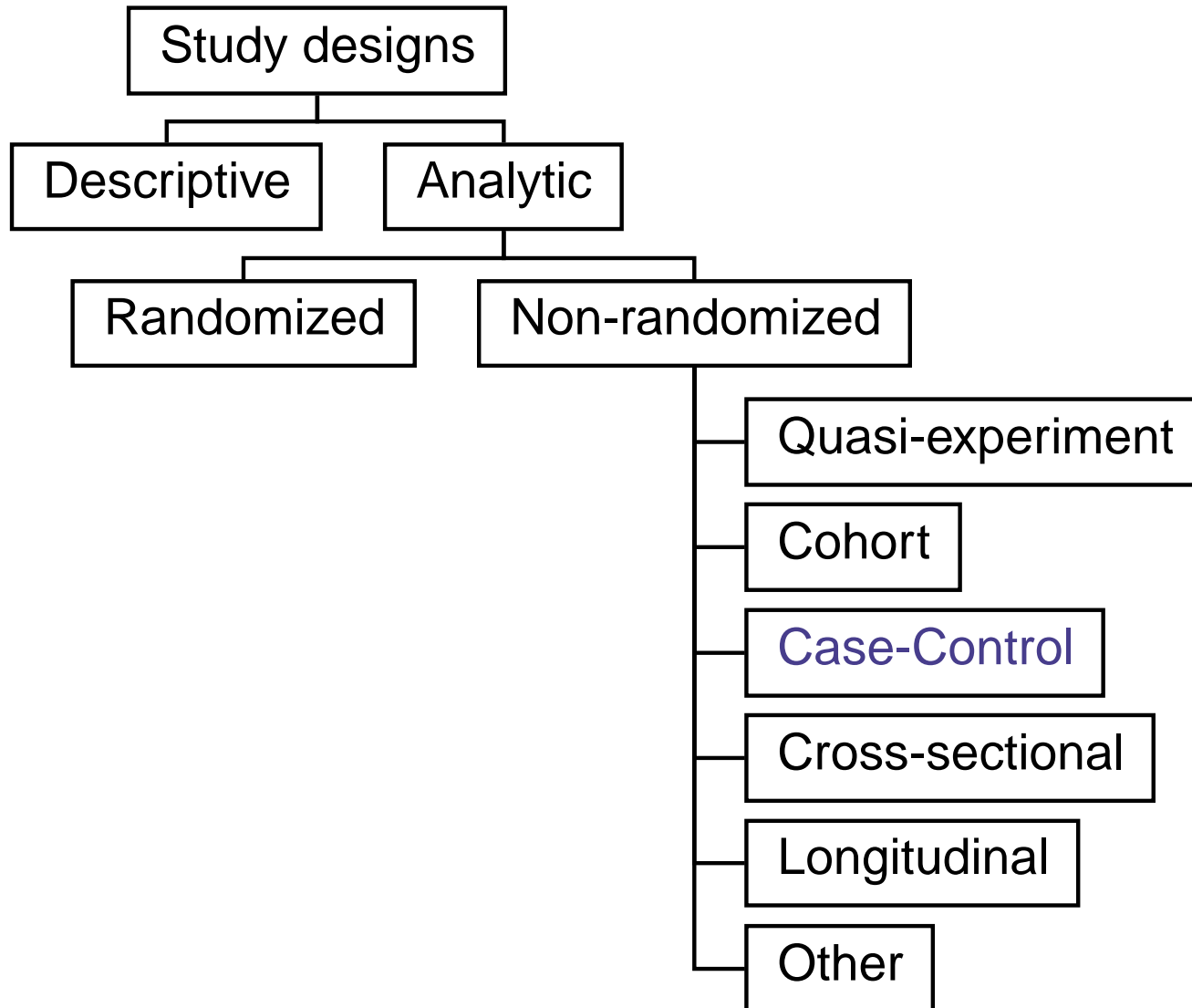# *Case-Control Studies*

Research Methods Mini-Course
July, 2005

Overview of study designs

Data layout common to many analytic studies
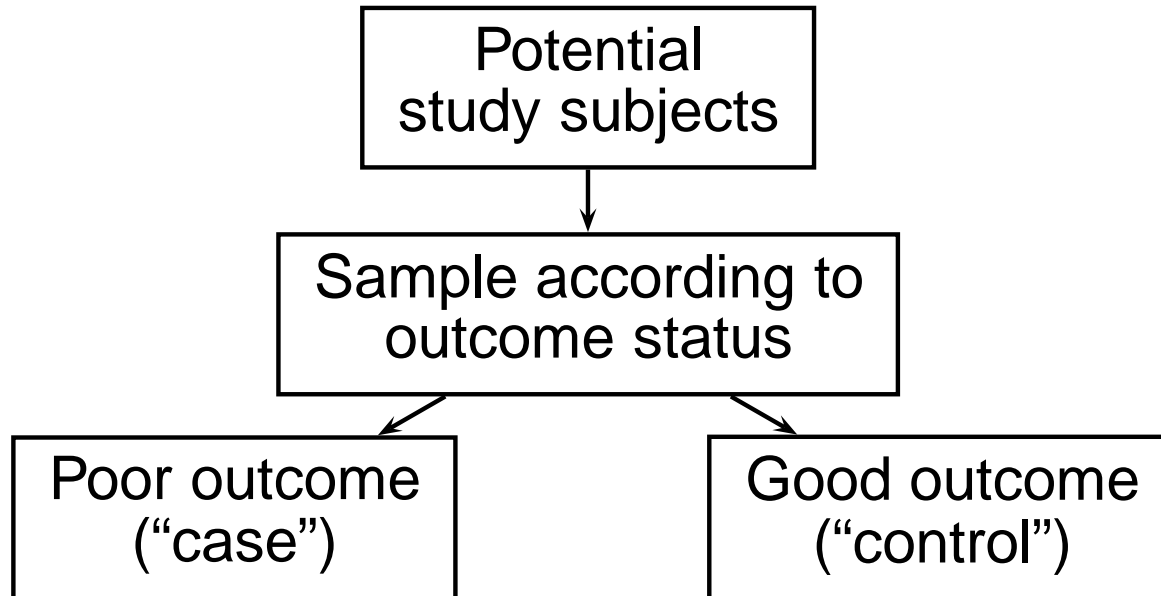
	Outcome	
Exposed?	Bad	Good
Yes	a	b
No	c	d

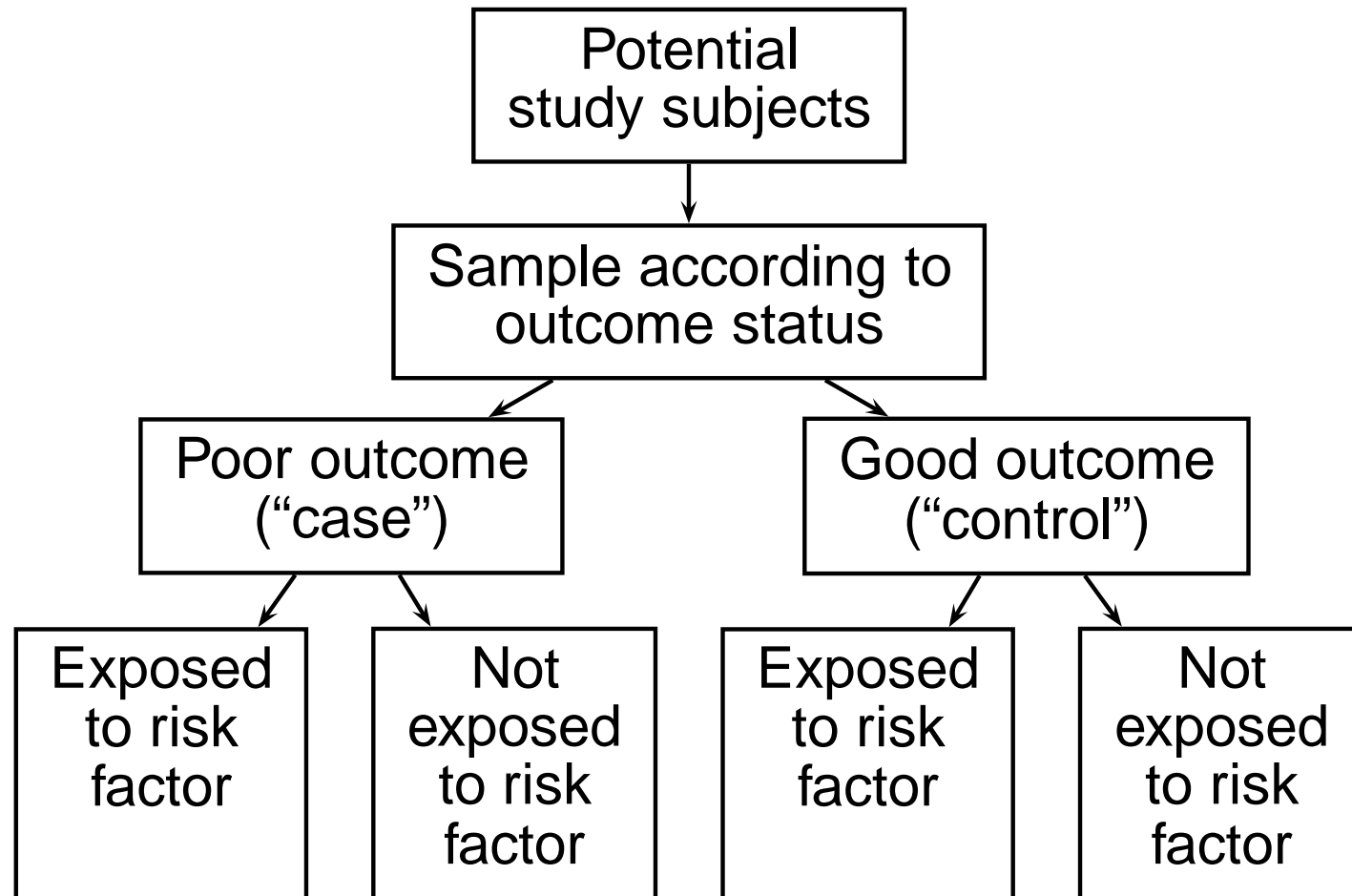where a, b, c, d are number of study subjects in each cell

Generic case-control study

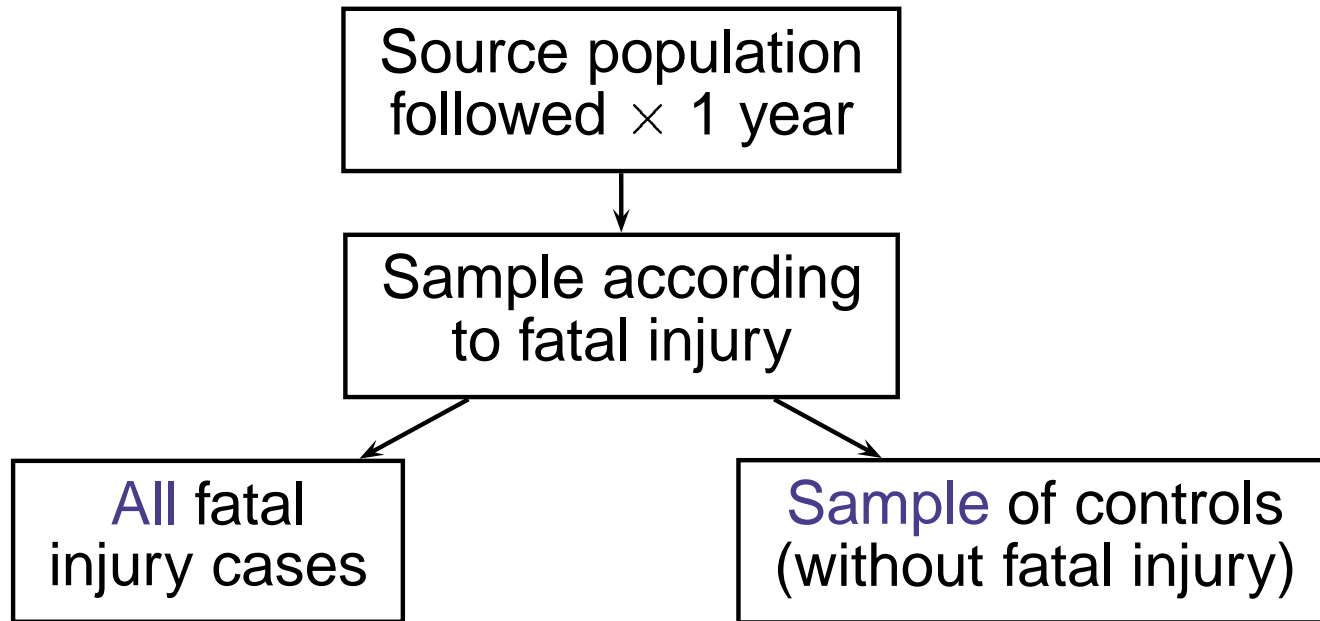Potential study subjects

Generic case-control study

Generic case-control study

Case-control study of statins and fatal injury

Source population
followed × 1 year

Case-control study of statins and fatal injury

Source population followed × 1 year

Sample according to fatal injury

All fatal injury cases

Sample of controls (without fatal injury)

Case-control study of statins and fatal injury
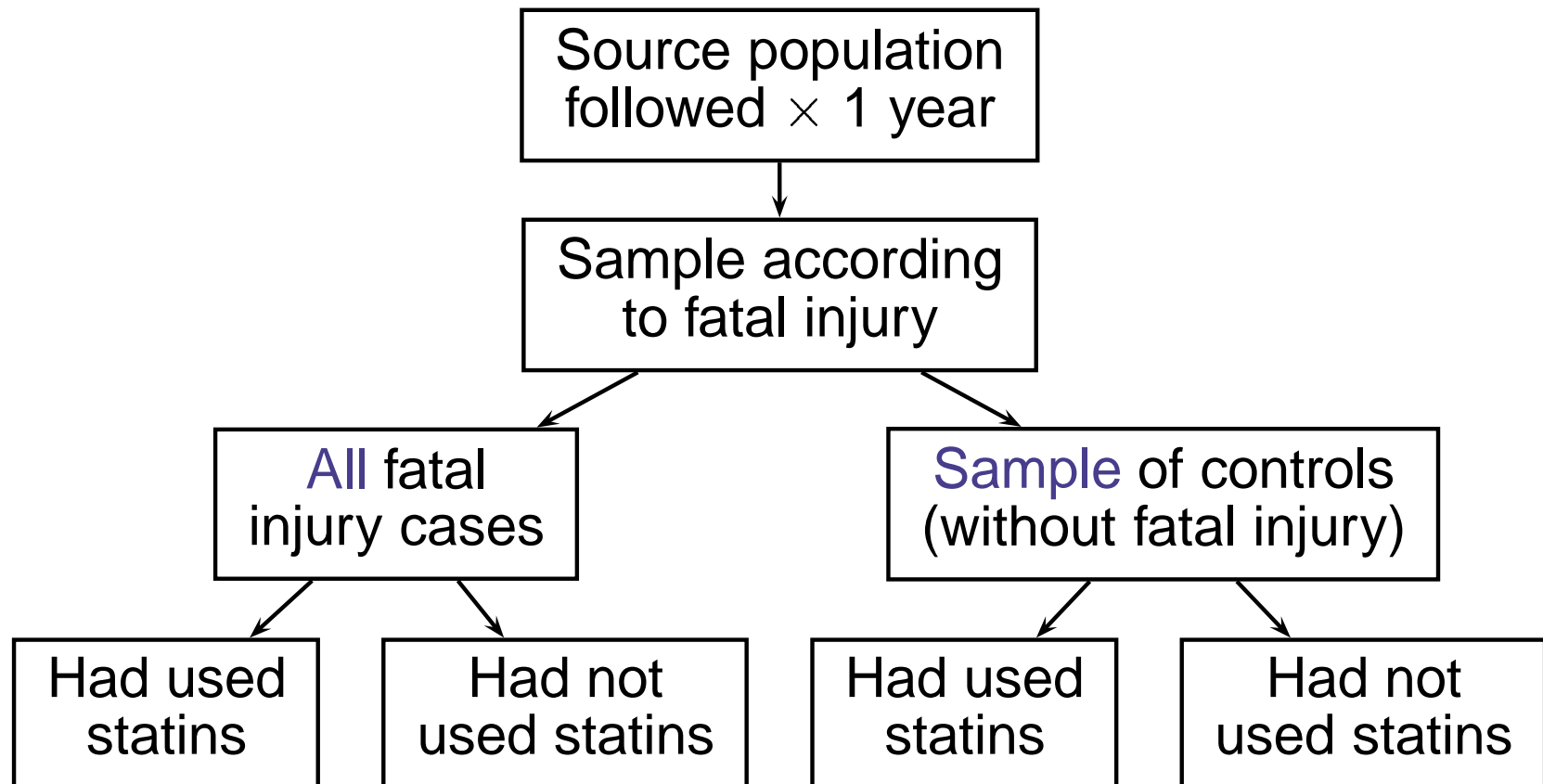
Source population
followed × 1 year

Sample according
to fatal injury

All fatal
injury cases

Sample of controls
(without fatal injury)

Had used
statins

Had not
used statins

Had used
statins

Had not
used statins

Key features of a case-control study

- Relative sizes of case and control groups usually do not reflect frequency of outcome in the source population. Typically:

  - All available cases included
  - Only a sample of many non-cases included

Key features of a case-control study

- Relative sizes of case and control groups usually do not reflect frequency of outcome in the source population. Typically:

  - All available cases included
  - Only a sample of many non-cases included

- Hence cannot directly estimate incidence from case-control data

Key features of a case-control study

- Relative sizes of case and control groups usually do not reflect frequency of outcome in the source population. Typically:
    - All available cases included
    - Only a sample of many non-cases included
- Hence cannot directly estimate incidence from case-control data
- Nonetheless, can nearly always estimate *ratio* of incidence in exposed to that in non-exposed (i.e., relative risk):

$RR \approx Odds\ Ratio = ad/bc$

(To be explained in EPI 512/513)

Case-control sampling: 1 control/case

Full population (unobserved):

	Fatal injury?		
	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

Case-control sampling: 1 control/case

Full population (unobserved):

Fatal injury?

	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

$$RR = \frac{20/1,020}{80/9,080}$$
$$= 2.23$$

Full population (unobserved):

Fatal injury?

	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

$$RR = \frac{20/1,020}{80/9,080}$$

$$= 2.23$$

Case-control sampling:

Statin		
No statin		

Full population (unobserved):

Fatal injury?

	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

$$RR = \frac{20/1,020}{80/9,080}$$
$$= 2.23$$

Case-control sampling:

Statin		
No statin		
100		

Case-control sampling: 1 control/case

Full population (unobserved):

Fatal injury?

	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

$$RR = \frac{20/1,020}{80/9,080}$$
$$= 2.23$$

Case-control sampling:

Statin		
No statin		
	100	100

Case-control sampling: 1 control/case

Full population (unobserved):

	Fatal injury?		
	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

$$RR = \frac{20/1,020}{80/9,080}$$
$$= 2.23$$

Case-control sampling:

Statin	20	
No statin	80	
	100	100

Case-control sampling: 1 control/case

Full population (unobserved):

Fatal injury?

	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

$$RR = \frac{20/1,020}{80/9,080}$$
$$= 2.23$$

Case-control sampling:

Statin	20	10
No statin	80	90
	100	100

Case-control sampling: 1 control/case

Full population (unobserved):

Fatal injury?

	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

$$RR = \frac{20/1,020}{80/9,080}$$
$$= 2.23$$

Case-control sampling:

Statin	20	10
No statin	80	90
	100	100

$$RR \approx OR$$
$$\approx \frac{20 \times 90}{10 \times 80}$$
$$\approx 2.25$$

Case-control sampling: 2 controls/case

Full population (unobserved):

Fatal injury?

	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

$$RR = \frac{20/1,020}{80/9,080}$$
$$= 2.23$$

Case-control sampling:

Statin	20	
No statin	80	
	100	200

Case-control sampling: 2 controls/case

Full population (unobserved):

Fatal injury?

	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

$$RR = \frac{20/1{,}020}{80/9{,}080}$$
$$= 2.23$$

Case-control sampling:

Statin	20	20
No statin	80	180
	100	200

Case-control sampling: 2 controls/case

Full population (unobserved):

Fatal injury?

	Yes	No	
Statin	20	1,000	1,020
No statin	80	9,000	9,080
	100	10,000	10,100

$$RR = \frac{20/1,020}{80/9,080}$$
$$= 2.23$$

Case-control sampling:

Statin	20	20
No statin	80	180
	100	200

$$RR \approx OR$$
$$\approx \frac{20 \times 180}{20 \times 80}$$
$$\approx 2.25$$

How many controls?
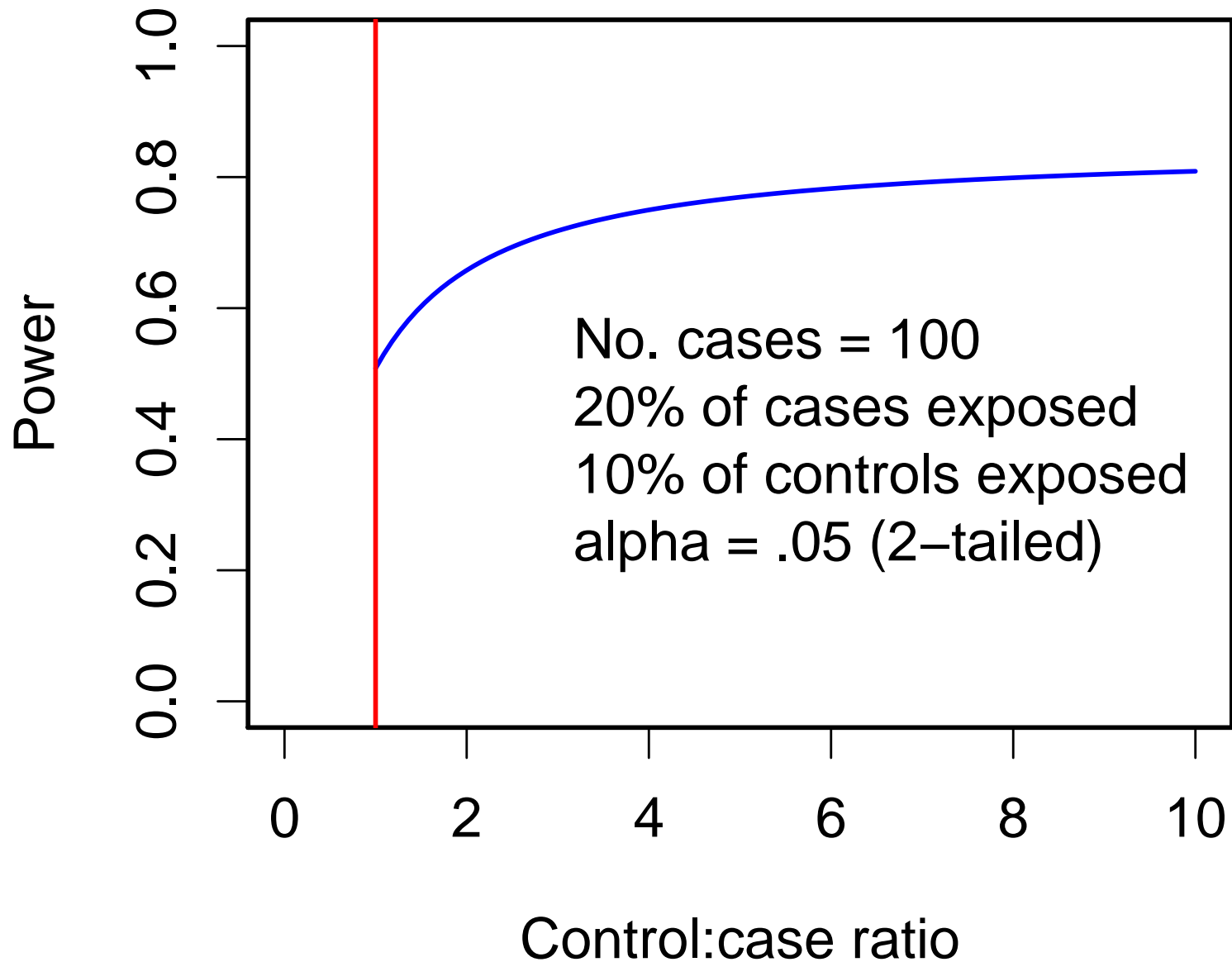
No. cases = 100
20% of cases exposed
10% of controls exposed
alpha = .05 (2–tailed)

Power

Control:case ratio

Strengths and weaknesses of case-control studies

Strengths

- Efficient for rare or delayed outcomes

Strengths

- Efficient for rare or delayed outcomes

- Usually relative quick and inexpensive

Strengths and weaknesses of case-control studies

Strengths

- Efficient for rare or delayed outcomes

- Usually relative quick and inexpensive

- Easy to study 2+ different exposures

Strengths and weaknesses of case-control studies

Strengths	Weaknesses
• Efficient for rare or delayed outcomes	• Cannot estimate incidence directly
• Usually relative quick and inexpensive	
• Easy to study 2+ different exposures	

Strengths and weaknesses of case-control studies

Strengths	Weaknesses
• Efficient for rare or delayed outcomes	• Cannot estimate incidence directly
• Usually relative quick and inexpensive	• Choice of appropriate controls can be difficult
• Easy to study 2+ different exposures	

Strengths and weaknesses of case-control studies

Strengths	Weaknesses
• Efficient for rare or delayed outcomes	• Cannot estimate incidence directly
• Usually relative quick and inexpensive	• Choice of appropriate controls can be difficult
• Easy to study 2+ different exposures	• Recall bias possible if exposure self-reported

Broad categories of case-control studies

- Population-based

- Clinic-based

Population-based case-control study

- Cases: all (or a representative sample of) cases in a *defined population*—e.g.:

Population-based case-control study

- Cases: all (or a representative sample of) cases in a *defined population*—e.g.:
    - Geopolitical area

Population-based case-control study

- Cases: all (or a representative sample of) cases in a *defined population*—e.g.:

  - Geopolitical area
  - HMO enrollees

Population-based case-control study

- Cases: all (or a representative sample of) cases in a *defined population*—e.g.:
    - Geopolitical area
    - HMO enrollees
    - Previously defined study cohort ("nested" case-control study)

Population-based case-control study

- Cases: all (or a representative sample of) cases in a *defined population*—e.g.:

    - Geopolitical area

    - HMO enrollees

    - Previously defined study cohort ("nested" case-control study)

- Controls: sample of non-cases drawn from the same defined population

Clinic-based case-control study

- Cases: those captured through a particular source of care

Clinic-based case-control study

- Cases: those captured through a particular source of care

- Controls: ?

Clinic-based case-control study

- Cases: those captured through a particular source of care

- Controls: ?

  - Other patients receiving care from same source?

Clinic-based case-control study

- Cases: those captured through a particular source of care

- Controls: ?

  - Other patients receiving care from same source?   (But their illnesses, too, may be related to the exposure)

Clinic-based case-control study

- Cases: those captured through a particular source of care

- Controls: ?

  - Other patients receiving care from same source? (But their illnesses, too, may be related to the exposure)

  - Spouse / sib / other relative?

Clinic-based case-control study

- Cases: those captured through a particular source of care

- Controls: ?

  - Other patients receiving care from same source?   (But their illnesses, too, may be related to the exposure)

  - Spouse / sib / other relative?   (But may share the same exposure)

Clinic-based case-control study

- Cases: those captured through a particular source of care

- Controls: ?

  - Other patients receiving care from same source?   (But their illnesses, too, may be related to the exposure)

  - Spouse / sib / other relative?   (But may share the same exposure)

  - Friends?

Clinic-based case-control study

- Cases: those captured through a particular source of care

- Controls: ?
  - Other patients receiving care from same source? (But their illnesses, too, may be related to the exposure)

  - Spouse / sib / other relative? (But may share the same exposure)

  - Friends? (But may skew toward gregarious volunteers)

Case-control studies by UW physician fellows

Scholar	Exposure	Outcome
Ed Boyko	Cigarette smoking	Ulcerative colitis

Case-control studies by UW physician fellows

Scholar	Exposure	Outcome
Ed Boyko	Cigarette smoking	Ulcerative colitis
Bruce Psaty	"Non-cardiac" chest pain in people with hypertension	Myocardial infarction

Case-control studies by UW physician fellows

Scholar	Exposure	Outcome
Ed Boyko	Cigarette smoking	Ulcerative colitis
Bruce Psaty	"Non-cardiac" chest pain in people with hypertension	Myocardial infarction
Ken Schellhase	Screening for Type 2 diabetes	End-organ complications

Case-control studies by UW physician fellows

Scholar	Exposure	Outcome
Ed Boyko	Cigarette smoking	Ulcerative colitis
Bruce Psaty	"Non-cardiac" chest pain in people with hypertension	Myocardial infarction
Ken Schellhase	Screening for Type 2 diabetes	End-organ complications
Heidi Blume	Chorioamnionitis, fever	Neonatal encephalopathy

The end