

# Case-Control Studies

Jerrilyn Cambron, DC, PhD  
Professor, Department of Research

# Hierarchy of scientific evidence: observational



# Case-control studies

- Start with subjects who are:
  - Cases (with disease)
  - Controls (without disease)
- Look backwards in time (*retrospective*)
- Compare past histories of possible risk factors between cases and controls

# Case-control studies

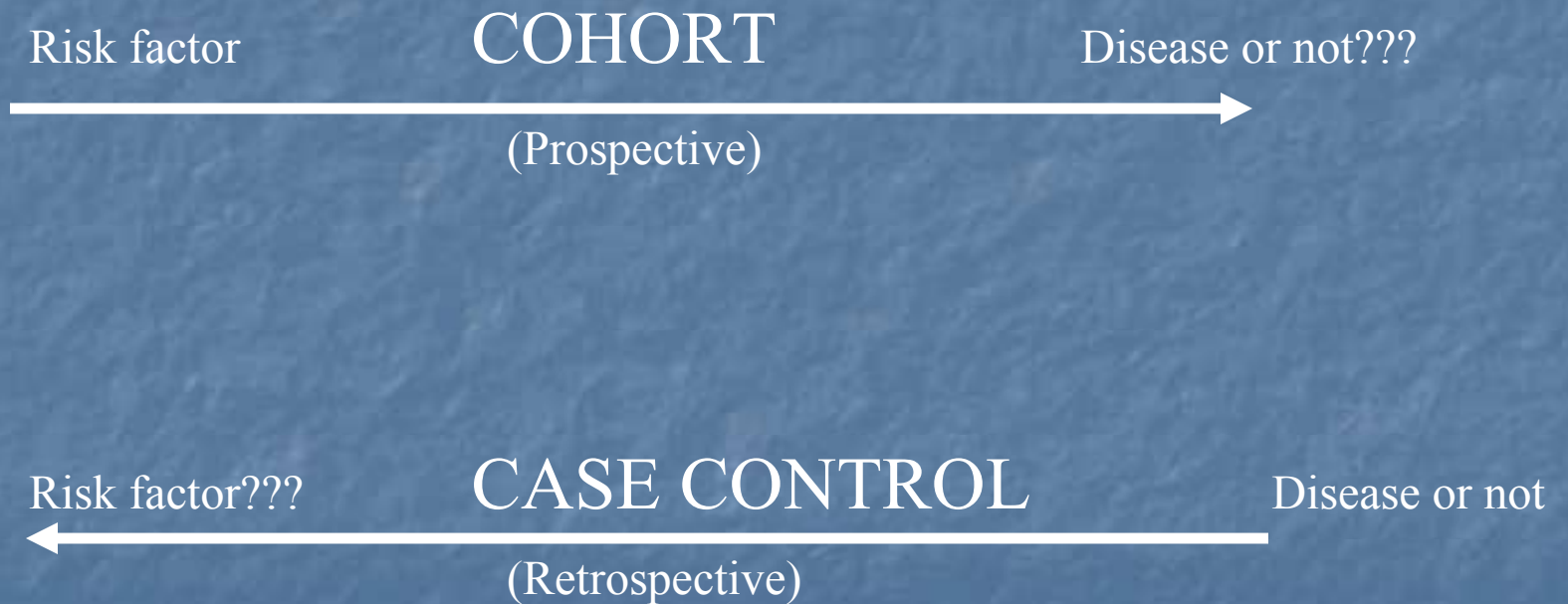
- Observational study starting with persons who have a disease (case) and who do not have the disease (control), and comparing histories of exposure to possible risk factors.
- The goal is to assess the relationship of potential risk factors to a specific disease.
- Easiest study design to assess *rare diseases*.

# Case-control vs. Cohort studies

- Calendar time is NOT the characteristic that distinguishes these two designs
- Rather, direction of calendar time is important



# Study timeline differences



# Nested case control study

- Study design where cases and controls are chosen from a population that was defined before the study began
- Still starts with cases and controls and looks back at risk factors
- Can be nested in a cohort study or general population as long as data were already prospectively being collected within the group

# Case selection

- Case definition must be very clear
- Incident (new) cases are most desirable so case can remember past exposure to risk factors
- Cases may come from hospitals, disease registries, doctor's offices, insurance records, surveillance, patient groups, etc.
- Ex: Cases might be all new cases of ovarian cancer diagnosed in Cook County 2000-2005 in women aged 21-79 years. We may look at Cook Co. area hospitals or tumor registries.



# Control selection

- Controls are people at risk of developing the disease but do not have the disease
- Selecting the controls is the key challenge for case control studies, and this is the major determinant of whether the conclusions are valid
- Multiple controls can be chosen for each case
- Ex: For a study on ovarian cancer diagnosed in Cook County, controls might be selected from gynecologists' offices in Cook Co. To be eligible, women must have healthy ovaries and be 21-79 years old.

# “Sampling controls with replacement”

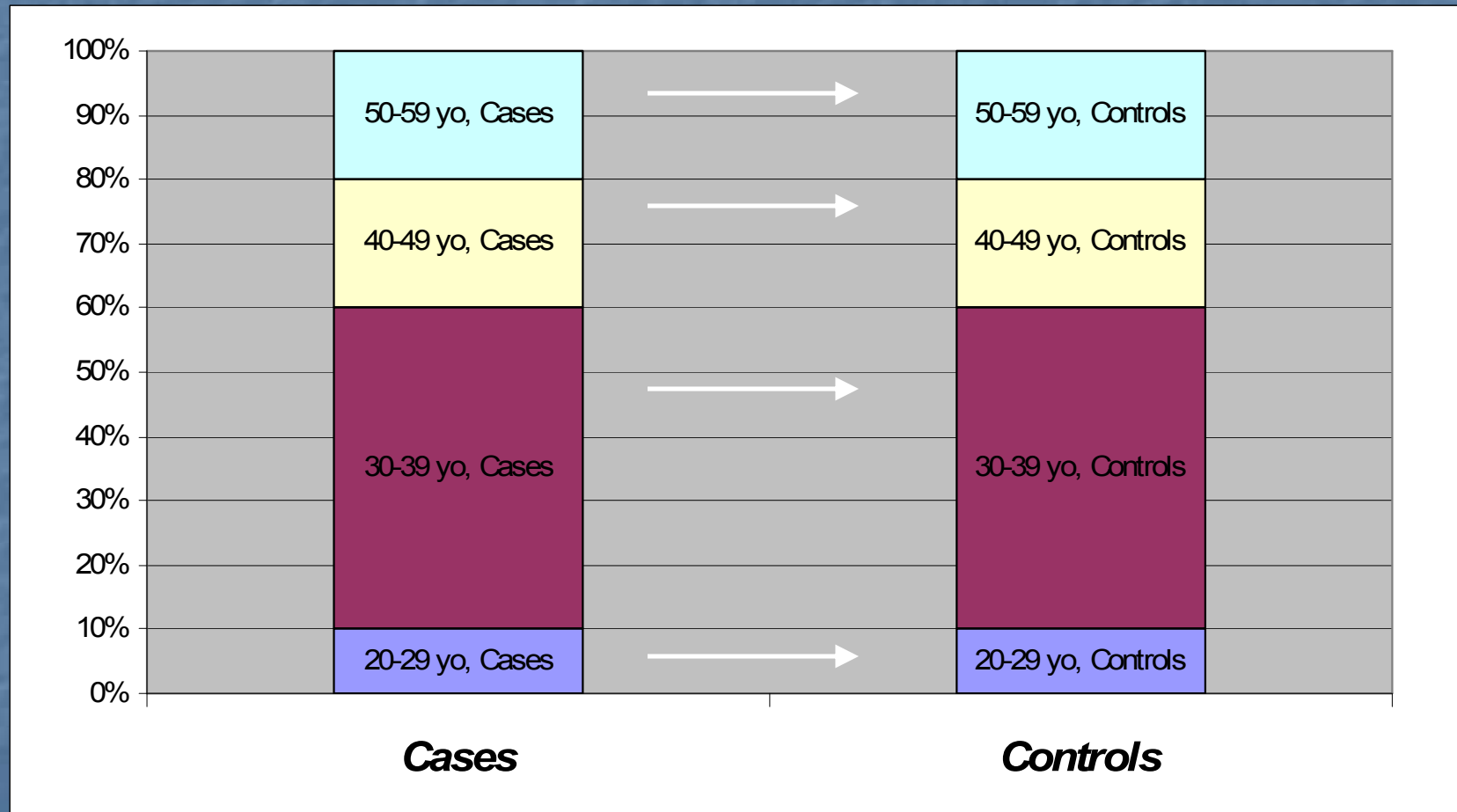
- This means that if a chosen ‘control’ subject is not eligible or interested in participating, the investigators replace that control with another subject

# Selecting controls

## 1. Frequency (group) matching

- Distribution of cases and controls are similar on a known confounding factor
- Ex: If 20% of cases are 60-64 years old, then 20% of controls are 60-64 years old
- Commonly used
  
- Advantages
  - Eliminates that variable as a confounder
  - More efficient use of data
- Disadvantages
  - Need to find cases before controls
  - Cannot assess association of matched variable with disease

# Frequency matching: Age of cases/controls



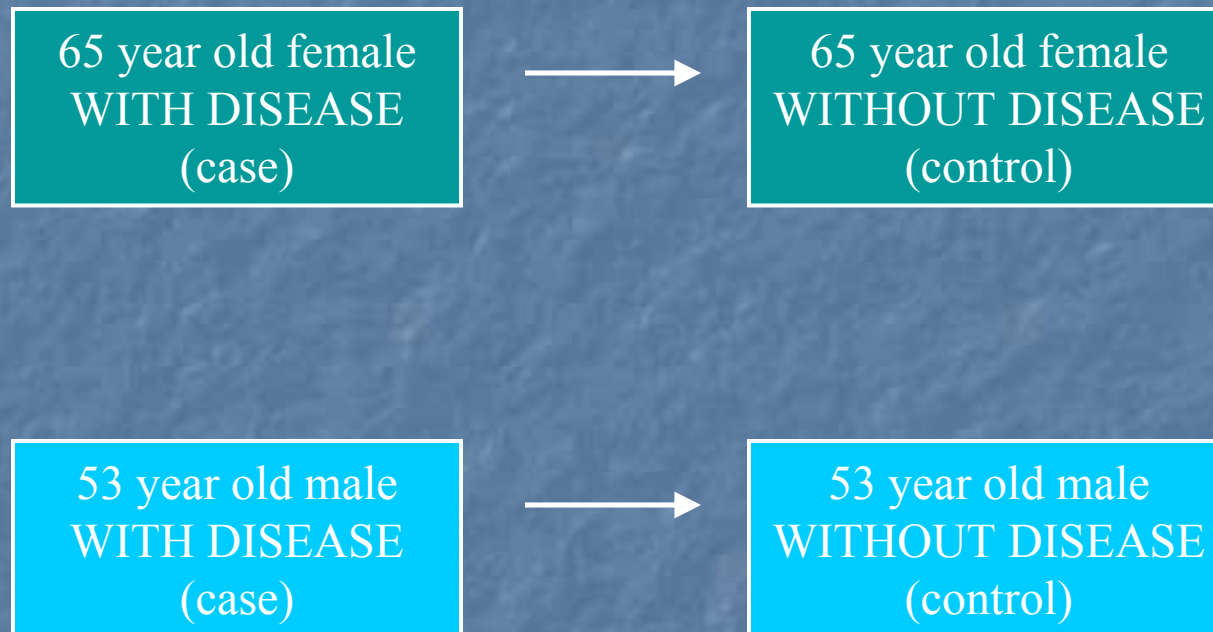
# Selecting controls

## 2. Pair (individual) matching

- Matches each case to a control on one or more confounding factors
- Ex: If a 45 year old female is a case, then a 45 year old female must be chosen as the matching control
  
- Advantages
  - Eliminates these variables as confounding factors
- Disadvantages
  - Can be more expensive and time consuming to find a match for each case
  - May not be able to find a match for each case and therefore lose information on that case



# Pair matching: Age and gender



# Selecting controls

## 3. No matching

- Investigators do not match on any factors
- Ex: If a 45 year old female is a case, then anyone can be chosen as the control such as a 90 year old male or 18 year old female
  
- Advantages
  - Easy and NOT time consuming to complete
  - Can have a control for every case
- Disadvantages
  - Groups may be uneven on certain factors leading to bias and 'garbage' results

# Sources of bias in case control studies

1. Recall bias- Cases may (try to) remember information better than controls

This is why prospective cohort studies are a stronger study design than retrospective case control.

Ex: In a study on tongue cancer, cases may 'remember' more details about chewing tobacco use in adolescence than the associated controls.

# Sources of bias in case control studies

2. Non-response and refusal bias- Cases and/or controls who refuse to participate may somehow be different than those who do participate

Ex: In a study on HIV/AIDS, cases may not want to participate if they have to admit that they have used injectable drugs or had unprotected sex.

# Sources of bias in case control studies

3. Selection of controls- Controls may not match the exposure within the general population

## Ex: Pancreatic cancer and coffee drinking

- Cases with pancreatic cancer were admitted to a hospital
- Hospital controls were used, including large number of controls admitted for gastrointestinal diseases
- Persons with GI diseases were less likely to be coffee drinkers than population at large causing a biased comparison



# Sources of bias in case control studies

4. Misclassification of exposure or disease- A mistake in classification of exposure or disease can cause spurious end results.
5. Confounding factors- Other measured or unmeasured factors may affect the outcome.

# Advantages of case control study

1. Well suited to rare diseases or those with long latency
2. Short time required to conduct
3. Relatively inexpensive
4. Requires relatively few subjects
5. Existing records can be used
6. No additional risk to subjects
7. Can look at multiple exposures for disease

# Disadvantages of case control study

1. Recalled information or existing records determine exposure; validation of information is difficult
2. Can study only one disease per study
3. Sometimes unable to control for extraneous variables
4. Difficult to select controls; selection bias
5. Difficult to study timing of events
6. Cannot calculate direct measure of risk (incidence)

# Analysis of case control studies

- *Odds ratio (OR)*- Very similar to relative risk
- Interpreted the same as relative risk
- RR not used because cannot calculate risk (incidence) in retrospective study

# Multivariate analyses

- An analytic method that allows the simultaneous study of two or more dependent variables (such as risk factors in a case control study)
- When multiple variables are included in an analysis at the same time, you are essentially 'controlling' for those variables or 'adjusting' the rates based on those variables



# For example...

*Table 4 Multivariate analysis*

Variable	All dissections, n = 51	
	OR (95% CI)	<i>p</i>
Pain before stroke/TIA	4.6 (2.1–10)	<0.001
SMT within 30 d	NS	
Illness within 30 d	2.3 (1.0–5.1)	0.042
Alcohol, current	2.7 (1.1–6.2)	0.023

# Questions to ask when reading a case control study

1. Were the criteria for cases adequately described? For controls?
2. Were cases and controls from the same population? Were they comparable?
3. Were risk factors adequately defined and reasonably measured?
4. Was there a potential for recall bias?
5. Were there any potentially associated factors that were not measured or accounted for?

Rothwell DM, Bondy SJ, Williams JI, Bousser MG. Chiropractic Manipulation and Stroke: A Population-Based Case-Control Study. *Stroke* 2001;32:1054-1060.

Smith WS, Johnston SC, Skalabrin EJ,  
Weaver M, Azari P, Albers GW, Gress  
DR. Spinal manipulative therapy is an  
independent risk factor for vertebral  
artery dissection. *Neurology*  
2003;60:1424–1428

Cassidy JD, Boyle E, Cote P, He Y,  
Hogg-Johnson S, Silver FL, Bondy SJ.  
Risk of Vertebrobasilar Stroke and  
Chiropractic Care: Results of a  
Population-Based Case-Control and  
Case-Crossover Study. Spine  
2008;33(4S):S176–S183



# Hierarchy of scientific evidence



# Any questions?