

Lesson 3: Measurement of Association for Contingency Tables

Nicky Wakim

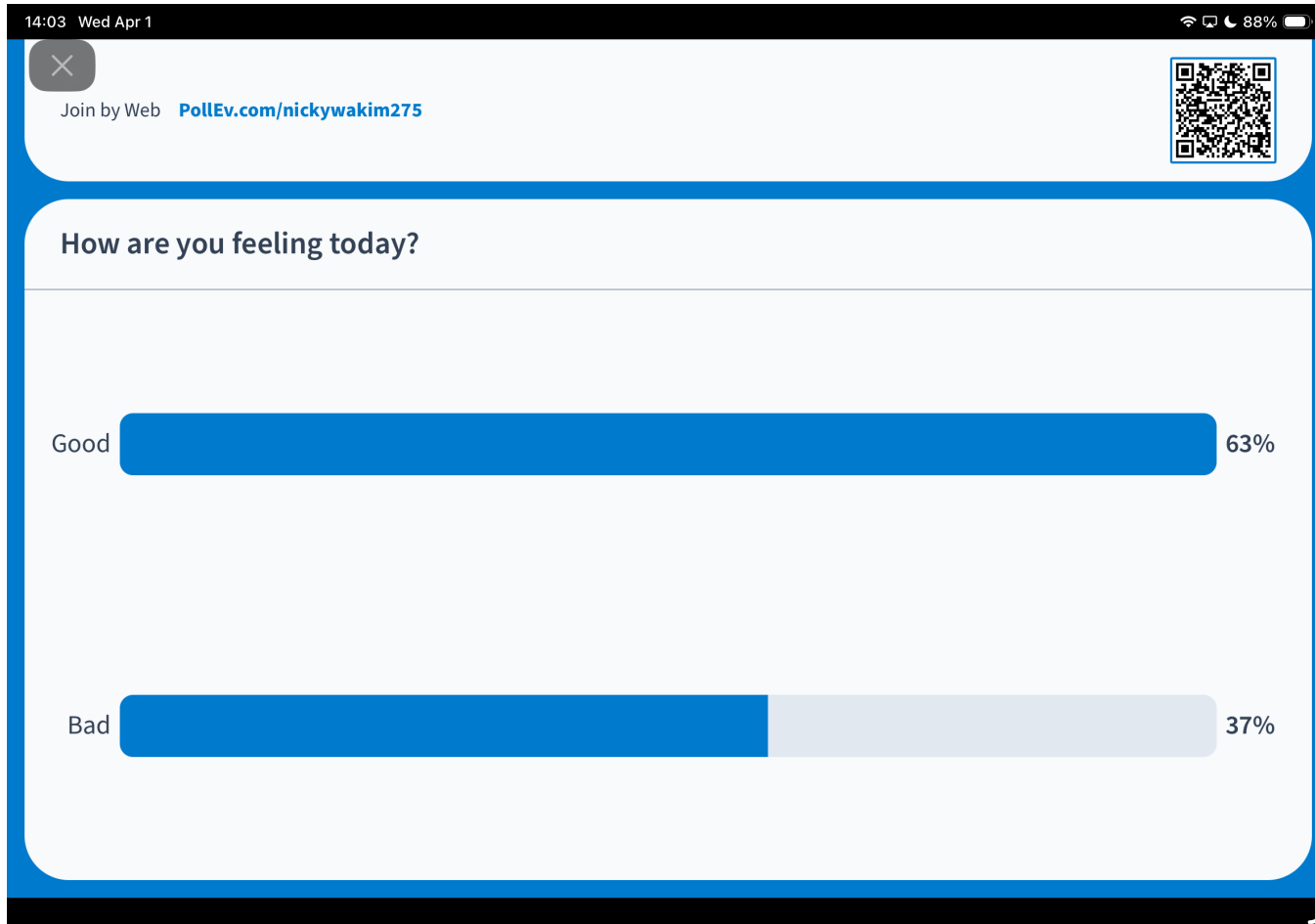
2026-04-01

Learning Objectives

1. Understand the difference between testing for association and measuring association
2. Estimate the risk difference (and its confidence interval) from a contingency table and interpret the estimate.
3. Estimate the risk ratio (and its confidence interval) from a contingency table and interpret the estimate.
4. Estimate the odds ratio (and its confidence interval) from a contingency table and interpret the estimate.

Poll Everywhere Question 1

Make sure to remember your answer!! We'll use this on Wednesday!



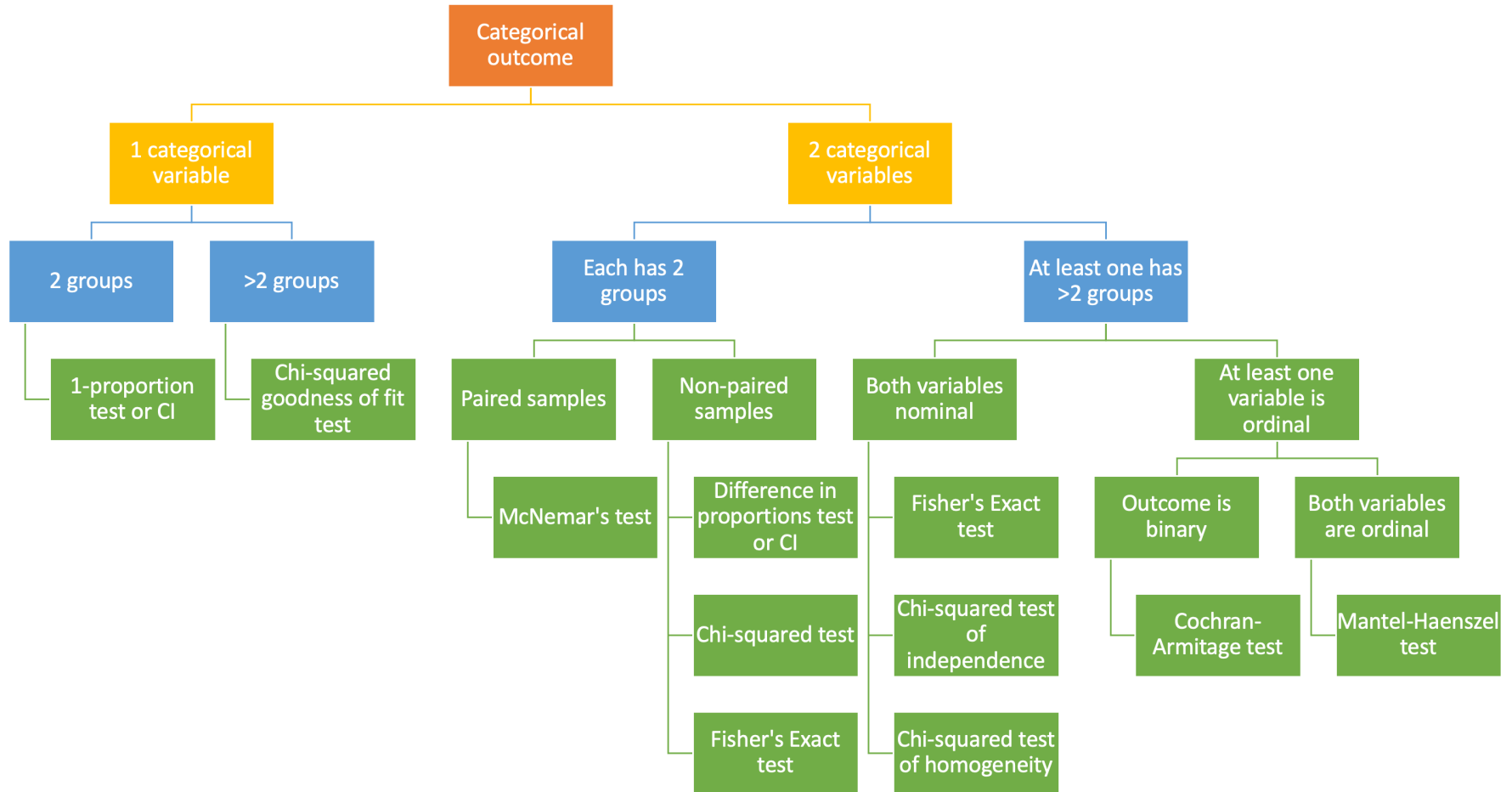
Learning Objectives

1. Understand the difference between testing for association and measuring association
2. Estimate the risk difference (and its confidence interval) from a contingency table and interpret the estimate.
3. Estimate the risk ratio (and its confidence interval) from a contingency table and interpret the estimate.
4. Estimate the odds ratio (and its confidence interval) from a contingency table and interpret the estimate.

Review of Test of Association (1/2)

- Last class: learned some tests of association for contingency tables
- For studies with two independent samples
 - General association
 - Chi-squared test
 - Fisher's Exact test
 - Test of trends
 - Cochran-Armitage test
 - Mantel-Haenszel test

Review of Test of Association (2/2)



Test of association *does not* measure association

- Test of association **does not** provide an effective measure of association.
- The p-value alone is not enough
 - $p\text{-value} < 0.05$ suggests there is a statistically significant association between the group and outcome
 - $p\text{-value} = 0.00001$ vs. $p\text{-value} = 0.01$ does not mean the magnitude of association is different
- But it does not tell **how different** the risks are between the two groups
- We want a measurement to quantify different risks across the groups

Measures of Association

- When we have a **2x2 contingency table** (binary outcome and explanatory variable) and **independent samples**, we have three main options to measure of association:

1. Risk difference (RD)

2. Relative risk (RR) *Risk ratio / prevalence ratio*

3. Odds ratio (OR)

Each measures association by comparing the proportion of successes/failures from each categorical group of our explanatory variable.

Before we discuss each further...

Let's define the cells within a 2x2 contingency table:

Explanatory Variable	Response Variable		Total
	Failure (0)	Success (1)	
Group 1	n_{10}	n_{11}	n_1
Group 2	n_{20}	n_{21}	n_2
Total	n_F	n_S	n

- Then we can define risk/prevalence: the proportion of “successes”

- Risk/proportion/prevalence of successful response for explanatory Group 1: $\widehat{\text{Risk}}_1 = \hat{p}_1 = \frac{n_{11}}{n_1}$
- Context and study design will determine the name

Learning Objectives

1. Understand the difference between testing for association and measuring association
2. Estimate the risk difference (and its confidence interval) from a contingency table and interpret the estimate.
3. Estimate the risk ratio (and its confidence interval) from a contingency table and interpret the estimate.
4. Estimate the odds ratio (and its confidence interval) from a contingency table and interpret the estimate.

Risk Difference (RD)

- Risk difference computes the absolute difference in risk for the two groups (from the explanatory variable)

- Point estimate:

$$\widehat{RD} = \hat{p}_1 - \hat{p}_2 = \frac{n_{11}}{n_1} - \frac{n_{21}}{n_2}$$

- With range of point estimate from $[-1, 1]$

- Approximate standard error:

$$SE_{\widehat{RD}} = \sqrt{\frac{\hat{p}_1 \cdot (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 \cdot (1 - \hat{p}_2)}{n_2}}$$

- 95% Wald confidence interval for $\widehat{RD}$ (given expected cell counts > 5):

$$\widehat{RD} \pm 1.96 \cdot SE_{\widehat{RD}}$$

Explanatory Variable	Response Variable		Total
	Failure (0)	Success (1)	
Group 1	n_{10}	n_{11}	n_1
Group 2	n_{20}	n_{21}	n_2
Total	n_F	n_S	n

Recall the Strong Heart Study

The [Strong Heart Study](#) is an ongoing study of American Indians residing in 13 tribal communities in three geographic areas (AZ, OK, and SD/ND). We will look at data from this study examining the **incidence of diabetes** at a follow-up visit and **impaired glucose tolerance** (ITG) at baseline (4 years apart).

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664



SHS Example: Risk Difference (1/6)

Risk difference

Compute the point estimate and 95% confidence interval for the diabetes risk difference between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

Needed steps:

1. Check expected cell counts > 5
2. Compute the risk difference
3. Compute 95% confidence interval
4. Interpret the estimate

SHS Example: Risk Difference (2/6)

Risk difference

Compute the point estimate and 95% confidence interval for the diabetes risk difference between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

1. Check expected cell counts > 5

★ Create the table and look at it:

```
1 SHS_ct = table(SHS$glucimp, SHS$case)
2 SHS_ct
```

	0	1
Impaired	334	198
Normal	1004	128

Check the expected cell counts:

```
1 library(epitools)
2 expected(SHS_ct)
```

	0	1
Impaired	427.774	104.226
Normal	910.226	221.774

> 5

Normal approx
(CI's)

SHS Example: Risk Difference (3/6)

Risk difference

Compute the point estimate and 95% confidence interval for the diabetes risk difference between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

2. Compute the risk difference

$$\widehat{RD} = \hat{p}_1 - \hat{p}_2 = \frac{n_{11}}{n_1} - \frac{n_{21}}{n_2} = \frac{198}{532} - \frac{128}{1132} = 0.3722 - 0.1131 = \underline{0.2591}$$

SHS Example: Risk Difference (4/6)

Risk difference

Compute the point estimate and 95% confidence interval for the diabetes risk difference between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

3. Compute 95% confidence interval

$$\begin{aligned}& \widehat{RD} \pm z_{(1-\frac{\alpha}{2})}^* \times SE_{\widehat{RD}} \\&= \widehat{RD} \pm z_{(1-\frac{\alpha}{2})}^* \times \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\&= 0.2591 \pm 1.96 \times \sqrt{\frac{0.3722(1-0.3722)}{532} + \frac{0.1131(1-0.1131)}{1132}} \\&= (0.2141, 0.3041)\end{aligned}$$

SHS Example: Risk Difference (5/6)

Risk difference

Compute the point estimate and 95% confidence interval for the diabetes risk difference between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

2/3. Compute risk difference and 95% confidence interval

library(fmsb)

```
1 fmsb::riskdifference(198, 128, 532, 1132)
```

	Cases	People at risk	Risk
<u>Exposed</u>	198.0000000	532.0000000	0.3721805
<u>Unexposed</u>	128.0000000	1132.0000000	0.1130742
<u>Total</u>	326.0000000	1664.0000000	0.1959135

Risk difference and its significance probability (H_0 : The difference equals to zero)

data: 198 128 532 1132

p-value < 2.2e-16

95 percent confidence interval:

0.2140779 0.3041346

CI

sample estimates:

[1] 0.2591062

diff

SHS Example: Risk Difference (6/6)

Risk difference

Compute the point estimate and 95% confidence interval for the diabetes risk difference between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

4. Interpret the estimate

The diabetes diagnosis ~~risk~~ *prevalence* difference between impaired and normal glucose tolerance is 0.2591 (95% CI: 0.2141, 0.3041). Since the 95% confidence interval does not contain 0, we have sufficient evidence that the ~~risk~~ *prevalence* of diabetes diagnosis for people with impaired versus normal glucose tolerance is different.

When is the risk difference misleading?

- The same risk differences can have very different clinical meanings depending on the risk for each group

• **Example:** for two treatments A and B, we know the risk difference (RD) is 0.009. Is it a meaningful difference?

- If the risk is 0.01 for Trt A and 0.001 for Trt B?

$$0.01 - 0.001 = 0.009 \rightarrow \frac{0.01}{0.001} = 10$$

- If the risk is 0.41 for Trt A and 0.401 for Trt B?

$$0.41 - 0.401 = 0.009$$

- Using the RD alone to summarize the difference in risks for comparing the two groups can be **misleading**

- The **ratio of risk** can provide an informative descriptive measure of the “relative risk”

Learning Objectives

1. Understand the difference between testing for association and measuring association
2. Estimate the risk difference (and its confidence interval) from a contingency table and interpret the estimate.
3. Estimate the risk ratio (and its confidence interval) from a contingency table and interpret the estimate.
4. Estimate the odds ratio (and its confidence interval) from a contingency table and interpret the estimate.

Relative Risk (RR) Risk ratio OR prevalence ratio

- Relative risk computes the **ratio** of each group's proportions of "success"

- Also called **risk ratio**

- Point estimate:

$$\widehat{RR} = \frac{\hat{p}_1}{\hat{p}_2} = \frac{n_{11}/n_1}{n_{21}/n_2}$$

- Range: $[0, \infty]$

- if $\hat{p}_1 = 0$ then $\widehat{RR} = 0$ regardless of $\hat{p}_2$
- if $\hat{p}_2 = 0$ then $\widehat{RR} = \infty$ regardless of $\hat{p}_1$

Explanatory Variable	Response Variable		Total
	Failure (0)	Success (1)	
Group 1	n_{10}	n_{11}	n_1
Group 2	n_{20}	n_{21}	n_2
Total	n_F	n_S	n

What do values of risk ratios mean?

Risk Ratio	Clinical Meaning
$\widehat{RR} < 1$	Risk is smaller in group 1 than in group 2 $\hat{p}_1 < \hat{p}_2$
$\widehat{RR} = 1$	Risks are the same (explanatory and outcome are independent) $\hat{p}_1 = \hat{p}_2$
$\widehat{RR} > 1$	Risk is greater in group 1 than in group 2 $\hat{p}_1 > \hat{p}_2$

Poll Everywhere Question 2

14:26 Wed Apr 1

Join by Web PollEv.com/nickywakim275

If we use the previous example with two scenarios with the same risk difference. What is the relative risk for each scenario:

1. The risk is 0.01 for Trt A and 0.001 for Trt B
2. The risk is 0.41 for Trt A and 0.401 for Trt B.

1 = 10; 2 = 1.0

RR = 10, 1.02

$$\begin{aligned} \textcircled{1} \quad \widehat{RR} &= \frac{0.01}{0.001} \\ &= 10 \\ \textcircled{2} \quad \widehat{RR} &= \frac{0.41}{0.401} \\ &= 1.02 \\ &\quad \frac{1}{0.401} \\ &= 2.494 \end{aligned}$$

Log-transformation of RR

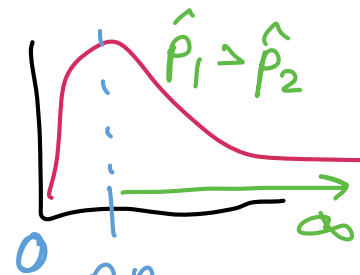
- Sampling distribution of the relative risk is **highly skewed** unless sample sizes are quite large
 - Log transformation** results in *approximately normal distribution*
 - Thus, **compute confidence interval** using **normally distributed, log-transformed RR**
 - Then we convert back to the RR
- We take the log (natural log) of RR: $\ln(\widehat{RR})$ or $\log(\widehat{RR})$
 - Whenever I say “log” I mean natural log (base e , very common in statistics)
- Then we need to find approximate standard error for $\ln(\widehat{RR})$

$$SE_{\ln(\widehat{RR})} = \sqrt{\frac{1}{n_{11}} - \frac{1}{n_1} + \frac{1}{n_{21}} - \frac{1}{n_2}}$$

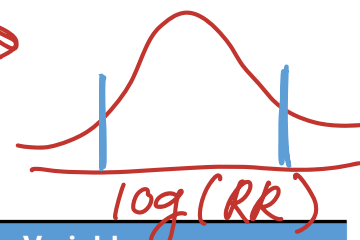
- 95% confidence interval for $\ln(\widehat{RR})$:

$$\ln(\widehat{RR}) \pm 1.96 \times SE_{\ln(\widehat{RR})}$$

$[0, \infty)$



$RR=1$
 $p_1 = p_2$



Explanatory Variable	Response Variable		Total
	Failure (0)	Success (1)	
Group 1	n_{10}	n_{11}	n_1
Group 2	n_{20}	n_{21}	n_2
Total	n_F	n_S	n

How do we get back to the RR scale?

- We **computed confidence interval** using normally distributed, log-transformed RR ($\ln(\widehat{RR})$):

$$\left(\underbrace{\ln(\widehat{RR}) - 1.96 \times SE_{\ln(\widehat{RR})}}_{\text{LCI}}, \underbrace{\ln(\widehat{RR}) + 1.96 \times SE_{\ln(\widehat{RR})}}_{\text{UCI}} \right)$$

- Now we need to exponentiate the CI to get back to interpretable values

- Take exponential of lower and upper bounds

- 95% confidence interval for RR: two ways to display equation

$$\left(e^{\ln(\widehat{RR}) - 1.96 \times SE_{\ln(\widehat{RR})}}, e^{\ln(\widehat{RR}) + 1.96 \times SE_{\ln(\widehat{RR})}} \right)$$

$$\left(\exp \left[\ln(\widehat{RR}) - 1.96 \times SE_{\ln(\widehat{RR})} \right], \exp \left[\ln(\widehat{RR}) + 1.96 \times SE_{\ln(\widehat{RR})} \right] \right)$$

ln RR CI:

 $(1, 3)$

RR CI:

$$(e^1, e^3)$$

Relative Risk (RR)

- Can you compute the estimated RRs for the previous example?

- If the risk for Trt A is 0.01 and Trt B is 0.001? $\widehat{RR} = 10$
- If the risk for Trt A is 0.41 and Trt B is 0.401? $\widehat{RR} = 1.02$

- When $\widehat{RR} = 1$...

- Risk is the same for the two groups
- In other words, the group and the outcome are independent

- When computing $\widehat{RR}$ it is important to identify which variable is the response variable and which is explanatory variable

- We may say "risk for Trt A" but this translates to the risk (or probability) of outcome success for those receiving Trt A

→ risk for TRT A $\neq$ risk of TRT A

SHS Example: Relative Risk (1/7)

Relative risk

Compute the point estimate and 95% confidence interval for the diabetes Relative risk between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

Needed steps:

1. Check expected cell counts > 5]
2. Compute the relative risk
3. Find confidence interval of log RR
4. Convert back to RR
5. Interpret the estimate

SHS Example: Relative Risk (2/7)

Relative risk

Compute the point estimate and 95% confidence interval for the diabetes Relative risk between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

1. Check expected cell counts > 5

```
1 library(epitools)
2 expected(SHS_ct)
```

```
      0      1
Impaired 427.774 104.226
Normal   910.226 221.774
```

SHS Example: Relative Risk (3/7)

Relative risk

Compute the point estimate and 95% confidence interval for the diabetes Relative risk between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

1. Compute the relative risk

$$\widehat{RR} = \frac{\hat{p}_1}{\hat{p}_2} = \frac{n_{11}/n_1}{n_{21}/n_2} = \frac{198/532}{128/1132} = \frac{0.3722}{0.1131} = 3.2915$$

Risk for impaired

Risk for normal (of diabetes)

SHS Example: Relative Risk (4/7)

Relative risk

Compute the point estimate and 95% confidence interval for the diabetes Relative risk between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

2. Find confidence interval of log RR

$$\begin{aligned} & \ln(\widehat{RR}) \pm 1.96 \times SE_{\ln(\widehat{RR})} \\ &= \ln(\widehat{RR}) \pm z_{(1-\frac{\alpha}{2})}^* \times \sqrt{\frac{1}{n_{11}} - \frac{1}{n_1} + \frac{1}{n_{21}} - \frac{1}{n_2}} \\ &= 1.1913 \pm 1.96 \times \sqrt{\frac{1}{198} - \frac{1}{532} + \frac{1}{128} - \frac{1}{1132}} \\ &= (0.9944, 1.3883) \end{aligned}$$

SHS Example: Relative Risk (5/7)

Relative risk

Compute the point estimate and 95% confidence interval for the diabetes Relative risk between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

3. Convert back to RR

$$(\exp(0.9944), \exp(1.3883)) \\ = (2.703, 4.0081)$$

SHS Example: Relative Risk (6/7)

Relative risk

Compute the point estimate and 95% confidence interval for the diabetes Relative risk between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

2/3/4. Compute risk ratio and 95% confidence interval

Create the table and look at it:

```
1 SHS_ct = table(SHS$glucimp, SHS$case)
2 SHS_ct
```

Calculate the risk ratio:

```
1 library(epitools)
2 riskratio(x = SHS_ct, rev = "rows")$measure
```

	0	1	
2 Impaired	334	198	denominator ($\hat{p}_2$)
1 Normal	1004	128	numerator ($\hat{p}_1$)

$$\frac{\hat{p}_{imp}}{\hat{p}_{norm}}$$

risk ratio with 95% C.I.			
	estimate	lower	upper
Normal	1.000000	NA	NA
Impaired	3.291471	2.702998	4.008061

SHS Example: Relative Risk (7/7)

Relative risk

Compute the point estimate and 95% confidence interval for the diabetes Relative risk between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

3. Interpret the estimate

The estimated risk of diabetes is 3.29 times greater for American Indians who had impaired glucose tolerance at baseline compared to those who had normal glucose tolerance (95% CI: 2.70, 4.01).

Additional interpretation of 95% CI (not needed): We are 95% confident that the (population) relative risk is between 2.70 and 4.01.

Since the 95% confidence interval does not include 1, there is sufficient evidence that the risk of diabetes differs significantly between impaired and normal glucose tolerance at baseline.

Learning Objectives

1. Understand the difference between testing for association and measuring association
2. Estimate the risk difference (and its confidence interval) from a contingency table and interpret the estimate.
3. Estimate the risk ratio (and its confidence interval) from a contingency table and interpret the estimate.
4. Estimate the odds ratio (and its confidence interval) from a contingency table and interpret the estimate.

$$\hat{p}_1 - \hat{p}_2$$

$\hat{p}_1$: risk of success
in outcome
for grp 1

$\hat{p}_2$: risk of success
for grp 2

$$\frac{\hat{p}_1}{\hat{p}_2}$$

Odds (building up to Odds Ratio)

★ odds is ratio of prop of succ to prop of failure

- For a probability of success p (or sometimes referred to as π), the odds of success is:

$$\widehat{\text{odds}} = \frac{\hat{p}}{1 - \hat{p}} = \frac{\hat{\pi}}{1 - \hat{\pi}}$$

prop of success / prop of failure

Explanatory Variable	Response Variable		Total
	Failure (0)	Success (1)	
Group 1	n_{10}	n_{11}	n_1
Group 2	n_{20}	n_{21}	n_2
Total	n_F	n_S	n

- Example: if $\hat{\pi} = 0.75$, then odds of success = $\frac{0.75}{0.25} = 3$

prop of success / (1 - 0.75)
- If odds > 1, it implies a success is more likely than a failure

- Example: for $\widehat{\text{odds}} = 3$, we expect to observe three times as many successes as failures
- If odds is known, the probability of success can be computed

$$\hat{\pi} = \frac{\widehat{\text{odds}}}{\widehat{\text{odds}} + 1}$$

odds = 0.333
 failure is 3x more likely than success

Odds Ratio (OR)

- Odds ratio is the ratio of two odds:

$$\widehat{OR} = \frac{\widehat{odds}_1}{\widehat{odds}_2} = \frac{\overset{\text{prop succ}}{\hat{p}_1} / \overset{\text{prop fail}}{(1 - \hat{p}_1)}}{\underset{\text{prop succ}}{\hat{p}_2} / \underset{\text{prop fail}}{(1 - \hat{p}_2)}} \begin{matrix} > \text{grp 1} \\ > \text{grp 2} \end{matrix}$$

Explanatory Variable	Response Variable		Total
	Failure (0)	Success (1)	
Group 1	n_{10}	n_{11}	n_1
Group 2	n_{20}	n_{21}	n_2
Total	n_F	n_S	n

- Range: $[0, \infty]$

- Interpretation: The odds of success for "group 1" is $\widehat{OR}$ times the odds of success for "group 2"
- value

- What do values of odds ratios mean?

Odds Ratio

Clinical Meaning

$$\widehat{OR} < 1$$

$$[0, 1)$$

Odds of success is smaller in group 1 than in group 2

$$\widehat{OR} = 1$$

Explanatory and response variables are independent

$$\widehat{OR} > 1$$

$$(1, \infty)$$

Odds of success is greater in group 1 than in group 2

odds equal for grp 1 & grp 2

Poll Everywhere Question 3

13:19 Mon Apr 6

75%



Join by Web PollEv.com/nickywakim275



Given $\hat{p}_2 = 0.8$, what is the range of possible values for our relative risk and our odds ratio?

RR: $[0, \infty)$, OR: $[0, \infty)$ 30%

RR: $[0, 1.25]$, OR: $[0, \infty)$ 27%

RR: $[0, \infty)$, OR: $[0, 1.25]$ 30%

RR: $[0, 1.25]$, OR: $[0, 1.25]$ 13%

$\hat{OR} \in [0, \infty)$

$$\hat{RR} = \frac{\hat{p}_1}{\hat{p}_2} = \frac{\hat{p}_1}{0.8}$$

$$0 \leq \hat{p}_1 \leq 1$$

$$\hat{p}_1 = 0 \Rightarrow \hat{RR} = \frac{0}{0.8} = 0$$

$$\hat{p}_1 = 1 \Rightarrow \hat{RR} = \frac{1}{0.8} = 1.25$$

$$\hat{OR} = \frac{\hat{p}_1}{1 - \hat{p}_1} = \frac{\hat{p}_1}{1 - \hat{p}_1}$$

odds for grp 2 is 4

$$\hat{p}_1 = 0 \Rightarrow \hat{OR} = \frac{0}{1-0} = 0$$

$$\hat{p}_1 = 1 \Rightarrow \hat{OR} = \frac{1}{0/4} = \infty$$

Odds Ratio (OR)

- Values of $\hat{OR}$ farther from 1.0 in a given direction represent stronger association
 - An $OR = 4$ is farther from independence than an $OR = 2$
 - An $OR = 0.25$ is farther from independence than an $OR = 0.5$
 - For $OR = 4$ and $OR = 0.25$, they are equally away from independence (because $\frac{1}{4} = 0.25$)

odds of succ.
 $OR = \frac{4 \text{ odds}_1}{1 \text{ odds}_2}$

- We take the inverse of the OR for **success** of **group 1** compared to **group 2** to get...

- OR for **failure** of **group 1** compared to **group 2**
 - OR for **success** of **group 2** compared to **group 1**
- inverse of*

$OR = \frac{1 \text{ odds}_2}{4 \text{ odds}_1}$
odds of fail

Log-transformation of OR

- Like RR, sampling distribution of the odds ratio is highly skewed
 - Log transformation results in approximately normal distribution
 - Thus, compute confidence interval using normally distributed, log-transformed OR
- Approximate standard error for $\ln(\widehat{OR})$:

$$SE_{\ln(\widehat{OR})} = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{10}} + \frac{1}{n_{21}} + \frac{1}{n_{20}}}$$

- 95% confidence interval for $\ln(\widehat{OR})$:

$$\ln(\widehat{OR}) \pm 1.96 \times SE_{\ln(\widehat{OR})}$$

↓

$$Z^*_{(1-\alpha/2)}$$

$\alpha = 0.05$

Explanatory Variable	Response Variable		Total
	Failure (0)	Success (1)	
Group 1	n_{10}	n_{11}	n_1
Group 2	n_{20}	n_{21}	n_2
Total	n_F	n_S	n

How do we get back to the OR scale?

- We computed confidence interval using normally distributed, log-transformed OR ($\ln(\widehat{OR})$):

$$\left(\overset{LCI}{\ln(\widehat{OR}) - 1.96 \times SE_{\ln(\widehat{OR})}}, \overset{UCI}{\ln(\widehat{OR}) + 1.96 \times SE_{\ln(\widehat{OR})}} \right) \text{ of log trans scale}$$

- Now we need to **exponentiate the CI to get back to interpretable values**
 - Take exponential of lower and upper bounds
- 95% confidence interval for RR: two ways to display equation

$$\left(e^{\ln(\widehat{OR}) - 1.96 \times SE_{\ln(\widehat{OR})}}, e^{\ln(\widehat{OR}) + 1.96 \times SE_{\ln(\widehat{OR})}} \right)$$

$$\left(\exp \left[\ln(\widehat{OR}) - 1.96 \times SE_{\ln(\widehat{OR})} \right], \exp \left[\ln(\widehat{OR}) + 1.96 \times SE_{\ln(\widehat{OR})} \right] \right)$$

SHS Example: Odds Ratio (1/7)

Odds ratio

Compute the point estimate and 95% confidence interval for the diabetes odds ratio between impaired and normal glucose tolerance.

↳ of diabetes

Needed steps:

1. Check expected cell counts > 5
2. Compute the odds ratio ✓
3. Find confidence interval of log OR ✓
4. Convert back to OR ✓
5. Interpret the estimate ✓

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

SHS Example: Odds Ratio (2/7)

Odds ratio

Compute the point estimate and 95% confidence interval for the diabetes Odds ratio between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

1. Check expected cell counts > 5

```
1 library(epitools) ✓  
2 expected(SHS_ct)
```

	0	1
Impaired	427.774	104.226
Normal	<u>910.226</u>	<u>221.774</u>

> 5 ✓

SHS Example: Odds Ratio (3/7)

Odds ratio

Compute the point estimate and 95% confidence interval for the diabetes Odds ratio between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

2. Compute the odds ratio

$$\hat{p}_1 = 198/532 = 0.3722, \hat{p}_2 = 128/1132 = 0.1131$$

$$\widehat{OR} = \frac{\hat{p}_1 / (1 - \hat{p}_1)}{\hat{p}_2 / (1 - \hat{p}_2)} = \frac{0.3722 / (1 - 0.3722)}{0.1131 / (1 - 0.1131)} = 4.6499$$

SHS Example: Odds Ratio (4/7)

Odds ratio

Compute the point estimate and 95% confidence interval for the diabetes Odds ratio between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

3. Find confidence interval of log OR

$$\begin{aligned} & \ln(\widehat{OR}) \pm 1.96 \times SE_{\ln(\widehat{OR})} \\ & = \ln(\widehat{OR}) \pm z_{(1-\frac{\alpha}{2})}^* \times \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \\ & = 1.5368 \pm 1.96 \times \sqrt{\frac{1}{198} + \frac{1}{334} + \frac{1}{128} + \frac{1}{1004}} \\ & = (1.2824, 1.7913) \end{aligned}$$

log(4.65) (pointing to 1.5368)
on log scale (pointing to the interval)

SHS Example: Odds Ratio (5/7)

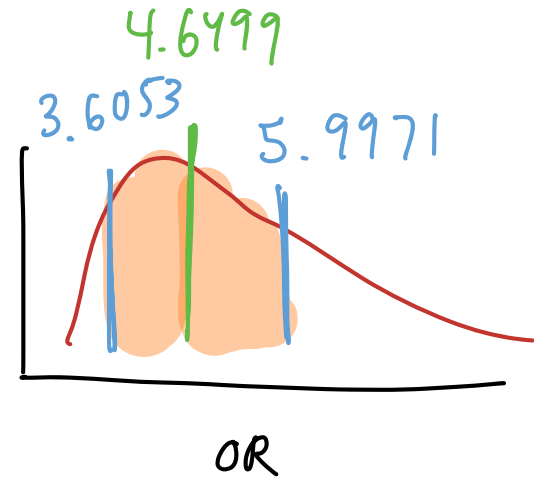
Odds ratio

Compute the point estimate and 95% confidence interval for the diabetes Odds ratio between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

4. Convert back to OR

$$\begin{aligned} & (\exp(1.2824), \exp(1.7913)) \\ &= (3.6053, 5.9971) \\ & \quad \downarrow \\ & 4.6499 \end{aligned}$$



SHS Example: Odds Ratio (6/7)

Odds ratio

Compute the point estimate and 95% confidence interval for the diabetes Odds ratio between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

2/3/4. Compute OR and 95% confidence interval

```
1 library(epitools)
2 SHS_ct = table(SHS$glucimp, SHS$case)
3 # no `rev` needed below bc we set the reference level in slide 32
4 oddsratio(x = SHS_ct, method = "wald")$measure
```

odds ratio with 95% C.I.

	estimate	lower	upper
Impaired	1.000000	NA	NA
Normal	0.215059	0.1667459	0.2773702

↪ rev = "rows"

OR comparing Normal to impaired
should be inverse of impaired to normal

0.215059
 $\frac{1}{0.215059} = 4.6499$

SHS Example: Odds Ratio (7/7)

Odds ratio

Compute the point estimate and 95% confidence interval for the diabetes Odds ratio between impaired and normal glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

3. Interpret the estimate

The estimated odds of diabetes for American Indians with impaired glucose tolerance at baseline is 4.65 times the odds for American Indians with normal glucose tolerance at baseline. (95% CI: 3.61, 6.00)

Additional interpretation of 95% CI (not needed): We are 95% confident that the odds ratio is between 3.61 and 6.00.

Since the 95% confidence interval does not include 1, there is sufficient evidence that the odds of diabetes differs significantly between impaired and normal glucose tolerance at baseline.

@ 1, odds are equal

Inverting an Odds Ratio

- Some people prefer interpretations of $OR > 1$ instead of an $OR < 1$
- The transformation can easily be done by inverse
 - Remember we discussed that $OR = 4$ is an equivalent a strong association as $OR = 0.25$ ($1/4$)
- OR comparing group 1 to group 2 = inverse of OR comparing group 2 to group 1

$$\begin{aligned}
 \text{odds}_1 & \\
 \underline{OR_{1v2}} &= \frac{\hat{p}_1 / (1 - \hat{p}_1)}{\hat{p}_2 / (1 - \hat{p}_2)} \\
 \text{odds}_2 & \\
 \frac{1}{\underline{OR_{1v2}}} &= \frac{1}{\frac{\hat{p}_1 / (1 - \hat{p}_1)}{\hat{p}_2 / (1 - \hat{p}_2)}} = \frac{\hat{p}_2 / (1 - \hat{p}_2)}{\hat{p}_1 / (1 - \hat{p}_1)} = OR_{2v1} \\
 & \quad \text{odds}_2 \quad \text{odds}_1
 \end{aligned}$$

Poll Everywhere Question 4

13:52 Mon Apr 6

64%

Join by Web PollEv.com/nickywakim275



Given the estimated odds ratio (4.65) that we just calculated in our example, select the following statements that are true.

The odds of diabetes for American Indians with impaired glucose tolerance is 4.65 times the odds for American Indians with normal glucose tolerance. ✓ 42%

The odds of diabetes for American Indians with normal glucose tolerance is 0.22 times the odds for American Indians with impaired glucose tolerance. ✓ 28%

The odds of diabetes for American Indians with normal glucose tolerance is 0.33 times the odds for American Indians with impaired glucose tolerance. 1%

Diabetes diagnosis is less likely for those impaired glucose tolerance than those with normal glucose tolerance. 0%

Diabetes diagnosis is less likely for those normal glucose tolerance than those with impaired glucose tolerance. ✓ 28%

4.65 odd ratio
of diabetes
comparing
impaired gluc
to normal gluc

$$\text{odds}_I > \text{odds}_N$$
$$\frac{\hat{p}_1}{1-\hat{p}_1} > \frac{\hat{p}_2}{1-\hat{p}_2}$$

SHS Example: Inversing Odds Ratio

Inversing odds ratio

Compute the point estimate and 95% confidence interval for the diabetes odds ratio between **normal** and **impaired** glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

Needed steps:

1. Inverse point estimate and confidence interval

$$\widehat{OR} = \frac{1}{4.6499} = 0.2151$$

The 95% Confidence interval is then

$$\left(\frac{1}{5.9971}, \frac{1}{3.6053} \right) = (0.1667, 0.2774)$$

SHS Example: Inversing Odds Ratio

Inversing odds ratio

Compute the point estimate and 95% confidence interval for the diabetes odds ratio between **normal** and **impaired** glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

Needed steps:

1. Inverse point estimate and confidence interval

```
1 library(epitools)
2 oddsratio(x = SHS_ct, method = "wald", rev = "rows")$measure
```

odds ratio with 95% C.I.

	estimate	lower	upper
<u>Normal</u>	1.000000	NA	NA
<u>Impaired</u>	4.649888	3.605289	5.997148

0.215059

SHS Example: Inversing Odds Ratio

Inversing odds ratio

Compute the point estimate and 95% confidence interval for the diabetes odds ratio between **normal** and **impaired** glucose tolerance.

Glucose tolerance	Diabetes		Total
	No	Yes	
Impaired	334	198	532
Normal	1004	128	1132
Total	1338	326	1664

Needed steps:

2. Interpret the estimate

The estimated odds of diabetes for American Indians with normal glucose tolerance at baseline is 0.22 times the odds for American Indians with impaired glucose tolerance at baseline.

Additional interpretation of 95% CI (not needed): We are 95% confident that the odds ratio is between 0.17 and 0.28.

Since the 95% confidence interval does not include 1, there is sufficient evidence that the odds of diabetes differs significantly between impaired and normal glucose tolerance at baseline.

Learning Objectives

1. Understand the difference between testing for association and measuring association
2. Estimate the risk difference (and its confidence interval) from a contingency table and interpret the estimate.
3. Estimate the risk ratio (and its confidence interval) from a contingency table and interpret the estimate.
4. Estimate the odds ratio (and its confidence interval) from a contingency table and interpret the estimate.