Confounding and Case/Control Studies

Anthony Harris, MD, MPH

Department of Epidemiology and Preventive Medicine

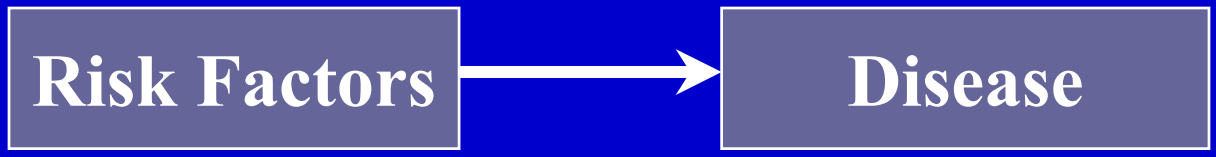
August 10, 2005

Learning objectives

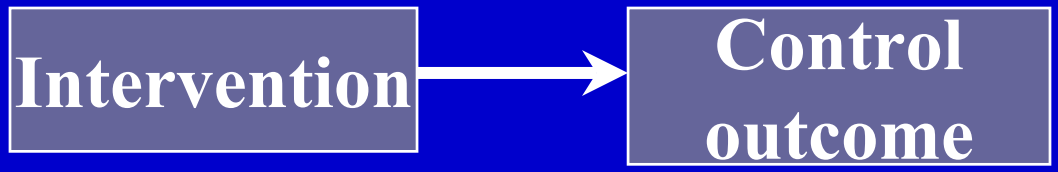
- Be able to define confounding
- Understand what effect it has on the association between exposure and outcome
- Understand ways to correct or adjust for confounding
- Differentiate bias from confounding

Why is confounding important?

It interferes with our search for causality



Disease Etiology



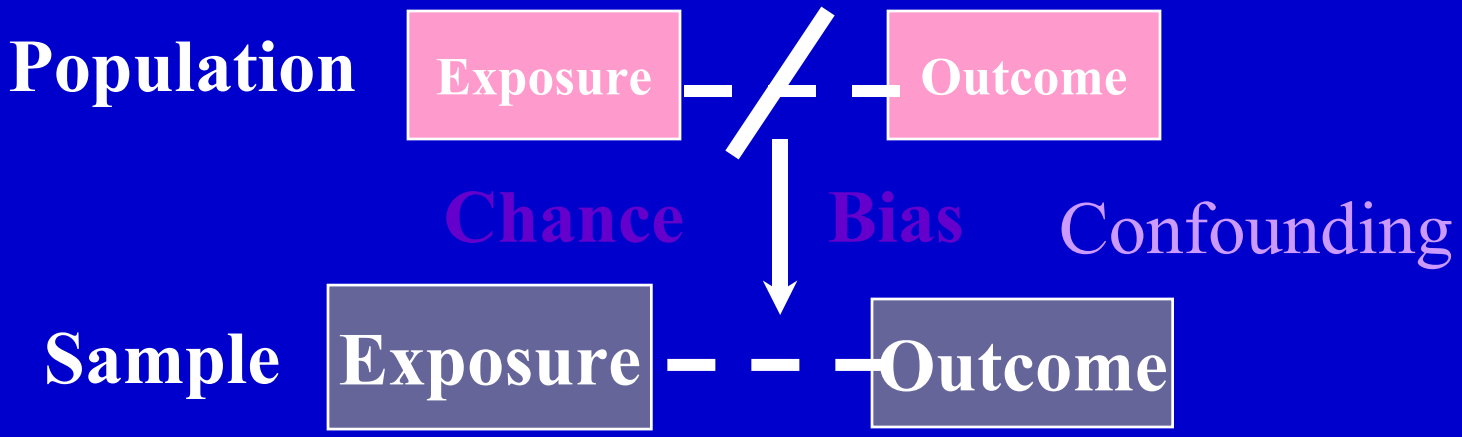
Treatments/Prevention

It complicates interpretation of study results

Study leads to associations

- We say that the study factor is associated with the study outcome when
  - In cohort study, RR not equal to 1
  - In case control study, OR not equal to 1

Spurious Associations



Association in Sample but not in Population

Hallmark: Lack of Repeatability

Spurious associations

- On previous slide, bias, chance, confounding in a study design,which is uncorrectable statistically, can lead to a statistically significant association but not a real or causal association

What is the difference between bias and confounding?

- One important difference is that there are statistical ways to adjust for confounding once a study is done but there are not statistical ways to control for bias

- One important similarity is that both bias and confounding can lead to a statistical association that is not a causal association

Bias

- Bias is a systematic error in the design or conduct of a study
- Bias is any systematic error in the design, conduct, or analysis of a study that results in a mistaken estimate of an exposure's effect on the risk of disease or the risk of outcome

Bias

- Systematic error results from flaws either in the method of selection of study participants or in the procedures for gathering relevant exposure and/or disease information

Bias

Bias = Systematic error = Lack of validity

Some Types of Bias

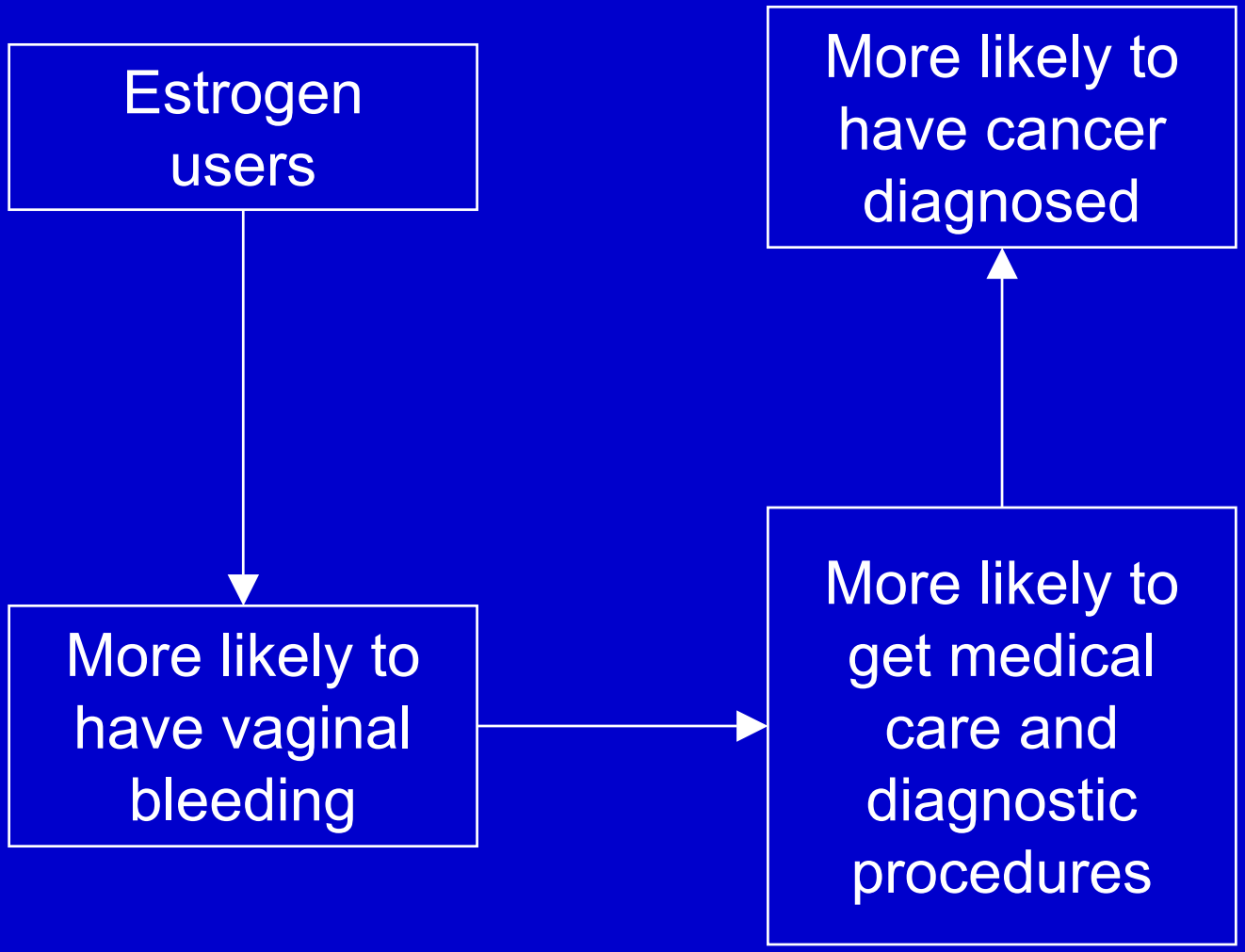
- Selection bias
  - self-selection, e.g., volunteers
  - selective losses or nonresponse
- Information bias
  - differential misclassification

Selection Bias

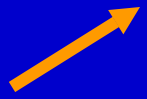
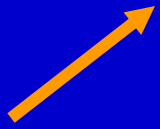
Example:

- Several case-control studies found a strong association between estrogen use and endometrial cancer
- Validity of findings was challenged on the basis of detection bias
- Selective surveillance of estrogen users (people on estrogen were more likely to bleed and thus be investigated)

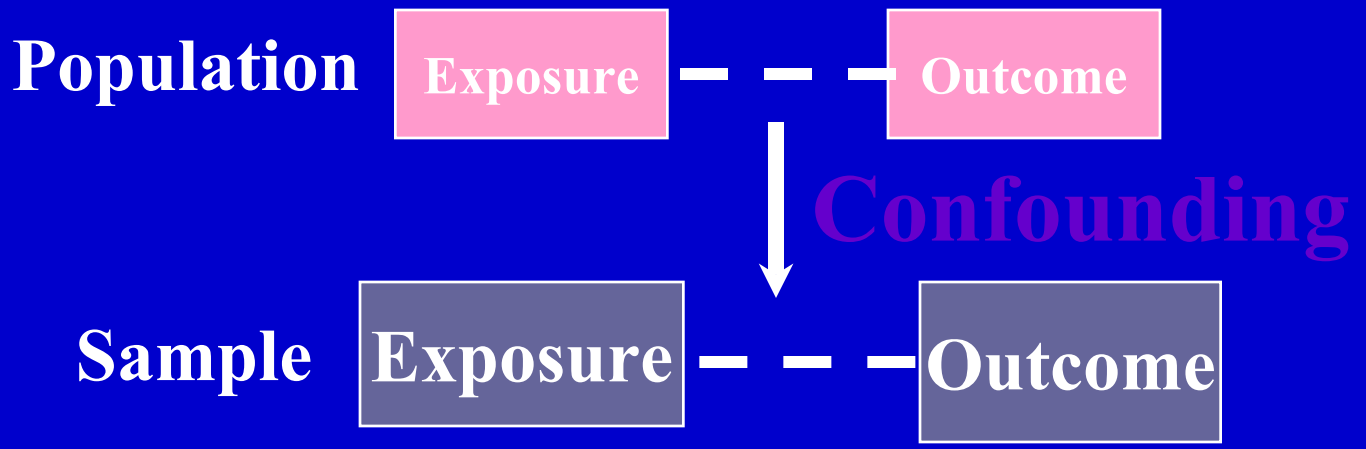
Selection Bias



2 by 2 table

	Cases of renal disease	Controls without renal disease
Used an NSAID		
Did not use an NSAID		

Confounding



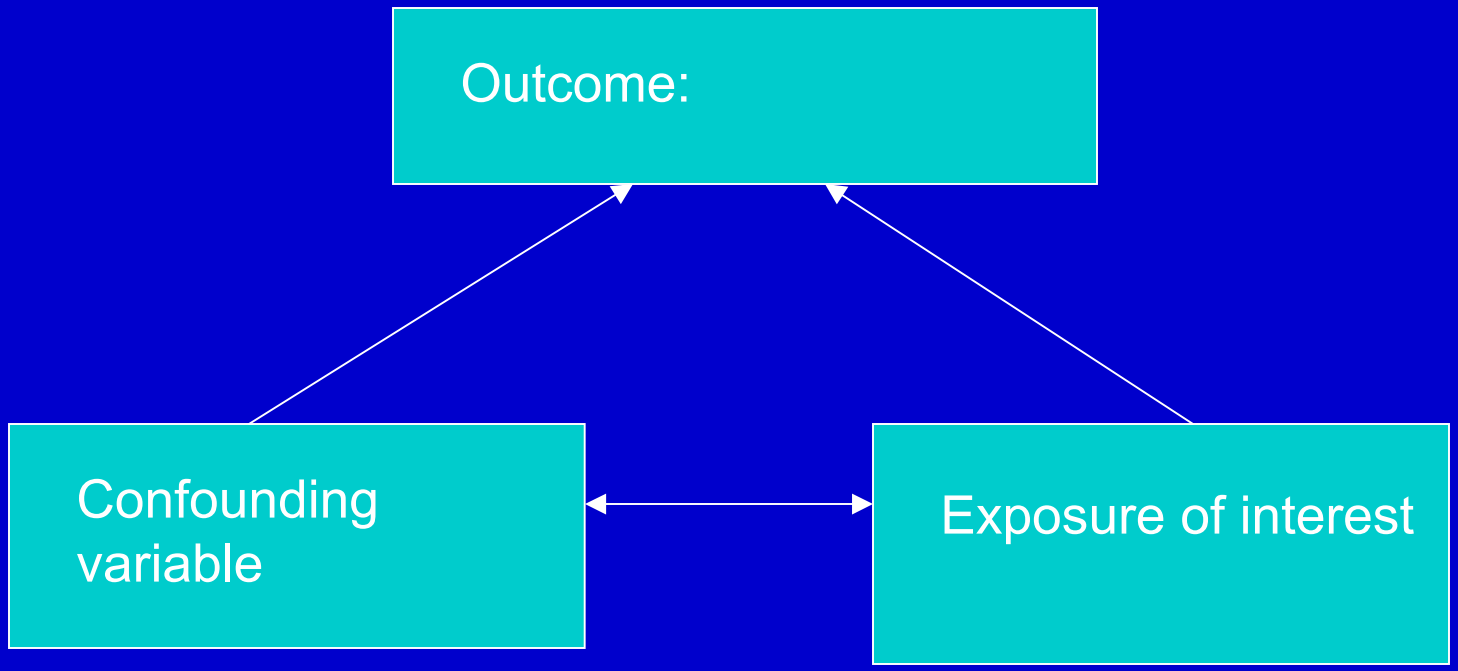
Association in Sample and in Population

"Guilt" by Association

What is a Confounder?

- A variable entangled with the study factor that masks the true relationship between the study factor and the outcome

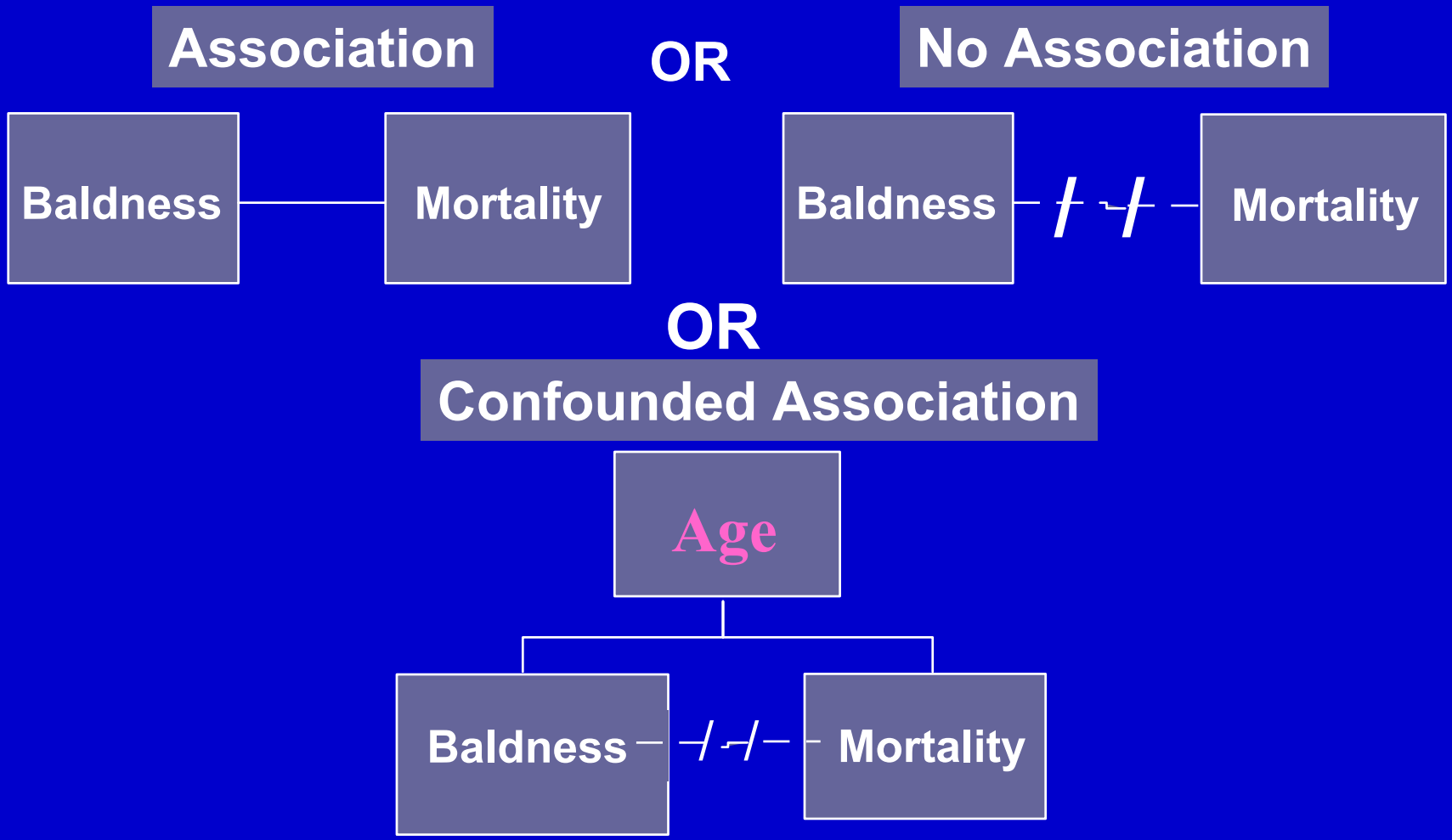
Confounding



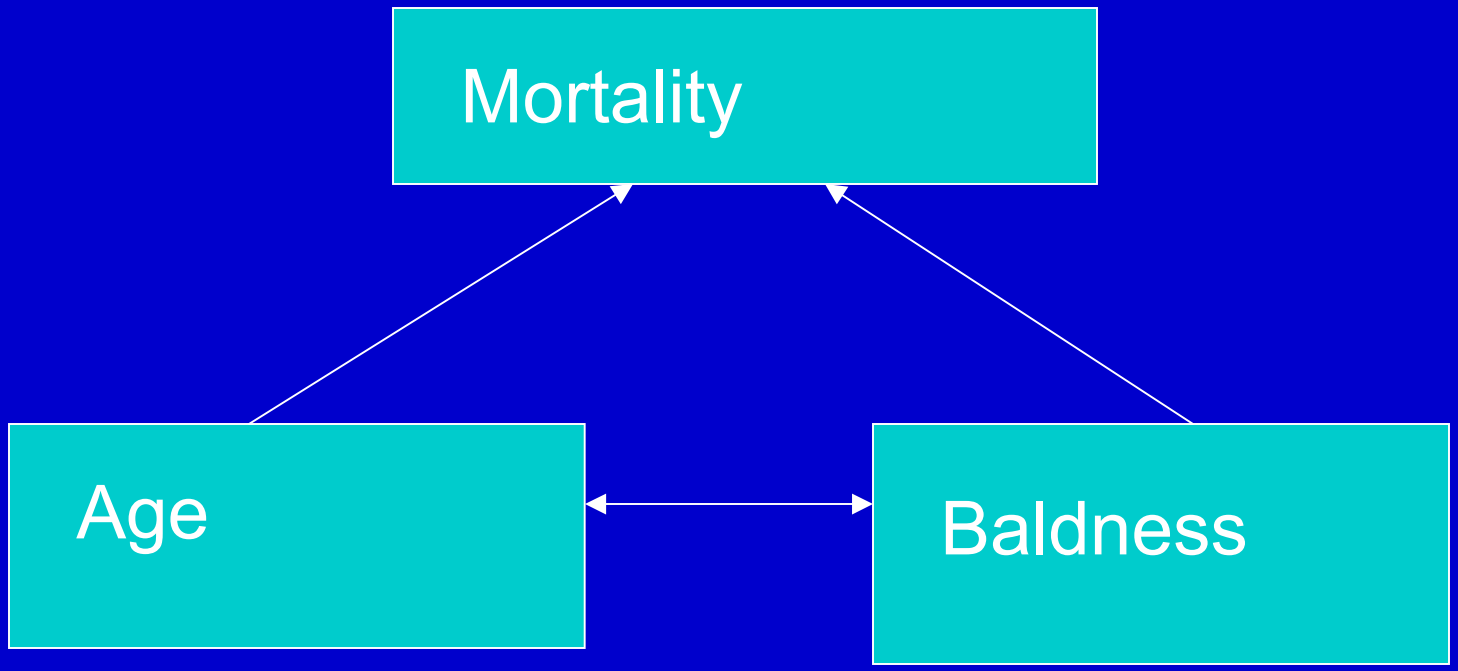
To get true effect of exposure of interest, we need to control for confounding variable

Examples of confounding

Is male baldness related to mortality?



Confounding



To get true effect of exposure of interest, we need to control for confounding variable

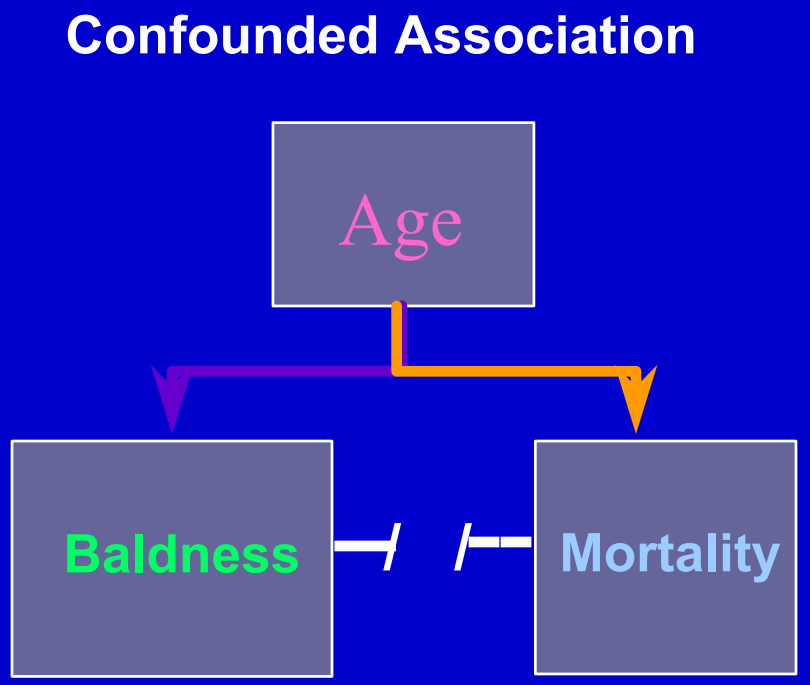
Criteria for Confounding

- Confounder must differ by levels of exposure variable

  Bald are older than non-bald

- Confounder associated with study Outcome

  Old are at greater risk of death than young

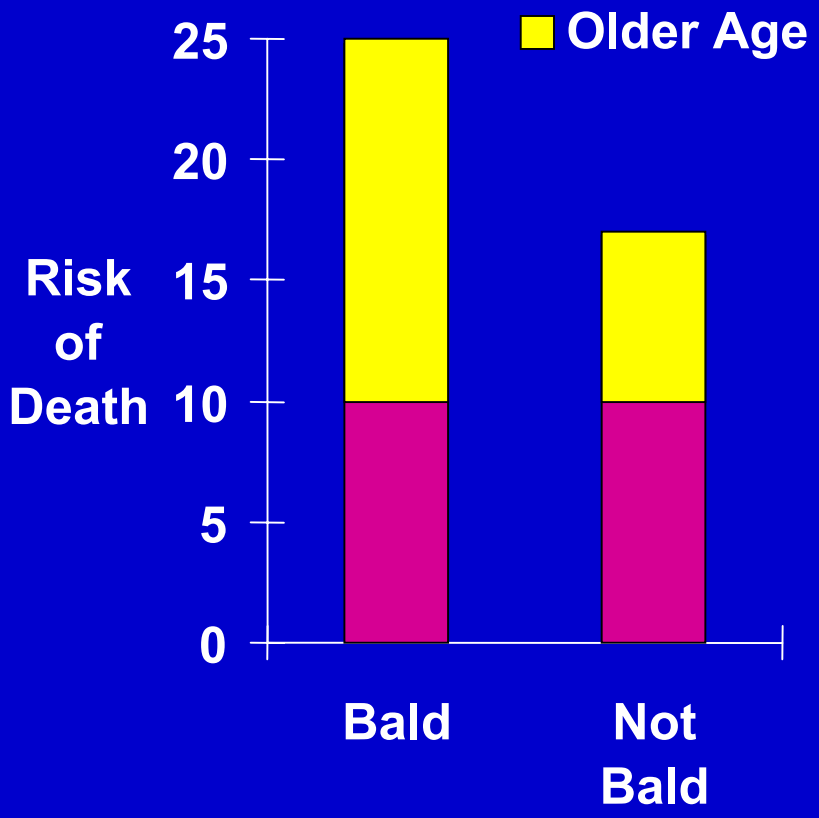


Because Bald are Older and Mortality Increases with age, the Bald appear to have greater risk of Mortality

What is the effect of confounding?

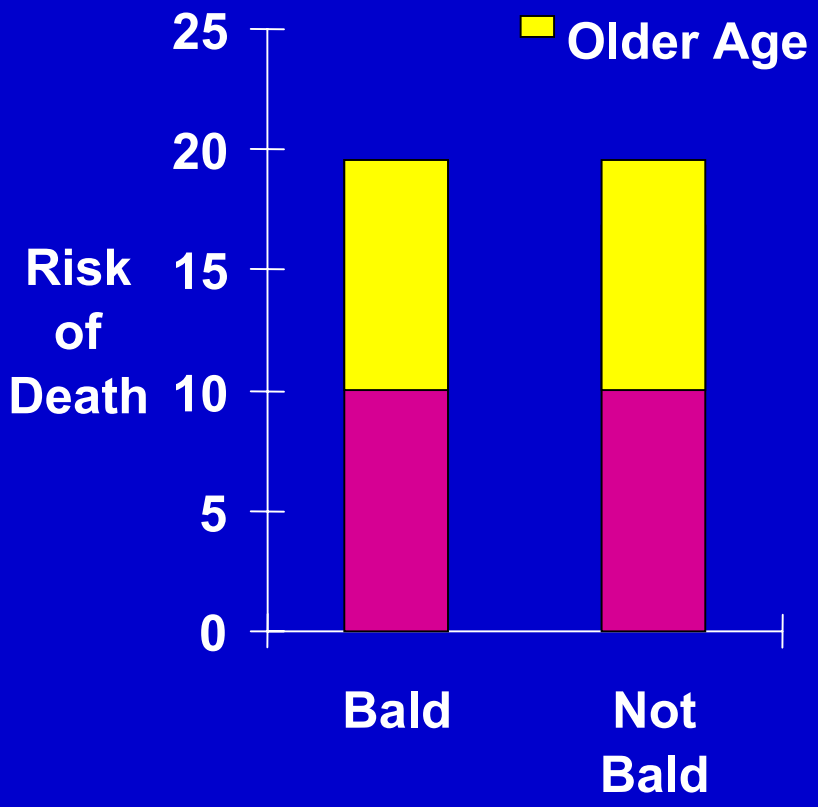
- Leads to wrong conclusions about the association of interest
  - Ignoring confounding, it appears that baldness is an independent risk factor for mortality
  - Recognizing that confounding is present, bald men only appear to be at increased risk because they are older

How does confounding work?



What can be done about it?

If we could equalize age, we could reduce/eliminate confounding



How can we adjust for confounding?

- Study design before the study starts
- Statistical analysis

How can we adjust for confounding?

- During Study Design
  - restrict (limit study to a narrow range e.g., age 45-54)
  - match (force balance) (used most often in case-control studies)
  - randomly allocate persons to study groups (RCT)
- After data are collected
  - stratify (e.g., compare old w/ old; young w/ young)
  - adjust (combine stratified results into 1 estimate)
    - answers the question: What would the association be if the groups did not differ by the confounder?

Stratify on levels of Confounder to avoid Confounding

- Compare association in Old to Old, Young to Young
  - RR_{old} =Death Rate$(Bald)_{old}$ to Death Rate $(NonBald)_{old}$
  - RR_{young} = Death Rate$(Bald)_{young}$ to Death Rate$(NonBald)_{young}$
- If $RR_{old} \sim= RR_{young}$
  - We can derive a summary or age-adjusted RR for entire Population
- Confounding exists if $RR_{old} = RR_{young}$ but not the same as the overall RR

Adjustment Methods

- Multivariate Regression Adjustment
  - statistically model relationships of exposure & confounders to disease, fix confounders

Case-control studies

Objectives

- Be able to recognize a case-control study
- Know the advantages and disadvantages of a case-control study versus other study designs
- Know the summary questions that allows the reader to assess the quality of a case-control study

Case-control studies

- Subjects are selected on the basis of whether they do (cases) or do not (controls) have a particular disease or outcome of interest.
- Cases and controls are then compared for their exposure or risk factor history

Case-Control Design

- Design in Words
  - Select representative cases (with study outcome)
  - Select comparable controls (without study outcome)
  - Look historically in both groups for exposure
  - Compare odds of exposure in cases & controls

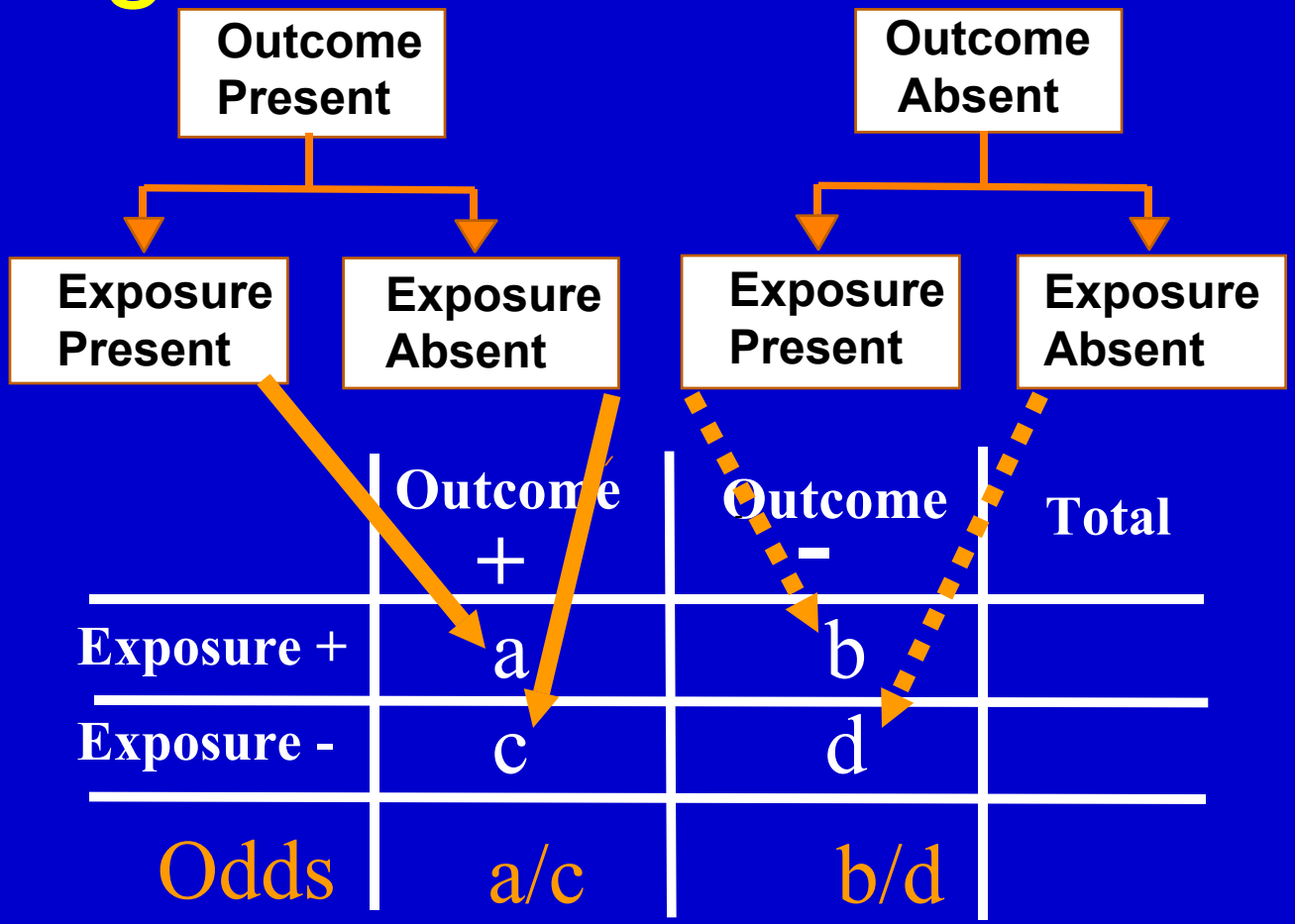
Strengths of case-control study

- Good for outcomes with long latent periods
- Inexpensive
- Optimal for evaluation of rare diseases or rare outcome
- Can examine multiple risk factors for a single disease or outcome

Weaknesses of case-control study

- Particularly prone to biases and confounding, I.e. statistical association but not “true” cause and effect
- Inefficient for the evaluation of rare exposures

Analysis of C/C Idealized Design



Example

- You are the chief of medicine at a prominent tertiary care hospital. You are interested in studying the effects of non-steroidal anti-inflammatory drugs (e.g. advil, motrin etc…) on renal failure using a case-control study.

Example

- You get one of your junior faculty to enroll 100 cases who present to your dialysis clinic with acute renal failure. You choose 100 random controls. You then determine that 30 of your cases had history of NSAID exposure and that 10 of your controls have history of NSAID exposure. What is the odds ratio?

2 by 2 table

	Cases of renal disease	Controls without renal disease
Used an NSAID	30	10
Did not use an NSAID	70	90

Odds ratio=(30/70)/(10/90)=3.85

What does this odds ratio mean in words?

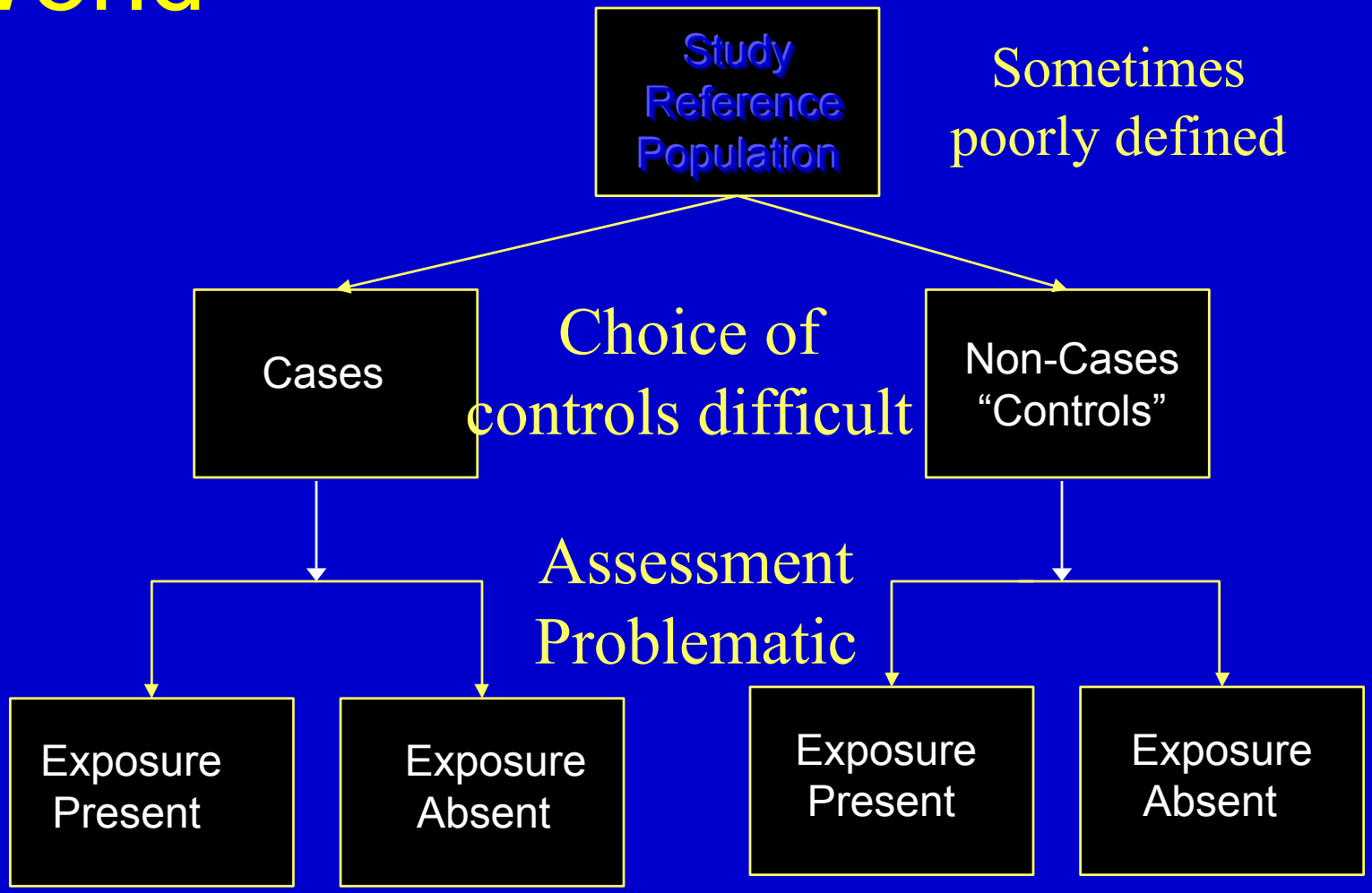
- Odds ratio of 3.85 means that the ratio of NSAID use among people who have renal failure compared to controls who do not have renal failure is 3.85
- If incidence of renal failure among NSAID users is rare (a fair assumption), it also means that NSAID users are 3.85 times more likely to develop renal failure than non-NSAID users

A useful property of OR

- Odds ratio approximately equals the relative risk (RR) when the disease being studied is rare
  - Rare disease assumption
  - If the disease is rare, OR ~= RR
    - rare means < 5%

Ways to avoid problems in case-control study designs

Study Design in the Real World



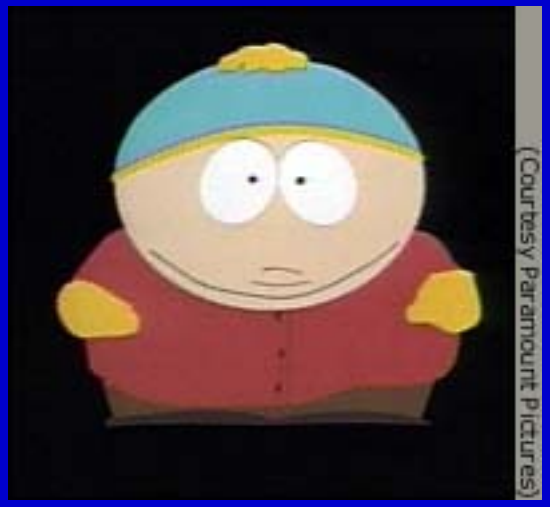
Choose controls carefully

Case-control studies

- Done when outcomes are rare events.
- Ideally if money was no issue would be cohort studies.
- Hence, case control should be viewed as an efficient cohort study.
- Controls should come from the base population of interest or the cohort of interest.

Case-control studies

- Controls should come from the members of the underlying cohort or source population for the cases during the time periods when they are eligible to become cases.


(Courtesy Paramount Pictures)

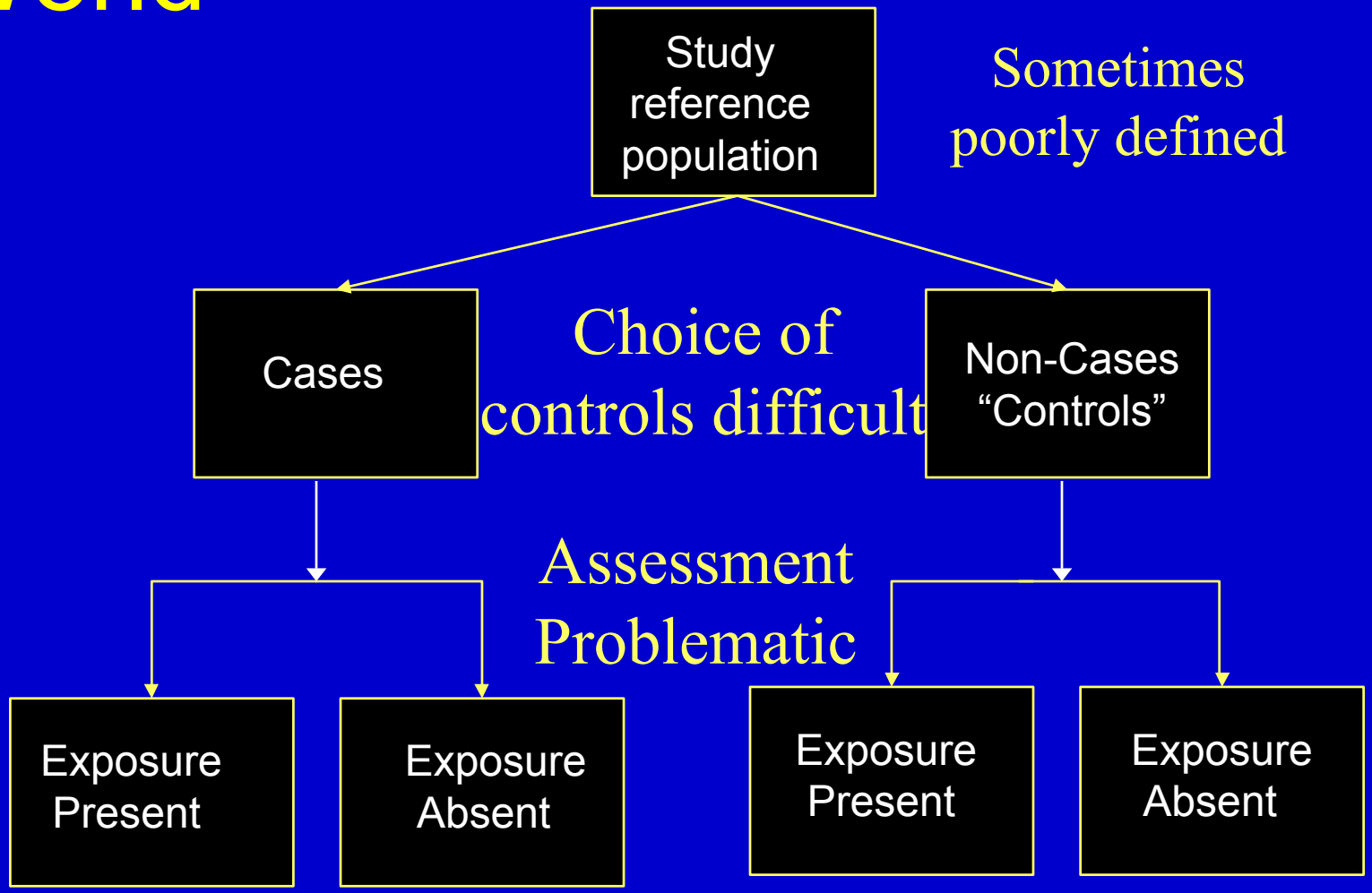
Examples of control group selection

- There is an outbreak of E. coli O157 in South Park. Cartman has renal failure and hemolytic anemia. Kenny has died. Other kids are affected. You are performing a case-control study, who should be my controls.

Case-control studies

- Important to control for confounding variables.

Study Design in the Real World



Points of Vulnerability (cont.)

- Ascertainment of exposure accurately
  - must try to reconstruct info from years ago
    - remedy-use concurrent records where possible
- Ascertainment of exposure without bias
  - subjects aware of disease status-> recall bias
    - cases more likely to “soul search”
    - interviewers may also probe harder in cases
    - remedy-assess exposure before final diagnosis, mask interviewers to outcome status

Points of vulnerability (cont.)

- Accuracy of outcome diagnosis
  - diagnosis usually made before study begins
  - inconsistent due to lack of standardization
- Bias in outcome diagnosis
  - if diagnosticians are more likely to make the diagnosis in exposed persons
- ~Remedy- impose a standard protocol for outcome diagnosis

Summary questions in assessing case-control studies?

- 1) Was there a pre-specified hypothesis defining a relationship between an exposure and an outcome?

- 2) Were the exposure and health outcome clearly and operationally defined?

- 3) Was the control group appropriate?

- 4) Was the measurement of exposure both in the cases and controls accurate and unbiased?

- 5) Was the measurement of the outcome both in the cases and controls accurate and unbiased?

- 6) Were the important confounding variables accounted for and controlled for in the statistical analysis?