Cohort Studies

Theoklis Zaoutis, MD, MSCE
Professor of Pediatrics and Epidemiology
University of Pennsylvania School of Medicine
Chief, Division of Infectious Diseases
The Children's Hospital of Philadelphia

Lecture Outline

- Definition
- Comparison to other study designs
- Advantages and Disadvantages
- Analysis
- Designing a cohort study
- Potential Bias and Study Design Issues
- Considerations in Exposures and Outcome Selection

Cohort

"A group of people who share a common experience or condition."

-Rothman

Examples of Cohorts:
Framingham population, Patients enrolled in Medicare

Cohort Study: Definitions

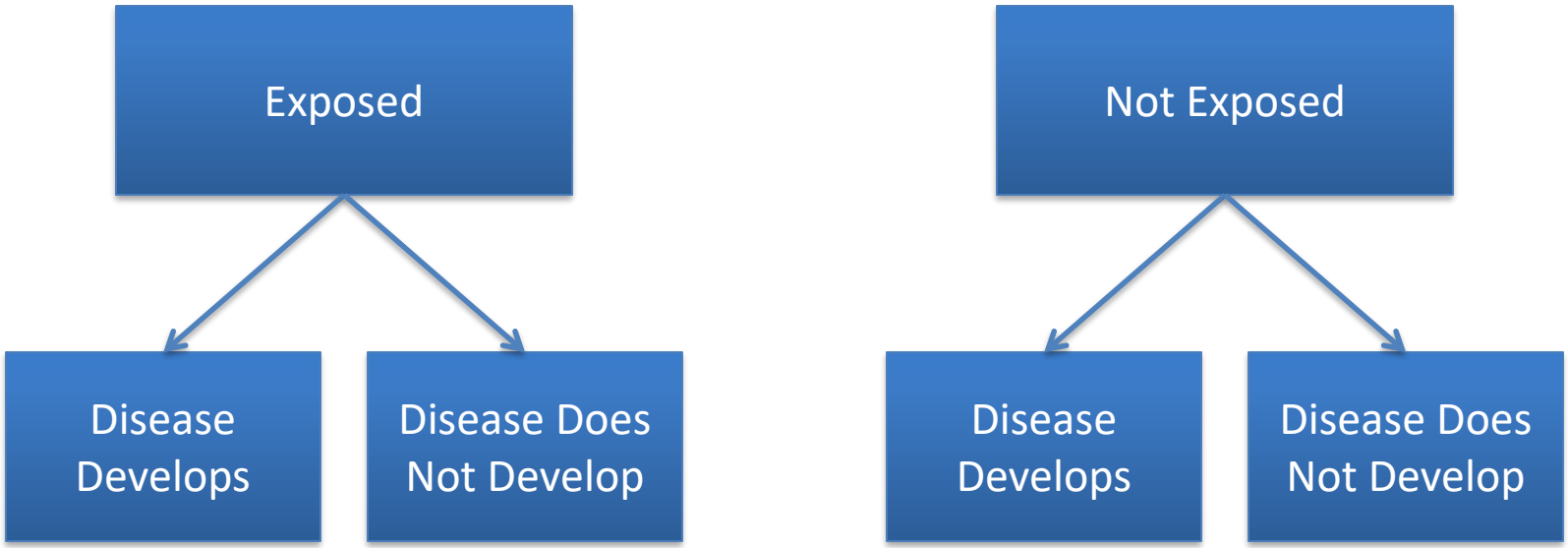
- Study to examine the incidence rate of an outcome among a group defined by a common exposure.
- Study to examine differences in outcome among at least two groups: one group with a risk factor or exposure compared to others without the risk factor or exposure (or with a different exposure).

Cohort Study: Goals

To determine:

1) the incidence of disease in one or more groups and/or
2) the relative association of an exposure with an outcome when compared between two or more groups.

Cohort Studies



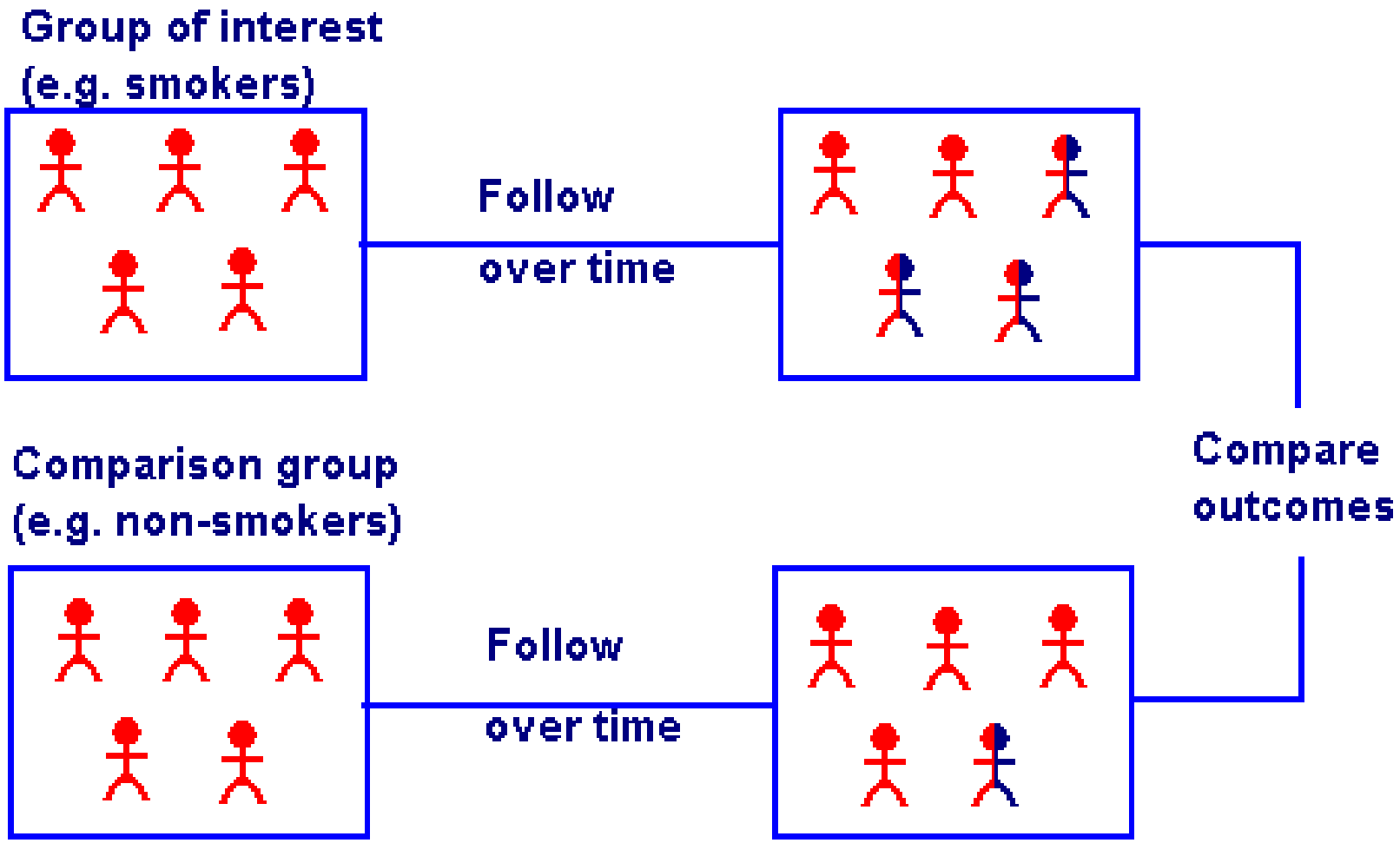
Gordis Figures 9-1 and 9-2

Cohort Studies: Defining Features

How patients are recruited:

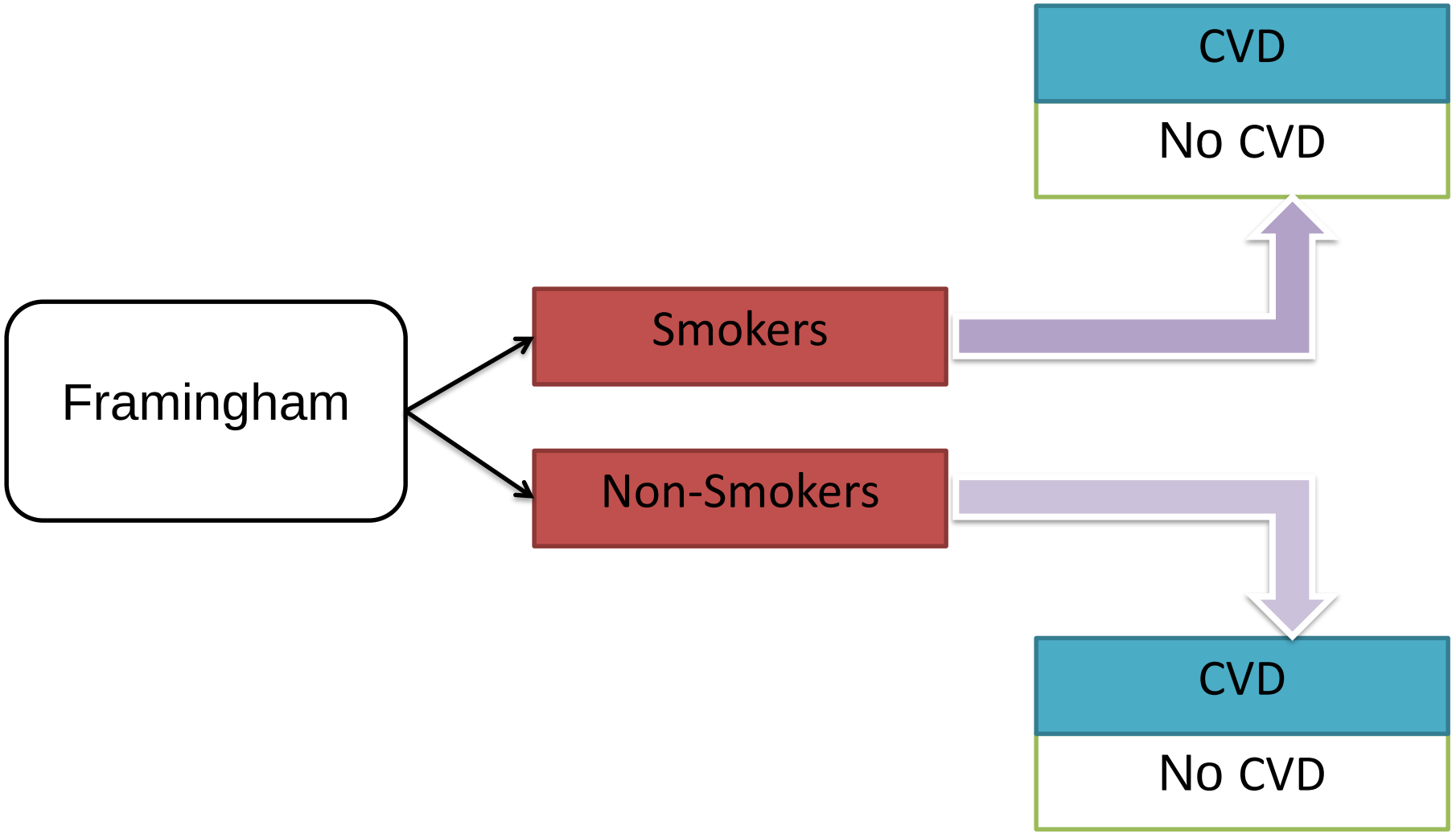
a) Individuals selected based on exposure or
b) Follow a single cohort of people (before exposure) and then designate exposed vs unexposed.

Cohort Study:
Selected based on Exposure

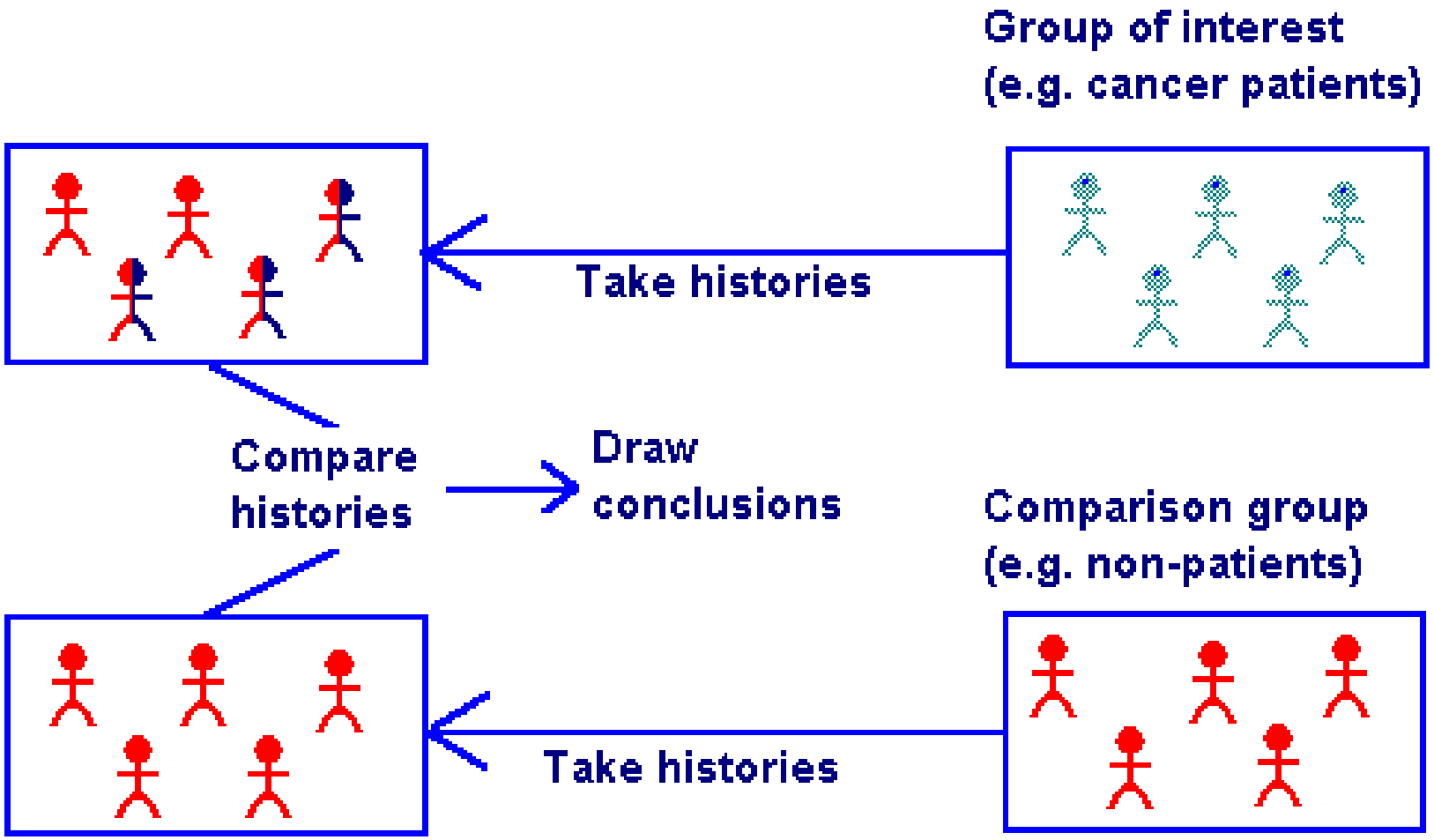


http://library.downstate.edu/EBM2/2400.htm

Framingham

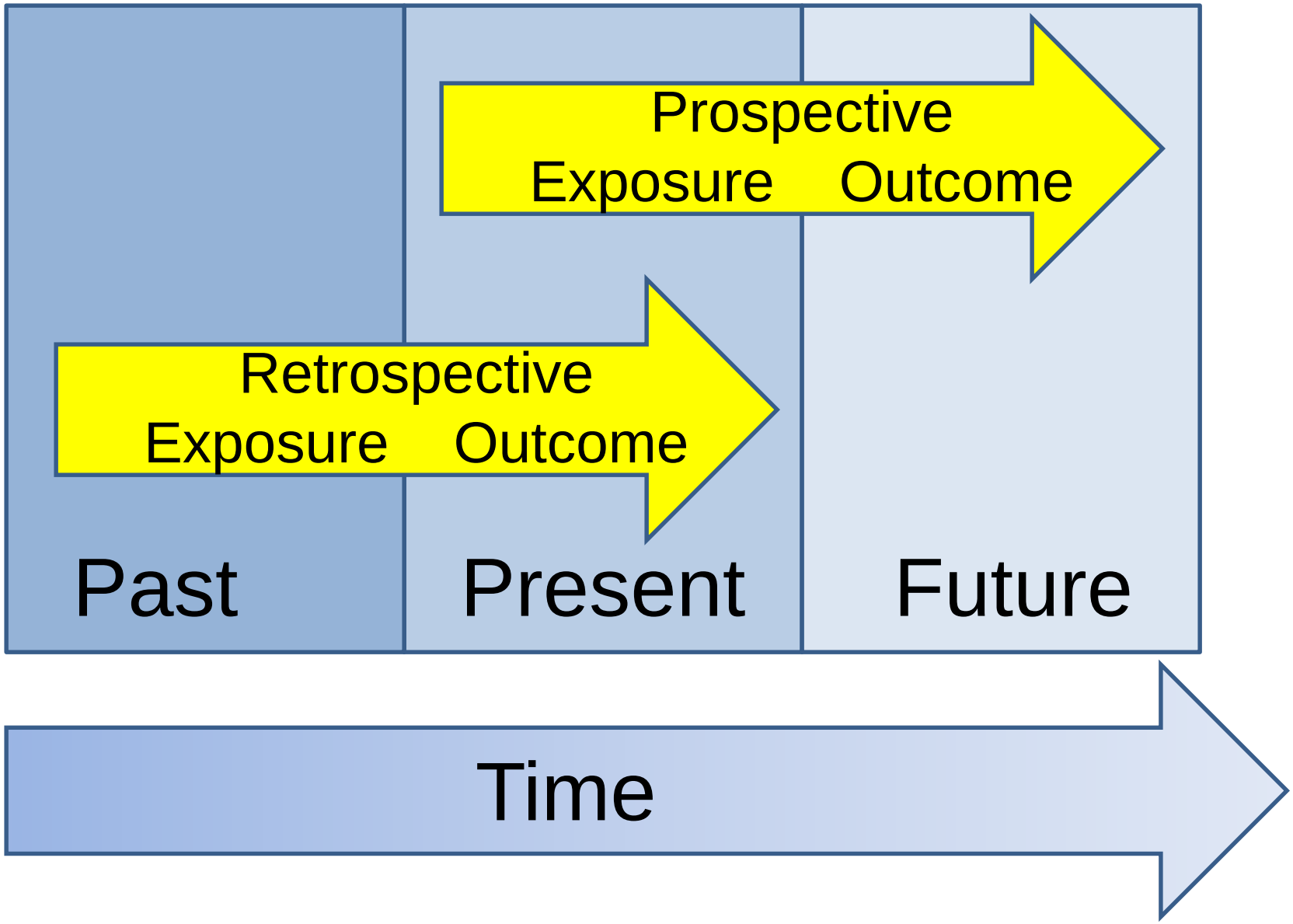


As Opposed to a Case-Control Study

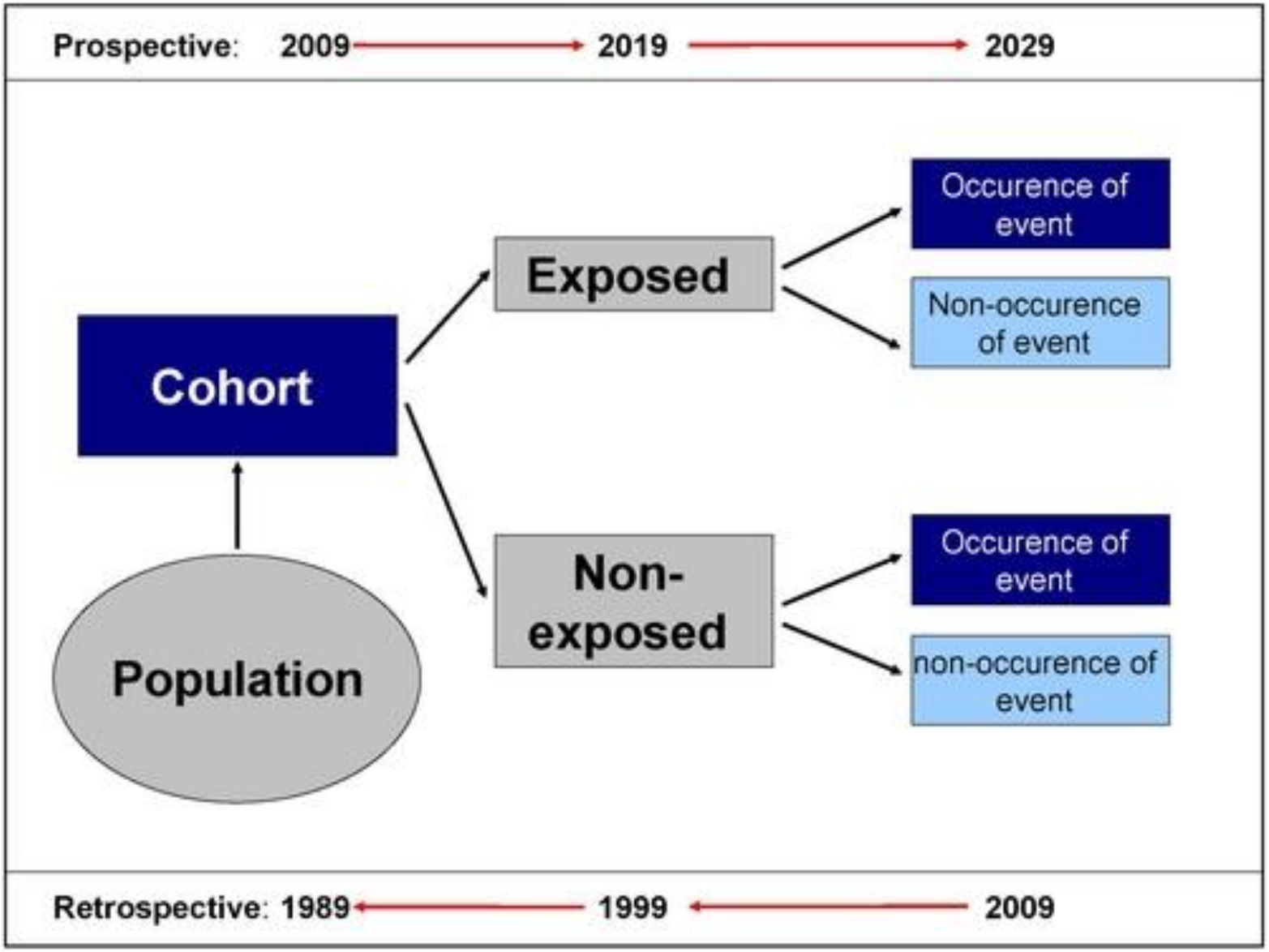


http://library.downstate.edu/EBM2/2500.htm

Prospective vs Retrospective



Prospective vs Retrospective Cohort Study



http://medanth.wikispaces.com/Cohort+Studies

Name the type of study . . .

- Patients in a community enrolled in a study to determine the risk of *C.difficile* (CDI). They are asked on the date of enrollment about their antibiotic status (receiving an antibiotic). Over the next ten years, they have yearly study visits to ascertain antibiotic exposure and the occurrence of CDI.

Name the type of study . . .

- Patients in a community enrolled in a study to determine the risk of *C.difficile* (CDI). They are asked on the date of enrollment about their antibiotic status (receiving an antibiotic). Over the next ten years, they have yearly study visits to ascertain antibiotic exposure and the occurrence of CDI.

Prospective Cohort Study

Name the type of study . . .

- In 2016 study aims to assess whether patients with MRSA infection disease are more likely to have received an antibiotic in the past year than those without an MRSA infection. Using data collected from 1995 to 2015 as part of routine primary care, the investigators identify patients with MRSA infection and a sample of the primary care patients who had not been diagnosed with MRSA infection. They compare the incidence of a new diagnosis MRSA during the period from 1995 to 2015 between the two groups, the one receiving antibiotics and the one without exposure to an antibiotic.

Name the type of study . . .

- In 2016 study aims to assess whether patients with MRSA infection disease are more likely to have received an antibiotic in the past year than those without an MRSA infection. Using data collected from 1995 to 2015 as part of routine primary care, the investigators identify patients with MRSA infection and a sample of the primary care patients who had not been diagnosed with MRSA infection. They compare the incidence of a new diagnosis MRSA during the period from 1995 to 2015 between the two groups, the one receiving antibiotics and the one without exposure to an antibiotic.

Retrospective Cohort Study

Cohort study compared to RCT

Cohort Study

- Exposure not determined by the investigator as part of the study design
- More external validity (more generalizable)

Randomized Control Trial

- Exposure determined by the investigator, usually with random assignment
- More internal validity (less systematic bias)

What issues arise because of the lack of randomization?

Cohort study compared to RCT

Cohort Study

- Exposure not determined by the investigator as part of the study design
- More external validity (more generalizable)

Randomized Control Trial

- Exposure determined by the investigator, usually with random assignment
- More internal validity (less systematic bias)

What issues arise because of the lack of randomization?

Imbalance in confounders

Selection bias

Advantages of Cohort Studies

- Can study multiple outcomes
- Can study uncommon exposures
- Ability to calculate incidence rates
- Can establish a temporal relationship and can help establish cause and effect

Disadvantages of Cohort Study

- More expensive
  - Large sample sizes for uncommon events
- Possibly biased outcome data
- Long
  - May take years to complete prospective study
  - Loss to follow up
  - Changes over time in criteria
  - Costly methods

What can you measure in a cohort study?

- Prevalence
- Incidence
- Relative Risk
- Risk Difference
- Attributable Risk and Risk Proportion
- Hazard Ratios/Time-to-Event Analysis

Incidence and Prevalence

$$\text{Incidence} = \frac{\text{Number of } \textit{new} \text{ cases of a disease over a period of time}}{\text{Number of people at risk of developing that disease during that time}}$$

$$\text{Prevalence} = \frac{\text{Number of } \textit{existing} \text{ cases of a disease at a specified point in time}}{\text{Number of people in the population at that time}}$$

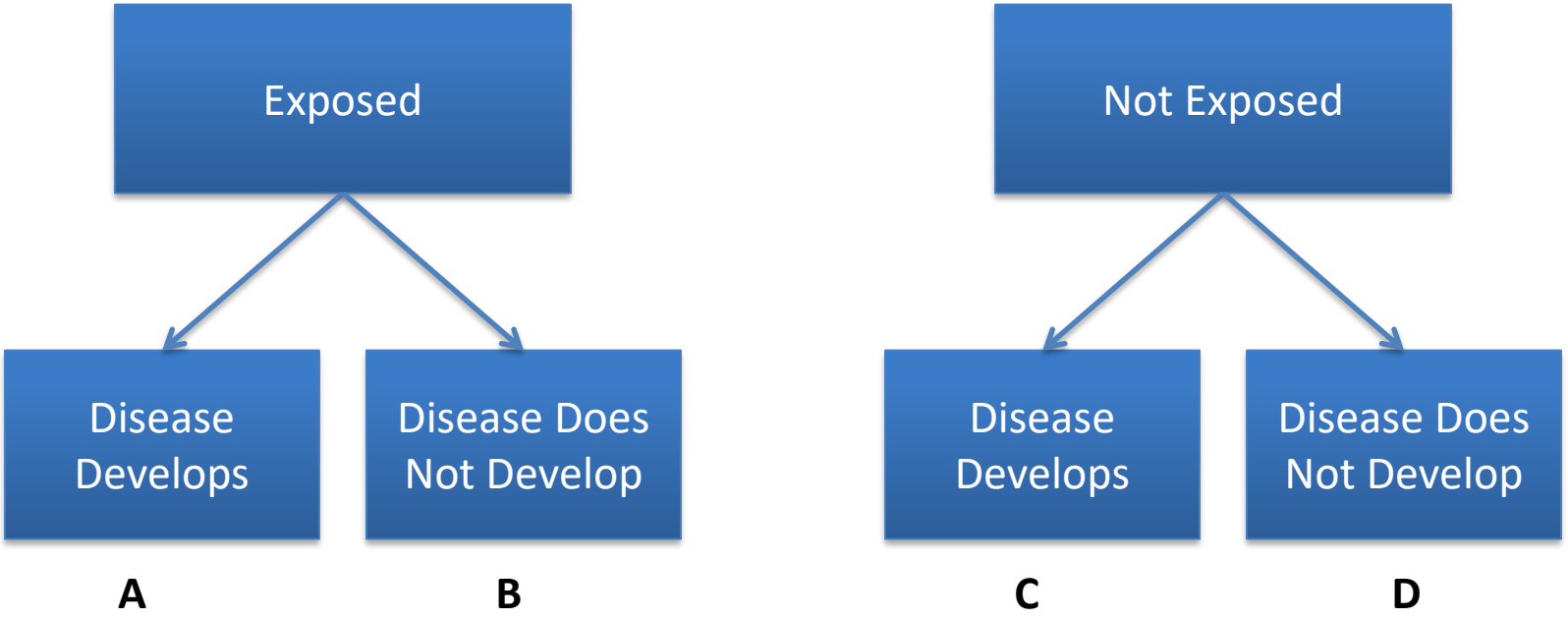
Incidence and Prevalence

- Point prevalence (at specific point)
- Period Prevalence (over time period)

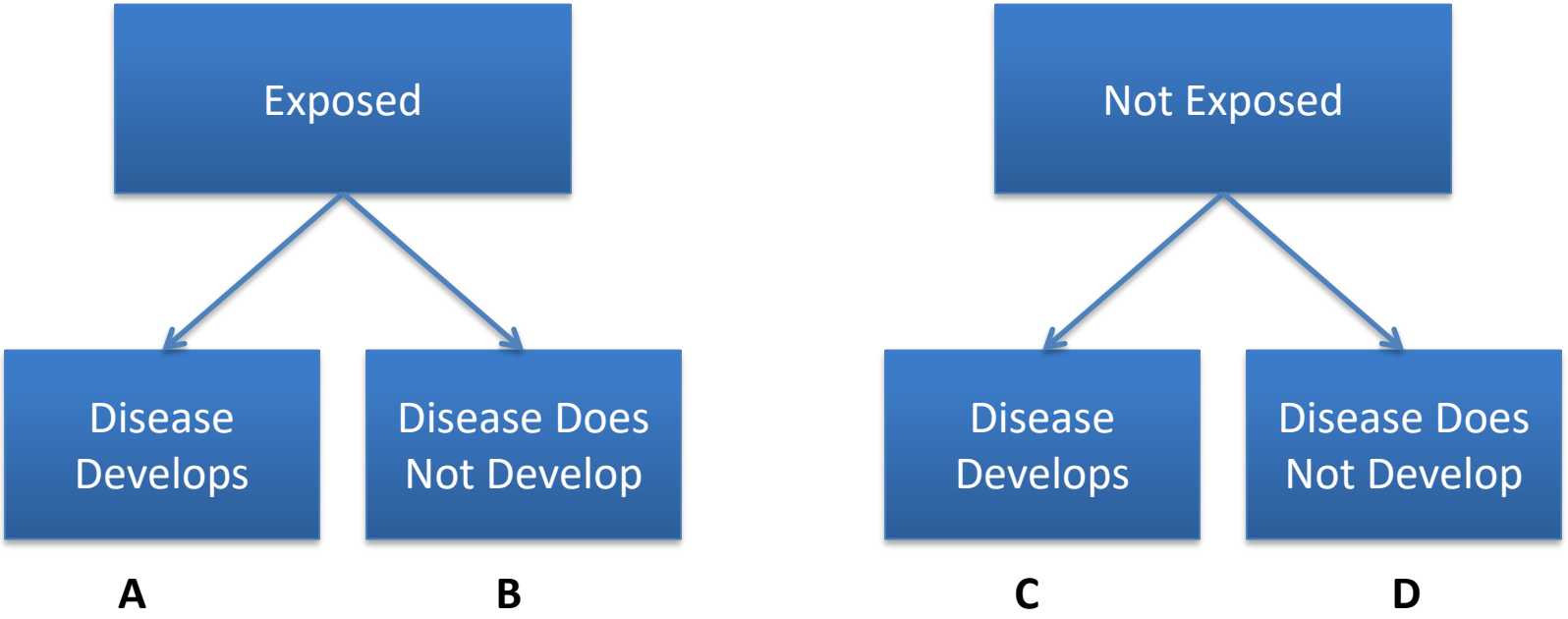
Relative Risk (RR)

- Ratio of incidence in the exposed compared to the unexposed (I_e/I_u)
  - Estimation of average risk, rates, or occurrence times.
  - Assumes everyone followed for their whole time in the cohort.
- RR: Relative Risk/Risk Ratio/Rate Ratio
  - RR<1 means the event is less likely to occur in the exposed group compared to the unexposed group
  - RR>1 means the event is more likely to occur in the exposed group compared to the unexposed group

Relative Risk



Relative Risk



	Disease	No Disease	Incidence Rate
Exposed	A	B	a/(a+b)
Not Exposed	C	D	c/(c+d)

Relative Risk

	Disease	No Disease	Incidence Rate
Exposed	A	B	a/(a+b)
Not Exposed	C	D	c/(c+d)

$$RR = I_e/I_u = [a/(a+b)]/[c/(c+d)]$$

Note: this assumes that each patient was followed for the same amount of time.

Calculate the Relative Risk

- A cohort study seeks to determine the association between antibiotics and resistance One hundred persons receiving antibiotic X and 100 not receiving antibiotic X are enrolled and followed for 5 years. There was no loss to follow up. At the end of the study, 30 persons who received antibiotics and 10 who did not developed resistance

Calculate the Relative Risk

	Resistance	No Resistance	5-year Cumulative Incidence
Antibiotic X	30	70	30/100
No Antibiotic X	10	90	10/100

$$RR = I_e/I_u = [0.3]/[0.1] = 3$$

Confidence Intervals

- "Range of relative risks within which the true relative risk for the entire theoretical population is most likely to lie." – Strom, Pharmacoepidemiology, Ch2
- What are the following likely to mean?
  - RR 1.0, 95% CI: (0.9-1.1)
  - RR 10.0, 95% CI: (8.9-11.1)
  - RR 5.0, 95% CI: (0.9-11.1)

Risk Difference

- AKA Absolute Risk Difference
- AKA Attributable Risk
- Difference in incidence between the experimental and control groups

- **$RD = I_e - I_u$**

Attributable Risk

- **Attributable risk (i.e. Risk Difference):**
  - absolute difference in incidence between exposed and unexposed
  - $I_e - I_u$
- **Population Attributable Risk:**
  - When the incidence in exposed is replaced with incidence in the total population we compute the Population Attributable Risk (PAR):
  - Incidence of disease in the population that is attributable to the exposure (i.e. would be reduced if the exposure was eliminated)
  - $I_{pop} - I_u$
- **Attributable proportion**
  - proportion of all cases in a total defined population which can be ascribed to an exposure
  - $(I_e - I_u)/I_e$
- **Population attributable risk proportion**
  - Proportional Reduction in incidence that would be observed if the population were entirely unexposed (compared to the current exposure pattern). P=prev of exposure in pop.
  - $(I_{pop} - I_u)/I_{pop} = P(RR-1)/[P(RR-1)+1]$

Attributable Risk Example

Calculate the yearly attributable risk of CDI on patients receiving antibiotic X.

The adjusted HR for CDI was 1.53.

The baseline incidence rate of CDI in the general population was 0.0115902 events per year.

Attributable Risk Example

Incidence in the Exposed

$I_u * HR = I_e \rightarrow 1.53 * 0.012 = 0.018$

Attributable Risk / Risk Difference

$= I_e - I_u \rightarrow 0.018 - 0.012 = 0.006$

Attributable Risk Proportion

$= (I_e - I_u)/I_e \rightarrow 0.006/0.018 = 33\%$

Relative Risk: Accounting for Time

- Accounts for different entry and dropout rates → varied duration of follow up (and time at risk)
- Assumes each year (or time increment) is equal

Relative Risk: Accounting for Time

	Event	Person Years	Cumulative Incidence Rate
Exposed	A	PYe	A/PYe
Not Exposed	C	PYu	C/PYu

$$RR = I_e/I_u = [A/PYe]/[C/PYu]$$

Calculate the Rate Ratio

- Investigators aimed to determine the association between antibiotics and development of IBD. They retrospectively examined 12,000 children exposed to antibiotics and 12,000 children who did not receive antibiotics **over an average of 5 and 5.3 years of follow up** respectively. 512 incident cases of IBD were detected: 358 among antibiotic recipients and 154 among non-recipients.

Loosely based on Kronman, et al. Pediatrics

Calculate the Rate Ratio

	IBD Dx	Person Years	Cumulative Incidence Rate
Antibiotics Yes	358	60,000	0.006
Antibiotics No	154	63,600	0.002

$$RR = I_e/I_u = [0.006]/[0.002] = 3$$

Part II: Logistics of a Cohort Study

When would you choose a cohort study?

- You hypothesize an exposure is associated with an outcome or particularly if multiple outcomes
  - One of the major advantages of cohort studies is the ability to study the relative risk of multiple different outcomes with a single exposure
- Not able to use an experimental study design

Planning your cohort study . . .

1. What's the question? Is a cohort study the appropriate study design?
2. What's the hypothesis? Form conceptual model.
3. What's the population? What's the underlying cohort?

Planning your cohort study . . .

4. Define the exposures
5. Define the comparator group
6. Define the outcomes
7. How long should patients be followed?
8. Sample sizes

Planning your cohort study . . .

9. Which variables need to be collected to address your hypotheses? Future studies?
   - The study is only as good as the data collected
10. How will subjects be followed?
   - Must have a very well planned study protocol so information is collected in the same way
11. Will there be internal validity? External validity (generalizability)?
12. Is it feasible?

Selecting a Study Population

Select individuals based on exposure and non-exposure

OR

Select defined population and follow for exposure and outcome

Different types of cohorts/populations

- Open population
  - Can move in and out of the "system" – e.g. commercial insurance plan
- Closed population
  - Only includes those present at the start of the study.
  - No new entries (and some define as no loss to follow-up)
  - E.g. Hiroshima bombing victims
- Prevalent disease/exposure cohort
  - Any patient with the disease or exposure, no matter how long they've had it, are eligible for inclusion
- Inception cohort
  - Only patients with incident exposure are enrolled in the "exposed group"
  - Referred to as "new user" design when studying medications

Open vs Closed Cohort

- Open cohort – allows the populations to grow over time
- Closed cohort – Population will shrink over time but don't have to worry about changing reasons for exposure over time (e.g. changing prescribing patterns)

Prevalent vs Inception Cohort

- Prevalent
  - Larger sample size
  - Mix of duration of exposure
- Inception cohort
  - Approximates the natural history of exposure
  - Eliminates risk of loss of susceptible subjects
    - People at highest risk may have the outcome event before the start of the study such that prevalent exposure group are at lower risk than newly exposed

Exposures

- Spend a lot of time really thinking about what you want to measure it and how you want to measure it.
- How will you define your exposure?
  - Cumulative exposure, average exposure, levels of exposure, intensity or level exposure up to the start of the study
  - How do you know your definition is valid?
  - Do the exposed in your exposure group represent the exposed in the general population? (Generalizability)

Disease as exposure

- When did the disease really begin? (lag time in diagnosis?)
  - How long do people have asymptomatic disease prior to diagnosis?
  - May be longer with diseases that tend not to cause profound symptoms at early stages
    - Cancer
    - Precancerous lesions
    - Diabetes
    - Hypertension
- Does “time at risk” or disease duration matter?

Time Varying Exposures

- One individual can contribute to more than one exposure group.

Selection of comparison group

- As important as selection of the exposed group – this will define your study question
  - Compare children with MRSA vs. children without a diagnosis of MRSA
  - Compare children with MRSA vs. children with VRE
- How do you know they are free of the exposure?
- Same inclusion/exclusion criteria for the exposed should be applied to unexposed
- How many unexposed subjects per exposed subject?
- Some cohort studies have no comparator group

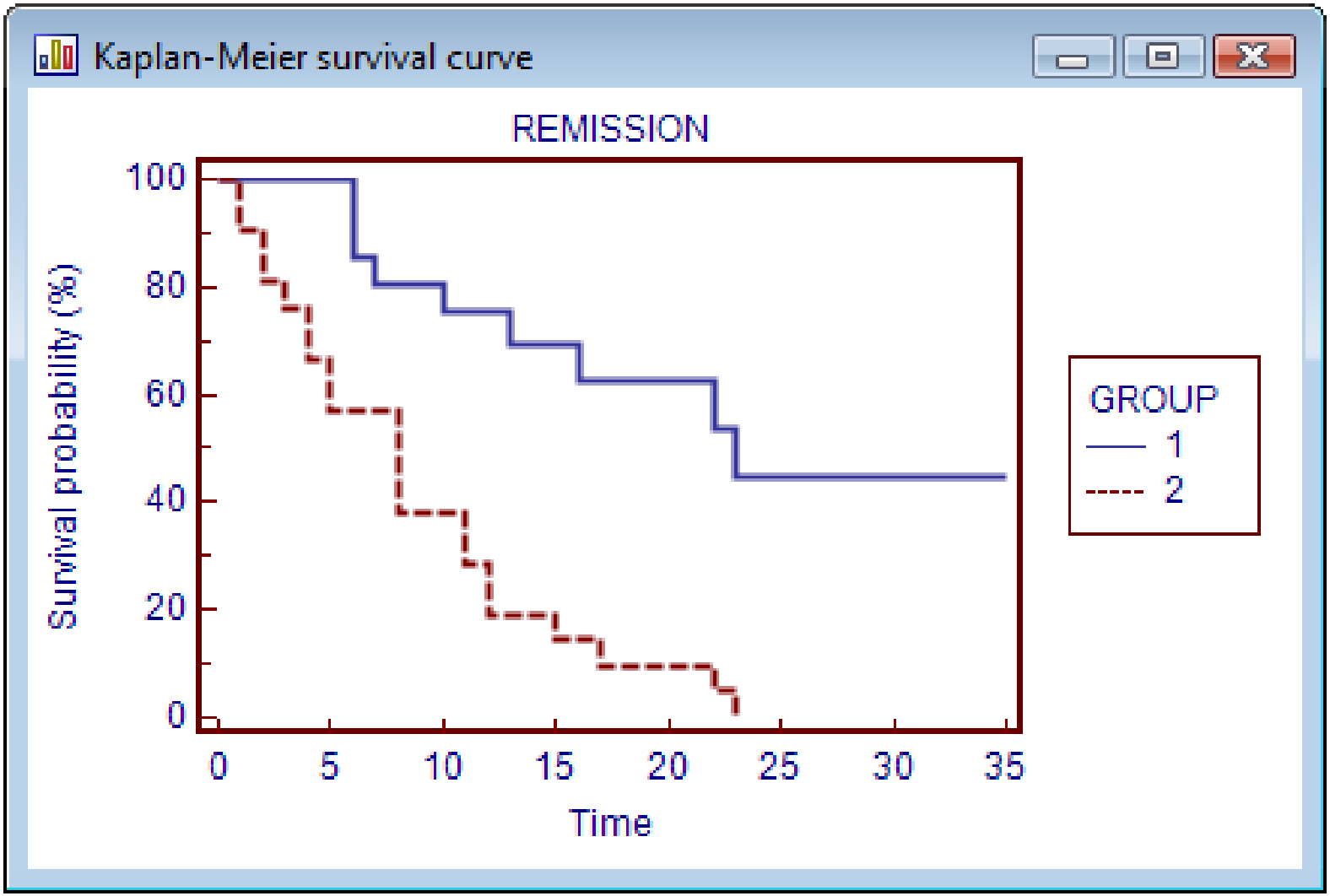
Outcomes

- Is the event likely to be captured?
- Is there any ambiguity in the outcome definition?
- How will you define time of onset of the outcome?
  - When does resistance start?
- Competing risks – reason for dropout may differ between groups and be related to the risk of the outcome in the future

Analytic Considerations

- How will you report your results?
  - (e.g. cumulative incidence, incidence rate, RR, IRR, HR, etc)
- Is unequal follow up time considered?
  - Special statistical methods that account for variable follow up

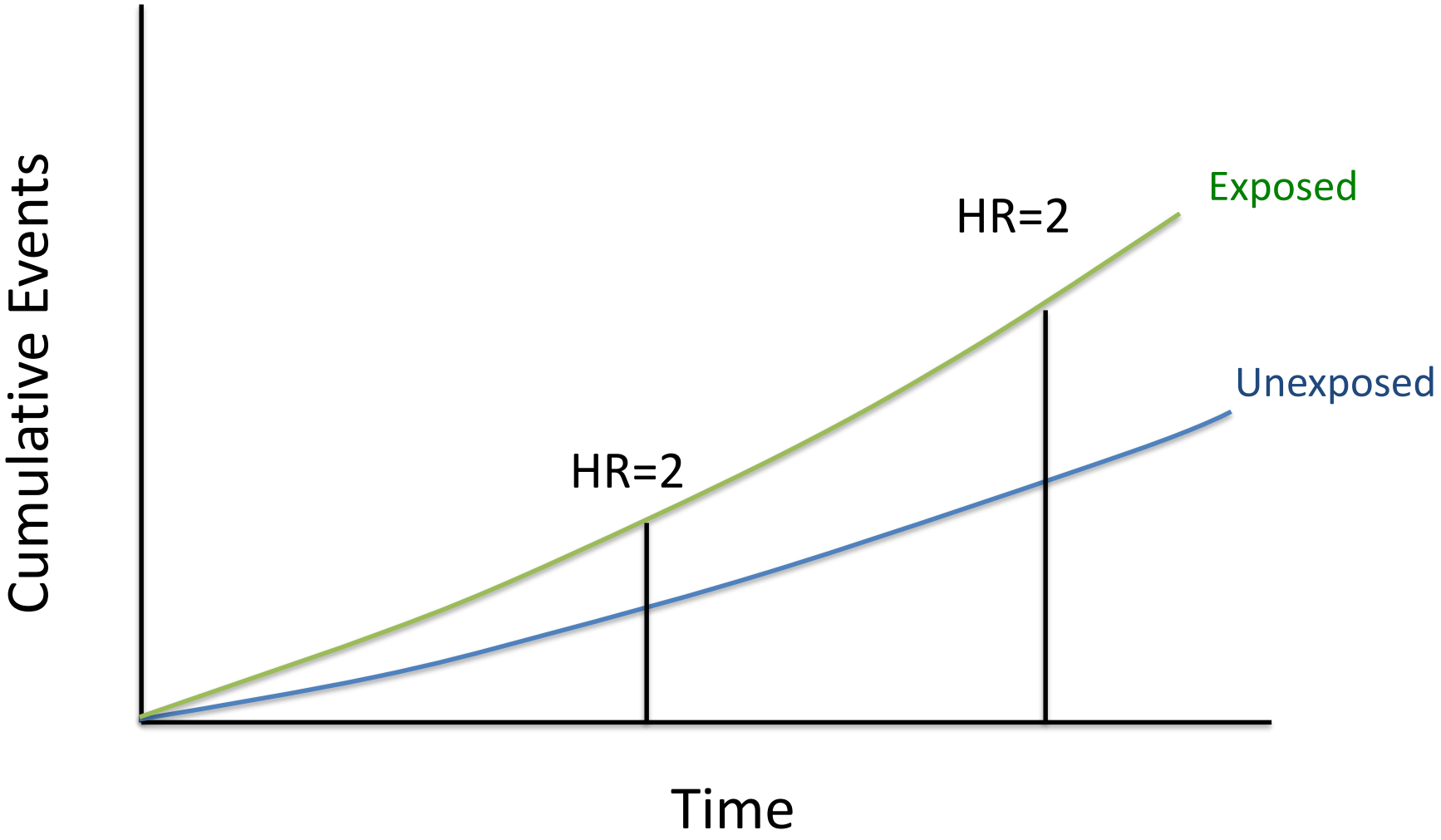
Time-to-Event Analysis: Kaplan Meier Curves



Time-to-Event Analysis: Hazard Ratios

- Ratio of risk in exposed to risk in unexposed
  - RR are cumulative over the entire study period
  - HR are instantaneous risk over the study time period: basically calculated at each event and then averaged.
  - Captures time to the event (event happening earlier in one group than another)
- Accounts for time in the cohort (person years)
  - Accounts for different entry and dropout rates (varying durations of follow up and thereby varying risks)
- Calculated using Cox proportional hazards models
- Assumes the ratio between the two groups is constant over time (proportional hazards assumption)

Time-to-event analyses: Hazard Ratios



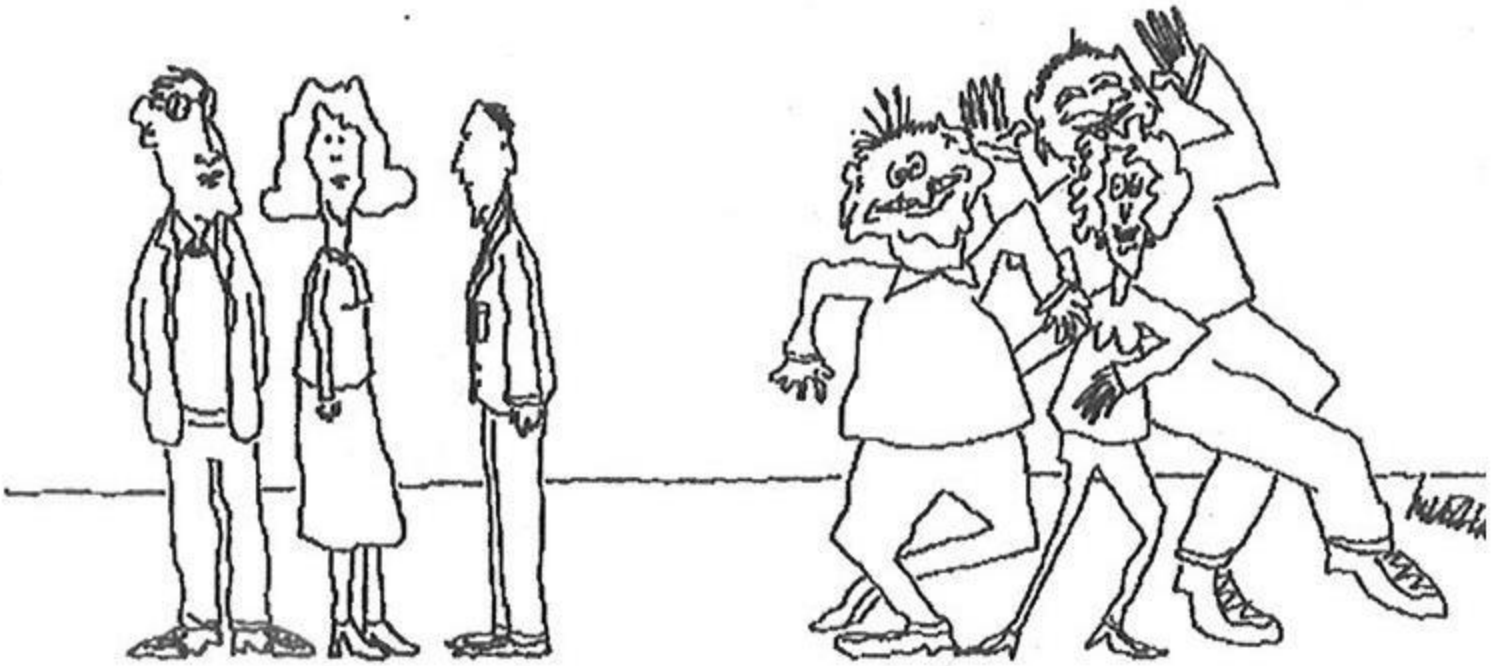
HR=2 means that at any given point in time, patients in the exposed group are twice as likely as patients in the unexposed group to develop the outcome.

Analysis

- How will you handle loss to follow up, withdrawals, and missing data?
- Multiple imputation
- Cohort effect: changes and variation in disease or health status of a study population over time – also generation effect.

Potential Bias in Cohort Studies

- Ascertainment bias/information bias
  - quality and extent of data may be different between two groups
- Knowing the hypotheses of the study may bias diagnosis or assessment of the outcome
- Non-response and loss to follow up
  - May refuse to participate more likely if have one or the other exposure
  - May drop out from one exposure preferentially


CONTROL GROUP
OUT OF CONTROL GROUP.