

Prism Manual

Categorical Data Analysis

Jingwen Gu
BCBB/OCICB/NIAID

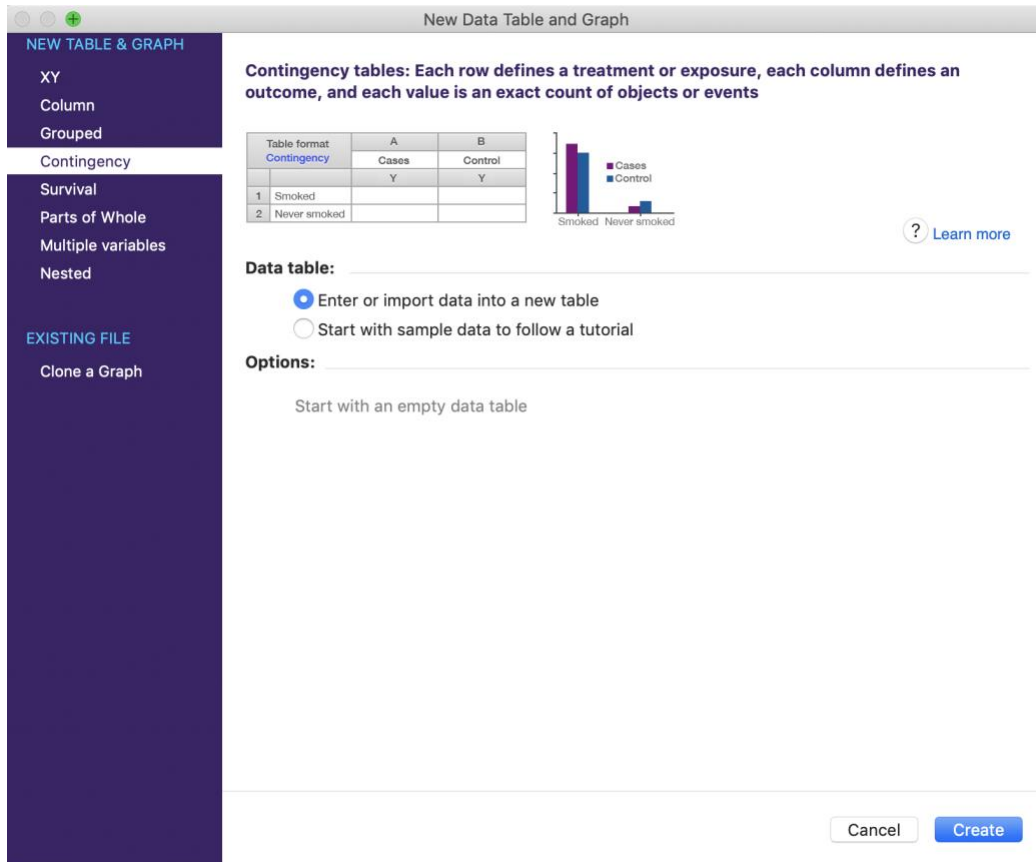
This lab will introduce the methods of conducting categorical data analysis using examples from Categorical data analysis workshop in Prism. For more analysis for categorical data analysis in Prism, please check chapter in [Prism statistics guide](#). To learn more about the examples and interpretation, please categorical data analysis lecture note provided by BCBB.

0. Outline:

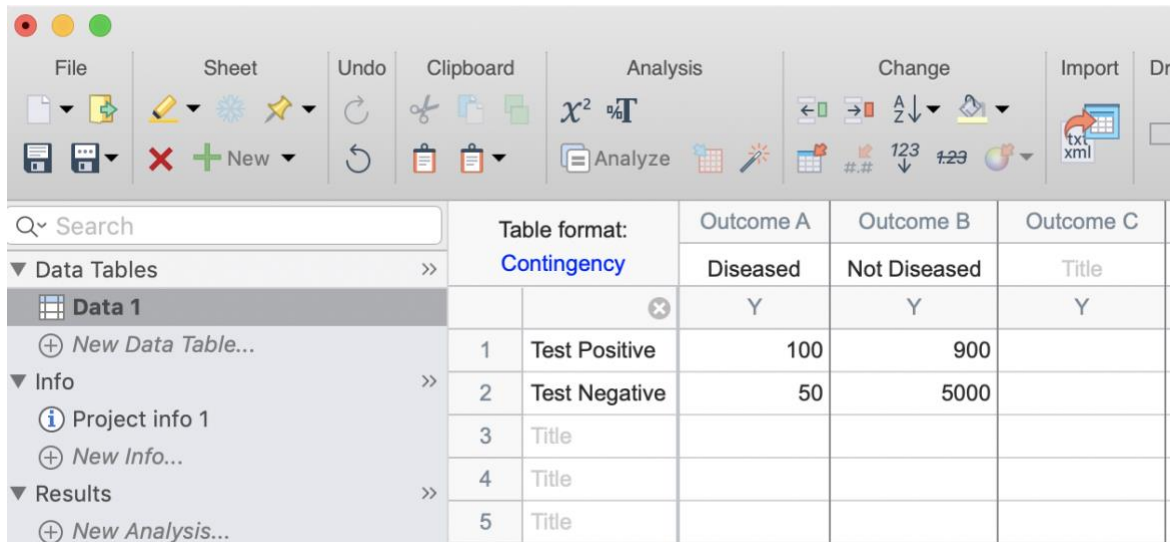
- 1) Positive predictive value and Negative predictive value
- 2) Sensitivity, Specificity
- 3) Odds ratio
- 4) Relative risk
- 5) Perform statistical testing for contingency table

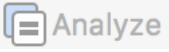
1. Positive predictive value and Negative predictive value

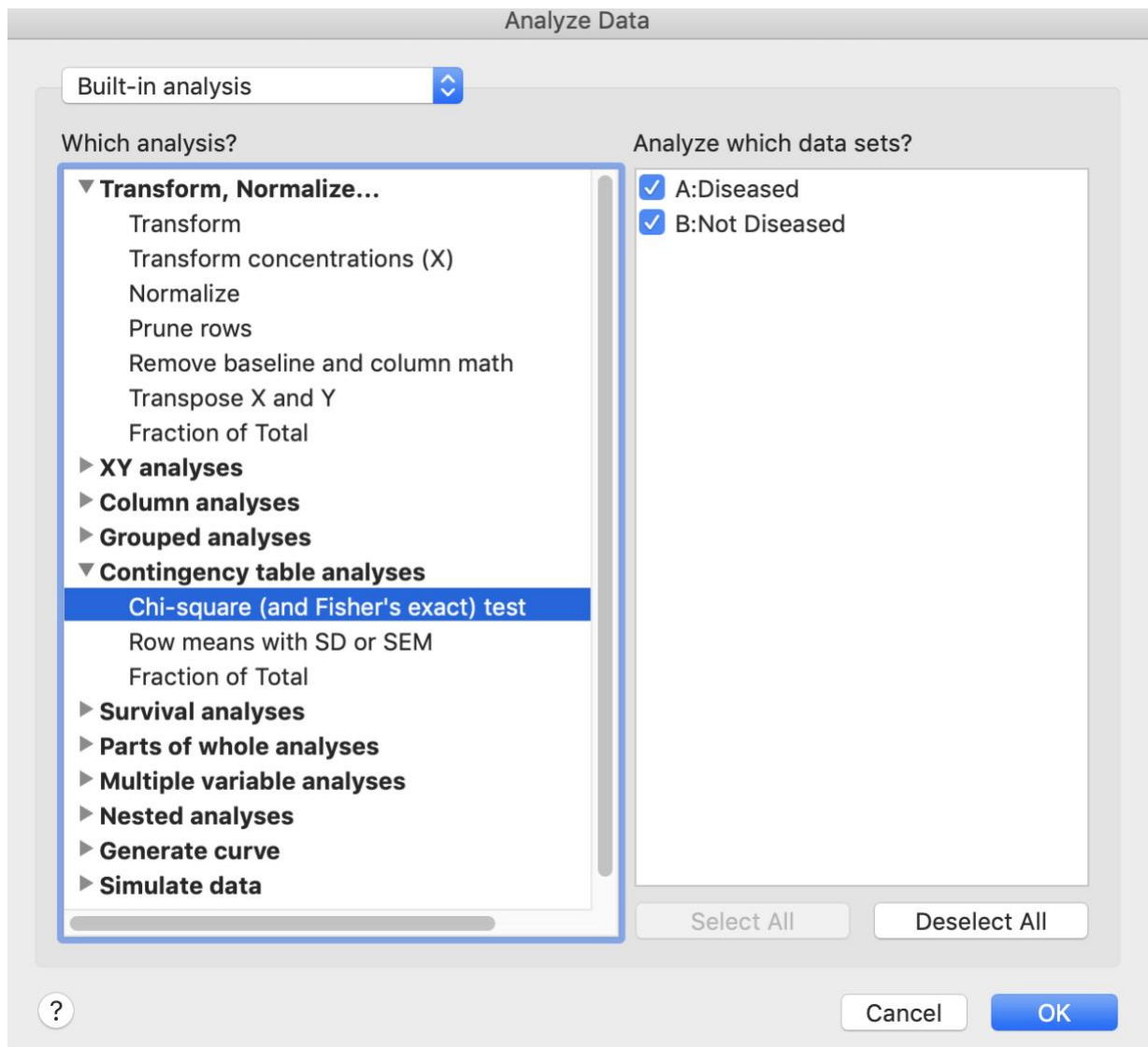
1.1 Copy the data 1 from Prism example.csv. Open Prism, Select **Contingency** under “New Table & Graph”; Select **Enter or import data into a new table**, then click **Create**.



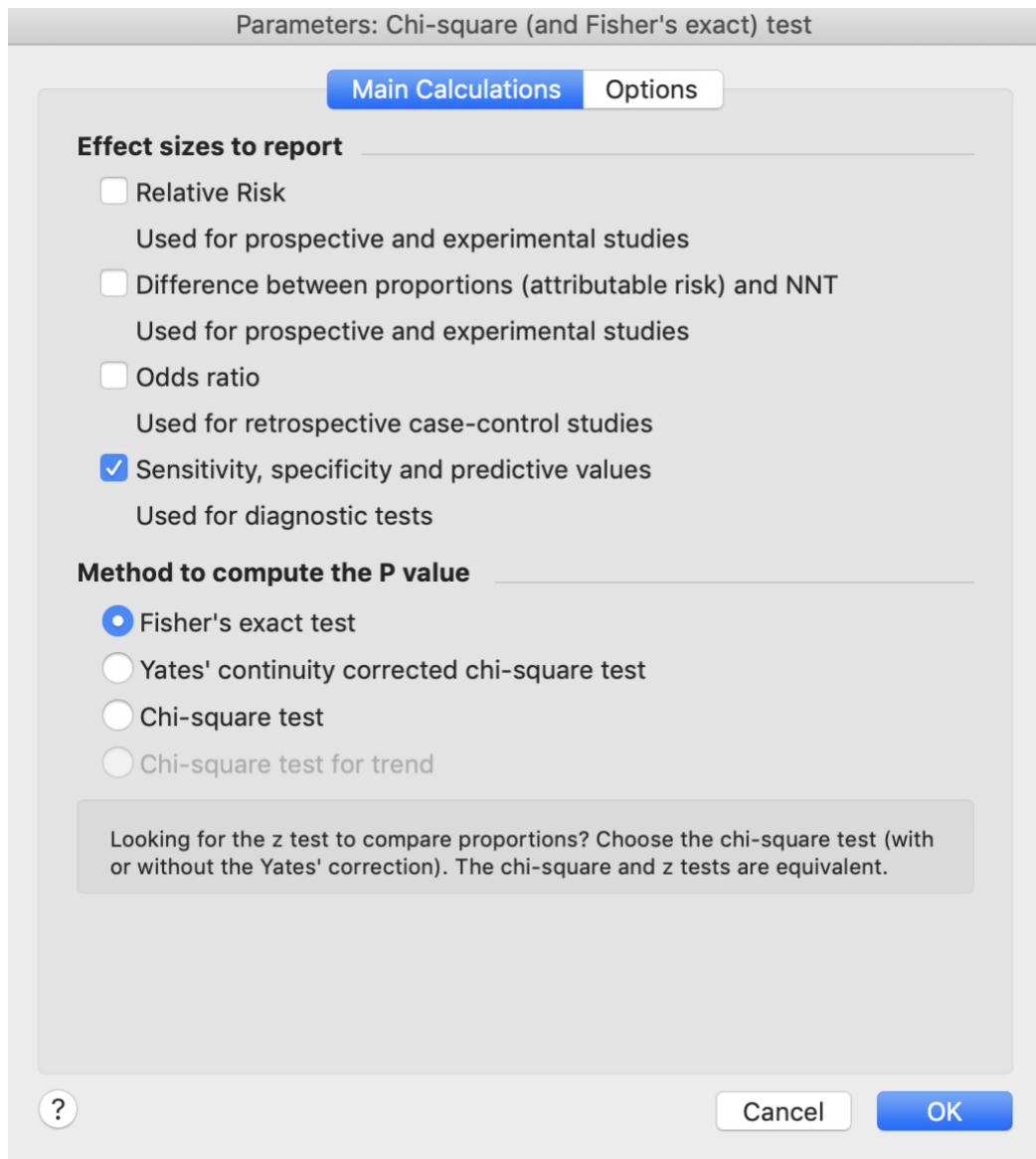
1.2 Copy data 1 from “Prism examples.xlsx” into the blank sheet.



1.3 Click **Analyze**  then select **Chi-square (and Fisher's exact) test** under Contingency table analyses list, then click **OK**.



1.4 Check **Sensitivity, specificity and predictive values** under Main Calculations menu.
Click **OK**.

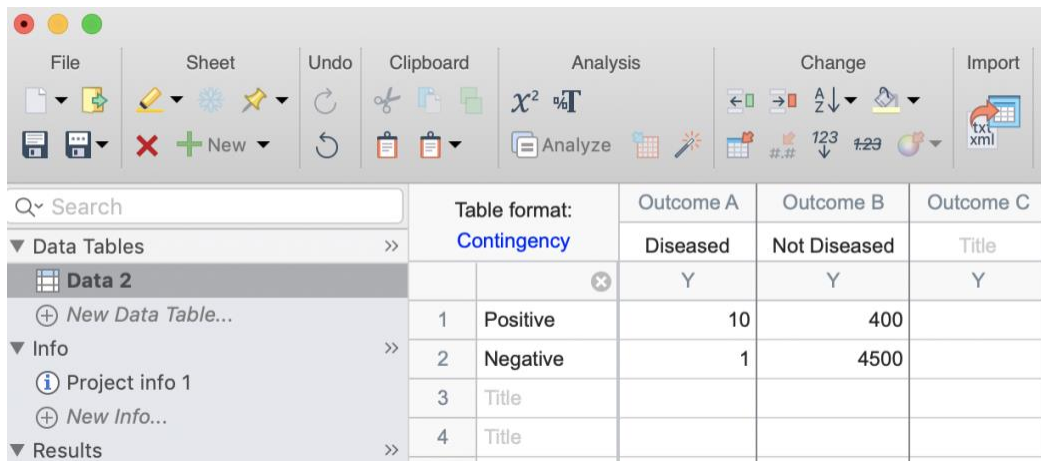


1.5 Prism will return the report with 95% confidence interval. They are circled in red below.

Contingency					
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	95% CI		
11	Sensitivity	0.6667	0.5879 to 0.737		
12	Specificity	0.8475	0.8381 to 0.856		
13	Positive Predictive Value	0.1000	0.08291 to 0.12		
14	Negative Predictive Value	0.9901	0.9870 to 0.992		
15	Likelihood Ratio	4.370			
16					
17	Methods used to compute CIs				
18	Sensitivity, specificity, etc.	Wilson-Brown			
19					
20	Data analyzed	Diseased	Not Diseased	Total	
21	Test Positive	100	900	1000	
22	Test Negative	50	5000	5050	
23	Total	150	5900	6050	
24					
25	Percentage of row total	Diseased	Not Diseased		
26	Test Positive	10.00%	90.00%		
27	Test Negative	0.99%	99.01%		
28					
29	Percentage of column total	Diseased	Not Diseased		
30	Test Positive	66.67%	15.25%		
31	Test Negative	33.33%	84.75%		
32					
33	Percentage of grand total	Diseased	Not Diseased		
34	Test Positive	1.65%	14.88%		
35	Test Negative	0.83%	82.64%		
36					

2. Sensitivity, specificity

2.1 Open a blank contingency sheet (see 1.1). **Copy data 2** from “Prism examples.xlsx”.



The screenshot shows the Prism software interface with a contingency table. The table has the following structure:

		Outcome A	Outcome B	Outcome C
Table format:		Diseased	Not Diseased	Title
		Y	Y	Y
1	Positive	10	400	
2	Negative	1	4500	
3	Title			
4	Title			

2.2 **Follow 1.3 and 1.4** (same way as calculating PPV and NPV). Prism will return result with 95% confidence interval.

Contingency				
1	Table Analyzed	Data 2		
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	95% CI	
11	Sensitivity	0.9091	0.6226 to 0.995	
12	Specificity	0.9184	0.9104 to 0.925	
13	Positive Predictive Value	0.02439	0.01330 to 0.04	
14	Negative Predictive Value	0.9998	0.9987 to 1.000	
15	Likelihood Ratio	11.14		
16				
17	Methods used to compute CIs			
18	Sensitivity, specificity, etc.	Wilson-Brown		
19				
20	Data analyzed	Diseased	Not Diseased	Total
21	Positive	10	400	410
22	Negative	1	4500	4501
23	Total	11	4900	4911
24				
25	Percentage of row total	Diseased	Not Diseased	
26	Positive	2.44%	97.56%	
27	Negative	0.02%	99.98%	
28				
29	Percentage of column total	Diseased	Not Diseased	
30	Positive	90.91%	8.16%	
31	Negative	9.09%	91.84%	
32				
33	Percentage of grand total	Diseased	Not Diseased	
34	Positive	0.20%	8.14%	
35	Negative	0.02%	91.63%	

3. Odds ratio

3.1 Open a blank contingency sheet (see 1.1). **Copy data 3** from “Prism examples.xlsx”.

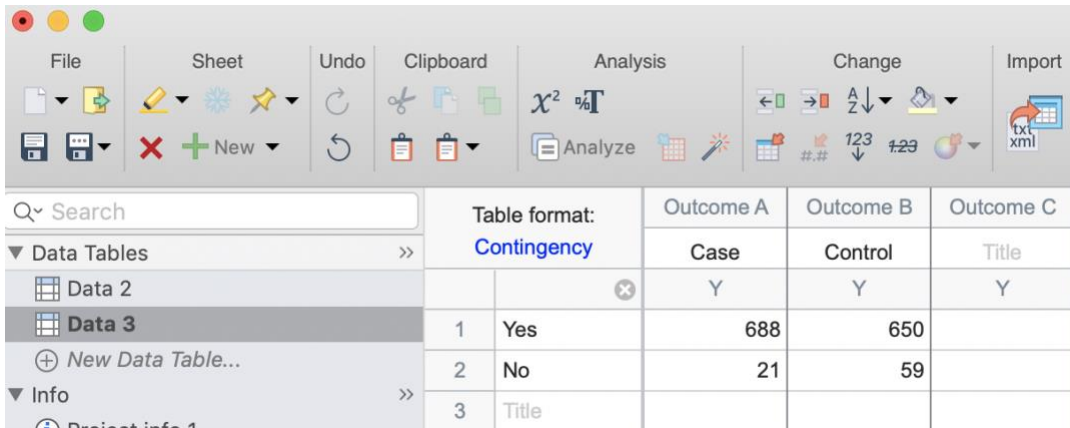
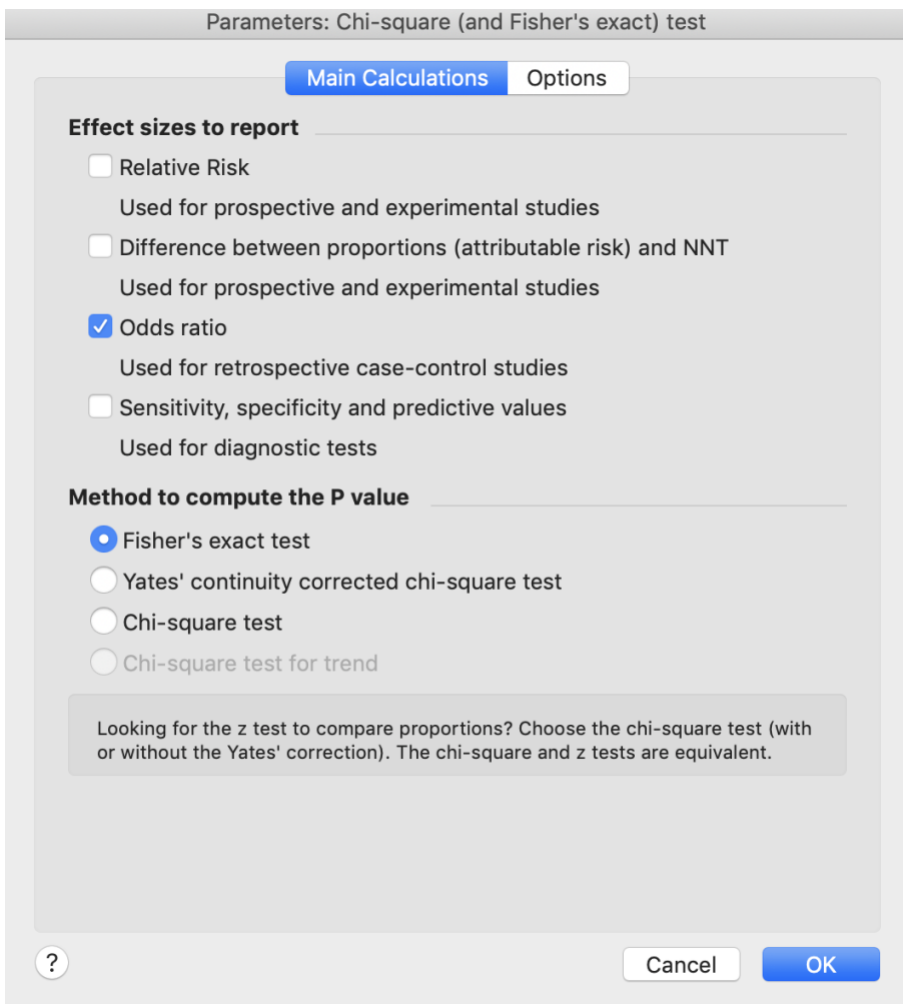


Table format:		Outcome A	Outcome B	Outcome C
Contingency		Case	Control	Title
		Y	Y	Y
1	Yes	688	650	
2	No	21	59	
3	Title			

3.2 Click **Analyze**, select **Chi-square (and Fisher's exact) test** under Contingency table analyses list, then click **OK**. Select **Odds ratio**. Then **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.

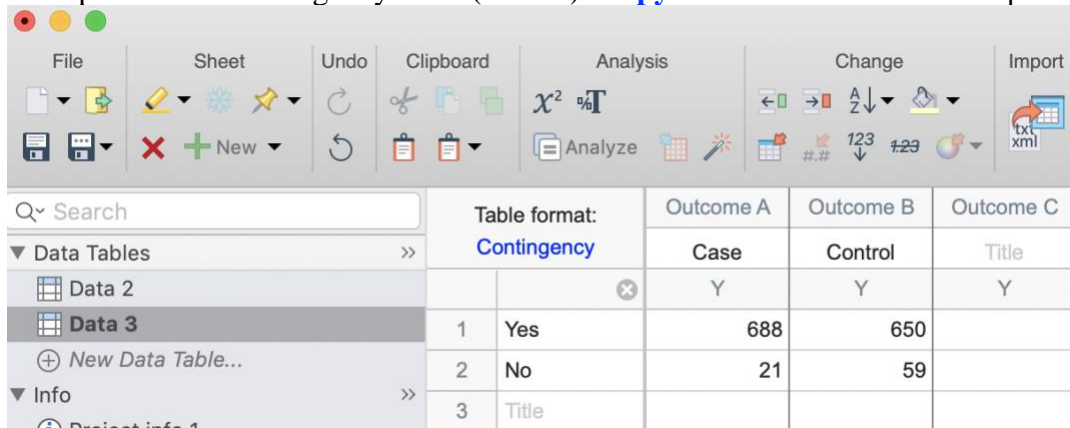
? Cancel OK

3.3 It will return result with 95% confidence interval. It is circled in Red below.

Contingency				
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	95% CI	
11	Odds ratio	2.974	1.819 to 4.900	
12	Reciprocal of odds ratio	0.3363	0.2041 to 0.54	
13				
14	Methods used to compute CIs			
15	Odds ratio	Baptista-Pike		
16				
17	Data analyzed	Case	Control	Total
18	Yes	688	650	1338
19	No	21	59	80
20	Total	709	709	1418
21				
22	Percentage of row total	Case	Control	
23	Yes	51.42%	48.58%	
24	No	26.25%	73.75%	
25				
26	Percentage of column total	Case	Control	
27	Yes	97.04%	91.68%	
28	No	2.96%	8.32%	
29				
30	Percentage of grand total	Case	Control	
31	Yes	48.52%	45.84%	
32	No	1.48%	4.16%	
33				

4. Relative Risk

4.1 Open a blank contingency sheet (see 1.1). **Copy data 4** from “Prism examples.xlsx”.



The screenshot shows the Prism software interface. The top menu bar includes File, Sheet, Undo, Clipboard, Analysis, Change, and Import. The Analysis menu is open, showing options for Chi-square and Fisher's exact test. The main window displays a contingency table with the following data:

Table format:		Outcome A	Outcome B	Outcome C
Contingency		Case	Control	Title
		Y	Y	Y
1	Yes	688	650	
2	No	21	59	
3	Title			

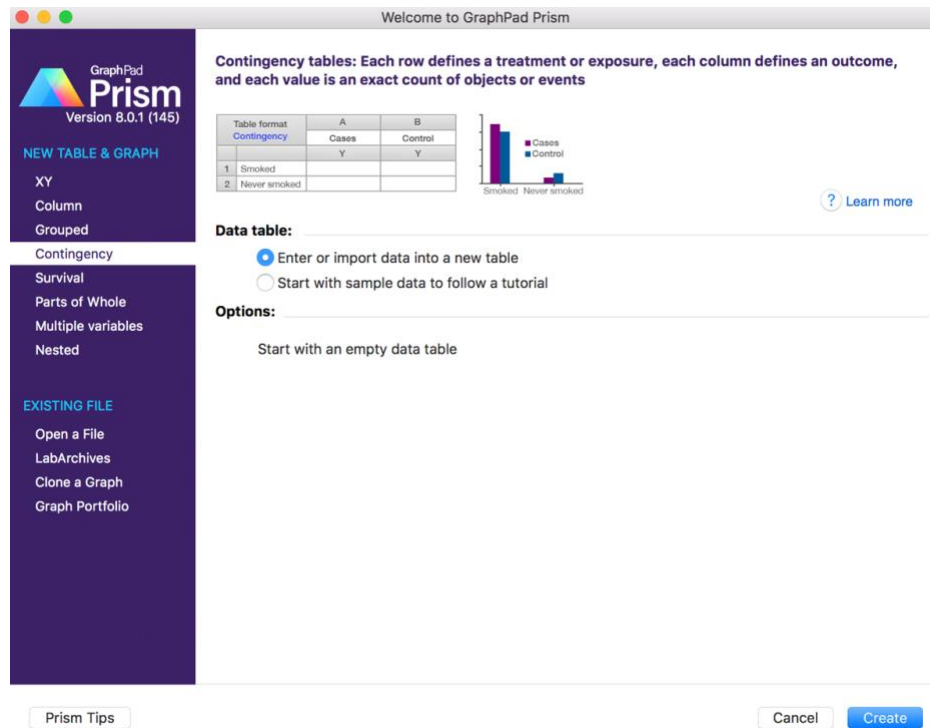
4.2 Click **Analyze**, select **Chi-square (and Fisher's exact) test** under Contingency table analyses list, then click **OK**. Select **Relative risk**. Then **OK**.

4.3 Result return with 95% confidence interval.

Contingency				
1	Table Analyzed	Data 1		
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	95% CI	
11	Relative Risk	0.5501	0.4339 to 0.697	
12	Reciprocal of relative risk	1.818	1.434 to 2.305	
13				
14	Methods used to compute CIs			
15	Relative Risk	Koopman asymptot		
16				
17	Data analyzed	MI	No MI	Total
18	Aspirin	104	10933	11037
19	Placebo	189	10845	11034
20	Total	293	21778	22071
21				
22	Percentage of row total	MI	No MI	
23	Aspirin	0.94%	99.06%	
24	Placebo	1.71%	98.29%	
25				
26	Percentage of column total	MI	No MI	
27	Aspirin	35.49%	50.20%	
28	Placebo	64.51%	49.80%	
29				
30	Percentage of grand total	MI	No MI	
31	Aspirin	0.47%	49.54%	
32	Placebo	0.86%	49.14%	
33				


5.1 Pearson Chi-square test

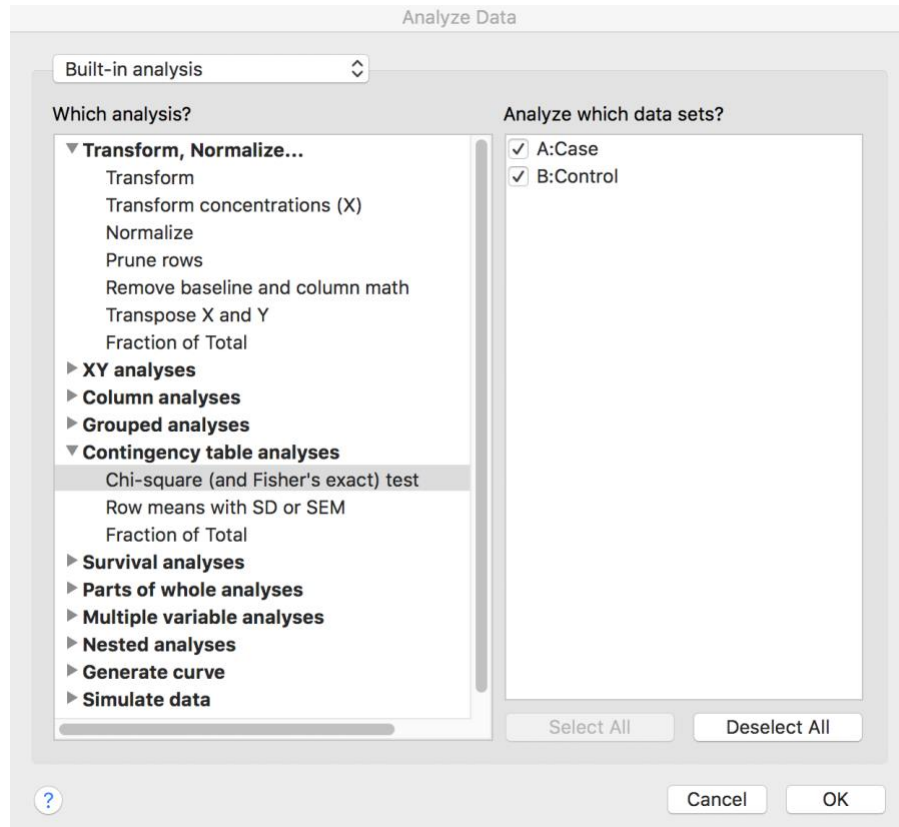
5.1.1 Select **Contingency** under “New Table & Graph”; Select **Enter or import data into a new table**, then click **Create**.



5.1.2 Open a blank contingency sheet (see 1.1). **Copy data 3** from “Prism examples.xlsx”.

Table format:		Outcome A	Outcome B
Contingency		Case	Control
	⊗	Y	Y
1	smoking	688	650
2	not smoking	21	59

5.1.3 Click **Analyze**  **Analyze** then select **Chi-square (and Fisher’s exact) test** under Contingency table analyses list. Check the columns you want to analyze on the right side, then click **OK**.



5.1.4 Select test type under Method to compute P value and other report statistic under Effect sizes to report. Click **Chi-square test**, then **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

5.1.5 Result includes p-value, marginal table, and marginal percentage. Chi-square statistic, degree of freedom and p-value are circled. Test result is significant, which means reject the null hypothesis that smoking and lung cancer are independent.

Contingency				
1	Table Analyzed	Data 3		
2				
3	P value and statistical significance			
4	Test	Chi-square		
5	Chi-square, df	19.13, 1		
6	z	4.374		
7	P value	<0.0001		
8	P value summary	****		
9	One- or two-sided	Two-sided		
10	Statistically significant (P < 0.05)?	Yes		
11				
12	Data analyzed	Case	Control	Total
13	smoking	688	650	1338
14	not smoking	21	59	80
15	Total	709	709	1418
16				
17	Percentage of row total	Case	Control	
18	smoking	51.42%	48.58%	
19	not smoking	26.25%	73.75%	
20				
21	Percentage of column total	Case	Control	
22	smoking	97.04%	91.68%	
23	not smoking	2.96%	8.32%	
24				
25	Percentage of grand total	Case	Control	
26	smoking	48.52%	45.84%	
27	not smoking	1.48%	4.16%	

5.2 Fisher's exact test

5.2.1 Follow 5.1.1 to open a blank data sheet. Copy **data 5** from “Prism examples.xlsx”.

Table format:		Outcome A	Outcome B
Contingency		Guess mik first	Guess tea first
	⊗	Y	Y
1	Pour milk first	3	1
2	Pour tea first	1	3

5.2.2 Select test type under Method to compute P value and other report statistic under Effect sizes to report. Click **Fisher's exact test**, then go to **Options**, select **One-sided** under Calculations options, then click **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

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

Main Calculations
Options

Calculations options

P values: One-sided Two-sided

Confidence Interval:

Method to calculate CI:

Relative risk:

Difference between proportions:

Odds ratio:

Sensitivity, specificity, etc.:

Output

Show this many significant digits (for everything except P values):

P value style: N=

Make these choices be the default for future analyses.

?
Cancel
OK

5.2.3 Review Result includes p-value, marginal table, and marginal percentage. p-value is circled. Test result is not significant, which means cannot reject the null hypothesis that pouring order and guessing order are independent.

Contingency				
1	Table Analyzed	Data 1		
2				
3	P value and statistical significance			
4	Test	Fisher's exact test		
5	P value	0.2429		
6	P value summary	ns		
7	One- or two-sided	One-sided		
8	Statistically significant (P < 0.05)?	No		
9				
10	Data analyzed	Guess mik first	Guess tea first	Total
11	Pour milk first	3	1	4
12	Pour tea first	1	3	4
13	Total	4	4	8
14				
15	Percentage of row total	Guess mik first	Guess tea first	
16	Pour milk first	75.00%	25.00%	
17	Pour tea first	25.00%	75.00%	
18				
19	Percentage of column total	Guess mik first	Guess tea first	
20	Pour milk first	75.00%	25.00%	
21	Pour tea first	25.00%	75.00%	
22				
23	Percentage of grand total	Guess mik first	Guess tea first	
24	Pour milk first	37.50%	12.50%	
25	Pour tea first	12.50%	37.50%	

5.3 McNemar test

McNemar test is available via Prism web. (<https://www.graphpad.com/quickcalcs/>)

5.3.1 Click in the website above, select Categorical data, select **McNemar's test to analyze a matched case-control study**, then click **Continue**.

1. Select category

2. Choose calculator

3. Enter data

4. View results

Analyze categorical data

- Confidence interval of a proportion or count.
- Chi-square. Compare observed and expected frequencies.
- Fisher's and chi-square. Analyze a 2x2 contingency table.
- McNemar's test to analyze a matched case-control study.
- Binomial and sign test. Compare observed and expected proportions.
- NNT (Number Needed to Treat) with confidence interval.
- Predictive values from sensitivity, specificity, and prevalence.
- Kappa. Quantify interrater agreement.

CONTINUE >

5.3.2 **Input # of pairs of case and control**. In our example, 2004 Election as case and 2008 Election as control (or they can be exchanged, result won't change), then click **Calculate**.

[1. Select category](#)

[2. Choose calculator](#)

3. Enter data

[4. View results](#)

McNemar's test to analyze a matched case-control study

McNemar's test is used to compare paired proportions. It can be used to analyze retrospective case-control studies, where each case is matched to a particular control. Or it can be used to analyze experimental studies, where the two treatments are given to matched subjects. [Read an example with explanation.](#)

Risk Factor?

Control Case # of pairs

No	Yes	<input type="text" value="16"/>	<input type="button" value="↕"/>
Yes	No	<input type="text" value="54"/>	<input type="button" value="↕"/>
Yes	Yes	<input type="text" value="175"/>	<input type="button" value="↕"/>
No	No	<input type="text" value="188"/>	<input type="button" value="↕"/>

Calculate

Use McNemar's test (and this calculator) only when you are analyzing matched pairs. Each value you enter above represents a number of PAIRS. The total number of subjects in the study is twice the total of the values you enter above.

Note that the calculations are based entirely on the first two numbers you enter. Enter the remaining two numbers in order to document your full results.

5.3.3 Results include summary, p-value, odds ratio and contingency table. The result is significant, we can reject the null hypothesis that 2004 and 2008 election are independent.

[1. Select category](#)

[2. Choose calculator](#)

[3. Enter data](#)

[4. View results](#)

Results of McNemar's test for a case-control study

Summary:

If there were no association between the risk factor and the disease, you'd expect the number of pairs where cases was exposed to the risk factor but control was not to equal the number of pairs where the control was exposed to the risk factor but the case did not. In this study, there were 70 discordant pairs (case and control had different exposure to the risk factor). There were 54 (77.143%) pairs where the control was exposed to the risk factor but the case was not, and 16 (22.857%) pairs where the case was exposed to the risk factor but the control was not.

P Value:

The two-tailed P value is less than 0.0001

By conventional criteria, this difference is considered to be extremely statistically significant.

The P value was calculated with McNemar's test with the continuity correction.

Chi squared equals 19.557 with 1 degrees of freedom.

The P value answers this question: If there is no association between risk factor and disease, what is the probability of observing such a large discrepancy (or larger) between the number of the two kinds of discordant pairs? A small P value is evidence that there is an association between risk factor and disease.

Odds ratio:

The odds ratio is 0.296, with a 95% confidence interval extending from 0.158 to 0.525

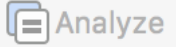
Review your data:

		Control		
		+	-	Total
Case	+	175	16	191
	-	54	188	242
Total		229	204	433

5.4 Chi-square Test for Trend

5.4.1 Follow 5.1.1 to open a blank data sheet. Copy [data 6 from "Prism examples.xlsx"](#).

Table format:		Outcome A	Outcome B	Outcome C
Contingency		Response	Not Reponse	Title
		Y	Y	Y
1	1	10	0	
2	2	9	1	
3	3	10	0	
4	4	7	3	
5	Title			

5.4.2 Click **Analyze**  then select **Chi-square (and Fisher's exact) test** under Contingency table analyses list. Select **Chi-square for trend**. Click Options. Here only two-sided test is default. Then click **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

5.4.3 Review results. P-value is 0.0593. This is two-sided test. We cannot reject the null hypothesis of no linear trend between response across increasing dosage at significance level 0.05.

Contingency		
1	Table Analyzed	Data 1
2		
3	P value and statistical significance	
4	Test	Chi-square test for trend
5	Chi-square, df	3.556, 1
6	P value	0.0593
7	P value summary	ns
8	One- or two-sided	NA
9	Statistically significant (P < 0.05)?	No
10		
11	Data analyzed	
12	Number of rows	4
13	Number of columns	2
14		

In the end, if you have any questions regarding to this topic, please contact me (jingwen.gu@nih.gov) or submit a request to BCBB (bioinformatics@niaid.nih.gov).