The Online R-FETPV 1st Module : Basic Epidemiology and Surveillance Data Analysis

5 April -28 May 2021

USAID FROM THE AMERICAN PEOPLE

Department of Livestock Development

Food and Agriculture Organization of the United Nations

Session Title: Inferential statistics in epidemiology

Sangeeta Rao, BVSc, MVSc, PhD
Colorado State University

Food and Agriculture Organization of the United Nations

Outline:

- Measures of association
- Chi-square
- Odds Ratio
- Risk Ratio
- Confidence limits

Survey of cattle TB in slaughter plants:

You are interested to investigate TB occurrence in cattle in slaughter plants in your region. Your diagnosis is based on examination of cattle for the 'presence of TB lesions' at post-mortem inspection.

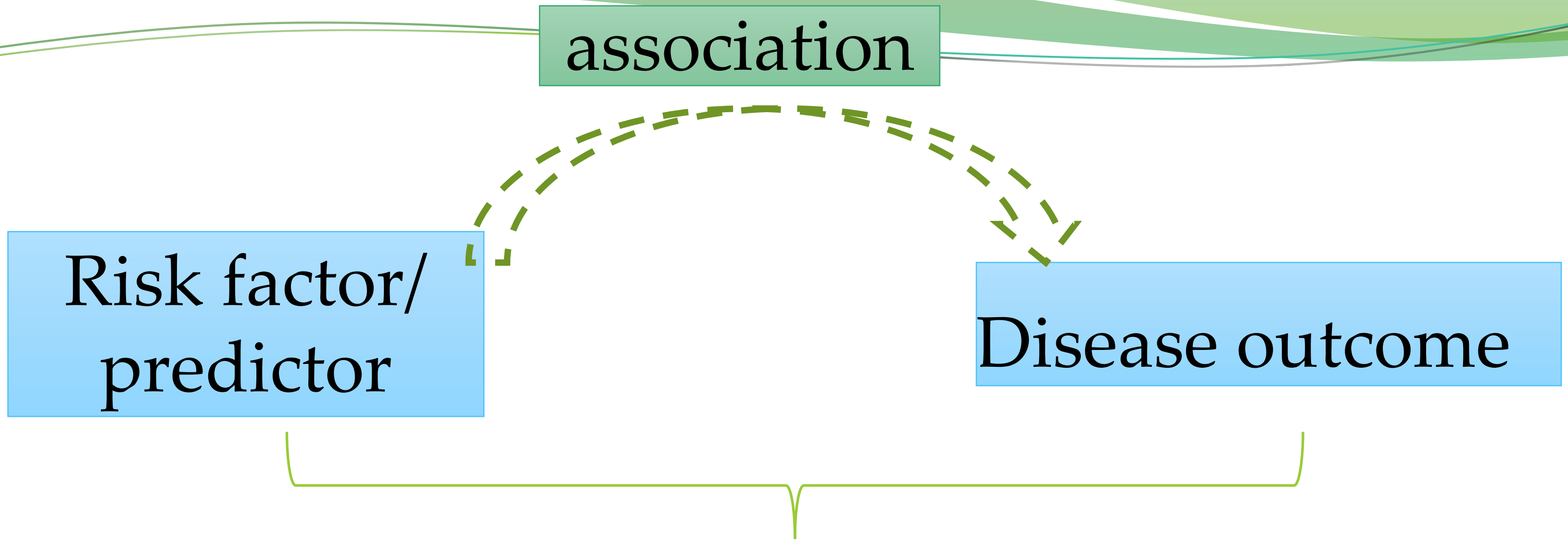
After following appropriate sampling procedures, you have planned to collect samples from 120 cattle randomly selected from one slaughter plant in your region and trace to their farms. You hypothesize that 'over-crowding' conditions at the farms to be a risk factor for the TB occurrence. Other information of breed, age etc. were also collected during the investigation. The investigation was performed at the animal level.

Shutterstock.com

Survey of cattle TB :

	TB lesion +	TB lesion -	Total
Over-crowding ‘Yes’			
Over-crowding ‘No’			
Total			

Measures of association



Requires Collection of data

Description of data

Drawing inferences on ‘association’ using appropriate analysis

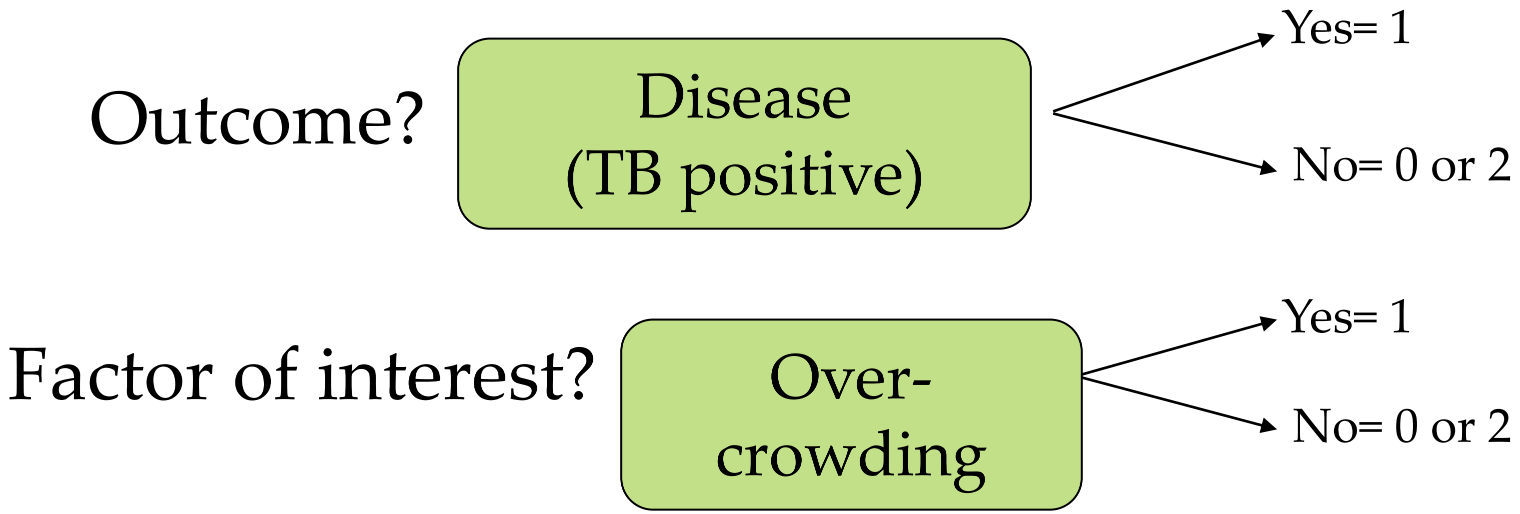
Measures of association

- Assess the strength of the association between an exposure and disease
- Compare disease measurements between groups of individuals
- Chi-square and Fisher's exact test
- Odds Ratio and Risk Ratio (measures of assessing likelihood/ risk of disease)

Statistical analysis of categorical data:

- Chi-square analysis ($\chi 2$)

Example:



For categorical data:

2x2 table (or rxn or contingency table):

TB

Outcome

Over-crowding **Factor**	Yes	No	
Yes	a	b	***r1***
No	c	d	***r2***
	c1	***c2***	***n***

$$Chi-square = \frac{[(a*d)-(b*c)]^2 * n}{c1*c2*r1*r2}$$

Note:
Each cell in table should have at least 5 counts to use Chi-square analysis.

Example

- You collect data on TB in total of 120 cattle.
- 80 had TB and 40 did not.
- Of the cattle with TB, 45 came from over-crowded farms and of the non-TB cattle, 8 came from over-crowded farms .
- Is over-crowding a significant factor for TB in cattle?

  or,

  Is TB more common in cattle from over-crowded farms than non-TB cattle?

		TB	
		yes	no
Over-crowding	yes	a	b
	no	c	d

		TB		
		yes	no	
Over-crowding	yes	45	8	53
	no	35	32	67
		80	40	120

- P-value?
- Statistical significance?
- Biological significance?

Note:
Each cell in table should have at least 5 counts to use Chi-square analysis.

Chi-square statistic (χ^2):

TB

		yes	no	
Over-crowding	yes	45 $_a$	8 $_b$	53 $_{r1}$
	no	35 $_c$	32 $_d$	67 $_{r2}$
		80 $_{c1}$	40 $_{c2}$	120 $_n$

$$Chi - square = \frac{[(a*d)-(b*c)]^2 * n}{c1*c2*r1*r2}$$

$$Chi - square = \frac{[(45*32)-(8*35)]^2 *120}{80*40*53*67} = 14.21$$

(the squared difference between the diagonals, times the total)
divided by
(the product of all the margins)

Interpretation:

$$Chi - square = \frac{[(45*32)-(8*35)]^2 * 120}{80*40*53*67} = 14.21$$

Compare the Chi-square statistic to these critical values
[If chi-square value > critical value, the association is significant]:
Critical values:
80%: 1.64
90%: 2.70
95%: 3.84

14.21> 3.84 at 95% confidence level, hence the TB in cattle is significantly associated with Over-crowding.

Odds Ratio

		Diseased (+)	Diseased (-)	
Exposed	(+)	Exposed & diseased	Exposed & not diseased	# exposed
	(-)	Unexposed & diseased	Unexposed & not diseased	# Unexposed
		# diseased	# not diseased	Total

$$\text{Odds of exposure among diseased} = \frac{\text{\# Exposed \& diseased}}{\text{\# Exposed \& not diseased}}$$

$$\text{Odds of exposure among not diseased} = \frac{\text{\# Unexposed \& diseased}}{\text{\# Unexposed \& not diseased}}$$

Odds Ratio

- What does it tell us?
  - The odds of disease in exposed compared to unexposed

- When can you use it?
  - When you have counts of diseased and exposed animals
    - Case-control studies

$$\text{OR} = \frac{\text{Odds of exposure in diseased}}{\text{Odds of exposure in non-diseased}}$$

- Interpretation
  - >1 means exposure associated with more disease
  - <1 means exposure associated with less disease (preventive)

Example

- A student was conducting research to ascertain relationship between biosecurity and HPAI infection in commercial poultry farms in a country with biosecurity guidelines. He selected 50 affected farms and 100 non affected farms for his study and found that 10 farms among the affected farms and 40 farms among non affected farm complied with the biosecurity guidelines. Find the relationship with the non-compliance and disease occurrence.

		Disease (+)	Disease (-)	
Exposure	(+)	(Exposed & diseased)	(Exposed & not diseased)	# Exposed
	(-)	(Unexposed & diseased)	(Unexposed & not diseased)	# Unexposed
		# diseased	# Not diseased	Total

		HPAI Outbreak (+)	HPAI Outbreak (-)	
Biosecurity guideline not followed	(+)	40	60	100
	(-)	10	40	50
		50	100	Total

$$OR = \frac{40/60 \text{(Odds of disease in exposed)}}{10/40 \text{(Odds of disease in unexposed)}}$$

$$OR = \frac{40 \times 40}{60 \times 10} = \mathbf{2.66}$$

The odds/ likelihood of HPAI outbreak was **2.7 times** higher in farms that did not follow biosecurity guidelines

Relative risk (Risk Ratio)

		Diseased (+)	Diseased (-)	
Exposed	(+)	Exposed & diseased	Exposed & not diseased	# Exposed
	(-)	Unexposed & diseased	Unexposed & not diseased	# Unexposed
		# diseased	# not diseased	Total

$$\text{Risk of disease among exposed} = \frac{\text{Exposed \& Diseased}}{\text{\# Exposed}}$$

$$\text{Risk of disease among unexposed} = \frac{\text{Unexposed \& Diseased}}{\text{\# Unexposed}}$$

Relative risk (Risk Ratio)

- What does it tell us?
  - Risk (or rate) of disease in exposed relative to unexposed
- When can you use it?
  - When you have prevalence or incidence data
    - Outbreak situations
    - Cohort studies

$$RR = \frac{\text{Risk of disease in exposed}}{\text{Risk of disease in unexposed}}$$

- Interpretation
  - >1 means exposure associated with more disease
  - <1 means exposure associated with less disease (preventive)

Example

- 300 cattle of a village was followed up for 3 years to observe development of fascioliasis . 100 cattle were allowed to graze in marshy land and 200 cattle were not allowed to graze in the marshy land. 27 cattle developed fascioliasis in the herd grazing in the marshy land and 48 cattle developed the disease among those not allowed to graze in marshy land. Calculate the risk ratio.

		Fascioliasis (+)	Fascioliasis (-)
Exposed	(+)		
Unexposed	(-)		

Fascioliasis

		(+)	(-)	
Exposed	(+)	27	73	100
Unexposed	(-)	48	152	200
		75	225	300

Risk of disease among exposed = 27/100= 0.27

Risk of disease among un-exposed = 48/200= 0.24

Relative Risk = 0.27 /0.24= 1.13

Cattle grazing in marshy land are 1.13 times at higher risk of getting the disease fascioliasis

Attributable risk (aka Risk difference)

- What does it tell us?
  - Risk (or rate) of disease in exposed due to exposure
- When can you use it?
  - When you have data on new cases in a population
    - Outbreak situations
    - Cohort studies

AR = Risk (rate) in exposed – risk (rate) in unexposed

$$\frac{\text{Exposed \& Diseased}}{\text{\# Exposed}} - \frac{\text{Unexposed \& Diseased}}{\text{\# Unexposed}}$$

Confidence Intervals for Odds Ratio or Risk Ratio

Which is the right question?

Is there a difference in risk or odds between groups?

or,

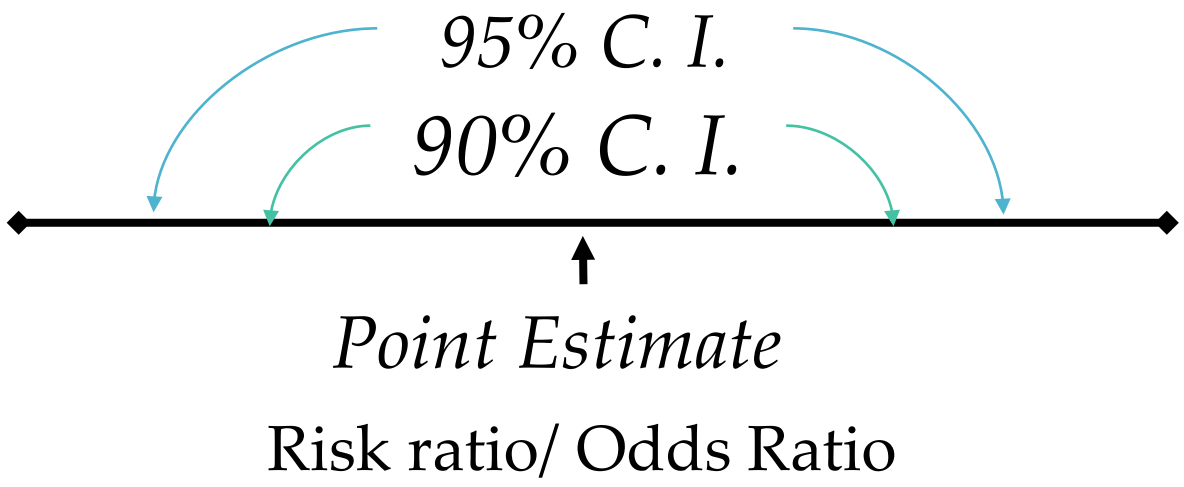
Is there a difference in risk or odds of disease between exposed and non-exposed groups?

Odds Ratio
or
Risk Ratio:

1.0	*– odds or risk is the same in both exposed and unexposed groups*
> 1.0	*– odds or risk is higher for exposed*
< 1.0	*– odds or risk is lower for exposed (protective)*

Confidence Interval or Confidence Limit:

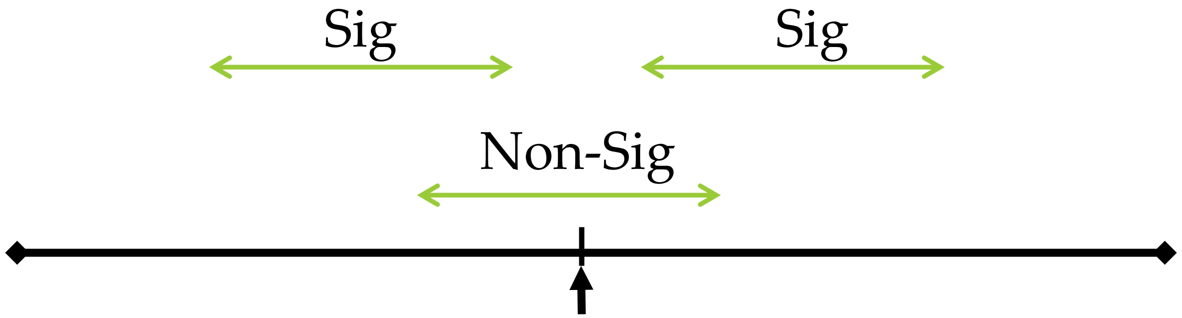
The potential range, around the estimate, in which the real value should lie



Interpretation of Confidence Interval of OR or RR:

CI overlapping ‘1’: non-significant

CI below or above ‘1’: significant

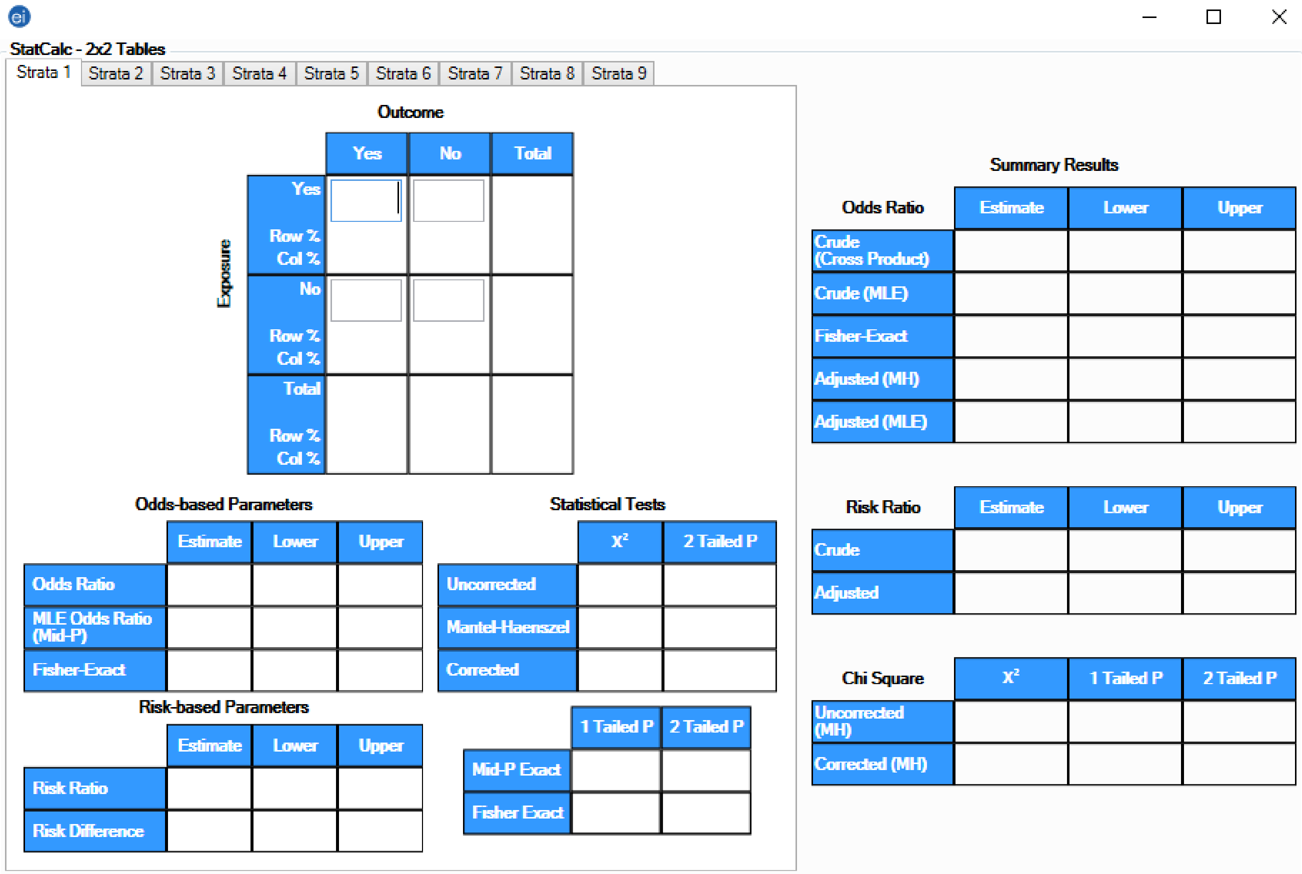


RR or OR=1

Risk ratio/ Odds Ratio

http://wwwn.cdc.gov/epiinfo/7/index.htm


epiinfo™
CREATE FORMS
Create surveys or questionnaires with field validation and skip logic.
ENTER DATA
Enter data, browse records, and search the database.
CREATE MAPS
Map data with coordinates, by choropleth, or dot density.
STATCALC
Statistical calculators for sample size, power, and more.
ANALYZE DATA
CLASSIC
Clean, transform, and analyze data with commands.
EPI INFO™ WEBSITE | ABOUT EPI INFO™
StatCalc
STATCALC
POPULATION SURVEY
COHORT OR CROSS-SECTIONAL
UNMATCHED CASE-CONTROL
CHI SQUARE FOR TREND
TABLES (2 x 2 x N)
POISSON (RARE EVENT VS. STD)
POPULATION BINOMIAL (PROPORTION VS. STD.)
MATCHED PAIR CASE CONTROL STUDY
EPI INFO™ WEBSITE | ABOUT EPI INFO™
LANGUAGE: en-US
VERSION:7.2.3.1


StatCalc - 2x2 Tables
Strata 1
Strata 2
Strata 3
Strata 4
Strata 5
Strata 6
Strata 7
Strata 8
Strata 9
Outcome
Yes
No
Total
Exposure
Row %
Col %
Odds-based Parameters
Estimate
Lower
Upper
Odds Ratio
MLE Odds Ratio (Mid-P)
Fisher-Exact
Risk-based Parameters
Risk Ratio
Risk Difference
Statistical Tests
X²
2 Tailed P
Uncorrected
Mantel-Haenszel
Corrected
1 Tailed P
Mid-P Exact
Fisher Exact
Summary Results
Odds Ratio
Crude (Cross Product)
Crude (MLE)
Adjusted (MH)
Adjusted (MLE)
Risk Ratio
Crude
Adjusted
Chi Square
Uncorrected (MH)
Corrected (MH)