

# Biostatistics Refresher

Douglas Whitaker, Ph.D.  
Mount Saint Vincent University

ISANS IPG/IMG Bridging  
27 April 2022

# Topics

- Core Ideas: Prevalence & Incidence
- Types of Studies
- Measures of Risk
- Statistical Inference *Big Idea*
- Measures for Screening and Diagnostic Tests

# Notes

- Biostatistics & Epidemiology are *big* topics
- The measures and approaches discussed here are the core ideas
- Many studies use more sophisticated techniques and/or adjustments
- Trying to reproduce results from a paper?
  - Sophisticated techniques or adjustments might prevent this
  - Rounding might make numbers ‘a little’ off
- Proportions vs. Percentages
  - Proportions are more suitable for calculations; between 0 and 1
  - Percentages are often reported; between 0% and 100%

# Core Ideas

Definitions

# Incidence and Prevalence

- **Prevalence:** the *proportion* of a population with the disease (overall)

$$\text{Prevalence} = \frac{\text{Number of subjects with the disease}}{\text{Total number of subjects in the population}}$$

- ~~Rate~~
- **Incidence:** measures new cases of a disease in a population in a specific time period
    - Reported as *incidence rate* or *incidence proportion* (also called *incidence risk* or *cumulative incidence*)

# Incidence

$$\text{Incidence Proportion} = \frac{\text{[number of new cases of disease over time]}}{\text{population without disease at baseline}}$$

$$\text{Incidence Rate} = \frac{\text{number of new cases of disease over time}}{\text{person-time at risk}}$$

*Example 2.3* Investigators seek to determine the incidence of influenza among nurses at three local hospitals. They identify 500 nurses who do not have influenza as of December 1, 2010, and follow them for the development of influenza over the next 3 months. They find that ten of the study nurses develop influenza during follow-up. What is the incidence rate of influenza?

*We don't yet have enough info!*

# Incidence

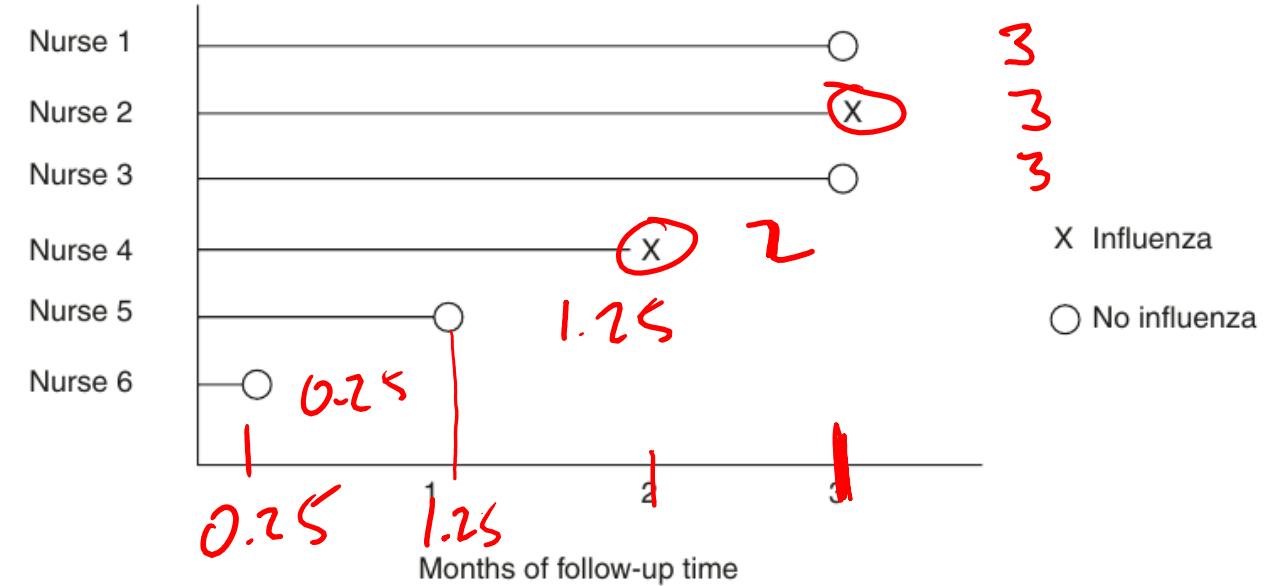
**Detour:** Instead of all 500 nurses, consider these 6.

What is the incidence rate for these 6 individuals?

2 new cases

12.5 months

= 0.16 cases/month



**Fig. 2.1** Diagram of individual risk times and disease status

Total

# Incidence

0.096 cases / year

*Example 2.3* Investigators seek to determine the incidence of influenza among nurses at three local hospitals. They identify 500 nurses who do not have influenza as of December 1, 2010, and follow them for the development of influenza over the next 3 months. They find that ten of the study nurses develop influenza during follow-up.

What is the incidence rate of influenza?

Need to know person-time at risk. Suppose it is 1200 months.

$$\frac{10 \text{ new cases}}{1200 \text{ months}} = 0.008 \text{ cases/month}$$

# Incidence

- Incidence proportions: more interpretable
- Incidence rates: used for comparisons of disease among different groups
- Incidence rate is often multiplied by a factor (e.g., 10 or 1000) to make it more easily interpretable.

# Morbidity and Mortality

- **Morbidity** refers to *illness* – not death
- **Mortality** refers to *death*

*Prevalence* measures morbidity in a population

*Mortality rate* measures deaths in the population

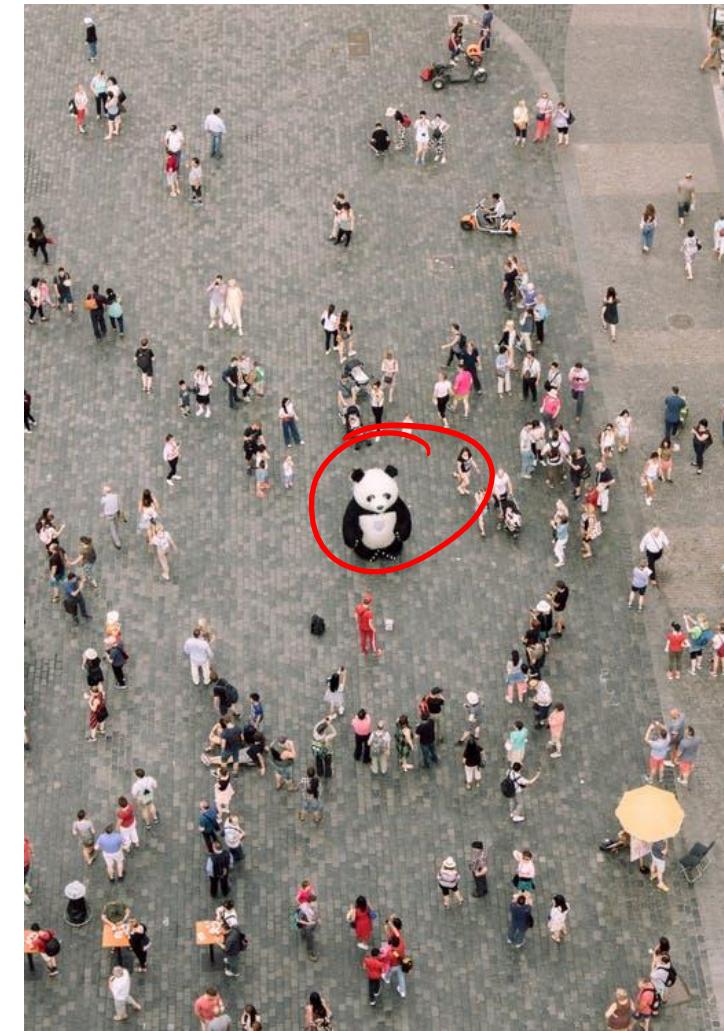
# Types of Studies

# Types of Studies: Overview

- Case Reports
- Cross-Sectional Studies
- Cohort Studies
- Case-Control Studies
- Randomized Trials

# Types of Studies: Case Reports

- Observational study ~~✓~~
- Describes experiences of individuals with a disease or condition
- **Good for:** new diseases, generating hypotheses, raising awareness
- **Limitations:**
  - Cannot calculate incidence
  - No comparison group
  - Often describe individuals that are *not typical*
  - Often small sample sizes – lots of variability



# Types of Studies: Cross-Sectional

- Measure the outcome of interest and exposures/factors *at the same time* from a group of interest.





# Types of Studies: Cross-Sectional

- Measure the outcome of interest and exposures/factors *at the same time* from a group of interest.

*Snapshot*

**Good for:** establishing associations between exposures/factors and disease prevalence

**Limitations:** cannot establish a temporal link (though if only one direction of a link is plausible you may still be able to justify a temporal link)

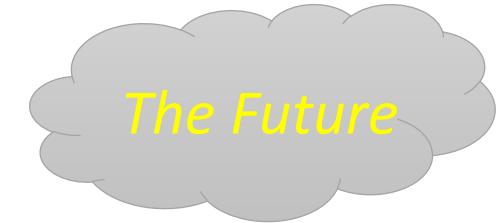
~~★~~ no causation

# Types of Studies: Cohort

Observational study that  
follows a group over time:

1. Exclude individuals with the disease
2. Define cohorts by measuring exposures
3. Follow over time to determine incidence

# Types of Studies: Cohort



Observational study that follows a group over time:

1. Exclude individuals with the disease
2. Define cohorts by measuring exposures
3. Follow over time to determine incidence



*The Present*



# Types of Studies: Cohort

Observational study that follows a group over time: ~~time~~

1. Exclude individuals with the disease
2. Define cohorts by measuring exposures
3. Follow over time to determine incidence

- **Good for:**
  - Temporal relationships
  - Studying multiple outcomes
- **Limitations:**
  - Confounding & bias
  - Not always efficient (rare or long latency)
- **Analysis:**
  - Relative risk
  - Attributable risk, population attributable risk

# Types of Studies: Case-Control

Observational study that compares groups of individuals

1. Identify individuals **with** the condition of interest
2. Separately, identify another group of individuals **without** the condition of interest.
3. Work backward to determine what factors may be influencing having the condition of interest.



*Group **with** condition  
"Cases"*



*Group **without** condition  
"Controls"*



# Types of Studies: Case-Control

Observational study that compares groups of individuals

1. Identify individuals **with** the condition of interest
2. Separately, identify another group of individuals **without** the condition of interest.
3. Work backward to determine what factors may be influencing having the condition of interest.



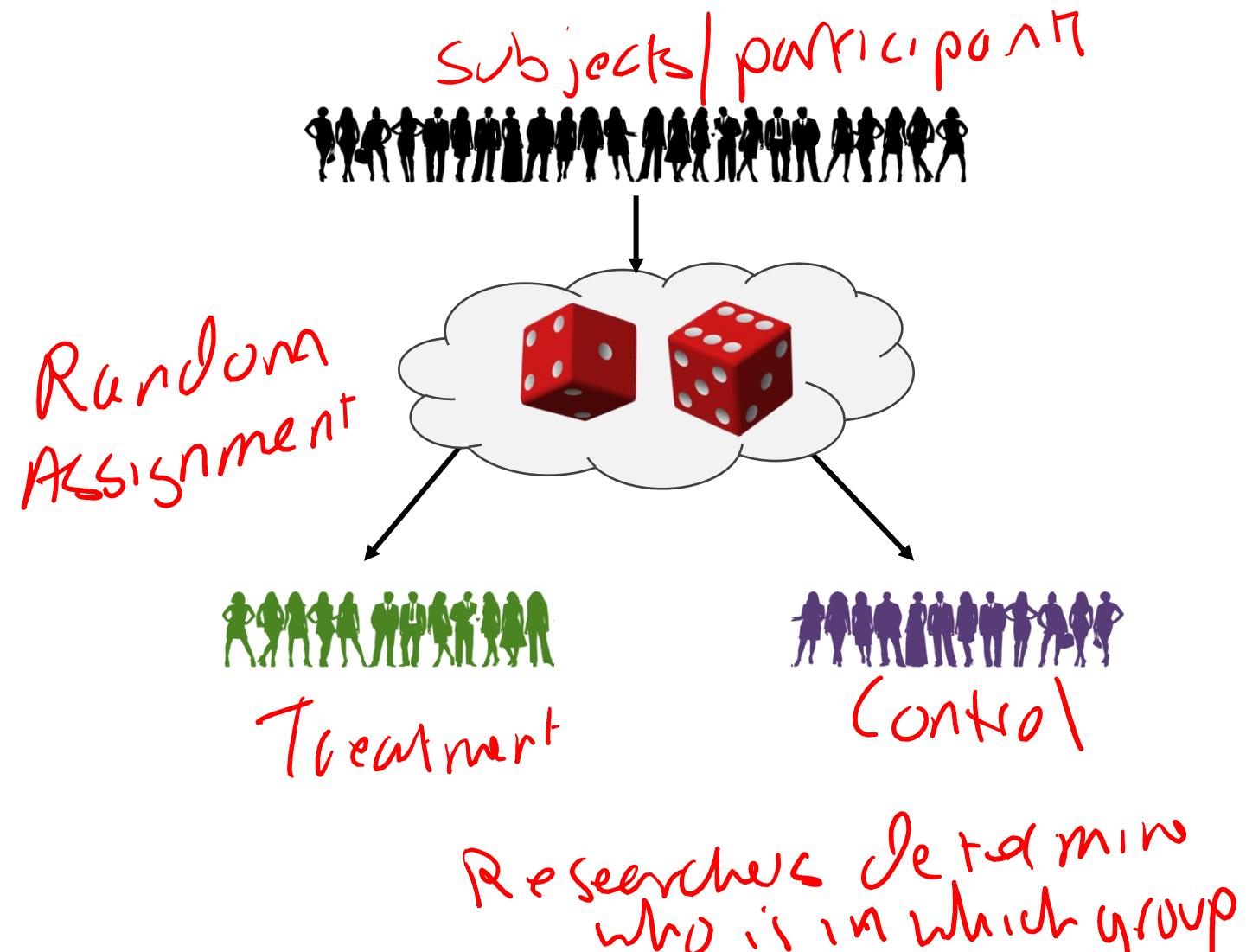
- **Good for:**
  - Studying rare diseases
  - Efficiency (using few participants)
  - Multiple risk factors
- **Limitations:**
  - Confounding & bias
  - Retrospective
  - Cannot calculate incidence
- **Analysis:**
  - Cannot use Relative Risk
  - Odds Ratio *instead of RR*

# Types of Studies: Randomized Trials

- Experimental, prospective study to compare treatments with controls
- Participants are *assigned* to an experimental condition (contrast with observational studies)
- **Good for:**
  - Establishing cause and effect (e.g., does a treatment perform better than control?)
- **Limitations:**
  - External validity issues (realism, monitored clinical environment, many exclusion criteria)

# Types of Studies: Randomized Trials

- Experimental, prospective study to compare treatments with controls
- Participants are *assigned* to an experimental condition (contrast with observational studies)





# Types of Studies: Randomized Trials

- Experimental, prospective study to compare treatments with controls
- Participants are *assigned* to an experimental condition (contrast with observational studies)
- **Good for:**
  - Establishing cause and effect (e.g., does a treatment perform better than control?)
- **Limitations:**
  - External validity issues (realism, monitored clinical environment, many exclusion criteria)

# Measures of Risk

# Risk

- Risk
- Absolute Risk Reduction (ARR)
- Number Needed to Treat (NNT)
- Number Needed to Harm (NNH)
- Relative Risk (RR)
- Relative Risk Reduction (RRR)
- Odds Ratio (OR)

2x2 two-way table

at one

		Event Occurred	
		Yes	No
Group	Treatment	a	b
	Control	c	d

↑  
factor  
of  
interest

These labels are arbitrary. Often exposed/unexposed is used. There is no firm rule about which variable is *rows* and which is *columns*. A description of the data and study design is key.

# Risk

- Risk means the *proportion* of people for which an event occurs during the study.
  - This is sometimes called the *incidence proportion* (different from the *incidence rate*).
- The proportion in each group for which the Event did occur (Yes):

$$p_{\text{Treatment}} = \frac{a}{a + b}$$

$$p_{\text{Control}} = \frac{c}{c + d}$$

↓ more common

		Event Occurred	
		Yes	No
Group	Treatment	a *	b
	Control	c *	d

# success  
 \_\_\_\_\_  
 # in the group

# Risk and Study Design

- Note that the study design is deeply connected to the types of analyses that can be done with a 2x2 table. *One group*
- Example: In a **Cohort Study**, the *Exposed group* and *Unexposed group* belong to a single cohort. The cohort is followed to determine *Status*.

		Status	
		Disease	Not Disease
Group	Exposed	a	b
	Unexposed	c	d

- Note these sums have meanings:
  - $a + b$  is the number of people in the *Exposed group* with the *Disease*
  - $a + c$  is the number of people in the *Cohort* with the *Disease*

These values can be used to estimate the prevalence of the *Disease*,

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

*# w.m. disease*  
*# total n*

# Status

	Disease	Not
Exposed	a	b
Unexposed	c	d

# Risk and Study Design

- Note that the study design is deeply connected to the types of analyses that can be done with a 2x2 table.
- Example: In a **Case Control Study**, participants are recruited because they **have the Disease** and or because they do **Not** it.



- Note these sums have meanings:
  - $a + c$  is the number of people in the study with the *Disease* (i.e. selected in the *Disease* group)
  - $a + b$  is NOT the number of people in the *Exposed* group... there wasn't an exposure group! *meaningless*

Because the participants were selected based on outcome (*Disease* or *Not*), we **cannot** get an estimate of the prevalence.

# Comparing groups

## Risk Difference

- Also known as:
  - Attributable Risk
  - Absolute Risk Increase (ARI)
  - Absolute Risk Reduction (ARR) (context)
- Compare the risks using subtraction

		Event Occurred	
		Yes	No
Group	Treatment	a	b
	Control	c	d

$$\text{Attributable Risk} = \text{Risk}_{\text{Control}} - \text{Risk}_{\text{Treatment}}$$

$$\text{Risk Difference} = p_{\text{Control}} - p_{\text{Treatment}}$$

- Be careful with order and context (usually structured so that the value is positive)

# Number Needed to Treat (NNT)

- Also known as Number Needed to Harm (NNH), depending on context
- Estimates the number of people who need to receive the treatment/intervention to prevent or cause one instance of the outcome
- Note the value should be positive and is usually rounded *whole number*

$$NNT = \frac{1}{\text{Attributable Risk}} = \frac{1}{\text{Risk}_{\text{Control}} - \text{Risk}_{\text{Treatment}}}$$

$$NNT = \frac{1}{p_{\text{Control}} - p_{\text{Treatment}}}$$

# Example (Neoplasms)



A recent study examined the Surveillance, Epidemiology, and End Results (SEER) database to examine the occurrence of myeloid neoplasms (MN) among breast cancer survivors who had received radiotherapy (RT) and who did not receive chemotherapy. Excerpts from the journal article and questions follow.

# Example (Neoplasms)



In the unadjusted analysis, there was an increased risk of subsequent MN among breast cancer patients who received RT compared to those who underwent surgery alone ... After 5 years of follow-up 5.0 (95% CI 2.1–12.2) of 1,000 who received RT and 3.7 (95% CI: 1.8–7.9) of 1,000 who did not receive RT had developed subsequent MN (absolute risk increase of A per 1,000 patients); corresponding to a number needed to harm of B. After 8 years of follow-up, the absolute risk increase was 1.7 per 1,000 patients corresponding to a number needed to harm of C, consistent with an increase in risk over time and longer follow-up. 1.3

RT                    no RT

1. What is the absolute risk increase (A)?

$$\frac{5.0 \text{ cases}}{1000 \text{ people}} - \frac{3.7 \text{ cases}}{1000 \text{ people}} = \frac{1.3 \text{ cases}}{1000 \text{ people}} = 0.0013$$

ARI

# Example (Neoplasms)



In the unadjusted analysis, there was an increased risk of subsequent MN among breast cancer patients who received RT compared to those who underwent surgery alone ... After 5 years of follow-up 5.0 (95% CI 2.1–12.2) of 1,000 who received RT and 3.7 (95% CI: 1.8–7.9) of 1,000 who did not receive RT had developed subsequent MN (absolute risk increase of A per 1,000 patients); corresponding to a number needed to harm of B. After 8 years of follow-up, the absolute risk increase was 1.7 per 1,000 patients corresponding to a number needed to harm of C, consistent with an increase in risk over time and longer follow-up.

2. What is the number needed to harm (NNH) after 5 years (B)?

$$\frac{1}{0.0013} = 769.23 \rightarrow 769 \text{ people}$$

# Example (Neoplasms)



In the unadjusted analysis, there was an increased risk of subsequent MN among breast cancer patients who received RT compared to those who underwent surgery alone ... After 5 years of follow-up 5.0 (95% CI 2.1–12.2) of 1,000 who received RT and 3.7 (95% CI: 1.8–7.9) of 1,000 who did not receive RT had developed subsequent MN (absolute risk increase of A per 1,000 patients); corresponding to a number needed to harm of B. After 8 years of follow-up, the absolute risk increase was 1.7 per 1,000 patients corresponding to a number needed to harm of C, consistent with an increase in risk over time and longer follow-up.

$$NNH = \frac{1}{ART}$$

3. What is the number needed to harm (NNH) after 8 years (C)?

$$\frac{1}{1.7/1000} = 588.24 \rightarrow 588 \text{ people}$$

## Example (Neoplasms)

B was 769 people  
C was 588 people



In the unadjusted analysis, there was an increased risk of subsequent MN among breast cancer patients who received RT compared to those who underwent surgery alone ... After 5 years of follow-up 5.0 (95% CI 2.1–12.2) of 1,000 who received RT and 3.7 (95% CI: 1.8–7.9) of 1,000 who did not receive RT had developed subsequent MN (absolute risk increase of A per 1,000 patients); corresponding to a number needed to harm of B. After 8 years of follow-up, the absolute risk increase was 1.7 per 1,000 patients corresponding to a number needed to harm of C, consistent with an increase in risk over time and longer follow-up.

588 vs 769

4. Explain why a *lower* value for #3 represents an *increase* in risk.

Fewer patients need to receive RT (588 vs. 769) to result in an additional MN case.

Zeidan AM, Long JB, Wang R, et al. Risk of myeloid neoplasms after radiotherapy among older women with localized breast cancer: A population-based study. Hills RK, ed. *PLOS ONE*. 2017;12(9):e0184747. doi:10.1371/journal.pone.0184747

Comparing risk for groups

## Relative Risk

**Division**

The relative risk (RR) is an incidence ratio.

$$RR = \frac{\text{Risk for Treatment Group}}{\text{Risk for Control Group}}$$

$$RR = \frac{\text{"Yes" proportion for Treatment Group}}{\text{"Yes" proportion for Control Group}}$$

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

~~We cannot calculate RR from case-control studies~~

		Event Occurred	
		Yes	No
Group	Treatment	a	b
	Control	c	d

Possible values for RR and their meaning:

$RR > 1$ : The Treatment group has a higher risk than the Control group easier to interpret

$RR = 1$ : The Treatment group has the same risk as the Control group

$RR < 1$ : The Treatment group has a lower risk than the Control group

~~Cannot be calculated from case-control studies~~

# Relative Risk



A study of patients presenting GeoSentinel clinics following rabies exposure for rabies post-exposure provalaxis (RPEP) examined the relationship between receiving rabies immunoglobulin (RIG) and where RPEP was started (in country of exposure or not). (Gautret et al., 2018)

		RIG given in country of exposure	
		Yes	No
Country of Exposure	Thailand	43	44
	Other	78	139

		RIG given in country of exposure	
		Yes	No
Country of Exposure	Indonesia	5	82
	Other	42	175

$$P_{\text{Thailand}} = 0.494$$

$$P_{\text{Other}} = 0.351$$

## Relative Risk

$$P_T - P_O = 0.494 - 0.351 = 0.135$$

Country of Exposure	RIG given in country of exposure	
	Yes	No
Thailand	43	44
Other	78	139

RIG given in country of exposure

$$RR = \frac{43 / (43 + 44)}{78 / (78 + 139)} = \frac{43 / 87}{78 / 217}$$



$\approx 1.38$

The relative risk for receiving RIG in the country of exposure for Thailand is 1.38. Reasonable interpretations of 1.38 include:

- Exposure to rabies in Thailand is associated with 1.38 times greater chance of receiving RIG compared with exposure in other countries.
- Exposure to rabies in Thailand is associated with a 38% greater chance of receiving RIG compared with exposure in other countries.

$$1.38 - 1.00 = 0.38 \rightarrow 38\%$$

above 100% is okay!

# Relative Risk

Country of Exposure	RIG given in country of exposure	
	Yes	No
Indonesia	5	82
Other	42	175

$$RR = \frac{5/(5+82)}{42/(42+175)} = \frac{5/87}{42/217}$$



$\approx 0.30$  less than 1

The relative risk for receiving RIG in the country of exposure for Indonesia is 0.30. Reasonable interpretations of 0.30 include:

- Exposure to rabies in Indonesia is associated with a 70% lower chance of receiving RIG compared with exposure in other countries.
- Exposure to rabies in a country *other than* Indonesia is associated with a 3.33 times greater chance of receiving RIG compared with exposure in Indonesia.

$$1.00 - 0.30 = 0.70 \rightarrow 70\% \quad \text{positive}$$

$$\frac{1}{RR} = \frac{1}{0.30} \approx 3.33$$

changing the order of comparison

# Relative Risk Reduction (RRR)

- This measure is a blend of two:
  - Relative Risk
  - Risk Difference (Attributable Risk)
- The *risk difference* is compared to the *risk for the control group* using division:

$$RRR = \frac{\text{Attributable Risk}}{\text{Risk for Control Group}}$$

$$RRR = \frac{\text{Risk}_{\text{Control}} - \text{Risk}_{\text{Treated}}}{\text{Risk}_{\text{Control}}} = \frac{p_{\text{Control}} - p_{\text{Treatment}}}{p_{\text{Control}}}$$

# Odds Ratio

$$\frac{\frac{a}{ab}}{\frac{b}{ab}} = \frac{a}{b}$$

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

The **odds ratio (OR)** compares *odds* for two groups.

*Odds* is a technical term for  $\frac{p_{success}}{p_{failure}}$  and NOT a synonym for "probability."

The odds ratio is more difficult to interpret than the relative risk.

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

\* Can be calculated from case-control studies... often a good estimate for relative risk.

$$OR = RR \times \left( \frac{1 - p_{Treatment}}{1 - p_{Control}} \right)$$

if  $p_T$  and  $p_C$  are small then  $RR$  can be large  
so  $OR \approx RR$

		Event Occurred	
		Yes	No
Group	Treatment	a	b
	Control	c	d

Possible values for *OR* and their meaning:

$OR > 1$ : The Treatment group has a *higher* odds than the Control group

$OR = 1$ : The Treatment group has the *same* odds as the Control group

$OR < 1$ : The Treatment group has a *lower* odds than the Control group

Can be calculated from case-control studies

# Relative Risk and Odds Ratio

$$\frac{3}{2}$$

$$\frac{2}{3}$$

1.5

$$1.5 - 1.0 = 0.5$$

0.67

$$1.0 - .67 = 0.33$$

- OR is more difficult to interpret... but we often want to (and can) interpret it *like* a RR
- Note that for both measures the reciprocal is also a valid value.
- That is,  $\frac{1}{RR}$  is a relative risk, and  $\frac{1}{OR}$  is an odds ratio.
- The relationship (less than, greater than) 1 will change.
- This might make the value easier or harder to interpret.
- Neither RR nor OR are symmetric about 1 (e.g., 1.5 vs. 0.67)
- Exposed group usually in the numerator (convention).

# Example (Malaria)



A prospective cohort study sought to determine the association between *P. vivax* and severe malaria (SM). Data from this study are reproduced below and are accompanied by questions.

		Severe Malaria?	
		Yes	No
Age	0 to < 2 years	173	846
	2 to < 5 years	207	2216

1. Can prevalence be calculated using these data? Why or why not?

Yes, it is a cohort study.





## Example (Malaria)

		Severe Malaria?	
		Yes	No
Age	0 to < 2 years	173	846
	2 to < 5 years	207	2216

2. What is the prevalence of Severe Malaria among children aged 0 to < 5 years?

$$\frac{173 + 207}{3442} = 0.11 \quad 11\%$$

3. What is the relative risk of Severe Malaria for participants under 2 years of age compared with those aged between 2 and 5?

under 2  
ages 2 to 5

$$\frac{173/(173 + 846)}{207/(207 + 2216)} = 1.99$$

It appears that children under age 2 are 1.99 times as likely as those who are between ages 2 and 5 to develop Sym.



## Example (Malaria)

		Severe Malaria?	
		Yes	No
Age	0 to < 2 years	173	846
	2 to < 5 years	207	2216

4. Can the odds ratio be calculated using these data? Why or why not?

Yes, odds ratio can always be calculated from a 2x2 table.

5. What is the odds ratio for Severe Malaria for participants under 2 years of age compared with those aged between 2 and 5?

$$\frac{173 \cdot 2216}{207 \cdot 846} = 2.19$$

# Example (Malaria)

		Severe Malaria?	
		Yes	No
Age	0 to < 2 years	173	846
	2 to < 5 years	207	2216



6. Compare the relative risk to the odds ratio.

1.99 vs. 2.19. They are similar values.

RR

OR



## Example (Malaria)

		Severe Malaria?	
		Yes	No
Age	0 to < 2 years	173	846
	2 to < 5 years	207	2216

7. Interpret the Relative Risk from #3. "About twice as likely"

- Children aged 0 to < 2 have a 1.99 times greater chance of having SM compared with those aged 2 to < 5 years.
- Children aged 0 to < 2 have a 99% greater chance of having SM compared with those aged to 2 < 5 years.  $1.99 - 1.0 = .99 = 99\%$
- Or, if taking reciprocal ( $\frac{1}{1.99} = .503$ ): Children aged 2 to < 5 years have a older 49.7% lower chance of having SM compared with those aged 0 to < 2 years.

8. Interpret the Odds Ratio from #5.

- The odds of having SM are 2.19 times greater for children aged 0 to < 2 compared with those aged 2 to < 5 years.

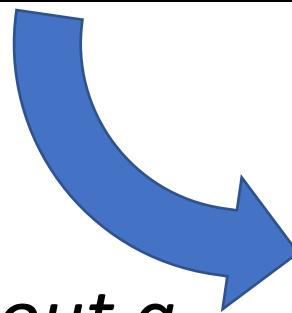
# Statistical Inference



# Statistical Inference



*... to draw conclusions about a larger group (population).*



# Statistical Inference

*Use data from a small group (sample) ...*



# Statistical Inference

Two major approaches:

- Hypothesis testing (p-values)
  - “Which of two descriptions of the population is better-supported by the data?”
- Confidence Intervals (CIs)
  - “Based on the data, what are plausible values for an unknown population value?”
- Underlying logic of statistical inference is *sophisticated and subtle* – be careful when reporting/documenting statistics (see NHST)

# Hypothesis Testing

- ★ • Null hypothesis ( $H_0$ ): usually “no effect” (or equivalent)
- Alternative hypothesis ( $H_A$ ): usually “there is some effect”
- Critical idea: *Circular reasoning*
  - We assume that the null hypothesis is true and collect evidence *against it*.
  - If we collect *enough* evidence against the null hypothesis to think it is wrong...
  - Then we think the alternative hypothesis is a better description of reality.
  - We do **NOT** accept or prove the null hypothesis true.
  - *At best* we “fail to reject” the null hypothesis.

# Hypothesis Testing

A p-value quantifies the strength of the evidence against the null hypothesis.

A p-value answers this question:

- *If the null hypothesis is true,*
- *what is the probability*
- *we would observe a test statistic as extreme*
- *(or more extreme)*
- *than the test statistic we actually did observe?*

*If p-value is small*

- ↳ 1) either we got really unlucky  
or  
2) The null hypothesis is wrong*

*then reject  $H_0$*

*conclude  $H_A$  is true*

# Hypothesis Testing

A p-value quantifies the strength of the evidence against the null hypothesis.

A p-value answers this question:

- If the null hypothesis is true, what is the probability we would observe a test statistic as extreme (or more extreme) than the test statistic we actually did observe?



A small p-value is **strong** evidence against the null hypothesis.

A **BIG** p-value is *weak* evidence against the null hypothesis.

- Caution: This feels counterintuitive to many people at first.

essentially no evidence  
Very little evidence against the null hypothesis

# Interpreting p-values

P-value of 0.50 provide essentially the same amount of evidence against  $H_0$

Borderline evidence

Weak evidence against the null hypothesis

0.10

above

nothing magical or special about this value → 0.05

Moderate evidence against the null hypothesis

0.025

Strong evidence against the null hypothesis

0.001

overwhelming

Very strong evidence against the null hypothesis

# Hypothesis Testing

*H<sub>0</sub>: There is no difference in average cost between High and Low dose groups*

*H<sub>A</sub>: There is a difference in average cost...→*

- A study compared healthcare costs for patients prescribed a High Dose of gabapentin to those prescribed a Low Dose. (Fleet et al., 2018)

Table 5. Average per person adjusted cost in 30-day follow-up period.

All Patients	Gabapentin	Emergency Visit
High Dose	\$64.75	\$103.70
Low Dose	\$33.53	\$90.30
Average Cost Differential	\$31.22	\$13.41 ↙
P-value	<0.0001	<0.0001

- If there was really no difference in average cost for Emergency Visits, the probability we would see an average difference of \$13.41 (or greater) is less than 0.0001.

# Hypothesis Testing

X   The p-value tells us nothing  
about the size of  
the effect.

- A study compared healthcare costs for patients prescribed a High Dose of gabapentin to those prescribed a Low Dose. (Fleet et al., 2018)

**Table 5. Average per person adjusted cost in 30-day follow-up period.**

All Patients	Gabapentin	Emergency Visit
High Dose	\$64.75	\$103.70
Low Dose	\$33.53	\$90.30
Average Cost Differential	\$31.22	\$13.41
P-value	<0.0001	<0.0001

- Because the p-value is so small (less than 0.0001), there is overwhelming evidence that there is some difference in average cost for Emergency Visits between High Dose and Low Dose patients.

# Hypothesis Testing



IS RR different from 1?

- Consider this excerpt of Table 3 from the rabies study earlier (Gautret et al., 2018)

Characteristic		RIG given in country of exposure (N = 87)	RIG not given in country of exposure (N = 217)	Relative risk of receiving RIG in the country of exposure (95% CI)	P value
Countries of exposure n (%)	Thailand	43 (49.4)	78 (39.5)	1.38 (1.0–1.8)	0.02
	Indonesia	5 (5.8)	42 (19.4)	0.30 (0.1–0.7)	0.01
	India	3 (3.4)	13 (6.0)	0.58 (0.2–2.0)	0.38
	China	1 (1.2)	12 (5.5)	0.21 (0.0–1.6)	weak 0.13
	Algeria	6 (6.9)	7 (3.2)	2.14 (0.7–6.2)	0.16
	Philippines	8 (9.2)	1 (0.5)	19.95 (2.5–157.2)	0.01
	Sri Lanka	8 (9.2)	5 (2.3)	3.99 (1.3–11.9)	0.013

- For which countries is there strong evidence that the RR is not equal to 1? Thailand, Indonesia, Philippines, Sri Lanka

# Contemporary Views of p-values

The new guidelines discuss many aspects of the reporting of studies in [*The New England Journal of Medicine*], including a requirement to replace P values with estimates of effects or association and 95% confidence intervals when neither the protocol nor the statistical analysis plan has specified methods used to adjust for multiplicity. **Journal editors and statistical consultants have become increasingly concerned about the overuse and misinterpretation of significance testing and P values in the medical literature.** Along with their strengths, P values are subject to inherent weaknesses...  
(Harrington et al., 2019) [emphasis added]

Harrington, D., D'Agostino, R. B., Gatsonis, C., Hogan, J. W., Hunter, D. J., Normand, S.-L. T., Drazen, J. M., & Hamel, M. B. (2019). New Guidelines for Statistical Reporting in the *Journal*. *New England Journal of Medicine*, 381(3), 285–286.  
<https://doi.org/10.1056/NEJM1906559>

# Contemporary Views of p-values

**In view of the prevalent misuses of and misconceptions concerning p-values, some statisticians prefer to supplement or even replace p-values with other approaches.** These include methods that emphasize estimation over testing, such as confidence, credibility, or prediction intervals; Bayesian methods; alternative measures of evidence, such as likelihood ratios or Bayes Factors; and other approaches such as decision-theoretic modeling and false discovery rates. All these measures and approaches rely on further assumptions, but they may more directly address the size of an effect (and its associated uncertainty) or whether the hypothesis is correct. (Wasserstein et al., 2016, p. 132) [emphasis added]

# Contemporary Views of p-values

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A p-value, or statistical significance, does not measure the size of an effect or the importance of a result.
6. By itself, a p-value does not provide a good measure of evidence regarding a model or hypothesis. (Wasserstein et al., 2016, pp. 131-132)

A [

# Contemporary Views of p-values

The ASA *Statement on P-Values and Statistical Significance* stopped just short of recommending that declarations of “statistical significance” be abandoned. We take that step here. **We conclude, based on our review of the articles in this special issue and the broader literature, that it is time to stop using the term “statistically significant” entirely.** Nor should variants such as “significantly different,” “ $p < 0.05$ ,” and “nonsignificant” survive, whether expressed in words, by asterisks in a table, or in some other way. (Wasserstein et al., 2019, p. 2) [emphasis added]

# Type I and Type II Errors

- If we reject  $H_0$  when  $H_0$  is true, we have committed a **Type I error**.
- If we fail to reject  $H_0$  when  $H_0$  is false, we have committed a **Type II error**.

		<u>Truth about the population</u>	
		$H_0$ true	$H_0$ false ( $H_a$ true)
<u>Conclusion</u> based on sample	Reject $H_0$	Type I error	Correct conclusion 😊
	Fail to reject $H_0$	Correct conclusion 😊	Type II error

# Type I and Type II Errors (Criminal Trial)



- If we reject  $H_0$  when  $H_0$  is true, we have committed a **Type I error**.
- If we fail to reject  $H_0$  when  $H_0$  is false, we have committed a **Type II error**.

		Actual Truth	
		<u>Innocent</u> $H_0$ true	<u>Committed Crime</u> $H_0$ false ( $H_a$ true)
Conclusion of trial	“Guilty” Reject $H_0$	Type I error	Correct  conclusion
	“Not Guilty” Fail to reject $H_0$	Correct  conclusion	Type II error

# Type I and Type II Errors (Criminal Trial)



- How can we reduce the chances that we send an innocent person to jail?
  - How can we lower the chances of a **Type I Error**?

*require more evidence to reject Ho*

- What are the consequences of this?

*Increase Type II error rate*

		Actual Truth	
		<i>Innocent</i> $H_0$ true	<i>Committed Crime</i> $H_0$ false ( $H_a$ true)
Conclusion of trial	“Guilty” Reject $H_0$	Type I error	Correct conclusion
	“Not Guilty” Fail to reject $H_0$	Correct conclusion	Type II error

# Type I and Type II Errors (Criminal Trial)



- How can we reduce the chances that we let a guilty person go free?
  - How can we lower the chances of a **Type II Error**?

*Require very little evidence to reject  $H_0$*

- What are the consequences of this?

*Type I errors ↑*

		Actual Truth	
		Innocent $H_0$ true	Committed Crime $H_0$ false ( $H_a$ true)
Conclusion of trial	“Guilty” Reject $H_0$	Type I error	Correct conclusion
	“Not Guilty” Fail to reject $H_0$	Correct conclusion	Type II error

# Type I and Type II Errors

- Replication studies and meta-analyses are important
- Importance of Type I & Type II Errors strongly depend on context
- Need to know statistical hypotheses (may not be explicitly written, but usually standard – consult references)
- **Power** is the ability to correctly reject the null hypothesis  
$$\text{Power} = 1 - P(\text{Type II Error})$$

# Confidence Intervals ~~☆~~ preferred to HT

- Provide an interval of values that are all *plausible* for a true, unknown population value based on the data.
- A confidence level is specified (often 95%).
- Want to be more confident (e.g., 95% -> 99%)?
- Interval gets wider.  
*At the same confidence level  
as  $n \uparrow$   
width ↓*
- Want a narrower (more precise) interval?
- Lower confidence level (e.g., 95% -> 90%).

# Confidence Intervals

- The confidence level (e.g., 95%) is a statement about the confidence in *the method* and not a statement of confidence about the specific interval calculated!
- That is, if you commit to making 95% confidence intervals over and over and over, about 95% of them will contain the true (unknown) value you want to identify.
- And therefore about 5% will *not* contain the true (unknown) value you want to identify. (That is, they will be wrong.)
- No way of knowing which are right and which are wrong.
- (Confidence in the *method*!)

# Confidence Intervals

- Confidence Intervals can be used in a way analogous to Hypothesis Testing.
- In Hypothesis Testing, there is often a single value in the null hypothesis.
  - Difference in values? Usually null hypothesis is “0”
  - Ratio of values? Usually null hypothesis is “1”
- Examine the CI:
  - if this value is *in* the CI, then it is plausible...
  - if it is plausible, that is *not* evidence against the null hypothesis...
  - “fail to reject” the null hypothesis
- If the value is *not* in the CI, it is not plausible, go with the alternative hypothesis

# Hypothesis Testing



- Consider this excerpt of Table 3 from the rabies study earlier (Gautret et al., 2018)

Characteristic		RIG given in country of exposure (N = 87)	RIG not given in country of exposure (N = 217)	Relative risk of receiving RIG in the country of exposure (95% CI)	P value
Countries of exposure n (%)	Thailand	43 (49.4)	78 (39.5)	1.38 (1.0–1.8)	0.02
	Indonesia	5 (5.8)	42 (19.4)	0.30 (0.1–0.7)	0.01
	India	3 (3.4)	13 (6.0)	0.58 (0.2–2.0)	0.38
	China	1 (1.2)	12 (5.5)	0.21 (0.0–1.6)	0.13
	Algeria	6 (6.9)	7 (3.2)	2.14 (0.7–6.2)	0.16
	Philippines	8 (9.2)	1 (0.5)	19.95 (2.5–157.2)	0.01
	Sri Lanka	8 (9.2)	5 (2.3)	3.99 (1.3–11.9)	0.015

- Which countries have a RR significantly different from 1?

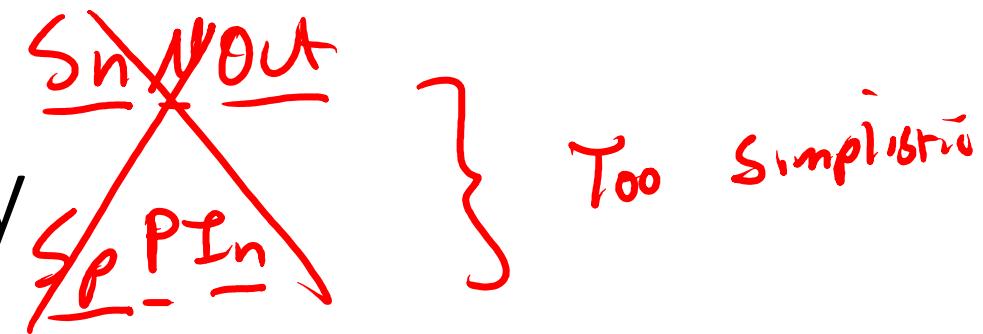
do not contain 1

do contain 1

# Diagnostic Rates

Sensitivity & Specificity

# Sensitivity and Specificity



- Screening diseases – need a gold standard
- “Accuracy” of a test is not good enough
- The **sensitivity** of a test is its ability to correctly identify patients who have the condition being tested for
  - Correct positives
- The **specificity** of a test is its ability to correctly identify patients who do not have the condition being tested for
  - Correct negatives

# Sensitivity and Specificity

TP = True Positive  
 TN = True Negative  
 FN = False Negative  
 FP = False Positive

		True Disease Status	
		Yes	No
Screening Result	Positive	TP	FP
	Negative	FN	TN

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{All Yes Disease}} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{All No Disease}} = \frac{TN}{TN + FP}$$

$$\text{False Positive Error Rate} = \frac{\text{False Positives}}{\text{All No Disease}} = \frac{FP}{TN + FP}$$

The rate at which true No disease patients are incorrectly flagged Positive.

$$\text{False Negative Error Rate} = \frac{\text{False Negatives}}{\text{All Yes Disease}} = \frac{FN}{TP + FN}$$

The rate at which true Yes disease patients are incorrectly flagged Negative.

# Sensitivity and Specificity

TP = True Positive  
 TN = True Negative  
 FN = False Negative  
 FP = False Positive

True Disease Status			
Yes	No		
Screening Result	Positive	TP	FP
	Negative	FN	TN

$$\text{Positive Predictive Value(PPV)} = \frac{\text{True Positives}}{\text{All Positives}} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

The proportion of patients with Positive results who Yes had the disease

$$\text{Negative Predictive Value(NPV)} = \frac{\text{True Negatives}}{\text{All Negatives}} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

The proportion of patients with Negative results who No, did not have the disease

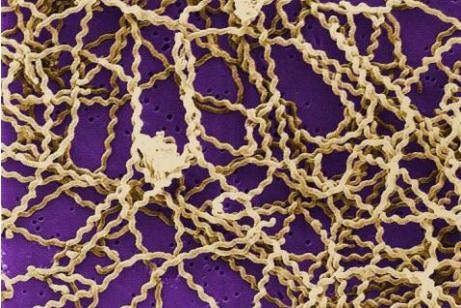
$$\text{Likelihood Ratio Positive (LR+)} = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$$

The higher this value is, the better a test is.  
(Should be much larger than 1.)

$$\text{Likelihood Ratio Negative (LR-)} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$$

The lower this value is, the better a test is.  
(Should be as close to 0 as possible.)

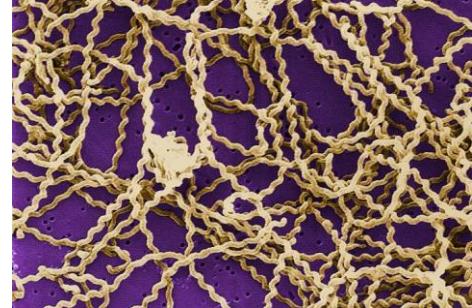
# Example (Leptospirosis)



A study was conducted to evaluate the performance of a Rapid Diagnostic Test (RDT) for leptospirosis in French tropical territories. PCR was used as a gold standard to classify patients as either having Leptospirosis or being Controls. Germane results are reproduced below, and questions follow.

True

		Leptospirosis	Control
IgM RDT assay <i>Diagnose</i>	Positive	168	14
	Negative	19	207



		Leptospirosis	Control
IgM RDT assay	Positive	168	14
	Negative	19	207

- Sensitivity

$$\frac{168}{168+19} = .898 \text{ or } 89.8\%$$

- Specificity

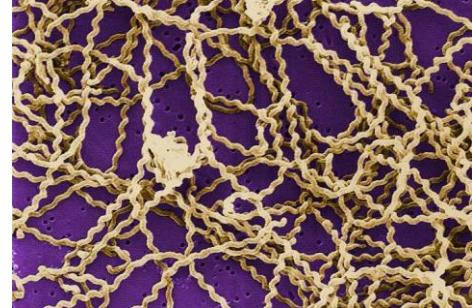
$$\frac{207}{207+14} = .937 \text{ or } 93.7\%$$

- False Positive Error Rate

$$1 - .937 = .063 \text{ or } 6.3\%$$

- False Negative Error Rate

$$1 - .898 = .102 \text{ or } 10.2\%$$



		Leptospirosis	Control
IgM RDT assay	Positive	168	14
	Negative	19	207

- Likelihood Ratio +

$$\frac{.898}{1-.937} = 14.182 \quad \text{much larger than 1}$$

so good

- Likelihood Ratio -

$$\frac{1-.898}{.937} = .108 \quad \text{close to 0}$$

so good

- Positive Predictive Value

$$\frac{168}{168+14} = .923 \text{ or } 92.3\%$$

- Negative Predictive Value

$$\frac{207}{207+19} = .916 \text{ or } 91.6\%$$

# References & Credits

# This presentation

- Slides for this presentation will be made available online
- A handout with additional practice problems is also online (with solutions)

**<https://tinyurl.com/2rbe225y>**

<https://github.com/douglaswhitaker/public-files/tree/main/biostatistics>

- These files will be available at this URL for at least 6 months
  - (If they move before that, there will be new links at that URL to take you to them.)

# Online Resources

- *The Little Handbook of Statistical Practice*
  - [StatisticalPractice.com](http://StatisticalPractice.com) redirects to <http://www.jerrydallal.com/LHSP/LHSP.HTM>
  - Old but good

# References

- Case, C. & Whitaker, D. (2016). Innocent until proven guilty: Exploring statistical power through simulations. *Mathematics Teacher*, 109(9), 686-692.  
<https://www.jstor.org/stable/10.5951/mathteacher.109.9.0686>
- Harrington, D., D'Agostino, R. B., Gatsonis, C., Hogan, J. W., Hunter, D. J., Normand, S.-L. T., Drazen, J. M., & Hamel, M. B. (2019). New Guidelines for Statistical Reporting in the Journal. *New England Journal of Medicine*, 381(3), 285–286. <https://doi.org/10.1056/NEJMMe1906559>
- Katz DL, Elmore JG, Wild DMG, Lucan SC, eds. *Jekel's Epidemiology, Biostatistics, Preventive Medicine, and Public Health*. 4th ed. Philadelphia, Pa. ; London: Saunders; 2014.
- Kestenbaum B. *Epidemiology and Biostatistics: An Introduction to Clinical Research*. Second edition. (Weiss NS, Shoben A, eds.). Cham: Springer; 2019.
- Noordzij M, Dekker FW, Zoccali C, Jager KJ. Measures of Disease Frequency: Prevalence and Incidence. *Nephron Clin Pract*. 2010;115(1):c17-c20. doi:10.1159/000286345
- Porta MS, Greenland S, Hernán M, Silva I dos S, Last JM, International Epidemiological Association, eds. *A Dictionary of Epidemiology*. Six edition. Oxford: Oxford University Press; 2014.
- Shapiro K, Brown SA. *RxPrep Course Book: A Comprehensive Course for the NAPLEX® and CPJE*. RxPrep, Inc.; 2015.

# Studies Used in Examples

- Christiansen SC, Næss IA, Cannegieter SC, Hammerstrøm J, Rosendaal FR, Reitsma PH. Inflammatory Cytokines as Risk Factors for a First Venous Thrombosis: A Prospective Population-Based Study. Greaves M, ed. *PLoS Med.* 2006;3(8):e334. doi:10.1371/journal.pmed.0030334
- Fleet JL, Dixon SN, Kuwornu PJ, et al. Gabapentin dose and the 30-day risk of altered mental status in older adults: A retrospective population-based study. Wu P-H, ed. *PLOS ONE.* 2018;13(3):e0193134. doi:10.1371/journal.pone.0193134
- Galappaththi-Arachchige HN, Holmen S, Koukounari A, et al. Evaluating diagnostic indicators of urogenital *Schistosoma haematobium* infection in young women: A cross sectional study in rural South Africa. Knight M, ed. *PLOS ONE.* 2018;13(2):e0191459. doi:10.1371/journal.pone.0191459
- Gautret P, Angelo KM, Asgeirsson H, et al. Rabies post-exposure prophylaxis started during or after travel: A GeoSentinel analysis. Gilbert AT, ed. *PLoS Negl Trop Dis.* 2018;12(11):e0006951. doi:10.1371/journal.pntd.0006951
- Genton B, D'Acremont V, Rare L, et al. Plasmodium vivax and Mixed Infections Are Associated with Severe Malaria in Children: A Prospective Cohort Study from Papua New Guinea. Rogerson S, ed. *PLoS Med.* 2008;5(6):e127. doi:10.1371/journal.pmed.0050127
- Goarant C, Bourhy P, D'Ortenzio E, et al. Sensitivity and Specificity of a New Vertical Flow Rapid Diagnostic Test for the Serodiagnosis of Human Leptospirosis. Büscher P, ed. *PLoS Negl Trop Dis.* 2013;7(6):e2289. doi:10.1371/journal.pntd.0002289
- Zeidan AM, Long JB, Wang R, et al. Risk of myeloid neoplasms after radiotherapy among older women with localized breast cancer: A population-based study. Hills RK, ed. *PLOS ONE.* 2017;12(9):e0184747. doi:10.1371/journal.pone.0184747

# Art Credits



1



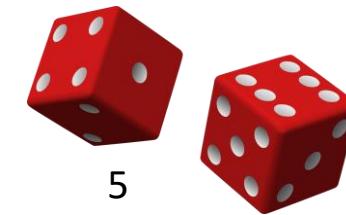
2



3



4



5



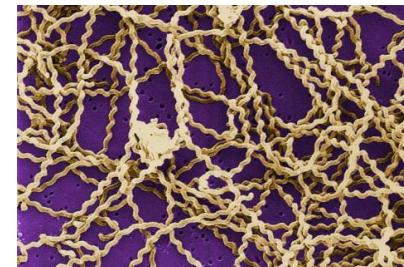
6



7



8



11



9

1. <https://www.pexels.com/photo/people-gathered-watching-a-panda-mascot-2346289/>
2. <https://pixabay.com/illustrations/girl-photographer-camera-snapshot-2803516/>
3. <https://pixabay.com/vectors/association-community-group-meeting-152746/>
4. <https://pixabay.com/vectors/people-group-crowd-line-silhouette-312122/>
5. <https://pixabay.com/vectors/dice-red-two-game-rolling-chance-25637/>
6. <https://www.pexels.com/photo/two-person-doing-surgery-inside-room-1250655/>
7. <https://www.pexels.com/photo/syringe-injectors-placed-on-yellow-surface-4210559/>
8. <https://www.pexels.com/photo/close-up-view-of-mosquito-9891863/>
9. <https://www.pexels.com/photo/bird-s-eye-view-of-group-of-people-1299086/>
10. <https://www.pexels.com/photo/judges-desk-with-gavel-and-scales-5669619/>
11. By CDC/ Rob Weyant - [http://phil.cdc.gov/PHIL\\_Images/20050308/22ad4ce53a1648feb011a7d6dd26fbb6/138\\_lores.jpg](http://phil.cdc.gov/PHIL_Images/20050308/22ad4ce53a1648feb011a7d6dd26fbb6/138_lores.jpg), Public Domain, <https://commons.wikimedia.org/w/index.php?curid=611941>