

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2008, The Johns Hopkins University and Sukon Kanchanaraksa. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.



JOHNS HOPKINS  
BLOOMBERG  
SCHOOL *of* PUBLIC HEALTH

## *Case-Control Studies*

---

Sukon Kanchanaraksa, PhD  
Johns Hopkins University



JOHNS HOPKINS  
BLOOMBERG  
SCHOOL *of* PUBLIC HEALTH

## *Section A*

---

Distinctions of a Case Control Study

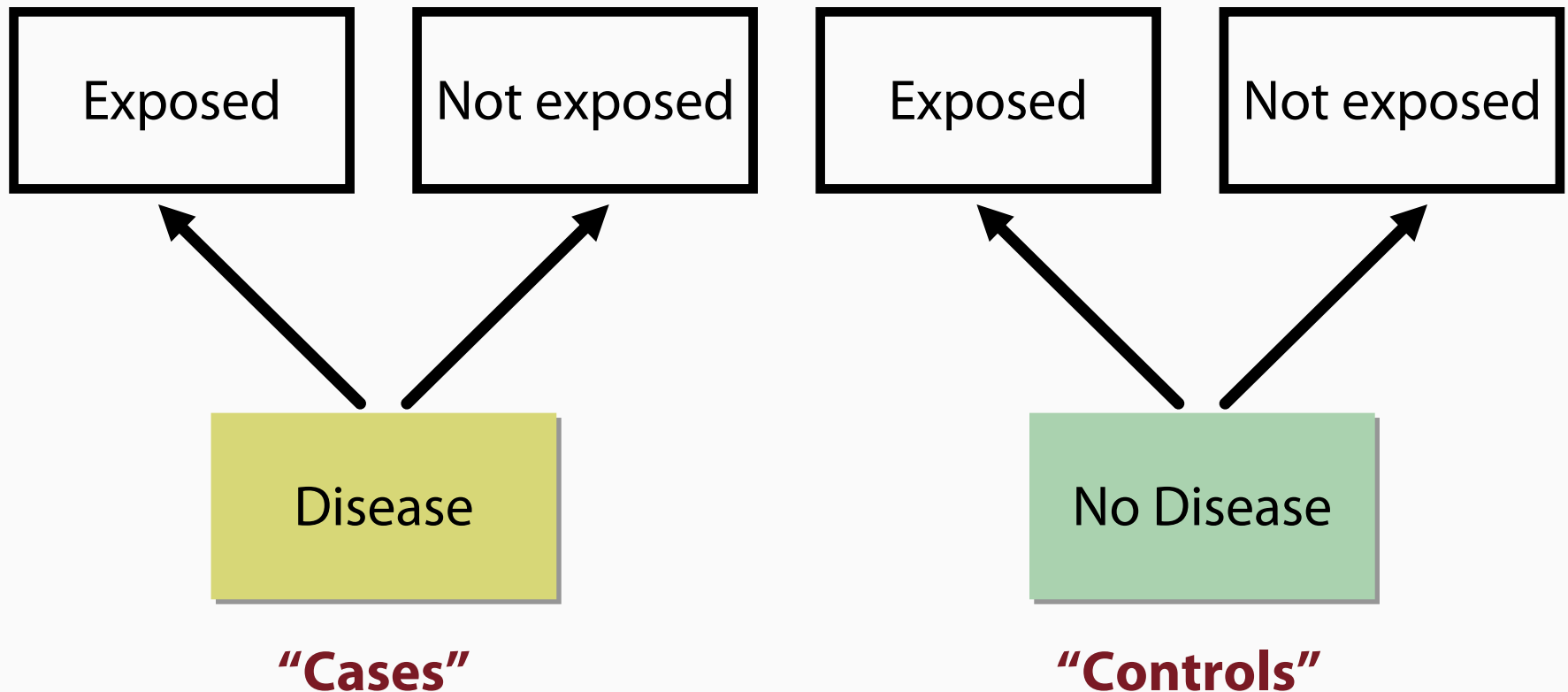
# *One of the Primary Goals of Epidemiological Investigation*

Exposure



Disease

# Designing a Case-Control Study



# Case-Control Study

First, select

	Cases (with disease)	Controls (without disease)
Totals	$a + c$	$b + d$

# Case-Control Study

First, select

Then,  
measure  
past  
exposure

	Cases (with disease)	Controls (without disease)
Were exposed	a	b
Were not exposed	c	d
Totals	a + c	b + d

# Case-Control Study

First, select

Then,  
measure  
past  
exposure

	Cases (with disease)	Controls (without disease)
Were exposed	a	b
Were not exposed	c	d
Totals	a + c	b + d

Proportion exposed  $\frac{a}{a + c}$   $\frac{b}{b + d}$



# Case-Control Study: Example of CHD and Smoking

Total

Cases CHD	Controls (without disease)
200	400

# Case-Control Study: Example

	Cases CHD	Controls (without disease)
Smoked cigarettes	112	176
Did not smoke cigarettes	88	224
Total	200	400

% Smoking cigarettes       $\frac{112}{200} = 56\%$        $\frac{176}{400} = 44\%$

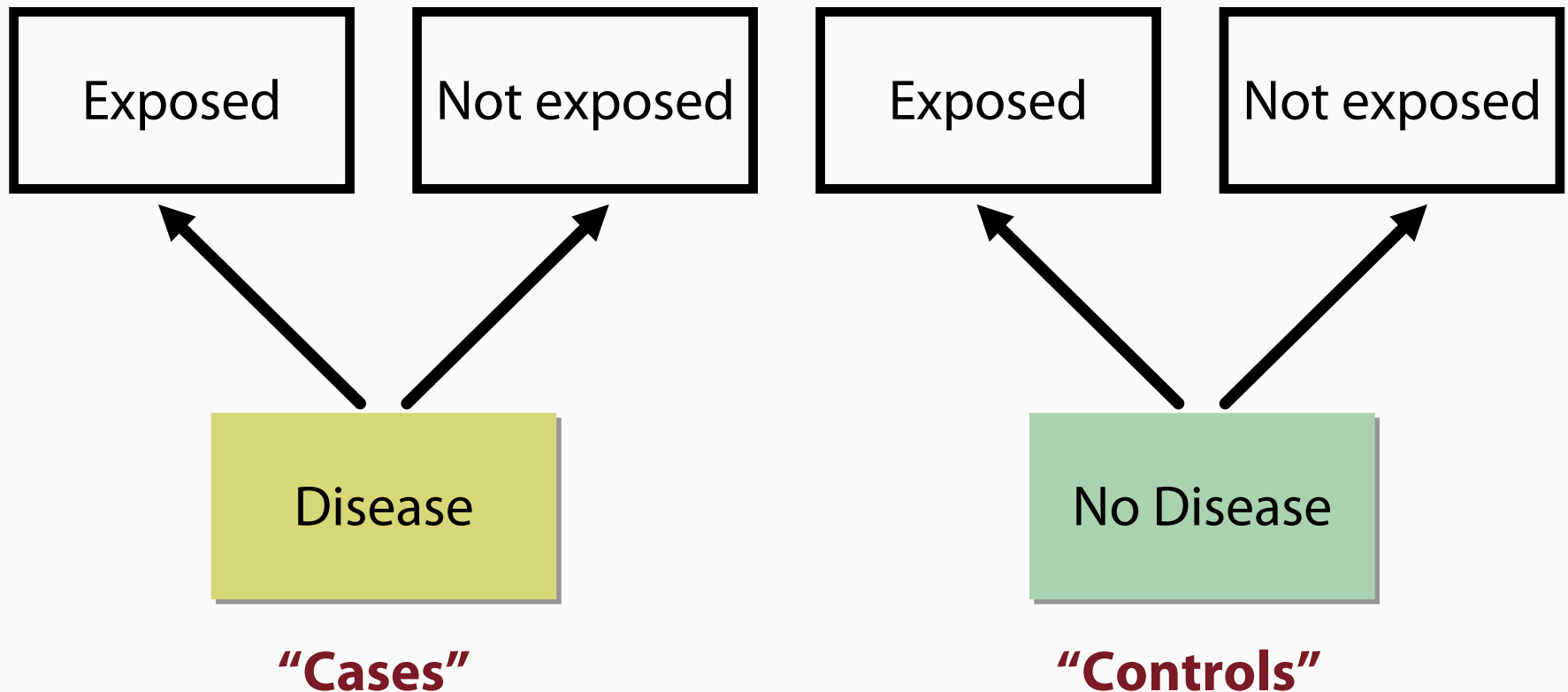
# Case-Control Study: Example

	Cases CHD	Controls A	Controls B
Smoked cigarettes	112	176	352
Did not smoke cigarettes	88	224	448
Total	200	400	800

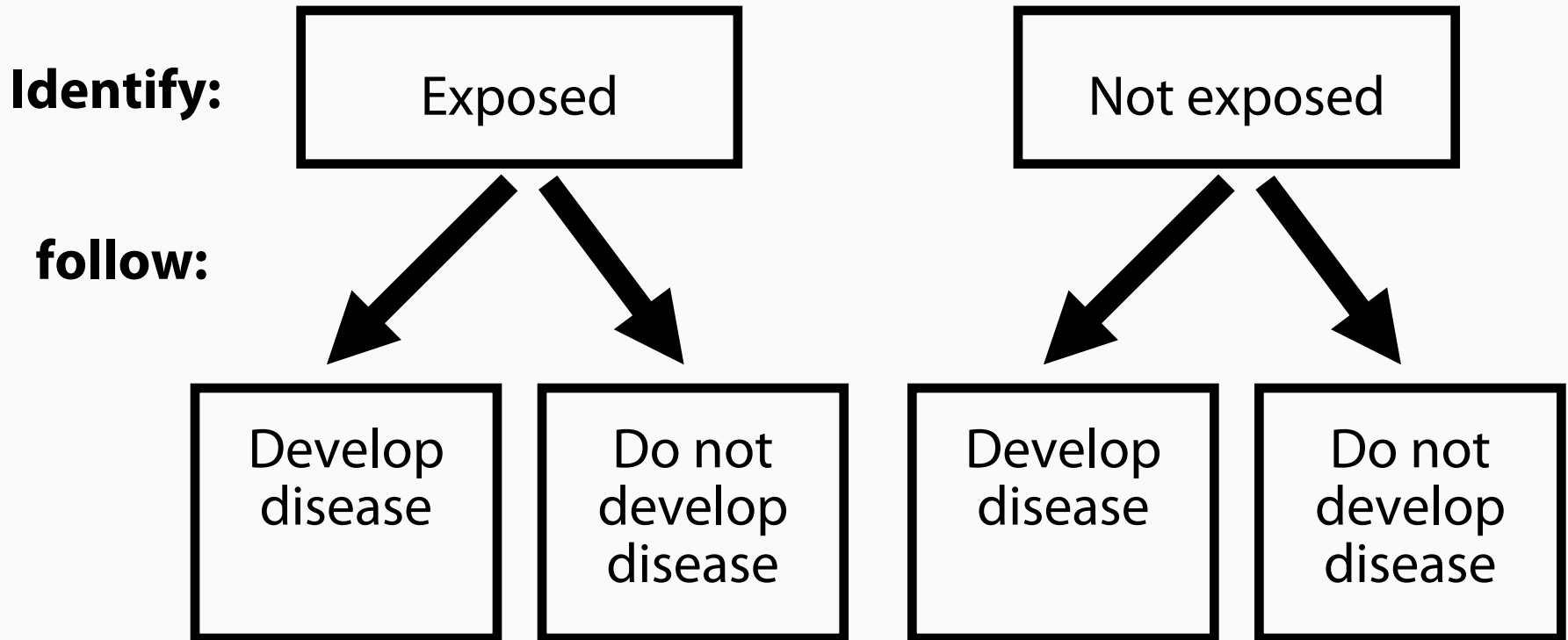
2 controls per case

4 controls per case

# Designing a Case-Control Study



# Designing a Cohort Study



# Example: Case-Control Study of Uranium Miners and Lung Cancer in Navajo Men

		Lung Cancer	
		Yes	No
Uranium Mining	Yes	23	0
	No	9	64
		32	64
% Exposed to Uranium		$\frac{23}{32} = 72\%$	$\frac{0}{64} = 0\%$

# Example: Case-Control Study of Lung Cancer and Tobacco Smoking

		Lung Cancer	
		Case	Control
Tobacco Smoking	Yes	597	666
	No	8	114
		605	780
% Exposed to Tobacco		$\frac{597}{605} = 99\%$	$\frac{666}{780} = 85\%$

# Case-Control Study: Example

**Distribution of 1465 Lung Cancer Patients and a Control Group According to Average Number of Cigarettes Smoked Daily Over the 10 Years Preceding Onset of the Present Illness**

Daily Average Cigarette	Males		Females	
	Lung Cancer Patients	Control Group	Lung Cancer Patients	Control Group
0	7	61	40	59
1-4	55	129	16	25
5-14	489	570	24	18
15-24	475	431	14	6
25-49	293	154	14	0
50+	38	12	0	0
Total	1357	1357	108	108



# *When Is a Case-Control Study Warranted?*

- A case-control study is usually conducted before a cohort or an experimental study to identify the possible etiology of the disease
  - It costs relatively less and can be conducted in a shorter time
- For a given disease, a case-control study can investigate multiple exposures (when the real exposure is not known)
- A case-control study is preferred when the disease is rare because investigators can intentionally search for the cases
  - A cohort study of rare disease would need to start with a large number of exposed people to get adequate number of cases

## *Issues in Selection of Cases*

- Cases should be homogeneous
  - Criteria or definition of cases must be well formulated and documented
  - If diagnostic tests are used to identify cases:
    - ▶ High-sensitivity tests (same as broad criteria or definition) will yield a higher number of false positives
    - ▶ Low-sensitivity tests (same as restrictive criteria or case definition), and thus high specificity, will result in a lower number of false positives
  - A mild form of disease may also include higher false positives than a severe form of disease

## *Issues in Selection of Cases*

- Incident cases (newly diagnosed) may be harder to come by, but they are closer in time to exposures and, unlike prevalent cases, they are not likely the survivors of the disease
- If cases are misclassified (include false positives), the findings may be false

## *Issues in Selection of Controls*

- Conceptually, controls should come from the same **population at risk of disease** from which cases develop
- But practically, controls are often selected to be similar to cases on key factors but without the disease—because it is difficult to define the population at risk of disease
- Different types of controls may be used, and they have different limitations

# Types of Controls

- Hospital controls
  - Have similar quality of information and are convenient to select, but they may have characteristics or diseases that led to hospitalization
- Dead controls
  - If cases are dead, information of past exposures will be given by surrogates, such as spouse or children
  - Dead controls share the same limitation
- Best friend or neighbor controls
  - May share similar characteristics (too similar?)
- Population controls
  - Random digit dialing (RDD) is often used

# *Selection of Cases and Controls in Macmahon's Case-Control Study of Pancreatic Cancer*

- Cases
  - Patients with histological diagnoses of pancreatic cancer in 11 Boston and Rhode Island hospitals from October, 1974, to August, 1979
- Controls
  - All other patients who were under the care of the same physician in the same hospital at the time of the interview of a patient with pancreatic cancer

# *Estimated Relative Risk of Pancreatic Cancer by Gender and Coffee Drinking Habits*

	Coffee Drinking (Cups/Day)			
	0	1–2	3–4	>5
Men	1.0	2.6	2.3	2.6
Women	1.0	1.6	3.3	3.1
Both sexes	1.0	2.1	2.8	3.2

# Estimated Relative Risk of Pancreatic Cancer by Cigarettes and Coffee Drinking

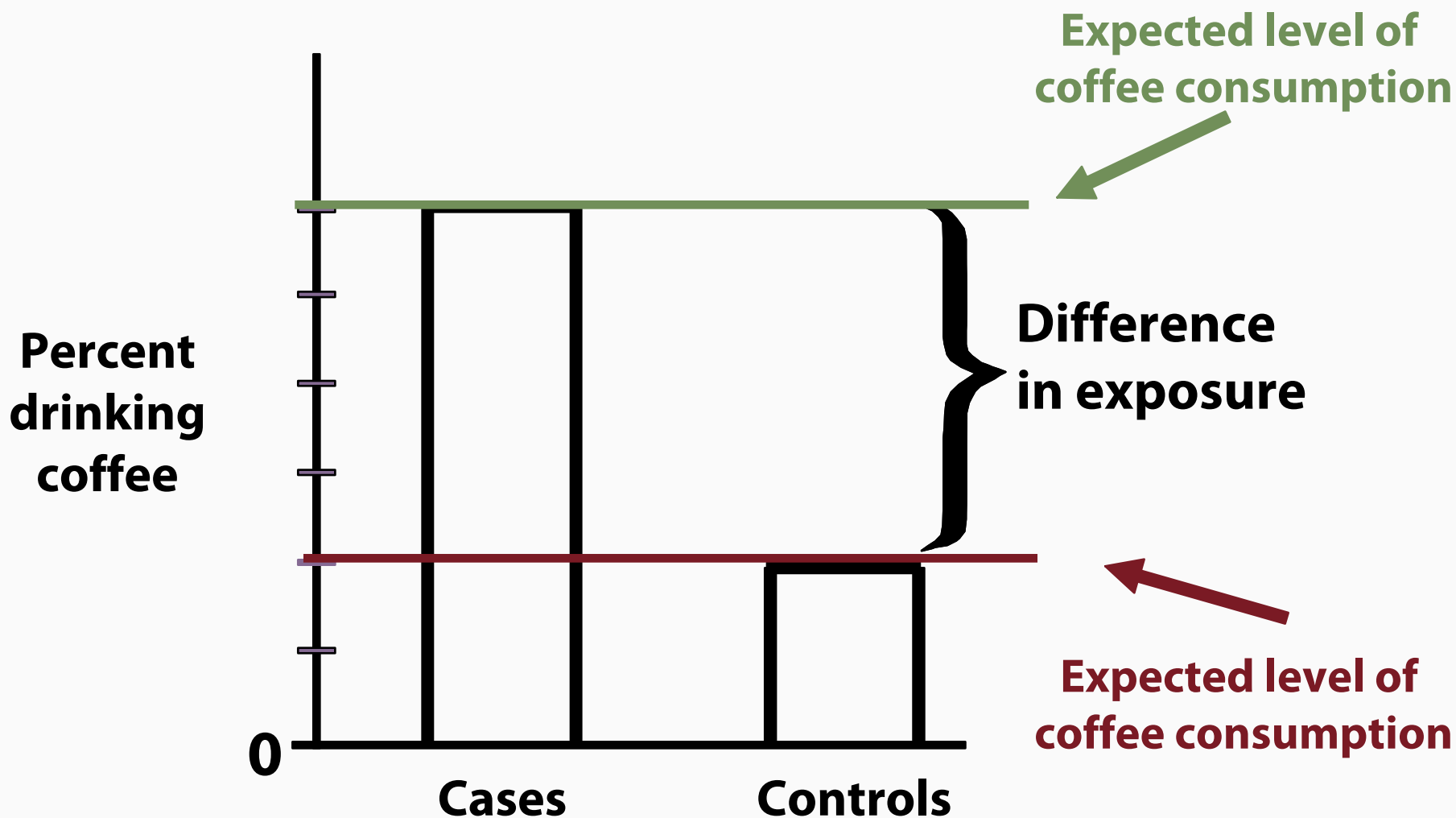
Cigarette Smoking	Coffee Drinking (Cups/Day)			Total
	0	1–2	>3	
Never	1.0	2.1	3.1	1.0
Ex-Smokers	1.3	4.0	3.0	1.3
Current Smokers	1.2	2.2	4.6	1.2
<b>Total</b>	<b>1.0</b>	<b>1.8</b>	<b>2.7</b>	



# *Selection of Controls in MacMahon's Case-Control Study of Pancreatic Cancer*

- Included
  - All other patients who were under the care of the same physician in the same hospital at the time of the interview of a patient with pancreatic cancer
- Excluded
  - Patients with pancreatic or hepatobiliary tract diseases known to be associated with smoking or alcohol consumption
  - Patients with cardiovascular disease, diabetes, respiratory or bladder cancer, and peptic ulcer
  - Non-whites; older than 79 years

# Interpreting the Results of a Case-Control Study of Coffee and Pancreatic Cancer





JOHNS HOPKINS  
BLOOMBERG  
SCHOOL *of* PUBLIC HEALTH

## *Section B*

---

Other Issues Involved with Case Control Studies

# Matching

- **Matching** is the process of selecting controls in a case-control study so that the controls are similar to the cases with regard to certain key characteristics—such as age, sex, and race
- Matching can be performed at an individual or group level
  - **Individual matching** (matched pairs)
  - **Group matching** (frequency matching)

## Group Matching: Example

- In a study of breast cancer and reproductive risk factors from the New Mexico Women's Health Study:

**“Controls were ascertained using random digit dialing with frequency matching on ethnicity, the three age groups (30–39, 40–64, and 65–74), and the seven health planning districts.”**

## *Problems with Matching*

- Matching on many variables may make it difficult to find an appropriate control
- We cannot explore the possible association of the disease with any variable on which the cases and controls have been matched

## *Recall of Information in a Case-Control Study*

- Information on some past exposures depends on memory of events from both cases and controls
  - Often inadequate or limited
- **Recall bias** occurs when the recall is better among cases than controls because of the presence of the disease
  - Consequently, a false association may be found between exposure and disease

# *An Artificial Association Resulting from Recall Bias*

A Study of Maternal Infections During Pregnancy and Congenital Malformations		
	Cases (Congenital Malformations)	Controls (No Malformations)
True incidence of infection	15%	15%
Percent of infections recalled	90%	10%
Infection rate as ascertained by interview	13.5%	1.5%



# Multiple Control Groups

- Because of the several limitations in the selection of controls, the use of **multiple control groups** may address some of the concerns
  - Use both living controls and dead controls
    - ▶ The use of surrogates to provide data
  - Hospital controls and community controls
    - ▶ Hospital controls may have some conditions that lead to frequent hospital visits
  - Non-disease controls and cancer controls
    - ▶ Recall of past exposure differs with outcome
- If findings are in agreement between groups, then they are likely to be valid

## Multiple Control Groups: Example

- In a case-control study of measles and subacute sclerosing panencephalitis (SSPE), two control groups were used
  - Playmates of the patients
  - Hospitalized patients in the same hospital
- Both matched on age, gender, and race to cases

## *Nested Case-Control Study*

- A **nested case-control study** is conducted within a defined cohort in which exposure data and population characteristics are available to some extent, often from the time of enrollment into the cohort

## *Nested Case-Control Study: Advantages*

- Both cases and controls are from a known, defined population at risk of disease
- Exposure data are collected prior to the diagnosis of the disease
- Information collected has not been influenced by the knowledge of disease status (no interviewer/information bias)
- Is less costly than a cohort study because fewer subjects and fewer tests and/or specimens are required

## *Nested Case-Control Study: Example*

- In a study to determine whether the *helicobacter pylori* infection was associated with the development of gastric cancer, 189 patients with gastric cancer and 189 cancer-free individuals were identified from 130,000 people from a HMO cohort followed since the 1960s
- *Helicobacter pylori* infection status was determined using **stored serum collected during the 1960s**
- The proportions of infection were 84% among the gastric cancer cases and 61% among the controls

# Review

- Differentiate between a retrospective cohort study and a case-control study
- Define
  - Matching
  - Controls
  - Nested case-control study