

# Measuring association in contingency tables

Patrick Breheny

April 4

## Hypothesis tests and confidence intervals

- Fisher's exact test and the  $\chi^2$ -test can be used to calculate  $p$ -values and assess evidence against the null
- This null hypothesis can be loosely described as saying that there is “no association” between treatment/exposure and the outcome, or more formally as the hypothesis that the two events are independent
- However, we also need to be able to measure and place confidence intervals on effect sizes – otherwise, we have no way of assessing the practical and clinical significance of the association
- So we want a confidence interval . . . but what do we want a confidence interval of?

## Difference in proportions

- One way of measuring the strength of an association for categorical data is to look at the difference in proportions
- For example, in Lister's experiment, 46% of the patients who received the conventional surgery died, but only 15% of the patients who received the sterile surgery died
- The difference in these percentages is 31%

## Flaws with the difference in proportions

- However, differences in proportions are not informative for rare events
- For example, in a well-known study that made front-page headlines in the *New York Times*, 0.9% of subjects taking aspirin suffered heart attacks, compared to 1.7% of placebo subjects
- The difference in proportions, 0.8%, doesn't sound front-page-of-the-*New-York-Times*-worthy

## The relative risk

- Instead, for proportions, we often describe the strength of an association using ratios
- If we summarize the study's findings by saying that the probability of suffering a heart attack was twice as large ( $1.7/0.9 = 1.9$ ) for the placebo group as for the aspirin group, this is much more attention-grabbing
- Similarly for Lister's experiment: the risk of dying from surgery is three times lower ( $46/15 = 3.1$ ) if sterile technique is used
- This ratio is called the *relative risk*, and it is often more informative than the difference of proportions, especially for rare events

## Flaws with the relative risk

- The relative risk is a good measure of the strength of an association, but it too has flaws
- One is that it's asymmetric
- For example, the relative risk of dying is  $46/15 = 3.1$  times greater with the nonsterile surgery, but the relative risk of living is only  $85/54 = 1.57$  times greater with the sterile surgery

## Relative risks and retrospective studies

- Another flaw is that it doesn't work with retrospective studies
- For example, consider the results of a classic case-control study of the relationship between smoking and lung cancer published in 1950:

	Cases	Controls
Smoker	688	650
Nonsmoker	21	59

- Is the probability of developing lung cancer given that a person smoked  $688/(688 + 650) = 51\%$ ?

## Relative risks and retrospective studies (cont'd)

- Absolutely not; this isn't even remotely accurate
- By design, this study included 709 people with lung cancer and 709 without; the fact that about 50% of smokers had lung cancer doesn't mean anything
- For retrospective and cross-sectional studies, then, we cannot calculate a relative risk
- This would require an estimate of the probability of developing a disease given that an individual was exposed to a risk factor, which we can only get from a prospective study
- Instead, retrospective studies give us the probability of being exposed to a risk factor given that you have developed the disease



# Odds

- A slightly different measure of association, the *odds ratio*, gets around both of these flaws
- Instead of taking the ratio of the probabilities, the odds ratio is a ratio of the odds of developing the disease given risk factor exposure to the odds given a lack of exposure
- The *odds* of an event is the ratio of the number of times the event occurs to the number of times the event fails to occur
  - For example, if the probability of an event is 50%, then the odds are 1; in speech, people usually say that “the odds are 1 to 1”
  - If the probability of an event is 75%, then the odds are 3; “the odds are 3 to 1”

# The symmetry of the odds ratio

- As advertised, the odds ratio possesses the symmetry that the relative risk does not
- For example, in Lister's experiment the odds of dying were  $6/34 = .176$  for the sterile group and  $16/19 = .842$  for the control group
- The relative odds of dying with the control surgery is therefore  $.842/.176 = 4.77$
- On the other hand, the odds of surviving were  $34/6 = 5.67$  for the sterile group and  $19/16 = 1.19$  for the control group
- The relative odds of surviving with the sterile surgery is therefore  $5.67/1.19 = 4.77$

## An easier formula for the odds ratio

- Summarizing this reasoning into a formula, if our table looks like

a	b
c	d

then

$$\widehat{\text{OR}} = \frac{ad}{bc}$$

- Because of this formula, the odds ratio was originally called the “cross-product ratio”

## There are two odds ratios

- Keep in mind that there are two odds ratios, depending on how we ordered the rows and columns of the table, and that they will be reciprocals of one another
- When calculating and interpreting odds ratios, be sure you know which group has the higher odds of developing the disease
- In Lister's experiment, the odds ratio for surviving with the sterile surgery was 4.77, but the odds ratio for surviving with the control surgery was  $1/4.77 = 0.210$
- NOTE: When writing about an odds ratio less than 1, it is customary to write, for example, that "the sterile procedure reduced the odds of death by 79%"

## Odds ratios and retrospective studies

- The symmetry of the odds ratio works wonders when it comes to retrospective studies
- So, in our case-control study of lung cancer and smoking, the odds ratio for smoking given lung cancer is

$$\begin{aligned}\widehat{\text{OR}} &= \frac{688 \cdot 59}{21 \cdot 650} \\ &= 2.97\end{aligned}$$

- However, this is also the odds ratio for lung cancer given smoking
- This is a minor miracle: we have managed to obtain a prospective measure of association from a retrospective study!
- Hence the popularity of the odds ratio: it can be used for any study design (prospective, retrospective, cross-sectional) that results in a 2x2 contingency table

## Interpretation of odds ratios

- We will focus on odds ratios in this course, although it is worth noting that many researchers consider relative risks easier to interpret, and prefer reporting them when possible
- Indeed, odds ratios are always larger in magnitude (i.e., further away from 1) than relative risks, something to keep in mind when interpreting clinical significance
- For example, we saw that for the Lister study, the relative risk was 3.1, while the odds ratio was 4.8
- An even more extreme example is the Nexium trial (recall that the healing rates in the two groups were 93% vs. 89%), where the relative risk is 1.04 but the odds ratio is 1.6

## The $\chi^2$ -test and measures of association

- Note that when the difference between two proportions equals 0, the relative risk equals 1 and the odds ratio equals 1
- Furthermore, when relative risk of disease given exposure equals 1, the relative risk of exposure given disease equals 1
- Indeed, all of these statements are equivalent to saying that exposure and disease are independent
- Thus, any one of these may be thought of as the null hypothesis of the  $\chi^2$ -test
- This is why the null hypothesis is often loosely described as there being no “association” between exposure and disease

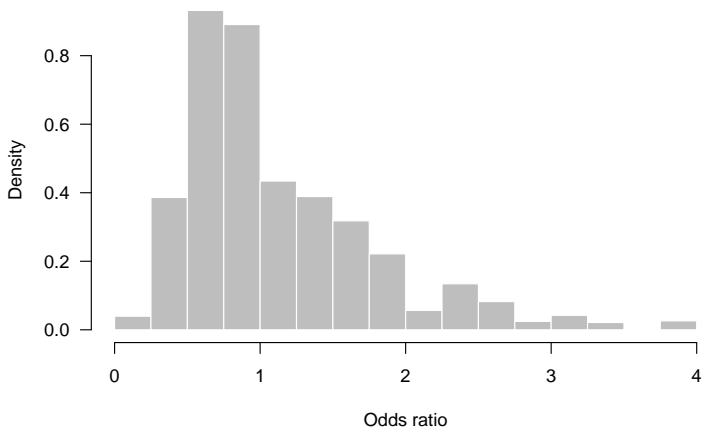
## The odds ratio and the central limit theorem

- So far in this class, we've calculated confidence intervals for averages and percentages
- For both statistics, the central limit theorem guarantees that if the sample size is big enough, their sampling distribution will look relatively normal
- The central limit theorem, however, does not directly establish that the distribution of odds ratio will be normally distributed, and indeed, it tends to be skewed to the right



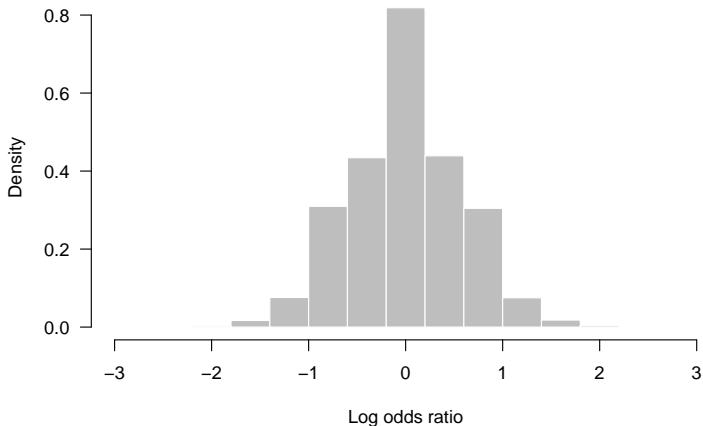
# Simulation: The sampling distribution of the odds ratio

$n = 25$  per group



# The log transform

However, look at the sampling distribution of the logarithm of the odds ratio:



## Confidence intervals for the log odds ratio

- The log of the odds ratio is quite normal-looking and amenable to finding confidence intervals for using the central limit theorem/normal distribution
- Thus, the procedure that has been developed for constructing approximate confidence intervals for the odds ratio actually constructs confidence intervals for the log of the odds ratio
- Getting a confidence interval for the odds ratio itself then requires an extra step of converting the confidence interval back to the odds ratio scale

# The standard error of the log odds ratio

It can be shown that a good estimate of the standard error of the log of the odds ratio is

$$SE_{\log OR} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

where  $a$ ,  $b$ ,  $c$ , and  $d$  are the four entries in the contingency table

## Confidence intervals for the odds ratio: procedure

The first three steps for constructing confidence intervals for the odds ratio should look familiar; the fourth will be new:

#1 Estimate the standard error of the log OR,

$$SE_{\log OR} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

#2 Determine the values that contain the middle  $x\%$  of the normal curve,  $\pm z_{x\%}$

#3 Calculate the confidence interval for the log OR:

$$(L, U) = \left( \log(\widehat{OR}) - z_{x\%} \cdot SE_{\log OR}, \log(\widehat{OR}) + z_{x\%} \cdot SE_{\log OR} \right)$$

#4 Convert the confidence interval from step 3 back into the odds ratio scale to obtain the confidence interval for the odds ratio:

$$(e^L, e^U)$$

NOTE: “log” above refers to the natural log (sometimes abbreviated “ln”), not the base-10 logarithm

## Confidence intervals for the odds ratio: example

To see how this works in practice, let's calculate a 95% confidence interval for the odds ratio of surviving with sterile surgery in Lister's experiment

#1 Estimate the standard error of the log odds ratio:

$$\begin{aligned} SE_{\log OR} &= \sqrt{\frac{1}{34} + \frac{1}{6} + \frac{1}{19} + \frac{1}{16}} \\ &= 0.56 \end{aligned}$$

#2 As usual, 1.96 contains the middle 95% of the normal distribution

#3 Recall that the sample odds ratio was 4.77, so the log of the sample odds ratio is 1.56; the 95% confidence interval for the log odds ratio is therefore

$$(1.56 - 1.96(0.56), 1.56 + 1.96(0.56)) = (0.47, 2.66)$$

## Confidence intervals for the odds ratio: example (cont'd)

#4 The 95% confidence interval for the odds ratio is therefore

$$(e^{0.47}, e^{2.66}) = (1.60, 14.2)$$

- Note that the confidence interval doesn't include 1; this agrees with our test of significance
- Note also that this confidence interval is asymmetric (its right half is much longer than its left half) – this would be impossible to achieve without the log transform
- Now we have an idea of the possible clinical significance of sterile technique: it may be lowering the odds of surgical death by a factor of about 1.6, or by a factor of 14, with a factor of around 5 being the most likely

## Reverse example

Note also the symmetry that arises if we had decided to calculate a confidence interval for the other odds ratio (the relative odds of dying on the sterile surgery):

- $\widehat{OR} = 0.21$
- $\log(\widehat{OR}) = -1.56$
- 95% CI for  $\log(\widehat{OR})$ : (-2.66, -0.47)
- 95% CI for  $\widehat{OR}$ : (0.07, 0.62)



## Supplemental: CIs for RRs

- The same procedure is used to construct confidence intervals for relative risks; the only difference is that

$$SE_{IRR} = \sqrt{\frac{b}{a(a+b)} + \frac{c}{c(c+d)}}$$

- So, in the Lister example, the 95% CI for the relative risk is [1.3, 6.9] (risk of death in control vs. sterile surgery)
- This slide is just an extra, for-your-information slide; for the purposes of quizzes/tests, I will only ask about CIs for odds ratios

# Summary

- There are three natural ways to measure the association present in a  $2 \times 2$  table:
  - Difference of proportions
  - Relative risk
  - Odds ratio
- One big advantage of the odds ratio is that it works equally well for both prospective and retrospective studies, unlike the other two
- Be able to construct confidence intervals for odds ratios