

Confidence Intervals & Their Use for Interpreting Significance of Results

Confidence Intervals (CI)

- The CI represents a statistically plausible **range for which there is a 95% chance that the true answer lies** in a valid study
 - Example: ARR = 5%; 95% CI (3% to 7%)
- “95 percent confidence” is conventional and provides for a 5% margin for chance effects
- **Can be used for any measure of outcomes**
- **More useful than p-values**
 - Helps quantify **uncertainty**
 - Help determine **statistical significance**
 - Helps determine **meaningful clinical benefit**
 - Helps deal with conclusivity of **non-significant findings** (Type II or beta error)

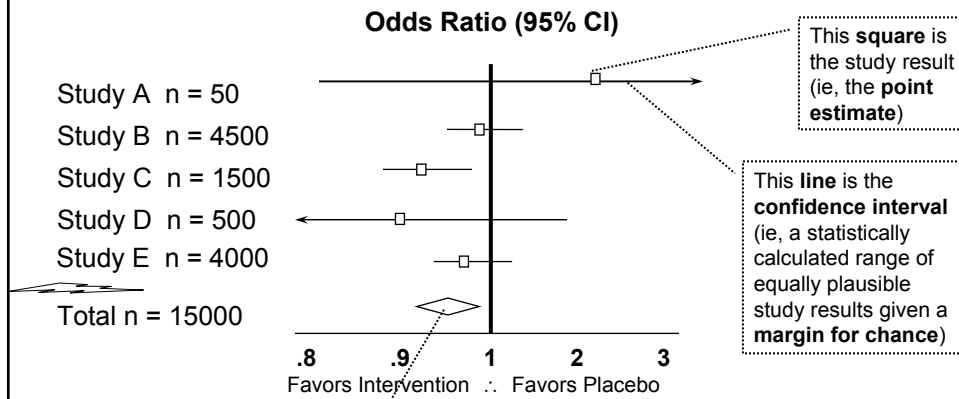


But First! How to Read a Forest Plot

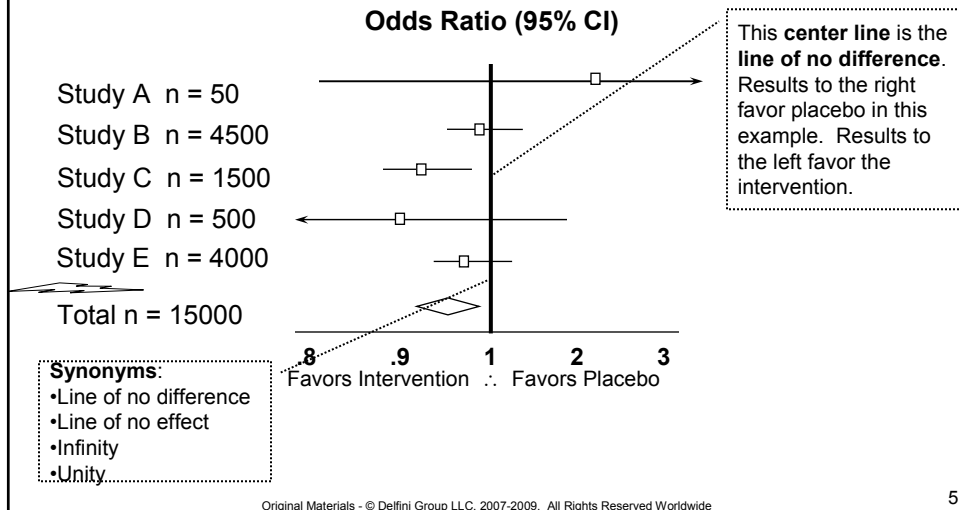


Graphic Display: Point Estimate, CI and Summary Diamond

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.

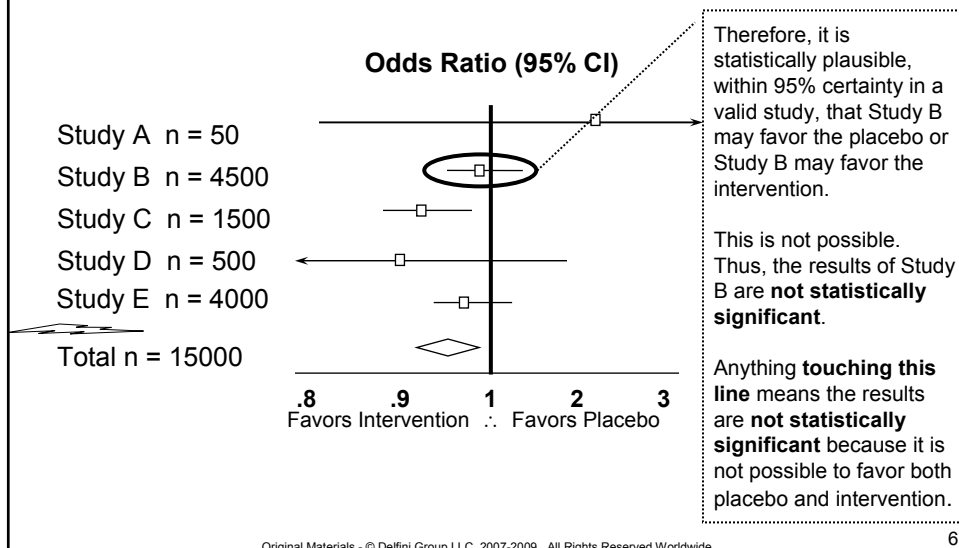


Favors Intervention ∴ Favors Placebo & The Line of No Difference



Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

Non-Statistical Significance



Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

Frequently CIs are Reported Numerically Only

Odds Ratio (95% CI)

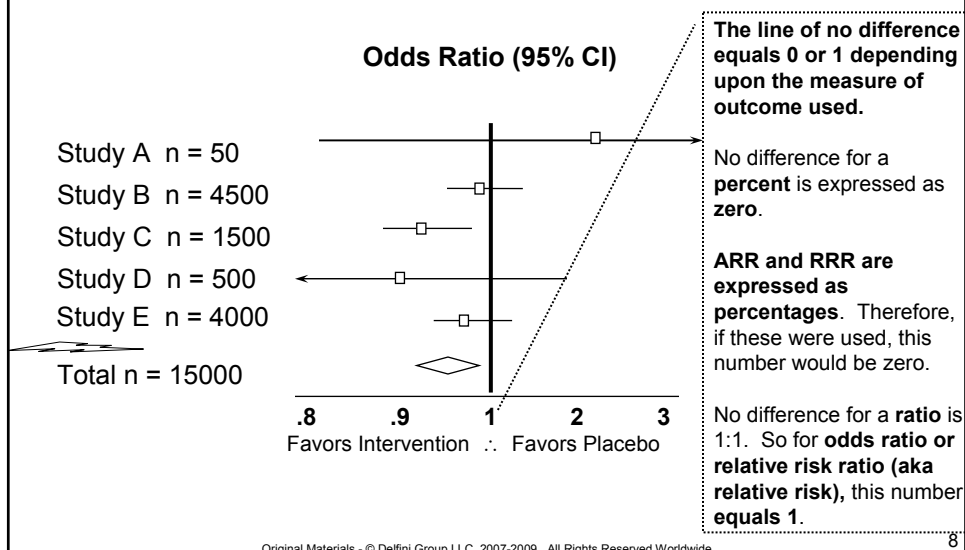
Study A n = 50	2.12 (0.82, 20.91)
Study B n = 4500	0.98 (0.96, 1.39)
Study C n = 1500	0.93 (0.88, 0.97)
Study D n = 500	0.91 (0.73, 1.71)
Study E n = 4000	0.97 (0.94, 1.33)
<hr style="width: 10%; margin-left: 0;"/> Total n = 15000	0.96 (0.92, 0.98)

Sampler – not meant to add up

Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

7

And So You Need to Know the Numerical Value for “No Difference”



8

Which of the Study Results 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)
<hr/> Total n = 15000	0.96 (0.92, 0.98)

Sampler – not meant to add up

Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

9

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 (1,3) meets your requirement for meaningful clinical benefit and, therefore, these results can be considered **conclusive** (given a 5% margin for the play of chance)



Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

10

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

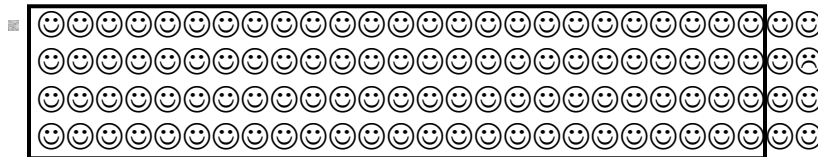
- Findings that are **not statistically significant** raise a question –
 - Is there **truly no difference** between the groups?
 - **Or** was the study **insufficiently powered** (Type II or beta error), meaning...

Power is About People

- **Are there enough people** to show a statistically significant difference between the groups if there is one?

All Readers Need to Know about Statistical Power

- Power is about the **number of people needed to show a statistically significant difference if there is one**
- Scenario
 - An event actually happens in 1 in 108 people, but we only studied 100 people, and we missed the one who had the outcome



Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

13

All Readers Need to Know about Power Calculations

- That unless you are a researcher, you can **ignore** them
- Power calculations are performed prior to a study **in an attempt to help investigators guess at the number of people they should enroll**
 - Power calculations are generally performed only for the primary outcome
 - They entail a lot of assumptions and provide for the play of chance

Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

14

Non-significant Findings Raise a Question

- Is there **TRULY no difference** between the groups?

OR

- Is there **TRULY a difference** between the groups, but there were **not enough people** to show it?
- Confidence intervals get you out of this bind; they can tell you when -
 - **results are conclusive that there is no benefit** (given the play of chance)
 - **results are inconclusive – benefit is possible**

Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

15

Using CIs to Address Uncertainty of Non-significant Findings

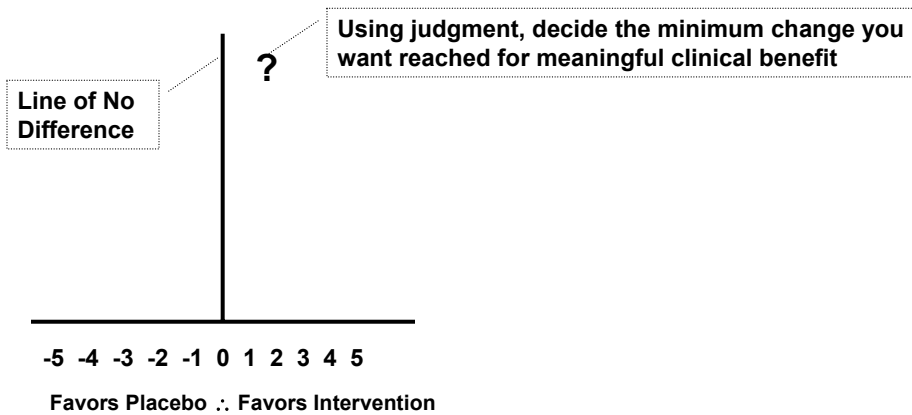
- For **non-significant findings**, is the confidence interval **wholly outside** your judgment for meaningful clinical benefit?
 - ARR 0.3, 95% CI (-0.1, 0.7) meets your requirements for **no possible meaningful clinical benefit** and results can be considered **conclusive of no difference between the groups** (given some margin for the play of chance)

Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

16

Applying Confidence Intervals to Decide Meaningful Clinical Benefit

- Endpoint = Reduction in mortality

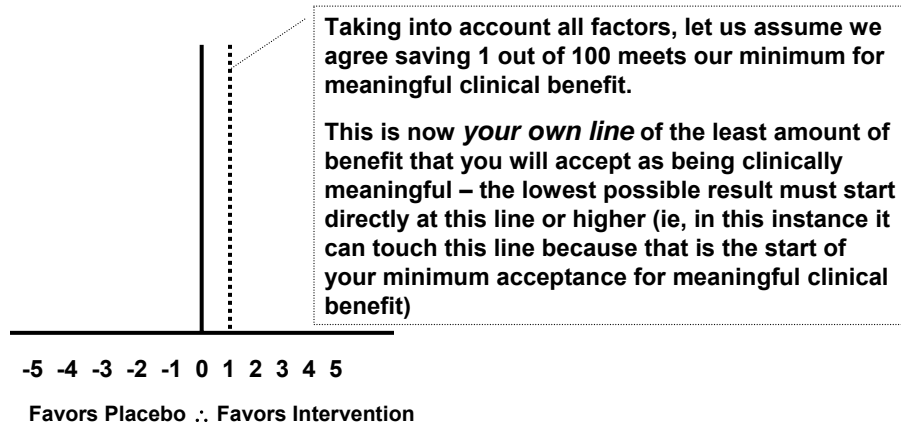


Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

17

Applying Confidence Intervals to Decide Meaningful Clinical Benefit

- Endpoint = Reduction in mortality

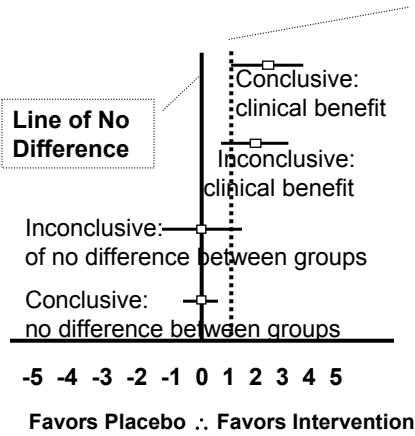


Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

18

Applying CIs to Decide Meaningful Clinical Benefit Size

- Endpoint = 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 --

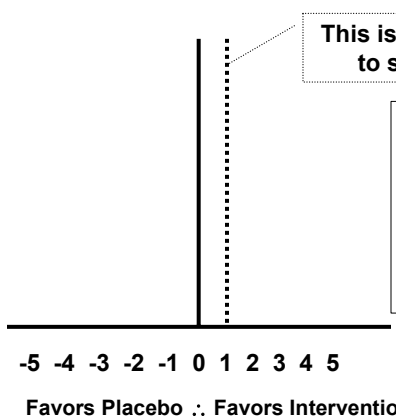
- If wholly within your margin, results can be considered conclusive for meaningful clinical benefit

- If overlapping your line, results are inconclusive

- If wholly outside, results can be considered conclusive for no difference between groups

Exercises: Evaluating Confidence Intervals

- Endpoint = Reduction in mortality



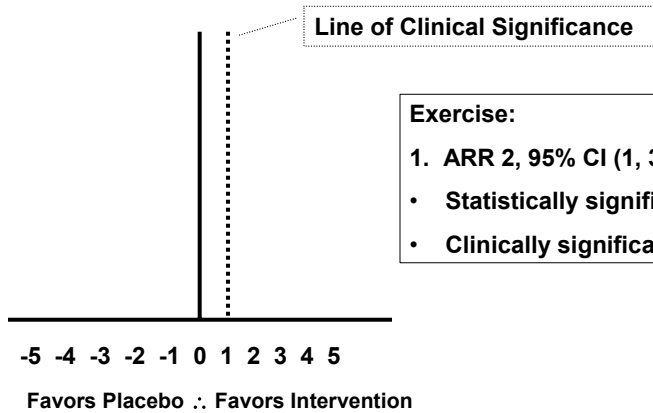
This is now your *own* line of the minimum you want to see reached for meaningful clinical benefit

Exercises:

1. ARR 2, 95% CI (1, 3)
2. ARR 2, 95% CI (0.5, 3.5)
3. ARR 2, 95% CI (-1, 5)
4. ARR 0, 95% CI (-0.75, 0.75)

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality

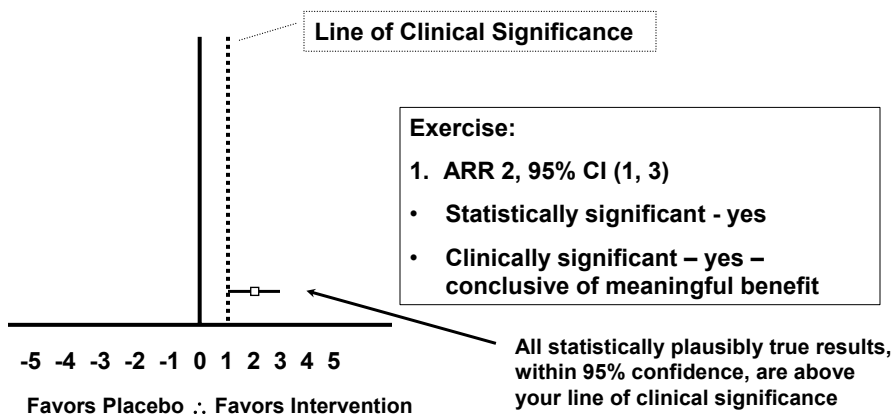


Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

21

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality

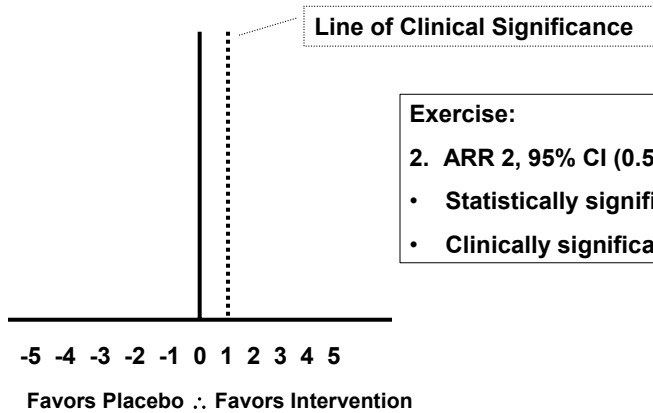


Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

22

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality



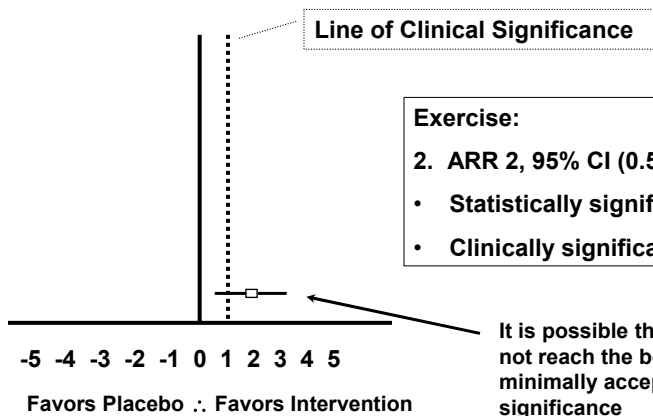
Exercise:

2. ARR 2, 95% CI (0.5, 3.5)

- Statistically significant or not?
- Clinically significant or not?

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality



Exercise:

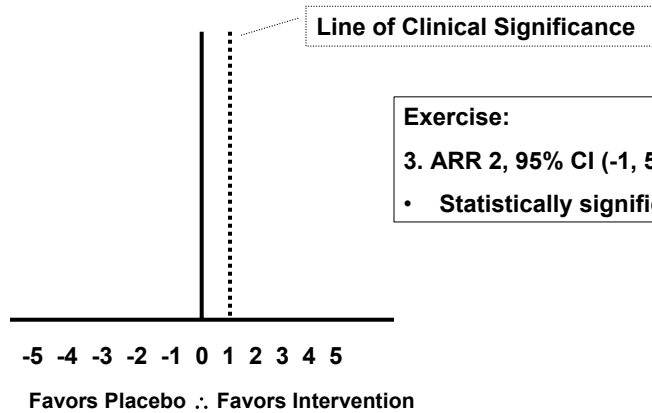
2. ARR 2, 95% CI (0.5, 3.5)

- Statistically significant - yes
- Clinically significant - inconclusive

It is possible that the true value does not reach the boundary for your minimally acceptable clinical significance

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality



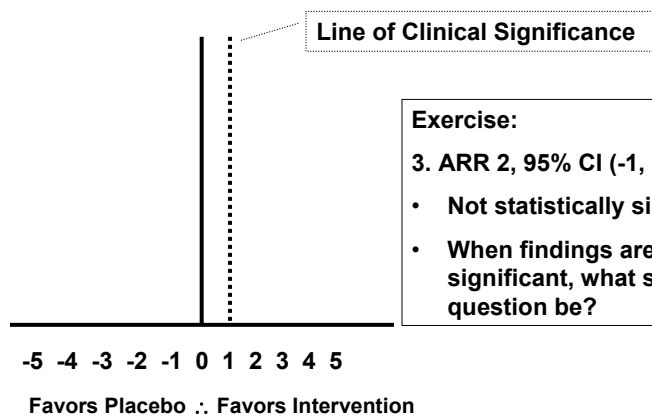
Exercise:

3. ARR 2, 95% CI (-1, 5)

- Statistically significant or not?

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality



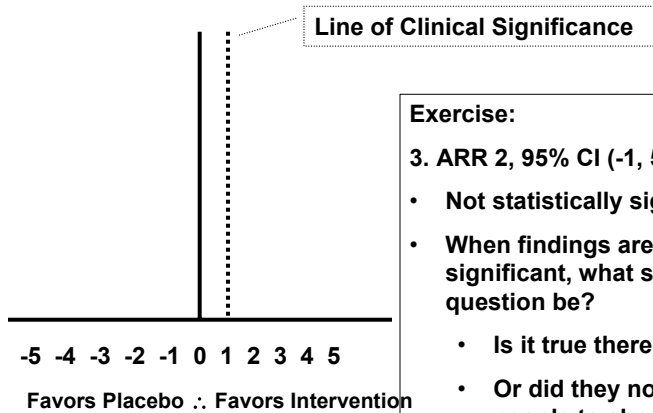
Exercise:

3. ARR 2, 95% CI (-1, 5)

- Not statistically significant
- When findings are not statistically significant, what should our next question be?

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality



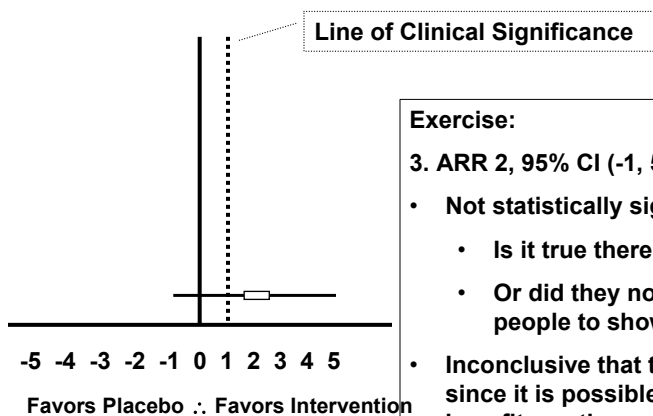
Exercise:

3. ARR 2, 95% CI (-1, 5)

- Not statistically significant
- When findings are not statistically significant, what should our next question be?
 - Is it true there is no difference?
 - Or did they not have enough people to show one?

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality



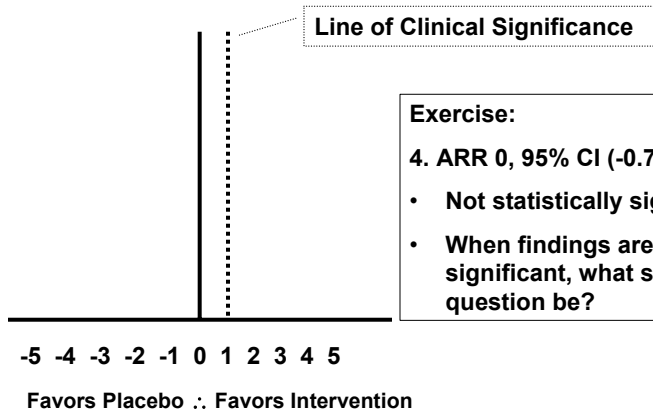
Exercise:

3. ARR 2, 95% CI (-1, 5)

- Not statistically significant
 - Is it true there is no difference?
 - Or did they not have enough people to show one?
- Inconclusive that there is no difference since it is possible there could be benefit meeting your criteria

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality



Exercise:

4. ARR 0, 95% CI (-0.75, 0.75)

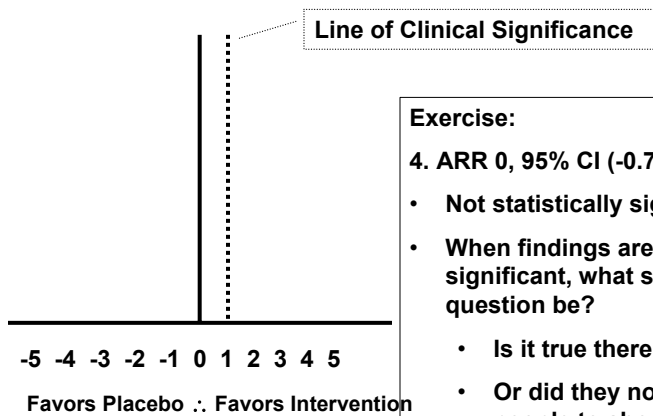
- Not statistically significant
- When findings are not statistically significant, what should our next question be?

-5 -4 -3 -2 -1 0 1 2 3 4 5

Favors Placebo ∴ Favors Intervention

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality



Exercise:

4. ARR 0, 95% CI (-0.75, 0.75)

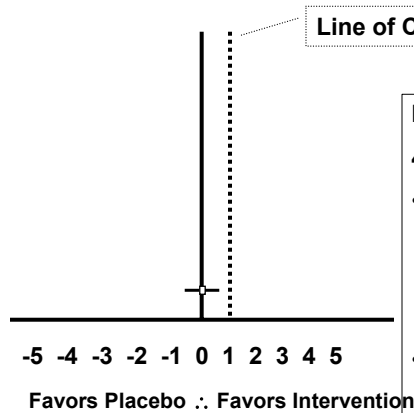
- Not statistically significant
- When findings are not statistically significant, what should our next question be?
 - Is it true there is no difference?
 - Or did they not have enough people to show one?

-5 -4 -3 -2 -1 0 1 2 3 4 5

Favors Placebo ∴ Favors Intervention

Evaluating Confidence Intervals

- Endpoint = Reduction in mortality



Line of Clinical Significance

Exercise:

4. ARR 0, 95% CI (-0.75, 0.75)

- Not statistically significant
 - Is it true there is no difference?
 - Or did they not have enough people to show one?
- You can conclude there is no difference since the range of plausibly statistically true results, within 95% confidence, do not reach your criteria

Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

31

Adverse Events 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, (95% CI) = 1.3, (-0.21 to 2.8) and since the true difference could be as great as 2.8% (ie, clinically relevant) the authors’ conclusion is misleading
 - Lack of statistically significant difference may be due to Type II error (meaning a power issue or not enough people to show a statistically significant difference if there is one)
 - In this case a systematic review reported a statistically significant increased bleeding rate with fondaparinux vs enoxaparin 96/3616 (2.7%) vs 63/3621 (1.7%), OR (95% CI) 1.54 (1.11–2.16), Bounemeux PMID: 14615118

Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

32

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.

Wording Example for Safety

- **Draft Safety Evidence Statement — LeClerc 96**
 - LeClerc 96 in a 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% (-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

When CIs are not Reported

- How much less is the p-value than .05? As it decreases, uncertainty around the values decreases also (see JAMA Users Guides to the Medical Literature for advice)
- If the outcomes are dichotomous, you can calculate them yourself if you wish →

Delfini Web Links for Confidence Interval Calculators



GraphPad Software
ANALYZE, GRAPH AND ORGANIZE YOUR 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

Create a 2x2 Table

From this →

Two x Two Table
% Alive v. Dead

	Control	Study
Alive	85	90
Dead	15	10

To this ↓

Alive Dead

Control Group 85 15

Study Group 90 10

Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

37

GraphPad Software
ANALYZE, GRAPH AND ORGANIZE YOUR DATA

QuickCalcs Online Calculators for Scientists

1. Select category 2. Choose calculator 3. Enter data

Number needed to treat (NNT)

How effective is a clinical treatment? Because many people find it hard to think about small fractions, these kind of data are better understood when converted to the Number Needed to Treat or NNT. Enter the number of patients in each group who had the "good" or "bad" outcome, and this calculator will convert to NNT and explain the results.

Enter the actual number of patients in each group. Don't enter fractions, percentages, or rates per 1000 or some other value.

Our Data

Two x Two Table
% Alive v. Dead

	Control	Study
Alive	85	90
Dead	15	10

Good Outcome Bad Outcome

control

experimental

Desired confidence level: 95% CI

Compute NNT

Good Outcome Bad Outcome

Study Group 90 10

Control Group 85 15

Desired confidence level: 95% CI

Our Results

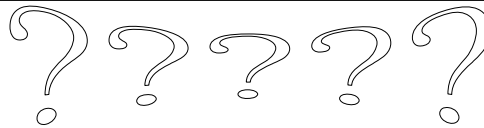
The difference, the absolute risk increase, is 5.00 percent.

The 95% confidence interval for this difference ranges from -4.14% to 14.14%.

Original Materials - © Delfini Group LLC, 2007-2009. All Rights Reserved Worldwide

38

Silent Mini-Review



In your workbook or on a piece of paper, record your answer for the following scenario?

- You read a study in which the investigators report non-significant differences for safety between Drug X and placebo. They claim that Drug X appears to be as safe as placebo.
- What is your next step?

Silent Mini-Review



In your workbook or on a piece of paper, record your answer for the following scenario?

- You read a study in which the investigators report non-significant differences for safety between Drug X and placebo. They claim that Drug X appears to be as safe as placebo.
- What is your next step?
Review the confidence intervals to discern if a possible value is clinically significant - if yes, the findings are actually inconclusive