Clinical vs Statistical Significance

Mehdi Moradinazar

Dept. of Epidemiology, Kermanshah University of Medical Sciences, Iran

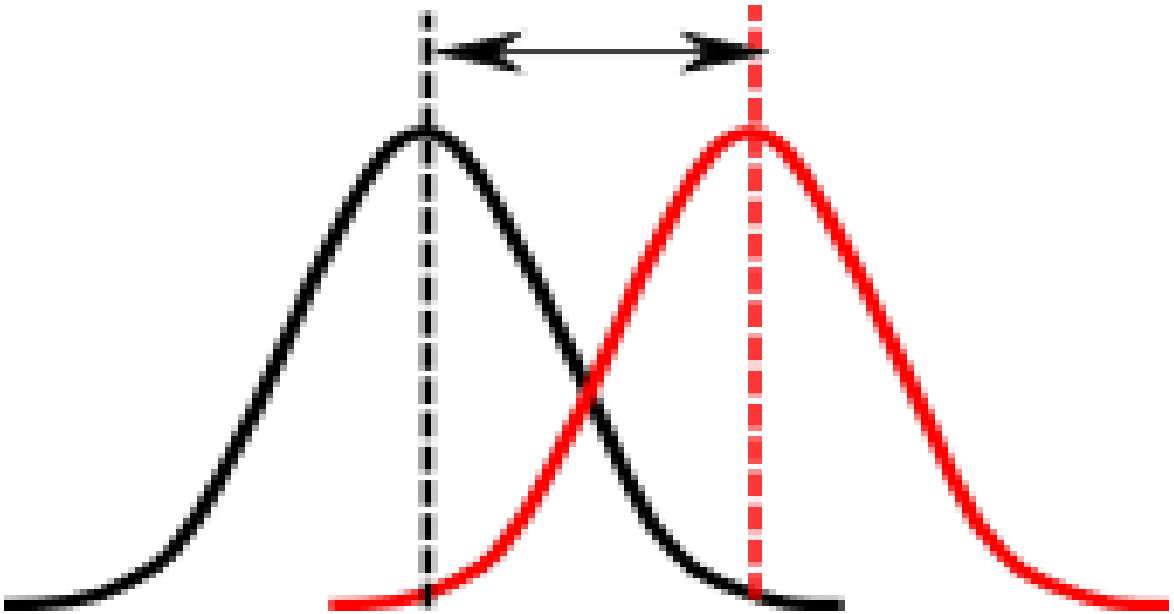
m.moradinazar@gmail.com
m.moradinazar@kums.ac.ir

دانشگاه علوم پزشکی و خدمات بهداشتی درمانی کرمانشاه

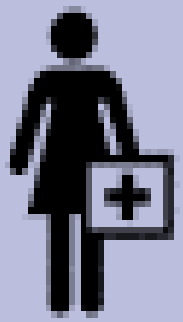
Specific objectives of this session:

- ***Statistical Significance***
  - ✓ ***P values***
  - ✓ ***Confidence interval***
  - ✓ ***Problems with Statistical Significance***
  - ✓ ***Problems with confidence interval***
- ***Clinical Significance***
  - ✓ ***Minimal clinically important difference***
  - ✓ ***What Is Effect Size?***
  - ✓ ***Problems with Clinical significance***
  - ✓ ***Calculation of clinical significance***

Is the mean value different between the two groups in a meaningful way?



Statistical vs Clinical Significance

Statistical Significance	Clinical Significance
Is this difference unlikely?	Is this difference important?
$p < .05$	?
	
Does it matter to a statistician?	Does it matter to a clinician?

Statistically significant

- Statistically significant indicates that the observed results are not likely due to random variability or chance alone.
- In statistical terms, it means that there is enough evidence to support the conclusion that the results are not merely a fluke but rather reflect a **true effect or relationship** in the data being studied.

Statistical Significance

- Statistical significance is a statement about the likelihood of findings being **due to chance**.

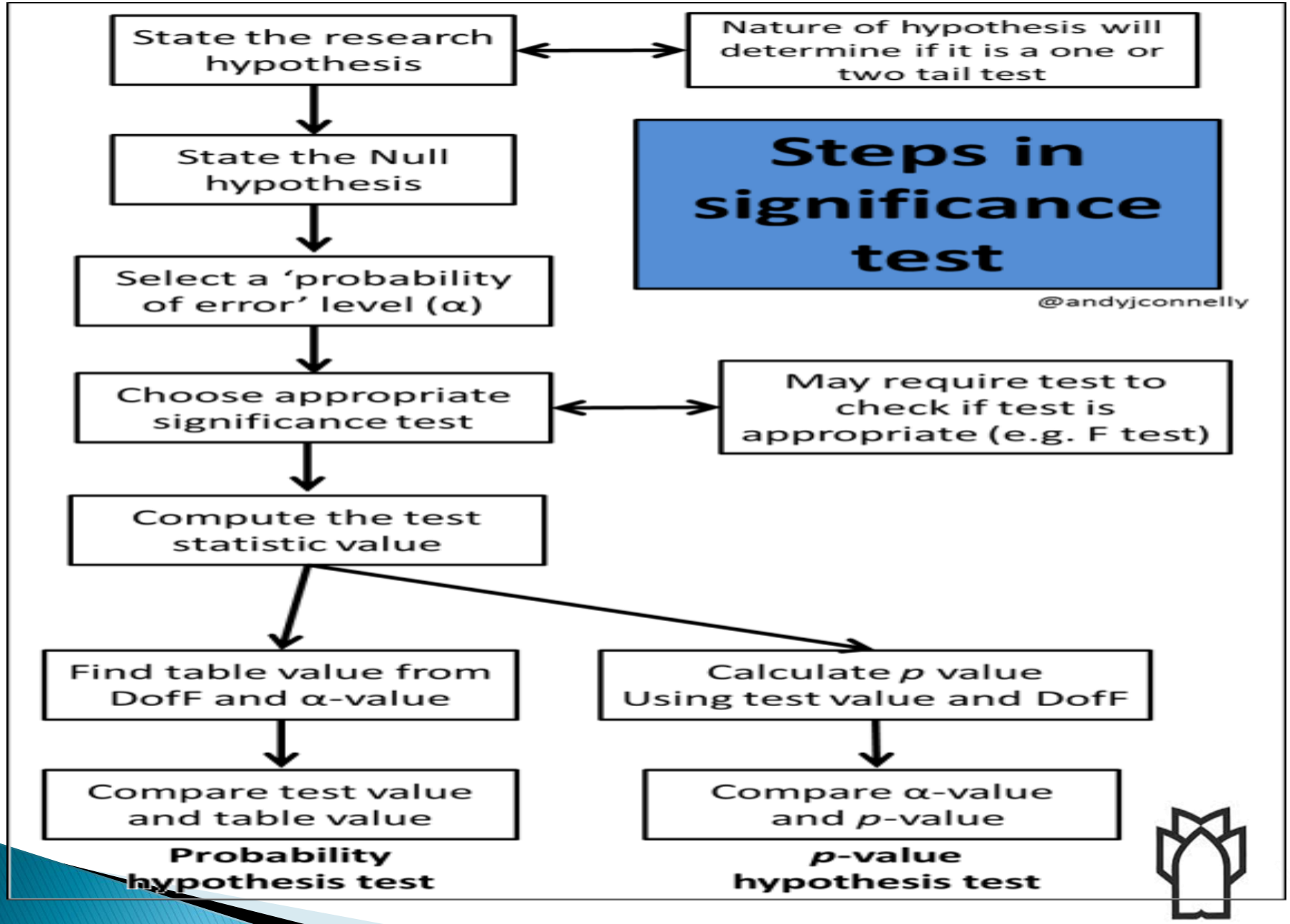
- Statistical significance is a determination that a relationship between two or more variables is caused by something other than **chance or randomly**.

- A level of significance is selected (most commonly $\alpha = 0.05$ or 0.01), which signifies the probability of incorrectly rejecting a true null hypothesis.

Alpha value

- The number alpha is the threshold value that we measure p-values against.
- It tells us how extreme observed results must be in order to reject the null hypothesis of a significance test.

- But in fact, the origin of this threshold is arbitrary, and in practice designs, often have lower or higher false-positive thresholds depending on design features such as adjustments for multiple comparisons or phase of development, respectively.


State the research hypothesis
Nature of hypothesis will determine if it is a one or two tail test
State the Null hypothesis
Steps in significance test
@andyjconnelly
Select a 'probability of error' level (α)
Choose appropriate significance test
May require test to check if test is appropriate (e.g. F test)
Compute the test statistic value
Find table value from DofF and α-value
Calculate p value Using test value and DofF
Compare test value and table value
Compare α-value and p-value
Probability hypothesis test
p-value hypothesis test

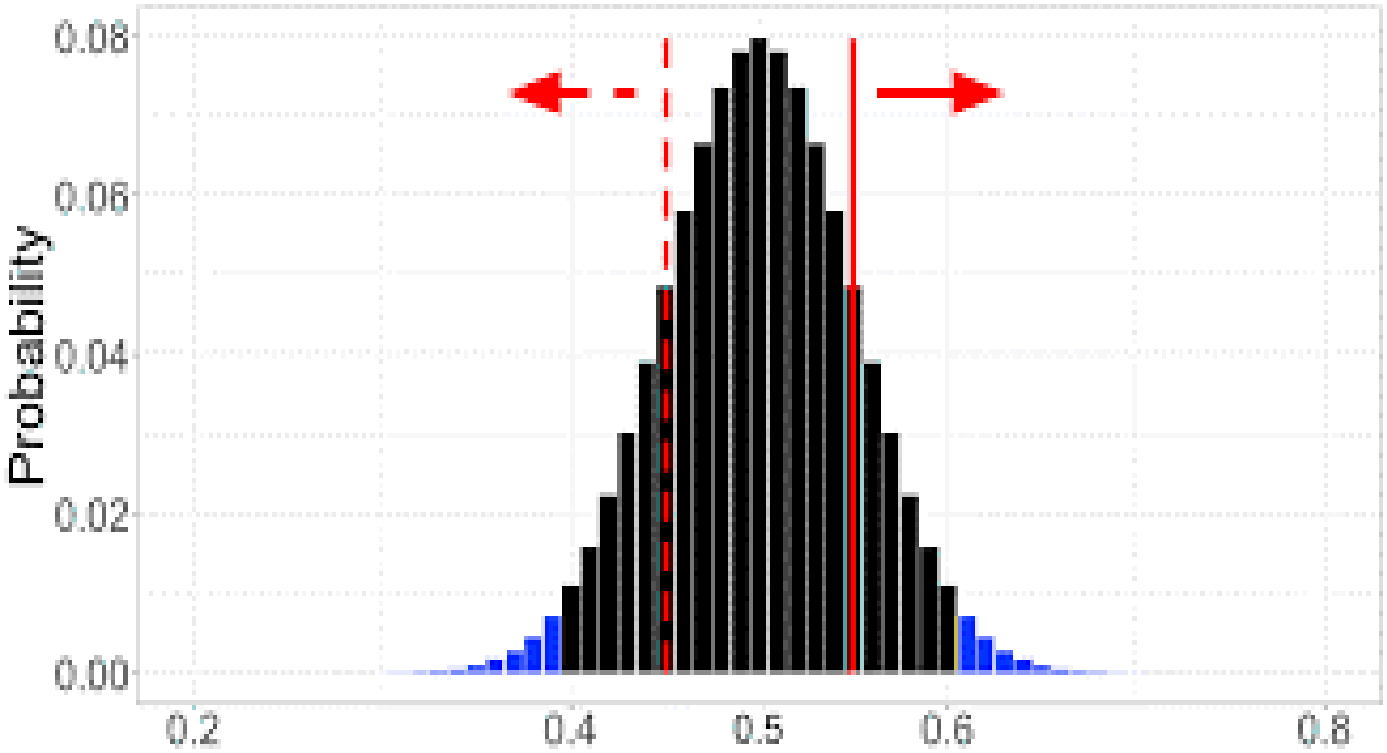
Hypothesis

- The null hypothesis states that there is **no relationship** between the two variables being studied (one variable does not affect the other).

	H0 is true	**H0 is false**
Reject H0	null hypothesis is true & was rejected (type I error) α	null hypothesis is false & was rejected (**correct conclusion**)
Accept H0	null hypothesis is true & was accepted (**correct conclusion**)	null hypothesis is false & was accepted (type II error) β

Probability value(p value)

- The p values should be interpreted as the probability that the study results (as good as or better than observed) occurred by chance when the null hypothesis is true.



P value

- the *P* value will not reveal the size of the effect.
- P values are considered to be confounded because of their dependence on sample size.
- Sometimes a statistically significant result means only that a huge sample size was used.

Statistically significant

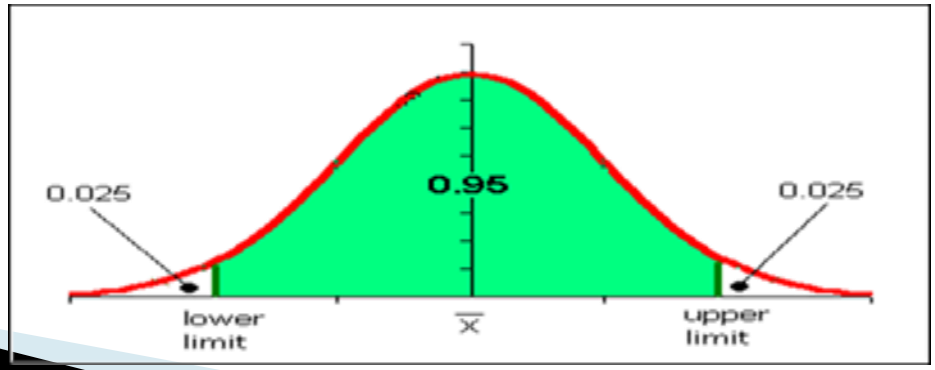
- Statistically significant
  - ✓It does not definitively prove that the null hypothesis is false. Instead, it suggests that the data provide enough evidence to cast doubt on the null hypothesis and support the alternative hypothesis, which posits that there is a significant effect or relationship.
  - ✓It does not provide absolute certainty about the truth or falsity of the null hypothesis.
- Non significant
  - ✓Results do not prove that the null hypothesis is true; they also give ***no evidence*** of the truth or falsity of the hypothesis the researcher has generated.
  - ✓There are two possible explanations for Non significant:
    - No mean difference
    - The sample size was not large enough

Problems with Statistical Significance

- 0.05 is arbitrary.
- The P value is not a probability of anything in reality.
- Statistically significant does not necessarily mean that the effect is real/true.
- One in 20 may be by chance and will mislead (type I error).
- Non significant does not mean no effect. Small studies will often report non significance even when the difference is real & important (type II error)
- **Two solutions:** *clinical significance* via *confidence interval*

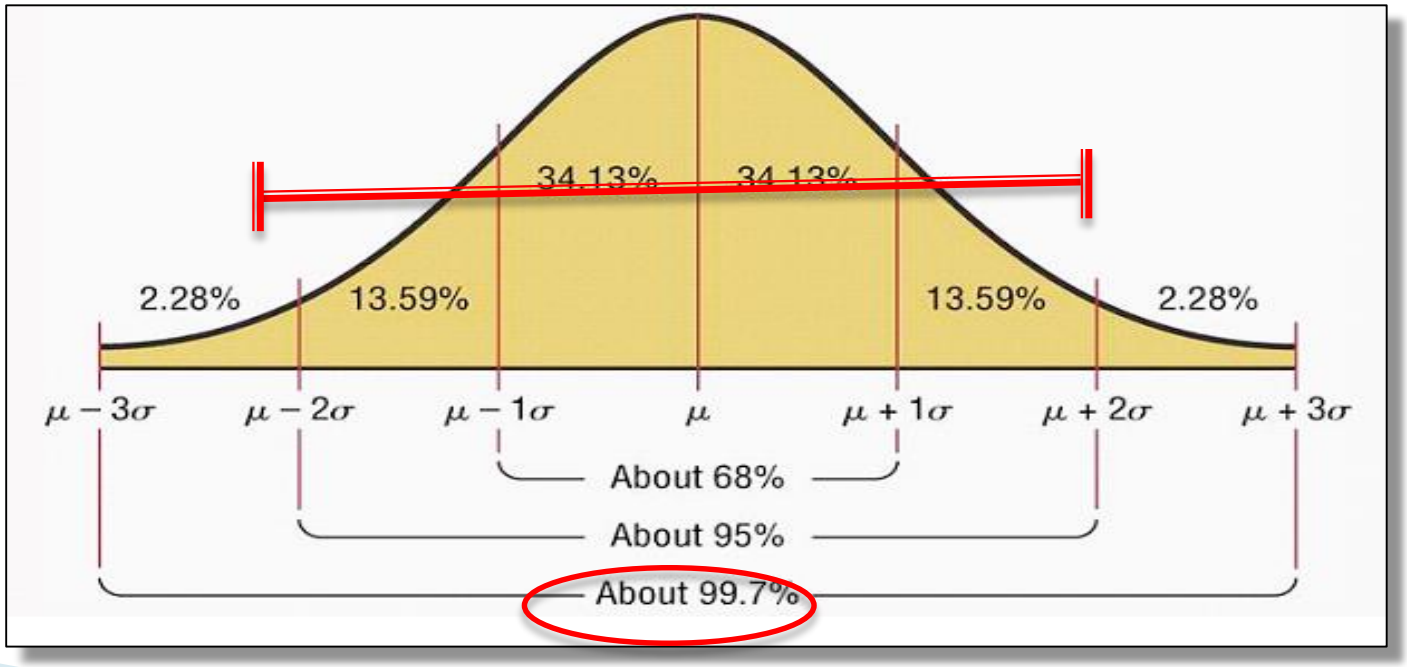
Confidence interval

- Confidence intervals are calculated from the same equations that generate p-values, so there is a relationship with P
- Confidence interval conveys more information than P values. It provides magnitude of effect as well as its variability.
- Confidence interval should be calculated for each variable especially if P values are insignificant
- Width of confidence interval is associated with sample size

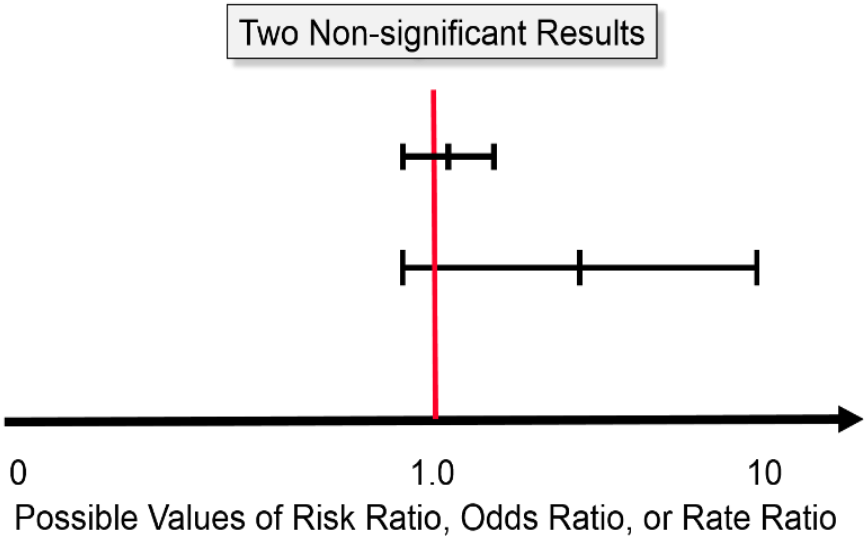


Confidence intervals

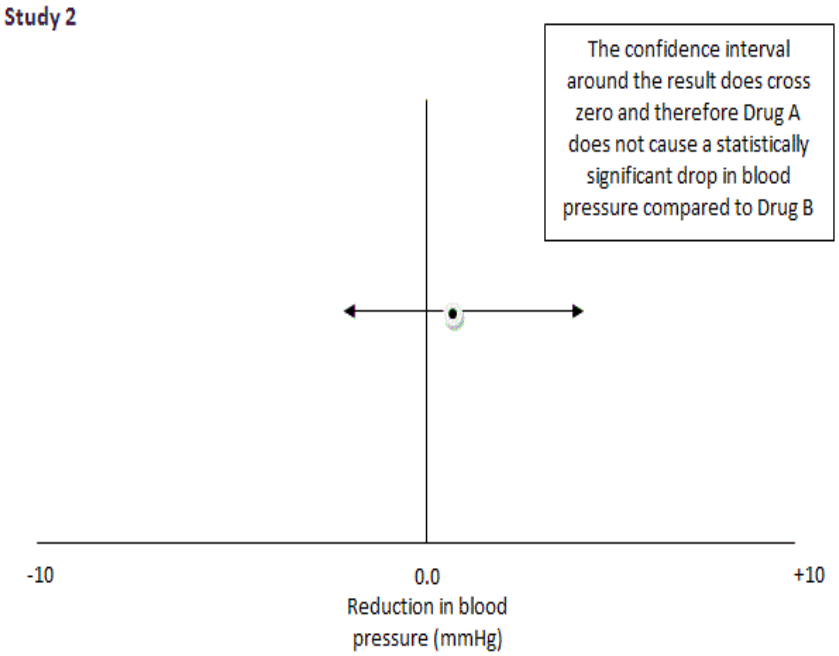
When you see a 95% confidence interval for a mean, think of it like this: if we'd collected 100 samples and calculated the mean for each sample, than for 95 of these samples the mean would fall within the confidence interval.



Confidence intervals significant



OR=3.5
95%CI: (0.8_10.2)



Mean difference: 2.7
95%CI: (-3.2_+5.8)

Clinical Significance

- Clinical significance refers to the changes due to treatment that are practically meaningful for a client.

- Statistical significance is denoted by p-values whereas practical significance is represented by effect sizes.

Methods to detect clinical significance

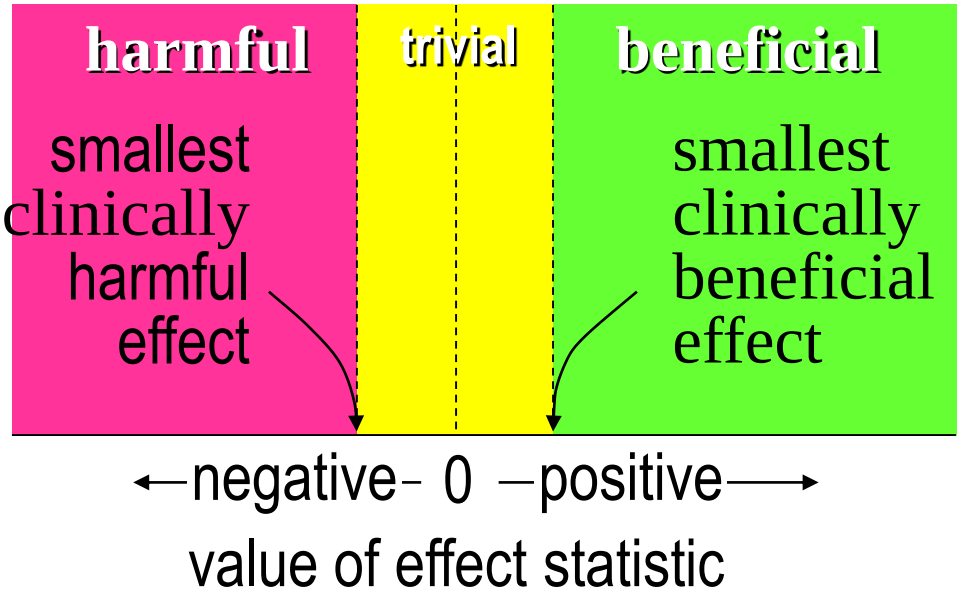
- **Minimal Clinically Important Difference (MCID):**
  - ✓ This method determines the smallest change in a treatment outcome that is perceived as beneficial by patients or healthcare providers.
- **Effect Size:**
  - ✓ Effect size measures the strength of a relationship between two variables, such as the treatment and the outcome. Cohen's d is a commonly used effect size measure.

$$d = \frac{M_2 - M_1}{\sqrt{\frac{SD_1^2 + SD_2^2}{2}}}$$

- **Clinical Judgment:**
  - ✓ Healthcare professionals may use their clinical expertise to assess whether a change in a particular outcome is meaningful or significant for a patient.
- **Patient-Reported Outcomes:**
  - ✓ Patient-reported outcome measures (PROMs) involve directly asking patients about their symptoms, quality of life, or functional status to determine clinical significance from the patient's perspective.

Clinical significance

- In clinical significance, we need to interpret confidence limits in relation to the smallest clinically beneficial and harmful effects.
  - These are usually **equal and opposite** in sign.
  - They define **regions** of beneficial, trivial, and harmful values.



Methods of determining the MID

- what's the smallest clinically important effect?

- What is meaningful may be subjective and may depend on the context.

- There are several techniques to calculate the MID. They fall into three categories:
  - distribution-based methods,
  - anchor-based methods
  - the Delphi method.

Common Methods for Determining Clinical Significance			
Method	Developed by (year)	Formula	Description
Jacobson-Truax Method (JT)	Jacobson, Follette, & Revenstorf (1984), revised by Jacobson & Truax (1991)	$\frac{(X_{post} - X_{pre})}{(2[S_{pre}(1-r_{xx})^{0.5}])^{0.5}}$	Determines cutoff points and Reliability Change Index (RCI)
Gulliksen-Lord-Novick Method (GLN)	Hsu (1999)	$\frac{[X_{post} - M_{pop}] - r_{xx}[X_{pre} - M_{pop}]}{S_{pop}(1 - r^2_{xx})^{0.5}}$	Alters JT by factoring in hypothesized group means
Edward-Nunnally Method (EN)	Speer (1992)	$[r_{xx}(X_{pre} - M_{pre}) + M_{pre}] \pm 2S_{pre}(1 - r_{xx})^{0.5}$	Alters JT by placing true score on a confidence interval
Hageman-Arrindell Method (HA)	Hageman & Arrindell (1999)	$\frac{(X_{post} - X_{pre})r_{dd} + (M_{post} - M_{pre})(1-r_{dd})}{((r_{dd})^{0.5})((2S^2_E)^{0.5})}$	Alters JT by calculating clinical significance index and reliability of change index; can calculate individual or group change
Hierarchical Linear Model (HLM)	Speer and Greenbaum (1995)	$B^*/V^{*1/2}$	Uses growth curve models to determine clinical change

Calculation of clinical significance

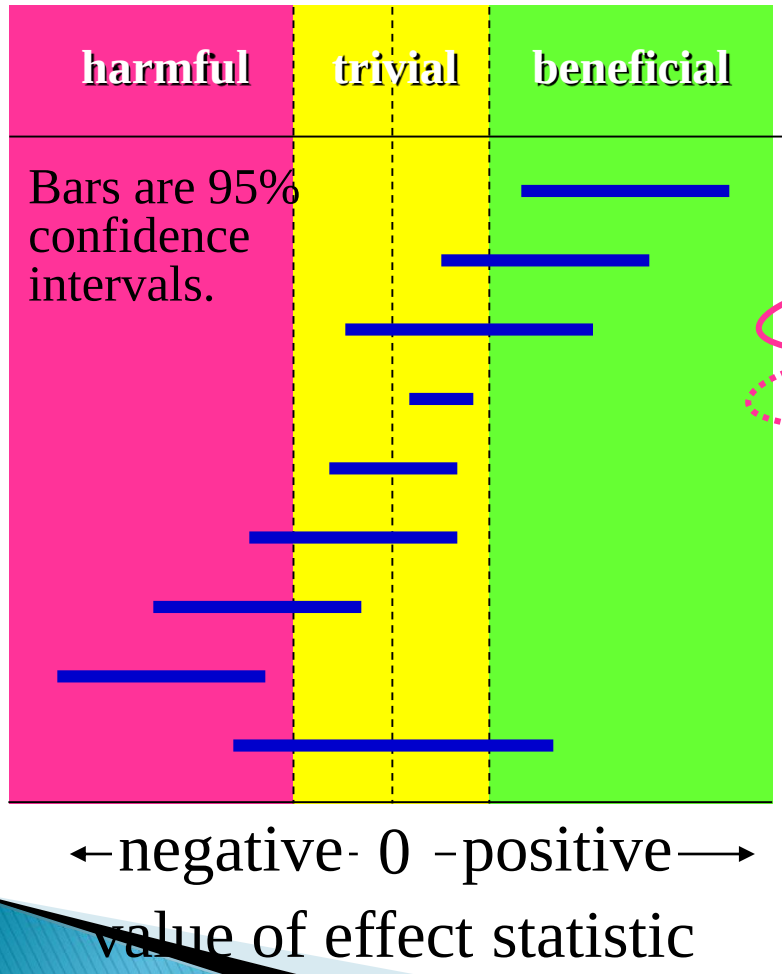
- there are many ways to calculate statistical significance and practical significance.

- Five common methods are
  - the Jacobson-Truax method
  - the Gulliksen-Lord-Novick method
  - the Edwards-Nunnally method
  - the Hageman-Arrindell method
  - hierarchical linear modeling

How to interoperate Effect Size

Index	Description[b]	Effect Size	Comments
Between groups			
Cohen's d[a]	$d = M_1 - M_2 / s$ $M_1 - M_2$ is the difference between the group means (M); s is the standard deviation of either group	Small 0.2 Medium 0.5 Large 0.8 Very large 1.3	Can be used at planning stage to find the sample size required for sufficient power for your study
Odds ratio (OR)	$\frac{\text{Group 1 odds of outcome}}{\text{Group 2 odds of outcome}}$ If OR = 1, the odds of outcome are equally likely in both groups	Small 1.5 Medium 2 Large 3	For binary outcome variables Compares odds of outcome occurring from one intervention vs another
Relative risk or risk ratio (RR)	Ratio of probability of outcome in group 1 vs group 2; If RR = 1, the outcome is equally probable in both groups	Small 2 Medium 3 Large 4	Compares probabilities of outcome occurring from one intervention to another
Measures of association			
Pearson's r correlation	Range, −1 to 1	Small ±0.2 Medium ±0.5 Large ± 0.8	Measures the degree of linear relationship between two quantitative variables
r^2 coefficient of determination	Range, 0 to 1; Usually expressed as percent	Small 0.04 Medium 0.25 Large 0.64	Proportion of variance in one variable explained by the other

- Putting the confidence interval and these regions together, we can make a decision about clinical significance.
  - ***Clinically decisive*** or ***clear*** is preferable to *clinically significant*.



Clinically decisive?	Statistically significant?
Yes: use it.	Yes
Yes: use it.	Yes
Yes: use it.	No
Yes: don't use it.	Yes
Yes: don't use it.	No
Yes: don't use it.	No
Yes: don't use it.	Yes
Yes: don't use it.	Yes
No: need more research.	No

Why statistical significance is impractical or unethical!

Interpreting the Probabilities

- You should describe outcomes in plain language in your paper.
- Therefore you need to describe the probabilities that the effect is beneficial, trivial, and/or harmful.
- Suggested scheme:

Probability	Chances	Odds	The effect… beneficial/trivial/harmful
<0.01	<1%	<1:99	is almost certainly not…
0.01–0.05	1–5%	1:99–1:19	is very unlikely to be…
0.05–0.25	5–25%	1:19–1:3	is unlikely to be…, is probably not…
0.25–0.75	25–75%	1:3–3:1	is possibly (not)…, may (not) be…
0.75–0.95	75–95%	3:1–19:1	is likely to be…, is probably…
0.95–0.99	95–99%	19:1–99:1	is very likely to be…
>0.99	>99%	>99:1	is almost certainly…

		Clinically Significant	
		Yes	**No**
Statistically Significant	**Yes**	Typically assume the groups, outcomes, or treatments **are different**	Consider that the sample size may be too large
	No	Consider that the sample size may be too small	Typically assume the groups, outcomes, or treatments **are not different**

In general

- Statistical significance and clinical significance will usually be in harmony in a well-designed clinical trial.
- Statistical significance is directly impacted by sample size.
- Clinicaly significance is not directly influenced by sample size.
- Statistical significance indicates the reliability of the study results
- clinical significance reflects its impact on clinical practice.

References

- Kieser M, Friede T, Gondan M. Assessment of statistical significance and clinical relevance. Statistics in medicine. 2013 May 10;32(10):1707-19.
- Houle TT, Stump DA. Statistical significance versus clinical significance. InSeminars in Cardiothoracic and Vascular Anesthesia 2008 Mar (Vol. 12, No. 1, pp. 5-6). Sage CA: Los Angeles, CA: SAGE Publications.
- Bhardwaj SS, Camacho F, Derrow A, Fleischer AB, Feldman SR. Statistical significance and clinical relevance: the importance of power in clinical trials in dermatology. Archives of dermatology. 2004 Dec 1;140(12):1520-3.
- van Rijn MH, Bech A, Bouyer J, van den Brand JA. Statistical significance versus clinical relevance. Nephrology Dialysis Transplantation. 2017 Apr 1;32(suppl_2):ii6-12.
- van Tulder M, Malmivaara A, Hayden J, Koes B. Statistical significance versus clinical importance: trials on exercise therapy for chronic low back pain as example. Spine. 2007 Jul 15;32(16):1785-90.
- Turk DC. Statistical significance and clinical significance are not synonyms!. The Clinical journal of pain. 2000 Sep 1;16(3):185-7.