500 Class 01 (Zoom)

https://thomaselove.github.io/500-2025/

2025-01-16

Agenda for Zoom Meeting

Thursday 2025-01-16 from 10 to 11 AM. Zoom details in your email and on Canvas.

- Welcome to 500
- Logistics of the Course, Assignments
- A Motivating Example: Aspirin and Mortality in Heart Patients
  - How can we avoid being misled?
  - Causal Effects as comparing potential outcomes
- Your Questions

Course Overview

- Randomized Experiments vs. Observational Studies
  - Randomization as the "fundamental basis for inference"
  - Observational Studies and Causal Effects
- Propensity Scores: Crucial Tools for Causal Models
  - Selection Bias: key issue for observational studies
  - Dealing with Bias (both overt and hidden)
  - Subclassification (stratification) and direct regression adjustment
  - Matching and weighting using the Propensity Score
  - Sensitivity Analysis
- Instrumental Variables and Other Techniques
- Using R, RStudio and Quarto to accomplish all of this

Paul Rosenbaum's 2023 book *Causal Inference*

My Expectations

- You are interested in learning about the effects of an intervention, treatment or policy on subjects when the treatments cannot be assigned at random.
- You have little interest in technical details of methods, but serious interest in designing, conducting and analyzing observational studies skillfully.
- You have access to software (specifically R) which you can use to obtain basic hypothesis testing, regression and logistic regression results.

This Year is Unusual

- Classes 1-8 (before Spring Break) involve:
  - a recorded 60-90 minute lecture (like this one)
  - a Zoom meeting to discuss the lecture and other issues from 10-11 AM on Thursday
- After Spring Break, starting with Class 9 (2025-03-20) we will meet in person from 8:30 AM to 11:15 AM in Wolstein Research Building room 1217 on the CWRU campus.
- For all sessions, Dr. Love is available after class for informal "office hours".
- Recordings of class sessions will be available through Zoom on Canvas when things work properly.

The Web Site

https://thomaselove.github.io/500-2025 is at the bottom of every slide

- Syllabus
- Calendar
  - links to class sessions, final word on all deadlines
- R and Data
  - Installing/Updating R, RStudio, R Packages
  - Data and Code
- Sources / References
  - Some things are **password-protected**.
- Links to Canvas, and to ways to Contact Us
- Assignments ...

Assignments / Deliverables

1. Course Project
   - Semester-long project, with your first proposal draft due 2025-02-13.
   - Second proposal draft 2025-02-27.
   - Final presentation of your work in class April 10, 17 or 24.
2. Observational Studies in Action
   - Present methods/results from a published article using propensity scores.
   - You'll present once as primary reviewer, once as second reviewer.
   - First step is to identify a study and claim it by 2025-02-19.
3. Labs
   - Lab 01 is due Thursday 2025-01-30 at 9 AM to Canvas.
   - There is a "Lab 0" worked example to look at first.
   - Deadlines and instructions for all labs are on the website.

There are no quizzes or examinations in 500.

A Key Goal for the Project and Course

- Help you learn how to tackle a problem, rather than just be able to perform particular statistical techniques.
  - Goal: think and solve problems when trying to infer causal effects from observational data
- But the need to think in statistical terms is omnipresent
  - Identifying researchable problems
  - Dealing with variation
  - Interplay of Design and Analysis
  - Preparing, writing and revising results, in a replicable way.

Section 1

A Motivating Example (Aspirin and Mortality)

Aspirin and Mortality in Heart Patients

Suppose you want to understand the effect of aspirin (acetylsalicylic acid: ASA) on mortality among patients undergoing stress echocardiography.

- What is the population?
- What is the outcome?
- What are the treatments?

Potential Outcomes and The Aspirin Study

Potential Outcomes and The Aspirin Study

Potential Outcomes and The Aspirin Study

This space intentionally left blank.

ASA and Mortality in Heart Patients

Suppose you want to understand aspirin's effect on all-cause five-year mortality among patients undergoing stress echocardiography.

- Comparing ASA to "No ASA"
- What are the potential outcomes here?

Potential Outcomes and The Aspirin Study

This space intentionally left blank.

Potential Outcomes and The Aspirin Study

Potential Outcomes and The Aspirin Study

Aspirin – "No Aspirin" Effect =

–

ASA and Mortality in Heart Subjects

- Suppose you want to study the effect of aspirin (acetylsalicylic acid: ASA) on all-cause mortality.
- You identify an interesting group of Subjects as those undergoing stress echocardiography.
    - Your goal is to compare ASA Subjects to "no ASA" Subjects

What would be the **ideal** study?

Step 1. Identify a large group of Subjects from the population at Time 0.

- We want to understand the causal effect of aspirin on all-cause five-year mortality among patients undergoing stress echocardiography.
- Having identified a set of patients, what is the ideal study?

Step 2?

ASA and Mortality: Ideal Study

Identify a large group of patients from population at Time 0

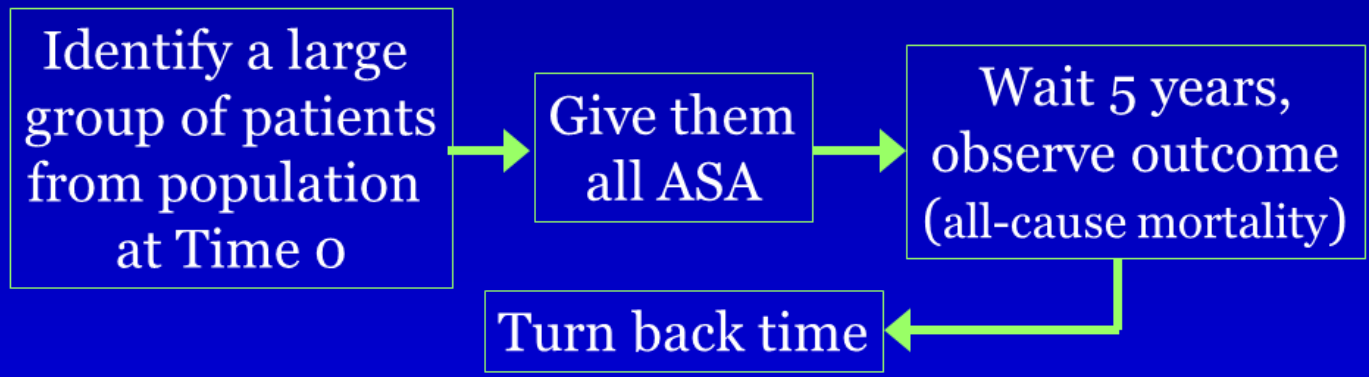
ASA and Mortality: Ideal Study

Identify a large group of patients from population at Time 0

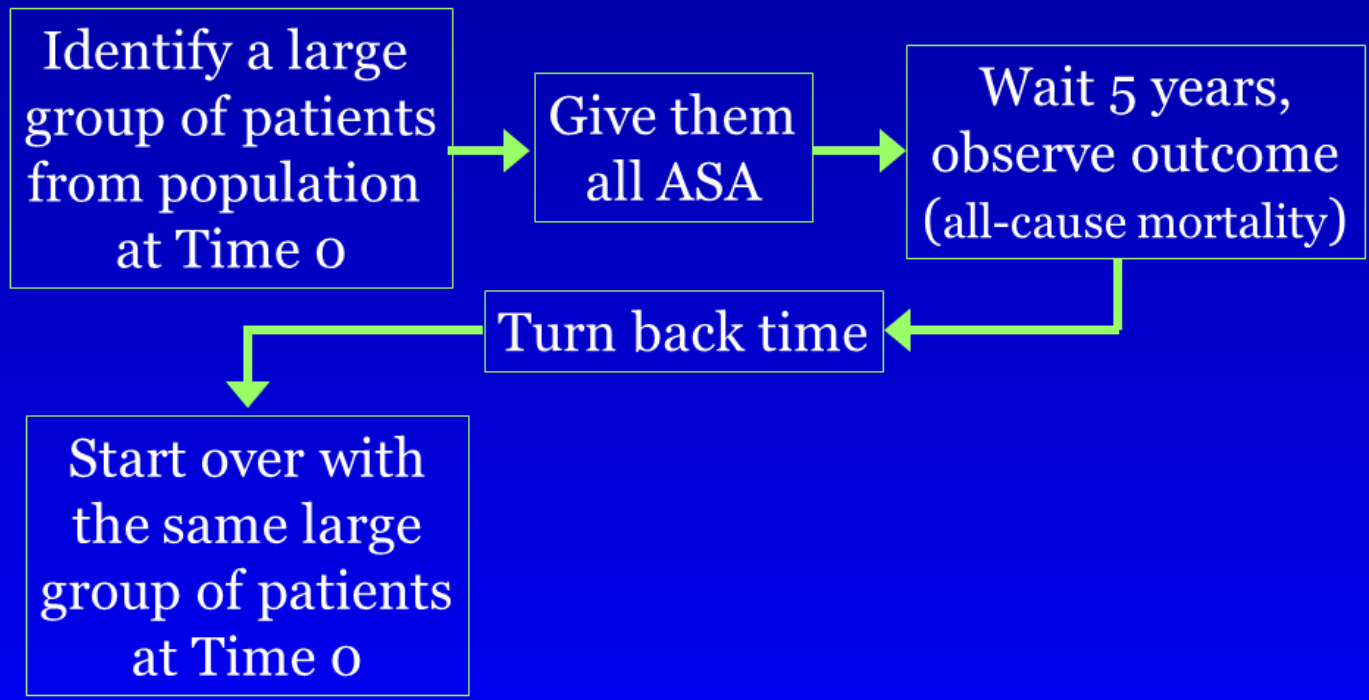
Give them all ASA

ASA and Mortality: Ideal Study

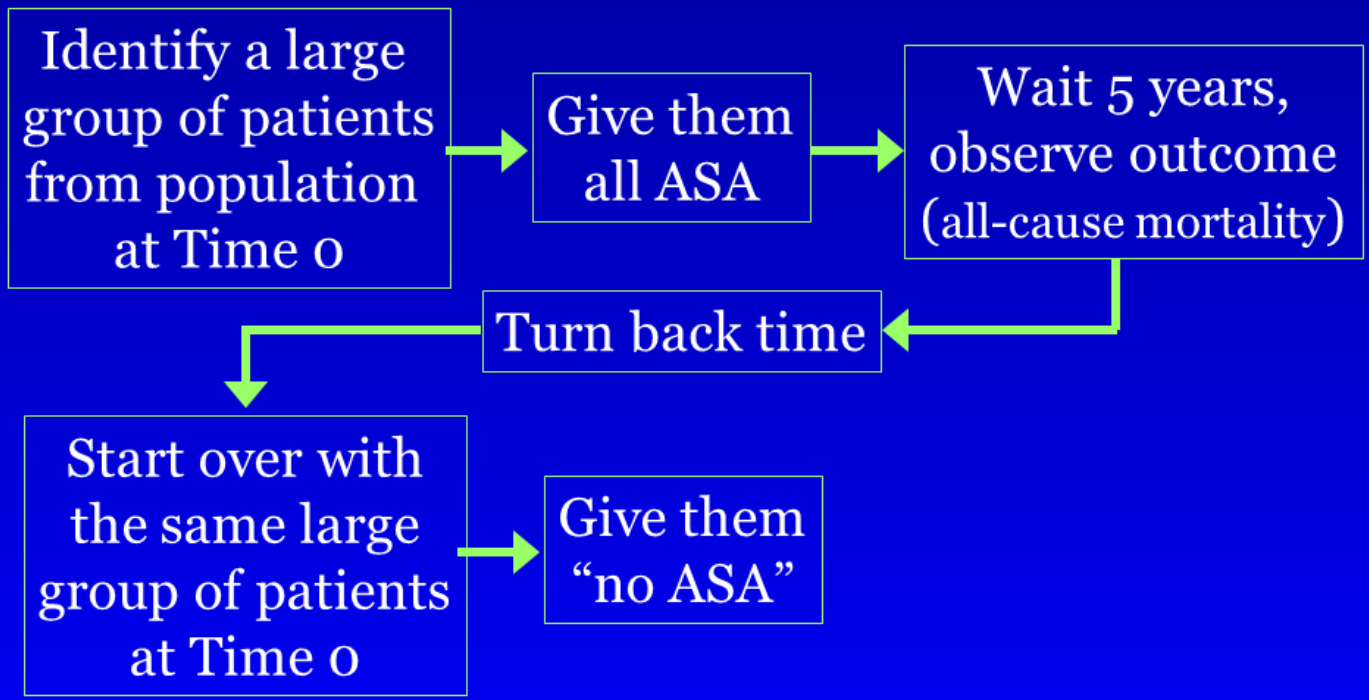
ASA and Mortality: Ideal Study



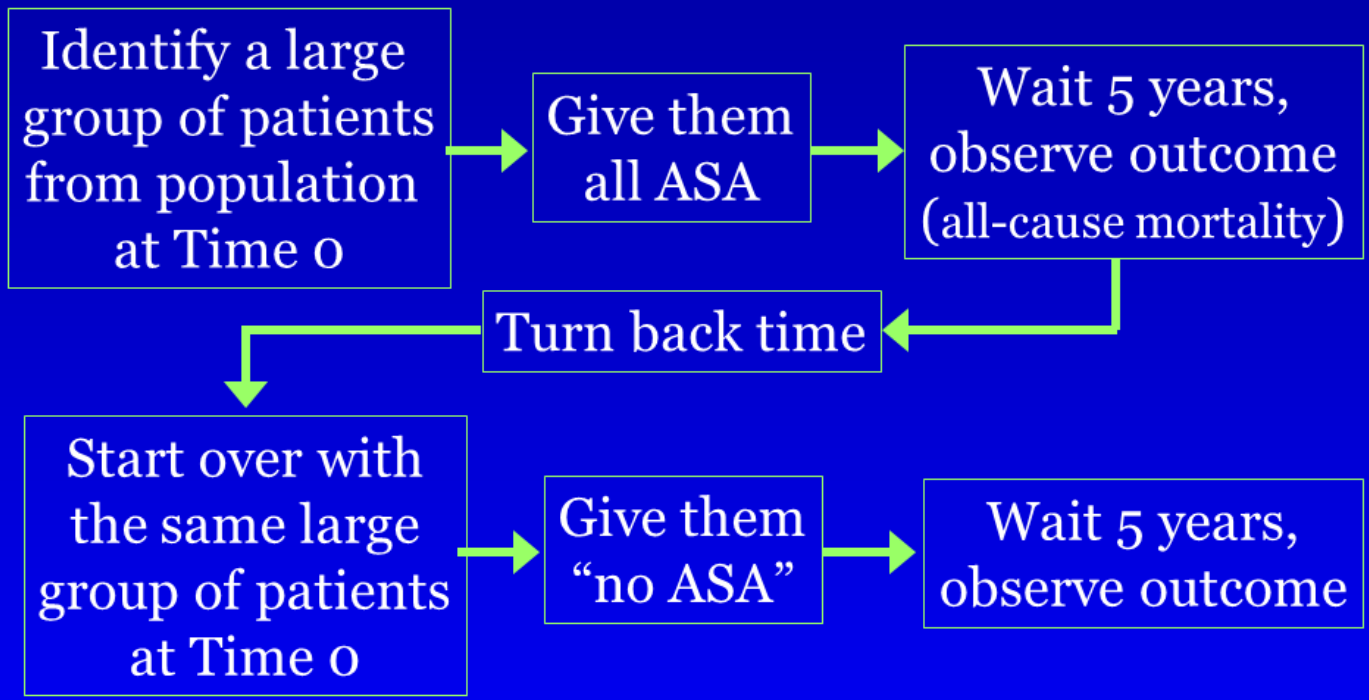
ASA and Mortality: Ideal Study



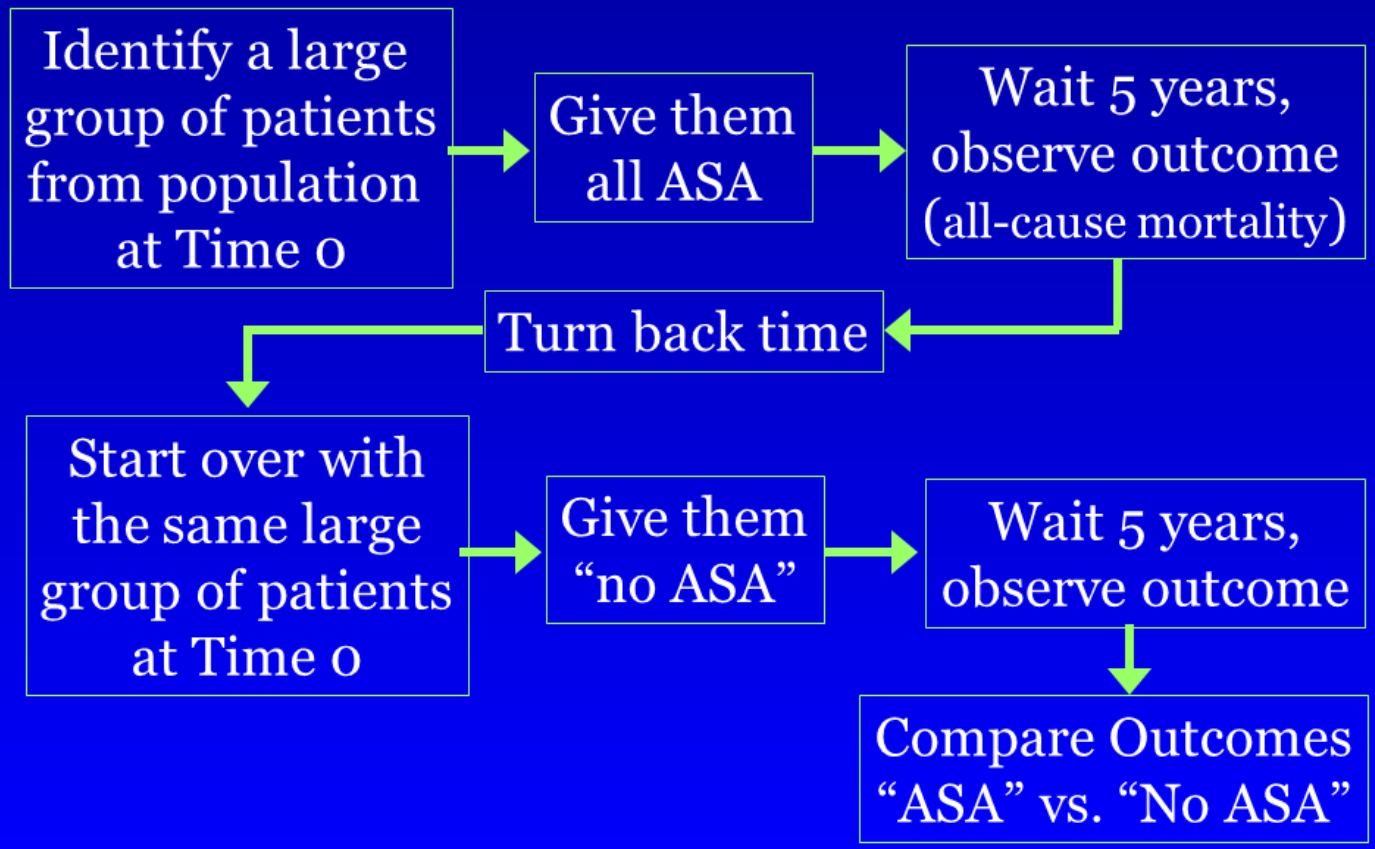
ASA and Mortality: Ideal Study



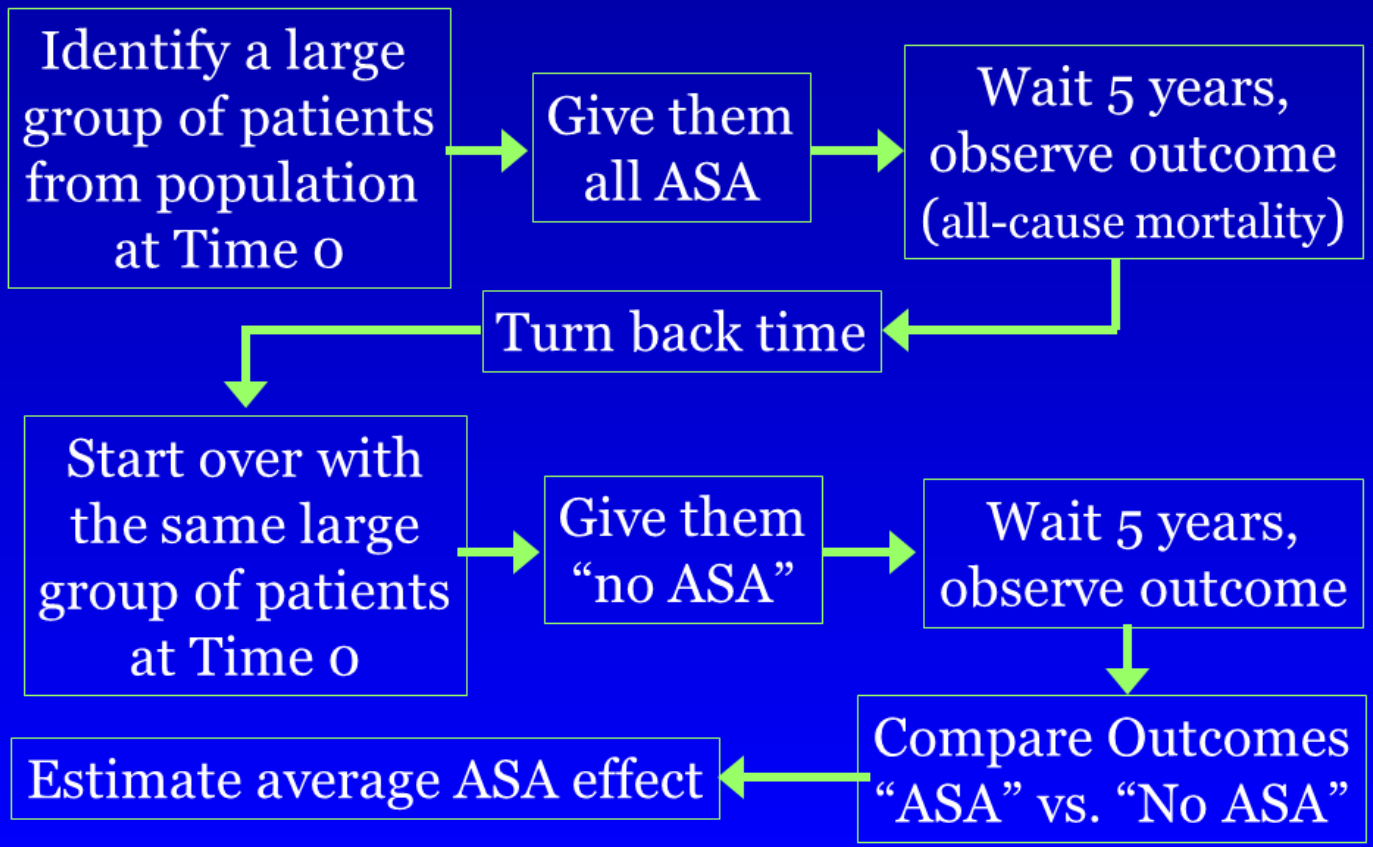
ASA and Mortality: Ideal Study



ASA and Mortality: Ideal Study



ASA and Mortality: Ideal Study



ASA & Mortality: Next Best Study

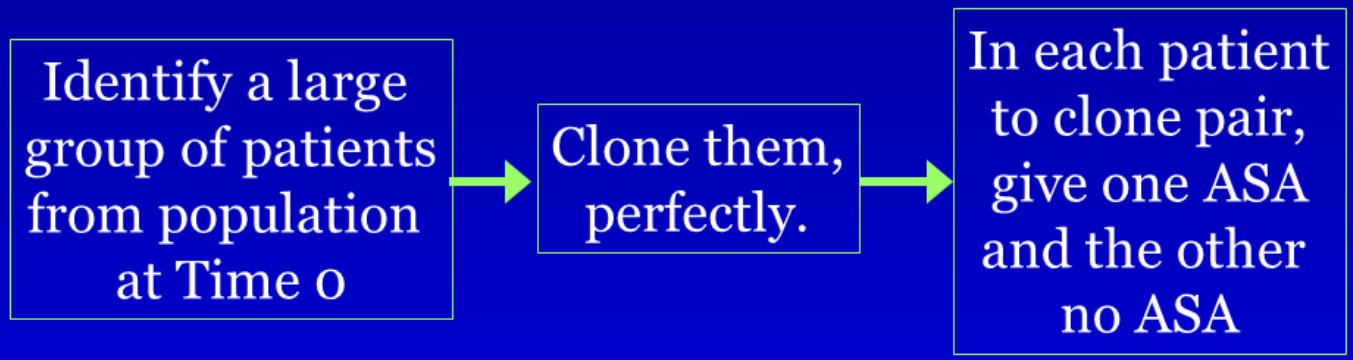
Identify a large group of patients from population at Time 0

ASA & Mortality: Next Best Study

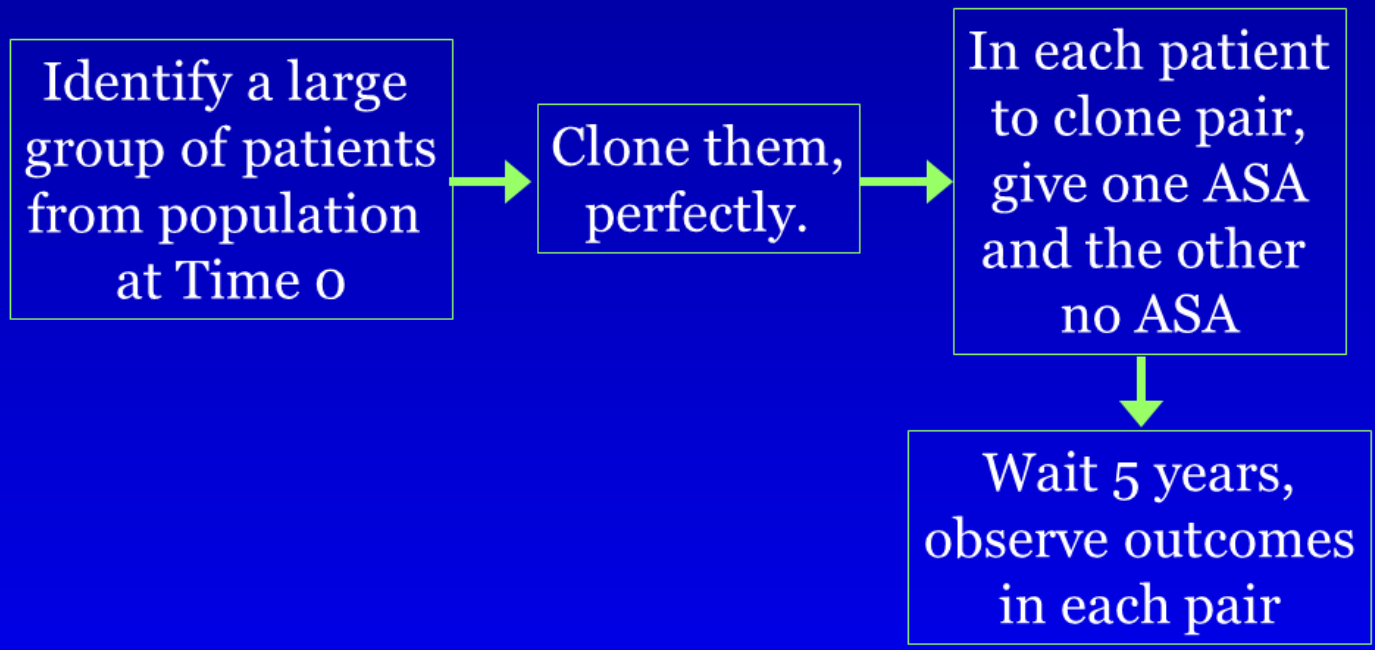
Identify a large group of patients from population at Time 0

Clone them, perfectly.

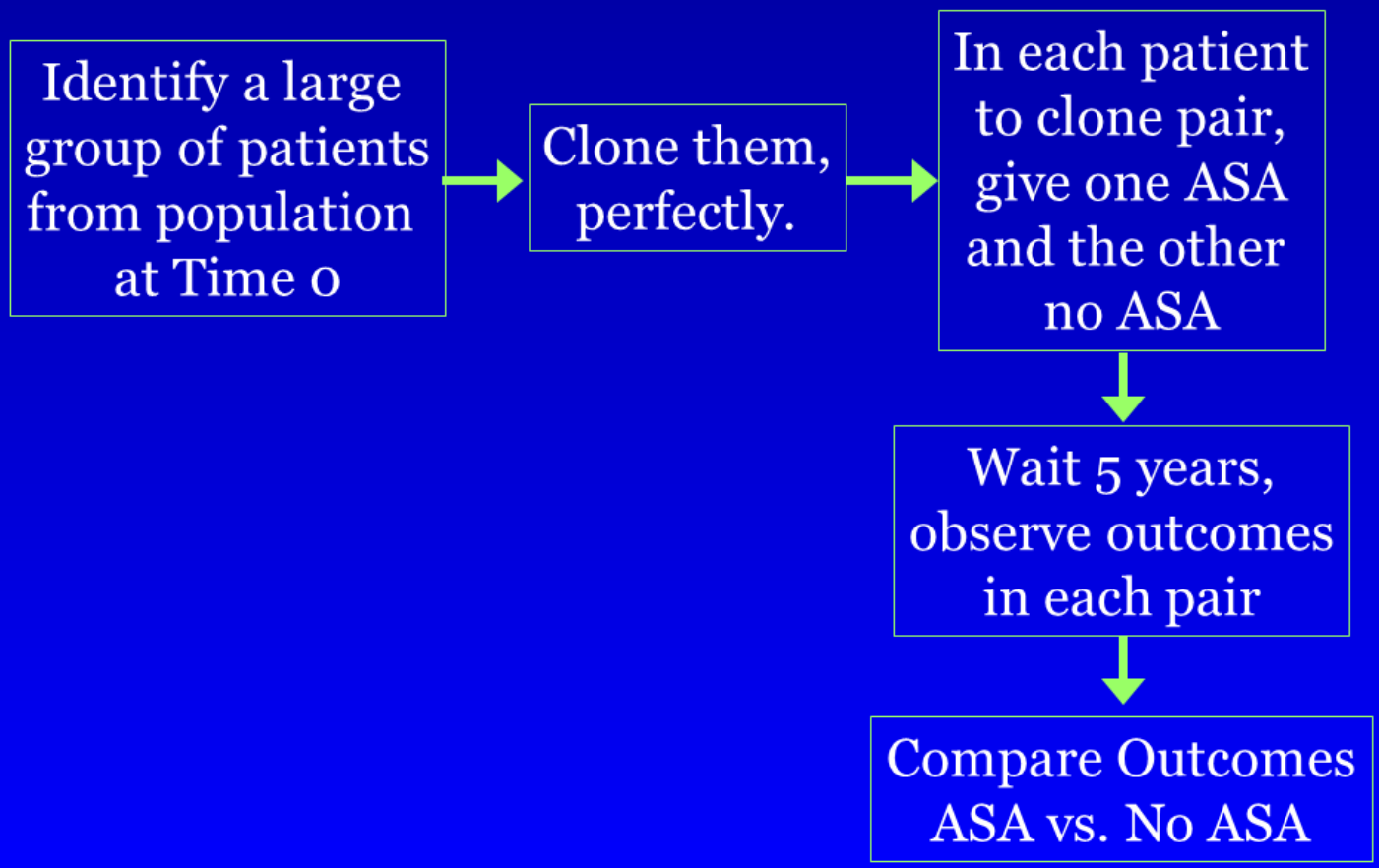
ASA & Mortality: Next Best Study



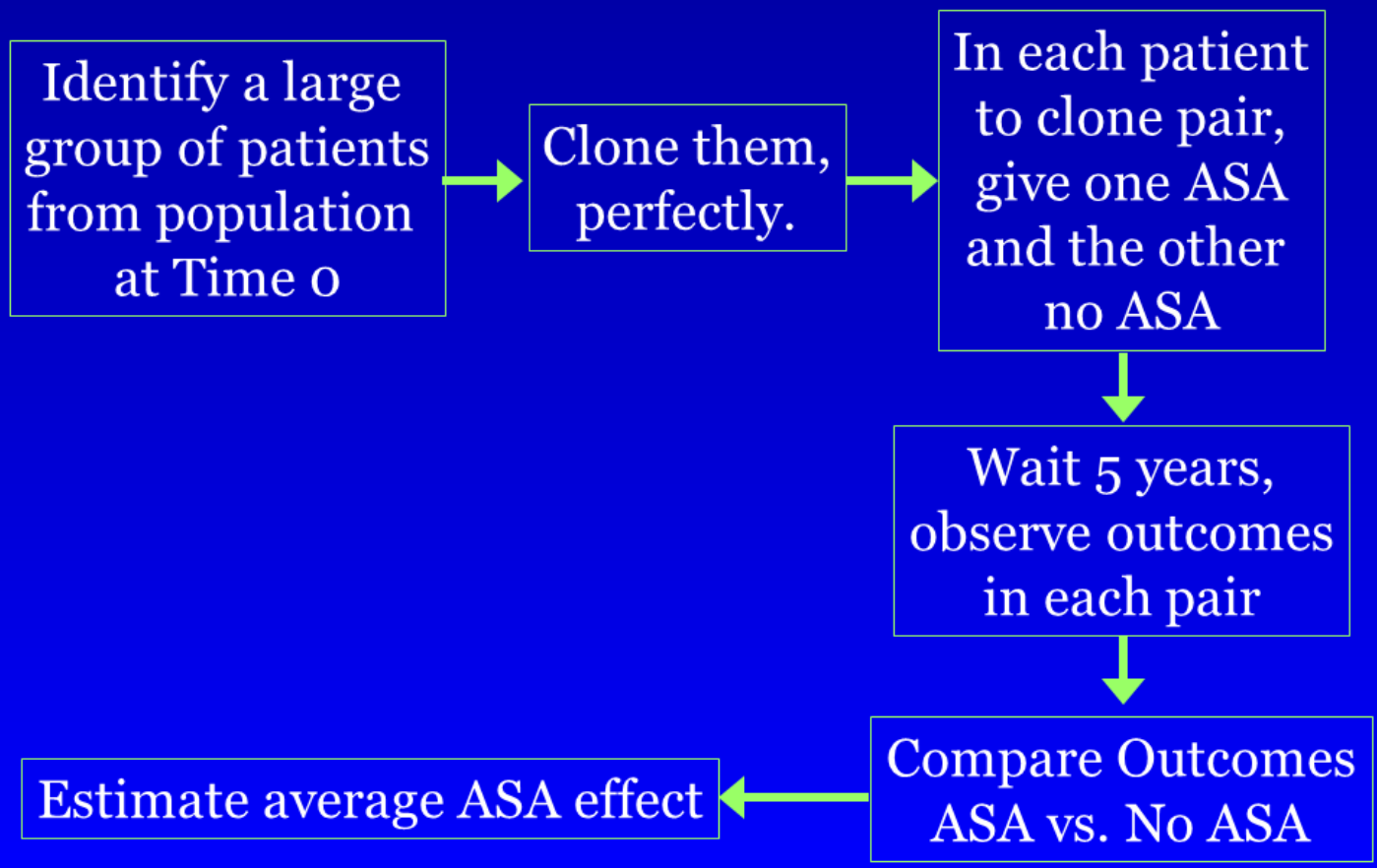
ASA & Mortality: Next Best Study



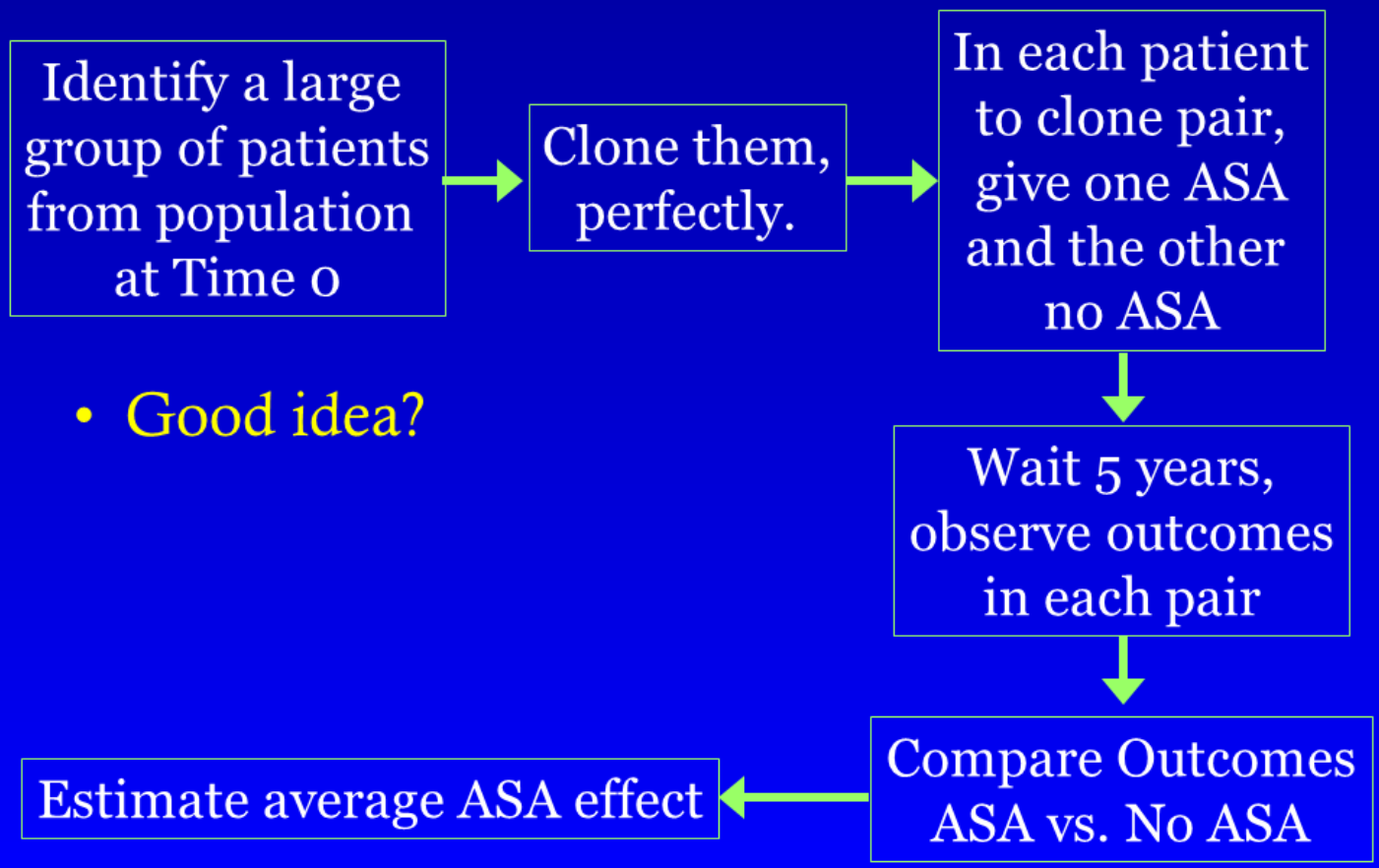
ASA & Mortality: Next Best Study



ASA & Mortality: Next Best Study

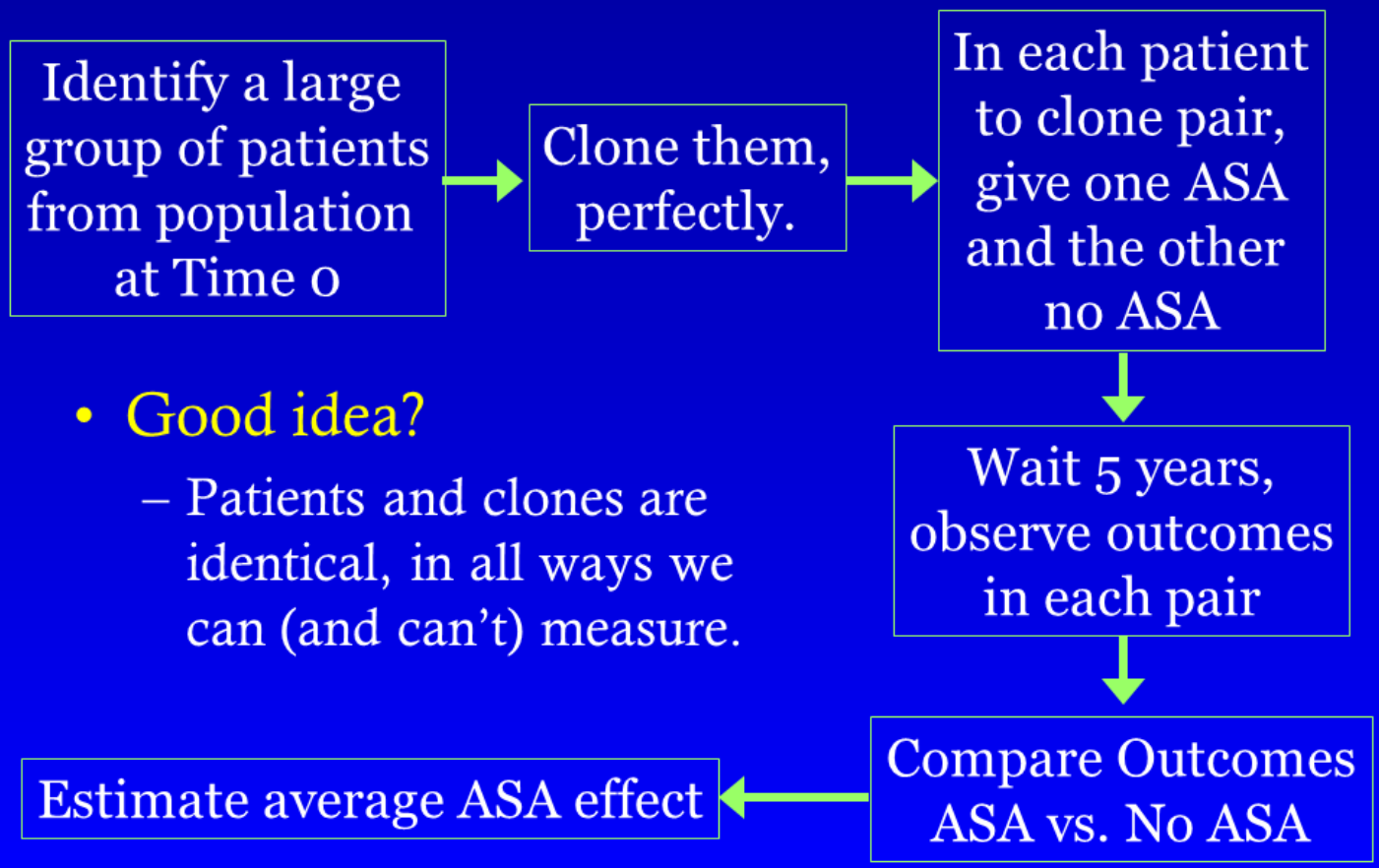


ASA & Mortality: Next Best Study



- Good idea?

ASA & Mortality: Next Best Study



- Good idea?
  - Patients and clones are identical, in all ways we can (and can't) measure.

ASA and Mortality in Heart Patients

- Designing the Study

We want to understand aspirin's effect on all-cause five-year mortality among patients undergoing stress echocardiography.

- OK.
- What's the best **practical** study?

ASA & Mortality: RCT

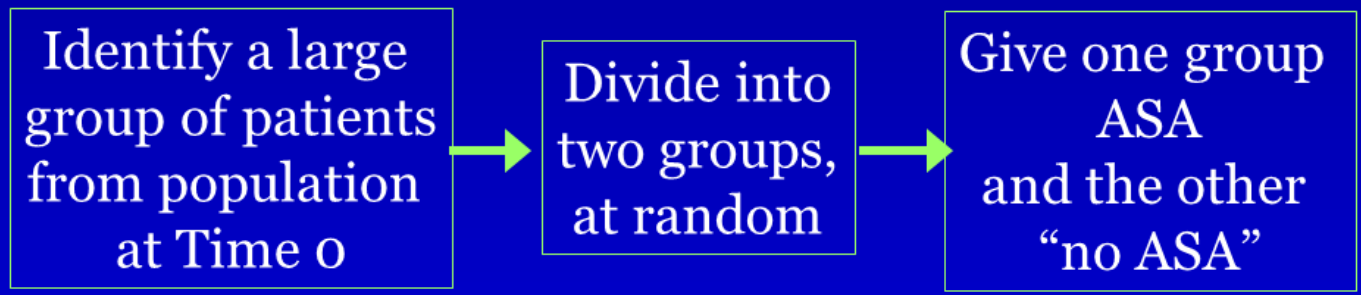
Identify a large group of patients from population at Time 0

ASA & Mortality: RCT

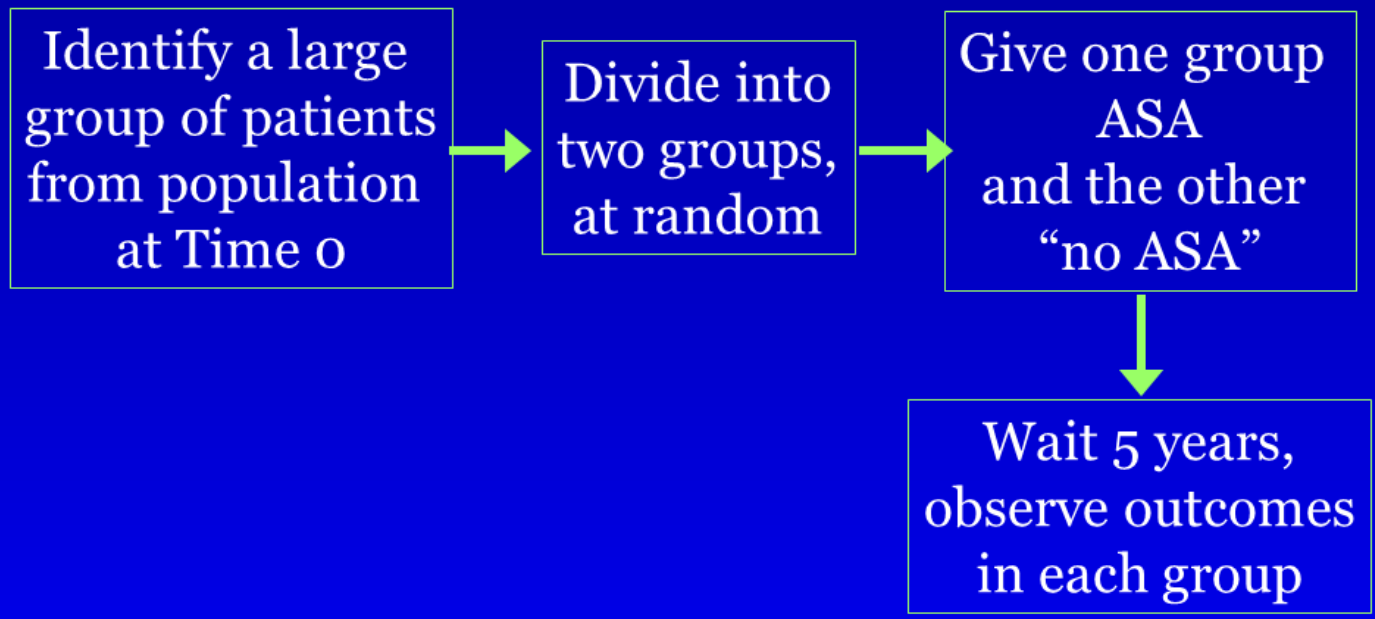
Identify a large group of patients from population at Time 0

Divide into two groups, at random

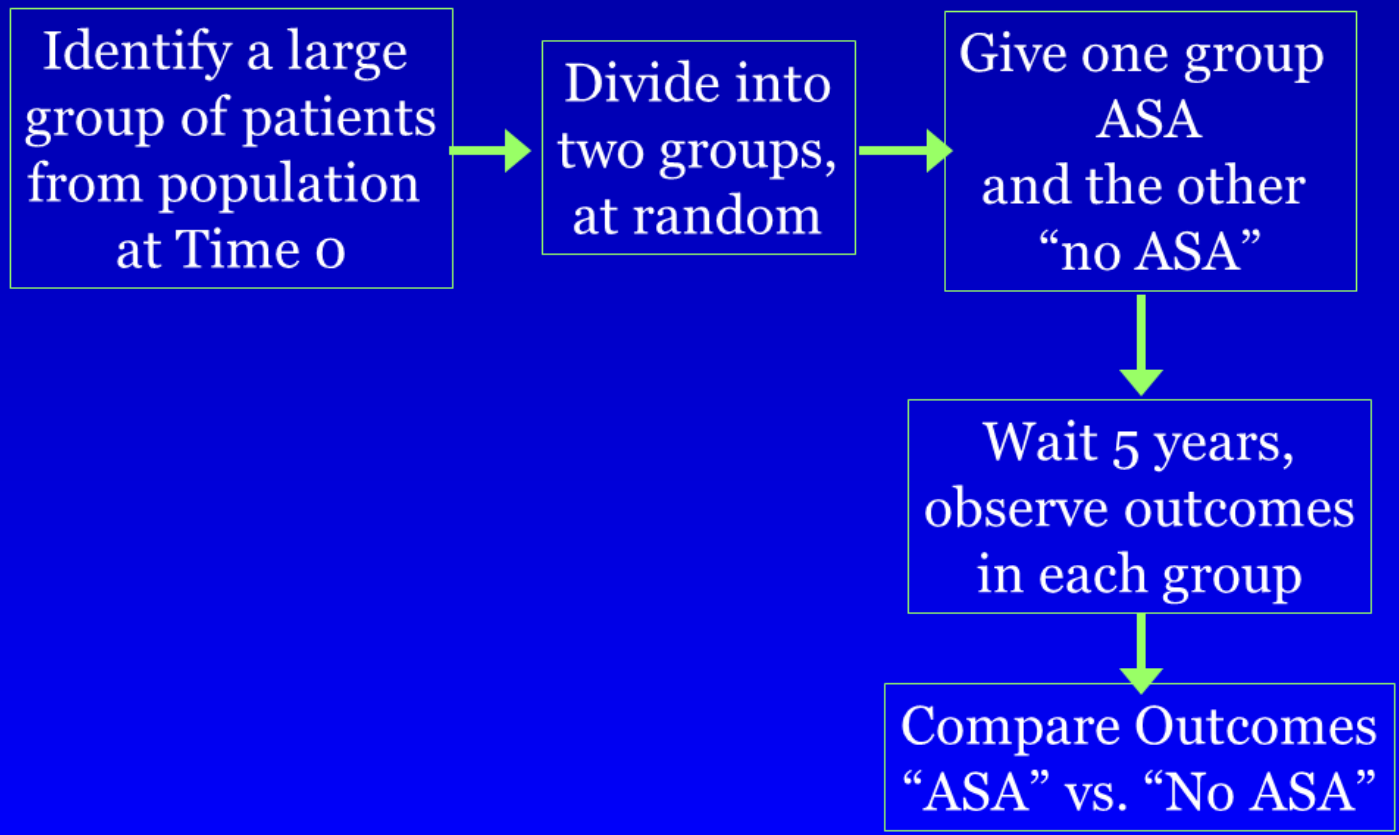
ASA & Mortality: RCT



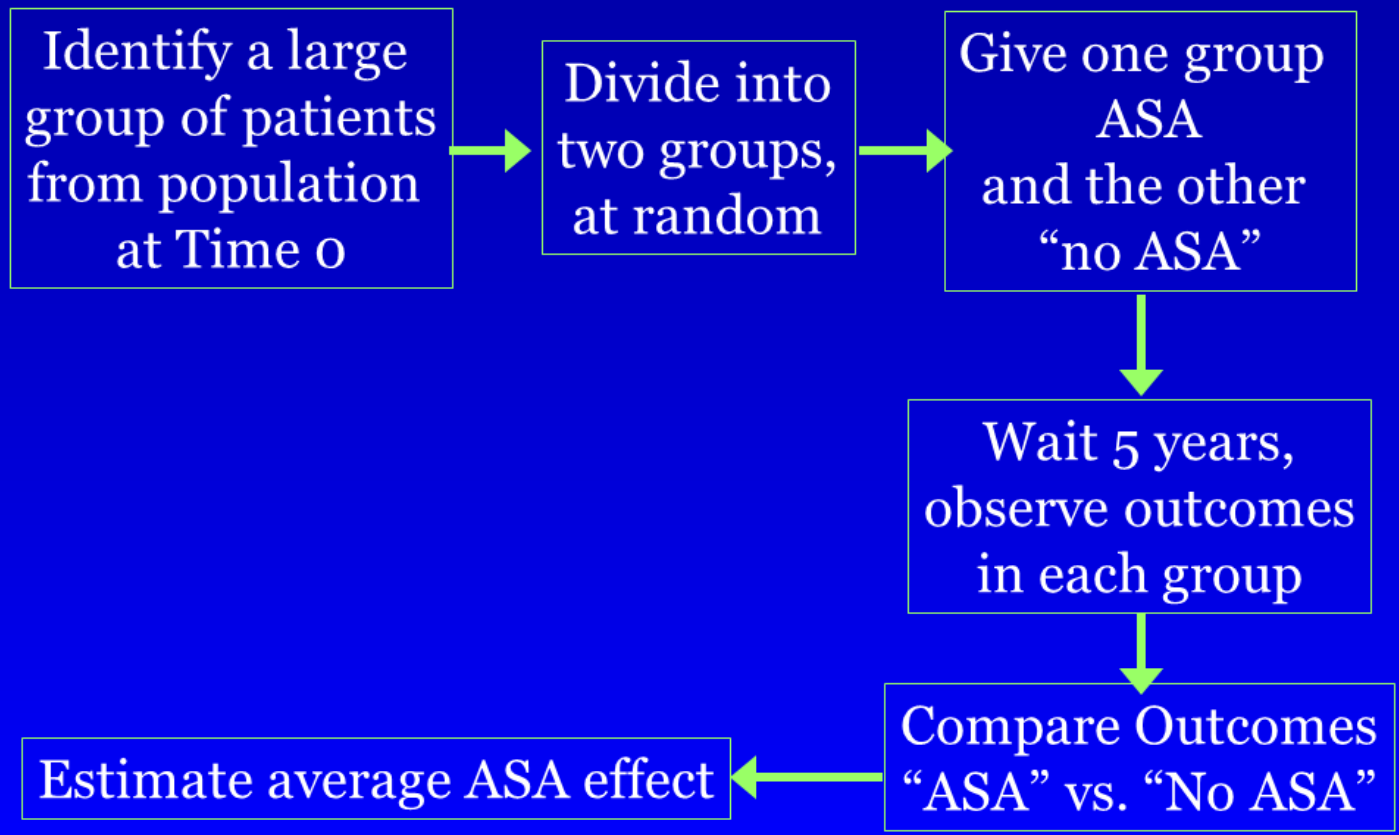
ASA & Mortality: RCT



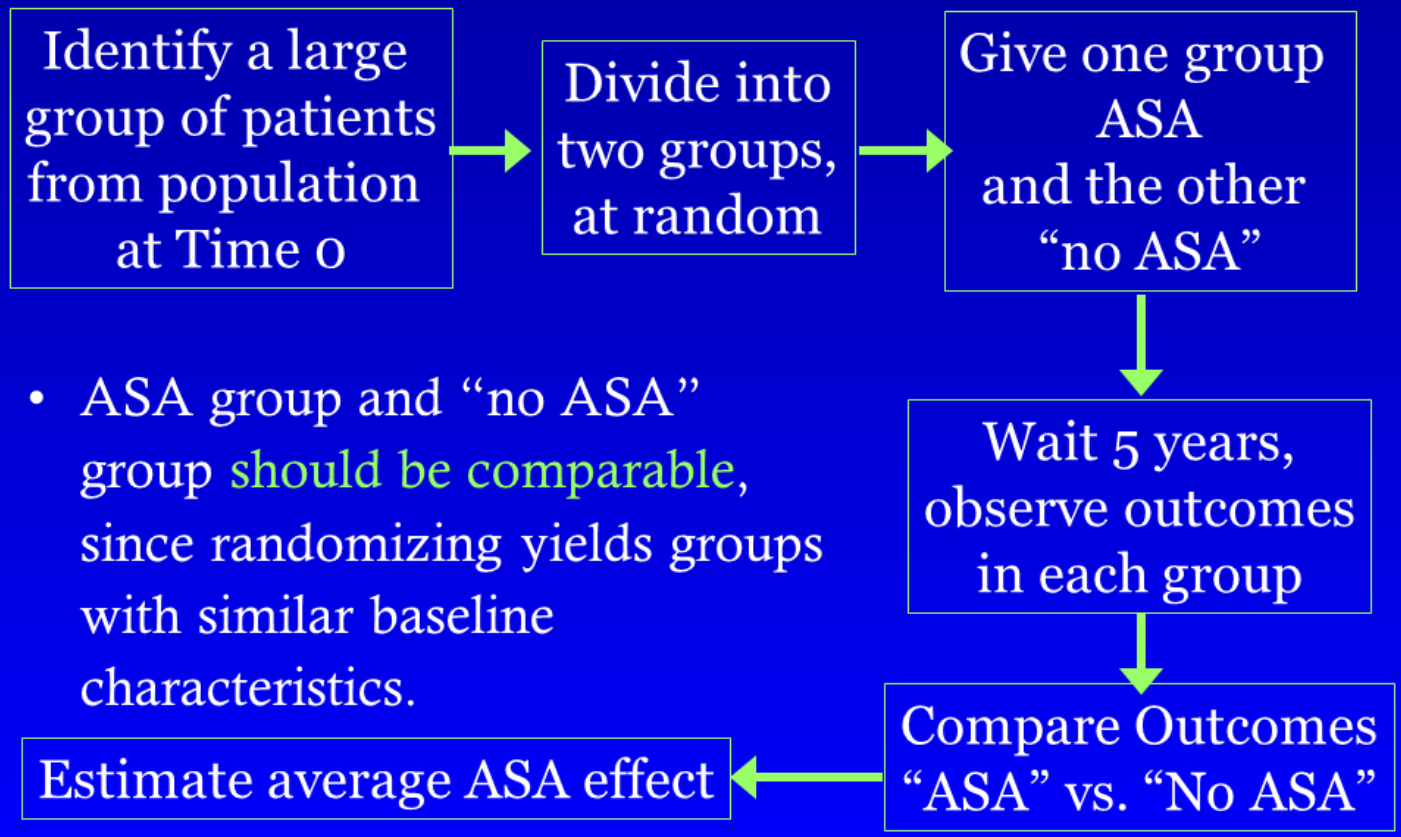
ASA & Mortality: RCT



ASA & Mortality: RCT



ASA & Mortality: RCT



- ASA group and "no ASA" group should be comparable, since randomizing yields groups with similar baseline characteristics.

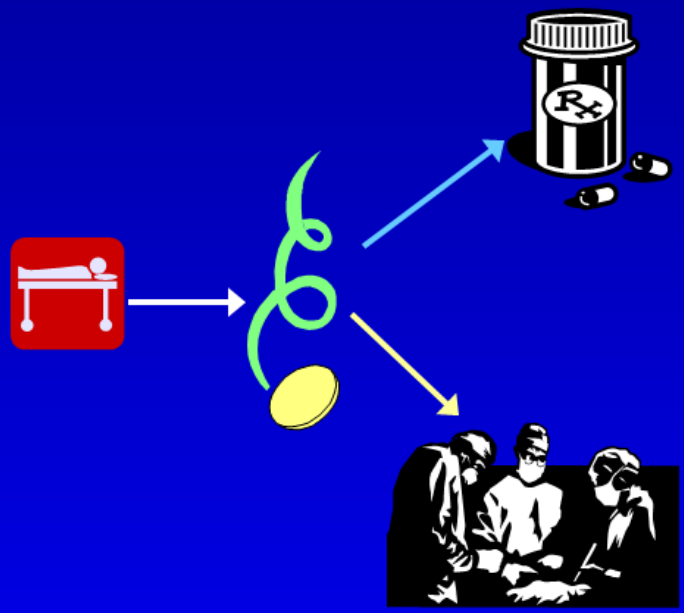
ASA and Mortality in Heart Patients

- Designing the Study

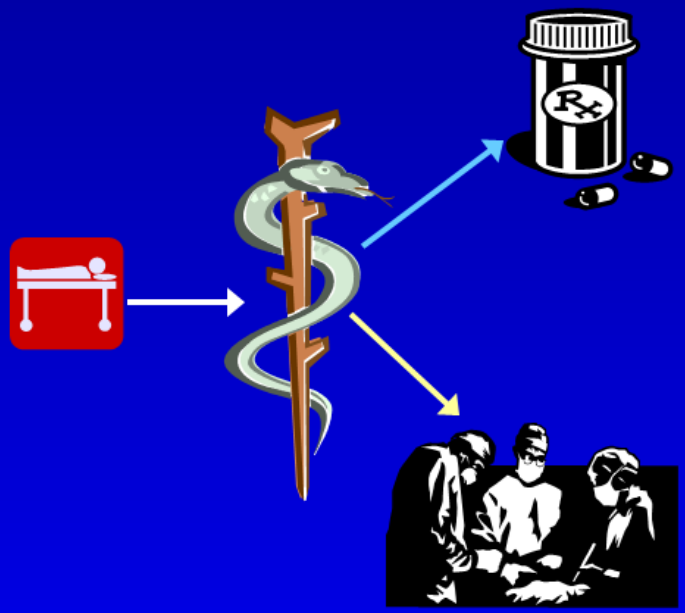
We want to understand aspirin's effect on all-cause five-year mortality among patients undergoing stress echocardiography.

- But what if we **cannot** do an RCT?

Randomized vs. Observational Studies



Randomization ensures that subjects receiving different treatments are comparable.



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

ASA & Mortality: **Observational Study**

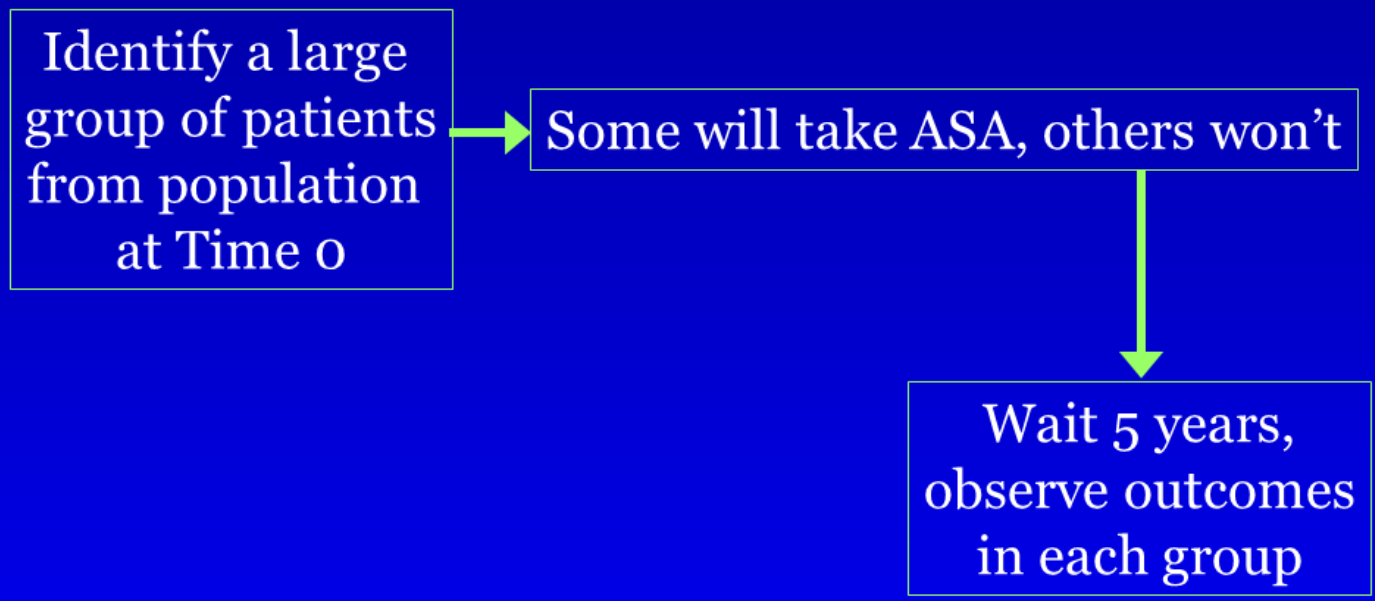
Identify a large group of patients from population at Time 0

ASA & Mortality: **Observational Study**

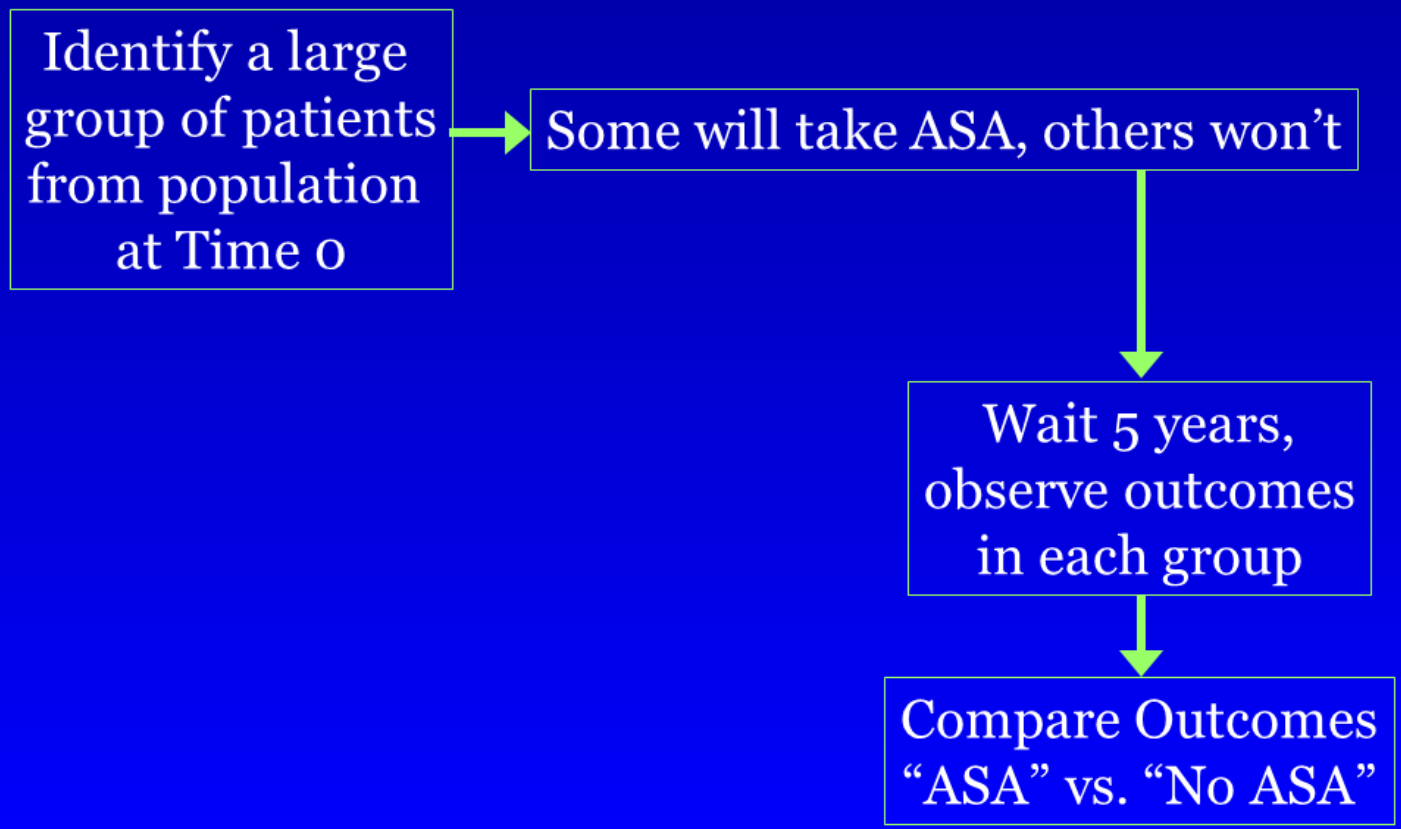
Identify a large group of patients from population at Time 0

→ Some will take ASA, others won't

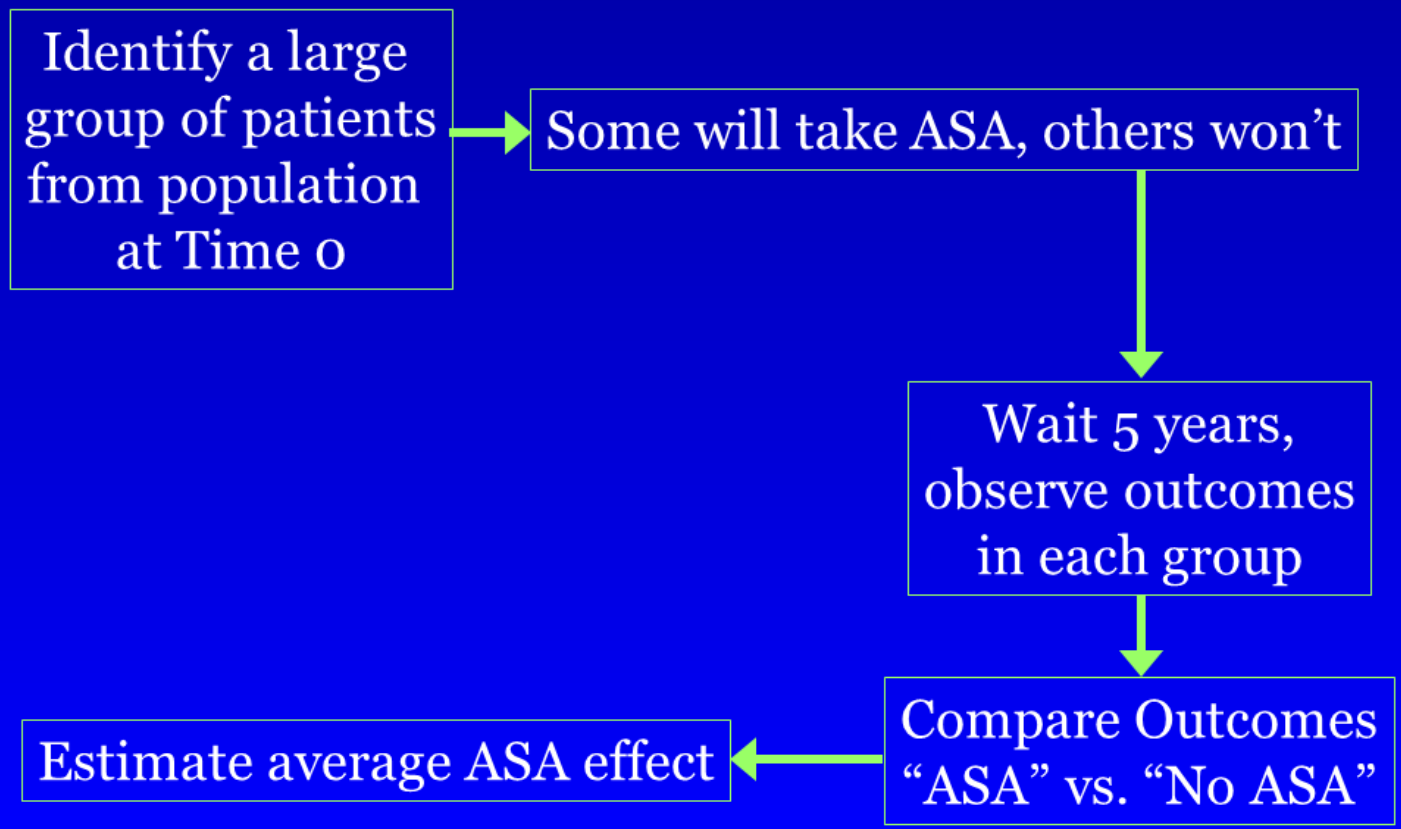
ASA & Mortality: **Observational Study**



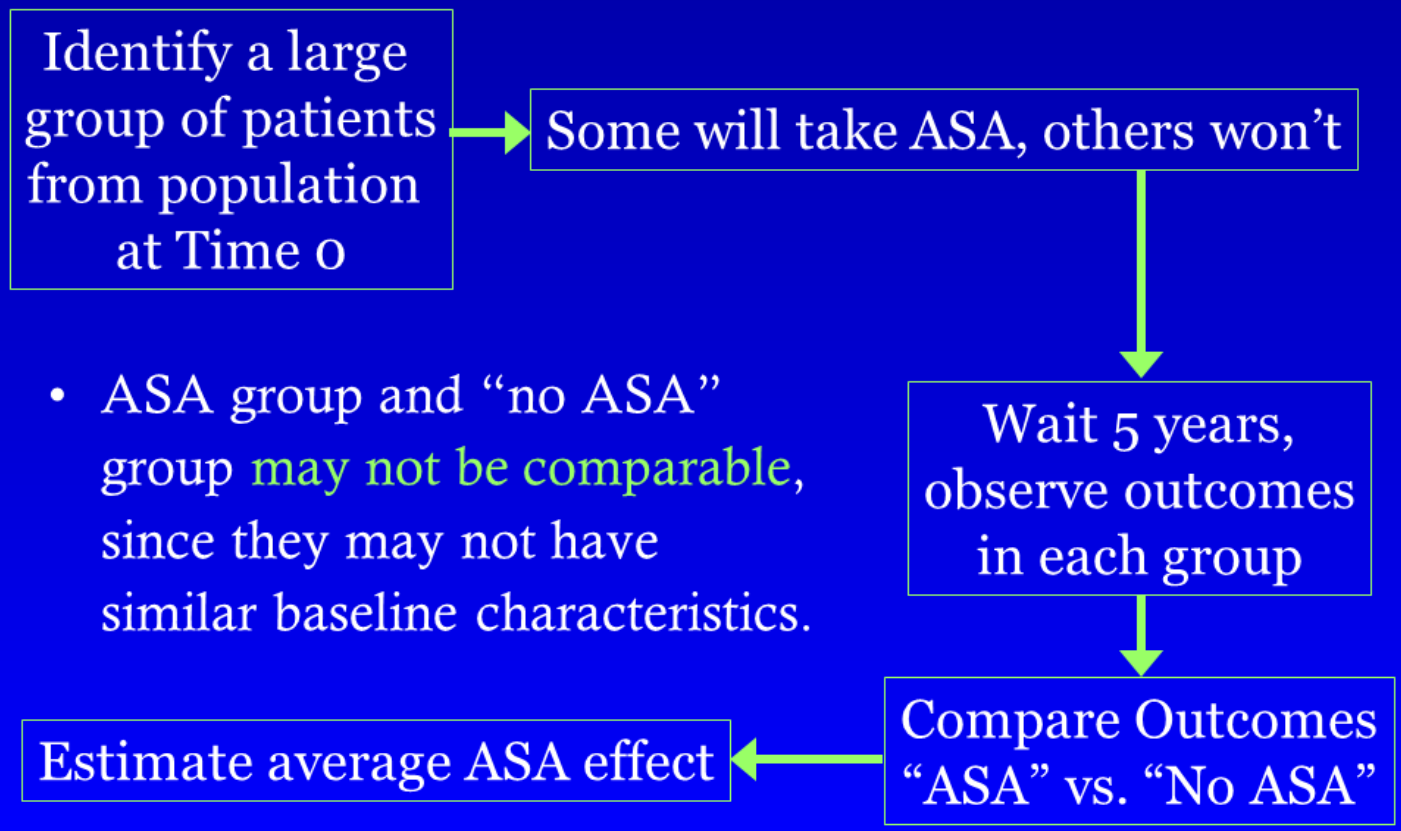
ASA & Mortality: **Observational Study**



ASA & Mortality: **Observational Study**



ASA & Mortality: **Observational Study**



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

How Do We Avoid Being Misled?

- What differentiates an observational study from a randomized controlled trial?
  - One key element: potential for selection bias.
- What is selection bias and what can we do about it?
  - Baseline characteristics of comparison groups are different in ways that affect the outcome.

We will often distinguish between overt and hidden bias.

- Overt Bias (seen in data - propensity scores can help)
- Hidden Bias (required data not collected - requires sensitivity analyses)

Aspirin Use and Mortality - The Real Study

6174 consecutive adults at CCF undergoing stress echocardiography for evaluation of known or suspected coronary disease[1].

- 2310 (37%) were taking aspirin (treatment).
- Main Outcome: all-cause mortality
  - Median follow-up: 3.1 years
- Univariate Analysis: 4.5% of aspirin patients died, and 4.5% of non-aspirin patients died.
  - Unadjusted Hazard Ratio: 1.08 (0.85, 1.39)

More on this study to come.

[1] Gum PA et al. Aspirin use and all-cause mortality among patients being evaluated for known or suspected coronary artery disease: A propensity analysis JAMA 2001 Sep 12; 286(10): 1187-94. https://pubmed.ncbi.nlm.nih.gov/11559263/

Section 2

Next Time...

Recorded Lecture 2

- How Can We Avoid Being Misled by Observational Studies?
  - What is **selection bias** and why should I care about it?
  - What can be done to deal with selection bias in observational studies?
- What is a propensity score, and how do we ...
  - estimate it,
  - see how well it's working, and
  - use it to estimate causal effects?

The slides I use will be posted to the Class 02 README.

Next Zoom - Thursday 2025-01-23

We'll meet via Zoom from 10 to 11:00 AM. Key things we'll discuss include:

- Lab 0
- Welcome to 500 survey
- Rosenbaum (2023) Chapters 1-3
- STROBE statement Items 13-21
- Abramson Chapter 2
- McGowan blog post
- Bradford Hill (1965)