Lab 14: Inference for relative risk and odds ratios

STAT218

This activity focuses on inference of relative risk and odds ratios. There are three objectives:

1. Learn to perform tests and compute interval estimates for relative risk
2. Learn to perform tests and compute interval estimates for odds ratios
3. Practice discerning which estimand is appropriate given study design

Note that *you cannot determine* the study design from inspecting the data.

## Examples

### Relative risk

Researchers recorded deaths due to cardiovascular disease (CVD) among a random sample of 3,112 individuals in American Samoa, where obesity was considered socially desirable, in order to investigate a potential link between obesity and CVD mortality.

Deaths NonDeaths  
Obese 16 2045  
NotObese 7 1044

Because the data are a random sample from the population of interest (Samoans), it is possible to estimate the probability of the outcome of interest.

* the outcome of interest is death due to CVD
* the groups are obese and non-obese individuals
* this is a prospective-type observational study
* relative risk is a more appropriate target for inference

Therefore we’ll perform inference on:

In R, this is a one-line command:

riskratio(obesity, rev = 'both')

$data  
 NonDeaths Deaths Total  
NotObese 1044 7 1051  
Obese 2045 16 2061  
Total 3089 23 3112  
  
$measure  
 NA  
risk ratio with 95% C.I. estimate lower upper  
 NotObese 1.000000 NA NA  
 Obese 1.165592 0.4810219 2.824415  
  
$p.value  
 NA  
two-sided midp.exact fisher.exact chi.square  
 NotObese NA NA NA  
 Obese 0.7561834 0.8277447 0.7340665  
  
$correction  
[1] FALSE  
  
attr(,"method")  
[1] "Unconditional MLE & normal approximation (Wald) CI"

The riskratio function has a specific formatting requirement:

* rows should be ordered so that the reference/control group appears first
* columns should be ordered so that the outcome of interes appears second

The rev = ... argument *reverses* the order of rows, columns, or both.

* rev = 'neither' (default) indicates order of rows and columns is correctly specified
* rev = 'rows' reverses the order of rows only
* rev = 'columns' reverses the order of columns only
* rev = 'both' reverses the order of rows and columns

In the example, rev = 'both' is used because the contingency table shows non-obese individuals (reference group) in the second row and deaths (outcome of interest) in the first column. Both should be reversed.

### Odds ratio

Researchers sampled 86 lung cancer patients (cases) and 86 healthy individuals (controls) and recorded smoking status.

Smokers NonSmokers  
Cancer 83 3  
Control 72 14

This is a case-control study and inherently retrospective, because the outcome of interest is developing lung cancer, but individuals are sampled based on outcome and the risk factor (smoking) is identified retrospectively.

* the outcome of interest is lung cancer
* the groupings are smokers and nonsmokers
* the study design is retrospective
* odds ratio is the more appropriate target for inference

In this case it is not possible to estimate the probability of the outcome of interest, so inference should focus on the odds ratio:

This, as well, is a one-line command in R:

$data  
 NonSmokers Smokers Total  
Control 14 72 86  
Cancer 3 83 86  
Total 17 155 172  
  
$measure  
 NA  
odds ratio with 95% C.I. estimate lower upper  
 Control 1.00000 NA NA  
 Cancer 5.37963 1.486376 19.47045  
  
$p.value  
 NA  
two-sided midp.exact fisher.exact chi.square  
 Control NA NA NA  
 Cancer 0.005116319 0.008822805 0.004948149  
  
$correction  
[1] FALSE  
  
attr(,"method")  
[1] "Unconditional MLE & normal approximation (Wald) CI"

The oddsratio function expects the same input format: the first row is the reference/control group, and the second column is the outcome of interest.

## Practice problems

1. An outbreak of cyclosporiasis was detected among residents of New Jersey. In a case-control study, investigators found that 21 of 30 case-patients and four of 60 controls had eaten raspberries.
   1. Identify the outcome of interest and groupings.
   2. Identify the study design.
   3. Based on the study design, is the odds ratio or relative risk a more appropriate inferential target?
   4. Perform inference according to your answer in (c); carry out a test and compute a confidence interval.

cyclosporiasis  
raspberries yes no  
 yes 21 4  
 no 9 56

1. Researchers conducted a prospective cohort study of 3,000 smokers and 5,000 nonsmokers to investigate the relationship between smoking and the development of coronary heart disease (CHD) over 1 year.
   1. Identify the outcome of interest and groupings.
   2. Identify the study design.
   3. Based on the study design, is the odds ratio or relative risk a more appropriate inferential target?
   4. Perform inference according to your answer in (c); carry out a test and compute a confidence interval.

chd  
smoker yes no  
 yes 84 2916  
 no 87 4913

1. Researchers gave 48 male bank supervisors attending a management institute hypothetical personnel files and asked them whether they would promote the applicant based on the file. The personnel files were identical except that 24 of them listed a male and 24 listed a female applicant. The assignment of managers to receive either a male or female applicant file was carried out at random.
   1. Identify the outcome of interest and groupings.
   2. Identify the study design.
   3. Based on the study design, is the odds ratio or relative risk a more appropriate inferential target?
   4. Perform inference according to your answer in (c); carry out a test and compute a confidence interval.

Gender Promoted NotPromoted  
1 Male 21 3  
2 Female 14 10

1. The Learning Early About Peanut allergy (LEAP) study recruited 530 children with risk factors for developing peanut allergies and randomly allocated peanut exposure and peanut avoidance regimens to each participant. At 5 years of age, an oral food challenge (OFC) test was administered to determine whether participants had developed allergies.
   1. Identify the outcome of interest and groupings.
   2. Identify the study design.
   3. Based on the study design, is the odds ratio or relative risk a more appropriate inferential target?
   4. Perform inference according to your answer in (c); carry out a test and compute a confidence interval.

treatment.group  
overall.V60.outcome Peanut Avoidance Peanut Consumption  
 FAIL OFC 36 5  
 PASS OFC 227 262

1. A 30-year study to investigate the effectiveness of mammograms versus a standard non-mammogram breast cancer exam was conducted in Canada with 89,835 female participants.12 During a 5-year screening period, each woman was randomized to either receive annual mammograms or standard physical exams for breast cancer. During the 25 years following the screening period, each woman was screened for breast cancer according to the standard of care at her health care center. At the end of the 25 year follow-up period, 1,005 women died from breast cancer. The results by intervention are summarized in mammogram.
   1. Identify the outcome of interest and groupings.
   2. Identify the study design.
   3. Based on the study design, is the odds ratio or relative risk a more appropriate inferential target?
   4. Perform inference according to your answer in (c); carry out a test and compute a confidence interval.

cancer.death  
mammogram yes no  
 yes 500 44425  
 no 505 44405

1. An observational study of 227,571 Medicare beneficiaries who initiated treatment with one of two diabetes medications recorded whether each patient reported cardiovascular problems.
   1. Identify the outcome of interest and groupings.
   2. Identify the study design.
   3. Based on the study design, is the odds ratio or relative risk a more appropriate inferential target?
   4. Perform inference according to your answer in (c); carry out a test and compute a confidence interval.

cardiovascular\_problems  
treatment no yes  
 Pioglitazone 154592 5386  
 Rosiglitazone 65000 2593

1. On March 31, 2021, Pfizer and BioNTech announced that “in a Phase 3 trial in adolescents 12 to 15 years of age with or without prior evidence of SARS-CoV-2 infection, the Pfizer-BioNTech COVID-19 vaccine BNT162b2 demonstrated 100% efficacy and robust antibody responses, exceeding those recorded earlier in vaccinated participants aged 16 to 25 years old, and was well tolerated.” These results are from a Phase 3 trial in 2,260 adolescents 12 to 15 years of age in the United States. In the trial, 18 cases of COVID-19 were observed in the placebo group (n = 1,129) versus none in the vaccinated group (n = 1,131).
   1. Identify the outcome of interest and groupings.
   2. Identify the study design.
   3. Based on the study design, is the odds ratio or relative risk a more appropriate inferential target?
   4. Perform inference according to your answer in (c); carry out a test and compute a confidence interval.

outcome  
group COVID-19 no COVID-19  
 vaccine 0 1131  
 placebo 18 1111

1. A random sample of 1629 U.S. adults obtained from NHANES data was used to investigate whether asthma is more common in women than in men.
   1. Identify the outcome of interest and groupings.
   2. Identify the study design.
   3. Based on the study design, is the odds ratio or relative risk a more appropriate inferential target?
   4. Perform inference according to your answer in (c); carry out a test and compute a confidence interval.

asthma  
sex yes no  
 female 49 781  
 male 30 769