Primer on 95% Confidence Intervals

Readers frequently face questions about the role of chance in a study's results. The traditional approach has been to consider the probability that an observed result is due to chance–the *P* value. However, *P* values provide no information on the results' precision—that is, the degree to which they would vary if measured multiple times. Consequently, journals are increasingly emphasizing a second approach: reporting a range of plausible results, better known as the 95% confidence interval (CI). This Primer reviews the concept of CIs and their relationship to *P* values.

Assessing the Role of Chance

There are two basic statistical methods used to assess the role of chance: hypothesis testing (which results in a *P* value–the subject of last issue's Primer) and 95% CIs. As shown in Figure 1, both use the same fundamental inputs.
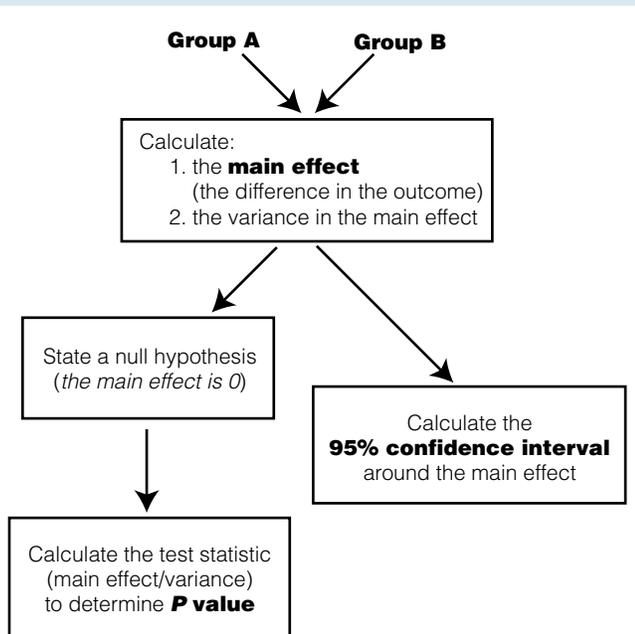

FIGURE 1. Statistical approach for comparing two groups.

Consider a study of a new weight loss program: Group A receives the intervention and loses an average of 10 pounds, whereas group B serves as a control and loses an average of 3 pounds. The main effect of the weight loss program is therefore estimated to be a 7-pound weight loss (on average).

But readers should recognize that the true effect of the program may not be exactly a 7-pound weight loss. Instead, the true effect is best represented as a range. What is the range of effects that might be expected just by chance? That is the question addressed by a 95% CI. In this example the study abstract might read:

> The mean weight loss was 10 pounds for patients in the intervention group and 3 pounds for patients in the control group, resulting in a mean difference of 7 pounds and a 95% CI of 2 to 12. In other words, 95% of the time the true effect of the intervention will be within the range from 2 to 12 pounds.

To conceptualize the more formal definition of a 95% CI, it is useful to consider what would happen if the study were repeated 100 times. Obviously, not every study would result in a 7-pound weight loss in favor of the intervention. Simply due to the
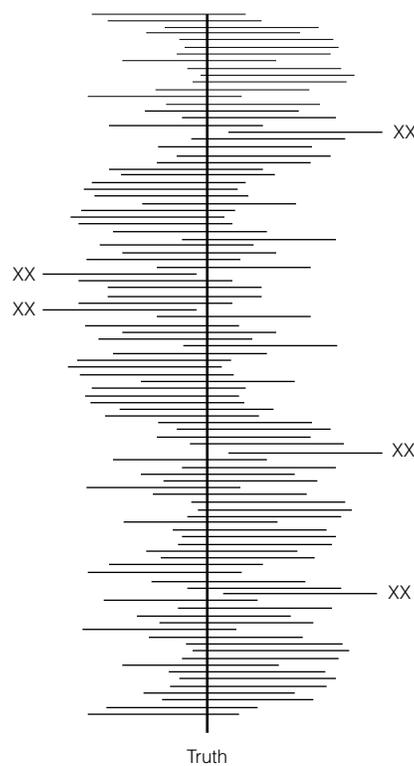

FIGURE 2. Every study can have a 95% CI.

play of chance, weight loss would be greater in some studies and less in others, and some studies might show that the controls lost more weight. As shown in Figure 2, we can generate a 95% CI for each study.

Note that for 95 out of 100 studies, the CI contains the truth (and 5 times out of 100 it does not). This example helps explain the formal definition of a 95% CI: "The interval computed from the
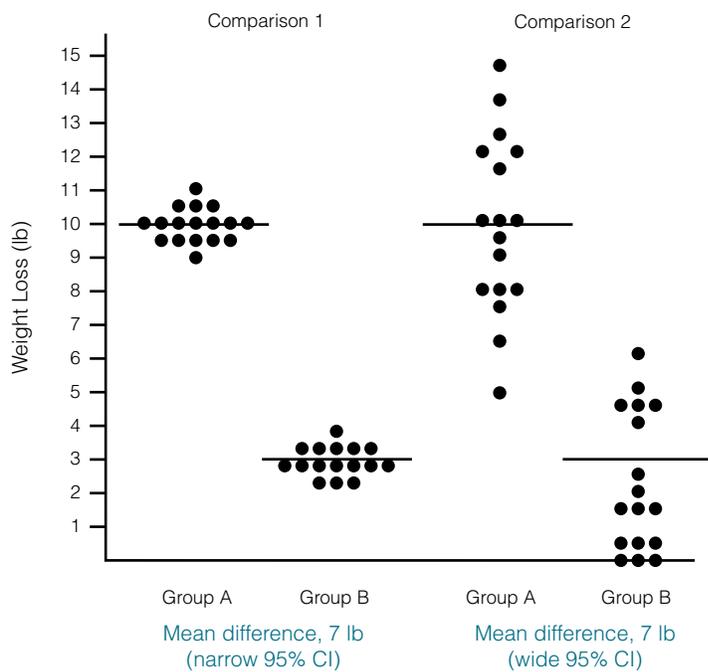
FIGURE 3. More diversity in weight loss equals a larger 95% CI. The horizontal bars represent group means.

Factors That Influence 95% CIs

Confidence intervals really are a measure of how precise an estimated effect is. The range of a CI is dependent on the two factors that cause the main effect to vary:

1) *The number of observations*. This factor is largely under the investigator's control. A 7-pound difference observed in a study with 500 patients in each group will have a narrower CI than a 7-pound difference observed in a study with 25 patients in each group.

2) *The spread in the data* (commonly measured as a standard deviation). This factor is largely outside the investigator's control. Consider the two comparisons in Figure 3. In both cases, the mean weight loss in group A is 10 pounds and the mean weight loss in group B is 3 pounds. If everybody in group A loses

sample data which, were the study repeated multiple times, would contain the true effect 95% of the time."

about 10 pounds and everybody in group B loses about 3 pounds, then the CI will be narrower (left part of figure) than if individual weight changes are spread all over the map (right part of figure).

Readers will occasionally encounter CIs calculated for other confidence levels (e.g., 90% or 99%). The higher the degree of confidence, the wider the confidence interval. Thus, a 99% CI for the 7-pound difference would have to be wider than the 95% CI for the same data.
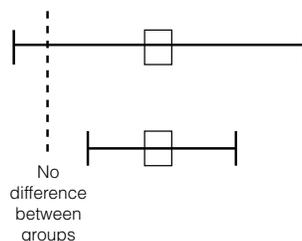
Relationship between 95% CIs and *P* values

Information about the *P* value is contained in the 95% CI. As shown in Figure 4, the *P* value can be inferred based on whether the finding of "no difference" falls within the CI.

So, given a CI of 2 to 12 pounds for the 7-pound difference, one could infer that the *P* value is less than 0.05. Alternatively,

If the 95% CI includes no difference between groups, then the *P* value is > 0.05.

If the 95% CI does not include no difference between groups, then the *P* value is < 0.05.

No difference between groups

FIGURE 4. Relationship between *P* value and 95% CI.

given a CI of –3 to 17 pounds for the 7-pound difference, one could infer that the P value is greater than 0.05. If the CI terminates exactly on no difference, such as 0 to 14 pounds, then the P value is exactly 0.05.

Remember that the value for no difference depends on the type of effect measure used. When the effect measure involves a subtraction, the value for the difference is 0. When the effect measure involves a ratio, the value for no difference is 1. As shown in Table 1, readers must pay careful attention to this in order to reliably interpret the CI.

Although P values and 95% CIs are related, CIs are preferred because they convey information about the range of plausible effects. In other words, the CI provides the reader with some sense of how precise the estimate of the effect is. This is a valuable dimension that is not contained within a P value.

But, like P values, 95% CIs do not answer two critical questions: 1) Is the result correct? 2) Is the observed effect "important"? To answer the first question, readers must seek other data and evaluate the possibility of systematic error (bias). To answer the second, they must rely on their own clinical judgment.

TABLE 1
Examples Demonstrating 95% CIs and P Values

EXAMPLE	EFFECT MEASURE	VALUE FOR NO DIFFERENCE	CI INCLUDES NO DIFFERENCE?	STATISTICALLY SIGNIFICANT? ($P < 0.05$)
The average weight loss was 7 lbs (95% CI, –3 to 17)	Difference in means	0	Yes	No
42% absolute reduction in the need for intubation (95% CI, 7% to 70%)	Difference in proportions	0	No	Yes
The relative risk for cancer was 2.3 for smokers compared with nonsmokers (95% CI, 1.8 to 3.0)	Relative risk	1	No	Yes
The odds ratio for readmission was 0.8 for managed care patients (95% CI, 0.3 to 1.2)	Odds ratio	1	Yes	No

A compendium of **ecp** *primers from past issues can be viewed and/or requested at http://www.acponline.org/journals/ecp/primers.htm.*