Confidence Intervals

Confidence Intervals (CI)

- CIs represent a range of **statistically plausible results** consistent with an outcome from a single study
- Example: ARR = 5%, 95% CI (3% to 7%)
- Can be used for any measure of outcomes
- Confidence intervals have some practical limitations similar to P-values
- Although the CIs can project a range of results consistent with the study results, they cannot tell you the truth of the outcomes

Confidence Intervals (CI)

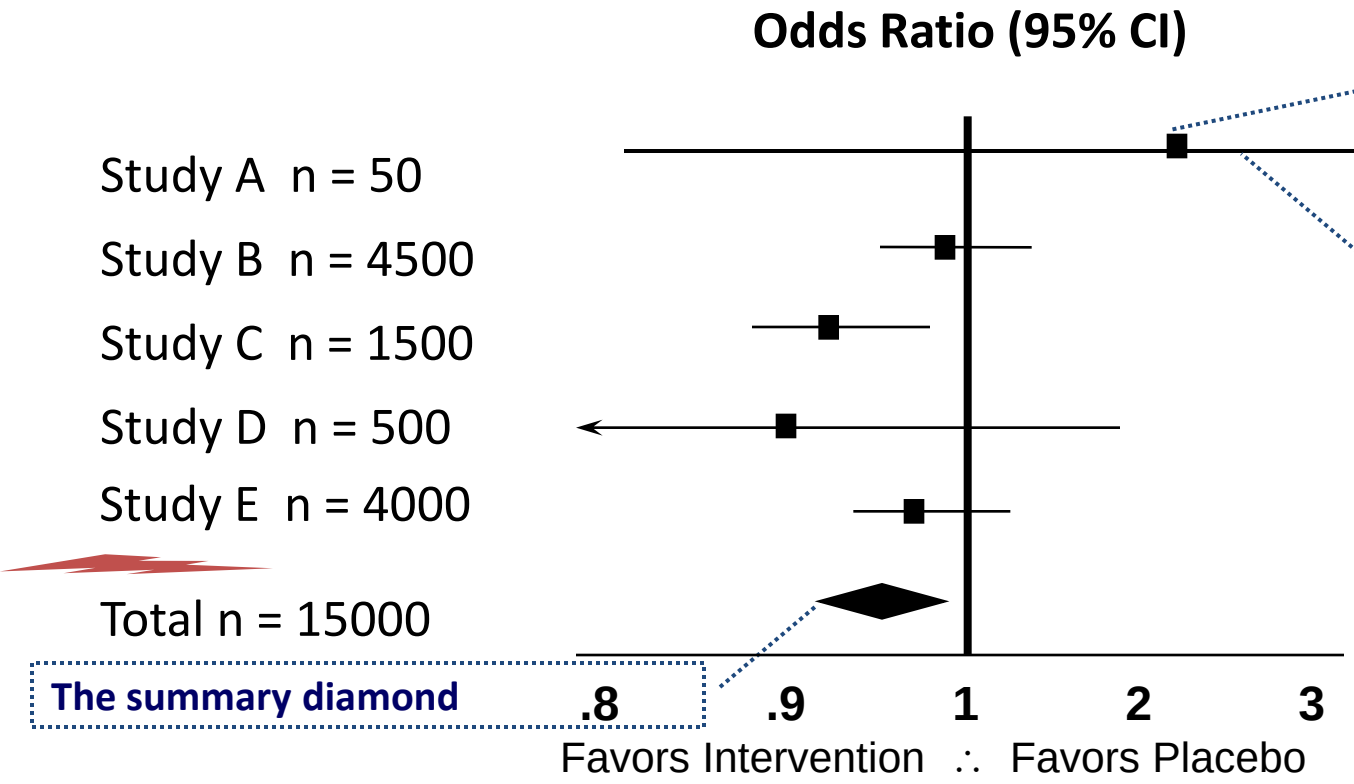
- We approach them as providing a possible range of plausible results for the larger population **IF the study results in the studied population are true**; however, point estimate is most likely to be right
- Affected by confidence level (e.g., 90% CI), sample size and effect size
  - Helps quantify **uncertainty**
  - Helps determine **meaningful clinical benefit**
  - Helps deal with conclusivity of **non-significant findings** (Type II or beta error)

But First! How to Read a Forest Plot

Virgin beech forest in Biogradska Gora, Montenegro
© Snežana Trifunović, 2007.

Graphic Display:
Point Estimate, CI and Summary Diamond

These are several studies reported in a meta-analysis (some studies are removed, so this is not meant to total correctly) — this is just a sampler.

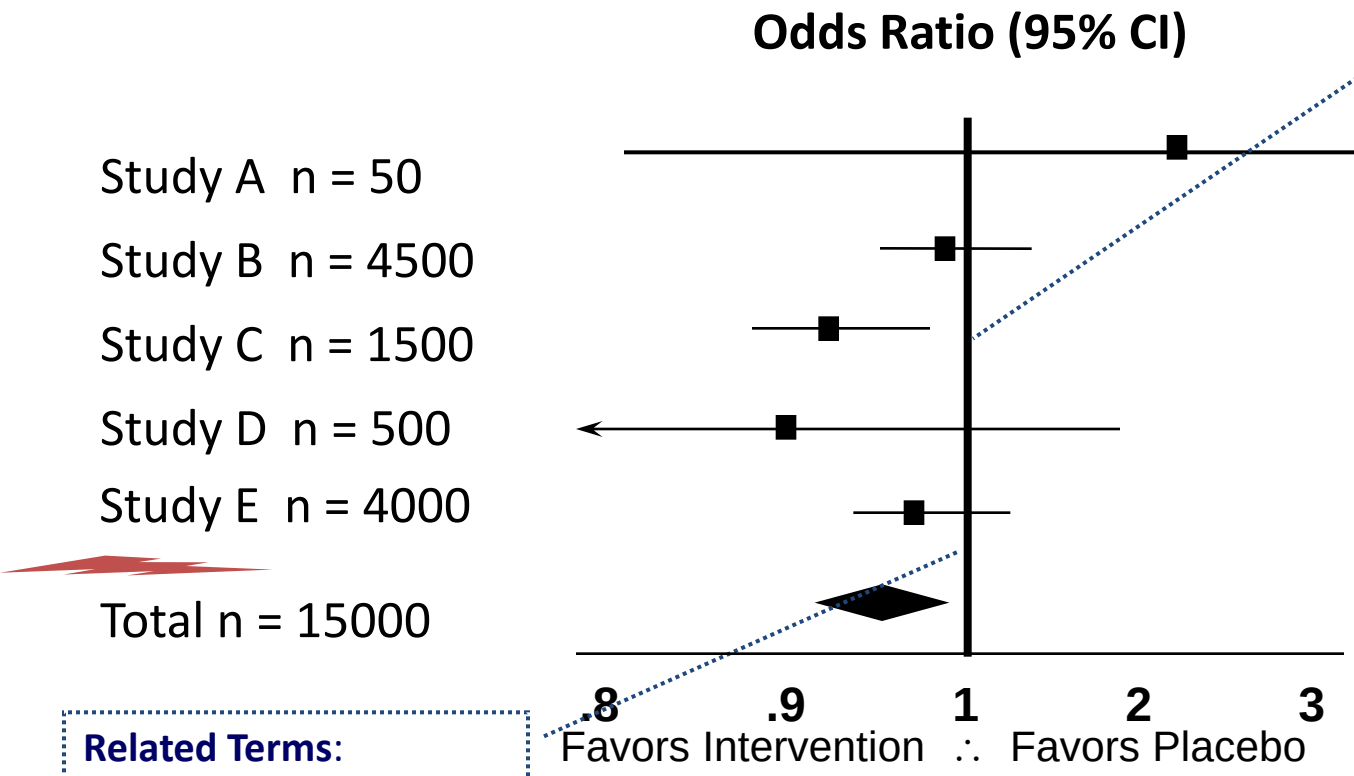


This **square** is the study result (ie, the **point estimate**)

This **line** is the **confidence interval** (ie, a statistically calculated range of equally plausible study results given a **margin for chance**)

The summary diamond

Favors Intervention ∴ Favors Placebo & The Line of No Difference



This **center line** is the **line of no difference**. Results to the right favor placebo in this example. Results to the left favor the intervention.

Related Terms:
- Line of no difference
- Line of no effect
- Infinity (NNT etc)
- Unity (ratios)

Non-Statistical Significance

Study A n = 50

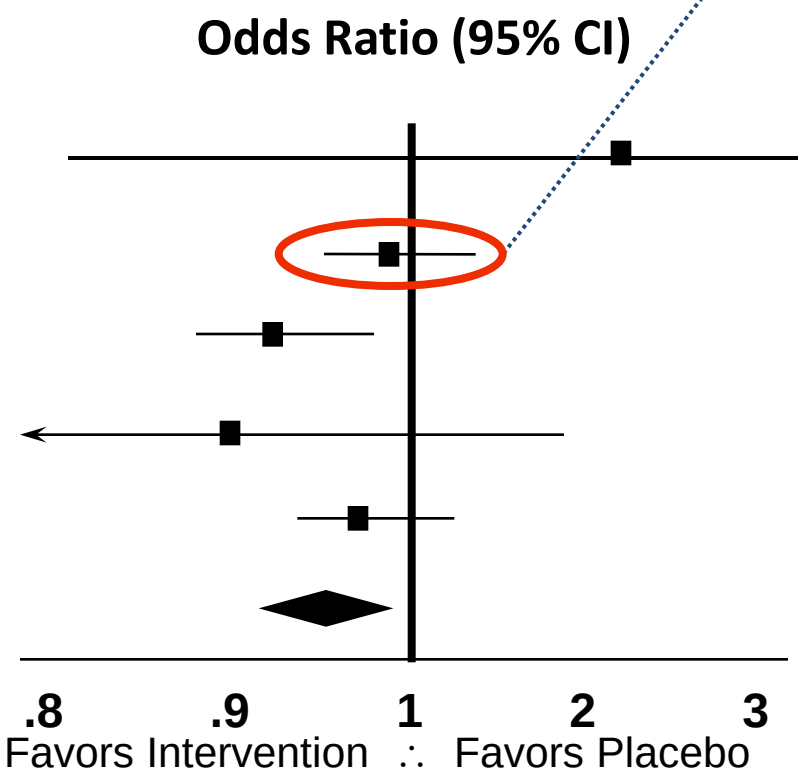
Study B n = 4500

Study C n = 1500

Study D n = 500

Study E n = 4000

Total n = 15000



Therefore, it is statistically plausible, within 95% certainty in a valid study, that Study B may favor the placebo or Study B may favor the intervention.

This is not possible. Thus, the results of Study B are **not statistically significant**.

Anything **touching this line** means the results are **not statistically significant** because it is not possible to favor both placebo and intervention.

Frequently CIs are Reported Numerically Only—
You need to determine the line of no difference to determine if result is statistically significant.

	Odds Ratio (95% CI)
Study A n = 50	2.12 (0.82, 20.91)
Study B n = 4500	0.98 (0.96, 1.39)
Study C n = 1500	0.93 (0.88, 0.97)
Study D n = 500	0.91 (0.73, 1.71)
Study E n = 4000	0.97 (0.94, 1.33)
Total n = 15000	0.96 (0.92, 0.98)

Sampler – not meant to add up

You Need to Know the Numerical Value for "No Difference"

Study A n = 50

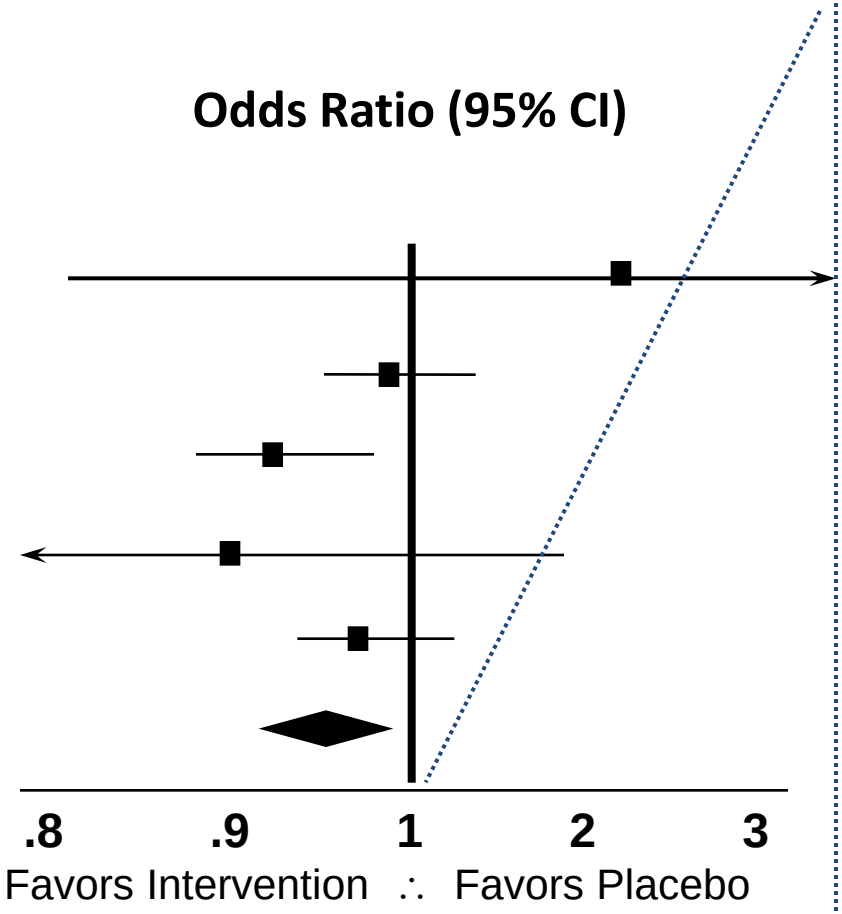
Study B n = 4500

Study C n = 1500

Study D n = 500

Study E n = 4000

Total n = 15000



The line of no difference equals 0 or 1 depending upon the measure of outcome used.

No difference for a **percent** is expressed as **zero**.

ARR and RRR are expressed as percentages. Therefore, if these were used, this number would be zero.

No difference for a **ratio** is 1:1. So for **odds ratio or relative risk ratio (aka relative risk),** this number **equals 1**.

Which of the studies below are statistically significant?

	Odds Ratio (95% CI)
Study A n = 50	2.12 (0.82, 20.91)
Study B n = 4500	0.98 (0.96, 1.39)
Study C n = 1500	0.93 (0.88, 0.97)
Study D n = 500	0.91 (0.73, 1.71)
Study E n = 4000	0.97 (0.94, 1.33)
Total n = 15000	0.96 (0.92, 0.98)

Sampler – not meant to add up

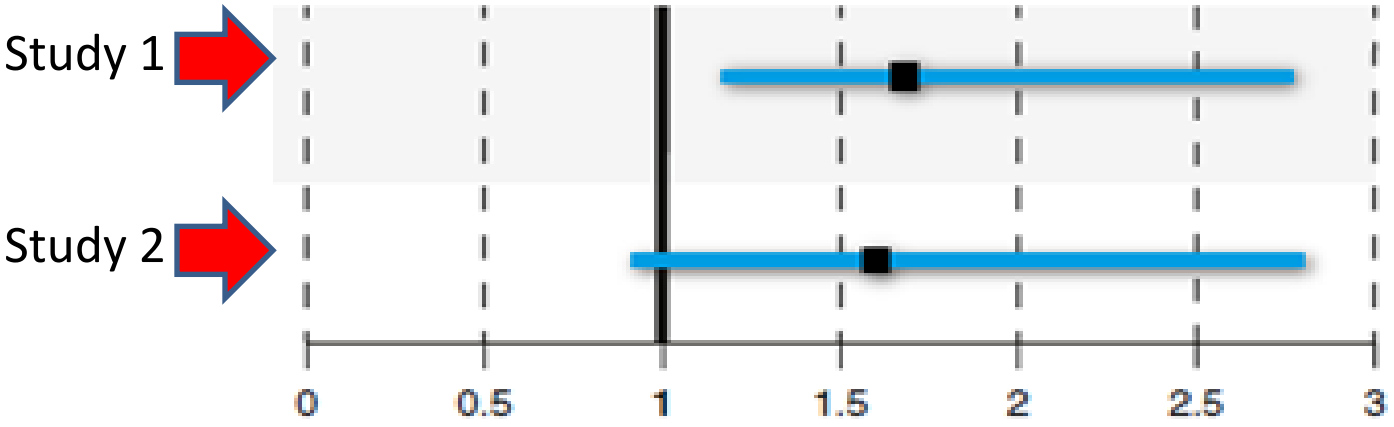
Confidence Intervals and Wording Issues

- Study 1: Discuss how you would explain the results of your stroke prevention study where the ARR for Drug A compared to Drug B was 1.3%, 95% CI (—0.90% to 1.80%)
- Study 2: Discuss how you would explain the results of your stroke prevention study where the ARR for Drug A compared to Drug B was 1.3%, 95% CI (0.50% to 1.50%)

How Statements About Confidence Intervals Can Cause Confusion

- Investigators often report different conclusions from very similar data
- Authors' wording becomes important when confidence intervals marginally cross or just fail to cross the line of no difference on a forest plot
- Examples 

Similar Relative Risks From 2 Studies Reporting Adverse CV Events With Drug A



Study 1—Authors' Conclusion: Our meta-analysis raises safety concerns about the potential for an increased risk of serious adverse cardiovascular events associated with the use of drug A.

Study 2—Authors' Conclusion: Drug A produced no significant increase in serious cardiovascular adverse events associated with its use

Useful Wording CIs

- In this large, valid, 36 month trial the relative risk for overall mortality with drug A compared to placebo was 0.9, 95% CI (0.80 to 1.05).
- Drug A may reduce risk by up to approximately 20% or increase risk by approximately 5%.

Bleeding Rate Comparisons Between 2 Drugs

- Example: Rate for major bleeding with drug A was 4.1% and for drug B it was 2.8%, ARI 1.3%, p=0.11, 95% CI (−1.8% to 2.8%)
- Watch out for statements like these:
  - "There was no difference in bleeding rates."
  - "There was no statistically different rates of bleeding rates in the two drugs."

Examples Of Misleading Yes/No Conclusions Based on CIs: Aspirin for Primary Prevention CV Disease

- Problem of using cutoff for statistical significance (e.g., $p<0.05$ or 95% CI not touching or crossing line of no difference)
  - Authors of Meta-analysis 1: With aspirin, modest, but nonsignificant, reductions were observed for all-cause mortality: OR 0.94, 95% CI (0.88 to 1.00)
  - Authors of Meta-analysis 2: Aspirin reduced all-cause mortality: RR 0.94, 95% CI (0.88 to 1.00)

Ref. Meta-analysis 1: Seshasai SR et al. PMID: 22231610.

Ref. Meta-analysis 2: Raju N et al. PMID: 21592450

Safety—More Misleading Conclusion and CIs

- Authors of RCTs may mislead readers when reporting adverse events, (eg, "Adverse effects were similar in both groups")
- Example: Lassen et al. PMID: 12049858
  - **Authors report, "The 2 groups did not differ in clinically relevant bleeding."**
    - Actual rates for major bleeding: 47/1140 (4.1%) fondaparinux vs 32/1133 (2.8%) enoxaparin, p=0.11
    - But CIs provide more information: ARI 1.3, 95% CI (—0.21 to 2.8) and since the true difference could be as great as 2.8% (i.e., clinically relevant) the authors' conclusion is misleading
    - Consensus of orthopedic surgeons: 2.8% difference is clinically meaningful

Good Wording Example

- ***Draft* Safety Evidence Statement — LeClerc 96**
  - LeClerc 96 in a valid study of 670 patients undergoing total knee replacement surgery reported bleeding rates of 1.8% for warfarin vs 2.1 % for enoxaparin,
    ARR 0.3%, 95% CI (−2.4% to 1.8%)
  - The evidence for a difference in bleeding rates between warfarin and enoxaparin is inconclusive based on a consideration of the rates and 95% CIs of study patients
    - Review of confidence intervals indicates that the difference in bleeding rates could have been as great as 2.4% favoring warfarin or up to 1.8% favoring enoxaparin

Wording May Mislead

- When you seethe word *"may" e.g.*, "drug A *may* reduce mortality…"
  - P-value is likely to be close to or greater than 0.05
  - CIs close to, touch or cross line of no difference
- Examples: ARR 0.3%, 95% CI (−2.4% to 1.8%)
  - "There was no difference in bleeding rates."
  - "There was no statistically different rates of bleeding in the two drugs."

Cochrane Handbook

9.7 Common errors in reaching conclusions

A common mistake when there is inconclusive evidence is to confuse 'no evidence of an effect' with 'evidence of no effect'. When there is inconclusive evidence, it is wrong to claim that it shows that an intervention has 'no effect' or is 'no different' from the control intervention. It is safer to report the data, with a confidence interval, as being compatible with either a reduction or an increase in the outcome. When there is a 'positive' but statistically non-significant trend authors commonly describe this as 'promising', whereas a 'negative' effect of the same magnitude is not commonly described as a 'warning sign'. Authors should be careful not to do this.

Using CIs to Establish Meaningful Clinical Outcomes

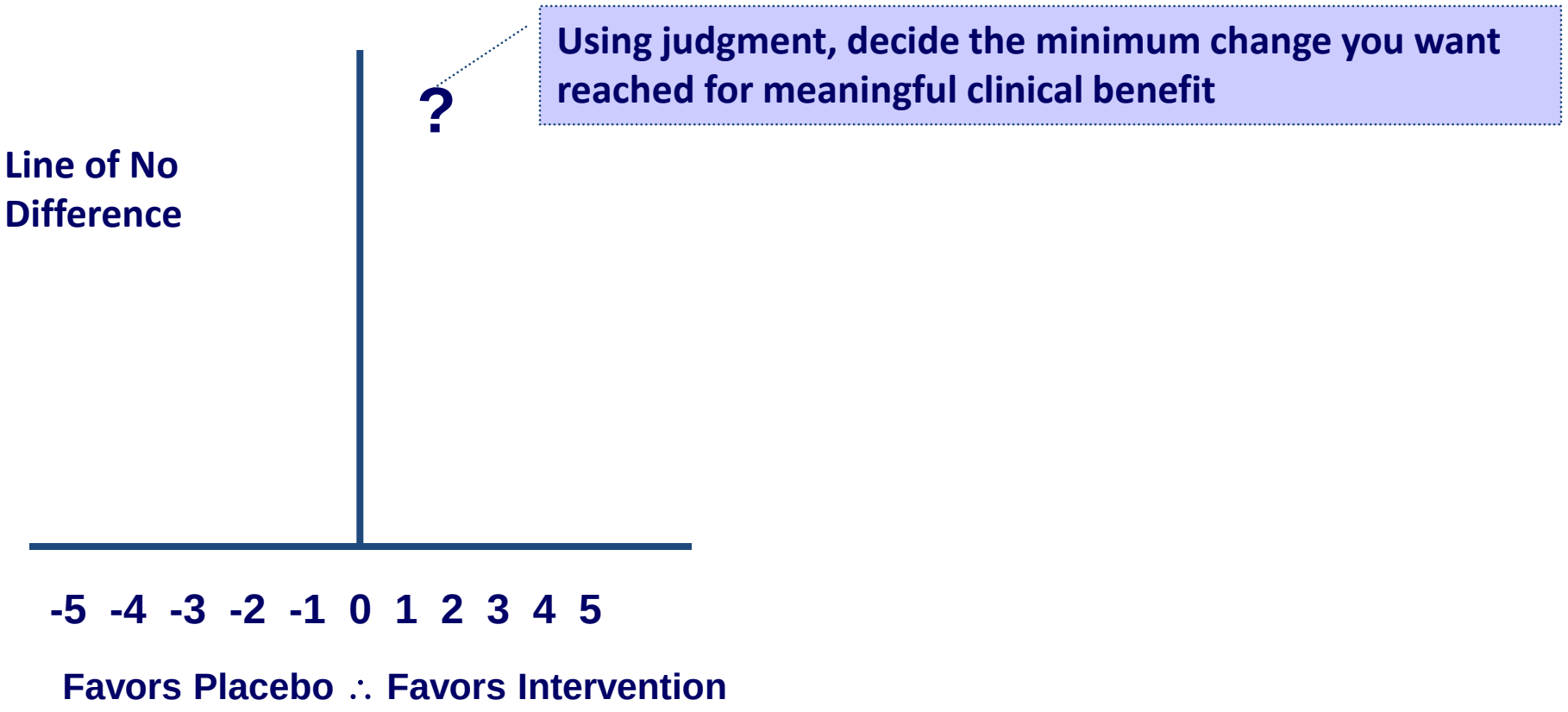
- For **statistically significant results**, is the confidence interval **wholly within** your judgment for meaningful clinical benefit?
  - Example: You decide you want to see **at least a 1 percent reduction in mortality** – this is a judgment
  - ARR 2, 95% CI (1 to 3) meets your requirement for meaningful clinical benefit and, therefore, these results can be considered **conclusive** (given a 5% margin for the play of chance)

Reminder About Non-Significant Findings: No Difference or Not Enough People?

- Findings that are **not statistically significant** raise a question –
  - Is there **truly no difference** between the groups?
  - **Or** was the study **insufficiently powered** (Type II or beta error), meaning…

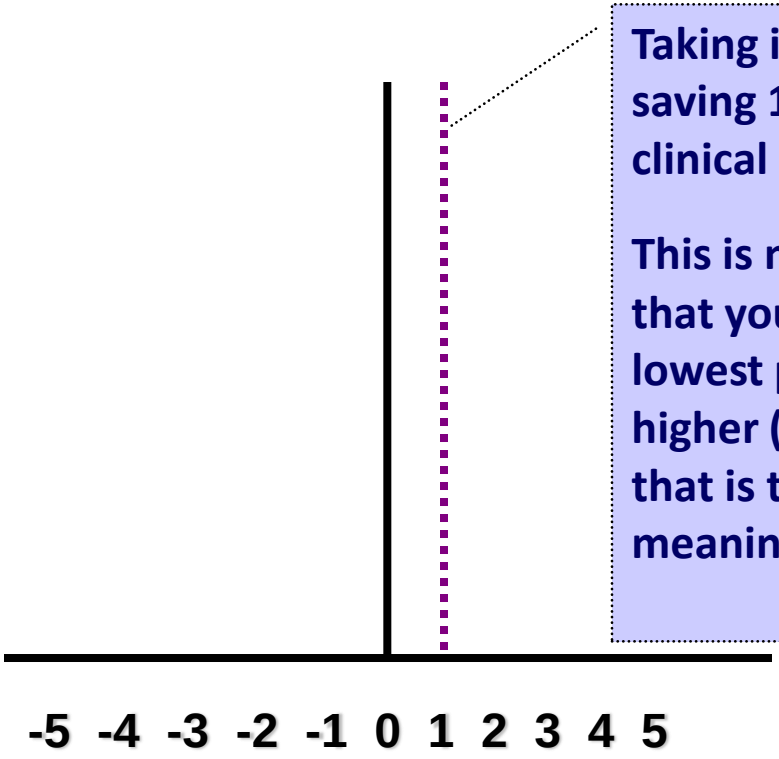
Applying Confidence Intervals to Decide Meaningful Clinical Benefit

- Endpoint = Reduction in mortality



Applying Confidence Intervals to Decide Meaningful Clinical Benefit

- Endpoint = Reduction in mortality

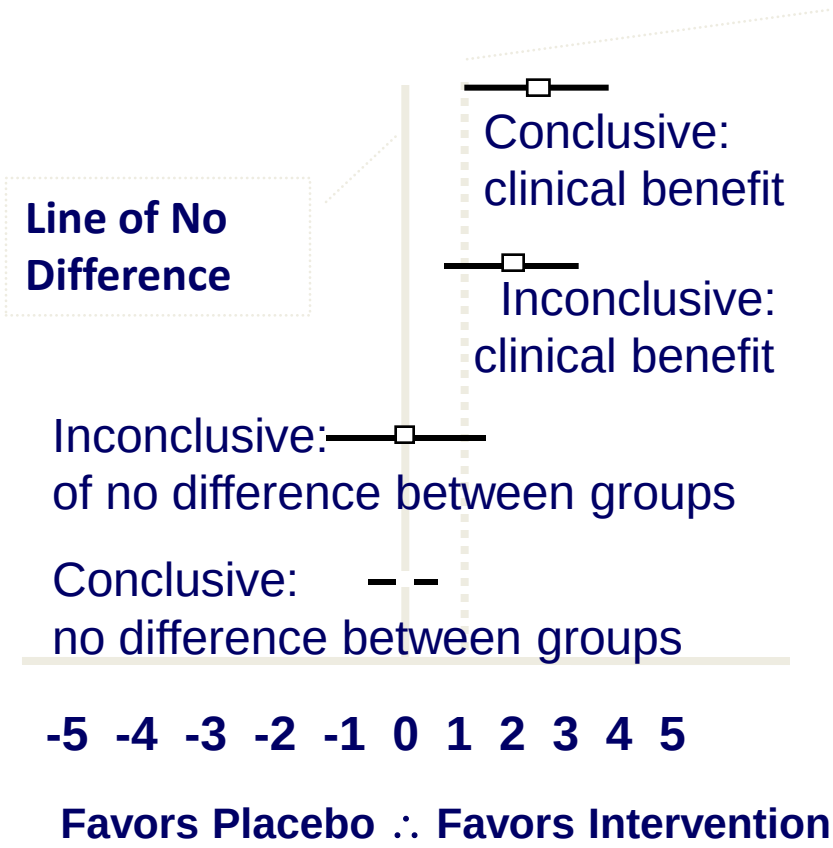


Taking into account all factors, let us assume we agree saving 1 out of 100 meets our minimum for meaningful clinical benefit.

This is now *your own line* of the least amount of benefit that you will accept as being clinically meaningful – the lowest possible result must start directly at this line or higher (ie, in this instance it can touch this line because that is the start of your minimum acceptance for meaningful clinical benefit)

Applying CIs to Decide Meaningful Clinical Benefit Size

Hypothetical Outcome: Reduction in Mortality

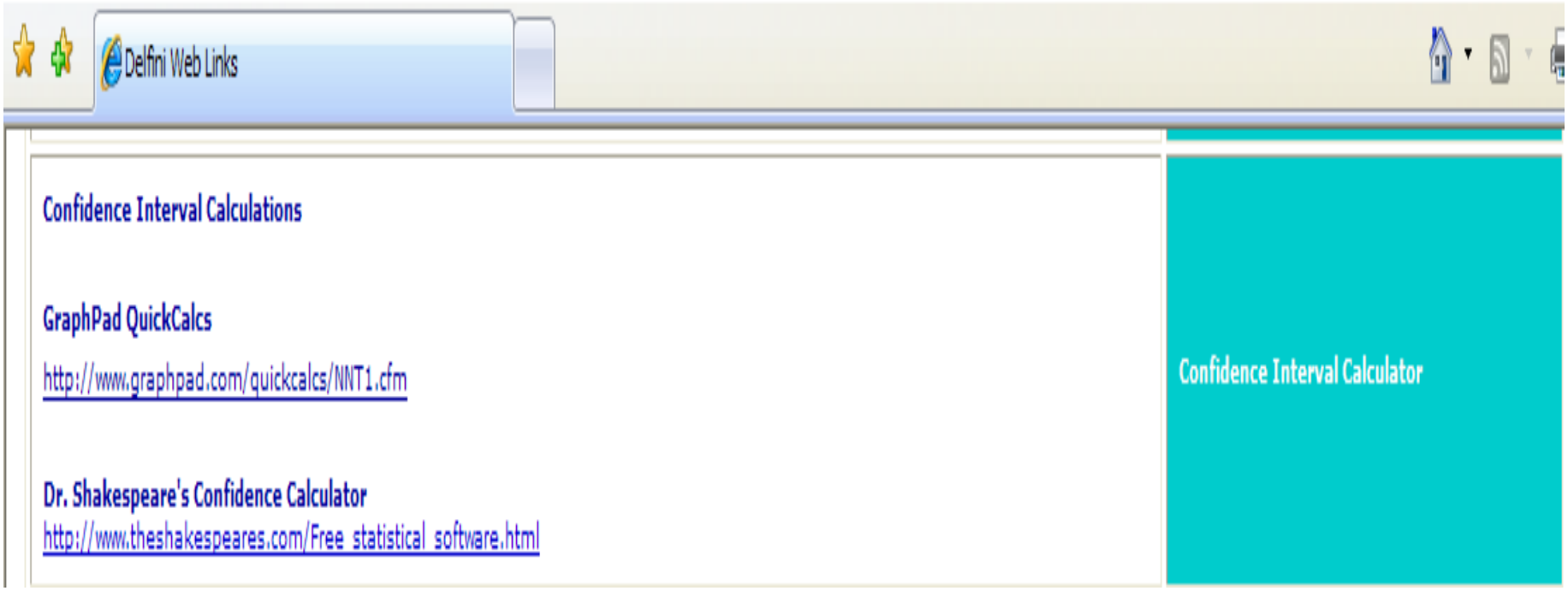


For valid studies, compare the confidence intervals to *your line* to evaluate the possibilities for or against what *you* define as meaningful clinical benefit

Given the margin for the play of chance --

- **If wholly within your margin, results can be considered conclusive for meaningful clinical benefit**
- **If overlapping your line, results are inconclusive**
- **If wholly outside, results can be considered conclusive for no difference between groups**

Delfini Web Links for Confidence Interval Calculators



From Study Data…

QuickCalcs Online Calculators for Scientists

1. Select category | 2. Choose calculator | **3. Enter data** | 4. View results

Analyze a 2x2 contingency table

Enter your data

Enter the number of subjects actually observed. Don't enter proportions, percentages or means.

Learn how to create a contingency table.

	Outcome 1	Outcome 2
Group 1		
Group 2		

	Alive	Dead
Control Group	85	15
Study Group	90	10

Questions on Forest Plots or CIs?