The Epidemiology Toolbox: Epi Concepts 101

Ebbing Lautenbach, MD, MPH, MSCE

University of Pennsylvania School of Medicine

Nothing to disclose

Perelman School of Medicine
University of Pennsylvania

Outline

- Definitions / Historical Perspective
- Measures of Disease Occurrence / Measures of Effect
- Types of Study Design
- Study Design Issues
- Summary

Epidemiology

- Definition: The study of the distribution and determinants of health and disease in populations
- Basic science of public health and preventive medicine

Epidemiology

- The study of the *distribution* and determinants of health and disease in populations

Measures of Disease Occurrence & Measures of Effect

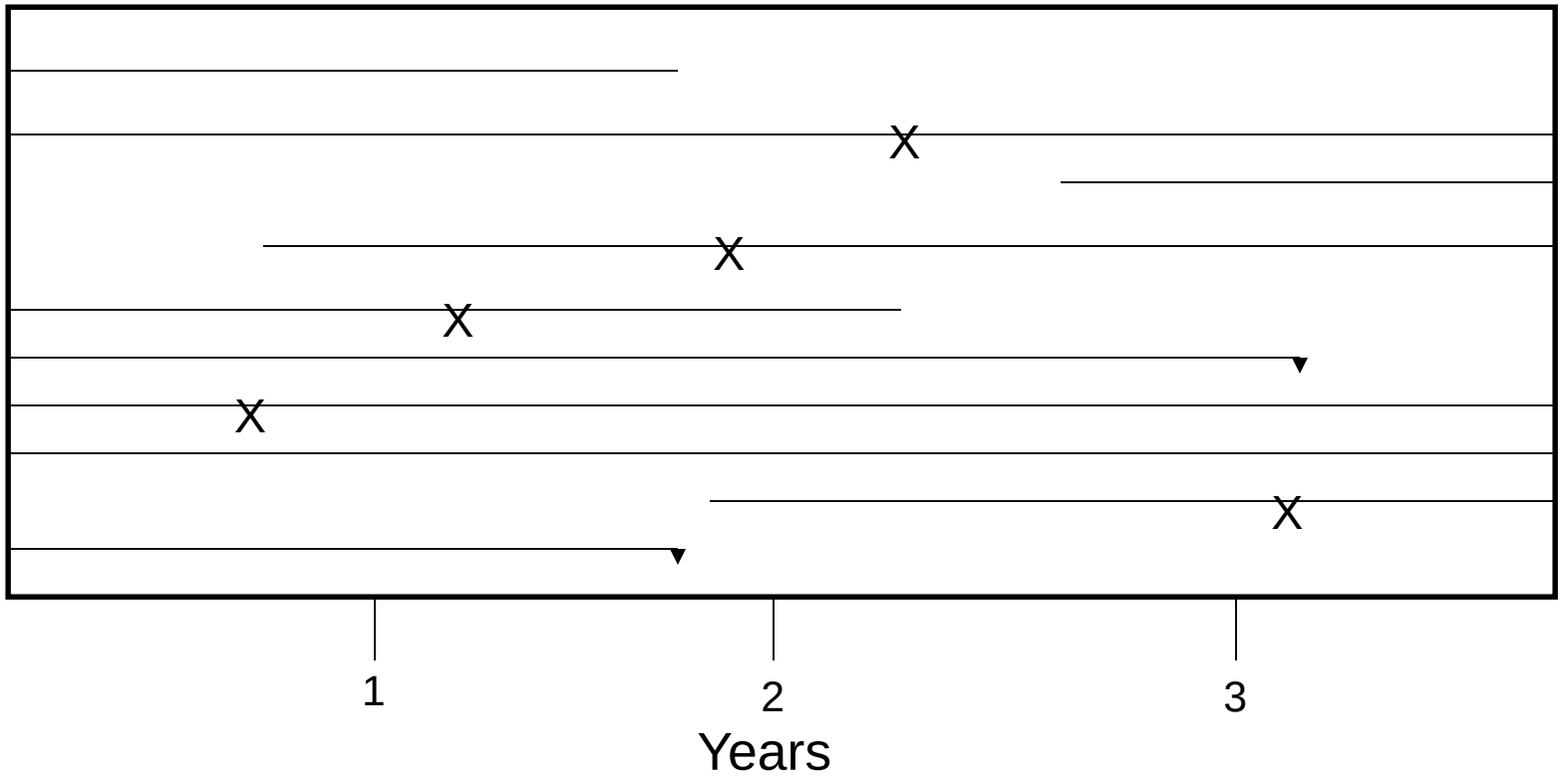
- Prevalence
- Cumulative incidence
- Incidence rate
- Relative risk
- Attributable risk

Prevalence

$$\text{Prevalence} = \frac{\text{number diseased individuals}}{\text{total population}}$$
(at a given point in time)

- Estimates the burden of disease
  - Useful in setting priorities, allocating resources
- Dependent on incidence and duration of disease

Prevalence



Time point	Prevalence
0.5 years	0/7
1.0 years	1/8
2.5 years	3/6
4.0 years	4/6

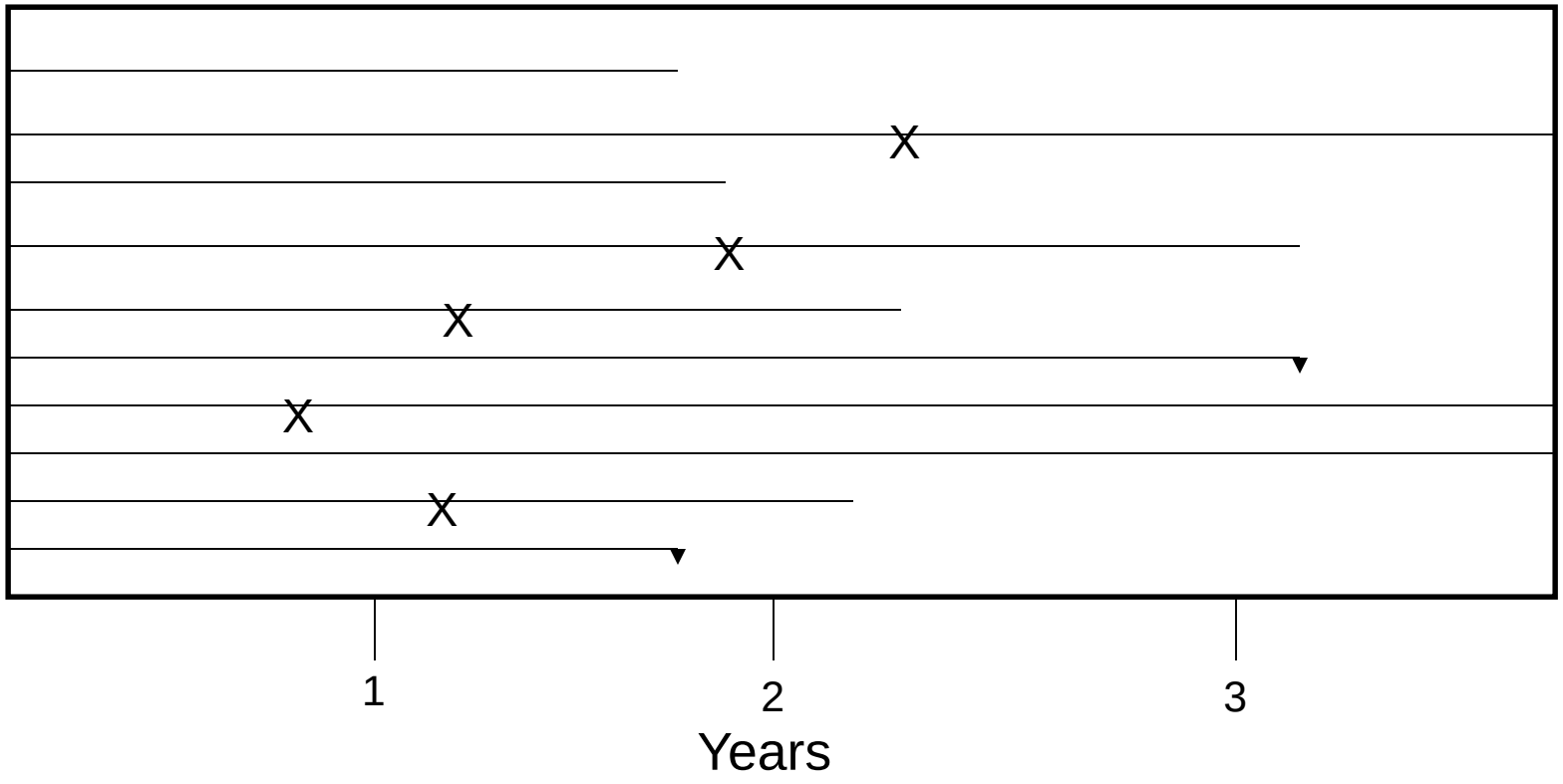
Cumulative Incidence

Cumulative incidence =

$$\frac{\text{number of new cases of disease between } t_0 \text{ and } t_1}{\text{total disease free individuals at risk of disease at } t_0}$$

- Assumes complete follow up
  - (use incidence rate when follow up incomplete)
- Must refer to a specific time period
- Does not tell you when in the time period a case occurred

Cumulative Incidence



Time point	Prevalence
0.5 years	0/10
1.0 years	1/10
1.5 years	3/10
2.0 years	?

Cumulative Incidence in Hospital Infections

- Cumulative incidence of HAIs
  - Implied time period is the course of hospitalization until a first event or until discharge without first event
  - However, patients do not all stay in hospital and remain at risk for exactly the same period of time.
  - Most HAIs are time related
  - Comparing cumulative incidence of HAIs among patient groups with differing lengths of stay may be misleading.
- Infections related to a point source
  - Generally not time related
    - Tuberculosis (from a contaminated bronchoscope)
    - Surgical site infections (from the operation)
  - In this case, cumulative incidence is excellent measure of incidence.

Incidence Rate

Incidence Rate (incidence density) =

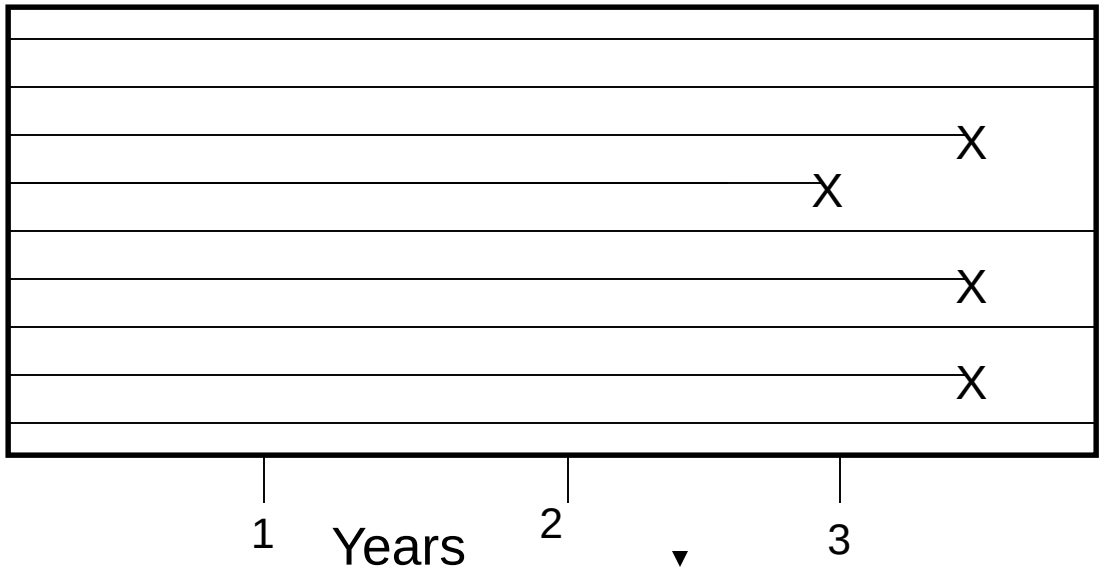
$$\frac{\text{number of new cases of disease during given time period}}{\text{total person-time of observation among individuals at risk}}$$

- Does not assume complete followup
- Time as a denominator (Units = time $^{-1}$)
  - Accounts for different entry/dropout rates
  - Assumes all time periods are equivalent

Incidence Rate in HAIs

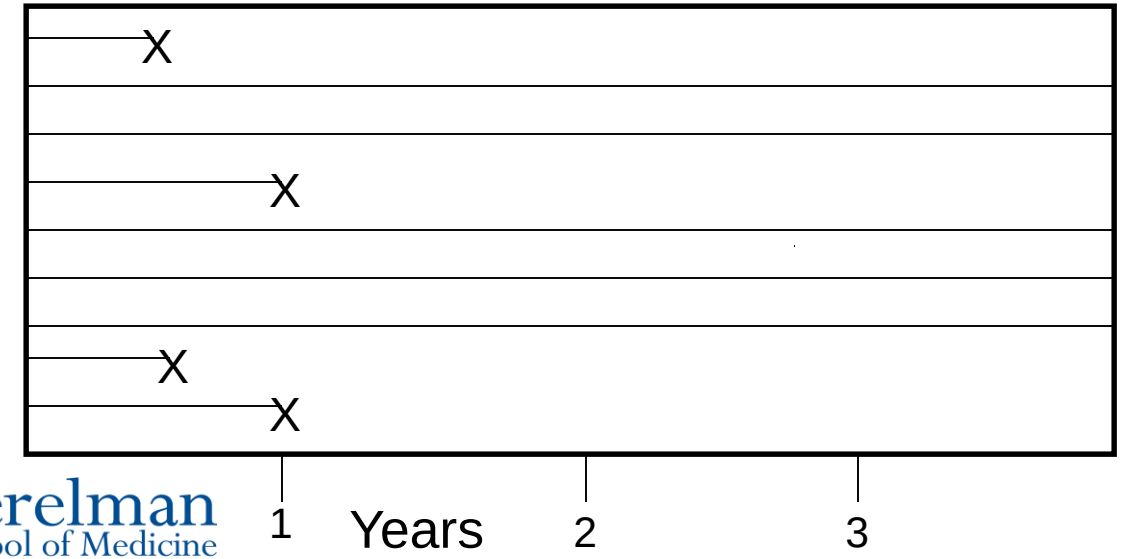
- Incidence rate valuable when comparing HAI rates in groups which differ in their time at risk (e.g., short-stay patients vs. long-stay patients)
  - The incidence rate (i.e., risk per day) is the most convenient way to correct for time
    - Separate the effect of time (duration of exposure) from the effect of daily risk
  - In hospital epidemiology, incidence rates usually expressed as the number of first events in a certain number of days at risk (e.g., HAIs per 1,000 hospital days,)
- Incidence rate is usually restricted to first events (e.g., the first episode of a specific HAI).
  - Second events are not statistically independent from first events in the same individuals (i.e., patients with a first event are more likely to suffer a second event).

CI vs IR



CI: 4/9

IR: 0.12/ptu



CI: 4/9

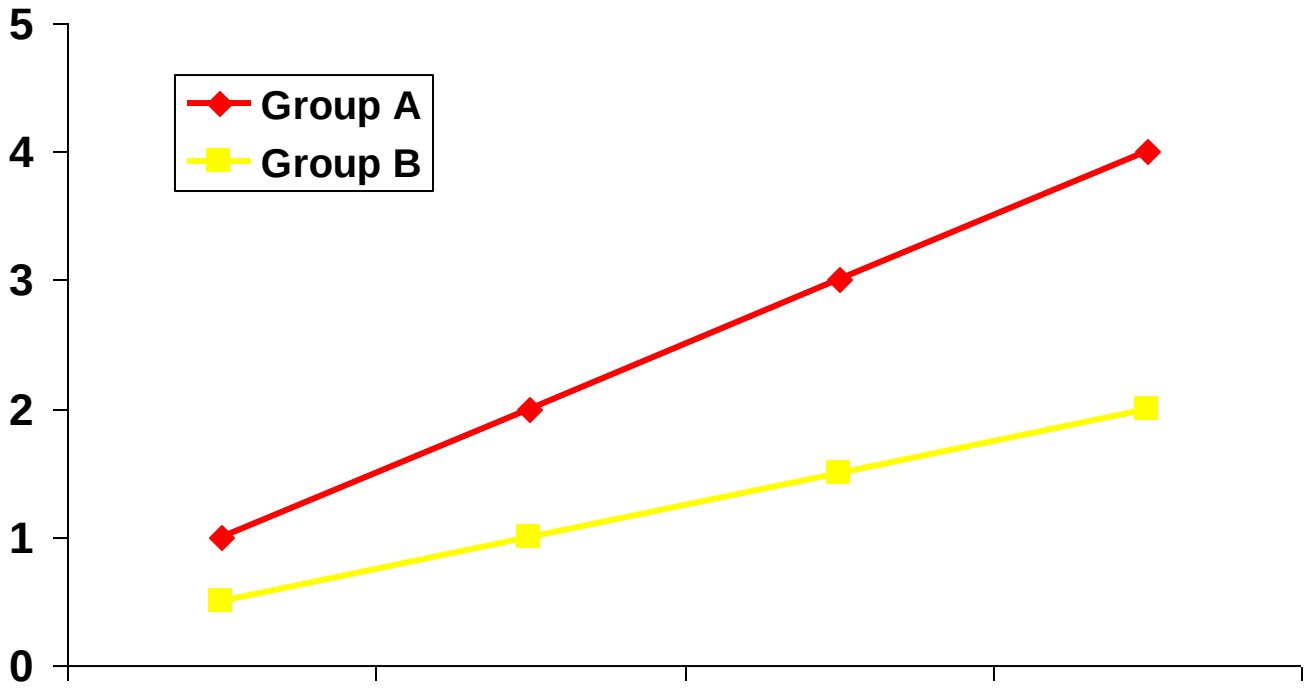
IR: 0.17/ptu

Relative Risk (RR)

$$RR = \frac{\text{Incidence of disease in the exposed } (I_e)}{\text{Incidence of disease among the unexposed } (I_0)}$$

- Attributable risk (Risk difference) = $I_e - I_0$
- Attributable proportion = $\frac{I_e - I_o}{I_e} = \frac{RR-1}{RR}$

RR vs Attributable Risk



Group A	1.0	2.0	3.0	4.0
Group B	0.5	1.0	1.5	2.0
RR	2.0	2.0	2.0	2.0
Att Risk	0.5	1.0	1.5	2.0

Epidemiology

- The study of the distribution and *determinants* of health and disease in populations

Study Design

"What is the question"

Options in Study Design

- Descriptive studies
  - Case report
  - Case series
  - Ecologic / Cross Sectional
- Analytic studies
  - Case-control study
  - Cohort study
  - Experimental study

Options in Study Design

- Descriptive studies
  - Case report
  - Case series
  - Ecologic / Cross Sectional

Case Report/Case Series

- Clinical description of a single patient or a small group of patients
- Advantages
  - Hypothesis generation
  - Diagnostic / therapeutic example
- Disadvantages
  - Lack of generalizability
  - No control group
    - Cant determine which factors are unique to patients

Case Report

The New England Journal of Medicine

ORIGINAL ARTICLE

BRIEF REPORT

Volume 345:1607-1610 November 29, 2001 Number 22

Index Case of Fatal Inhalational Anthrax Due to Bioterrorism in the United States

Larry M. Bush, M.D., Barry H. Abrams, M.D., Anne Beall, B.S., M.T., and Caroline C. Johnson, M.D.

Perelman School of Medicine
University of Pennsylvania

Cross Sectional Study

- Survey of a sample of the population in which the status of individuals with respect to exposure and/or disease is assessed at the same point in time.
- Advantages
  - Support for or against hypothesis
- Disadvantages
  - Do not capture concept of elapsed time
  - No information about transitions between health states

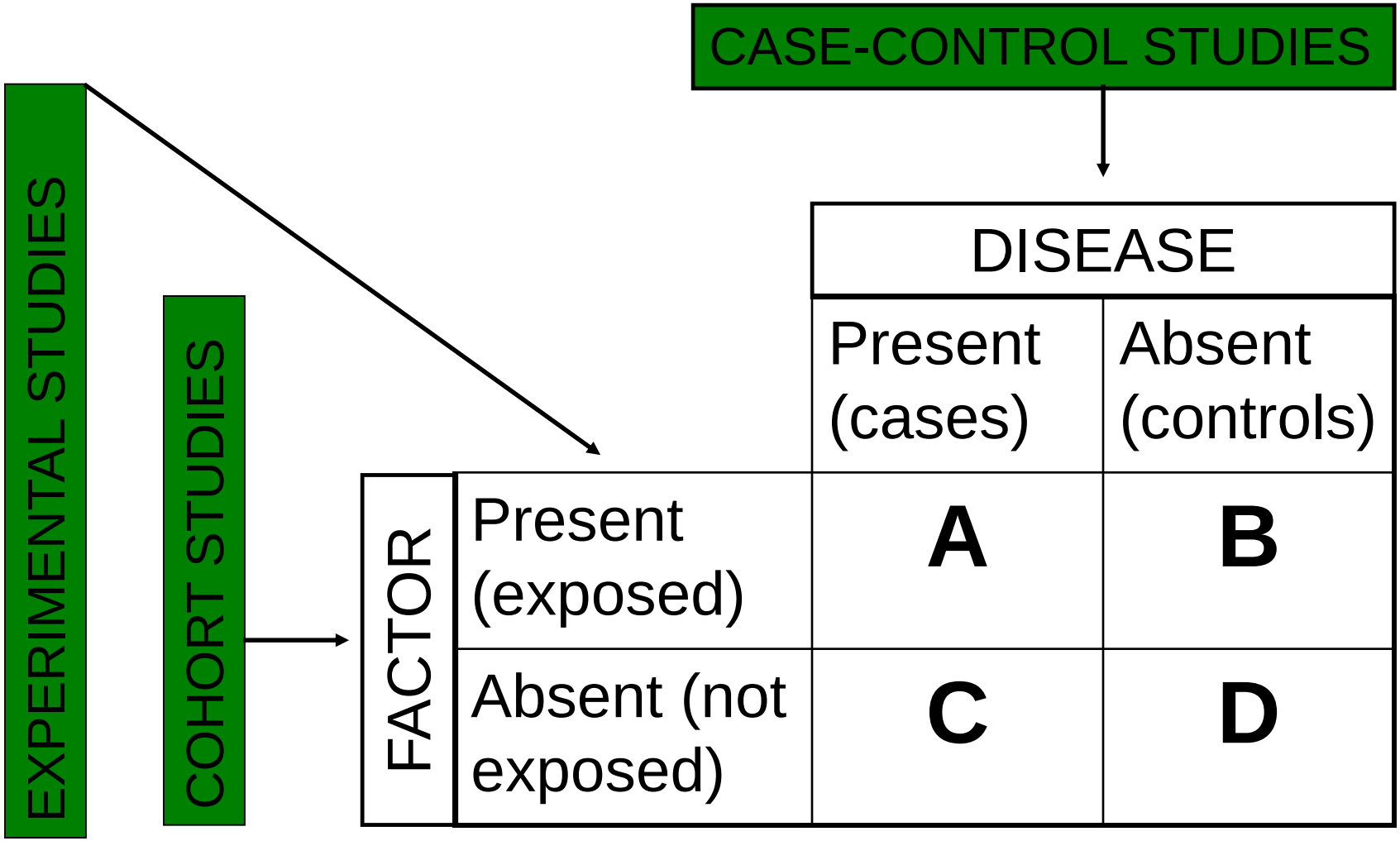
Ecologic Studies

- Compare geographic and/or time trends of an illness to trends in risk factors
  - Aggregate data (population based)
    - Birth / Death rates
- Advantages
  - Rapid/easy support for or against hypothesis
- Disadvantages
  - Cannot differentiate among those hypotheses consistent with the data
  - No patient level data

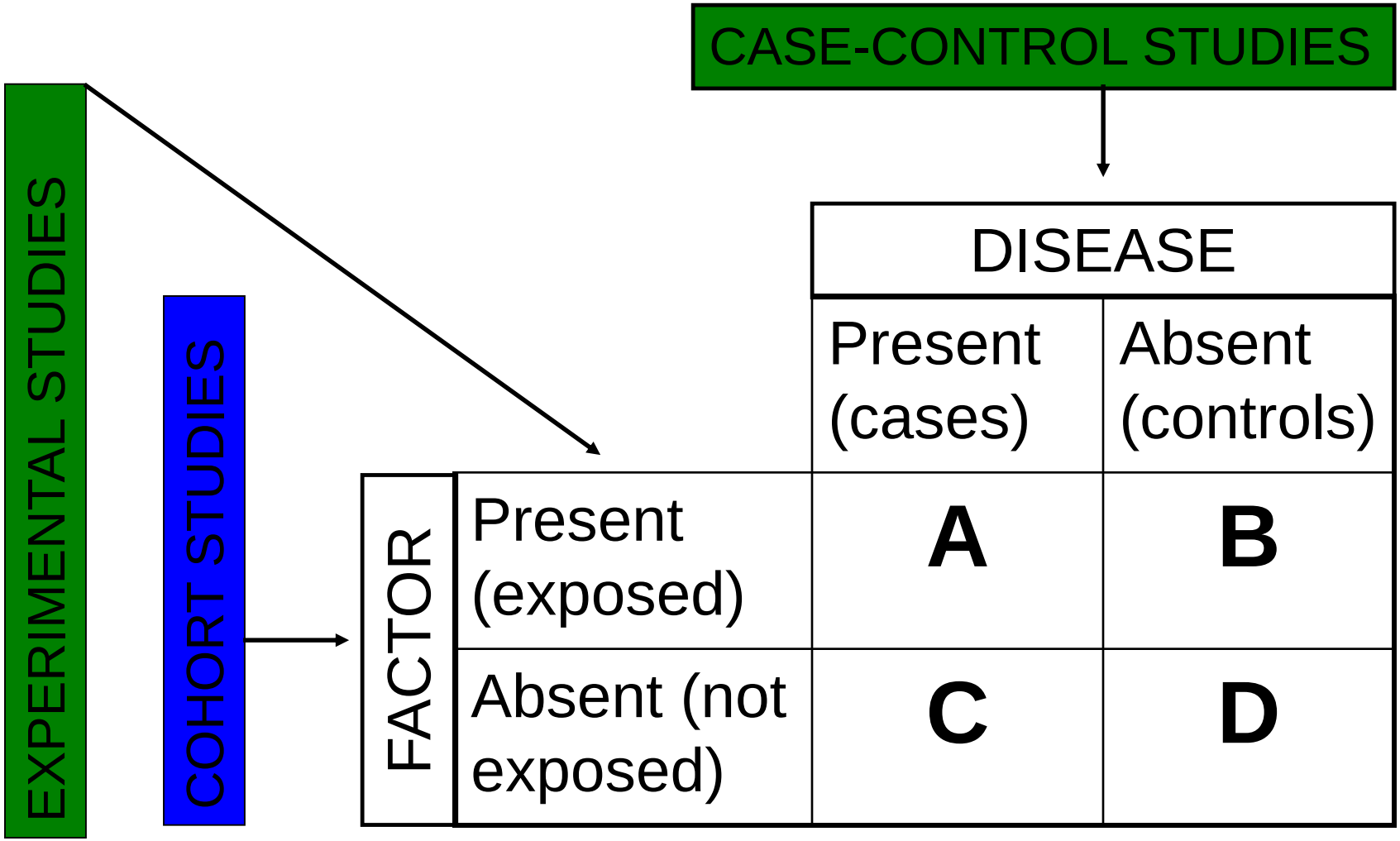
Options in Study Design

- Analytic studies
  - Case-control study
  - Cohort study (prospective/retrospective)
  - Experimental study
    - Randomized controlled trial
    - Quasi-Experimental Study
    - Cluster Randomized Trial

Study Design

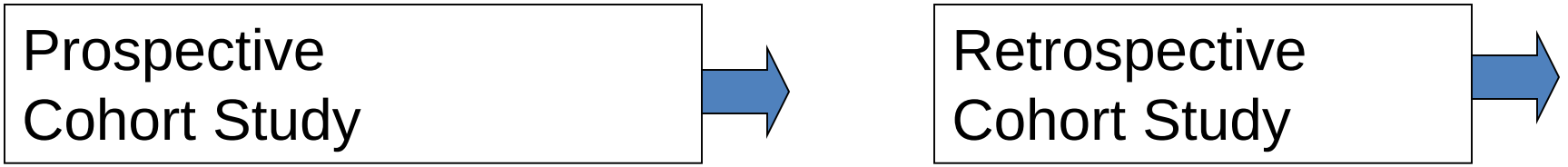


Study Design



Perelman School of Medicine
University of Pennsylvania

Prospective vs Retrospective



Exposure ⟶ Disease

Time

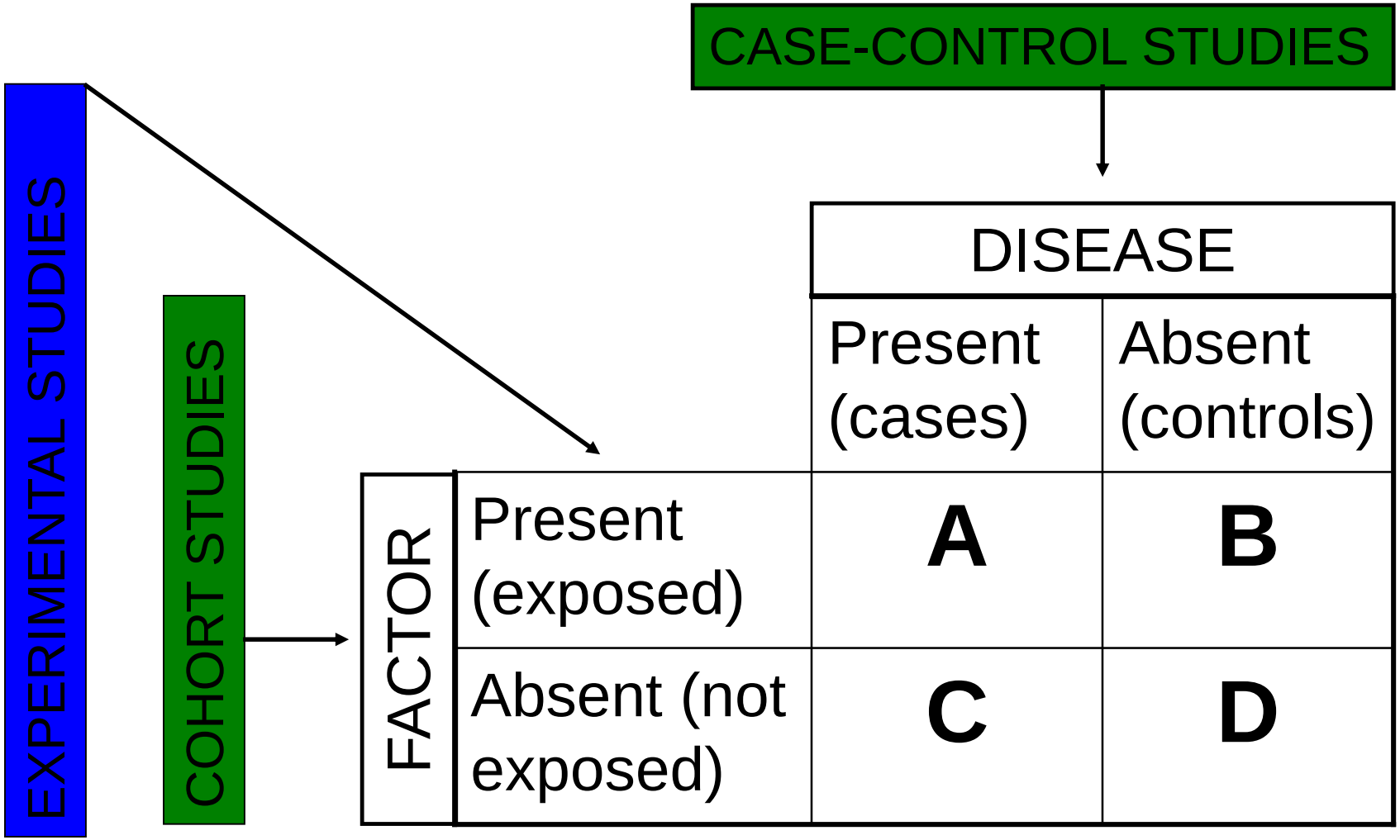
Cohort study

- A study comparing patients with a risk factor/exposure to others without the risk factor/exposure for differences in outcome
- Advantages
  - The study of any number of outcomes from a single risk factor/exposure
  - Incident rates available
    - Can calculate RR
  - Lack of bias in exposure data

Cohort study

- Disadvantages / Limitations
  - Potentially biased outcome data
  - Large sample size need for rare diseases
  - Long follow up needed
    - Subject to loss to follow up
    - Costly
    - Criteria and methods may change over time

Study Design



Experimental Study (RCT)

- A study in which the risk factor/exposure of interest is controlled by the investigator
  - Usually randomized
- Role
  - Most convincing demonstration of causality
  - Control of confounding
- Limitations
  - Logistic
  - Ethical

Quasi-Experimental Study

- (a.k.a.- non-randomized pre-post intervention design
- Evaluate intervention without using RCT
- The most basic type:
  - Collect baseline data
  - Implementation intervention
  - Collect same data as during baseline period
- Many different variations of quasi-experimental
  - 1) institution of multiple pretests
    - (i.e., collection of baseline data on more than one occasion)
  - 2) repeated interventions
    - (i.e., instituting and removing the intervention on sequentially);
  - 3) inclusion of a control group
    - (i.e., a group on which baseline and subsequent data is collected but on which no intervention is implemented).

Perelman School of Medicine
University of Pennsylvania

Harris AD, *Clin Infect Dis*, 2004;38:1586

Quasi-Experimental Study

- Advantages
  - Use when RCT not ethical
  - Use when intervention must be instituted rapidly (e.g., outbreak)
  - Use when RCT not logistically feasible
    - Broad interventions difficult to randomize to individual patients or hospital floors/units.
- Disadvantages
  - Difficult to control for potential confounding variables
    - e.g., patient severity of illness, quality of medical and nursing care
  - Regression to the mean
    - Use of a control group
  - Maturation effects
    - Seasonal variation

Harris AD, *Clin Infect Dis*, 2004;38:1586

Cluster Randomized Trials (I)

- Randomization by group
  - Hospital, practice site, unit
- Greater external validity
  - One intervention implemented per site
  - Broader patient/clinician eligibility
- More “real world”
  - Built into workflow of clinical care

Cluster Randomized Trials (II)

- Implementation easier
  - Clinicians/administrators
  - Fewer IRB issues (e.g., waiver of consent)
- Avoids issues of contamination
  - Particularly relevant for infectious diseases
- Statistical issues
  - Unit of analysis?

Challenges in Antibiotic Use / Antibiotic Resistance Research

- **<u>Competing Risks</u>**
  - Primary endpoint of interest is measure(s) of antibiotic use
  - Other important outcomes: repeat provider visit, emergency department visit, length of stay, mortality
  - Significant distortion issues due to competing risks when considered as outcomes separately
  - Outcomes must be interpreted in context of each other

Evans SR, *Clin Infect Dis* 2015;61:800-6

Challenges in Antibiotic Use / Antibiotic Resistance Research

- **<u>Issues with Non-Inferiority Designs</u>**
  - Doesn't address whether one approach is better
  - More susceptible to biases and manipulation
    - Lower scientific integrity
  - Implies preservation of previously demonstrate effect (i.e., vs placebo)
  - Effectiveness of the "control" may change over time
  - Acceptance of non-inferiority margin

Perelman School of Medicine
University of Pennsylvania

Evans SR, *Clin Infect Dis* 2015;61:800-6

Challenges in Antibiotic Use / Antibiotic Resistance Research

- **Individual vs Group Assessment**
  - Some patients experience benefit while some patients experience harm
    - Degree of overlap of these two groups often unclear
  - If little overlap: focus intervention on those who experience benefit but not harm
  - If great overlap: determine net effect (benefits vs risks)
  - Traditional analytic approaches treat these benefit and harm outcomes separately
  - Need novel approaches to evaluate net effect in individuals

Perelman School of Medicine
University of Pennsylvania

Evans SR, *Clin Infect Dis* 2015;61:800-6

Desirability of Outcome Ranking (DOOR)

- Ranking of trial participants by their overall outcome
- "Outcomes used to analyze patients rather than using patients to analyze outcomes"
- Define ordinal overall clinical outcome: Example
  - Clinical benefit (symptoms/function) without adverse effects (AEs)
  - Clinical benefit with some AEs
  - Survival without clinical benefit or AEs
  - Survival without clinical benefit but with AEs
  - Death
- Number of definition of categories tailored to disease
- Consensus regarding the definition is key

Perelman School of Medicine
University of Pennsylvania

Evans SR, *Clin Infect Dis* 2015;61:800-6

Response Adjusted for Duration of Antibiotic Risk (RADAR)

- Version of DOOR tailored for studies comparing antibiotic use strategies
- Subjects assigned a DOOR ranking using 2-step process
  - Better overall clinical outcome receives a higher rank
  - When two patients have the same overall clinical outcome, the patient with the shorter duration of antibiotic use receives a high rank
- Clinical outcome trumps duration of antibiotic use
- Adherence incorporated into the DOOR ranking
- Duration of antibiotic use most common measure
  - Others: broad vs narrow spectrum; oral vs IV

Perelman School of Medicine
University of Pennsylvania

Evans SR, *Clin Infect Dis* 2015;61:800-6

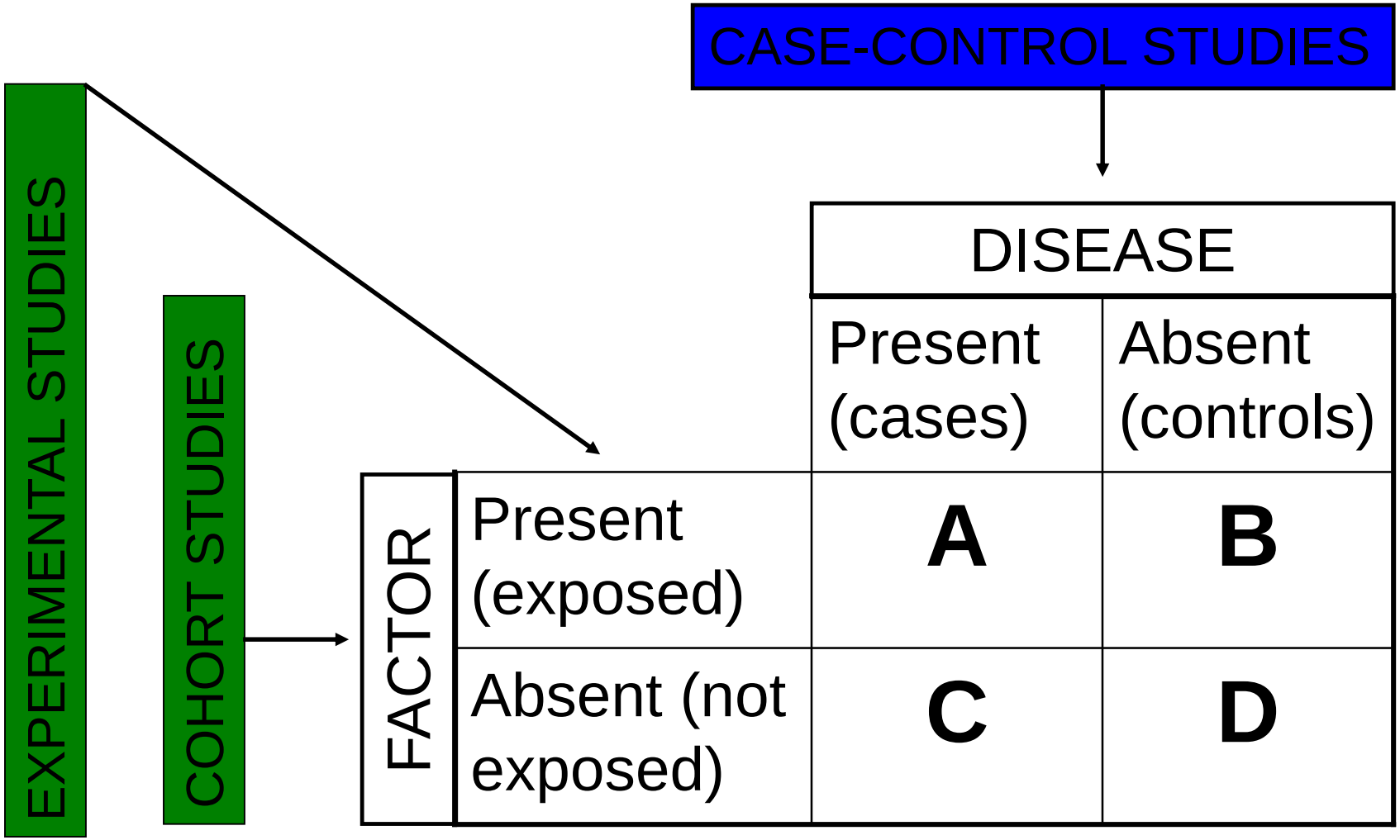
DOOR/RADAR Analysis

- Distributions of DOORs compared between strategies
  - Non-parametric testing – Wilcoxon Rank Sum test
- Sample size based on superiority testing
  - Null hypothesis: no difference in DOOR between groups
  - Alternative: new strategy has higher DOOR (i.e., >50%)
    - Magnitude of superiority based on minimum clinical importance
- Sample sizes lower than comparable non-inferiority studies

Perelman School of Medicine
University of Pennsylvania

Evans SR, *Clin Infect Dis* 2015;61:800-6

Study Design



Perelman School of Medicine
University of Pennsylvania

Case-Control Studies

- A study comparing patients with an outcome to others without the outcome for differences in risk factors/exposures
- Advantages
  - Study of any number of risk factors for a single outcome
  - Can study a rare event
  - Less costly and time-consuming than a cohort study

Selection of Cases

- May be restricted to any group of diseased individuals
- Arise from a theoretical source population
  - A diseased person not selected (or eligible) as a case is presumed to have arisen from a different source population
- Must be chosen independently of exposure

Selection of Controls

- Controls should be representative of the theoretical source population that gave rise to the cases
- Must be chosen independently of exposure
- Controls are NOT selected because they have characteristics similar to cases
  - McMahon et al, NEJM, 1981
    - "coffee consumption and pancreatic cancer"

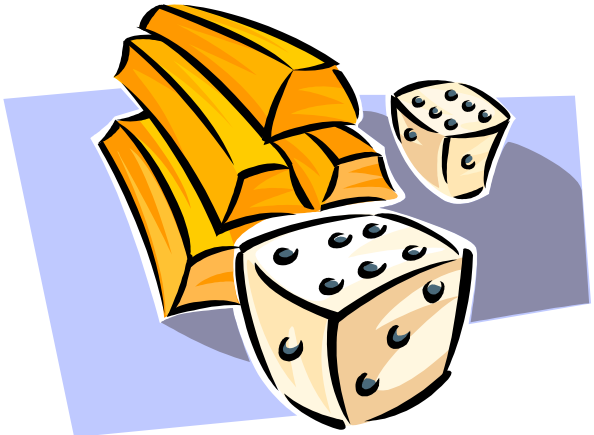
Case-Control Studies

- Disadvantages
  - Can study only one outcome
  - Information bias (multiple types)
  - Selection bias
  - Can’t calculate incidence / RR

Risk vs Odds

- Risk: ratio of a part to the whole
- Odds: ratio of a part to the remainder

- Rolling dice
  - Risk of rolling a 6: 1/6 = 16.7%
  - Odds of rolling a 6: 1/5 = 20.0%



- Odds always higher than risk

RR vs OR (Cohort Study)

		DISEASE	
		Present (cases)	Absent (controls)
FACTOR	Present (exposed)	**A**	**B**
	Absent (not exposed)	**C**	**D**

Risk of disease among the exposed = A / (A+B)

Risk of disease among the unexposed = C/ (C+D)

$$\text{Relative Risk (RR)} = \frac{A / (A+B)}{C/ (C+D)}$$

Perelman School of Medicine
University of Pennsylvania

RR vs OR (Case-Control)

Odds = Risk / (1-Risk)

		DISEASE	
		Present (cases)	Absent (controls)
FACTOR	Present (exposed)	**A**	**B**
	Absent (not exposed)	**C**	**D**

Odds of exposure given disease = A / C

Odds of exposure given no disease = B / D

$$\text{Disease Odd Ratio} = \frac{AD}{BC}$$

Perelman School of Medicine
University of Pennsylvania

RR vs OR

		DISEASE	
		Present (cases)	Absent (controls)
FACTOR	Present (exposed)	**A**	**B**
	Absent (not exposed)	**C**	**D**

When disease is rare, B>>A, and D>>>C

$$\text{Relative Risk (RR)} = \frac{A / (\cancel{A}+B)}{C/ (\cancel{C}+D)} \sim \frac{AD}{BC} = \text{Odds ratio (OR)}$$

Perelman School of Medicine
University of Pennsylvania

Bias

- Definition: systematic error in collecting or interpreting data
- Particularly likely to occur if there is uncertainty about the question being asked
- Potential for bias must be addressed in the design of the study

Bias

- Selection bias
  - Distortion in the estimate of effect resulting form the manner in which subjects are selected for the study
  - Case Control
    - Non response (refusals, too sick, not at home, moved away, can't speak English)
  - Cohort
    - Non participation; loss to follow up
  - Impact of selection bias?

Bias

- Information bias
  - Distortion in the estimate of effect due to measurement error or misclassification of subject on one or more variables.
  - Case control
    - Memory, communication, knowledge, motivation, social desirability, threatening/personal questions
  - Cohort
    - Ascertainment of disease more vigorously pursued in one group than in another
  - Differential or non-differential

Bias

- Potential for bias does not mean that there actually is bias
- Existence of bias does not mean that the bias is severe enough to cause concern

Study Effect	Direction of Bias	Implication
Yes	Toward Null	Real effect even stronger
No	Toward Null	Might have missed real effect
Yes	Away from Null	Spurious conclusion
No	Away from Null	Really nothing going on

How To Control Bias

- Careful study design
- Can't adjust for it in analysis
- Blinding
  - Bias may occur if everyone knows which treatment the patient is receiving
    - Patient: psychological benefit from knowing he/she is on new treatment
    - Treatment team: closer observation, more ancillary care
    - Evaluator: may record more favorable result
    - Statistician?

Confounding

- Estimate of the effect of the exposure of interest is distorted because it is mixed with the effect of an extraneous factor
- Confounder: associated with both the exposure and the outcome
  - Not a consequence of the exposure

How to Address Confounding

- Gather accurate measurements of potential confounding variables
  - Stratified analysis
  - Multivariable analysis
- Randomization
  - Should make groups the same with regard to *known* and *unknown* confounders

Confounding by Indication

- Major concern in non-randomized stewardship studies
  - Why do patients receive different treatments/strategies?
    - Measured and unmeasured factors
  - Approaches
    - Multivariable modeling
    - Propensity score analysis
    - Instrumental variables

Multivariable Modeling

- Ascertainment of known potential confounders
- Inclusion of confounders in multivariable model
- Independent effect of the exposure/treatment
- Good when you have a large number of outcomes

Propensity Score Analysis

- Develop statistical model to predict receipt of treatment
- Patients then stratified by propensity score
- Treatment effect estimated within each stratum and averaged across strata
- Can see how propensity score distributed across groups
  - Often limited data at extremes
- Good when small number of outcomes

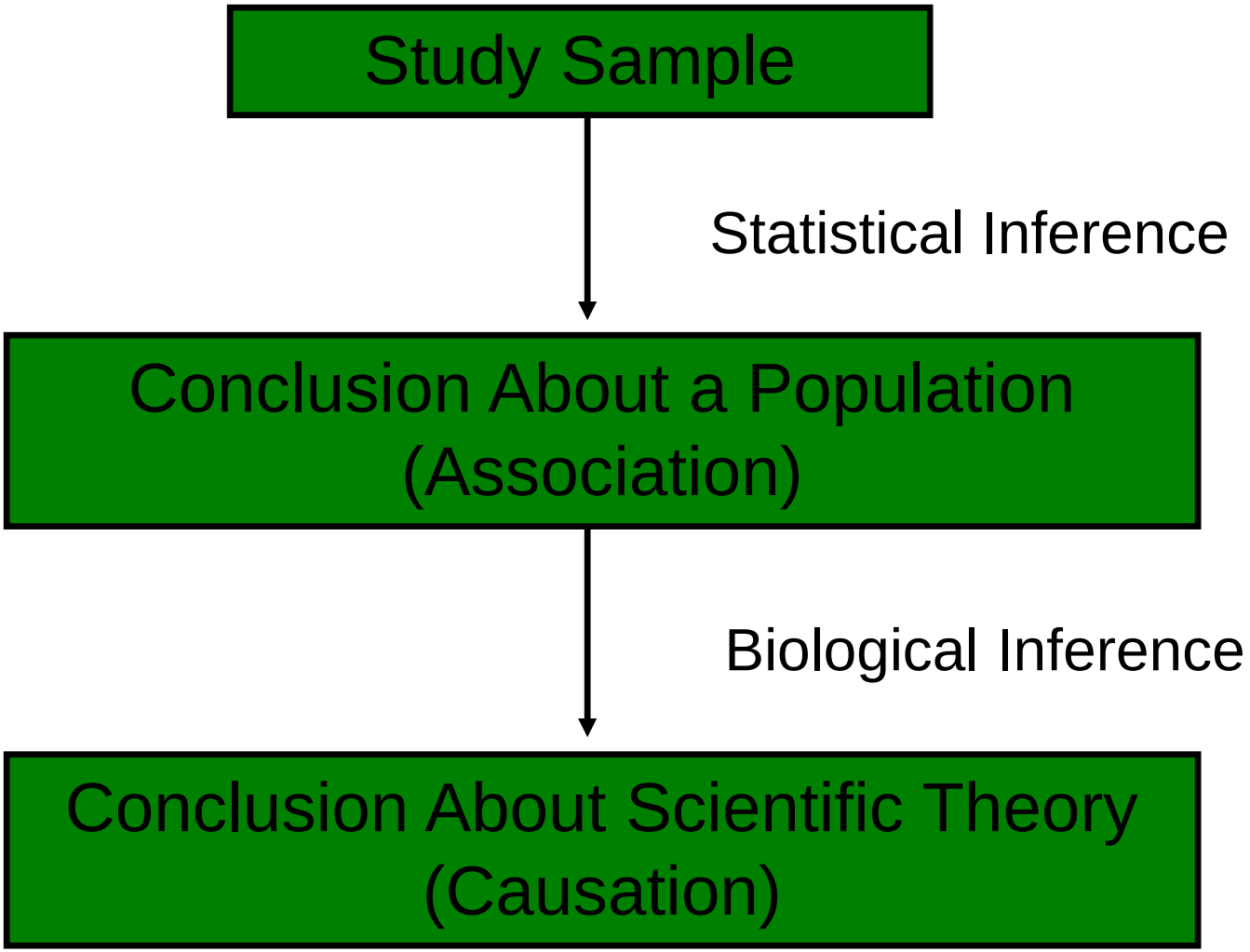
Instrumental Variables

- External cause of the intervention but is by itself unrelated to the outcome
  - "Natural randomization"
  - Policy change, geographic differences
- Likelihood of intervention a proportion (not yes/no)
- Can help account for measured and unmeasured confounding
- Not always available

Significance

- P value
  - Likelihood that results occurred by chance
  - Reflects both sample size & magnitude of the difference between the groups
- OR/RR (95%CI)
  - Range within which the true magnitude of the effect lies with certain degree of assurance
  - Statistical significance
  - Variability (sample size)
    - Particularly useful in negative studies

Scientific Method



Causality

- Strength
  - Study design
  - Quantitative strength
  - Dose-response relationship
- Coherence with existing information
- Time sequence
- Specificity
- Consistency

* none is necessary or sufficient