

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

Cohort Studies

Sukon Kanchanaraksa, PhD
Johns Hopkins University

Design of a Cohort Study

Identify:

Exposed

Not exposed

follow:

Develop disease

Do not develop disease

Develop disease

Do not develop disease

Cohort Study

First,
identify

		Totals		
Exposed			$a + b$	
Not exposed			$c + d$	

Cohort Study

Then, follow to see whether

First,
identify

	Disease develops	Disease does not develop	Totals	
Exposed	a	b	a + b	
Not exposed	c	d	c + d	

Cohort Study

Then, follow to see whether

Calculate
and compare

First,
identify

	Disease develops	Disease does not develop	Totals	Incidence of disease
Exposed	a	b	a + b	$\frac{a}{a + b}$
Not exposed	c	d	c + d	$\frac{c}{c + d}$

$$\frac{a}{a + b} = \text{Incidence in exposed} \quad \frac{c}{c + d} = \text{Incidence in not exposed}$$

Cohort Study

Then, follow to see whether

Calculate

First, select

	Develop CHD	Do not develop CHD	Totals	Incidence of disease
Smoke cigarettes	84	2916	3000	$\frac{84}{3000}$
Do not smoke cigarettes	87	4913	5000	$\frac{87}{5000}$

$$\frac{84}{3000} = 0.028 = \text{Incidence in 'smoke cigarettes'}$$

$$\frac{87}{5000} = 0.0174 = \text{Incidence in 'not smoke cigarettes'}$$

Design of a Cohort Study

Begin with:

Defined population

Non-randomized

Identify :

Exposed

Not exposed

follow:

Develop
disease

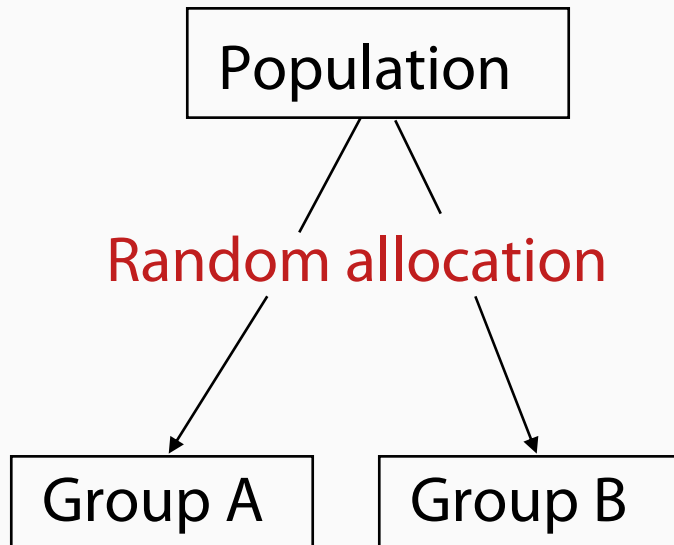
Do not
develop
disease

Develop
disease

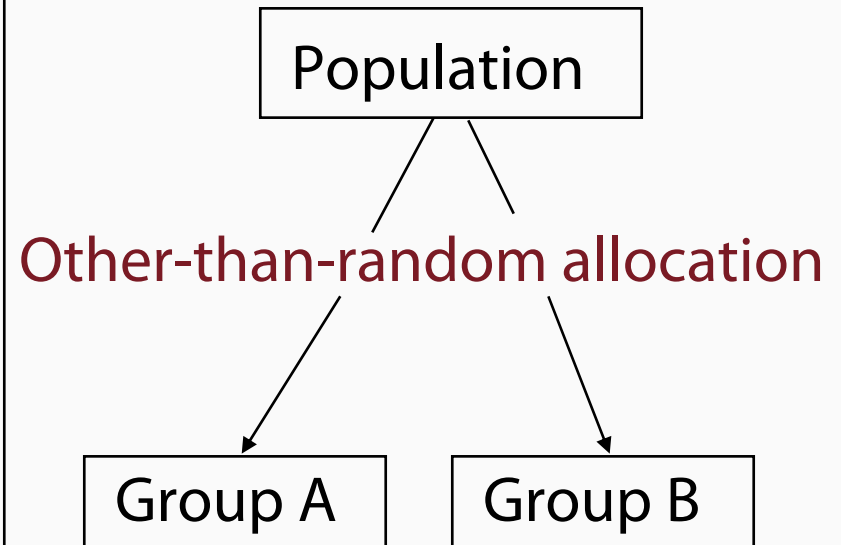
Do not
develop
disease

Comparison of Experimental vs. Observational Study

Experimental
(Randomized trial)

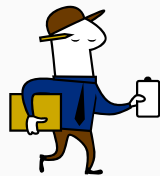


Observational
(Cohort study)



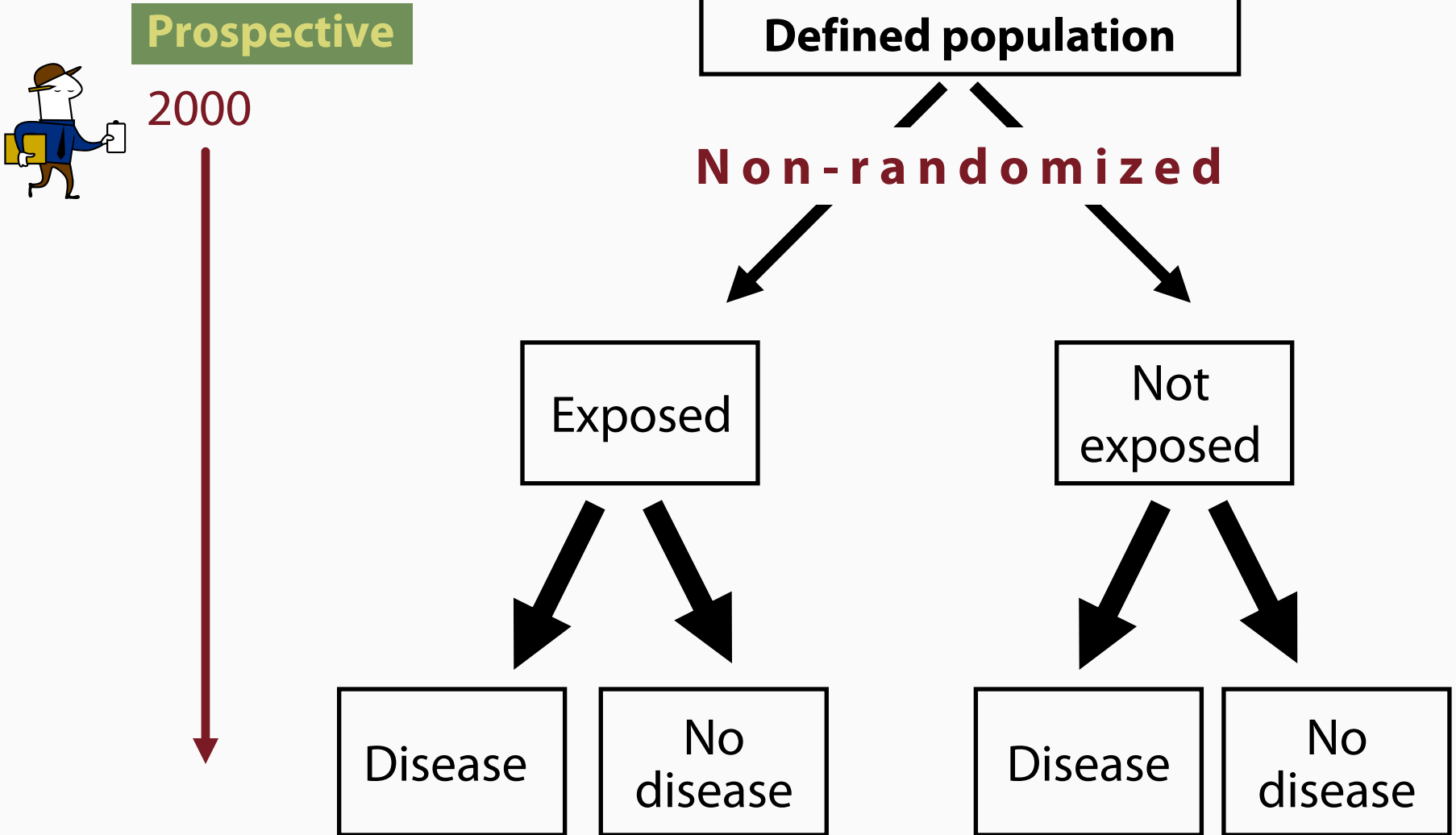
Types of Cohort Studies

- **Prospective cohort study**
 - Concurrent cohort study or longitudinal study
- **Retrospective cohort study**
 - Non-concurrent cohort or historical cohort study

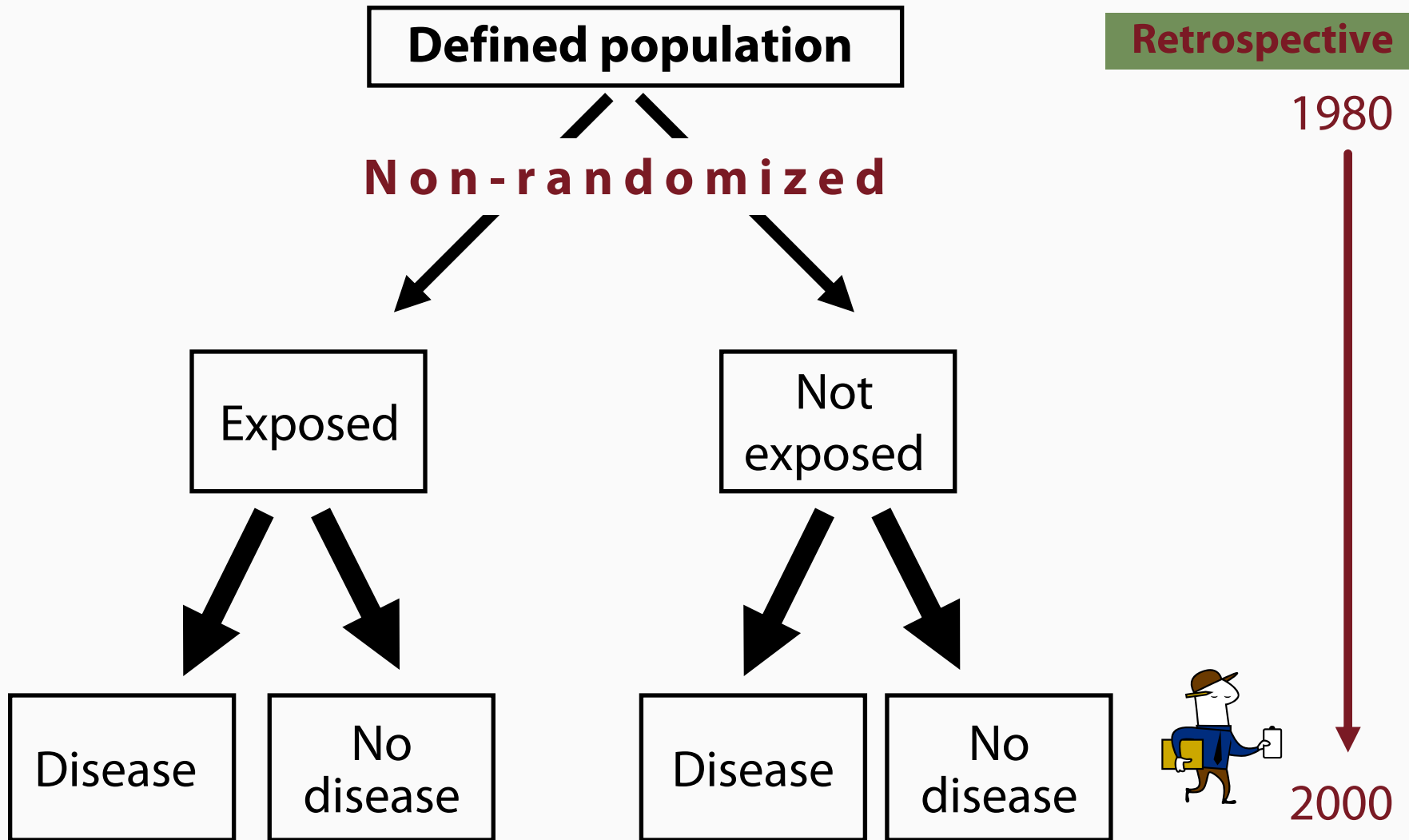


= Investigator

Time Frames for a Hypothetical Prospective Cohort Study Conducted in 2000



Time Frames for a Hypothetical Retrospective Cohort Study Conducted in 2000



Differentiating between Prospective and Retrospective

■ **Prospective cohort study**

— Investigator

- ▶ Starts the study (from the beginning) with the identification of the population and the exposure status (exposed/not exposed groups)
- ▶ Follows them (over time) for the development of disease
- ▶ Takes a relatively long time to complete the study (as long as the length of the study)

Differentiating between Prospective and Retrospective

■ **Retrospective cohort study**

- Investigator
 - ▶ Uses existing data collected in the past to identify the population and the exposure status (exposed/not exposed groups)
 - ▶ Determines at present the (development) status of disease
- Investigator spends a relatively short time to:
 - ▶ Assemble study population (and the exposed/not exposed groups) from past data
 - ▶ Determine disease status at the present time (no future follow-up)

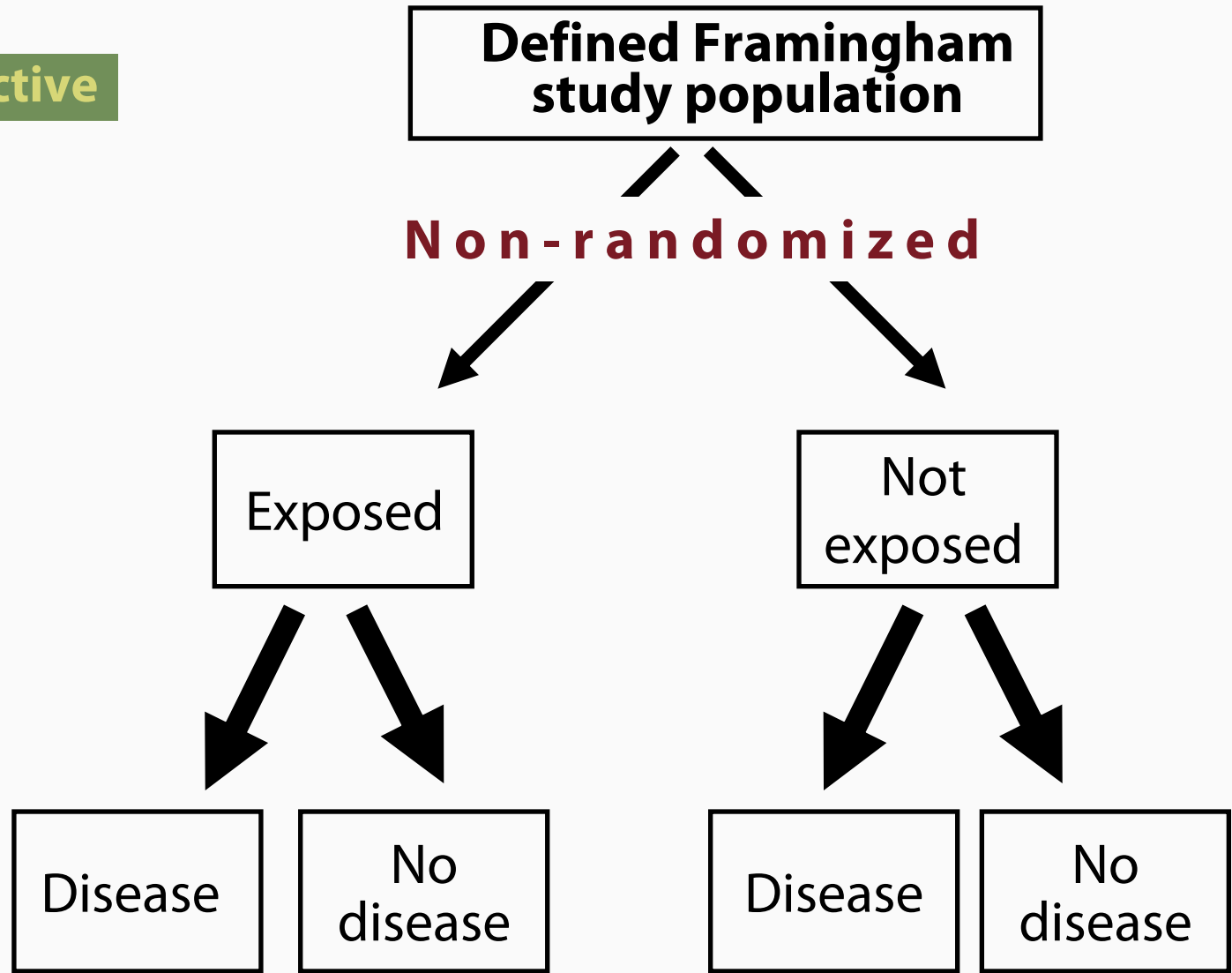
Combined Prospective and Retrospective Cohort Study

- Investigator uses existing data collected in the past to:
 - Identify the population and the exposure status (exposed/not exposed groups)
 - **Follow them into the future** for the development of the disease
- Investigator
 - Spends a relatively short time to assemble study population (and the exposed/not exposed groups) from past data
 - Will spend additional time following them into the future for the development of disease

Example of a Prospective Cohort Study: Framingham Study

Prospective

1948



Framingham Study

Objectives

To study the impact of several factors on incidence of cardiovascular diseases

Exposures

Blood pressure, smoking, body weight, diabetes, exercise, etc.

Multiple Outcomes

Coronary heart disease, stroke, congestive heart failure, peripheral arterial disease

Framingham Study as a Cohort Study

- The study started with a defined population
 - Investigators (USPHS and NHLBI) started by identifying a new population and did not use existing data to identify the population and the exposure groups
- There were several hypotheses to be tested
 - Different exposures and different outcomes
- For each exposure, investigators identified the “exposed” and the “not exposed” groups
- For each exposure, the participants were followed for the development of disease
- Different exposures were studied, as well as different diseases

Derivation of the Framingham Study Population

	Men	Women	Total
Random sample	3074	3433	6507
Respondents	2024	2445	4469
Volunteers	312	428	740
Respondents free of CHD*	1975	2418	4393
Volunteers free of CHD	307	427	734
Total Free of CHD	2282	2845	5127

* CHD = coronary heart disease

Follow-Up of Participants

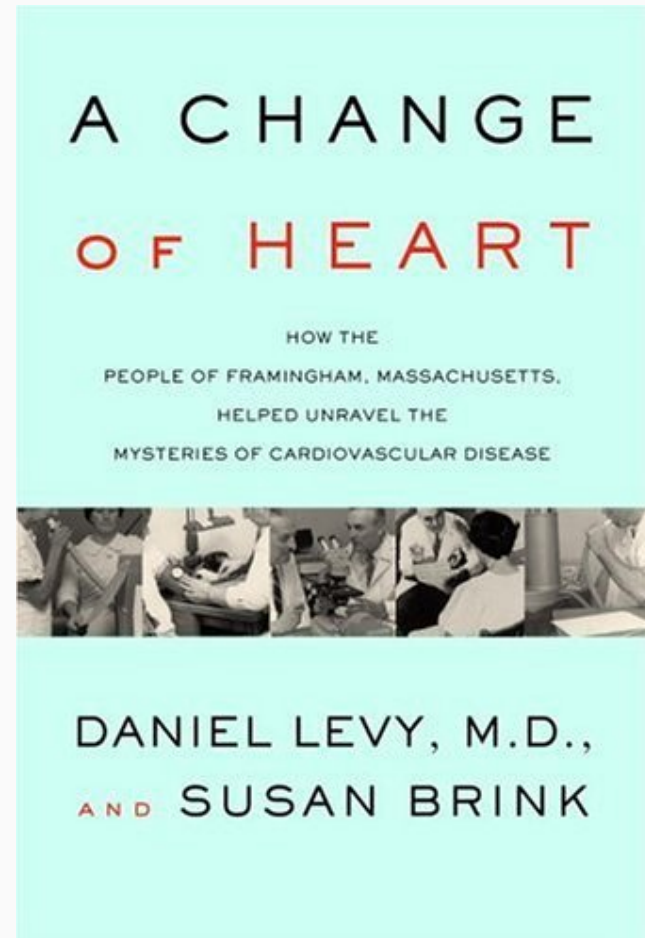
- Risk factors and the development of cardiovascular events were evaluated every two years by medical history, medical record review, and physical examination
- All diagnoses were verified without knowledge of risk factors by Framingham examiners who reviewed medical records and death certificates
- Approximately three percent of the subjects were lost to follow-up for mortality during the first 45 years of the study

Timeline of Milestones from the Framingham Study

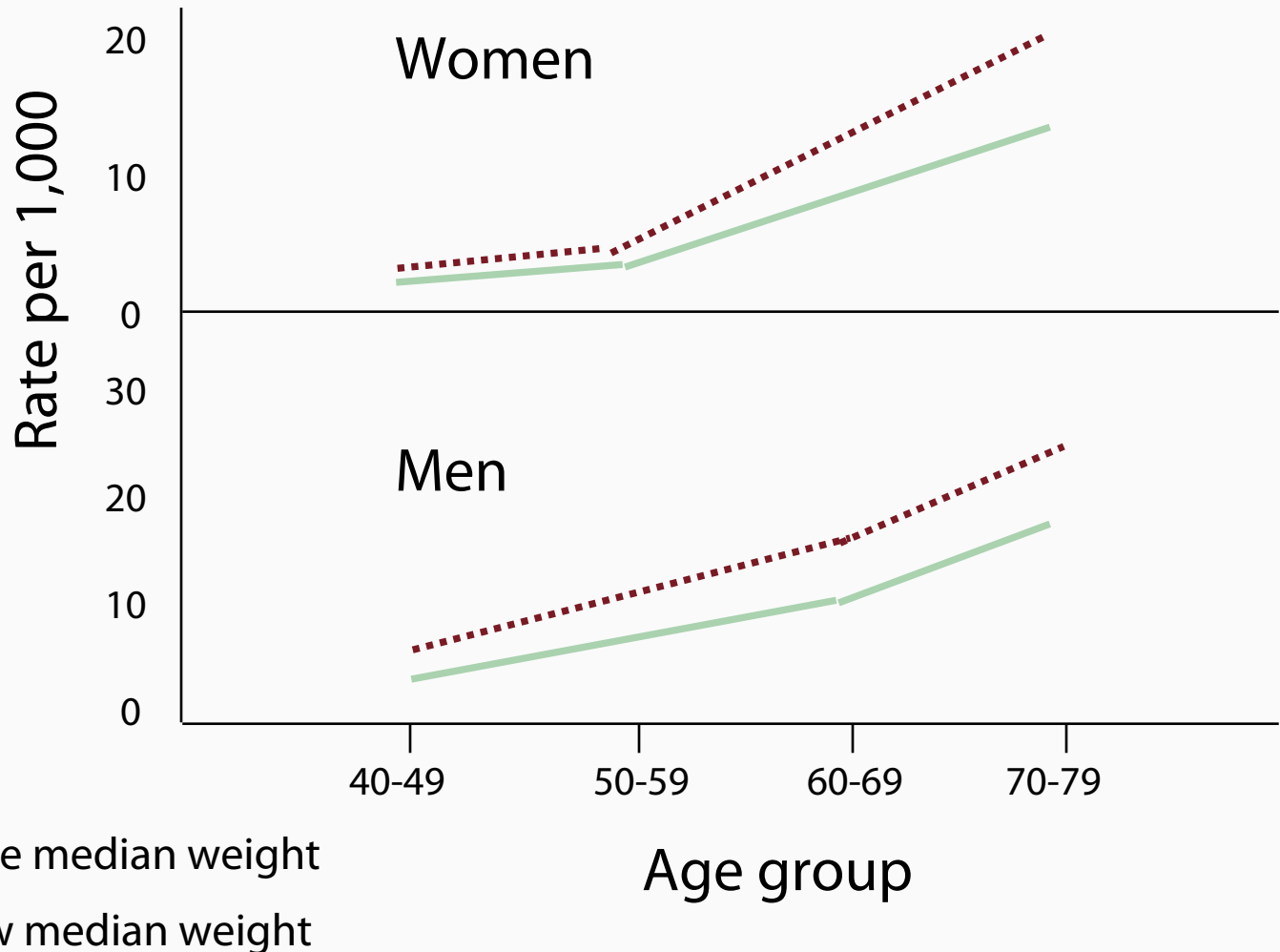
- **1948:** start of the Framingham Heart Study
- **1960:** cigarette smoking found to increase risk of heart disease
- **1961:** cholesterol, blood pressure, and ECG abnormalities found to increase risk of heart disease
- **1965:** first Framingham Heart Study report on stroke
- **1967:** physical activity found to reduce risk of heart disease; obesity to increase the risk
- **1970:** high blood pressure found to increase the risk of stroke
- **1974:** diabetes found to be associated with cardiovascular disease
- More milestones: www.framingham.com/heart/timeline.htm

Book

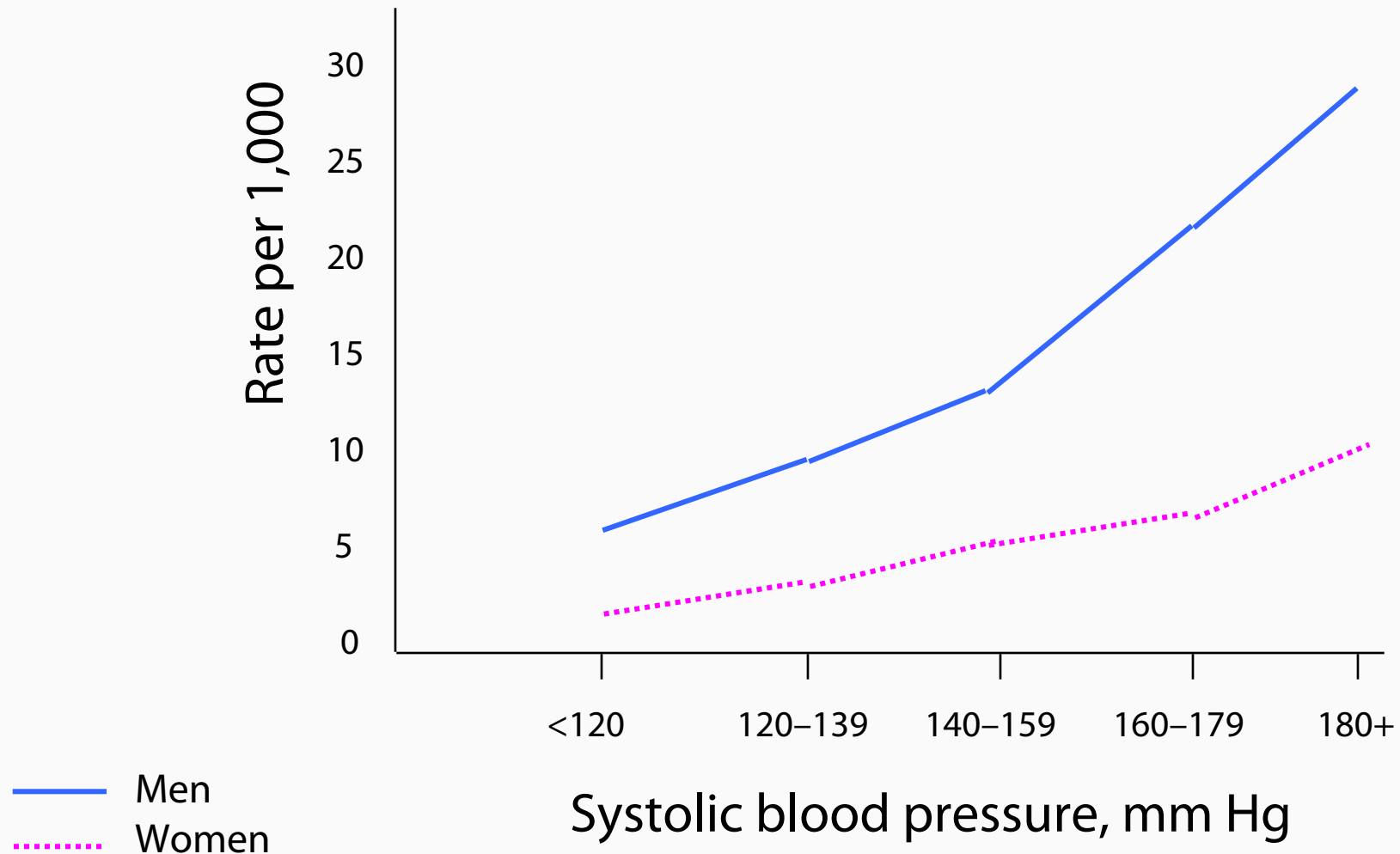
- *A Change of Heart: How the People of Framingham, Massachusetts, Helped Unravel the Mysteries of Cardiovascular Disease*



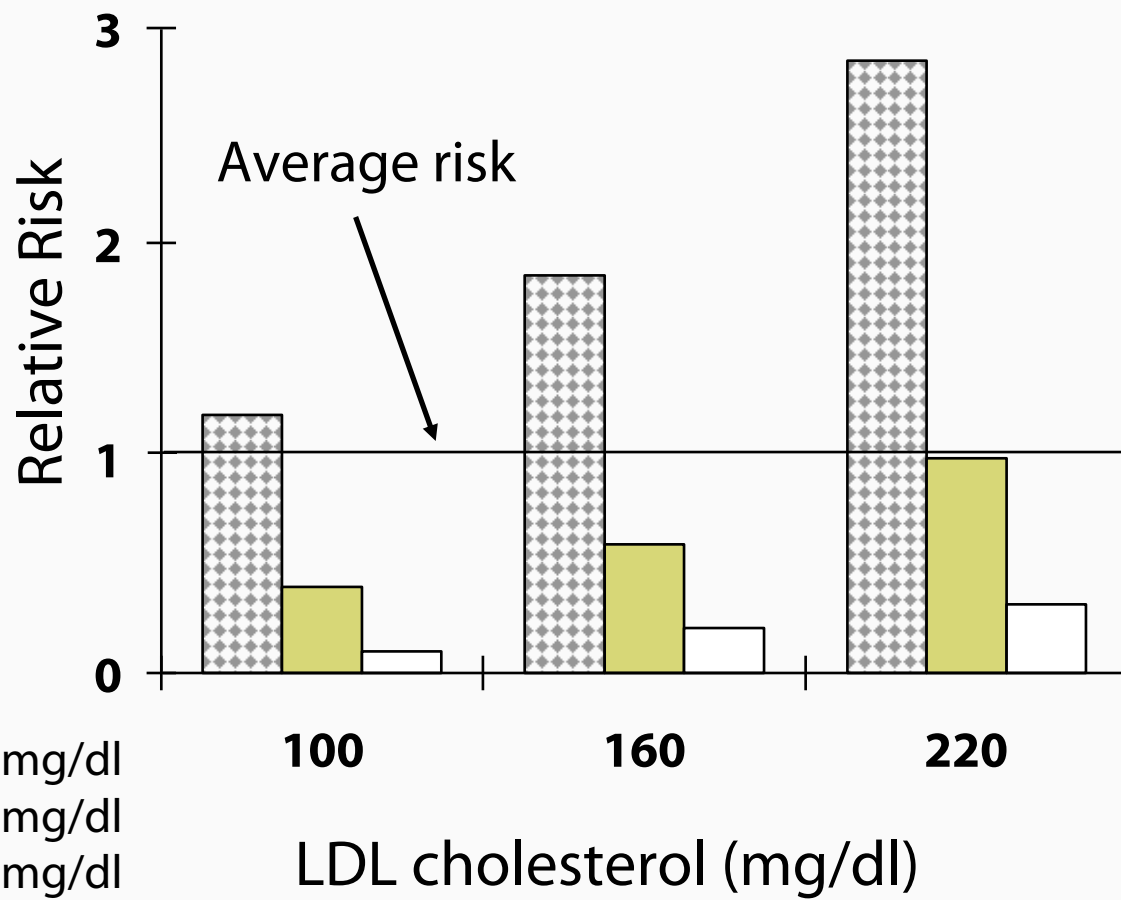
Average Annual Incidence of Coronary Heart Disease by Weight, Gender, and Age Group



Average Annual Incidence of Coronary Heart Disease by Systolic Blood Pressure



CHD Risk Assessment Based on Relationship Between HDL and LDL Cholesterol Men 50–70 Years



Types of Potential Bias in Cohort Studies

■ **Selection bias**

- Select participants into exposed and not exposed groups based on some characteristics that may affect the outcome

■ **Information bias**

- Collect different quality and extent of information from exposed and not exposed groups
- Loss to follow-up differs between exposed and not exposed (or between disease and no disease)

■ **Misclassification bias**

- Misclassify exposure status or disease status

When Is a Cohort Study Warranted?

- When the (alleged) exposure is known
- When exposure is rare and incidence of disease among exposed is high (even if the exposure is rare, determined investigators will identify exposed individuals)
- When the time between exposure and disease is relatively short
- When adequate funding is available
- When the investigator has a long life expectancy

Review

- What are the differences in the study design between prospective cohort study and retrospective cohort study?
- What are the differences in the study design between randomized clinical trial study and cohort study?
- Why is cohort study preferred for studying rare exposure?