Confidence Intervals and P-Values

What They Mean — and What They Do Not

Most APPs were taught that $p < 0.05$ means a finding is real. It does not. A p-value and a confidence interval are different tools that answer different questions. Knowing the difference changes how you read every clinical trial.

Part 1 — The P-Value

A p-value is the probability of observing results this extreme — or more extreme — **if the null hypothesis were true**.

> **Example:** A trial reports HR = 0.78, $p = 0.03$ for a new beta-blocker vs. placebo in HFrEF.
>
> This means: *If the drug had zero effect*, there is only a 3% chance of seeing a hazard ratio this far from 1.0 by chance. It does **not** mean there is a 97% probability the drug works.

What the p-value tells you	What it does not tell you
Whether to reject the null hypothesis at a given threshold	Whether the null hypothesis is true or false
How unlikely the data are *if* there is no effect	The probability that the finding is real
Statistical significance	Clinical importance
That the result was unlikely by chance	That the effect size is meaningful

The threshold is arbitrary. $p < 0.05$ is a convention, not a law of nature. A result with $p = 0.049$ and $p = 0.051$ are essentially the same finding.

Part 2 — The Confidence Interval

A 95% CI is the range in which the true effect plausibly lies, based on your sample. Across many repeated studies using the same method, 95% of the calculated intervals would contain the true population value.

> **Common misread:** “There is a 95% probability the true value is in this interval.”
>
> **Correct read:** “This interval was calculated by a method that captures the true value 95% of the time across repeated sampling.”

Reading CIs in practice:

Measure	Crosses the null?	Interpretation
RR, OR, HR (ratios)	CI crosses **1.0**	Not statistically significant
RR, OR, HR (ratios)	CI entirely below or above 1.0	Statistically significant
Mean difference, NNT	CI crosses **0**	Not statistically significant
Mean difference, NNT	CI does not include 0	Statistically significant

What the CI tells you	What it does not tell you
The range of plausible effect sizes	The exact true effect
Precision (narrow CI = more precise estimate)	Whether the effect is clinically meaningful
Statistical significance (does it cross the null?)	Whether the study was well designed
Both direction and magnitude of the effect	That the point estimate is correct

Part 3 — Statistical vs. Clinical Significance

Large trials have statistical power. That is a feature. It also means they can detect effects too small to matter clinically.

> **Example:** A trial of 40,000 patients reports a blood pressure reduction of 1.2 mmHg, $p = 0.001$, 95% CI [0.5–1.9 mmHg].
>
> Statistically significant? **Yes.** Clinically meaningful? **No.** A 1.2 mmHg reduction has no measurable impact on cardiovascular outcomes in practice.

Before citing a “significant” result, ask:

- What is the **absolute risk reduction** (ARR), not just the relative risk?
- What is the **number needed to treat** (NNT)?
- Is the effect size large enough to change your management?
- Does the confidence interval include clinically trivial values even at its upper bound?

CLINICAL RULE

A p-value tells you whether to reject the null hypothesis.

A confidence interval tells you the range of plausible effect sizes.

The CI is almost always more useful. Before citing $p < 0.05$, ask: is the effect size clinically meaningful?