# Measures of Effect for 2 by 2 Tables

BIOS 6611

CU Anschutz

Week 5

## 2 by 2 Contingency Tables Refresher

## 2 by 2 Contigency Tables

|  | Disease | | |
|---|---|---|---|
| Exposure | Yes | No | |
| Yes | a | b | $a+b = n_1$ |
| No | c | d | $c+d = n_2$ |
| | $a+c = m_1$ | $b+d = m_2$ | |

We are interested in the probability of disease for exposed and unexposed subjects.

Let $p_1 = \frac{a}{n_1}$ be the estimate for the probability of disease among exposed.

Let $p_2 = \frac{c}{n_2}$ be the estimate for the probability of disease among unexposed.

# Measures of Effect for 2 by 2 Contingency Tables

3 ways to describe behavior of $p_1$ and $p_2$:

1. Risk Difference (RD)

2. Risk Ratio (RR)

3. Odds Ratio (OR)

*Note: $p_1$, $p_2$, RD, RR, and OR are statistics. They use the data to estimate the true probability parameters.*

# Risk Difference

# Risk Difference (aka Attributable Risk)

- **Risk Difference (RD)**

$$RD = p_1 - p_2$$

- Appropriate for cohort and cross-sectional studies

  ▶ Cohort study interpretation: difference in incidence rates between exposed and unexposed individuals.

  ▶ Cross-sectional study interpretation: Difference in the prevalence rates between exposed and unexposed individuals.

- $RD > 0$ means increased risk by disease for those exposed, $RD < 0$ means decreased risk

- $-1 < RD < 1$

## Standard Error of Risk Difference

- Standard error (SE) for the RD is

$$SE(RD) = SE(p_1 - p_2) = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

  where $p = \frac{a+c}{n_1+n_2}$

- The sampling distribution of $RD$ is approximately normal, so the $1 - \alpha$ CI is

$$CI_{1-\alpha}(RD) = RD \pm SE(RD) \times Z_{1-\alpha/2}$$

## Risk Difference Example

Ex: Calculate the $RD$ and $SE(RD)$ for the following table:

| Drinking | Lung Cancer | | |
|---|---|---|---|
| Status | Yes | No | |
| Heavy | 33 | 1667 | 1700 |
| Non | 27 | 2273 | 2300 |
| | 60 | 3940 | 4000 |

## Risk Difference Example

Ex: Calculate the *RD* and *SE(RD)* for the following table:

| Drinking | Lung Cancer | | |
|---|---|---|---|
| Status | Yes | No | |
| Heavy | 33 | 1667 | 1700 |
| Non | 27 | 2273 | 2300 |
| | 60 | 3940 | 4000 |

$$RD = \frac{33}{1700} - \frac{27}{2300} \approx 0.0077$$

$$SE(RD) = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$= \sqrt{\frac{33 + 27}{1700 + 2300}\left(1 - \frac{33 + 27}{1700 + 2300}\right)\left(\frac{1}{1700} + \frac{1}{2300}\right)}$$

$$\approx 0.0039$$

$$CI_{1-0.05}(RD) = 0.0077 \pm 1.96 \times 0.0039$$

$$= (0.000056, 0.015)$$

## RD Interpretation

**Interpretation:** There were 0.7 additional lung cancer cases (95% CI: (0.00006,0.015)) per 100 subjects in the heavy-drinker group compared to the non-heavy drinker group.

The 95% confidence interval does not contain zero, so this is a statistically significant result. However, when we look at the CI, we see it gets very close to zero.

Reporting the risk difference reflects the probability of getting lung cancer in the first place, whereas the risk ratio and odds ratio do not.

# Risk Ratio

# Risk Ratio (special case of Relative Risk)

- **Risk Ratio (RR)**

$$RR = \frac{p_1}{p_2}$$

- Appropriate for cohort and cross-sectional studies

- Assuming a causal effect between exposure and outcome, $RR = 1$ means no relationship, $RR > 1$ indicates a risk factor, $RR < 1$ indicates a protective factor

- $0 < RR < \infty$

## Standard Error of Risk Ratio

- Sampling distribution of the natural log of the RR is approximately normal

- The standard error of $\log(RR)$ is

$$SE[\log(RR)] = SE\left[\log\left(\frac{p_1}{p_2}\right)\right] = \sqrt{\frac{b}{an_1} + \frac{d}{cn_2}}$$

- We calculate the CI for $\log(RR)$ and then exponentiate the bounds to get CI for RR:

$$CI_{1-\alpha}(RR) = \exp\{\log(RR) \pm Z_{1-\alpha/2}SE[\log(RR)]\}$$

# Risk Ratio Example

Ex: Calculate the *RR* and 95% CI for the following table:

| Drinking Status | Lung Cancer Yes | No | |
|---|---|---|---|
| Heavy | 33 | 1667 | 1700 |
| Non | 27 | 2273 | 2300 |
| | 60 | 3940 | 4000 |

## Risk Ratio Example

Ex: Calculate the *RR* and 95% CI for the following table:

| Drinking | Lung Cancer | | |
|----------|-----|------|------|
| Status | Yes | No | |
| Heavy | 33 | 1667 | 1700 |
| Non | 27 | 2273 | 2300 |
| | 60 | 3940 | 4000 |

$$
\begin{aligned}
RR &= \frac{33/1700}{27/2300} \approx 1.65 \\
CI_{1-0.05}(RR) &= \exp\{\log(RR) \pm Z_{1-0.05/2} \times SE[\log(RR)]\} \\
&= \exp\left\{ \log(1.65) \pm 1.96 \times \sqrt{\frac{1667}{33 \times 1700} + \frac{2273}{27 \times 2300}} \right\} \\
&= \exp\{(-0.00397, 1.005)\} \\
&= (0.996, 2.733)
\end{aligned}
$$

## RR Interpretation

**Interpretation:** Those who drink heavily have 1.65 times (95% CI: (0.996, 2.733)) the risk of lung cancer as those who do not drink heavily.

The CI contains one, so this is not a significant result. However, again, it is very close to one.

The RD and RR cannot be estimated for case-control studies. Why?

In case-control studies, you cannot measure incidence, because you start with predefined groups of diseased people and non-diseased people. So, you cannot calculate relative risk.

Which brings us to...

# Odds Ratio

## Odds Ratio

- **Odds Ratio (OR)**

$$OR = \frac{p_1/(1 - p_1)}{p_2/(1 - p_2)} = \frac{ad}{bc}$$

- Appropriate for cohort, cross-sectional, *and* case-control studies

- $OR = 1$ means the exposure does not impact the outcome, $OR > 1$ indicates a risk factor, $OR < 1$ indicates a protective factor

- $0 < OR < \infty$

- The OR approximates the RR for rare diseases ("low" incidence)

# Standard Error of Odds Ratio

- Sampling distribution of $\log(OR)$ is approximately normal

- The standard error for the OR is

$$SE[\log(OR)] = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

- CI is formed by exponentiating the bounds:

$$CI_{1-\alpha}(OR) \;\; = \;\; \exp\{\log(OR) \pm Z_{1-\alpha/2} \times SE[\log(OR)]\}$$

## Odds Ratio Example

Ex: Calculate the *OR* and *CI* for the following table:

| Drinking Status | Lung Cancer Yes | No | |
|---|---|---|---|
| Heavy | 33 | 1667 | 1700 |
| Non | 27 | 2273 | 2300 |
| | 60 | 3940 | 4000 |

## Odds Ratio Example

Ex: Calculate the *OR* and *CI* for the following table:

| Drinking Status | Lung Cancer Yes | No | |
|---|---|---|---|
| Heavy | 33 | 1667 | 1700 |
| Non | 27 | 2273 | 2300 |
| | 60 | 3940 | 4000 |

$$
\begin{aligned}
OR &= \frac{33 \times 2273}{27 \times 1667} \\
&\approx 1.665 \\
CI_{1-0.05}(OR) &= \exp\left\{ \log(1.665) \pm Z_{1-0.05/2} \times \sqrt{\frac{1}{33} + \frac{1}{1667} + \frac{1}{27} + \frac{1}{2273}} \right. \\
&= (0.9982, 2.7823)
\end{aligned}
$$

Note, very close to RR because this is case of low incidence!

## OR Interpretation

**Interpretation:** The odds of lung cancer in heavy drinkers is 1.665 (CI:(0.9982, 2.7823)) times the odds in non-heavy drinkers.

Again, CI contains 1 (marginally), so results are not significant.

# Summary