Understanding confidence intervals helps you make better clinical decisions

By Zhaomin He, PhD, and Ellen Fineout-Overholt, PhD, RN, FNAP, FAAN

How to tell if you should base your care on a particular research finding

PERHAPS YOU DIDN'T LEARN about the confidence interval (CI) in your formal education or you don't hear the term in daily conversation. Confidence interval just doesn't roll of the tongue of a staff nurse quite like *blood pressure* or *urine output* does.

But knowing the importance of the CI allows you to interpret research for its impact on your practice. Evidence-based decision making is central to healthcare transformation. To make good decisions, you must know how to interpret and use research and practice evidence. Evaluating research means determining its validity (were the researchers' methods good ones?) and reliability (can clinicians get the same results the researchers got?).

CI and the degree of uncertainty

In a nutshell, the CI expresses the degree of uncertainty associated with a sample statistic (also called a study estimate). The CI allows clinicians to determine if they can realistically expect results similar to those in research studies when they implement those study results in their practice. Specifically, the CI helps clinicians identify a range within which they can expect their results to fall most of the time.

Used in quantitative research, the CI is part of the stories that studies tell in numbers. These numeric stories describe the characteristics, or *parameters*, of a population; populations can be made up of individuals, communities, or systems. Collecting information from the whole population to find answers to clinical questions is practically impossible. For instance, we can't possibly collect information from all cancer patients. Instead, we collect information from smaller groups within the larger population, called *samples*. We learn about population characteristics from these samples through a process called *inference*.

To differentiate sample values from those of the population (*parameters*), the numeric characteristics of a sample most commonly are termed *statistics*, but also may be called *parameter estimates* because they're estimates of the population. Inferring information from sample statistics to population parameters can lead to errors, mainly because statistics may differ from one sample to the next. Several other terms are related to this opportunity for error—probability, standard error (SE), and mean. (See *What are probability, standard error, and mean?*)

Calculating the CI

Used in the formula to calculate the upper and lower boundaries of the CI (within which the population

What are probability, standard error, and mean?

Because errors can occur when we infer information from sample statistics to population parameters, we make a plan to allow mistakes—but limit them to a reasonable level. This is called the *statistical significance level*, or probability of error—usually called the *probability*, or *p* value.

Typically, the *p* value is set at 0.05, but researchers may use *p* values of 0.10, 0.05, 0.01, and 0.001. In high-stake situations, such as life and death, the number of mistakes should be controlled more stringently by using a smaller *p* value—for example, $p = 0.001$.

Let's dig a little deeper: A probability of 0.05 means that if you keep drawing samples from the same population and get statistics for each sample, sample statistics are unlikely to represent the population parameters only five times out of 100. In those five times, the statistics occur by chance, which is error. In other words, 95 times out of 100, the sample statistics are likely to be a good representation of the population parameters. A *p* value of 0.05 or smaller gives researchers confidence in their study results and in inferring information from the sample statistics to population parameters.

Standard error and mean

Other terms important for understanding confidence intervals are *standard error* (SE) and *mean*. The SE tells us how accurately the sample statistics reflect the population parameters. It's a measure of dispersion for sample means.

For a sample of a certain number, the SE informs us about how the mean of that sample agrees with the population mean. Each individual's data in the sample doesn't represent the population parameter well, but the mean of the sample is more likely to do so. So the larger the sample size, the more accurately the sample mean is expected to represent the population parameter. In this case, as sample size increases, the SE is expected to get smaller. The smaller the SE, the more similar the sample statistics are to the population parameter. Thus, it follows that a large SE value means the sample statistics differ markedly from the population parameter and are less likely to represent the parameter.

parameter is expected to fall), the SE reveals how accurately the sample statistics reflect population parameters. Choosing a more stringent probability, such as 0.01 (meaning a CI of 99%), would offer more confidence that the lower and upper boundaries of the CI contain the true value of the population parameter.

Not all studies provide CIs. For example, when we prepared this article, our literature search found study after study with a probability (*p*) value) but no CI. However, studies usually report SEs and means. If the study you're reading doesn't provide a CI, here's the formula for calculating it:

95% CI: $X = \bar{X} \pm (1.96 \times \text{SE})$, where X denotes the estimate and $\bar{X}$ denotes the mean of the sample

To find the upper boundary of the estimate, add 1.96 times the SE to $\bar{X}$. To find the lower boundary of the estimate, subtract 1.96 times the SE from $\bar{X}$. *Note:* 1.96 is how many standard deviations from the mean are required for the range of values to contain 95% of the values.

Be aware that values found with this formula aren't reliable with samples of less than 30. But don't despair; you can still calculate the CI—although explaining that formula is beyond the scope of this article. Watch the video at https://goo.gl/AuQ7Re to learn about that formula.

Real-world decision-making: Where CIs really count

Now let's apply your new statistical knowledge to clinical decision making. In everyday terms, a CI is the range of values around a sample statistic within which clinicians can expect to get results if they repeat the study protocol or intervention, including measuring the same outcomes the same ways. As you critically appraise the reliability of research ("Will I get the same results if I use this research?"), you must address the precision of study findings, which is determined by the CI. If the CI around the sample statistic is narrow, study findings are considered precise and you can be confident you'll get close to the sample statistic if you implement the research in your practice. Also, if the CI does *not* contain the statistical value that indicates no effect (such as 0 for effect size or 1 for relative risk and odds ratio), the sample statistic has met the criteria to be statistically significant.

The following example can help make the CI concept come alive. In a systematic review synthesizing studies of the effect of tai chi exercise on sleep quality, Du and colleagues (2015) found tai chi affected sleep quality in older people as measured by the Pittsburgh Sleep Quality Index (mean difference of -0.87; 95% CI [-1.25, -0.49]). Here's how clinicians caring for older adults in the community would interpret these results: Across the studies reviewed, older people reported better sleep if they engaged in tai chi exercise. The lower boundary of the CI is -1.25, the study statistic is -0.87, and the upper boundary is -0.49. Each limit is 0.38 from the sample statistic, which is a relatively narrow CI. Keep in mind that a mean difference of 0 indicates there's no difference; this CI doesn't contain that value. Therefore, the sample statistic is statistically significant and unlikely to occur by chance. Because this was a systematic review and tai chi exercise has been established as helping people sleep, based on the sample statistics and the CI, clinicians can confidently include tai chi exercises among possible recommendations for patients who have difficulty sleeping.

(continued on page 58)

(continued from page 37)

with other advocates. And above all, be persistent.

Take on large-scale advocacy

The Miss America brouhaha demonstrates the value of diverse, large-scale advocacy. After the offensive *The View* episode, nurses took to social media to protest the hosts' remarks. Virginia nurse Amanda Claybrook's Change.org petition got more than 50,000 signers; ANA and other nursing groups released statements expressing dismay. As a result, *The View* lost major sponsors. The Truth About Nursing also started a petition and discussed with *The View*'s publicity director specific ways the show could make amends.

Meanwhile, the mainstream media ran stories and op-eds about the incident. Within a week, *The View* apologized for the hosts' remarks and invited nurses onto the show "to share firsthand what these hard-working nurses do on a daily basis." Two nursing professors appeared on the show to explain some aspects of nursing. *The Dr. Oz Show* devoted an entire episode to nursing, although information about nursing skills was sadly limited. Media creators' responses to nurses' protests and petitions show the media can be persuaded to let nurses play a role in shaping program content. (See *Scrubbing out MTV's "Scrubbing In"*)

Praise the media for what it does well. Submit feedback to media outlets; phone calls and hard-copy letters probably have the biggest impact, but email, social media, and hashtags also are good.

Of course, no one knows nursing as well as nurses themselves do. Ideally, we should create our own media and tell our own stories, in vehicles ranging from broadcast TV shows to modest websites and blogs. But that's a topic for another time.

Meanwhile, The Truth About Nursing offers many ideas on how to improve nursing's image at www.TruthAboutNursing.org/action and in our book, *Saving Lives: Why the Media's Portrayal of Nursing Puts Us All at Risk*. We hope you'll join us in taking a stand for nurses and patients. ★

The authors of *Saving Lives: Why the Media's Portrayal of Nursing Puts Us All at Risk*, Sandy and Harry Summers are the executive director and senior advisor, respectively, of The Truth About Nursing, a 501(c)(3) nonprofit organization. You can reach them at www.truthaboutnursing.org.

(continued from page 40)

Now you can apply your knowledge of CIs to make wise decisions about whether to base your patient care on a particular research finding. Just remember—when appraising research, consistently look for the CI. If the authors report the mean and SE but *don't* report the CI, you can calculate the CI using the formula discussed earlier. ★

As you critically appraise the reliability of research, you must address the precision of study findings, which is determined by the CI.

The authors work at the University of Texas at Tyler. Zhaomin He is an assistant professor and biostatistician of nursing. Ellen Fineout-Overholt is the Mary Coulter Dowdy Distinguished Professor of Nursing.

Selected references

Du S, Dong J, Zhang H, et al. Taichi exercise for self-rated sleep quality in older people: a systematic review and meta-analysis. *Int J Nurs Stud*. 2015;52(1):368-79.

Fineout-Overholt E. EBP, QI, and research: strange bedfellows or kindred spirits? In: Hedges C, Williams B, eds. *Anatomy of Research for Nurses*. Indianapolis, IN: Sigma Theta Tau International; 2014:23-44.

Fineout-Overholt E, Melnyk BM, Stillwell SB, Williamson KM. Evidence-based practice, step by step: critical appraisal of the evidence: part II: digging deeper—examining the "keeper" studies. *Am J Nurs*. 2010;110(9):41-8.

Kahn Academy. Small sample size confidence intervals [video]. https://goo.gl/AuQ7Re

Melnyk BM, Fineout-Overholt E. ARCC (Advancing Research and Clinical practice through close Collaboration): a model for system-wide implementation and sustainability of evidence-based practice. In: Rycroft-Malone J, Bucknall T, eds. *Models and Frameworks for Implementing Evidence-Based Practice: Linking Evidence to Action*. Indianapolis, IN: Wiley-Blackwell & Sigma Theta Tau International; 2010.

O'Mathúna DP, Fineout-Overholt E. Critically appraising quantitative evidence for clinical decision making. In: Melnyk BM, Fineout-Overholt E, eds. *Evidence-Based Practice in Nursing and Healthcare: A Guide to Best Practice*. 3rd ed. Philadelphia: Lippincott Williams and Wilkins; 2015:81-134.

Plichta, SB, Kelvin E. *Munro's Statistical Methods for Health Care Research*. 6th ed. Philadelphia, PA: Lippincott, Williams & Wilkins; 2013.