How do I interpret a confidence interval?

Sheila F. O'Brien[1,2] *and Qi Long Yi*[1]

A 95% confidence interval (CI) of the mean is a range with an upper and lower number calculated from a sample. Because the true population mean is unknown, this range describes possible values that the mean could be. If multiple samples were drawn from the same population and a 95% CI calculated for each sample, we would expect the population mean to be found within 95% of these CIs. CIs are sensitive to variability in the population (spread of values) and sample size. When used to compare the means of two or more treatment groups, a CI shows the magnitude of a difference between groups. This is helpful in understanding both the statistical significance and the clinical significance of a treatment. In this article we describe the basic principles of CIs and their interpretation.

From the [1]Canadian Blood Services and the [2]School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, Ontario, Canada.

Address reprint requests to: Sheila F. O'Brien, Epidemiology and Surveillance, Canadian Blood Services, 1800 Alta Vista Drive, Ottawa, ON, Canada K1G 4J5; e-mail: sheila.obrien@blood.ca.

Supported by Canadian Blood Services.

Received for publication March 31, 2016; and accepted March 31, 2016.

doi:10.1111/trf.13635

TRANSFUSION 2016;56;1680–1683

In the public media we are bombarded with "facts" and "statistics" such as "17.8% of US adults smoke cigarettes" or "29% of US adults have high blood pressure." Actually, we don't really know if any of these numbers are correct. Generally, they come from a sample of people that were asked questions or from whom a measurement was taken and it is assumed that if you were to measure everybody else it would be the same. Maybe it would be the same (or close to it) or, then again, maybe not. It depends on how representative the sample was in the first place, on how large the sample was, and on chance.

We often want to know statistics about a whole group of people or things but it is not practical to measure everyone. Instead we draw a representative sample, measure the people in the sample and calculate the mean. Now we know the mean of the sample, but we still don't know the mean of the population. In fact, we are never going to know the mean of the population because we are never going to measure everybody. However, it seems reasonable that the true population mean would be somewhere around the mean of the sample. We can construct a confidence interval (CI) around the sample mean to take into account some of the uncertainty.

WHAT IS A CI?

A 95% CI is a range with an upper and lower number that is calculated from a sample where the true value is unknown. One could also calculate 90% CIs or 99% CIs, but in this article we focus on 95% CIs because they are the most commonly used. CIs can be calculated for all sorts of statistics such as odds ratios or percentages, but for simplicity we focus on the mean.

WHAT INFORMATION GOES INTO A CI?

CIs assume random sampling. That is to say, everyone in the population had an equal chance of being selected for the sample. In the case of clinical trials, rather than random sampling, patients who meet the study criteria are randomly assigned to treatment groups. The calculation of a CI takes into account how much variability

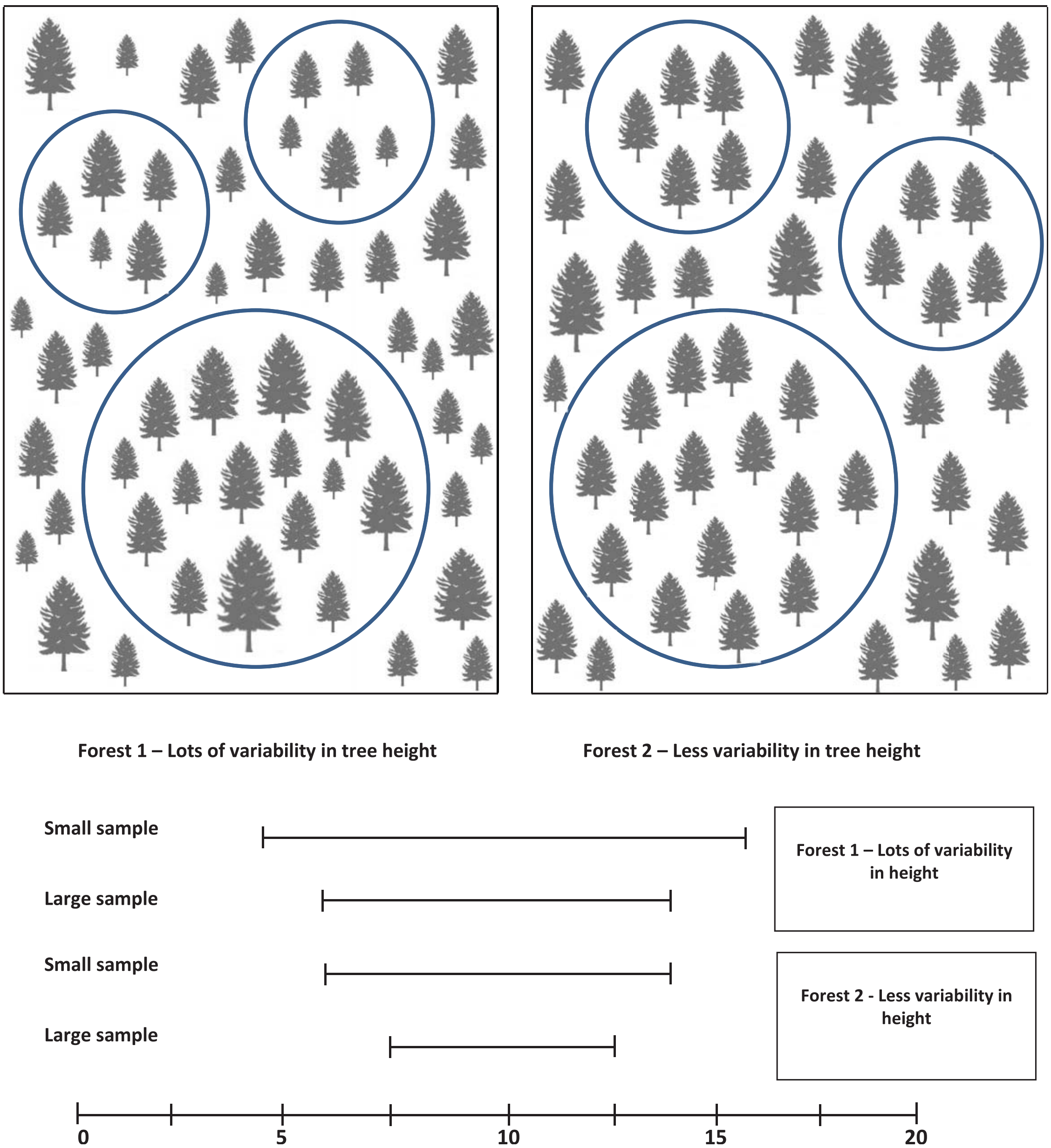

Fig. 1. Impact of variability of data and sample size on 95% CIs. The width of the 95% CI decreases with less variability in the sample and with increasing sample size. Forest 1 = lots of variability in tree height; Forest 2 = less variability in tree height.

there is between the individual values as well as the sample size.[1]

Two populations of trees are shown in Fig. 1. The trees on the left (Forest 1) have considerable variability in their heights, whereas the trees on the right (Forest 2) are of more similar height. If we drew a sample from each of these populations, the CI for the trees on the left would be wider than the one for the trees on the right to take into account the greater spread of tree heights. If we increase the sample size the CI would be smaller from either of these populations, but smallest for the sample from Forest 2 because of the combined effect of less variability in tree heights and larger sample. This relationship of CI width with the variability in the population and the sample size is shown in the lower half of Fig. 1.

INTERPRETING A CI

If we could measure everyone in the population, the mean would be one number and one number only. We therefore know that the mean of the population is a fixed

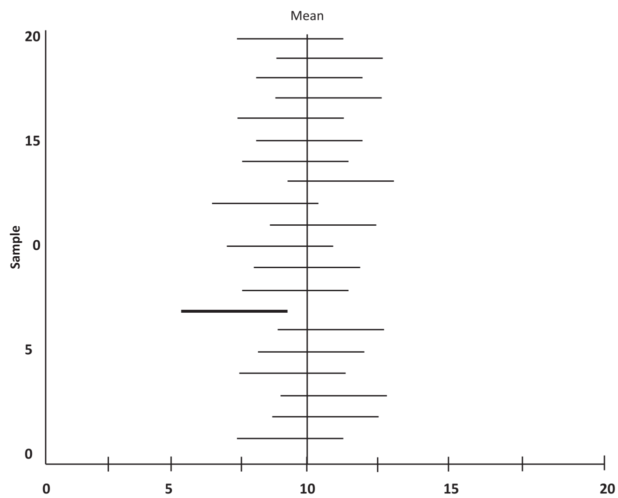

Fig. 2. 95% CIs of 20 randomly selected samples from the same population. If you construct 95% CIs for the mean of each of 20 samples, approximately 19 in every 20 samples will have a CI that includes the mean of the population.

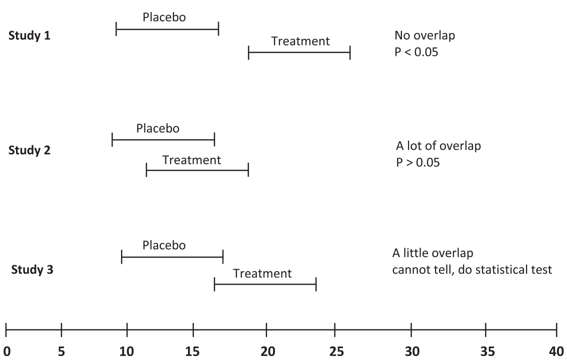

Fig. 3. Interpreting statistical significance from 95% CIs.

number. The concept of probability relates to sampling and estimating a range of plausible values. The true population mean either is within the CI or is not. If we drew multiple samples and constructed a 95% CI for each, we would expect that 95% of these CIs would contain the (unknown) population mean.[2] This is shown in Fig. 2 where 20 random samples were drawn from the same population and the CIs constructed. First, it shows that for every sample the CI is a little different. Second, most of the CIs include the population mean but there is one that does not. It shows that for any one sample there is a pretty good chance that the population mean is somewhere between the upper and lower limits calculated from the sample, but every now and again we could draw an unusual sample and the population mean would not be within the upper and lower limits.

COMPARING THE MEANS OF SAMPLES USING CIs

Sometimes we draw random samples from different groups because we want to know if they are different for a particular parameter. In clinical trials we often want to compare the means of two or more patient groups who received different treatments (e.g., placebo vs. treatment). Statistical significance is generally indicated by a p value of less than 0.05. However, CIs for each group (or for the difference between groups) are more informative than a p value because they give information about how much variability there is, and the readers can judge for themselves how much difference there is and how clinically relevant it is likely to be. For more information about hypothesis testing and p values, see our previous "How do I . . .?" paper.[3]

95% CIs can also be used as a quick way of checking for statistical significance (see Fig. 3). If the intervals do not overlap at all, we can be pretty sure that the samples were drawn from different populations (or that a treatment in a clinical trial had a real effect) and that the p value will be less than 0.05. If they overlap by quite a lot, they probably are not statistically different. However, when there is some overlap but not a lot, it is not possible to tell if there is a significant difference. It is then necessary to use a t test.[4]

COMMON QUESTIONS OR PROBLEMS WITH INTERPRETATION

Can I say that we are "95% confident" that the true mean lies within the interval?

To be fair this is how most researchers interpret a CI.[5,6] On a rather technical point this is not exactly correct. The correct interpretation is that if samples of the same size were repeatedly drawn from the population and their 95% CIs constructed, 95% of the CIs would be expected to include the true population mean. Just the same, if you interpret your results as being 95% confident that the CI includes the mean it should not prevent an appropriate conclusion, and so we believe that it will not get you into too much trouble.

Does the 95% CI mean that 95% of the population data falls within the CI?

No. As shown in Fig. 1, the CI is sensitive to sample size so if you draw a large sample the CI will be smaller, if you draw a small sample the CI will be large for exactly the same population of trees.

Is the CI invalid if the sample selection is not completely random?

The theory behind CIs assumes random sampling. In truth, most of the time there is some bias in the sample. For example, not everyone in the sample will agree to be measured and often a list of all members of the population from which to draw the sample does not exist. In clinical trials some people who were assigned to a group may withdraw from the study. The reader needs to evaluate how likely the sampling method would be to bias the results.

What is the advantage of a CI over a p value?

A CI gives more information than a p value. It allows the researcher to compare the magnitude of a difference. This is important because the difference between two populations or two treatment groups can be statistically significant but not clinically significant (not clinically relevant). The CI also gives an indication as to whether statistical significance or nonsignificance may be simply a function of choice of sample size.[6,7]

CONFLICTS OF INTEREST

The authors have disclosed no conflicts of interest.

REFERENCES

1. Ramsey FL, Schafer DW. The statistical sleuth: a course in methods of data analysis. Belmont (CA): Wadsworth Publishing Company; 1977.
2. Pandis N. Statistical inference with confidence intervals. Am J Orthod Dentofacial Orthop 2015;147:632-4.
3. O'Brien SF, Osmond L, Yi QL. How do I interpret a p value? Transfusion 2015;55:2778-82.
4. Andrade C. A primer on confidence intervals in psychopharmacology. J Clin Psychiatry 2015;76:228-31.
5. Laing CM, Rankin JA. Odds ratios and confidence intervals: a review for the pediatric oncology clinician. J Pediatr Oncol Nurs 2011;28:363-7.
6. Page P. Beyond statistical significance: clinical interpretation of rehabilitation research literature. Int J Sports Phys Ther 2014;9:726-36.
7. Sim J, Reid N. Statistical inference by confidence intervals: issues of interpretation and utilization. Phys Ther 1999;79:186-95.