By Ellen Fineout-Overholt, PhD, RN, FNAP, FAAN, Bernadette Mazurek Melnyk, PhD, RN, CPNP/PMHNP, FNAP, FAAN, Susan B. Stillwell, DNP, RN, CNE, and Kathleen M. Williamson, PhD, RN

EVIDENCE-BASED PRACTICE Step by Step

Critical Appraisal of the Evidence: Part I

An introduction to gathering, evaluating, and recording the evidence.

This is the fifth article in a series from the Arizona State University College of Nursing and Health Innovation's Center for the Advancement of Evidence-Based Practice. Evidence-based practice (EBP) is a problem-solving approach to the delivery of health care that integrates the best evidence from studies and patient care data with clinician expertise and patient preferences and values. When delivered in a context of caring and in a supportive organizational culture, the highest quality of care and best patient outcomes can be achieved.

The purpose of this series is to give nurses the knowledge and skills they need to implement EBP consistently, one step at a time. Articles will appear every two months to allow you time to incorporate information as you work toward implementing EBP at your institution. Also, we've scheduled "Chat with the Authors" calls every few months to provide a direct line to the experts to help you resolve questions. Details about how to participate in the next call will be published with September's *Evidence-Based Practice, Step by Step*.

In May's evidence-based practice (EBP) article, Rebecca R., our hypothetical staff nurse, and Carlos A., her hospital's expert EBP mentor, learned how to search for the evidence to answer their clinical question (shown here in PICOT format): "In *hospitalized adults* (P), how does a *rapid response team* (I) compared with *no rapid response team* (C) affect *the number of cardiac arrests* (O) and *unplanned admissions to the ICU* (O) during a *three-month period* (T)?" With the help of Lynne Z., the hospital librarian, Rebecca and Carlos searched three databases, PubMed, the Cumulative Index of Nursing and Allied Health Literature (CINAHL), and the Cochrane Database of Systematic Reviews. They used keywords from their clinical question, including *ICU*, *rapid response team*, *cardiac arrest*, and *unplanned ICU admissions*, as well as the following synonyms: *failure to rescue*, *never events*, *medical emergency teams*, *rapid response systems*, and *code blue*. Whenever terms from a database's own indexing language, or controlled vocabulary, matched the keywords or synonyms, those terms were also searched. At the end of the database searches, Rebecca and Carlos chose to retain 18 of the 18 studies found in PubMed; six of the 79 studies found in CINAHL; and the one study found in the Cochrane Database of Systematic Reviews, because they best answered the clinical question.

As a final step, at Lynne's recommendation, Rebecca and Carlos conducted a hand search of the reference lists of each study they retained looking for any relevant studies they hadn't found in their original search; this process is also called *the ancestry method*. The hand search yielded one additional study, for a total of 26.

RAPID CRITICAL APPRAISAL

The next time Rebecca and Carlos meet, they discuss the next step in the EBP process—critically appraising the 26 studies. They obtain copies of the studies by printing those that are immediately available as full text through library subscription or those flagged as "free full text" by a database or journal's Web site. Others are available through interlibrary loan, when another hospital library shares its articles with Rebecca and Carlos's hospital library.

Carlos explains to Rebecca that the purpose of critical appraisal isn't solely to find the flaws in a study, but to determine its worth to practice. In this rapid critical appraisal (RCA), they will review each study to determine

- its level of evidence.
- how well it was conducted.
- how useful it is to practice.

Once they determine which studies are "keepers," Rebecca and Carlos will move on to the final steps of critical appraisal: evaluation and synthesis (to be discussed in the next two installments of the series). These final steps will determine whether overall findings from the evidence review can help clinicians improve patient outcomes.

Rebecca is a bit apprehensive because it's been a few years since she took a research class. She

shares her anxiety with Chen M., a fellow staff nurse, who says she never studied research in school but would like to learn; she asks if she can join Carlos and Rebecca's EBP team. Chen's spirit of inquiry encourages Rebecca, and they talk about the opportunity to learn that this project affords them. Together they speak with the nurse manager on their medical–surgical unit, who agrees to let them use their allotted continuing education time to work on this project, after they discuss their expectations for the project and how its outcome may benefit the patients, the unit staff, and the hospital.

Learning research terminology. At the first meeting of the new EBP team, Carlos provides Rebecca and Chen with a glossary of terms so they can learn basic research terminology, such as *sample*, *independent variable*, and *dependent variable*. The glossary also defines some of the study designs the team is likely to come across in doing their RCA, such as systematic review, randomized controlled trial, and cohort, qualitative, and descriptive studies. (For the definitions of these terms and others, see the glossaries provided by the Center for the Advancement of Evidence-Based Practice at the Arizona State University College of Nursing and Health Innovation [http://nursingandhealth.asu.edu/evidence-based-practice/resources/glossary.htm]

and the Boston University Medical Center Alumni Medical Library [http://medlib.bu.edu/bugms/content.cfm/content/ebmglossary.cfm#R].)

Determining the level of evidence. The team begins to divide the 26 studies into categories according to study design. To help in this, Carlos provides a list of several different study designs (see *Hierarchy of Evidence for Intervention Studies*). Rebecca, Carlos, and Chen work together to determine each study's design by reviewing its abstract. They also create an "I don't know" pile of studies that don't appear to fit a specific design. When they find studies that don't actively answer the clinical question but

Hierarchy of Evidence for Intervention Studies

Type of evidence	Level of evidence	Description
Systematic review or meta-analysis	I	A synthesis of evidence from all relevant randomized controlled trials.
Randomized controlled trial	II	An experiment in which subjects are randomized to a treatment group or control group.
Controlled trial without randomization	III	An experiment in which subjects are nonrandomly assigned to a treatment group or control group.
Case-control or cohort study	IV	Case-control study: a comparison of subjects with a condition (case) with those who don't have the condition (control) to determine characteristics that might predict the condition. Cohort study: an observation of a group(s) (cohort[s]) to determine the development of an outcome(s) such as a disease.
Systematic review of qualitative or descriptive studies	V	A synthesis of evidence from qualitative or descriptive studies to answer a clinical question.
Qualitative or descriptive study	VI	Qualitative study: gathers data on human behavior to understand *why* and *how* decisions are made. Descriptive study: provides background information on the *what*, *where*, and *when* of a topic of interest.
Expert opinion or consensus	VII	Authoritative opinion of expert committee.

Adapted with permission from Melnyk BM, Fineout-Overholt E, editors. Evidence-based practice in nursing and healthcare: a guide to best practice [forthcoming]. 2nd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams and Wilkins.

may inform thinking, such as descriptive research, expert opinions, or guidelines, they put them aside. Carlos explains that they'll be used later to support Rebecca's case for having a rapid response team (RRT) in her hospital, should the evidence point in that direction.

After the studies—including those in the "I don't know" group—are categorized, 15 of the original 26 remain and will be included in the RCA: three systematic reviews that include one meta-analysis (Level I evidence), one randomized controlled trial (Level II evidence), two cohort studies (Level IV evidence), one retrospective pre-post study with historic controls (Level VI evidence), four preexperimental (pre-post) intervention studies (no control group) (Level VI evidence), and four EBP implementation projects (Level VI evidence). Carlos reminds Rebecca and Chen that Level I evidence—a systematic review of randomized controlled trials

or a meta-analysis—is the most reliable and the best evidence to answer their clinical question.

Using a critical appraisal guide. Carlos recommends that the team use a critical appraisal checklist (see *Critical Appraisal Guide for Quantitative Studies*) to help evaluate the 15 studies. This checklist is relevant to all studies and contains questions about the essential elements of research (such as, purpose of the study, sample size, and major variables).

The questions in the critical appraisal guide seem a little strange to Rebecca and Chen. As they review the guide together, Carlos explains and clarifies each question. He suggests that as they try to figure out which are the essential elements of the studies, they focus on answering the first three questions: *Why was the study done? What is the sample size? Are the instruments of the major variables valid and reliable?* The remaining questions will be addressed later on in the critical

appraisal process (to appear in future installments of this series).

Creating a study evaluation table. Carlos provides an online template for a table where Rebecca and Chen can put all the data they'll need for the RCA. Here they'll record each study's essential elements that answer the three questions and begin to appraise the 15 studies. (To use this template to create your own evaluation table, download the *Evaluation Table Template* at http://links.lww.com/AJN/A10.)

EXTRACTING THE DATA
Starting with level I evidence studies and moving down the hierarchy list, the EBP team takes each study and, one by one, finds and enters its essential elements into the first five columns of the evaluation table (see Table 1; to see the entire table with all 15 studies, go to http://links.lww.com/AJN/A11). The team discusses each element as they enter it, and tries to determine if it meets the criteria of the critical

Table 1. Evaluation Table, Phase I

First Author (Year)	Conceptual Framework	Design/Method	Sample/Setting	Major Variables Studied (and Their Definitions)	Measurement	Data Analysis	Findings	Appraisal: Worth to Practice
Chan PS, et al. *Arch Intern Med* 2010;170(1):18-26.	None	SR Purpose: effect of RRT on HMR and CR • Searched 5 databases from 1950-2008, and "grey literature" from MD conferences • Included only studies with a control group	N = 18 studies Setting: acute care hospitals; 13 adult, 5 peds Average no. beds: NR Attrition: NR	IV: RRT DV1: HMR DV2: CR				
McGaughey J, et al. *Cochrane Database Syst Rev* 2007;3: CD005529.	None	SR (Cochrane review) Purpose: effect of RRT on HMR • Searched 6 databases from 1990-2006 • Excluded all but 2 RCTs	N = 2 studies 24 adult hospitals Attrition: NR	IV: RRT DV1: HMR				
Winters BD, et al. *Crit Care Med* 2007;35(5): 1238-43.	None	SR Purpose: effect of RRT on HMR and CR • Searched 3 databases from 1990-2005 • Included only studies with a control group	N = 8 studies Average no. beds: 500 Attrition: NR	IV: RRT DV1: HMR DV2: CR				
Hillman K, et al. *Lancet* 2005; 365(9477): 2091-7.	None	RCT Purpose: effect of RRT on CR, HMR, and UICUA	N = 23 hospitals Average no. beds: 340 • Intervention group (n = 12) • Control group (n = 11) Setting: Australia Attrition: none	IV: RRT protocol for 6 months • 1 AP • 1 ICU or ED RN DV1: HMR (unexpected deaths, excluding DNRs) DV2: CR (excluding DNRs) DV3: UICUA	HMR CR rates of UICUA			Note: • Criteria for activating RRT

Shaded columns indicate where data will be entered in future installments of the series.
AP = attending physician; CR = cardiopulmonary arrest or code rates; DNR = do not resuscitate; DV = dependent variable; ED = emergency department; HMR: hospital-wide mortality rates; ICU = intensive care unit; IV = independent variable; MD = medical doctor; NR = not reported; Peds = pediatric; RCT = randomized controlled trial; RN = registered nurse; RRT = rapid response team; SR = systematic review; UICUA = unplanned ICU admissions.

appraisal guide. These elements—such as purpose of the study, sample size, and major variables—are typical parts of a research report and should be presented in a predictable fashion in every study so that the reader understands what's being reported.

Usually the important information in a study can be found in the abstract.

As the EBP team continues to review the studies and fill in the evaluation table, they realize that it's taking about 10 to 15 minutes per study to locate and enter the information. This may be because when they look for a description of the sample, for example, it's important that they note how the sample was obtained, how many patients are included, other characteristics of the sample, as well as any diagnoses or illnesses the sample might have that could be important to the study outcome. They discuss with Carlos the likelihood that they'll need a few sessions to enter all the data into the table. Carlos responds that the more studies they do, the less time it will take. He also says that it takes less time to find the information when study reports are clearly written. He adds that usually the important information can be found in the abstract.

Rebecca and Chen ask if it would be all right to take out the "Conceptual Framework" column, since none of the studies they're reviewing have conceptual frameworks (which help guide researchers as to how a study should proceed). Carlos replies that it's helpful to know that a study has no framework underpinning the research and

suggests they leave the column in. He says they can further discuss this point later on in the process when they synthesize the studies' findings. As Rebecca and Chen review each study, they enter its citation in a separate reference list so that they won't have to create this list at the end of the process. The reference list will be shared with colleagues and placed at the end of any RRT policy that results from this endeavor.

Carlos spends much of his time answering Rebecca's and Chen's questions concerning how to phrase the information they're entering in the table. He suggests that they keep it simple and consistent. For example, if a study indicated that it was implementing an RRT and hoped to see a change in a certain outcome, the nurses could enter "change in [the outcome] after RRT" as the purpose of the study. For studies examining the effect of an RRT on an outcome, they could say as the purpose, "effect of RRT on [the outcome]." Using the same words to describe the same purpose, even though it may not have been stated exactly that way in the study, can help when they compare studies later on.

Rebecca and Chen find it frustrating that the study data are not always presented in the same way from study to study. They ask Carlos why the authors or journals wouldn't present similar information in a similar manner. Carlos explains that the purpose of publishing these studies may have been to disseminate the

findings, not to compare them with other like studies. Rebecca realizes that she enjoys this kind of conversation, in which she and Chen have a voice and can contribute to a deeper understanding of how research impacts practice.

As Rebecca and Chen continue to enter data into the table, they begin to see similarities and differences across studies. They mention this to Carlos, who tells them they've begun the process of synthesis! Both nurses are encouraged by the fact that they're learning this new skill.

The MERIT trial is next in the stack of studies and it's a good trial to use to illustrate this phase of the RCA process. Set in Australia, the MERIT trial[1] examined whether the introduction of an RRT (called a medical emergency team or MET in the study) would reduce the incidence of cardiac arrest, unplanned admissions to the ICU, and death in the hospitals studied. See Table 1 to follow along as the EBP team finds and enters the trial data into the table.

Design/Method. After Rebecca and Chen enter the citation information and note the lack of a conceptual framework, they're ready to fill in the "Design/Method" column. First they enter *RCT* for randomized controlled trial, which they find in both the study title and introduction. But MERIT is called a "cluster-randomised controlled trial," and *cluster* is a term they haven't seen before. Carlos explains that it means that hospitals, not individuals or patients, were randomly assigned to the RRT. He says that the likely reason the researchers chose to randomly assign hospitals is that if they had randomly assigned individual patients or units, others in the hospital might have heard about the RRT and potentially influenced the outcome.

To randomly assign hospitals (instead of units or patients) to the intervention and comparison groups is a cleaner research design.

Keep the data in the table consistent by using simple, inclusive terminology.

To keep the study purposes consistent among the studies in the RCA, the EBP team uses inclusive terminology they developed after they noticed that different trials had different ways of describing the same objectives. Now they write that the purpose of the MERIT trial is to see if an RRT can reduce *CR*, for cardiopulmonary arrest or code rates, *HMR*, for hospital-wide mortality rates, and *UICUA* for unplanned ICU admissions. They use those same terms consistently throughout the evaluation table.

Sample/Setting. A total of 23 hospitals in Australia with an average of 340 beds per hospital is the study sample. Twelve hospitals had an RRT (the intervention group) and 11 hospitals didn't (the control group).

Major Variables Studied. The independent variable is the variable that influences the outcome (in this trial, it's an RRT for six months). The dependent variable is the outcome (in this case, HMR, CR, and UICUA). In this trial, the outcomes didn't include do-not-resuscitate data. The RRT was made up of an attending physician and an ICU or ED nurse.

While the MERIT trial seems to perfectly answer Rebecca's PICOT question, it contains elements that aren't entirely relevant, such as the fact that the researchers collected information on how the RRTs were activated and provided their protocol for calling the RRTs. However, these elements might be helpful to the EBP team later on when they make decisions about implementing an RRT in their hospital. So that they can come back to this information, they place it in the last column, "Appraisal: Worth to Practice."

After reviewing the studies to make sure they've captured the essential elements in the evaluation table, Rebecca and Chen still feel unsure about whether the information is complete. Carlos reminds them that a system-wide practice change—such as the change Rebecca is exploring, that of implementing an RRT in her hospital—requires careful consideration of the evidence and this is only the first step. He cautions them not to worry too much about perfection and to put their efforts into understanding the information in the studies. He reminds them that as they move on to the next steps in the critical appraisal process, and learn even more about the studies and projects, they can refine any data in the table. Rebecca and Chen feel uncomfortable with this uncertainty but decide to trust the process. They continue extracting data and entering it into the table even though they may not completely understand what they're entering at present. They both realize that this will be a learning opportunity and, though the learning curve may be steep at times, they value the outcome of improving patient care enough to continue the work—as long as Carlos is there to help.

In applying these principles for evaluating research studies to your own search for the evidence to answer your PICOT question, remember that this series can't contain all the available information about research methodology. Fortunately, there are many good resources available in books and online. For example, to find out more about *sample size*, which can affect the likelihood that researchers' results occur by chance (a random finding) rather than that the intervention brought about the expected outcome, search the Web using terms that describe what you want to know. If you type *sample size findings by chance* in a search engine, you'll find several Web sites that can help you better understand this study essential.

Be sure to join the EBP team in the next installment of the series, "Critical Appraisal of the Evidence: Part II," when Rebecca and Chen will use the MERIT trial to illustrate the next steps in the RCA process, complete the rest of the evaluation table, and dig a little deeper into the studies in order to detect the "keepers." ▼

Ellen Fineout-Overholt is clinical professor and director of the Center for the Advancement of Evidence-Based Practice at Arizona State University in Phoenix, where Bernadette Mazurek Melnyk is dean and distinguished foundation professor of nursing, Susan B. Stillwell is clinical associate professor and program coordinator of the Nurse Educator Evidence-Based Practice Mentorship Program, and Kathleen M. Williamson is associate director of the Center for the Advancement of Evidence-Based Practice. Contact author: Ellen Fineout-Overholt, ellen.fineout-overholt@asu.edu.

REFERENCE

1. Hillman K, et al. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. *Lancet* 2005;365(9477): 2091-7.