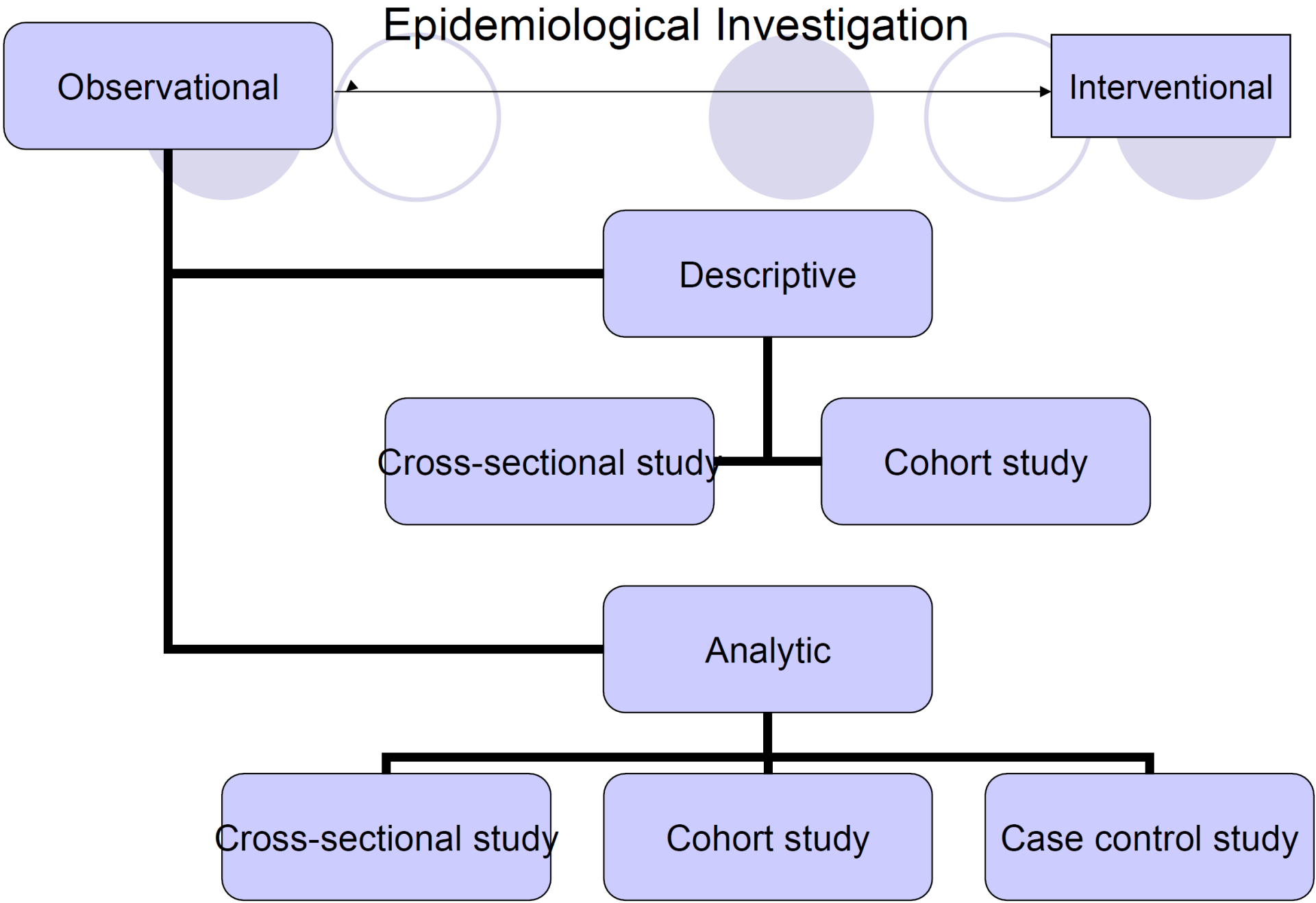


Measures of association

MWAI



2X2 or contingency table

| Exposure | Outcome | | Total |
|-----------------|----------------|-----|--------------|
| | Yes | No | |
| Yes | a | b | a+b |
| No | c | d | c+d |
| Total | a+c | b+d | a+b+c+d |

Presentation of data in Case control or Cohort studies with count denominators

2x2 table

- Contains two rows and two columns
- Each represents the presence or absence of exposure or disease.
 - ***a= number of individuals who are exposed and have outcome***
 - ***b= number who are exposed and do not have outcome***
 - ***c=number who are not exposed and have outcome***
 - ***d= number who are not exposed and do not have outcome***
- The margin represent total number in each row and column calculated by adding relevant cells

2x2 table

| | |
|-------|---|
| $a+b$ | the total number of individuals exposed |
| $c+d$ | the total number of individuals not exposed |
| $a+c$ | the total number of individuals with outcome |
| $b+d$ | the total number of individuals without outcome |

The sum of all four cells $a+b+c+d$ is the total sample size, represent T or N

Measures of Association

- Quantifies the relationship between exposure and outcome among two groups
 - Odds Ratio
 - Risk Ratio
- Ratio - ***how much more one group is to develop a disease than the another***

Important Jargon

- **Exposure (E)** \equiv an explanatory factor; the independent variable i.e. any potential health determinant;
- **Outcome (D)** \equiv the response; the dependent variable i.e. any health-related outcome;
- **Measure of association (syn. measure of effect)** \equiv a statistic that quantifies the relationship between an exposure and a disease
- **Measure of potential impact** \equiv a statistic that quantifies the potential impact of removing a hazardous exposure

Relative Risk/Risk Ratio

- Relative Risk/Risk Ratio (RR) → relative effect associated with exposure or the “risk”
 - Cohort studies
 - “Risk” refers to the probability of occurrence of an event or outcome

Risk Ratio

$$RR = \frac{R_1}{R_0}$$

where

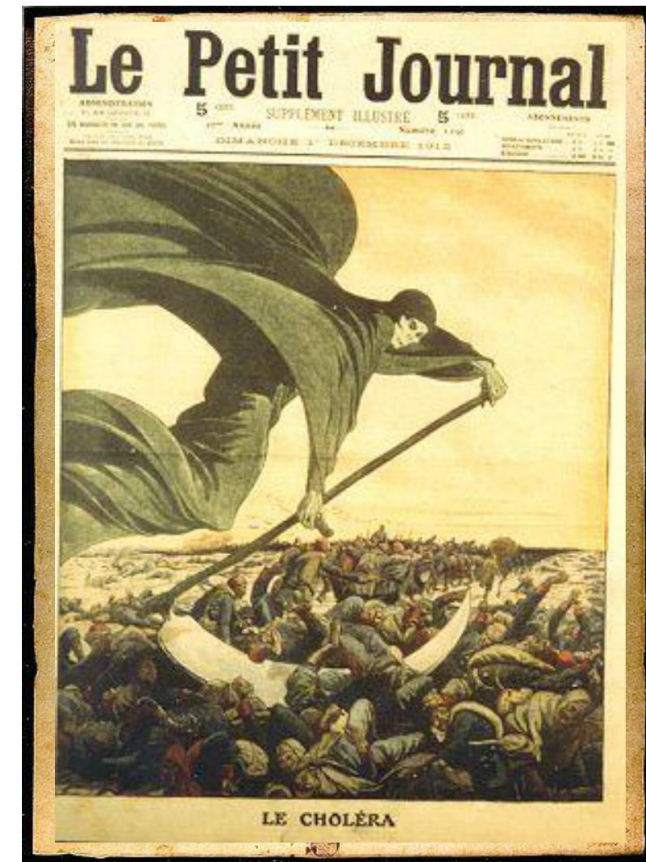
$R_1 \equiv$ risk in the exposed group

$R_0 \equiv$ risk in the non-exposed group

Example

Fitness & Mortality (Blair et al., 1995)

- Is improved fitness associated with decreased mortality?
- Exposure \equiv improved fitness
(1 = yes, 0 = no)
- Disease \equiv death
(1 = yes, 0 = no)
- Mortality rate, group 1:
 $R_1 = 67.7$
- Mortality rate, group 0:
 $R_0 = 122.0$



RR?

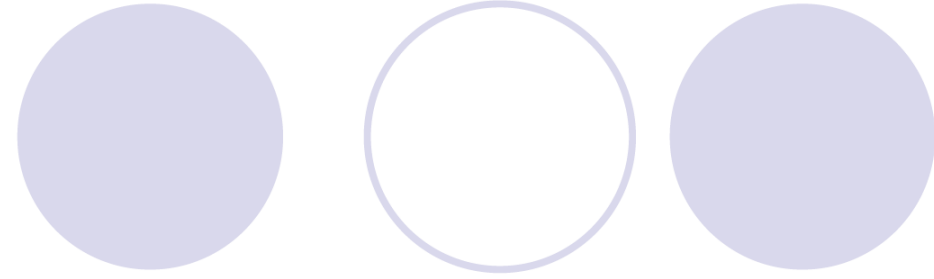
$$RR = \frac{R_1}{R_0}$$

where

$R_1 \equiv$ risk in the exposed group

$R_0 \equiv$ risk in the non-exposed group

Example Relative Risk



What is the effect of improved fitness on mortality in relative terms?

$$RR = \frac{R_1}{R_0} = \frac{67.7}{122.0} = 0.55$$

The effect of the exposure is to cut the risk *almost* in half.

2-by-2 Table Format

| | Outcome + | Outcome - | Total |
|------------|-----------|-----------|-------|
| Exposure + | A_1 | B_1 | N_1 |
| Exposure - | A_0 | B_0 | N_0 |
| Total | M_1 | M_0 | N |

$$R_1 = \frac{A_1}{N_1}$$

$$R_0 = \frac{A_0}{N_0}$$

Food borne Outbreak Example

Exposure \equiv eating a particular dish

Outcome \equiv gastroenteritis

| | Outcome + Outcome - | | Total |
|------------|---------------------|----|-------|
| Exposure + | 63 | 25 | 88 |
| Exposure - | 1 | 6 | 7 |
| Total | 64 | 31 | 95 |

Food borne Outbreak Example

Exposure \equiv eating a particular dish

Outcome \equiv gastroenteritis

| | Outcome + Outcome - | | Total |
|------------|---------------------|----|-------|
| Exposure + | 63 | 25 | 88 |
| Exposure - | 1 | 6 | 7 |
| Total | 64 | 31 | 95 |

$$R_1 = \frac{A_1}{N_1} = \frac{63}{88} = 0.7159$$

$$R_0 = \frac{A_0}{N_0} = \frac{1}{7} = 0.1429$$

Food borne Outbreak Data

| | Outcome+ | Outcome – | Total |
|------------|----------|-----------|-------|
| Exposure + | 63 | 25 | 88 |
| Exposure – | 1 | 6 | 7 |
| Total | 64 | 31 | 95 |

$$RR = \frac{R_1}{R_0} = \frac{63/88}{1/7} = \frac{0.7159}{0.1429} = 5.01$$

Exposed group had 5 times the risk

How to report RR

- The risk of food borne disease is 5X higher in the exposed group compared to the non exposed group.
- Exposure is associated with 400% increase in the risk of food borne disease.

Note:: Remember to report the 95% Confidence interval for each

ODD Ratio

- An *odds ratio* (OR) → is a measure of association between an exposure and an outcome.
- “Odds” refers to the probability of occurrence of an event/probability of the event not occurring.¹
- The OR represents the *odds* that an outcome will occur given a particular exposure, compared to the *odds* of the outcome occurring in the absence of that exposure.
 - Case control studies

¹ For <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4640017/>

$$\text{ODDS} = \frac{\text{Number with Outcome (D+)}}{\text{Number without Outcome (D-)}}$$

$$\text{ODDS RATIO} = \frac{\text{odds in exposed } \left(\frac{A_1}{B_1} \right)}{\text{odds in non exposed } \left(\frac{A_0}{B_0} \right)}$$

| | D+ | D- | Total |
|-------|-------|-------|-------|
| E+ | A_1 | B_1 | N_1 |
| E- | A_0 | B_0 | N_0 |
| Total | M_1 | M_0 | N |

$$OR = \frac{A_1 / B_1}{A_0 / B_0} = \frac{A_1 B_0}{B_1 A_0}$$

“Cross-product ratio”

Odds Ratio, Example

Milunsky et al, 1989, Table 4

NTD = Neural Tube Defect

| | NTD+ | NTD- |
|-------------|------|--------|
| Folic Acid+ | 10 | 10,703 |
| Folic Acid- | 39 | 11,905 |

$$OR = \frac{A_1 B_0}{B_1 A_0} = \frac{10 \cdot 11,905}{10,703 \cdot 39} = 0.29$$

How to report OR

- The odds of neural tube effect (outcome) given folic acid consumption (exposure) is 0.29X lower relative to lack of folic acid.
- The odds of neural tube effect (outcome) given folic acid consumption (exposure) is 71% lower relative to lack of folic acid.

Note:: Remember to report the 95% Confidence interval for each

Risk Ratio vs Odds Ratio

- In **case-control studies**, and in **cohort studies** in which the outcome occurs in less than 10% of the unexposed population, the OR provides a reasonable approximation of the RR
- RR represent the **risk of an outcome among the exposed group** compared to the risk of the outcome in the non exposed group
- **OR** represents the **odds that an outcome will occur given a particular exposure**, compared to the odds of the outcome occurring in the absence of that exposure.

What if?

- Want to deal with confounding

What if?

- Want to deal with confounding during analysis
 - Regression (only works if you can identify and measure the confounders)
 - Stratify
 - adjustment (usually distorted by choice of standard)

Logistic regression

- **Logistic regression** - a regression modelling technique for producing Odds Ratios (ORs); models the log odds of a binary “outcome”
 - *Effect of T.B infection on death in HIV positive patients crude(unadjusted) OR; 95% CI and hypothesis tests*
 - *Effect of mothers education on childs' measles immunisation status*
 - *Effect of ethnicity on risk of death from breast cancer*
 - *Effect of gender on being a high wage earner*

Why model log odds?

- The reason for modelling the log odds rather than risk or odds is that the log odds can take any value, positive or negative, whereas risks are constrained to lie between 0 and 1.
- When using statistical models it is easier to model a quantity which is unconstrained than one which is constrained.
- This avoids the possibility of predicting impossible values (like risks which are negative or greater than 1) from the model.
- Modelling log odds is referred to as **logistic regression**, and the models are referred to as **logistic models**.

Example

Calculate the prevalence, odds and log odds of Microfilariae infection in the forest and savannah areas

| Microfilariae Infection | Savannah | Forest | Total |
|-------------------------|----------|--------|-------|
| Positive | 267 | 213 | 480 |
| Negative | 281 | 541 | 822 |
| Total | 548 | 754 | 1302 |

| | savannah | forest | overall |
|-----------------|----------|--------|---------|
| Risk/prevalence | | | 63.1 |
| Odds | | | 1.712 |
| Log odds | | | 0.538 |

What are the odds ratio and log odds ratio?

| | savannah | forest | overall |
|--------------------|----------|--------|---------|
| Risk/prevalence(%) | 51.3 | 71.8 | 63.1 |
| Odds | 1.052 | 2.540 | 1.712 |
| Log odds | 0.052 | 0.932 | 0.538 |

The Odds ratio = 2.41 Whilst the log odds ratio? = 0.881

| area | odds | log odds of disease |
|------------|-----------------------------|-------------------------|
| 0=savannah | 1.052 | 0.051 |
| 1=forest | $1.052 \times 2.41 = 2.536$ | $0.051 + 0.881 = 0.932$ |

Summarise results in a model

log odds = Baseline + Area

where Baseline = log (odds in savannah) = $(0.051 + 0.881 \times 0) = 0.051$
Area = log odds for individuals in the forest and 0 individuals in the savannah
= $(0.051 + 0.881 \times 1) = 0.932$

Example in R - <https://bit.ly/2OTXImY>

- Calculate the prevalence, odds and log odds of Microfilariae infection in the forest and savannah areas
- `onch <- read.csv("onchall.csv")` *# Read in CSV data*
- `m1 <- glm(mf~area, data=onch, family=binomial)` *# Run model*
- `summary(m1)`