

# **Overview of Prism**

## **Sensitivity vs Specificity**

## **Contingency Tables**

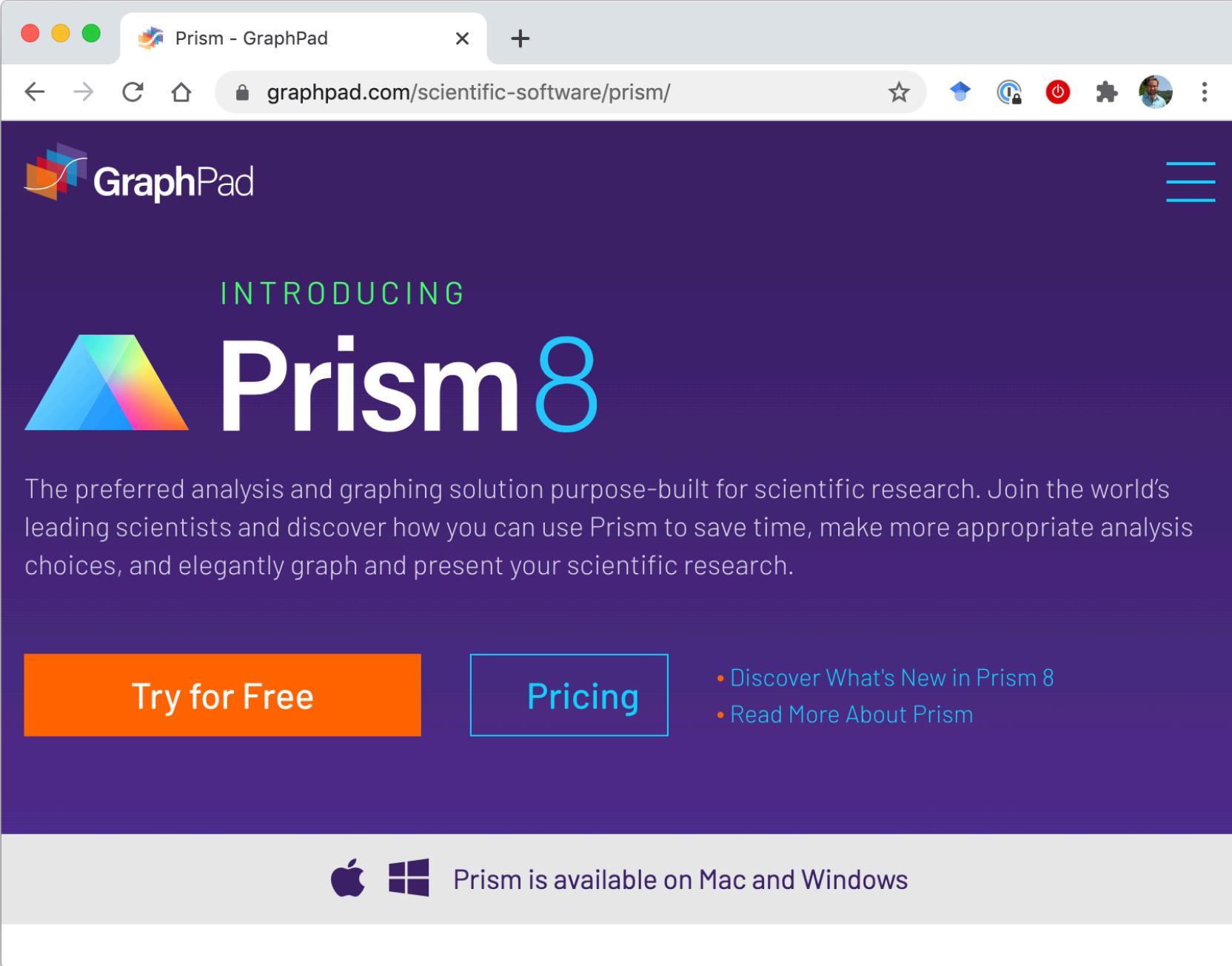
## **Fisher's Exact Test**

---



Practical Statistics for Experimental Biologists  
Lecture 1  
Monday, 27 July 2020  
10:00am - 12:00pm

## Overview of Prism



The screenshot shows a web browser window for 'Prism - GraphPad' at [graphpad.com/scientific-software/prism/](https://graphpad.com/scientific-software/prism/). The page has a dark purple header with the GraphPad logo and a navigation bar. The main content features a large 'INTRODUCING Prism 8' heading with a stylized 'A' icon. Below the heading is a descriptive paragraph about Prism 8. At the bottom, there are 'Try for Free', 'Pricing', and 'Discover What's New in Prism 8' buttons, along with a note about Mac and Windows compatibility.

INTRODUCING

# Prism 8

The preferred analysis and graphing solution purpose-built for scientific research. Join the world's leading scientists and discover how you can use Prism to save time, make more appropriate analysis choices, and elegantly graph and present your scientific research.

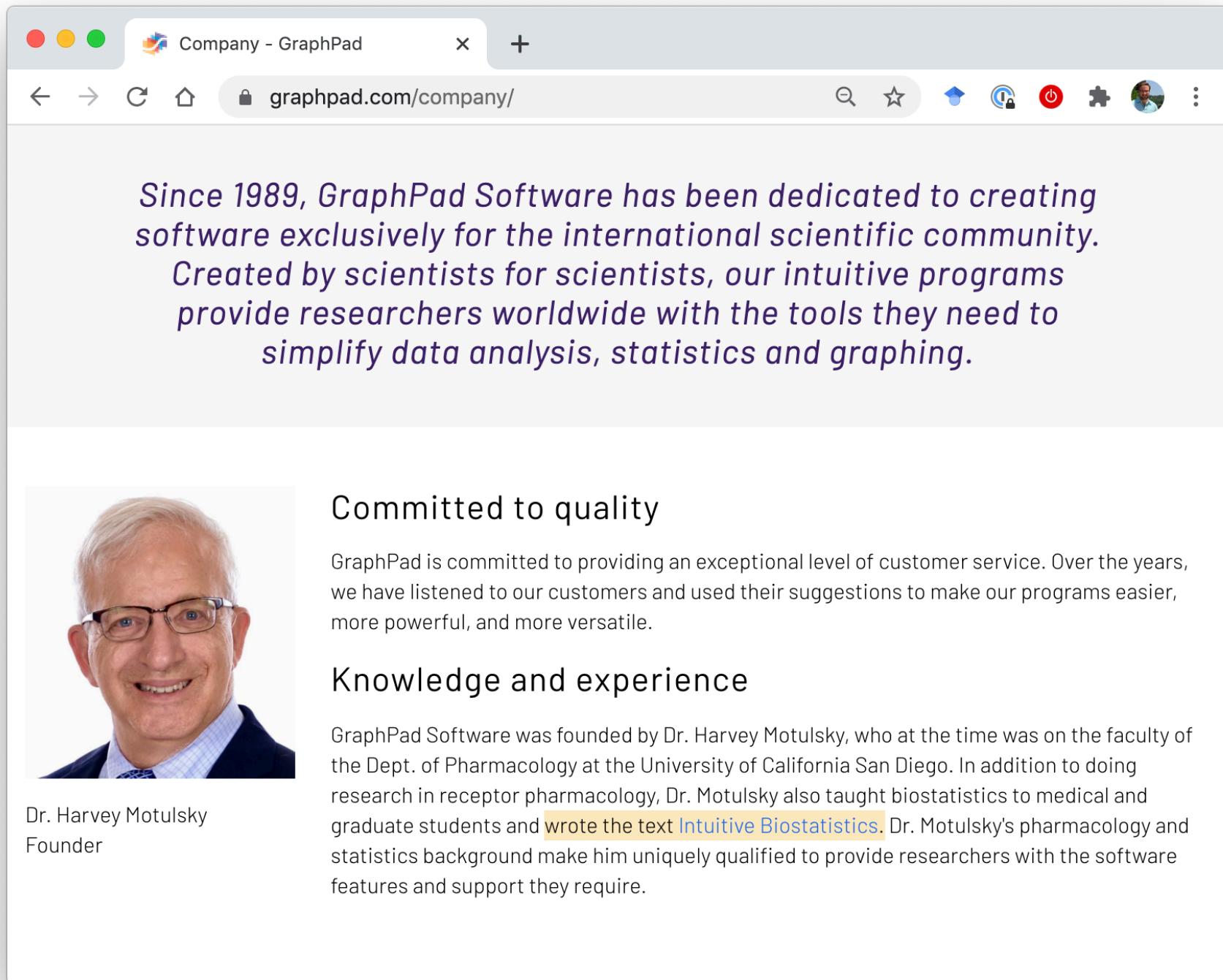
Try for Free

Pricing

- Discover What's New in Prism 8
- Read More About Prism

Prism is available on Mac and Windows

# Harvey Motulsky

A screenshot of a web browser window showing the GraphPad Company website. The title bar says "Company - GraphPad". The address bar shows "graphpad.com/company/". The main content area contains a quote in purple text: "Since 1989, GraphPad Software has been dedicated to creating software exclusively for the international scientific community. Created by scientists for scientists, our intuitive programs provide researchers worldwide with the tools they need to simplify data analysis, statistics and graphing." Below this, on the left, is a portrait of Dr. Harvey Motulsky, a man with white hair and glasses, wearing a suit. To his right are two sections with headings and text: "Committed to quality" and "Knowledge and experience".

Since 1989, GraphPad Software has been dedicated to creating software exclusively for the international scientific community. Created by scientists for scientists, our intuitive programs provide researchers worldwide with the tools they need to simplify data analysis, statistics and graphing.



Dr. Harvey Motulsky  
Founder

**Committed to quality**

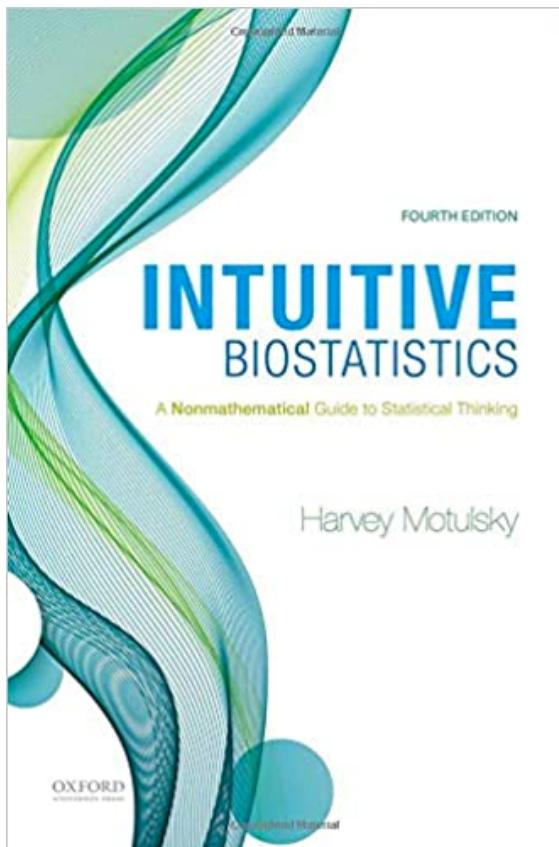
GraphPad is committed to providing an exceptional level of customer service. Over the years, we have listened to our customers and used their suggestions to make our programs easier, more powerful, and more versatile.

**Knowledge and experience**

GraphPad Software was founded by Dr. Harvey Motulsky, who at the time was on the faculty of the Dept. of Pharmacology at the University of California San Diego. In addition to doing research in receptor pharmacology, Dr. Motulsky also taught biostatistics to medical and graduate students and wrote the text [Intuitive Biostatistics](#). Dr. Motulsky's pharmacology and statistics background make him uniquely qualified to provide researchers with the software features and support they require.

## Main reference book

---



Motulsky, 2017  
Intuitive Biostatistics  
4th Edition

“Intuitive Biostatistics is both an introduction and review of statistics. Compared to other books, it has:

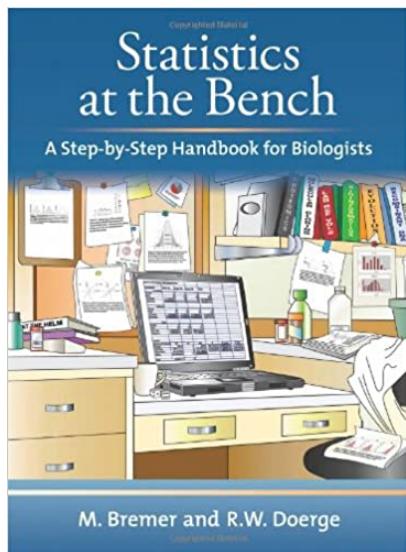
- Breadth rather than depth. It is a guidebook, not a cookbook.
- Words rather than math. It has few equations.
- Explanations rather than recipes. This book presents few details of statistical methods and only a few tables required to complete the calculations....

I wrote Intuitive Biostatistics for three audiences:

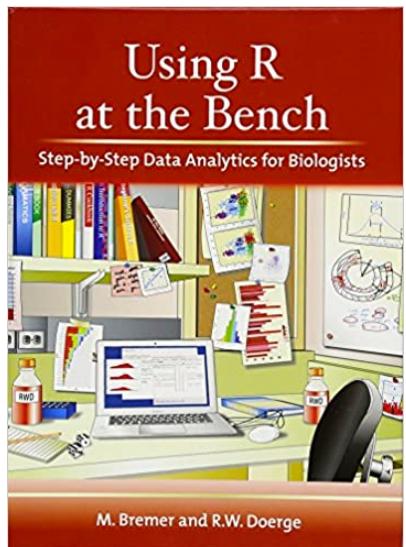
- Medical (and other) professionals who want to understand the statistical portions of journals they read. These readers don't need to analyze any data, but need to understand analyses published by others. I've tried to explain the big picture, without getting bogged down in too many details.
- Undergraduate and graduate students, post-docs and researchers who will analyze data. This book explains general principles of data analysis, but it won't teach you how to do statistical calculations or how to use any particular statistical program. It makes a great companion to the more traditional statistics texts and to the documentation of statistical software.
- Scientists who consult with statisticians. Statistics often seems like a foreign language, and this text can serve as a phrase book to bridge the gap between scientists and statisticians. Sprinkled throughout the book are “Lingo” sections that explain statistical terminology, and point out when statistics gives ordinary words very specialized meanings (the source of much confusion).”

## Other useful books

---



Bremmer & Doerge, 2009

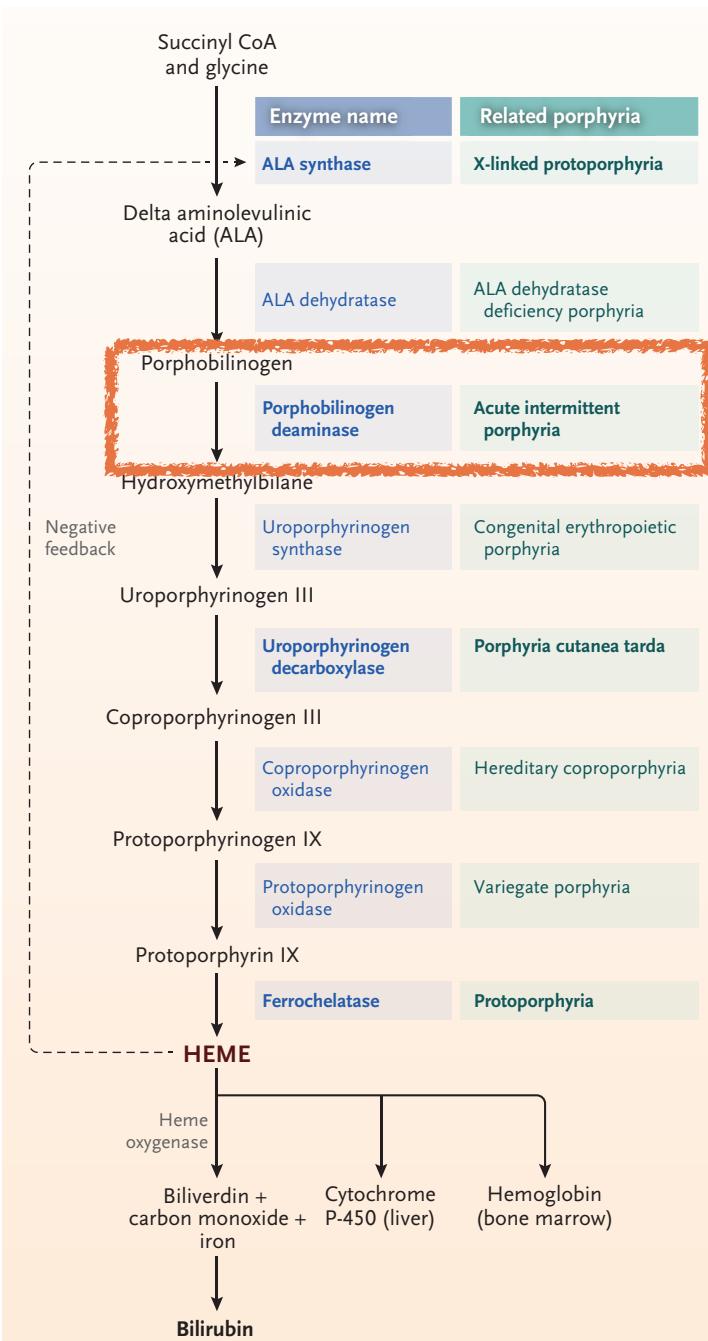


Bremmer & Doerge, 2015

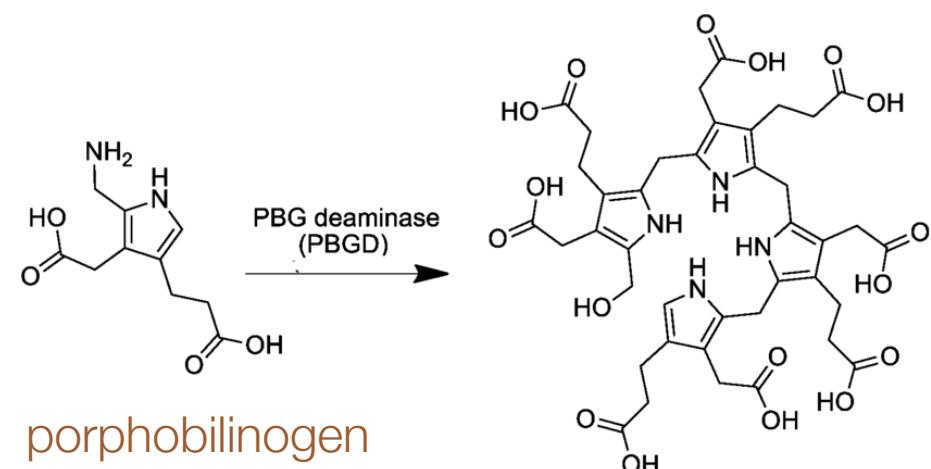
- Present basic statistical equations (without derivation).
- Best read linearly, not just as references (despite the titles).
- A good refresher for those who have had some statistics training.
- Does not provide as much intuition or practical guidance as Motulsky.
- For novices I recommend reading this after Motulsky's book.
- 2009 book provides recipes to use in Microsoft Excel (best to avoid doing this)
- 2015 book provides recipes to use in R. R is much more powerful than GraphPad Prism, but it's also much easier to mess up statistical calculations in R.
- I recommend using Prism unless you have confidence in your understanding of statistical equations.

**Contingency table: sensitivity vs. specificity**

# Porphyria is a class of diseases caused by impaired heme synthesis



We focus on Acute Intermediate Porphyria, which is caused by loss-of-function mutations in porphobilinogen deaminase and leads to a build-up of porphobilinogen.



## AIP is not a pleasant disease

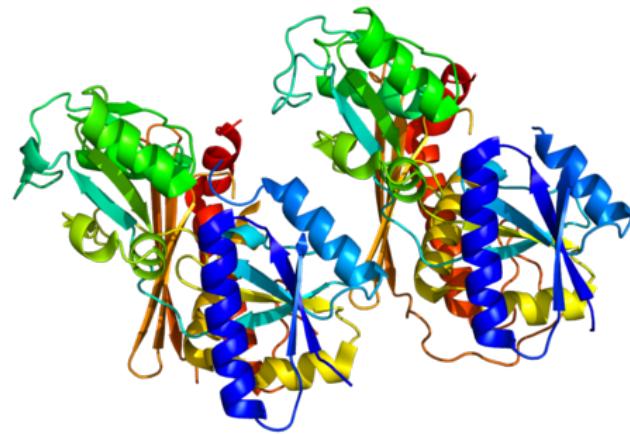
---

"The typical patient with an attack of acute intermittent porphyria is a previously healthy young woman who has had several days of severe fatigue and an inability to concentrate, followed by progressively worsening abdominal pain, nausea, vomiting, and subtle neurologic signs."

## Screening tests are low-cost non-invasive tests given to healthy individuals

---

There is a screening test for AIP, based on the measurement of reduced levels of porphobilinogen deaminase (PBGD) activity in urine or serum.



### Question:

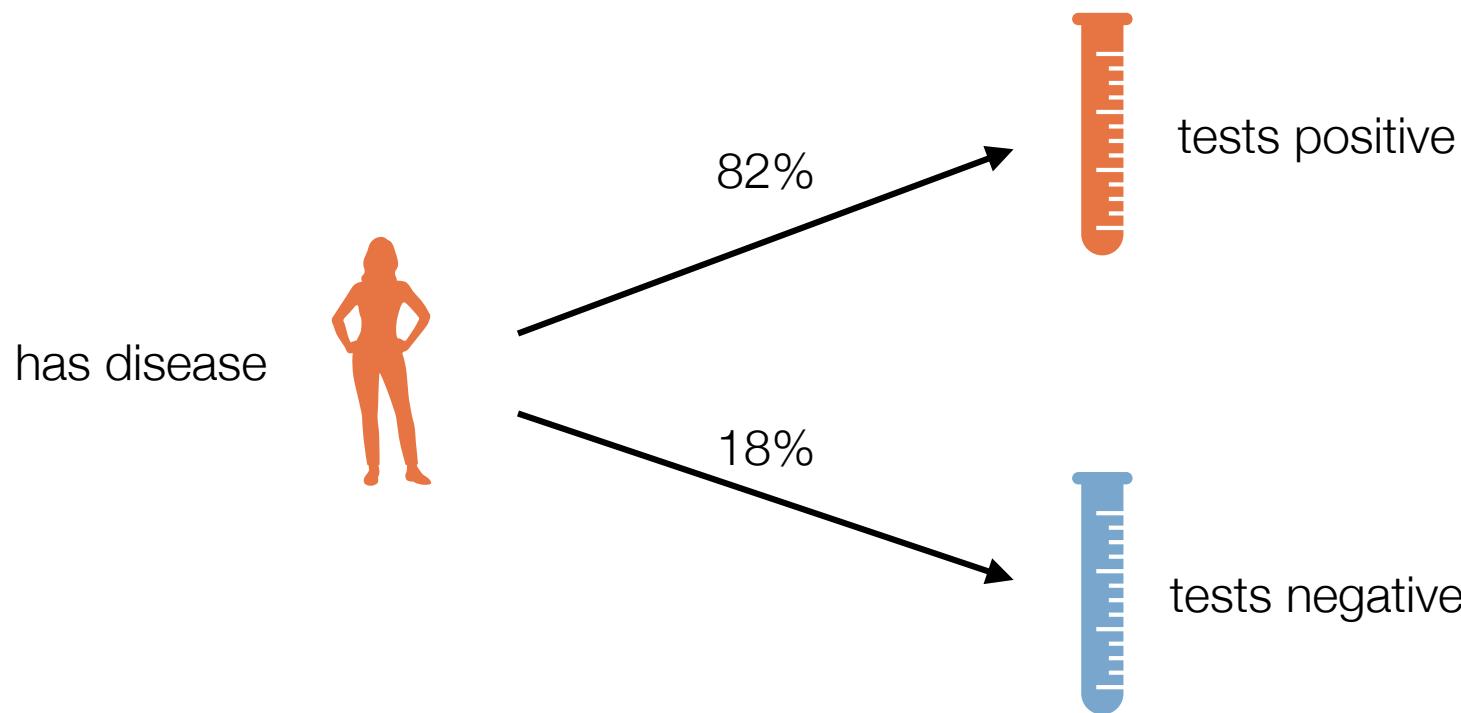
If you test positive for AIP in this screening test, what is the probability that you actually have AIP?

**Sensitivity** is the probability of testing positive given that the subject has the disease.

---

For the AIP test:

$$\text{Sensitivity} = p(\text{test}^+ \mid \text{disease}^+) = 82\%$$

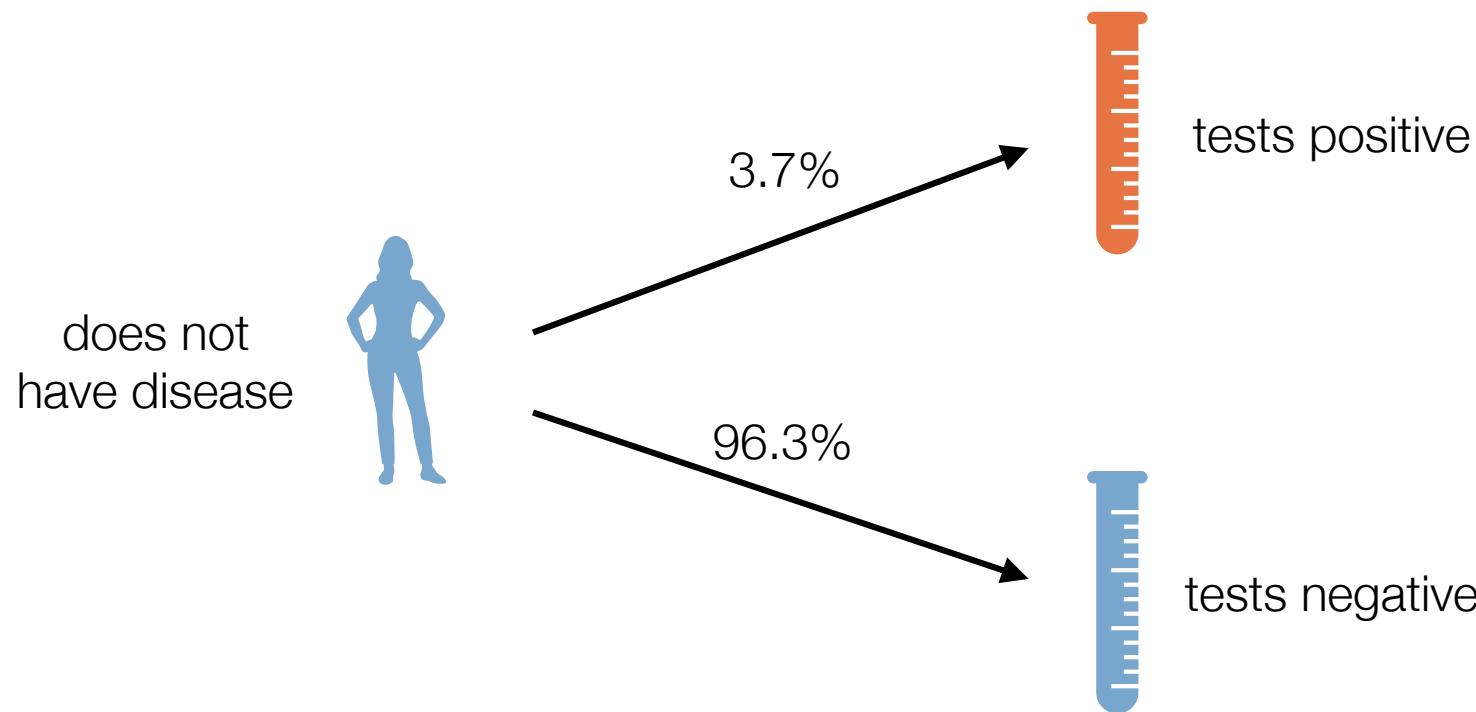


**Specificity** is the probability of a negative test given that the subject does not have the disease.

---

For the AIP test:

$$\text{Specificity} = p(\text{test}^- \mid \text{disease}^-) = 96.3\%$$



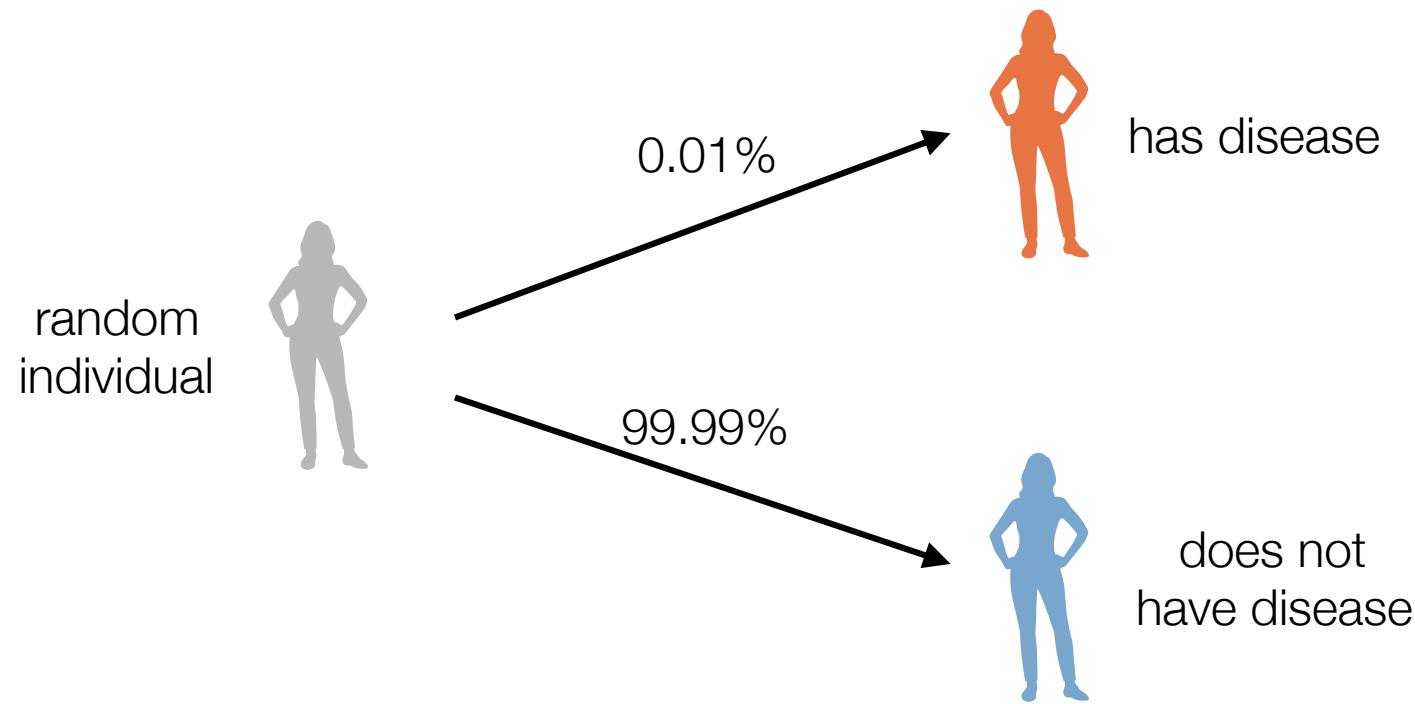
Prevalence is the fraction of individuals in a population who have a disease.

---

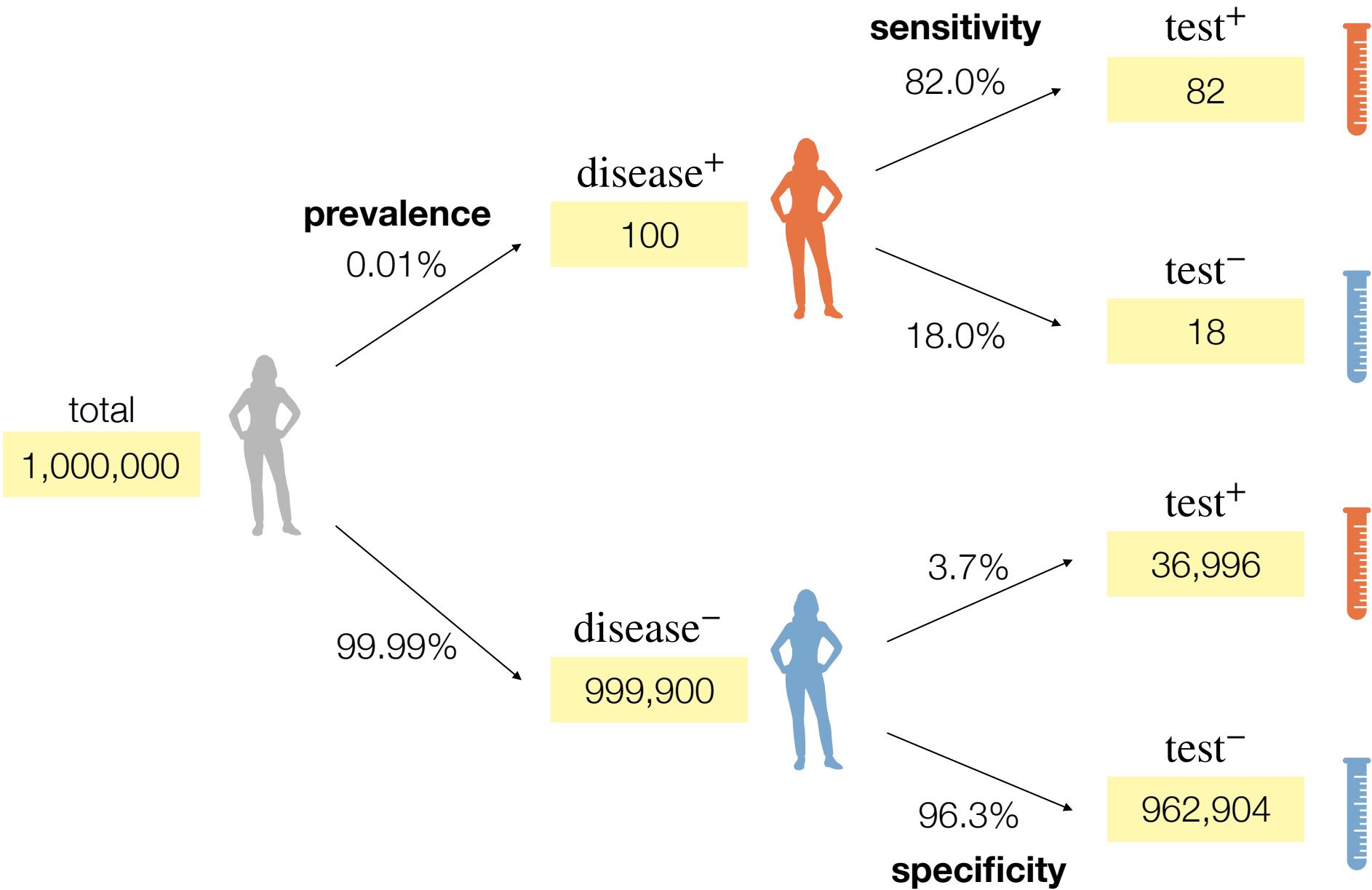
Understanding the results of a medical screening test requires also knowing the prevalence of a disease

For AIP:

$$\text{Prevalence} = p(\text{disease}^+) = 0.01\%$$



## Consider the expected outcome in 1,000,000 randomly chosen individuals



## Contingency tables summarize these results

---

**Contingency table** showing the expected results of the AIP test on 1,000,000 random individuals



		disease <sup>+</sup>	disease <sup>-</sup>
		True positive (TP)	False positive (FP)
test <sup>+</sup>	disease <sup>+</sup>	82	36,996 Type I error
	disease <sup>-</sup>	18	962,904 Type II error

**What person who tests positive truly cares about is the positive predictive value.**

		disease <sup>+</sup>	disease <sup>-</sup>
		test <sup>+</sup>	test <sup>-</sup>
test <sup>+</sup>	disease <sup>+</sup>	82 (TP)	36,996 (FP)
	disease <sup>-</sup>	18 (FN)	962,904 (TN)

**Positive predictive value (PPV):**

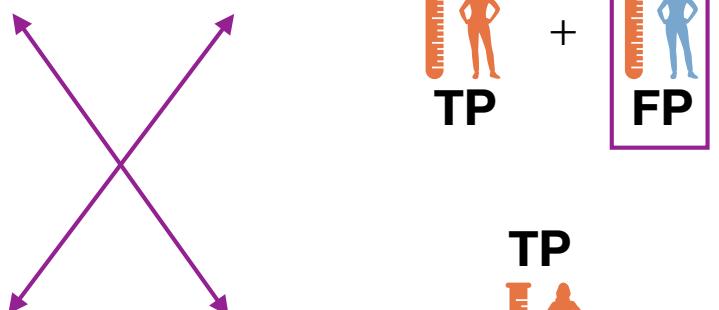
$$p(\text{disease}^+ | \text{test}^+) = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{82}{82 + 36,996} = 0.22\% \text{ (!!!)}$$

TP  
  
TP + FP

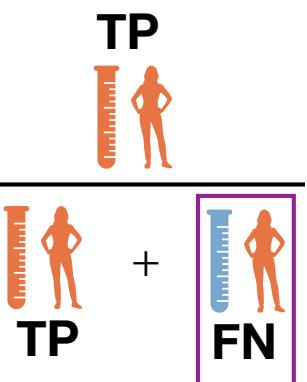
**Even if you test positive, the probability of you having AIP is still very, very low.**

## PPV is often far less than sensitivity in screening tests for rare diseases

**PPV:**  $p(\text{disease}^+ | \text{test}^+) = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{82}{82 + 36,996} = 0.22\%$



**sensitivity:**  $p(\text{test}^+ | \text{disease}^+) = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{82}{82 + 18} = 82\%$



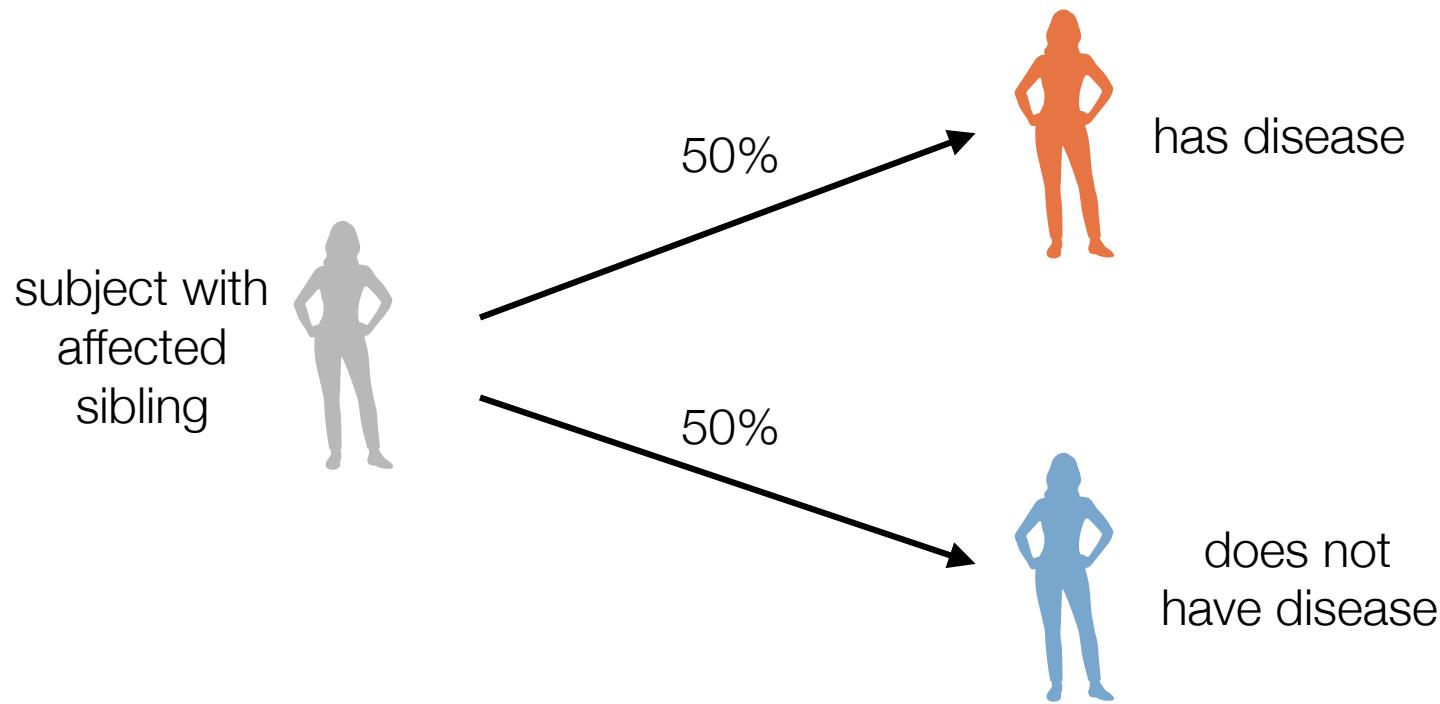
**PPV ≪ sensitivity** because   $\gg$   because   $\gg$   disease<sup>-</sup>  $\gg$  disease<sup>+</sup>

## Porphyria is an autosomal dominant disease

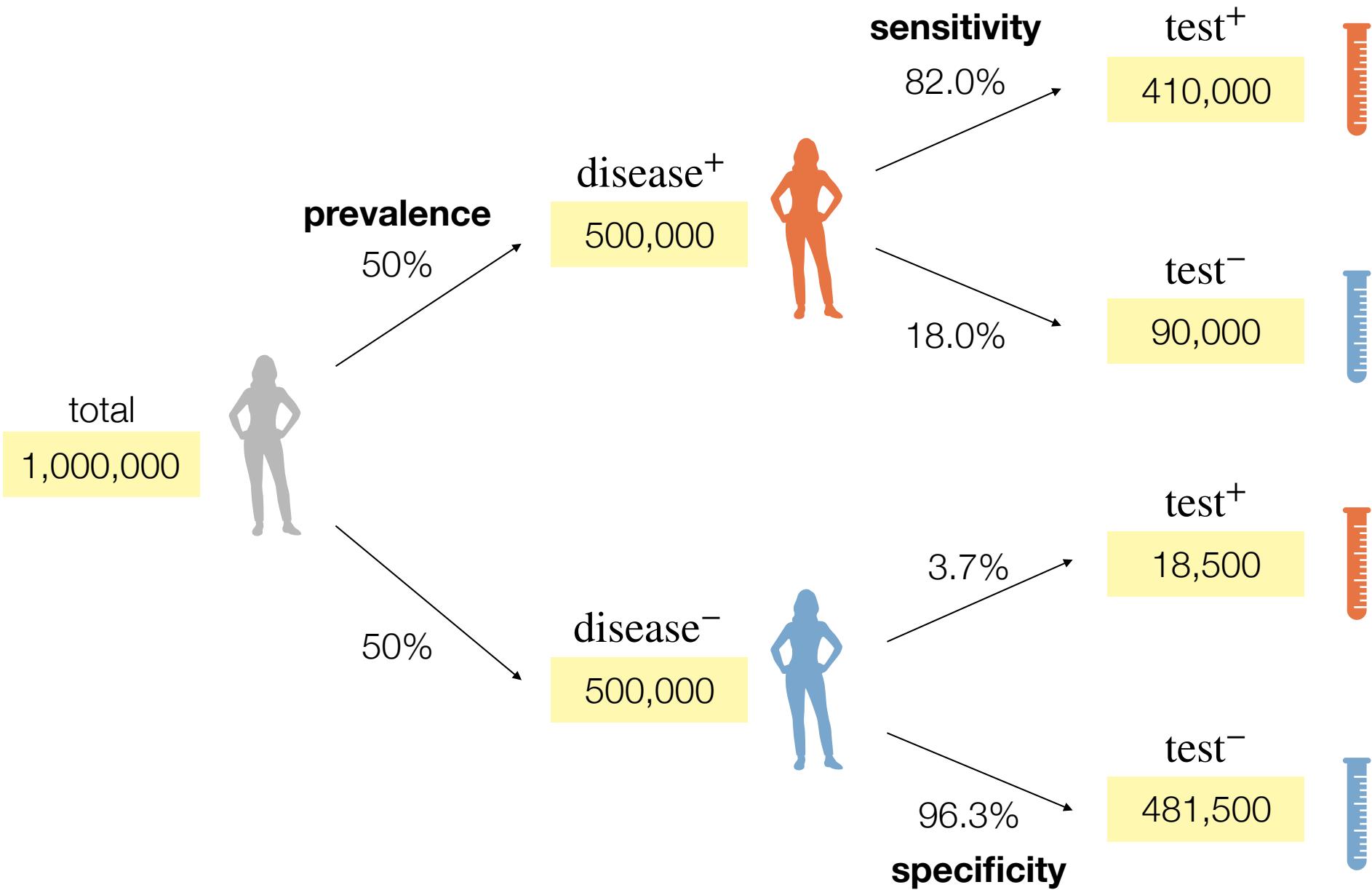
---

If a subject's sibling has AIP,  
there is a 50% chance that they do too.

$$\text{prevalence} = p(\text{disease}^+) = 50\%$$



## Consider the expected outcome in 1,000,000 individuals with affected siblings



**What person who tests positive truly cares about is the positive predictive value.**

		disease <sup>+</sup>	disease <sup>-</sup>
		test <sup>+</sup>	test <sup>-</sup>
test <sup>+</sup>	disease <sup>+</sup>	410,000 (TP)	18,500 (FP)
	disease <sup>-</sup>	90,000 (FN)	481,500 (TN)

**Positive predictive value (PPV):**

$$p(\text{disease}^+ | \text{test}^+) = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{41,000}{41,000 + 18,500} = 95.7\%$$

TP  
  
TP + FP

**Just knowing that you sibling has AIP increases the PPV of the test enormously.**

## In medicine, there is a difference between screening tests and diagnostic tests.

The influence of population is a key reason that doctors distinguish between screening tests and diagnostic tests

	Screening tests	Diagnostic tests
Purpose	To detect potential disease indicators	To establish presence/absence of disease
Target population	Large numbers of asymptomatic, but potentially at risk individuals	Symptomatic individuals to establish diagnosis, or asymptomatic individuals with a positive screening test
Test method	Simple, acceptable to patients and staff	maybe invasive, expensive but justifiable as necessary to establish diagnosis
Positive result threshold	Generally chosen towards high sensitivity not to miss potential disease <b>implies many FPs!</b>	Chosen towards high specificity (true negatives). More weight given to accuracy and precision than to patient acceptability
Positive result	Essentially indicates suspicion of disease (often used in combination with other risk factors) that warrants confirmation	Result provides a definite diagnosis
Cost	Cheap, benefits should justify the costs since large numbers of people will need to be screened to identify a small number of potential cases	Higher costs associated with diagnostic test maybe justified to establish diagnosis.

## The relationship between prevalence, sensitivity, specificity, and PPV is clarified by considering “odds”

---

posterior odds	likelihood ratio	prior odds
$\frac{p(\text{disease}^+   \text{test}^+)}{p(\text{disease}^-   \text{test}^+)}$	$= \frac{p(\text{test}^+   \text{disease}^+)}{p(\text{test}^+   \text{disease}^-)} \times \frac{p(\text{disease}^+)}{p(\text{disease}^-)}$	
$\frac{\text{PPV}}{1 - \text{PPV}}$	$\frac{\text{sensitivity}}{1 - \text{specificity}}$	$\frac{\text{prevalence}}{1 - \text{prevalence}}$
(what you care about)	(property of test)	(property of population)

$\left[ 0.0022 = \frac{0.22\%}{99.78\%} \right]$	$=$	$\left[ 22.2 = \frac{82.0\%}{3.7\%} \right]$	$\times$	$\left[ 10^{-4} = \frac{0.01\%}{99.99\%} \right]$	random individual
$\left[ 22.2 = \frac{95.7\%}{4.3\%} \right]$	$=$	$\left[ 22.2 = \frac{82.0\%}{3.7\%} \right]$	$\times$	$\left[ 1 = \frac{50\%}{50\%} \right]$	sibling of affected individual

The base rate fallacy describes the human tendency to discount prior information

---

$$\text{posterior odds} = \text{likelihood ratio} \times \text{prior odds}$$

**base rate fallacy:** If presented with related base rate information (i.e. generic, general information) and specific information (information pertaining only to a certain case), the mind tends to ignore the former and focus on the latter.

## In all fairness, it can be very hard to quantify prior odds.

---

The “population” an individual comes from, and thus prior odds, are greatly affected by many hard-to-quantify factors

- Has the individual had any relevant symptoms?
- Does the individual have a relevant family history?
- What is the individual’s ethnicity (ancestry)?
- What is the individual’s sex?
- Has the individual been tested before? How?

Prior odds aren’t a property of an individual per se, but rather one’s state of knowledge about that individual.

**Prior odds (and thus posterior odds) quantify subjective uncertainty.**

## **Statistics is divided into two schools: Frequentist and Bayesian.**

---

Frequentist statistics avoids calculations involving prior odds.

It therefore yields results that are prone to misinterpretation due the base rate fallacy.

However, frequentist statistics is used heavily in biological research, so you have to learn it anyway.

Frequentist statistics is still useful and informative if you know what to watch out for.

Bayesian statistics explicitly accounts for prior odds.

It therefore requires prior information that is often hard to quantify.

Bayesian statistics is central to modern machine learning and more advanced areas of quantitative biology.

Experimental researchers in biology tend not use Bayesian statistics, so in this specific course won't discuss it much.

Welcome to GraphPad Prism

GraphPad Prism Version 8.4.3 (471)

**NEW TABLE & GRAPH**

- XY
- Column
- Grouped
- Contingency
- Survival
- Parts of a graph
- Multiple variables
- Nested

**EXISTING FILE**

- Open a File
- LabArchives
- Clone a Graph
- Graph Portfolio

Contingency tables: Each row defines a treatment or exposure, each column defines an outcome, and each value is an exact count of objects or events

Table format

	A	B
	Cases	Control
1	Smoked	
2	Never smoked	

Bar chart showing the distribution of smoking status (Smoked vs. Never smoked) between Cases (purple) and Control (blue) groups.

Smoked      Never smoked

?

Learn more

**Data table:**

- Enter or import data into a new table
- Start with sample data to follow a tutorial

**Select a tutorial data set:**

- Chi-square test of prospective data (aspirin and MI)
- Fishers exact test of retrospective data (smoking and cancer)
- Sensitivity and specificity (HIV)
- Chi-square test for trend

Prism Tips

Cancel

Create



Untitled

Table format: **Contingency**

		Outcome A	Outcome B	Outcome C	Outcome D	Outcome E	Outcome F	Outcome G
	HIV antigen	Y	Y	Y	Y	Y	Y	Y
1	p24 antigen +	48	0					
2	p24 antigen -	8	382					
3	Title							
4	Title							
5	Title							
6	Title							
7	Title							
8	Title							
9	Title							
10	Title							
11	Title							
12	Title							
13	Title							
14	Title							
15	Title							
16	Title							
17	Title							
18	Title							
19	Title							
20	Title							
21	Title							
22	Title							
...								

**How the data are organized**  
The columns represent presence or absence of HIV antigen among patients with symptoms suggestive of HIV infection. The rows represent the results of a simpler test. The values are the number of subjects in each group. Data from: Daar et. al., Annals of Internal Medicine, 134:25-29 (2001).

**The goal**  
To quantify the sensitivity (what fraction of people with the disease are identified by the test) and specificity (what fraction of healthy people have a negative test result), with confidence intervals.

**How to analyze the data**

1. Click Analyze
2. Choose "Chi-square (and Fisher's exact) test" from the list of analyses for contingency tables.
3. Click OK.
4. Choose Fisher's exact test and check the option to compute the sensitivity, specificity and predictive values.

**?** Step by step instructions for analyzing contingency tables

Search: Sensitivity and specificity (HIV)

Row 1, A: HIV antigen

lec1\_aip\_test.pzfx — Edited

Search

**Data Tables**

- Sensitivity and specificity (AIP)
- New Data Table...

**Info**

- Project info 1
- New Info...

**Results**

- New Analysis...
- Graphs
- Sensitivity and specificity (AIP)
- New Graph...

**Layouts**

- New Layout...

Family

- Sensitivity and specificity (AIP)
- Sensitivity and specificity (AIP)

Table format: Contingency

		Outcome A	Outcome B	Outcome C
	Contingency	AIP disease +	AIP disease -	Title
1	PBGD test +	82	36996	Y
2	PBGD test -	18	962904	Y
3	Title			
4	Title			
5	Title			
6	Title			
7	Title			
8	Title			
9	Title			
10	Title			
11	Title			
12	Title			
13	Title			
14	Title			
15	Title			
16	Title			
17	Title			
18	Title			
19	Title			
20	Title			

## Create New Analysis

### Data to analyze

Table: Sensitivity and specificity (AIP)

### Type of analysis

Which analysis?

- ▼ **Transform, Normalize...**
  - Transform
  - Transform concentrations (X)
  - Normalize
  - Prune rows
  - Remove baseline and column math
  - Transpose X and Y
  - Fraction of Total
- **XY analyses**
- **Column analyses**
- **Grouped analyses**
- ▼ **Contingency table analyses**
  - Chi-square (and Fisher's exact) test** 
  - Row means with SD or SEM
  - Fraction of Total
- **Survival analyses**
- **Parts of whole analyses**
- **Multiple variable analyses**
- **Nested analyses**
- **Generate curve**
- **Simulate data**
- **Recently used**

Analyze which data sets?

- A:AIP disease +
- B:AIP disease -

Select All

Deselect All



Cancel

OK

## Parameters: Chi-square (and Fisher's exact) test

Main Calculations Options

### Effect sizes to report

Relative Risk

Used for prospective and experimental studies

Difference between proportions (attributable risk) and NNT

Used for prospective and experimental studies

Odds ratio

Used for retrospective case-control studies

Sensitivity, specificity and predictive values

Used for diagnostic tests

### Method to compute the P value

Fisher's exact test

Yates' continuity corrected chi-square test

Chi-square test

Chi-square test for trend

Looking for the z test to compare proportions? Choose the chi-square test (with or without the Yates' correction). The chi-square and z tests are equivalent.



Cancel

OK

lec1\_aip\_test.pzfx — Edited

Contingency

1	Table Analyzed	Sensitivity and specificity (AIP)		
2				
3	P value and statistical significance			
4	Test	Fisher's exact test		
5	P value	<0.0001		
6	P value summary	****		
7	One- or two-sided	Two-sided		
8	Statistically significant ( $P < 0.05$ )?	Yes		
9	Effect size	Value	95% CI	
10	Sensitivity	0.8200	0.7333 to 0.8830	
11	Specificity	0.9630	0.9626 to 0.9634	
12	Positive Predictive Value	0.002212	0.001782 to 0.002744	
13	Negative Predictive Value	1.000	1.000 to 1.000	
14	Likelihood Ratio	22.16		
15	Methods used to compute CIs			
16	Sensitivity, specificity, etc.	Wilson-Brown		
17				
18				
19				
20	Data analyzed	AIP disease +	AIP disease -	Total
21	PBGD test +	82	36996	37078
22	PBGD test -	18	962904	962922
23	Total	100	999900	1000000
24				

Contingency of Sensitivity and specificity (AIP)

Row 1, Column A

Effect size

Value

95% CI

Sensitivity

0.8200

0.7333 to 0.8830

Specificity

0.9630

0.9626 to 0.9634

Positive Predictive Value

0.002212

0.001782 to 0.002744

Negative Predictive Value

1.000

1.000 to 1.000

Likelihood Ratio

22.16

Methods used to compute CIs

Sensitivity, specificity, etc.

Wilson-Brown

Data analyzed

AIP disease +

AIP disease -

Total

PBGD test +

82

36996

37078

PBGD test -

18

962904

962922

Total

100

999900

1000000

**Contingency table: prospective study**

## Does taking aspirin daily affect one's chance of myocardial infarction (MI)

---

	MI	no MI
placebo	189	10,845
aspirin	104	10,933

NEJM 318: 262-264 (1988)

### **Null hypothesis:**

Aspirin usage has no effect on MI risk

### **Alternative hypothesis:**

Aspirin increases or decreases MI risk.

### **Statistical test:**

Fisher's exact test

## Statistical test: Fisher's exact test

---

	column 1	column 2
row 1	a	b
row 2	c	d

### Mathematical formalization:

Is there a statistical dependence between the row an observation falls in and the column that observation falls in?

### Null hypothesis:

There is no statistical dependence:  $p(\text{row}, \text{column}) = p(\text{row}) \times p(\text{column})$

### Alternative hypothesis:

There is a statistical dependence:  $p(\text{row}, \text{column}) \neq p(\text{row}) \times p(\text{column})$

W Fisher's exact test - Wikipedia

en.wikipedia.org/wiki/Fisher%27s\_exact\_test

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

# Fisher's exact test

From Wikipedia, the free encyclopedia

**Fisher's exact test** is a [statistical significance](#) test used in the analysis of [contingency tables](#).<sup>[1][2][3]</sup> Although in practice it is employed when [sample](#) sizes are small, it is valid for all sample sizes. It is named after its inventor, [Ronald Fisher](#), and is one of a class of [exact tests](#), so called because the significance of the deviation from a [null hypothesis](#) (e.g., [P-value](#)) can be calculated exactly, rather than relying on an approximation that becomes exact in the limit as the sample size grows to infinity, as with many statistical tests.

Fisher is said to have devised the test following a comment from [Muriel Bristol](#), who claimed to be able to detect whether the tea or the milk was added first to her cup. He tested her claim in the "[lady tasting tea](#)" experiment.<sup>[4]</sup>

**Contents** [hide]

- [1 Purpose and scope](#)
- [2 Example](#)
- [3 Controversies](#)
- [4 Alternatives](#)
- [5 See also](#)
- [6 References](#)
- [7 External links](#)

WIKIPEDIA  
The Free Encyclopedia

Main page  
Contents  
Current events  
Random article  
About Wikipedia  
Contact us  
Donate  
Contribute  
Help  
Community portal  
Recent changes  
Upload file  
Tools  
What links here  
Related changes  
Special pages  
Permanent link  
Page information

Welcome to GraphPad Prism

GraphPad Prism Version 8.4.3 (471)

**NEW TABLE & GRAPH**

- XY
- Column
- Grouped
- Contingency
- Survival
- Parts of a graph
- Multiple variables
- Nested

**EXISTING FILE**

- Open a File
- LabArchives
- Clone a Graph
- Graph Portfolio

Contingency tables: Each row defines a treatment or exposure, each column defines an outcome, and each value is an exact count of objects or events

Table format

	A	B
	Cases	Control
1	Smoked	
2	Never smoked	

Bar chart showing the distribution of smoking status (Smoked vs. Never smoked) between Cases (purple) and Control (blue) groups.

Smoked      Never smoked

?

Learn more

**Data table:**

- Enter or import data into a new table
- Start with sample data to follow a tutorial

**Select a tutorial data set:**

- Chi-square test of prospective data (aspirin and MI)
- Fisher's exact test of retrospective data (smoking and cancer)
- Sensitivity and specificity (HIV)
- Chi-square test for trend

Prism Tips

Cancel

Create

Untitled — Edited

Table format: **Contingency**

		Outcome A	Outcome B	Outcome C	Outcome D	Outcome E	Outcome F	Out
		Myocardial Infarction	No MI	Title	Title	Title	Title	
1	Placebo	189	10845					
2	Aspirin	104	10933					
3	Title							
4	Title							
5	Title							
6	Title							
7	Title							
8	Title							
9	Title							
10	Title							
11	Title							
12	Title							
13	Title							
14	Title							
15	Title							
16	Title							
17	Title							
18	Title							
19	Title							
20	Title							
21	Title							
22	Title							
23	Title							

**How the data are organized**  
 This is a prospective study. The two rows represent two treatments assigned randomly to subjects. The two columns represent two alternative outcomes. The values are the number of subjects in each category. Data from: New England Journal Medicine 318: 262-264 (1988).

**Goals**

- To assess whether the discrepancy between incidence of myocardial infarction is more than expected by chance.
- To quantify the relative risk, with its 95% confidence interval.

**How to analyze the data**  
 Click Analyze, choose "Chi-square (and Fisher's exact) test" from the list of analyses for contingency tables, and then choose the chi-square test and computation of relative risk in the dialog. Click below for more detailed instructions, and to learn about contingency tables.

[Step by step instructions for analyzing contingency tables](#)

Search: Prospective (aspirin and MI)

Row 1, A: Myocardial I

## Create New Analysis

### Data to analyze

Table: Prospective (aspirin and MI)

### Type of analysis

Which analysis?

- ▼ **Transform, Normalize...**
  - Transform
  - Transform concentrations (X)
  - Normalize
  - Prune rows
  - Remove baseline and column math
  - Transpose X and Y
  - Fraction of Total
- **XY analyses**
- **Column analyses**
- **Grouped analyses**
- ▼ **Contingency table analyses**
  - Chi-square (and Fisher's exact) test** 
  - Row means with SD or SEM
  - Fraction of Total
- **Survival analyses**
- **Parts of whole analyses**
- **Multiple variable analyses**
- **Nested analyses**
- **Generate curve**
- **Simulate data**
- **Recently used**

Analyze which data sets?

- A:Myocardial Infarction
- B:No MI

Select All

Deselect All



Cancel

OK

## Parameters: Chi-square (and Fisher's exact) test

Main Calculations Options

### Effect sizes to report

- Relative Risk
  - Used for prospective and experimental studies
- Difference between proportions (attributable risk) and NNT
  - Used for prospective and experimental studies
- Odds ratio
  - Used for retrospective case-control studies
- Sensitivity, specificity and predictive values
  - Used for diagnostic tests

### Method to compute the P value

- Fisher's exact test
- Yates' continuity corrected chi-square test
- Chi-square test
- Chi-square test for trend

Looking for the z test to compare proportions? Choose the chi-square test (with or without the Yates' correction). The chi-square and z tests are equivalent.



Cancel

OK



Untitled — Edited

Contingency

Table Analyzed: Prospective (aspirin and MI)

P value and statistical significance:

Test: Fisher's exact test

P value: <0.0001

P value summary: \*\*\*\*

One- or two-sided: Two-sided

Statistically significant (P < 0.05)?: Yes

Effect size:

Relative Risk: 1.818

Reciprocal of relative risk: 0.5501

Attributable risk (P1 - P2): 0.007706

NNT (reciprocal of attrib. risk): 129.8

Methods used to compute CIs:

Relative Risk: Koopman asymptotic score

Attributable risk (P1 - P2): Newcombe/Wilson with CC

Data analyzed:

	Myocardial Infarction	No MI	Total
Placebo	189	10845	11034
Aspirin	104	10933	11037
Total	293	21778	22071

## Results

---

- P value:  $< 0.0001$  (\*\*\*\*), is highly significant, so we **reject the null hypothesis**, concluding that Aspirin affects MI risk.
- **Relative risk:** 1.8 [1.4 to 2.3] meaning that NOT taking Aspirin increases risk of MI.
- **Reciprocal of relative risk:** 0.55 [.43 to .70] meaning that taking Aspirin reduces risk of MI.
- **Attributable risk:** 0.77% [0.46% to 1.08%] quantifies how much the probability of MI decreases due to taking Aspirin
- **Number Needed to treat (NNT):** 130 [92 to 215] quantifies how many individuals would need to take Aspirin in order for one to avoid a MI event.

**Caveats:** Quantifications of risk apply only to MI events during the observational period used in the study; they do not quantify lifetime risk which of course will be higher.

**Contingency table: retrospective study**

## Does smoking affect one's risk of lung cancer

---

	lung cancer	control
smoker	688	658
nonsmoker	21	59

Doll & Hill, British Med. J. (1950)

### **Null hypothesis:**

Smoking does not affect lung cancer risk

### **Alternative hypothesis:**

Smoking increases or decreases lung cancer risk

### **Statistical test:**

Fisher's exact test

Welcome to GraphPad Prism

GraphPad Prism Version 8.4.3 (471)

NEW TABLE & GRAPH

XY

Column

Grouped

Contingency

Survival

Paired

Whole

Multiple variables

Nested

EXISTING FILE

Open a File

LabArchives

Clone a Graph

Graph Portfolio

Contingency tables: Each row defines a treatment or exposure, each column defines an outcome, and each value is an exact count of objects or events

Table format

	A	B
	Cases	Control
1	Smoked	
2	Never smoked	

Bar chart showing the distribution of smoking status (Smoked vs. Never smoked) between Cases (purple) and Control (blue) groups.

Smoked      Never smoked

?

Learn more

Data table:

Enter or import data into a new table

Start with sample data to follow a tutorial

Select a tutorial data set:

Chi-square test of prospective data (aspirin and MI)

Fishers exact test of retrospective data (smoking and cancer)

Sensitivity and specificity (HIV)

Chi-square test for trend

Prism Tips

Cancel

Create



Untitled — Edited

Q Search

▼ Data Tables >> **Retrospective (smoking and canc)**

+ New Data Table...

▼ Info >> **Project info 1**

+ New Info...

▼ Results >> **New Analysis...**

▼ Graphs >> **Retrospective (smoking and canc)**

+ New Graph...

▼ Layout >> **New Layout...**

Family >>

Retrospective (smoking and canc)

Retrospective (smoking and canc)

Table format: **Contingency**

Outcome A    Outcome B    Outcome C    Outcome D    Outcome E    Outcome F    Outcome G

Cases (lung cancer)    Control    Title    Title    Title    Title    Title

		X	Y	Y	Y	Y	Y	Y
1	Smoked		688	650				
2	Never smoked		21	59				
3	Title							
4	Title							
5	Title							
6	Title							
7	Title							
8	Title							
9	Title							
10	Title							
11	Title							
12	Title							
13	Title							
14	Title							
15	Title							
16	Title							
17	Title							
18	Title							
19	Title							
20	Title							
21	Title							
22	Title							

How the data are organized

This is a retrospective case-control study. The two columns represent two groups of subjects. The two rows represent two alternative exposures (smoking or not). The values are the number of subjects who fall into each category. Data are the first to show a relationship between smoking and cancer (Doll and Hill, British Med. J, 1950, 739-748).

Goals

- To assess whether the relationship between cancer and smoking is more than expected by chance.
- To quantify the odds ratio with its 95% confidence interval.

How to analyze the data

Click Analyze, choose "Chi-square (and Fisher's exact) test" from the list of analyses for contingency tables, and then choose the Fisher's exact test and check the option to compute the odd's ratio in the dialog. Click below for more detailed instructions, and to learn about contingency tables.

Step by step instructions for analyzing contingency tables

## Create New Analysis

### Data to analyze

Table: Retrospective (smoking and cancer)

### Type of analysis

Which analysis?

- ▼ **Transform, Normalize...**
  - Transform
  - Transform concentrations (X)
  - Normalize
  - Prune rows
  - Remove baseline and column math
  - Transpose X and Y
  - Fraction of Total
- **XY analyses**
- **Column analyses**
- **Grouped analyses**
- ▼ **Contingency table analyses**
  - Chi-square (and Fisher's exact) test** 
  - Row means with SD or SEM
  - Fraction of Total
- **Survival analyses**
- **Parts of whole analyses**
- **Multiple variable analyses**
- **Nested analyses**
- **Generate curve**
- **Simulate data**
- **Recently used**

Analyze which data sets?

- A:Cases (lung cancer)
- B:Control

Select All

Deselect All



Cancel

OK

## Parameters: Chi-square (and Fisher's exact) test

Main Calculations Options

### Effect sizes to report

Relative Risk

Used for prospective and experimental studies

Difference between proportions (attributable risk) and NNT

Used for prospective and experimental studies

Odds ratio

Used for retrospective case-control studies

Sensitivity, specificity and predictive values

Used for diagnostic tests

### Method to compute the P value

Fisher's exact test

Yates' continuity corrected chi-square test

Chi-square test

Chi-square test for trend

Looking for the z test to compare proportions? Choose the chi-square test (with or without the Yates' correction). The chi-square and z tests are equivalent.



Cancel

OK

Untitled — Edited

Contingency

1	Table Analyzed	Retrospective (smoking and cancer)
2		
3	P value and statistical significance	
4	Test	Fisher's exact test
5	P value	<0.0001
6	P value summary	****
7	One- or two-sided	Two-sided
8	Statistically significant ( $P < 0.05$ )?	Yes
9		
10	Effect size	Value
11	Odds ratio	2.974
12	Reciprocal of odds ratio	0.3363
13		
14	Methods used to compute CIs	
15	Odds ratio	Baptista-Pike
16		
17	Data analyzed	Cases (lung cancer)
18	Smoked	688
19	Never smoked	21
20	Total	709
21		

Contingency of Retrospective (smoking and cancer)

Row 1, Column A

## Results

---

- P value: < 0.0001 (\*\*\*\*), is highly significant, so we **reject the null hypothesis**, concluding that smoking and cancer are associated.
- **Odds ratio:** 3.0 [1.8 to 4.9] meaning that smoking is associated with a nearly 3-fold higher odds of getting cancer.
- **Reciprocal of odds ratio:** 0.34 [.20 to .55] NOT smoking is associated with a nearly 3-fold decrease in the odds of getting cancer.

**Caveats:** These results are from a retrospective study, so we can't conclude that smoking causes cancer, only that it is associated with cancer.

## Relative risk vs. Odds ratio

		Cancer (event)	No Cancer (no event)	Total
Smoker	$a$	$b$	$a+b$	
	$c$	$d$	$c+d$	
Total	$a+c$	$b+d$		

**Risk** is the probability of an event

Risk for smokers:  $a/(a + b)$

Risk for nonsmokers:  $c/(c + d)$

**Relative risk:** 
$$\frac{a/(a + b)}{c/(c + d)}$$

**Odds** is the probability of an event divided by the probability of no event

Odds for smokers:  $a/b$

Odds for nonsmokers:  $c/d$

**Odds ratio:** 
$$\frac{a/b}{c/d}$$

Odds is not affected by the relative number of events vs. no events, and is preferable when this ratio reflects the design of the study, not natural phenomena.

**Questions?**

**Duration of material  
10:00a-11:42a**