

Room Assignments for Midterm Exam

Friday, September 25, 2015

10:00am – 12:00pm

**If your last name begins
with:**

Go to room:

A – B

W2030

C – N

Sommer Hall

O – T

Sheldon Hall

U – Z

W5030

Midterm exam

- **Friday, September 25**
- 10 am-12 noon
- Room assignments are posted
- You must complete the exam on your own. The exam is closed notes and books. You must adhere to the academic ethics code.
- **Bring a calculator.** You may not use the calculator on your cell phone or laptop.
- 20 multiple choice questions

Midterm exam, continued

- Exams are administered via Courseplus; you are required to bring a laptop to class in order to take the exam
- *Charge your laptop the night before the exam!*
- You will have a paper copy of the exam that you may use as scratch paper; this exam must be turned in at the end of the exam

Midterm exam, continued

- Exams are administered via Courseplus; you are required to bring a laptop to class in order to take the exam
- *Charge your laptop the night before the exam!*
- You will have a paper copy of the exam that you may use as scratch paper; this exam must be turned in at the end of the exam

Midterm exam, continued

- Bring to the exam:
 - Pen or pencil
 - Calculator (you are not allowed to use the calculator function on your cellphone or tablet.
You cannot use Excel or any other computer software.)
 - Charged laptop computer

Technical Considerations

- Be kind to our wireless network (and help the wireless network be kind to you)!!!
 - Leave tablets, cell phones in a locker or at home
 - OR-
 - Turn them completely off
 - OR-
 - Put them in airplane mode

Technical Considerations, continued

- Do not all log on to Courseplus immediately at 10am, but log on as needed in order to input your answers into Courseplus

Technical Considerations, continued

- Use the JHSPH secured wireless network (JHSPH WPA)
- Do not use the JHSPH Guest network (users on the Guest network are timed out frequently)

LAPTOP RECOMMENDATIONS

MANAGED PRINT SERVICES

MULTIMEDIA STUDIO

MYJHSPH PORTAL SYSTEM

TELECOMMUNICATIONS

WIRELESS NETWORK

IT HELP DESK

Office: W3014, Wolfe St.
Phone: 410-955-3781 (5-3781)

Hours: Monday - Friday
8:30am – 5:00pm
(Walk in and Phone in)

Home > Offices and Services > Information Technology > Wireless Network

Wireless Network

Information Technology provides a wireless network for laptops of JHSPH faculty, staff and students. The interface will enable you to access JHSPH networking resources such as e-mail, the Internet, and networked printers without connecting any cables or modems within the JHSPH environment. At the beginning of each semester there are special sessions for incoming students. All other laptop user will need to contact Client Services in W3014 to have their laptop set up to connect to the wireless network.

GUEST WIRELESS ACCESS

Johns Hopkins Bloomberg School of Public Health has expanded its wireless network to include complimentary guest access. The goal is to provide quick, easy, and convenient Internet access to all of the School's visitors and guests. This new service is an extension to our current network. We will still provide our secure and full-featured wireless access to authenticated Faculty, Staff, and Students of the school.

Midterm Exam

- During the exam:
 - The only application that should be open on your computer is an Internet browser, and the only tab that should be open on that browser is the examination in Courseplus

I ❤️ Epi

Midterm Exam

- During the exam:
 - The only application that should be open on your computer is an Internet browser, and the only tab that should be open on that browser is the examination in Courseplus

I ❤️ Epi



I ❤️ Epi

Midterm Exam

- During the exam:
 - The only application that should be open on your computer is an Internet browser, and the only tab that should be open on that browser is the examination in Courseplus
 - *Any failure to follow this rule will be considered an academic ethics violation*

Midterm Exam

- During the exam:
 - In addition to recording your answers in Courseplus, you should record your answers on the paper copy of the exam, as a precaution in case of technical difficulties
 - Should you experience technical difficulties (e.g., loss of power, loss of internet connectivity), you will need to re-enter your answers into Courseplus
 - *Only the answers recorded in Courseplus will be counted toward your grade*

Midterm Exam

- At the end of the exam:
 - All exam materials (including the paper copy of the exam) must be given to a proctor
 - **You must show a proctor the screen verifying that you have submitted your answers and completed the exam prior to leaving the testing room***

Jennifer Deal ▾

[Go to Faculty Tools](#)

You have completed the exam titled "Midterm Exam."

An email has been sent to jdeal1@jhu.edu containing your score.

© 2015 The Johns Hopkins University. All rights reserved.

Copyright to this collective work of materials is owned by The Johns Hopkins University.

Copyright to individual contributions may be retained by contributing authors.

Midterm Exam

- Grades will be distributed via Courseplus once all exams (including make-up exams) have been completed

Midterm exam, continued

The exam will cover material up to and including
TODAY, Wednesday, September 23

1) Lectures

- Through Dr. Celentano's lecture on Mon Sept 21:
Epidemiologic Study Design II
- In class & online

2) Assignments & Activities:

- Outbreak Investigation
- Measuring Disease Frequency
- Validity & Reliability

Midterm exam, continued

- The emphasis of the mid-term exam questions will be on concepts/methods covered in greater detail or covered multiple times in the course lectures, Activities, Assignments, textbook and other course materials.

Midterm exam, continued

- **Practice Questions:**
 - The practice questions posted in Courseplus are from past mid-term exams. The mid-term will be similar in format and depth.

Midterm exam, continued

- **Office Hours:**
 - Please take advantage of office hours for addressing questions (office hours will not be formal review sessions).
 - 12:30-1:30 pm, TODAY, Wednesday 9/23
 - W2303
 - W4013
 - 12:30-1:30pm, Thursday 9/24
 - W4013
 - W4019

Room Assignments for Midterm Exam

Friday, September 25, 2015

10:00am – 12:00pm

**If your last name begins
with:**

Go to room:

A – B

W2030

C – N

Sommer Hall

O – T

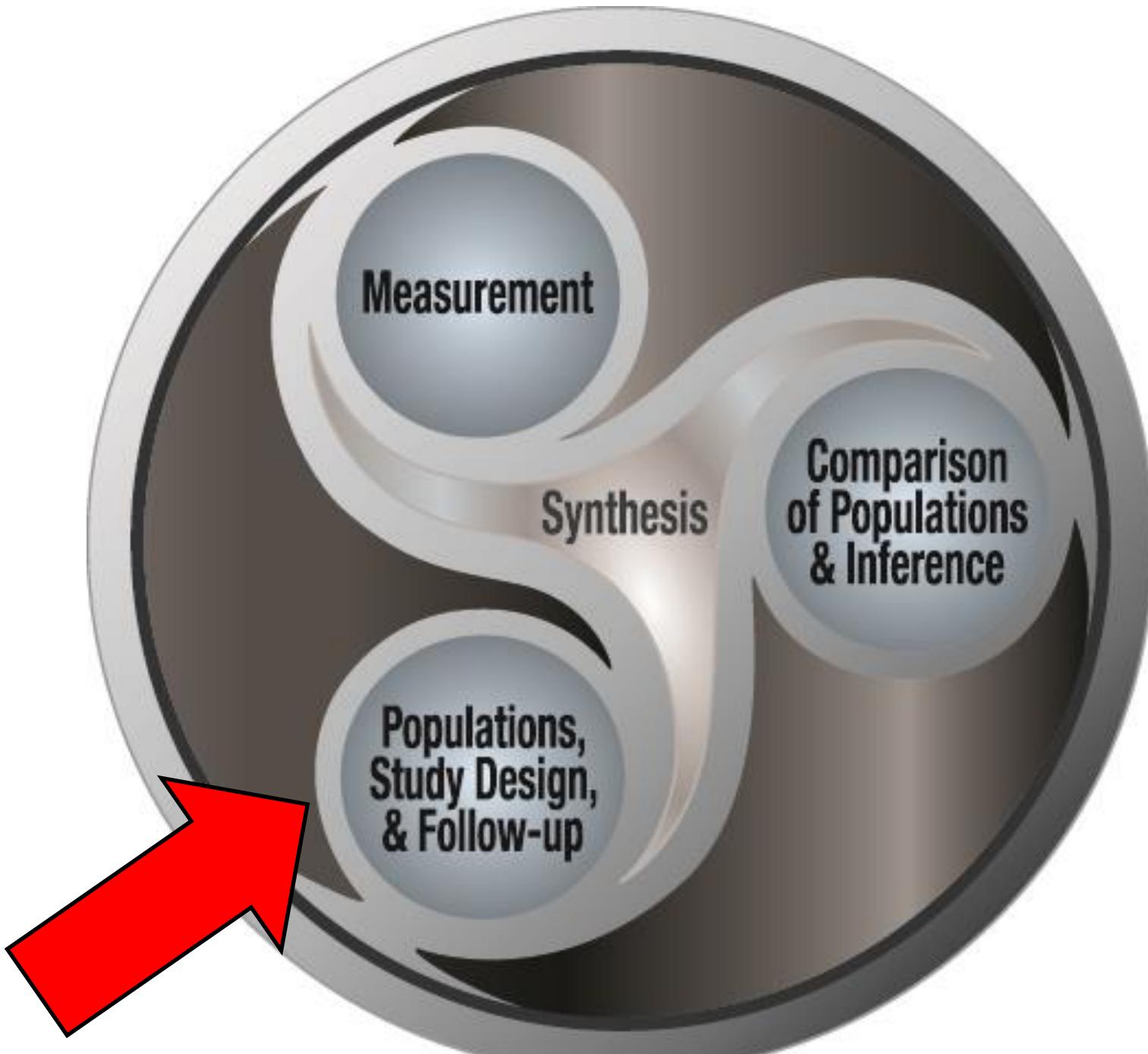
Sheldon Hall

U – Z

W5030

Objectives

- Summarize the course content in preparation for the midterm exam
- The review will NOT provide a list of topics that will appear on the exam and will NOT answer question by question the practice exam questions.



Populations



JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Epidemiologic Inference in Public
Health I
340.721

Populations & Epidemics

David D Celentano, ScD, MHS



Populations

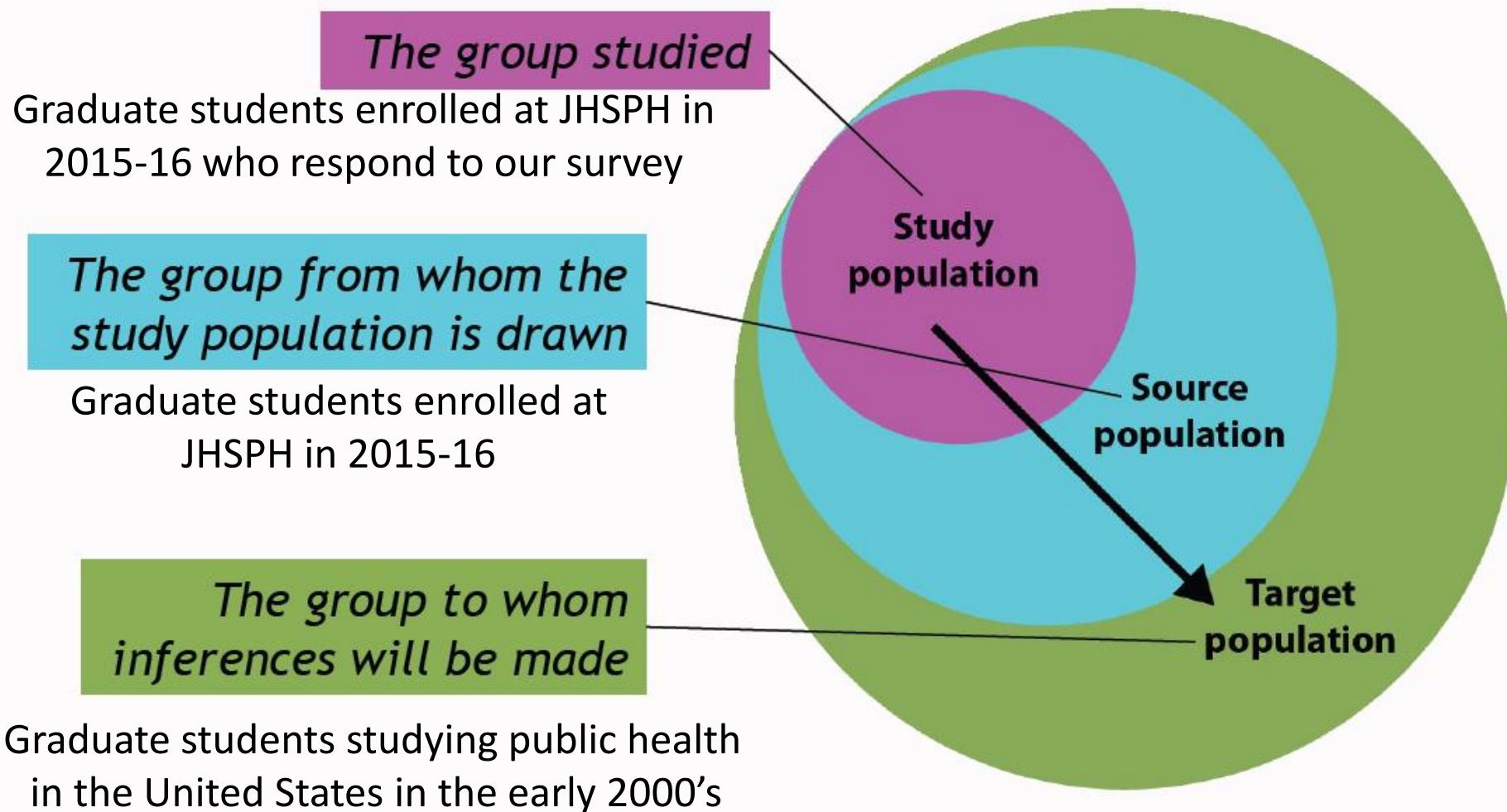
David Celentano, ScD, MHS
Johns Hopkins University

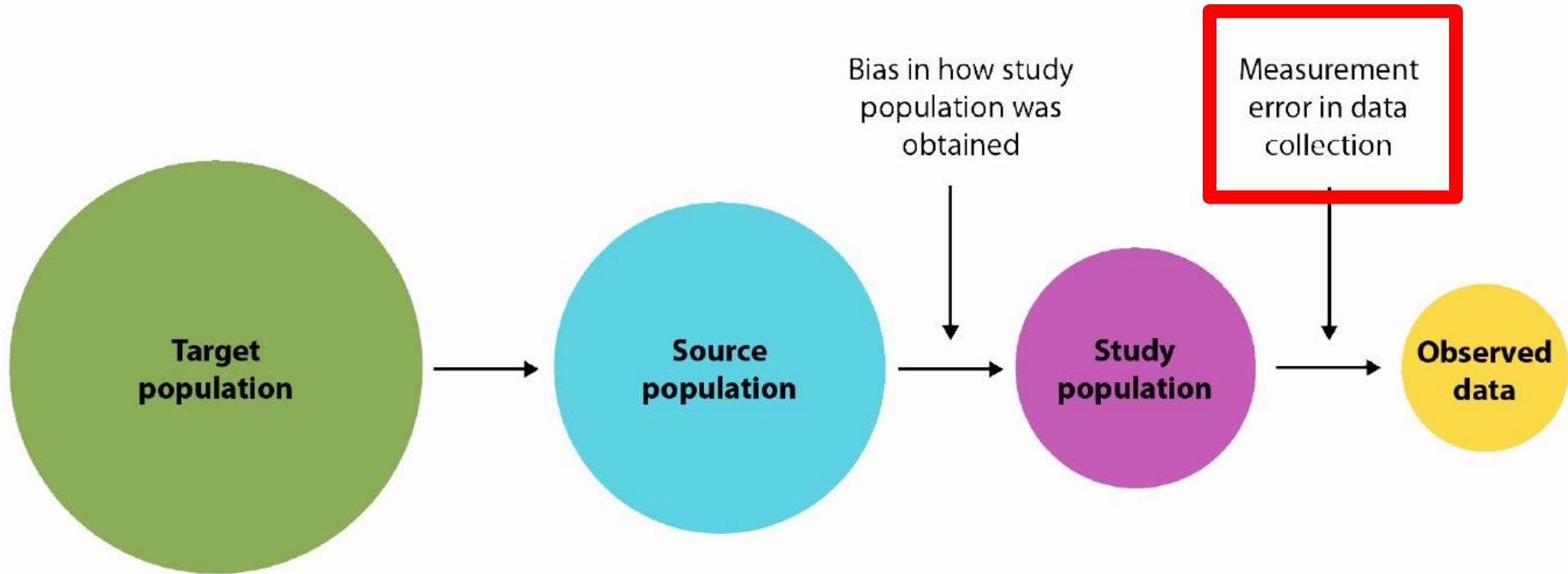


Person, Place, Time!

- In epidemiological studies, we usually describe the membership criteria by three characteristics:
 1. ***Person***: age, sex, race, ethnicity, occupation
 2. ***Place***: country, state, city, place of employment
 3. ***Time***: season, calendar year, life stage

Target, Source, Study Population

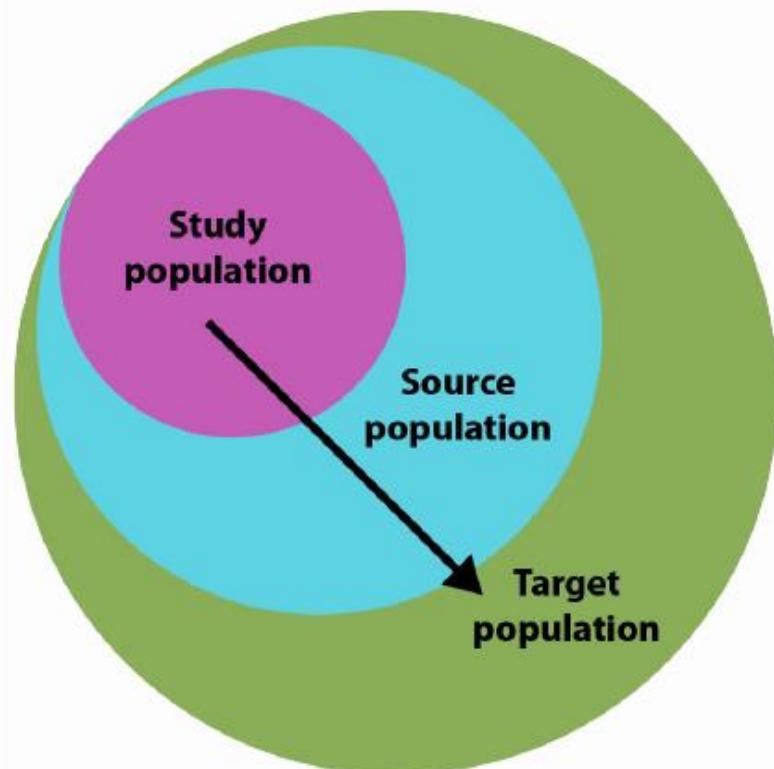




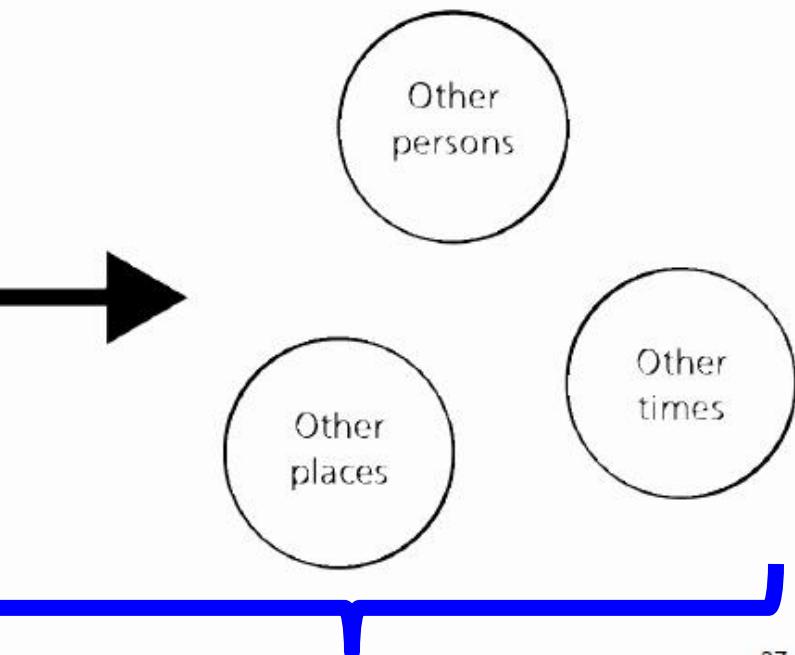
- Degree to which a study is free from bias or systematic error
- Soundness of study design, conduct, and analysis in answering the question that it posed for the study participants
- Prerequisite for external validity

External Validity

- “Generalizability”
- “Degree to which the results of the study may apply, be relevant, or be generalized to populations or groups that did not participate in the study”



INTERNAL VALIDITY



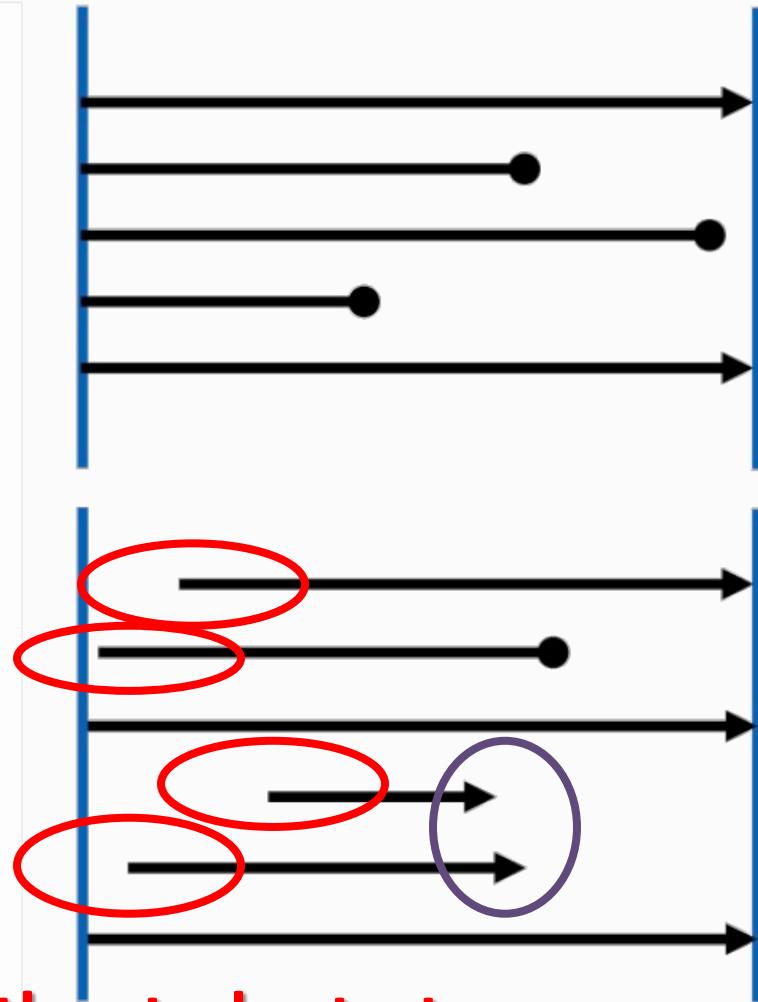
EXTERNAL VALIDITY

Populations in Time

Open and Closed Populations

- Closed populations
 - ▶ Individuals on a flight to Baltimore
 - ▶ Patients admitted to the ICU at JHH on a specific day

- Open populations
 - ▶ Population of Baltimore in 2013
 - ▶ Students in doctoral programs at JHSPH



Enter after the study starts

Exit before the study ends

Epidemics & Outbreak Investigation

- Assignment & Activity: *Outbreak Investigation*



Epidemiologic Inference in Public
Health I
340.721

Populations & Epidemics

David D Celentano, ScD, MHS



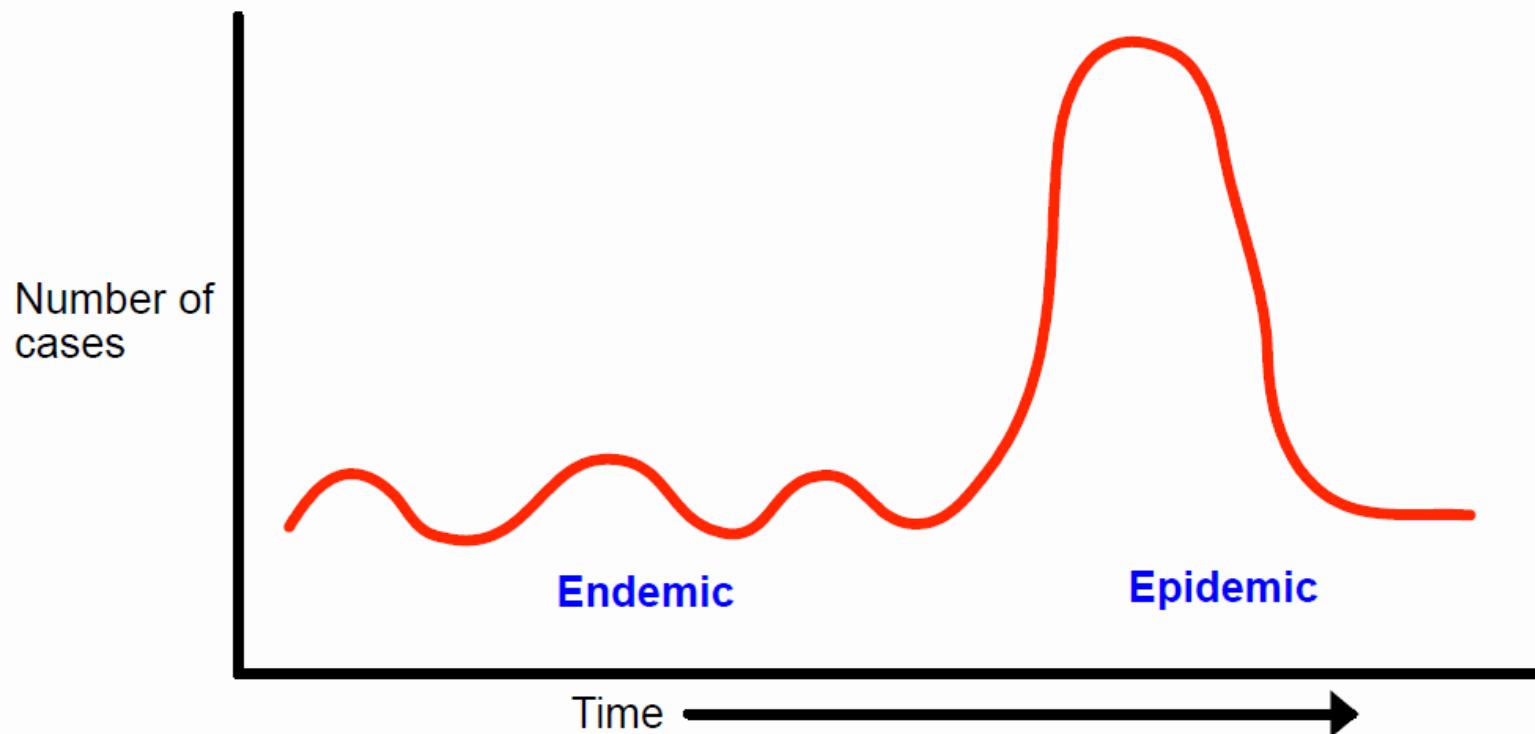
Epidemics

David Celentano, ScD, MHS
Johns Hopkins University



Endemic, Epidemic, and Pandemic

- **Endemic:** the habitual presence of a disease within a given geographic area
- **Epidemic:** the occurrence in a community or region of a group of illnesses of similar nature, clearly in excess of normal expectancy, and derived from a common or propagated source
- **Pandemic:** a world-wide epidemic



Attack Rate

Number of cases among people *at risk*

—————
Total number of people *at risk*



- Compare risk of disease in groups with different exposures
- Assumptions
 - ▶ All persons in denominator were exposed
 - ▶ All persons in denominator were susceptible
 - ▶ All cases were detected
 - For example, no subclinical cases

Case definition!

Question 1

Which of the following statements is(are) true of using a **less strict** case definition as compared to a more strict case definition? (SELECT ALL THAT APPLY)

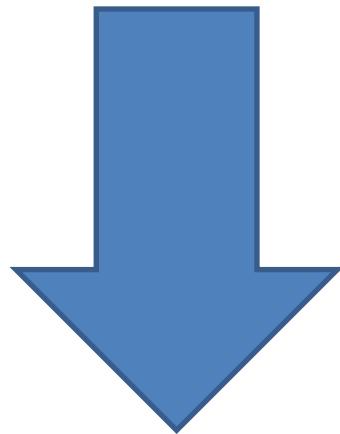
- a. More individuals who truly are ill will be counted as cases
- b. More individuals who truly are ill will be counted as non-cases
- c. More individuals who truly are not ill will be counted as non-cases
- d. More individuals who truly are not ill will be counted as cases

“Each case definition will differ in its ability to properly classify those individuals who are truly sick and those individuals who are truly not sick”

(From the in-class discussion of the Outbreak Investigation Activity)

“Each case definition will differ in its ability to properly classify those individuals who are truly sick and those individuals who are truly not sick”

(From the in-class discussion of the Outbreak Investigation Activity)



Validity = the ability of a test to distinguish between individuals who have a disease and individuals who do *not* have the disease

Question 1

Which of the following statements is(are) true of using a **less strict** case definition as compared to a more strict case definition? (SELECT ALL THAT APPLY)

- a. More individuals who truly are ill will be counted as cases
- b. More individuals who truly are ill will be counted as non-cases
- c. More individuals who truly are *not* ill will be counted as non-cases
- d. More individuals who truly are *not* ill will be counted as cases



Maximize
SENSITIVITY

More strict case definition = Maximize SPECIFICITY

Attack Rate



PROPORTION!

Ratio

- ▶ Division of two unrelated numbers

Proportion *What proportion of the population affected?*

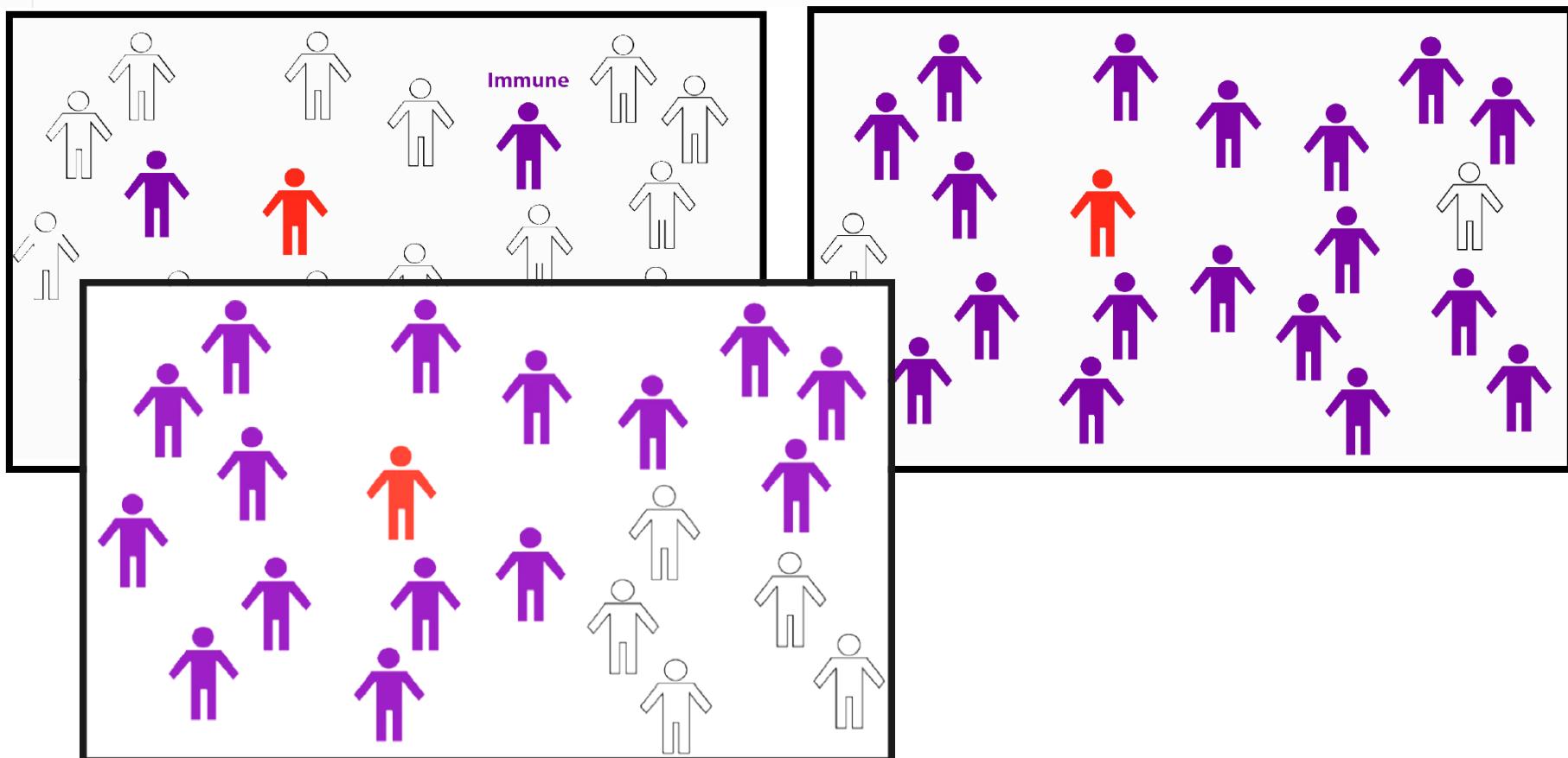
- ▶ Division of two numbers
- ▶ Numerator is subset of the denominator

Rate *How fast is disease occurring?*

- ▶ Division of two numbers
- ▶ Time is in the denominator

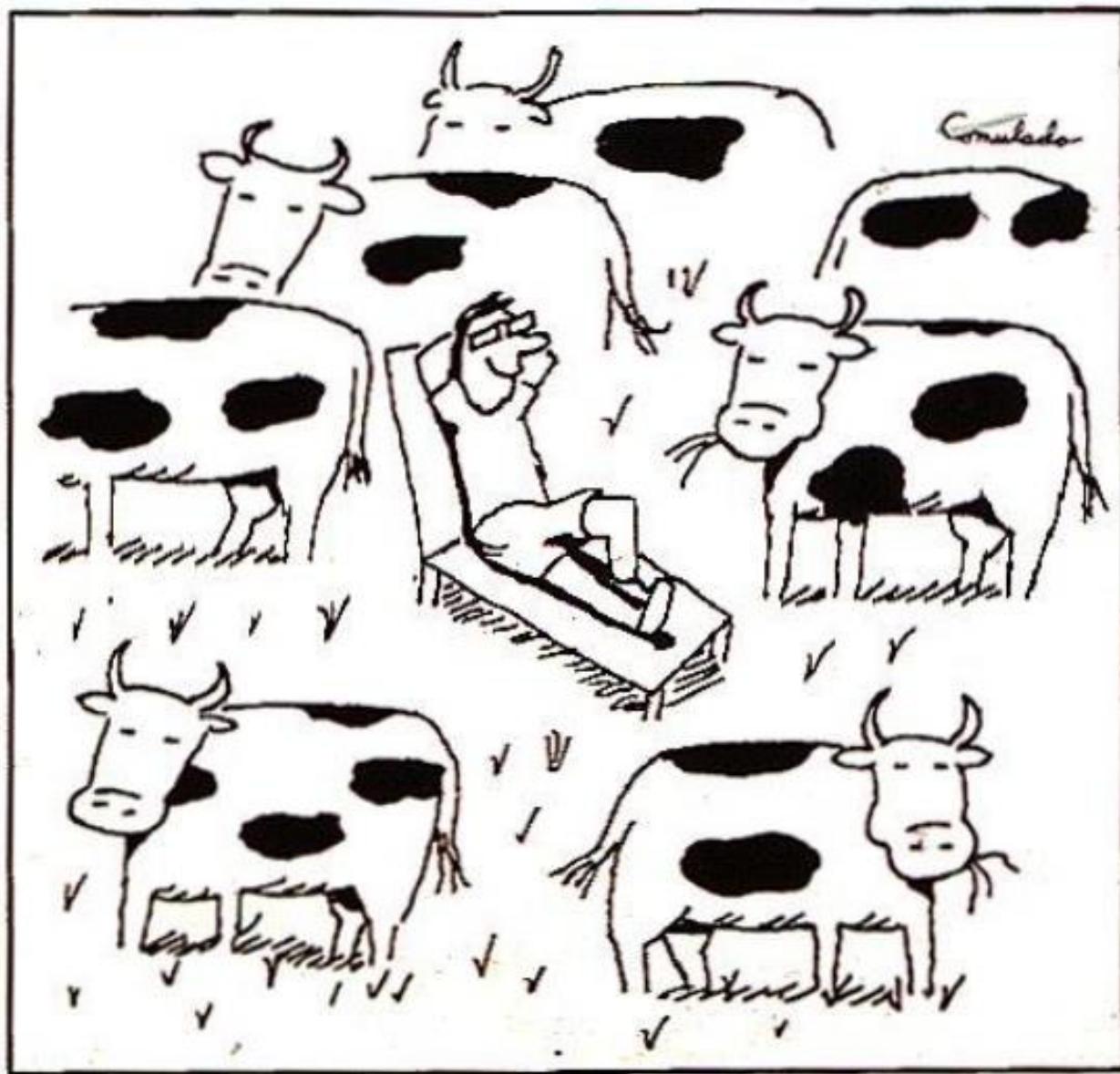
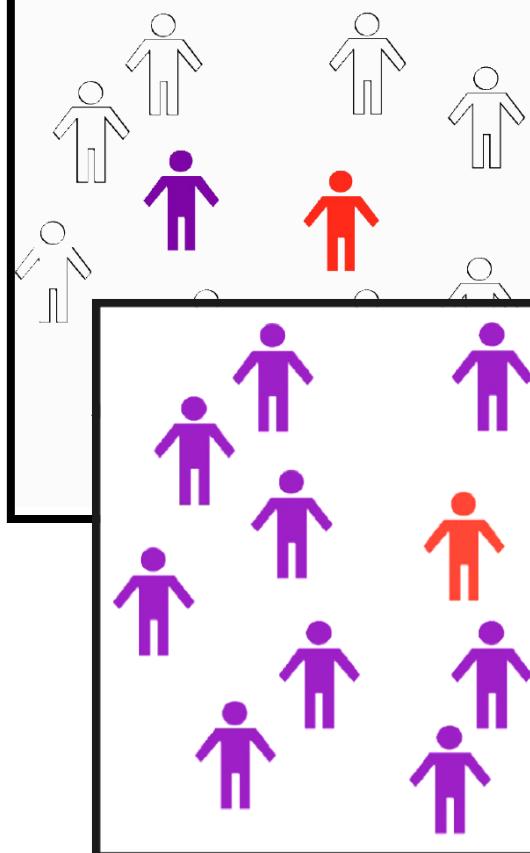
Herd Immunity

- The resistance of a group to attack by a disease to which a large portion of members are immune, thus lessening the likelihood of a patient with a disease coming into contact with a susceptible individual



Herd Immunity

- The resistance of a population to an infection is called herd immunity. If enough people in a population are immune, then those who are not immune are less likely to come into contact with a susceptible person.



BOB MISUNDERSTANDS THE CONCEPT OF "HERD IMMUNITY".

Three Important Variables in the Investigation of an Epidemic

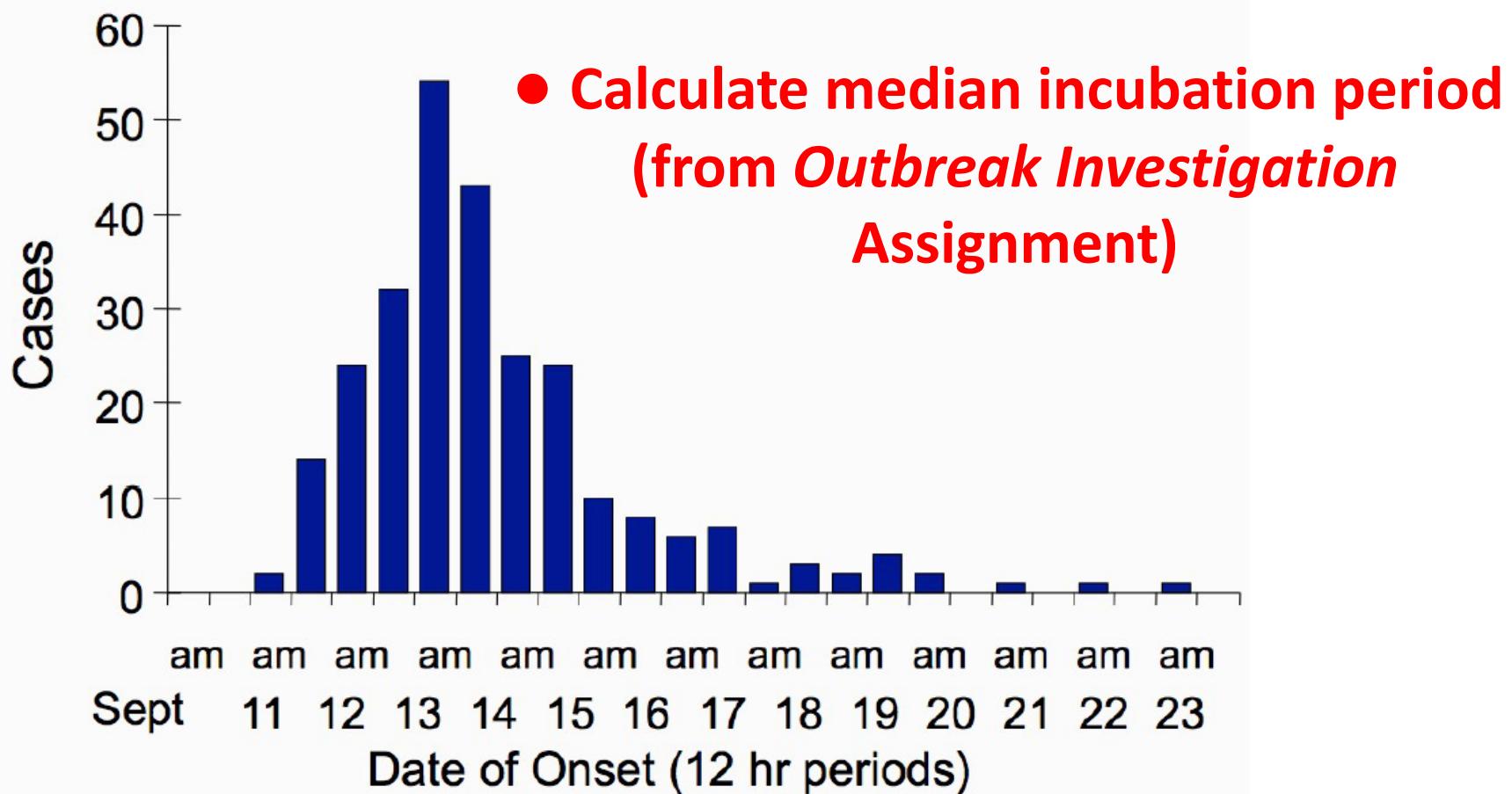
1. Time of exposure
2. Time of disease onset
3. Incubation period

Epidemic Curves

- Distribution of the times of onset of a disease
- In a single exposure, common vehicle epidemic, the epidemic curve represents the distribution of incubation periods

Epidemic Curve

Diarrheal Illness in Passengers Aboard the M/S Viking Sun,
By Onset, September 11-23, 1976



Also from Outbreak Investigation Assignment:

- Determine most likely infective food

***Cross-tabulations use attack rates to help isolate
INDEPENDENT factors associated with disease***

Example:

	# ILL	#WELL	ATTACK RATE
ATE Chicken ONLY	13	21	$13/(13+21)*100 = 38.2\%$
ATE Vegetables ONLY	84	124	$84/(84+124)*100 = 40.4\%$
ATE BOTH	45	55	
ATE Neither	12	120	$12/(12/+120)*100 = 9.1\%$

Which was the contaminated food? Calculate a RR and compare to the null value (RR=1.0).

$$\text{RR Chicken Only} : 38.2 / 9.1 = 4.2$$

$$\text{RR Vegetables Only} : 40.4 / 9.1 = 4.4$$

BOTH Chicken and Vegetables!

Steps in Investigating an Acute Outbreak

1. Define the epidemic
2. Examine the distribution of the cases by:
 - ▶ Time
 - ▶ Place
 - ▶ Person
3. Look for combinations of relevant variables
4. Develop hypotheses
5. Test hypotheses
6. Recommend control measures
 - ▶ Control present outbreak
 - ▶ Prevent similar outbreaks

Measures of Morbidity and Mortality

- Assignment & Activity: *Measuring Disease Frequency*

JOHNS HOPKINS
BLOOMBERG SCHOOL of PUBLIC HEALTH

Measures of Morbidity and Mortality, Part 1

JOHNS HOPKINS
BLOOMBERG SCHOOL of PUBLIC HEALTH

Measures of Morbidity and Mortality, Part 2

Jennifer Deal, PhD
Johns Hopkins University

Measures of Morbidity & Mortality

Jennifer A. Deal, PhD

■ Prevalence

No. of cases of a disease present in the population at a specified time
No. of persons in the population at that specified time

→ *Existing cases, measure of burden of disease*

Prevalence

No. of cases of a disease present in the population at a specified time
No. of persons in the population at that specified time

→ *Existing cases, measure of burden of disease*

Incidence

No. of NEW cases of a disease occurring
in the population at a specified time

No. of persons who are at risk of developing the disease
during that specified time

2 types of denominators:

(1.) No. of people =

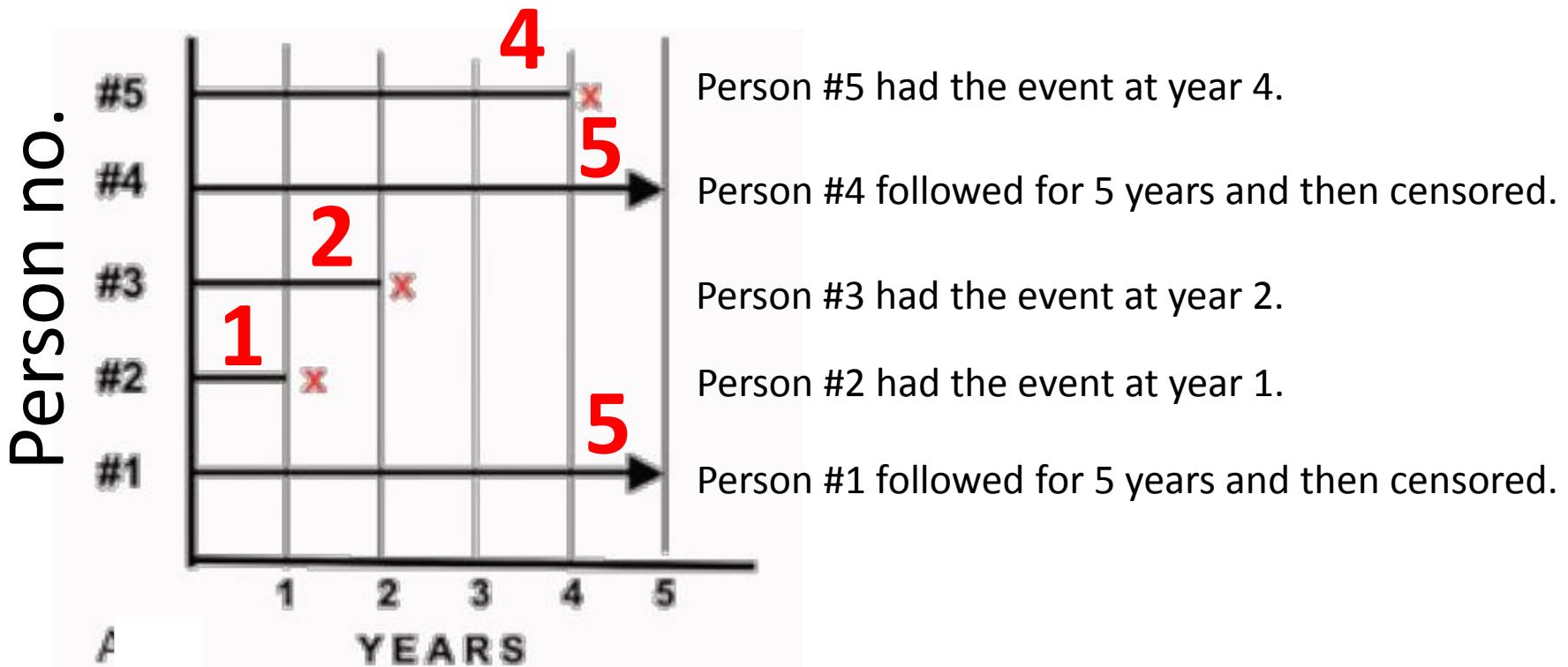
CUMULATIVE INCIDENCE (risk)

(2.) Person-time =

INCIDENCE RATE

→ *New cases, measure of risk of disease*

Person-time



$$5 + 1 + 2 + 5 + 4 = 17 \text{ person-years (py)}$$

$$\text{Cumulative incidence} = \frac{3}{5} = 0.6$$

$$\text{Incidence rate} = \frac{3}{17 \text{ py}} = 17.6 \text{ per 100 py}$$

(person-time discussed in 1st lecture)

Relationship between Incidence and Prevalence

- Prevalence = incidence x duration of disease

A Hypothetical Example of Chest X-Ray Screening

1,000 residents screened in each city (Hitown and Lotown):

Point prevalence
per 1,000 =

100 Hitown

60 Lotown



Prevalence is lower in Lotown

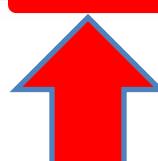
Relationship between Incidence and Prevalence

- Prevalence = incidence x duration of disease

A Hypothetical Example of Chest X-Ray Screening

1,000 residents screened in each city (Hitown and Lotown):

Point prevalence per 1,000 =	Incidence
100 Hitown	= 4/year
60 Lotown	= 20/year



**Prevalence is lower
in Lotown**

**But risk of disease
is higher in Lotown – why?**

Relationship between Incidence and Prevalence

- Prevalence = incidence x duration of disease

A Hypothetical Example of Chest X-Ray Screening

1,000 residents screened in each city (Hitown and Lotown):

Point prevalence per 1,000 =	Incidence	x Average duration
100 Hitown	= 4/year	X 25 years
60 Lotown	= 20/year	X 3 years



**Prevalence is lower
in Lotown**



**But risk of disease
is higher in Lotown – why?**

3 years

25 years

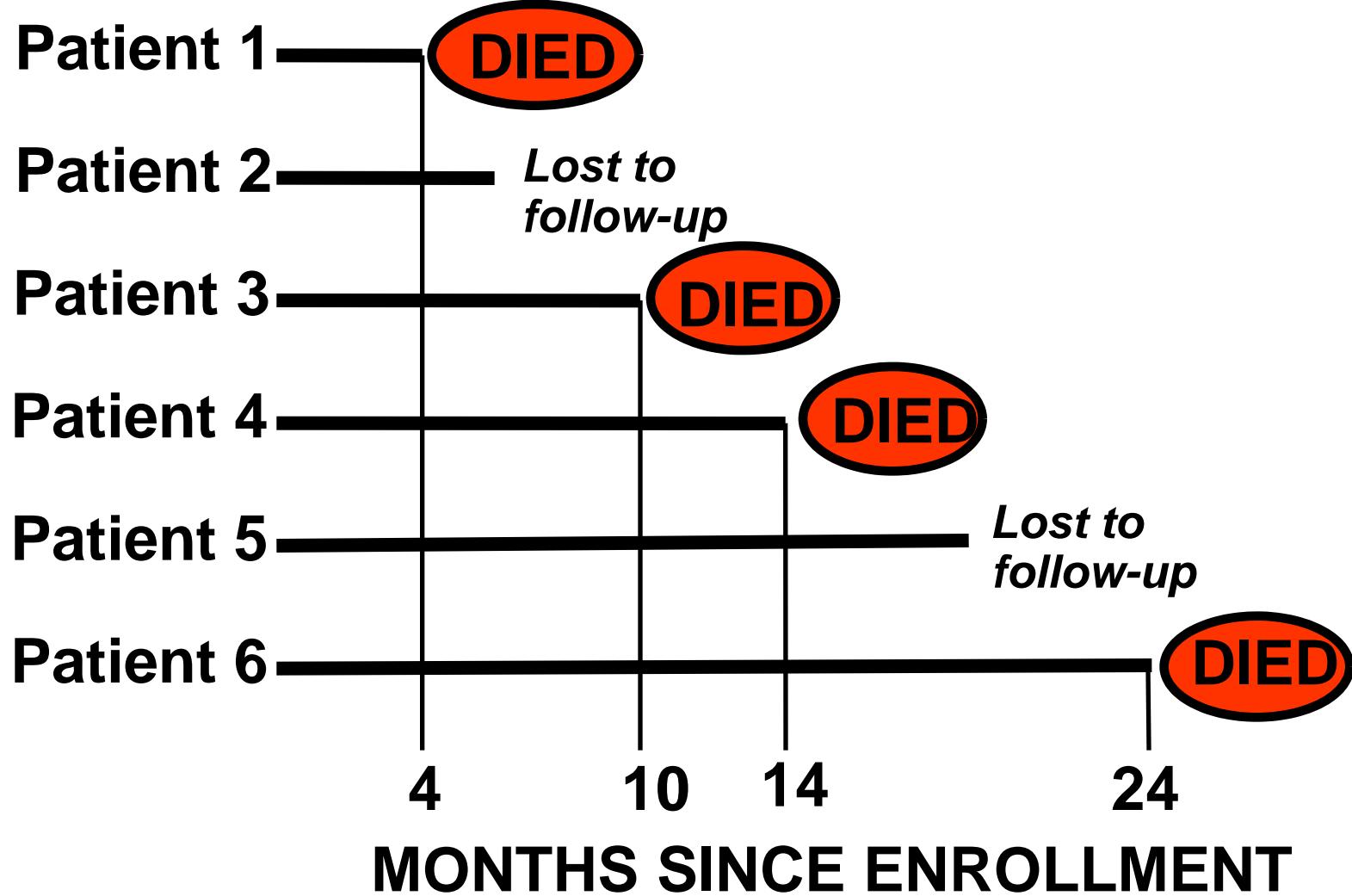
X

3 years

X

**Because disease
duration is shorter**

Kaplan-Meier Calculation



CALCULATING SURVIVAL USING THE KAPLAN-MEIER METHOD

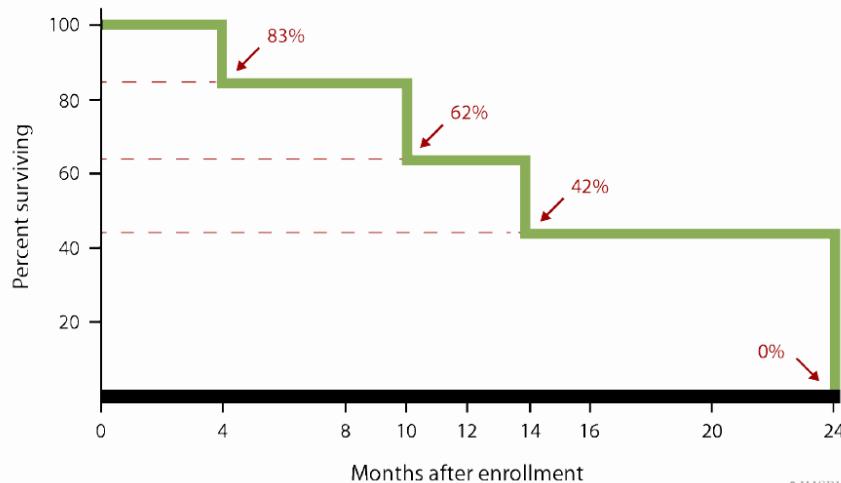
(1) Times to deaths from starting treatment (months)	(2) Number alive at each time	(3) Number who died at each time	(4) Proportion who died at that time <u>Column (3)</u> Column (2)	(5) Proportion who survived at that time 1-Column (4)	(6) Cumulative proportion who survived to that time (Cumulative Survival)
4	6	1	0.167	0.833	0.833
10	4	1	0.250	0.750	0.625
14	3	1	0.333	0.667	0.417
24	1	1	1.000	0.000	0.000

Why use Kaplan-Meier Method?

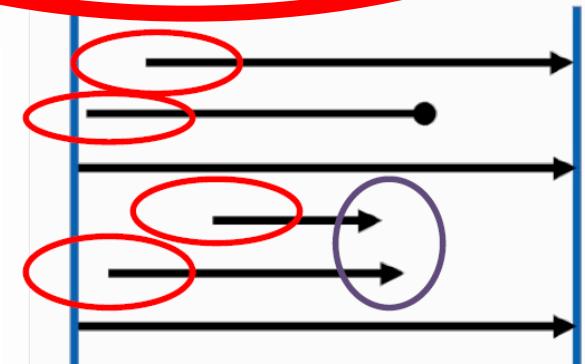
Measuring risk: Cumulative Incidence

- Closed cohort
 - ▶ Simple calculation of a proportion

Kaplan-Meier Survival Curve



In an open cohort!



Assumptions for Kaplan-Meier

- No changes have occurred in survivorship over calendar time (no secular trends)
- Participants who are lost to follow-up experience the same survivorship as those participants who remain in the study

■ Mortality Rate

- Annual, all-cause

$\frac{\text{Total no. of deaths from all causes in 1 year}}{\text{No. of persons in the population at midyear}}$

■ Cause-specific Mortality Rate

$\frac{\text{No. of deaths from a given cause in 1 year}}{\text{No. of persons in the population at midyear}}$

■ Case-Fatality

$\frac{\text{No. of individuals dying during a specified period of time after disease onset or diagnosis}}{\text{No. of individuals with the specified disease}}$

In Population X:

- 2,000 individuals at the midpoint of 1900
- 100 total deaths during 1900
- 300 smallpox cases during 1900
- 10 smallpox deaths during 1900

*Assume no Smallpox cases in this population occurred before 1900

Mortality Rate :

$$(100 \text{ total deaths} / 2,000) * 1000 = 50 \text{ per 1,000}$$

Smallpox mortality rate :

$$(10 \text{ smallpox deaths} / 2,000) * 1000 = 5 \text{ per 1,000}$$

Smallpox case-fatality :

$$10 \text{ smallpox deaths} / 300 \text{ smallpox cases} = 3.3\%$$

Comparing Mortality (or incidence)

- 1) Direct adjustment
- 2) Indirect adjustment (SMR)

Comparing Mortality (or incidence)

1) Direct adjustment

- I want to compare mortality rates in Population A and Population B
- But age distributions differ in the 2 populations, and age is a strong predictor of mortality
 - For example, residents of Population A are on average older than residents of Population B
- If Populations A and B had the same age distribution, how would mortality rates compare?

Direct Age-Adjustment: Comparison of Age-Specific Death Rates in the Two Time Periods

Age group	Early period			Later period		
	Population	Number of deaths	Death rate per 100,000	Population	Number of deaths	Death rate per 100,000
All ages	900,000	862	96	900,000	1,130	126
30-49	500,000	60	12	300,000	30	10
50-69	300,000	396	132	400,000	400	100
70+	100,000	405	406	200,000	700	350

e.g., $\frac{60 \text{ deaths}}{500,000} \times 100,000 = 12$

Start by calculating the rates in our populations

Direct Age-Adjustment: Carrying Out an Age-Adjustment Using the Total of the Two Populations as the Standard

Age group	Standard population	“Early” rate per 100,000	Expected number of deaths using “early” rate	“Later” rate per 100,000	Expected number of deaths using “later” rate
All	1,800,000				
30-49	800,000	12	96	10	80
50-69	700,000	132	924	100	700
70+	300,000	406	1,218	350	1,050

e.g.,

$$\frac{12 \text{ deaths}}{100,000} \times 800,000 = 96$$



Then apply the rates to the standard population
To get the expected number of deaths



Direct Age-Adjustment: Carrying Out an Age-Adjustment Using the Total of the Two Populations as the Standard

Age group	Standard population	“Early” rate per 100,000	Expected number of deaths using “early” rate	“Later” rate per 100,000	Expected number of deaths using “later” rate
All	1,800,000				
30-49	800,000	12	96	10	80
50-69	700,000	132	924	100	700
70+	300,000	406	<u>1,218</u>	350	<u>1,050</u>
Total number of deaths expected in standard population			2,238		1,830
Age-adjusted rates, “Early” = $\frac{2,238}{1,800,000} = 124.3$			“Later” = $\frac{1,830}{1,800,000} = 101.7$		

Key Points: Direct Adjustment

- Direct adjustment useful for *comparison* of incidence or mortality rates in two populations *independent* of a given factor
- Factor can be age, race, sex, education, etc.

Key Points: Direct Adjustment

- Adjusted rates are a weighted average of factor-specific rates (where the weight is the number of people in each strata for that factor in a standard population)
- Adjusted rates are not the true (actual) rates in the population

Choice of Standard Population

**Question 4: Compare the WHO World Standard Population to the Standard Population in Table 3.
How do they compare?**

- Notice that the absolute numbers for each population differ
- If another standard population were used that had a very different age distribution from the first, then different inferences could be made because different weights would be given to different age groups that may contribute more or less to the overall incidence.

Comparing Mortality (or incidence)

- 1) Direct adjustment
- 2) Indirect adjustment (SMR)
 - Do not have data on deaths by strata, but we do know distribution of age or other factors in the total population
 - Study mortality in occupationally exposed population (or where number of events in each strata is small)

Standardized Mortality Ratio (SMR)

Indirect Adjustment

“The number of deaths occurring in a given population (occupation) expressed as the percentage of the number of deaths that might have been expected to occur if the given population (occupation) had experienced within each age group the same rate as that of the standard population.”

- $$\text{SMR} = \frac{\text{Observed number of deaths per year}}{\text{Expected number of deaths per year}}$$

Computation of an SMR for Tuberculosis (TB), all Forms, for White Miners Ages 20-59 Years, US, 1950

Age (years)	Estimated population of White miners (1)	Death rate per 100,000 for TB for males in general population (2)	Expected deaths from TB in White miners if they had the same risk as general population (3) = (1) x (2)	Observed deaths from TB in White miners (4)
20-24	74,598	X 12.26	= 9.14	10
25-29	85,077	X 16.12	= 13.71	20
30-34	80,845	X 21.54	= 17.41	22
35-44	148,870	X 33.96	= 50.00	98
45-54	102,649	X 56.82	= 58.32	174
55-59	42,494	X 75.23	= <u>31.96</u>	<u>112</u>
			181.09	436



$$\text{SMR (for 20-59 years old)} = \frac{436}{181.09} = 241$$

We apply the death rate in the general population to our population

Adjustment

- To compare two populations

- ▶ Direct adjustment
- ▶ Indirect adjustment (SMR, SIR)

Adjusted rates are useful for comparing populations but are *not* the true rates

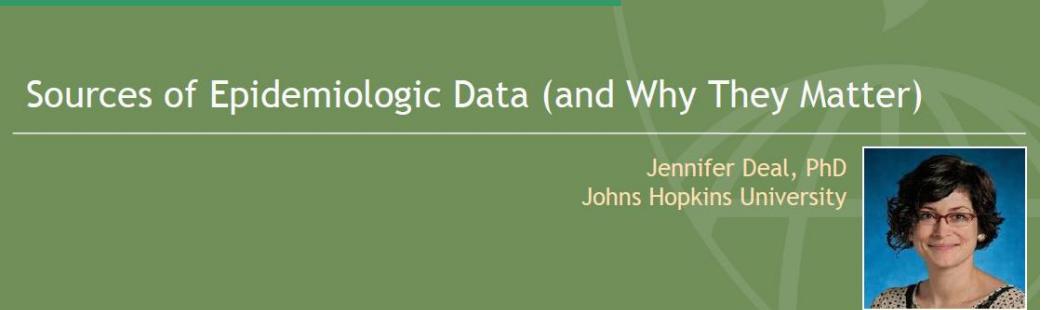
Adjusted rates are dependent on the standard population used for the adjustment (e.g., 1940 vs. 2000)

Surveillance

Epidemiologic Inference in Public Health I
340.721

**Validity & Reliability
Surveillance**

Jennifer A. Deal, PhD



Sources of Epidemiologic Data (and Why They Matter)

Jennifer Deal, PhD
Johns Hopkins University



“Public health surveillance is the continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice.”

http://www.who.int/topics/public_health_surveillance/en/

“Such surveillance can:

- serve as an early warning system for impending public health emergencies;
- document the impact of an intervention, or track progress towards specified goals; and
- monitor and clarify the epidemiology of health problems, to allow priorities to be set and to inform public health policy and strategies.”

Active vs. Passive Surveillance

■ ACTIVE Surveillance:

- System in which project staff are recruited to carry out a surveillance program (for example, to periodically visit health care facilities in order to identify new cases of disease)
- Data is collected

■ ACTIVE Surveillance:

Advantages

- Can be highly sensitive**
- Can collect more detailed information**
- May be more representative**

Disadvantages

- Costly/resource intensive**
- Labor intensive**
- Difficult to sustain over time**

■ PASSIVE Surveillance:

- System in which available data on reportable diseases are used; responsibility for the reporting often falls to the health care provider
- Uses existing data

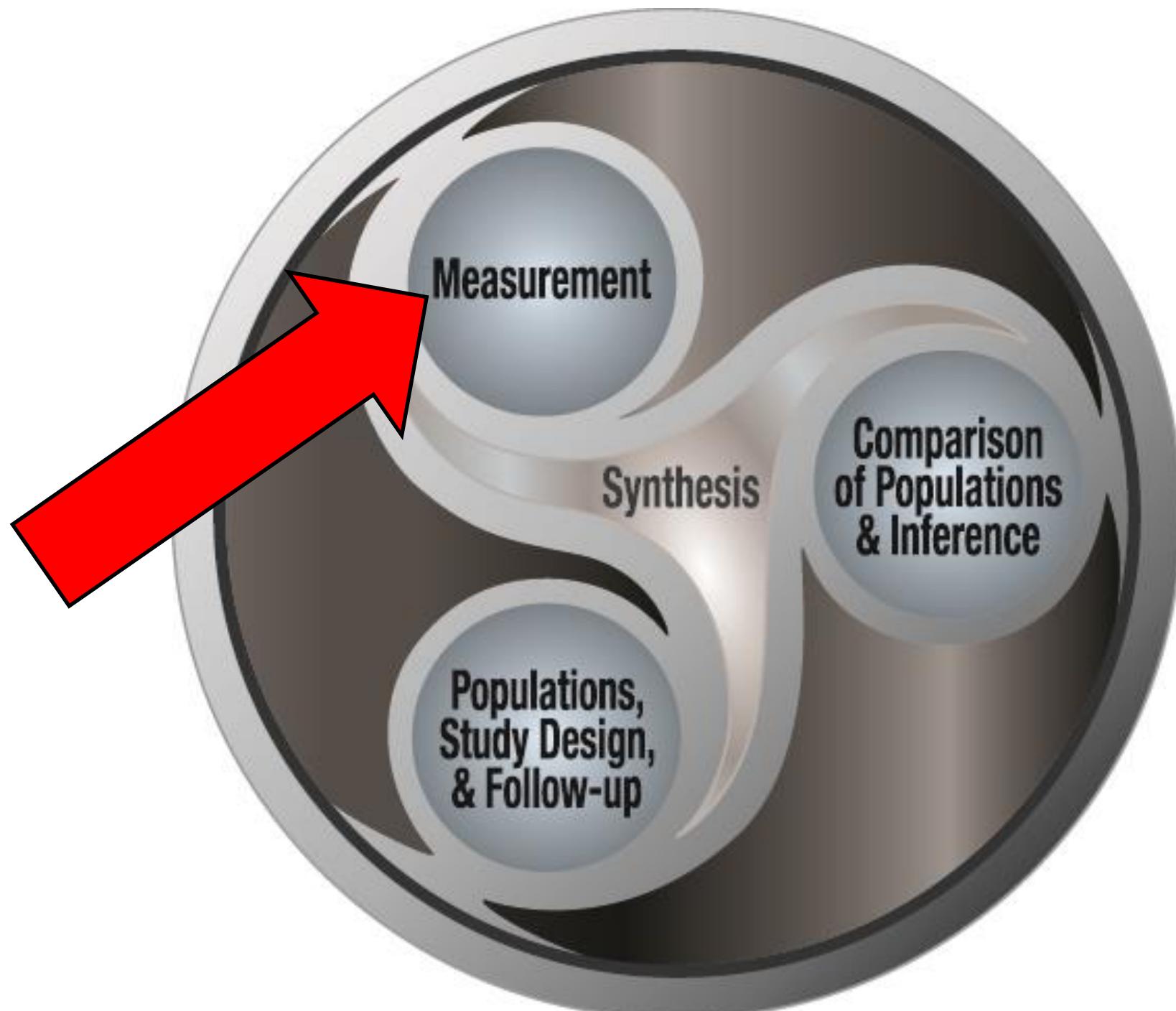
■ PASSIVE Surveillance:

Advantages

- Less costly**
- Easier to design and carry out (can cover large areas)**
- Useful to monitor trends over time**

Disadvantages

- Low sensitivity**
- Amount of data available is limited**
- May not be representative**



Measurement - Validity

- Assignment & Activity: *Validity & Reliability*

Epidemiologic Inference in Public Health I
340.721

**Validity & Reliability
Surveillance**

Jennifer A. Deal, PhD

Validity of Diagnostic and Screening Tests

Jennifer Deal, PhD
Johns Hopkins University



- Validity = the ability of a test to distinguish between individuals who have a disease and individuals who do not have the disease
- (1) Sensitivity = ability of a test to correctly identify individuals who have the disease
- (2) Specificity = *ability of a test to correctly identify individuals who do NOT have the disease*

Sensitivity =
the ability of the test
to correctly identify
individuals who
have the disease

		Have the Disease	Do <u>NOT</u> have the Disease	Totals
		Test Positive	100	180
		Test Negative	800	820
Totals		100	900	1,000

Sensitivity =
 $\frac{80}{100} = 0.80 = 80\%$

	Have the Disease	Do <u>NOT</u> have the Disease	Totals
Test Positive	80	100	180
Test Negative	20	800	820
Totals	100	900	1,000

Specificity =
the ability of the test
to correctly identify
individuals who do
NOT have the disease

Specificity =

$$\frac{800}{900} = 0.89 = 89\%$$

	Have Disease	Do <u>NOT</u> have Disease
Test Positive	True Positive (TP): Have disease & test positive	False Positive (FP): Do <u>NOT</u> have disease but test positive
Test Negative	False Negative (FN): Have disease but test negative	True Negative (TN): Do <u>NOT</u> have disease & test negative

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Comparing Cutpoints for Continuous Measures

Cutpoint of ≥ 80 mg/dL:

- Sensitivity = 100% (no FN)
- Specificity = low (lots of FP)

Cutpoint of ≥ 200 mg/dL:

- Sensitivity = low (lots of FN)
- Specificity = 100% (no FP)

Consequences of FP:

- Emotional cost
- Financial cost to re-test
- More invasive test

Consequences of FN:

- Missed opportunity to treat

Recall....

Thought experiment

- You are a public health worker in a clinic in the field
 - You have 2 screening tests for a disease:
 - 1st test: Sensitivity=70%, Specificity=80%
 - 2nd test: Sensitivity=90%, Specificity=90%
 - It is much more important to minimize false positives than to catch all cases of a disease
- Maximize
SPECIFICITY

Use of multiple tests: an example

(Assume a Population of 10,000 People with a Disease Prevalence = 5%)

■ TEST 1

- Sensitivity = 70%
- Specificity = 80%

		No		Totals
		Disease	No Disease	
Test Positive	Disease	350	1,900	2,250
	No Disease	150	7,600	7,750
Totals	500	9,500	10,000	

■ TEST 2

- Sensitivity = 90%
- Specificity = 90%

		No		Totals
		Disease	No Disease	
Test Positive	Disease	315	190	505
	No Disease	35	1,710	1,745
Totals	350	1,900	2,250	



NET SENSITIVITY:

- ▶ Proportion of those with the disease who test positive on BOTH Test 1 and Test 2
- ▶ $\frac{\text{Test Positive on BOTH Tests}}{\text{Total With Disease}}$
- ▶ $= \frac{315}{500} = 0.63 = 63\%$
- ▶ Net sensitivity is DECREASED

NET SPECIFICITY:

- ▶ Proportion of those without the disease who test negative on EITHER Test 1 or Test 2
- ▶
$$\frac{\text{Negative on Test 1} + \text{Negative on Test 2}}{\text{Total Without Disease}}$$
- ▶ $= \frac{7,600 + 1,710}{9,500} = 98\%$
- ▶ Net specificity is INCREASED

Defining PPV and NPV

Predictive value = the probability of disease given the results of the test

- (1) *Positive Predictive Value (PPV)* =
the probability that a person with a positive test does have the disease

- (2) *Negative Predictive Value (NPV)* =
the probability that a person with a negative test does NOT have the disease

	Have the Disease	Do <u>NOT</u> have the Disease	Totals
Test Positive	80	100	180
Test Negative	20	800	900
Totals	100	900	

PPV =

the probability that a person with a positive test does have the disease

PPV =

$$\frac{80}{180} = 0.44 = 44\%$$

	Have the Disease	Do <u>NOT</u> have the Disease	Totals
Test Positive	80	100	180
Test Negative	20	800	820
Totals	100	900	

□ **NPV =**

the probability that a person with a negative test does NOT have the disease

□ **NPV =**

$$\frac{800}{820} = 0.98 = 98\%$$

Example: Prevalence and PPV IV (Assume Sensitivity=99% and Specificity=95%)

■ Prevalence = 1%

		No		Totals
		Disease	No Disease	
Test Positive	Disease	99	495	594
	No Disease	1	9,405	9,406
Totals	100	9,900	10,000	

$$\text{PPV} = \frac{99}{594} = 17\%$$

■ Prevalence = 5%

		No		Totals
		Disease	No Disease	
Test Positive	Disease	495	475	970
	No Disease	5	9,025	9,030
Totals	500	9,500	10,000	

$$\text{PPV} = \frac{495}{970} = 51\%$$

Example: Specificity and PPV

(Assume Prevalence=20% and Sensitivity=50%)

■ Specificity = 50%

		No	Totals
		Disease	
Test Positive	Disease	100	500
	No Disease	400	500
Test Negative	Disease	100	200
	No Disease	400	800
Totals		200	1,000

$$\text{PPV} = \frac{100}{500} = 20\%$$

■ Specificity = 90%

		No	Totals
		Disease	
Test Positive	Disease	100	180
	No Disease	80	500
Test Negative	Disease	100	200
	No Disease	720	820
Totals		200	1,000

Smaller denominator

$$\text{PPV} = \frac{100}{180} = 56\%$$

Fewer FP

Measurement - Reliability

- Assignment & Activity: *Validity & Reliability*

*Epidemiologic Inference in Public
Health I*
340.721

**Validity & Reliability
Surveillance**

Jennifer A. Deal, PhD

Reliability

Jennifer Deal, PhD
Johns Hopkins University

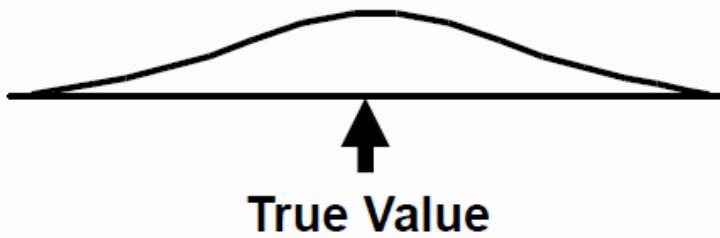


Validity

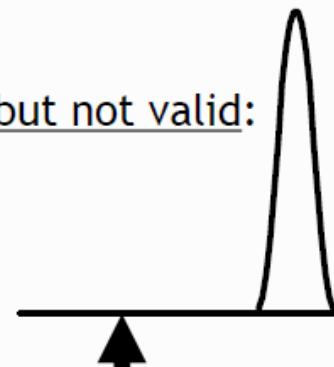
- Are the test results correct?

- *Analogous to accuracy*

Valid but not reliable:



Reliable but not valid:

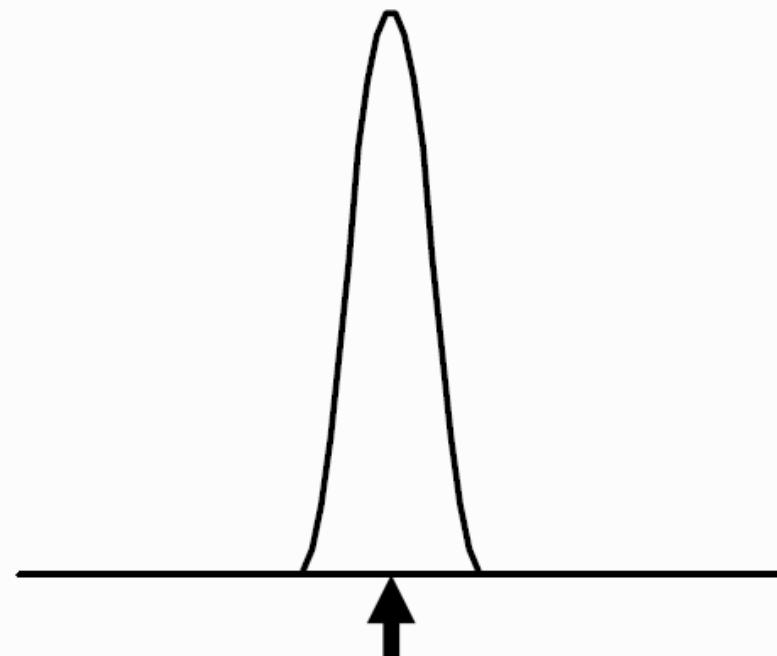


Reliability

- Are the test results the same when the same test is repeated in the same individual under similar conditions?

- *Analogous to precision*

Valid AND Reliable Test Results



True Value

Percent Agreement

		Reading No. 1			
		Abnormal	Suspect	Doubtful	Normal
Reading No. 2					
Abnormal	A	B	C	D	
Suspect	E	F	G	H	
Doubtful	I	J	K	L	
Normal	M	N	O	P	

$$\text{Percent Agreement} = \frac{A + F + K + P}{\text{Total Readings}} \times 100\%$$

Percent Positive Agreement

		<u>OBSERVER 1</u>	
		Positive	Negative
<u>OBSERVER 2</u>	Positive	a	b
	Negative	c	d

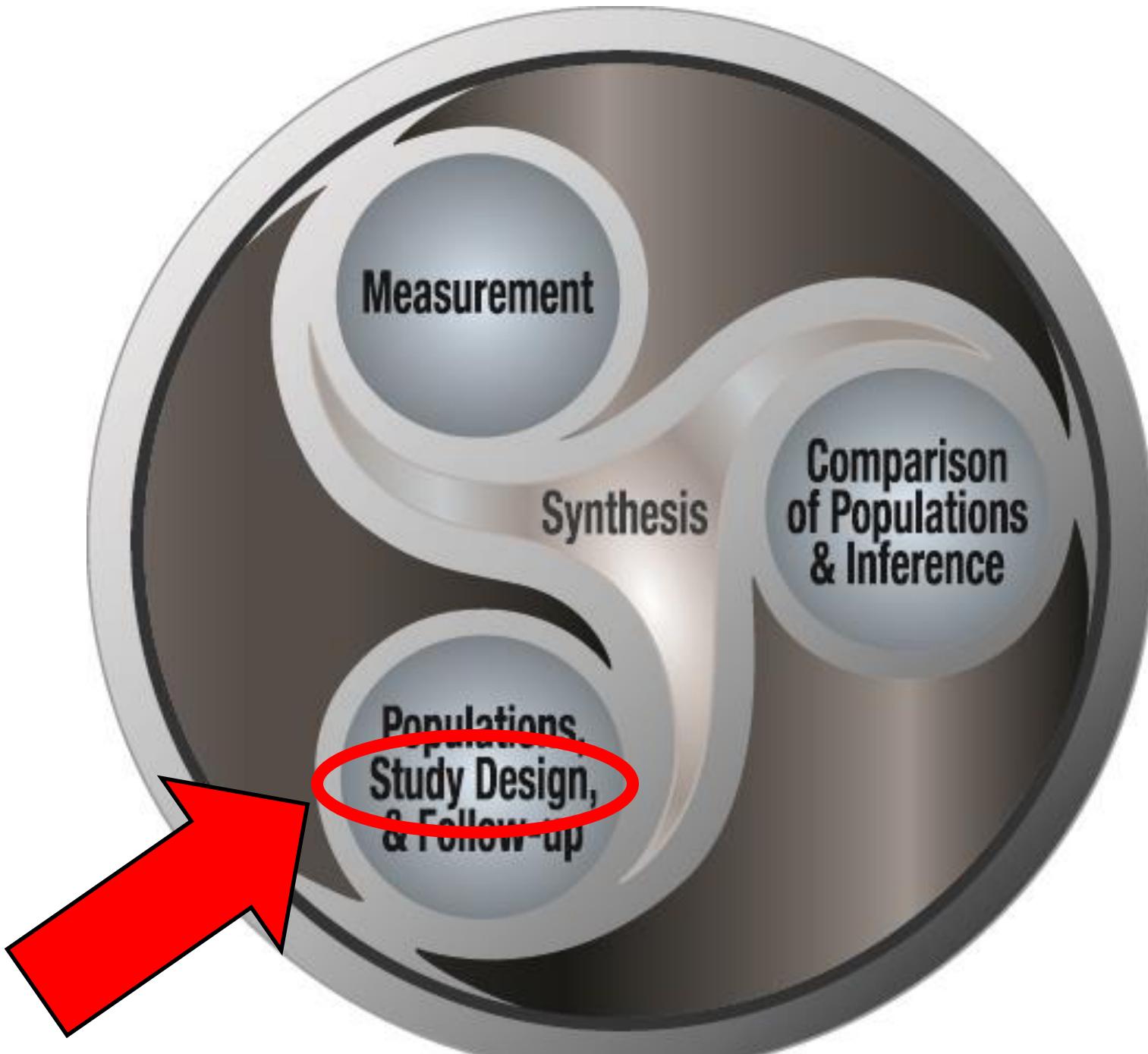
- Percent agreement when at least one test is positive
$$= \frac{a}{a + b + c}$$
- Tests that both observers classified as negative are removed from the calculation (*both the numerator and the denominator!*)

Value of Kappa Strength of Agreement

0.0	No agreement better than chance alone
<0.20	Poor
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Good
0.81 - 1.00	Very Good

Kappa Statistic

- Sometimes observers will agree solely by chance
- The Kappa statistic allows us to calculate the level of agreement *independent of chance*
- *What is the agreement between observers beyond what would be expected by chance alone?*



Study Design



JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Descriptive Study Designs: Case-Series, Cross-Sectional, and Ecologic Studies



Study Designs: Case-Control Studies and Other Nested Designs



Study Designs: Cohort Studies



Analytic Studies and Clinical Trials

David Celentano, ScD, MHS
Johns Hopkins University



Epidemiologic Inference in Public Health I
340.721

Epidemiologic Study Design I

David D. Celentano, ScD, MHS

Epidemiologic Study Design II

David D. Celentano, ScD, MHS

Summary of Study Designs

Main Types of Epidemiologic Studies

Study type	Characteristics
Experimental	<ul style="list-style-type: none">Studies prevention and treatment of diseaseInvestigator actively manipulates which groups receive the study agent
Observational	<ul style="list-style-type: none">Studies causes, prevention and treatment for diseasesInvestigator watches as nature takes its course
Cohort	<ul style="list-style-type: none">Examines multiple health effects of an exposureSubjects defined by exposure levels and follow for disease occurrence
Case-control	<ul style="list-style-type: none">Typically examines multiple exposures in relation to a diseaseSubjects are defined as cases and controls and exposure histories compared
Cross-sectional	<ul style="list-style-type: none">Examine relationship between exposure and disease prevalence in a defined population at one point in time
Ecological	<ul style="list-style-type: none">Examines relationship between exposure and disease with population-level data rather than individual data

Experimental Study Designs

Main Types of Epidemiologic Studies

Study type	Characteristics
Experimental	<ul style="list-style-type: none">• Studies prevention and treatment of disease• Investigator actively manipulates which groups receive the study agent

= Randomized Trial

Design of a **Randomized Trial**

Start
with:

Defined Population

Then randomize
to treatment:

Randomize

New
Treatment

Current/No
Treatment

Then
follow
up:

Improve

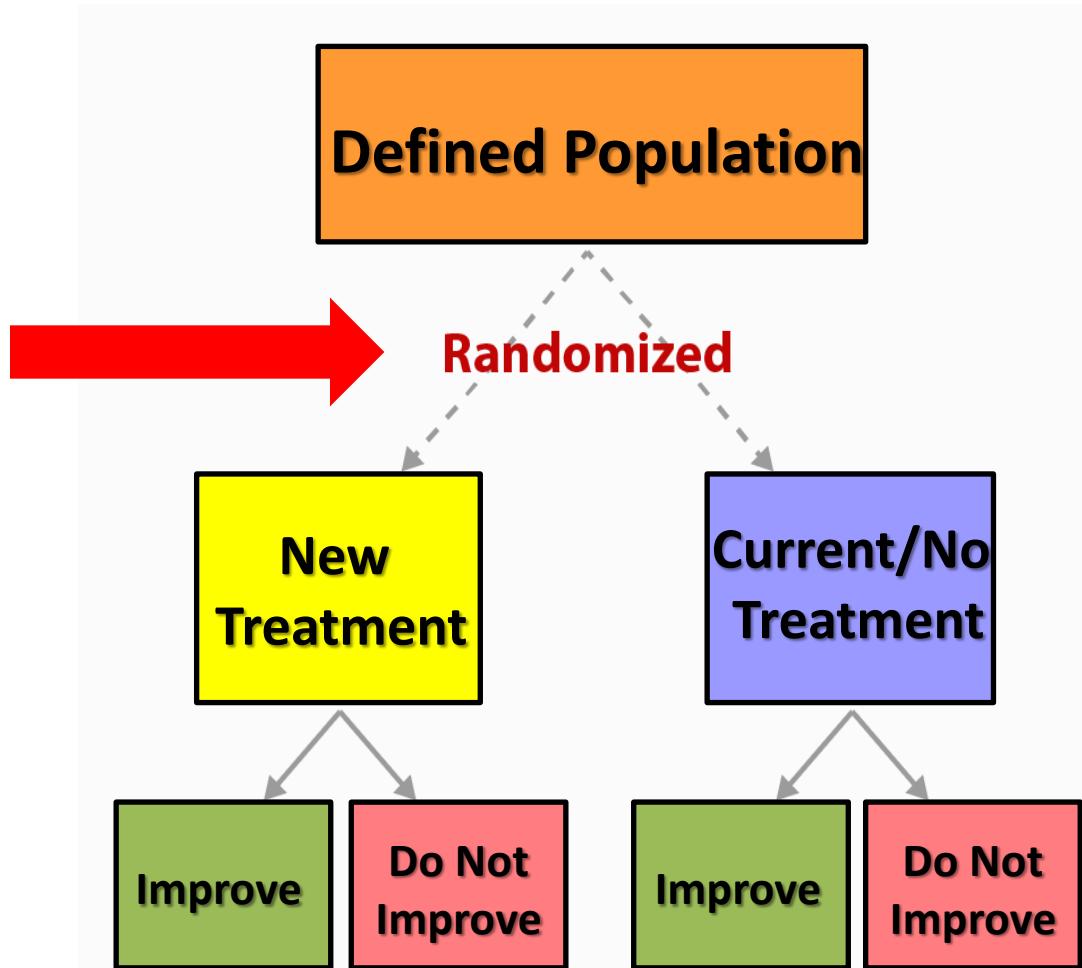
Do Not
Improve

Improve

Do Not
Improve

What is the primary purpose of randomization?

- To prevent bias in the choice of treatment



Observational Study Designs

Main Types of Epidemiologic Studies

Study type	Characteristics
<u>Observational</u>	<ul style="list-style-type: none">Studies causes, prevention and treatment for diseasesInvestigator watches as nature takes its course
1. Cohort	<ul style="list-style-type: none">Examines multiple health effects of an exposureSubjects defined by exposure levels and follow for disease occurrence
2. Case-control	<ul style="list-style-type: none">Typically examines multiple exposures in relation to a diseaseSubjects are defined as cases and controls and exposure histories compared
3. Cross-sectional	<ul style="list-style-type: none">Examine relationship between exposure and disease prevalence in a defined population at one point in time
4. Ecological	<ul style="list-style-type: none">Examines relationship between exposure and disease with population-level data rather than individual data

Observational Study Designs

Main Types of Epidemiologic Studies

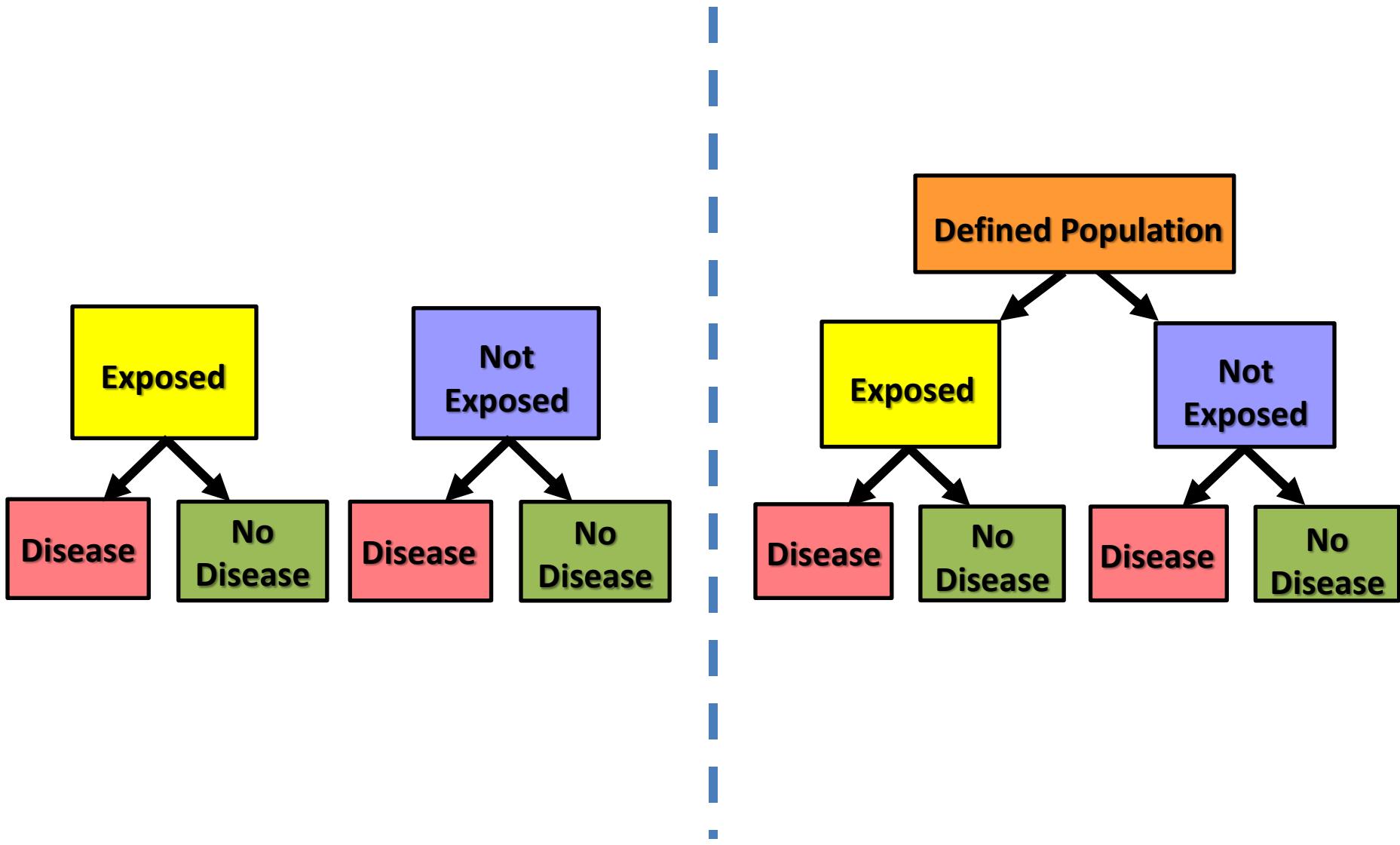
Study type	Characteristics
Observational	<ul style="list-style-type: none">Studies causes, prevention and treatment for diseasesInvestigator watches as nature takes its course
1. Cohort	<ul style="list-style-type: none">Examines multiple health effects of an exposureSubjects defined by exposure levels and follow for disease occurrence
2. Case-control	<ul style="list-style-type: none">Typically examines multiple exposures in relation to a diseaseSubjects are defined as cases and controls and exposure histories compared
3. Cross-sectional	<ul style="list-style-type: none">Examine relationship between exposure and disease prevalence in a defined population at one point in time
4. Ecological	<ul style="list-style-type: none">Examines relationship between exposure and disease with population-level data rather than individual data



Population-level data

Individual-level data

Design of a **Cohort Study**



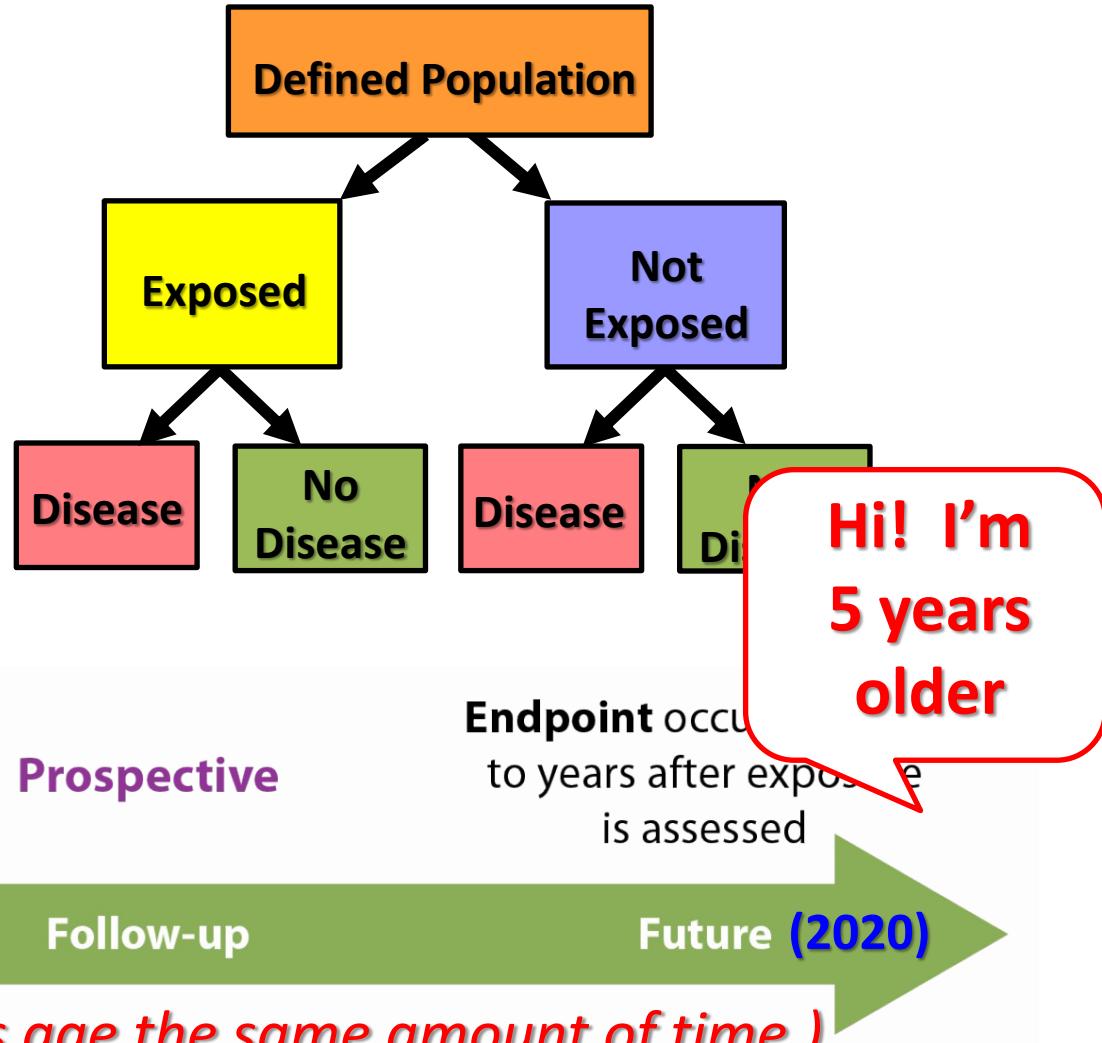
Design of a *Prospective* Cohort Study

Today (Sept 2015):

Today (Sept 2015):

**Future
(for example, Sept 2020):**

Measure **exposure** or
collect biological
samples now



(Investigator and participants age the same amount of time.)

Design of a *Retrospective* Cohort Study

Past (Sept 2010):

Past (Sept 2010):

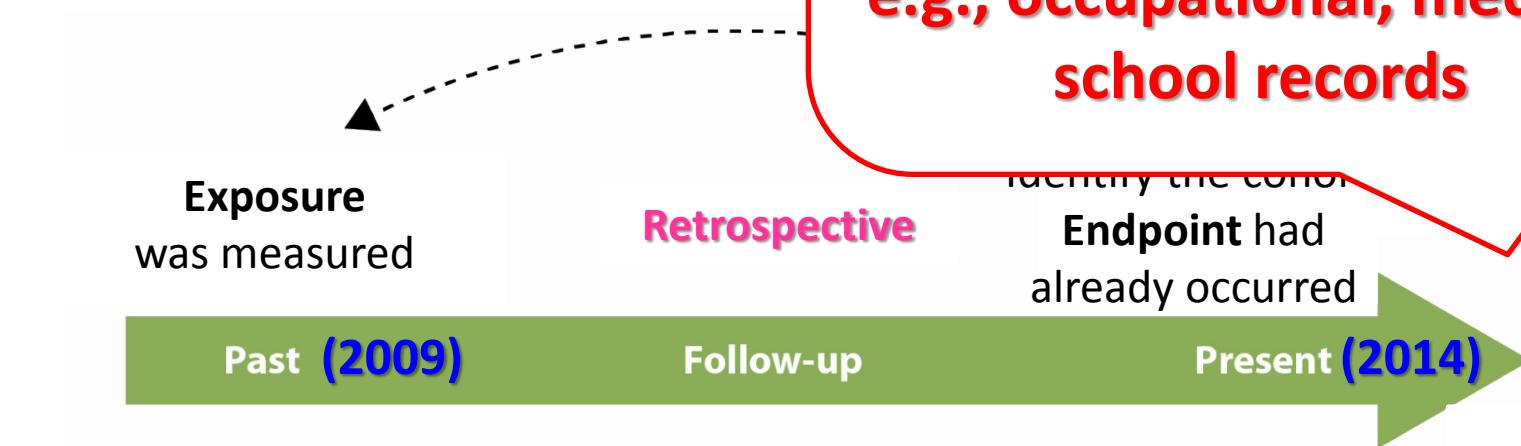
Anytime between Sept 2010
and the Present (Sept 2015):

Exposure
was measured

Retrospective

Past (2009)

Follow-up



Defined Population

Hi! Materials were collected
in the past for non-research
reasons and I'm now finding
and assembling them.
e.g., occupational, medical,
school records

Retrospective vs. Prospective Cohort Studies

RETROSPECTIVE



Exposure measured **Then follow up for the outcome**



PROSPECTIVE



Exposure measured **Then follow up for the outcome**

Is this a prospective or retrospective cohort study?

Dr. Platz asks her research question in Dr. Deal's cohort in 2014



Measure **exposure** or collect biological samples now



Present

Follow-up

Future

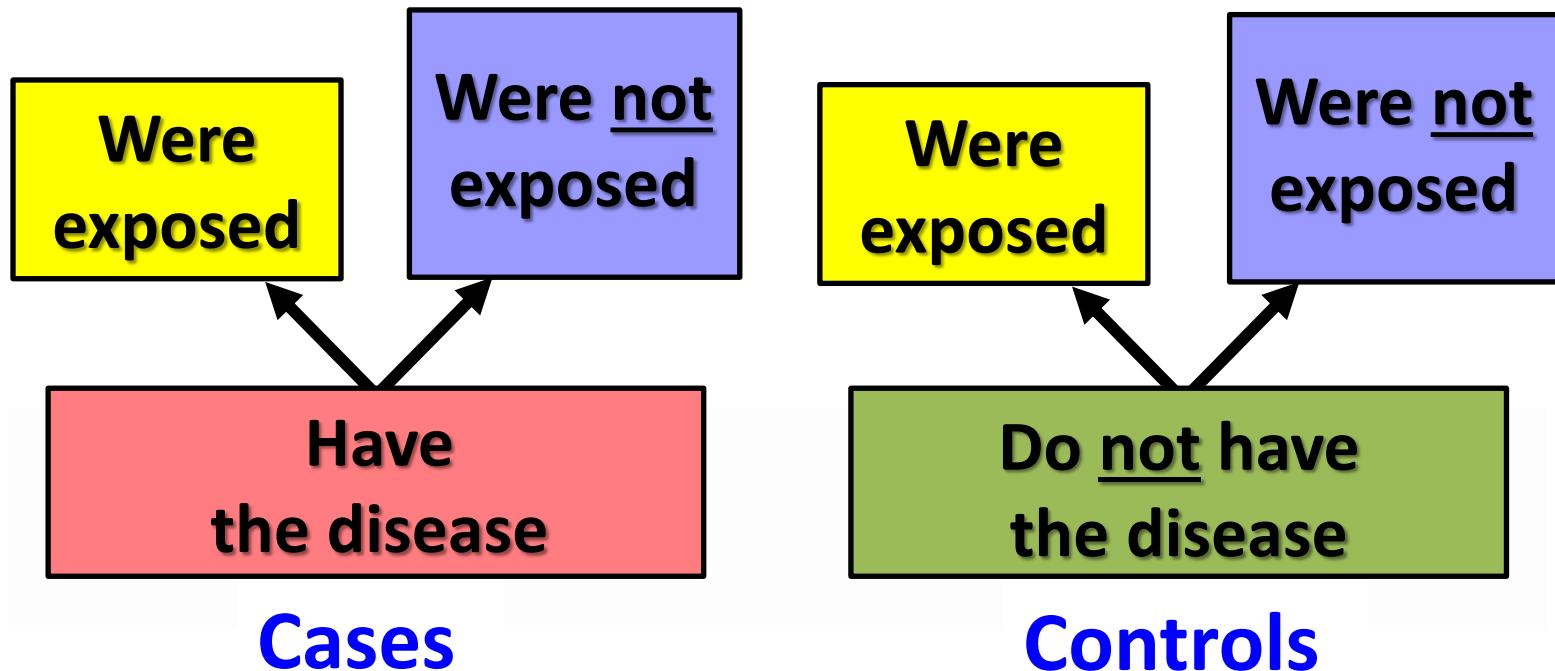
Dr. Deal started a cohort in 2000

Endpoint occurs months to years after exposure is assessed

Answer: Prospective cohort study

Design of a Case-control Study

Then
determine
exposure
history:



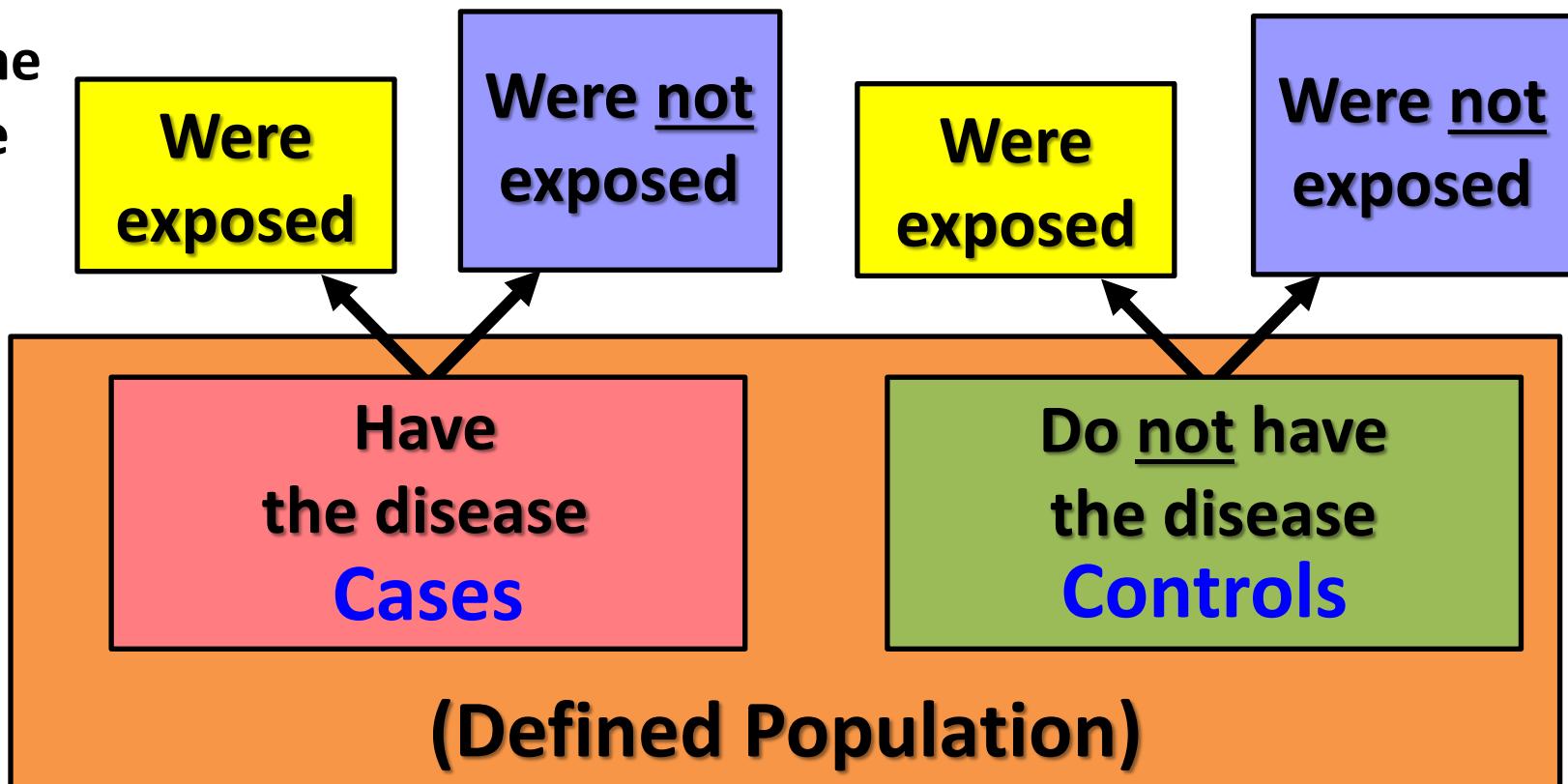
Start
with:

Cases

Controls

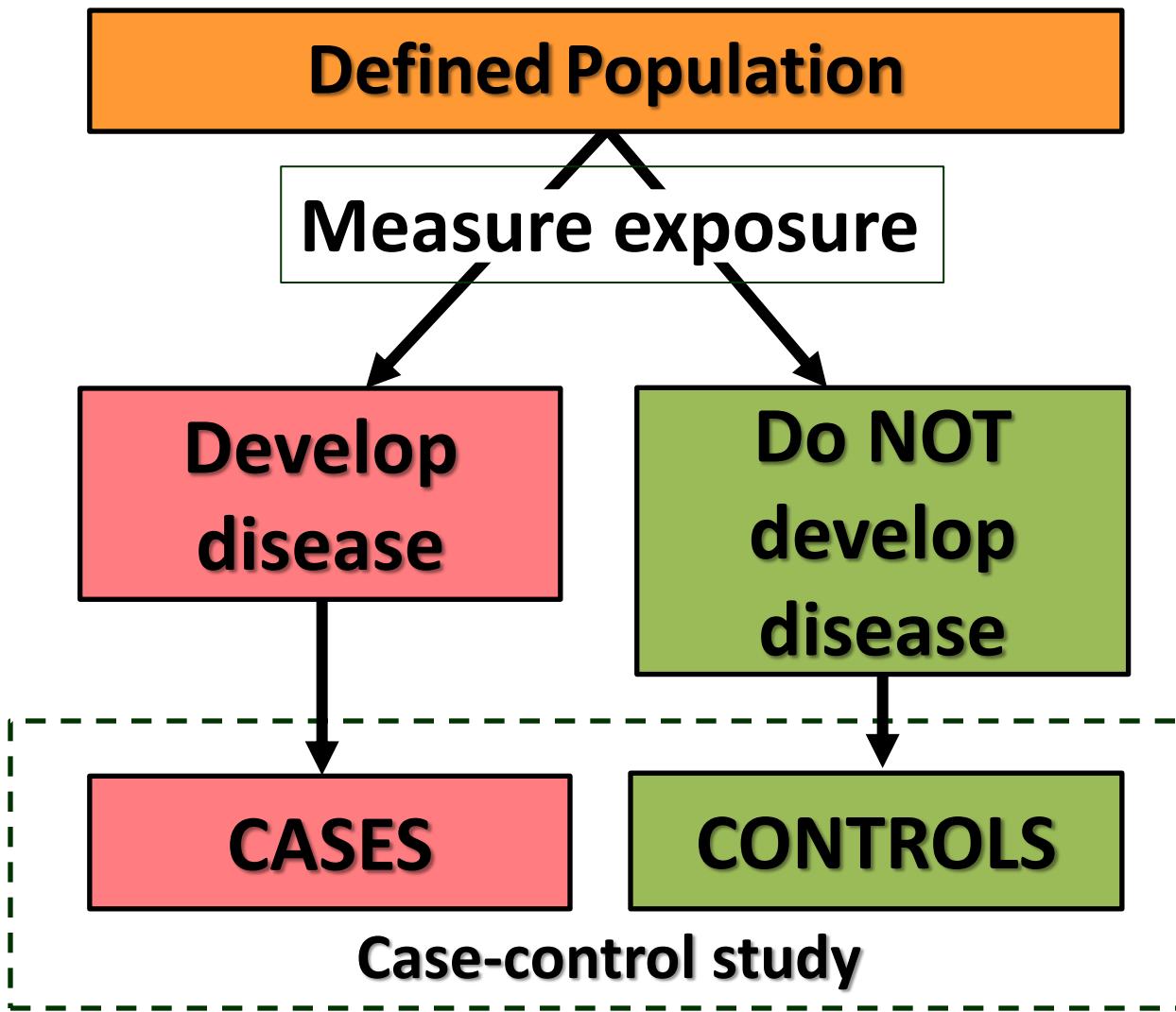
Design of a Case-control Study

Then
determine
exposure
history:



(Controls should come from the same source population as the cases – if they don't, inferences may not be valid)

Design of a case-control study nested in a cohort



Design of a Cross-Sectional Study

Begin with:

Defined Population

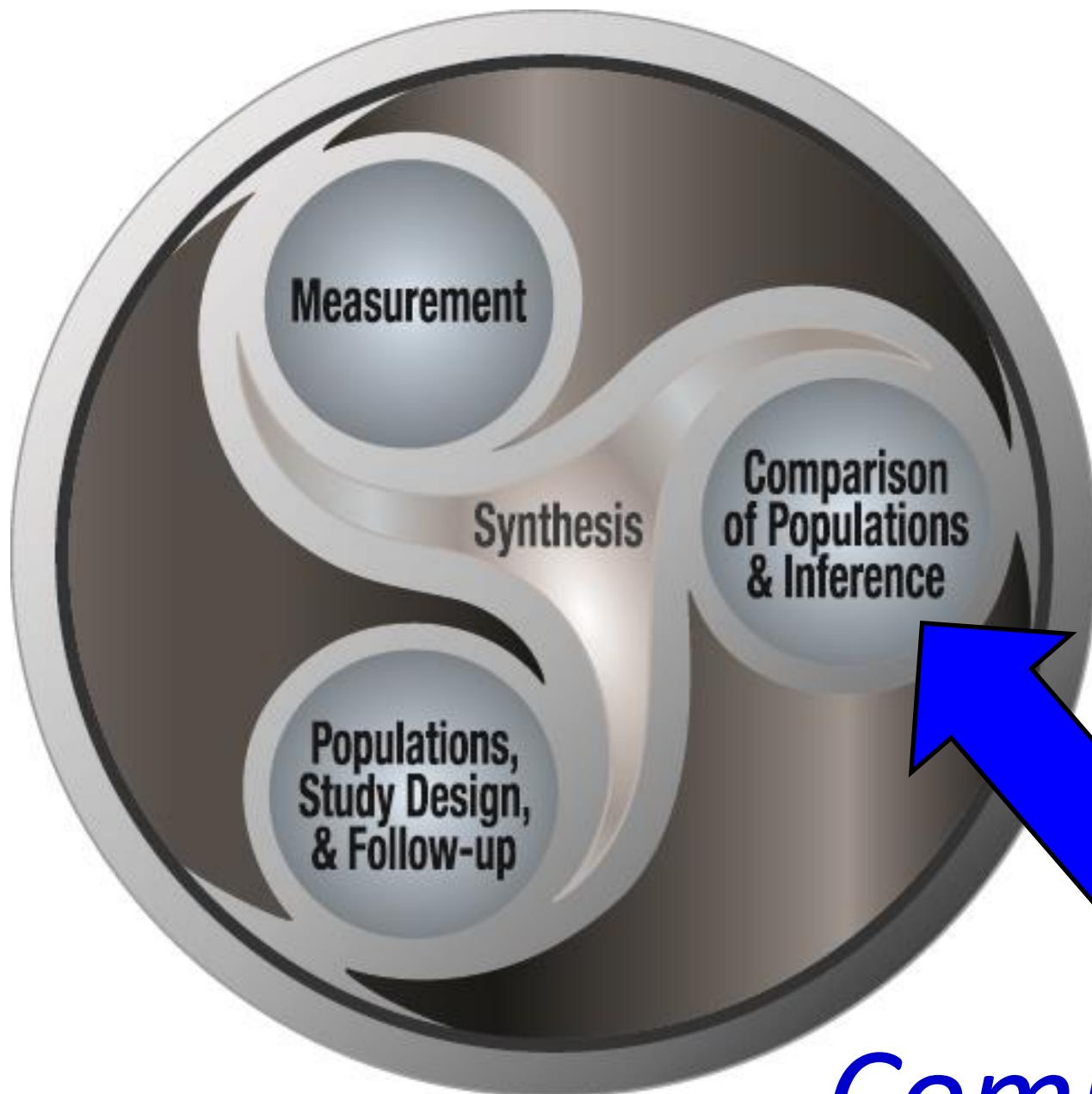
**Gather data on exposure and outcome (or 2 exposures)
e.g., height and weight**

**Exposed;
have disease**

**Exposed;
do not have
disease**

**Not exposed;
have disease**

**Not exposed;
do not have
disease**



Coming soon!

GOOD LUCK!