Statistical significance using Confidence Intervals

Objectives:

- Able to understand the concept of confidence intervals.
- Able to apply the concept of statistical significance using confidence intervals in analyzing the data.
- Able to interpret the concept of 95% confidence intervals in making valid conclusions.

Team Members: Jawaher Abanumy - Monera AlAyuni- basel almeflh

Team Leaders: Rawan Alwadee & Mohammed ALYousef

Revised By: Maha Alghamdi

Dr. shaffi ahmed

Resources:

- 436 Lecture Slides + Notes

Important – Notes - KAPLAN Behavioral Science and Social Sciences

436researchteam@gmail.com

Editing file

Feedback link

Statistical Significance using Confidence Intervals

Confidence Interval
"Interval Estimation"

- Formula:
  $CI = p \pm Z\alpha \times SE$
- Most commonly used values:
  - CI **90%** corresponds to α = **0.10** & Z value = **1.645**
  - CI **95%** corresponds to α = **0.05** & Z value = **1.96**
  - CI **99%** corresponds to α = **0.01** & Z value = **2.576**
- Characteristics:
  Width of the confidence interval depends on:
  1. Sample size
  2. Variability
  3. Degree of confidence
- **Interpretation** (statistical & clinical significance)
- Examples & Application

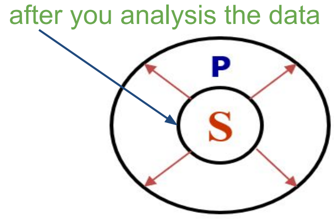
Sampling variability
"**Standard Error**"

- Sample **statistic** vs sample **variability**
- Formula: **$SEM = \frac{SD}{\sqrt{n}}$**
- Effect of variability (properties of error):
  Error **increases** with
  - Smaller sample size
  - Larger standard deviation
  - Larger z value

P-values and CIs

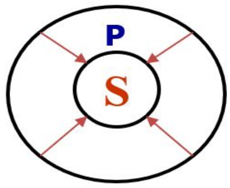
- Duality between P-values and CIs
- Comparison: p-value vs CI

Association studies: the value of no difference = on RR. OR.



Hypothesis Testing (p value) & Estimation (confidence interval)

Sample is assumed to be representative to the population.
In research: measurement are always done in the sample, the results will be applied to population.



Statistic and Parameter:(statistic=sample,,parameter=population)

- An observed value drawn from the sample is called a statistic
- The corresponding value in population is called a parameter
- We measure, analyze, etc statistics and translate them as parameters

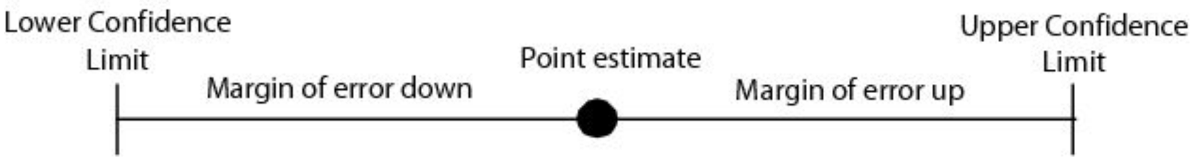
There is no need to do the whole population in the study. We depend on a sample and from the sample calculations we estimate population parameters

Estimation

Two forms of estimation:

- **Point estimation** = single value, e.g., (mean, proportion, difference of two means, difference of two proportions, OR, RR etc.,) E.g. Class mean marks, number of students who do daily exercise, prevalence of diabetes and incidents of hypertension > all are one value so considered point estimates

- **Interval estimation** = range of values ⇒ confidence interval (CI). A confidence interval consists of:



interval estimation : contant 3 thing, margin error down ,point estimate,margin of error up

Confidence intervals

- P values give no indication about the clinical importance of the observed association P value will tell you whether the results are statically significant or not, whether the risk factor is associated with outcome or not but it won't tell you the direction and it won't tell you the magnitude
- Relying on information from a sample will always lead to some level of uncertainty.
- Confidence interval is a range of values that tries to quantify this uncertainty:
  - For example , 95% CI means that under repeated sampling 95% of CIs would contain the true population parameter

  Uncertainty: Not sure whether the results you got from the study is close to population parameter or not because of sampling variations, measurement errors, selection bias or whatever mistake in methodology.
  CI reveals whatever the miss up you done during methodology

> Confidence intervals are a way of admitting that any measurement from a sample is only an estimate of the population. Although the estimate given from the sample is likely to be close, the true values for the population may be above or below the sample values. A confidence interval specifies how far above or below a sample-based value the population value lies within a given range, from a possible high to a possible low. Reality, therefore, is most likely to be somewhere within the specified range.

Computing confidence intervals (CI)

- **General formula:**

(Sample statistic) ± [(confidence level) × (measure of how high the sampling variability is)]

- **Sample statistic:** Whatever you calculate from the sample

observed magnitude of effect or association (e.g., odds ratio, risk ratio, single mean, single proportion, difference in two means, difference in two proportions, correlation, regression coefficient, etc.,)

- **Confidence level:**

varies – 90%, 95%, 99%. For example, to construct a 95% CI, $Z_{a\backslash 2}$ =1.96

Multiplied by measure of how the sampling variability is (each sample will be different value) so it means how much error will happen when we take different samples.

Most of the times we use 95% confidence according to the normal distribution concept. So we assume that 95% is the normal range and anything beyond that we consider abnormal

- **Sampling variability**:

Standard error (S.E.) of the estimate is a measure of variability

If we take different samples from the population each sample will give different values. So sampling variability is calculated from how much the error will happen among different sample means and what different sample proportions

Don't get confused with the terms of STANDARD DEVIATION (Variability within a sample) **and STANDARD ERROR** (Variability among samples (more than one group))

Example:

Data: X = {6, 10, 5, 4, 9, 8}; N = 6

Mean:

$$\bar{x} = \frac{\Sigma x}{N} = \frac{42}{6} = 7$$

Variance:

$$s^2 = \frac{\sum(\bar{x} - x)^2}{N} = \frac{28}{6} = 4.67$$

Standard Deviation:

$$s = \sqrt{s^2} = \sqrt{4.67} = 2 \cdot 16$$

on average the 6 value are deviating from the mean by 2.16

x	$x - \bar{x}$	$(x - \bar{x})^2$
1	-1	6
9	3	10
4	-2	5
9	-3	4
4	2	9
1	1	8
Total: 28		Total:42

Statistical Inference is based on Sampling Variability

Sample Statistic – we summarize a sample into one number; e.g., could be a mean, a difference in means or proportions, an odds ratio, or a correlation coefficient

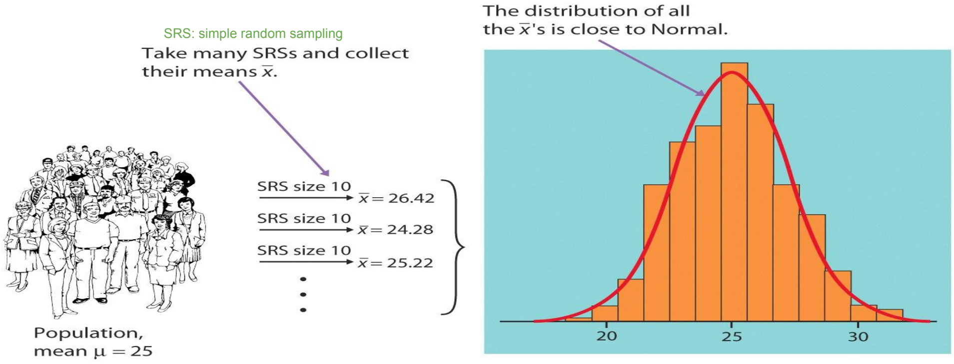
E.g.: average blood pressure of a sample of 50 Saudi men

E.g.: the difference in average blood pressure between a sample of 50 men and a sample of 50 women

Sampling Variability – If we could repeat an experiment many, many times on different samples with the same number of subjects, the resultant sample statistic would not always be the same (because of chance!). Remember: measure of sampling variability is nothing but standard error

Standard Error – a measure of the sampling variability

436 TEAMS 436 Research Team



We have different samples and different values but some are close to the population and some are away so how much the variabilities among the different sample means? This is nothing but standard error

Standard error of the mean(very important in exam)

- Standard error of the mean (sem): $s_{\bar{x}} = sem = \frac{s}{\sqrt{n}}$

You will be give a scenario and asked to calculate standard error or confidence interval (see below)

- Comments:
  - ✓ n = sample size
  - ✓ even for large s, if n is large, we can get good precision for sem
  - ✓ always (SE) smaller than standard deviation (s) Because we are dividing standard deviation by the root of sample size
  - ✓ If the sample size is small standard error will increase
  - ✓ Standard error is always positive.

In a representative sample of 100 observations of heights of men, drawn at random from a large population, suppose the sample mean is found to be 175 cm (sd=10cm) .
- Can we make any statements about the population mean ?

- We cannot say that population mean is 175 cm because we are uncertain as to how much sampling fluctuation has occurred.
- What we do instead is to determine a range of possible values for the population mean, with 95% degree of confidence.
- This range is called the 95% confidence interval and can be an important adjuvant to a significance test.

❖ **In the example, n =100 ,sample mean = 175, S.D., =10, and the S. Error =10/√100 = 1.**

Using the general format of confidence interval : **Statistic ± confidence factor x Standard Error of statistic**

Therefore, the 95% confidence interval is,

175 ± 1.96 * 1 = 173 to 177” 1.96 is Z value

That is, if numerous random sample of size 100 are drawn and the 95% confidence interval is computed for each sample, the population mean will be within the computed intervals in 95% of the instances.

That means if someone repeat this exercise 100 times 95 of the mean sample hight values will fall in this interval (173-177)

- If exercise repeated 100 times on same sample -> 95 times will be 173-177 almost in this range.
- another interpretation: I’m 95% confident that population mean hight is 173-177.

This is called "forest plot”

1 doesn't cross population mean and 19 cross = 95% .
So this is the concept of confidence interval here we repeated 20 times and 19 of them contain population value only one time doesn't contain.

6.4 Smaller samples 53

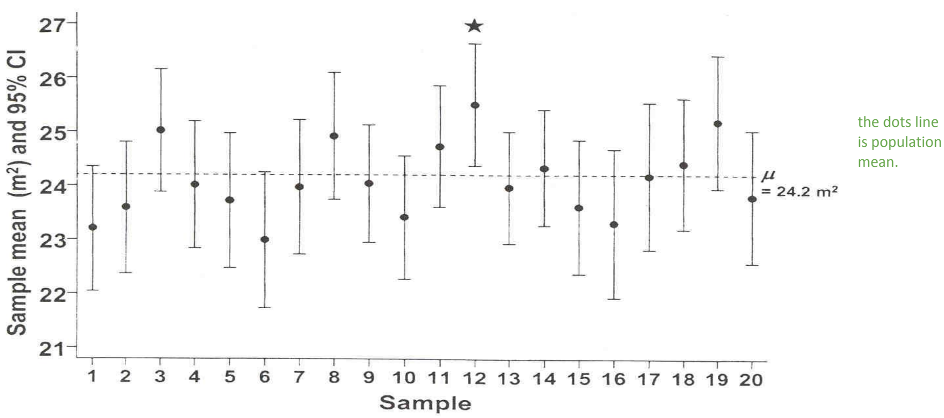


Fig. 6.2 Mean sprayable areas, with 95% confidence intervals, from 20 samples of 100 houses in a rural area. The star indicates that the CI does not contain the population mean

Confidence intervals

- The previous picture shows 20 confidence intervals for μ (ميو).
- Each 95% confidence interval has fixed endpoints, where μ might be in between (or not).
- There is no probability of such an event!

Confidence intervals

- Suppose α =0.05, we cannot say: "with probability 0.95 the parameter μ lies in the confidence interval."
- We only know that by repetition, 95% of the intervals will contain the true population parameter (μ)
- In 5 % of the cases however it doesn't. And unfortunately we don't know in which of the cases this happens.
- That's why we say: with confidence level 100(1 − α) % μ lies in the confidence interval."

Different Interpretations of the 95% confidence interval

- "We are 95% sure that the TRUE parameter value is in the 95% confidence interval"
- "If we repeated the experiment many many times, 95% of the time the TRUE parameter value would be in the interval"

Most commonly used CI:

- CI 90% corresponds to α 0.10
- CI 95% corresponds to α 0.05
- CI 99% corresponds to α 0.01

Note:

- p value → only for analytical studies You have 2 groups and you make comparison
- CI → for descriptive and analytical studies In descriptive you are not comparing anything and you are not having any hypotheses testing

How to calculate CI

General Formula: **CI = p ± Zα x SE**

- p = point of estimate, a value drawn from sample (a statistic)
- Zα = standard normal deviate for α, if α = 0.05 → Zα = 1.96 (~ 95% CI)

Example 1

100 KKUH students → 60 do daily exercise (p=0.6)

❖ What is the proportion of students do daily exercise in the KSU ? So I want to know what is the parameter value of KSU students and to get that I calculate CI

$$SE(p)CI = \sqrt{\frac{pq}{n}}$$

$$95\%CI = 0 \cdot 6 \pm 1.96 \sqrt{\frac{0.6x0.4}{100}}$$

0.6 = who exercise
0.4 = who don't exercise

$$0.6 \pm 1.96x0.5\backslash 10$$

$$0.6 \pm 0.1 = 0.5; 0.7$$

Statement 1: we are 95% confident that the population parameter of KSU students who exercise will be between 50% to 70%
Statement 2 : if someone reported this study 100 times then 95 times the values will be 50% to 70%

Example 2: CI of the mean

100 newborn babies, mean BW = 3000 (SD = 400) grams

❖ what is 95% CI?   95% CI = $\bar{x} \pm 1.96$ (SEM)

$$SEM = SD/\sqrt{n}$$

$$95\%CI = 3000 \pm 1.96\,(\frac{400}{\sqrt{100}})$$

$$= 3000 \pm 80 = (3000 - 80)\,;(3000 + 80)$$

$$= 2920; 3080$$

Interpretation : we are 95% confident that the population parameter of newborns will be anywhere between these two values

Examples 3: CI of difference between proportions (p1-p2) **we are comparing here two groups**

- 50 patients with drug A, 30 cured (p1=0.6)
- 50 patients with drug B, 40 cured Better (p2=0.8)

30/50 = 0.6
40/50 = 0.8
Difference = 0.8 - 0.6 = 0.2 = 20%

$$95\%CI(p_1 - p_2) = (p_1 - p_2) \pm 1.96xSE(p_1 - p_2)$$

$$SE(p_1 - p_2) = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

$$= \sqrt{\frac{(0.6 \times 0.4)}{50} + \frac{(0.8 \times 0.2)}{50}} = \sqrt{0.008} = 0.09$$

$$\Rightarrow 95\%CI(p_1 - p_2) = [0.2 - (0.09 * 1.96)];[0.2 + (0.09 * 1.96)]$$

$$= 0.024, 0.3764 = 2.4\% to 37.6\%$$

Is the difference of 20% statically significant ? The null hypothesis is 0 so are they getting 0 in this example?
Minimum is 2.4 and maximum is 37.6 so at any case we won't get 0 so this study is statically significant.
Interpretation: if someone do this type of study 100 times then 95 times will be between 2.4% to 37.6%
Worst scenario they'll get 2.4 not zero.

Example 4: CI for difference between 2 means

Mean systolic BP:

50 smokers	= 146.4 (SD 18.5) mmHg
50 non-smokers	= 140.4 (SD 16.8) mmHg
→ $\bar{x}_1 - \bar{x}_2$	= 6.0 mmHg
95% CI ($\bar{x}_1 - \bar{x}_2$)	= ($\bar{x}_1 - \bar{x}_2$) ± 1.96 x SE ($\bar{x}_1 - \bar{x}_2$)
SE ($\bar{x}_1 - \bar{x}_2$)	= S x $\sqrt{(1/n_1 + 1/n_2)}$

$$s = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 \; - 2)}}$$

$$s = \sqrt{\frac{(49 \ * 18.6) + 49 \ * 16.2}{98}} = 17.7$$

$$SE(\bar{x}_1 - \bar{x}_2) = 17.7 \times \sqrt{\frac{1}{50} + \frac{1}{50}} = 3.53$$

Point estimate

$$95\%CI = 6.0 \pm \left(1.96 \times 3.53\right) = -1.0; 13.0$$

Is this study (comparing between smokers and non-smokers) statically significant? No
Because it includes 0.
So if someone repeat he may get -1 , he may get 0 or he may get 13 so 0 is included so it's not statically significant.

436 TEAMS
436 Research Team

Other commonly supplied CI

Relative risk	(RR)
Odds ratio	(OR)
Sensitivity, specificity	(Se, Sp)
Likelihood ratio	(LR)
Relative risk reduction	(RRR)
Number needed to treat	(NNT)

CHARACTERISTICS OF CI'S

- The (im) precision (Reliability, repeatability, reproducibility) of the estimate is indicated by the width of the confidence interval.
- The wider the interval the less precision

The width of C.I. Depends on:

- SAMPLE SIZE
- VARIABILITY
- DEGREE OF CONFIDENCE

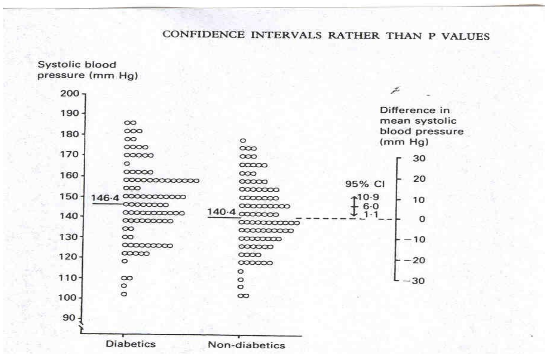


FIG 2.1—Systolic blood pressures in 100 diabetics and 100 non-diabetics with mean levels of 146·4 and 140·4 mm Hg respectively. The difference between the sample means of 6·0 mm Hg is shown to the right together with the 95% confidence interval from 1·1 to 10·9 mm Hg.

Normal distribution bell shaped curve.
Is it statically significant ? Yes because it does not include 0.
Upper limit: 10.9 / difference: 6 / lower limit: 1.1

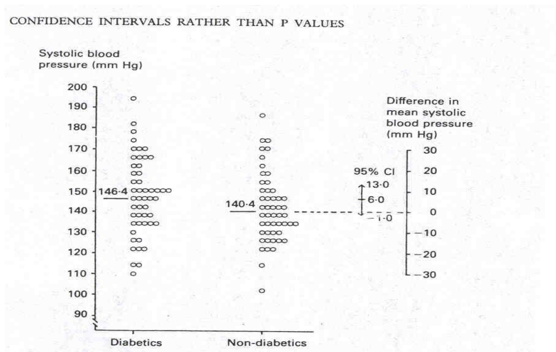


FIG 2.2—As fig 2.1 but showing results from two samples of half the size—that is, 50 subjects each. The means and standard deviations are as in fig 2.1, but the 95% confidence interval is wider, from −1·0 to 13·0 mm Hg, owing to the smaller sample sizes.

By reducing the sample size to half the bell shape is gone, high variability even though the accurate mean difference is same, statistical significance disappeared and the CI increase width
n= 50
So when we decrease the sample size the width will increase

EFFECT OF VARIABILITY

- Properties of error

1. Error increases with smaller sample size
   For any confidence level, large samples reduce the margin of error

2. Error increases with larger standard Deviation
   As variation among the individuals in the population increases, so does the error of our estimate

3. Error increases with larger z values
   Tradeoff between confidence level and margin of error

Not only 95%….

90% confidence interval:

NARROWER than 95% $\bar{x} \pm 1.65 sem$

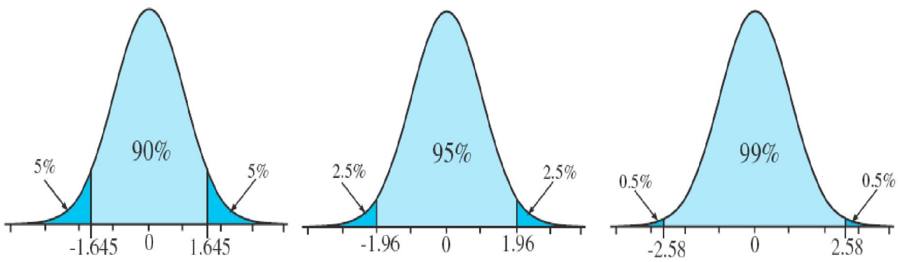
99% confidence interval:

WIDER than 95% $\bar{x} \pm 2.58 sem$

Or we can say: 95% is narrower than 99%. (نفس المعنى بس طريقة الصياغه غير)

Common Levels of Confidence

Confidence level 1 – α	Alpha level α	*Z* value $z_{1-(\alpha/2)}$
.90	.10	1.645
.95	.05	1.960
.99	.01	2.576



As the alpha value and Confidence level increase the width of CI increases

CONFIDENCE INTERVALS RATHER THAN P VALUES

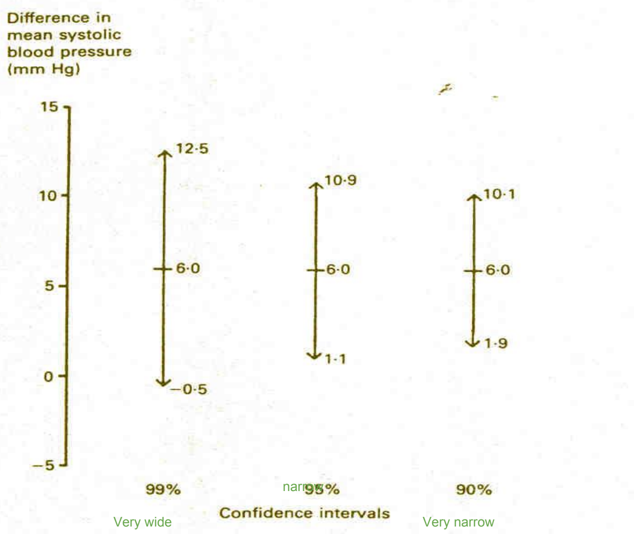


FIG 2.3—Confidence intervals associated with differing degrees of "confidence" using the same data as in fig 2.1.

Application of confidence intervals

Example: The following finding of non-significance in a clinical trial on 178 patients.

- Chi-square value = 1.74 (p > 0.1) (non –significant)

i.e. there is no difference in efficacy between the two treatments.

- The observed difference is: 75% - 66% = 9%

and the 95% confidence interval for the difference is: - 4% to 22%

- This indicates that compared to treatment B, treatment A has, at best an appreciable advantage (22%) and at worst , a slight disadvantage (- 4%).
- This inference is more informative than just saying that the difference is non significant.

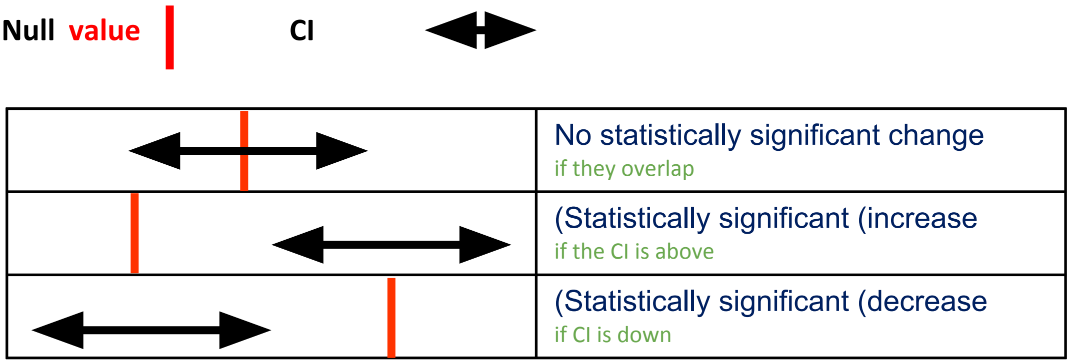
Treatment	Success	Failure	Total
A	76 (75%)	25	101
B	51(66%)	26	77
Total	127	51	178

Interpretation of Confidence intervals

- Width of the confidence interval (CI)
  - A narrow CI implies high precision
  - A wide CI implies poor precision (usually due to inadequate sample size*)
- Does the interval contain a value that implies no change or no effect or no association?
  - CI for a difference between two means: Does the interval include 0 (zero)? no significant difference
  - CI for a ratio (e.g, OR, RR): Does the interval include 1? no association

*so if the sample size increases the confidence interval will become more precise and narrower.

Interpretation of Confidence intervals Important



Duality between P-values and CIs

- If a 95% CI includes the null effect, the P-value is >0.05 (and we would fail to reject the null hypothesis)
- If the 95% CI excludes the null effect, the P-value is <0.05 (and we would reject the null hypothesis)

Interpreting confidence intervals

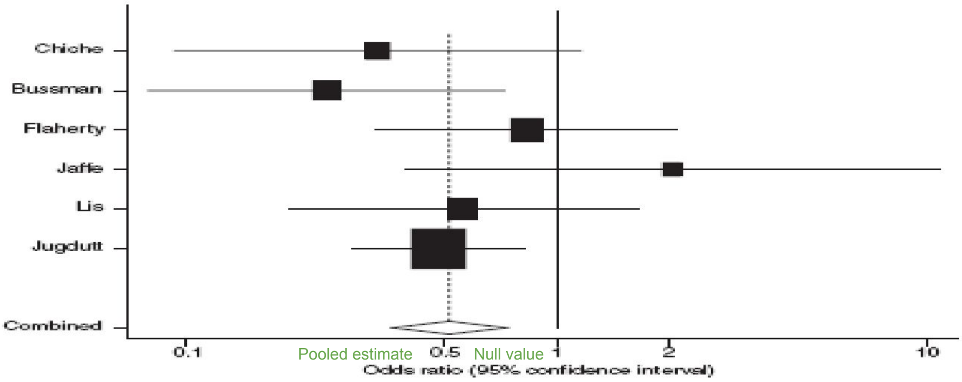
Trial	Number dead / randomized: Intravenous nitrate	Number dead / randomized: Control	Risk Ratio	95% C.I.	P value
Chiche	3/50	8/45	0.33	(0.09,1.13)	0.08
Wide interval: suggests reduction in mortality of 91% and an increase of 13%					
Flaherty	11/56	11/48	0.83	(0.33,2.12)	0.70
Jaffe	4/57	2/57	2.04	(0.39,10.71)	0.40
Reduction in mortality as little as 18%, but little evidence to suggest that IV nitrate is harmful					
Jugdutt	24/154	44/156	0.48	(0.28, 0.82)	0.007

1-0.09=0.91=91%

- Non-statically significance because RR less than 1
- Remember: in RR we look for 1 not 0

Table adapted from Whitley and Ball. Critical Care; 6(3):222-225, 2002

Figure 1 Forest plot



Individual and combined odds ratios and 95% confidence intervals for six intravenous nitrate trials.

There are 6 studies 4 of them reported that there is no effect IV nitrate in acute myocardial infarction patients in reduction of mortality. While 2 studies reported that IV nitrate is effective in reducing mortality. So there is controversy so they did systematic review and mixed everything and got the Pooled estimate. Pooled estimate doesn't cross the null value. So pooled estimate is showing that IV nitrate is not harmful it's making a good effect in reducing the mortality

Which of the following odds ratios for the relationship between various risk factors and heart disease are statistically significant at the .05-significance level? Which are likely to be clinically significant?

	Statistically significant?	**Clinically significant?**
A. Odds ratio for every 1-year increase in age: 1.10 (95% CI: 1.01—1.19)	Not include 1 ✓	✓
B. Odds ratio for regular exercise (yes vs. no): 0.50 (95% CI: 0.30—0.82)	Not include 1 ✓	✓
C. Odds ratio for high blood pressure (high vs. normal): 3.0 (95% CI: 0.90—5.30) Interpretation of odds ratio OR: the odds of having a heart disease is 3 times more in patients who does not maintain their blood pressure in comparison to those who do maintain it.		✓
D. Odds ratio for every 50-pound increase in weight: 1.05 (95% CI: 1.01—1.20)	✓	

Comparison of p values and confidence interval

- p values (hypothesis testing) gives you the probability that the result is merely caused by chance or not by chance*, it does not give the magnitude and direction of the difference
- Confidence interval (estimation) indicates estimate of value in the population given one result in the sample, it gives the magnitude and direction of the difference

*So if I do a study and obtain a result: if $p < 0.05$, it means the probability of getting this result by chance is **less than** 5%.

P-values versus Confidence intervals

- P-value answers the question...
  - "Is there a statistically significant difference between the two treatments?“ (or two groups)
- The point estimate and its confidence interval answers the question...
  - "What is the size of that treatment difference?", and "How precisely did this trial determine or estimate the treatment difference?"

Summary of key points

- A P-value is a probability of obtaining an effect as large as or larger than the observed effect, assuming null hypothesis is true
  - Provides a measure of strength of evidence against the Ho
  - Does not provide information on magnitude of the effect
  - Affected by sample size and magnitude of effect: interpret with caution!
- Confidence interval quantifies
  - How confident are we about the true value in the source population
  - Better precision with large sample size
  - Much more informative than P-value
- Keep in mind clinical importance when interpreting statistical significance!