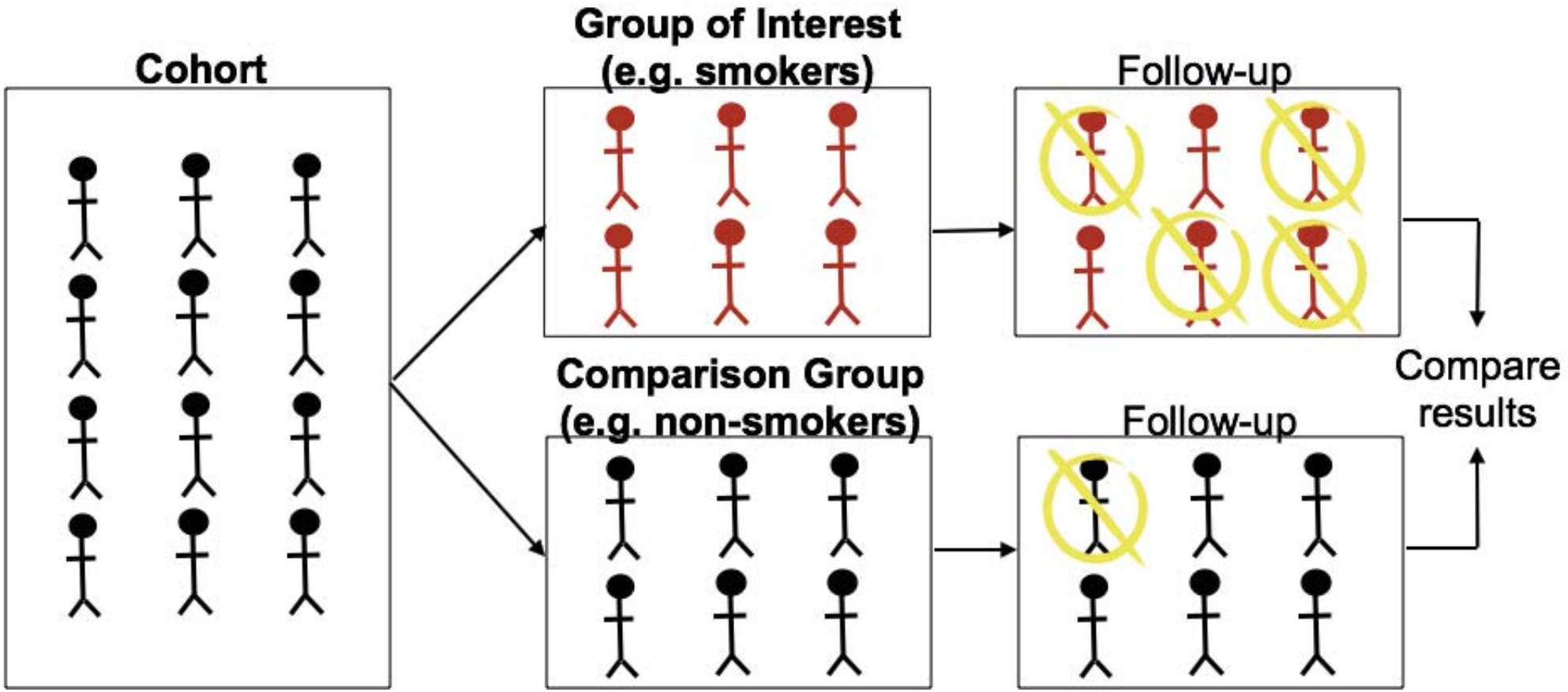
Abortion

Week 6

Objectives

1. Learn more details about the cohort study design
2. Comprehend confounding and calculate unbiased estimates
3. Critically evaluate how abortion is related to issues that derived from sex-linked biology and gender



Cohort

Synonyms: follow-up study, longitudinal study

Type

Open (dynamic)

- Defined by a changeable characteristic
- Exposure status may change over time
- Outcome measure
    - Incidence rate (IR) since variable follow-up duration

Fixed

- Defined by an irrevocable event
- Exposure defined at start of follow-up, no new enrollees
- Outcome measure
    - Cumulative incidence (CI) (if loss to follow-up loss is low)
    - IR (if loss to follow-up is high)

Closed

- Defined by an irrevocable event
- Exposure defined at start of follow-up, no new enrollees
- No losses during short follow-up
- Outcome measure
    - CI

Timing

Retrospective

- Investigator does not wait for outcomes to develop
- Various benefits and determinants compared to prospective
- Less control of quality and quantity of the data
- Less time consuming
- Less expensive
- Completely dependent on available data
- Potential good starting point for scientific inquiry

Prospective

- Investigator waits for outcomes to develop
- Various benefits and determinants compared to retrospective
- More control of the quality and quantity of the data
- Less potential for bias
- Less unavailable data
- More time consuming
- More expensive

Ambidirectional: Elements of both

Nature of Cohort

General

- Nothing special about exposure
- Often selected on geography (Framingham) or profession (Nurses')
- Appropriate when prevalence of exposure is not too high or low

Special Exposure

- Higher prevalence of exposure (good for rare exposure)

Advantages

- Correct temporal sequence: exposure → outcome
- Good exposure status information
- Efficient for rare exposures
- Can study several outcomes associated with a single exposure
- Can minimize bias in exposure ascertainment (prospective cohorts)
- Can directly measure incidence of disease among exposed and non-exposed subjects

Disadvantages

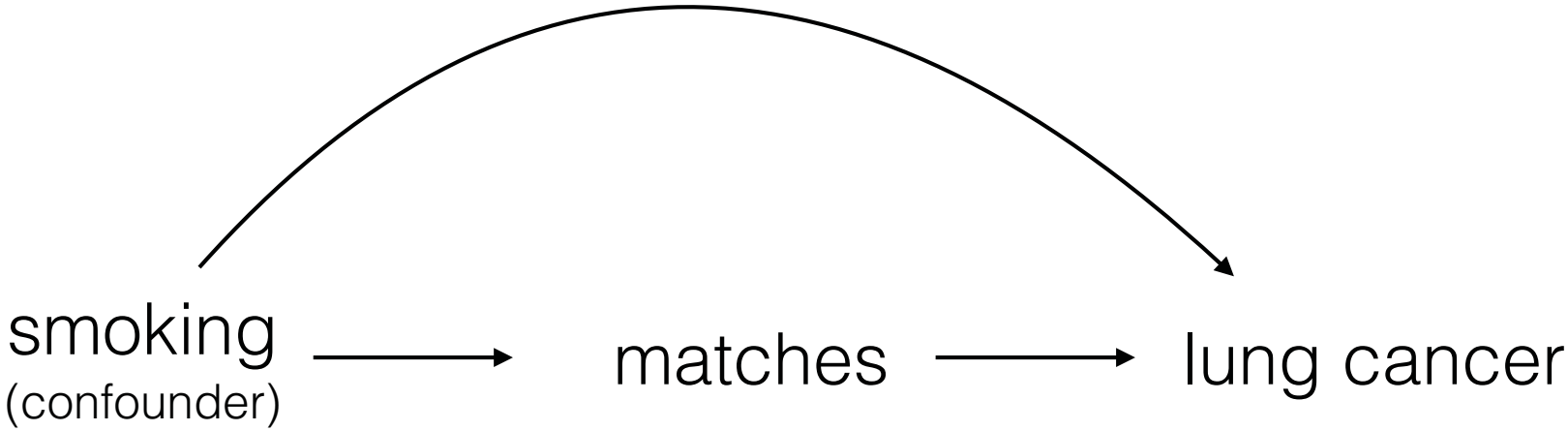
- Inefficient for studying rare diseases
- Time-consuming (prospective cohorts)
- Must minimize loss to follow-up
- Requires pre-recorded information on exposure and confounders (retrospective)

Confounding

A confounder is a factor which because of its relationship with the exposure and disease will distort the relative risk

- Will depend on the relationships of the factors in your study
- Confounding is a nuisance factor
- Need to remove the effect of the confounder to understand the exposure/disease relationship – want to control for confounding
    - Need to collect information on potential confounders – or at least known risk factors for outcome.

Can demonstrate visually with Direct Acyclic Graph (DAG)



	lung cancer	no lung cancer
matches	25	20
no matches	125	130
	150	150

OR=1.3

Two possible paths:
Direct effect of matches on lung
“Backdoor path” from matches to lung through smoking

Problem with confounding is that the exposed and unexposed groups differ.

We want to look at the effect of the exposure on disease in the scenario where the exposed and unexposed do not differ.

Solution: Adjust (or otherwise account) for potential confounder

Overall

	lung cancer	no lung cancer
matches	25	20
no matches	125	130
	150	150

OR=1.3

Smokers

	lung cancer	no lung cancer
matches	20	10
no matches	80	40
	100	50

OR=1.0

Non-Smokers

	lung cancer	no lung cancer
matches	5	10
no matches	45	90
	50	100

OR=1.0

Weighted estimate: OR=1.0

Confounding Definition

- Confounder must have a **different distribution** in the exposed and unexposed groups.
- Confounder must have a **direct effect** on the disease in **absence** of exposure.
- Confounder should NOT be in the **causal pathway** between exposure and disease.
- Important note: Something that is a confounder in one population may not be a confounder in another population.

Methods to Control for Potential Confounders

In the **design** of the study

- Randomization
- Restriction
- Matching

In the **analysis** of the study

- Matched analysis
- Stratification (e.g., pooling)
- Multivariate analysis

Design: Randomization

Randomization to allocate exposure

- Can only be done in experimental studies
- Control of confounding by known as well as unknown confounding factors, as long as the sample is big enough
- The control of unknown confounders is unique to this design feature

Design: Restriction

Restrict subjects to one level/stratum of the confounding factor(s)

- For example, perform your study just in men if you are worried about confounding by sex/gender
- Limitation: Limits generalizability

Design: Matching

Match the study groups so they have identical levels of the confounder

- Exact matching (or individual matching)
- Frequency matching
- Limitations
  - Individual matching can be difficult to do
  - Lose many potential participants
  - Can’t examine matched factor

Analysis: Matched

- But note
  - Because of the potential for overmatching, special type of test needed if you matched individually in the design
- Biostatistics test
  - McNemar's test

Analysis: Stratification

- Want to look at the effect where the exposed and unexposed do not differ by levels of confounder
- Stratum-specific estimates by levels of the confounders are unconfounded
- Need to combine the unconfounded stratum-specific estimates into one relative risk which is also unconfounded
    - Can do with pooling or standardization

Analysis: Stratification

A weighted average of stratum-specific relative risks

Approach

- **Divide** the data into groups (strata) according to categories of your potential confounder
- **Calculate** stratum-specific relative risks
- Pool information over all stratum by calculating a **weighted average** of stratum-specific relative risks to compare to the crude estimate
- The weights should reflect the amount of information in each stratum (e.g., sample size)

Crude Analysis

		Disease: Yes	Disease: No	
Exposure	**Yes**	a	b	a+b
	No	c	d	c+d
		a+c	b+d	

RR_{crude}

Stratified Analysis by Level of Confounding Factor(s)

Stratum 1

		Disease: Yes	Disease: No	
Exposure	**Yes**	a	b	a+b
	No	c	d	c+d
		a+c	b+d	

$RR_{stratum1}$

Stratum 2

		Disease: Yes	Disease: No	
Exposure	**Yes**	a	b	a+b
	No	c	d	c+d
		a+c	b+d	

$RR_{stratum2}$

$RR_{adjusted}$

Confounding: RR_{crude} vs $RR_{adjusted}$

To Obtain Weighted/ Adjusted RR

Mantel-Haenszel estimators

- Weighted average of RRs of a series of tables: RR_i

Weights reflect amount of "information" within each stratum

- Weight increases with
  - Total number in table
  - Balance in exposed-nonexposed
  - Increased risk of outcome

Mantel-Haenszel estimators

Cumulative incidence data

		Disease: Yes	Disease: No	Total # people
Exposure	**Yes**	a	b	a+b (N_1)
	No	c	d	c+d (N_0)
		a+c	b+d	T

$$RR_{MH} = \frac{\sum w_i RR_i}{\sum w_i} = \frac{\sum \frac{a_i N_{0i}}{T_i}}{\sum \frac{c_i N_{1i}}{T_i}} \quad \text{(if } w_i \neq 0\text{)}$$

Incidence rate data

		Disease: Yes	Total # p-yrs
Exposure	**Yes**	a	N_1
	No	c	N_0
		a+c	T

where $$w_i = (T_i)\left(\frac{N_{1i} N_{0i}}{T_i T_i}\right)\left(\frac{c_i}{N_{0i}}\right) = \frac{c_i N_{1i}}{T_i}$$

Stratification Example

Gender and mortality among patients with heart disease

- Potential confounding by age

Crude Analysis

		Mortality Yes	Person-yrs
Exposure	**Males**	90	2,465
	Females	131	3,946
		221	6,411

RR = (90/2465p-y)/(131/3946p-y) = 1.1

Stratification Example

Stratified Analysis

Age <65

	Mortality Yes	Person-yrs
Males	14	1,516
Females	10	1,701
	24	3,217

$RR_{age<65} = (14/1516)/(10/1701) = 1.57$

Age 65+

	Mortality Yes	Person-yrs
Males	76	949
Females	121	2,245
	197	3,194

$RR_{age65+} = (76/949)/(121/2245) = 1.49$

MH estimate

$$RR_{MH} = \frac{\sum \frac{a_i N_{0i}}{T_i}}{\sum \frac{c_i N_{1i}}{T_i}} = \frac{\frac{(14)(1701)}{3217} + \frac{(76)(2245)}{3194}}{\frac{(10)(1516)}{3217} + \frac{(121)(949)}{3194}} = 1.50$$

Stratification Example

Conclusions

- Age-adjusted RR (1.5) differs from crude RR (1.1)
- There is confounding by age
- Report relative risk adjusted for age

Mantel-Haenszel estimators

Case-control data

		Disease		
		Case	**Control**	**Total**
Exposure	**Yes**	a	b	a+b
	No	c	d	c+d
		a+c	b+d	T

$$RR_{MH} = \frac{\sum w_i OR_i}{w_i} = \frac{\sum \frac{a_i d_i}{T_i}}{\sum \frac{b_i c_i}{T_i}} \qquad \text{(if } w_i \neq 0\text{)}$$

where $w_i = \frac{b_i c_i}{T_i}$

Mantel-Haenszel Limitations

Can be done as a univariate analysis

- One variable at a time

Cumbersome with multiple confounders

- Results in multiple tables with small numbers (sparse data) in some of the cells
- Reduces power

Analysis: Multivariable Analysis

- Use mathematical modeling (regression models) to control for many confounders simultaneously
  - Many types, basic structure of formula is a line:
    - $Y = a(X) + b$
  - Outcome = intercept term (b) + a1(exposure) + a2(first confounder) + a3(second confounder) + … + ai (last confounder)

Analysis: Multivariable Analysis

a1: coefficient of the exposure

- Effect of the exposure on the outcome, adjusting/ controlling for the differences in all the confounding factors included in the model.

Example: Mortality = b + a1 [gender (exposure)] + a2 [age (confounder)]

- a1 represents effect of gender on mortality, controlling for differences in age

Confounding Summary

Confounder is a factor which, because of its relationship with the exposure and disease, will distort the relative risk

- Will depend on the relationships of the factors in your study
- Confounding is a nuisance factor
- You need to remove the effect of the confounder to understand the exposure/disease relationship
    - Want to control for confounding
- Need to collect information on potential confounders – or at least known risk factors for outcome

WGS.151 Gender, Health, and Society
Spring 2016