Statistical significance using Confidence Intervals

Dr.Shaik Shaffi Ahamed Ph.d.,
Professor
Dept. of Family & Community Medicine

Learning Objectives

(1)Able to understand the concept of confidence intervals.

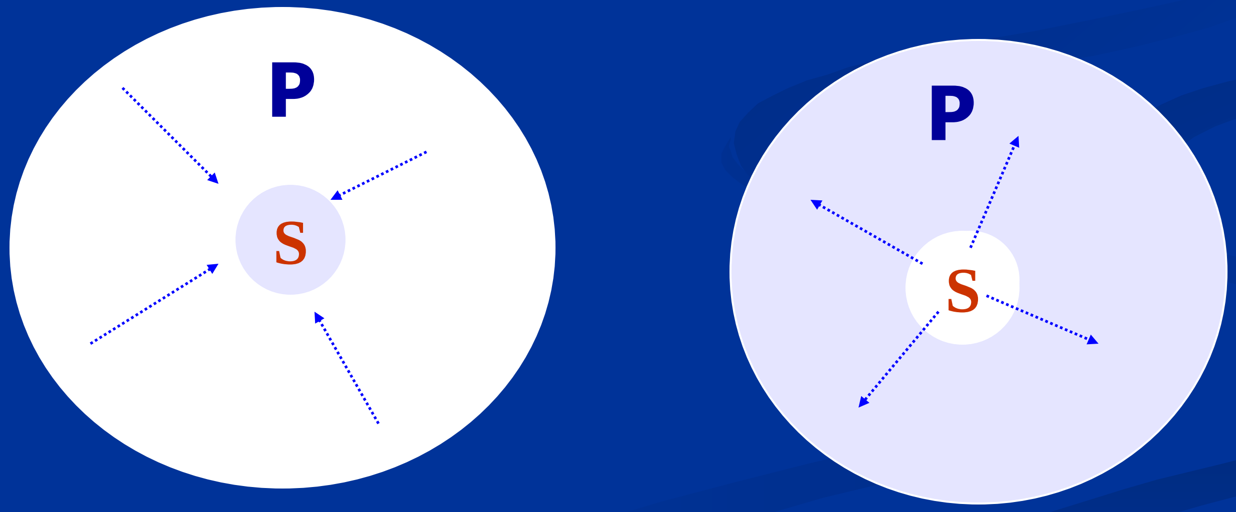
(2)Able to apply the concept of statistical significance using confidence intervals in analyzing the data.

(3)Able to interpret the concept of 95% confidence intervals in making valid conclusions.

HYPOTHESIS TESTING(p value)
&
ESTIMATION (Confidence Interval)

Sample is assumed to be representative to the population.

In research: measurement are <u>always</u> done in the sample, the results will be applied to population.



Statistic and Parameter

- 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

Estimation

Two forms of estimation

- **Point estimation** = single value, e.g., (mean, proportion, difference of two means, difference of two proportions, OR, RR etc.,)
- **Interval estimation** = range of values ⇒ **confidence interval (CI).** A confidence

Confidence intervals

- P values give no indication about the clinical importance of the observed association
- 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

Computing confidence intervals (CI)

- General formula:

(Sample statistic) $\pm$ [(confidence level) $\times$ (measure of how high the sampling variability is)]

- Sample statistic: 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, $Z_{\alpha/2}=1.96$

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

- Don’t get confuse with the terms of

STANDARD DEVEIATION

and

STANDARD ERROR

Example:

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

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

Mean:

$$\bar{X} = \frac{\sum X}{N} = \frac{42}{6} = 7$$

Variance:

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

Standard Deviation:

$$s = \sqrt{s^2} = \sqrt{4.67} = 2.16$$

Statistical Inference is based on Sampling Variability

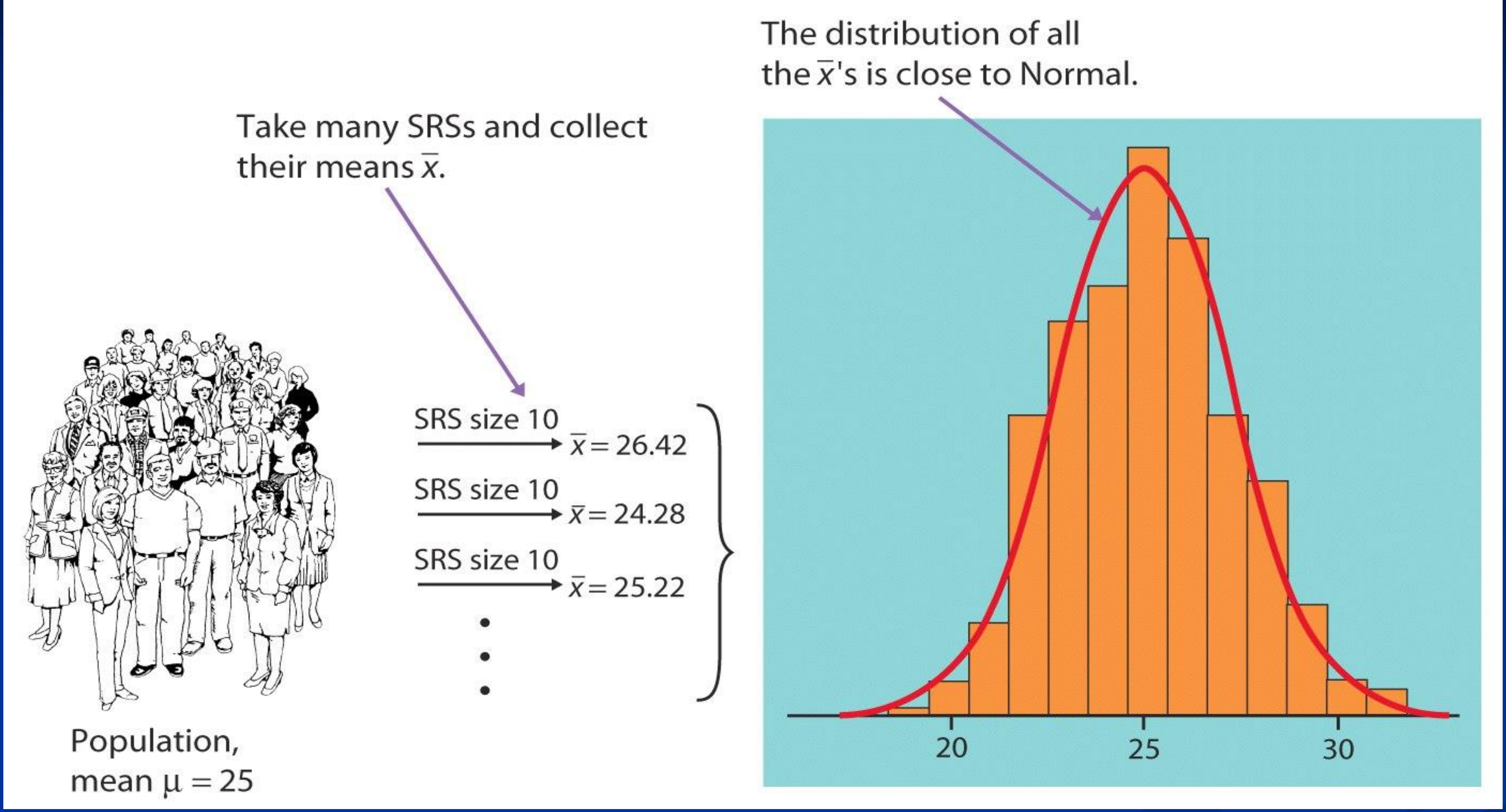
Sample Statistic – we summarize a sample into one number; e.g., could be a mean, a difference in means or proportions, an odds ratio, or a correlation coefficient

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!).

Standard Error – a measure of the sampling variability



The distribution of all
the $\bar{x}$'s is close to Normal.
Take many SRSs and collect
their means $\bar{x}$.
SRS size 10
$\bar{x} = 26.42$
SRS size 10
$\bar{x} = 24.28$
SRS size 10
$\bar{x} = 25.22$
20
25
30
Population,
mean $\mu = 25$

STANDARD ERROR OF THE MEAN

- Standard error of the mean (sem):

$$s_{\bar{x}} = sem = \frac{s}{\sqrt{n}}$$

- Comments:
  - n = sample size
  - even for large s, if n is large, we can get good precision for sem
  - always smaller than standard deviation (s)

-- **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,

$$175 \pm 1.96 * 1 = 173 \text{ to } 177”$$

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.

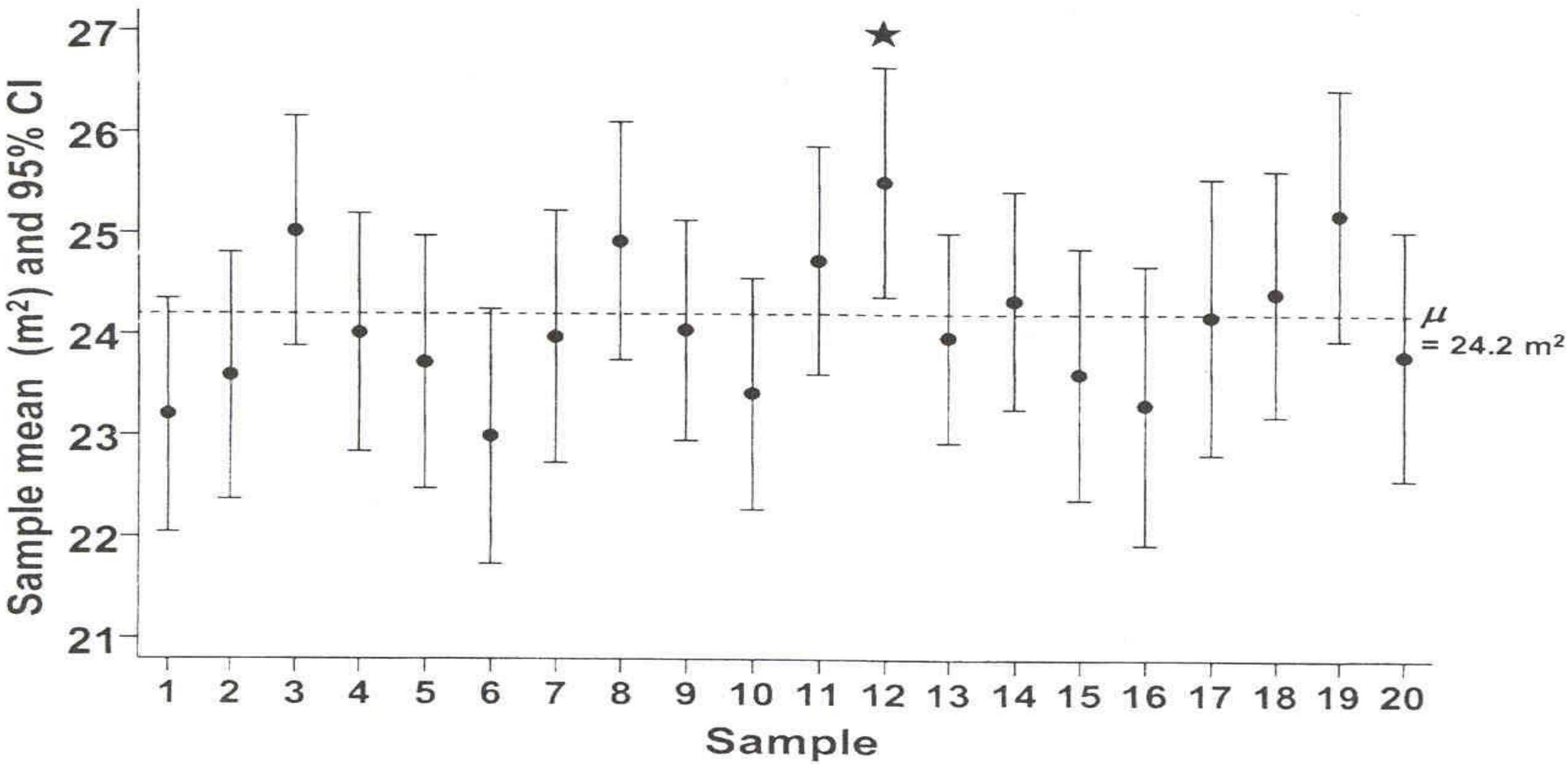


Fig. 6.2 Mean sprayable areas, with 95% confidence intervals, from 20 samples of 100 houses in a rural area. The star indicates that the CI does not contain the population mean

Confidence intervals

- The previous picture shows 20 confidence intervals for μ.
- 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 $\alpha = 0.05$, 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 $100(1 - \alpha)$ % 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 p 0.10
- CI 95% corresponds to p 0.05
- CI 99% corresponds to p 0.01

Note:

- p value → only for analytical studies
- CI → for descriptive and analytical studies

How to calculate CI

General Formula:

$$CI = p \pm Z\alpha \times SE$$

- p = point of estimate, a value drawn from sample (a statistic)
- $Z\alpha$ = standard normal deviate for α, if $\alpha = 0.05$ → $Z\alpha = 1.96$ (~ 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 ?

Example 1

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

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

$$= 0.6 \pm 1.96x\,0.5/10$$

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

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)

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

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

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

$$= 2920; 3080$$

Examples 3: CI of difference between proportions (p1-p2)

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

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

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

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

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

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

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$(\bar{x}_1 - \bar{x}_2)$ = $(\bar{x}_1 - \bar{x}_2) \pm 1.96 \times SE(\bar{x}_1 - \bar{x}_2)$

SE$(\bar{x}_1 - \bar{x}_2)$ = $S \times \sqrt{(1/n_1 + 1/n_2)}$

Example 4: CI for difference between 2 means

$$s = \sqrt{\frac{(n_1 - 1)s_1^{\,2} + (n_2 - 1)s_2^{\,2}}{(n_1 + n_2 - 2)}}$$

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

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

$$95\%CI = 6.0 \pm (1.96X3.53) = -1.0;13.0$$

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 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

---- VAIRABILITY

---- DEGREE OF CONFIDENCE

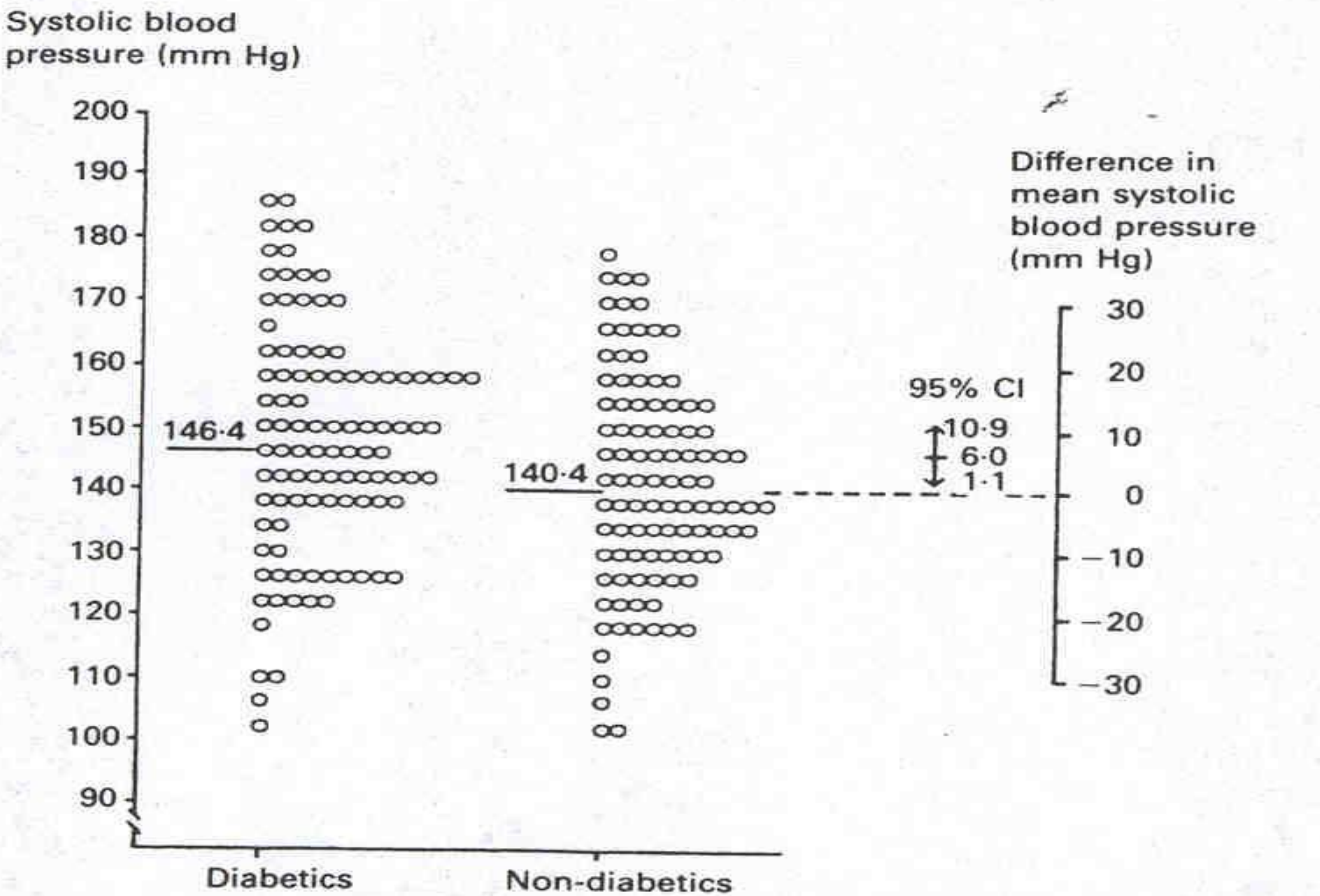


FIG 2.1—Systolic blood pressures in 100 diabetics and 100 non-diabetics with mean levels of 146·4 and 140·4 mm Hg respectively. The difference between the sample means of 6·0 mm Hg is shown to the right together with the 95% confidence interval from 1·1 to 10·9 mm Hg.

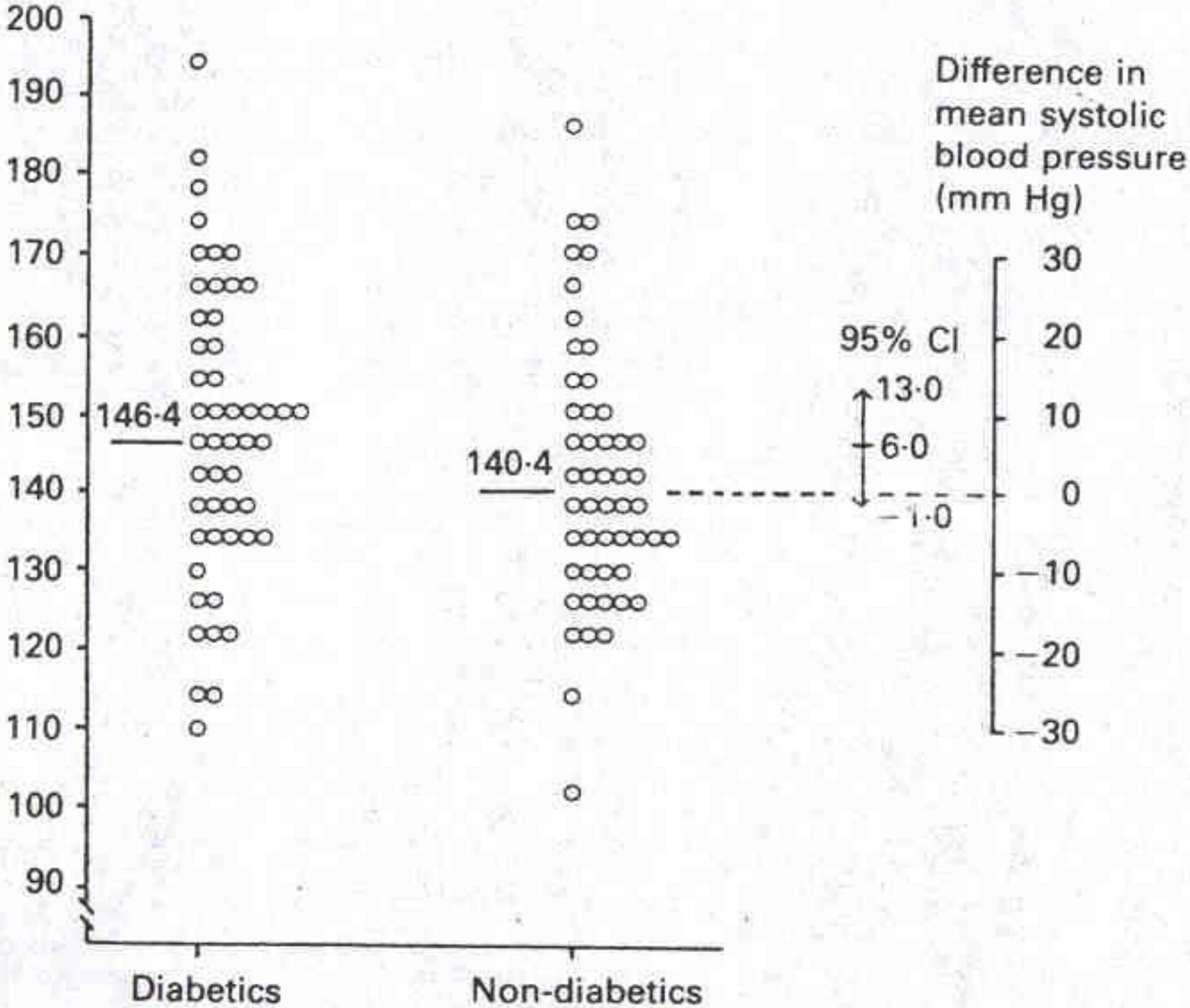


FIG 2.2—As fig 2.1 but showing results from two samples of half the size—that is, 50 subjects each. The means and standard deviations are as in fig 2.1, but the 95% confidence interval is wider, from −1·0 to 13·0 mm Hg, owing to the smaller sample sizes.

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%

$$\bar{x} \pm 1.65 sem$$

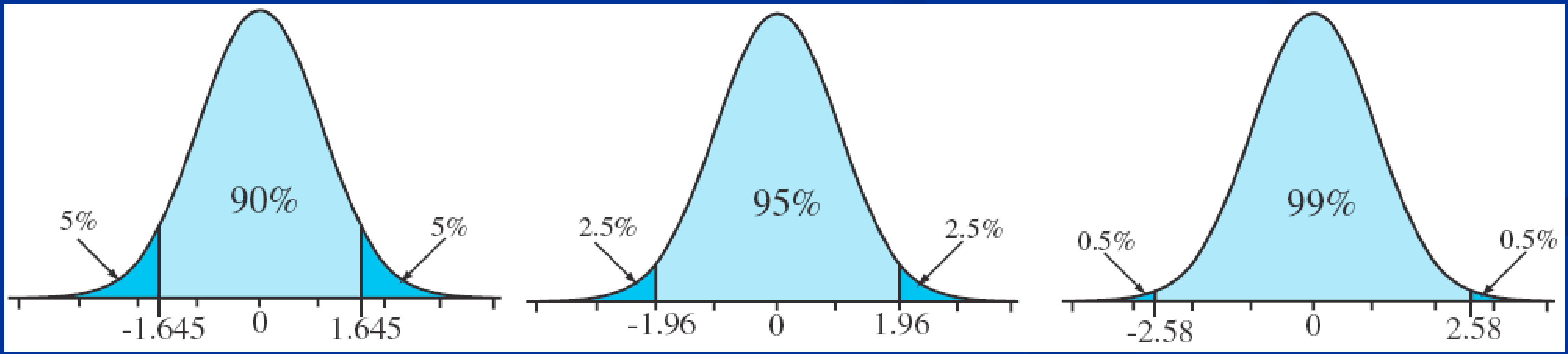
- 99% confidence interval:

  WIDER than 95%

$$\bar{x} \pm 2.58 sem$$

Common Levels of Confidence

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



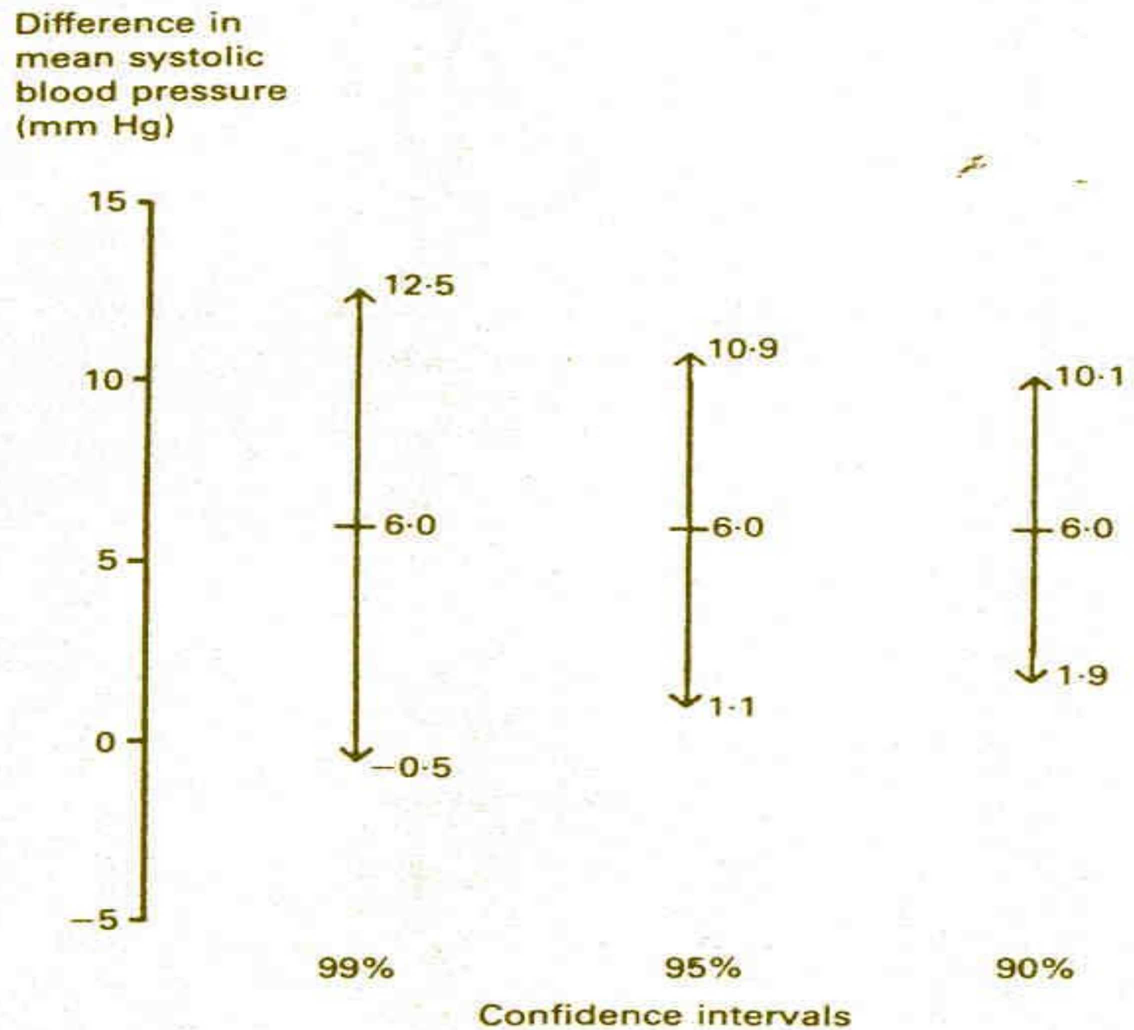


FIG 2.3—Confidence intervals associated with differing degrees of "confidence" using the same data as in fig 2.1.

APPLICATION OF CONFIDENCE INTERVALS

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

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

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%).

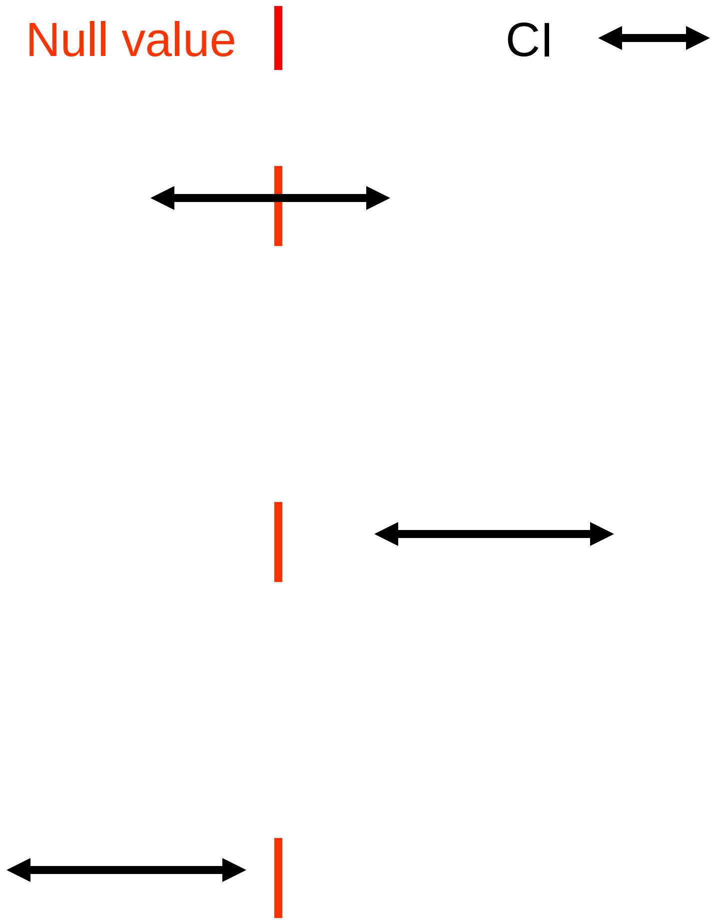
--- This inference is more informative than just saying that the difference is non significant.

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)?
  - CI for a ratio (e.g, OR, RR): Does the interval include 1?

Interpretation of Confidence intervals



No statistically significant change

Statistically significant (increase)

Statistically significant (decrease)

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%

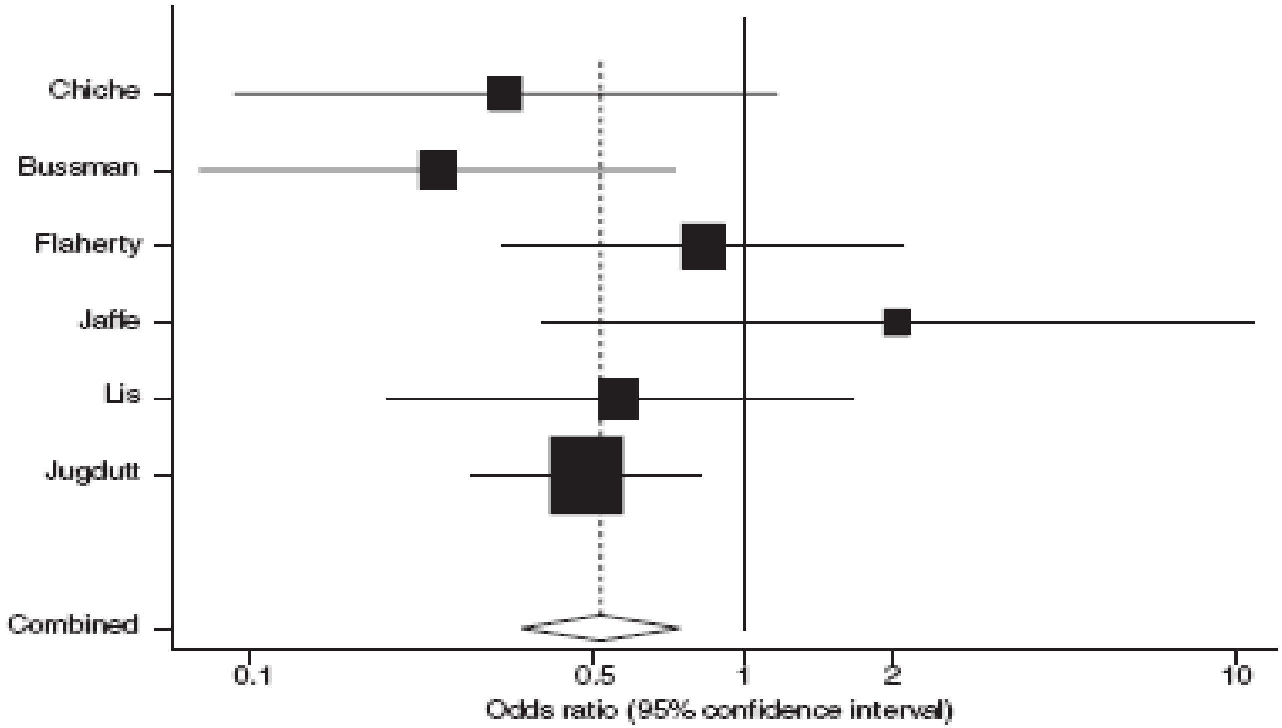
Trial	Intravenous nitrate	Control	Risk Ratio	95% C.I.	P value
Flaherty	11/56	11/48	0.83	(0.33,2.12)	0.70
Jaffe	4/57	2/57	2.04	(0.39,10.71)	0.40

Reduction in mortality as little as 18%, but little evidence to suggest that IV nitrate is harmful

Trial	Intravenous nitrate	Control	Risk Ratio	95% C.I.	P value
Jugdutt	24/154	44/156	0.48	(0.28, 0.82)	0.007

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

Figure 1



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

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)	✓	✓
B. Odds ratio for regular exercise (yes vs. no): 0.50 (95% CI: 0.30—0.82)	✓	✓
C. Odds ratio for high blood pressure (high vs. normal): 3.0 (95% CI: 0.90—5.30)		✓
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

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 H_o
  - Does not provide information on magnitude of the effect
  - Affected by sample size and magnitude of effect: interpret with caution!

Summary of key points

- 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!

Thanks