

## Making Sense of Biostatistics: Clinical Significance

By J. Rick Turner

In the previous two columns, we discussed hypothesis testing and statistical significance. While consideration of statistical significance is informative and necessary in the clinical arena, it is not sufficient. Gardner and Altman (1986) noted that "presenting p-values alone can lead to them being given more merit than they deserve. In particular, there is a tendency to equate statistical significance with medical importance or biological relevance." Statistical significance must not be equated with medical importance or biological relevance. In large trials where an accurate estimate of the treatment effect is obtained, it is possible to obtain a statistically significant result when the treatment effect is small in magnitude, and very likely not clinically meaningful. Therefore, in drug development, confidence intervals (CIs) are used to convey the degree to which results are clinically significant. Accordingly, CIs are typically presented in regulatory documentation and scientific papers.

Before discussing CIs, let's clarify the information provided by a statistically significant finding. If we find in a clinical trial that an investigational antihypertensive drug lowers blood pressure statistically significantly more than a placebo (i.e., we get the result  $p < 0.05$ ), this result provides compelling evidence that the drug is effective. However, the p-value does not indicate how much more the drug lowers blood pressure than placebo. That is, it does not convey the magnitude of the treatment effect. The treatment effect might be relatively small, say 3 mmHg, or relatively large, say 15 mmHg. While a statistically significant result could be obtained in both cases, the clinical significance of these two results differs considerably.

It is important to note that, while there is a precise delineation between a result being statistically significant and not being statistically significant, there is no such precise delineation between being and not being clinically significant. Assessment of clinical significance is made by clinicians using all of their medical expertise and experience. However, it is very likely that different clinicians would have very similar views concerning the clinical significance of a particular result. There are two scenarios in which clinicians are likely to decide that a treatment effect is not of clinical significance. The first is when it is too small to be important in the real world. The second is when the CIs placed around the treatment effect are particularly wide. The first scenario is self-explanatory, so we'll focus on the second.

The treatment effect found in a single trial is only an estimate of the "true" treatment effect in the general population. If another trial were to be conducted in exactly the same manner using a different sample of participants, the treatment effect obtained in that trial almost certainly would not be identical. The question of interest therefore becomes: How similar would it likely be? Put another way, how confident can we be that the treatment effect in the original trial is representative of the true but unknown population treatment effect? We answer this question by constructing CIs and placing them around the treatment effect obtained in our single trial. This treatment effect is called the point estimate, and it is now surrounded by the lower limit of the CI and the upper limit of the CI.

Two CIs are commonly used in drug development research: 95% CIs and 99% CIs. We will not address the computations necessary to calculate CIs in detail. It is sufficient here to note the following:

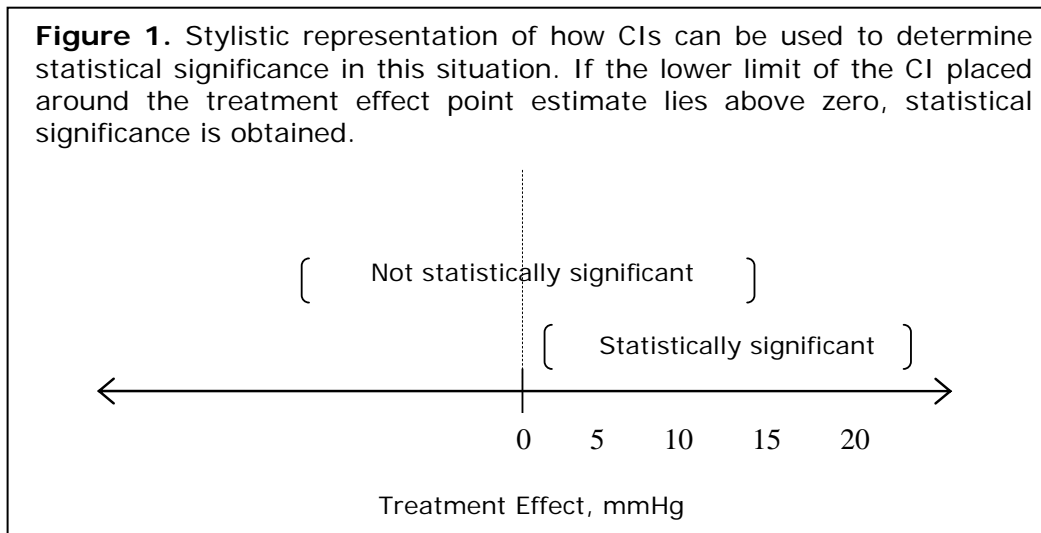
- The standard error of the point estimate is calculated. The greater the variation in the overall data set, the greater the standard error will be.
- This standard error is multiplied by a precision coefficient. The precision coefficient used in the calculation of 95% CIs in large trials is 1.96, and the precision coefficient used in the calculation of 99% CIs is 2.58.
- The value given by this multiplication is subtracted from the point estimate to determine the lower limit of the CI, and added to the point estimate to determine the upper limit.
- In this case, therefore, the lower and upper limits of the CI are placed symmetrically around the point estimate (which is not always the case for CIs).
- 99% CIs placed around a given point estimate will always be wider than 95% CIs. That is, the lower and upper limits of a 99% CI will be farther away from the point estimate.

Imagine that we conducted a trial of a new antihypertensive drug and found a treatment effect point estimate of 15 mmHg; that is, the drug lowered blood pressure 15.0 mmHg more than placebo. Imagine also that the 95% CI was calculated and represented as: (12.0, 15.0, 18.0). How is this CI interpreted? We can say that we are 95% confident that the true but unknown population treatment effect lies between 12.0 mmHg and 18 mmHg, and our best estimate is 15.0 mmHg. In other words, the true treatment effect could be as low as 12 mmHg or as high as 18 mmHg. Imagine now that a different result had been obtained in the trial. In this second scenario, the treatment effect point estimate was also 15 mmHg, but the 95% CI was: (5.0, 15.0, 25.0). In this case, we interpret the CI as saying that we are 95% confident that the true but unknown population treatment effect lies between 5.0 mmHg and 25 mmHg, and our best estimate is 15.0 mmHg. In other words, the true treatment effect could be as low as 5 mmHg or as high as 25 mmHg.

Even though the treatment effect point estimate is the same in both scenarios, 15 mmHg, the clinical implications of the results may be quite different. The lowest estimate of the drug's efficacy in the first scenario is 12.0 mmHg, while the lowest estimate of the drug's efficacy in the second scenario is 5.0 mmHg. Yes, it is true that the highest estimate of the drug's efficacy in the first scenario (25.0 mmHg) is greater than that in the second scenario (15.0 mmHg), but, in this case, it is the potential for lower efficacy that would receive more attention. While the lowest possible treatment effect in the first scenario may well still be considered of clinical significance, it is possible that, in the real world in which other antihypertensive drugs already exist, the lowest possible treatment effect in the second scenario may not be considered of clinical significance. Such a decision may lead to termination of the drug's development program. Therefore, in the drug development arena, consideration of statistical significance alone is not sufficient; consideration of clinical significance is also necessary.

It is also worth noting here that the calculation of CIs in this manner also tells us if the treatment effect is statistically significant. It does not provide a precise p-value, and so the exact degree of statistical significance is not known. However, if the lower limit of the 95% CI placed around a treatment effect lies above zero, it is possible to state that the treatment effect is statistically significant at the  $p < 0.05$  level. Similarly, if the lower limit of the 99% CI placed around a treatment effect lies above zero, it is possible to state that the treatment effect is statistically significant at the  $p < 0.01$  level. Why is the value of zero so important? If the drug were truly to be no more efficacious than the placebo, the expected treatment would be zero. If the value zero is not included in a 95% CI (that is, the lower limit of the CI lies above zero), we can have a reasonable degree of confidence that the treatment effect is statistically significantly greater than zero. Figure 1 provides a stylistic representation of this

statement. This ability of CIs to provide information on both clinical significance and statistical significance makes them extremely important in this arena.



## Reference

Gardner, M.J. and Altman, D.G., 1986, Estimation rather than hypothesis testing: Confidence levels rather than *p*-values. In Gardner, M.J. and Altman, D.G. (Eds), *Statistics with Confidence*, British Medical Association.

## Author

J. Rick Turner, PhD, PGCE, MICR is Chairman of the Department of Clinical Research, Campbell University School of Pharmacy, RTP Campus. Contact him at 1.919.463.7869 or TurnerR@campbell.edu.