# Basic Probability

**Prof. dr. Siswanto Agus Wilopo, M.Sc., Sc.D.**
**Department of Biostatistics, Epidemiology and Population Health**
**Faculty of Medicine**
**Universitas Gadjah Mada**

# Learning Objectives

**In this lecture, you learn:**

- Basic probability concepts and definitions
- Conditional probability
- To use Bayes' Theorem to revise probabilities
- Various counting rules


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Important Terms

- **Probability** – the chance that an uncertain event will occur (always between 0 and 1)
- **Event** – Each possible outcome of a variable
- **Simple Event** – an event that can be described by a single characteristic
- **Sample Space** – the collection of all possible events


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Assessing Probability

**There are three approaches to assessing the probability of an uncertain event:**

**1. *a priori* classical probability**

$$\text{probability of occurrence} = \frac{X}{T} = \frac{\text{number of ways the event can occur}}{\text{total number of elementary outcomes}}$$

**2. empirical classical probability**

$$\text{probability of occurrence} = \frac{\text{number of favorable outcomes observed}}{\text{total number of outcomes observed}}$$

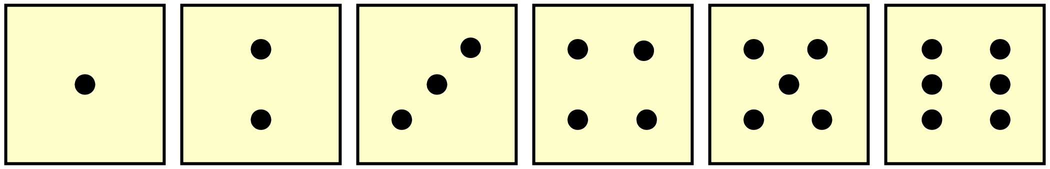
**3. subjective probability**

**an individual judgment or opinion about the probability of occurrence**

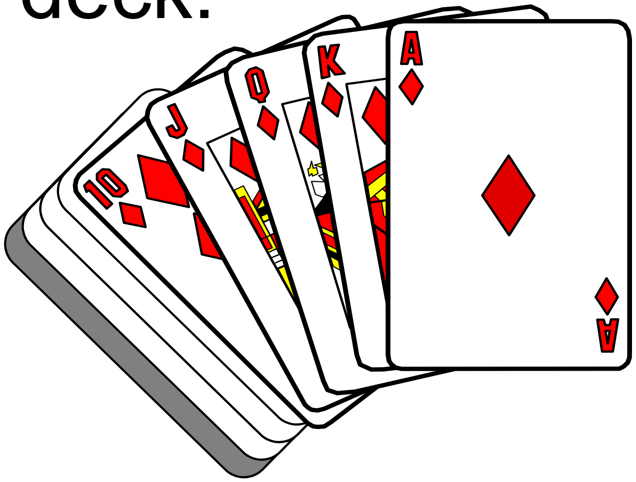
# Sample Space

The Sample Space is the collection of all possible events

e.g. All 6 faces of a die:



e.g. All 52 cards of a bridge deck:






sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Events

- **Simple event**
  - An outcome from a sample space with one characteristic
  - e.g., A red card from a deck of cards
- **Complement of an event A (denoted A')**
  - All outcomes that are not part of event A
  - e.g., All cards that are not diamonds
- **Joint event**
  - Involves two or more characteristics simultaneously
  - e.g., An ace that is also red from a deck of cards


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

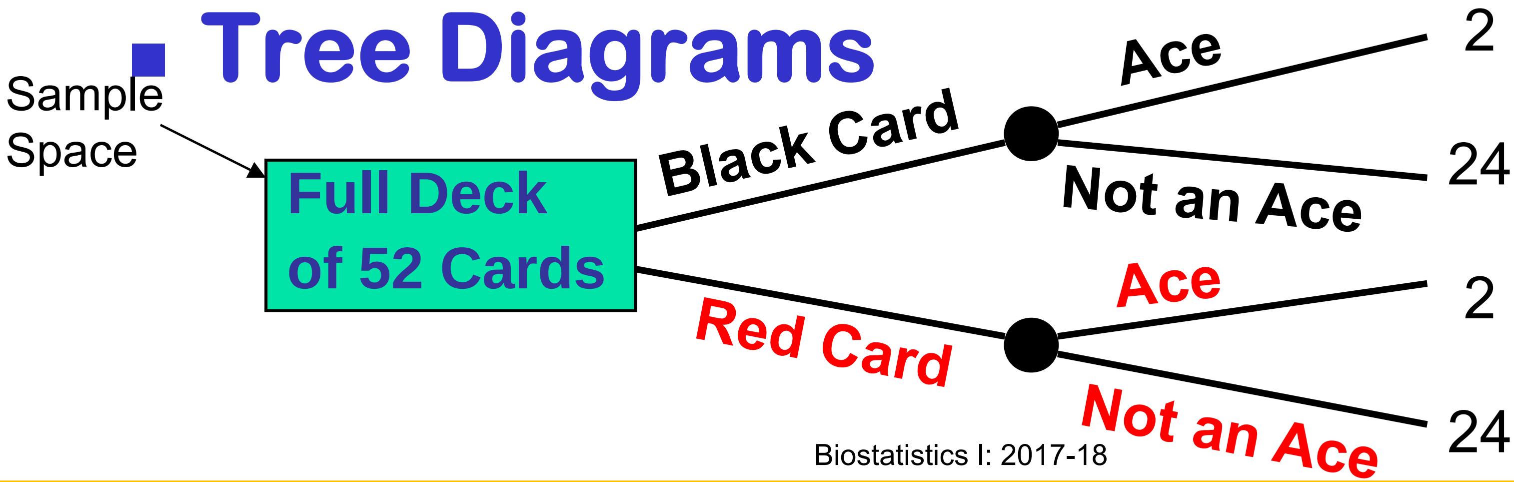
# Visualizing Events

- ## Contingency Tables

| | Ace | Not Ace | Total |
|---|---|---|---|
| Black | 2 | 24 | 26 |
| Red | 2 | 24 | 26 |
| Total | 4 | 48 | 52 |

Sample Space

- ## Tree Diagrams

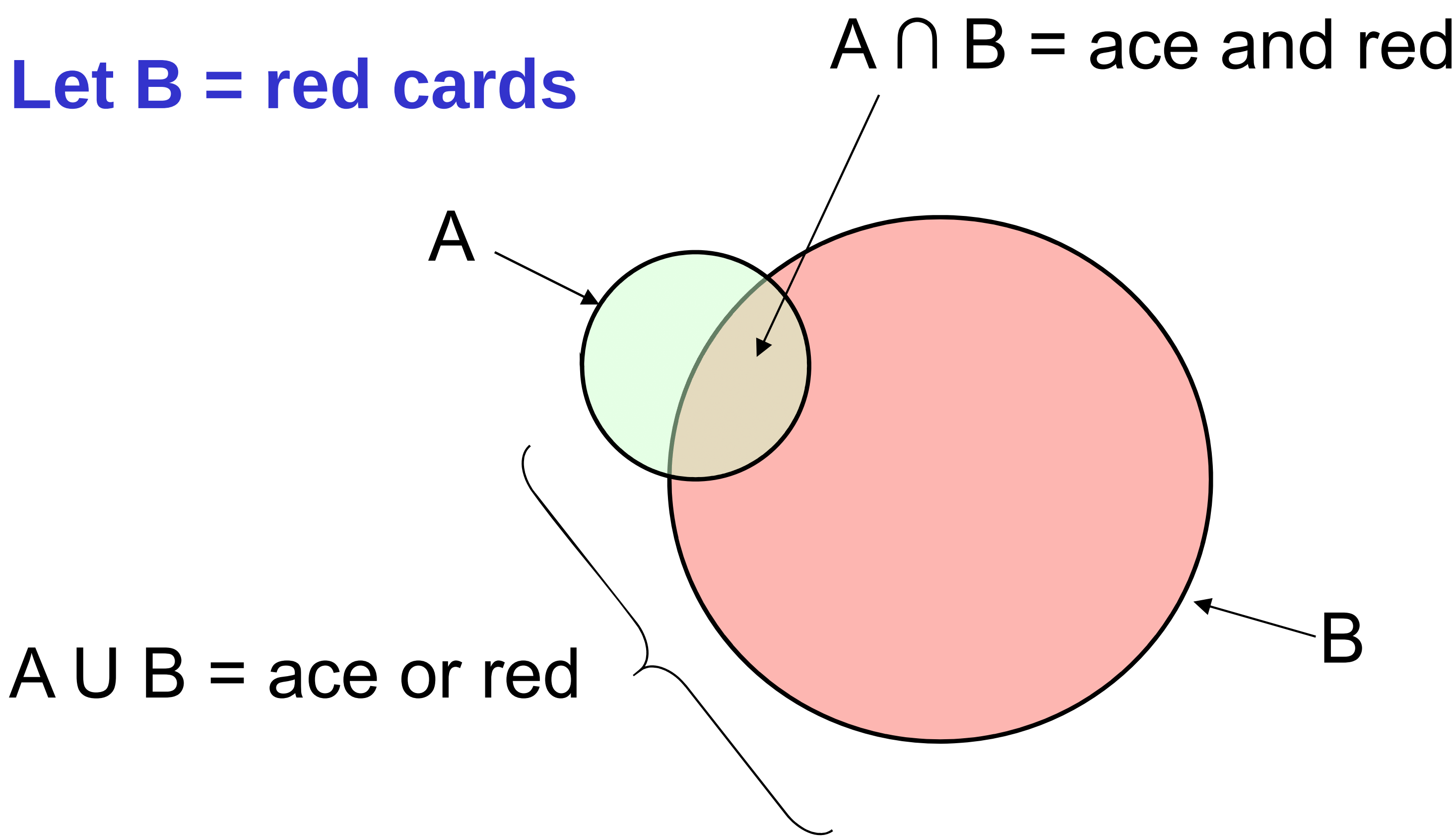




sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Visualizing Events

- **Venn Diagrams**
  - **Let A = aces**
  - **Let B = red cards**







sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Mutually Exclusive Events

- **Mutually exclusive events**
  - **Events that cannot occur together**

**example:**

| |
|---|
| **A = queen of diamonds;  B = queen of clubs** |

- **Events A and B are mutually exclusive**


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Collectively Exhaustive Events

- Collectively exhaustive events
  - One of the events must occur
  - The set of events covers the entire sample space

example:

| |
|---|
| A = aces; B = black cards;<br>C = diamonds; D = hearts |

  - Events A, B, C and D are collectively exhaustive (but not mutually exclusive – an ace may also be a heart)
  - Events B, C and D are collectively exhaustive and also mutually exclusive



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Probability

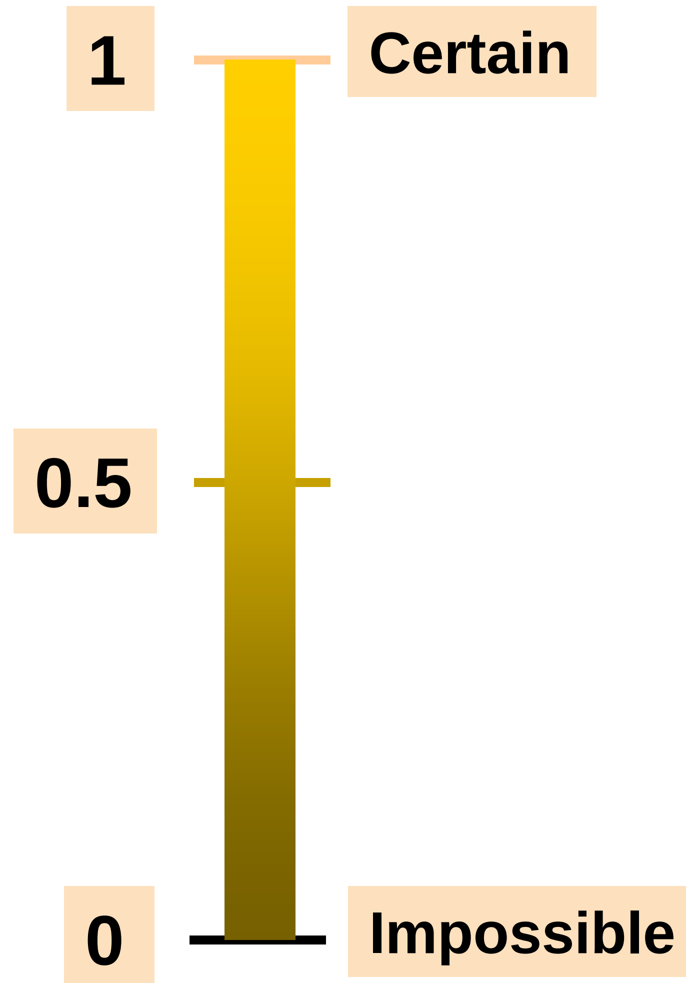
- **Probability is the numerical measure of the likelihood that an event will occur**
- **The probability of any event must be between 0 and 1, inclusively**

$$0 \le P(A) \le 1 \quad \text{For any event A}$$

- **The sum of the probabilities of all mutually exclusive and collectively exhaustive events is 1**

$$P(A) + P(B) + P(C) = 1$$

If A, B, and C are mutually exclusive and collectively exhaustive






sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Computing Joint and Marginal Probabilities

- **The probability of a joint event, A and B:**

$$P(A \text{ and } B) = \frac{\text{number of outcomes satisfying A and B}}{\text{total number of elementary outcomes}}$$

- **Computing a marginal (or simple) probability:**

$$P(A) = P(A \text{ and } B_1) + P(A \text{ and } B_2) + \cdots + P(A \text{ and } B_k)$$

  - Where $B_1$, $B_2$, …, $B_k$ are k mutually exclusive and collectively exhaustive events



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Joint Probability Example

**P(Red and Ace)**

$$= \frac{\text{number of cards that are red and ace}}{\text{total number of cards}} = \frac{2}{52}$$

| Type | Color | | Total |
|---|---|---|---|
| | Red | Black | |
| Ace | 2 | 2 | 4 |
| Non-Ace | 24 | 24 | 48 |
| Total | 26 | 26 | 52 |


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Marginal Probability Example

**P(Ace)**

$$= P(\text{Ace and Red}) + P(\text{Ace and Black}) = \frac{2}{52} + \frac{2}{52} = \frac{4}{52}$$

| Type | Color | | Total |
|---|---|---|---|
| | **Red** | **Black** | |
| **Ace** | **2** | **2** | **4** |
| **Non-Ace** | **24** | **24** | **48** |
| **Total** | **26** | **26** | **52** |



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Joint Probabilities Using Contingency Table

| Event | Event B$_1$ | Event B$_2$ | Total |
|---|---|---|---|
| A$_1$ | P(A$_1$ and B$_1$) | P(A$_1$ and B$_2$) | P(A$_1$) |
| A$_2$ | P(A$_2$ and B$_1$) | P(A$_2$ and B$_2$) | P(A$_2$) |
| Total | P(B$_1$) | P(B$_2$) | 1 |

Joint Probabilities

Marginal (Simple) Probabilities



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# General Addition Rule

General Addition Rule:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

If A and B are mutually exclusive, then

P(A and B) = 0, so the rule can be simplified:

$$P(A \text{ or } B) = P(A) + P(B)$$

For mutually exclusive events A and B



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# General Addition Rule Example

P(Red or Ace) = P(Red) +P(Ace) - P(Red and Ace)

$$= 26/52 + 4/52 - \boxed{2/52} = 28/52$$

Don't count the two red aces twice!

| Type | Color | | Total |
|---|---|---|---|
| | Red | Black | |
| Ace | 2 | 2 | 4 |
| Non-Ace | 24 | 24 | 48 |
| Total | 26 | 26 | 52 |



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Computing Conditional Probabilities

- **A conditional probability is the probability of one event, given that another event has occurred:**

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)}$$

The conditional probability of A given that B has occurred

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

The conditional probability of B given that A has occurred

Where P(A and B) = joint probability of A and B

P(A) = marginal probability of A

P(B) = marginal probability of B



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Conditional Probability Example

- Of the cars on a used car lot, 70% have air conditioning (AC) and 40% have a CD player (CD). 20% of the cars have both.

- **What is the probability that a car has a CD player, given that it has AC ?**

**i.e., we want to find P(CD | AC)**



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Conditional Probability Example

*(continued)*

- Of the cars on a used car lot, **70%** have air conditioning (AC) and **40%** have a CD player (CD).
  **20%** of the cars have both.

| | CD | No CD | Total |
|---|---|---|---|
| **AC** | **0.2** | **0.5** | **0.7** |
| **No AC** | **0.2** | **0.1** | **0.3** |
| **Total** | **0.4** | **0.6** | **1.0** |

$$P(CD \mid AC) = \frac{P(CD \text{ and } AC)}{P(AC)} = \frac{0.2}{0.7} = 0.2857$$



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Conditional Probability Example

*(continued)*

- Given AC, we only consider the top row (70% of the cars). Of these, 20% have a CD player. 20% of 70% is about 28.57%.

| | CD | No CD | Total |
|---|---|---|---|
| AC | 0.2 | 0.5 | 0.7 |
| No AC | 0.2 | 0.1 | 0.3 |
| Total | 0.4 | 0.6 | 1.0 |

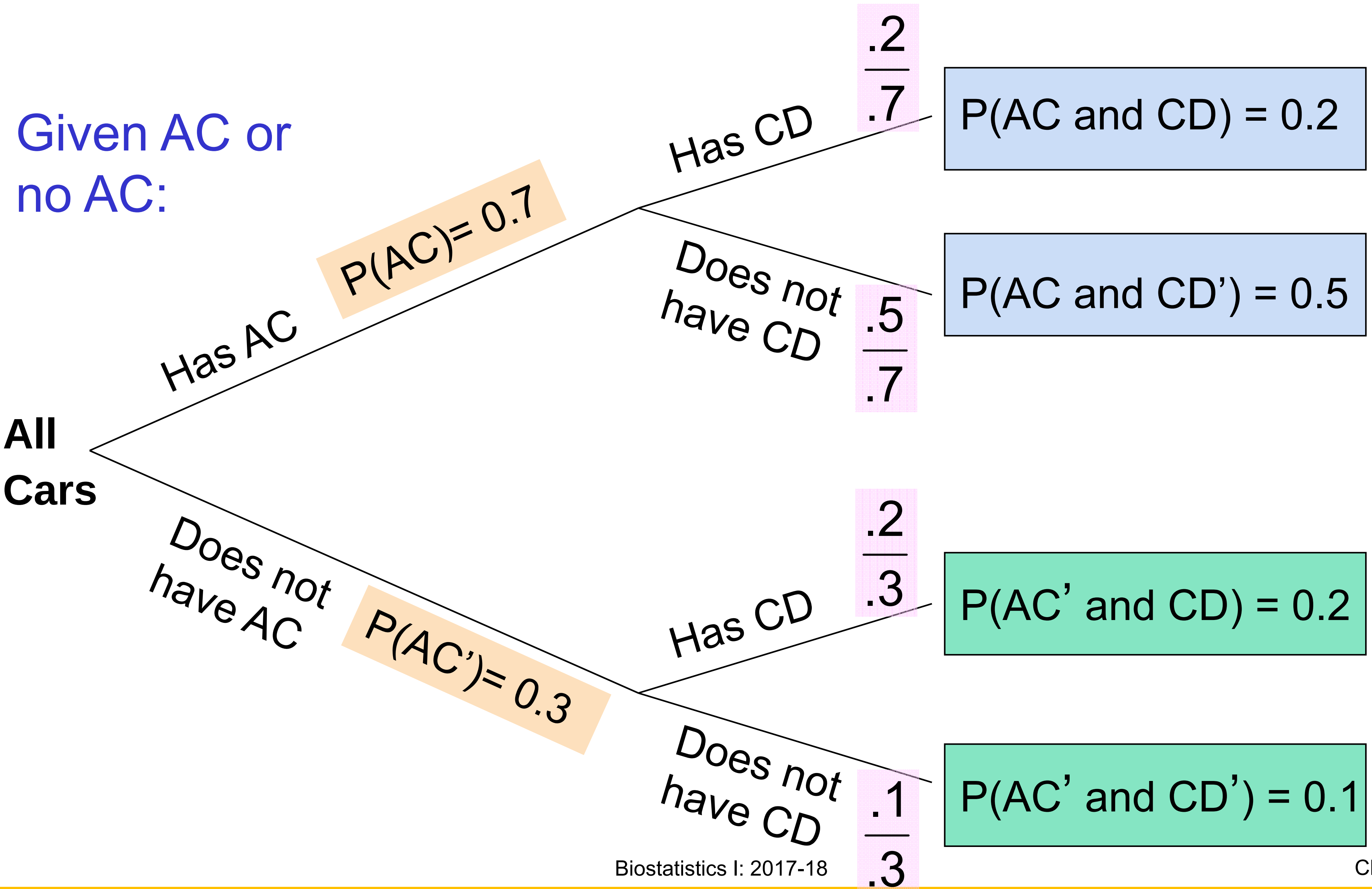
$$P(\text{CD} \mid \text{AC}) = \frac{P(\text{CD and AC})}{P(\text{AC})} = \frac{0.2}{0.7} = 0.2857$$



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Using Decision Trees

Given AC or no AC:



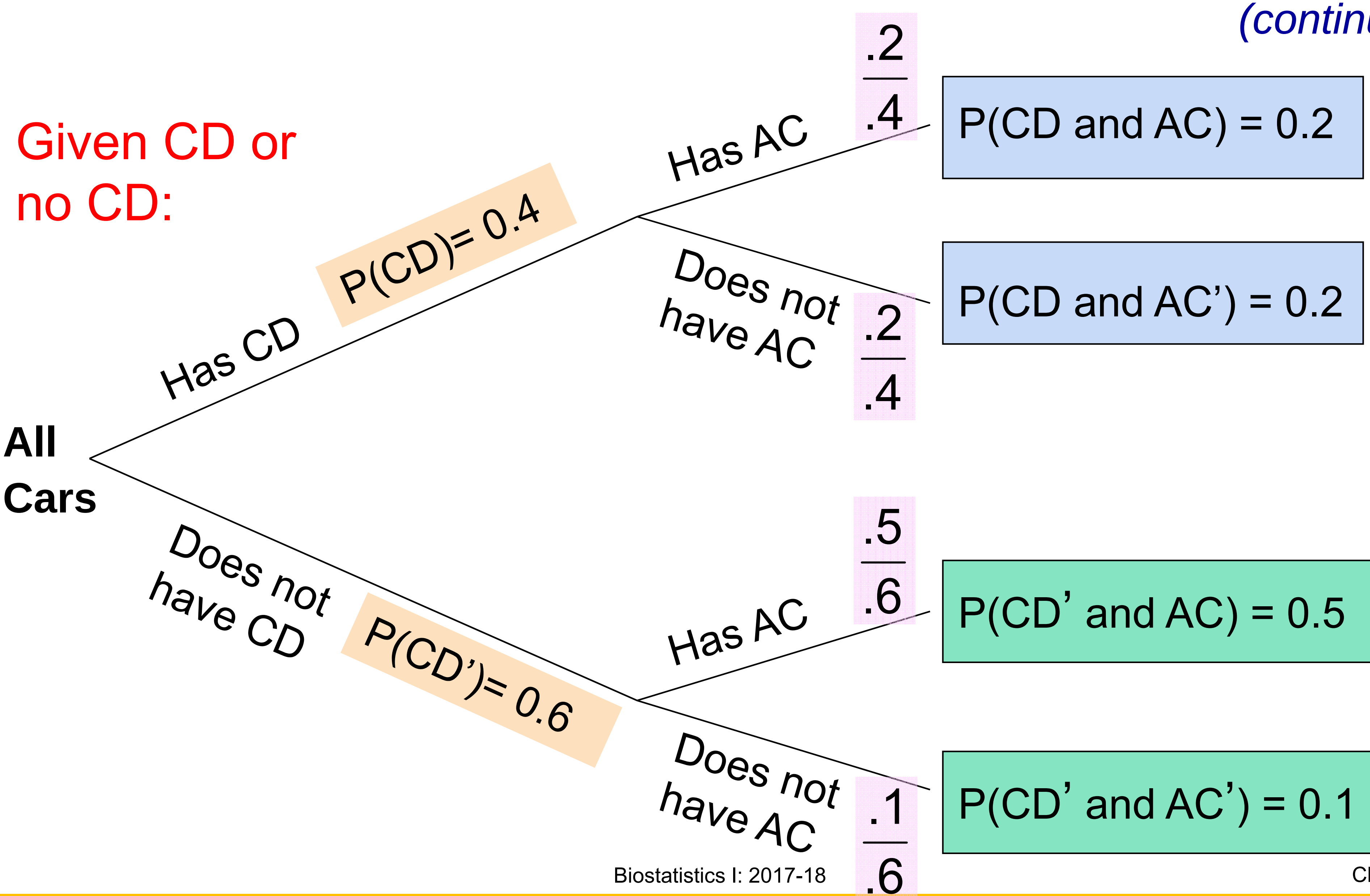



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Using Decision Trees

*(continued)*

Given CD or no CD:






sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Statistical Independence

- Two events are **independent** if and only if:

$$\boxed{P(A \mid B) = P(A)}$$

- Events A and B are independent when the probability of one event is not affected by the other event


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Multiplication Rules

- **Multiplication rule for two events A and B:**

$$P(A \text{ and } B) = P(A|B)P(B)$$

**Note:** If A and B are independent, then $P(A|B) = P(A)$ and the multiplication rule simplifies to

$$P(A \text{ and } B) = P(A)P(B)$$


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Bayes' Theorem

Bayes' theorem is used to revise previously calculated probabilities after new information is obtained

$$P(B_i \mid A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_k)P(B_k)}$$

- **where:**

  **$B_i$ = $i^{th}$ event of k mutually exclusive and collectively exhaustive events**

  **A = new event that might impact $P(B_i)$**



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Marginal Probability

- **Marginal probability for event A:**

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_k)P(B_k)$$

  - **Where $B_1$, $B_2$, …, $B_k$ are k mutually exclusive and collectively exhaustive events**

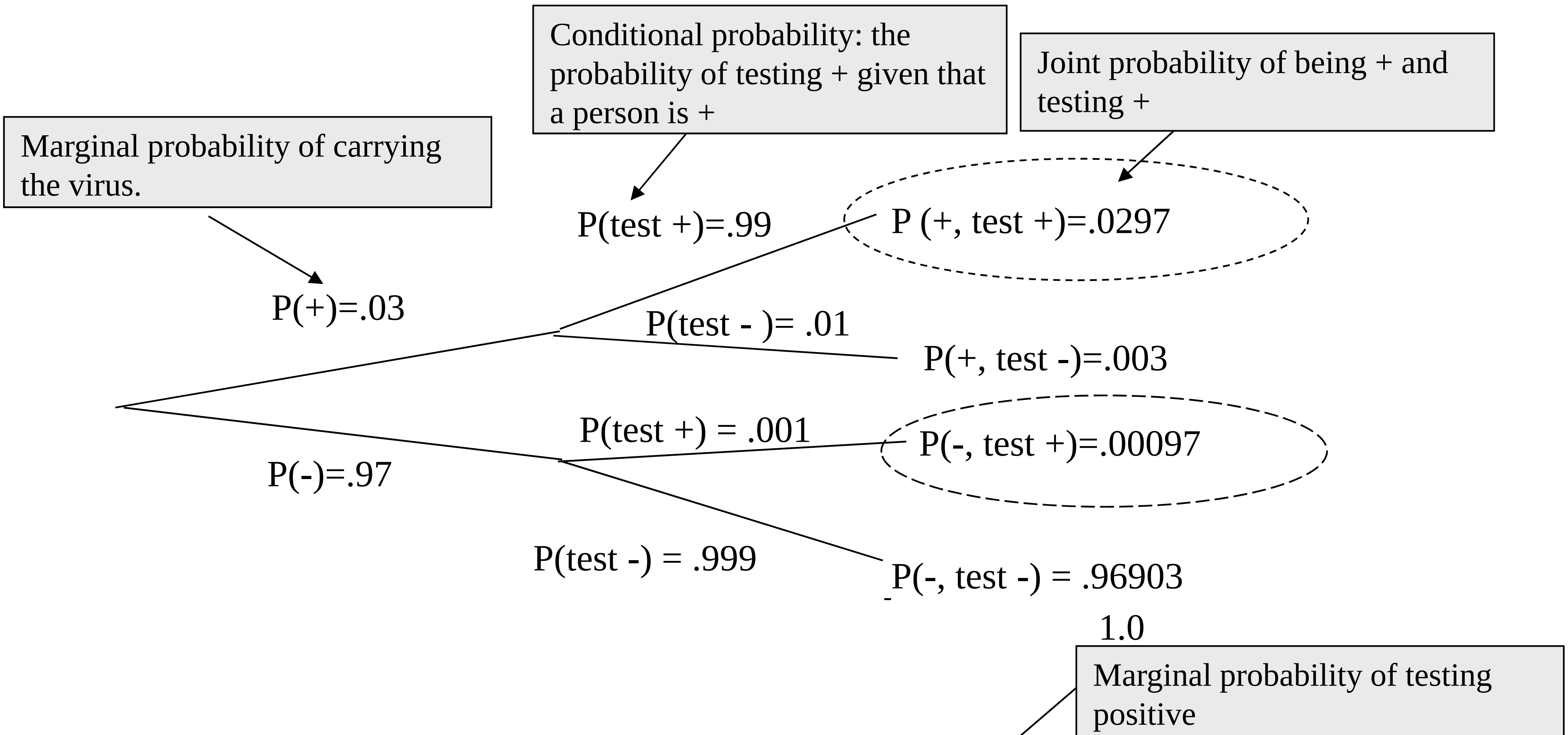

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Practice problem

If HIV has a prevalence of 3% in Jayapura, and a particular HIV test has a false positive rate of .001 and a false negative rate of .01, what is the probability that a random person selected off the street will test positive?


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Answer




$$\therefore P(\text{test }+)=.0297+.00097=.03067$$

$$P(+\&\text{test}+)\neq P(+)*P(\text{test}+)$$

$$.0297 \neq .03*.03067\ (=.00092)$$

$\therefore$ Dependent!




sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Law of total probability

$$P(test+) = P(test+/HIV+)P(HIV+) + P(test+/HIV-)P(HIV-)$$

One of these has to be true (mutually exclusive, collectively exhaustive). They sum to 1.0.

$$P(test+) = .99(.03) + .001(.97)$$


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

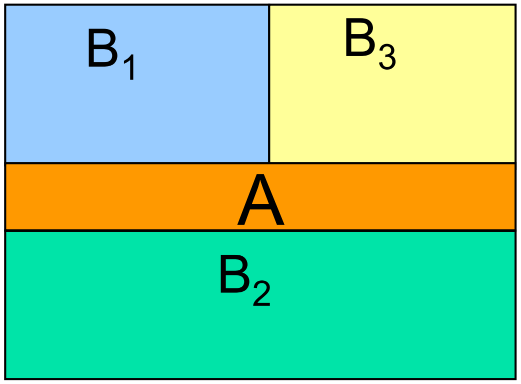
# Law of total probability

- **Formal Rule: Marginal probability for event A=**

$$\boxed{P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_k)P(B_k)}$$

$$\sum_{i=1}^{k} B_i = 1.0 \text{ and } P(B_i \& B_j) = 0 \text{ (mutually exclusive)}$$

- Where:




$$\boxed{P(A) = (50\%)(25\%) + (0)(50\%) + \cdots + (50\%)(25\%) = 25\%}$$

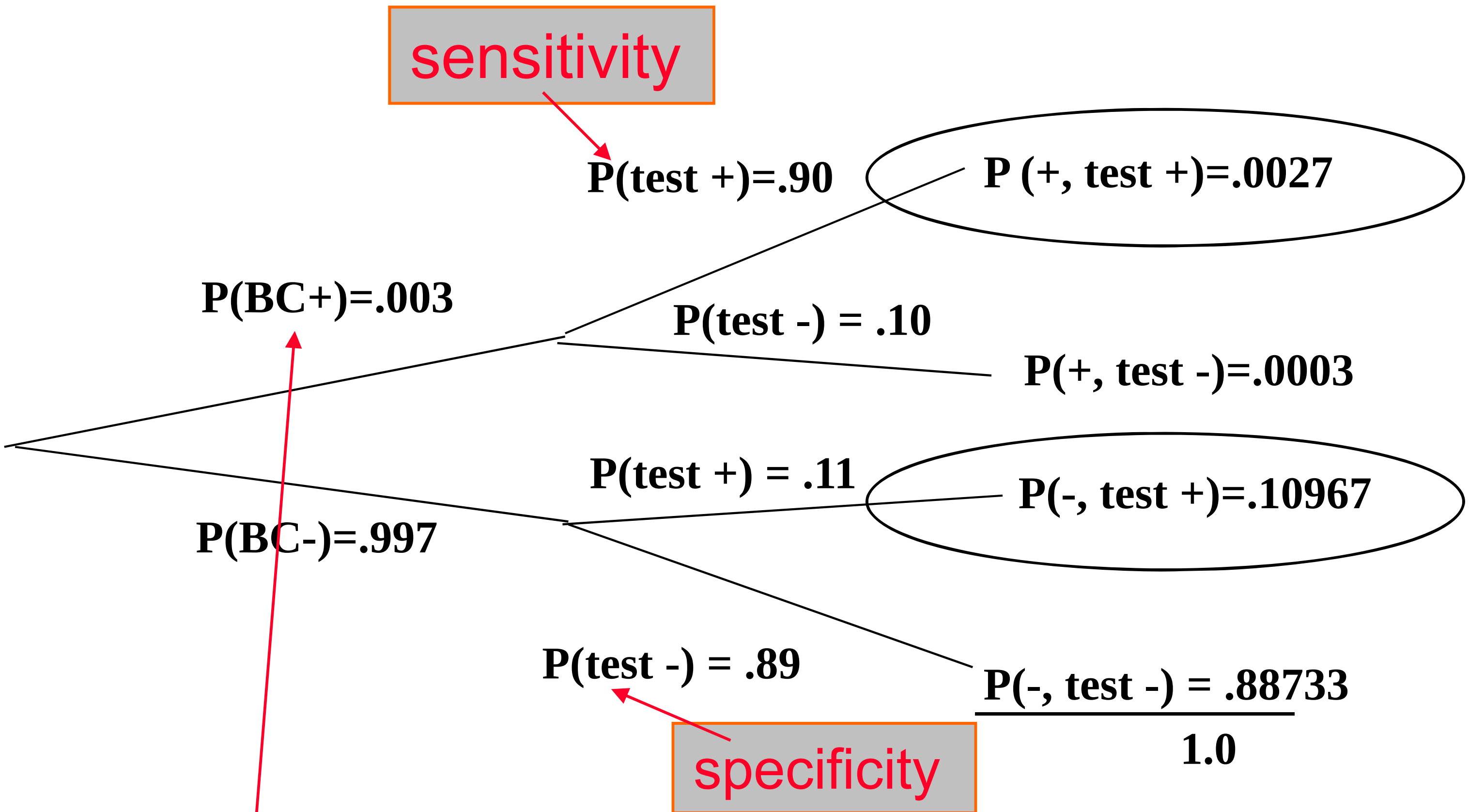


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Example 2

- A 54-year old woman has an abnormal mammogram; what is the chance that she has breast cancer?
- Mammogram sensitivity=0.90 and specificity =0.89

# Example: Mammography




Marginal probabilities of breast cancer….(prevalence among all 54-year olds)

P(BC/test+)=.0027/(.0027+.10967)=2.4%



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# BAYES' RULE



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Bayes' Rule: derivation

- **Definition:**

  **Let A and B be two events with $P(B) \neq 0$. The conditional probability of A given B is:**

$$P(A/B) = \frac{P(A \& B)}{P(B)}$$

The idea: if we are given that the event B occurred, the relevant sample space is reduced to B {P(B)=1 because we know B is true} and conditional probability becomes a probability measure on B.


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Bayes' Rule: derivation

$$P(A/B) = \frac{P(A\&B)}{P(B)}$$

**can be re-arranged to:**

$$P(A\&B) = P(A/B)P(B)$$

and, since also:

$$P(B/A) = \frac{P(A\&B)}{P(A)} \qquad \therefore P(A\&B) = P(B/A)P(A)$$

$$P(A/B)P(B) = P(A\&B) = P(B/A)P(A)$$

$$P(A/B)P(B) = P(B/A)P(A)$$

$$\therefore P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$

# Bayes' Rule:

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$

**OR**

$$P(A/B) = \frac{P(B/A)P(A)}{P(B/A)P(A) + P(B/\sim A)P(\sim A)}$$

From the "Law of Total Probability"



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Bayes' Rule:

- Why do we care??
- Why is Bayes' Rule useful??
- It turns out that sometimes it is very useful to be able to "flip" conditional probabilities. That is, we may know the probability of A given B, but the probability of B given A may not be obvious. An example will help…

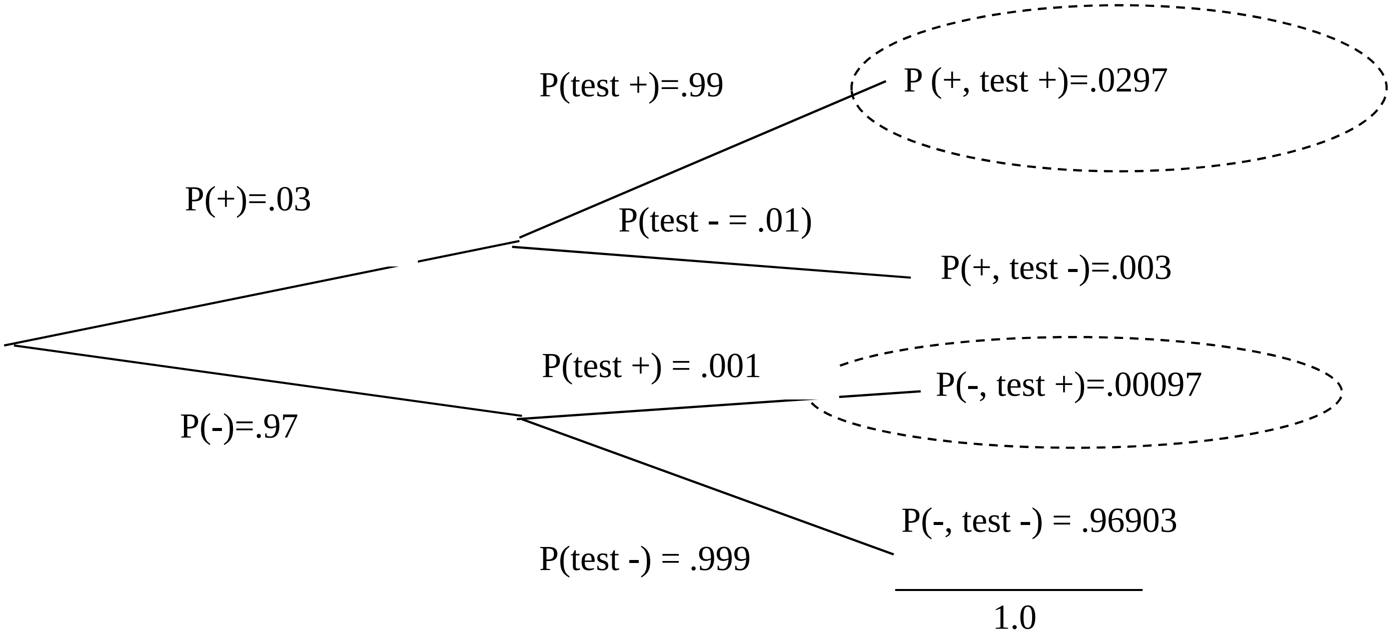

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# In-Class Exercise

- If HIV has a prevalence of 3% in Jayapura, and a particular HIV test has a false positive rate of .001 and a false negative rate of .01, what is the probability that a random person who tests positive is actually infected (also known as “positive predictive value”)?


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Answer: using probability tree




A positive test places one on either of the two “test +” branches.
But only the top branch also fulfills the event “true infection.”
Therefore, the probability of being infected is the probability of being on the top branch given that you are on one of the two circled branches above.

$$P(+/test+) = \frac{P(test+\&true+)}{P(test+)} = \frac{.0297}{.0297+.00097} = 96.8\%$$


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Answer: using Bayes' rule

$$P(true+/test+) = \frac{P(test+/true+)P(true+)}{P(test+/true+)P(true+) + P(test+/true-)P(true-)} =$$

$$\frac{.99(.03)}{.99(.03) + .001(.97)} = 96.8\%$$



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# In-class exercise

**An insurance company believes that drivers can be divided into two classes—those that are of high risk and those that are of low risk. Their statistics show that a high-risk driver will have an accident at some time within a year with probability .4, but this probability is only .1 for low risk drivers.**

a) **Assuming that 20% of the drivers are high-risk, what is the probability that a new policy holder will have an accident within a year of purchasing a policy?**
b) **If a new policy holder has an accident within a year of purchasing a policy, what is the probability that he is a high-risk type driver?**


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Answer to (a)

**Assuming that 20% of the drivers are of high-risk, what is the probability that a new policy holder will have an accident within a year of purchasing a policy?**

**Use law of total probability:**

**P(accident)=**

**P(accident/high risk)*P(high risk) +**

**P(accident/low risk)*P(low risk) =**

**.40(.20) + .10(.80) = .08 + .08 = .16**



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Answer to (b)

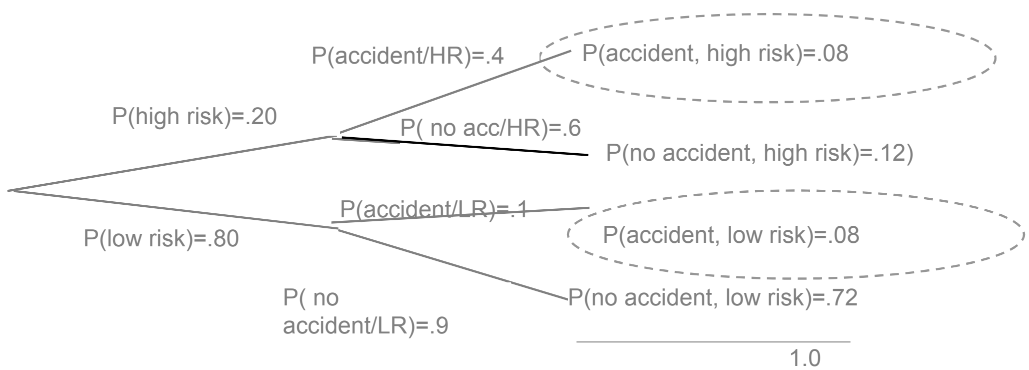
**If a new policy holder has an accident within a year of purchasing a policy, what is the probability that he is a high-risk type driver?**

**P(high-risk/accident)=**

**P(accident/high risk)*P(high risk)/P(accident)**

**=.40(.20)/.16 = 50%**

**Or use tree:**




P(high risk/accident)=.08/.16=50%

# Conditional Probability for Epidemiology:

## The odds ratio and risk ratio as conditional probability


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# The Risk Ratio and the Odds Ratio as conditional probability

In epidemiology, the association between a risk factor or protective factor (exposure) and a disease may be evaluated by the "risk ratio" (RR) or the "odds ratio" (OR).

Both are measures of "relative risk"—the general concept of comparing disease risks in exposed vs. unexposed individuals.


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Odds and Risk (probability)

Definitions:

Risk = P(A) = cumulative probability (you specify the time period!)

For example, what's the probability that a person with a high sugar intake develops diabetes in 1 year, 5 years, or over a lifetime?

Odds = P(A)/P(~A)

For example, "the odds are 3 to 1 against a horse" means that the horse has a 25% probability of winning.

Note: An odds is always higher than its corresponding probability, unless the probability is 100%.



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Odds vs. Risk=probability

| If the risk is... | Then the odds are... |
|---|---|
| ½ (50%) | **1:1** |
| ¾ (75%) | **3:1** |
| 1/10 (10%) | **1:9** |
| 1/100 (1%) | **1:99** |

Note: An odds is always higher than its corresponding probability, unless the probability is 100%.



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Introduction to the 2x2 Table

| | Exposure (E) | No Exposure (~E) | |
|---|---|---|---|
| Disease (D) | a | b | a+b = P(D) |
| No Disease (~D) | c | d | c+d = P(~D) |
| | a+c = P(E) | b+d = P(~E) | |

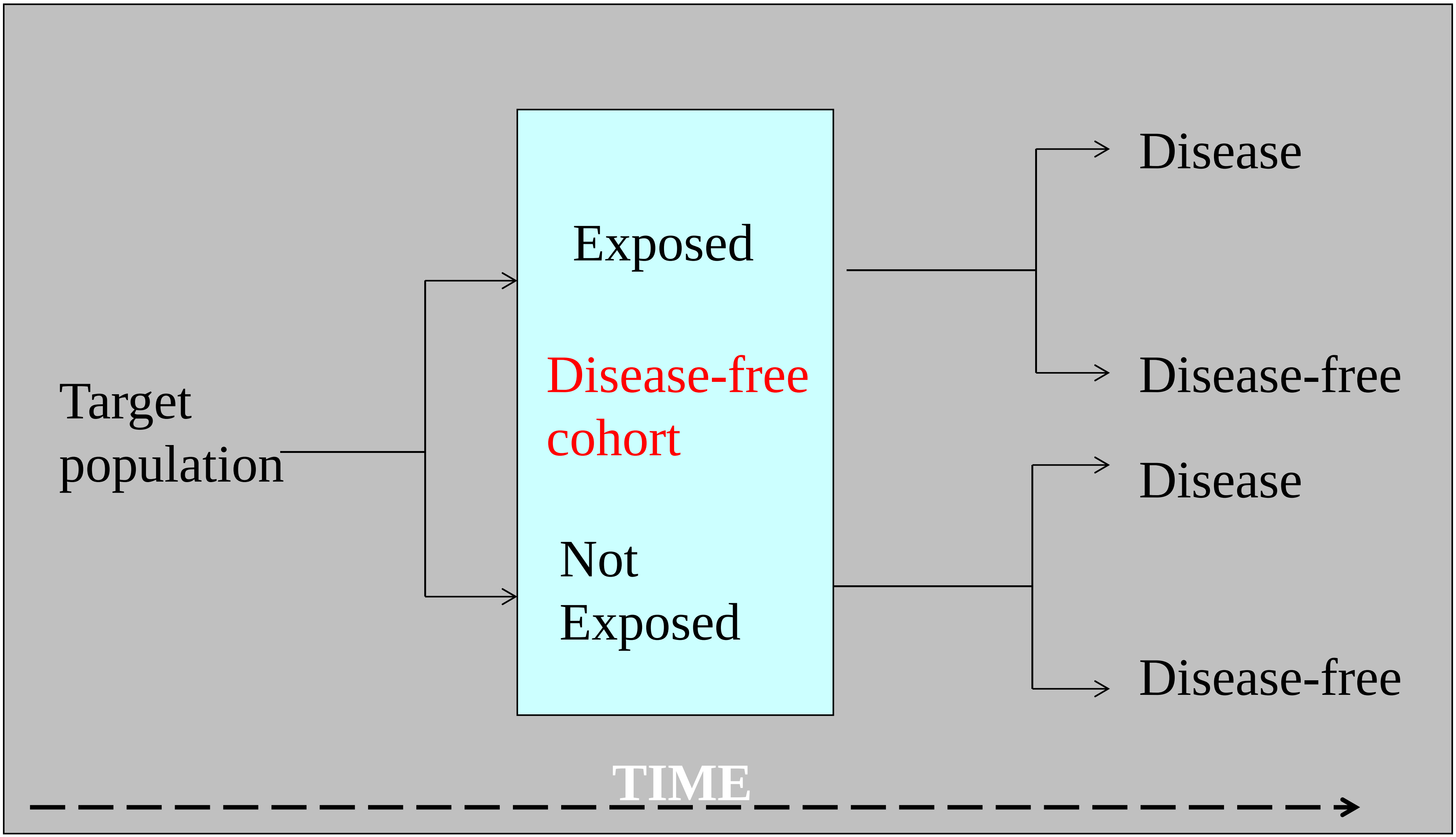
Marginal probability of exposure

Marginal probability of disease



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Cohort Studies






sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# The Risk Ratio, or Relative Risk (RR)

| | Exposure (E) | No Exposure (~E) |
|---|---|---|
| Disease (D) | a | b |
| No Disease (~D) | c | d |
| | a+c | b+d |

risk to the exposed

$$RR = \frac{P(D/E)}{P(D/\sim E)} = \frac{\boxed{a/(a+c)}}{\boxed{b/(b+d)}}$$

risk to the unexposed



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Hypothetical Data

| | High Systolic BP | Normal BP |
|---|---|---|
| Congestive Heart Failure | 400 | 400 |
| No CHF | 1100 | 2600 |
| | 1500 | 3000 |

$$RR = \frac{400 / 1500}{400 / 3000} = 2.0$$


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# The odds ratio…

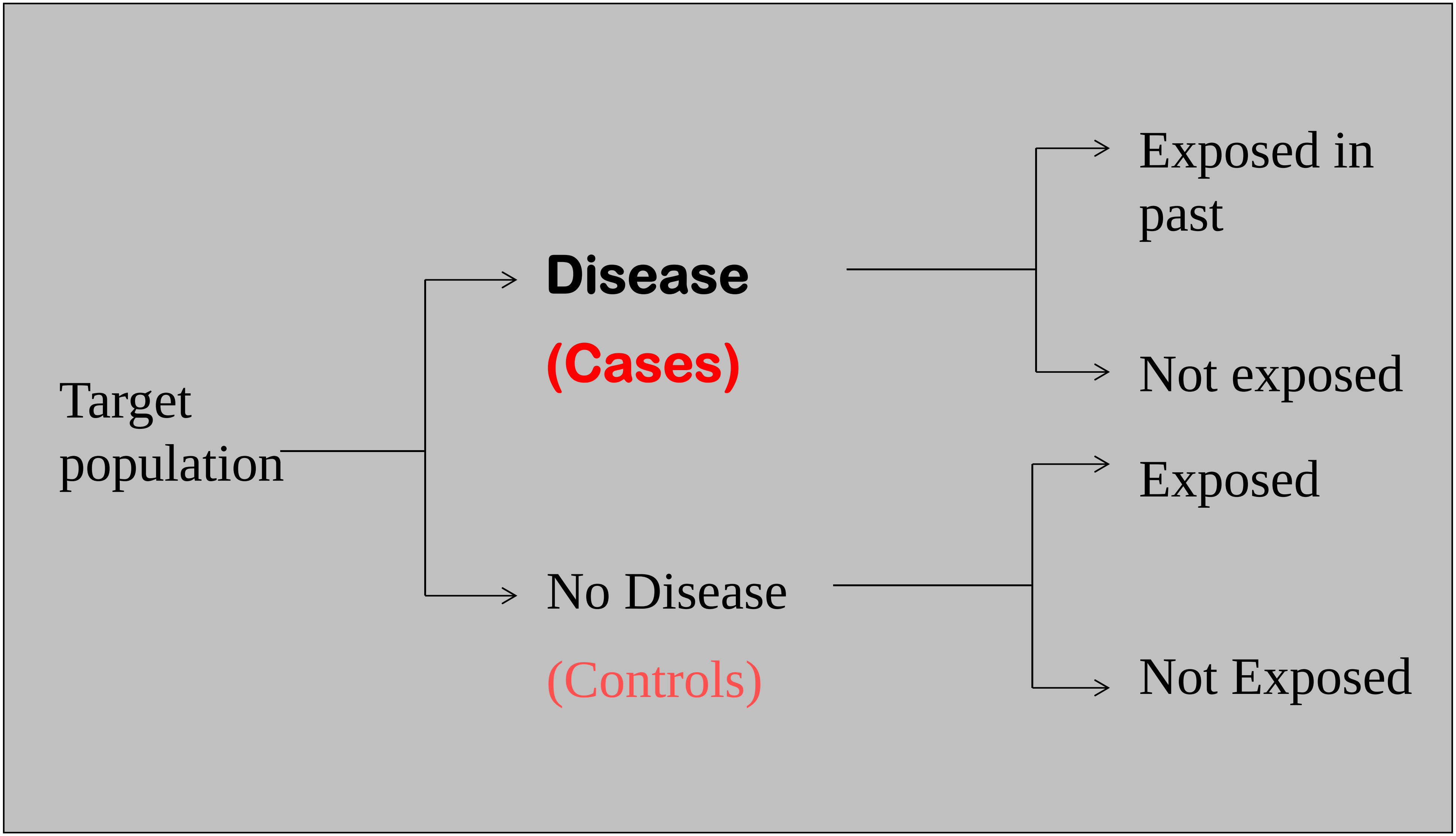
- This risk ratio seems like the perfect measure of relative risk. Why not stop here? Why introduce the more complicated odds ratio??

- We cannot calculate a risk ratio from a case-control study. Case-control studies are a popular study design in epidemiology, because they are useful for studying rare diseases.

- In a case-control study, we sample *conditional on disease status*, so we cannot calculate risk of disease.


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Case-Control Studies





sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# The Odds Ratio (OR)

| | Hep C + | Hep C - | |
|---|---|---|---|
| **Cases: Liver cancer** | **90** | **10** | **100** |
| **Controls** | **30** | **70** | **100** |

What are P(D/E) and P(D/~E) here?

*We can't tell, because, by design, we have fixed the proportion of liver cancer cases in this sample at 50% simply by selecting half controls and half cases.*

All these data give us is: P(E/D) and P(E/~D).


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# The Odds Ratio (OR)

| | Hep C + | Hep C - | |
|---|---|---|---|
| Cases: Liver cancer | 90 | 10 | 100 |
| Controls | 30 | 70 | 100 |

by Bayes' Rule…

Luckily, P(E/D) [=the quantity you have] → $P(E/D) = \frac{P(D/E)P(E)}{P(D)}$

Unfortunately, our sampling scheme precludes calculation of the marginals: P(E) and P(D), but turns out we don't need these if we use an odds ratio because the marginals cancel out!



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

$$\frac{\;\dfrac{P(E/D)}{P(\sim E/D)}\;}{\;\dfrac{P(E/\sim D)}{P(\sim E/\sim D)}\;}$$

Odds of exposure in the cases

Odds of exposure in the controls

Bayes' Rule

$$\frac{\;\dfrac{P(D/E)\cancel{P(E)}}{\cancel{P(D)}}\;}{\;\dfrac{P(D/\sim E)\cancel{P(\sim E)}}{\cancel{P(D)}}\;} \Bigg/ \frac{\;\dfrac{P(\sim D/E)\cancel{P(E)}}{\cancel{P(\sim D)}}\;}{\;\dfrac{P(\sim D/\sim E)\cancel{P(\sim E)}}{\cancel{P(\sim D)}}\;} =$$

$$\frac{\;\dfrac{P(D/E)}{P(\sim D/E)}\;}{\;\dfrac{P(D/\sim E)}{P(\sim D/\sim E)}\;}$$

Odds of disease in the exposed

Odds of disease in the unexposed

What we want!



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# The Odds Ratio (OR)

| | Exposure (E) | No Exposure (~E) |
|---|---|---|
| Disease (D) | a | b |
| No Disease (~D) | c | d |

$$OR = \frac{\boxed{\frac{a}{b}}}{\boxed{\frac{c}{d}}} = \frac{ad}{bc} = \frac{\boxed{\frac{a}{c}}}{\boxed{\frac{b}{d}}}$$

Odds of exposure for the cases.

Odds of exposure for the controls

Odds of disease for the exposed

Odds of disease for the unexposed



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# The rare disease assumption

$$OR = \frac{\dfrac{P(D/E)}{\cancel{P(\sim D/E)}_{\to 1}}}{\dfrac{P(D/\sim E)}{\cancel{P(\sim D/\sim E)}_{\to 1}}} \approx \frac{P(D/E)}{P(D/\sim E)} = RR$$

When a disease is rare:
P(~D) = 1 - P(D) ≅ 1



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Example

| | Hep C + | Hep C - |
|---|---|---|
| Cases: Liver cancer | 90 | 10 |
| Controls | 30 | 70 |

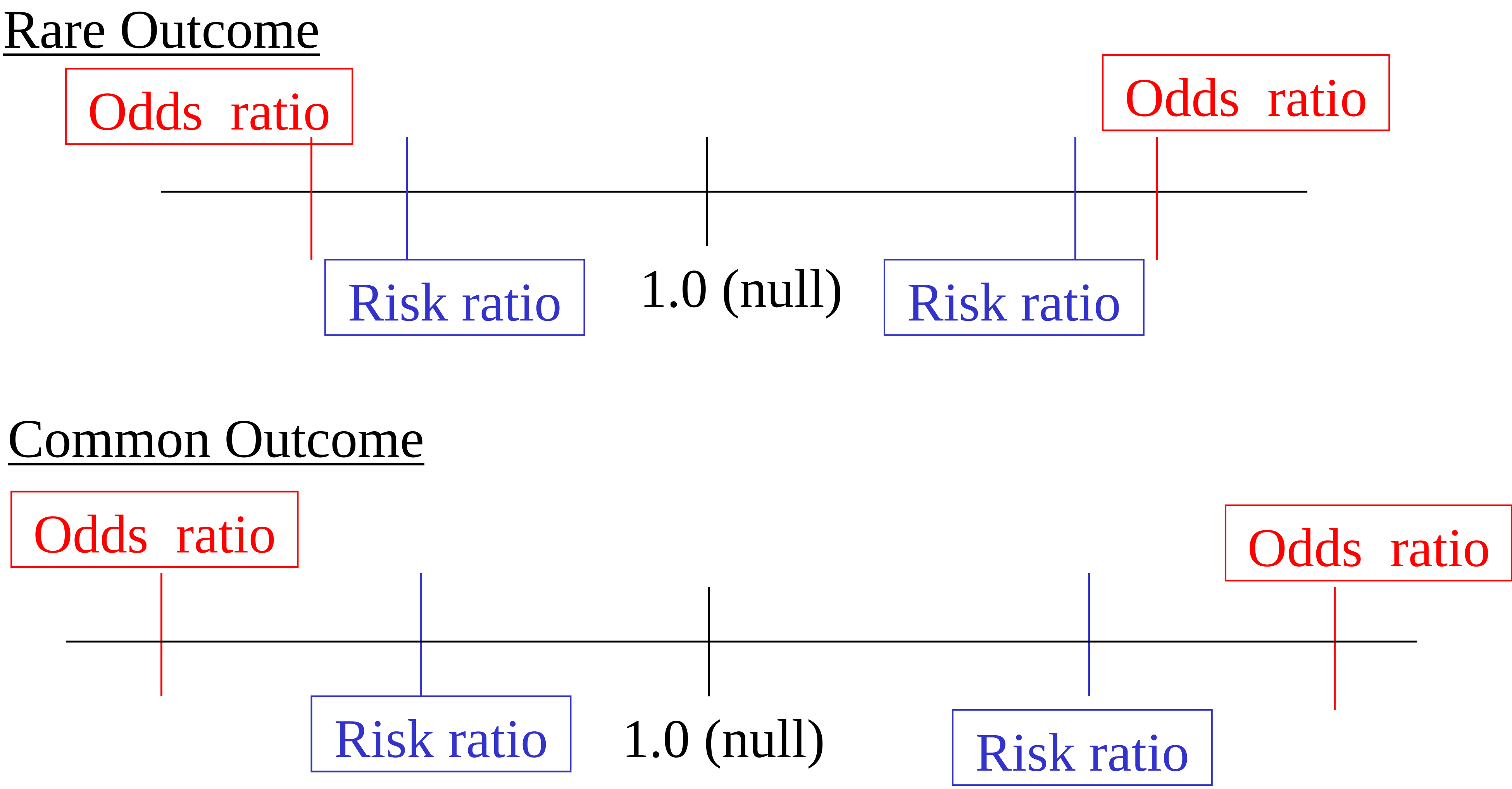
**OR = 90*70/10*30 =21.0**

Note: This indicates that those with Hep C infection have a 21-fold increase in their *odds* of developing liver cancer (not in their risk!).

- The odds ratio will always be bigger than the corresponding risk ratio if RR >1 and smaller if RR <1 (the harmful or protective effect always appears larger)
- The magnitude of the inflation depends on the prevalence of the disease.


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# The odds ratio vs. the risk ratio






sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# In-Class Exercise:

1. Suppose the following data were collected on a random sample of subjects (the researchers did not sample on exposure or disease status).

| | Neck pain | No Neck Pain |
|---|---|---|
| Own a cell phone | 143 | 209 |
| Don’t own a cell phone | 22 | 69 |

- Calculate the odds ratio and risk ratio for the association between cell phone usage and neck pain.

# Answer

| | Neck pain | No Neck Pain |
|---|---|---|
| Own a cell phone | 143 | 209 |
| Don’t own a cell phone | 22 | 69 |

- OR = (69*143)/(22*209) = 2.15
- RR = (143/352)/(22/91) = 1.68


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# In-Class Exercise:

- 2. Suppose the following data were collected on a random sample of subjects (the researchers did not sample on exposure or disease status).

| | Brain tumor | No brain tumor |
|---|---|---|
| Own a cell phone | 5 | 347 |
| Don't own a cell phone | 3 | 88 |

Calculate the odds ratio and risk ratio for the association between cell phone usage and brain tumor.


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Answer

| | Brain tumor | No brain tumor |
|---|---|---|
| Own a cell phone | 5 | 347 |
| Don't own a cell phone | 3 | 88 |

- OR = (5*88)/(3*347) = .42267
- RR = (5/352)/(3/91) = .43087


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Counting Rules

- Rules for counting the number of possible outcomes

- Counting Rule 1:
  - If any one of k different mutually exclusive and collectively exhaustive events can occur on each of n trials, the number of possible outcomes is equal to

$$\boxed{k^n}$$


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Counting Rules

*(continued)*

- **Counting Rule 2:**
  - **If there are $k_1$ events on the first trial, $k_2$ events on the second trial, … and $k_n$ events on the $n^{th}$ trial, the number of possible outcomes is**

$$\boxed{(k_1)(k_2)\ldots(k_n)}$$

  - **Example:**
    - You want to go to a park, eat at a restaurant, and see a movie. There are 3 parks, 4 restaurants, and 6 movie choices. How many different possible combinations are there?
    - Answer: (3)(4)(6) = 72 different possibilities



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Counting Rules

*(continued)*

- **Counting Rule 3:**
  - **The number of ways that n items can be arranged in order is**

$$n! = (n)(n-1)\ldots(1)$$

  - **Example:**
    - Your restaurant has five menu choices for lunch. How many ways can you order them on your menu?
    - Answer: $5! = (5)(4)(3)(2)(1) = 120$ different possibilities



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Counting Rules

*(continued)*

- **Counting Rule 4:**
  - **Permutations: The number of ways of arranging X objects selected from n objects in order is**

$$ {}_nP_x = \frac{n!}{(n-X)!} $$

  - **Example:**
    - Your restaurant has five menu choices, and three are selected for daily specials. How many different ways can the specials menu be ordered?
    - Answer: $nPx = \frac{n!}{(n-X)!} = \frac{5!}{(5-3)!} = \frac{120}{2} = 60$ different possibilities

# Counting Rules

*(continued)*

- **Counting Rule 5:**
  - **Combinations: The number of ways of selecting X objects from n objects, irrespective of order, is**

$$\boxed{{}_nC_x = \frac{n!}{X!(n-X)!}}$$

  - **Example:**
    - Your restaurant has five menu choices, and three are selected for daily specials. How many different special combinations are there, ignoring the order in which they are selected?
    - Answer: ${}_nC_x = \frac{n!}{X!(n-X)!} = \frac{5!}{3!(5-3)!} = \frac{120}{(6)(2)} = 10$ different possibilities



sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Chapter Summary

- Discussed basic probability concepts
  - Sample spaces and events, contingency tables, simple probability, and joint probability
- Examined basic probability rules
  - General addition rule, addition rule for mutually exclusive events, rule for collectively exhaustive events


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health

# Chapter Summary

- Defined conditional probability
  - Statistical independence, marginal probability, decision trees, and the multiplication rule
- Discussed Bayes' theorem
- Examined counting rules


sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Biostatistics, Epidemiology and Population Health