



# Determining the level of evidence: Experimental research appraisal

BY AMY GLASOFER, DNP, RN, NE-BC, AND ANN B. TOWNSEND, DRNP, RN, ANP-C, CNS-C

**Abstract:** To support evidence-based nursing practice, the authors provide guidelines for appraising research based on quality, quantity, and consistency. This is the first of a three-part series.

**Keywords:** evidence-based practice, evidence hierarchy, meta-analysis, quasi-experimental research, randomized controlled trial, RCT, systematic review

NURSES HAVE A RESPONSIBILITY to use evidence-based practices in their patient care. To ensure their actions will produce the desired outcomes, nurses must use the strongest evidence available to support patient care.<sup>1</sup> However, determining what qualifies as “strong” evidence can be challenging.

According to the Agency for Healthcare Research and Quality, the eviden-

tial strength includes three elements: quality, quantity, and consistency.<sup>2</sup>

- *Quality* is the most challenging element nurses must evaluate to determine the strength of evidence for a topic. Quality refers to the methods used to ensure that results are valid and not influenced by bias or occurring by chance.<sup>2</sup> One component of quality is the level of the evidence.

- **Quantity** is evaluated by considering the number of studies on a topic, the size of the studies, and the impact of studied treatments.
- **Consistency** is the easiest of these elements to understand; for evidence to be strong, similar findings should be reported across multiple sources.<sup>2</sup>

This three-part series will provide basic guidance for appraising evidence. However, this is only one step in the evidence-based practice (EBP) process, which includes complexities that this series will not address. Many resources exist for nurses to develop their critical appraisal skills and strengthen their understanding of the EBP process. For example, the *American Journal of Nursing* published a 12-article series outlining a step-by-step approach to EBP.<sup>3</sup>

Various evidence hierarchies exist to evaluate the level of evidence.<sup>1</sup> To apply these hierarchies, nurses must have a working knowledge of research design. This first installment in the series provides nurses with a basic understanding of research design to appraise a source's level of evidence. This article will review appraisal of experimental research, which includes randomized con-

trolled trials (RCTs; Level 1) and quasi-experimental research (Level 2). Future installments will address nonexperimental research appraisal (Level 3) and the leveling of non-research evidence (Levels 4 and 5).

## The evidence pyramid

One way to understand evidence hierarchies is to consider crime scene evidence. Different types of crime scene evidence are weighed differently when one tries to determine an individual's guilt or innocence. For example, DNA evidence is superior to eyewitness testimony because witnesses are susceptible to bias and DNA is more objective.<sup>4</sup> A determination of guilt is more likely if DNA evidence is present or if multiple eyewitnesses offer consistent reports than if testimony of only one eyewitness is presented. DNA might be on the top level of a criminal evidence hierarchy, and eyewitness testimony could be found lower down.<sup>4</sup>

The same is true of clinical evidence, but rather than determining guilt or innocence, nurses must determine if cause and effect exists. To objectively draw a conclusion, nurses must use the strongest evidence available.

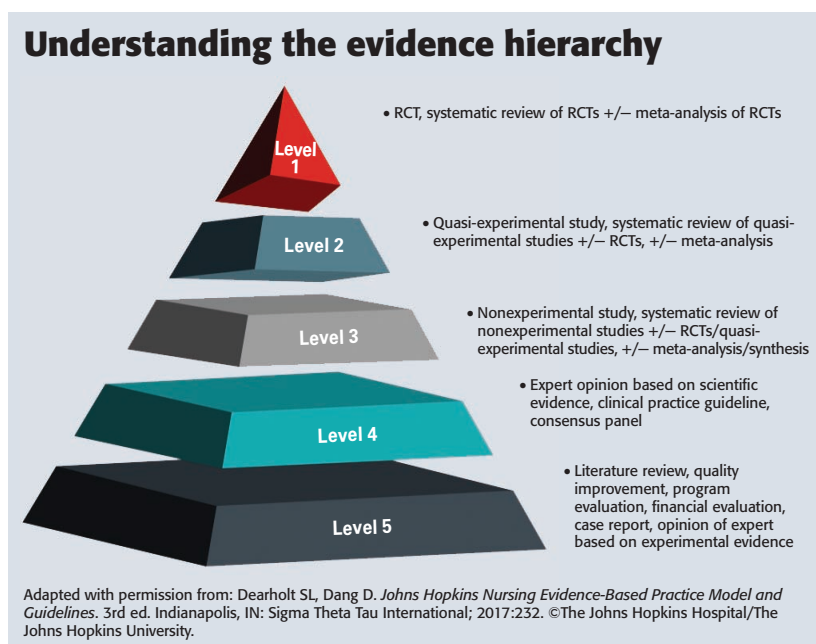
Imagine the evidence levels arranged by research design (see *Understanding the evidence hierarchy*). The top of the pyramid, Level 1, represents the strongest evidence. As researchers move through the pyramid from Level 1 down, the study designs become less rigorous, which may influence the results through the introduction of bias or conclusion errors. Pyramids vary between organizations and disciplines, but they all follow these basic principles.

Level of evidence hierarchies include the Joanna Briggs Institute levels of evidence and the Oxford Centre for Evidence-Based Medicine model.<sup>5,6</sup> This article will use the Johns Hopkins hierarchy of evidence.<sup>7</sup>

## Level 1: RCTs, systematic reviews, and meta-analyses

According to the Johns Hopkins hierarchy of evidence, the highest level of evidence is an RCT, a systematic review of RCTs, or a meta-analysis of RCTs.<sup>7</sup> In an RCT, the study must meet three criteria: random or "by chance" assignment of participants into two or more groups, an intervention or treatment applied to at least one of the groups, and a control group that does not receive the same treatment or intervention. The methodologies used in Level 1 evidence reduce bias and help identify cause-and-effect relationships.<sup>8</sup>

Consider the following sample research question: *What is the effect of caffeine on nursing medication errors?* To answer this question using an RCT, researchers would first recruit a sample of nurses. The study must have institutional review board approval and obtain informed consent from the participants, and the study should follow guidelines set forth by the EQUATOR (Enhancing the QUALity and Transparency Of health Research) Network, an international initiative dedicated to improving the reliability and value of published health research literature.<sup>9</sup> Each



participating nurse is assigned by chance to the caffeine (intervention) group or the no-caffeine (control) group. Researchers must ensure that the two groups are the same regarding any other factor that might impact medication errors aside from the intervention (such as patient acuity or nurse experience), or take these other factors into account in the data analysis and conclusion. In doing so, they can conclude that any statistically significant differences in medication errors between the groups are a result of the caffeine and not chance.

Although one DNA sample provides strong evidence in a criminal investigation, multiple DNA samples confirming the same suspect are even stronger. Systematic reviews and meta-analyses of RCTs follow the same reasoning. A meta-analysis evaluates multiple research studies. When all the studies included are RCTs, the findings are more powerful than any one RCT on its own.

A systematic review uses a rigorous process to identify, appraise, and synthesize the evidence on a particular topic.<sup>1</sup> A meta-analysis takes it one step further and conducts a statistical analysis of the synthesized data to obtain a statistic representing the effect of the intervention across multiple studies.<sup>1</sup> So, a systematic review on the effect of caffeine and medication errors would include a rigorous review of every RCT on the topic that met specific inclusion criteria, and a meta-analysis would provide a summary statistic on the size of the effect or the influence of caffeine on medication errors.

Just as DNA evidence can be flawed, RCTs, systematic reviews, and meta-analyses can also have limitations. In the example, researchers are seeking volunteers to participate. The voluntary participants could be very different from the nurses who choose not to participate. If so, study findings might not apply to nurses in general. Consider these factors:



**Just as DNA evidence can be flawed, RCTs, systematic reviews, and meta-analyses can also have limitations.**

- Nurses in both groups might improve their practice because they know they are being observed, resulting in fewer medication errors across both groups.
- Nurses assigned to the control group may perform poorly because they are in withdrawal from their typical caffeine intake.
- Nurses in the control group could be unhappy that they were assigned to the noncaffeine group and behave differently.

Certain strategies are designed to eliminate some sources of bias. For example, in this case the researchers could “blind” or “mask” the participants so they do not know to which group they were randomly assigned. To achieve this, researchers would not tell the nurses which group they are in and give both groups coffee

(caffeinated to the intervention group and decaffeinated to the control group).

However, even in a well-designed RCT, the reader must be critical of the findings. The same is true of systematic reviews and meta-analyses, as they are only as strong as the thoroughness of the review and the findings of the weakest study included in the analysis.

## **Level 2: Quasi-experimental research**

Fingerprints can be an important source of crime scene evidence, although they are not as reliable as DNA.<sup>10</sup> Fingerprint comparisons require expert review. Because expert judgment introduces greater bias and uncertainty than DNA evidence, fingerprints might be considered one level below DNA in the crime scene evidence hierarchy.<sup>10</sup>

In the Johns Hopkins hierarchy, Level 2 contains quasi-experimental research studies as well as systematic reviews of both RCTs and quasi-experimental studies with or without meta-analysis.<sup>7</sup> This group of evidence is still experimental because it involves manipulation or an intervention introduced by the research. However, it is termed *quasi-experimental* because it lacks one or two of the three criteria required for a true experimental design. Examples of quasi-experimental designs used in nursing research are the nonequivalent control group design, the pre-posttest design, and the interrupted time series design.<sup>7</sup>

Consider the sample research question. Instead of randomly assigning nurses to the caffeine or noncaffeine groups, researchers could compare two units in a nonequivalent control group design. One could be the caffeine unit, and the other could be the noncaffeine unit. Or researchers could give one group of nurses no caffeine for a time, and then give them caffeine during another period as in an interrupted time series design.



Researchers would observe medication errors throughout, comparing one study period to the other. Further still, researchers could have only one group receive caffeine and make no comparison. In these examples, assignment is no longer random and there could be alternative explanations for the difference in medication error rates seen between the groups.

When comparing two different units, patient or nursing populations may be dissimilar, fewer medications may be given on one unit than another, processes for medication administration may differ, or any of a multitude of other factors may impact the study outcomes. Similarly, when researchers compare the same group at two different time periods, an unrelated change in practice, patient population, or acuity could explain results. And when there is no comparison group, researchers have no basis for determining if medication errors are associated with caffeine consumption.

No matter how well executed a quasi-experimental study is, nurses must be less certain of its results compared with results from an RCT. The same is true of systematic reviews with or without meta-analysis that include quasi-experimental studies. A review is only as strong as the weakest study included. Therefore, reviews that include quasi-experimental studies are not as strong as those that include only RCTs. The quasi-experimental design will always fall lower than an RCT in an evidence hierarchy, regardless of the model consulted.

Despite this limitation, researchers will continue to use quasi-experimental designs. Quasi-experimental research can be simpler to carry out in practice, and often feasibility trumps rigor.

### Scrutinize the evidence

Nurses endeavoring to provide evidence-based care may find themselves acting as detectives. Although



**The quasi-experimental design will always fall lower than an RCT in an evidence hierarchy, regardless of the model consulted.**

they may be tempted to reach a conclusion when they encounter a piece of evidence that matches their suspicions, the investigation must go deeper. Nurses are obligated to find a sufficient number of sources that arrive at similar conclusions.

Although no magic number indicates sufficient evidence, fewer sources are needed when synthesizing higher-quality evidence. One element of quality is the level of evidence, which is based on how the design minimizes the impact of bias and chance on the conclusions drawn. Regardless of the evidence hierarchy used, RCTs and systematic reviews with or without meta-analysis exist at or near the highest level of evidence, with quasi-experimental research follow-

ing closely behind. Nurses must use their critical appraisal skills to determine when a study has employed an experimental design, is using a control group, or has assigned participants to groups randomly to support the quest to provide evidence-based patient care. Upcoming installments of this series will discuss Levels 3, 4, and 5, which include nonexperimental research, and sources of non-research evidence. ■

### REFERENCES

1. Melnyk BM, Fineout-Overholt E. *Evidence-Based Practice in Nursing & Healthcare: A Guide to Best Practice*. 4th ed. Philadelphia, PA: Wolters Kluwer Lippincott Williams & Wilkins; 2019.
2. West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence. Evidence report/technology assessment No. 47. AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality. 2002. [www.thecre.com/pdf/ahrq-system-strength.pdf](http://www.thecre.com/pdf/ahrq-system-strength.pdf).
3. *American Journal of Nursing*. Evidence-based practice, step-by-step. 2019. <https://journals.lww.com/ajnonline/pages/collectiondetails.aspx?TopicalCollectionId=10>.
4. Thomson Reuters. How DNA evidence works. 2019. <https://criminal.findlaw.com/criminalprocedure/how-dna-evidence-works.html>.
5. Joanna Briggs Institute. JBI grades of recommendation. 2013. [https://joannabriggs.org/sites/default/files/2019-05/JBI-grades-of-recommendation\\_2014.pdf](https://joannabriggs.org/sites/default/files/2019-05/JBI-grades-of-recommendation_2014.pdf).
6. Oxford Centre for Evidence-Based Medicine. The Oxford 2011 Levels of Evidence. 2011. [www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf](http://www.cebm.net/wp-content/uploads/2014/06/CEBM-Levels-of-Evidence-2.1.pdf).
7. Dearholt SL, Dang D. *Johns Hopkins Nursing Evidence-Based Practice Model and Guidelines*. 3rd ed. Indianapolis, IN: Sigma Theta Tau International; 2017.
8. Polit DF, Beck CT. *Nursing Research: Generating and Assessing Evidence for Nursing Practice*. 10th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2017.
9. Equator Network. Enhancing the QUALity and Transparency Of health Research. [www.equator-network.org](http://www.equator-network.org).
10. Servick K. Reversing the legacy of junk science in the courtroom. 2016. [www.sciencemag.org/news/2016/03/reversing-legacy-junk-science-courtroom](http://www.sciencemag.org/news/2016/03/reversing-legacy-junk-science-courtroom).

Amy Glasofer is a nurse scientist at Virtua Center for Learning in Mt. Laurel, N.J., and Ann B. Townsend is an adult NP with The Nurse Practitioner Group, LLC.

The authors have disclosed no financial relationships related to this article.

This article was originally published as Glasofer A, Townsend AB. Determining the level of evidence: experimental research appraisal. *Nurs Crit Care*. 2019;14(6):22-25.

DOI-10.1097/01.NURSE.0000697156.46992.b2