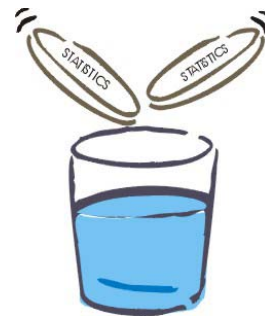


Statistics in Divided Doses



July 2005 No 8

Confidence intervals

Contents

- Some revision of confidence intervals
- Applying confidence intervals
- Factors affecting the size of the confidence interval
- Confidence intervals and P values
- Interpretation of confidence intervals

Some revision of confidence intervals

What do we already know about confidence intervals?

Observations from samples of subjects in clinical trials are used to draw **inferences** about the population from which those samples are drawn (*SiDD 2*). Due to the effects of **random variation** between the subjects and measurement errors, such observations have an inherent level of uncertainty, which can usually be quantified by calculating the relevant **confidence interval** (*SiDD 3, 6, 7*).

What is the purpose of a confidence interval?

Confidence intervals indicate the precision (or imprecision) with which a **study sample estimates** the true **population value**. They have an important role whenever we wish to apply the results of a clinical study to the general population.

What is meant by 'study sample estimates'?

The 'study sample estimates' are obtained by analysing the data provided by a study sample. An example would be the mean reductions in systolic blood pressure in two groups of hypertensive patients receiving different drug treatments. In practice, the primary effect of interest would be the **mean difference** in systolic blood pressure reduction, since this would provide an estimate of the comparative effectiveness of the two treatments.

What is meant by the 'population value'?

This is the average response that would be obtained if the intervention being studied were applied to every person in the population from which the study sample was drawn.

What is meant by a 95% confidence interval (CI)?

This is the interval, computed from the study sample data, within which we can expect the population value to lie **with a 95% level of probability** (i.e. we can be

95% **confident** that the population value lies within this interval).

Example - Comparing antihypertensives

Two groups of men, diagnosed as having a rare type of hypertension, were randomised to receive either drug A or drug B in a clinical trial designed to compare the antihypertensive effects of the two drugs. The results are shown in *Table 1*.

Table 1

Observations	Drug A	Drug B
Number of subjects	50	50
Mean reduction (mmHg) in systolic BP	45	35
Standard deviation (SD) (mmHg)	20	16

The difference in systolic blood pressure reduction is 10mmHg in favour of drug A. In order to calculate the 95% CI we first need to calculate the standard error of the difference (SED) between the two drugs (*SiDD 5*);

$$\text{SED} = \sqrt{\left(\frac{SD_1^2}{n_1}\right) + \left(\frac{SD_2^2}{n_2}\right)}$$
$$\text{SED} = \sqrt{\left(\frac{20^2}{50}\right) + \left(\frac{16^2}{50}\right)} = \sqrt{\left(\frac{400}{50}\right) + \left(\frac{256}{50}\right)}$$
$$= \sqrt{(8 + 5.12)} = 3.6$$

Using this SED and a *t* value of 1.984 (from Tables of the Student *t* distribution for 98 degrees of freedom - our total sample size minus 2) (*SiDD 6*) the 95% CI, is:

Difference in means - [1.984 x SED] to
Difference in means + [1.984 x SED]

$$10 - (1.984 \times 3.6) \text{ to } 10 + (1.984 \times 3.6)$$
$$= 2.85 \text{ to } 17.14$$

= 3 to 17 (to nearest whole numbers)

We can conclude with 95% certainty that the true (population) **additional** reduction in systolic blood pressure achieved with drug A compared with drug B is between 3mmHg and 17mmHg (i.e. 95% CI 3 to 17mmHg).

Does this result indicate that values outside the interval of 3mmHg to 17mmHg are excluded?

No, they are simply less likely. The real difference in the effects of the two drugs **could** be less than 3mmHg or more than 17mmHg, but the probability that this is the case is less than 5%.

Applying confidence intervals

Can a confidence interval be attached to any estimate?

Confidence intervals can be calculated for most types of statistical analysis. They can be applied to means and their differences as well as to medians and their differences. They can also be applied to other types of measures for example, proportions, correlations, regression coefficients and survival curves.

Confidence intervals are central to all forms of meta-analysis.

Are some parts of the 95% CI more likely to contain the population value than others?

Yes. In the above example, the population difference in blood pressure reduction is more likely to be near the centre of the confidence interval than towards its outer limits. The sample mean (which is at the exact centre of the confidence interval) is the best available estimate of the population value. However, when interpreting the confidence interval, we cannot exclude the possibility of the population value lying at or near the ends of the confidence interval.

When should we use confidence intervals?

Confidence intervals should be presented whenever we are using the results of a study to draw inferences about the average response to a treatment in the parent population. In the above example, two drugs were studied in patients with hypertension of defined severity and type. The results give an estimate of the likely average response in the many thousands of patients in the general population with the same disorder. When determining the clinical effect, the estimated difference in blood pressure is judged in terms of its size. The confidence interval in our example is compatible with both a clinically important (17mmHg) and a clinically minimal (3mmHg) reduction in systolic blood pressure.

Why do we usually choose a 95% interval for confidence intervals?

Convention! Intervals can be calculated for any desired level of confidence; 95% is usually chosen because it conforms to the customary acceptance of a 5% P value as the threshold for statistical significance.

Factors affecting the size of the confidence interval

What factors affect the size of the confidence interval?

- the size of each sample
- the variability (scatter) in the responses of the subjects in each treatment group
- the level of confidence required.

How does sample size affect the confidence interval?

Larger samples yield smaller confidence intervals. This agrees with our everyday experience and our intuitive feeling that larger samples are more informative and reliable than smaller ones. The benefit of further increasing the sample size reduces as the sample size increases; i.e. increasing the sample size from 20 to 40 produces a greater reduction in the width of a confidence interval than increasing the sample size from 520 to 540.

Is the confidence interval influenced by a difference in sample sizes within a clinical trial?

Yes - the narrowest confidence interval is obtained if the study subjects are allocated equally to each group. In practice, a small difference in the sample sizes will have little effect but a large disparity will tend to increase the confidence interval for a given difference between the two groups.

How does sample variability affect the confidence interval?

The less that individual values within samples differ from each other (i.e. the more consistent the responses to each treatment), the narrower will be the confidence interval for each sample mean, and for the difference between sample means.

Why does the level of confidence we choose affect the width of the confidence interval?

A confidence interval is determined by first calculating the standard error of the difference (SED) between the two groups then adding and subtracting a multiple of the SED to the mean difference; the value of this multiple is obtained by referring to Tables of the Student t distribution (SiDD 6 and the worked example above).

Example - to calculate the **90% CI** around the mean difference in systolic blood pressure reduction (10mmHg) in our earlier example;

- First multiply the SED (3.6) by the value of t corresponding to a significance level of 0.10 at 98 degrees of freedom i.e. 1.661.
- Second add and subtract this multiple to and from the mean difference, giving a 90% CI of 4mmHg to 16mmHg.

Table 2 shows the 90%, 95%, 99% and 99.9% CIs around our treatment estimate of 10mmHg.

Table 2

Chosen confidence level (%)	Level of significance, P value (2-sided)	Calculated confidence interval (mmHg)	t value
90	0.1	4 to 16	1.661
95	0.05	3 to 17	1.984
99	0.01	1 to 19	2.627
99.9	0.001	-2 to 22	3.393

Clearly, as we increase the confidence level, the confidence interval broadens i.e. the more certain we want to be that our interval includes the population value, the wider the confidence interval will be.

Confidence intervals and P values

Is there a link between confidence intervals and P values?

There is very close link between confidence intervals and hypothesis (significance) tests. The P value from a significance test is the probability that the observed (or even larger) difference between the two study samples could have occurred if the true population difference is zero (i.e. no difference at all).

Thus, when $P < 0.05$;

- the probability of the observed difference occurring when the true difference is zero is 5% or less
- we can be 95% confident that the true population difference is inside the sample's 95% confidence interval
- if we continually repeated our study, 95% of the samples would produce confidence intervals that include the true population difference and 5% of the samples would not.

If the population value were zero,

- 5% of our studies on average would produce 95% confidence intervals that do not include zero
- the chance of a sample from the one study we carry out not including zero, is 5% or less.

The confidence interval provides the same information as the P value, but as well as indicating that the population value is unlikely to be zero, it also indicates the range of plausible values for the population difference.

In many situations, the mathematical formula used to calculate 95% CI is exactly the same as that used to calculate the appropriate significance test statistic.

I am rather confused – please clarify by means of an example!

Using our example of drug A vs. drug B in hypertensive men, we calculated that the 95% CI was 3mmHg to 17mmHg. This excludes zero (which would correspond to no difference in blood pressures); the P value from a significance test will be less than 0.05 (see Table 2).

In Table 2 the 99% CI also excluded zero so the P value will be less than 0.01. However, the 99.9% confidence interval included zero, so the P value is greater than 0.001.

We can check this easily by computing the Student *t* test for unpaired samples (*SiDD6*).

$$t = \frac{\text{difference in mean systolic blood pressure}}{\text{standard error of that difference}}$$

$$\text{Thus, } t = \frac{10}{3.6} = 2.778.$$

Referring to a Table of the Student *t* distribution with 98 degrees of freedom, we find that a value of $t = 2.778$ corresponds to a P value of 0.007 which, as

predicted by the confidence intervals, is between 0.01 and 0.001.

How should we report these results?

We could report them, in summary version, as follows¹;

- Mean difference in systolic BP reduction = 10mmHg
- 95% CI 3 to 17mmHg
- $t = 2.778$
- degrees of freedom = 98
- $P = 0.007$.

Are confidence intervals sometimes misused in clinical trial reports?

Unfortunately, yes. A common error is to present both sample means with their respective confidence intervals and to examine the degree to which the intervals overlap (*SiDD 3*). Although this provides some indication of comparability, the correct approach is to calculate the difference between the sample means, the standard error of this difference, and then the confidence interval at a chosen level of probability (usually 95%). Comparing the confidence intervals of the two sample means becomes increasingly unreliable if the standard error of the two samples differ markedly.

Interpretation of confidence intervals

Do confidence intervals allow us to interpret trial results that show no difference in effect?

Yes. For example, repeating the calculation in our example with 20 patients in each group we now find that the 95% CI has increased to -2mmHg to 22mmHg. How do we interpret such data?

Firstly, the confidence interval now includes zero, which equates to no difference in effect between the two drugs. We are unable to reject the null hypothesis as $P > 0.05$ i.e. the outcome is not statistically significant and convention indicates that we must conclude that the drugs are equally effective.

How reliable is this conclusion? That depends on what size of difference in the effects of the two drugs would be considered clinically significant. Many clinicians would agree that an average difference in the effects of two anti-hypertensive drugs of 10mmHg or greater can be considered to be important (clinically significant). In this case, the 95% CI suggests that the difference between the two drugs we tested could be as great as 22mmHg in favour of drug A. Thus, we cannot discount the possibility of a clinically significant difference between the drugs, even though the difference was not statistically significant.

Would it be correct to describe such results as negative?

No. The correct interpretation is that the results of this trial are **inconclusive**.

¹ Altman DG *et al.* Statistics with confidence. 2nd edition, 2000; London; BMJ Books.