STATISTICAL INFERENCE AND HYPOTHESIS TESTING PART III

Nasir Mushtaq, PhD, MBBS
Associate Professor
Department of Biostatistics and Epidemiology
Hudson College of Public Health

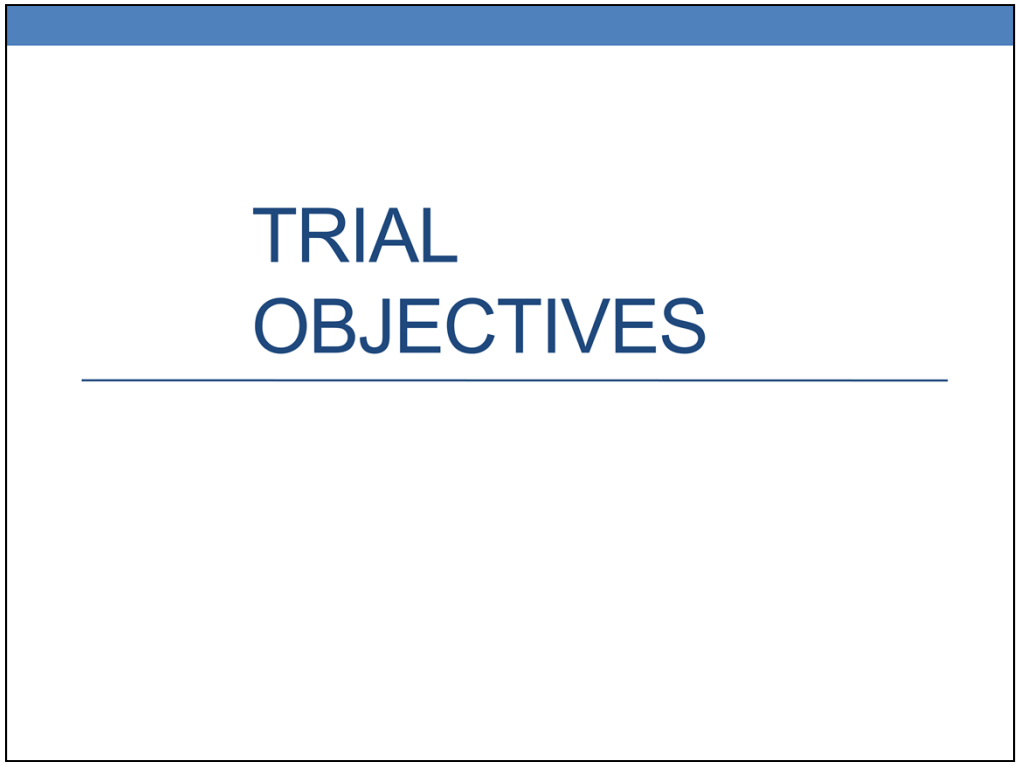
Department of Family and Community Medicine
OU-TU School of Community Medicine

In the third part of this series entitled Statistical Inference and Hypothesis Testing, we will discuss types of hypotheses, namely, superiority, non-inferiority, and equivalence hypotheses.

Objectives

- Distinguish superiority, non-inferiority, and equivalence hypotheses

After viewing this module, you will be able to distinguish among superiority, non-inferiority, and equivalence hypotheses.



Let’s discuss different types of trial objectives.

Hypothesis of Superiority

A trial with the primary objective of showing that response to the investigational product is superior to a comparative agent (active or placebo control)

- Is treatment A better than treatment B?
- Are these two groups different in regards to response time?

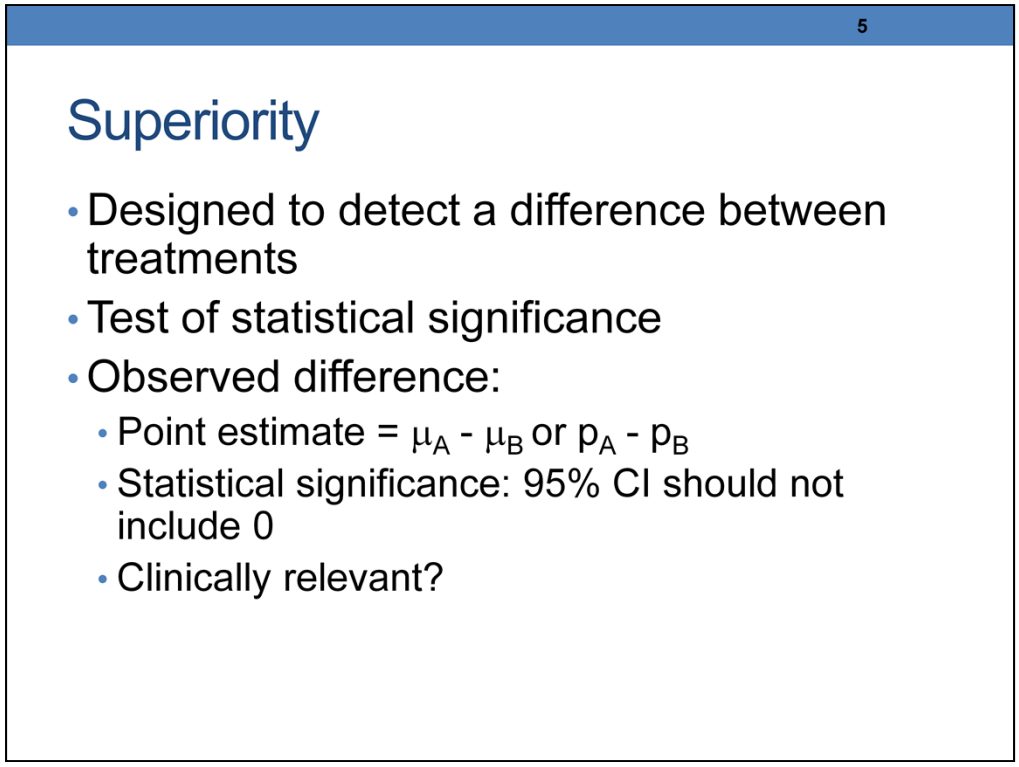
We will first consider superiority hypotheses.

We utilize superiority hypotheses in trials with the primary objective of showing that response to the investigational product is superior to a comparative agent (active or placebo control). In these settings, we are hypothesizing that there is a difference between groups.

Superiority hypotheses would be relevant for the following types of questions:

Is treatment A better than treatment B?

Are these two groups different in regards to response time?



Basic review of Superiority testing

Superiority can be established in placebo-controlled trials, active control trials, dose-response trials

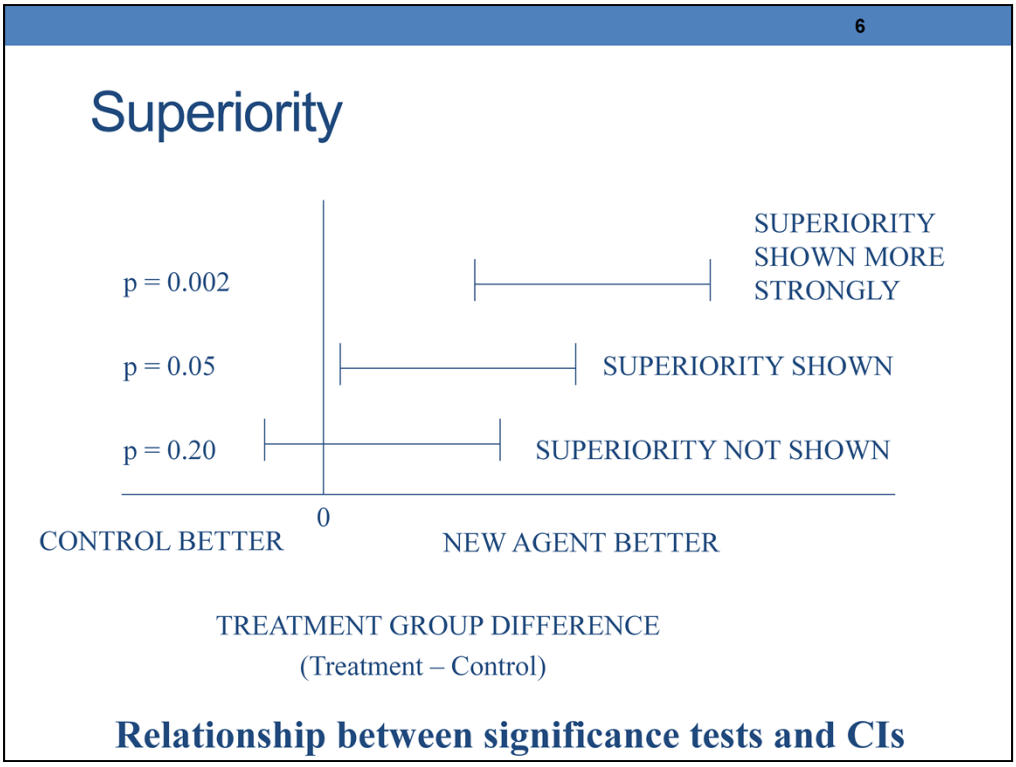
Our first step is to conduct a test of statistical significance to evaluate whether the results are consistent with the assumption of there being no difference in the clinical effect of 2 treatments (i.e., consistent with the null hypothesis).

We will calculate a p-value, the probability that the observed difference – or a larger one – could have arisen by chance assuming that no difference really existed. The smaller the p-value, the more implausible the null hypothesis (i.e., the assumption that there really is no difference between the treatments.

If the p-value is less than or equal to alpha or if the confidence interval does not include 0, we reject the null hypothesis and conclude that there is a significant difference between groups.

Note that clinical relevance requires separate considerations: a statistically significant difference may not be clinically relevant. To determine if a difference in

clinically-significant, we need to compare the confidence interval to the threshold used to define a clinically-important difference.



In this slide, we provide interpretations of the results for different scenarios.

Differences to the right, calculated as the treatment minus control groups, provide evidence that the treatment (new agent) is better than the control while differences to the left provide evidence that the control is better.

In the bottom figure, the p-value is greater than the alpha level of 0.05 and the confidence interval includes 0. We conclude that there is no significant difference between groups and that superiority is not shown.

In the middle figure, the p-value is equal to alpha and the confidence interval does not include 0. We conclude that there is a significant difference between groups and that superiority is shown.

In the top figure, the p-value is less than alpha and the confidence interval is shifted well above 0. We conclude that there is a significant difference between groups and that superiority is shown even more strongly given the shift farther from 0.

Is this OK?

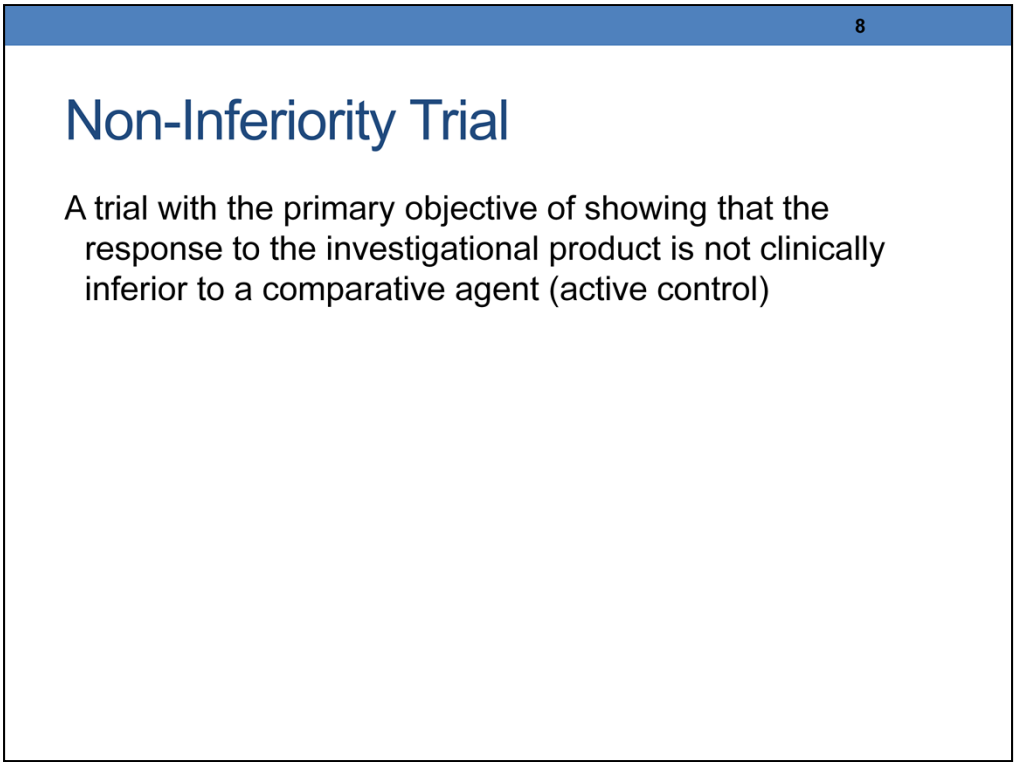
Perform a superiority study and fail to find statistical significance. Conclude groups are equivalent by default.

Failure to show a significant difference is not the same as proving equivalence!

Now, consider a problem where we perform a superiority study and fail to find statistical significance (p-value > alpha level).

Can we conclude that the groups are equivalent by default?

No! Failure to show a significant difference is not the same as proving equivalence. It may be that we have very small sample sizes and low power to detect true differences. Given the small sample sizes, our estimation will not be precise and therefore, we will not be able to establish equivalence.



The second type of hypothesis that we will consider is a non-inferiority hypothesis.

A non-inferiority hypothesis is appropriate for a trial with the primary objective of showing that the response to the investigational product is not clinically inferior to a comparative agent (active control).

For example, we might want to test if a home-based exercise therapy program is non-inferior to a clinic-based exercise therapy program in improving physical outcomes following cardiac surgery. The home-based program may be more feasible for patients and less expensive to implement; however, we would want to make sure that the effect of the home-based program is not inferior to the clinic based program before prescribing home-based therapy over clinic-based therapy.

Non-inferiority Example

Example:

The primary goal for this study was to determine whether fluconazole would be as effective (or nearly as effective) as amphotericin B in preventing the relapse of cryptococcal meningitis in patients with AIDS. It was thought that the reduced toxicity and oral administration of fluconazole might give it an advantage over amphotericin B, even if fluconazole was slightly less effective.

Fluconazole vs amphotericin B in prevention of relapse of cryptococcal meningitis (Powderly, Saaf et al, NEJM (1992))

As an example from the literature, consider the following NEJM article.

The primary goal for this study was to determine whether fluconazole would be as effective (or nearly as effective) as amphotericin B in preventing the relapse of cryptococcal meningitis in patients with AIDS. It was thought that the reduced toxicity and oral administration of fluconazole might give it an advantage over amphotericin B, even if fluconazole was slightly less effective.

Equivalence Margin (δ)

- This margin is the largest difference between groups that can be judged as still being clinically non-inferior
- The margin should be specified in the protocol
- The choice of the equivalence margin should be justified clinically

In order to make any inference related to non-inferiority, we need to first define a margin of equivalence.

This margin is the largest difference between groups that can be judged as still being clinically non-inferior, for example, the amount of physical functioning that if lost, would not be of clinical concern.

The margin should be specified in the protocol and the choice of the equivalence margin should be justified clinically.

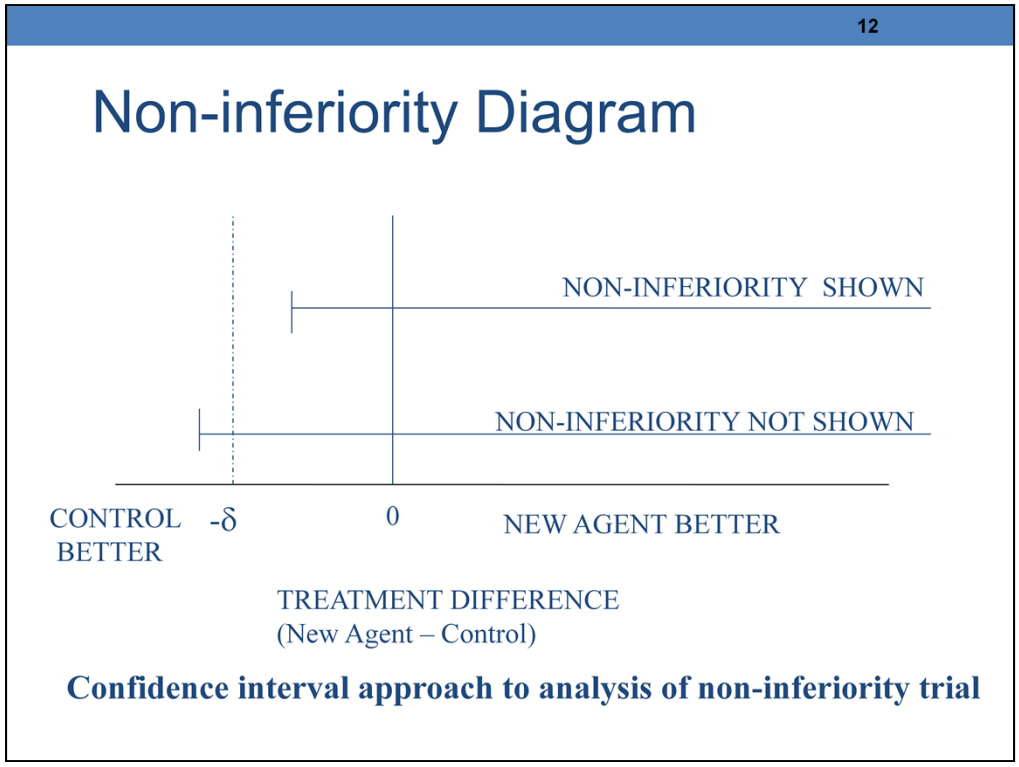
Non-Inferiority Trial Statistical Approach

Statistical analysis is generally based on the use of confidence intervals. A one-sided confidence interval should be used for non-inferiority trials.

1-sided 97.5% CI should lie entirely to the right of the value $-\delta$ where δ is the margin of non-inferiority or equivalence.

The statistical analysis for a non-inferiority hypothesis is generally based on the use of confidence intervals. A one-sided confidence interval should be used for non-inferiority trials because we are interested in establishing that outcomes are not worse under the investigational arm compared to the standard of care. We are only interested in establishing differences in one direction, not both.

To determine statistical significance, we will determine if the 1-sided 97.5% CI lies entirely to the right of the value -□ where □ is the margin of non-inferiority or equivalence.



This slide show a schematic of the types of intervals that we will calculate to evaluate non-inferiority.

In this case, we would judge the new therapy to be non-inferior if the lower limit of the confidence interval lies above the –delta value (corresponding to the margin of equivalence).

In the bottom example, the interval includes the value (–delta) and therefore, non-inferiority is not established.

In the top example, the interval does not include the value (–delta) and therefore, non-inferiority is established because the interval lies entirely above the value (-delta). The interval includes 0 and therefore, the difference is not statistically significant. In addition, we can see that the loss expected under the new treatment (negative difference values) is expected to be small and not of clinical importance because the negative values that are included in the interval still lie above (-delta) and are therefore, not of clinical importance.

Equivalence Trial

A trial with the primary objective of showing that the response to two or more treatments differs by an amount which is clinically unimportant.

This is usually demonstrated by showing that the true treatment difference is likely to lie between a lower and an upper equivalence margin (δ) of clinically acceptable differences.

The third type of hypothesis that we will consider is an equivalence hypothesis.

A trial with the primary objective of showing that the response to two or more treatments differs by an amount which is clinically unimportant, either in the negative or positive direction, is an equivalence study.

Equivalence is usually demonstrated by showing that the true treatment difference is likely to lie between a lower and an upper equivalence margin (□) of clinically acceptable differences; meaning, we do not expect to see differences between the groups on average that are clinically important.

Equivalence Trial Statistical Approach

Statistical analysis is generally based on the use of confidence intervals (CI).

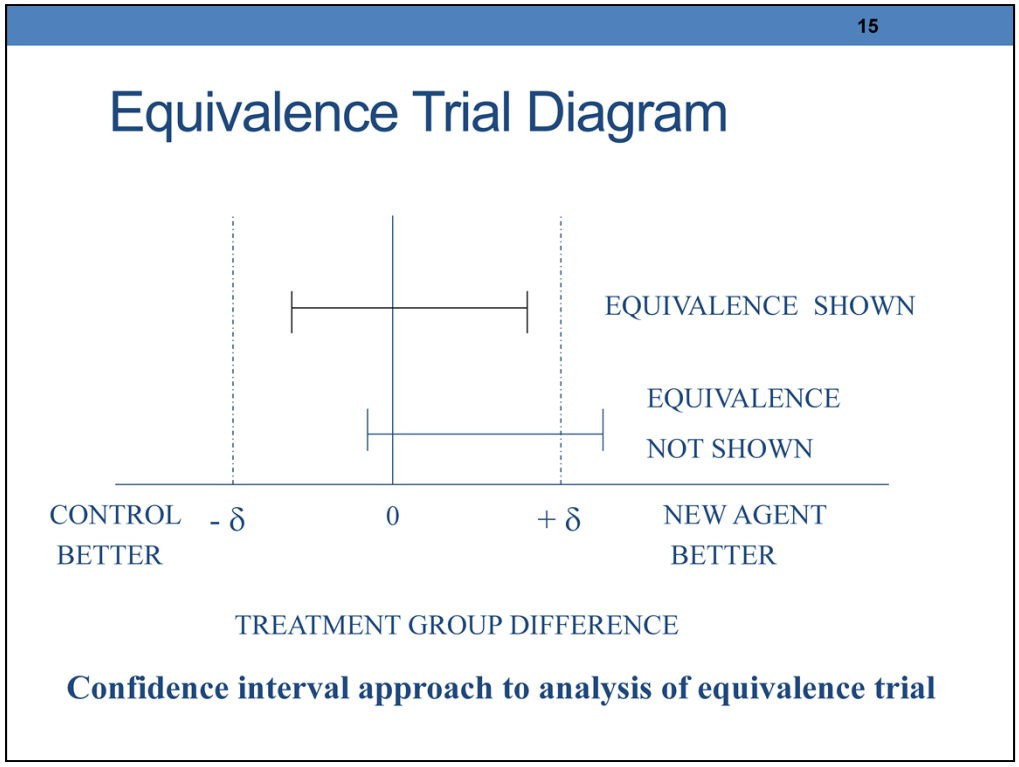
For equivalence trials, two-sided confidence intervals should be used.

Equivalence is inferred when the entire confidence interval falls within the equivalence margins (-δ to +δ).

For an equivalence trial, statistical analysis is generally based on the use of confidence intervals (CI).

For equivalence trials, two-sided confidence intervals should be used because we want to determine if differences in either direction are clinically important.

Equivalence is inferred when the entire confidence interval falls within the equivalence margins (-□ to +□). This result would indicate that expected differences are small and are not large enough to be judged clinically important.

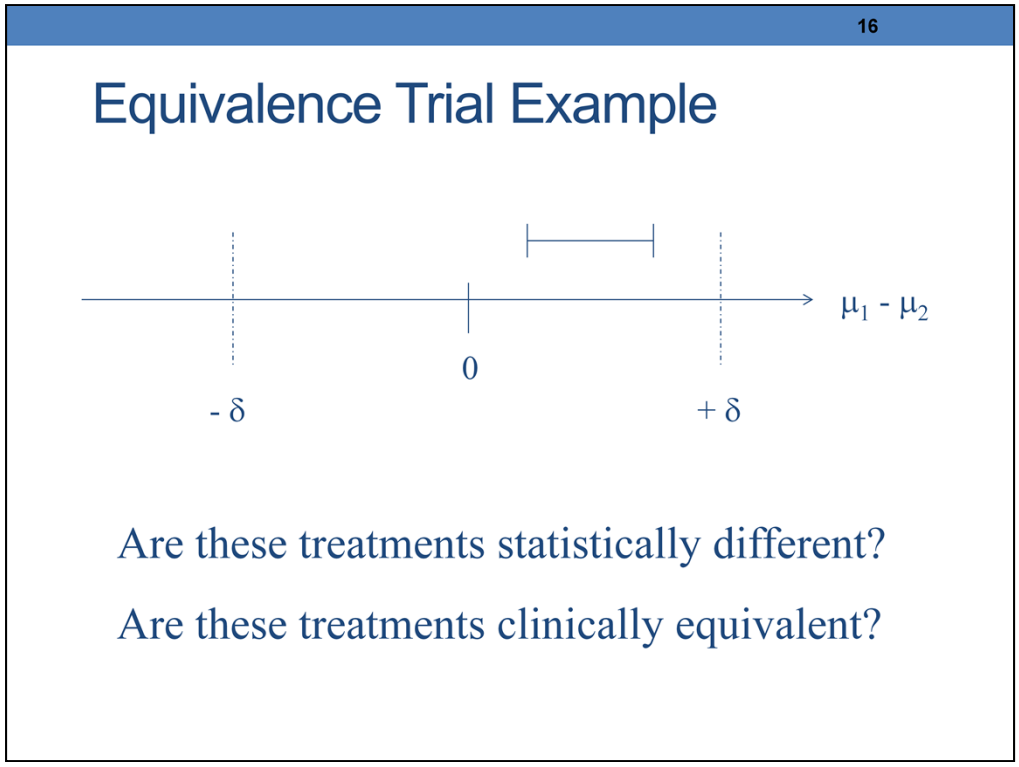


This slide show a schematic of the types of intervals that we will calculate to evaluate equivalence.

In this case, we would judge the new therapy to be equivalent if the lower limit of the confidence interval lies above the (–delta) value and the upper limit of the confidence interval lies below the (+delta) value (corresponding to the margin of equivalence).

In the bottom example, the interval includes the value (+delta) and therefore, equivalence is not established.

In the top example, the interval does not include the value (–delta) or the value (+delta) and therefore, equivalence is established because the interval lies entirely within the margin of equivalence (+/-delta). The interval includes 0 and therefore, the difference is not statistically significant. In addition, we can see that the loss or gain expected under the new treatment (negative or positive difference, respectively) is expected to be small and not of clinical importance because the values that are included in the interval still lie within (+/-delta) and are therefore, not of clinical importance.



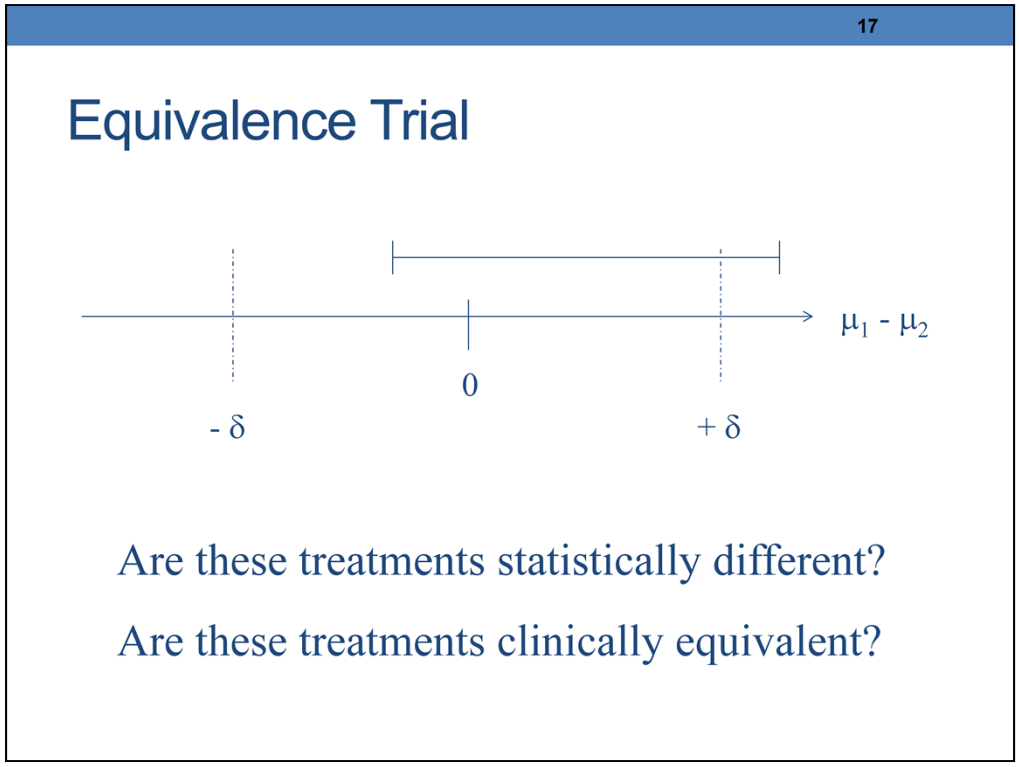
In this example, based on the defined margin of equivalence (+/- delta) and the confidence interval,

Are the treatments statistically different?

Yes, the confidence interval does not include 0.

Are the treatment clinically equivalent?

- Yes, the confidence interval lies entirely within the interval (+/- delta). Although the difference is statistically significant, the difference is not clinically important.



In this example, based on the defined margin of equivalence (+/- delta) and the confidence interval,

Are the treatments statistically different?

No, the confidence interval includes 0.

Are the treatment clinically equivalent?

- No, the confidence interval does not lie entirely within the interval (+/- delta). Instead, the upper limit lies above (+delta). We cannot conclude statistical significance or clinical equivalence based on these results.

Summary

- Distinguish superiority, non-inferiority, and equivalence hypotheses
- A non-significant difference, based on the p-value alone, does not necessarily establish equivalence

In summary, we have discussed differences among superiority, non-inferiority, and equivalence hypotheses and have discussed statistical approaches for testing these hypotheses. Remember that testing of non-inferiority and equivalence hypotheses rely on confidence intervals. A non-significant difference, based on the p-value alone, does not necessarily establish equivalence; instead, we need to determine if the confidence interval lies within (+/- delta) in order to declare equivalence.