Case-control studies

Alfredo Morabia

Division d'épidémiologie Clinique, Département de médecine communautaire, HUG
Alfredo.Morabia@hcuge.ch
www.epidemiologie.ch

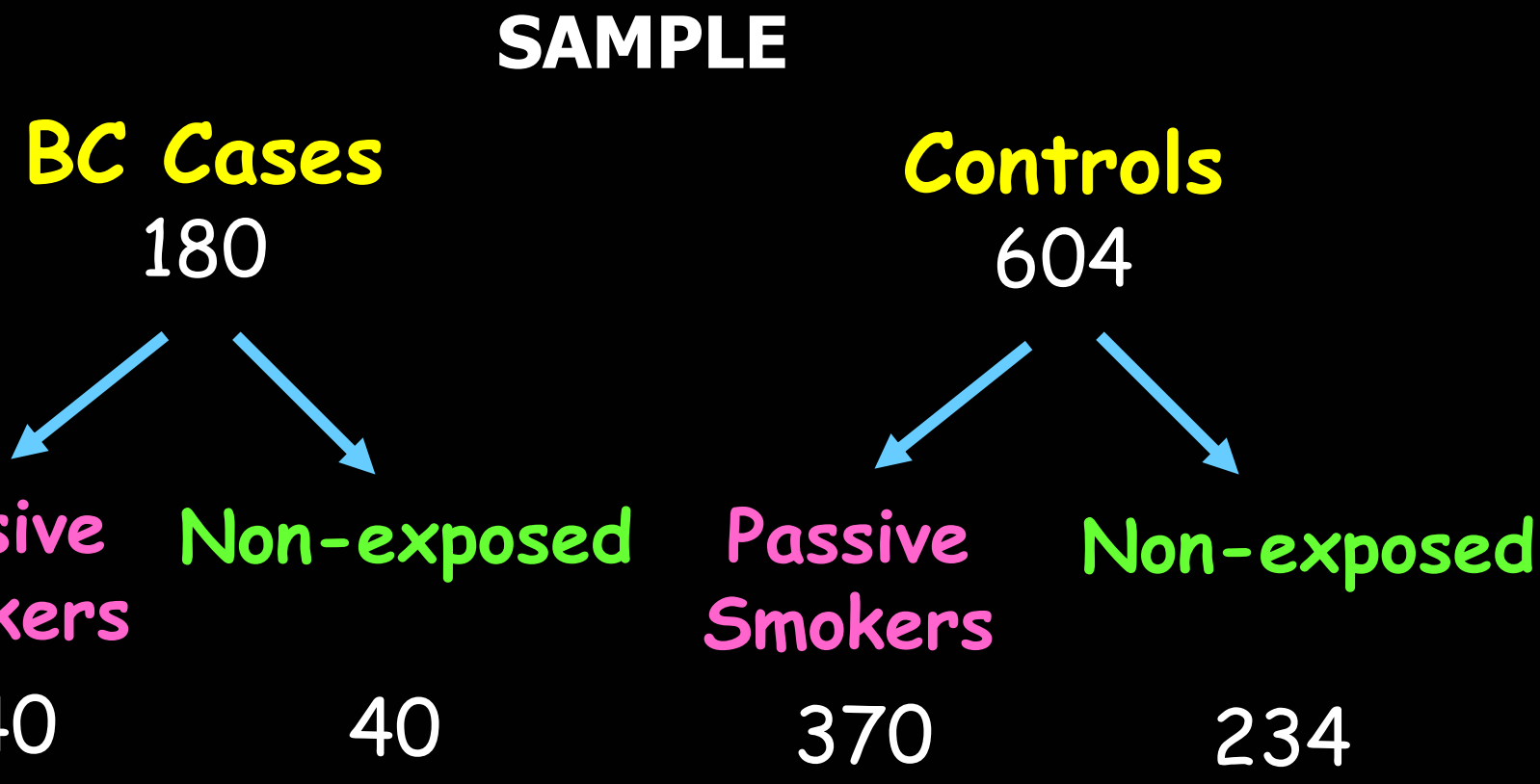
Outline

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

1. Example: Passive Smoking & Breast Cancer

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

Case-Control Design



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

Smoking	Cases n	Controls n
Unexposed	40	234
Passive	140	370

The proportion of passive smoker cases is…

1. $\left(\frac{40}{234}\right)$
2. $\left(\frac{140}{40}\right)$
3. $\left(\frac{140}{180}\right)$
4. $\left(\frac{370}{234}\right)$
5. $\left(\frac{370}{604}\right)$

Prevalence of Passive Smoking

Smoking	Cases		Controls	
	n	%	n	%
Unexposed	40	22.2	234	38.7
Passive	140	**77.8**	370	**61.3**

Prevalence of Passive Smoking

Smoking	Cases		Controls	
	n	%	n	%
Unexposed	40	22.2	234	38.7
Passive	140	**77.8**	370	**61.3**

The odds of passive smoking in CASES is…

1. $\left(\frac{140}{40}\right) = 3.5$
2. $\left(\frac{77.8}{22.2}\right) = 3.5$
3. $\left(\frac{140}{180}\right) = 77.8$
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 =	**3.5**	**3.5**

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 =	**1.6**	**1.6**

AR in case-control study?

Recall

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

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

RR in case-control study?

RR = 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

Group	Odds	Odds Ratio
Cases	3.5	$\left(\frac{3.5}{1.6}\right) = 2.2$
Controls	1.6	$\left(\frac{1.6}{1.6}\right) = 1.0$ Reference Group

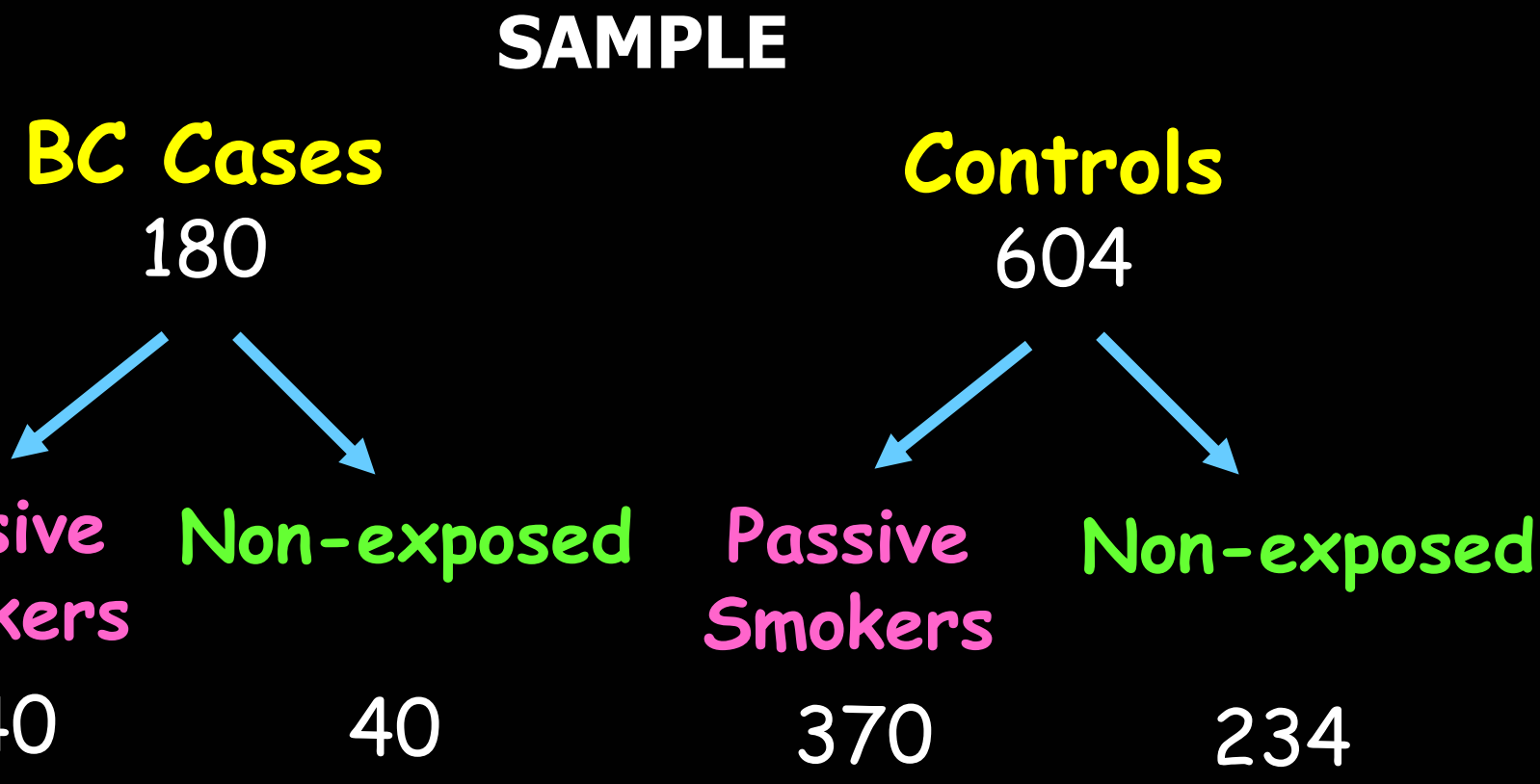
Interpretation of the Odds Ratio (1)

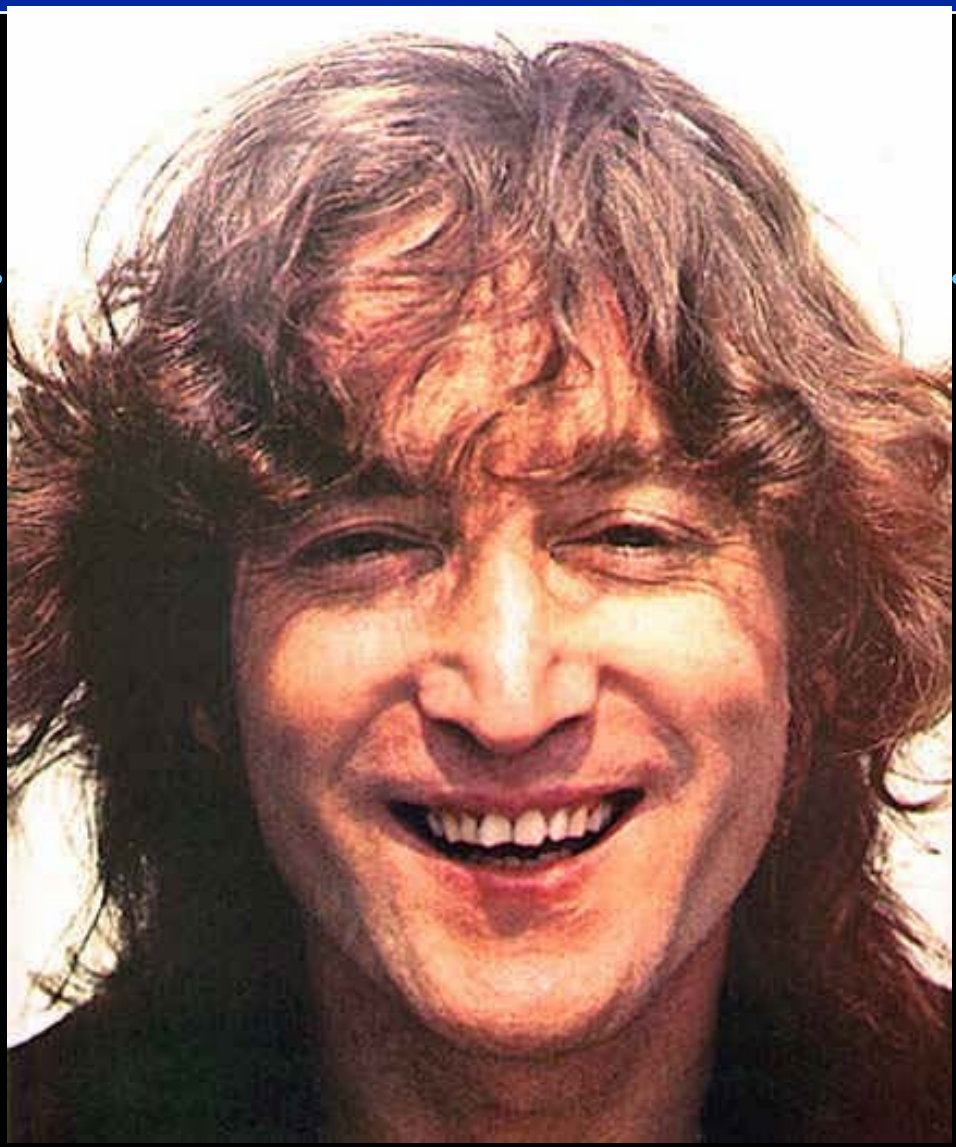
- The odds of being a passive smoker are 2.2 greater in breast cancer cases than in population controls.

Alternatively:

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

Case-Control Design



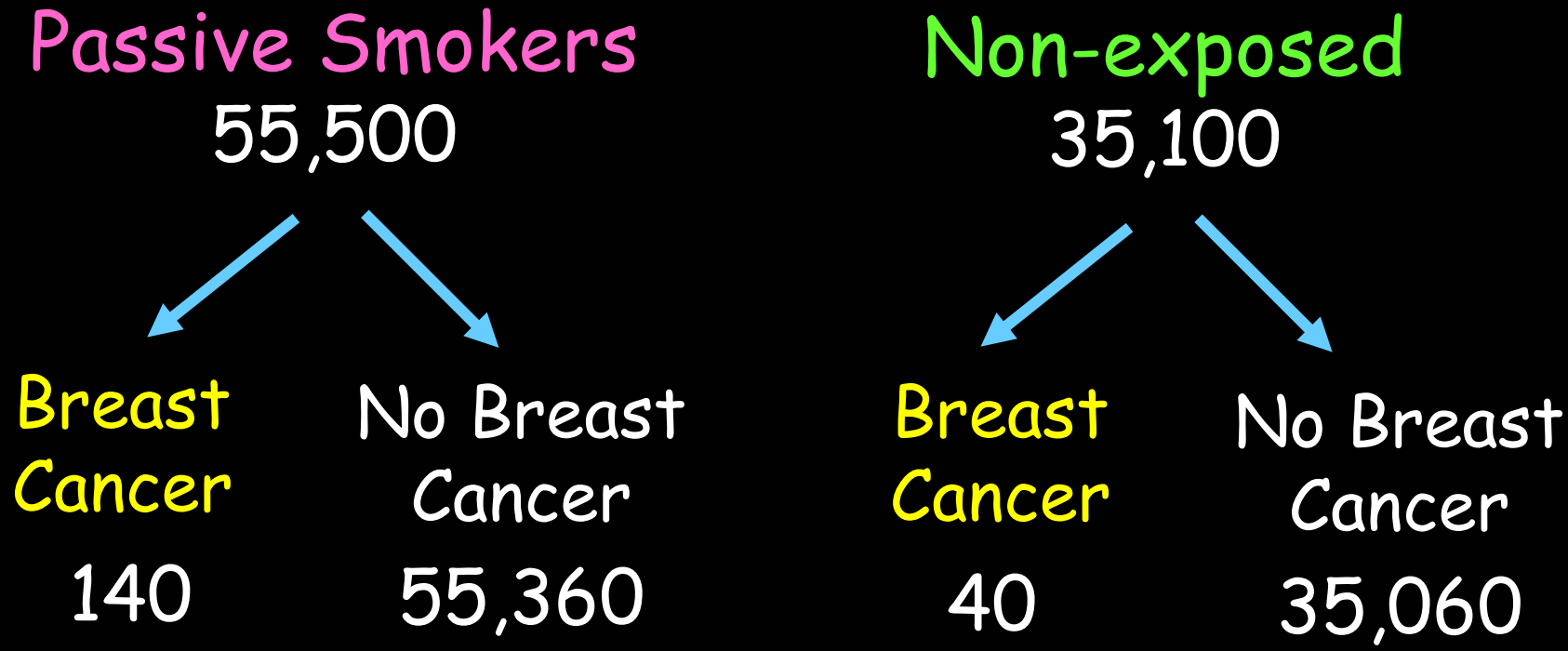


Imagine ...

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

Cohort Design (Risk period: 2 yrs)

Female Population of Geneva

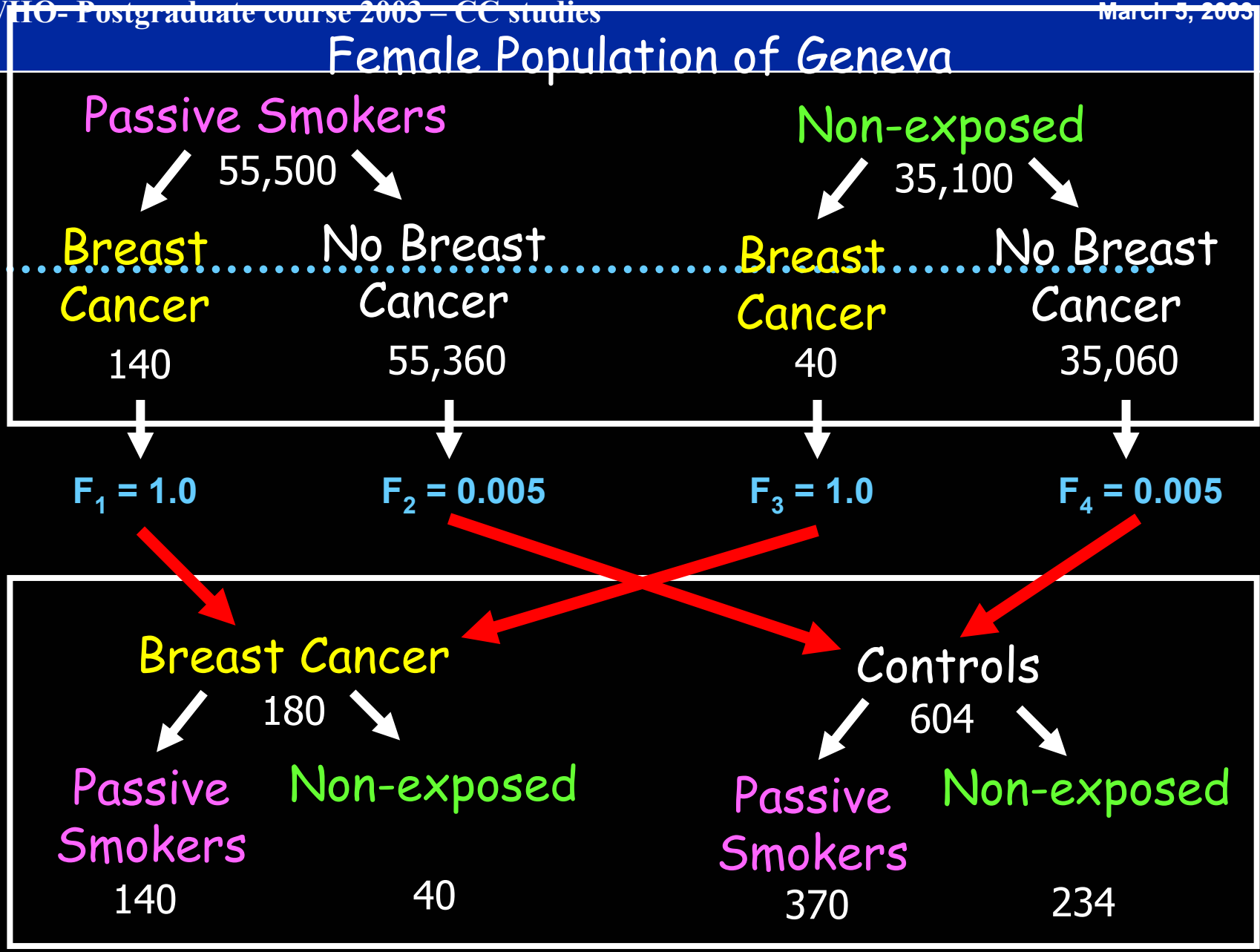


Odds Ratio of Breast Cancer

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

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_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.

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