The Correct Interpretation of Confidence Intervals

Sze Huey Tan[1,2], *MSc, BEng*, Say Beng Tan[1,2,3], *PhD CStat*

[1]Division of Clinical Trials & Epidemiological Sciences, National Cancer Centre Singapore
[2]Centre for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore
[3]Singapore Clinical Research Institute

Confidence intervals (CI) are a key output of many statistical analyses, and have a critical role to play in the interpretation of estimates of parameters. Although they are commonly used, CIs are often misunderstood and incorrectly interpreted. In this article, we discuss how CIs should correctly be interpreted and also highlight some common misunderstandings associated with them.

RELATIONSHIP BETWEEN CONFIDENCE INTERVALS AND P-VALUES

CIs and p-values are closely related although they provide different information. While p-values are the outcome of hypothesis tests and indicate whether or not the sample data provide sufficient evidence to reject the null hypothesis (e.g. whether an observed difference between 2 treatments is because the 2 treatments really have different levels of efficacy, or whether it is due to random chance), CIs describe how uncertain our estimated difference is. Simply, CI indicates whether the estimate is a precise one or only a very "rough" estimate. For example, if the objective of a study is to test whether there is a difference in systolic blood pressure between 2 groups of people, the p-value only tells if there is a statistically "significant" or statistically "non-significant" difference in the systolic blood pressure between the 2 groups. It does not provide an estimate of what this difference is. To get such an estimate, we need to compute the mean difference in the systolic blood pressure, along with the associated CI, which tells us how precise (i.e. how "good") this estimate is in predicting the population level. A narrower CI will indicate a more precise estimate, while a wider CI indicates a less precise estimate.

If the 95% CI for the DIFFERENCE between the 2 groups contains the value 0, this means that the p-value will be greater than 0.05. Conversely, if the 95% CI does not contain the value 0, then the p-value will be strictly less than 0.05. The same applies when comparing groups using a ratio, such as an odds ratio or risk ratio. When using a RATIO instead of a DIFFERENCE, the situation of no difference between the 2 groups will be indicated by a value of 1 instead of 0. If the ratio equals to 1, the 2 groups are equal. Hence, if the 95% CI of the ratio contains the value 1, the p-value will be greater than 0.05. Alternatively, if the 95% CI does not contain the value 1, the p-value is strictly less than 0.05.

MANY VALUES OF CONFIDENCE INTERVALS AND HOW TO CALCULATE CONFIDENCE INTERVALS

CIs can be presented as 90% CI, 95% CI, 99% CI or any percentage (between 0% and 100%) CI of interest. Although 95% CI are commonly used in many applications, the choice of whether to use a 90% or 95% CI is somewhat arbitrary, and depends on the level of "confidence" that the investigator wishes to convey in his or her estimate. The calculation method of the CI depends on the estimate of interest (mean, median, proportion, odds ratio, etc). Typically, the expression of a CI for a mean will be:

[(sample mean) – (constant) x (SEM)] to
[(sample mean) + (constant) x (SEM)])

where "constant" takes the value of 2.58, 1.96 and 1.64 depending on whether we use the 99%, 95%

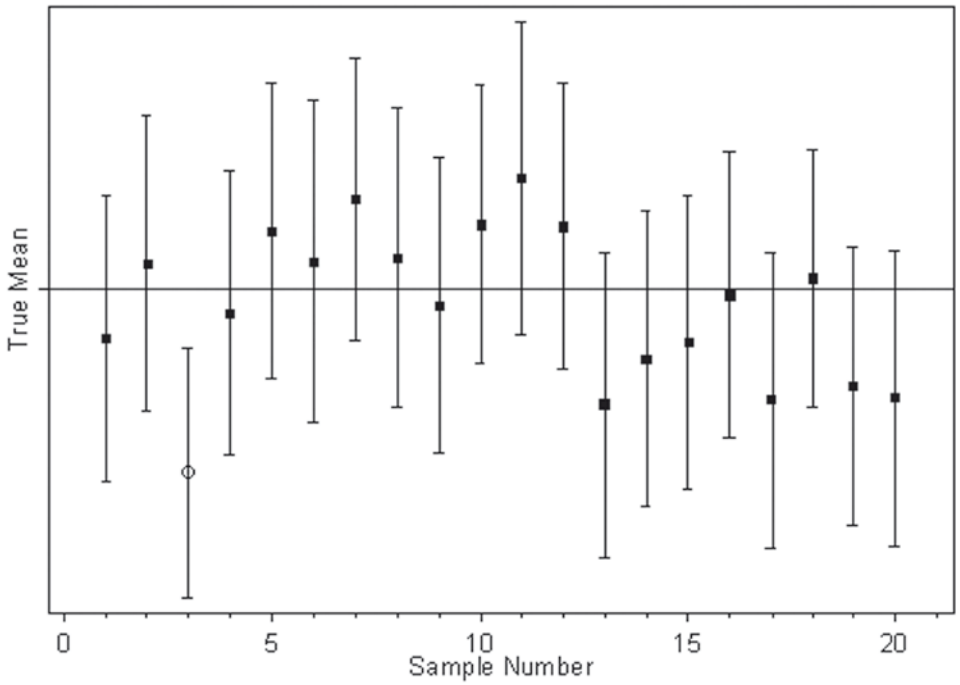

Fig. 1. 95% CI for the population mean for 20 independent samples drawn from the population.

and 90% CI respectively and SEM is the standard error of the mean. For a single study, 90% CI, 95% CI and 99% CI can be calculated and they differ in their width, with a 90% CI having a narrower width than a 99% CI. Using a hypothetical data with 100 observations of systolic blood pressure with a mean of 123mmHg, the corresponding 90% CI, 95% CI and 99% CI are (117.7 to 128.4), (116.6 to 129.5) and (114.5 to131.6) respectively. As illustrated, the width of the 90% CI is narrower than the 99% CI. Besides the choice of confidence level, the width of CIs is also related to the sample size of the study; a larger study sample size will give a narrower CI. Using the same example, if the study sample size is increased from 100 to 1,000 observations, the corresponding 95% CI will be (121.4 to 126.5), which is narrower than the 95% CI for the sample with 100 observations (116.6 to 129.5).

COMMON MISUNDERSTANDING OF THE TRUE MEANING OF CONFIDENCE INTERVALS

A common misunderstanding about CIs is that for say a 95% CI (A to B), there is a 95% probability that the true population mean lies between A and B. This is an incorrect interpretation of 95% CI because the true population mean is a fixed unknown value that is either inside or outside the CI with 100% certainty. As an example, let us assume that we know that the true population mean systolic blood pressure and it is 120mmHg. A study conducted gave us a mean systolic blood pressure of 105mmHg with a 95% CI of (95.5 to 118.9 mmHg). Knowing that the true population mean is 120mmHg it would be incorrect to say that there is a 95% probability that the true population mean lies in the 95% CI of (95.5 to 118.9mmHg) because we are certain that the 95% CI calculated did not contain the true population mean. A 95% CI simply means that if the study is conducted multiple times (multiple sampling from the same population) with corresponding 95% CI for the mean constructed, we expect 95% of these CIs to contain the true population mean. Fig. 1 illustrates this concept for a study conducted 20 times using 20 different samples from the population. Nineteen (95%) out of the 20 95% CI constructed contain the true population mean. Note that the 95% CI with a hollow circle did not contain the true population mean; therefore every CI either contains or does not contain the true population mean.

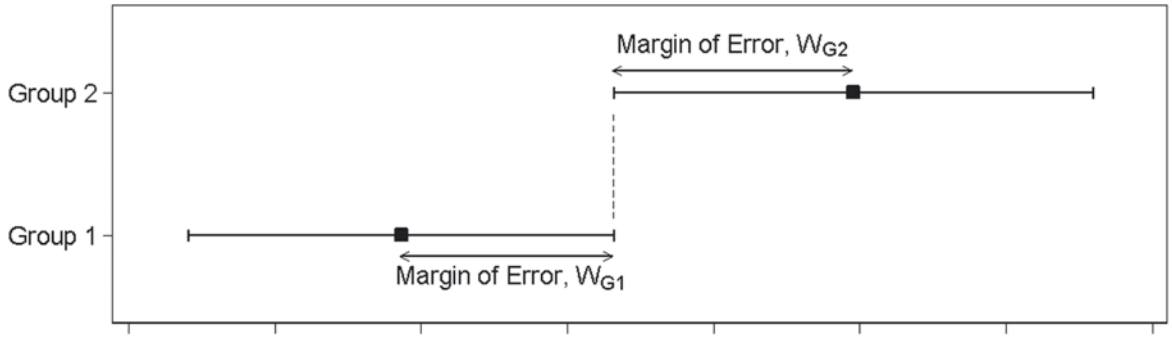

Fig. 2. Two independent groups with 95% CIs just touching.

COMMON MISINTERPRETATION OF OVERLAPPING CONFIDENCE INTERVALS WHEN COMPARING GROUPS

Another common misunderstanding occurs with the interpretation of the results when comparing 2 groups' population means using their 95% CIs. It is often misinterpreted that if the upper limit of the 95% CI of one group just touches the lower limit of the 95% CI of the other group (as illustrated in Fig. 2), the p-value for the difference between the 2 groups is 0.05. Contrary to what may be expected, the corresponding p-value on the difference between the 2 groups is NOT 0.05. In fact, the p-value is approximately 0.01 if both group sizes are at least 10 and the margins of error do not differ by more than a factor of 2[1]. The margin of error is the range from the sample estimate (the filled square) to the upper or lower limit of the CI as shown in Fig. 2. Even when the CIs are overlapping slightly, the p-value may still be less than 0.05. Hence just because two 95% CIs overlap, it DOES NOT necessarily imply that there is no statistically significant difference between the 2 groups. Taking as an example using hypothetical data where the mean and corresponding 95% CI of systolic blood pressure for 2 treatment groups were 123mmHg (95% CI: 116.6 to 129.5mmHg) and 135mmHg (95% CI: 126.6 to 142.6mmHg). As seen, there was slight overlap of the 2 CIs and it may be tempting to conclude that there was no significant difference in systolic blood pressure between the 2 treatment groups. This is however incorrect as a formal T-test conducted or a 95% CI on the difference between the 2 groups would give a p-value of 0.03 and 95% CI of (1.4 to 21.8 mmHg) which does not include the value 0. The better way to make use of CIs to conclude whether 2 groups are different or not would be to compute the 95% CI on the difference between the 2 groups and see whether this includes the value 0 or not.

Confidence intervals contain key information that is necessary for the proper interpretation of many statistical analyses. It is thus essential to understand and interpret confidence intervals correctly as a failure to do so could result in incorrect or misleading conclusions being drawn.

REFERENCE

1. Cumming G, Finch S. Inference by eye: Confidence intervals, and how to read pictures of data. Am Psychol. 2005 Feb-Mar;60(2):170–80.