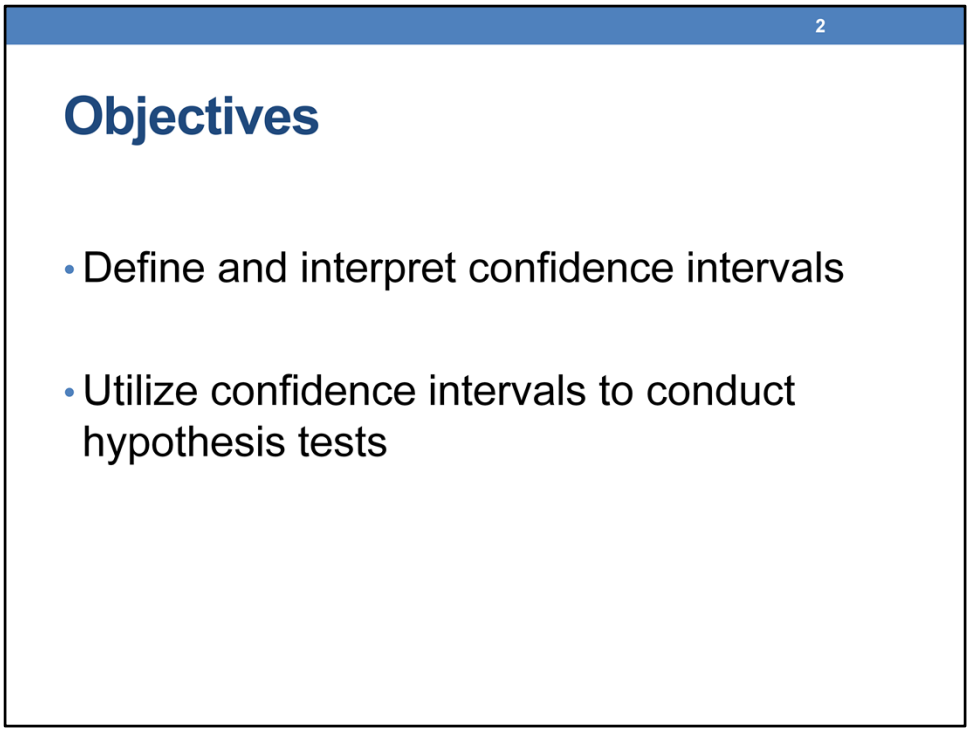
STATISTICAL INFERENCE AND HYPOTHESIS TESTING PART II

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

Welcome, in this second part of the series entitled Statistical Inference and Hypothesis Testing, we will discuss statistical concepts related to confidence intervals and estimation.



After viewing this module, you will be able to:

Define and interpret confidence intervals

Utilize confidence intervals to conduct hypothesis tests

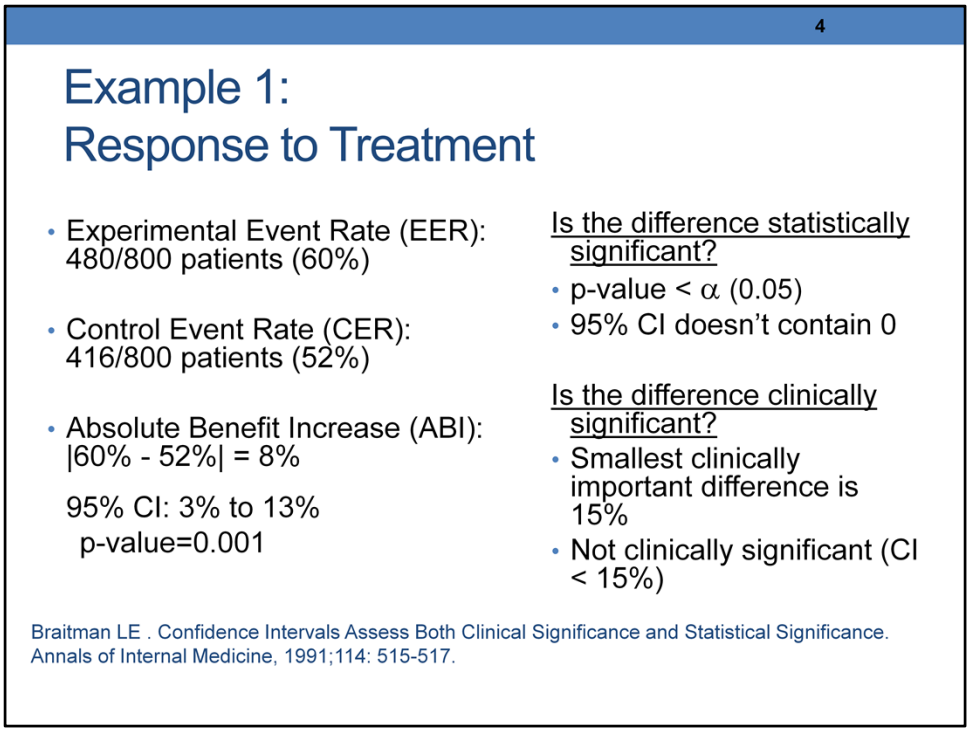
Confidence Interval

An interval estimate consisting of a range of values (with a lower and upper bound) constructed to have a specific probability (the confidence) of including the population parameter with repeated sampling.

A confidence interval is defined as an interval estimate consisting of a range of values (with a lower and upper bound) constructed to have a specific probability (the confidence) of including the true population parameter with repeated sampling.

Examples of a true population parameter are the difference in mean total cholesterol measures for patients receiving a statin versus placebo therapy, or the difference in proportions of patients undergoing surgery type A versus B who survive more than 30 days following traumatic injuries.

We couple a point estimate of the parameter with a confidence interval. The confidence interval is used to quantify the uncertainty and variability in our estimate of the true parameter.



Let's consider an experiment to compare a response rate between an experimental group and a placebo control group.

The response rate is 60% among the experimental patients and is 52% among the control group.

The absolute benefit increase is 8%. A corresponding confidence interval for the difference in event rates is 3% to 13%. The p-value testing the null hypothesis that the difference in response rates is 0 is 0.001.

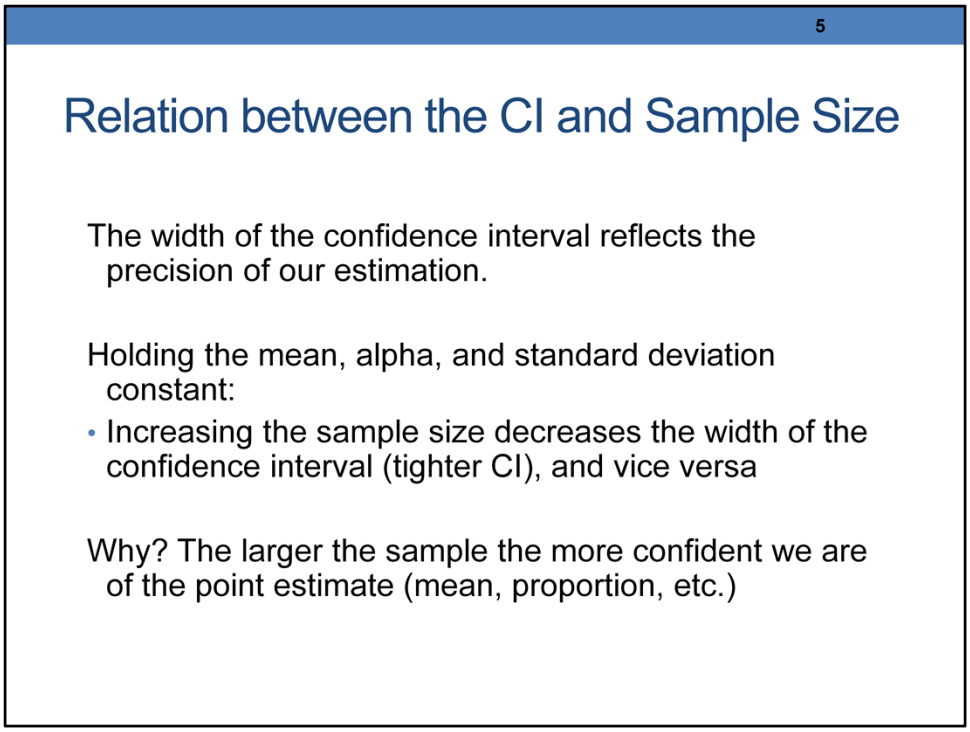
Based on an assumed alpha level of 0.05, we will reject the null hypothesis because the p-value = 0.001 < 0.05 = alpha-level. We conclude that there is a significant difference between the intervention and control groups in terms of the response rate.

We can also use the confidence interval to perform the hypothesis test. In this case, the interval 0.03 to 0.13 does not include 0 and therefore, we can conclude that the true different in response rates is significantly different from 0.

If the confidence interval for a difference does not include zero, then the difference is statistically significant.

Now, let's discuss clinical significance. If we state a priori that there needs to be at

least a 15% difference in response rates before the difference would be clinically important. We can compare the confidence interval to the threshold of 15% and will conclude that because the interval does not lie beyond 15%, the difference is not clinically significant.

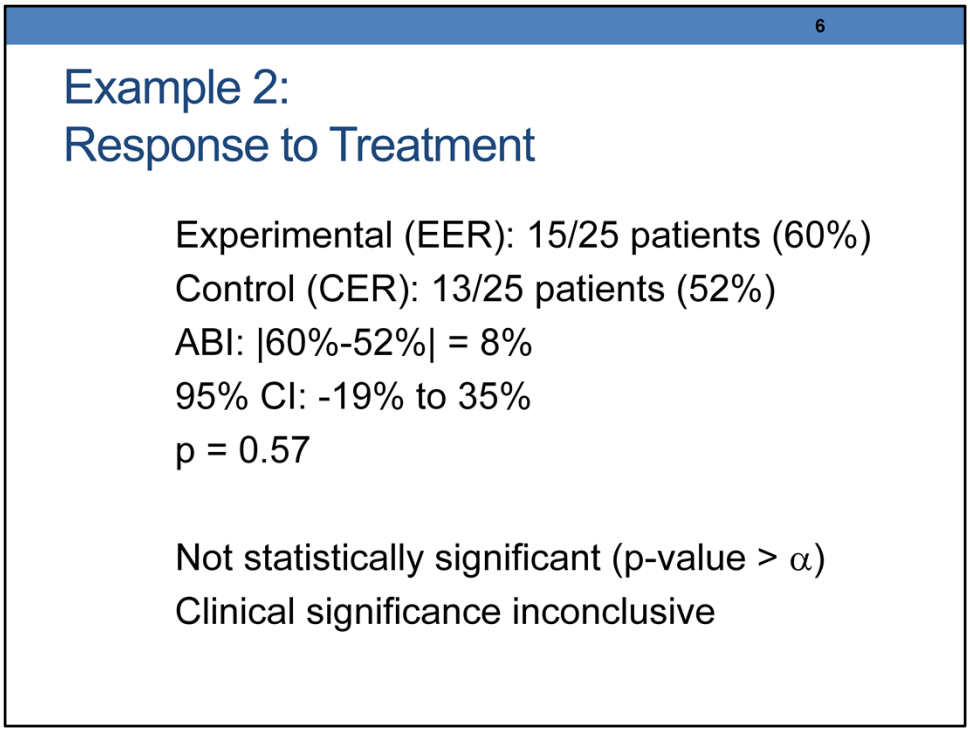


The width of the confidence interval reflects the precision of our estimation. The narrower the confidence interval, the more precise the estimate is.

In practice, there are several factors that influence the width of the confidence interval.

Holding the mean, alpha, and standard deviation constant, increasing the sample size decreases the width of the confidence interval, and vice versa.

Why? The larger the sample the more confident we are of the point estimate (mean, proportion, etc.). Our estimation is more precise when analyzing data from larger sample sizes.



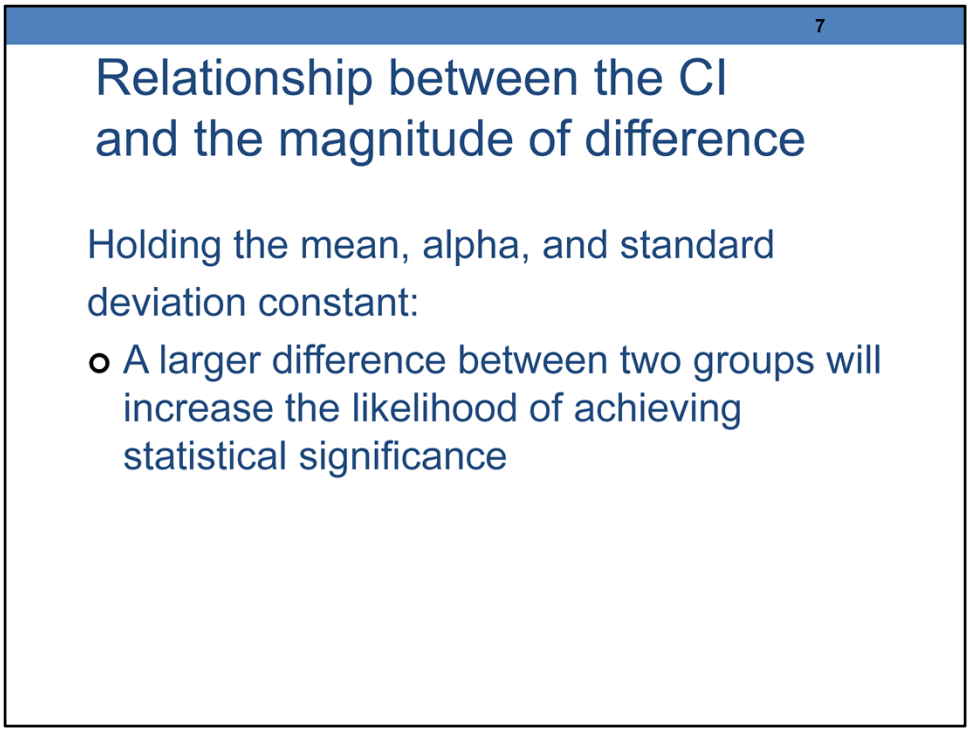
Let's consider a second example where we compare the response rate between an experimental and a control group. In this example, the response percentages are the same as in the first example; however, the sample size in each group is only 25 instead of 800 as in the first example.

The Experimental event rate is (EER): 15/25 patients (60%)
The Control event rate is (CER): 13/25 patients (52%)
The absolute benefit increase is: |60%-52%| = 8%
The resulting confidence interval is 95% CI: -19% to 35% and the p-value = 0.57.

We see that with the smaller sample size, the confidence interval is much wider. It is still centered at a difference of 8% as in the first example, but is much wider in this example.

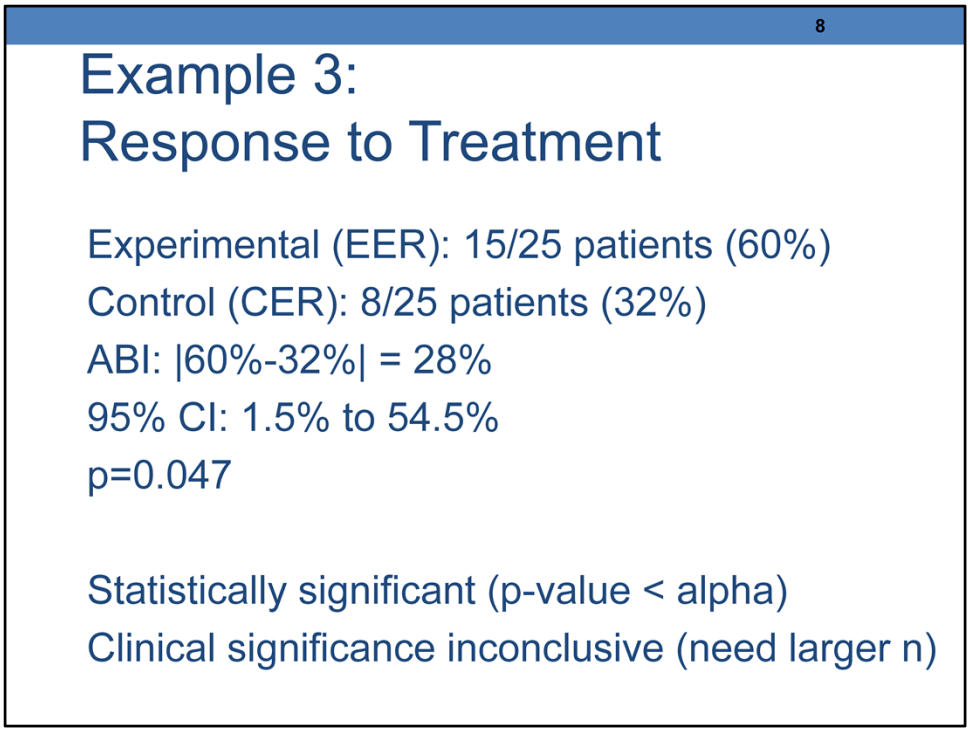
Based on the p-value, we conclude that the difference is not statistically significant (p-value > □). We also reach this same conclusion based on the confidence interval. The confidence interval includes 0 and therefore, the difference between the groups is not statistically significant.

Furthermore, given that the result is not statistically significant, the result is clinically inconclusive.



The magnitude of the difference impacts the location of the center of the confidence interval.

If we hold the mean, alpha level and standard deviation constant, a larger difference between the two groups results in a confidence interval that is shifted away from the null hypothesis value of no difference and increases the likelihood of achieving statistical significance.



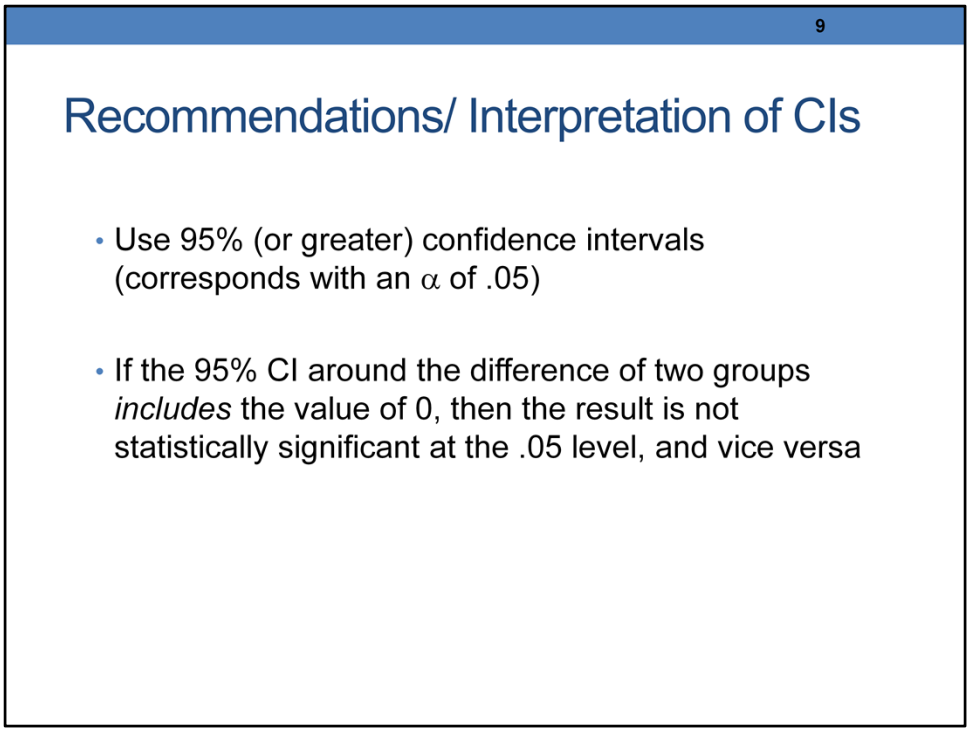
Let's consider a third example where we compare the response rate between an experimental and a control group. In this example, the per-group sample size remains at 25 while the difference in response rates increases from 8% to 28% in absolute value..

The Experimental event rate is (EER): 15/25 patients (60%)
The Control event rate is (CER): 8/25 patients (32%)
The absolute benefit increase is: |60%-32%| = 28%
The resulting confidence interval is 95% CI: 1.5% to 54.5% and the p-value = 0.047.

We see that the confidence interval does not include 0 and therefore, the difference is statistically significant.

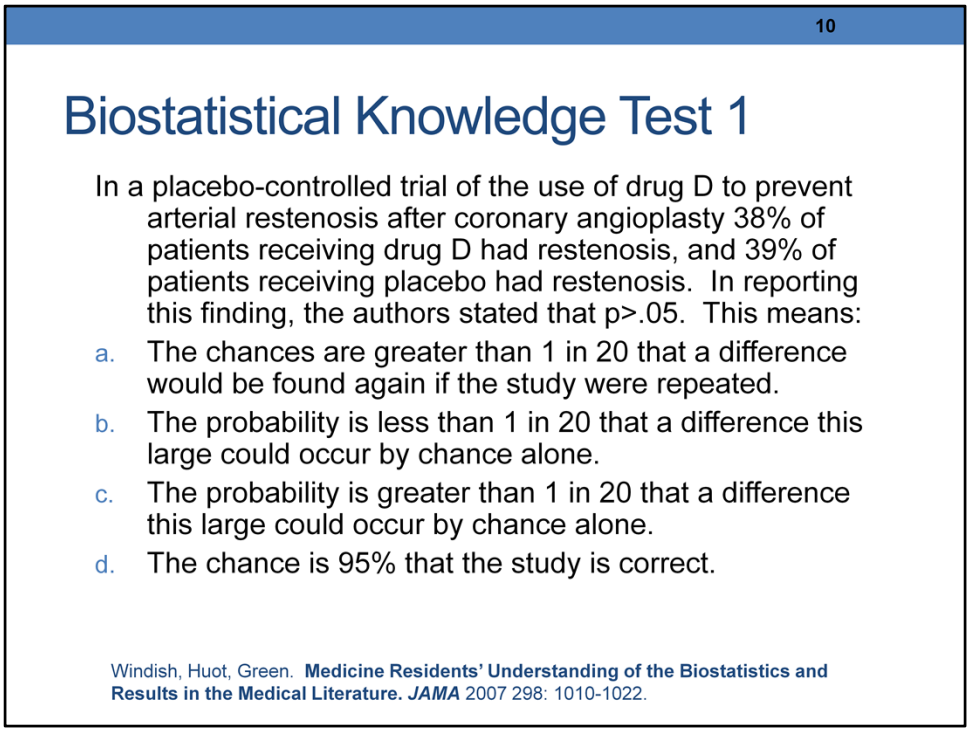
Based on the p-value, we conclude that the difference is statistically significant (p-value < □). We also reach this same conclusion based on the confidence interval. The confidence interval does not include 0 and therefore, the difference between the groups is statistically significant.

Given the small sample size, the confidence interval is still very wide and includes values lower than 15%; therefore, the result is still clinically inconclusive.



In practice, we typically report 95% confidence intervals, which correspond to alpha levels of 5%. If we report results using a 99% confidence interval, the interval will be wider than that from a 95% confidence level (i.e., we are more confident in including the true parameter and therefore, the interval must be wider). The alpha level for a 99% confidence interval is 1%.

When considering difference in means or proportions, if the 95% CI around the difference of two groups includes the value of 0, then the result is not statistically significant at the .05 level, and vice versa.



Now, let's consider an example problem.

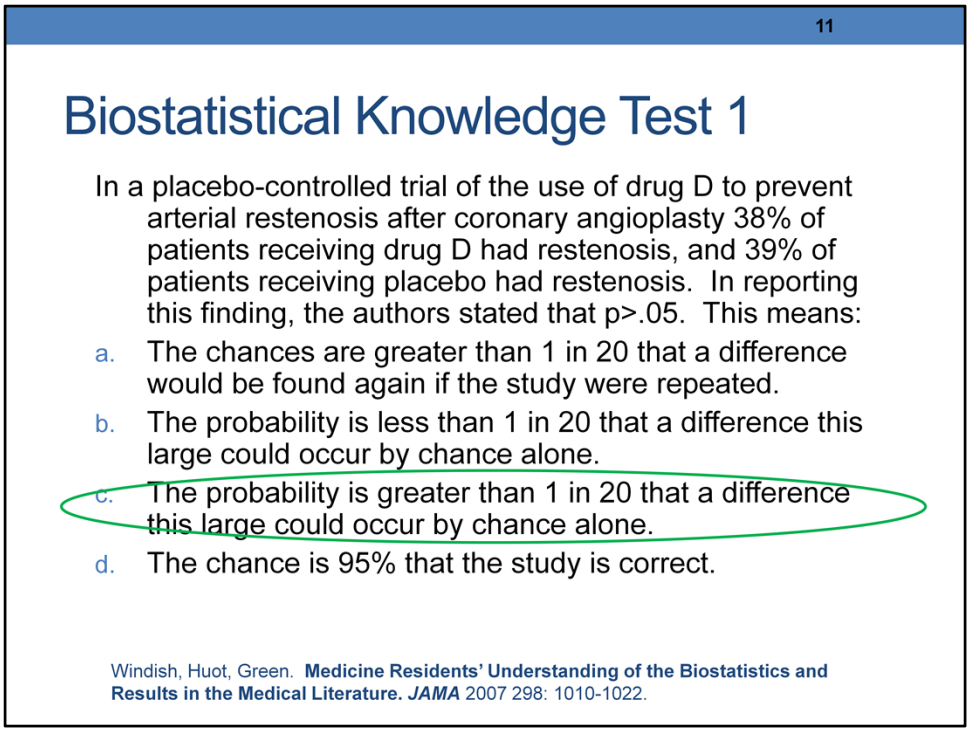
In a placebo-controlled trial of the use of drug D to prevent arterial restenosis after coronary angioplasty 38% of patients receiving drug D had restenosis, and 39% of patients receiving placebo had restenosis. In reporting this finding, the authors stated that p>.05. This means:

The chances are greater than 1 in 20 that a difference would be found again if the study were repeated.

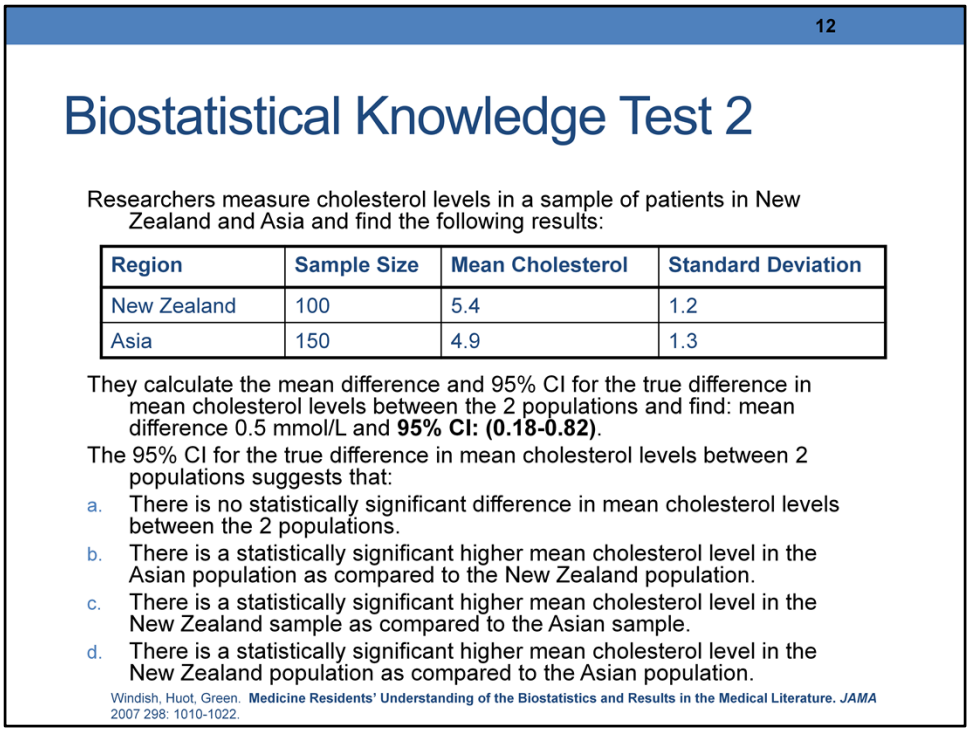
The probability is less than 1 in 20 that a difference this large could occur by chance alone.

The probability is greater than 1 in 20 that a difference this large could occur by chance alone.

The chance is 95% that the study is correct.

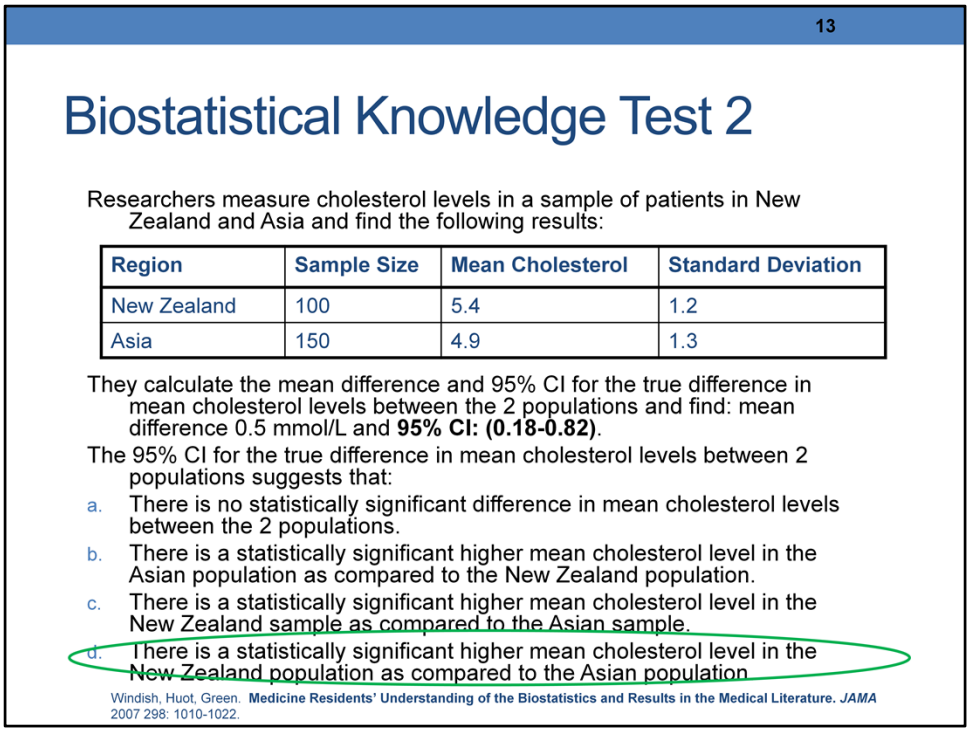


The correct answer is c: The probability is greater than 1 in 20 that a difference this large could occur by chance alone.



Let's consider a second question.

Researchers measure cholesterol levels in a sample of patients in New Zealand and Asia and find the following results presented on this slide. They calculate the mean difference and 95% CI for the true difference in mean cholesterol levels between the 2 populations (New Zealand minus Asia) and find: mean difference 0.5 mmol/L and 95% CI: (0.18-0.82).
The 95% CI for the true difference in mean cholesterol levels between 2 populations suggests that:
There is no statistically significant difference in mean cholesterol levels between the 2 populations.
There is a statistically significant higher mean cholesterol level in the Asian population as compared to the New Zealand population.
There is a statistically significant higher mean cholesterol level in the New Zealand sample as compared to the Asian sample.
There is a statistically significant higher mean cholesterol level in the New Zealand population as compared to the Asian population.



The correct answer is: There is a statistically significant higher mean cholesterol level in the New Zealand population as compared to the Asian population.

We estimate that the mean in New Zealand is 0.5 mmol/L higher than in Asia (based on the manner in which the difference was calculated with the New Zealand mean minus the Asia mean). The confidence interval ranges from 0.18 to 0.82. Given that the interval does not include 0, the result is statistically significant.

Example: Antihypertensive Treatment

- Aim: Compare two antihypertensive strategies for lowering blood pressure
  - Double-blind, randomized study
  - 5 mg Enalapril + 5 mg Felodipine ER to 10 mg Enalapril
  - 6-week treatment period
  - 217 patients
  - *AJH*, 1999;**12**:691-696

Now, let's consider an example comparing the mean blood pressure between two treatment groups – one that received single agent Enalapril and one that received dual therapy with Felodipine plus Enalapril.

Example

- Randomized clinical trial
- After 6 weeks of therapy, changes (baseline minus post-treatment) in DPB were determined.
- How can we determine if the efficacy differs between the single agent and combined therapies?

	Diastolic Blood Pressure (mm Hg)	
	Mean Change	**Standard Deviation**
Enalapril + Felodipine (n=109)	10.6	8.1
Enalapril (n=108)	7.4	6.9

The changes in blood pressure, relative to baseline, are shown in this table. Among patients receiving combination therapy, the mean change is 10.6 mmHg (reduction of 10.6 mmHg at post-treatment relative to baseline) and among those receiving single agent Enalapril, the mean change is 7.4 mmHg.

How can we determine if the efficacy differs between the single agent and combined therapies?

Example

- Statement from *AJH*
  - "The group randomized to 5 mg enalapril + 5 mg felodipine ER had a significantly greater reduction in trough DBP after 6 weeks of blinded therapy (10.6 mm Gh) than the group randomized to 10 mg enalapril (7.4 mm Hg, P<0.01)."
  - What does P<0.01 mean?
    - Assuming that the 2 therapies are equally effective, there is less than a 1% chance that we would have observed treatment differences as large or larger than we did.

In the manuscript, the authors state: "The group randomized to 5 mg enalapril + 5 mg felodipine ER had a significantly greater reduction in trough DBP after 6 weeks of blinded therapy (10.6 mm Hg) than the group randomized to 10 mg enalapril (7.4 mm Hg, P<0.01)."

What does P<0.01 mean?

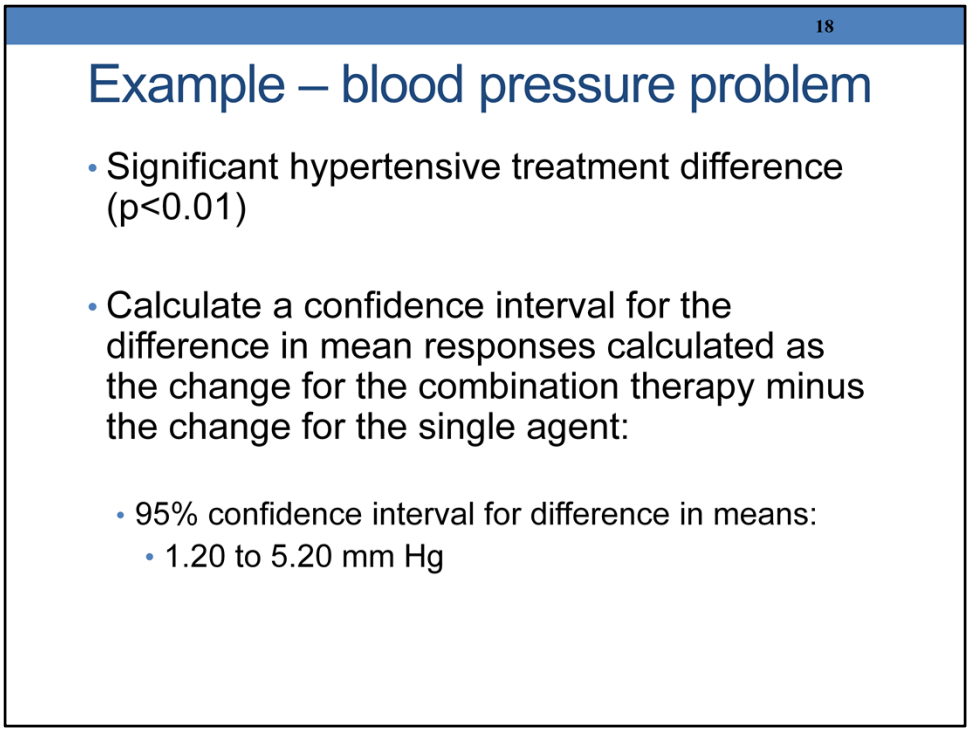
Assuming that the 2 therapies are equally effective, there is less than a 1% chance that we would have observed treatment differences as large or larger than we did.

Estimation vs. Hypothesis Testing

- Significant p-value not the same as clinical significance
- P-value may not give all information needed for interpretation: how large is effect?
- Confidence intervals give more information
- Some authors display both p-values and CIs

We have now learned two methods for testing hypotheses, one based on p-values and the other based on confidence intervals.

Keep in mind that a significant p-value is not the same as clinical significance. A P-value may not give all information needed for interpretation, for example, with a p-value alone, we do not know how large the effect is. Confidence intervals provide more information because we can see a range of reasonable estimates for the true parameter. In practice, some authors display both p-values and Cis.

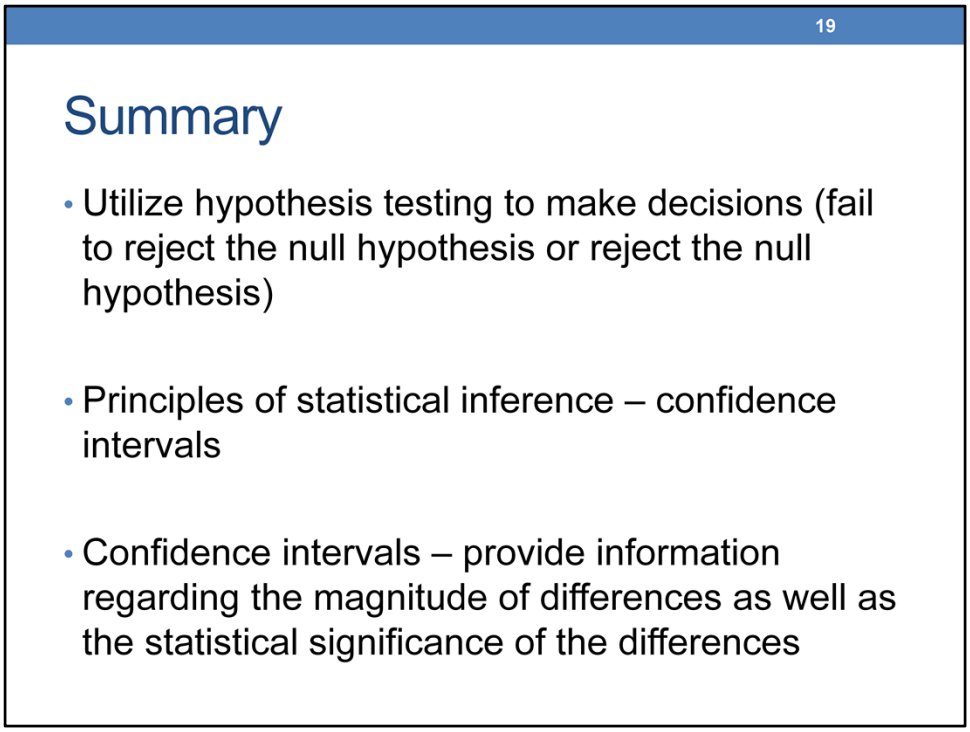


As an example, from the hypertension problem, we conclude that the effect of the dual agent therapy is significantly different from the effect of the single agent therapy because the p-value of <0.01 is lower than our assumed alpha level of 0.05.

Furthermore, we can calculate a confidence interval for the difference in mean responses calculated as the change for the combination therapy minus the change for the single agent:

95% confidence interval for difference in means:
1.20 to 5.20 mm Hg

Given that the confidence interval does not contain 0 and lies entirely above 0, we conclude that the combination therapy is more effective than the single agent therapy in lowering blood pressure values. We can infer the direction of the effect based on the sign of the difference calculated as the reduction under combination therapy minus the reduction under the single agent.



In summary, we can utilize hypothesis testing methods, either based on p-values or confidence intervals, to make decisions to fail to reject the null hypothesis or reject the null hypothesis.

In practice, we will couple a point estimate of a parameter with a confidence interval where the confidence interval provides information regarding the variability and uncertainty of our estimate.

Confidence intervals are useful in practice as they provide information regarding the magnitude of differences as well as the statistical significance of the differences