Risk Ratios: Relative Risk and Odds Ratio

STAT 244NF: Infectious Disease Modeling

Marie Ozanne

9/21/2021

Agenda

- Wrap up R Lab 2.
- Relative risk
- Odds ratio
- Brief introduction - logistic regression

R Lab 2 wrap-up

- Fitting `glm` and what went wrong
- Interpretation of the results.
- Connection to what we did before.

Relative risk

Definition

The *relative risk* (RR) is a type of risk ratio that can be useful to compare probabilities of disease, or some other health condition, in two different groups or in two different situations. The formula for RR is defined in terms of a ratio of conditional probabilities:

$$RR = \frac{P(\text{disease}|\text{exposed})}{P(\text{disease}|\text{unexposed})}$$

(Read $P(disease|exposed)$ as the "probability of disease given exposed").

What are the two groups that are being compared in a relative risk?

In what type of study (cohort, case-control, or cross-sectional) would you expect to have data where participants are organized into these groups?

Example

In September 2003, the Connecticut Department of Public Health received notification of an outbreak of skin infections due to Methicillin-resistant Staphylococcus aureus (MRSA) among members of a college (American) football team. To understand this outbreak better, investigators performed a retrospective cohort study to examine skin injuries, hygienic practices, and known risk factors for MRSA. Specifically, they considered these injuries, practices, and risk factors between the time of arrival at the football camp and the outbreak announcement (Begier et. al., 2004).

The incidence curve represented in the paper is included and labeled Figure 1; details about the number of case patients with different risk factors are included in Table 2 ((c) 2004 by the Infectious Diseases Society of America).

What makes this a retrospective cohort study?

Using data from Table 2, how would we calculate a simple estimate of the relative risk of developing MRSA for those players who suffered from turf burn?

(a) Fill out the table below to get started:

	Yes MRSA	No MRSA
Yes turf burn		
No turf burn		

(b) Match the table to values in the table to the probabilities needed to calculate relative risk.

Odds ratio

$$OR = \frac{P(\text{disease}|\text{exposed})/[1 - P(\text{disease}|\text{exposed})]}{P(\text{disease}|\text{unexposed})/[1 - P(\text{disease}|\text{unexposed})]}$$

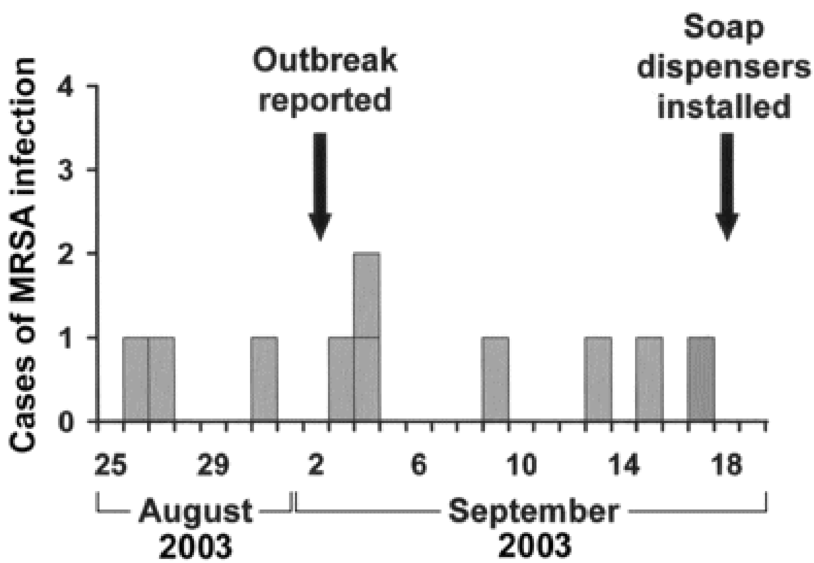
Equivalently,

$$OR = \frac{P(\text{exposure}|\text{diseased})/[1 - P(\text{exposure}|\text{diseased})]}{P(\text{exposure}|\text{nondiseased})/[1 - P(\text{exposure}|\text{nondiseased})]}$$

These two definitions are mathematically equivalent. Using what you know about conditional, joint, and marginal probabilities, as well as complements, show that these two definitions of the OR are equivalent.

To calculate an OR defined in this way, what kind of study (or studies) would we need to generate our data?

Figure 1 Date of onset of initial skin infections due to methicillin-resistant Staphylococcus aureus (MRSA) among ...



Clin Infect Dis, Volume 39, Issue 10, 15 November 2004, Pages 1446–1453, https://doi.org/10.1086/425313

Table 2 Additive interaction of player position with turf burns and body shaving, 2 markers of interruption of skin ...

Risk factor	High-risk position[a]	No. of case patients	Total no. of players	MRSA attack rate, % of players	Attributable risk of MRSA infection, % of players
Body shaving[b]					
No	No	0	53	0.0	Reference
Yes	No	2	17	11.8	11.8
No	Yes	3	12	25.0	Reference
Yes	Yes	5	8	62.5	37.5
Turf burn[c]					
No	No	0	32	0.0	Reference
Yes	No	2	38	5.3	5.3
No	Yes	1	8	12.5	Reference
Yes	Yes	7	12	58.3	45.8

NOTE. We were unable to assess multiplicative interaction because the attack rate in the reference group is 0. MRSA, methicillin-resistant *Staphylococcus aureus*.

[a] Cornerback or wide receiver.

[b] Shaving body areas other than the face (excluding shaving the head with hair clippers to produce a crew cut).

[c] Abrasions due to artificial grass.

Clin Infect Dis, Volume 39, Issue 10, 15 November 2004, Pages 1446–1453, https://doi.org/10.1086/425313

Relative risk vs. Odds ratio

- Relative risk and odds ratios both attempt to explain the same phenomenon.
- The relative risk is generally considered to be more intuitive, but the odds ratio has better statistical properties.
- Often people think of the odds ratio as a close approximation to the relative risk. However, this is only true under certain circumstances. Specifically, OR is a close approximation under the *rare disease assumption.*
- In order for the *rare disease assumption* to apply (in case-control studies), the prevailing rule of thumb is that the prevalence of disease in the population must be no more than 10%. There are a number of complicating factors related to this assumption, but we are going to use this rule of thumb in this class. Details of these complications are beyond the scope of this class.
- Under the *rare disease assumption*,
- $P(\text{disease}|\text{exposed}) \approx 0$
- $P(\text{disease}|\text{unexposed}) \approx 0$

Show that under the rare disease assumption, $RR \approx OR$. OR =

Let us think about an infection outbreak that grows exponentially. Suppose we can only identify people with and without disease and then look at their exposures, so we are limited to a case-control study. Also suppose we can identify these cases (and controls) in real time, starting at the beginning of the outbreak. In what time interval would you expect the odds ratio to be a good approximation of the relative risk?

Limitations of ratios

- Both RR and OR are ratios, so they do not reflect the magnitude of the probabilities used to calculate them.
- RR and OR should be placed in context when thinking about the risk associated with particular exposure.

(Chronic disease example) Suppose in a study of breast cancer, which investigates the effects of hormone use in postmenopausal people, those who had used hormone therapy for 5-9 years have an odds of developing invasive breast cancer that is 1.46 times the odds for those that have never used hormones. It has also been reported that the probability a 60-year old woman will develop breast cancer in the next 10 years is 3.59%. What is the probability of developing breast cancer given hormone therapy for 5-9 years?

- $OR = 1.46$
- $P(\text{cancer}|\text{no use}) = 0.0359$

Connection to regression models

Poisson regression and RR

- If we have data from a cohort study, and we fit our count data using a Poisson regression model, then we can estimate the relative risk of disease (or some other condition) for an exposure.
- Recall the form of the Poisson (log-linear) model:

$$\log(\lambda) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

- If we have adjusted the rate based on population size or the length of reporting time (e.g., R Lab 2), then we have

$$\log(\lambda) = \text{offset} + \beta_1 X_1 + \cdots + \beta_p X_p$$

- *Make sure you know the rationale for the offset term, even if you are not entirely comfortable deriving it.*

Sample R code

```
## Sample intercept-only glm fit
pois_glm_fit0 <- glm(y ~ 1, data=data, family=poisson(link = "log"))
## General sample glm fit for Poisson regression
pois_glm_fit <- glm(y ~ x_1 + ... + x_p, data=data, family=poisson(link = "log"))
## Print summary of fit; includes parameter estimates
summary(pois_glm_fit)
## 95% confidence intervals for estimates:
confint(pois_glm_fit)
## 95% confidence intervals for probabilties
exp(confint(pois_glm_fit))
```

Logistic regression and OR

- With data from a case-control or cohort study (or more generally, data collected in an appropriate way from two groups), we can fit out count data using a Binomial logistic regression model. Then we can estimate the odds ratio.

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

Sample R code

```
## Sample intercept-only glm fit
logit_glm_fit0 <- glm(y ~ 1, data=data, family=binomial(link = "logit"))
## General sample glm fit for Poisson regression
logit_glm_fit <- glm(y ~ x_1 + ... + x_p, data=data, family=binomial(link = "logit"))
## Print summary of fit; includes parameter estimates
summary(logit_glm_fit)
## 95% confidence intervals for estimates:
confint(logit_glm_fit)
## 95% confidence intervals for odds
exp(confint(logit_glm_fit))
```