

# Case-control studies

---

**Hans Wolff**

**Service d'épidémiologie clinique  
Département de médecine communautaire**

[Hans.Wolff@hcuge.ch](mailto:Hans.Wolff@hcuge.ch)

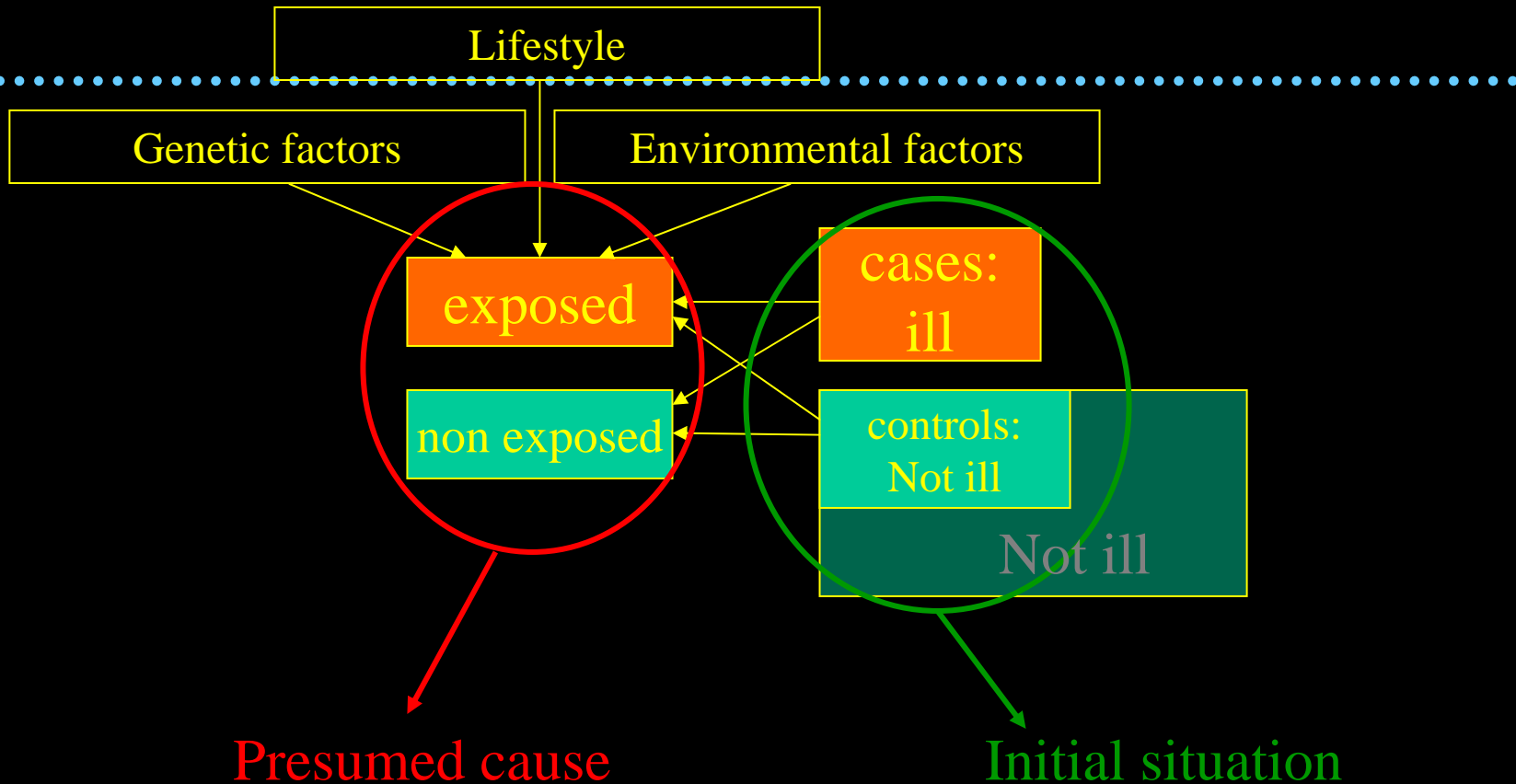


# Outline

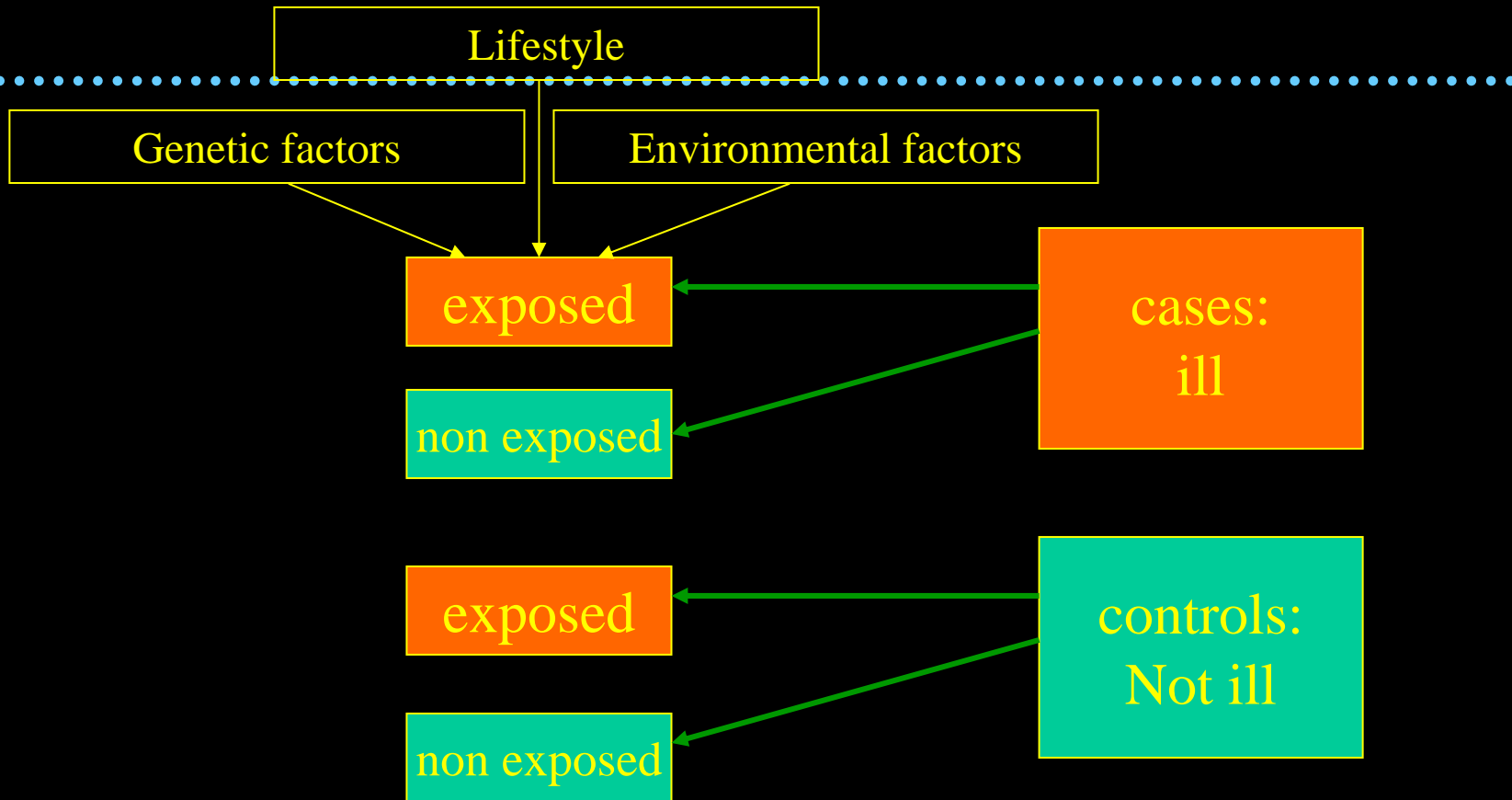
---

- Case-control study
- Relation to cohort study
- Selection of controls
- Sampling schemes of controls

# Case-control studies (CCS)



# Case-control studies (CCS)



**Did they were exposed or not ?**

# 1. Example: Passive Smoking & Breast Cancer

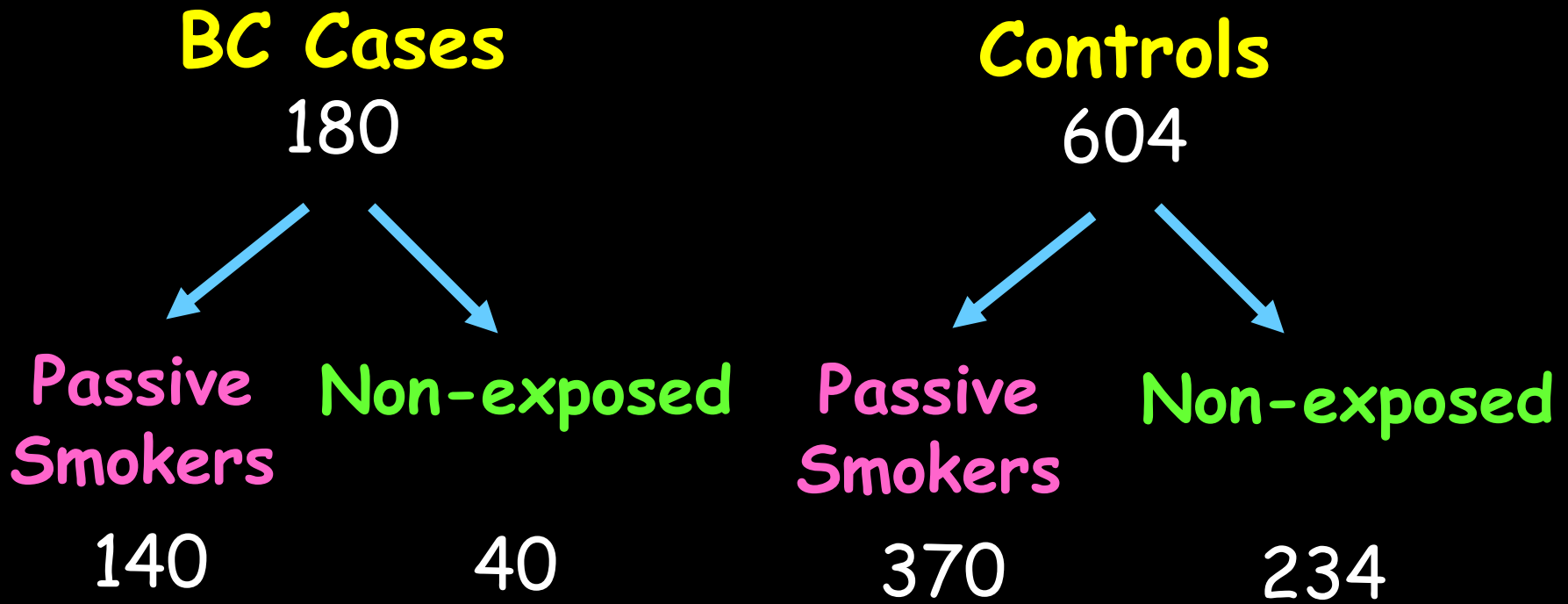
---

<b>Smoking</b>	<b>Cases</b>		<b>Controls</b>		<b>Odds Ratio</b>
	n	%	n	%	
<b>Unexposed</b>	40	22.2	234	38.7	<b>1.0</b>
<b>Passive</b>	140	77.8	370	61.3	<b>2.2</b>

# Case-Control Design

---

## SAMPLE



# Presence or absence of disease ...

---

... is fixed by design in case-control studies.

- Cases have the disease
- Controls don't.
- We can NOT compute a risk of disease
- We CAN compute prevalence of exposure in cases and controls

# Passive Smoking & Breast Cancer

---

- **Cases:** all incident breast cancer in Geneva
- **Controls:** random sample of the Geneva female population
- **Exposure:** questionnaire on lifetime history of exposure to passive smoke



# Have you ever been exposed?

---

- ... to passive smoking at least 1 hour per day for at least 1 year? (Yes / No)
- At home ? At work ? During leisure time ?
- If yes, describe each episode of exposure
  - Duration, who, size of the room, etc...
  - *Unexposed* = never active, never passive

# What should be always true for a case-control study?

---

1. Cases and controls are randomized with respect to exposure.
2. Cases are a representative sample of all cases in the general population
3. Controls are a representative sample of the general population
4. Cases and controls have the same population of origin
5. Always start with some cases, then identify their valid controls

## Fundamental conditions for the validity of this case-control design

---

Cases and controls :

- Originate from Geneva resident, <75 y.
- are sampled independently of their exposure to passive smoke

### **Solution:**

- All incident cases over a given time period
- Controls are a random sample of population

# Case Definition

---

- Incident (= newly diagnosed)
- Between 1/1/92 and 12/31/93
- Resident of Geneva
- Aged < 75 yrs
- Identified: all pathology labs of Geneva

# Control Definition

---

- Never diagnosed with breast cancer
- Between 1/1/92 and 12/31/93
- Resident of Geneva
- Aged < 75 yrs
- Stratified random sample
  - Population controls
    - **Why not use hospital controls?**

# Prevalence of Passive Smoking

---

<b>Smoking</b>	<b>Cases</b> n	<b>Controls</b> n
<b>Unexposed</b>	40	234
<b>Passive</b>	140	370

# The proportion of passive smoker cases is...

---

1.  $\left( \frac{40}{234} \right)$

4.  $\left( \frac{370}{234} \right)$

2.  $\left( \frac{140}{40} \right)$

5.  $\left( \frac{370}{604} \right)$

3.  $\left( \frac{140}{180} \right)$

# Prevalence of Passive Smoking

---

<b>Smoking</b>	<b>Cases</b>		<b>Controls</b>	
	n	%	n	%
<b>Unexposed</b>	40	22.2	234	38.7
<b>Passive</b>	140	<b>77.8</b>	370	<b>61.3</b>



# The odds of passive smoking in **CASES** is...

---

$$1. \left( \frac{140}{40} \right) = 3.5$$

$$3. \left( \frac{140}{180} \right) = 77.8$$

$$2. \left( \frac{77.8}{22.2} \right) = 3.5$$

$$4. \left( \frac{140}{77.8} \right) = 1.8$$

5. **Answers 1 or 2**

# Odds of Passive Smoking in CASES

---

Smoking history	N	%
Unexposed	40	22.2
Passive	140	77.8
Total	180	100.0
Odds =	$140/40=$	$77.8/22.2=$
Odds =	<b>3.5</b>	<b>3.5</b>

# Odds of Passive Smoking in CONTROLS

---

Smoking history	N	%
Unexposed	234	38.7
Passive	370	61.3
Total	604	100.0
Odds =	$370/234=$	$61.3/38.7=$
Odds =	<b>1.6</b>	<b>1.6</b>

# AR in case-control study?

---

Recall

$$AR_{\text{duration}} = \text{Risk (E+)} - R(\text{E-})$$

Since risk cannot be computed directly from a case-control study, AR cannot be computed either.

## RR in case-control study?

---

$$RR = \text{Risk (E+)} / R(E-)$$

Since risk cannot be computed directly from a case-control study, RR cannot be computed either

# Odds Ratio of Passive Smoking

---

<b>Group</b>	<b>Odds</b>	<b>Odds Ratio</b>
--------------	-------------	-------------------

---

**Cases**

3.5

$$\left( \frac{3.5}{1.6} \right) = 2.2$$

**Controls**

1.6

$$\left( \frac{1.6}{1.6} \right) = 1.0$$

**Your interpretation?**

**Reference  
Group**

# Interpretation of the Odds Ratio (1)

---

- The odds of being a passive smoker are 2.2 greater in breast cancer cases than in population controls.

## Alternatively:

- The odds of breast cancer is 2.2 greater in those exposed to passive smoke than in unexposed.
  - WHY ?



*Imagine ...*

.....  
you could have  
done the perfect  
cohort study  
instead of the  
case-control  
study



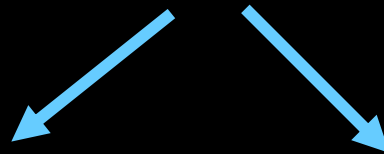
# Cohort Design (Risk period: 2 yrs)

---

## Female Population of Geneva

Passive Smokers

55,500



Breast  
Cancer

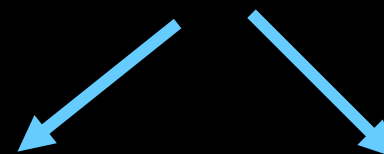
140

No Breast  
Cancer

55,360

Non-exposed

35,100



Breast  
Cancer

40

No Breast  
Cancer

35,060

# Odds Ratio of Breast Cancer

---

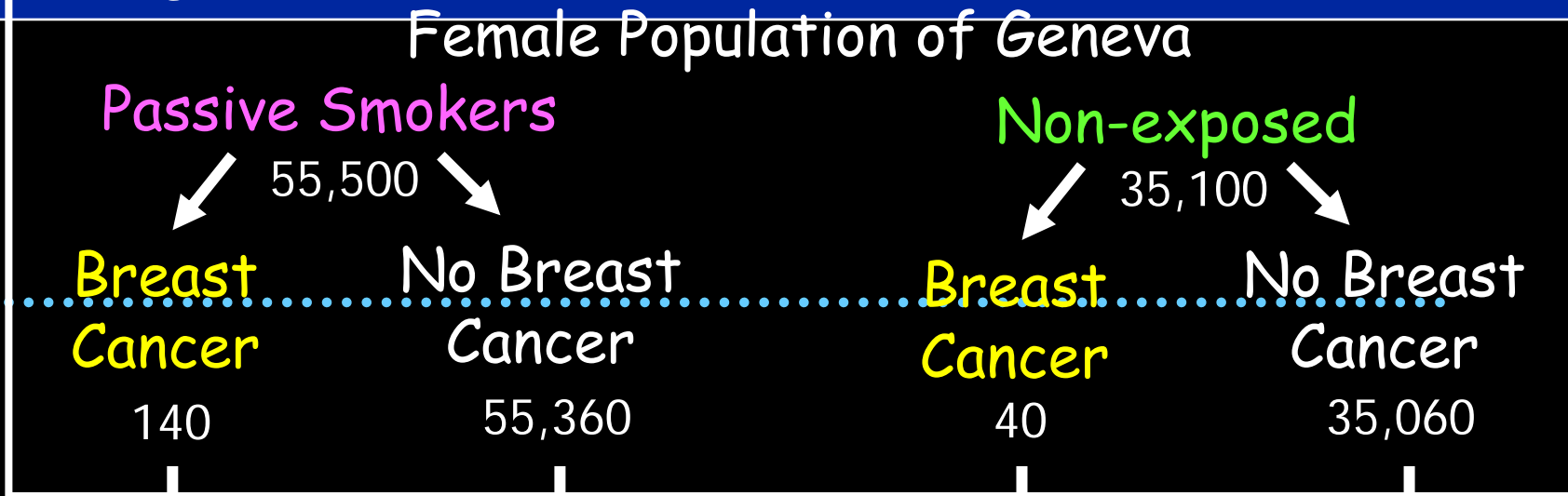
<b>Breast Cancer</b>	<b>Passive Smokers</b>	<b>Unexposed</b>
<b>Present (A)</b>	140	40
<b>Absent (B)</b>	55,360	35,060
<hr/>		
<b>Odds (A/B)</b>	0.00253	0.00114
<b>Odds Ratio</b>	<b>2.2</b>	1.0 (ref)

**Your interpretation?**

# Identity of Odds Ratio

---

- **Case-control study:**
  - Odds ratio of passive smoking = 2.2
- **Cohort study:**
  - Odds ratio of breast cancer = 2.2
    - Same interpretation
- **Identical Odds Ratio in the cohort and in the case-control studies.**

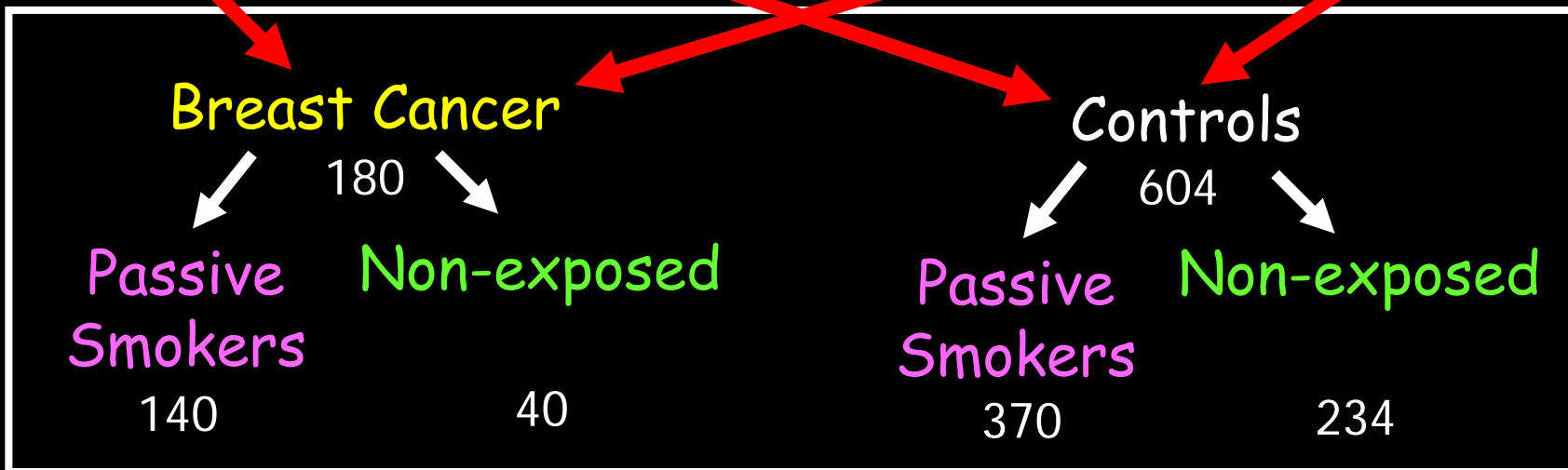


$F_1 = 1.0$

$F_2 = 0.005$

$F_3 = 1.0$

$F_4 = 0.005$



$F_n$  = fraction included into the sample

# Relation of Case-Control to Cohort Studies

---

- In a case-control study:
  - **CASES** are sampled among people in the unexposed and passive smokers cohorts who did develop breast cancer
  - **CONTROLS** are sampled among people in the unexposed and passive smokers cohorts who did **not** develop breast cancer

# Odds Ratio and Relative Risk

---

- Relative Risk =  $\left( \frac{140 / 55,500}{40 / 35,100} \right) = 2.2$

Note effect of rare disease on denominators

- Odds Ratio =  $\left( \frac{140 / 55,360}{40 / 35,060} \right) = 2.2$

## Interpretation of the Odds Ratio (2)

---

- The **ODDS** of breast cancer is 2.2 greater in those exposed to passive smoke than in unexposed.

### Alternatively:

- The **RISK** of breast cancer is 2.2 greater in those exposed to passive smoke than in unexposed.

# Equivalence OR and RR

---

The OR is a good estimation for the RR if :

the prevalence of the illness is low (<10%)



# Comparison of the OR and RR

## Illness with low prevalence

	Cases (M+)	Controls (M-)	n
Exposed (E+)	2	98	100
non-exposed (E-)	1	99	100
Total	3	197	

$$RR = \frac{2/100}{1/100} = 2 \quad OR = \frac{2/1}{98/99} = 2.02$$

# Comparison of the OR and RR

## Illness with high prevalence

	Cases (M+)	Controls (M-)	n
Exposed (E+)	50	50	100
Non-exposed (E-)	25	75	100
Total	75	125	

$$RR = \frac{50 / 100}{25 / 100} = 2 \quad OR = \frac{50 / 25}{50 / 75} = 3$$

# Advantages of Case-Control Studies (1)

---

- Less expensive ...
- Require smaller sample sizes ...
- Shorter duration ... than prospective study
- Study multiple risk factors for 1 disease
- Easily reproduced in different populations by different investigators

# Disadvantages of Case-Control Studies (1)

---

- Information about exposure is often obtained after the diagnosis is done
  - Example: diet, physical activity
- Dependent on the subject's memory, which may be affected by the disease

## Disadvantages of Case-Control Studies (2)

---

- Population of origin for cases is difficult to define precisely.
  - Difficult to identify appropriate control group
- Does not provide estimate of risks and attributable risk