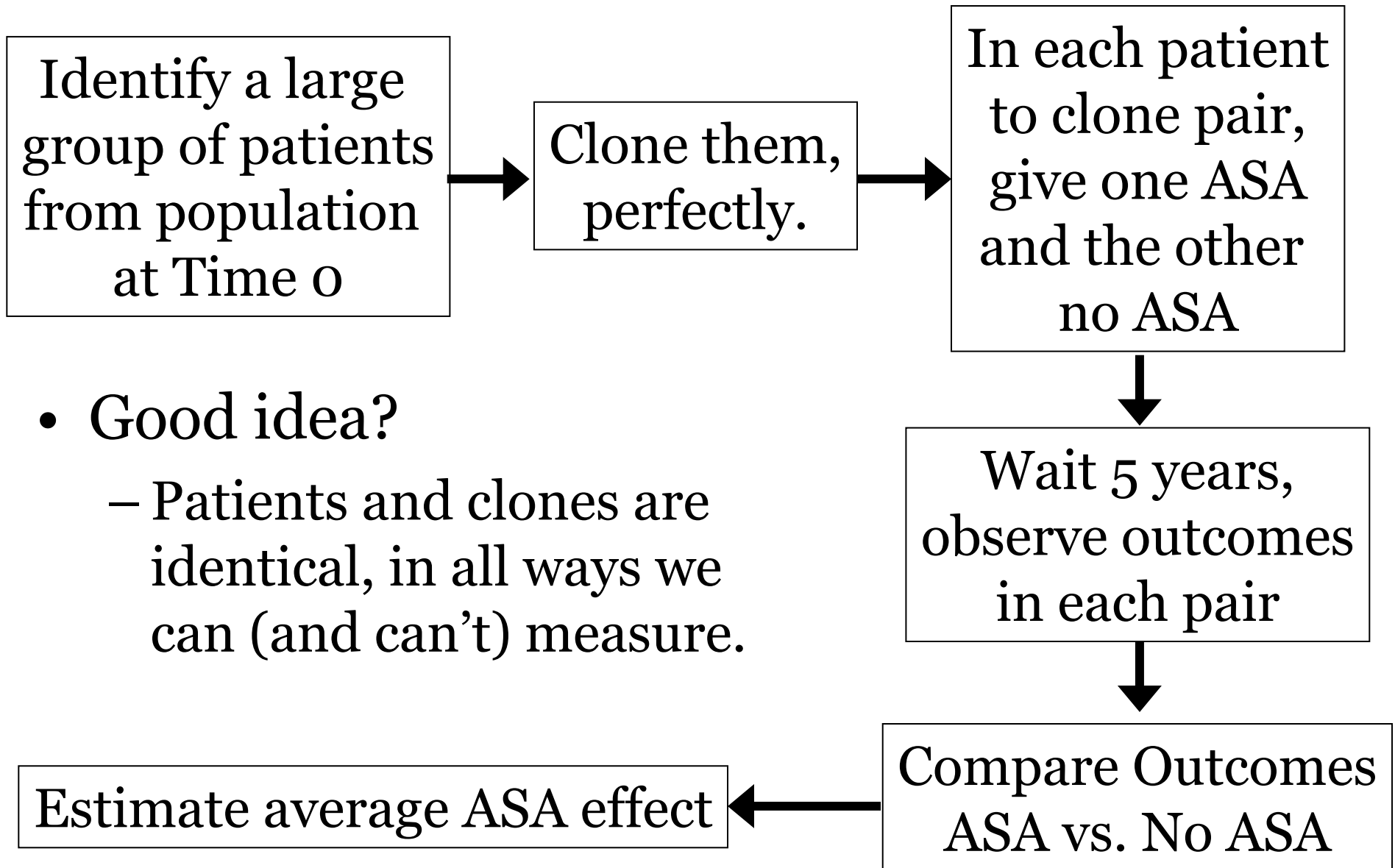
ASA and Mortality in Heart Patients

- Suppose you want to know aspirin's effect on all-cause 5-yr mortality among patients undergoing stress echocardiography.
 - Comparing “ASA” to “no ASA”
- Having identified a set of patients, what would be the ideal study?
- What would be the best practical study?

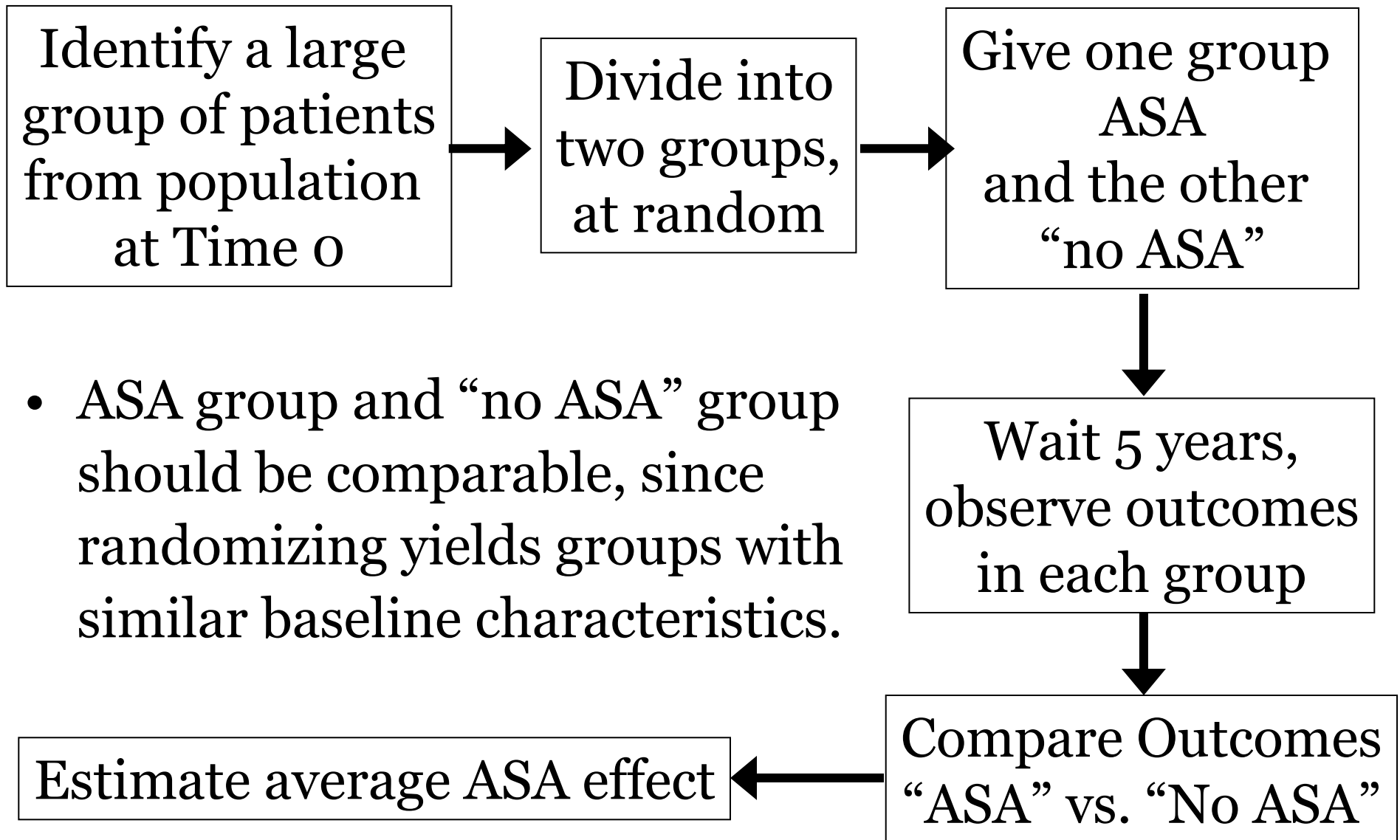
ASA and Mortality: Ideal Study



ASA & Mortality: Second Best Study



ASA & Mortality: RCT



ASA & Mortality: An Observational Study

Identify a large group of patients from population at Time 0

Some will take ASA, others won't

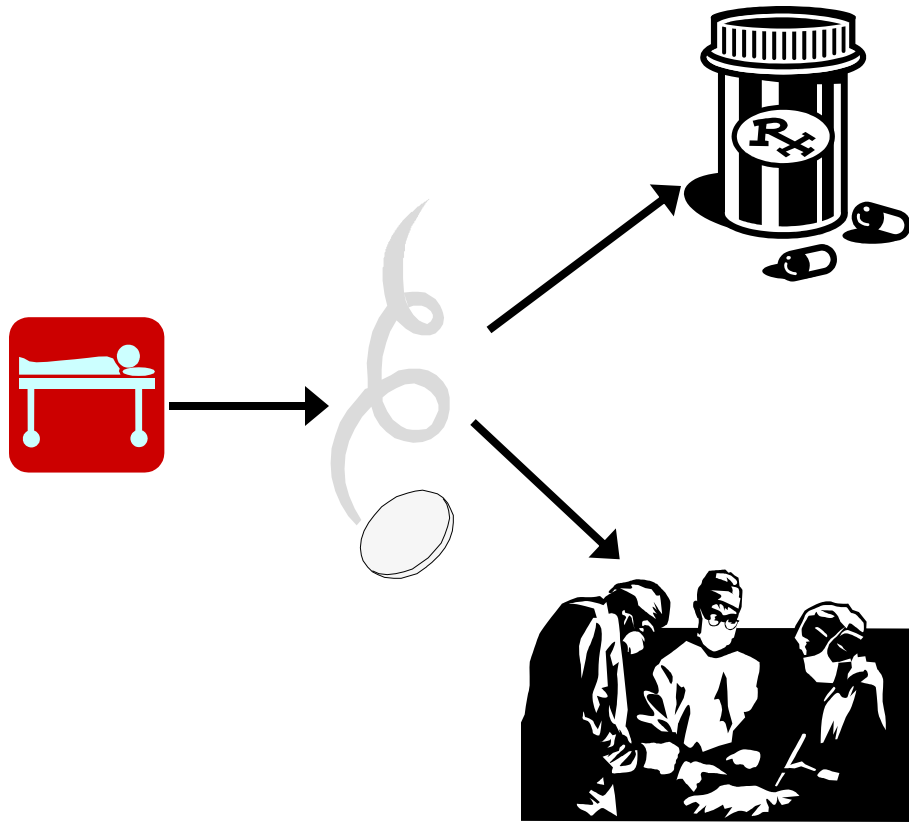
- ASA group and “no ASA” group may not be comparable, since they may not have similar baseline characteristics.

Wait 5 years, observe outcomes in each group

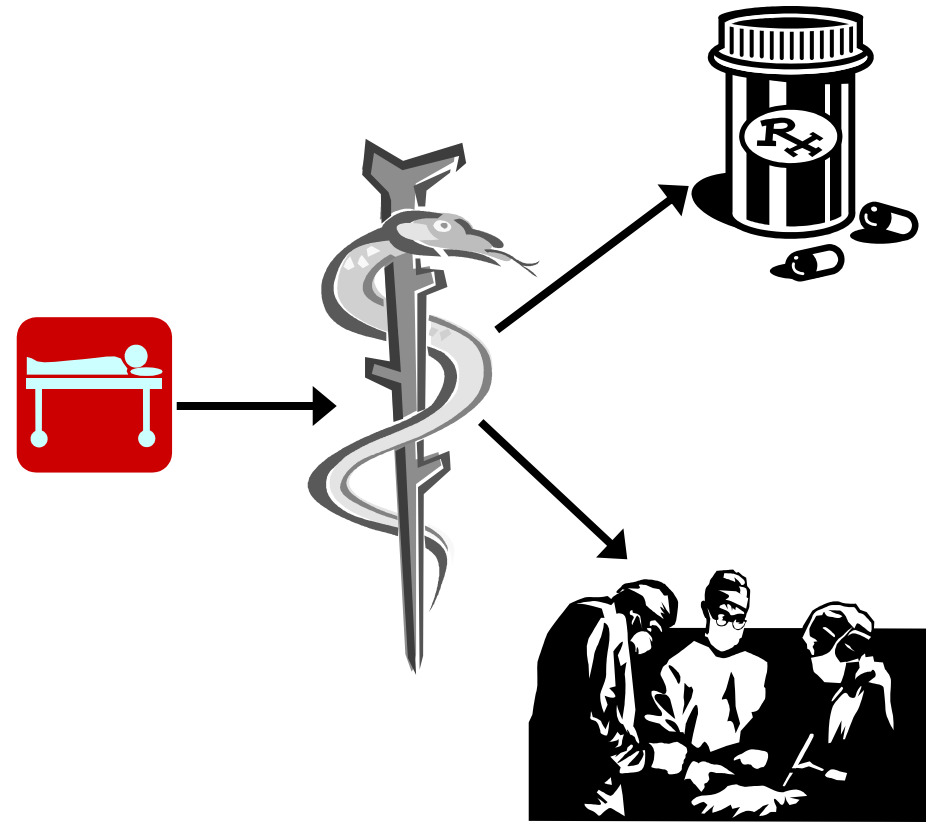
Compare Outcomes “ASA” vs. “No ASA”

Estimate average ASA effect

Randomized vs. Observational Studies



Randomization ensures that subjects receiving different treatments are comparable.

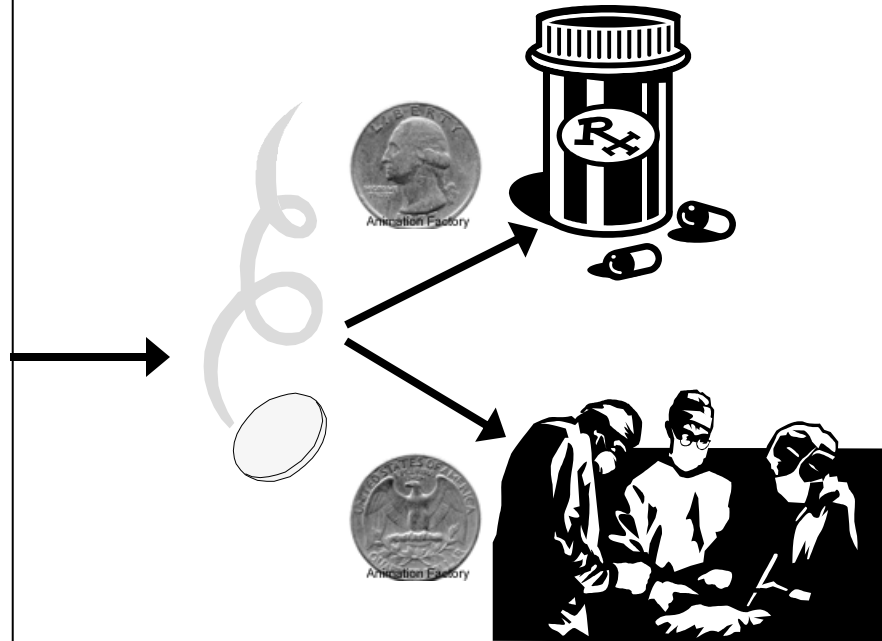
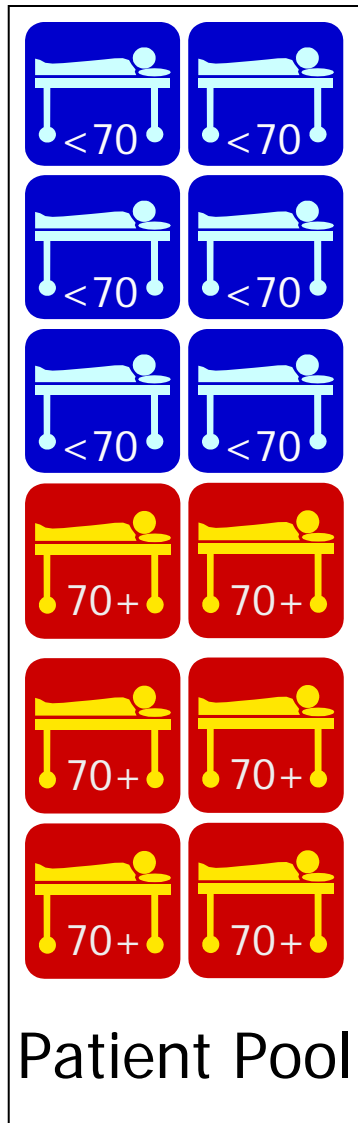


In observational studies, the researcher does not randomly allocate the treatments.

Randomization Tends To Balance Out Baseline Characteristics

- Randomization provides us with reason to expect that the two treatment groups will have similar distributions of
 - characteristics we observe or measure AND
 - characteristics we don't (or can't) measure.

“Balancing” in RCTs

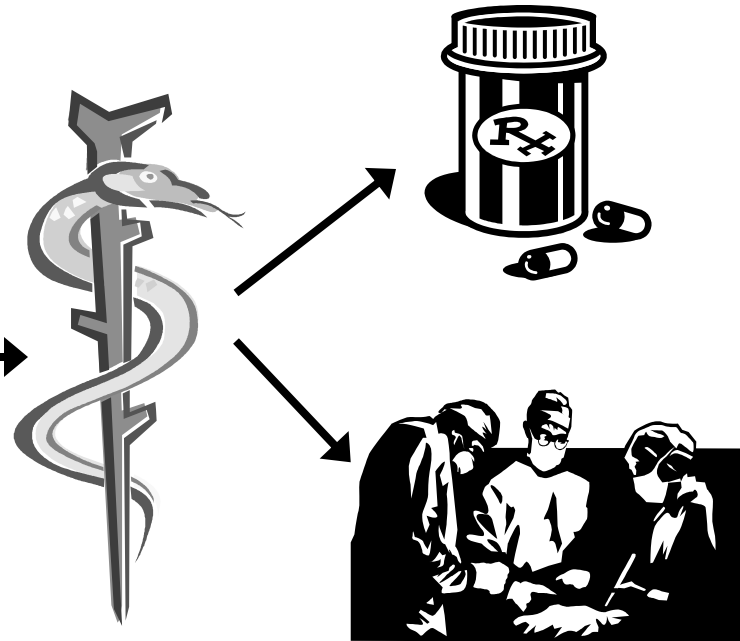
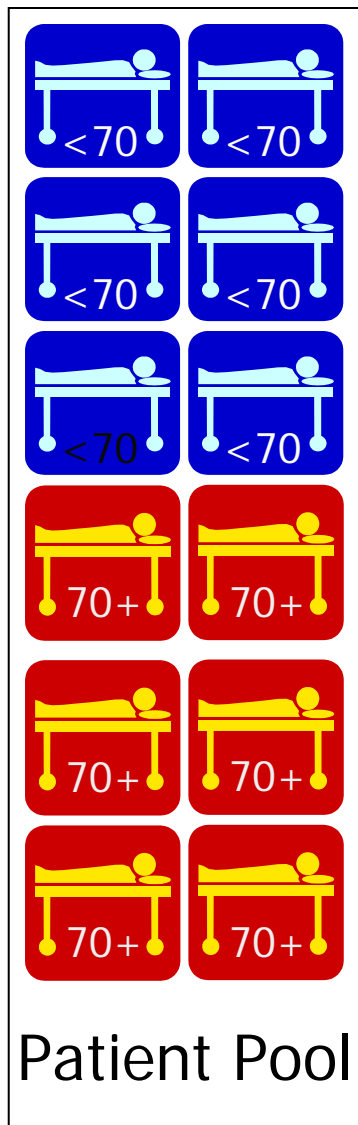


- Suppose we let the coin decide who gets surgery and who gets medicine.
- Randomization balances the baseline characteristics of the two groups.

The Big Problem in Observational Studies: Selection Bias

- We want to compare groups receiving the two treatments who looked similar prior to the treatment assignment.
- If there are differences between groups before treatment assignment that affect outcomes, we have selection bias.

Selection Bias in Observational Studies



- Suppose physician decides “only patients <70 will do well with surgery.”
- Enormous selection bias whenever age is related to the outcome of interest.

Advantages of Smart Observational Studies

- Address chief criticism of RCTs - limited generalizability / external validity
- Enable examination of exposure in “real life”
- May enable examination of effect size and “entrenched practices”
- Broader array of exposures and outcomes can be explored with an observational study than in controlled experiments.

**Pulmonary Artery
Catheterization
(Right Heart Catheterization)
in Critically Ill Patients**

**Or ... How a well-designed
observational study can bring
strong evidence to bear on
“conventional wisdom”**

Why Not Do A Randomized Controlled Trial of RHC?

- Procedure is very popular – Equipoise?
 - RHC directly measures cardiac function – lots of reasons to think this would be helpful.
 - Physician makes the decision – some could not ethically participate.
 - RHC effect sizes on survival, other outcomes may be small – need large, powerful RCT.
- 1989-94 Observational Study (SUPPORT)

Characteristics of RHC Patients

- RHC patients were more likely to ...
 - Be male, have private insurance, enter the study with ARF, MOSF or CHF
- RHC patients were less likely to ...
 - Be over 80 years old, have cancer, have a DNR order in the first 24 hours of hospitalization
- RHC patients had significantly...
 - Fewer comorbid conditions, more abnormal results of vital signs, WBC count, albumin, creatinine, etc.
 - Lower model probability of 2-month survival

Some Characteristics used to predict Propensity for RHC use

- Age, Sex, Race
- Education, Income
- Insurance type
- Primary and secondary disease category
- 12 categories of admission diagnosis
- ADL & DASI 2 weeks before admission
- DNR status on day 1
- Cancer (none, local, metastatized)
- 2-month survival model
- Weight, temperature, BP, heart rate, resp. rate
- 13 categories of comorbid illness
- Body chemistry (pH, WBC, PaCO₂, etc.)

Panel (7 specialists in clinical care) specified important variables related to the decision to use or not use a RHC.

Outcomes & Possible Explanations

- After selection bias adjustment using propensity score methods, RHC pts had
 - Significantly lower (i.e. worse) survival
 - Higher hospital costs
 - Longer ICU stays
- RHC may lead directly to worse outcomes, OR ...
 - RHC may indicate style of care that leads to higher mortality and costs, or
 - Change in therapy in response to RHC may lead to worse outcomes, or
 - RHC might be beneficial, but bias adjustments were inadequate.

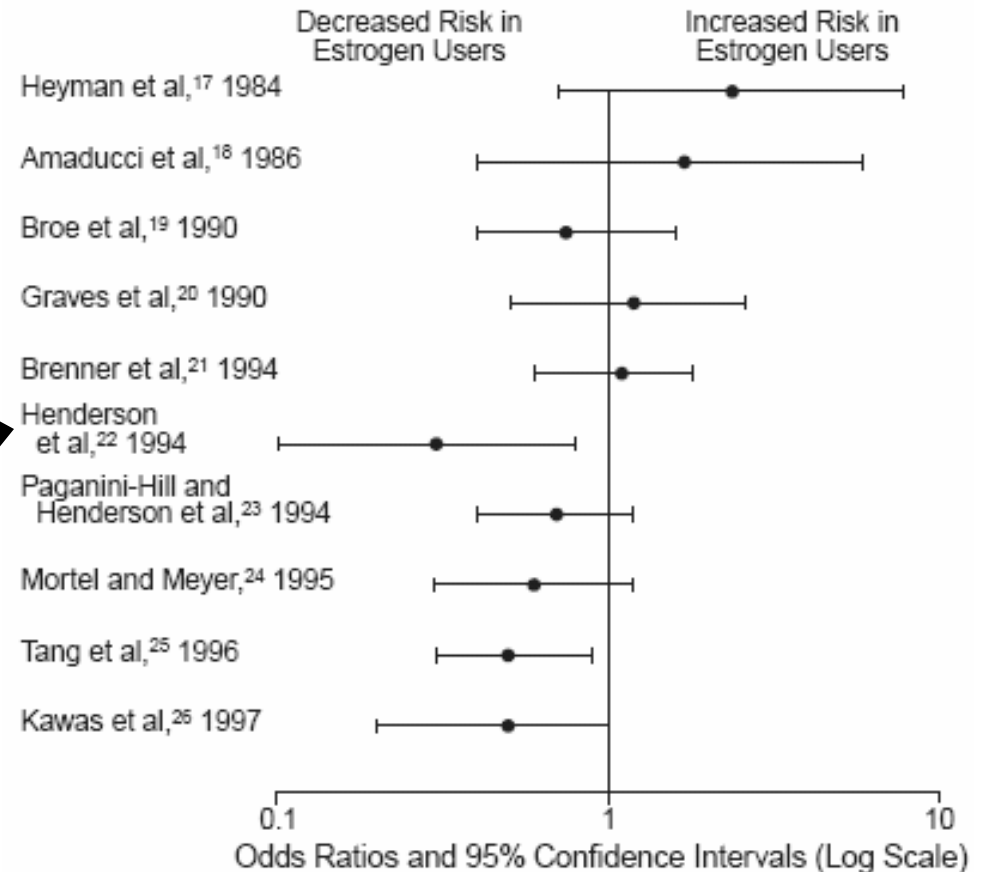
RCT is needed!

Hormone Replacement Therapy

**And Dementia
And Cardiovascular Risk**

Hormone Replacement Therapy and Dementia (1998 Meta-Analysis)

- Estrogen associated with 29% decreased risk of dementia
- Promising results for Alzheimer's disease →
- BUT these studies are for the most part small, short duration, non-randomized, and uncontrolled.

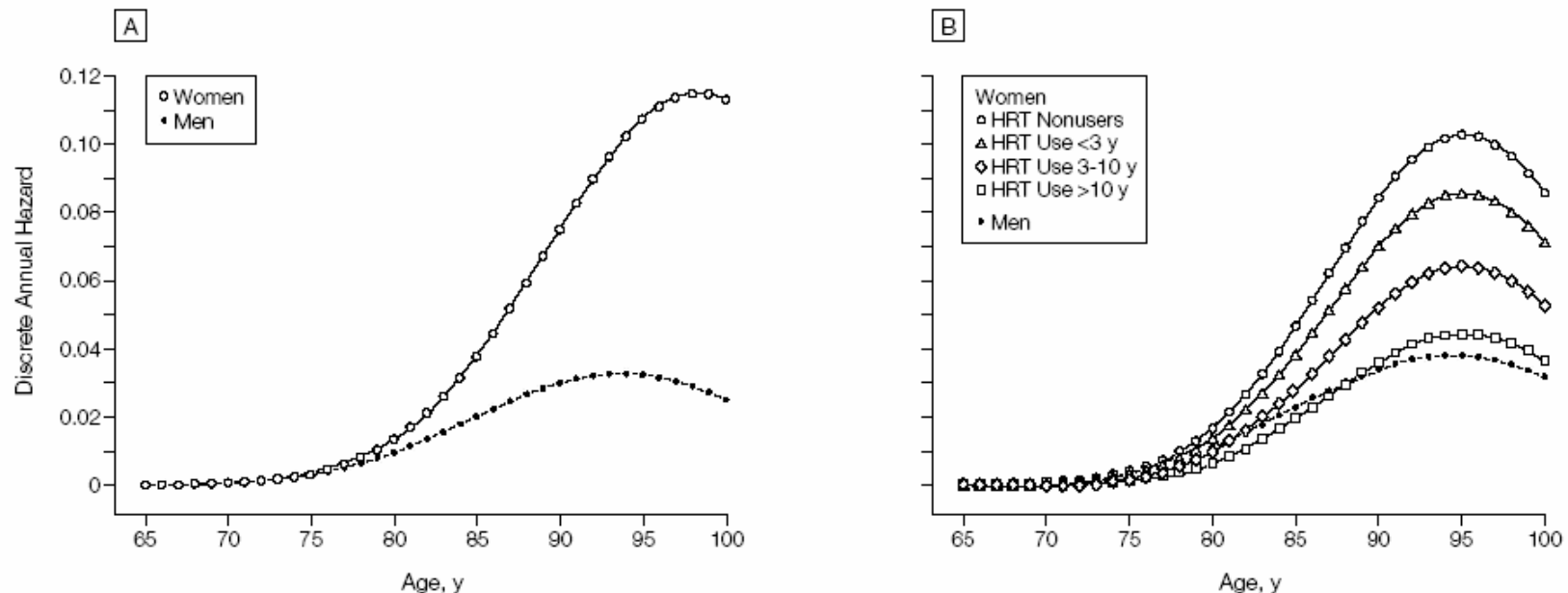


Cache County Memory Study

Prospective study of incident dementia

Adjustments: age, education, APOE ϵ 4 alleles

Figure 2. Estimated Discrete Annual Hazard of Alzheimer Disease for Men and Women by Age, and by Duration of Hormone Replacement Therapy Use for Women



Both figures indicate risks estimated for an individual with the mean value of 13 years of education and no ϵ 4 alleles at APOE. A, The curves depict the annual hazards predicted by fitting the base model including an age-by-sex interaction term. The annual hazard for Alzheimer disease (AD) appears similar for men and women before 80 years of age but diverges rapidly afterward with an excess risk found in women. B, The curves depict the annual hazards predicted by fitting model 7 of Table 3 to the women with available hormone replacement therapy (HRT) exposure information and, in filled circles, the corresponding annual hazards for men after omitting the terms for HRT. There were 35 instances of incident AD among 1357 men. Ordinate values for women differ slightly from those in panel A due to omission of women lacking HRT exposure information, several of whom experienced incident dementia.

Zandi et al. (2002)

[WHIMS] Women's Health Initiative Memory Study: RCT

- 4532 post-menopausal women age 65+ [RCT]
- E+P HRT increased risk (hazard ratio 2.05) for probable dementia.
 - Treating 434 women age 65+ with combination HRT would cause 1 new dementia case.
- NS impact on mild cognitive impairment
- NS baseline differences
 - Age, Education, Smoking, DM, Prior HRT or aspirin use, 3MSE score
- Significant differences comparing (E&P v Plac)
 - Stroke Hx (1.0 vs. 1.9)
 - Statin use (12.0 vs. 9.8)
 - Adherence (E&P < Plac)

HRT & Cardiovascular Disease

1980s & 1990s
Observational
Studies &
Meta-Analyses
HRT decreases
CHD risk by
35 to 50%

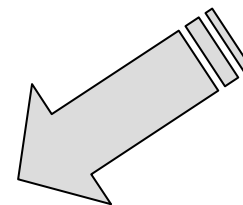
NHS [OS]
59337 women
Baseline age 30-55
Follow up to 16 yrs
RR for major CHD
of HRT vs Estrogen
0.60

1998 – HERS [RCT]
2763 women
Median age 67
Follow mean 4 yrs
RR for major CHD
of HRT is **0.99**

2002 – WHI [RCT]
16608 women
ages 50-79
Follow mean 5 yrs
HR for CHD
of HRT is **1.29**
Excess risk
7 more CHD events
8 more strokes
8 more PEs
8 more inv Br Ca
per 10,000
person-years

Col decision model (**1997**):
Predicted 2-3 yr ↑ Life Exp

Col SHORT-TERM decision model (**2004**):
With menopausal symptoms, ↓ survival but ↑ QALE.
Without symptoms, ↓ LE and QALE



Selection Bias?

Comparing NHS (OS) to WHI (RCT)

Selection Bias #1: Healthy User Effect

- Women with healthy behaviors may select to use postmenopausal hormones. (prevention bias)

In the NHS,

- HRT users tended to have better CV risk profiles
- HRT users were generally better educated
- Perhaps women taking HRT / ERT were “compliant” and such people have lower CHD risk.
- HRT users have more contact with physicians, and are perhaps more health conscious, generally.

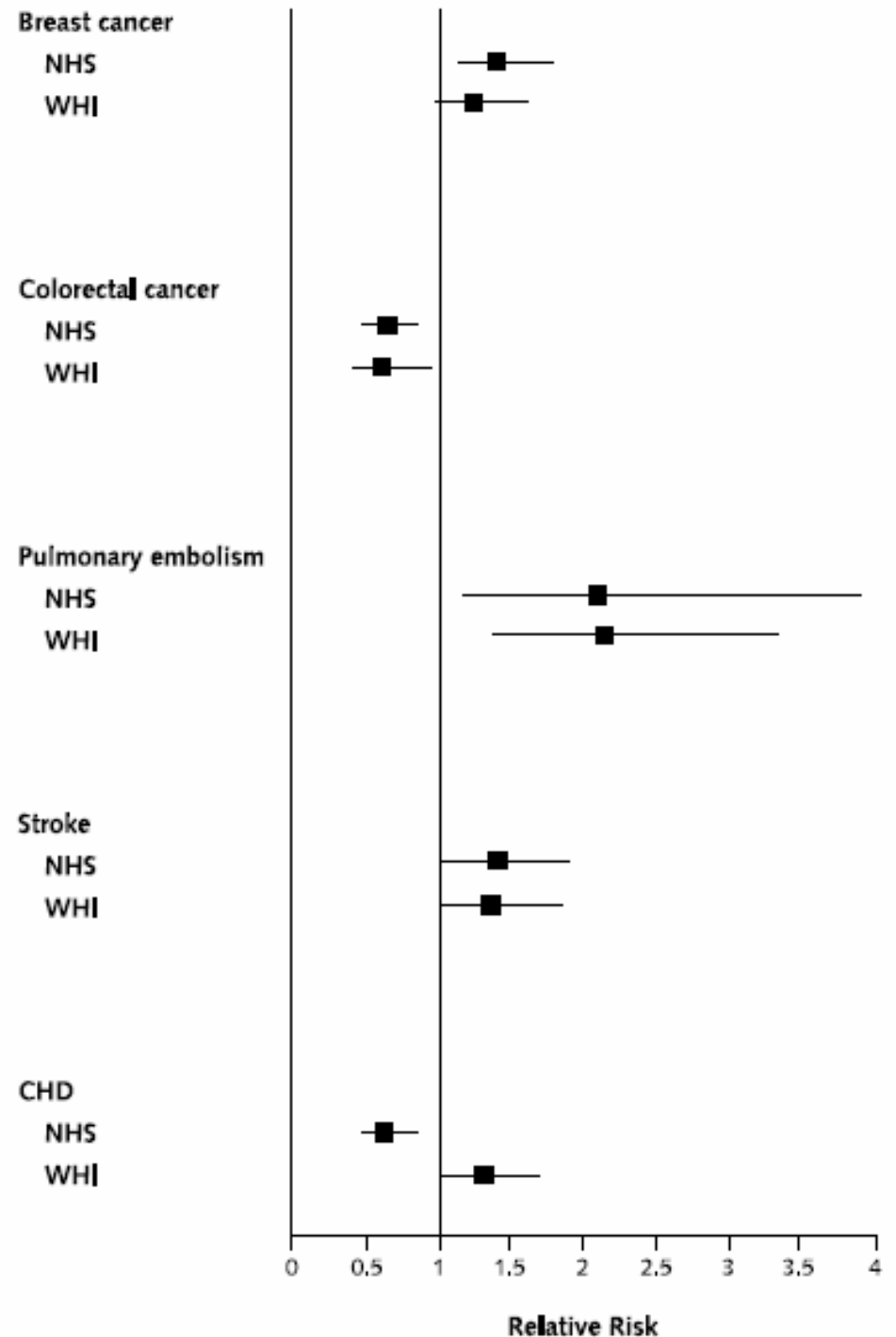
Is Selection Bias The Whole Answer Here?

Selection biases should affect all clinical outcomes with modifiable risk factors.

Biases selecting women at lower CHD risk should also select women at lower risk for stroke, PE...

Figure 1. Pairwise comparison of the relative risk for various neoplastic and vascular clinical outcomes associated with menopausal hormone therapy use reported in the Nurses' Health Study (NHS) and the Women's Health Initiative (WHI) trial.

Col and Pauker (2003)



Cox-2 Inhibitors and Cardiovascular Risk

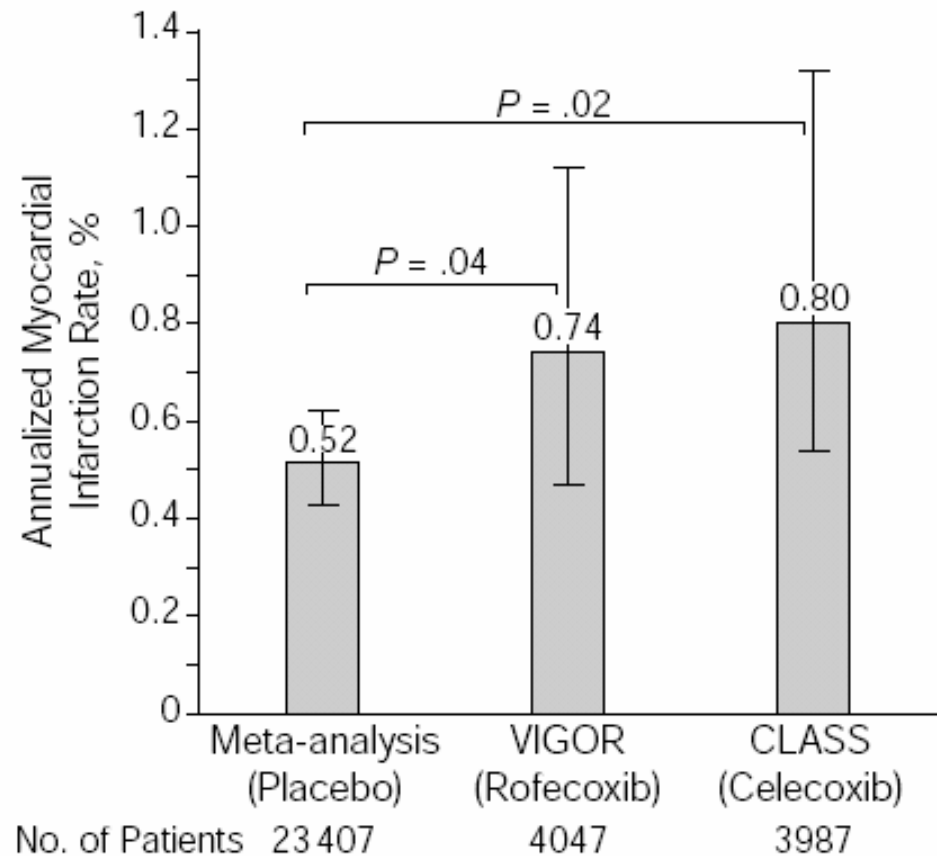
**2002: “no evidence that HRT
increases or decreases [MI risk].”**

**2005: “continued reason for concern
about adverse [CV effects] of both
Vioxx and Celebrex
(though less so for Celebrex)”**

Summarizing Cox-2 Results As of Aug 2001

- Meta-analyses of placebo
- CLASS trial (Celebrex)
- VIGOR trial (Vioxx)

Figure 3. Comparison of MI Rates Among Subjects Receiving Placebo vs Rofecoxib or Celecoxib



MI indicates myocardial infarction. Error bars indicate 95% confidence intervals.

CRESCENT RCT: Effect of Celebrex, Vioxx & Naproxen on 24 hr SBP in patients with Type 2 DM, HTN and OA

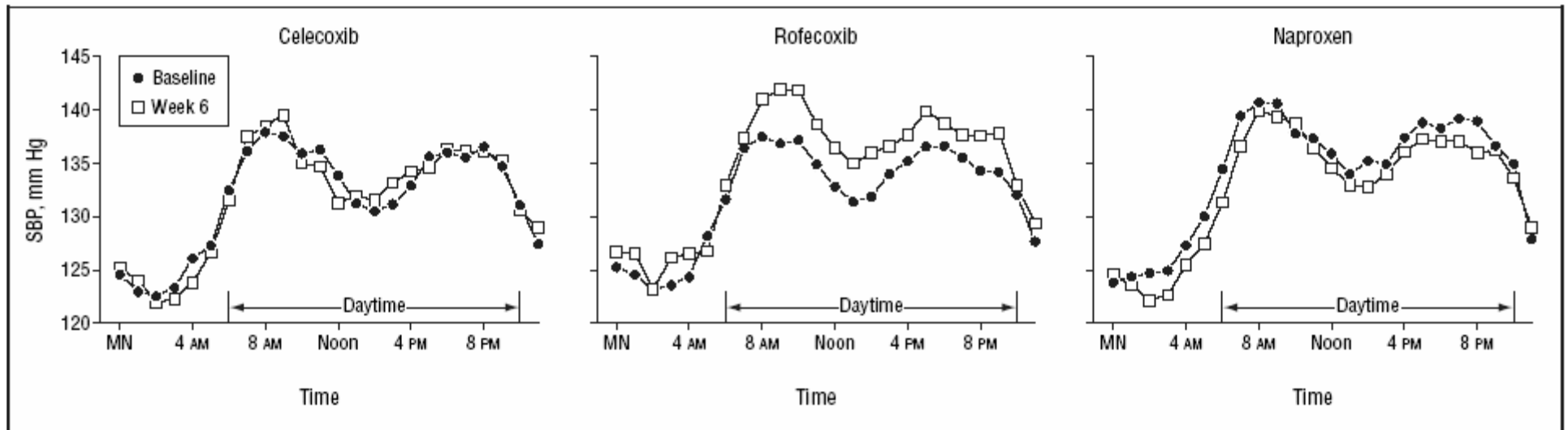


Figure 3. Hourly means of ambulatory systolic blood pressures (SBPs) over 24 hours at baseline and week 6 for celecoxib, rofecoxib, and naproxen. MN indicates midnight. Ambulatory BP monitoring was initiated at approximately 9 AM \pm 2 hours. The morning dose of study medication was administered within 5 minutes of initiating the ambulatory BP monitoring session. A consistent increase from baseline in ambulatory systolic pressure was observed only in the rofecoxib treatment group.

Maryland Medicaid Enrollees (2005)

- 1005 on Cox-2 inhibitors, 5245 on prescription non-naproxen NSAID
 - ≥ 60 d but not for prior 6 m
- Adjusted for demographics, indications for Cox-2 inhibitors and CV risk factors.
- Result: NS effect of Cox-2 on post-treatment CV thrombotic events

OR = 1.09 (0.90, 1.33)

Characteristic	COX-2 Inhibitor (n = 1005)	Other NSAID, Excluding Naproxen (n = 5245)
Demographic		
Age, y		
18-29	6	17
30-39	12	25
40-49	29	30
50-59	28	20
60-69	19	7
≥ 70	6	1
Race		
White	47	40
Black	43	54
Other	10	6
Female sex	74	69
Indications†		
Gastrointestinal problems	18	7
Rheumatoid arthritis	6	1
Osteoarthritis	22	6
Acute pain	46	22
Back pain	23	12
Cardiovascular risk factors‡		
Hypertension	56	40
Diabetes	28	18
Tobacco, alcohol, or drug abuse	19	25
Hyperlipidemia	34	22
Obesity	17	13
Renal	5	4
Cardiovascular event	15	6
Unadjusted rate of primary outcome: qualified cardiovascular event (APTC end point) after treatment	19	11

Canadian Retrospective Cohort NSAID-naïve subjects age 66 +

~15K Vioxx, 19K Celebrex,
5K non-selective
NSAIDs, 100K “random”
controls

RR of CHF Admission:

Vioxx vs. Control **1.8**

non-sel vs. Control **1.4**

Celebrex vs. Control **1.0**

35+ variables controlled

- Admissions
- Procedures
- Drug use in prior year
- Age, Sex, Long-term care,
Low income status

- Unaccounted for
 - Weight, Smoking,
 - OTC meds, Alcohol,
 - Dietary salt, Adherence

Switching to Cox-2 Inhibitors among U.S. RA and OA patients

- 6637 patients from 433 rheumatologists completed 2 detailed questionnaires about
 - July – Dec 1998
 - Jan – June 1999
- 1517 pts received Cox-2 inhibitors in period 2
- How did they differ from 5120 who didn't?
- Disease severity (most important: Visual Analog pain scale, also global severity, fatigue)
- GI protective drugs
- previous GI events
- age
- service utilization
- quality of life measures
- smoking status

**What Can (and Should)
We Do With
Observational Studies?**

Dealing With Selection Bias

- Analytical adjustments for Observed Bias
 - Matching, Stratification, Weighting, and Regression approaches are most common
- Goal: compare treated subjects to controls with same (or similar) baseline characteristics.
- Assuming no hidden bias (heroically), we use standard statistical comparisons of the groups after adjustment.

What should always be done in an OS ... and often isn't?

1. Collect data so as to be able to model selection
2. Demonstrate need for adjustment - selection bias
3. Carefully record intervention time - adjust only for things present before or at time of intervention.
4. Ensure baseline characteristic overlap [comparability]
5. Check baseline characteristic balance after adjustment
6. Specify relevant post-adjustment population with care
7. Estimate treatment effect in light of adjustment
8. Estimate sensitivity of results to potential hidden bias

How Can We Avoid Being Misled?

1. What differentiates an observational study from a randomized controlled trial?
 - One key element: potential for selection bias.
2. What is selection bias, and why should I care about it?
 - Baseline characteristics of comparison groups are different in ways that affect the outcome.
3. What can be done to deal with selection bias in observational studies?
 - Propensity score methods for overt bias.
 - Sensitivity analyses to deal with hidden bias.